University of Alberta

Characterization of Glycoproteins and Oligosaccharides Using Mass Spectrometry

by

Messele Fentabil

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Chemistry

Examining Committee

John S. Klassen, PhD

Professor, Department of Chemistry, University of Alberta

Charles A. Lucy, PhD

Professor, Department of Chemistry, University of Alberta

Christine M. Szymanski, PhD

Professor, Department of Biological Sciences, University of Alberta

Abstract

This thesis describes the application of mass spectrometry (MS) to glycoprotein and oligosaccharide analysis. Glycosylated proteins are involved in cell-cell and cell-matrix recognition. Applications of trypsin and proteinase K to hydrolyze glycoproteins into glycopeptides that are compatible with MS and MS/MS analysis are investigated. For successful site-specific analysis of glycans, glycopeptides with short peptide (3-8 residues) are needed. Although trypsin is an important enzyme for protein identification, proteinase K is superior for site-specific glycan analysis due to its potential to hydrolyze every glycoprotein to short glycopeptides.

The gas-phase dissociation pathways, kinetics and energetics of protonated oligosaccharides are described. The oligosaccharides dissociate via cleavage at the glycosidic linkages during thermal activation. Using double resonance experiments, it was established that oligosaccharides undergo sequential and parallel fragmentation reactions. Furthermore, dissociation of product ions to secondary ions was confirmed. Arrhenius activation parameters, E_a and A for protonated alpha- and beta-linked D-glucopyranose oligosaccharides are reported.

Acknowledgments

From the outset, I would like to thank my supervisor, Dr. John Klassen for giving me the opportunity to work in his lab, and his guidance. I have learnt a lot from you, John, and thank you. Dr. Charles Lucy and Dr. Jeff Stryker, I am happy for having both of you in my committee, and I appreciate all the help I received from you. Dr. Christine Szymanski, thank you very much for your smart scientific ideas and input to my thesis. My gratitude goes to my collaborators, Dr. Mario Feldman, Dr. Amirreza Faridmoayer and Isabell Hug. It was fun to work with you.

Jacqueline Jhingree and all the Klassen group, thank you very much for your support and the good times we had together.

I would like to thank my wife, Zinash Gela, and my son, Nahum Fentabil, for their understanding and untinting support. My mommy, Sado Tefera, I have no words to convey my heartfelt gratitude for your unreserved love and commitment. Thank you!

I owe lots of debt from lots of people and I just want to say thank you all. Especial thanks to Molla Tefera (my high school chemistry teacher), Gezahegni Getachew (my elder brother), Sister Letezigu Teklehaimanot and Mintiwab Hunegnaw. Finally, I would like thank my all my friends. Especially, Ayisheshim Kebie Ayisheshim (Kebe) and Amsalu Yigzaw Anagaw, thank you for being with me in the good as well as tough times.

Table of Contents

List of Figures

List of Abbreviations

A-factor	Pre-exponential factor
Asn	Asparagine
Asp	Aspartic acid
BIRD	Blackbody infrared radiative dissociation
CE	Capillary electrophoresis
CID	Collision induced dissociation
DTT	Dithiothreitol
E_a	Activation Energy
ECD	Electron capture dissociation
EDD	Electron detachment dissociation
E.coli	*Escherichia coli*
ES	Electrospray
ESI	Electrospray ionization
FT-ICR	Fourier transform ion cyclotron resonance
FA	Formic acid
FWHM	Full-width half maximum height
Fuc	L-fucose
Gal	D-galactose
GalNAc	N-acetylgalactosamine
GlcNAc	N-acetylglucosamine
Glu	Glutamic acid

HPLC	High performance liquid chromatography
i.d.	Inner diameter
IRMPD	Infrared multiphoton dissociation
LC	Liquid chromatography
MALDI	Matrix-assisted laser desorption ionization
MS	Mass spectrometry
m/z	Mass-to-charge ratio
nanoES	Nanoflow electrospray
NMR	Nuclear magnetic resonance
NM	*Neisseria meningitidis*
o.d.	Outer diameter
PA	*Pseudomonas aeruginosa*
PAGE	Polyacrylamide gel electrophoresis
PTM	Post translational modification
QToF	Quadrupole time of flight
RF	Radio frequency
RHF	Restricted Hartree-Fock
RP	Reversed Phase
SDS	Sodium dodecyl sulphate
Ser	Serine
S/N	Signal-to-noice ratio
TFA	Ttrifluoroacetic acid
Thr	Threonine

ToF   Time of flight

Chapter 1

Introduction to Mass Spectrometry and Glycoprotein Analysis

1.1 General Introduction

In the post genomic era, focus has shifted from genome sequencing to evaluate the role of the gene products, proteins, in organisms [1]. In order to understand the contributions that these proteins have in organisms, the further layer of biological complexity, namely post-translational modifications (PTM), requires study [2-4]. One of the most common post-translational modification of proteins is glycosylation, which is the addition of carbohydrates on to proteins. More than half of all proteins in eukaryotes are expected to be decorated with carbohydrates [5], which potentially influence the physical properties and bioactivities of these proteins. The biological roles of glycosylation vary from conformational stability and protection against degradation of a single protein to molecular and cellular recognition in development, growth and cellular communication of cells [6-12]. Additionally, the carbohydrate portion of glycoproteins is often recognized by pathogenic bacteria, viruses and toxins as binding sites. Thus understanding the molecular details of glycoproteins is essential for further understanding of biological systems and development of glycoprotein-based vaccines and therapeutics.

A variety of analytical techniques are used to understand the structures of glycoproteins [13]. These techniques include nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography, separation techniques, and mass

spectrometry. X-ray crystallography is the most prominent technique for elucidating the 3D structure of biomolecules. However, its application to glycoproteins is limited due to problems associated with crystallization and the flexible nature of glycans [14]. NMR spectroscopy is an important alternative to X-ray crystallography. It provides precise structural details about glycans as well as glycosylation sites, usually after digestion of the glycoproteins with proteases [15,16]. However, NMR spectroscopy requires relatively large sample sizes (milligram). Separation techniques such as capillary electrophoresis (CE) [17,18] and high-pH anion-exchange chromatography (HPAEC) pulsed amperometric detection (PAD) have been playing important roles for glycoprotein analysis [19,20]. For example, Guttman devised a method for sequencing of the glycan part of glycoproteins by gel CE and exoglycosidase digestions [18]. Nevertheless, mass spectrometry using electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI) is becoming the most widely applicable tool for glycoprotein analysis to which other techniques are ancillary [2]. However, glycoproteins using MS is challenging due to the complex nature of the glycan and sample heterogeneity.

This thesis focuses on analysis of glycoproteins and oligosaccharides using mass spectrometry (MS). All of the MS techniques presented in this thesis centers around the use of tandem mass spectrometry (MS/MS) for glycoprotein and oligosaccharide analysis. There are many excellent reviews in the topic of MS and MS/MS [21-27] and thus I do not intend to cover all areas in detail in this chapter. Rather I will focus on the most relevant topics to my thesis work.

Specifically, the technology of quadrupole time-of-flight (QToF) and Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers will be introduced. Glycoprotein analysis methods related to MS and MS/MS will be described. Finally, the scope of my thesis will be given.

1.2 Mass Spectrometry

In general, mass spectrometry involves the formation of gas-phase ions by various ionization techniques followed by their separation according to their mass-to-charge ratio (m/z) and detection using a mass analyzer. Many different ionization methods and mass analyzers exist and deciding which combination is best suited for any given analysis depends on the nature of the sample and the type of information required. In the analysis of glycoproteins, soft ionization techniques mainly ESI and MALDI have been playing important roles. There are a number of different types of mass analyzers such as magnetic sector [21,22], quadrupole [23], ion trap [24], time of flight (ToF) [25], FT-ICR [26], and recently, orbitrap [27]. In the present study, electrospray ionization together with QToF and FT-ICR MS are used for analysis of glycoproteins and oligosaccharides.

1.2.1 Electrospray Ionization

ESI is one of the soft ionization techniques, which enable the gentle conversion of samples from solution into gas-phase ions gently. The ESI process, as described by Kebarle and coworkers [28-30], involves the production of

charged droplets from electrolyte dissociation in a solvent; the shrinkage of the droplets by solvent evaporation accompanied with droplet fission for formation of very small, highly charged droplets from which gaseous ions are produced. Shown in **Fig**. 1.1 is a diagram describing the ESI process in the positive ion mode [29]. Applying a potential difference of 1-6 kV between the capillary and the counter electrode generates an electric field, which induces a charge accumulation at the liquid surface situated at the end of the capillary. This induces charge separation of electrolytes in the solution. Positive charges drift towards the liquid surface, leading to the formation of a liquid cone referred to as a Taylor cone. At a sufficiently high electric field, the liquid cone becomes unstable and emits a thin liquid filament which then breaks into small positively charged droplets. With solvent evaporation, these droplets start to shrink into smaller droplets. As the charge density on the droplet surface increases to the Rayleigh limit, which is the point at which the Coulombic repulsion of the surface charges is equal to the surface tension of the droplets, the droplets undergo fission. This fission leads to the production of small, highly charged droplets.

Two mechanisms have been proposed to account for the formation of gas-phase ions from the very small and highly charged droplets: the charge residue model (CRM) [31]; and the ion evaporation model (IEM) [32]. In the CRM, as proposed by Dole, the droplet continues to break until one solvent-free molecule that carries the remaining charge is left. The charge residue model better explains ionization of larger molecules such as proteins. On the other hand, the IEM proposes that ions are desorbed or ejected directly from the small droplets into the
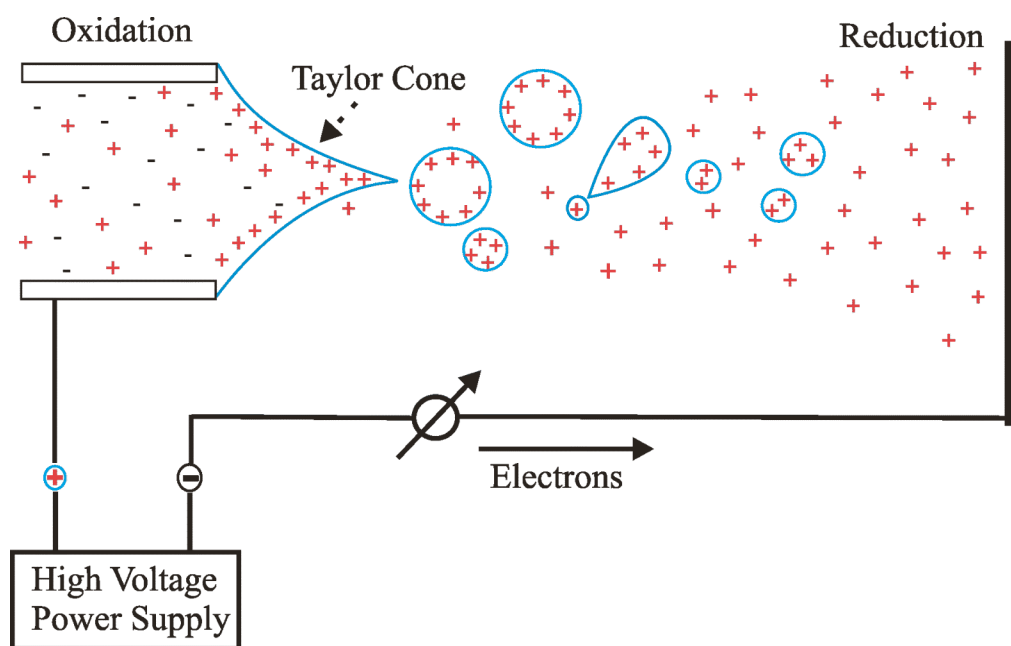
Figure 1.1 Illustration of the ES process. (Adapted from reference 29)

ambient gas in parallel with droplet fission [32]. According to the IEM, ions at the droplet surface will leave the droplets and become gas-phase ions. The gas-phase ion formation of an analyte will depend on the relative surface population of the analyte on the droplet surface. The rate constant for ion evaporation is expected to increase with the surface activity of the ion. In other words, surface active molecules ionize better than hydrophilic analytes during ESI. IEM better explains ionization of smaller analytes.

1.2.2 QToF MS

Fig. 1.2 shows the schematic diagram of the ESI QToF Premier^{TM} mass spectrometer from Waters, which is used for analysis of glycopeptides. It is a hybrid orthogonal acceleration time-of-flight mass spectrometer, which enables automated accurate mass measurement of precursor and fragment ions to yield high confidence in structure elucidation. It is hyphenated to the nano high pressure liquid chromatography (HPLC). The electrospray ionization of the analytes takes place as a result of imparting a high voltage to the eluent as it emerges from the emitter. The instrument combines the high transmission efficiency of the ZSpray^{TM} ion-source. This prevents the sample cone from getting dirty than the previous designs of QToF (e.g. QSTAR). The built-in NanoLockSpray^{TM} capability enables routine accurate mass measurement in both MS and MS/MS modes (< 30 ppm). The ToF analyzer uses a high voltage pulse to orthogonally accelerate the ions down the flight tube and a reflectron to reflect them back towards the multi channel plate (MCP) detector (V-optics). A mass

6
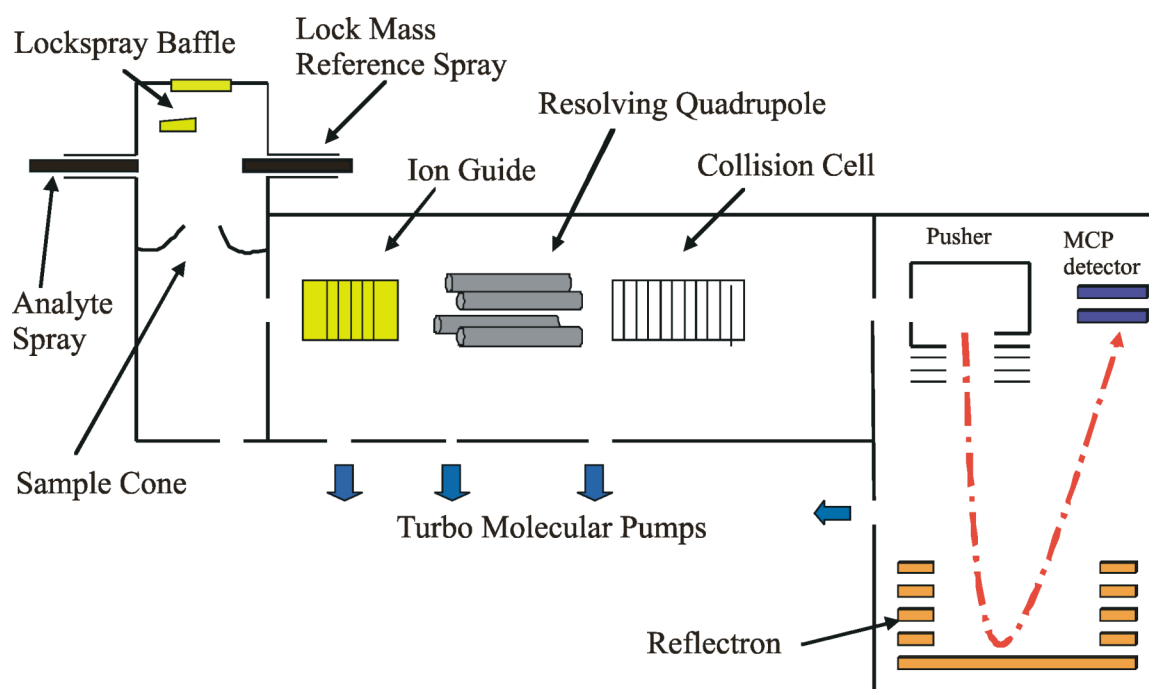
Figure 1.2 Schematic diagram of ESI QToF mass spectrometer from Waters

spectrum can be generated with an average resolution of 10, 000 (FWHM) for peptides of mass ~1000 Da.

The Premier QToF hybrid mass spectrometer provides both quadrupole (MS1) and ToF (MS2) mass analyzers with an intermediate collision cell for fragmentation if required. This combination allows ions to be selected, individually fragmented and then measured to a high degree of mass accuracy by the reflectron ToF. When the quadrupole resolving DC is off (RF only), ions can pass through the quadrupole and be accurately measured by the ToF in what is known as the ToF MS acquisition. When the quadrupole resolving DC is switched on, it can isolate a precursor ion of interest, which can be subjected to MS/MS dissociation using collision induced dissociation (CID). When the instrument is set to automatically switch between ToF MS and MS/MS modes, the ions are detected during the ToF MS scan and ions of interest will be subjected to MS/MS automatically, this is known as data directed analysis (DDA).

In the literature, there are several abbreviations used for describing the quadrupole time-of-flight mass spectrometer (e.g., QTOF, QqTOF, Qq-oaTOF, QToF, etc.). In this thesis, I use QToF.

1.2.3 Liquid Chromatography

After glycoprotein digestion, many peptides and glycopeptides are generated. Analyzing glycopeptides in a complex mixture directly by MS is not usually successful as glycopeptides are susceptible to ion suppression. This is mainly due to poor surface activity of glycopeptides and smaller abundance of

glycopeptides than their counter peptides of the glycoprotein due to heterogeneity of the glycans. Thus, efficient enrichment of glycopeptides is important before MS analysis. As a result, separation techniques such as liquid chromatography (LC) play important roles.

LC separation is done on the differences in how the peptides and glycopeptides interact with a stationary phase packed in a column. There are different stationary phases available and the surface chemistry of the stationary phase determines what separation mechanism is involved. The most common technique used in LC MS analysis of peptides and glycopeptides is reversed-phase (RP) LC. In reversed phase separation, the stationary phase surface contains a non-polar functional group such as C_{18}. Peptides and glycopeptides elution is done using solvent mixtures with gradual decrease in polarity (e.g. reducing the water content while increasing acetonitrile) in gradient elution chromatography. The presence of acetonitrile in RPLC elution makes it compatible with ES process because the surface tension of the charged droplets will be less than pure water. In addition, RPLC offers high separation efficiency compared to other LC techniques such as ion exchange. However, RPLC alone may not be sufficient for separating a complex mixture of peptides and glycopeptides. Thus partial fractionation of the complex mixture prior to RP-HPLC MS and MS/MS analysis is useful. The nano RP-HPLC variant of the LC (flow rate from 1-1000 nL/min) increases ionization efficiency and thus sensitivity. Because the evaporation of the solvent as well as the surface area of the droplets increases, these enhance electrospray ionization of the effluent [33]. In nanoflow, desolvation of the

droplet is relatively easy and the spray cone is not as large as in the microflow. Thus, the analyte sampling efficiency in nanospray is higher than in microspray. Also the sample consumption for nanoflow ES is significantly less than microflow ES. These are the main reasons nanospray combined with nanoLC, and commonly used for proteomic and recently for glycoproteomic analysis.

1.2.4 Collision Induced Dissociation (CID)

CID is one of the most popular methods for peptide and glycopeptide ion fragmentation in the gas-phase. In CID, the molecular or precursor ions are accelerated by an electrical field to high kinetic energy in the vaccum. Then they collide with neutral gas species (e.g. He, N_2, or Ar) in the collision cell of the mass spectrometer (see **Fig. 1.2**). During the collision, part of the kinetic energy is converted into internal energy of the precursor ion. This results in decomposition of the precursor ion into smaller fragment ions. These fragment ions can then be analyzed by a mass spectrometer. CID can be performed at either high or low collision energies. Low energy CID (10-100 eV) is widely used in most mass spectrometers (triple quadrupole, ion trap, QToF) for proteomic and glycoproteomic analysis. The QToF Premier used for glycopeptide analysis is equipped with CID technique using Ar at a pressure of $\sim 10^{-3}$ mbar as a collision gas and the acceleration voltage varies from 20-70 V according to m/z of the ions.
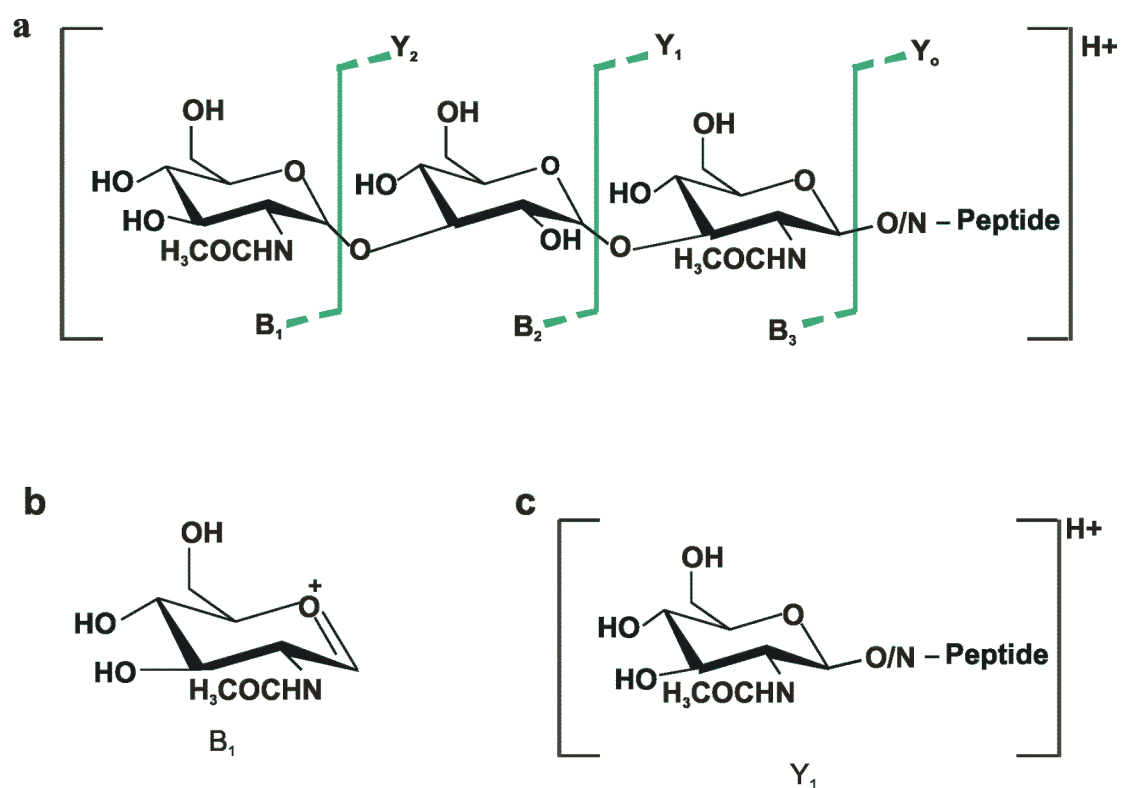
Figure 1.3 a) CID fragmentation of a protonated glycopeptides. b) an example of B ions and c) Y ions

The fragmentation pattern of glycopeptides using low energy CID is dominated by fragment ions resulted from cleavage of the glycosidic bonds (see **Fig**. **1.3**). Several bonds along the glycopeptide backbone can possibly be broken during CID. The most common ion types are the B and Y ions, which refer to the fragmentation at the glycosidic bond with charge retention on the nonreducing or reducing end, respectively (see **Fig. 1.3**). The nomenclature differentiates fragment ions according to which end of the fragment retains a charge after fragmentation and where the bond breakage occurs. If the charge associated with the glycopeptide ion is retained on the glycan terminal side of the glycosidic bond, the fragment ion is named a B ion (see **Fig**. **1.3b**). If the charge retains on the glycopeptide-terminal of the broken glycosidic bond, this ion is then named a Y ion (See **Fig**. **1. 3b**). Subscripts are used to designate the specific glycosidic bond fragmented to the product ions.

1.2.5 FT-ICR MS

FT-ICR MS was used for ion detection during the mechanistic study of glycosidic bond dissociation of oligosaccharides. It is a form of trapping MS, exhibiting the highest resolving power and mass accuracy (1 ppm) among available MS instruments. The principle of the FT-ICR mass analyzer is based on ion cyclotron motion, which arises from the interaction of a moving ion with homogeneous magnetic field. **Fig. 1.4** illustrates the cyclotron motion of a positive ion in a homogeneous magnetic field directed in to the plane of the page.
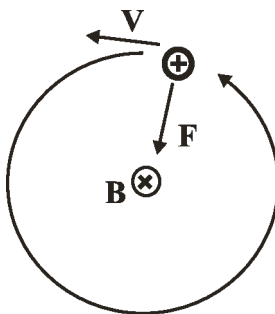
Figure 1.4 Cyclotron motion of a positive ion of charge q moving at velocity v in the presence of a constant magnetic field, B, which is pointing into the page. The ion moving to the left experience a downward force, F = q(v x B), resulting in a counterclockwise orbit.

During circular motion of ions, there should be a centripetal force given by equation (eq 1.1) .

$$F_c = mv^2/r \qquad \text{eq 1.1}$$

where F_c is the centripetal force on the ion, m is the mass, v is tangential velocity and r is the radius of curvature. Similarly, the magnetic force on an ion moving perpendicular to the homogeneous magnetic field will experience a magnetic force (Lorenz force) given by:

$$F_m = qvB \qquad \text{eq 1.2}$$

where F_m is the magnetic force, q is the charge on the ion, v is the velocity of the ion and B is the strength of the magnetic field. For a circular motion of ions in an ion cyclotron resonance, F_m and F_c should be equal. From the combination of those two equations, the cyclotron frequency of ions is described in equation

$$\omega_c = zeB/m \qquad \text{eq 1.3}$$

where z is the number of fundamental charges on the ion, e is the elementary charge, B is the magenetic field strength, and m is the mass of the ion. To obtain the cyclotron frequency (f) in Hertz (Hz) the results in radian per second has to be divided by 2π (i.e $w_c = 2\pi f$). A notable feature of equation 1.3 is that all ions of a given m/z rotate at the same frequency, independent of their velocities. The ultrahigh resolution achieved by FT-ICR MS is a direct result of the insensitivity of the cyclotron frequency to the kinetic energy distribution.

