APPLICATION OF MASS SPECTROMETRY IN ENDOPOLYGALACTURONASES GLYCOSYLATION SITE MAPPING AND CARBOHYDRATE STRUCTURE ELUCIDATION

by

MIN XIE

(Under the Direction of RON ORLANDO)

ABSTRACT

Bioanalytical mass spectrometry is a rapidly developing field due to its wide range of applications. Especially MALDI-MS and ESI-MS have been the essential analytical tools in the areas of proteomics and glycomics. In order to understand the protein functionalities, the extensive post-translational modification events such as glycosylation have to be characterized. The complete characterization of glycosylation is composed of intact protein analysis, glycosylation site mapping and the carbohydrate structure/composition analysis. In each and every step of analysis, mass spectrometry methods can provide detailed information needed.

The focus of the research presented in this dissertation is the utilization of sensitive mass spectrometric techniques, mainly MALDI and ESI MS, to characterize the protein glycosylation, the most complex post-translational modification of proteins. The characterization of *Botrytis cinerea* endopolygalacturonases, BcPG3 and BcPG6, and *Aspergillus niger* endopolygalacturonase PGC were described. The glycoproteins are first analyzed by MALDI to obtain their molecular masses, followed by trypsin digestion and the resulting peptides are

analyzed by LC-ESI MS. The location of glycosylation is identified by using a stepped orifice voltage technique and the structures of the carbohydrates were obtained from ESI MS/MS data. For BcPG3 and BCPG6, multiple *N*-linked glycosylation sites are identified which are occupied by high mannose *N*-linked glycans, while one *N*-linked glycosylation site is identified from *A. niger* PGC and it is also occupied by high mannose *N*-linked glycan. In addition, *O*-linked glycosylations are identified from both BcPG3 and PGC using a convention ESI MS based on the heterogeneity of glycoforms. The location of *O*-linked glycosylation sites on PGC is further identified from electron-capture dissociation (ECD) MS/MS, which confirms that a single mannose is attached to the residue Thr5.

INDEX WORDS: MASS SPECTROMETRY; MATRIX-ASSISTED LASER DESORPTION/IONIZATION, MALDI; ELECTROSPRAY IONIZATION, ESI; TANDEM MASS SPECTROMETRY, MS/MS; LIQUID CHROMATOGRAPHY, LC; LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY,LC/MS; QTOF2; STEPPED ORIFICE VOLTAGE TECHNIQUE; ELECTRON CAPTURE DISSOCIATION, ECD; PROTEIN GLYCOSYLATION; CARBOHYDRATE STRUCTURE; *N*-LINKED GLYCOSYLATION; *O*-LINKED GLYCOSYLATION; ENDOPOLYGALACTURONASE, EPG;

APPLICATION OF MASS SPECTROMETRY IN ENDOPOLYGALACTURONASES GLYCOSYLATION SITE MAPPING AND CARBOHYDRATE STRUCTURE ELUCIDATION

by

MIN XIE

BS, SHANDONG NORMAL UNIVERSITY, CHINA, 1994

MS, INSTITUTE OF CHEMISTRY, CHINESE ACADEMY OF SCIENCES, CHINA, 1997

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2004

APPLICATION OF MASS SPECTROMETRY IN ENDOPOLYGALACTURONASES GLYCOSYLATION SITE MAPPING AND CARBOHYDRATE STRUCTURE ELUCIDATION

by

MIN XIE

Major Professor: Ron Orlando

Committee: Jonathan Amster
Carl Bergmann
Lucia Babcock

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2004

DEDICATION

This dissertation is gratefully dedicated to:

My husband,
Arthur An,

without whose love, unfailing support and tenacity,
not a word would have been written;

My parents,
Xiangqing Xie and Xiangyun Wang,

for their endless encouragement and patience,
for teaching me about research with curiosity and constancy

.

ACKNOWLEDGEMENTS

Many people have inspired, guided, helped, and laughed with (and sometimes at) me during the five years I spent at the University of Georgia, and I would like to thank them all for a great graduate school experience.

First and foremost, I am indebted to my major advisor, Dr. Ron Orlando, for his guidance and patience, for his helpful criticisms, as well as his support and encouragement when they were most needed. Dr. Orlando helped me appreciate how to do research, which I will gladly carry forward with me in my future work.

In the years that this dissertation has been in progress, it benefited tremendously from interim reviews from my advisors. I would like to thank especially Dr. Carl Bergmann, for his direction and assistance in refining the scope of my research, and for his great suggestions and careful reviews of this dissertation. He and Dr. Orlando stood by me each time things were rough and their magic words helped me to keep persistence. In many ways, this dissertation could not be completed without their endeavor.

In addition, I would like to thank Drs. Jon Amster and Lucia Babcock for being on my advisory committee, also for the effort they put into reading this dissertation under significant time constraints. Their thoughtful criticisms and suggestions ultimately made this a sounder and more coherent document.

The research I carried out would not have been possible without the help from Dr. Kumar Kolli. I would like to thank him for the often thankless job of managing the day-to-day details of the instruments. He is the one taught me how to deal with the instrument step by step,

and he is one who I was kept calling when I had problems with the instruments during the nights and weekends. I am also grateful to my labmates, we shared many ups and downs during the course of the research, and I'm thankful for their emotional support and friendship.

This thank-you would not be complete without mention of my family. Although this dissertation is a milestone, it would mean nothing were it not against the backdrop of enormous support, love and unwavering faith in me throughout the long process of being in graduate school and far away from them. This dissertation owes a great deal to my parents and my brother who have encouraged me, for as long as I remember, to pursue whatever goals I set for my life. For this and much more, I am forever in their debt.

The last thank you is saved for my husband, Arthur, who loves me for who I am. After he moved to Chicago last year, his absence from my side makes me realize more than ever that how much he meant to me. His love and support is unfailing and his patience and understanding is appreciated more than he knows. Arthur, my parents and brother, they are the semantic behind what I do. It is to them that I dedicate this dissertation.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

Chapter 1. Introduction

A mass spectrometer is an analytical instrument that can determine the molecular mass of chemical compounds by separating molecular ions according to their mass-to-charge ratio (m/z). MS holds a special place in modern analytical techniques because it measures an intrinsic property of a molecule, its mass, with very high sensitivity and therefore can be used in a wide range of applications. Since the early 1900s, mass spectrometry (MS) has played a prominent role in the fields of physics, chemistry, geology, nuclear science, material science, the petroleum industry, forensic science, and environmental science. However, in the past, the contribution of MS to biological sciences remained limited mainly because of restrictions on the size of the molecules that could be assayed. This situation did not change until the late 1980s, when breakthroughs in two ionization methods, electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI), were achieved. These ionization techniques can form ions from polar substances in liquids or solids, which helps avoid chemical derivatization and/or destructive sample vaporization. During the decade of the 1990s, new types of mass spectrometers were designed to accommodate the inherent sensitivities of these high efficiency ion sources. As a result, the measurement of biomolecular masses with high resolution and accuracy can now be performed routinely. These advances in the instrumentation and techniques have revolutionized protein chemistry by fundamentally changing the analysis of proteins. This trend is especially obvious in the field of proteomics, where both qualitative and quantitative analyses of proteins rely heavily on MS. Proteins are often present as part of a complex mixture in immunoprecipitates or cell lysates, and the components of interest are often in low abundance relative to the "housekeeping" proteins present in the same sample. The coupling of chromatographic and electrophoretic techniques with mass spectrometry has produced powerful

tools for the simultaneous separation and detection of complex mixtures in the pico- and femtomole ranges. As a result, the applications of MS in biochemistry have grown substantially.

A survey of the current literature on mass spectrometry provides ample testament to its practical and potential usefulness. Mass spectrometry is applicable to all molecules, and is capable of determining the structures of most classes of unknown compounds, volatile or nonvolatile, polar or nonpolar. The current applications of MS in protein analysis are focused on several major areas. First, it is the method of choice for protein identification by providing mechanisms for peptide fingerprint mapping and peptide sequencing, the data from which can then be integrated with protein or gene databases. Recent advances in the utilization of multidimensional liquid chromatography combined with tandem mass spectrometry has provided new avenues for large scale protein sequencing[1-3], and proteomic analysis has become a relatively simple task. Secondly, MS is the preferred technique for the study of protein post-translational modifications. Protein identification alone provides only limited insight into protein function. Various covalent modifications often occur, either during or after assembly of the polypeptide chain, termed co- and post-translational modifications, respectively. By introducing other chemical groups into the makeup of a protein, these modifications may alter physical and chemical properties, folding, conformation distribution, stability, activity, and consequently function of the proteins[4;5]. Identifying the type and location of protein modification is the first step in understanding their regulatory potential. Obtaining this type of structural information is a much more complex task than protein sequencing, and normally requires multi- dimensional analysis. The study of protein conformational changes by MS [6;7] is the third area that has developed rapidly. MS is increasingly used to investigate the folding-unfolding dynamics of proteins and the non-covalent interactions of proteins with other biomolecules, a field

traditionally studied by infrared (IR),[8-10] ultraviolet (UV) absorption spectroscopy,[9;11;12] X-ray crystallography[9;10;13] and nuclear magnetic resonance (NMR) spectroscopy in conjunction with hydrogen/deuterium exchange.[9;10;13-16] MS offers several distinct advantages in terms of sensitivity, preservation of protein stability and extended mass range, thus providing information that is complementary to that obtained by other techniques.[17;18] MS has also been widely used to characterize recombinant proteins for the purpose of quality control in industries. Due to its sensitivity and rapid analysis, MS has expanded its role in clinical, immunological and cancer-related studies, as well as in the profiling of bacteria and viruses.

The overall utility of various MS techniques can be demonstrated in the structural study of glycoproteins. Pathogenic fungi such as *Aspergillus niger,* a post-harvest pathogen, and *Botrytis cinerea* secrete a number of enzymes that digest or lead to the digestion of pectic polysaccharides present in the plant cell wall. Endopolygalacturonases (EPGs) are the first cell wall degrading enzymes detected in the infection process. EPGs hydrolyze the de-methylated backbone of the pectin consisting of α 1, 4-galacturonic acid units.[19] The cell wall fragments released by this hydrolysis reaction may have dual effects. On one hand, they may provide nutrients for fungal growth. On the other hand, they may activate an array of plant defense responses, among which is the production of polygalacturonase inhibiting proteins (PGIPs). The function of any single endoPG during the infection process is not yet clear. Biochemical studies revealed that differential expression of EPG genes is dependent on the type of host tissue, the stage of infection and the temperature. Further, the EPGs may have different substrate specificities.[20-22] All this data indicates these enzymes are complementary to each other and have a concerted action on pectin. EPGs are expressed by numerous fungi and bacteria. The host plant itself may also secrete EPGs, especially during fruit ripening. Because EPGs are

capable of the rapid break down of plant cell walls, these enzymes along with other pectin-degrading enzymes play a major role in food industries such as the fruit-juice and wine industry.

Mass spectrometry has become an important tool in the analysis of EPG glycosylation sites.[15;23-25] Information on the glycosylation of EPGs obtained by MS has been used to generate mutant EPGs with one or more glycosylation sites missing. Various mutants have been analyzed in order to evaluate the role of glycosylation in virulence. The glycosylation of several over-expressed EPGs including PGC from *A. Niger,* BcPG3 and BcPG6 from *Botrytis cinerea* was studied in this dissertation by various mass spectrometric methods. The complete characterization of glycoprotein and glycopeptide heterogeneity with respect to *N*- and *O*-linked glycosylation site mapping and carbohydrate structures is presented.

Reference List

1. McCormack A L, Schieltz D M, Goode B, Yang S, Barnes G, Drubin D & Yates J R III, *Anal. Chem.* **1997, *69***, 767-76.

2. Link AJ, Eng J K, Schieltz D M, Carmack E, Mize G J, Morris D R, Garvik B M & Yates J R III *Nat. Biotechnol.* **1999,** *17*, 676-82.

3. Washburn M P, Wolters D & Yates J R III *Nat .Biotechnol.* **2001,** *19*, 242-47.

4. Goochee CF, Gramer MJ, Anderson DC, Bahr JB, Rasmussen JR, *Bio. Technology* **1991,** *9*, 1347-55.

5. Dwek RA, *Chem.Rev* **1996,** *96*, 683-720.

6. Smith DL, Zhang Z, *Mass Spectrom.Review* **1994,** *13*, 411-29.

7. Smith DL, Deng Y, Zhang Z, *J.Mass Spectrom.* **1997,** *32*, 132-46.

8. Osbome HB and Nabedryk-Viala E, *Meth.Enzymol* **1982,** *88*, 676-82.

9. Creighton TE, *Protein structures and molecular properties,* Freeman: New York, 1984.

10. Creighton TE, *Biochemistry J.* **1990,** *270*, 1-16.

11. Englander JJ, Calboun DB, Englander SW, *Anal.Biochem.* **1979,** *92*, 517-24.

12. Thomson JA.,Shirly BA, Grimsley GR, Pace CN, *J.Biol.Chem.* **1988,** *294*, 11614-20.

13. Woods RJ, *Current Opinion in Structural Biology* **1995,** *5*, 591-98.

14. Bai Y., Sosnick T.R., Mayne L., Englander S.W. , *Science* **1990,** *249*, 755.

15. Green ED, Adelt G, Baenziger JU, Wilson S and Van Halbeek H, *J.Biol.Chem.* **1988,** *263*, 18253-68.

16. Koerner TA, Prestegard JH. Yu RK, *Methods Enzymol* **1987,** *138*, 38-59.

17. Katta V. and Chait B.T., *Rapid Commun.Mass Spectrom.* **1991,** *5*, 214-17.

18. Katta V. and Chait B.T., *J.Am.Chem.Soc.* **1993,** *115*, 6317-21.

19. Rombouts FM, *Microbial Enzymes and Bioconversions, Economic Microbiology*, Rose, A. H., Ed.; Academic Press: New York USA, 1980.

20. Esquerre-Tugaye M-T, Boudart G., Duma B, *Plant Physiol.Biochem.* **2000,** *38*, 157-63.

21. Herron SR, Benen J., Scavetta R D, Visser J, and Jurnak F., *PNAS* **2000,** *97*, 8762-69.

22. de Vries R P, Kester HCM, Poulsen CH, Benen JAE, Visser J, *Carbohydrate Research* **2000,** *327* , 401-10.

23. Korner R, Limberg G., Dalgaard M., J. *Journal of Mass Spectrometry* **1998,** *33*, 836-42.

24. Yang Y, Bergmann C, Benen J, Orlando R, *Rapid Commun.Mass Spectrom* **1997,** *11*, 1257-62.

25. Shimizu T, Miyairi K, and Okuno T, *Eur.J.Biochem.* **2000,** *267*, 2380-88.

Chapter 2. Literature Review

2.1. Bioanalytical Mass Spectrometry

The analysis and characterization of biomolecules presents numerous challenges to the analytical chemist. It requires a combination of instrumental attributes such as speed, sensitivity, wide linear dynamic range, low detection limits, as well as the capability of coupling to separation instrumentation for on-line analysis. Sensitive and rapid analytical MS techniques provide unique tools for situations where limited sample quantity does not permit detailed structural characterization by other techniques.

The first essential step in MS analysis is to convert the analyte molecules into a gas phase ionic species. The choice of ionization methods of a compound is contingent on the nature of the sample under investigation. For biomolecules, the most widely applied ionization techniques are matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI). Next, the gas phase ions are separated in a mass analyzer according to their m/z ratio, detected by a suitable detector, and displayed in the form of a mass spectrum.

A large number of biological molecules are not ionizable by traditional ionization methods such as electron ionization (EI) and chemical ionization (CI). Both EI and CI are limited by their inability to handle nonvolatile compounds. Most biological molecules are polar compounds, nonvolatile and thermally unstable. Thus, these molecules cannot be easily vaporized without extensive decomposition. The major breakthroughs in the analysis of biopolymers came with the introduction of electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) in the late 1980s. Both techniques are sensitive, capable of analyzing molecules with a wide range of masses, and allow observation of the intact biopolymer at 100 kDa or higher. These attributes, along with steady improvement in their mass analyzers, have made MALDI-MS and ESI-MS the methods of choice for biomolecule characterization.

2.1.1. Matrix-Assisted Laser Depsorption/Ionization – Mass Spectrometry (MALDI-MS)

The landmark work in the development of MALDI was achieved in 1988 by two research groups, Karas and Hillenkamp[1] and Tanake and co-workers[2]. In both cases, the key to success was embedding biomolecules in a suitable matrix that strongly absorbs radiation from a UV laser. Their work showed that proteins with molecular masses in excess of 35 kDa could be detected using a time-of-flight mass spectrometer. After that, new laser wavelengths and correspondingly optimized matrices were introduced. Today, proteins with molecular masses higher than 100 kDa are being routinely analyzed with sub-picomole sensitivity.[3]

In practice, the analyte molecules are first mixed with an excess of matrix molecules (~ 1:1000-1:10,000 analyte:matrix), which are then dried to form an analyte-matrix crystal. The crystal is then irradiated by intense pulses of a laser (typical 1-20 ns) striking the surface at an incident angle of between 37 and 75 degrees. Ultraviolet (UV) lasers such as nitrogen lasers (λ = 337 nm) or Nd:YAG lasers (λ = 266 or 355 nm) are the most commonly used because of their ease of operation and low price. The matrix is an indispensable element of the MALDI process. The matrix molecules must exhibit strong absorption at the laser wavelength used, which permits a large amount of energy to be absorbed by the matrix and subsequently transferred to the analyte in a controllable manner. In addition, by mixing analyte with a large excess of matrix, any intramolecular interactions between the analyte molecules are minimized. The matrix molecules are excited following the absorption of energy. The absorbed energy initially causes heating of the crystal, and then induces the expansion of the matrix into the gas phase, containing the intact analyte in the expanding matrix plume. Ionization of the analyte can occur at any time during this process (Figure 2.1). MALDI produces predominately singly charged molecular ions, although larger proteins can produce multiply charged ions as well. This feature makes the

analysis of mixtures much more straightforward. In addition, Na^+ and K^+ adducts are a common feature of MALDI spectra.

The MALDI ionization mechanisms are still not fully understood [4]. The suggested ionization pathways include gas phase photoionization, excited-state proton transfer, ion-molecular reactions, desorption of preferred ions and cluster ionization[5-7]. The "cluster ionization model" explains the MALDI phenomenon in a straightforward and simple manner and fits most biomolecules containing charged functional groups. In this model, the initial species formed as a result of laser desorption are tiny clusters consisting of matrix, analyte and the ionic species (such as Na^+ and K^+) embedded in the matrix crystal. These clusters are excited as a result of the laser energy absorbed by the matrix and contain enough excess energy to evaporate. Then the charged clusters are generated by either charge separation (excess of anions or cations) or photochemical processes.

MALDI Time-of-Flight Mass Spectrometer (MALDI-TOF MS)

Current applications of MALDI overwhelmingly use time-of-flight (TOF) analyzers because the pulsed nature of a laser beam matches well with the pulsed scanning mode of TOFMS. MALDI has also been demonstrated on magnetic-sector [8], ion-trap[9] and Fourier-transform ion-cyclotron resonance (FTICR) mass spectrometers [10].

TOF is one of the simplest types of mass analyzers and can be classified as either a linear or reflector instrument. It separates the ions based on their differences in speed of travel in a field-free region. A linear instrument (Figure 2.2) consists of an ion source where the MALDI process takes place, a flight tube where the ions of different masses are separated from one another, and a detector. After they are generated by the MALDI process, the ions are expelled

out of the ion source under the influence of a strong electric field. Theoretically, all the ions with the same charge have the same final kinetic energy upon entering the flight tube. An ion with mass m and charge of q = *ze* has a kinetic energy of:

$$E_k = qV_s = zeV_s = \frac{1}{2}mv^2$$

The time it needs to fly the distance of d (the length of field free region) is given by:

$$t = d/v$$

Thus,

$$t^2 = \frac{m}{z}\left(\frac{d^2}{2V_s e}\right)$$

This equation shows that m/z can be calculated from a measurement of t^2. It also shows that all other factors being equal, ion travel time is only related to its m/z ratio. Ions with greater m/z will travel slower than the lighter ions and will therefore reach the detector later. Based on the equation, the mass range of a TOF instrument has no upper limit, and samples above 300 kDa have been observed by MALDI-TOF[11;12]. Another advantage of TOF instruments is the high transmission efficiency. All the ions are produced by a short pulse (~ several ns) and temporal separation of them allows all of them to be directed towards the detector together as well defined ion packets. Therefore, all the ions formed are, in principle, analyzed. This is the reason why TOF instruments have such high sensitivity and proteins with 100-200 amol amounts have been detected.[13]

In reality, ions with same m/z may be formed in different regions of the ion source at different time within one laser shot, and the initial velocities imparted to ions are not always the same. Therefore the ions have a broad kinetic energy distribution. Secondly, when the MALDI process occurs, kinetic energy is lost by collisions with the neutral plume, and further mass-dependent energy dispersion results. Thirdly, ions formed in the source are extracted

immediately by a continuously applied voltage. Those with the same m/z may arrive at the detector at slightly different times due to their kinetic energy difference, also resulting in peak broadening. The resulting poor resolution is the major limitation of linear TOF. The level of kinetic energy distribution can be reduced by introducing a time delay between ion formation and extraction, resulting in enhanced resolution [14;15]. The ions are first allowed to expand into a field-free region in the source and after a certain delay (hundreds of ns) a voltage pulse is applied to extract the ions from the source. This mode of operation is known as delayed extraction and was first introduced by Wiley and McLaren in 1955[16]. During the time delay, the ions drift with the initial velocity gained from the desorption event. For ions of the same m/z, those with greater initial velocity drift faster than ions with less initial velocity; and therefore, move farther away from the desorption point. After the delay time expires, a voltage pulse is applied to accelerate the ions. Those ions that drift slower obtain more energy from the electric field than those that drift faster. Thus, the initially less energetic ions (less initial velocity) receive more kinetic energy to catch up with the initially more energetic ions (greater initial velocity). If the delay time and voltage pulse are chosen correctly, the two ions will arrive at some point in space at exactly the same time. Ideally, this focal point is the detector surface (Figure 2.3). Delayed extraction will also reduce the collisions the ions undergo on their way out of the source, thereby reducing additional peak broadening caused by metastable decomposition.

Another method for the correction of the kinetic energy distribution is to use a reflector[17;18], which is essentially an electrostatic ion mirror. Since the back of the reflector is held at a potential slightly higher than the source accelerating potential, ions traveling in the reflector will stop at some point, turn around and then be reaccelerated back out of the reflector and towards a second detector. When ions with the same mass enter the reflector, the ones with

higher kinetic energy will enter sooner and go deeper into the reflector before they stop and turn around. The less energetic ions enter later, but do not go as deep before they turn around. Because of their longer path-length in the reflector, the more energetic ions will leave later and have their flight time retarded, allowing the other ions to catch up (Figure 2.4). Optics can be adjusted to bring ions of a given m/z ratio to be space-time focused, thus improving the resolution. Ions are usually reflected at a small angle less than 2° with respect to the incoming ions[19;20], which is intended to avoid any further velocity dispersion along the y-axis. If two molecules of the same mass have different axial velocities when they are ionized, the reflector will compensate for an energy difference of up to 20 electron volts. An additional contribution to improving the mass resolution comes from the longer ion path-length in the reflectron TOF instrument. The gain in resolution, however, is at the expense of sensitivity because transmission losses occur when ions pass through the reflector. It is noteworthy that the improvement in resolution by a reflector is most noticeable at masses of ~3000 or below [21]. For larger molecules, where the resolution is no longer sufficient to provide the separation of isotopes, the natural width of the unresolved isotopic envelope has a larger influence on the apparent resolution.

MALDI has historically been considered a "soft" ionization technique that produces almost exclusively intact protonated pseudomolecular ion species. Studies have shown that molecular ions with excessive internal energy obtained during ionization will undergo further fragmentation.[22] Prompt ion fragmentation can occur at the time of ion formation (in-source decay), or after the extraction, during their travel in the flight tube (post-source decay, PSD). Sometimes the metastable ions contain structurally important information and can be analyzed either by using delayed extraction to allow the ions in the source to fragment in a relatively short

period of time (<100 ns) prior to the extraction[23;24], or by a reflector instrument for the metastable ions produced by post-source decay. All PSD fragment ions will have the same velocity as the precursor, thus their time of flight will be the same in the linear mode. But the kinetic energy of a fragment ion is only a fraction of that of the precursor ion. This difference in kinetic energy can be distinguished in the reflector since fragment ions will not penetrate as deep into reflector, and so will have a shorter flight time relative to the precursor ion. Metastable ion decay is very useful for protein and peptide structure studies [25;26] since the fragment ions typically include both neutral molecule losses (such as water, ammonia and portions of the amino acid side chains) and random cleavage at peptide bonds.

2.1.2. Electrospray ionization Mass Spectrometry (ESI-MS)

Electrospray ionization (ESI) is a method to produce gaseous ionized molecules from a liquid solution by creating a fine spray of droplets in the presence of a strong electric field. The phenomenon of generating gas-phase ions of macromolecules by spraying a solution from the tip of an electrically charged capillary was observed in the 1960s by Dole *et al* [27;28]. Building on this early work, the pioneering work of interfacing ESI with a mass spectrometer was done by Fenn in the 1980s[29;30] and many of the current ESI techniques derive from this development.

ESI is one of the atmospheric pressure ionization (API) methods. The samples are ionized at atmospheric pressure and then transferred into the mass spectrometer. In order to couple the atmospheric pressure source compartment with the mass analyzer compartment, which is kept at low pressure (<10^{-8} torr), focusing lenses (nozzle and skimmer) with very small openings between the compartments and several different stages of high-capacity pumps are used (Figure 2.5).