Fig. 1.5 illustrates how a mass spectrum is generated from this ion cyclotron motion. First, the addition of an RF pulse equal to the cyclotron (resonance) frequencies of the ions is transmitted by the excitation plates which

lie parallel to the magnetic field. The trajectory of a group of excited ions of the same m/z at their cyclotron frequency is given in **Fig**. **1.5a**. As these coherently orbiting ions pass another pair of opposing electrodes (detection plates), which are also parallel to the magnetic field, they induce an alternating current called an image current (**Fig**. **1.5b**). The amplitude of this transient signal is proportional to the number of ions in the cell and its frequency is the same as the cyclotron frequency of the ions. The transient signal is amplified, digitized, and stored for processing. A Fourier transform of the time domain signal is performed. Eq 1. 3 is used to generate a mass spectrum from the resultant frequency components of the signal.

A schematic diagram of the nanoESI FT-ICR MS instrument used in the present work is given in **Fig**. **1.6**. A solution containing the analyte of interest, typically mixed in a 2.5 mM ammonium acetate buffer, is placed in a borosilicate capillary (1.0 mm o.d., 0.68 mm i.d.), pulled to 4-7 μm o.d. at one end, forming the nanoES tip. A platinum wire is inserted into the distal end of the nanoES tip and a voltage (± 800-1000V) is applied. Droplets and gaseous ions produced by the nanoES are sampled into the mass spectrometer through a heated metal capillary. Gaseous ions are then transmitted through the skimmer and accumulated in the hexapole for a certain period of time to enhance the signal-to-noise (S/N) ratio. The accumulated ions are ejected from the hexapole and accelerated into a 9.4 T superconducting magnet. Here the ions are decelerated and eventually trapped by a combination of electric and magnetic fields in the FT-

a

b

$$\omega_c = zeB/m$$

m/z

time

ω_c
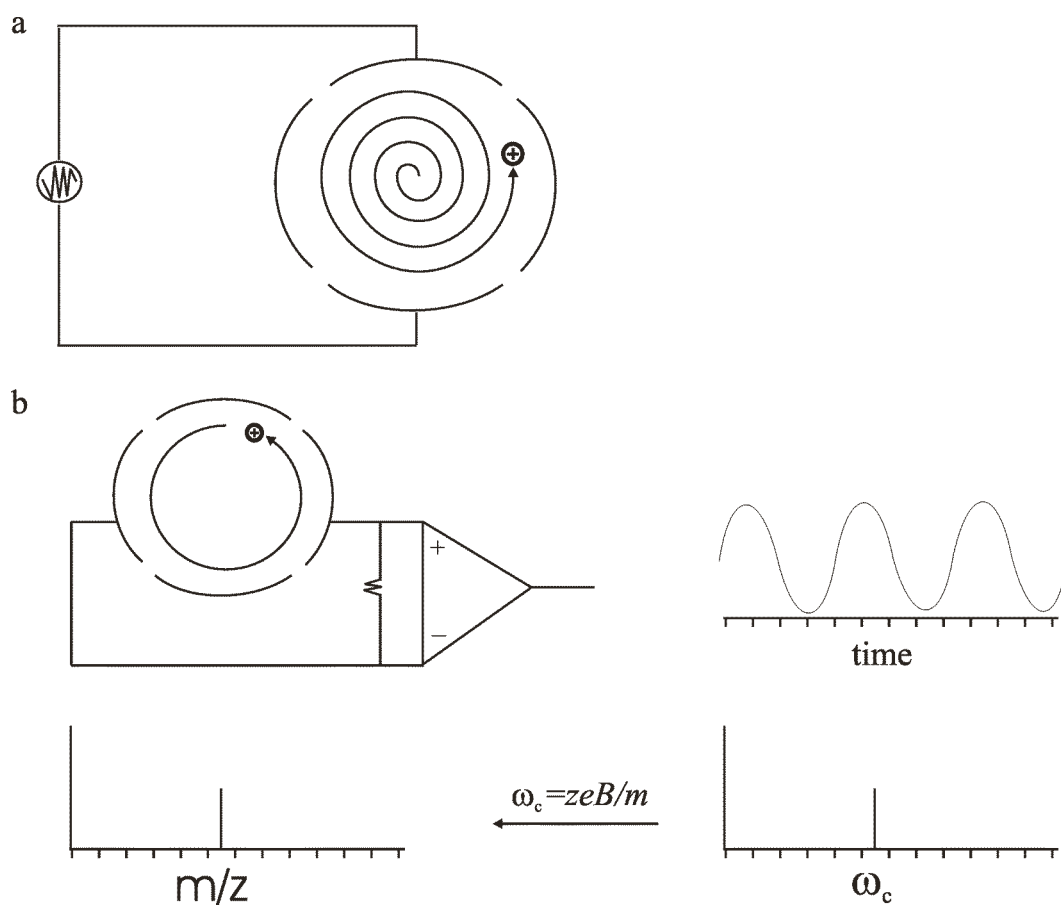
Figure 1.5 Illustration of (a) ion excitation and (b) the generation of a mass spectrum from the measured image current for ions in an FT-ICR/MS mass analyzer cell.
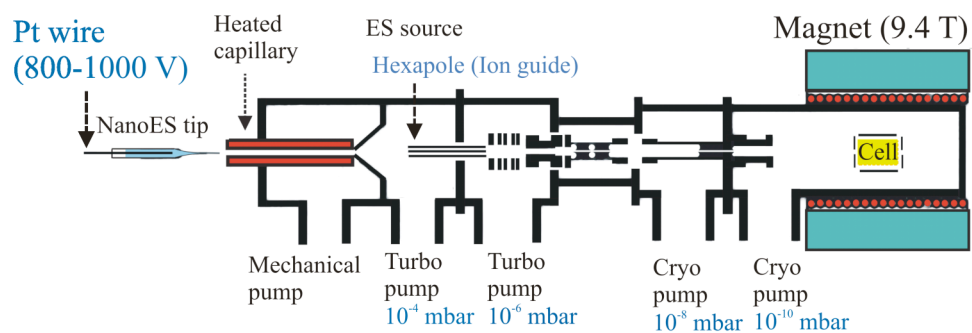
Figure 1.6 Schematic of the ESI FT-ICR MS instrument used in this study.

ICR cell for detection. The pressure for the instrument is typically mainitained at 5×10^{-10} mbar by the differential pumping system.

1.2.6 Blackbody Infrared Radiative Dissociation (BIRD)

Ions trapped within the ion cell of the FT-ICR MS can be fragmented prior to detection using a variety of collisional, radiative or electron-mediated activation techniques. In the present study, the blackbody infrared radiative dissociation (BIRD) [34] technique was used to perform time-resvolved thermal dissociation experiments. Briefly, in the BIRD technique, ions trapped within the ion cell, which is essentially a collison-free environment (pressure $< 10^{-10}$ mbar), are activated by the absorption of blackbody infrared photons emitted from the walls of the ion cell. The general chemical equation for unimolecular dissociation reactions initiated by the blackbody infrared photons is shown below for the hypothetical ion **A** [35-37].

$$\mathbf{A} + \mathbf{h}\nu \underset{k_{-1, rad}}{\overset{k_{1, rad}}{\rightleftharpoons}} \mathbf{A*} \overset{k_d}{\longrightarrow} \mathbf{P}$$
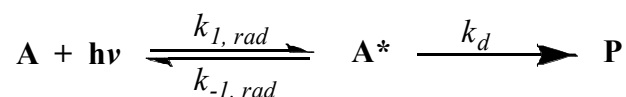
where $\mathbf{A}^*$ is the activated species and **P** is the reaction products, k_1 and k_{-1} are the energy dependent rate constants for absorption and emission of radiation, respectively, and k_d is the energy dependent rate constant for dissociation. When the dissociation rate is slow compared to the rates of energy transfer, the rapid energy exchange limit (REX) is achieved and the internal energy distribution of the reactant ion will resemble that of a Boltzmann distribution. Using steady-state

approximation, the unimolecular dissociation rate constant (k) of the process is given by:

$$k = k_d \left(k_{1,rad} / (k_{-1,rad} + k_d) \right) \qquad \text{eq. 1.4}$$

From the temperature dependence of the unimolecular dissociation constant (k), Arrhenius activation parameters were obtained from the slope of the plot of ln k against 1/T, eq 1.5.
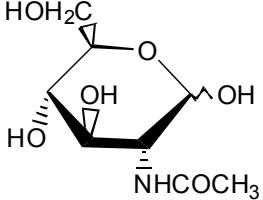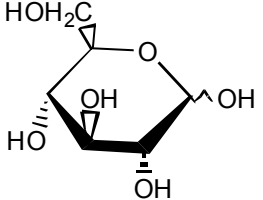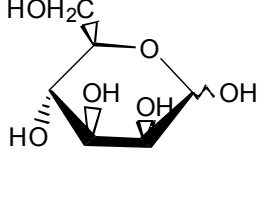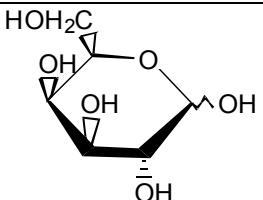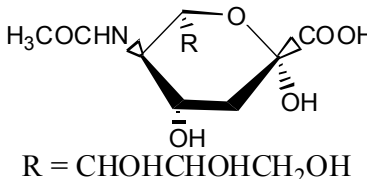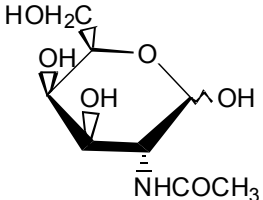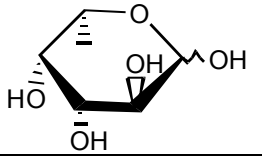
$$\ln k = -E_a/RT + \ln A \qquad \text{eq. 1.5}$$

1.3 Glycoprotein Anaysis Using Mass Spectrometry

Proteins decorated with one or more glycans during or after their synthesis in cells are glycoproteins. Glycans are carbohydrates covalently attached to glycoproteins by a process called glycosylation. In eukaryotes, glycosylated proteins are ubiquitous components of extracellular matrices and cellular surfaces. Eukaryotic glycans are composed of several monosaccharide units (Table 1.1). Glycosylation was previously considered to be restricted to eukaryotes. However, there have been increasing reports of glycoproteins in bacteria [10]. Although in most cases, the precise functions of the glycans attached to bacteria proteins have not been determined, glycoproteins may play roles in adhesion, stabilization of the proteins against proteolysis, and evasion of the host immune response.

Generally, there are two main types of glycosylation which are found in all domains of life: *N*-linked and *O*-linked. N-glycans are attached to asparagine residues in the consensus sequence Asn-Xaa-(Ser/Thr) in eukaryotes and Asp/Glu-Xaa-Asn-Xaa-(Ser/Thr) in bacteria [38], where Xaa is any of the

Table 1.1 Common monosaccharides of eukaryotic glycoproteins. Rersidual mass is dehydrated product of the monosaccharide.

Name	Structure	Mass	
		Molar (g/mol)	Residual
N-acetyl glucosamine (GlcNAc)		221.1	203.1
Glucose (Glc)		180.1	162.1
Mannose (Man)		180.1	162.1
Galactose (Gal)		180.1	162.1
Sialic acid	R = CHOHCHOHCH$_2$OH	309.3	291.2
N-acety galactosamine (GalNAc)		221.1	203.1
Fucose (Fuc)		164.1	146.1

standard amino acids except proline. For *O*-linked glycoproteins, the glycan is attached to serine or threonine residues. However, there is no consensus sequence defined although a high density of hydroxylated residues are common around the *O*-linked glycosylation sites.

Assembly of *N*-linked glycans on proteins of eukaryotes occurs in three major stages: i) formation of a lipid-linked precursor oligosaccharide ($Glc_3Man_9GlcNAc_2$) in the lumen of endoplasmic reticulum; ii) *en block* transfer of the oligosaccharide to nascent polypeptides during protein translation by oligosaccharyl transferases and iii) processing of the oligosaccharide (trimming), both removal of the original sugars and addition of new sugars at the non-reducing end. The third step results in complex oligosaccharide structures of the glycan although there is a common pentasaccharide core of *N*-glycans ($Man_3GlcNAc_2$). This prior knowledge of the biosynthetic pathway of *N*-glycans complements analysis of *N*-glycosylated proteins of eukaryotes. However, *O*-glycans of eukaryotes and the glycans of bacteria glycoproteins in general are much more diverse in structure than the *N*-glycans of eukaryotes, which makes their analysis more challenging.

There are two major strategies for MS-based characterization of glycoproteins. The first is analyzing the glycan after releasing them from glycoproteins using enzymatic and chemical techniques (Scheme 1.1). *N*-linked glycans of eukaryote glcoproteins can be released using endoglycosidases such as

```
                                              Glycoproteins
   Cleavage &
   Blocking of
   S-S Bridges                    │ Reduction and Carboxymethylation of
                                  │  Glycoproteins

                                              Glycoproteins

                                  │ Trypsin Digesion of Glycoproteins
   Cleavage into                  │
   Glycopeptides                  │ Sep pak C₁₈

                                      Peptides / Glycopeptides

                                  │ N-Glycosidase F digestion


                          -Separation of N-glycans Sep pak C₁₈ purification

                  N-glycans                              Peptides / O-glycopeptides
   Release of
   Glycans and
   Derivatisation     │ Permethylation                    │ Reductive elimination

              Permethylated N-Glycans                              O-Glycans

                                                           │ Permethylation

                                                     Permethylated O-Glycans


   Analysis
                             MALDI ToF and/or MALDI ToF/ToF
```

Scheme 1.1 *N*- and *O*-linked glycan analysis strategy suitable for glycoproteins of eukaryotes

peptide *N*-glycosidase F (PNGase F) or endoglycosidase H (Endo H) [39-43].

Chemical approaches such as beta-elimination have been successfully used to

released *O*-linked oligosaccharides [44,45]. After purification, the released

oligosaccharides are usually derivatized e.g. permethylated to improve the signal-

to-noise ratio (S/N) of the mass spectrum [45-47]. This is mainly due to

improvement of surface activity and proton affinity of the glycans during

permethylation. Sequential treatment with specific exoglycosidases of released

glycans, and monitoring the mass changes (by measuring molecular mass of the

glycans before and after treatment) gives information on the identity of the

monosaccharides as well as the linkage types [13]. Various tandem MS

techniques, such as collision induced dissociation (CID) [48], photo dissociation

(e.g. infrared multiphoton dissociation (IRMPD) [49,50]), electron capture

dissociation (ECD) [51], electron detachment dissociation (EDD) [52] and post-

source decay (PSD) [53] are regarded as useful tools for structural

characterization of the released glycans. Lots of effort has been exerted to obtain a

variety of structural data, including sequence and stereochemistry information

such as anomeric configuration [53,54] and linkage types [55-57], through the

MS/MS analysis. The success of MS/MS to provide structural information about

the precursor ion requires the ability to relate the fragments observed in a mass

spectrum back to the original structure of the ion.

The alternative MS approach is site-specific analysis of glycans on

glycoproteins (Scheme 1.2). In this strategy, glycoproteins will usually be

hydrolyzed using proteases. The resulting glycopeptides are subjected to MS

Cleavage &
Blocking of
S-S Bridges

Glycoproteins

Reduction and Carboxymethylation of
Glycoproteins

Glycoproteins

Cleavage into
Glycopeptides

Trypsin Digesion of Glycoproteins

Peptides / Glycopeptides

MS and LC-MS & MS/MS
(QToF, QSTAR, QTRAP, etc)
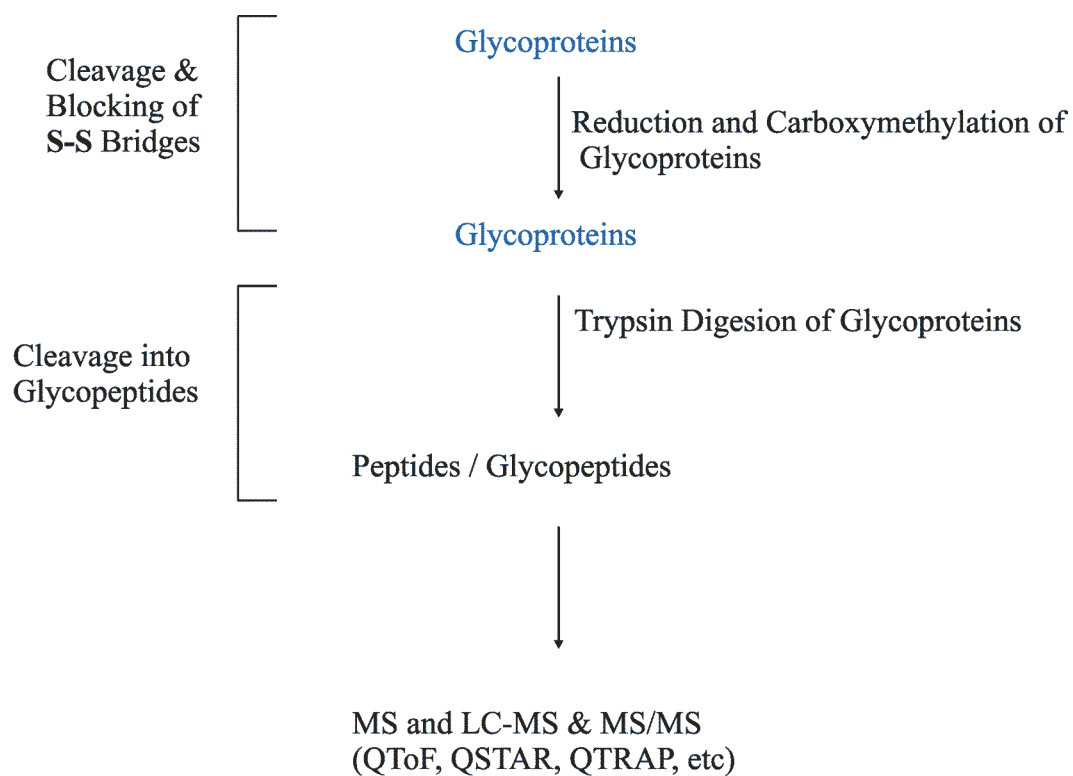
Scheme 1.2 *N-* and *O-*linked glycan analysis strategy suitable for glycoproteins.

and/or tandem MS analysis for defining the glycosylation site(s) as well as the structure of the glycan, simultaneously [58-63]. The proteases used for glycoprotein hydrolysis can be specific or nonspecific in terms of what they act upon. Specific enzymes, such as trypsin usually produce a peptide mixture with predictable cleavage specificity. This cleavage specificity can be exploited by database search algorithms to facilitate identification of the peptide sequence. As a result, digestion of glycoproteins with specific enzymes facilitates application of MS for glycopeptide analysis, because the peptides identity can be predicted theoretically. This strategy is successful especially with matrix assisted laser desorption ionization (MALDI) based MS techniques since less complicated spectra from singly charged ions are obtained [64,65]. However, tryptic peptides often suppress the mass spectrum signal of glycopepetides during MALDI MS analysis [65]. Similarly, analyzing glycopeptides in a complex mixture using ESI based MS techniques is not usually successful because glycopeptides are susceptible to ion suppression. In my thesis, applications of trypsin and proteinase K for site-specific analysis of glycans on glycoproteins are investigated to optimize *"in situ"* analysis of glycans.

1.4 Scope of Thesis

The work described in this thesis focuses on analysis of glycoproteins and oligosaccharides using mass spectrometry. In Chapter 2, MS and MS/MS analysis of glycoproteins after their hydrolysis with proteases is described. To know the glycosylation site as well as the glycan structure on glycoproteins, MS and

MS/MS analysis of glycopeptides is a common strategy. However, obtaining full information about the structure of glycans as well as the glycosylation site is challenging mainly due to the complex nature of the glycans. In the present work, application of trypsin and proteinase K to hydrolyze glycoproteins into glycopeptides that are suitable for MS and MS/MS analysis is investigated. It was found that for successful site-specific analysis of glycans on glycoproteins, glycopeptides with short peptide (3-8 residues) are needed. As a result, although trypsin is important enzyme for protein and glycoprotein identification work, proteinase K is superior for site-specific glycan analysis due to its potential to hydrolyze every glycoprotein to short glycopeptides. This has been successfully demonstrated using model glycosylated proteins as well as real samples.

Chapter 3 discussion focuses on the dissociation pathways, kinetics and energetics of protonated oligosaccharide dissociation in the gas-phase using the BIRD technique. Tandem MS approaches involve isolating an ion of interest in the gas-phase in time or space, dissociating it to fragment ions and accurately determining the mass of the product ions. Thus, the success of MS/MS to provide structural information about the precursor ion requires the ability to relate the fragments observed in a mass spectrum back to the original structure of the ion. Chapter 3 presents a detailed study of gas-phase dissociation of protonated oligosaccharides using BIRD FT-ICR MS. Dissociation of maltooligosaccharides, (α-D-Glc$_p$-1$\rightarrow$(4-α-D-Glc$_p$-1)$_n\rightarrow$4- D-Glc$_p$, n = 3,4,5,6) and cellohexaose, (β-D-Glc$_p$-1$\rightarrow$(4-β-D-Glc$_p$-1)$_n\rightarrow$4- D-Glc$_p$, n = 4) were studied using the BIRD technique. Over the range of temperatures investigated (70 ^{o}C to 160 ^{o}C), it was

26

found that the protonated oligosaccharides dissociate via cleavage at the glycosidic linkages forming B and Y ions. Using double resonance experiments, it was established that an oligosaccharide undergoes sequential and parallel fragmentation reactions. In other words, the fragmentation of protonated oligosaccharide ions in the gas-phase follows both sequential and parallel dissociation pathways. Furthermore, dissociation of product ions further to secondary ions was confirmed. Arrhenius activation parameters, E_a and A for protonated alpha-and beta-linked oligosaccharides are reported in the thesis. The thesis ends with conclusions (Chapter 4) where I also briefly comment on future work related to my research.

1.6 Literature Cited

1. Pandey, A.; Mann, M. *Nature* **2000**, *405*, 837-848.

2. Dell, A. Morris, H. R. *Science* **2001**, *291*, 2351-2356.

3. Bertozzi, C.R.; Kiessling, L. L. *Science* **2001**, *291*, 2357-2364.

4. Apweiler, R.; Hermjakob, H., Sharon, N. *Biochim. Biophys. Acta* **1999**, *1473*, 4-8.

5. Spiro, R. G. *Glycobiology*, **2002**, *12(4)* 43-56.

6. Rudd, p.M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A. *Science*, **2001**, *291*(*5512*), 2370—2376.

7. Dennis, J. W.; Granovsky, M.; Warren, C. E.; *Bioessays*, **1999**, *21(5),* 412-421.

8. Fukui, S.; Feizi, T.; Galustian, C.; Lawson, A. M.; Chai, W. G. *Nat. Biotechnol.* **2002**, *20,* 1011-1017.

9. Lowe, J. B.; Barth, J. D. Annu. *Rev. Biochem.* **2003**, *72,* 643-691.

10. Szymanski, C. M.; Wren, B. W. *Nat. Rev. Microbiol.* **2005**, *3*, 225-237.

11. Varki, A. *Glycobiology*, **1993**, *3*, 97-130.

12. Wurm, F. M.; *Nat. Biotech.,* **2004**, *22*, 1393- 1398.

13. Mechref, Y.; Novotny, M. V. *J. Chromatog. B.,* **2006**, *841,* 65-78.

14. Zhang, X.; Rosenthal, P. B.; Formanowski, F.; Fitz, W.; Wong, C.; Meier-Ewert, H.; Skehel, J. J.; Wiley, D. C. *Biol. Crystallogr.* **1999**, *55(5)*, 945-961.

15. Ortner, K.; Sivanandam, V. N.; Buchberger, W.; Muller N. *Anal. & Bioanal. Chem.* **2007**, *388(1),* 173-177.

16. Young, N. M.; Brisson, J. R.; Kelly, J.; Watson, D. C.; Tessier, L.; Lanthier, P. H.; Jarrell, H. C.; Cadotte, N.; St Michael, F.; Szymanski, C. M., *J. Biol. Chem.,* **2002**, *277*, 42530-42539.

17. Guttman, A., *Electrophoresis*, **1997**, *18*, 1136-1141.

18. Guttman, A. *Nature*, **1996**, *380*, 461-462.

19. Townsend, R. R.; Hardy, M. R. *Glycobiology Rev.,* **1991**, *1*, 139-147.

20. Hardy, M. R.; Townsend R. R.; *Methods Enzymol.* **1994**, *230*, 208-225.

21. Cottrell, J. S.; Greathead, R. J.; *Mass Spectrom. Rev.* **1986**, *5,* 215-247.

22. Burgoyne, T. W.; Hieftje, G. M. *Mass Spectrom. Rev.* **1998**, *15,* 241-259.

23. Dawson, P. H. *Quadrupole Mass Spectrometry and its Applications:* New York, 1976.

24. March, R. E. *J. Mass Spectrom.* **1997**, *32*, 351-369.

25. Wiley, W. C.; McLaren, I. H. *Rev. Scientific Instruments* **1955**, *26,* 1150-1157.

26. Amster, I. J. *J. Mass Spectrom.* **1996**, *31,* 1325-1337.

27. Hu, Q. Z.; Noll, R. J.; Li, H. Y.; Makarov, A.; Hardman, M.; Cooks, R. G. *J. Mass Spectrom.* **2005**, *40,* 430-443.

28. Kebarle, P.; Verkerk, U. H. *Mass Specrom. Rev.* **2009**, *28*, 898-917.

29. Kebarle, P.; Tang, L. *Anal. Chem.* **1993**, *65*, A972-A986.

30. Tang, L.; Kebarle, P. *Anal. Chem.* **1993**, *65*, 3654-3668.

31. Dole, M.; Mack, L. L.; Hines, R. L. *J. Chem. Phys.* **1968**, *49*, 2240-2249.

32. Irbarne, J. V.; Thomson, B. A. *J. Chem. Phys.* **1976**, *64*, 2287-2294.

33. Ivanov, A. R.; Zang, L.; Karger, B. L. *Anal. Chem.* **2003**, *75*, 5306-5316.

34. Perrin, J. *Ann. Phys.* **1919**, *11*, 5-108.

35. Dunbar, R. C. *Mass Spectrom. Rev.* **2004**, *23*, 127-158.

36. Price, W. D.; Schnier, P. D.; Jockusch, R. A.; Strittmatter, E. F.; Williams, E. R. *J. Am. Chem. Soc.* **1996**, *118*, 10640-10644.

37. Price, W. D.; Jockusch, R. A.; Williams, E. R. *J. Phys. Chem.* **1997**, *119*, 11988-11989.

38. Kowarik, M.; Young, N. M.; Numao, S.; Schulz, B. L.; Hug, I.; Callewaert, N.; Mills, D. C.; Watson, D. C.; Hernandez, M.; Kelly, J. F.; Wacker, M.; Aebi, M. *EMBO J.* **2006**, *25,* 1957-66.

39. Kita, Y.; Miura, Y.; Furukawa, J.; Nakano, M.; Ohno, M.; Takimoto, A.; Nishimura, S.; *Mol. & Cellular Proteomics*, **2007**, *6*, 1437-1445.

40. An, H. J.; Ninovuevo, M.; Aguilan, J.; Liu, H.; Lebrilla, C. B.; Alvarenga, L. S.; Mannis, M. J.; *J. Proteome Res.* **2005**, *4,* 1981-1987.

41. Goetz. J. A.; Novotny, M. V.; Mechref, Y. *Anal. Chem.*, **2009**, *81,* 9546-9552.

42. Haslam, S. M.; Coles, G. C.; Morris, H. R.; Dell, A. *Glycobiology*, **2000**, *10,* 223-229.

43. Kuster B.; Mann, M. *Anal. Chem.*, **1999**, *71,* 1430-1440.

44. Mirgorodskaya, E.; Hassan, H.; Clausen, H.; Roepstorff, P. *Anal. Chem.*, **2001**, *73*, 1263-1269.

45. Lei, M.; Mechref, Y.; Novotny, M. V. *J. Am. Soc. Mass Spectrom.,* **2009**, *20*, 1660-1671.

46. Ciucanu, I.; Kerek, F. *Carbohydr. Res.* **1984**, *131*, 209-213.

47. Yu, S. Y., Wu, S. W.; Khoo, K. K. *J. Glycoconj.,* **2006**, *23*, 355-369.

48. Harvey D. J. *Mass Spectrom. Rev.* **1999**, *18,* 349-450.

49. Budnik, B. A.; Haselmann, K. F.; Elkin, Y. N.; Gorbach, V. I.; Zubarev, R. A. *Anal. Chem.* **2003**, *75,* 5994-6001.

50. Adamson, J. T.; Hakansson, K. *Anal. Chem.* **2007**, *79(7),* 2901-2910.

51. Leavell, M. D.; Leary, J. A. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 528-536.

52. Smith, G.; Leary, J. A. *J. Am. Soc. Mass Spectrom.* **1996**, *7,* 953-957.

53. Gaucher, S. P.; Leary, J. A. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 269-272.

54. Laine, R. A.; Pamidimukkala, K. M.; French, A. D.; Hall, R. W.; Abbas, S. A.; Jain, R. K.; Matta, K. L. *J. Am. Chem. Soc.* **1988**, *110,* 6931-6931.

55. Lain, R. A.; Yoon, E.; Mahier, T. J.; Abbas, S.; Delappe, B.; Jain, R.; Matta, K. *Biol. Mass Spectrom.* **1991**, *20*, 505-514.

56. Yoon, E. S.; Laine, R. A. *Biol. Mass Spectrom.* **1992**, *21*, 479-485.

57. Suzuki, H.; Kameyama, A.; Tachibana, K.; Narimatsu, H.; Fukui K. *Anal. Chem.* **2009**, *81(3),* 1108-1120.

58. Brull, L. P.; Kovacik, V.; Thomas-Oates, J. E.; Heerma, W.; Haverkamp, J. *Rapid Commun. Mass Spectrom.* **1998**, *12,* 1520-1532.

59. Kubota, K.; Sato, Y.; Suzuki, Y.; Goto-Inoue, N.; Toda, T.; Suzuki, M.; Hisanaga, S.; Suzuki, A.; Endo, T. *Anal. Chem.,* **2008**, *80(10)*, 3693-3698.

60. Yu, S. Y., Wu, S. W.; Khoo, K. K. *J. Glycoconj.,* **2006**, *23*, 355-369.

61. Jiang, H. Desaire, H.; Butney, V. Y.; Bousfield, G. R., *J. Am. Soc. Mass Spectrom.,* **2004**, *15(5),* 750-758.

62. Zhang, Y.; Go, E. P.; Desaire, H. *Anal. Chem.,* **2008**, *80*, 3144-3158.

63. Wu, C. C.; Yates, J. R.; *Nature Biotechnology*, **2003**, *21*, 262-267.

64. Wuhrer, M.; Hokke, C. H.; Deelder, A. M.; *Rap. Commun. Mass Spectrom*, **2004**, *18*, 1714-1748.

65. An, H. J.; Peavy, T. R.; Hedrick, J. L.; Lebrilla, C. B., *Anal. Chem.*, **2003**, *75*, 5628-5637.

66. Andersen, M. T.; Mysling, S.; Hojrup, P. *Anal. Chem.,* **2009**, *81,* 3933-3943.

Chapter 2

Trypsin versus Proteinase K for Site-Specific Analysis of Glycans on Glycoproteins Using Mass Spectrometry*

2.1 Introduction

Proteins decorated [1] with one or more glycans during or after their synthesis in cells are glycoproteins. Glycosylation is one of the abundant co/post-translational protein modifications in cells. Its biological roles vary from conformational stability and protection against degradation of a single protein to molecular and cellular recognition [1-3]. In eukaryotes, glycosylated proteins are ubiquitous components of extracellular matrices and cellular surfaces. The glycan moieties of eukaryotic glycoproteins are implicated in a wide range of cell-cell and cell-matrix recognition events that are required for biological processes, such as immune recognition and cancer development. The carbohydrate portion of glycoproteins is often recognized by pathogenic bacteria, viruses and toxins as binding sites [4-6]. Many therapeutic proteins produced by the biotechnology industries are glycoproteins such as antibodies [7]. The need for faster and more sensitive methods for characterizing glycoproteins is particularly important to understand their molecular details.

Mass spectrometry is at the forefront of glycoprotein analysis to provide information about glycosylation site(s) as well as structure of the glycans [8,9]. Glycans are generally attached to serine or threonine (*O*-linked), or to asparagine

(*N*-linked) residues of glycoproteins [10,11]. There are two major strategies for MS based characterization of glycoproteins [12]. The first is analyzing the glycan after releasing them from glycoproteins using enzymatic and chemical techniques. *N*-linked glycans of eukaryote glcoproteins can be released using endoglycosidases such as peptide *N*-glycosidase F (PNGase F) or endoglycosidase H (Endo H) [13-16]. Chemical approaches such as beta-elimination have been successfully used to release *O*-linked oligosaccharides [17,18]. After purification, the released oligosaccharides are usually derivatized e.g. permethylated to improve the signal-to-noise ratio (S/N) of the mass spectrum [19,21]. The advantage of this strategy is that MS characterization of the glycan using different MS techniques will be relatively easier for heterogeneighty due to the polypeptide parts is removed. However, the challenges associated with releasing the glycans and the longer steps of sample preparation, and most importantly the loss of the glycosylation site information limit application of the approach.

The alternative MS approach is site-specific analysis of glycans on glycoproteins [12]. In this strategy, glycoproteins will usually be hydrolyzed using proteases. The resulting glycopeptides are subjected to MS and/or tandem MS analysis to define the glycosylation site(s) as well as the structure of the glycan, simultaneously [20-25]. The proteases used for glycoprotein hydrolysis can be specific or nonspecific in terms of what they act up on. Specific enzymes such as trypsin usually produce a peptide mixture with predictable cleavage specificity [23]. This cleavage specificity can be exploited by database search

algorithms to facilitate identification of the peptide sequence. As a result, digestion of glycoproteins with specific enzymes facilitates application of MS to glycopeptide analysis, because the peptides identity can be predicted theoretically. The mass of the glycan will be defined by cross checking the experimental glycopeptide mass and the predicted peptide mass. This strategy is called "glycomapping". Glycomapping is successful especially with matrix assisted laser desorption ionization (MALDI) based MS techniques since less complicated spectra from singly charged ions are obtained [26,27]. However, tryptic peptides often suppress the mass spectrum signal of glycopeptides during MALDI MS analysis [28]. This is mainly due to heterogeneity and poor ionization of glycopeptides [22].

Trimming down peptides using nonselective proteases had been demonstrated in previous studies [28-32] to selectively detect glycopeptides using MALDI MS techniques. The nonspecific protease, pronase E, had been applied for *N*-linked glycoproteins of eukaryotes [30] and bacteria [32]. Lebrilla and co-workers described that pronase E digestion of glycoproteins yielded primarily glycopeptides and amino acids [31]. Since large non-glycosylated peptides were no longer present, glycopeptides were readily identified using MALDI FT-ICR MS as they were essentially the only molecular species with signals above m/z 1000. Similarly, Larsen *et al* used trypsin and proteinase K in combination for *N*-linked eukaryotic glycoproteins [32]. Trypsin was used for *in-gel* digestion of the glycoproteins and part of the tryptic digest was subjected to proteinase K digestion. The remaining tryptic digest was directly taken to the MS and the mass

spectrum was used for database searching to identify the glycoproteins. In the study, glycopeptides from proteinase K hydrolyzed ovalbumin were analyzed by MALDI ToF MS. The glycopeptides were easily detected in the higher mass region (> 1000 Da) since the nonglycopeptides were hydrolyzed to short peptides.

There are also few studies on the application of nonspecific proteases for "*in situ*" analysis of glycans using ESI-based MS techniques [32-34]. Generally, ESI MS forms multiply charged glycopeptide ions which will be subjected to tandem MS analysis. During MS/MS of the glycopeptide ions, the mass spectra peaks of the fragment ions from dissociation of amide and glycosidic bonds will be used to identify the peptide primary structure as well as the glycan sequence, simultaneously. In 2004, Desaire *et al.* used proteinase K for digestion of ovine luteinizing hormone during the survey of ESI-MS for glycoprotein analysis [34]. The purified glycopeptides were analyzed using quadrupole QToF, triple quadrupole and ion trap mass spectrometers equipped with CID. The authors found a tripeptide carrying several heterogeneous glycans of the hormone. The glycosylation site of the hormone was easily determined from MS/MS data since there were only three amino acid residues in the detected glycopeptide ions. At the same time, more than 20 glycan structures were found on the tripeptide. The authors' conclusion was that QToF mass spectrometer showed the lowest sample consumption for site-specific analysis of glycoproteins using ESI MS.

Glycopeptide analysis using proteinase K seemed appealing for obtaining site-specific glycan analysis since the peptide part has few residues. However,

trypsin is widely applied for glycoprotein analysis mainly due to its matured digestion protocols [13,35]. Many glycoprotein studies involve trypsin. In the present study, proteinase K hydrolysis of glycoproteins is optimized and compared with application of trypsin for site-specific analysis of glycan on glycoproteins from bacteria and eukaryote using a nano RP-HPLC ESI QToF mass spectrometer. It was found that for successful site-specific analysis of glycans, glycopeptides with short (3-8 residues) peptide are needed. As a result, although trypsin is an important enzyme for glycoprotein identification, proteinase K is superior for site-specific glycan determination due to its potential to hydrolyze every glycoprotein to short glycopeptides. This has been successfully demonstrated using model glycosylated proteins from bacteria (AcrA) and eukaryote (chicken egg albumin). Finally, proteinase K was used to analyze *O*-glycosylated pilins from *Neisseria meningitidis* (NM) and *Pseudomonas aeruginosa* (PA) produced in *Escherichia coli (E. coli)* using glycol-engineering [36]. The glycosylation site of NM pilin was directly determined for the first time using direct analysis. Ser[63] was found to be the glycosylation site of the NM pilin. Similarly, PA pilin was found glycosylated on Ser[148].

2.2 Materials and Methods

2.2.1 Materials

Chicken egg albumin and proteinase K were purchased from Sigma-Aldrich Canada (Markaham, ON, Canada). The reagents water, acetonitrile and formic acid were LC-MS OPTIMA grade from Fisher Scientific Canada

(Edmonton, Canada). Sequencing grade modified trypsin was from Promega (Fisher Scientific, Canada) and C_{18} ZipTips were from Millipore (MilliPore, Bedford, Manchester). His-tagged *N*-glycosylated AcrA was produced using *E. coli* strain obtained from Dr. Mario Feldman's lab (Biological Sciences, University of Alberta) following standard protocols [36]. Pilins of *Neisseria meningitidis* and *Pseudomonas aeruginosa* were obtained from Dr. Feldman's lab.

2.2.2 SDS - PAGE

One-dimensional sodium dodecyl sulphate (SDS) polyacrylamide gel electrophoresis (PAGE) was performed using the Bio-Rad mini gel apparatus according to Laemmli [37] with slight modification. Briefly, the proteins were dissolved in SDS sample buffer (0.5 M Tris-HCl, pH 6.8, 10% glycerol, 2% SDS, 20 mM dithiothreitol (DTT), 0.05% bromophenol blue), boiled for 5 min, and applied to a 10% separation gel. Electrophoresis was carried out at a constant voltage of 125 V per gel piece. The separated proteins were visualized by staining with Coomassie Brilliant Blue [38]. The band corresponding to the protein of interest were excised and transferred to 1.5 ml eppendorf vials for digestion.

2.2.3 *In-gel* Proteinase K Digestion

The bands corresponding to the target glycoprotein were excised from gels and transferred to 1.5 ml eppendorf vials. The gel pieces were washed with 200 μl of 50 mM NH_4HCO_3 in 50% acetonitrile/water. After repeated washing for destaining, the samples were dehydrated with pure acetonitrile. Reduction of the

disulfide bonds and alkylation of the thiols were performed by hydration of the gel pieces with aqueous solution of 30-100 μl of 10 mM DTT and 30-100 μl of 55 mM iodoacetamide (IAA), respectively. After dehydration of the gels with 200 μl of pure acetonitrile, the gels were saturated with 10 μl of proteinase K (0.2 unit) and left at 4 $^{\circ}$C for 10 min for hydration. The samples were taken out from the fridge and 50 μl of 50 mM NH_4HCO_3 was added to the samples, and they were placed in 37 $^{\circ}$C for 6 to 12 hours for digestion. During optimization of the digestion protocol, from 2 to 0.02 units of proteinase K was applied and the digestion time was varied from 6 to 48 hrs. Glycopeptides were extracted with 50% acetonitrile/water containing 0.1% trifluoroacetic acid (TFA) from the gel pieces after digestion and dried using Speed-Vac (ThermoSavant, Holbrook, USA). The dried samples were re-solubilized with 0.1% TFA aqueous solution. Then they were purified using C_{18} ZipTips as discussed below before MS analysis.

2.2.4 *In-gel* Trypsin Digestion

In-gel digestion using trypsin was performed following the protocol of Shevchenko with slight modification [39]. Briefly, 2 μg sequence grade trypsin aliquoted in 10 μl aqueous solution containing 50 mM NH_4HCO_3 was added to gel pieces containing protein of interest in 1.5 ml eppendorf vials. Once the digestion was done, the samples were processed following the procedures described above for the proteinase K digest.

2.2.5 Desalting and Fractionation Using C_{18} ZipTips

C$_{18}$ ZipTips used for desalting and fractionation were first washed with 50% acentonitrile/water. Then the tips were equilibrated with aqueous solution of 0.1% TFA. The digests in aqueous solutions of 0.1% TFA were loaded on to the tips. Then the samples were repeatedly washed with 0.1% TFA in aqueous solution. Trapped glycopeptides and peptides were eluted with 10%, 20% and 50% acetonitrile/water solution containing 0.1% TFA in series and dried by SpeedVac to remove the acetonitril. The dried samples were re-solubilized with 0.1% formic acid (FA) in water and analyzed by nano reversed phase (RP) ultra high performance liquid chromatography (UPLC) (nanoACQUITY UPLC^{TM}, Waters, Milford, U.K.) ESI MS and MS/MS.

2.2.6 Nano RP-UPLC ESI QToF MS and MS/MS

Quadrupole orthogonal acceleration time of flight (QToF, Premier) mass spectrometer (Waters, Manchester, U.K.) which is in hyphenation with UPLC was used for analysis of the glycopeptides. The samples were prepared following the *in-gel* digestion protocol discussed above. The digests were solubilized with 0.1% formic acid (FA) in water for nano RP-UPLC ESI MS and MS/MS analysis. The column used was an Atlantis dC$_{18}$ microcolumn 15 cm length, 75 μm i.d and 3 μm particle size. Using columns with these parameters made the chromatography HPLC not UPLC, and in the thesis it is designated as HPLC.

For the nano RP-HPLC MS and MS/MS analysis of the digest, 2 μl of the sample solutions were injected on a precolumn on which the samples were trapped and on-column washed for 2 min. Then the samples were separated using

a binary solvent gradient. The solvents were 100% OPTIMA LC/MS grade water with 0.1% FA (Solvent A) and 100% acetonitrile with 0.1% FA (Solvent B). The flow rate was 350 nl/min. The solvent gradient for eluting glycopeptides from the column was 2% B for 4 min, 2-10% B from 4-30 min, 10-20% B from 30-45 min, 20-40% B from 45-65 min, 40-95% B from 65-67 min, 95% for 2 min, 95-2% B from 67-70 min and 2% B for 2 min. The analytes eluting out from the column were online electrosprayed to the orthogonal acceleration time of flight mass spectrometer. MS and MS/MS of ions were acquired with the data directed acquisition technique. During MS/MS, the collision energies (20 eV to 70 eV) were varied according to mass and charge state of the ions by using different collision energy functions. The instrument was calibrated with leucine Enkephalin and (Glu1)-Fibrinopeptide B, a mixed mass standard, and the mass to charge ratio was corrected by the LockSpray every minute throughout the data acquisition. The capillary and cone voltages were 3.8 kV and 40 V, respectively. Data acquisition and analysis were done using the MassLynx software (MassLynx V4.1, Waters). To identify elution of glycopeptides, putative glycans resulted from in-source fragmentation of glycopeptides such as peaks at m/z 204 (N-acetylhexosamine, HexNAc) and 366 (N-acetylhexosamine-hexose, HexNAc-Hex) corresponding to oxonium ions were searched using the algorithm in MassLynx.

2.3 Results and Discussion

2.3.1 Characterization of Model Glycoproteins

Model glycoproteins from bacteria and eukaryote were hydrolyzed using proteinase K and trypsin for MS and MS/MS analysis. Well studied *N*-linked glycoproteins: AcrA and ovalbumin were used to evaluate application of the enzymes for *N*-linked glycoproteins. The glycopeptides were desalted, fractionated and subjected to MS and tandem MS to obtain site-specific glycan information.

2.3.1.1 Trypsin Digested AcrA

AcrA was originally found to be *N*-glycosylated in the Gram negative bacteria, *Campylobacter jejuni* [40]. It is one of the few well studied *N*-glycoproteins of bacteria [40-42]. It has two *N*-glycosylation sites: [123]Asn and [273]Asn. Using glycoengineering, this glycoprotein was produced in *E.coli* [40,41]. Two oligosaccharides were found on the protein (AcrA) when it was expressed in *E. Coli*: GalNAcα1-4GalNAcα1-4(Glcβ1-3)GalNAcα1-4GalNAcα1-4GalNAcα1-3-Bac (Glc(GalNAc)$_5$Bac) and GalNAcα1-4GalNAcα1-4(Glcβ1-3)GalNAcα1-4GalNAcα1-4GalNAcα1-3-GlcNAc (Glc(GalNAc)$_5$GlcNAc) where GalNAc is N-acetylgalactosamine, Glc is glucose, GlcNAc is N-acetyglucosamine and Bac is 2,4-diacetamido-2,4,6-trideoxyglucopyranose (bacillosamine). The specific enzyme, trypsin was used for digestion of glycosylated AcrA. Trypsin cleaves peptide bonds at the carboxylic side of arginine and lysine. The theoretical trypsin digestion of AcrA (with www.expasy.org) shows that [121]DFNR[124] (550 Da) and
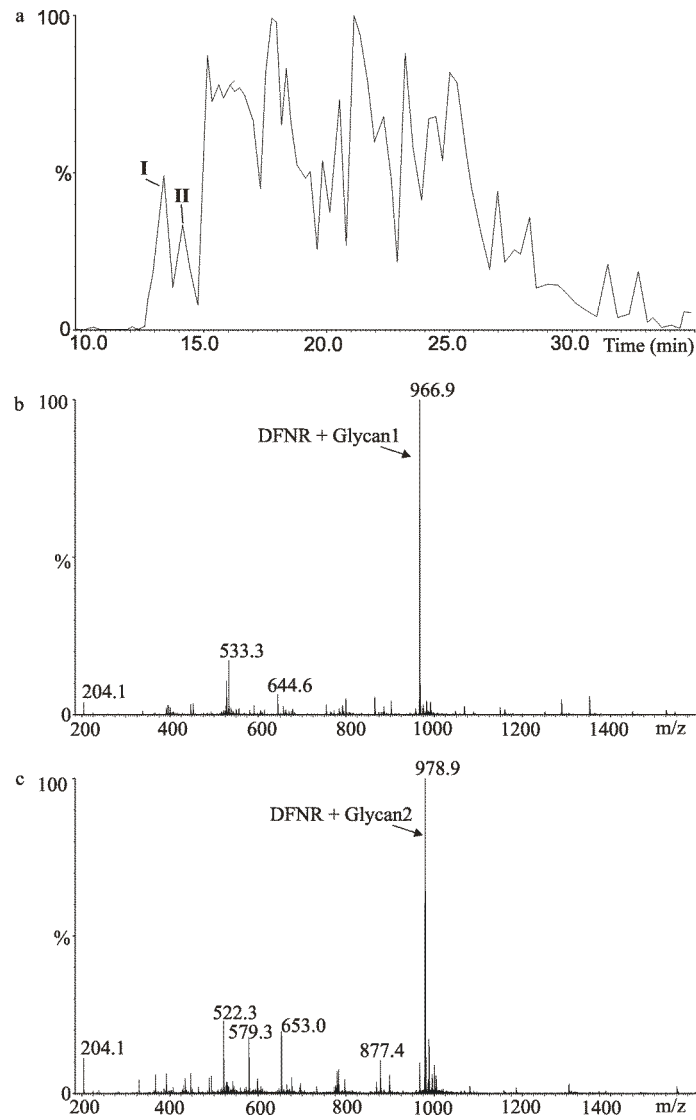
Figure 2.1 Base peak ion chromatogram and MS of tryptic digest of AcrA. a) Base peak ion chromatogram of the digest. Peaks **I** and **II** correspond to the glycopeptides. MS extracted from the chromatographic peaks b) **I** and c) **II**

modified with the two oligosaccharides. With this prior knowledge and prediction, the glycoprotein was subjected to hydrolysis and MS analysis. [267]AVFDNNNSTLLPGAFATITSEGFIQK[292] (2755 Da) are the two expected tryptic peptides of the glycan carriers [41,42]. Each of the two peptides would be modified with the two oligosaccharides. With those prior knowledge and prediction, the glycoprotein was subjected to hydrlysis and MS analysis.

After the digest had been desalted and fractionated as it is described in the experimental section, it was analyzed by RP-HPLC ESI QToF MS and MS/MS. The base peak ion chromatogram and MS of the tryptic digest of AcrA are shown in **Fig**. **2.1**. In the base peak ion chromatogram (**Fig**. **2.1a**), there are two peaks labeled as **I** and **II**. The MS data extracted from the peaks are shown in **Fig**. **2.1b** and **2.1c**. The base peaks in each mass spectrum of **Fig**. **2.1b** and **2.1c** are doubly charged, $[M + 2H]^{2+}$ glycopeptide ions at m/z 966.9 and 979.3, respectively. The mass of the ions matches with glycopeptides of the tryptic peptide, DFNR modified with the expected two oligosaccharides: $Bac(GalNAc)_5Glc$ (m/z 966.9) and $GlcNAc(GalNAc)_5Glc$ (m/z 979.3).