ESI is a combination of two independent processes, electrospray dispersion of charged droplets and gas-phase ion formation. When the analyte solution passes through a capillary that is held at high potential (3-4 KV), a mist of highly charged droplets is generated at the outlet of the capillary under the effect of a high electric field. The potential gradient between the capillary tip and the counter electrode establishes an electric field, causing the droplets to travel towards the mass analyzer. The imposed electric field will also partially penetrate the liquid flowing through the capillary. Depending on the polarity of the electric field, ions of the same charge (e.g. positive ions in the positive mode) in the solution will drift toward the liquid surface, and the oppositely charged ions (negative ions) will redistribute accordingly to form an electric double layer. The resulting accumulated charge at the surface leads to destabilization of the surface because ions at the surface are drawn toward the counter electrode, yet are not able to escape the surface, and thus form a Taylor Cone. It is important to note that the instability of the cone is profoundly influenced by the surface tension of the fluid, with high surface tension requiring a high onset voltage to initiate charged-droplet emission. Higher voltages lead to an increased probability of electric discharge. Corona discharge also increases with decreasing pressure, which is why atmospheric pressure is preferred for ESI. Once the charged droplets are released from the capillary, they are accelerated toward the counter electrode. While traveling towards the counter electrode, the droplet size is reduced due to solvent evaporation. However, the charge remains with the droplet, so the charge density is increased. At a limit, known as the Rayleigh limit, when surface charges overcome the surface tension which holds the droplets together, coulombic repulsion occurs and leads to disintegration of the droplets into smaller ones.

The mechanism of transforming these small droplets into gas-phase ions is still under debate. Several models have been described to explain ion formation. The early model proposed by Dole in 1968 [31], namely the charged residue model, assumed one droplet containing one single ion due to the continuous depletion of droplet size. Further evaporation of such a droplet leads to exposure of a solvent-free gas-phase ion. Another widely accepted model is the ion evaporation model proposed by Iribarne and Thomson in 1976 [32;33]. When a droplet evaporates, ions move towards the droplet surface and the charge sites on the analyte eventually replace the droplet surface charges. With further droplet evaporation, ions are emitted from the surface. While this model explained the formation of multiply-charged ions observed for almost all ESI spectra, it has difficulties explaining certain experimental observations, such as the fact that ionization efficiency is not related to solvent acidity and volatility. Another model, ion emission from a Taylor cone, was then proposed by Siu *et al.* in 1993 [34] to address these issues. One thing that is worth noting is that there is evidence showing that electrochemical reactions do indeed occur and result in the production of ions not originally present in the solution[35]. Most of the time, however, this does not have profound consequences on the analytes being investigated.

The transfer of ions to the gas-phase is not an energetic process, and the evaporation effectively cools the ions. Therefore, ions generated by ESI contain insufficient internal energy to cause fragmentation unless using tandem mass spectrometry on selected precursor ions. On the other hand, conditions can be established in the ESI source to achieve effective collisional activation. For example, applying a suitable potential difference between the nozzle and skimmer allows ions to achieve kinetic energy consistent with the applied potential and therefore undergo activating collisions.

Electrospray Ionization Mass Spectrometry

ESI has been routinely interfaced with quadrupole mass spectrometers, ion-trap mass spectrometers[36], FT-ICR mass spectrometers[37;38] and TOF mass spectrometers [39;40] . Quadrupole mass analyzers have been used in conjunction with electron ionization sources since the 1950s and are the most common mass spectrometers in existence today[41;42]. Quadrupoles are tolerant of relatively poor vacuums (~5 x 10^{-5} torr), which makes them well-suited to electrospray ionization since the ions are produced under atmospheric pressure conditions. In addition, quadrupoles are capable of routinely analyzing up to a m/z of 3000, the common range for multiply charged ions produced by ESI.

A quadrupole mass spectrometer consists of four accurately aligned parallel rods that are arranged symmetrically in a square array. The field within the square is created by supplying a positive direct current (DC) potential and a superimposed radiofrequency (RF) potential to one pair of rods, and the other pair receives a negative DC potential and an RF potential with the same magnitude, but 180° out of phase (Figure 2.6). This arrangement creates an oscillating field that will constrain ion motion. The ions accelerated along the *z*-axis enter the quadrupole and maintain their velocity along this axis. However, they are subjected to acceleration along the *x* and *y* axes under the forces induced by the electric field. For an ion of any mass, the distance it travels along the *x* and *y* axes is a function of the applied DC and RF voltage. By tuning the DC and RF voltage, the ion will not travel too far along either the *x* or *y* axis and will be able to pass the quadrupole without touching the rods (stable trajectories). Otherwise, the ion discharges itself against a rod and is not detected (unstable trajectories)[43].

The mass range and mass resolution of a quadrupole mass spectrometer are both dependant on the length and diameter of the rods, the supply voltage, the RF frequency and the

initial kinetic energy of the ions. In principle, the upper mass limit can be increased by increasing the amplitude of the RF signal, decreasing its frequency, or using short-diameter rods. However, there are practical limitations in optimizing these parameters. As a result, an upper m/z limit of a quadrupole usually is 4000 Da. Typically quadrupole mass spectrometers operate at unit mass resolution (m/Δm), which is sufficient to separate two peaks one mass unit apart. Some efforts have been made to increase resolution by means of increasing the duration of ions in the field. Such efforts include lowering the accelerating potential (to reduce the velocity) and reflecting the ions back and forth within the quadrupole. A resolution of up to 22,000 (full width at half maximum, FWHM) has been achieved by this method.[44] In some situations, it is advantageous to operate the quadrupole in the RF-only mode where the resolution becomes zero and a quadrupole behaves as a wideband mass filter. As a consequence, ions of a wide mass range can be contained within the field. The RF-only quadrupole has been used as an intermediate reaction region in quadrupole-based tandem mass spectrometers[45] such as used a collision-induced dissociation cell, as pre- and post-filters for high-performance quadrupole mass analyzers or as an ion containment region in ESI-TOF instruments and such interfaced established the base for quadrupole-TOF tandem mass spectrometer.

Traditionally, coupling ESI with TOF MS has some obvious difficulties since the ESI source generates ions in a continuous fashion, while TOF operates in a pulsed scan mode. This problem has been successfully solved by using orthogonal injection, where ions from the ESI source enter a storage region in the TOF instrument perpendicular to the spectrometer axis and are injected into the flight path by a pulse applied to the electrodes[39]. Alternatively, such an ion storage region can be replaced by an RF-only quadrupole set in between the ESI source and the TOF instrument [39;46]. Ions enter the quadrupole with velocities acquired in the ESI process, and

lose the energy by collisions with the gas molecules in the quadrupole. As they pass through the quadrupole, the ion motion is constrained by the RF field. Thus, the quadrupole works as a velocity damping device for ions[39;47].

2.2 Applications of Mass Spectrometry in Protein Glycosylation Characterization

Following the establishment of genomics and proteomics, it is becoming increasingly evident that another emerging paradigm in the molecular basis for cellular phenotype is glycomics[48-50] − a field that is dedicated to studying how a cell's array of carbohydrates interact with a wide range of proteins and impinge on cellular phenotype in the context of cell-cell and cell-tissue interactions. This area of research has to deal with an inherent level of complexity not seen in other areas of applied biology, and is the reason that mass spectral analysis of oligosaccharides and glycoproteins has lagged behind that of proteins and oligonucleotides. On the other hand, it is because of their complex and often dynamic structures that oligosaccharides have widely varying biological activities.

For glycoproteins, carbohydrate side chains have profound influence in modulating the physiochemical (e.g. solubility and stability) and biological (immunologic and proteolytic stability) properties of proteins[51]. Carbohydrate side chains serve as recognition markers for cell-cell and cell-molecule interactions, and as receptors for viruses, bacteria and parasites[52]. In addition, sugar units may orient glycoproteins in membranes and help determine the destination of a glycoprotein. Glycosylation can also affect the stability and secondary structure of peptides and proteins.

2.2.1 Structural features of glycans and glycoproteins

Carbohydrates have the potential for great structural diversity. Unlike amino acids that combine in a linear unbranched fashion to form peptides, monosaccharides can be linked through a variety of hydroxyl groups on the sugar to form a range of glycosidic linkages. In addition, there are numerous stereochemical centers, an anomeric center, and extensive branching that may exist in the carbohydrate chain[53]. Furthermore, a glycoprotein exists as a mixture of closely-related compounds consisting of individual glycoforms. This molecular diversity arises from variations in oligosaccharide structure at individual glycosylation sites (micro-heterogeneity) and in some cases variable occupation of potential glycosylation sites (macroheterogeneity). In combination with other potential modifications, the result may be a complex population of individual proteins. Characterization of this diverse molecular population represents a considerable analytical challenge.

The different chains of oligosaccharide side chains are covalently bound to proteins by a series of separated glycosyltransferase reactions. The three most important chemical linkage types are *N*-glycosidic (forming *N*-linked glycoproteins), *O*-glycosidic (forming *O*-linked glycoproteins) and ethanolamine phosphate (occurring in glycosylphosphatidylinositols, GPI-anchors). Figure 2.7 shows *N*- and *O*-glycosidic linkages.

In *N*-linked glycoproteins, the oligosaccharide moiety is always bound to the side chain of an asparagine residue present in a sequon (amino acid sequence) composed of Asn-Xaa-Ser (Thr), where "Xaa" is any amino acid except proline. All *N*-linked oligosaccharides have a common inner-core structure that consists of two *N*-acetylglucosamine (*N*-GlcNAc) and three mannose (Man) residues, with the terminal GlcNAc β-linked to Asn. This can be explained on the basis of the biosynthesis of N-oligosaccharides, in which *N*-linked glycans are derived from a

single lipid-bound saccharide precursor[54]. Branching out from the core structure are oligosaccharide chains of diverse structures, achieved by the attachment of different saccharides, leading to multiple branched structures. They can further be classified into three types: high mannose type, complex type and hybrid type (Figure 2.8). The high mannose type is the simplest structure containing only α-mannosyl residues bound to the core structure. In the complex type, each one of the bisecting cores formed at the central α-mannose residue is elongated with Gal-GlcNAc disaccharide moieties, and additional branching results in several antennas. Furthermore, the side chains usually terminate in galactose, fucose or neuraminic acid residues. Hybrid-type structures possess structural elements of both high mannose and complex types. They contain more than three mannose residues but also GlcNAc and Gal side chains, and the branches often terminate with *N*-acetylneuraminic acid.

In *O*-linked glycoproteins, the glycan moiety is linked to the hydroxyl group of a Ser/Thr residue, or hydroxylysine and hydroxyproline in some cases. Different O-linked glycan structures are found in nature (summarized in Table 1). Unlike *N*-linked glycans, there is no common core structure for O-glycans. Core structures may be short and terminated by sialic acid, or may be elongated[54;55]. Beside the common *O*-linked GalNAc, additional types of O-linked glycans have been described, such as *O*-linked mannosylation in yeasts and *O*-linked arabinosylation in plant glycoproteins, and digalalctosy l ,2,4-diacetamido-2,4,6 trideoxyhexose on pilines of *Megingococcus* species[54;56].

2.2.2 Analysis of Glycoproteins by MALDI-MS and ESI-MS

Among post-translational modifications glycosylation, because of its potential complexity, is the only one that requires structural characterization of the modifying moiety

beyond noting its presence. A general scheme for complete characterization of a glycoprotein involves three tasks:

- Determination of the molecular mass of a glycoprotein;
- Identification of the presence, type and sites of glycosylation (glycosylation site mapping);
- Structural characterization of the carbohydrate side chains (carbohydrate fingerprinting).

Since the carbohydrate moieties are not encoded in DNA sequences, their structural elucidation can only be obtained by direct analytical measurements. In addition, as a result of both the structural complexity mentioned above and the limited sample quantities that are usually encountered in biological systems, the analysis can be challenging, and most of the time, requires more than one dimension of analysis. Combined with other complimentary techniques, MS often provides a total analytical solution for all three tasks. A general strategy used for complete analysis of a glycoprotein is illustrated in Figure 2.9.

2.2.2.1 Determination of the molecular mass of a glycoprotein

The information obtained from the intact glycoprotein provides a global assessment of glycosylation. This was previously achieved primarily by sodium dodecyl sulfate-polyacrylamine gel electrophoresis (SDS-PAGE) and size-exclusion chromatography, which were typically limited by errors greater than 20% for a protein of MW > 20 kDa. For glycoproteins, they usually appear as broad, fuzzy bands as a result of the heterogeneity resulting from the various glycoforms[57]. Currently, this task is routinely achieved by both MALDI-TOF and ESI MS.

The ESI produces multiply charged ions of a protein in the detectable *m/z* range, from which the molecular mass of a glycoprotein can be deduced. ESI-MS is potentially capable of distinguishing each glycoform and the non-glycosylated protein. However, with the increase in the number of glycosylation sites, the number of probable glycoforms also increases. Such a heterogeneous mixture can yield a highly complicated spectrum, making it difficult to retrieve useful information. Therefore, MALDI-TOF is preferred over ESI MS for MW determination.

The principle advantages of using MALDI are the relative simplicity in the interpretation of the resulting spectra, and its robustness with respect to tolerance of buffers and salts in the sample. The quality of the MALDI spectra largely depends on both the matrix and the analyte itself. The most common matrices suitable for glycoprotein analysis are 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid), ferrulic acid (especially useful for glycoproteins > 100 kDa) and trihydroxyacetophenone (THAP). In addition, the buffer or salt content of the sample will affect the matrix-sample crystallization process. Although MALDI is comparatively tolerant to a certain level of salts, chaotropes and detergents, practically, these contaminants should be avoided in order to achieve optimal signals. A great deal of effort has gone into optimizing sample cleanup prior to analysis. For example, a two-layer sample preparation method was demonstrated to improve MALDI performance in analyzing complex proteins containing higher concentrations of contaminants [58;59]. Recent studies showed that reliable MALDI spectra for samples with levels of SDS up to ~1% can be obtained by this method[60;61].

Alternatively, modifying the surface of the MALDI target improves the quality of MALDI spectra significantly. Traditionally, a polymeric membrane, such as porous polyethylene [62], or poly(vinylidine difluoride)[63] or nitrocellulose[64] was deposited on the MALDI target in order to improve the binding of proteins. Modification of the target surface with self-

assembled monolayers also appeared to be successful [65]. Since lectins have binding specificities for various carbohydrates, lectin affinity MALDI probes were introduced by Bundy et al[66;67]. A lectin-derived surface can be achieved by several methods, for example the lectin can be directly immobilized to a gold foil via a self-assembled monolayer. Alternatively, after biotinylation, lectins can be immobilized onto a streptavidin-coated surface, such as glass slides[68]. At this time, lectin-derived probes have been mainly tested with isolated carbohydrates. As the same lectin-carbohydrate binding mechanism is also valid for glycoproteins, it is believed such affinity probes are also suitable for glycoprotein analysis. The downside of affinity probes comes from their high specificity, e.g. the binding between the lectin and carbohydrate requires the presence of a certain carbohydrate structure. Those that do not contain the required structures will not be able to bind.

The MALDI-TOF spectrum of a glycoprotein usually exhibits a broader peak than the deglycosylated protein, and it was suggested that this broadening reflected glycoform heterogeneity[69]. The contribution of metastable fragmentation has been evaluated for human interferon-γ and found not to contribute significantly to the peak width[70]. Apart from obtaining the molecular mass, MALDI-TOF in conjugation with glycosidase treatments is quite useful in determining the total mass of carbohydrate on a glycoprotein[71-73].

2.2.2.2. Glycosylation Site Mapping

Most strategies to locate the glycosylation sites within a glycoprotein rely on the digestion of the glycoprotein with proteases and then analysis of the resulting glycopeptides. This is much easier if prior knowledge of the primary structure of the glycoprotein is available. Otherwise, peptide sequence information is required. Using collisional induced dissociation

(CID) in tandem mass spectrometry to analyze the amino acid sequences of peptides is a valuable tool. Post-source decay (PSD) in MALDI TOF is also used, to a less extent, to sequence peptides.

CID MS/MS can be performed on an array of instruments including sector, triple quadrupole, ion-trap, TOF- TOF and hybrid tandem MS such as the quadrupole-TOF (QTOF). The precursor ion is selected in the first analyzer, and the collisional dissociations occur in a collision cell with a neutral gas (N_2 or Ar) as collision gas. The fragments are analyzed by the second analyzer. The CID characteristics of a peptide are well documented[74;75]. The site of protonation directs fragmentation to yield the amino acid sequence-specific fragment ions. Although protonation preferentially occurs at the N-terminus or the basic residues, the charge can be randomly localized on any one of the amide bonds. Thus, any of the peptide bonds can be cleaved, forming different fragment ion series such as y ions and b ions (Figure 2.10). It should be noted that on a practical level, it is virtually impossible to obtain an entire protein sequence just based on MS techniques. Nevertheless, CID MS/MS and PSD are still the most effective techniques for peptide sequencing.

Potential *N*-linked glycosylation sites can be deduced based on the consensus sequon Asn-Xaa-Ser(Thr) present within the protein sequence, however *O*-linked glycosylation could occur at any Ser or Thr. In addition, the presence of the sequon does not guarantee the attachment of the N-glycan to it. Therefore, it becomes necessary to locate the glycopeptides within the peptide mixtures.

ESI MS has proven to be a powerful tool for characterizing glycopeptides, based on its ability to couple with front-end separation techniques such as high performance liquid chromatography (HPLC), capillary electrophoresis (CE), and affinity chromatography. The most

straightforward approach using LC-ESI MS to analyze glycopeptides is comparative peptide mapping, in which, the comparison of peptides before and after the release of carbohydrate side chains is conducted[73]. However, O-linked glycosylation is not readily detected by this approach because of the lack of an O-glycosidase capable of removing all the various types of O-glycans potentially present in the peptide backbone.

A more specific procedure to selectively detect glycopeptides is to monitor the carbohydrate-specific marker ions generated by either in-source fragmentation[76] or by CID[77]. Detection of oxonium ions combined with precursor ion scan has been widely used to provide detailed insight into site-specific distribution of N-linked glycopeptides. While reports show that O-linked glycopeptides can also be similarly identified, this approach has to date proven more practical for N-glycosylation site mapping.

Alternatively, tandem mass spectrometers using neutral-loss scan provides a method for glycopeptide analysis. Neutral-loss scan consists of selecting a neutral fragment and detecting all the fragmentations leading to the loss of that neutral fragment. It is carried out by scanning precursor ions while monitoring only those fragment ions that result from a constant-neutral loss. Thus, it requires both mass spectrometers are scanned synchronously while maintaining a constant mass difference between the selected masses for all scans. This is normally achieved in a QqQ tandem mass spectrometer, in which, both Q1 and Q3 are scanned over the desired mass range and Q2 serves as a collision cell. It is possible that an unmodified peptide produces CID fragments at the same nominal mass as the carbohydrate oxonium ions, thus the higher the resolution, the more selective the measurements. This can be achieved by using a hybrid RF/DC quadrupole-linear ion trap instrument, which provides the capabilities of a QqQ and the improved sensitivity in product ion scan[78;79]. Neutral-loss scan has been applied to the analysis

of phosphorylation[80] and recently, CID energy-gradient neutral-loss scan was successfully applied for glycopeptide site mapping[81]. In this approach, instead of applying a preset collision energy for all glycopeptides as well as peptides, the collisional energy was programmed to ramp from 0 to 100 ev. The concept behind using such an energy gradient for glycopeptide CID is that one collisional energy value may be optimal for one type of glycan, but not for another. Thus, using a preset collisional energy may result in the failure to detect some glycans.

While precursor ion scan has been used to identify glycopeptides in tryptic digest mixtures ionized by nano ESI [82], it performs best in conjunction with a front-end separation. In most cases, this is accomplished with the peptides being separated by reverse phase HPLC using a C18 column interfaced to the mass spectrometer. Recently, the use of hydrophilic interaction liquid chromatography (HILIC) to capture glycopeptides was investigated[83]. Several glycosylation sites were identified from fetuin based on HILIC-ESI MS alone. Another strategy for characterizing glycopeptides within a mixture is using lectin chromatography to enrich the glycopeptides. However, in order to obtain a complete picture of all glycopeptides present in a complex mixture, serial lectin columns may be required [82;84]. The major drawbacks to this method are that it is time-consuming, requires large quantities of initial samples and it is usually performed off-line. Nevertheless, serial lectin chromatography is still often the method of choice in glycomics studies.

Although claims have been made that precursor ion scan and neutral loss scan are equally effective for *O*-linked glycosylation, the approaches mentioned above have been almost exclusively used for mapping *N*-linked glycosylation,. There are several reasons for this. Unlike *N*-glycans, *O*-glycans usually have short chains consisting of one or several monosaccharide residues. Since the glycosidic linkages are labile and easily cleaved during CID

or in-source fragmentation, selective detection by oxonium ion monitoring has proven difficult. In addition, the lack of available *O*-glycosidase enzymes makes comparative peptide mapping before and after deglycosylation impractical.

One approach to locating *O*-glycosylation sites is converting the thermally labile glycosidic residues to a stable functional group after the chemical release of the *O*-glycans. Alkali-catalyzed ß-elimination is an established method for the release of *O*-glycans from glycopeptides or glycoproteins[85]. The major drawback for this method is that peptide backbones are often cleaved under such conditions. Recently, it has been demonstrated that improvements in the ß-elimination reaction conditions, such as controlling the reaction buffer, time and temperature, can minimize the peptide backbone cleavage. The utilization of ammonia hydroxide[86] or alkylamine (methylamine or ethylamine)[87] has proven to provide a much safer set of conditions for preserving the peptide backbone structure, while still allowing the release of *O*-glycans. Furthermore, using alkylamine introduces methylamine or ethylamine into the elimination products, which are stable during MALDI-PSD and ESI-MS/MS. This method thus has great potential for *O*-glycosylation site mapping. Two similar techniques were developed for the mapping of O-GlcNAc sites, in which O-glycopeptides were ß-eliminated followed by addition of either dithiothreitol[88] or 2-aminoethylcysteine[89]. In the latter case, the labeled peptide is then cleaved by lysine endopeptidase. One potential problem associated with ß-elimination is that phosphorylation, which also occurs at Ser and Thr residues, is also accessible to ß-elimination. Though the changes in molecular mass are unique for phosphorylation and *O*-glycosylation, the resulting spectrum can become complicated to analyze. However, ß-elimination followed by labeling is still a promising approach for the mapping of O-glycosylation sites, and needs more thorough study.

Partial acid hydrolysis followed by MS analysis is another possible choice for the study of *O*-linked glycosylation[90]. Performing hydrolysis in the vapor-phase using pentafluoropropionic acid and hydrochloric acid was investigated on mannosyl- and mucin-type glycopeptides[90;91]. Extensive nonspecific polypeptide backbone cleavage with minimal cleavage of glycosidic bonds was observed. Though the O-glycosylation sites were eventually determined, the data interpretation was time-consuming due to the unpredicted peptide backbone cleavages. Therefore, this method is limited to the analysis of one purified O-glycopeptide containing a limited number of O-glycosylation sites.

One interesting approach for the identification of glycosylation sites using nonspecific proteases was proposed [92]. Unlike trypsin, the most popular protease, the nonspecific proteases cleave the peptide backbone nonspecifically and usually produce relatively small peptides and glycopeptides. For example, the *N*-glycosylated dipeptide (NL) and tripeptide (NLT) were observed with MALDI-PSD after protease E digestion[92]. Protease K digestion combined with precursor ion scanning on various tandem mass spectrometers was also studied recently[93]. To successfully assign the glycosylations, high resolution mass spectrometers is preferred such as MALDI – FT ICR MS [94]. So far, the use of nonspecific enzymes has proven an effective method for locating glycosylation sites, although most work has been done on glycoproteins carrying one or two glycosylation sites. It is questionable whether it is a practical approach for multiple glycosylation sites, since the spectra would be quite complicated.

Probably the most important advances in MS instrumentation for the study of glycosylation is the development of unique fragmentation techniques compatible with FT ICR MS such as Electron Capture Dissociation (ECD)[95;96]. ECD provides an alternative fragmentation technique to the traditional collisional or infrared-induced excitation of gas-phase

molecular ions. The mechanism of ECD is still poorly understood, but it is based on the dissociative recombination of multiply protonated polypeptide molecules with low-energy electrons. In general, the thermal electrons are captured by the polycations, followed by charge neutralization, which leads to an excited radical species that rapidly (within 10^{-11} s) undergoes bond cleavage[97]. The fragment ions produced by ECD are mainly N-terminal C' and C-terminal $Z^{\cdot}$ fragments, complementary to those obtained by CID. Secondary losses from termini and side chains occur, but usually with much lower abundance than the small neutral losses typically observed in traditional MS/MS. Essentially, the normally labile post-translational modification bonds, such as in glycosylation and phosphorylation, are retained during ECD. Therefore, ECD has been applied to study phosphorylation [98], as well as *N*-[99]and *O*-glycosylation [100]. The sensitivity of ECD is lower than traditional CID due to charge loss during electron capture. It is reported that the overall efficiency of ECD for peptides is 20 to 50% [101]. So far, ECD has only been used in FTICR MS. This is because it usually requires at least several milliseconds to ensure electron capture by most precursor ions [101], and this exceeds the residence time of ions in many types of mass spectrometers.

2.2.2.3 Glycan structure elucidation

In the age of glycomics, the separation, sequencing, linkage and composition analysis of oligosaccharides are the major tasks. Mass spectrometry combined with carbohydrate specific chromatography separations, such as serial lectin chromatography, play very important roles in this rapidly developing area. Yet, the analysis of oligosaccharides by MS has lagged significantly behind its application in protein analysis. This is due to the essential difficulties associated with the structural complexity of oligosaccharides, and the relatively low response in

the MS (compared to peptides). For peptides, the charge is largely associated with the basic residues, such as N-terminus, Arginine and Lysine in positive ion mode. Any tryptic peptide would contain at least two positive charges. However, neutral oligosaccharides lack groups with high proton affinity and are usually ionized with lower efficiency in positive mode. As a consequence, the signals in the MS are much weaker compared to the signals from peptides. These oligosaccharides which contain sialic acid, the situation is even worse since these residues preferentially retain the negative charges, resulting in a signal distribution between the positive and negative ions. Despite the analytical challenges, the application of MS to oligosaccharides has been developing rapidly. Some developments have been reviewed recently[102].

Derivatizing the intact oligosaccharides with stable functional groups have been widely used to achieve the goal of increasing the sensitivity of the mass spectrometric analysis [103]. Most of the derivatization methods involve covalent attachment of an appropriate moiety by reductive amination. Usually, a basic moiety is attached to the anomeric carbon (C1 position) at the reducing end of oligosaccharides, and therefore, the probability of protonation is greatly improved (Figure 2.11). In addition, the labeling compounds are usually aromatic in structure so that it is possible to separate the labeled oligosaccharides by HPLC while monitoring the effluents with a UV detector. The common compounds used for derivatization have been reviewed[104]. As an alternation to reductive amination, the formation of oximes is another common method for labeling the reducing end of oligosaccharides. An example is the labeling of oligosaccharides with a basic peptide, resulting in an increase in the sensitivity on MALDI-TOF MS by 50~100 times as compared to the native forms [105].

Several methods can be used to obtain oligosaccharide sequence information. The classic one is sequential enzymatic digestion and determination of the resulting oligosaccharide

molecular weight by MALDI MS after each step. The shifts in mass, combined with knowledge of the specificity of the enzymes, determine the stereochemical sequence and linkage. Tandem mass spectrometry, such as CID, or PSD-TOF[106], has also proven to be a powerful tool for oligosaccharide structural analysis,. CID works not only on oligosaccharides themselves, but also on glycopeptides since the glycosidic linkages are preferentially cleaved by CID. Both low energy CID (such as in QQQ or QTOF[107] instruments) and high energy CID[108] (such as in TOF-TOF instruments) have been used to study oligosaccharide structures. The nomenclature for oligosaccharide product ions generated by tandem mass spectrometry follows the rules laid out by Domon and Costello[109] (Figure 2.12). Mass spectra from low energy CID are dominated by the ions resulting from glycosidic linkage cleavages (Y and B ion series), thus allowing the sequence of oligosaccharides to be determined. To determine the linkage position, ions from cross ring cleavages are desired and can be obtained by using high collisional energy in CID[110]. Recently developed TOF-TOF instruments have CID energies of 1000 eV (QTOF is below 100 eV), and cross ring cleavages can be observed along with much weaker y and b ions. Therefore, both the linkage position and the sequence can be deduced accordingly. Recently, using the TOF-TOF instrument characterization of oligosaccharide structures was demonstrated [111]. The principle down side to using high energy CID is the uncontrollable fragmentation, resulting in certain unpredicted ions, which complicates the spectrum.

2.3. Plant pathogenic proteins: Pectin degrading enzymes, their structures and characterization

Most plant pathogens produce an array of extracellular enzymes that degrade the cell wall and aid the penetration and colonization of plant tissue. Cell walls, as part of the first line of the

plant defense system, are primarily composed of complex polysaccharides. To be successful in attacking the host cell, a mixture of cell wall degrading enzymes is secreted by pathogens which hydrolyze cell walls by cleaving the glycosidic bonds. Pectin degradation enzymes (PDEs) are a family of enzymes capable of hydrolyzing pectins, a major component of primary walls. These enzymes include polygalacturonases, pectin/pectate lyases and pectin methylesterases. The functional diversity of PDEs has led to the need for a greater understanding of their structures. This session reviews the basic background of plant primary cell wall structure, functions of PDEs, and their structural characterization.

Most plant cells are enclosed by a more or less rigid cell wall, with the exception for a few algae, some protists, and endosperm cells. The plant cell wall is a complex structure, which serves a variety of functions, from protecting cellular contents and giving rigidity to the plant structure, to regulating the life cycle of the plant organism.

The cell wall is formed from fibrils of cellulose molecules, embedded in a water-saturated matrix of polysaccharides and structural glycoproteins (Figure 2.13). Many plant cell walls have three layers, the middle lamella, primary wall and secondary wall. The middle lamella is the first layer formed during cell division. It is shared by adjacent cells, and is composed of pectic compounds and protein. The primary cell wall is formed after the middle lamella and consists of a rigid skeleton of cellulose microfibrils cross linked by hemicellulosic polymers and embedded in a gel-like pectin matrix. The secondary cell wall is produced inside the primary wall after cell enlargement is complete. In between the primary and secondary cell walls, are embedded conduits called plasmodesmata, which penetrate both layers and provide s pathway for transporting cytoplasmic molecules from one cell to another. Besides the three layers, cell walls also contain functional proteins including oxidative enzymes (peroxidases),

hydrolytic enzymes (pectinases and cellulases) and expansins. Those enzymes satisfy a variety of functions, from initiating reactions that form the structural networks of molecules and helping protect plants against fungal invasions by degrading the cell walls of the fungi, to degrading the plant cell walls to induce fruit to ripen and leaves to fall in autumn.

The main components of primary cell walls are members of two polysaccharide networks, one consisting of cellulose and hemicellulose (e.g. xyloglucan in dicotyledons and arabinoxylan in graminaceous monocotyledons), and the other consisting of pectic polysaccharides (Figure 2.13)[112-114]. Beside the complex polysaccharides, other components in the primary wall are structural proteins (hydroxyproline-rich glycoproteins and extensins), phenolic esters (ferulic and coumaric acids), ionically and covalently bound minerals [115;116] and enzymes.

Pectins are characterized by a backbone of 1,4-linked α-D-galacturonic acid (GalpA) residues. In the primary cell wall, they have three domains: the homogalacturonans (HG), and the highly branched rhamnogalacturonans I and II (RG-I and RG-II) [112;117;118]. HG is a linear chain of 1,4-linked α-D-galactopyranosyluronic acid (GalpA) residues in which some of the carboxyl groups are methyl esterified. In some cases, it may also be partially O-acetylated at C-3 or C-2[117]. The backbone of RG-I contains repeating disaccharide [→4)-α-D-Gal*p*A-(1→2)-α-L-Rhap-(1→]. The Gal*p*A residues typically are not substituted with mono or oligosaccharide side chains, and in contrast, 20–80% of the rhamnosyl (Rhap) residues are substituted at C-4 with neutral and acidic oligosaccharide side chains[119]. The other substituted galacturonan, RG-II, is more closely structurally related to HG since its backbone is composed of 1,4-linked α-D -Gal*p*A residues. Four structurally different oligosaccharide side chains are linked to the RG-II backbone[120]. It is often postulated that in primary walls HG, RG-I and RG-II are covalently

linked to one another and models have been proposed for the distribution of RG-I and HG in pectin. [121]

Most plant pathogens secrete a mixture of enzymes including pectinases that can hydrolyze the polysaccharide network within primary walls. Depending on the organisms producing them, these enzymes can have different functionalities in nature. When produced by pathogenic microorganisms, these enzymes serve as virulence factors. On the other hand, when produced endogenously (by the plant), they can act as enzymes important for fruit ripening [122;123], or serve as enzymes involved in the decay of plant tissue[124-126]. During infection, pathogens secrete one or multiple forms of a variety of pectin degrading enzymes, including various polygalacturonases (PGs), pectin methylesterases (PME) and pectin/pectate lyases (PL/PnL). Each of them has different substrate activities as shown in Figure 2.14. PME demethylates pectin to pectate which can subsequently be degraded by PGs. PGs (exoPG and endoPG) are most efficient at degrading the backbone of de-methylated pectin. Pectin and pectate lyases catalyse a β-elimination resulting in Δ4,5 unsaturated GalA at the non-reducing end. Pectate lyase requires Ca2+ to functioning properly, which is not seen for pectin lyase[127]. The different components in the pectic compound network suggest that a broad range of pectin degrading enzymes is required for total degradation of the pectin.

Endopolygalacturonases (PGs) are among the first array of enzymes released by pathogenic fungi [128-130] to break down the pectin by hydrolyzing the α-1,4 glycosidic bonds between adjacent α-D- galacturonosyl residues present in homogalacturonan chains [131;132]. Depending on the enzyme, the degradation proceeds either by a strict endo-mode or by an endo-/exo-mode of cleavage[133]. The released pectic fragments are mostly composed of linear oligogalacturonic acids (OGAs). The extent of degradation varies according to the level of

methylesterification of α-D- galacturonosyl residues. The relationship between such fragments and defense responses have been reviewed extensively[117;134] There is evidence that early responses such as plasma membrane depolarization, ion fluxes, and cytosol acidification are induced in suspension-cultured cells after treatment with OGAs [135]. Whether endoPGs elicit defense responses via the release of pectic fragments when expressed *in planta* is still unclear.

EPGs have been described from numerous species of phytopathogenic fungi, bacteria and insects [136]. Targeted gene disruption of pectinase genes was performed to evaluate the role of these enzymes in pathogenicity and a reduced virulence was observed in several fungi that cause tissue maceration. Shieh et al. showed that an endopolygalacturonase from the opportunistic pathogen *Aspergillus flavus* is involved in the infection of cotton balls[137]. The reduced virulence of the fungus on tomato and apple after the deletion of one out of six endopolygalacturonase genes of *Botrytis cinerea* was reported by ten Have et al[137;138]. However, it was shown that endopolygalacturonase can also play a major role in the biotrophic development of pathogens. For example, endoPG is a true pathogenicity factor in the *Claviceps purpurea*/rye interaction[137;139], suggesting that polygalacturonase could facilitate biotrophic development without causing extensive maceration of infected tissues.

Despite their biological importance, a great deal of work remains to be done to characterize endoPG structures. Usually, a group of EPGs are expressed during fungal infection. The amino acid sequences for endoPG isozymes secreted by the same fungus are often similar, based on their gene sequence. For instance, six out of seven different endopolygalacturonase encoding genes from *A. Niger* have been characterized [140], along with their biochemical properties [131;141;142]. Even though these isozymes share 60-70% amino acid sequence, their post-translational modifications were determined to be substantially different. Understanding the

structure of EPGs and their post-translational modifications is the first step to deciphering their complex biochemical properties. Mass spectrometry has proven to be an effective tool to study the structural features of certain endoPGs[143-145]. Other analytical tools, such as NMR spectroscopy[146;147], and liquid chromatography [148] have also been used to study endoPG structures.

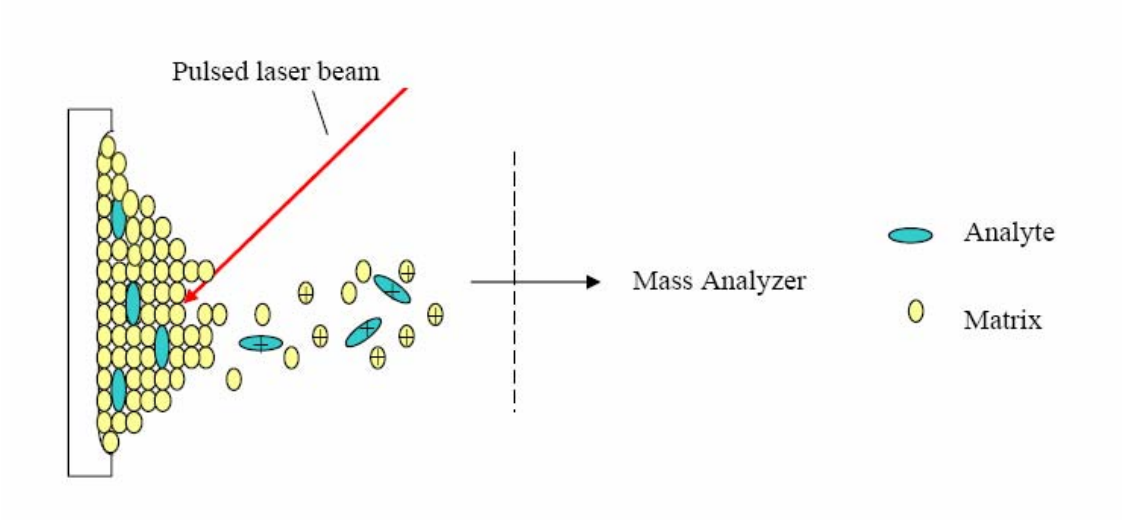


Figure 2.1: Cartoon diagram of Matrix-assisted Laser Desorption/Ionization (MALDI)

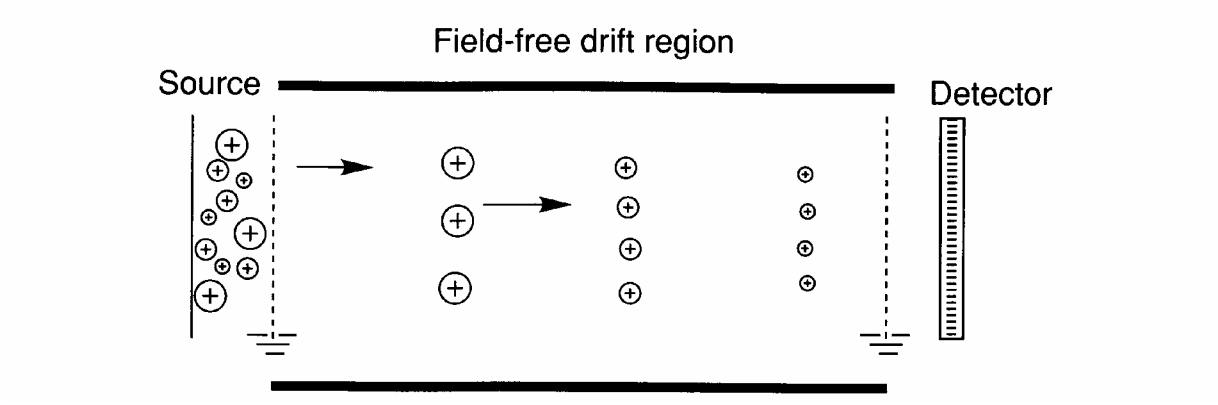


Figure 2.2: Principle of mass separation by a linear time-of-flight mass analyzer

Ions are separated on the basis of their m/z: high mass ions (big circles) travel more slowly than the lighter ions (small circles)

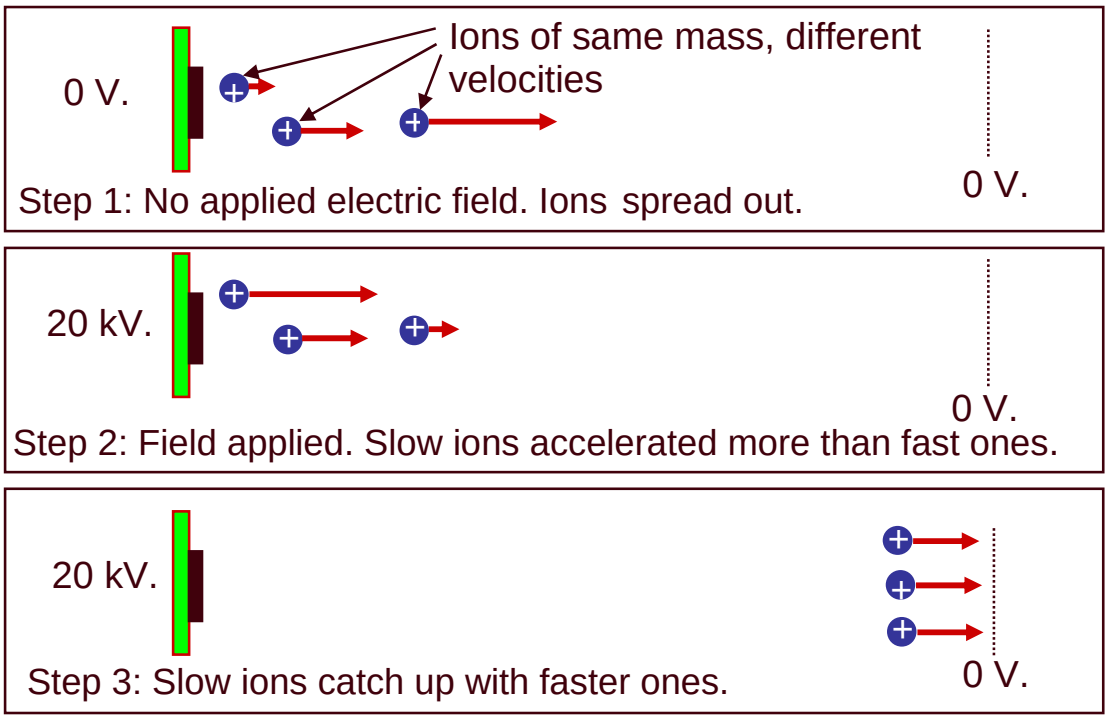


Figure 2.3: Schematic diagram of delayed extraction

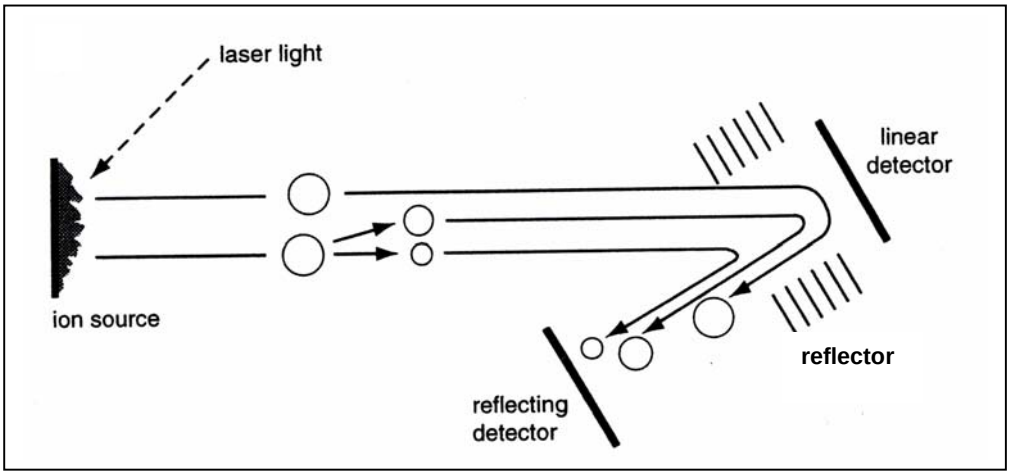


Figure 2.4: A diagram of a reflectron time-of-flight mass spectrometer.

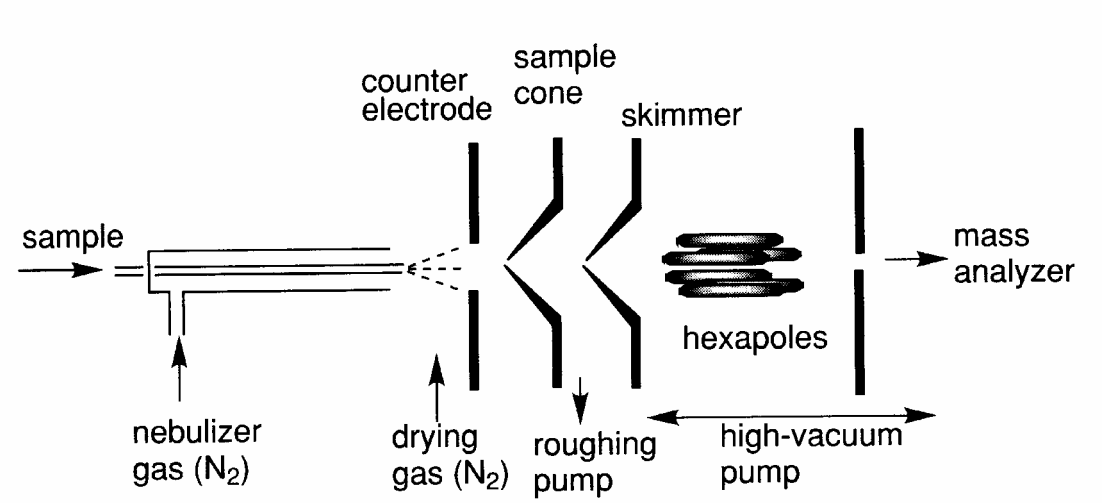


Figure 2.5: A diagram of electrospray ionization source, in which ions are sampled through a small-orifice skimmer

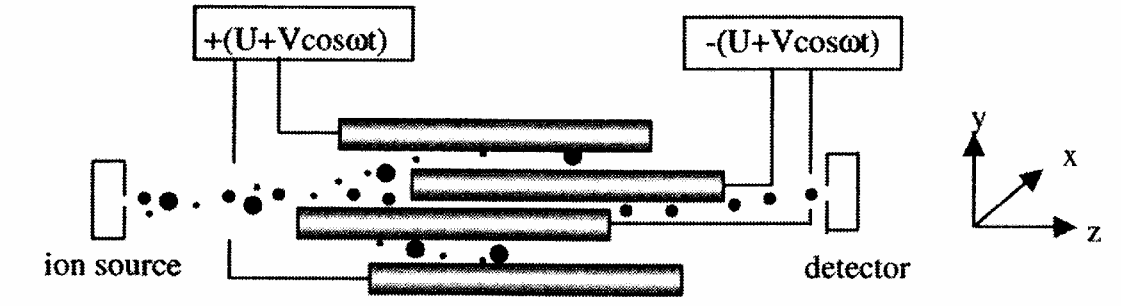


Figure 2.6: A diagram of quadrupole mass analyzer.

At a certain values of DC potential U and RF potential V, ions of a specific m/z are made to have stable trajectories in the *x* and *y* directions, and travel in the *z* direction within the quadrupole field; all other ions are lost due to the unstable trajectories.

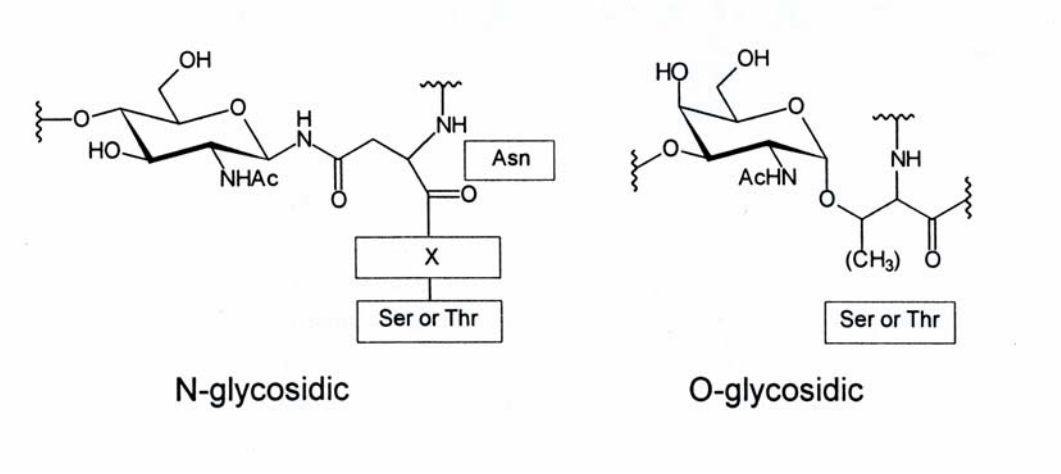


Figure 2.7: *N*- and *O*-linked glycosidic bonds

N-linked glycosidic bond formed on the side chain of Asn within a tripeptide sequon Asn-Xaa-Ser or Thr; *O*-linked glycosidic bonds formed on the –OH group of Ser or Thr side chain

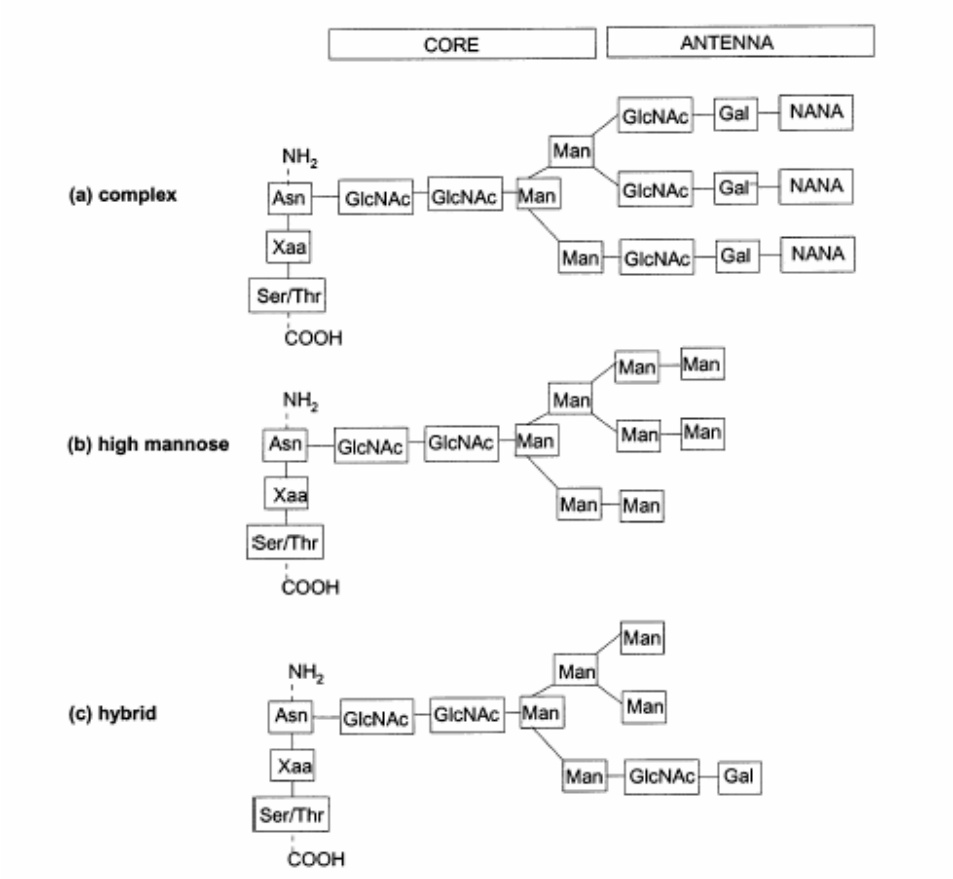


Figure 2.8: Structure of three main types of N-linked glycans (high mannose, complex and hybrid)