To confirm the identity of the glycopeptides, the ions were subjected to MS/MS analysis. MS/MS data of the ion at m/z 966.9 (**Fig**. **2.2a**) show oxonium ions of putative glycan peaks at m/z 204 (HexNAc), 366 (HexHexNAc) and 407 ($HexNAc_2$), which implies hexoses and N-acetylhexosamines are present in the glycopeptide. From the same data, a series of singly charged peaks with mass

43
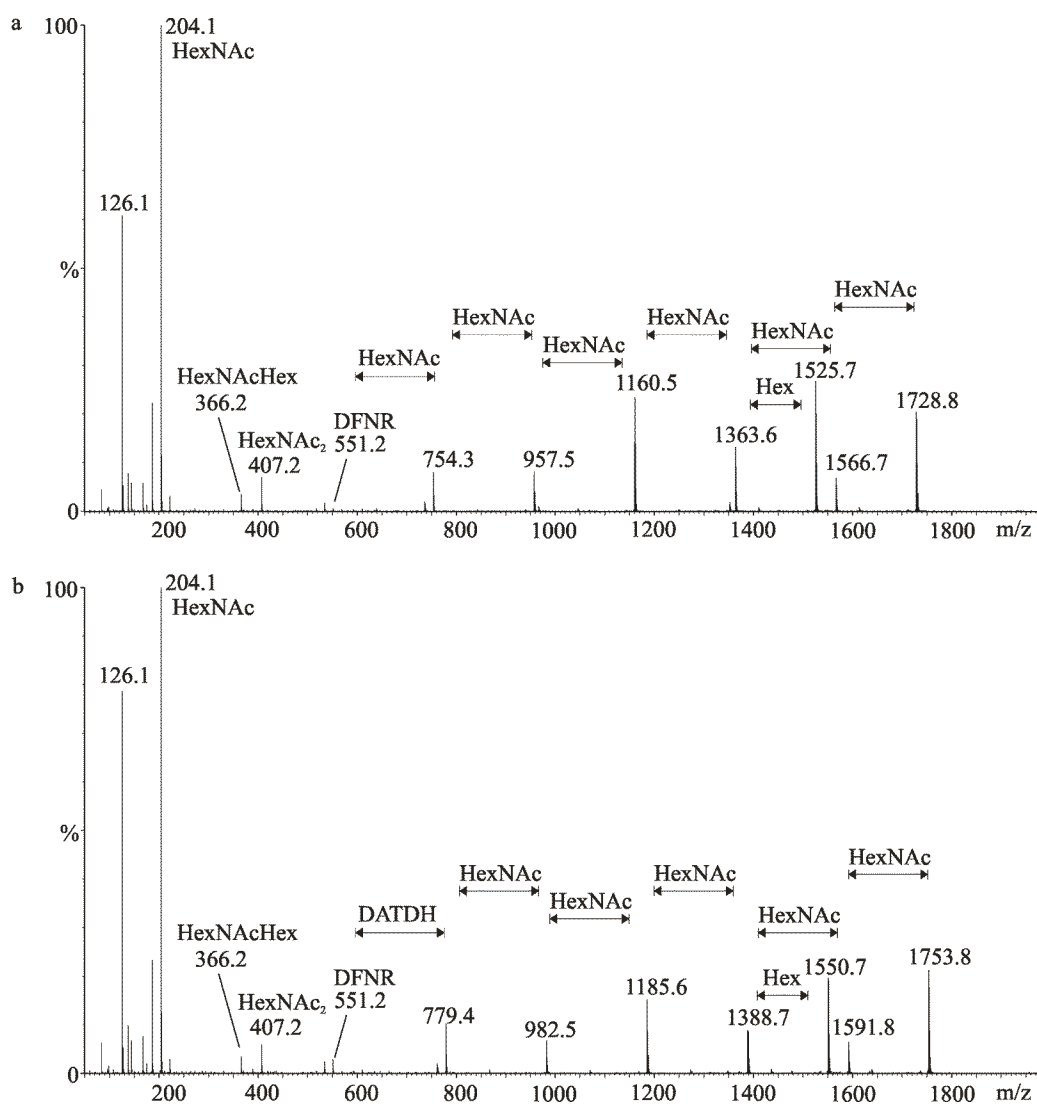
Figure 2.2 MS/MS of the doubly charged ions at m/z 966.7 and 979.3 obtained from trypsin digested AcrA a) Fig. 3a is MS/MS of the ion at m/z 966.7. b) The MS/MS of the glycopeptide ion at m/z 979.3

differences of 203 (HexNAc) and 162 (Hex), which matches with the expected

oligosaccharide structure of

GalNAcα1-4-GalNAcα1-4(Glcβ1-3)GalNAcα1-4GalcNAcα1-3GlcNAc-attached

to a peptide at m/z 551.2 (see **Fig**. **2.2a**) were detected. The branching of the

glycan corresponding to the mass difference between 1363.6 and 1525.7 is a

hexose, and 1363.6 and 1566.7 is an *N*-acetylhexosamine. On the same peptide,

another oligosaccharide was found (doubly charged ion at m/z 979.3). The

MS/MS spectrum of this ion is shown in **Fig**. **2.2b**. The data confirms that

DATDH(HexNAc)$_5$Hex was attached to the peptide, DFNR. The sugar at the

reducing end of the oligosaccharide has a mass of 228, which matches with the

mass of the expected glycan, Bac. This MS/MS data (**Fig**. **2.2a** and **2.2b**)

confirmed the presence of the expected two oligosaccharides attached to the

peptide, DFNR. In the entire chromatogram, these were the only glycopeptides

detected, although there were other glycopeptides expected from the other

glycosylation site, Asn[273].

The second glycosylation site is within the tryptic peptide,
[267]AVFDNNNSTLLPGAFATITSEGFIQK[292]. Together with the expected

oligosaccharide on it, the mass of tryptic glycopeptide will be more than 4,000 Da.

Similarly, tryptic glycopeptides of ovalbumin, whose tryptic glycopeptide mass is

> 4, 000 Da, were not detected as it is described below. Generally, for big

glycopeptides (m/z > 4 kDa), the nano RPLC ESI QToF MS failed to detect the

glycopeptides. This can lead to an erroneous conclusion. It is believed that the

reason for the negative results is that the glycopeptides elute from HPLC together

with surface active peptides that suppress their ionization. To circumvent the challenge, application of a nonspecific enzyme, proteinase K was investigated.

2.3.1.2 Proteinase K Digested AcrA

Proteinase K digestion of proteins results in short peptides due to its activity on several residues on proteins [43,44]. This is an advantage for trimming the peptides and glycopeptides to obtain glycopeptides with appropriate mass for RP-HPLC based MS analysis. However, the nonspecific nature of the enzyme and its extensive autolysis increases heterogeneity of the digest. Moreover, heterogeneity of glycopeptides due to non-specific hydrolysis around the glycosylation site will decrease the abundance of each entity of the glycopeptides. In the study, application of the enzyme for the study of glycoproteins was investigated and compared with application of trypsin.

Glycosylated AcrA was *in-gel* digested using proteinase K as described in the Experimental Section. After the digest had been desalted and fractionated, it was analyzed by RP-HPLC ESI QToF MS. The base peak ion chromatogram and MS data of the proteinase K digested AcrA are shown in **Fig**. **2.3**. The four peaks labeled in the ion chromatogram (**Fig**. **2.3a**), correspond to glycopeptides. The MS data extracted from each of the peaks are shown in **Fig 2.3b**, **c**, **d** and **e**). The base peaks in each mass spectrum (**Fig**. **2.3b**, **c**, **d** and **e**) correspond to doubly charged glycopeptide ions at m/z 1046.5, 1010.0, 1059.0 and 1023.0, respectively. The ions were subjected to low energy CID to confirm their identity. The MS/MS data (**Fig**. **2.4**) of the ions show that the four glycopeptides correspond to two

46
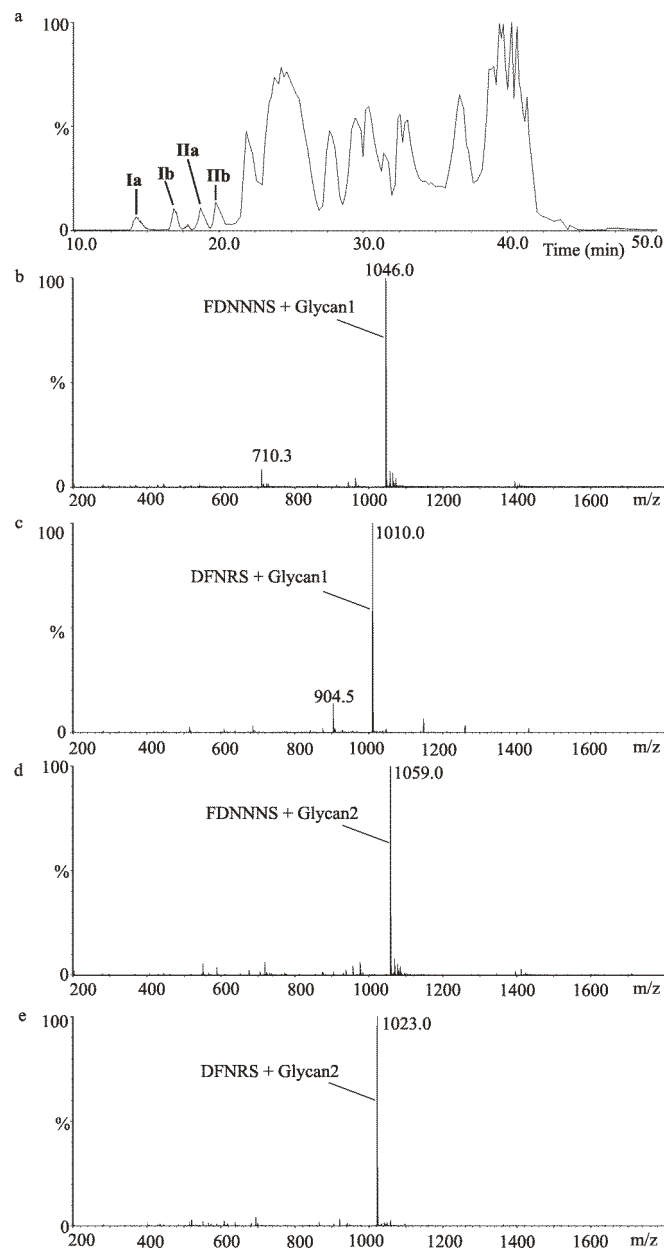
Figure 2.3 Base peak ion chromatogram and MS of proteinase K digested AcrA

a) Base peak ion chromatogram of the digest. The peaks labelled with **Ia**, **Ib**, **IIa** and **IIb** correspond to the four glycopeptides. MS extracted from the peaks b) **Ia** c) **Ib** d) **IIa** e) **IIb** are shown.

47

a

100

%

0

204.1
HexNAc

126.1

HexNAcHex

HexNAc₂

366.2

407.2

FDNNS
710.3

605.3

HexNAc

913.4

HexNAc

HexNAc

1116.4

HexNAc

1319.6

HexNAc

Hex

1522.7

1684.8

m/z

b

100

%

0

204.1
HexNAc

126.1

HexNAcHex

HexNAc₂

366.2

407.2

DFNRS
638.3

HexNAc

HexNAc

841.5

HexNAc

1044.6

HexNAc

1247.6

1450.8

HexNAc

Hex

HexNAc

1612.8

1653.8

1815.9

m/z

c

100

%

0

204.1
HexNAc

126.1

HexNAcHex

HexNAc₂

366.2

407.2

605.3

DATDH

FDNNS
710.3

HexNAc

938.5

HexNAc

1141.5

HexNAc

1344.6

Hex

1547.8

1709.9

m/z

d

100

%

0

204.1
HexNAc

126.1

HexNAcHex

HexNAc₂

366.2

407.2

DATDH

DFNRS
638.3

HexNAc

866.5

1069.6

HexNAc

HexNAc

1272.7

HexNAc

1475.8

Hex

HexNAc

1637.8

1678.9

1840.9

m/z

48

48

Figure 2.4 Tandem MS data of glycopeptides from proteinase K digested AcrA. MS/MS of the doubly charged ions at m/z a) 1046.0, [FDNNNS + HexNAc$_6$Hex] b) 1010.0, [DFNRS + HexNAc$_6$Hex] c) 1059.0, [FDNNNS + DATDH(HexNAc$_5$)Hex] d) 1023.0, [DFNRS + DATDH(HexNAc$_5$)Hex].

peptides modified with two oligosaccharides. For example, MS/MS of the ion at m/z 1010.0 (**Fig. 2.4b**) shows the oxonium ions of putative glycan peaks at m/z 204 (HexNAc), 366 (HexHexNAc) and 407 (HexNAc$_2$). This implies that hexoses and *N*-acetylhexosamines are parts of the glycopeptide. There are a series of singly charged peaks with mass differences of 203 (HexNAc) and 162 (Hex), which show the sequence of the oligosaccharide attached to a peptide at m/z 638.3. The glycan sequence matches with the expected structure of one of the glycans, GalNAcα1-4GalNAcα1-4(Glcβ1-3)-GalNAcα1-4GalNAcα1-4GalcNAcα1-3GlcNAc-attached to a peptide (m/z 638.3). Identifying the peptide (m/z 638.3) was challenging since theoretical prediction of the peptides is not possible during nonspecific protease hydrolysis. However, the MS/MS data together with MassLynx software (to search for peptides from the primary structure of the protein matching the acquired mass) confirmed the identity of the peptide as DFNRS, which included one of the *N*-glycosylation sites of AcrA, Asn[123]. On this peptide, the other oligosaccharide, Bac(GalNAc)$_5$Glc, was also found (m/z 1023.0) (**Fig**. **2.4e**). The MS/MS data of the ion shows that the glycan part matched with the expected glycopeptides. Thus the two expected oligosaccharides were detected on the peptide, DFNRS.

Similarly, MS/MS of the doubly charged ion at m/z 1046.5 shows that (HexNAc)$_6$Hex was attached to a peptide at m/z 710.3 (see **Fig 2.4a**). The identity of the peptide was confirmed to be FDNNNS, which included the second glycosylation site, Asn[273]. This peptide also carried the second oligosaccharide, Bac(GalNAc)$_5$Glc (m/z 1059.0, **Fig**. **2.4c**). It is important to remember that from

trypsin hydrolysis only one of the sites (the one with short peptide, DFNR) was detected. However, during proteinase K hydrolysis of AcrA, both glycosylation sites with their glycans were detected. To obtain optimum digestion glycoproteins with proteinase K, 0.2 unit of the ezyme for 6 to 12hrs is found the best contion.

The MS and MS/MS data of proteinase K hydrolyzed AcrA demonstrated that site-specific insight into heterogeneity of carbohydrates on glycoproteins can be obtained using RP-HPLC ESI QToF MS. Since the peptides are short, low energy CID is powerful enough to fragment the glycopeptides during tandem MS analysis. Similarly, the identity of the peptide can be determined easily. Discovery of glycoproteins in bacteria is becoming an area of interest. Thus proteinase K hydrolysis of glycoproteins for MS-based analysis will be an important strategy for future discovery of glycoproteins in the bacteria world.

2.3.1.3 Proteinase K Digested Ovalbumin

To demonstrate applicability of proteinase K for eukaryotic glycoproteins, chicken egg albumin was chosen as a model. Ovalbumin is one of the best characterized eukaryotic *N*-linked glycoproteins and was the first for which covalent attachment of carbohydrates to proteins was confirmed [45,46]. The protein was reported to have a single *N*-glycosylation site, Asn[292] [30]. On this site, various glycans have been found which include high mannose and hybrid *N*-glycans [46]. High mannose glycans are mannose rich *N*-glycans (5 to 9 mannoses on the inner core pentasaccharide, $Man_3GlcNAc_2$) matured from the pre-assembled precursor ($Glc_3Man_9GlcNAc_2$).

Ovalbumin was *in-gel* digested using proteinase K. The digest was analyzed using nano RP-HPLC ESI QToF mass spectrometer after purification. Representative MS/MS data of glycopeptides resulting from proteinase K digested ovalbumin are shown in **Fig 2.5**. For example, MS/MS data of a doubly charged ion at m/z 1073.3 shows glycan fragments (peaks at m/z 204-HexNAc, 366-HexHexNAc and 528-Hex_2HexNAc). A peptide ion at m/z 767.4 and several peaks with mass differences of 203 Da and 162 Da were also detected. It is important to mention that all the peaks detected in the MS/MS data are singly charged. The peptide identity on which the glycans attached was confirmed to be EKYNLT using MassLynx software together with the MS/MS data. Thus by cross checking the mass of the intact glycopeptide with its MS/MS data, together with MassLynx software, the oligosaccharide was determined to be $HexNAc_2Hex_6$ attached to the hexapeptide, EKYNLT. This oligosaccharide matched with the previously reported abundant high mannose glycan of ovalbumin: $GlcNAc_2Man_6$ [49]. Similarly, the MS/MS data from the doubly charged ions at m/z 1094.4 (**Fig. 2.5b**) and 1196.0 (**Fig 2.5c**) shows glycans of $HexNAc_3Hex_5$ and $HexNac_4Hex_5$, respectively, assembled on the same peptide, EKYNLT. The glycans match with the previously reported oligosaccharides of the glycoprotein: $GlcNAc_3Man_5$ and $GlcNAc_4Man_5$ [49], respectively.

In addition to the above three oligosaccharides, several glycans were detected on the same glycosylation site (**Fig. 2.6**). By using MS, MS/MS data, MassLynx and GlycoMode softwares together with previous studies [49], the
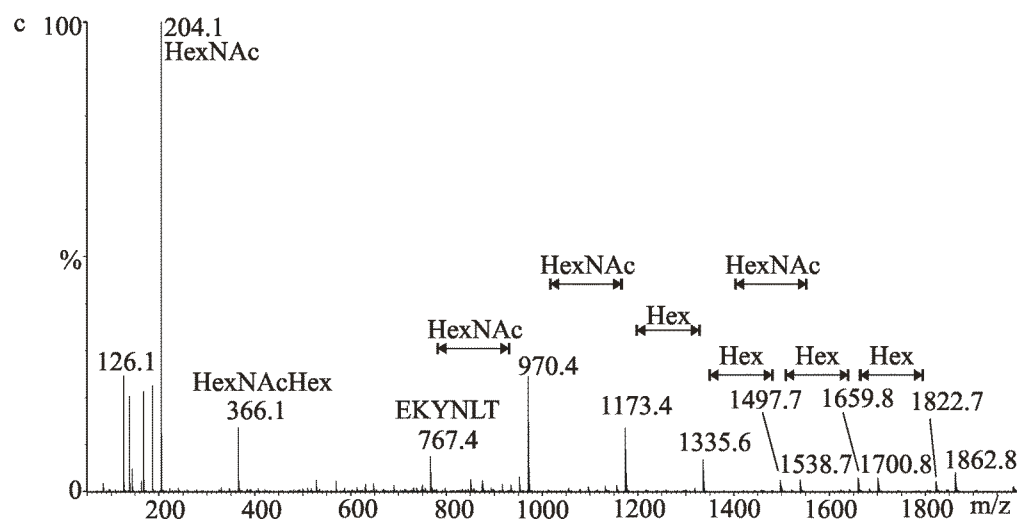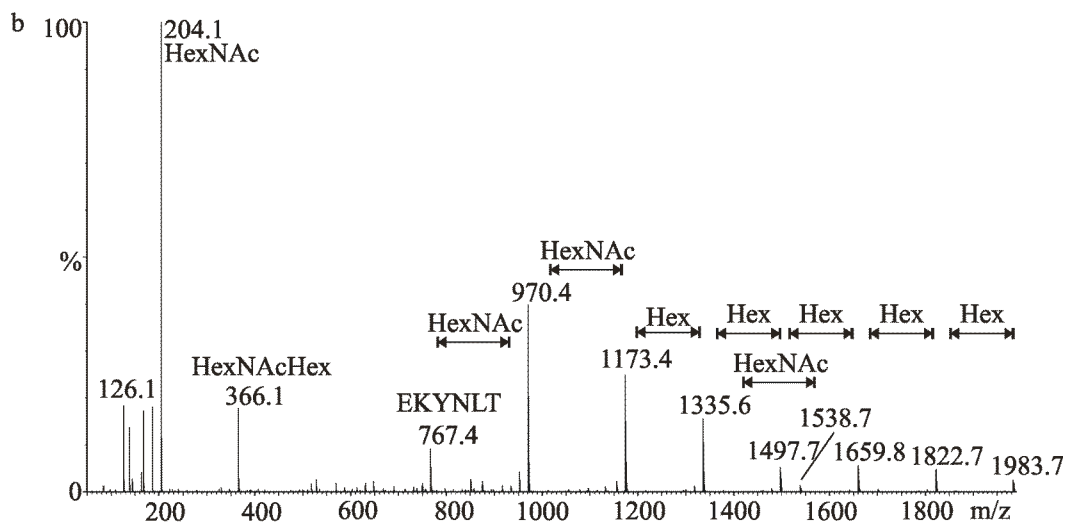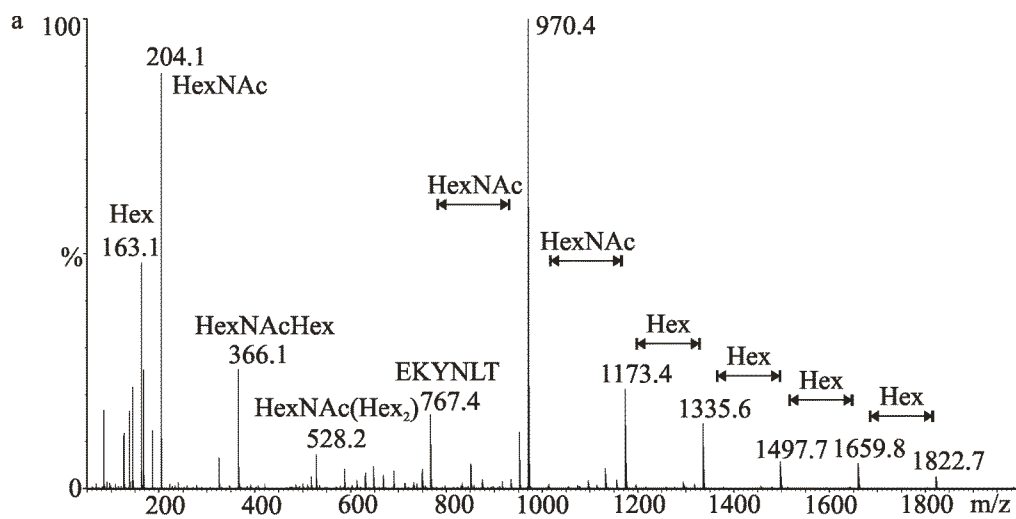
Figure 2.5 MS/MS data obtained from proteinase K digested ovalbumin. MS/MS of doubly charged ions at m/z a) 1073.3, a peptide EKYNLT modified with a HexNAc$_2$Hex$_6$. b) 1094.4, EKYNLT modified with HexNAc$_3$Hex$_5$. c) 1196.0, EKYNLT modified with HexNAc$_4$Hex$_5$.
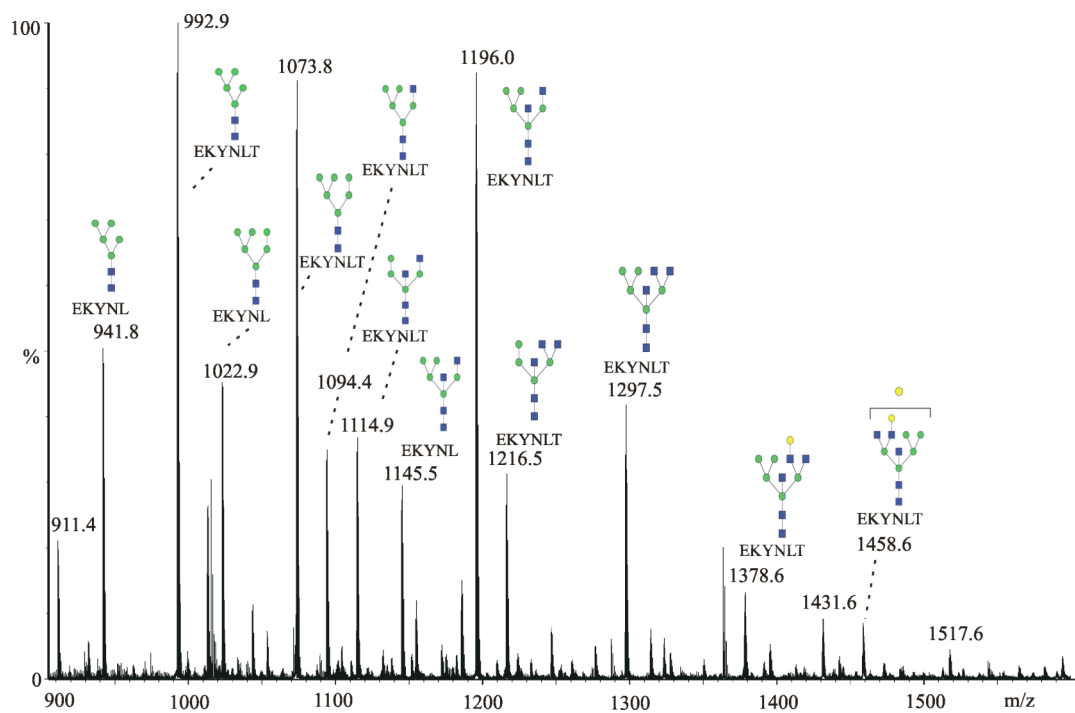
Figure 2.6 Average MS data obtained from proteinase K *in-gel* digested ovalbumin. Most of the peaks are annotated based on previous studies of the glycans of ovalbumin [46].

peaks were annotated. $GlcNAc_2Man_5$, $GlcNAc_2Man_6$, $GlcNAc_3Man_5$, $GlcNAc_4Man_5$, $GlcNAc_4Man_4$, $GlcNAc_5Man_5$ and $GlcNAc_5Man_6$ are among the abundant glycans of ovalbumin detected on its single glycosylation site, Ans^{292}.