Man = mannose; GlcNAc = N-acetylglucosamine
Gal = galactose; GlcNAc = N-acetylgalactosamine;
NANA = sialic acid

Table 2.1: Different types of O-linked glycans

Type of *O*-linked sugar	Structure and peptide linkage	Glycoprotein
Mucin-type	(R-)GalNAcα1-Ser/Thr	Secreted and Membrane-bound
O-linked GlcNAc	GlcNAc1-Ser/Thr	Nuclear and cytoplasmic
O-linked fucose	(Siaα2-6Galβ1-4GlcNAcβ1-3)Fucα1-Ser/Thr	EGF protein domains
O-linked glucose	(Xylα1-3Xylα1-3)Glcβ1-Ser	EGF protein domains
O-linked galactose	Glcα1-2Galβ1-*O*-Lys	Collagen
O-linked mannose	1) Siaα2-3Galβ1-4GlcNAcβ1-2/3Man1-Ser/Thr	1) Bovine α-dystroglycan, rat brain proteoglycans
	2) (Manα1-2)$_n$Manα1-Ser/Thr	2) Yeast glycoproteins
O-linked arabinose	Arabinose-*O*-Pro	Vegetal proteins

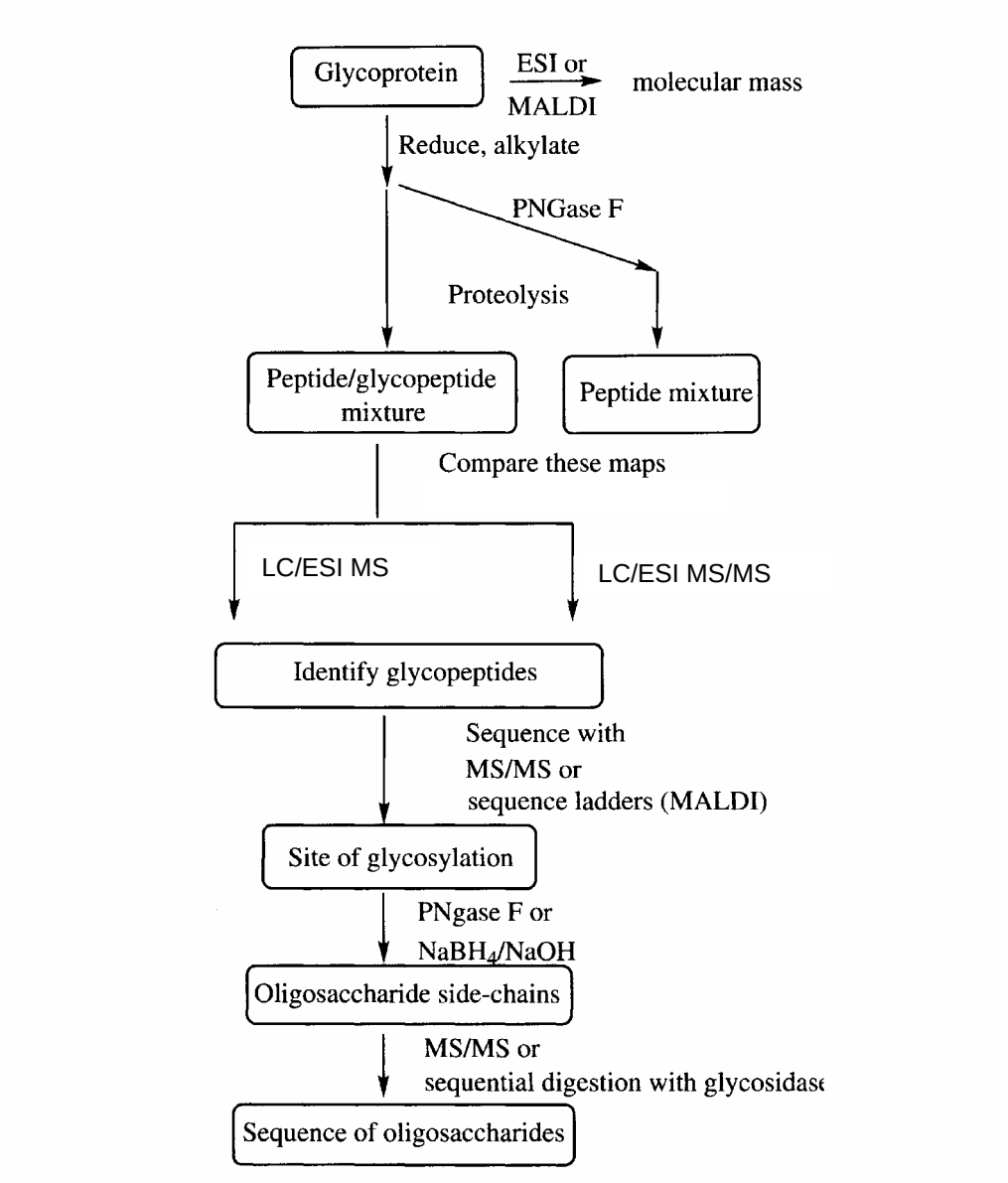


Figure 2.9: A general protocol for glycoprotein characterization

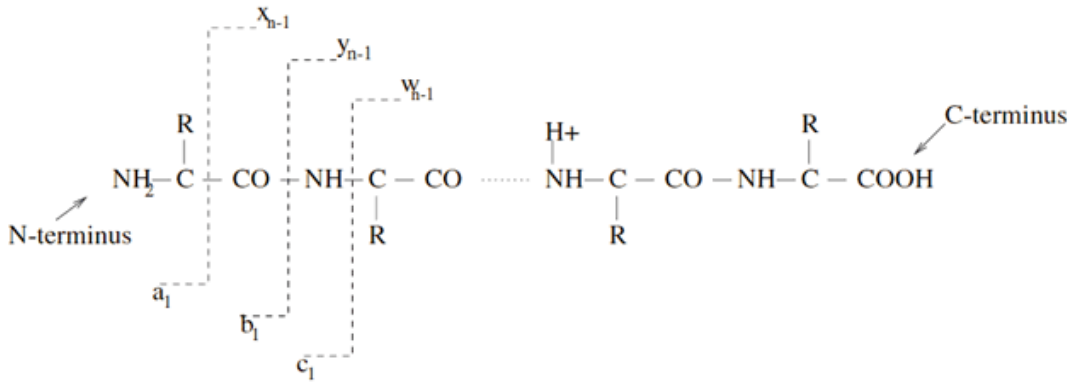


Figure 2.10: Types of peptide fragments ions generated by Tandem MS

Fragments will only be detected if they carry at least one charge. If this charge is retained on the N terminal fragment, the ion is classed as either *a*, *b* or *c*. If the charge is retained on the C terminal, the ion type is either *x*, *y* or *z*. A subscript indicates the number of residues in the fragment.

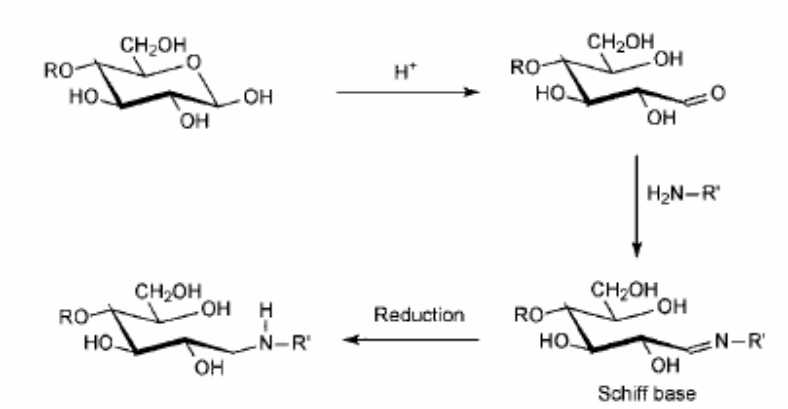


Figure 2.11: Scheme showing the preparation of reducing terminal derivatives by reductive amination. R = Remainder of carbohydrate, R' = amine.

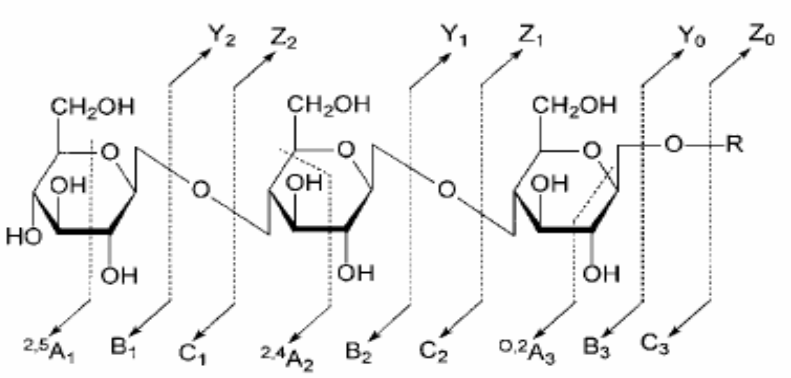


Figure 2.12. Nomenclature for oligosaccharides fragment ions generated by tandem mass spectrometry (From Domon & Costello, 1988)

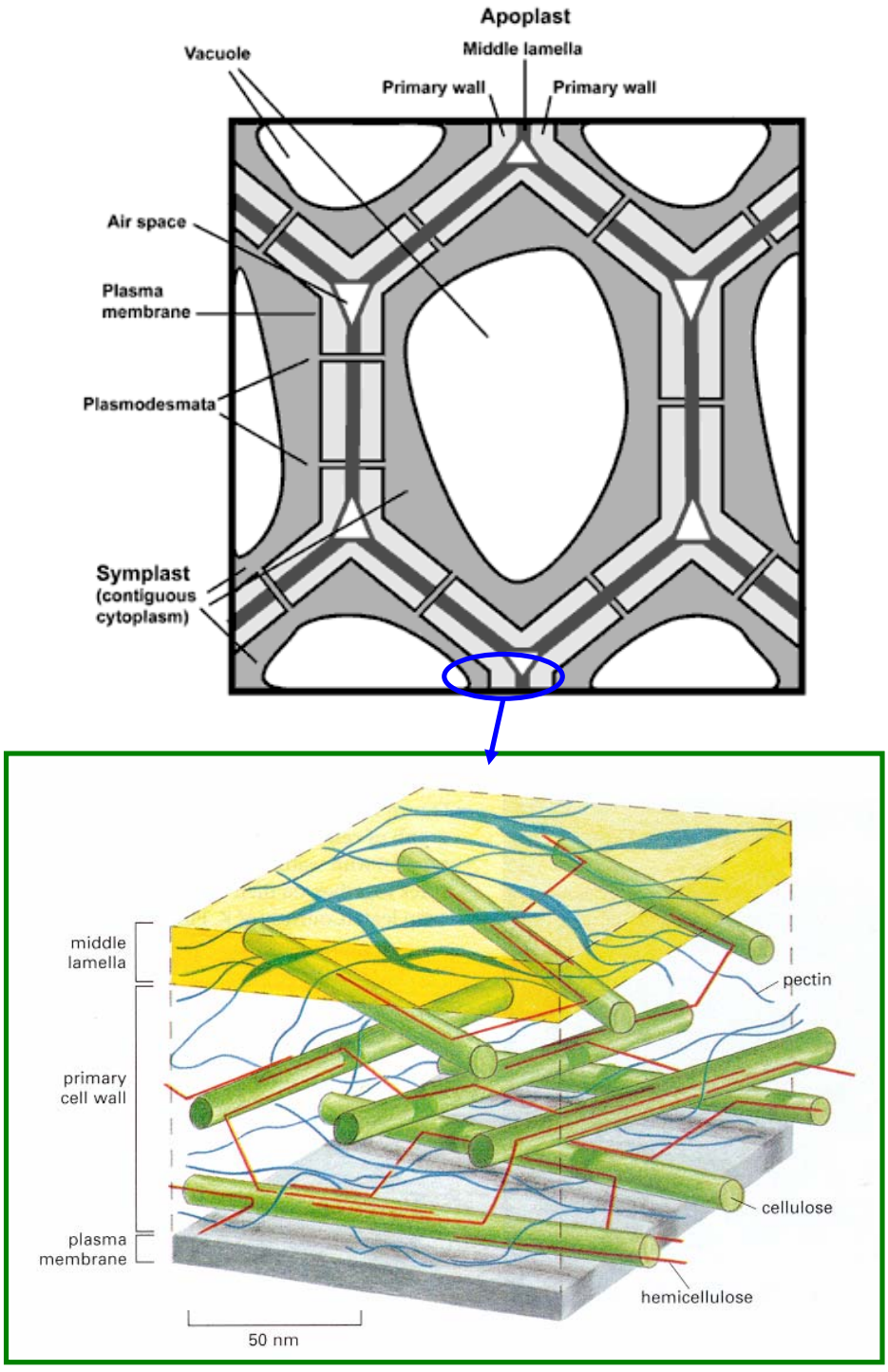


Figure 2.13. Primary plant cell structure

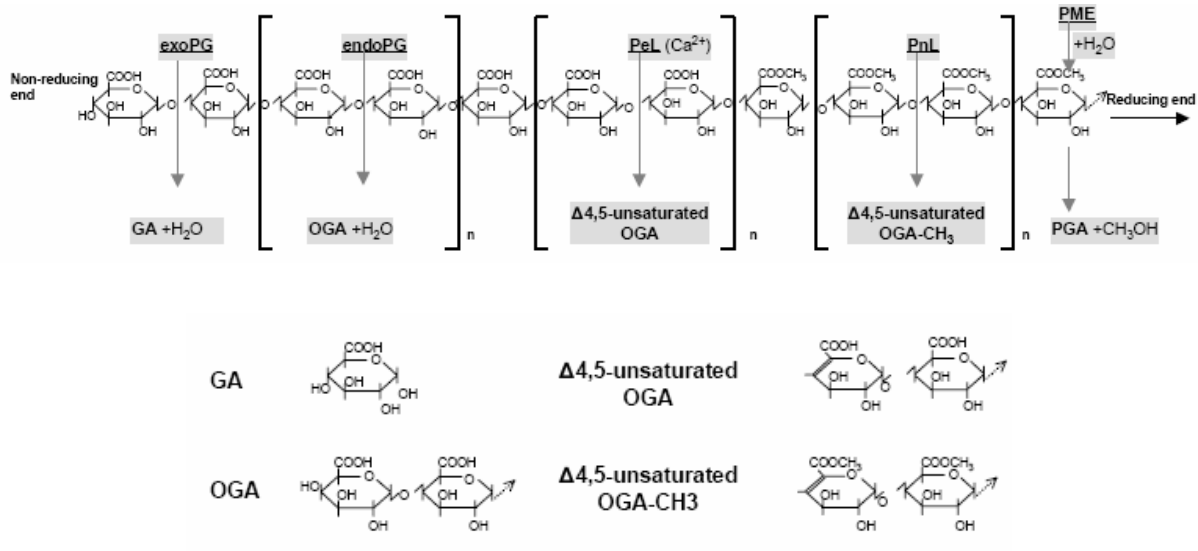


Figure 2.14. Cartoon of pectinase activities on polygalacturonic acid (PGA) residues

Polygalacturoanse (exoPG and endoPG) hydrolyze PGA at α-1,4 glycosidic bond resulting monogalacturonic acid (GA) and oligosaccharide galacturonic acid (OGA), respectively; Pectate and pectin lyase (PeL and PnL) perform β-elimination on the methylated PGA, resulting in OGA with a Δ 4,5 unsaturated bond at the non-reducing end; Pectin methylesterase (PME) demethylates pectin resulting in PGA.

References

1. F.W.McLafferty; F.Turecek *Interpretation of Mass Spectra,* 34 ed.; University Science Books: Mill Valley, CA, 1993.

2. M.Rosenstock; M.B.Wallenstein; A.L.Wahrhaftig; H.Eyring *Proc.Natl.Acad.Sci.* **1952,** *38*, 667-78.

3. R.A.Marcus *J.Chem.Phys.* **1952,** *20*, 359-62.

4. A.E.Ashcroft *Ionization Methids in Organic Mass Spectrometry,* The Rogal Society of Chemistry: Cambridge, England, 1997.

5. M.S.B.Munson; F.H.Field *J.Am.Chem.Soc.* **1966,** *88*, 2621-30.

6. M.Karas and F.Hillenkamp *Anal.Chem.* **1988,** *60*, 2299-301.

7. K.Tanaka; H.Waki; H.Ido; S.Akita; T.Yoshida *Rapid, Commun.Mass Spectrom.* **1989,** *2*, 151-53.

8. B.Spengler; R.J.Cotter *Anal.Chem.* **1990,** *62*, 793.

9. R.Zenobi; R.Knochenmuss *Mass Spectrom.Rev.* **1998,** *17*, 337-66.

10. M.L.Vestal *Chem.Rev.* **2001,** *101*, 361-75.

11. M.Karas; R.Kruger *Chem.Rev.* **2003,** *103*, 427-39.

12. M.Karas; M.Gluckmann; J.Schafer *J.Mass Spectrom.* **2000,** *35*, 1-12.

13. R.S.Annan; H.J.Kochling; J.A.Hill; and K.Biemann *Rapid, Commun.Mass Spectrom.* **1992,** *6*, 298-302.

14. J.Qin and B.Chait *J.Am.Chem.Soc.* **1995,** *117*, 5411-12.

15. *Anal.Chem.* **1995,** *67*, 2498-509.

16. D.C.Imrie; J.M.Pentney; J.S.Cottell *Rapid, Commun.Mass Spectrom.* **1995,** *9*, 1923.

17. M.Moniatte; F.G.van der Goot; J.T.Buckley *FEBS Lett.* **1996,** *384*, 269.

18. P.Onnerfjord; J.nilsson; L.Wallman *Anal.Chem.* **1998,** *70*, 4755.

19. S.M.Brown; J.J.Lennon *Anal.Chem.* **1995,** *67*, 1950-54.

20. S.M.Colby; T.B.King; J.P.Reilly *Rapid Commun.Mass Spectrom.* **1994,** *8*, 865-68.

21. W.C.Wiley and J.B.Mclaren *Rev.Sci.Instrum.* **1955,** *16*, 1150.

22. B.A.mamyrin; V.I.Karataev; D.V.Schmikk *Sov.Phhys.JETP* **1973,** *37*, 45-48.

23. B.A.mamyrin 131 ed.; 1994.

24. R.J.Cotter *Time-of-flight Mass Spectrometer:Instrumentation and Applications in Biological Research,* ACS ed.; Washington DC, 1997.

25. R.J.Cotter *Anal.Chem.* **1999,** *71*, 445A-51A.

26. B.Spengler, D. Kirsch and R. Kaufmann *Rapid Commun.Mass Spectrom.* **1991,** *5*, 198.

27. S.M.Brown; J.J.Lennon *Anal.Chem.* **1995,** *67*, 1950-54.

28. J.K.Olthof; I.A.Lys; R.J.Cotter *Rapid Commun.Mass Spectrom.* **1988,** *2*, 171-75.

29. R.kaufmann; D.Kirsch; B.Spengler *Int.J.Mass Spectrom.Ion Proc.* **1994,** *131*, 355-85.

30. J.C.Rouse; W.Yu; S.A.martin *J.Am.Soc.Mass Spectron.* **1995,** *6*, 822-35.

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

32. L.L.Mack; P.Kralick; A.Rheude; M.J.Dole *J.Chem.Phys.* **1970,** *52*, 4977-86.

33. M.Yamashita; J.B.Fenn *J.Phys.Chem* **1984,** *88*, 4671-75.

34. J.B.Fenn; M.Mann; C.K.Wang; S.F.Wong; C.M.Whitehouse *Science* **1989,** *246*, 64-71.

35. M.Dole; L.L.Mack; R.L.Hines; et al *J.Chem.Phys.* **1968,** *49*, 2240-49.

36. J.V.Iribarne and B.A.Thompson *J.Chem.Phys* **1976,** *64*, 2287.

37. B.A.Tho, pson and J. V. Iribarne *J.Chem.Phys* **1979,** *71*, 4451.

38. Sui, Guvremont et al *Org.Mass Spectrom* **1993,** ***28***, **579**.

39. Juan Fernandez de la Mora, Gary J. Van Berkel Christie G. Enke Richard B. Cole Manuel Martinez-Sanchez and John B. Fenn *J.Mass Spectrom.* **2000,** *35*, 935-52.

40. R.Korner; M.Wilm; K.Morand; M.Schubert; M.Mann *J.Am.Soc.Mass Spectron.* **1996,** *7*, 150-56.

41. B.E.Winger; S.A.Hofstadler; J.E.Bruce; H.R.Udseth; R.D.Smith *Nature* **1993,** *379*, 466-69.

42. R.D.Smith; J.E.Bruce; Q.Wu; X.Cheng; S.A.Hofstadler; et al *Mass Spectrometry in the Biological Sciences*, A.L.Burlingame; S.A.Carret al, Eds.; Humana Press: Totowa N.J., 1996.

43. A.N.Verentchikov; W.Ens; K.G.Standing *Anal.Chem.* **1994,** *66*, 126-33.

44. O.A.Mirgorodskaya; A.A.Shevchenko; A.I.Miroshnikov *Anal.Chem.* **1994,** *66*, 99-107.

45. P.H.Dawson *Mass Spectrom.Rev.* **1986,** *5*, 1-37.

46. P.H.Dawson *Quadrupole Mass Spectrometry and its Application,* Woodbury NY, 1994.

47. R.E.march; R.J.Hughes *Quadrupole Storage Mass Spectrometry,* Wiley: New York, 1989.

48. M.H.Amad; R.S.Houk *J.Am.Soc.Mass Spectron.* **2000,** *11*, 407-15.

49. I.V.Chernushevich, A. V. Loboda and B. A. Thomson *J.Mass Spectrom.* **2001,** *36*, 849-65.

50. A.Dodonov; V.Kozlovsky; A.Loboba; et al *Rapid Commun.Mass Spectrom.* **1997,** *11*, 1649-56.

51. A.N.Krutchinsky; I.V.Chernushevich; V.L.Spicer; W.Ens; K.G.Standing *J.Am.Soc.Mass Spectron.* **1998,** *9*, 569-79.

52. Feizi Tan; M.Barbara *Current Opinion in Structural Biology* **2003,** *13*, 602-04.

53. Hirabayashi, Jun; and Kasai, Ken-ichi *Trends in Glycoscience and Glycotechnology* **2000,** *12*, 1-5.

54. K.R.Love and P.H.Seeberger *Angew.Chem.Int.Ed.* **2002,** *41*, 3583-86.

55. C.F.Goochee; M.J.Gramer; D.C.Anderson; J.B.Bahr; J.R.Rasmussen *Bio technology* **1991,** *9*, 1347-55.

56. R.A.Dwek *Chem.Rev* **1996,** *96*, 683-720.

57. J.F.Robyt *Essentials of Carbohydrate Chemistry*, Springer: New York, 1998; Chapter 9.

58. Ajit Varki; R.Cummings; J.Esko; H.Freeze; G.Hart; J.Marth *Essentials of Glycobiology,* Cold Spring Harbor laboratory Press: Cold Spring Harbor, New York, 1999.

59. P.E.Van Den Steen; P.M.RUdd; M.R.Wormald; R.A.Dwek; G.Opdenakker *Trends in Glycoscience and Glycotechnology* **2000,** *12*, 35-49.

60. E.Stimson; M.Virji; K.Makepeace; A.Dell; H.R.Morris; et al *Mol.Microbiol.* **1995,** *17*, 1201-14.

61. C.Fenseleau; M.Kelly *Biochemical and Biotechnological Application of Electrospray Ionization Mass Spectrometry*, A.P.Snyder, Ed.; Americam Chemical Society: Washington D C, 1995.

62. Y.Q.Dai, R. M. Whittal *Dai, Y.Q., Whittal, R.M.* **1996,** *68*, 2721-25.

63. Y.Q.Dai, R. M. Whittal *Anal.Chem.* **1999,** *71*, 1087-91.

64. Y.Q.Dai, R. M. Whittal *Dai, Y.Q., Whittal, R.M.* **1996,** *68*, 2721-25.

65. N.Zhang, A. Doucette and L. Li *Anal.Chem.* **2001,** *73*, 2968-75.

66. J.A.Blakledge; A.J.Alexander *Anal.Chem.* **1995,** *67*, 843.

67. M.M.Vestling *Anal.Chem.* **1994,** *66*, 471.

68. K.K.Mock *Rapid Commun.Mass Spectrom.* **1992,** *6*, 233.

69. M.E.Warren, A. H. Brockman R. Orlando *Anal.Chem.* **1998,** *70*, 3757-61.

70. J.Bundy, C. Fenselau *Anal.Chem.* **1999,** *71*, 1460-63.

71. J.Bundy, C. Fenselau *Anal.Chem.* **2001,** *73*, 751-57.

72. C.Afonso, C. Fenselau *Anal.Chem.* **2003,** *75*, 694-97.

73. A.H.Brockman, R. Orlando *Anal.Chem.* **1995,** *67*, 4581-85.

74. F.Hillenkamp; M.Karas; R.C.Beavis; B.T.Chait *Anal.Chem.* **1991,** *60*, 2299-301.

75. E.Mortz; T.Sareneva; I.Julkunen; P.Roepstorff *J.Mass Spectrom.* **1996,** *31*, 1109-18.

76. J.Zaia; R.E.Boynton; A.McIntosh; et al *J Biol Chem* **1997,** *272*, 14120-26.

77. E.S.Jacoby; A.T.Kicman; P.Laidler; R.K.Iles *Clin Chem* **2000,** *46*, 1796-803.

78. J.Colangelo, V. Licon J. Benen J. Visser C. Bergmann R. Orlando *Rapid Commun.Mass Spectrom.* **1999,** *13*, 2382-87.