2.3.2 Proteinase K Digested *O*-glycosylated Pilins

O-glycosylation has been found in several bacterial species. In *Neisseiria meningitidis* (NM), *Neisseria gonorrhoeae* and some strains of *Pseudomonas aeruginosa* (PA), type IV pilins (structural components of type IV pili) are *O*-glycosylated proteins [47,48]. Pili are filamentous polymeric appendages that protrude from the bacterial surface and play a key role in bacterial virulence. Although the presence of glycosylated pilins has been known for many years, the mechanisms of glycosylation are just starting to be understood [35]. Recombinant glycosylated NM and PA pilins were produced in *E. coli* using glycoengineering. The products were characterized using RP-HPLC ESI QToF to determine their glycosylation sites as well as the glycan structures.

Originally, PA pilin was reported to be glycosylated on Ser^{148} in the *Pseudomonas aeruginosa* 1244 strain [47]. In this study, PA pilin was *in-gel* digested using proteinase K and the resulting peptides and glycopeptides were analyzed by MS and MS/MS using the RP-HPLC ESI QToF mass spectrometer. Two doubly charged ions at m/z 1,012.8 and 1,000.3 were detected in the MS spectra (**Fig**. **2.7**). These glycopeptides were subjected to MS/MS. The MS/MS data presented in **Fig**. **2.8** confirmed the identity of these glycopeptides, which

correspond to (HexNAc)$_6$Hex and DATDH(HexNAc)$_5$Hex attached to the *P. aeruginosa* pilin peptide [144]NCPKS[148] containing either DATDH or HexNAc at the reducing end. This variability in the glycan structure was reported previously [49,50] for recombinant *N*-glycoproteins expressing the *C. jejuni pgl* cluster in *E. coli*. Ser[148] is the only hydroxylated amino acid in this peptide; therefore, we concluded that this amino acid is the glycosylation site in our system. These two glycopeptides were the only glycopeptides identified in our MS analysis of the purified PA pilin. The glycosylation attachment site was therefore conserved when the *P. aeruginosa* system was transferred into *E. coli*.

Figure 2.7 Base peak ion chromatogram and MS of proteinase K digested PA pilin a) Base peak ion chromatogram of the digest eluted from C_{18} ZipTips using 10% acetonitrile. The peaks **I** and **II** are glycopeptides. b) MS extracted from peak **I** of the chromatogram c) MS extracted from peak **II** of the chromatogram.

Figure 2.8 MS/MS of glycopeptides of proteinase K digested PA pilin

a) MS/MS of the glycopeptide [NCPKS + HexNAc$_6$Hex] ion at m/z 1000.3.

b) MS/MS of the glycopeptide [NCPKS + DATDH(HexNAc)$_5$Hex] ion at m/z 1012.3.

Similarly, previous work identified the NM pilin glycosylation site as being contained within the tryptic peptide ^{45}SAVTEYYLNHGEWPGNNTSAGVATSSEIK^{73} [51-54]. However, the exact site of glycosylation remained to be identified with direct analysis [55]. Thus site-specific analysis of the glycoproteins were performed to define the glycosylation site as well as the glycan structures of the recombinant glycoproteins. The purified NM glycosylated pilin was digested with trypsin and the purified digest was analyzed using MALDI ToF MS. The data showed similar result as previous studies (data not shown) [ ]. However, the tryptic glycopeptide was not detected in the the RP-HPLC ESI QToF mass spectrometer. NM pilin was digested with proteinase K and the resulting peptides were analyzed by the RP-HPLC ESI QToF mass spectrometer. The base peak ion chromatogram, MS and MS/MS data of proteinase K digested NM pilin are shown in **Fig**. **2.9**. In the MS data a doubly charged glycopeptide peak at *m/z* 905.8 was detected. The MS/MS of this glycopeptide, presented in **Fig**. **2.9b**, shows that the glycan DATDH(HexNAc)$_5$Hex was attached to the pentapeptide, ^{63}SAGVA^{67}. The tandem MS data showed b and y ions of the peptide which were additional evidence to confirm the identity of the glycan carrier. The only hydroxylated residue of the peptide on which the glycan can be attached is Ser^{63}. From this set of experiments, we concluded that Ser^{63} was the site of *O*-glycosylation in NM pilin.
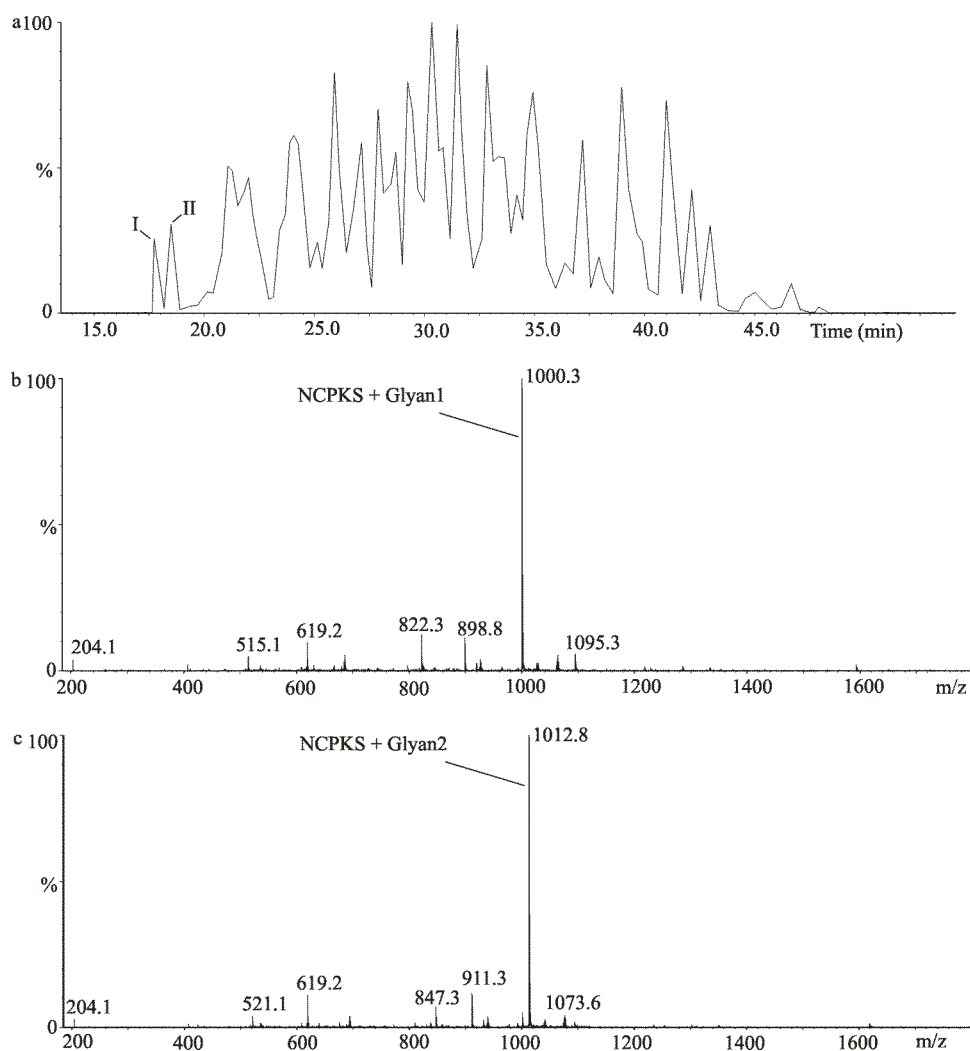
61
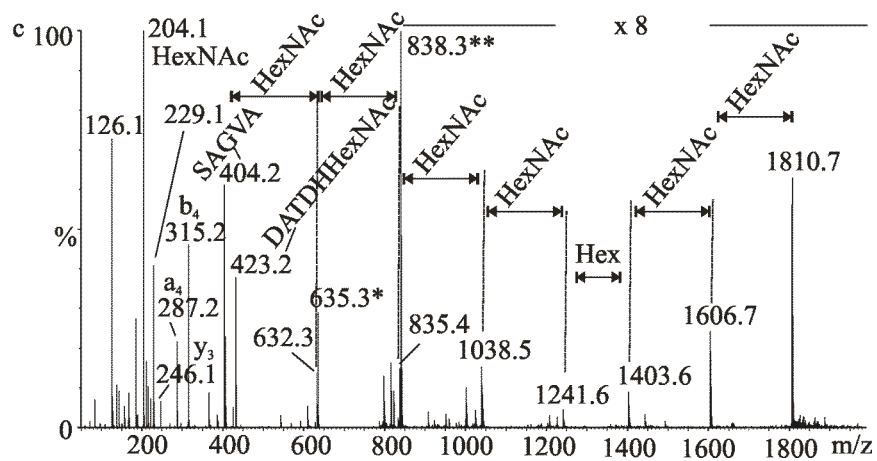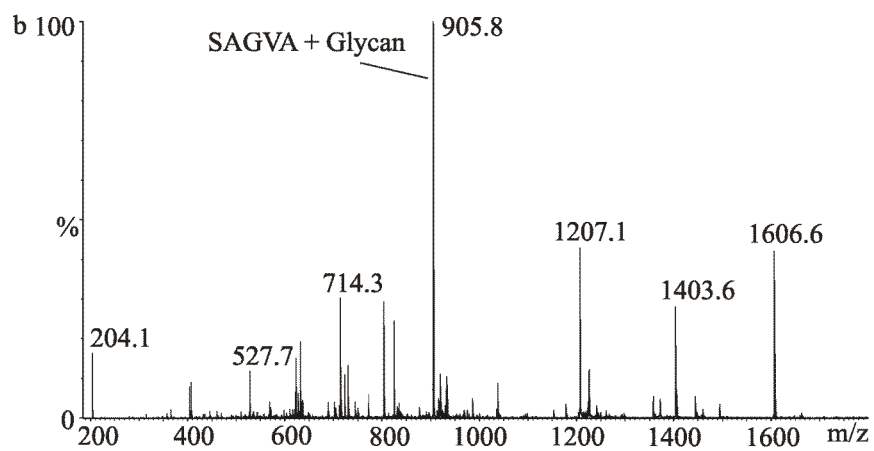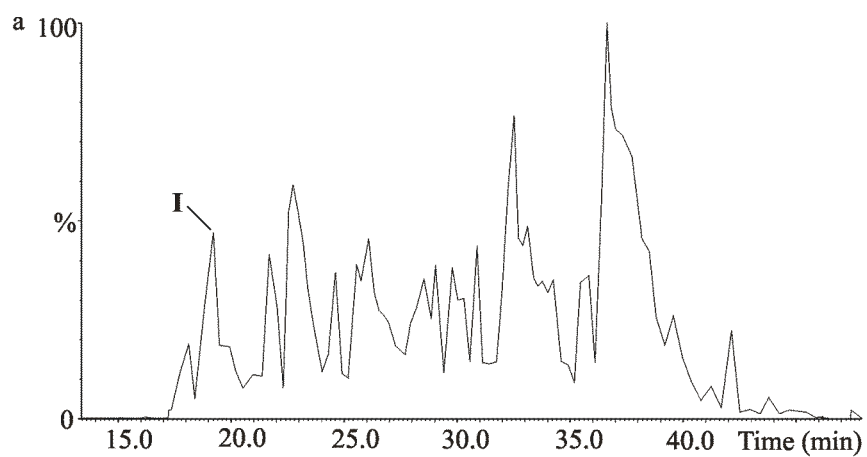
Figure 2.9 Base peak ion chromatogram and MS of proteinase k digested NM pilin a) Base peak ion chromatogram of the digest eluted from C_{18} ZipTips using 10% acetonitrile. b) MS extracted from peak **I** of the chromatogram whose base peak is a doubly charged glycopeptide at m/z 905.8. c) MS/MS of the glycopeptide (m/z 905.8).

2.4 Conclusions

Application of trypsin and proteinase K to hydrolyze glycoproteins to glycopeptides, which will be analyzed by MS and MS/MS is investigated. Although, trypsin is the enzyme of choice for protein identification, proteinase K is superior for site specific analysis of glycoproteins. For example, from the proteinase K digest of glycosylated AcrA, it was possible to detect all of the expected glycopeptides. The MS/MS data of those glycopeptides confirmed that the expected oligosaccharides: $GlcNAc(GalNAc)_5Glc$ and $Bac(GalNAc)_5Glc$ are attached to the peptides: DFNRS and DFNNNRS. Similarly, digestion of chicken egg albumin with proteinase K enabled us to identify most of the glycans reported so far on the protein "*in-situ*". $GlcNac_2Man_5$, $GlcNAc_2Man_6$, $GlcNAc_3Man_5$ and $GlcNAc_4Man_5$ are among the abundant oligosaccharides of the ovalbumin on its single glycosyaltion site, [292]Asn. However from tryptic digest of AcrA, only two of the glycopeptides (one of the glycosylation site was occupied by two oligosaccharides) were detected using the nano RP-HPLC ESI QToF mass spectrometer. Analysis of the tryptic digest of ovalbumin failed to detect any glycopeptide.

Generally, for site-specific analysis of glycans on glycoproteins using RP-HPLC ESI QToF MS, keeping the peptide part short is needed. Identification of glycosylation sites of *N*-linked and *O*-linked glycoproteins became easier since the number of residues of the glycan carrier is few (3 to 8 amino acid). It made glycopeptide desalting and separation on reversed phase C_{18} columns practical without permethylation of glycans. By keeping the peptide mass low, quality of

both MS and MS/MS data of glycopeptides have been improved. Identification of glycosylation sites of *N*-liked and *O*-linked glycoproteins of bacteria and eukaryote became easier since the number of residues of the glycan carrier is few (3 to 8 amino acides). Proteinase K digested *O*-glycosylated pilins of *N. meningitids* and *P. aerogunosa* were analyzed using RP-HPLC ESI QToF mass spectrometer. The glycosylation site of NM pilin was found to be Ser[63].

2.5 Literature Cited

1. Rudd, p.M.; Elliott, T.; Cresswell, P.; Wilson, I. A. Dwek, R. A. *Science* **2001**, *29*, 2370-2376.

2. Dennis, J. W.; Granovsky, M.; Warren, C. E. *Bioessays* **1999**, *21,* 412-421.

3. Szymanski, C. M.; Wren, B. W. *Nat. Rev. Microbiol*. **2005**, *3*, 225-237.

4. Varki, A. *Glycobiology* **1993**, *3*, 97-130.

5. Chandrasekaran, A.; Srinivasan, A.; Raman, R.; Viswanathan, K.; Raguram, S.; Tumpey, T. M; Sasisekharan, V.; Sasisekharan, R. *Nat. Biotechnol.* **2008**, *26*, 107-113.

6. Stevens, J.; Blixt, O.; Tumpey, T. M.; Taubenberger, J. K.; Paulson, J. C.; Wilson, I. A. *Science* **2006**, *312(5772),* 404-410.

7. Wurm, F. M. *Nat. Biotech.,* **2004** *22*, 1393- 1398.

8. Morelle, W.; Michalski, J. C. *Nat. Protoc*. **2007**, *2*, 1585-1602.

9. Harvey, D. J. *Proteomics* **2001**, *1*, 311-328.

10. Spiro, R. G. *Glycobiology* **2002**, *12,* 43-56.

11. Goldberg, D.; Bern, M.; Parry, S.; Sutton-Smith, M.; Panico, M.; Morris, H. R.; Dell, A. *J. Proteome Res.* **2007**, *6,* 3995-4005.

12. Dell, A.; Morris, H. R. *Science* **2001**, *291,* 2351-2356.

13. Haslam, S. M.; North, S. J.; Dell, A. *Current Opinion in Structural Biology* **2006**, *16*, 584-591.

14. Kita, Y.; Miura, Y.; Furukawa, J.; Nakano, M.; Ohno, M.; Takimoto, A.; Nishimura, S. *Mol. & Cellular Proteomics* **2007**, *6*, 1437-1445.

15. An, H. J.; Ninovuevo, M.; Aguilan, J.; Liu, H.; Lebrilla, C. B.; Alvarenga, L. S.; Mannis, M. J. *J. Proteome Res.* **2005**, *4,* 1981-1987.

16. Goetz. J. A.; Novotny, M. V.; Mechref, Y. *Anal. Chem.* **2009**, *81,* 9546-9552.

17. Mirgorodskaya, E.; Hassan, H.; Clausen, H.; Roepstorff, P. *Anal. Chem.* **2001**, *73*, 1263-1269.

18. Lei, M.; Mechref, Y.; Novotny, M. V. *J. Am. Soc. Mass Spectrom.* **2009**, *20,* 1660-1671.

19. Kubota, K.; Sato, Y.; Suzuki, Y.; Goto-Inoue, N.; Toda, T.; Suzuki, M.; Hisanaga, S.; Suzuki, A.; Endo, T. *Anal. Chem.* **2008**, *80*, 3693-3698.

20. Yu, S. Y., Wu, S. W.; Khoo, K. K. *J. Glycoconj.* **2006**, *23*, 355-369.

21. Jiang, H. Desaire, H.; Butney, V. Y.; Bousfield, G. R. *J. Am. Soc. Mass Spectrom.* **2004**, *15,* 750-758.

22. Zhang, Y.; Go, E. P.; Desaire, H. *Anal. Chem.,* **2008**, *80(9*), 3144-3158.

23. Wu, C. C.; Yates, J. R. *Nature Biotechnology* **2003**, *21*, 262-267.

24. Wuhrer, M.; Hokke, C. H.; Deelder, A. M. *Rap. Commun. Mass Spectrom.* **2004**, *18*, 1714-1748.

25. Yang, Z.; Hancock, W. S.; Chew, T. R.; Bonilla, L. *Proteomics*, **2005**, *5*, 3353-3366.

26. Kuster B.; Mann, M. *Anal. Chem*., **1999**, *71(7),* 1430-1440.

27. Andersen, M. T.; Mysling, S.; Hojrup, P. *Anal. Chem.* **2009**, *81(10),* 3933-3943.

28. Krokhin, O.; Ens, W.; Standing, K. G.; Wilkins, *J. Rap. Commun. Mass Spectrom.* **2004**, *18*, 2020-2030.

29. Juhasz, P.; Martin, S. A. *Int. J. Mass Spectrom. Ion Processes* **1997**, *164*, 217-230.

30. An, H. J.; Peavy, T. R.; Hedrick, J. L.; Lebrilla, C. B. *Anal. Chem.* **2003**, *75*, 5628-5637.

31. Larsen, M. R.; Hojrup, P.; Roepstorff, P. *Mol. Cell Proteomics* **2005**, *4*, 107-119.

32. Liu, X.; McNally, D. J.; Nothaft, H.; Szymanski, C. M.; Brisson, J.; Li, J. *Anal. Chem,* **2006**, *78,* 6081-6087.

33. Jiang, H.; Desaire, H., Desaire, H.; Butnev, V. Y.; Bousfield, G. R. *J. Am. Soc. Mass Spectrom.* **2004**, *15,* 750-758.

34. Weiskopf, A. S., Vouros, P., Harvey, D. J. *Anal. Chem.* **1998**, *70*, 4441-4447.

35. Karlsson, N. G., Packer, N. H. *Anal. Biochem.* **2002**, *305*, 173-185.

36. Novotny, M. V.; Mechref, Y. *J. Sep. Sci.* **2005**, *28*, 1956-1968.

37. Farydmoayer, A.; Fentabil, M. A.; Mills, D. C. Klassen, J. S.; Feldman, M. F. *J. Bacteriology* **2007**, *189(22),* 8088-8098.

38. Laemmli, U. K. *Nature* **1970**, *227*, 680-685.

39. Lowry, O. H.; Rosebrough, N. J.; Farr, A. L. *J. Biol. Chem.* **1951**, *193*, 265-275.

40. Shevchenko, A.; Jensen, O. N.; Wilm, M.; Vorm, O.; Mortensen, P.; Shevchenko, A.; Boucherie H.; Mann, M. *Proc. Natl. Acad. Sci.* **1996**, *93*, 14440-14445.

41. Young, N. M.; Brisson, J. R.; Kelly, J.; Watson, D. C.; Tessier, L.; Lanthier, P. H.; Jarrell, H. C.; Cadotte, N.; St Michael, F.; Szymanski, C. M. *J. Biol. Chem.* **2002**, *277*, 42530-42539.

42. Wacker, M.; Linton, D.; Hitchen, P. G.; Nita-Lazar, M.; Haslam, N. M.; North, S. J.; Panico, M.; Morris, H. R.; Dell, A.; Wren, B. W.; Aebi, M. *Science* **2002**, *298*,1790-1793.

43. Feldman, M. F.; Wacker, M.; Hernandez, M.; Hitchen, P. G.; Marolda, C. L.; Kowarik, M.; Morris, H. R.; Dell, A.; Valvano, M. A.; Aebi, M. *Proc. Natl. Acad. Sci.* **2005**, *102*, 3016-3021.

44. Ebeling, W.; Hennrich, N.; Klockow, M.; Metz, M.; Orth, H. D.; Lang, H. *Eur. J. Biochem.* **1974**, *47*, 91-97.

45. Bezouska, K.; Sklenar, J.; Novak, P.; Halada, P.; Havlicek, V.; Kraus, M.; Ticha, M.; Jonakova, V., *Protein Sci.,* **1999**, *8,* 1551-1556.

46. Neuberger, A. *J. Biochem.,* **1938**, *32*, 1435-1451.

47. Harvey, D. J.; Wing, D. R.; Kuster, B.; Wilson, I. B. *J. Am. Mass Spectrom.* **2002**, *11*, 564-575.

48. Castric, P. *Microbiology* **1995**, *141*, 1247-1254.

49. Virji, M.; Saunders, J. R.; Sims, G.; Makepeasce, K.; Maskell, D.; Ferguson, D. J. *Mol. Microbiology* **1993**, *10*, 1013-1028.

50. Logan, S. M. *Microbiology* **2006**, *152*, 1249-1262.

51. Ohtsubo, K.; Marth, J. D. *Cell* **2006**, *126*, 855-867.

52. DiGiandomenico, A.; Matewish, M. J.; Bisaillon, A.; Stehle, J. R.; Lam, J. S.; Castric, P. *Mol. Microbiol.* **2002**, *46,* 519-530.

53. Power, P. M.; Seib, K. L.; Jennings, M. P. *Biochem. Biophys. Res. Commun.* **2006**, *347*, 904-908.

54. Aas, F. E.; Vik, A.; Vedde, J.; Koomey, M.; Egge-Jacobsen, W. *Mol. Microbiol.* **2007**, *65*, 607-624.

55. Stimson, E.; Virji, M.; Makepeace, K.; Dell, A.; Morris, H. R.; Payne, G.; Saunders, J. R.; Jennings, M. P.; Barker, S.; Panico, M. *Mol. Microbiol.* **1995**, *17*, 1201-1214.

56. Hegge, F. T.; Hitchen, P. G.; Aas, F. E.; Kristiansen, H.; Lovold, C.; Egge-Jacobsen, W.; Panico, M.; Leong, W. Y.; Bull, V.; Virji, M.; Morris, H. R.; Dell, A.; Koomey, M. *Proc. Natl. Acad. Sci.* **2004**, *101*, 10798-10803.

Chapter 3

Blackbody Infrared Radiative Dissociation of Protonated Oligosaccharides Using FT-ICR MS

3.1 Introduction

With the development of soft ionization techniques, such as matrix assisted laser desorption ionization [1] and electrospray ionization [2,3], mass spectrometry has become an essential technique for structure elucidation of biomolecules such as proteins [4-6], oligonucleotides [7,8] and oligosaccharides [9-17]. Tandem MS is usually necessary to providing sequence and other detailed structural information. Tandem MS approaches involve isolating an ion of interest in the gas-phase in time or space, dissociating it to fragment ions and accurately determining the mass of the product ions. Thus, the success of MS/MS to provide structural information about the precursor ion requires the ability to relate the fragments observed in a mass spectrum back to the original structure of the ion. Fundamental understanding of the gas-phase dissociation mechanisms of biomolecules will facilitate identification of the precursor ion structure and the rational development of tandem MS techniques.

Various tandem MS techniques, such as collision induced dissociation (CID) [10], photo dissociation (e.g. infrared multiphoton dissociation (IRMPD) [18,19]), electron capture dissociation (ECD) [20], electron detachment dissociation (EDD) [21] and post-source decay (PSD) [22] have been applied to structural characterization of oligosacchrides. Effort has been exerted to obtain a

variety of structural data, including sequence and stereochemistry information such as anomeric configuration [23,24] and linkage types [25-27], through the MS/MS analysis. Most of the MS/MS studies of oligosaccharides are performed on metal coordinated oligosaccharides. For example, the sodium cation forms a stable complex with oligosaccharides, which are extensively studied [28-30]. Suzuki et al established general rules for fragmentation of sodium adduct of several oligosaccharides using computational and experimental data [28]. According to the study, the dissociation of glycosidic bonds involves transfer of protons from the hydroxyl group on C_2 to the glycosidic bond (see scheme 1).
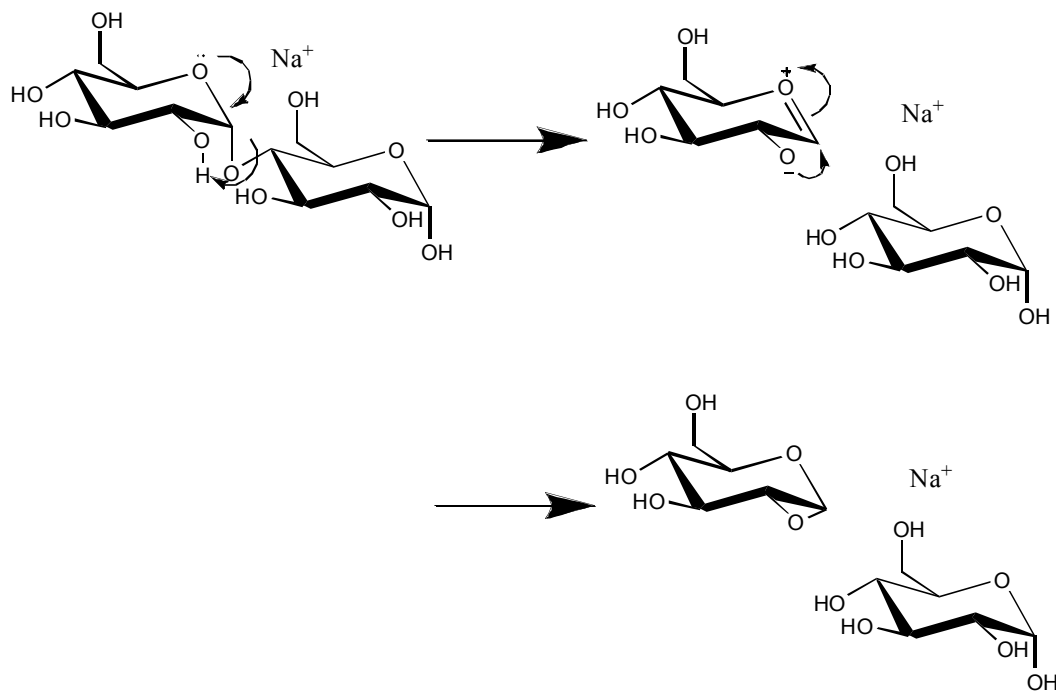
Scheme 3.1 Reaction mechanisms of fragmentation of the glycosyl bond of maltose.

However, the dissociation pathways, mechanisms and energetics of protonated oligosaccharides during MS/MS analysis remains less understood.

In this Chapter study, Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) is used to understand dissociation pathways, kinetics and energetics of protonated oligosaccharides. It provides long trapping times of ions (minutes – hours) and and the vicinity of the ion cell is at ultra high vacuum (pressure down to 10^{-10} mbar). As a result, a variety of tandem MS techniques can be employed to dissociate the ions trapped in the ICR cell [20-23, 40]. One of the tandem MS techniques used with FT-ICR MS is blackbody infrared radiative dissociation (BIRD) [40, 41]. It is a thermal activation technique. The pressure in the surrounding of ICR cell ($< 10^{-9}$ mbar) is so low that collision activation as a means for dissociation can be neglected [43]. As a result, BIRD has emerged as an important tool to study kinetics and energetics of thermal gas-phase ion dissociation [44-50]. In the present work, it was applied to protonated oligosaccharides to understand their dissociation pathways, kinetics and energetics.

The gas-phase dissociation of protonated maltooligosaccharides, (α-D-Glc$_p$-1$\rightarrow$(4-α-D-Glc$_p$-1)$_n\rightarrow$4- D-Glc$_p$, n = 3,4,5,6) and cellohexaose, (β-D-Glc$_p$-1$\rightarrow$(4-β-D-Glc$_p$-1)$_n\rightarrow$4- D-Glc$_p$, n = 4) were studied. Over the range of temperatures investigated (70 ^{o}C to 160 ^{o}C), the protonated oligosaccharides dissociate via cleavage at the glycosidic linkages and/or exclusive loss of water molecules forming B and Y ions according to the nomenclature of Domon and Costello [36]. The dissociation pathways through which the B and Y ions formed

71

were investigated in detail using double resonance experiments. Arrhenius activation parameters, activation energy (E_a) and pre-exponential factor (A) for protonated alpha and beta linked oligosaccharides were obtained. To our knowledge, this is the first report of dissociation kinetic and pathway study of protonated oligosaccharides.

3.2 Materials and Methods

3.2.1 Oligosaccharides

Maltopentaose (α-D-Glc$_p$-(1→4)-α-D-Glc$_p$-(1→4)- α-D-Glc$_p$-(1→4)- α-D-Glc$_p$-(1→4)-Glc$_p$), Mal$_5$), maltohexaose (α-D-Glc$_p$-(1→4)-α-D-Glc$_p$-(1→4)- α-D-Glc$_p$-(1→4)-α-D-Glc$_p$-(1→4)-α-D-Glc$_p$-(1→4)Glc$_p$, Mal$_6$) and maltoheptaose (α-D-Glc$_p$-(1→4)-α-D-Glc$_p$-(1→4)-α-D-Glc$_p$-(1→4)-α-D-Glc$_p$-(1→4)-α-D-Glc$_p$-(1→4)-α-D-Glc$_p$-(1→4)-Glc$_p$, Mal$_7$) were purchased from Sigma-Aldrich Canada (Oakville, Canada). Maltooctaose (α-D-Glc$_p$-(1→4)-α-D-Glc$_p$-(1→4)-α-D-Glc$_p$-(1→4)-α-D-Glc$_p$-(1→4)-α-D-Glc$_p$-(1→4)-α-D-Glc$_p$-(1→4)-α-D-Glc$_p$-(1→4)-Glc$_p$, Mal$_8$) and cellohexaose (β-D-Glc$_p$-(1→4)-β-D-Glc$_p$-(1→4)- β-D-Glc$_p$-(1→4)-β-D-Glc$_p$-(1→4)-β-D-Glc$_p$-(1→4)Glc$_p$, Cel$_6$) were purchased from CarboSynth (Berkshire, U.K.). All of the oligosaccharides were used without further purification. The oligosaccharides were dissolved in MilliQ Plus^{TM} water (Millipore, Bedford, MA) which has 2.5 mM ammonium acetate whose pH was adjusted to 6.8 by using NaOH. ES was performed on aqueous solutions containing 50 μM oligosaccharide and 2.5 mM ammonium acetate.

3.2.2 Mass Spectrometry

All experiments were performed on a 9.4 tesla Bruker Apex II Fourier transform ion cyclotron resonance mass spectrometer (Billerica, MA) equipped with a nanoflow electrospray ion source. Detailed descriptions of the instrument and the experimental and instrumental parameters used in the nanoES-MS measurements and the BIRD experiments are given elsewhere [49-51]. Briefly, nanoES tips were pulled from borosilicate glass tubes (1.0 mm o.d. and 0.78 mm i.d.) using a P-97 micropipette puller (Sutter Intruments, Novato, CA). A platinum (Pt) wire, inserted into the open end of the nanoES tip, was used to apply an approximately 1000 V potential to the nanoES solution. The tip was positioned 1-2 mm from a stainless steel sampling capillary using a microelectrode holder. The solution flow rate typically ranged from 20 to 50 nL/min depending on the diameter of the nanoES tip and the electrosray voltage.

The gaseous ions produced by nanoES were introduced into the mass spectrometer through a heated stainless steel capillary (0.43 mm i.d.) maintained at a temperature (external) of approximately 70 ^{o}C. Ions were accumulated in an external hexapole for 4 to 8 s, subsequently transferred to the ICR cell via a series of electrostatic optics. During MS/MS of ions, the precursor ion of interest was isolated and trapped in the ICR cell while the other ions were ejected by applying radio frequency (RF). The temperature of the ion cell, which established the reaction temperature, was controlled with two external flexible heating blankets placed around the vacuum tube near the ion cell. The reaction temperatures were equilibrated for 3 to 4 hours before the experiments were started. The reaction

delay time (before ion excitation and detection but after ion injection into the FT-ICR cell) was varied within the experiment to observe the degree of blackbody infrared dissociation that occurred while the protonated oligosaccharides were trapped in the cell. Mass spectra were acquired by an SGI R5000 computer running the Bruker Daltonics XMASS software, version 5.0. On average 20 scans, containing 256 K data points per scan, were acquired per spectrum.

3.2.3 Theoretical Calculations

Initial structures of alpha and beta linked D-glucopyranose disaccharides were sketched using the HyperChem (Hyper Cube plc.) database and model builder. Different conformers of the saccharides were searched using the AMBER force field in HyperChem. Each of the conformers was taken for geometry optimization using semiemperical calculations (AM1) in HyperChem. The minimum energy conformer was chosen for AM1 geometry optimization in Gaussian. All the Gaussian computations were performed using Gaussian 03 available in University of Alberta cluster server.

3.3 Results and Discussion

3.3.1 BIRD of Protonated Oligosaccharides

Protonated oligosaccharides were produced by electrospraying aqueous ammonium acetate (2.5 mM) solution containing 50 µM oligosaccharide. The major ions detected in the mass spectra correspond to a protonated as well as NH_4^+, Na^+ and K^+ adducts of the oligosaccharides. **Fig. 3.2a** is a representative

MS data acquired from a solution of Mal_6. The species observed were [Mal_6 + H – H_2O]$^+$ (m/z 973.2), [Mal_6 + H]$^+$ (m/z 991.2), [Mal_6 + NH_4]$^+$ (m/z 1008.2), [Mal_6 + Na]$^+$ (m/z 1013.2) and [Mal_6 + K]$^+$ (m/z 1029.2). Strong protonated ion peaks were detected in the presence of ammonium acetate using the nano ES. For example, the base peak in **Fig. 3.2a** is the protonated ion of maltohexaose. Similar mass spectra were obtained for all of the oligosaccharides (Mal_5, Mal_7, Mal_8 and Cel_6) investigated (supporting data).

BIRD of the protonated oligosaccharides was carried out at temperatures of 78 to 160 ^{o}C for reaction times of 0.5 to 90 s. **Fig. 3.2b** shows isolated protonated peak of Mal_6. This ion was subjected to fragmentation using different temperatures and reaction times. **Fig. 3.2c** shows a representative BIRD mass spectrum acquired at 100 ^{o}C and reaction time of 7 s. B (B_1, B_2, B_3, B_4, B_5 and B_6) and Y-ions (Y_4 and Y_5) were the major product ions. Generally, during BIRD of the protonated oligosaccharides, the B and Y type ions are the major product ions. Similar observation was reported by Harvey et al that during CID activation of protonated oligosaccharides, B and Y ions were observed, exclusively [51]. These ions are formed from glycosidic bond dissociation (see **Fig. 3.1**). Since BIRD is highly selective for low energy fragmentation [36] at low temperatures, the glycosidic bond dissociation is most likely the lowest energy fragmentation pathway.
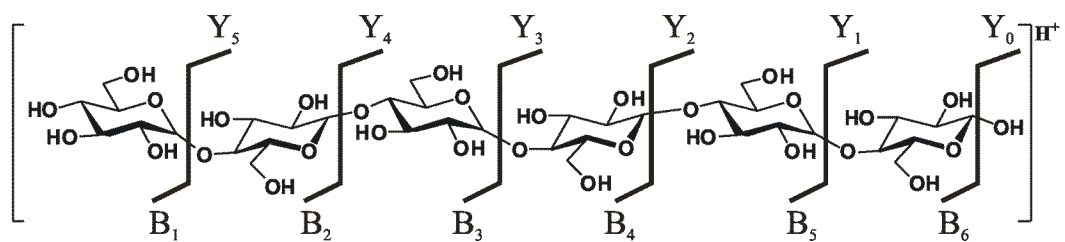
Figure 3.1   B and Y ions according to the Domen and Costello nomencleature [34] obtained during BIRD of protonated oligosaccharides.
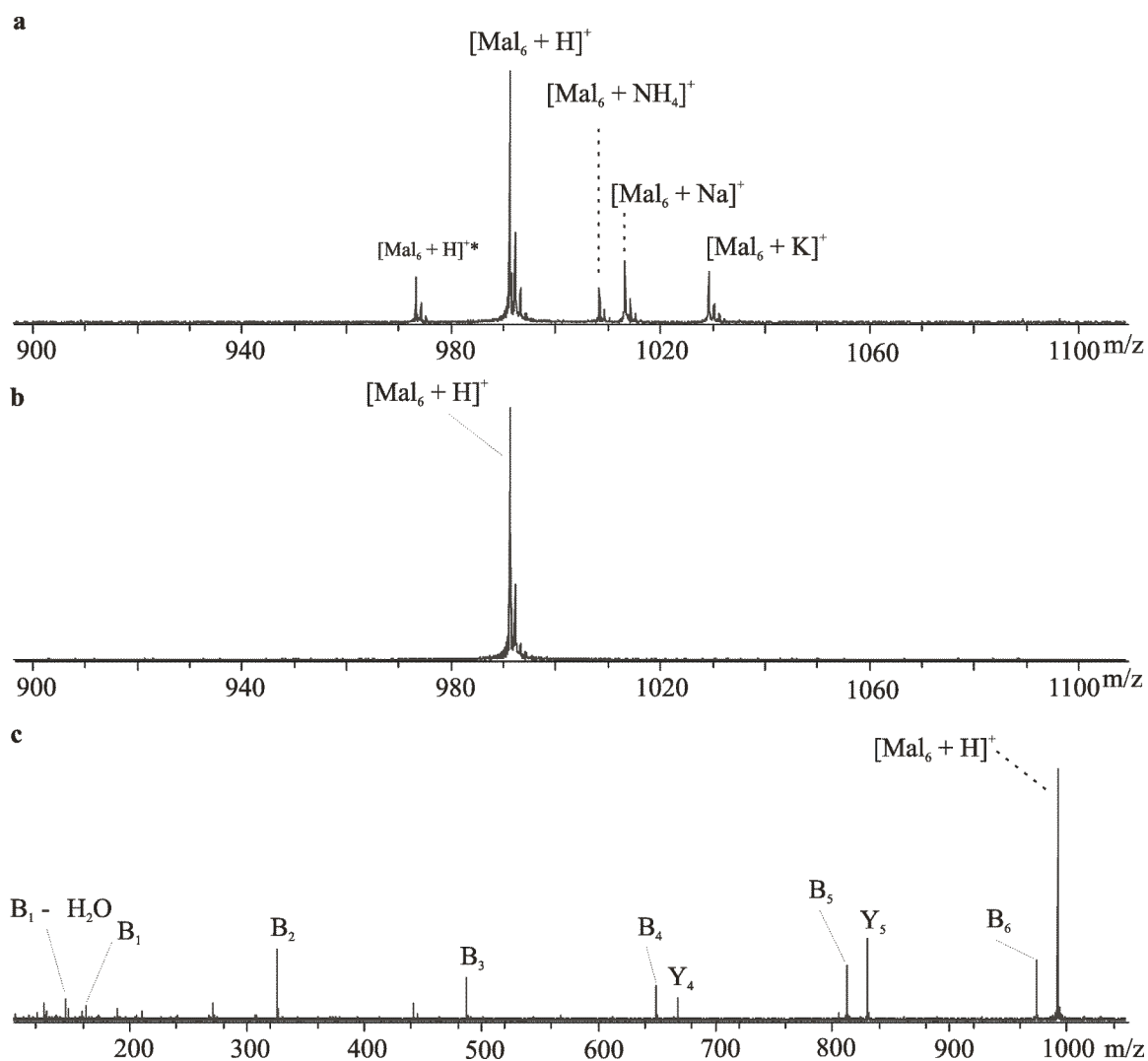
Figure 3.2 NanoES-FT-ICR MS of Mal$_6$ a) MS spectrum of Mal$_6$, the peaks at m/z 973.1, 991.1, 1008.1 and 1029.1 correspond to the [Mal$_6$ + H − H$_2$O]$^+$, [Mal$_6$ + H]$^+$, [Mal$_6$ + NH$_4$]$^+$, [Mal$_6$ + Na]$^+$, [Mal$_6$ + K]$^+$, respectively. b) MS spectrum of the protonated Mal$_6$ ion, [Mal$_6$ + H]$^+$ at 991.1 was isolated. c) BIRD spectrum of [Mal$_6$ + H]$^+$ using a temperature at 100 ^{o}C and reaction time delay of 7 s.

3.3.2 Dissociation Pathways

The dissociation pathways through which the B and Y ions formed during activation of protonated oligosaccharides were investigated. The ions may form in a competitive or consecutive manner from a single precursor ion or there may be multiple structures. As will be described latter, based on our kinetic studies, the possibility of multiple structures is excluded. The B and Y ion formation from a single product ion is investigated in detail. For example, the different B-type product ions of Mal_6 can form directly from the dissociation of the precursor ion in a parallel reaction or by the sequential fragmentation of the product ions to secondary ions. Dissociation of Mal_6 is described in detail to illustrate the dissociation pathways of protonated oligosaccharides. The MS/MS data of protonated Mal_6 obtained at 117 ^{o}C and reaction time 6.5 s is shown in **Fig**. **3.3a**. The B_6 (m/z 973.2) ion resulted from H_2O loss of the precursor ion (m/z 991.2). Similarly, the Y_5 ion (m/z 829.1) resulted from the loss of a monomer at the nonreducing end. Y_5 cannot be a secondary ion of B_6 because the mass difference between B_6 and Y_5 (973.2 − 829.1 = 144.1 Da) is different from the expected residual mass of the lost monosaccharide (162.0 Da). Thus, it is possible to conclude that at least these two product ions formed from the precursor ion in a competitive reaction (see Scheme 3.2).
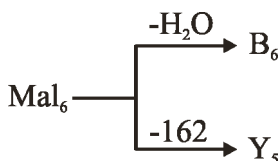
Scheme 3.2. B_6 and Y_5 ions of Mal_6 form in a competitive reaction.
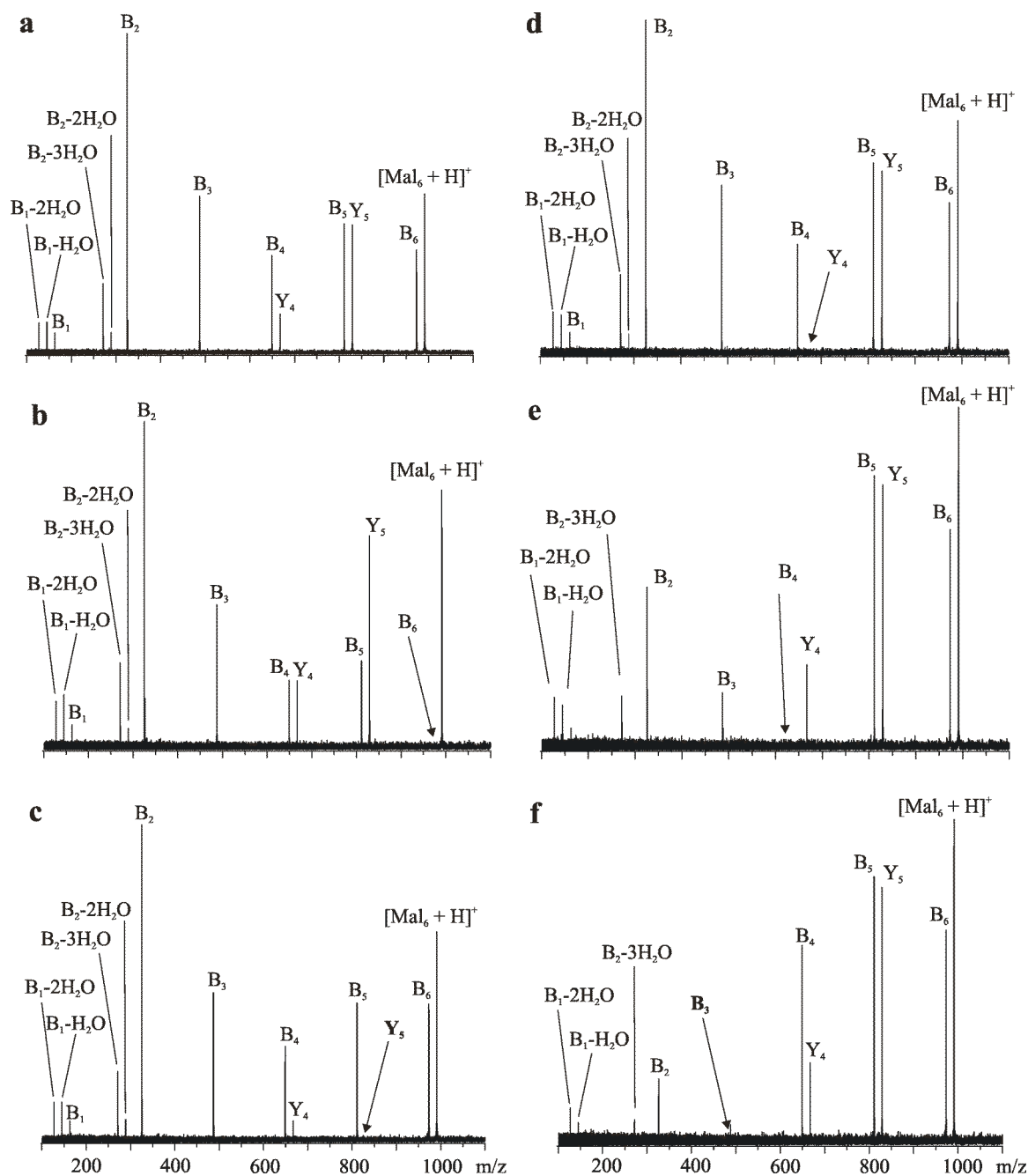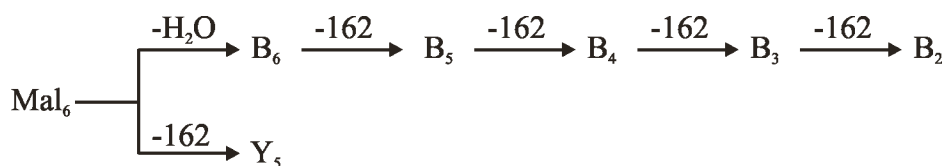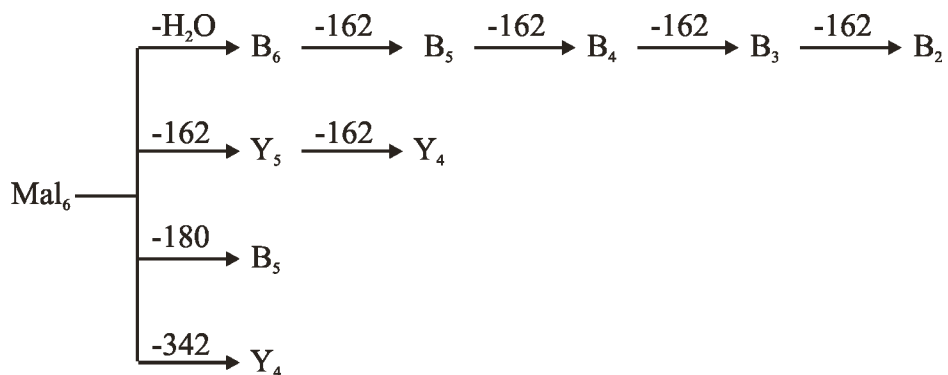
Figure 3.3 Mass spectra of Mal_6 at 117 ^oC and 6.5 s. a) MS/MS of Mal6 without RF pulse. Product ions of interest were removed by applying RF in all the other data: b) B_6 ion (m/z 1135.2); c) Y_5 ion (m/z 991.l); d) Y_4 ion (m/z 667); e) B_4 ion (m/z 649); and f) B_3 ion (m/z 487).

To further investigate the dissociation pathways, double resonance experiments were performed. Double resonance experiments provide insight about dissociation of product ions further to secondary ions. It is performed by exploiting the power of the FT-ICR MS to remove product ion of interest during its formation from the precursor ion. By applying an RF-pulse corresponding to a product ion of interest, the target product ion is removed while it forms. As a result, the effect of the removal of the target product ion on the intensities of the other product ions is investigated. Mass spectra of protonated Mal_6 ions involving double resonance experiments are shown in **Fig**. **3.3b,c,d,e,f**. In **Fig**. **3.3b**, the B_6 ion was removed during its production from the precursor ion throughout the data acquisition. As a result, the relative intensity of several peaks changed (see Table 3.1). For example, the relative abundance of B_5 has changed from 0.8 to 0.3, which is calculated based on ratio of the product ion to the precursor ion. This implies that B_6 is a source of B_5 ion. Similarly, removal of B_6 ion affected the intensity of B_4, B_3 and B_2 ions (see Table 3.1). Based on those observations, the dissociation of the B ions further to smaller B ions was identified to be a sequential loss of one sugar (see scheme 3.3).

$$Mal_6 \xrightarrow[]{-H_2O} B_6 \xrightarrow{-162} B_5 \xrightarrow{-162} B_4 \xrightarrow{-162} B_3 \xrightarrow{-162} B_2$$
$$Mal_6 \xrightarrow{-162} Y_5$$

Scheme 3.3. The series of B ions form in a sequential loss of monosaccharides from the bigger B ions. B_6 and Y_5 ions of Mal_6 form via competitive reactions.