79. P.Edman *Acta Chim.Scand* **1950,** *4*, 238-93.

80. D.H.Williams; C.Bradley; G.Bojesen; S.Santikarn; K.F.Faull; P.Rpepstorff *Proc.Natl.Acad.Sci.* **1989,** *103*, 5700-04.

81. C.Dass; D.M.Desiderio *Anal.Biochem.* **1987,** *163*, 52-66.

82. S.A.Carr; M.J.Huddleston; M.F.Bean *Protein Sci.* **1993,** *2*, 183-96.

83. J.J.Conboy; J.D.Henion *J.Am.Soc.Mass Spectron.* **1992,** *3*, 804-14.

84. Jiang, Hui; Desaire, Heather; Butnev, Vladimir Y.; Bousfield, George R. *Journal of the American Society for Mass Spectrometry* **2004,** *15*, 750-58.

85. H.Steen; B.Kuster; M.Mann *J.Mass Spectrom.* **2001,** *14*, 1793-800.

86. M.A.Ritchie; N.Johnson; J.B.Hoyes; A.Millar; R.Caruthers; C.Jones; J.Langridge 2003 2-10-2003.

87. J.Qu; Q.Liang; G.Luo; Y.Wang *Anal.Chem.* **2004,** *76*, 2239-47.

88. M.Wilm; G.Neubauer; M.Mann *Anal.Chem.* **1996,** *68*, 527-33.

89. J.Bunkenborg; B.J.Pilch; A.V.Podtelejnikov; J.R.Wisniewski *Proteomics* **2004,** *4*, 465.

90. P.Hagglund, J.Bunkenborg, F.Elortza, O.N.Jensen, and P.Roepstorff. A New Strategy for Identification of N-Glycosylated Proteins and Unambiguous Assignment of Their Glycosylation Sites Using HILIC Enrichment and Partial Deglycoslation. Journal of Proteome Research Outcoming. 2004.

Ref Type: In Press

91. G.J.Rademaker; J.Haverkamp; J.Thomas-Oates *Organic Mass Spectrometry* **1993,** *28*, 1536-41.

92. Rademaker, Geert Jan; Pergantis, Spiros A.; Blok-Tip, Leonore; Langridge, James I.; Kleen, Astrid; Thomas-Oates, Jane E. *Analytical Biochemistry* **1998,** *257*, 149-60.

93. F.Hanisch; M.Jovanovic; J.Peter-Katalinic *Analytical Biochemistry* **2001,** *290*, 47-59.

94. L.Wells; K.Vosseller; R.N.Cole; J.M.Cronshaw; M.J.Matunis; G.W.Hart *Molecular & Cellular Proteomics* **2002,** *1*, 791-804.

95. G.M.Hathaway, J. Z. a. F. R. 2003 .

96. O.Vorm; M.Mann *J.Am.Soc.Mass Spectrom.* **1994,** *5*, 955-58.

97. Mirgorodskaya, Ekaterina; Hassan, Helle; Wandall, Hans H.; Clausen, Henrik; Roepstorff, Peter *Analytical Biochemistry* **1999,** *269*, 54-65.

98. P.Juhasz; S.A.Martin *International Journal of Mass Spectrometry and Ion Processes* **1997,** *169/170*, 217-30.

99. Hui Jiang; Heather Desaire; V.Y.Butnev; G.R.Bousfield *J.Am.Soc.Mass Spectrom.* **2004,** *15*, 750-58.

100. H.J.An; T.R.Peavy; J.L.Hedrick; C.B.labrilla *Anal.Chem.* **2003,** *75*, 5628-73.

101. R.A.Zubarev, N. L. Kelleher F. W. McLafferty *J.Am.Chem.Soc.* **1998,** *120*, 3265-66.

102. R.A.Zubarev, N. A. Kruger E. K. Fridriksson M. A. Lewis D. M. Horn B. K. Carpenter F. W. McLafferty *J.Am.Chem.Soc.* **1999,** *121*, 2857-62.

103. Zubarev RA, Haselmann KF Budnik BA Kjeldsen F Jensen F *Eur J Mass Spectrom* **2002,** *8*, 337-49.

104. A.Stensballe, O. N. Jensen J. V. Olsen K. F. Haselmann R. A. Zubarev RA *Rapid Commun.Mass Spectrom.* **2000,** *14*, 1793-800.

105. Hakansson K, Cooper HJ Emmett MR Costello CE Marshall AG Nilsson C *Anal.Chem.* **2001,** *73*, 4530-36.

106. E.Mirgorodskaya, P. Roepstorff R. A. Zubarev *Anal.Chem.* **1999,** *71*, 4431-36.

107. Hakansson K, Emmett MR Hendrickson CL Marshall AG *Anal.Chem.* **2001,** *73*, 3605-10.

108. J.Zaia *Mass Spectrom.Rev.* **2004,** *23*, 161-227.

109. Yuji Sato, Minoru Suzuki Takashi Nirasawa Akemi Suzuki and Tamao End *Anal.Chem.* **2000,** *72*, 1207-16.

110. Harvey, David J. *Journal of Chromatography A* **1996,** *720*, 429-46.

111. Y.Zhao; S.B.H.Kent; B.T.Chait *Proc.Natl.Acad.Sci.USA* **1997,** *94*, 1629-33.

112. D.J.Harvey; R.H.Bateman; M.R.Green *J.Mass Spectrom.* **1997,** *32*, 167-87.

113. D.J.Harvey; R.H.Bateman; R.S.Bordoli *Rapid Commun.Mass Spectrom.* **2000,** *14*, 2135-42.

114. M.Claeys; H.Van den Heuvel; S.Chen; et al *J.Am.Soc.Mass Spectrom.* **1996,** *7*, 173-81.

115. Linsley, K. B.; Chan, S. Y.; Chan, S.; Reinhold, B. B.; Lisi, P. J.; Reinhold, V. N. *Analytical Biochemistry* **1994,** *219*, 207-17.

116. B.Domon; C.E.Costello *Biochemistry* **1988,** *27*, 1534-43.

117. B.L.Cillece-Castro; A.L.Burlingame *Methods Enzymol* **1990,** *193*, 689-712.

118. E.Stepens; S.L.Maslen; L.G.Green; D.H.Williams *Anal.Chem.* **2004,** *76*, 2343-54.

119. Willats WG *Plant Mol Biol* **2001,** *(47)*, 9-27.

120. Albersheim, P. *Plant Biochemistry*, Bonner and Varner eds, Ed.; 3 ed.; Academic Press: New York, 1976.

121. Carpita, N. C.; Gibeaut, D. M. *The Plant Journal* **1993,** *3*, 1-30.

122. Malcolm A.O'Neill, Dennis Warrenfeltz Keith Kates Patrice Pellerin Thierry Doco Alan G. Darvill and Peter Albersheim *THE JOURNAL OF BIOLOGICAL CHEMISTRY* **2004,** *Vol. 271, September 13, 1996*, 22923-30.

123. Masaru Kobayashi, Hironobu Nakagawa Tomoyuki Asaka and Toru Matoh *Plant Physiol* **1999,** *119*, 199-204.

124. Brent L.Ridley *Phytochemistry* **2001,** *2001 vol. 57 (dec)*, 929-67.

125. Carpita, N. C.; Gibeaut, D. M. *The Plant Journal* **1993,** *3*, 1-30.

126. O'Neill, M. A. A. P. D. A. *Methods in Plant Biochemistry, Vol. 2*, Dey, D. M., Ed.; Academic Press: London, 1990.

127. Mazeau, K. Perez S. *carbohydr.Res.* **1998,** *311*, 203-17.

128. Matsuhai, S. Nishikawa N. Negishi T. Hatanaka C. *J.Liquid Chromatoge.* **1993,** *16*, 3203-15.

129. Tateishi, A.; Inoue, H.; Shiba, H.; Yamaki, S. *Plant and Cell Physiology* **2001,** *42*, 492-98.

130. Wakabayashi, K. *Journal of Plant Research* **2000,** *113*, 231-37.

131. Lamikanra, O.; Juaraez, B.; Watson, M. A.; Richard, O. A. *Journal of the Science of Food and Agriculture* **2003,** *83*, 702-08.

132. Wang, Z. Y.; Macrae, E. A.; Wright, M. A.; Bolitho, K. M.; Ross, G. S.; Atkinson, R. G. *Plant Molecular Biology* **2000,** *42*, 317-28.

133. Zhou, T.; Xu, S. Y.; Sun, D. W.; Wang, Z. *Journal of Food Engineering* **2002,** *54*, 17-22.

134. Tateishi, A.; Inoue, H.; Shiba, H.; Yamaki, S. *Plant and Cell Physiology* **2001,** *42*, 492-98.

135. Wakabayashi, K. *Journal of Plant Research* **2000,** *113*, 231-37.

136. Lamikanra, O.; Juaraez, B.; Watson, M. A.; Richard, O. A. *Journal of the Science of Food and Agriculture* **2003,** *83*, 702-08.

137. Wang, Z. Y.; Macrae, E. A.; Wright, M. A.; Bolitho, K. M.; Ross, G. S.; Atkinson, R. G. *Plant Molecular Biology* **2000,** *42*, 317-28.

138. Zhou, T.; Xu, S. Y.; Sun, D. W.; Wang, Z. *Journal of Food Engineering* **2002,** *54*, 17-22.

139. English, P. D. Maglothin A. Keegstra K. and Albersheim P. *Plant Physiol* **1972,** *49*, 293-97.

140. D'Ovidio, R.; Mattei, B.; Roberti, S.; Bellincampi, D. *Biochimica et Biophysica Acta-Proteins and Proteomics* **2004,** *1696*, 237-44.

141. Karr, A. L. Albersheim P. *Plant Physiol.* **1970,** *46*, 69-78.

142. Parenicova, L. Benen. J. Kester H. Visser J. *Eur.J.Biochem.* **1998,** *251*, 72-80.

143. Tarchevsky, I. A. *Applied Biochemistry and Microbiology* **2001,** *37*, 441-55.

144. Cook BJ, Clay RP Bergmann CW Albersheim P Darvill AG *Mol Plant-Microbe Interact* **1999,** *12*, 703-11.

145. CF, Hahn MG Pl*ant Mol Biol 2*0**04, 26**, 1379-411.
146. Mathieu Y, Kurdjian A Xia H Guern J Koller A Spiro M-D O'Neill M Albersheim P Darvill A Pl*ant J 19***91, 1,** 333-43.

147. Lang, C.; Dornenburg, H. Ap*pl.Microbiol.Biotechnol. 2*0**00, 53**, 366-75.

148. M.Shieh, R. L. Brown M. P. Whitehead J. W. Carey P. J. Cotty T. E. Cleveland and R. A. Dean Ap*pl.Environ.Microbiol. 19***97, 63**, 3548-52.

149. A.ten Have, W. Mulder J. Visser and J. A. L. van Kan Mo*l.Plant-Microbe Interact. 19***98, 11**, 1009-16.

150. B.Oeser, etc Mo*lecular Biology of Fungal Development,* H.D.Osiewacz, Ed.; Marcel Dekker: New York , 2002.

151. Bussink, H. J.; Buxton, F. P.; Fraaye, B. A.; de Graaff, L. H.; Visser, J. Eur*.J Biochem. 19***92, 20***8,** 83-90.

152. Parenicova, L.; Benen, J. A. E.; Kester, H. C. M.; Visser, J. Bio*chemical Journal 2*0**00, 34***5,** 637-44.

153. de Vries, R. P.; Jansen, J.; Aguilar, G.; Parenicova, L.; Joosten, V.; Wulfert, F.; Benen, J. A. E.; Visser, J. Fe*bs Letters 2*0**02, 53***0,** 41-47.

154. Colangelo, J.; Orlando, R. Ra*pid Commun.Mass Spectrom. 2*0**01, 15**, 2284-89.

155. Colangelo, J.; Licon, V.; Benen, J.; Visser, J.; Bergmann, C.; Orlando, R. Ra*pid Commun.Mass Spectrom. 19***99, 13**, 1448-53.

156. Yang, Y.; Bergmann, C.; Benen, J.; Orlando, R. Ra*pid Commun.Mass Spectrom. 19***97, 11**, 1257-62.

157. Biely, P.; Benen, J.; Heinrichova, K.; Kester, H. C.; Visser, J. FE*BS Lett. 19***96, 38***2,** 249-55.

158. Kittur, F. S.; Vishu Kumar, A. B.; Tharanathan, R. N. Ca*rbohydr.Res.* 2**03, 33***8,* 1283-90.

159. Ishii, T.; Ichita, J.; Matsue, H.; Ono, H.; Maeda, I. Ca*rbohydr.Res.* 2**02, 33***7,* 1023-32.

Chapter 3. Glycosylation Site Mapping and Carbohydrate Structure Elucidation of Recombinant Endopolygalacturonase 3 and 6 from *Botrytis Cinerea* [1]

[1] Xie M., Bergmann C., Benen J. and Orlando R.
Submitted to Rapid Commun. Mass Spectrom.

Abstract

The fungus *Botrytis cinerea* is a ubiquitous plant pathogen that infects more than 200 different plant species and causes substantial economic losses in a wide range of agricultural crops and harvested products. Endopolygalacturonases (EPGs) are among the first array of cell wall degrading enzymes secreted by the fungus in the infection process. Up to 13 EPG isozymes have been described for *B. cinerea*. The presence of multiple *N*-linked glycosylation signals in six of them (BcPG1~6) were predicted by their amino acid sequences as deduced from their genes. In this work, the glycosylation sites and the attached oligosaccharide structures on BcPG3 and BcPG6 were analyzed. The molecular masses of both proteins were determined by matrix-assisted laser desorption/ ionization time-of-flight mass spectrometric (MALDI-TOF MS) analysis. The intact BcPG3 has a molecular weight of approximately 49320 Da, about 6050 Da higher than the corresponding protein backbone. And BcPG6 has a molecular weight of approximately 46730 Da, about 10900 Da higher than that of the protein backbone. The number of potential *N*-linked glycosylation sites present on both proteins was determined from their protein sequences to be 4 and 7 respectively. Occupancy of these glycosylation sites and the attached carbohydrate structures were analyzed by tryptic digestion followed by liquid chromatography-electrospray ionization mass spectrometry (LC-ESI-MS) using stepped orifice potential techniques applied at the nano ESI source. Three high-mannose type sugar chains were present in BcPG3 tryptic digested peptide mixtures. The glycosylated asparagines were determined to be Asn307 (T11 peptide), Asn341 (T13 peptide) and Asn351 (T14). No sugar chain was found attached to the potential *N*-linked site of Asn112. Five out of seven potential N-linked sites present in BcPG6 were determined to be occupied by high-mannose type oligosaccharides. Four of them were readily determined to be attached at Asn237 (T9 peptide),

Asn146 (T6 peptide), Asn198 (T7 peptide) and Asn256 (T11 peptide), respectively. Another was located at the T8 peptide, which contained two potential N-linked sites, Asn224 and Asn227, side by side (SNNN^{224}VTN^{227}ITFK). The glycosylation site in this peptide was determined to be at Asn224 rather the Asn227 by LC-ESI-MS/MS of an N-glycanase treated sample. Asn 58 on T3 was found to not be glycosylated.

Introduction:

The fungus *Botrytis cinerea* is a wide spread plant pathogen that affects many vegetables and fruits, as well as a large number of shrubs, trees, flowers, and weeds by causing grey mould or soft rot on the host tissues [1;2]. *B. cinerea* mainly infects dicotyledons and non-graminaceous monocotyledons, and the infection not only occurs during the growth phase of crops but also during storage of the harvested products[3;4]. Thus, it is responsible for significant economic losses. *B. cinerea* is a necrotrophic pathogen[2] and feeds on dead plant material, it can be envisaged that plant cell death is not as effective against this fungus as it is against a biotroph. Under certain circumstances, *B. cinerea* is able to grow through the thick layer of dead cells and fortified cell walls after a seemingly latent period. By then the fungus grows so vigorously that defense responses from the plant seem to be ineffective. The lesion will expand rapidly resulting in infestation of the complete organ or whole plant.

A great deal of research has focused on understanding the *B.cinerea* infection process[5-7], which in turn directs the development of effective prevention strategies. Histological studies on various plants suggest that the mechanism of *B. cinerea* infection differs depending on the hosts. However, the initial stage of infection often starts with the degradation of the cell wall by an array of cell-wall-degrading enzymes [8].

B. cinerea encounters many cell wall components during the infection process and accordingly secretes a great number of enzymes that can degrade cell walls. Among these cell-wall-degrading enzymes, endopolygalacturonases (EC 3.2.1.15) (EPGs) are in the first array of enzymes secreted. EndoPGs catalyze the hydrolysis of the de-esterified cell wall polygalacturonic acid via cleavage of the α-1, 4-glycosidic linkage. For *B. cinerea,* 13 EPG isozymes have been described[9] although the possible functions for each in the pathogenesis of *B. cinerea* are not fully elucidated. However, a clear association is observed between the expression of some *B. cinerea* EPG isozymes and soft-rot symptoms in the host [10]. The mutants produced from targeted replacement of the *Bcpg1* gene showed a reduction in pathogenicity of *B. cinerea* on apple fruit and tomato leaves[11]. A recent study of the chemically treated endopolygalacturonase-1 protein from *B. cinerea* strain T4 indicated that this protein can also trigger plant defense responses, and certain extracellular proteins may be secreted to inhibit its activity[12]. The results then suggest that the EPG1 protein from *B. cinerea* can be both a virulence and avirulence factor. The expression pattern of several other *B. cinerea* genes encoding EPG were also studied and the results indicate that each isozyme is differentially expressed depending on the stage of infection, the host [13], the temperature and pH value of the medium.

EPGs together with other pectinases and cellulases, are an essential part of the fruit juice industry and are used to help extract, clarify and modify juices. The enzymes used to date have generally been crude preparations, but the demand for large quantities of purer enzymes with specific, well-defined and predictable properties is increasing greatly. While improvements in enzyme purification are important, the focus has turned to DNA recombinant technology using heterologous host organisms and growth media. It has been well established that changing the

cell culture environment and the cell type in which a protein is produced may change protein post-translational modifications. This is especially true for the degree of glycosylation and the attached oligosaccharide structures[14]. The attachment of carbohydrate moieties to the protein via either an asparagine residue (*N*-linked) or serine/threonine residue (*O*-linked) has been demonstrated to be a most important post-translational modification. Glycosylation affects the biological properties of proteins, such as protein half-life, targeting, cell-cell and cell-matrix interactions. Therefore, it is important to fully characterize the glycosylation on a recombinant pectinase to ensure its properties have not been changed unintentionally, as well as to gain regulatory approval for it being used in the food industry.

EPG genes from *B. cinerea* strain BC-B05.10 were overexpressed in Pichia. Several recombinant EPGs (BcPG 1-6) were produced and their amino acid sequences were deduced from the predicted open reading frames present in the genomic sequences. Multiple *N*-linked glycosylation signals were present in all of them (BcPG1-6), which suggests that those enzymes might be produced as glycosylated proteins, as observed for pectinases produced from fungi *A. Niger, A. nidulans and S. purpuraum*. For example, glycosylation was observed on several homologously expressed EPGs from *A. niger*. *A. niger* EPG I and II both contain *N*-linked oligosaccharide chains with high mannose type structures[15;16]. Other *A. niger* EPGs, PGC and PGA, exhibit high mannose *N*-linked oligosaccharides as well as *O*-linked mannoses[15;17]. EPG I from *S. purpuraum* is a mixture of three components that contain different numbers of high mannose type oligosaccharide chains as well as disulfide bonds[18].

In this work, the glycosylation sites and the attached oligosaccharide structures of two recombinant *B. cinerea* EPGs, BcPG3 and BcPG6 were analyzed. The purified proteins were first analyzed by gel electrophoresis and both showed broad bands on the 12% tris-glycine gel.

Glycoproteins normally show broad bands on SDS-PAGE due to their multiple isoforms (glycoforms). This also suggested that both BcPG3 and BcPG6 were most probably glycoproteins. The molecular masses of both proteins were identified by Matrix Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-MS), and both have much higher molecular masses as compared to that calculated from amino acid sequences, again indicating that post-translational modification occurred on both proteins. To determine the presence of glycosylation, endoglycosidase digestion was performed on both proteins and the molecular weights of the resulting products were again examined by MALDI-MS to determine the difference, if any, resulting from the digestion. The location of glycosylation sites was determined from the Liquid Chromatography/Mass Spectrometric (LC-MS) analysis of the trypsin digested proteins. To selectively detect the glycopeptides, a stepped-orifice voltage technique was applied at the Electrospray Ionization (ESI) source. The oligosaccharide structures were then deduced from LC/MS data.

Experimental Materials

The overexpressed EPG 3 and 6 (BcPG3 and BcPG6) from *B.cinerea* strain BC-B05.10 were the kind gift from J. Benen. The purified enzymes were assayed for their activity in the medium and their purity was tested by SDS-PAGE. Purified enzymes with the concentration of 3.7mg/mL for PG3 and 3.8mg/mL for PG6 were stored in 50mM sodium acetate buffer (pH=5.0) under -20°C. Sodium azide (0.01% w/v) was added as an antimicrobial agent.

Sequencing grade trypsin was purchased from Promega (Madsion, WI, USA). 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid) was purchased from Sigma (St. Louis, MO, USA). Recombinant N-glycanase (PNGase-F) from *Chryseobacterium [Flavobacterium]*

menigosepticum was purchased from Prozyme (San Leandro, CA, USA). Recombinant Endoglycosidase H (Endo H) from Streptomyces plicatus (overexpressed in E.coli) was purchased from New England Biolabs (Beverly, MA, USA). All other chemicals were purchased from Fisher (Pittsburg, PA, USA)

MALDI-MS

All MALDI-MS experiments were performed on a Hewlett-Packard (Palo Alto, CA, USA) G2025A time-of-flight mass spectrometer. The instrument was equipped with a nitrogen laser source (λ=337nm), and operated at a pressure below 1×10^{-6} Torr. The data were collected in linear mode, and the accelerating voltage and extractor voltage were 28kV and 7kV respectively. Two sinapinic acid matrix solutions were prepared. One was prepared by dissolving sinapinic acid in a solution of 40% (v/v) acetone in water to a concentration of 10mg/mL and the other was 4mg/mL sinapinic acid in a solution of 80% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid (TFA) in water. Samples were either mixed with the matrix solution and vacuum dried on the MALDI target, or applied on target by a two-layer deposition method. In the latter case, a 10mg/mL matrix solution was first applied on the target and allowed to dry, and samples were then added on top of the first layer. Once the samples were semidry, another layer of 4mg/mL matrix solution was applied, followed by vacuum drying. A mixture of known protein standards was used to calibrate the instrument externally.

Modified Enzyme-Linked Immunosorbent Assay (ELISA)

DIG Glycan differentiation kit (Roche Molecular Biochemical Inc., Mannheim, Germany) was used to detect the carbohydrate present on PGC. The manufacture's protocol was followed

with the following modifications. An ELISA plate was used to bind the proteins. 0.1µg recombinant PGC diluted to 150ul in ELISA coating buffer as well as the positive controls (diluted standards provided in the kit) and a negative control (ELISA coating buffer) were loaded onto different spots on the plates and incubated at 4ºC for 24 hr. 250ul of blocking buffer (1% BSA in TBS) were loaded into sample wells afterwards, and the plate was incubated at room temperature for 3 hr to ensure complete blocking and eliminate non-specific binding. Five digoxigenin labeled lectins with different terminal carbohydrate residue binding specificities were used, Galanthus nivalis agglutinin (GNA), Sambucus nigra agglutinin (SNA), Maackia amurensis agglutinin (MAA), Peanut agglutinin (PNA) and Datura stramonium agglutinin (DSA). After removing the unbound lectins by extensive washing, 200µL of alkaline phosphate labeled anti-digoxigenin solution were added to form the carbohydrate-lectin-antilectin complex, which was detected using a colorimetric reaction by adding 200µL of 1mg/mL para-nitrophenylphosphate (PNPP) solution.

Endoglycosidase H digestion of BcPG3 and BcPG6 proteins

10ug of each recombinant protein was first dissolved in 50mM sodium citrate buffer, and then recombinant Endo-H was added to the reaction buffer at a ratio of 1:20 (w/w) enzyme to substrate. The solution was incubated for 24h at 37°C with gentle shaking. The enzyme was deactivated by adding 0.1%TFA to lower the pH. Endo H cleaves between the two *N*-acetylglucosamine (GlcNAc) residues within the core structure of an *N*-linked high mannose oligosaccharide. After digestion, the deglycosylated proteins were analyzed by MALDI-MS.

Trypsin digestion of BcPG3 and BcPG6 proteins

30µg of each recombinant protein was first dried and then redissolved in 30µL of 20mM ammonia biocarbonate, pH 7.8. Digestion was carried out with sequencing grade trypsin using a ratio of 1:30 (w/w) enzyme to substrate for 24 hours at 37°C with gentle shaking. An aliquot of peptide mixture was further digested using PNGase-F (see below) and the rest was analyzed by MALDI-MS, LC/MS (nanoESI and ESI) and LC/MS-MS.

LC nanoESI /MS and MS-MS analysis of trypsin digested BcPG3 and BcPG6 peptides

Capillary HPLC/ESI-MS experiments were performed with a hybrid quadrupole/time-of-flight mass spectrometer (QTOF-2) equipped with a Z-spray ESI source (Micromass, Manchester, UK). Glycopeptides were selectively detected by selected ion monitoring for sugar oxonium ions. Separation of peptides was performed using a Waters® CapLC^{TM} system (Model 920, Milford, MA, USA) set to deliver mobile phases at 1 µL/min. 5 µL of the trypsin digested peptide mixtures were injected onto a C18 capillary column (150mm length, 150µm inner diameter) directly interfaced with the QTOF-2, and separation was achieved using mobile phases of 0.1% (v/v) formic acid in milli-Q water as A, and 0.085% (v/v) formic acid in methanol as B. Prior to injection, the column was equilibrated with buffer A containing 5% buffer B. The gradient started 10 min after injection for desalting, and increased linearly from 5% to 60% buffer B over 60 min. The column effluent was directly analyzed by the mass spectrometer in the positive ion mode.