However, there is still a strong peak of B_5 (see **Fig. 3.3b**) when B_6 was removed. Thus, the B_5 ion must be formed from more than one source. Other possible sources are the precursor ion Mal_6 (m/z 991.2) and the product ion, Y_5 (m/z 829.1). **Fig. 3.3c** shows an MS/MS data while the product ion, Y_5, was removed. The data show that the contribution of Y_5 to the B_5 ion is insignificant (Table 3.1). This indicates that B_5 forms from Mal_6 by the loss of monosaccharide (180 Da) from the reducing end. Since the Y_5 itself results from loss of a monosaccharide from the nonreducing end of Mal_6, both the reducing and nonreducing ends are liable to dissociation. Furthermore, the ratio of Y_4 ion (Table 3.1) was changed significantly when Y_5 was removed, which implies that Y_4 is a secondary ion of Y_5. However, there is still the peak of the Y_4 ion in the absence of Y_5. Thus Y_4 as well as Y_5 are formed by two parallel reaction pathways from the precursor ion. The Mal_6 ion can lose di-saccharides from the nonreducing end to form Y_4 ion. Based on the observations, a new schematic diagram is proposed (Scheme 3.4).

$$Mal_6 \begin{cases} \xrightarrow{-H_2O} B_6 \xrightarrow{-162} B_5 \xrightarrow{-162} B_4 \xrightarrow{-162} B_3 \xrightarrow{-162} B_2 \\ \xrightarrow{-162} Y_5 \xrightarrow{-162} Y_4 \\ \xrightarrow{-180} B_5 \\ \xrightarrow{-342} Y_4 \end{cases}$$

Scheme 3.4. The dissociation pathway of Mal_6 ion follows both competitive as well as sequential reactions.

Table 3.1 Ratio of intensities of product ion peaks to Mal_6 calculated based on the mass spectra shown in **Fig**. 3.3. The **X** represents the product ion removed using double resonance technique. The first row of the data represents the ratios calculated before any product ion was removed. The instrumental parameters, reaction time and temperature were kept similar for acquiring all the mass spectra of the other data, but RF pulse applied corresponding to the product ion of interest. The asterisk (*) and (**) represents one and two water molecule loss from the B ions.

B_1*	B_1	B_2**	B_2*	B_2	B_3	B_4	Y_4	B_5	Y_5	B_6	Mal_6
0.2	0.1	0.4	0.1	2.1	1.1	0.6	0.3	0.8	0.8	0.7	1
0.2	0.1	0.3	0.1	**1.3**	**0.6**	**0.3**	0.3	**0.3**	0.8	X	1
0.2	0.1	0.3	0.1	2.3	0.8	0.5	**0.1**	0.7	X	0.7	1
0.2	0.1	0.3	0.1	**1.0**	**0.4**	**0.2**	0.3	X	0.8	0.7	1
0.2	0.1	0.4	0.1	1.5	0.8	0.5	X	0.8	0.8	0.7	1
0.1	0.1	0.2	0.0	**0.5**	**0.2**	X	0.3	0.8	0.8	0.7	1
0.1	0.0	**0.1**	0.0	**0.2**	X	0.6	0.3	0.8	0.8	0.7	1
0.0	0.0	**0.0**	0.0	X	1.1	0.7	0.3	0.8	0.8	0.7	1

During dissociation of the protonated malto-oligosaccharides (Mal_5, Mal_6, Mal_7 and Mal_8) the relative intensity of the B_2 ion increased with reaction temperature and time. For example, the base peak in **Fig. 3.3a** corresponds to the B_2 ion of Mal_6. It remained as a base peak in the absence of the B_6 ion although there was a significant decrease in its relative intensity. As described above, every ion has a potential to dissociate further into secondary ions. However, the B_2 ion remained the major species at longer reaction times and higher temperatures. This implied that the B_2 oxonium ion is very stable. However, removal of the B_3 ion (**Fig. 3.3d**) diminished the intensity of B_2. The significant effect of B_3 implied that it is the major source of B_2 ion via loss of monosaccharide. However, the presence of the B_2 ion in the absence of B_3 implied that it forms by loss of more than one sugar unit from bigger ions (e.g. B_5, B_6, Y-ions). In **Fig. 3.3e** and **f**, oxonium ions of the monosaccharide (B_1) together with dehydrated peaks (B_1-H_2O, B_1-$2H_2O$) are depicted. The presence of those dehydrated ions implied that the dehydration reaction of the product ions became significant when the product ions are getting smaller in mass.

In conclusion, these observations confirmed that many of the glycosydic bonds of oligosaccharides are prone to dissociation. Generally, the precursor ion could lose mono, di or tri-saccharides depending on the size of the oligosaccharide, competitively. Dissociation of product ions further to secondary ions was shown. In other words, the fragmentation of protonated oligosaccharide ions in the gas-phase followed both sequential and parallel dissociation reaction pathways.

3.3.2 Dissociation Kinetics and Arrhenius Parameters

The temperature dependent first-order rate constant (k) for dissociation of glycosidic bonds was determined from the change in the natural log of the normalized abundance (Ab_{norm}) of the protonated oligosaccharides with reaction time (eq 3.1). Ab_{norm} was calculated using eq 3.2, where Ab_R is the abundance of the reactant ion and ΣAb_p is the sum of the abundance of all product ions, including those produced by the secondary reactions. Ab_{norm} was calculated using only the most abundant isotopic ion for the reactant and product ions. In principle, Ab_{norm} should be calculated using the sum of the abundance of all the isotopes of ions. However, the BIRD spectra exhibit poor signal-to-noise ratio (S/N) for the low abundance isotope peaks in the lower mass region of the mass spectrum and so it is difficult to accurately quantify the signal. Daneshfar et al showed that the rate constants and activation energies should be independent of the method used to calculate the Ab_{norm} [48]

$$\text{Ln } (Ab_{norm}) = -kt \qquad \qquad \text{eq 3. 1}$$

$$Ab_{norm} = Ab_R /(Ab_R + \Sigma Ab_P) \qquad \qquad \text{eq 3.2}$$

Illustrative kinetic plots obtained for Mal_5, Mal_6, Mal_7, Mal_8 and Cel_6, at different temperatures are shown in **Fig. 3.4.** The plots are linear and exhibit near
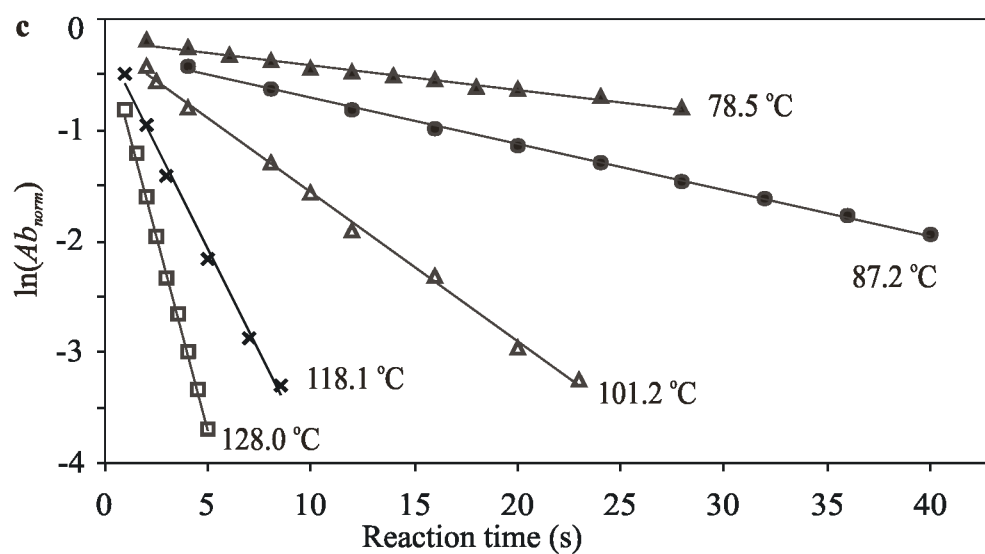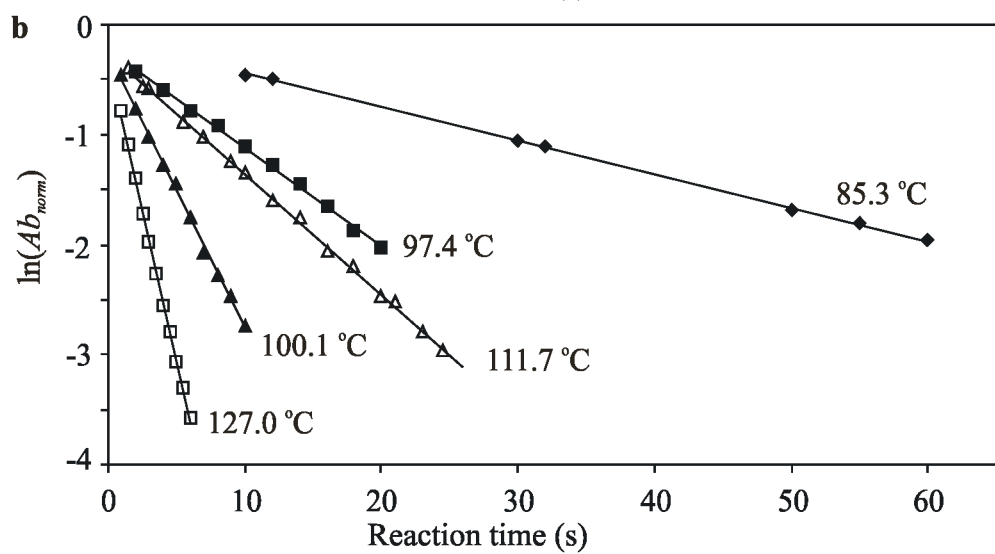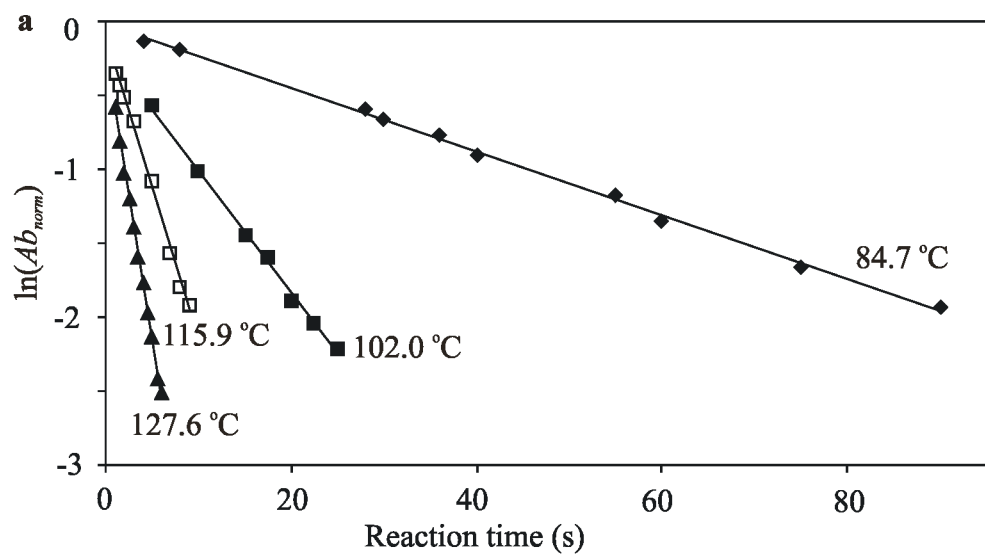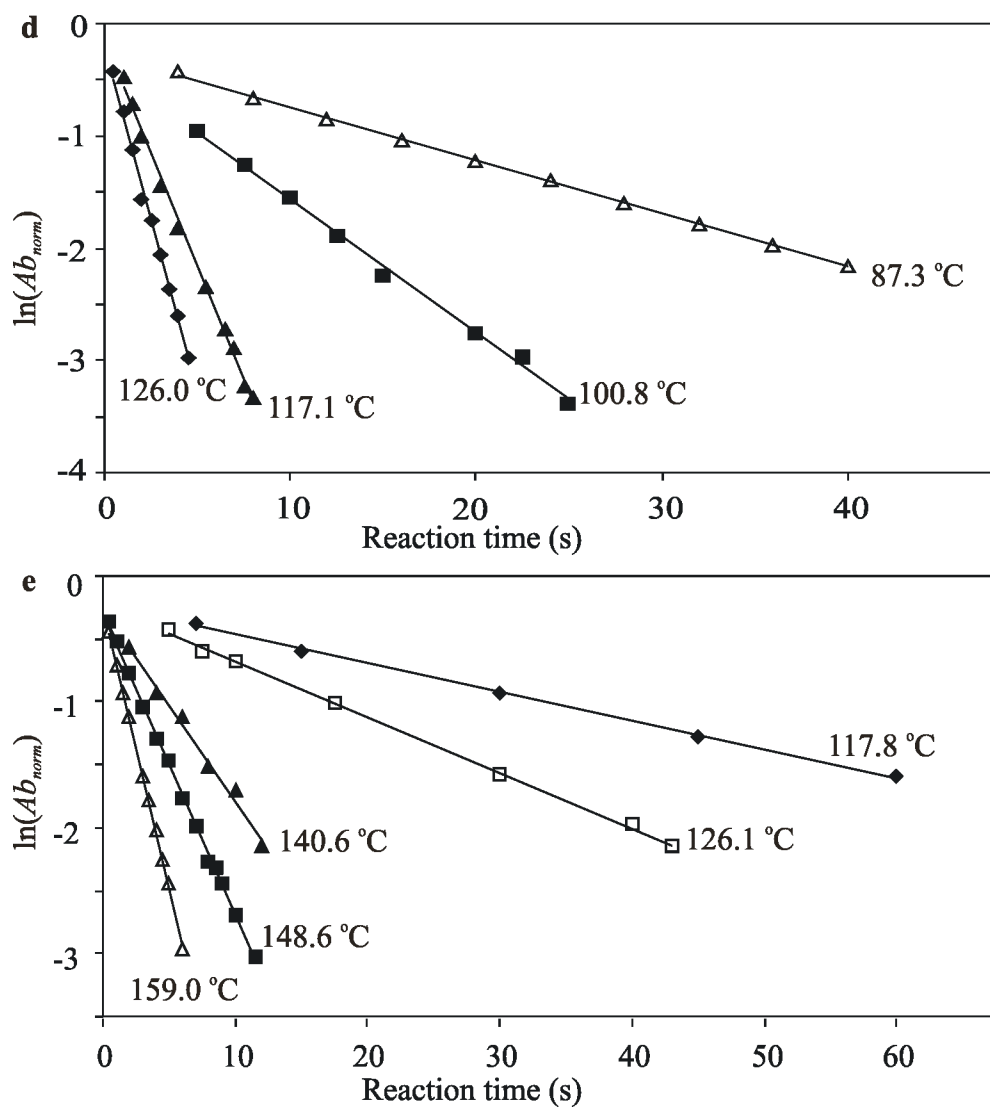
Figure 3.4 Plots of the natural logarithm of the normalized abundance (Ab_{norm}) of the a) Mal_5 b) Mal_6 c) Mal_7 d) Mal_8, e) Cel_6 ions versus reaction time measured by BIRD at the temperatures indicated. Ab_{norm} was calculated using eq 3.2.

zero intercepts, as expected for simple first-order reactions. The first order rate constant (k) was determined from the slope of a linear least squares fit of the kinetic data obtained at each temperature investigated. The linear fits have correlation coefficients $R^2 \sim 0.99$. The linear plots imply that there is single structure of the oligosaccharides. From the temperature dependence of the unimolecular dissociation constant, Arrhenius activation parameters were obtained from the slope of the plot of ln k against 1/T. Arrhenius plots constructed from the temperature dependent rate constants for the dissociation of protonated oligosaccharides are shown in **Fig. 3.5**. For all of the ions investigated, linear plots were obtained. The Arrhenius activation energy (E_a) and pre-exponential factor (*A*) were obtained from the slopes and y-intercepts, respectively, of the plots and presented in Table 3.2. The E_a values of the protonated maltooligosaccharides are similar in magnitude: 19.9 (Mal_8), 19.8 (Mal_7), 19.6 (Mal_6), 19.3 (Mal_5) kcal/mol. The activation energy values are very close to each other regardless of their size. This result implies that the molecules are close to the rapid exchange limit (REX). It was previously reported that for peptides the size of the ion, which is the major factor for rapid exchange limit (REX) [46], should be more than 1.6 kDa, for achieving the REX. Achieving the REX limit is important to maintain the Boltzmann distribution of the ions so that the E_a values are equivalent to E_a values that will be obtained using high-pressure methods by direct application of the Arrhenius equation. The Ea value for the beta linked oligosaccharide, Cel_6 is 24.6 kcal/mo,l which $\sim$ 5 kcal/mol higher than the maltooligosaccharides.
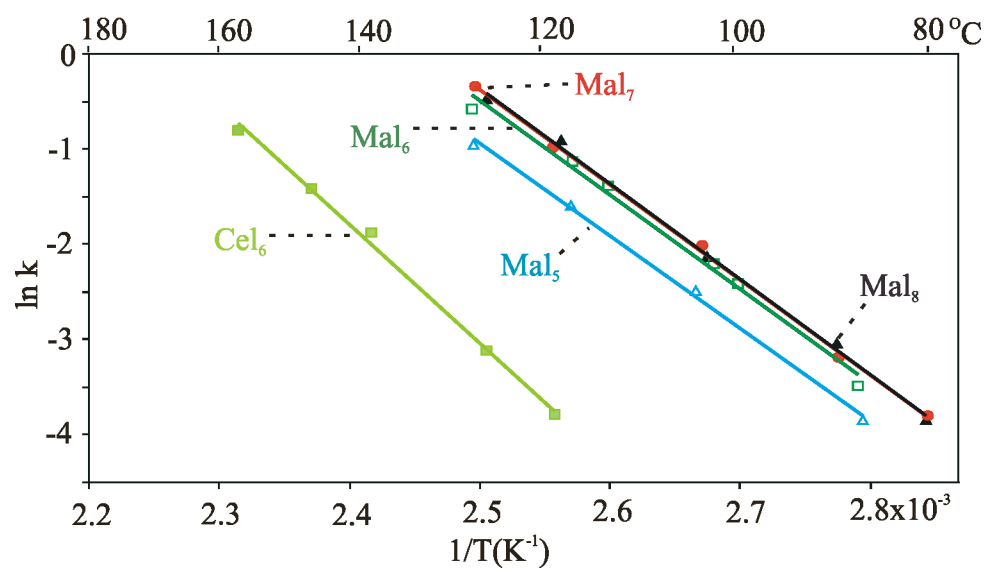
Figure 3.5 Arrhenius plot of dissociation of oligosaccharides

Information about the dynamics and mechanism of the dissociation is contained in the Arrhenius pre-exponential factor. The experimental A values for the maltooligosaccharides are determined to be $\sim 10^{10}$ s^{-1}. Also included in Table 3.1 are values for the entropy of activation ($\Delta S^{\ddagger}$), which corresponds to the difference in entropy between the transition state and the reactant, which can be calculated from the corresponding A-factor using eq 3.3,

$$A = (ekT/h) \exp(\Delta S^{\ddagger}/R) \qquad \text{eq 3. 3}$$

where k is the Boltzmann constant, T is the temperature, h is the Plank's contant, R is the ideal gas contant. A negative $\Delta S^{\ddagger}$ indicates a loss of entropy in the transition state (i.e., a tight transition state); a positive value indicates a gain in entropy (i.e., a loose transition state); and a value close to zero corresponds to a reaction that is entropically neutral. The experimental $\Delta S^{\ddagger}$ values are negative, which implies that the dissociation of the protonated oligosaccharides in average has a tight transition state.

Table 2. Arrhenius activation parameters for the dissociation of protonated

oligosaccharides.

Oligosaccharide	E_a (kcal/mol) [a]	A (s^{-1}) [a]	$\Delta S^{\ddagger}$ (cal mol^{-1}K^{-1})
Mal$_5$	19.3 ± 0.6	$10^{10.1 \pm 0.4}$	-15
Mal$_6$	19.6 ± 1.0	$10^{10.5 \pm 0.6}$	-13
Mal$_7$	19.8 ± 0.4	$10^{10.7 \pm 0.3}$	-13
Mal$_8$	19.9 ± 0.5	$10^{10.7 \pm 0.3}$	-12
Cel$_6$	24.6 ± 0.9	$10^{12.1 \pm 0.5}$	-6

a. Errors correspond to one standard deviation.

3.3.4 Effect of alpha and beta linkage on gas-dissociation

The dissociation of the protonated maltohexaose and cellohexaose, which are composed of six D-glucopyranose units with alpha and beta $1 \rightarrow 4$ linked glycosidic bonds, respectively, were studied. **Fig**. **3.6a** and **b** are representative BIRD data for the two oligosaccharides acquired at 126 ^{o}C and 5 s reaction. The product ions are similar. However, the relative intensities of the peaks are significantly different. For example, the precursor ion is a base peak in the Cel_6 spectrum (see **Fig**. **3.6**). In the case of Mal_6, the precursor ion peak is very weak although the reaction conditions are similar. This implies that dissociation of Mal_6 is faster than Cel_6 at similar fragmentation condition. **Fig. 3.6c** presents the kinetics of their dissociation at 126 ^{o}C. At this temperature, the rate of dissociation of maltohexaose is $\sim$ 3 times that of cellohexaose. Thus, the dissociation of glycosidic bonds of alpha linked glucose units is faster than the beta linked. Suzuki et al, reached at a similar conclusion from investigation of the CID of the sodium adducts of alpha and beta linked disaccharides of glucose [30].

To explain the differences in the rate of fragmentation of alpha and beta-linked protonated oligosaccharides, a detailed analysis of the mechanisms through which the product ions (B and Y) form is necessary. In Scheme 3.5, the formation of Y_5 and B_5 ions from the protonated maltohexaose is shown. The transfer of a proton from the hydroxyl group on C_2 to the glycosidic bond is required for the loss of the neutral monosaccharide (epoxide) from the nonreducing end. **Fig**. **3.7** shows the geometries of alpha and beta-linked disaccharides of D-glucopyranose
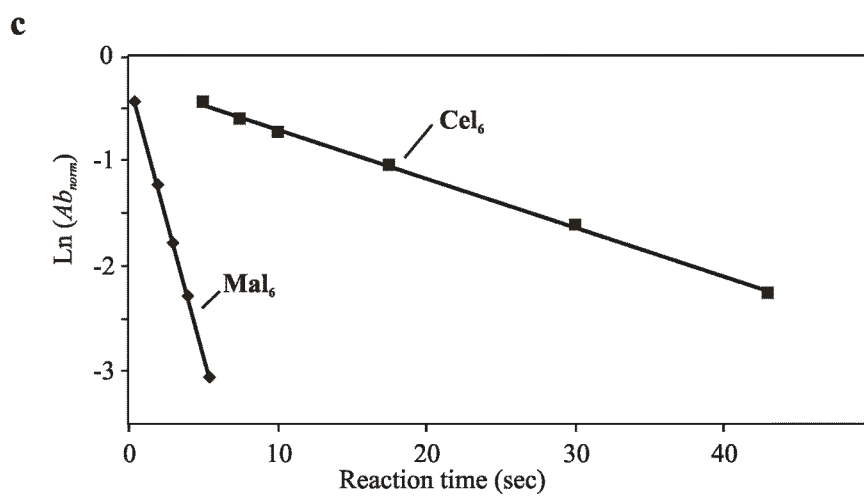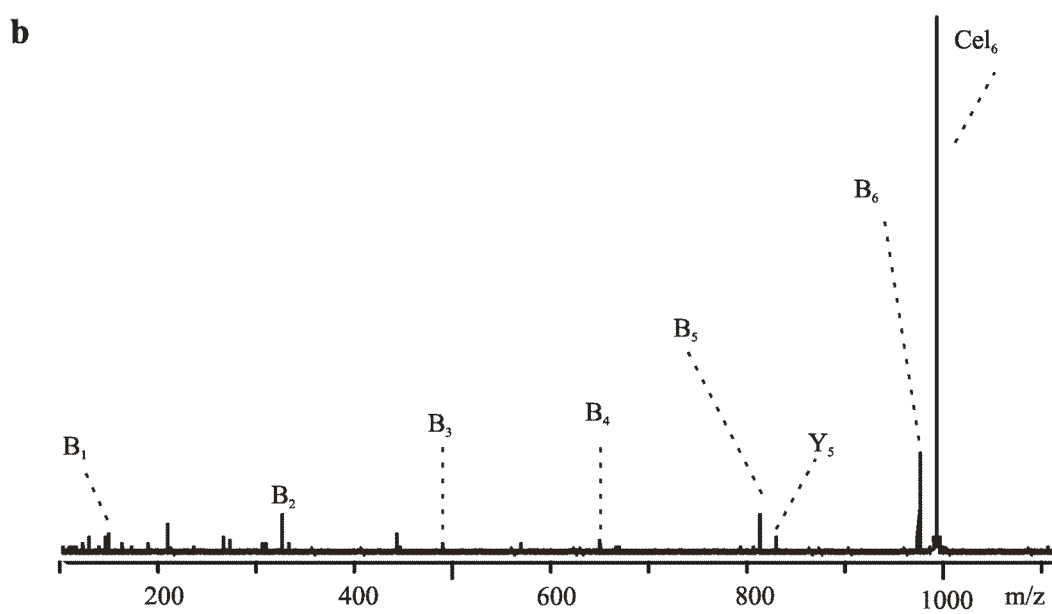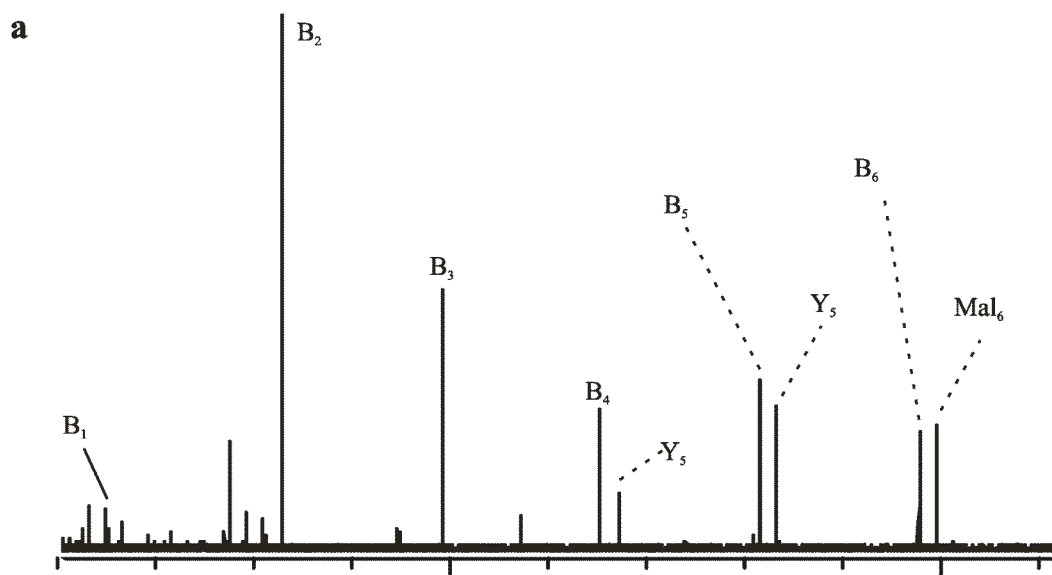
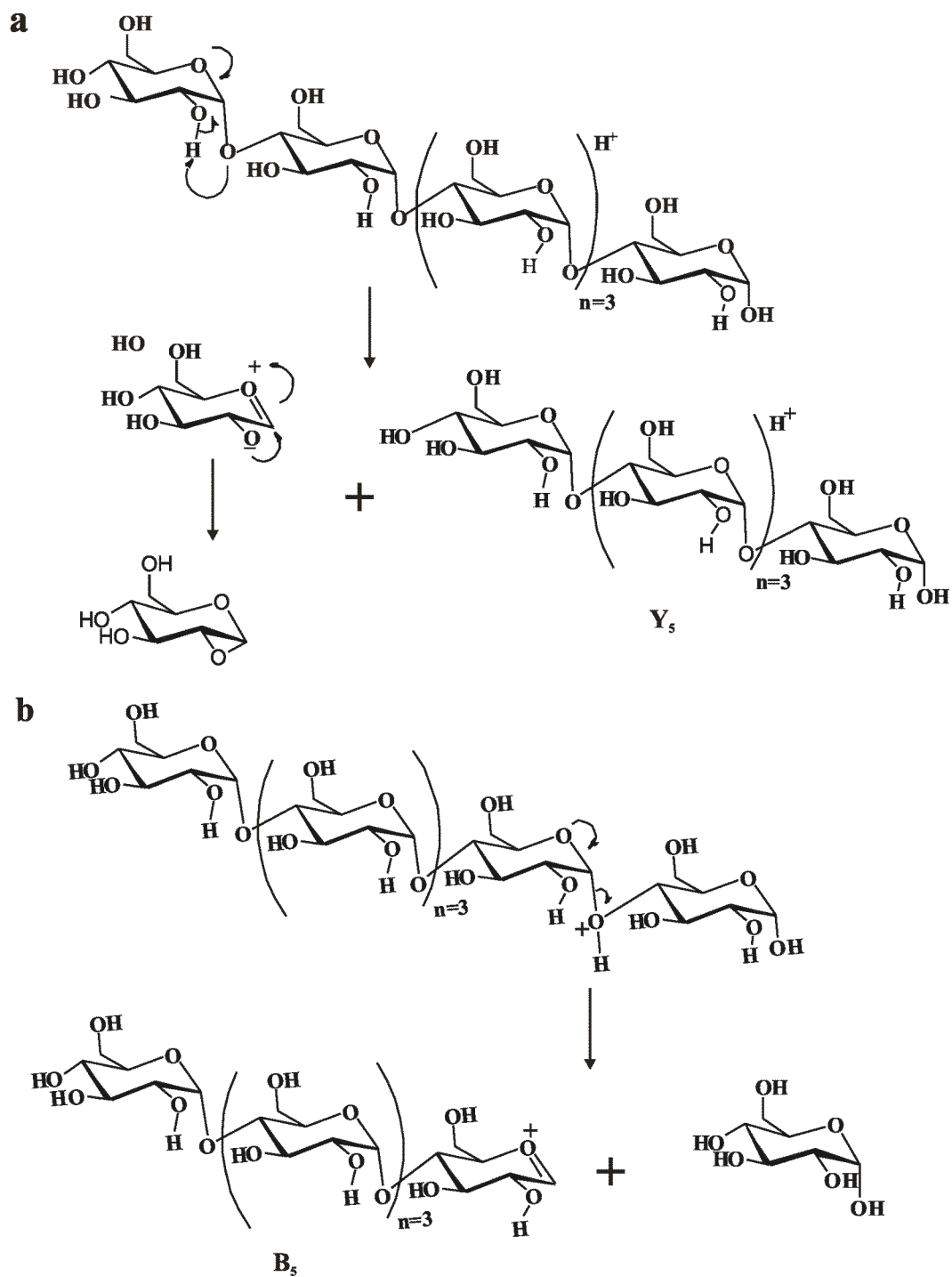**Figure** 3.6 BIRD spectrum of cellohexaose (a) and maltohexaose (b) using a temperature at 126 $^{\circ}$C and reaction time delay of 5 s. c) Dissociation kinetics of Mal$_6$ and Cel$_6$ at 126 $^{\circ}$C.

Scheme 3.5 Formation of a) Y_5 and b) B_5 ions from the protonated Mal_6 ion.

optimized at the restricted Hartree-Fock (RHF)/AM1 level using Gaussian 03. The distance between the proton on the C_2 hydroxyl and the glycosidic bond oxygen is 3.39 A for α-D-Glc$_p$-(1→4)- α-D-Glc$_p$. However, in the case of beta-linkage, the distance between the proton and the oxygen is 3.60 A, which is longer than the alpha linkage. It is believed that the proton transfer will be faster in alpha-linked oligosaccharides than in beta-linked. This implies that the formation of Y_5 from Mal$_6$ is easier than Cel$_6$. Similar, arguments can be made about the other Y ions. However, the B ion formation does not involve the proton transfer (see Scheme 3.5b). Anomeric effect may be of taken into consideration to explain the B ion formation [30]; however, detailed theoretical study is required to figure out the structures of the transition states for the B ion formation.
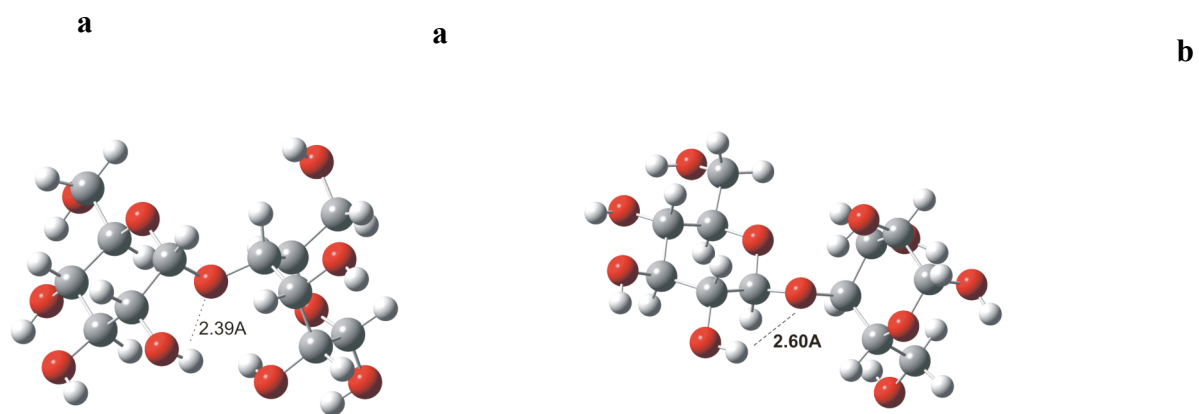
Figure 3.7 Optimized geometries of alpha (a) and beta (b) linked disaccharides of

D-glucopyranose

3.4 Conclusions

The thermal decomposition of a series of protonated oligosaccharides have been studied using the BIRD technique. It is confirmed that many of the glycosidic bonds of oligosaccharides are liable to dissociation. Dissociation of product ions further to secondary ions was shown. It is established that the fragmentation of protonated oligosaccharide ions in the gas-phase followed both sequential and competitive dissociation pathways. The present study reports the first Arrhenius parameters for protonated oligosaccharide dissociation. The activation energy values of malto-oligosaccharides, which are alpha linked D-glucopyranose units, and cellohexaose are ~20 and 25 kcal/mol. The respective pre-exponential factors are ~10^{10} s^{-1} and 10^{12} s^{-1}. Although it requires detailed theoretical study to identify the transition state, it is proposed that Y ion formation involves a proton transfer from the C_2 hydroxyl group to the glycosidic bond.

.

3.5 Literature Cited

1. Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 151-153.

2. Kebarle, P.; Tang, L. *Anal. Chem.* **1993**, *65,* A972-A986.

3. Kebarle, P.; Verkerk, U. H. *Mass Spectrom. Rev.* **2009**, *28,* 898-917.

4. Aebersold, R.; Mann, M. *Nature,* **2003**, *422,* 198-207.

5. El-Aneed, A.; Cohen, A.; Banoub, *J. Spectroscopy Rev.,* **2009**, *44,* 210-230.

6. Becker J. S.; Jakubowski, N. *Chem. Soc. Rev.* **2009**, *38*, 1969-1983.

7. Kirpekar, F.; Nordhoff, E.; Larsen, L. K.; Kristiansen, K.; Roepstorff, P.; Hillenkamp, F. *Nucleic Acids Res.* **1998**, *26,* 2554-2559.

8. Beck, J. L.; Colgrave, M. L.; Ralph, S. F.; Sheil, M. M. *Mass Spectrom. Rev.* **2001**, *20,* 61-87.

9. Zala, J. *Mass Spectrom. Rev* **2004**, *23,* 161-227.

10. Harvey D. J. Mass *Spectrom. Rev.* **1999**, *18,* 349-450.

11. Dell, A. Morris, H. R. *Science* **2001**, *291*, 2351-2356.

12. Park, Y. M.; Lebrilla, C. B. *Mass Spectrom. Rev.* **2005**, *24*, 232-264.

13. Mechref, Y. Novotny, M. V. *Chem. Rev.* **2002**, *102*, 321-369.

14. Harvey, D. J., Mass *Spectrometry Review* **2009,** *28*, 273-361.

15. Bertozzi, C.R.; Kiessling, L. L. *Science* **2001**, *291*, 2357-2364

16. Dove, A. *Nat. Biotechnol.* **2001**, *19*, 913-917.

17. Helenius, A.; Aebi, M. *Science* **2001**, *291*, 2364-2369.

18. Shi, S.D.-H.; Hendrickson, C. L.; Marshall, A. G.; Seigel, M. M.; Kong, F.; Carter, G. T. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1285-1290.

19. Xie, Y. M.; Lebrilla, C. B. *Anal. Chem.* **2003**, *75*, 1590-1598.

20. Budnik, B. A.; Haselmann, K. F.; Elkin, Y. N.; Gorbach, V. I.; Zubarev, R. A. *Anal. Chem.* **2003**, *75,* 5994-6001.

21. Adamson, J. T.; Hakansson, K. *Anal. Chem.* **2007**, *79,* 2901-2910.

22. Leavell, M. D.; Leary, J. A. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 528-536.

23. Smith, G.; Leary, J. A. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 953-957.

24. Gaucher, S. P.; Leary, J. A. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 269-272.

25. Laine, R. A.; Pamidimukkala, K. M.; French, A. D.; Hall, R. W.; Abbas, S. A.; Jain, R. K.; Matta, K. L. *J. Am. Chem. Soc.* **1988**, *110*, 6931-6931.

26. Lain, R. A.; Yoon, E.; Mahier, T. J.; Abbas, S.; Delappe, B.; Jain, R.; Matta, K. *Biol. Mass Spectrom.* **1991**, *20*, 505-514.

27. Yoon, E. S.; Laine, R. A. *Biol. Mass Spectrom.* **1992**, *21*, 479-485.

28. Suzuki, H.; Kameyama, A.; Tachibana, K.; Narimatsu, H.; Fukui K. *Anal. Chem.* **2009**, *81,* 1108-1120.

29. Brull, L. P.; Kovacik, V.; Thomas-Oates, J. E.; Heerma, W.; Haverkamp, J. *Rapid Commun. Mass Spectrom.* **1998**, *12,* 1520-1532.

30. Franz, A. H.; Lebrilla, C. B. *J. Am. Soc. Mass Spectrom.* **2002**, *13,* 325-337.

31. Harvey, D. J.; Mattu, T. S.; Wormald, M. R.; Royle, L.; Dwek, R. A.; Rudd, P. M. *Anal. Chem.* **2002**, *74,* 734-740.

32. Cancilla, M. T.; Wong, A. W.; Voss, L. R.; Lebrilla, C. B. *Anal. Chem*. **1999**, *71,* 3206-3218.

33. Guan S.; Marshall, A. G.; Scheppele, S. E. *Anal. Chem.* **1996**, *68,* 46-71.

34. Domon, B.; Costello, C. E. Glycoconjugate J. **1988**, *5*, 397-409.

35. Liu, Y.; Pan, Y. *Applied Spectroscopy Rev.* **2009**, *44,* 231-244.

36. Feng, X.; Siegel, M. M.; *Anal. Bio. Chem.* **2007**, *389*, 1341-1363.

37. Loo, J. A. *Mass Spectrom. Rev.* **1997**, *16*, 1-23.

38. Price, W. D.; Schnier, P. D.; Williams, E. R. *Anal. Chem.* **1996**, *68*, 859-866.

39. Dunbar, R. C. *Mass Spectrom. Rev.* **2004**, *23,* 127-158.

40. Perrin, J. *Ann. Phys.* **1919**, *11*, 5-108.

41. Stevens, S. M.; Dunbar, R. C.; Price, W. D.; Sena, M.; Watson, C. H.; Nichols, L. S.; Riveros, J. M. Richardson, D. E. Eyler, J. R. *J. Phys. Chem.* **A 2002**, *106*, 9686-9694.

42. Wyttenbach, T.; Bowers, M. T. *Annu. Rev. Phys. Chem.* **2007**, *58,* 511-533.

43. Kitova, E. N.; Kitove, P. I.; Bundle, D. R.; Klassen, J. S. *Glycobiology,* **2001**, *11*, 605-611.

44. Kaufmann, R.; Chaurand, P.; Kirsch, D.; Spengler, B. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1199-208.

45. Deng, L.; Sun, N.; Kitova, E. N.; Klassen, J. S. *Anal. Chem.* **2010**, *82,* 2170-2174.

46. Price, W. D.; Williams, E. R. *J. Phys. Chem.* **A 1997**, *101*, 8844-8852.

47. Klassen, J. S.; Schnier, P. D; Williams, E. R. *J. Am. Chem. Soc. Mass Spectrom.* **1998**, *9,* 1117-1124.

48. Daneshfar, R.; Klassen, J. S. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 55-64.

49. Amster, I. J. *J. Am. Soc. Mass Spectrom.* **1996**, *31*, 1325-1337.

50. Buchanan, M. V.; Hettich, R. L. *Anal. Chem.* **1993**, *65,* 245 A -259 A.

51. Marshall, A. G.; Grosshans, P. B. *Anal. Chem.* **1991**, *63*, 215 A -229 A.

52. Hofmeister, G. E.; Zhou, Z.; Leary, J. A. *J. Am. Chem. Soc.* **1991**, *113*, 5964-5970.

53. Harvey, D. J. Bateman, R. H.; Green, M. R. J. Mass Spectrom. **1997**, *32*, 167-187.

54. Fukui, K.; Kameyama, A.; Mukai, Y.; Takahashi, K.; Ikeda, N.; Akiyama, Y.; Narimatsu, H. *Carbohydr. Res.* **2006**, *341,* 624-633.

55. Yamagaki, T.; Fukui, K.; Tachibana, K. *Anal. Chem.* **2006**, *78*, 1015-1022.

Chapter 4

Conclusions and Future Work

The objectives of this thesis were to describe application of mass spectrometry for characterization of glycoproteins and oligosaccharides. Characterization of glycoproteins involves defining glycosylation sites and the glycan structure on the glycoproteins. Although there are lots effort to obtain glycosylation site and the structure of glycan "in-situ" using MS, it remains challenging. Part of this thesis presented analysis of glycopeptides resulted from trypsin and proteinase K hydrolysis of glycoproteins. The other part of the thesis focuses on mechanistic study of the dissociation process of oligosaccharides. Application of tandem MS techniques for obtaining sequence and other detailed structural information is a challenging task due to the complex structure of oligosaccharides. Application of blackbody infrared radiative dissociation to establish dissociation pathways, kinetics and Arrhenius parameters of the dissociation of protonated oligosaccharides is presented.

In Chapter 2 of the thesis, the application of trypsin and proteinase K to hydrolyze glycoproteins into glycopeptides, which will be subjected to MS and MS/MS analysis to define glycosylation site and sequence of the glycan, is described. Proteinase K hydrolysis of glyocproteins was optimized and compared with the application of trypsin. The resulting glycopeptides of the glycoproteins from bacteria and eukaryotes were analyzed using nano RP-HPLC ESI QToF

mass spectrometer to define the glycosylation site as well as the glycan structures. For successful site-specific analysis of glycoproteins, glycopeptides with short peptide (3-8 residues) are needed. As a result, although trypsin is important enzyme for protein and glycoprotein identification work, proteinase K is superior for site-specific glycoprotein analysis due to its potential to hydrolyze every glycoprotein to short glycopeptides. This has been successfully demonstrated using model glycosylated proteins from bacteria (AcrA) and eukaryotes (chicken egg albumin). For example, MS and MS/MS analysis of proteinase K hydrolysed AcrA showed four glycopeptides. The four glycopeptides are $(HexNAc)_6Hex$ and $DATDH(HexNAc)_5Hex$ attached to two different peptides. The peptides were identified to be DFNRS and DFNNNRS, which carried the expected glycosylation sites of AcrA, Asn^{123} and Asn^{273}, respectively [5]. The oligosacchrides, $(HexNAc)_6Hex$ and $DATDH(HexNAc)_5Hex$ matched to GalNAcα1-4GalNAcα1-4(Glcβ1-3)GalNAcα1-4GalNAcα1-4GalNAcα1-3GlcNAc- and GalNAcα1-4GalNAcα1-4(Glcβ1-3)-GalNAcα1-4GalNAcα1-4GalcNAcα1-3Bac- based on our MS/MS data and the literature [6]. Similarly, digestion of chicken egg albumin with proteinase K enabled us to identify most of the glycans attached to the protein "*in-situ*". $GlcNAc_2Man_5$, $GlcNAc_2Man_6$, $GlcNAc_3Man_5$ and $GlcNAc_4Man_5$ are among the abundant oligosaccharides of ovalbumin on its single glycosyaltion site, Asn^{292}. On the other hand, during the nano RP-HPLC ESI QToF MS and MS/MS analysis of the trypsin digested AcrA, only two glycopeptides were detected. Identity of the glycopeptides were confirmed to be the two oligosaccharides $(HexNAc)_6Hex$ and

DATDH(HexNAc)$_5$Hex) attached to the peptide, DFNR. However, the other expected glycopeptides, (HexNAc)$_6$Hex and DATDH(HexNAc)$_5$Hex) attached to [267]AVFDNNNSTLLPGAFATITSEGFIQK[292], were not detected. Similarly, analysis of trypsin digested ovalbumin failed to detect any glycopeptides, which were expected to have big tryptic glycopeptides (> 4000 Da).

Proteinase K was used to hydrolyze *O*-glycosylated pilins from *Neisseria meningitidis* (NM) and *Pseudomonas aeruginosa* (PA) produced in *E. coli* using glycoengineering. Glycosylation site of NM pilin was directly determined for the first time. It was found that Ser[63] is the glycosylation site of the NM pilin. Similarly, the glycosylation site of PA pilin was confirmed to be [148]Ser. Proteinase K hydrolysis of glycoproteins, especially *O*-glycosylated proteins and glycoproteins of bacteria, will be an important approach to obtain handy glycopeptides for MS and MS/MS for site-specific analysis.

In Chapter 3, the gas-phase dissociation pathways, kinetics and Arrhenius parameters of protonated oligosaccharides during MS/MS analysis are described. Tandem MS approaches involve isolating an ion of interest in the gas-phase in time or space, dissociating it to fragment ions and accurately determining the mass of the product ions. Thus, the success of MS/MS to provide structural information about the precursor ion requires the ability to relate the fragments observed in a mass spectrum back to the original structure of the ion. Suzuki *et al.* established general rules for fragmentation of sodium adducts of several oligosaccharides using computational and experimental data [7]. According to the

study, the dissociation of glycosidic bonds involves transfer of protons from the hydroxyl group on C_2 to the glycosidic bond. However, the dissociation pathways, mechanisms and energetics of protonated oligosaccharides remain poorly understood. Chapter 3 presents detailed study of gas-phase dissociation of protonated oligosaccharides using BIRD FT-ICR MS. Dissociation of maltooligosaccharides, (α-D-Glc$_p$-1$\rightarrow$(4-α-D-Glc$_p$-1)$_n$$\rightarrow$4- D-Glc$_p$, n = 3,4,5,6) and cellohexaose, (β-D-Glc$_p$-1$\rightarrow$(4-β-D-Glc$_p$-1)$_n$$\rightarrow$4- D-Glc$_p$, n = 4) were studied using the BIRD technique. Over the range of temperatures investigated (70 ^{o}C to 160 ^{o}C), the protonated oligosaccharides dissociate via cleavage at the glycosidic linkages forming B and Y ions. Using double resonance experiments, an oligosaccharide undergoes sequential and parallel fragmentation reactions. In other words, the fragmentation of protonated oligosaccharide ions in gas-phase follows both sequential and parallel dissociation pathways. Furthermore, dissociation of product ions further to secondary ions was confirmed. Arrhenius activation parameters, E_a and A for protonated alpha-and beta-linked oligosaccharides are reported in the thesis. It was found that in the case of alpha-linked protonated oligosaccharides, the E_a values are ~20 kcal/mol. The A values of the maltooligosaccharides are ~10^{10} s^{-1}. E_a and A values for protonated beta 1$\rightarrow$4 linked protonated oligosaccharide (cellohexaose) are ~25 kcal/mol and ~10^{12} s^{-1}, respectively. These values imply that the fragmentation of alpha 1$\rightarrow$4 linked protonated oligosaccharides of glucose is energetically less stable than the beta-linked. It was found that the rate of dissociation alpha 1$\rightarrow$4 linked glucose is faster than beta linked at the temperatures studied. To our knowledge, this is the

first report of dissociation pathways, kinetics and energetics of protonated oligosaccharides. A possible extension of this study is to investigate dissociation of other oligosaccharides such as N-glycans and other biologically interesting oligosaccharides. There are open questions about dissociation of alpha and beta linked oligosaccharides of mannoses, galactoses and other monosaccharides. The kinds of detailed investigations of the dissociation mechanisms will complement application of tandem MS techniques to identification of linkage types and composition of biologically important oligosaccharides.

Literature Cited

1. Jiang, H.; Desaire, H., Desaire, H.; Butnev, V. Y.; Bousfield, G. R., *J. Am. Soc, Mass Spectrom.,* **2004**, *15,* 750-758.

2. Larsen, M. R.; Hojrup, P.; Roepstorff, P. *Mol. Cell Proteomics* **2005**, *4,* 107-119.

3. Hitchen, P. G.; Dell, A. *Microbiology,* **2006**, *152*, 1575-1580.

4. Dell, A.; Morris, H. R. *Science*, **2001**, *291,* 2351-2356.

5. Wacker, M.; Linton, D.; Hitchen, P. G.; Nita-Lazar, M.; Haslam, N. M.; North, S. J.; Panico, M.; Morris, H. R.; Dell, A.; Wren, B. W.; Aebi, M., *Science,* **2002**, *298*,1790-1793.

6. Young, N. M.; Brisson, J. R.; Kelly, J.; Watson, D. C.; Tessier, L.; Lanthier, P. H.; Jarrell, H. C.; Cadotte, N.; St Michael, F.; Szymanski, C. M., *J. Biol. Chem.,* **2002**, *277*, 42530-42539.

7. Suzuki, H.; Kameyama, A.; Tachibana, K.; Narimatsu, H.; Fukui K. *Anal. Chem.* **2009**, *81,* 1108-1120.