The QTOF-2 mass spectrometer was operated alternatively in both normal orifice voltage and high orifice voltage modes. In the latter experiment, the orifice voltage was set at a value of 80V to generate "in-source" fragmentation and produce carbohydrate oxonium ions such as *m/z*

163 (hexose) and 204 (*N*-acetylhexosamine). The corresponding mass range of this function was set at m/z 50-400. Another function with an m/z range of 400-2500 was also acquired at a normal orifice voltage, which was set at a value of 35V to avoid "in-source" fragmentation. The instrument was set to collect data for both functions by switching from one to another scan to scan, with a scan duration of 1 s for each. The glycopeptides were identified in the low energy condition based on the detection of the carbohydrate fragments generated during the high energy condition. Collisional-induced dissociation (CID) MS/MS was performed as described later. The ESI voltage was set at 3100 V and the desolvation temperature was 225°C. The instrument was externally calibrated using 1.5pmole Glu^1-Fibrino-peptide B prior to sample analysis.

LC ESI/MS analysis of trypsin digested BcPG3

Approximately 10 µg of a tryptic digest of BcPG3 was separated by a C18 microbore column (1mm X 150mm, 5µm particle) interfaced with the QTOF-2. The buffers were delivered by an Agilent 1100 HPLC system (Palo Alto, California) at a flow rate of 50µl/min. The nanospray source on QTOF-2 was replaced by a conventional ESI source to accommodate the high flow rate and the instrument parameters were optimized accordingly. A gradient started with 10 min desalting (5% B) followed by increasing buffer B from 5% - 70% B over 70min. During LC-MS analysis, no UV detector was used. The QTOF-2 mass spectrometer was operated in the positive MS mode using 10 V of collisional energy and 35 V of orifice voltage for a scan function with full mass range of m/z 250-2500.

PNGase-F digestion of tryptic BcPG6 peptide mixuture

5μL of the trypsin digested BCPG6 peptide mixture was diluted in phosphate buffer (50mM sodium phosphate, pH 7.5) and further digested with recombinant PNGase-F using a ratio of 1:20 (w/w) enzyme to substrate for 24h at 37°C with gentle shaking. The doubly digested mixture was analyzed by LC/MS-MS.

LC/MS-MS analysis of tryptic and PNGase-F digested peptides

LC/MS-MS experiments were carried out with the same instrument settings as mentioned earlier utilizing a C18 capillary column interfaced to the QTOF-2. The tryptic peptide mixtures as well as the BcPG6 doubly digested (trypsin and PNGase-F) peptide mixture were separated by increasing buffer B from 10% to 70% over 60 min at a flow rate of 1 μL/min. CID-MS/MS was carried out in the positive ion mode with argon as the collision gas. A serial of collisional energies were used depending on the charge states and m/z ratio of the parent ions under investigation in order to achieve the optimum degree of fragmentation for them. Precursor selection was based upon a threshold of 7 counts/s and up to four precursor ions were selected for CID acquisitions of 2 seconds each during each MS acquisition. Data was acquired from m/z 400-2200 for MS mode (precursor scan) and m/z 50-2500 for MS/MS mode (product ion scan).

Results and Discussion

The recombinant BcPG3 and BcPG6 proteins from B. Cinerea were first analyzed by gel electrophoresis to determine their purity. Both proteins showed similar migration distances on a 12% Tris-Glycine gel with a broader band for the BcPG6 sample (Figure 3.1). More accurate molecular masses of both proteins were obtained by MALDI-MS analysis and the mass spectra revealed that molecular masses for BcPG3 and BcPG6 were approximately 49320 Da and 46730 Da, respectively (Fig. 3.2a and Fig. 3.3a). However, the calculated average masses based on the mature amino acid sequences were much less than the experimental masses (43262 And 35824 Da). The molecular mass difference for BcPG3 was approximately 6050 Da, and for BcPG6, the difference was about 10900 Da. The mass difference indicated post-translational modifications occurred on both proteins, and a such large mass discrepancy suggested that the post-translational modification was most probably present in the form of glycosylation. The weak signal for BcPG6 and broad peak suggested that extreme micro-heterogeneity exists on this protein.

The attachments of carbohydrate moieties to proteins are either through the amide on the side chain of asparagine residues (*N*-linked glycosylation) or through the hydroxyl group on the side chain of serine and/or threonine residues (*O*-linked glycosylation). The *N*-linked glycans share the same pentasaccharide core $Man_3GlcNAc_2$, and are normally found as extended structures of this core within a heterogeneous population. For this reason, glycosylation typically results in larger increments in protein molecular mass than other post-translational modifications, which usually result in molecular mass increases of less than 200 Da per site. A tripeptide consensus sequence (sequon) of Asn-X-Ser/Thr (X is any amino acid except proline) has to be present in order for *N*-linked glycosylation to occur. However, the consensus sequence

itself doesn't guarantee the presence of *N*-linked glycosylation, thus any Asn residue located within such a sequon is commonly known as a potential *N*-linked glycosylation site. An examination of BcPG3 and BcPG6 amino acid sequences indicated four potential *N*-linked glycosylation sites exist within BcPG3 and seven within BcPG6. *O*-linked glycosylation doesn't require any special amino acid sequon, thus it could occur on any Ser and Thr residues.

The presence of glycosylation on both BcPG3 and BcPG6 was determined with a modified ELISA method using five lectins. Four standard glycopeptide controls and one negative control all reacted appropriately. Only one lectin, GNA, reacted with BcPG3 and BcPG6. GNA is specific for terminal mannose residues and its positive reaction suggested the presence of high mannose type *N*-linked glycans. These results indicated that high mannose type *N*-linked glycans were attached to both BcPG3 and BcPG6.

To analyze the heterogeneity of glycosylation, both proteins were treated with an *N*-glycosidase and the resulting products were analyzed by MALDI-MS and 12% Tris-Glycine gel electrophoresis, respectively. PNGase-F and Endo-H were chosen as enzymes for digestion. PNGase-F cleaves the entire N-glycan from the Asn residue, while Endo-H cleaves the two core GlcNAc residues in high-mannose type and some hybrid type *N*-linked oligosaccharides. PNGase-F treated proteins were analyzed by gel electrophoresis (Figure 3.1), while MALDI-MS was used to analyze Endo-H treated ones (Figure 3.2b and 3.3b). Both experiments showed a decrease in molecular mass for BcPG3 and BcPG6 proteins. Based on MALDI-MS data, the mass of BcPG3 decreased from 49320 Da to 44830 Da (4490 Da deduction), while the mass decrease for BcPG6 was approximately 9550 Da (from 46730 Da to 37180 Da). Both data were in good agreement with gel electrophoresis results (Fig. 3.1). This indicated that both proteins

are heavily *N*-glycosylated. This is confirmed by the obvious reduction in band-width observed with gel electrophoresis.

To determine the number of glycans as well as their sites of attachment, both proteins were subjected to trypsin digestion and the resulting peptide mixtures were separated by reverse phase HPLC followed by nanoESI-MS analysis. The stepped orifice voltage method was used to selectively identify glycopeptides. Accordingly, mixed scan functions were used to promote the "in-source" fragmentation of carbohydrate oxonium ions under high orifice conditions, and the mass spectra for the intact glycopeptides were obtained by summing the carbohydrate-containing peaks from the TIC acquired under low orifice voltage, in which the energy applied to the peptides was not sufficient to promote any fragmentation in the orifice region. The mass spectrometer was set to monitor the m/z values within the range of the specific carbohydrate oxomium ions (m/z 50-400), or the range of intact peptides ions (m/z 400-2500). The retention times of the intact glycopeptides in the total ion chromatograms (TIC) (obtained from scans of intact peptide) were determined based on that of the oxonium ions, which were obtained by extracting their ion currents from another TIC (obtained from low cone voltage condition). Therefore, glycopeptides present in the tryptic mixture were selectively detected. The analysis of trypsin digested BcPG3 and BcPG6 is shown in Figure 3.4 and 3.5. The ion currents for *m/z* 163 from both proteins revealed elution profiles of multiple *N*-linked glycopeptide-containing fractions. It is noteworthy that false positives may occur when using carbohydrate oxonium ions to selectively identify glycopeptides[19]. To further confirm the presence of the glycopeptides, two more oxonium ions currents, *m/z* 204 (HexNAc) and *m/z* 366 (GlcNAc-hexose), were extracted along with *m/z* 163. As a result, several possible false positives were rejected for both proteins. However, even after the removal of false positives, the numbers of oxonium ion peaks

are still more than the number of potential *N*-linked sites. This could be explained by heterogeneity within the glycoforms, in which a variety of sugar chains attached to the same N-linked glycosylation site results in different retention times for the glycoforms on the C18 column.

The first carbohydrate-containing fraction from BcPG6 protein eluted from the column at 19.5 - 21 min. The corresponding summed spectrum showed a series of peaks spaced by *m/z* 81, which indicated that the mass difference between the adjacent peaks is 162 Da. This is a character of high mannose type *N*-linked glycans since high mannose type contains only mannoses residues bound to the core structure, thus the 162 Da mass difference indicates the difference of one mannose residue. On the other hand, the monosaccharide residues presented at the other two types include sialic acids and GlcNAc, which will result in different peak spacing other than m/z 162 (or the m/z values corresponding to the multiply charged ions). One peak in the spectrum was a singly charged ion at m/z 1247.7, which equals the T9 peptide backbone molecular mass (1246.6 Da) (Figure 3.6a). In addition, there is one potential N-linked glycosylation site within T9 peptide sequence: NSELVN^{237}SSNGAR. Therefore, the observed series of glycopeptides were assumed to be a group of glycoforms (with difference number of mannose residues) attached to peptide T9 at Asn 237 position. To confirm this hypothesis, the most abundant glycopeptide peak at m/z 1638.2 (2+) was analyzed. First, its corresponding molecular mass was calculated as 3274.4 Da. Then this calculated mass was compared with the peptide masses obtained from *in silico* digestion performed by Masslynx software (version 3.5) and no match was found. This result further indicated that the presence of glycosylation. The difference between the mass of this glycoform (3274.4 Da) and that of T9 peptide backbone (1246.2 Da) was calculated to be 2028.2, which is corresponded to the mass of the attached

glycan. The glycan composition of high mannose structures follows a simple molecular formula: $GlcNAc_2Man_x$, and according to the subtracted mass of 2028.2 Da, x was calculated to be 10. Therefore, the ion with m/z 1638.2 was determined to be a glycosylated T9 peptide attached to a high mannose structure with the composition of $GlcNAc_2Man_{10}$.

Additional confirmation of the structural identity of this glycan and its attachment site was obtained by LC-CID-MS/MS on the precursor ion of m/z 1638.2. The CID was performed by a QTOF-2 instrument, in which, the precursor ions were selected in Q1 and then fragmented in the collision cell (a RF only quadrupole), and the resulting fragment ions were analyzed by the reflectron TOF analyzer. The MS/MS data of a glycopeptide contains different features from those of a peptide, which are dominated by the fragment ions generated from peptide backbone cleavage. Figure 3.6b showed the MS/MS spectrum for the precursor ions with m/z 1638.2 (2+). It clearly indicated that carbohydrate oxonium ions are dominant in the low mass region, which are B ions (based on Domon and Costello nomenclature[20]) resulting from the cleavage of the glycosidic linkage within the oligosaccharide chain. A series of doubly charged Y ions were observed within the high mass region with peak spacing of either 81 or 101.6. These mass shifts indicate the sequential loss of ten mannose residues followed by one GlcNAc residue from the glycopeptide backbone. The formation of such Y ions is favored for glycopeptides because the glycosidic bond is a labile linkage, and as such, it is more readily cleaved than the peptide backbone. The loss of mannoses confirmed the presence of a high mannose structure and the GlcNAc residue most likely was previously presented within the core structure. No further loss of GlcNAc residue was observed, which indicated that the other GlcNAc residue belonging to the core structure should still be attached to peptide backbone and since it is the one directly linked to Asn, it is expected to be more stable than the one next to it during the CID. The

fragment ion corresponding to the peptide attached with one GlcNAc residue was showed in the spectrum as the peak at m/z 725.7 (1450.4 Da), and the resulting mass from subtracting a mass of 204, which is 1246.4 Da, should be corresponded to the peptide backbone and it indeed is match with T9 peptide. Therefore, T9 peptide was determined as the peptide attached by the glycans at Asn237. In addition, the observed 204.2 Da mass difference along with the loss of ten mannoses and one GlcNAc completed the information needed for structure determination of the attached glycan, which obviously is $GlcNAc_2Man_{10}$, same as the one calculated from the LC-MS data mentioned earlier. The same conclusion from both MS and CID MS/MS data suggests that MS data is sufficient to determine a high mannose structure present at a peptide.

Other glycoforms observed within this series were also analyzed in the similar manner, and the results lead to the same conclusion, that is, T9 peptide is attached to a group of high mannose structures with varied numbers of mannose.

Another group of glycopeptides were identified from the effluent corresponding to ion peaks at 24 - 25.5 min and the summed spectrum was shown in Figure 3.7a. One series of peaks with a spacing of m/z 54 was observed, while another series of peaks were spaced *m/z* 81 apart. Such 162 Da mass difference is the same as seen from T9 glycoforms discussed earlier, which is a good indication for the presence of high mannose structure. As the predominant peak within the doubly charged ion series, m/z 1639.8 corresponded to a calculated molecular mass of 3277.6 Da, which was also seen as a triply charged ion at m/z 1093.5. This suggests that both series of ions originated from the same peptide. The molecular mass of 3277.6 was first compared to the peptide masses obtained from *in silico* digestion and didn't find a match. Then, it was compared specifically to the backbone masses of peptides containing potential *N*-linked glycosylation sites (apart T9). Among the four potential glycopeptides, the backbone masses for T3 (1689.8), T8

(1250.6) and T11 (1975.0) are below 3277.6 Da, which are then used for the following data analysis. First, the mass of the peptide backbone was subtracted from the mass of observed glycoform (3277.6 Da), and the resulting mass was then used to check whether fit in the molecular formula of $GlcNAc_2Man_x$, where x has to be an integer. After calculating, only the mass difference between the T8 peptide and the glycoform provided a valid molecular formula of $GlcNAc_2Man_{10}$. Therefore, the ion with m/z 1639.8 was determined as T8 peptide attached with a high mannose structure containing ten mannose residues. This result was further confirmed by LC-CID-MS/MS analysis of T8 glycopeptide ions (data not show). Investigation of the peptide backbone sequence indicates that T8 contains two potential N-linked glycosylation sites (SNN**N**VT**N**IT**F**K). However, the glycan composition result reveals the presence of only two GlcNAc residues, which suggested that there is one *N*-linked glycan core structure attached to T8 peptide and any one of the two sites could be occupied. As the loss of carbohydrate residues during the CID MS/MS experiment occurs prior to the glycopeptide backbone fragmentation, it is impossible to determine the glycosylation site from CID MS/MS.

To determine the glycosylation site of T8 peptide, i.e. to reveal whether the oligosaccharide is attached to either the NIT or the NVT segment, the tryptic peptide pool was first subjected to PNGase-F digestion. PNGase-F cleaves the glycosidic linkage between the *N*-linked oligosaccharide and the Asn residue, and converts the Asn residue to an Asp. Thus, the molecular mass of the glycopeptide backbone will increase by 1 Da for every *N*-linked glycosylation site. For T8 peptide, the resulting molecular ion is expected to have a mass of 1252.6 Da since only one glycosylation site is present. The de-glycosylated peptide sequence was then analyzed by LC-CID MS/MS to identify the emerging Asp residue.

During each MS scan, four precursor ions were selected for CID. The data obtained from MS scan contain the information of the precursor ions, from which, the ion with the expected mass of 1252.6 Da can be extracted and its spectrum is shown in Figure 3.7b. A singly charged precursor ion with *m/z* 1252.6 and the corresponding doubly charged ion with *m/z* 626.8 were both observed, with a 1 Da increment as compared to the T8 peptide sequence ($(M+1)^+$ = 1251.6 Da). This confirmed that indeed, only one Asn residue out of the two potential sites was actually occupied. The detailed sequence information of this precursor ion (*m/z* 1252.6) was obtained from MS/MS scans (Figure 3.7c). Fragmentation of the peptide backbone produced the well-defined series of y and b ions, with several a and z ions. The assignment of y1-y7 ions was in excellent agreement with the T8 segment sequence of VTN^{227}ITFK. The y8 (m/z 937.57), y9 (m/z 1051.59) and y10 (m/z 1165.59) ions all shifted by 1 Da as compared to the T8 segment sequences. Since the Asp residue can only occur on either Asn 224 or Asn 227 position within the T8 sequence, the agreement of the y7 ion with the T8 segment sequence indicated that the Asn at position 227 did not undergo any change following the treatment with PNGase-F. On the other hand, the m/z for the y8 ion corresponding to segment N^{224}VTN^{227}ITFK increased by 1 Da, indicating that the Asn at position 224 was converted to Asp after treatment. This also explained the 1 Da increment for the y9 and y10 ions. This Asn224 attachment site was also confirmed from the B ions series. By combining these experiments on deglycosylated glycopeptide with profiling the released oligosaccharide with CID-MS/MS, we have been able to obtain both linkage sites as well as the oligosaccharide structure.

Other glycopeptides present in BcPG6 were also detected and the attached glycan structures were identified in the similar manner. Their detailed information is listed in Table 1 and their summed mass spectra are shown in Figures 3.8 and 3.9, respectively. Glycopeptides

corresponding to ion peaks both at 26-27 min and 27.5-28.5 min originate from a single peptide backbone T9-10 (one miscleavage from trypsin). Effluent at 26-27 min contains glycopeptides of T9-10 + $GlcNAc_2Man_{8-13}$ (Figure 3.8a), while effluent at 27.5 – 28.5 min contains glycopeptide of T9-10 + $GlcNAc_2Man_{7-11}$ (Figure 3.8b). As a rule, peptides containing shorter oligosaccharide chain are more hydrophobic than those with longer oligosaccharide chains, and thus have a longer retention time on the C18 column.

Figure 3.9a shows the summed mass spectrum of glycosylated T11 peptide, which was eluted from the C18 column at 32-33.5 min. The oligosaccharide compositions for the glycoforms were determined to be $GlcNAc_2Man_{9-13}$ in the similar manner used to study the glycans attached to T8 peptide. Another series of 4+ charged ions was found at the lower mass region in the same spectrum. The dominant peak was at m/z 974.5 (4+), which was then determined to be a T3 peptide with a high mannose $GlcNAc_2Man_{11}$ structure coeluting with the T11 glycopeptides.

Another potential *N*-linked glycosylation site is located at T6, which was not observed during LC/MS analysis due to its high molecular mass (7180.02 Da). To identify this peptide, the tryptic digested BPG6 peptide mixture was analyzed by MALDI-MS. A peak at 7178.9 Da was identified as the T6 peptide backbone, well within the 0.1% mass error which is typical for a linear MALDI instrument (data not shown). The presence of this peak indicated that the *N*-linked site on this peptide was vacant.

Utilizing the sugar oxonium ions to selectively identify glycopeptides has proven to be effective for the detection of *N*-linked glycopeptides. As seen from the above results, for the BcPG6 protein, five out of seven potential *N*-linked sites were determined to be occupied by high-mannose structures. The combined mass from all *N*-linked oligosaccharides was

calculated to be 10630 Da based on the most abundant glycoform present at each site. This value is in good agreement with the molecular mass difference between the experimental and calculated, which suggested that BcPG6 does not possess other post-translational modifications (PTMs). To further determine the existence of the *O*-linked glycopeptide, peptide fingerprint mapping was performed based on the precursor ion data acquired from CID-MS/MS experiments on the trypsin and PNGase-F doubly digested peptides. After removing the *N*-linked oligosaccharides, the resulting peptides T3, T7, T8, T9 and T11 were all observed with a 1 Da mass increase, which is expected and discussed earlier. The detected molecular masses for peptides T16 and T13 were 2 Da less than that determined from *in silico* digestion. And, both peptides contain two Cys residues, thus the loss of 2 Da resulted from the internal disulfide bonds formed between the Cys residues. All of the other tryptic peptides, except T6, were identified by LC-MS and T6 was identified from MALDI MS data mentioned above. The coverage of the entire sequence confirmed that no other PTMs are present on BcPG6. The *N*-linked glycosylations and the disulfide bonds present in the BcPG6 amino acid sequence are shown in Figure 3.10.

The glycosylation sites and the carbohydrate structure for BcPG3 peptide mixtures were also analyzed. Examination of the BcPG3 amino acid sequence revealed four potential *N*-linked glycosylation sites located at T5, T11, T13 and T14, respectively (Figure 3.11). The occupancy of these potential sites was identified using the stepped orifice voltage technique with the help of the marker ions (Figure 3.4) and the data were analyzed in the similar manner for BcPG6 sample.

A 3+ charged ion with *m/z* of 2189.3 was observed in the mass spectra obtained from low cone voltage condition at the retention time of 43.5-44.5 min, a region which contains no m/z 163 and m/z 204 oxonium ions. In addition, the molecular mass of this peptide was

calculated to be 6564.9 Da, matching the *in silico* digested molecular mass of the T5 peptide. This indicated that T5 peptide is not glycosylated. Thus, the potential *N*-linked glycosylation site present on T5 is vacant. Following the similar manner as studying the glycopeptides of BcPG6, the occupancy of the remaining three peptides (T11, T13 and T14) containing potential *N*-linked glycosylation sites are all confirmed from the LC-MS data, and they all contain high mannose structures. The detailed information of the glycoforms and the glycan composition are listed in Table 2.

Figure 3.12 showed two averaged mass spectra corresponding to the ion peaks from 36.5-37.7 min and 38-39.2 min, respectively. Both showed a series of peaks with spacing of *m/z* 54, indicating a high mannose structure. The predominant peak in Figure 3.12a was *m/z* 1700.4, corresponding to a calculated molecular mass of 5098.2 Da. It was then determined to be T14 + $GlcNAc_2Man_{11}$. In Figure 3.12b, this peak shifted to the higher mass region of *m/z* 1727.3, and the corresponding molecular mass was calculated to be 5178.9 Da with an 80.6 Da increment as compared to T14 + $GlcNAc_2Man_{11}$. This mass difference could result from phosphorylation (+ 79.98 Da) or sulfation (+ 80.06 Da). Phosphorylation usually occurs at Ser/Thr residues or, in rare cases, at Tyr residues, and the sulfation usually occurs at Tyr residues. Therefore, a total of six potential phosphorylation/sulfation sites (2 Ser, 2 Thr and 2 Tyr) are present on the T14 sequence (M<u>S</u>NI<u>S</u>D<u>Y</u>GIDVQQD<u>Y</u>LNGGP<u>T</u>GEP<u>T</u>NGVK). An attempt to deduce the attachment site by CID-MS/MS was not successful due to the size of the T14 peptide. However, the possibility of phosphorylation can not be ruled out and needs to be confirmed by further experiments.

The combined mass from all *N*-linked oligosaccharides was calculated as 5112 Da based on the most abundant glycoform present at each site, which is in good agreement with the results

obtained from the Endo-H digestion experiment, remembering that one GlcNAc residue is left on the Asn after Endo-H digestion at each of the three *N*-linked glycosylation sites. However, the molecular mass difference between the experimental and calculated data was 6050 Da. Therefore, while the *N*-linked glycans account for a majority of the mass discrepancy, the remaining ~940 Da mass difference suggests that BcPG3 may possess *O*-linked glycosylation because it is the only PTM besides *N*-linked glycosylation could provide large enough mass increase in the range of several hundreds. The *O*-linked glycans have comparatively shorter chain and simpler structure than *N*-linked glycans, but the analysis of *O*-glycosylation proven to be difficult than *N*-glycosylation. First of all, sites of *N*-linked glycosylation can be readily predicted from the amino acid sequence. In contrast, *O*-linked glycans can attach to any Ser/Thr residues. Also, the large size of the *N*-linked glycans result that they are well-separated from each other on the protein sequence, but the *O*-linked glycans are smaller in size and frequently occur in a short region of Ser/Thr rich peptide. In addition, the stereo chemical factor of *O*-linked glycans could be another reason to explain the difficulty to generate the carbohydrate oxonium ions by stepped orifice potential technique.

To investigate the presence of *O*-linked glycans, peptide mapping was first performed to locate the potential O-linked peptides. 14 out of 16 tryptic peptides were identified, with the exception of T16 (1317.5 Da) and T8 (7073.5 Da). The observed masses for T10 (m/z 1113.2 (4+)) and T15 (m/z 1101.2 (3+)) peptides were 2 Da less than the calculated ones (m/z 1112.7 (4+), 1101.9 (3+) respectively). Sequence data revealed that both peptides contain 2 Cys residues, indicating the presence of disulfide bonds. Thus, *O*-glycosylation should only be present on either T8 or T16.

One glycopeptide, identified by LC/MS using a conventional ESI source, was different from the *N*-linked glycopeptides detected. The retention time of this peptide was 10-13 min, much earlier than the rest of the peptides (30-60 min). Figure 3.13 shows the summed mass spectrum corresponding to this peptide. Similar to what was seen for the *N*-linked glycopeptides, a series of peaks with an 81 Da peak spacing was observed. The predominant peak within this series was at m/z 658.4, corresponding to a calculated molecular mass of 1115.8 Da, 2 Da less than the T16 peptide (1117.5 Da). 2 Cys residues present in the T16 sequence (TSTCNYPSTGCPS) indicate the possibility of a disulfide bond. A total of 4 glycoforms, m/z 739.4, m/z 810.4, m/z 901.5 and m/z 982.5, were observed from the doubly charged ion series. They corresponded to 1, 2, 3 and 4 hexoses attached to T16, respectively. These *O*-linked hexoses could be located at any one of the six Ser/Thr residues within the T16 sequence. The coupling of LC with a conventional ESI source does have several advantages as compared to the nanospray LC-MS setting. The *O*-glycopeptide can be purified afterwards based on the determined retention time. Also, this setup allows much larger quantities of sample to be analyzed as compared to nanospray, thus the quantity of the *O*-glycopeptide collected will allow for further experiments without the necessity of repeated purification. Thus, run-to-run errors can be minimized, the problems caused by extensive sample handling are greatly reduced and the overall analysis time is shortened.

Protein O-glycosylation has potential functions including the stabilization of secondary structure, protection of the sensitive domains from proteolytic attack, and serving as ligands in cell-cell recognition. Changes of *O*-glycosylation may parallel changes in protein functions. *O*-glycosylation on fungal endoPGs was recently observed for *A. Niger* endopolygalacturonase III (PGC) and endopolygalacturonase A (PGA), and the *O*-linked glycans were mannoses[17].

Interestingly, while *O*-glycosylation on both of these *A. Niger* EPGs occurs at the N-terminal, it occurs at the C-terminal for *B. cinerea* EPGs. Also, while Both *B. cinerea* endoPGs studied contain multiple *N*-linked glycosylation sites, EPGs from *A. Niger* such as, EPG I, EPG II, PGC and PGA, all contain a single N-linked glycosylation site.

Conclusions

Three high-mannose type *N*-linked sugar chains were present on BcPG3. The glycosylated asparagines were determined to be Asn307 (T11 peptide), Asn341 (T13 peptide) and Asn351 (T14). No sugar chain was found attached to the other potential N-glycosylation site at Asn112. One potential phosphorylation or sulfation was found at T14, but the attachment site could not be determined from MS/MS data. Using MS/MS on the T14 glycopeptide digested with a secondary protease (e.g. chymotrypsin) may provide further information on this post-translational modification. T16 was determined to be *O*-glycosylated with maximum of 4 hexoses based on the microheterogeneity observed.

Five out of 7 potential *N*-linked sites present at BcPG6 were determined to be occupied by high-mannose type oligosaccharides. Four of them were attached at Asn237 (T9 peptide), Asn146 (T6 peptide), Asn198 (T7 peptide) and Asn256 (T11 peptide), respectively. The other one was located at the T8 peptide, which contained two potential *N*-linked sites, Asn224 and Asn227, side by side (SNN$\underline{\text{N}}^{224}VT\underline{\text{N}}^{227}$ITFK). Asn224 rather than Asn227 was later determined by LC-ESI-MS/MS to be the occupied site.

Peptide mapping of both proteins revealed that 2 disulfide bonds were formed in BcPG6 between the Cys residues in T13 and T16, and 3 disulfide bonds were formed between Cys residues in T10, T15 and T16 on BcPG3. The existence of disulfide bridges are indicated by X-

ray crystallography for other fungal EPGs. For example, the three-dimensional structure of *E. carotobora* EPG II are determined by X-ray crystallography reveals two disulfide bridges[21;22]. Disulfide bridges were also revealed by mass spectrometric data for *S. purpureum* EPGI proteins[18]. In addition, the alignment of several fungal EPG show conserved Cys residues. One hypothesis being proposed based on those studies, is that the C-terminal cysteine residues in fungal endoPGs form disulfide bonds for the purpose of thermostability. Our study of both BcPG3 and BcPG6 indicated disulfide bonds indeed formed between the C-terminal Cys residues.

Acknowledgements:

We would like to thank Dr. Jacques Benen for providing recombinant BcPG3 and BcPG6 proteins to make this work possible. The grand for this study is from NIH and NSF.

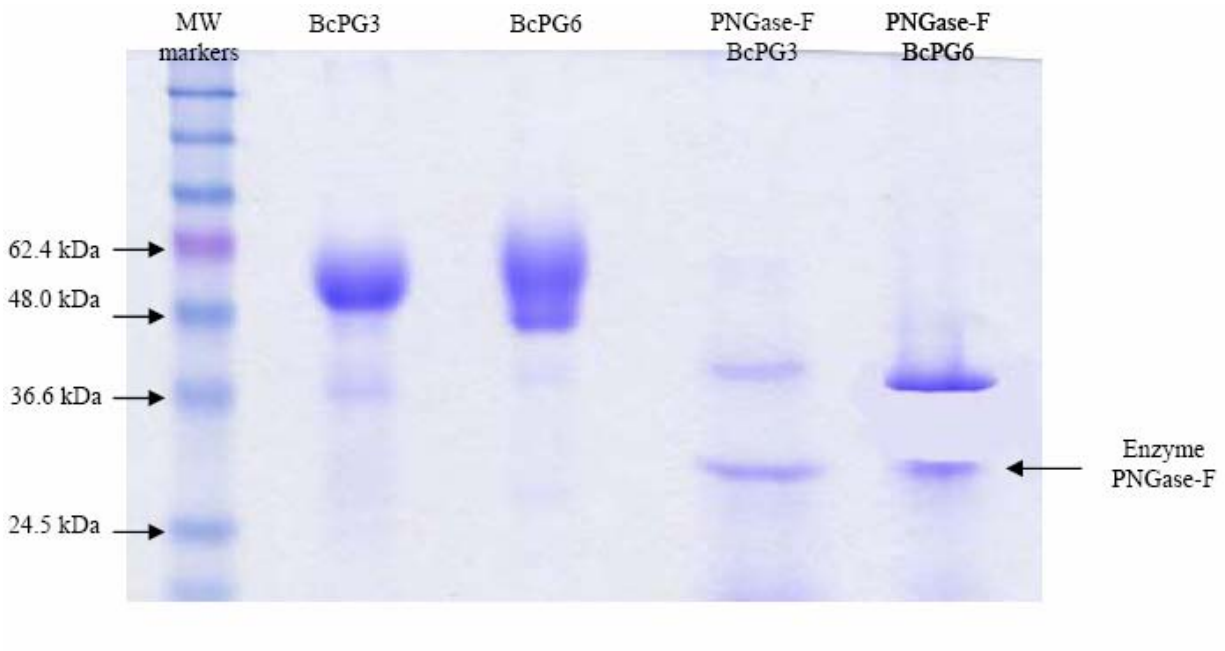


Figure 3.1. 12% Tris-glycine gel analysis of BcPG3 and BcPG6 before and afer PNGase-F treatment

Both proteins showed broad bands before the removal of N-linked glycan, and much thinner bands afterwards. The mass decrease for both samples indicated the existence of *N*-glycosylation.

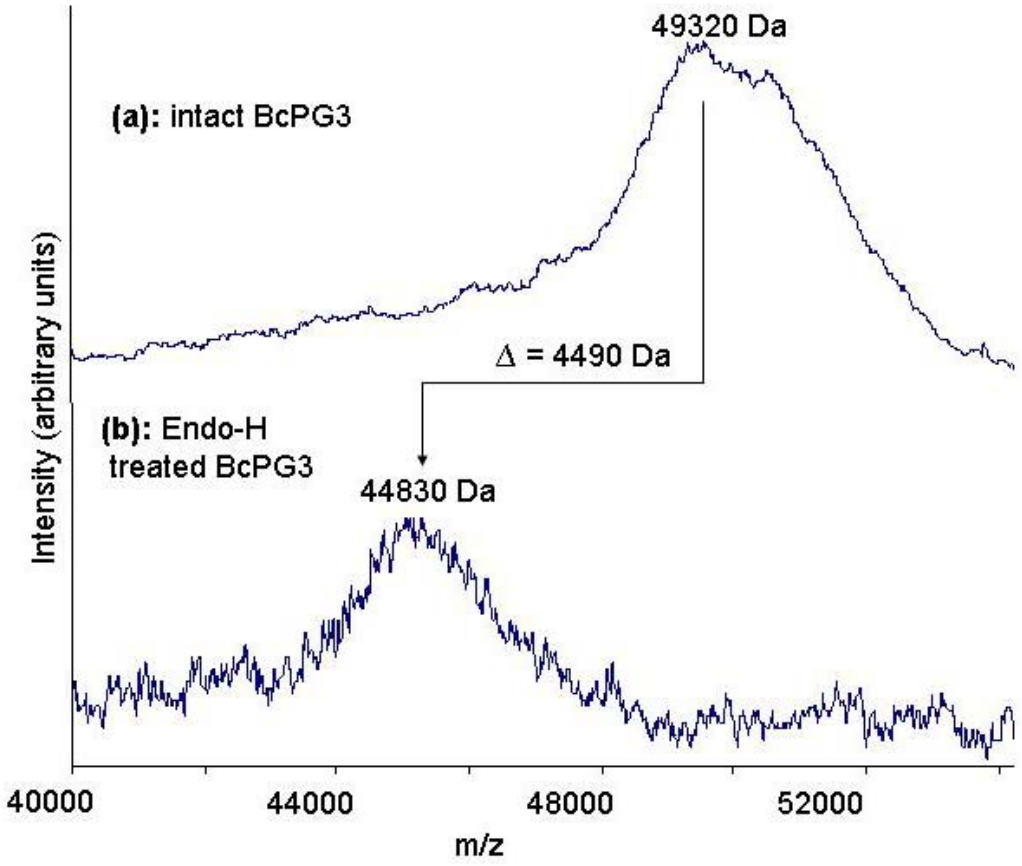


Figure 3.2. Molecular weights of (a) intact BcPG3 and (b) Endo-H treated BcPG3

The mass difference between (a) and (b) is 4490 Da, indicating *N*-linked glycosylation

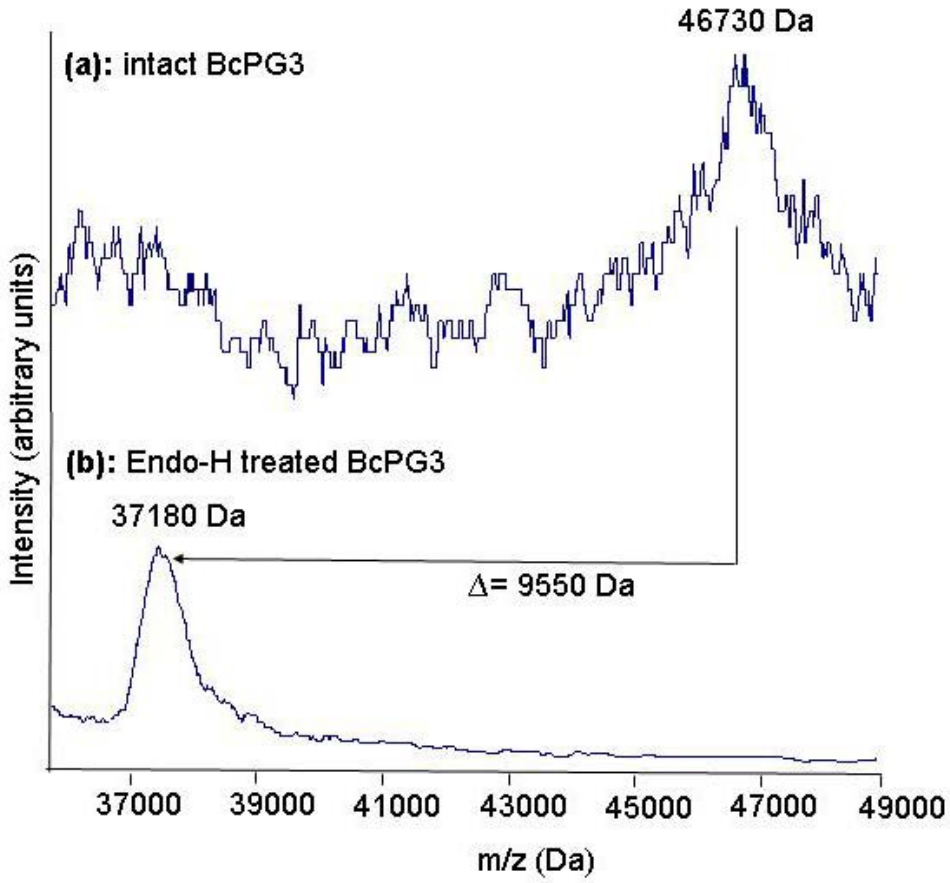


Figure 3.3. Molecular weights of (a) intact BcPG6 and (b) Endo-H treated BcPG6

The mass difference between (a) and (b) is 9550 Da, indicating extensive *N*-linked glycosylation

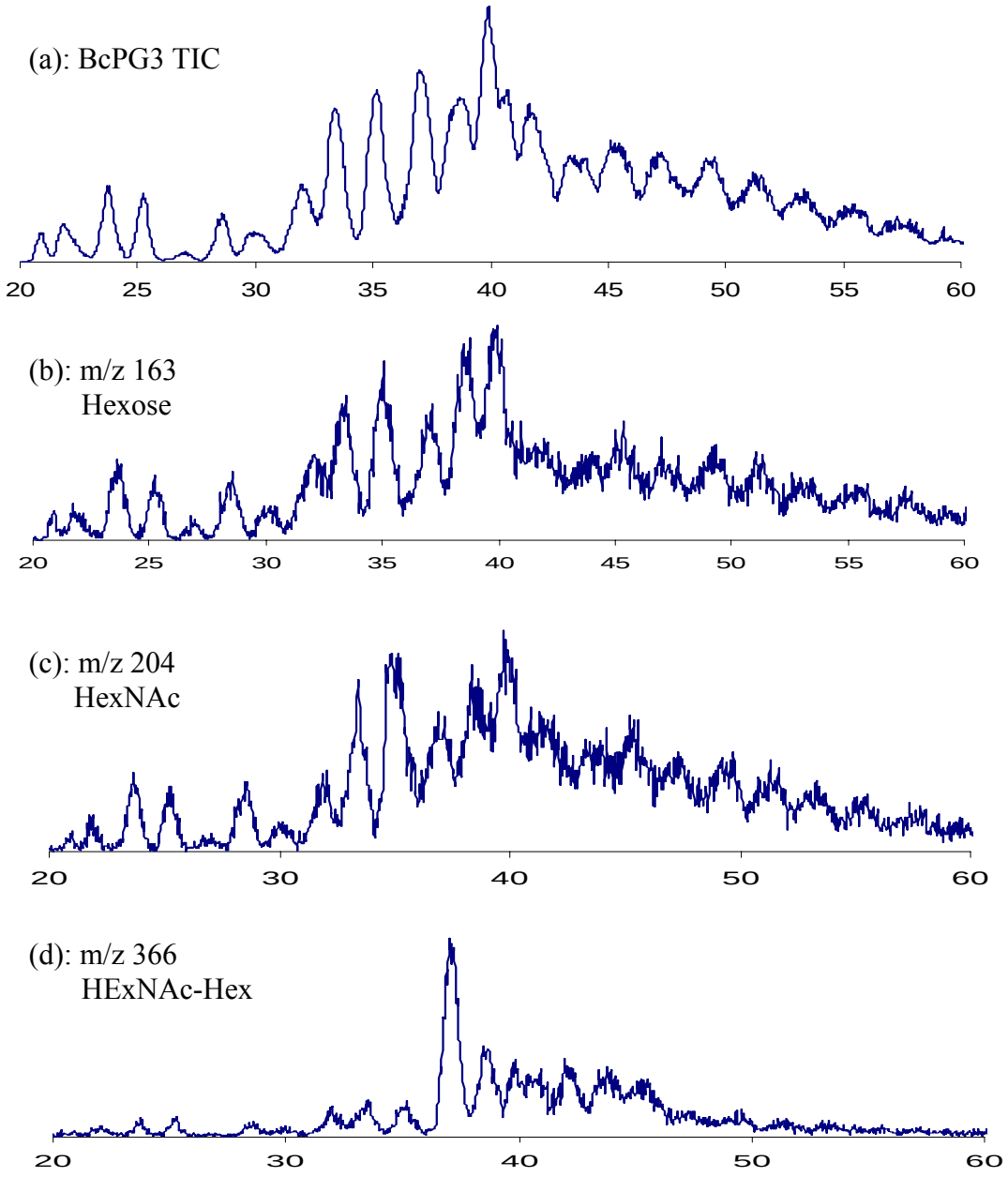


Figure 3.4. Nanoelectrospray LC/MS analysis of a 15 pmole injection of tryptic digest of BcPG3. (a). Total ion chromatogram (TIC) (m/z 400-2300) obtained by using low orifice voltage; (b). Extracted ion current from high cone voltage TIC for m/z 163 (Hexose); (c) m/z 204 (HexNAc); and (d) m/z 366 (HexNAc-Hex) from TIC (m/z 50-400) obtained by using high orifice voltage

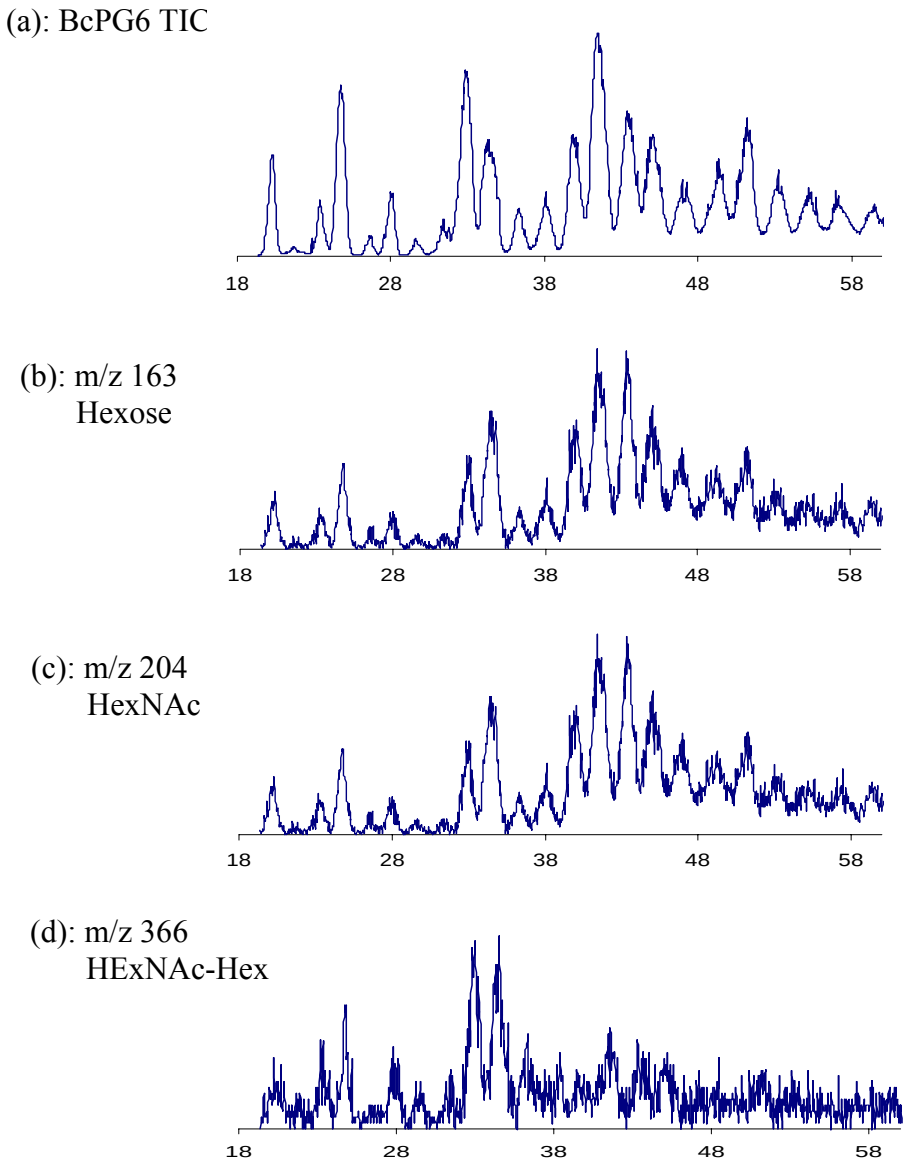


Figure 3.5. Nanoelectrospray LC/MS analysis of a 15 pmole injection of tryptic digest of BcPG6. (a). Total ion chromatogram (TIC) (m/z 400-2300) obtained by using low orifice voltage; (b). Extracted ion current from high cone voltage TIC for m/z 163 (Hexose); (c) m/z 204 (HexNAc); and (d) m/z 366 (HexNAc-Hex) from TIC (m/z 50-400) obtained by using high orifice voltage

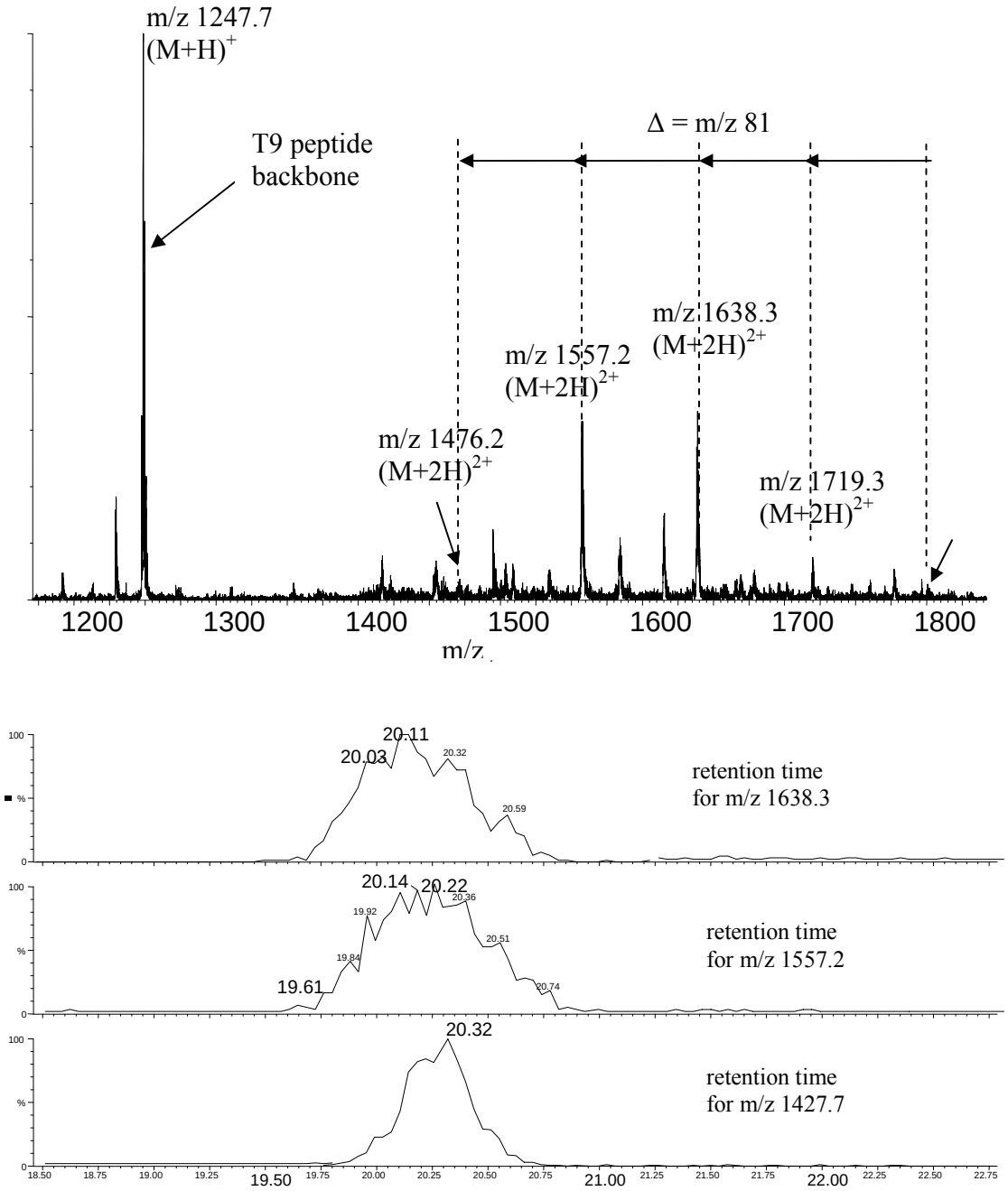


Figure 3.6a: Mass Spectrum of glycosylated BcPG6 T9 peptide
(peptide backbone molecular ion =1247.7 Da)

The corresponding retention time for T9 peptide and the glycoforms are also showing.
N-linked glycan is attached to NSELVN^{237}SSNGAR.

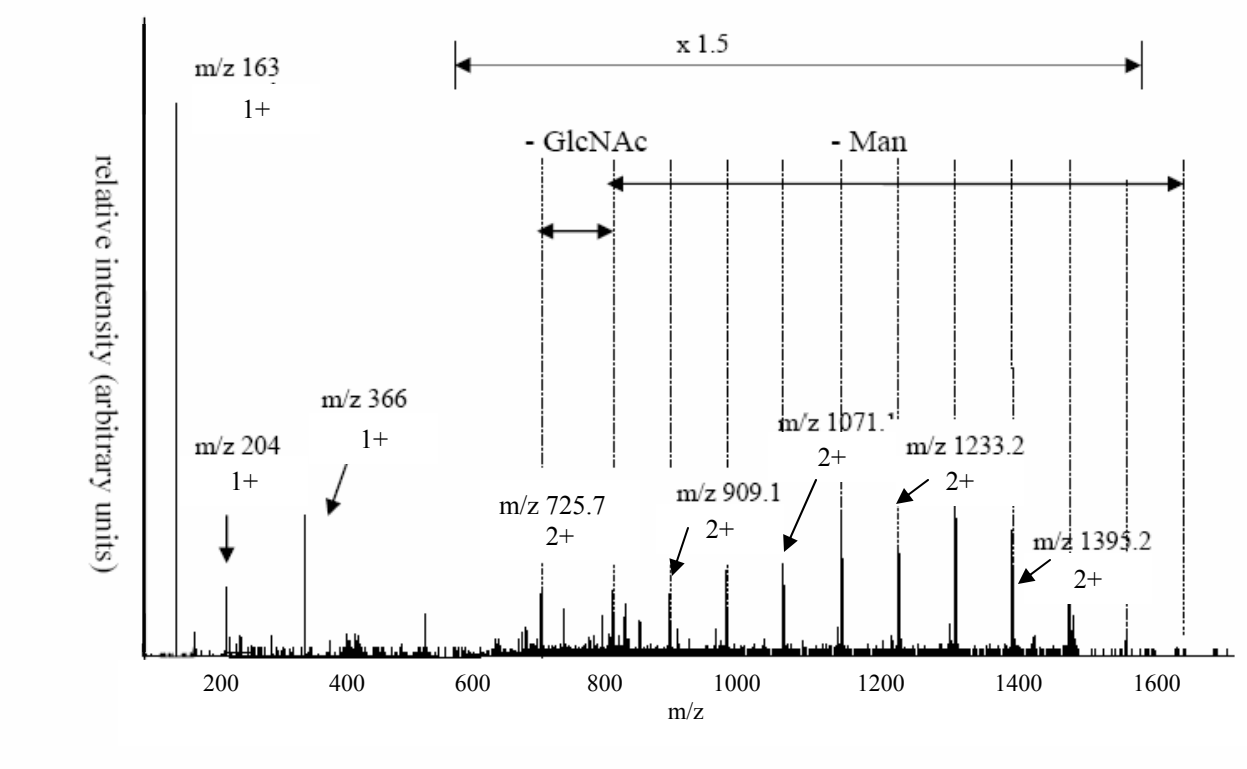


Figure 3.6b. MS/MS spectrum of glycosylated BcPG6 T9 peptide at m/z 1638.3

The loss of a series of mannose followed by GlcNAc indicates the high-mannose structure. For precursor ion at m/z 1638.3, the attached glycan is determined to be $GlcNAc_2Man_{10}$.

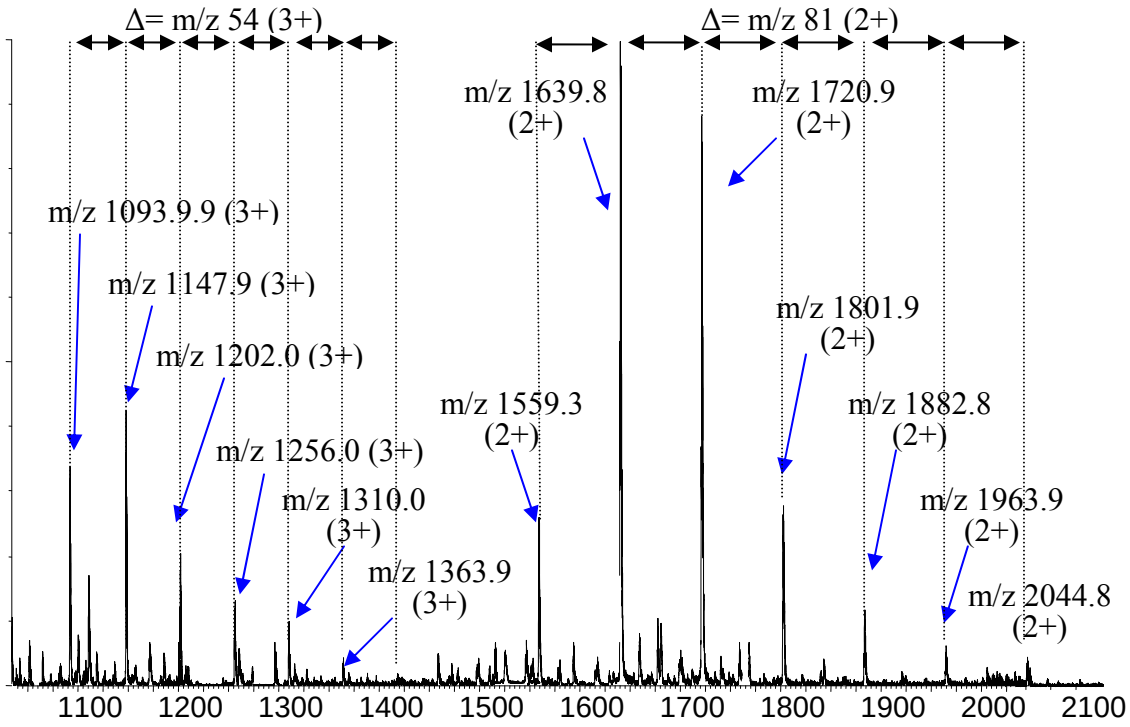


Figure 3.7a: Nanoelectrospray LC/MS spectrum of BcPG6 glycosylated T8 peptide (two potential N-linked sites are present: SNN**N**^{224}VT**N**^{227}ITFK)

Two series of glycoforms were observed: doubly charged peaks spacing with an 81 Da and triply charged peaks spacing with an 54 Da. Ion peaks of m/z 1639.8 (2+) and m/z 1093.9 (3+) correspond to T8 attached with $GlcNAc_2Man_{10}$.

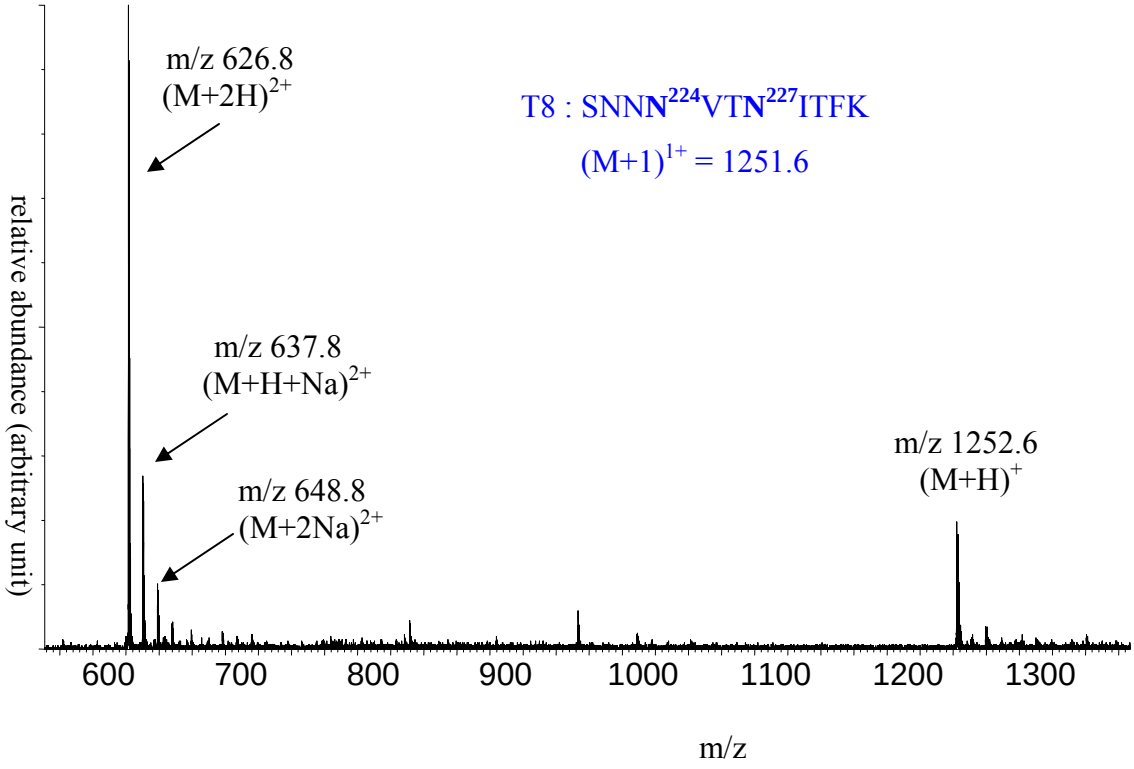


Figure 3.7b. LC/MS spectrum of 10pmol PNGase-F treated tryptic BcPG6 T8 peptides

PNGase-F digestion converts Asn to Asp residue, therefore there is 1 Da increase of the resulting peptide of each N-linked site. molecular mass of resulting T8 peptide increased 1 Da to m/z 1252.6 (1+), indicates only one N-linked site located on either Asn224 or Asn227

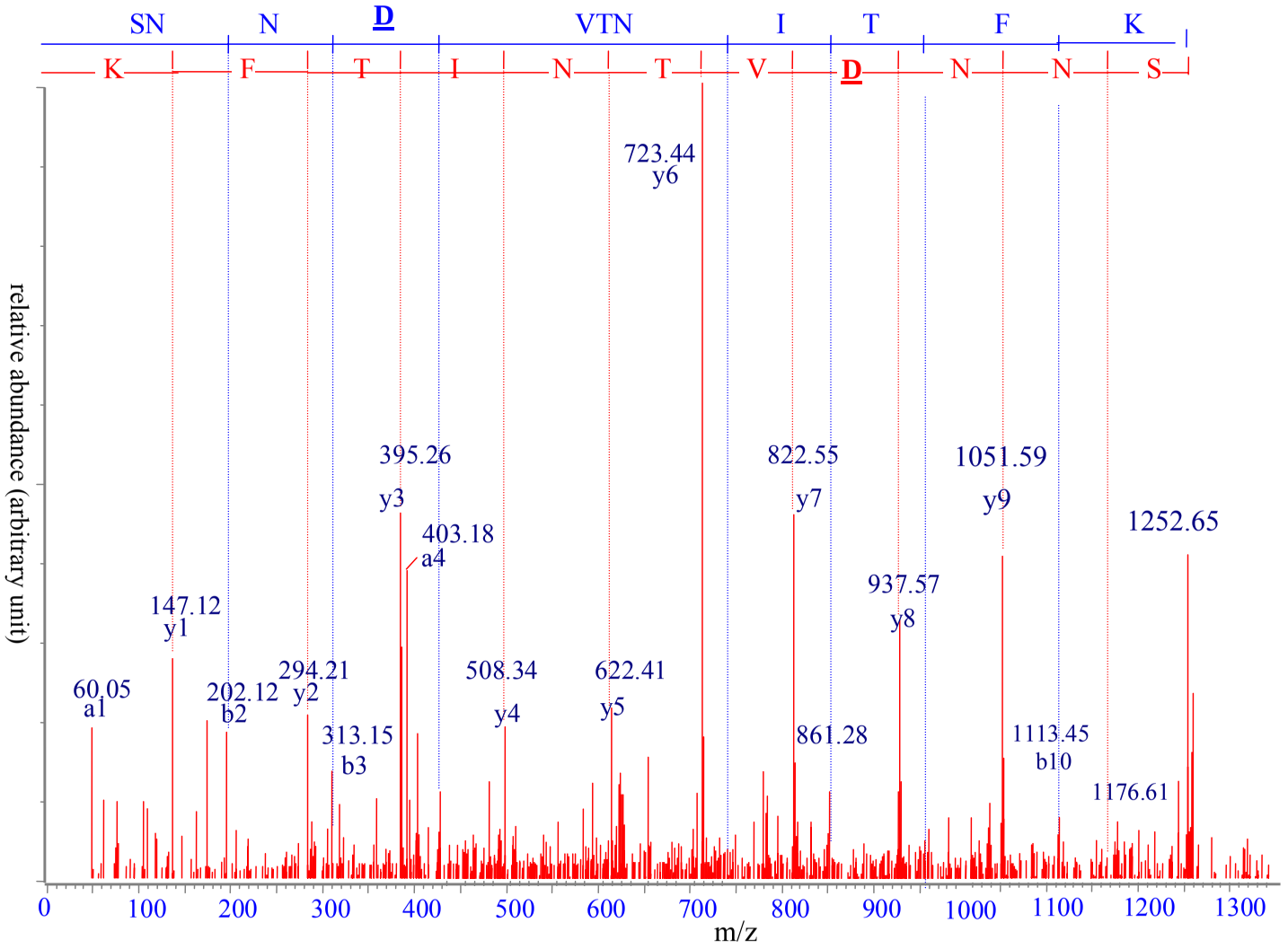


Figure 3.7c. LC/MS-MS spectrum of precursor 1252.65 (1+) from PNGase-F treated BcPG6 T8 peptide

The sequence of PNGase-F treated T8 peptide is obtained from both Y ions and B ions to be SNNDVTNITFK. The Asp observed in this sequence was originally attached to oligosaccharides.

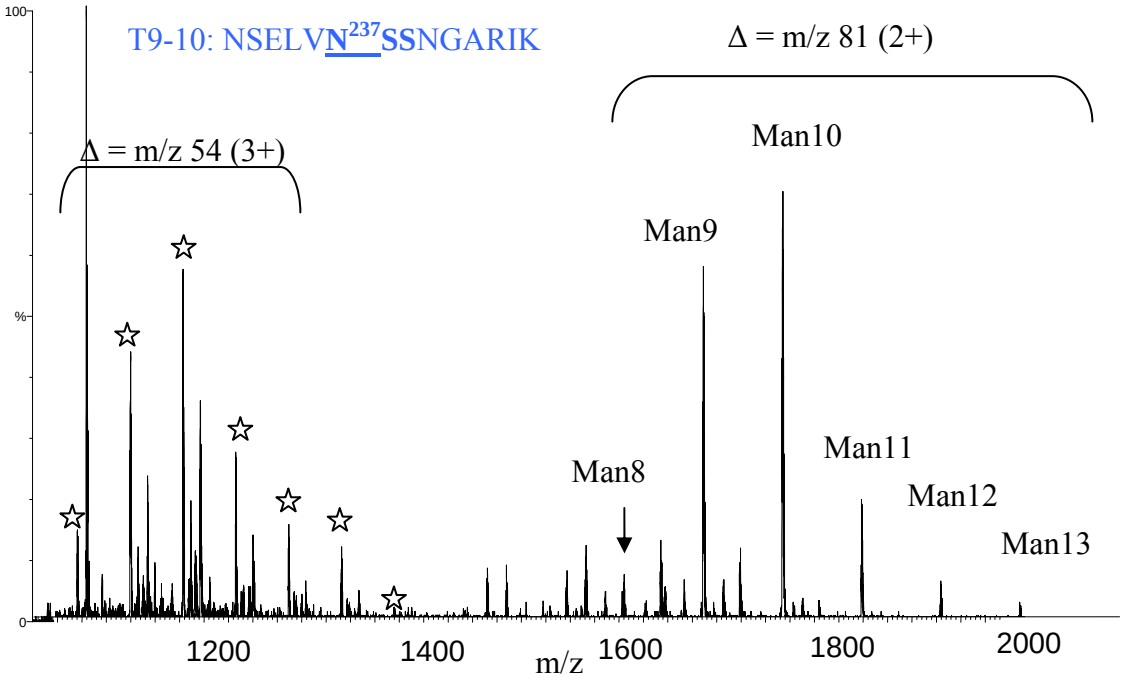


Figure 3.8a: LC/MS spectrum of glycosylated T9-10 peptide at retention time of 26-27 min.

The dominated glycoform is $GlcNAc_2Man_{10}$
(the observed m/z value of each ion is described in Table 1)

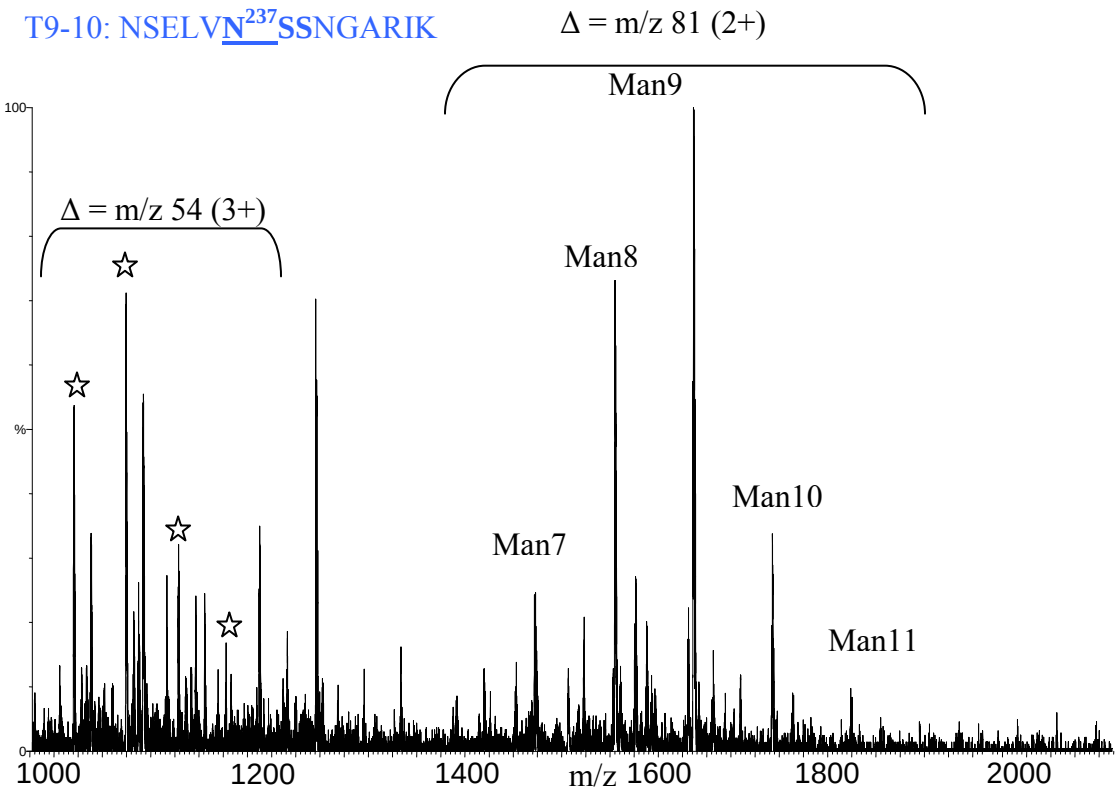
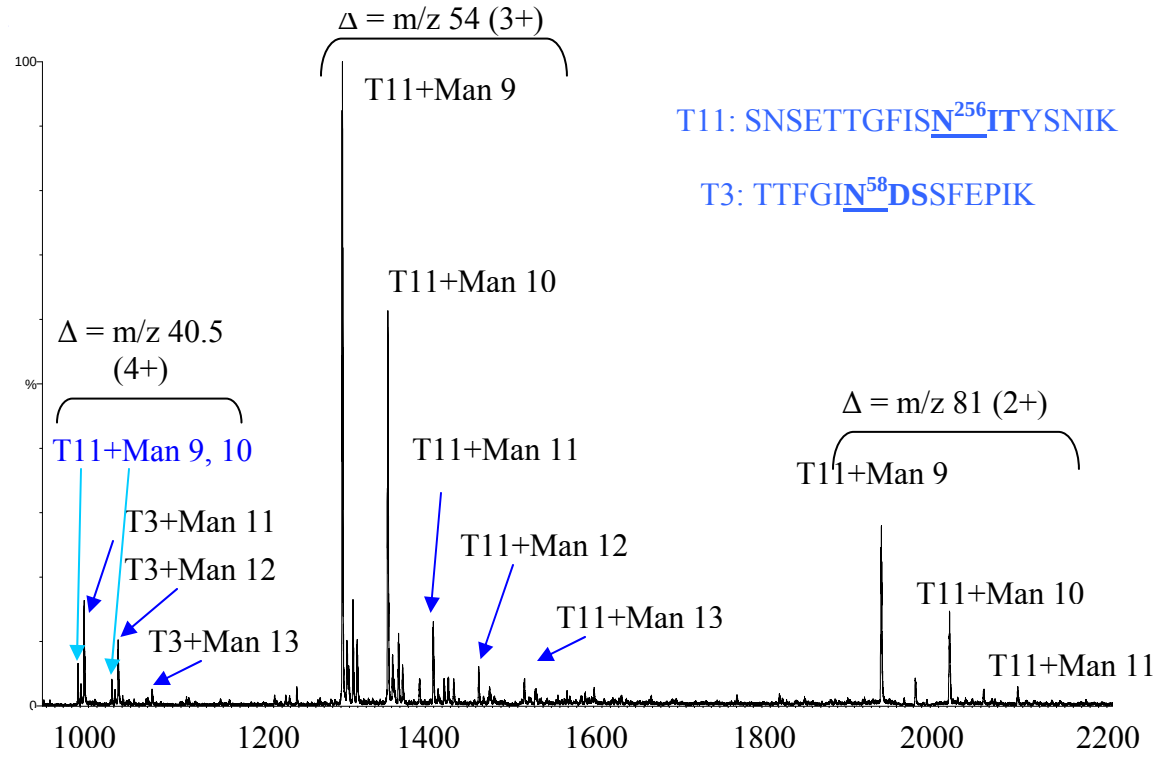


Figure 3.8b: LC/MS spectrum of glycosylated T9-10 peptide
at retention time of 27.5-28.5 min.

The dominated glycoform is $GlcNAc_2Man_9$
(the observed m/z value of each ion is described in Table 1)

(a). Peaks from 32-33.5 min:



(b). Peaks from 36-37 min:

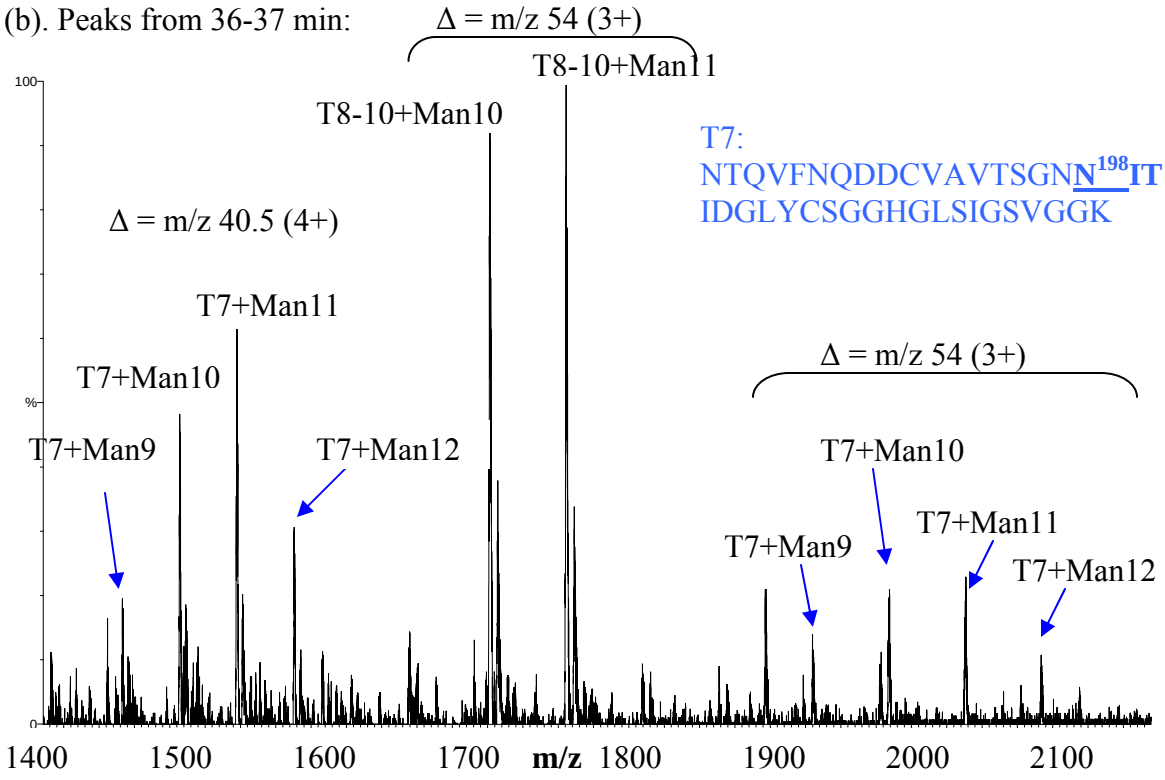


Figure 3.9. LC/MS spectra of glycopeptides corresponding to ion peaks from (a) 32-33.5 min and (b) 36-37 min (The observed m/z value of each ion is described in Table 1)

Table 3.1: The glycopeptides present in BcPG6 and the composition of the attached oligosaccharides

Tryptic Peptide	peptide mass (Da)	Observed glycoform (m/z)			oligosaccharides composition
		M^{2+}	M^{3+}	M^{4+}	
T8 SNN**N**VT NITFK	1250.6		1363.5		$GlcNAc_2Man_{15}$
		1964. 2	1309.6		$GlcNAc_2Man_{14}$
		1883.1	1255.6		$GlcNAc_2Man_{13}$
		1802.0	1201.5		$GlcNAc_2Man_{12}$
		1720.9	1147.6		$GlcNAc_2Man_{11}$
		1639.8	1093.5		$GlcNAc_2Man_{10}$
		1558.8	1039.5		$GlcNAc_2Man_9$
T9 NSELV**N**SS NGAR	1247.7	1800.3			$GlcNAc_2Man_{12}$
		1719.3			$GlcNAc_2Man_{10}$
		1638.3			$GlcNAc_2Man_9$
		1476.2			$GlcNAc_2Man_8$
T9-10 NSELV**N**SS NGARIK	1487.8	2004.6	1336.2		$GlcNAc_2Man_{13}$
		1923.4	1282.3		$GlcNAc_2Man_{12}$
		1841.8	1228.2		$GlcNAc_2Man_{11}$
		1760.8	1174.2		$GlcNAc_2Man_{10}$
		1679.8	1120.2		$GlcNAc_2Man_9$
		1599.3	1066.5		$GlcNAc_2Man_8$
		1518.2			$GlcNAc_2Man_7$
T11 SNSETTGFIS **NIT**YSNIK	1975.0		1497.1		$GlcNAc_2Man_{13}$
			1443.0		$GlcNAc_2Man_{12}$
		2084.1	1389.0		$GlcNAc_2Man_{11}$
		2002.1	1335.0		$GlcNAc_2Man_{10}$
		1921.0	1281.0		$GlcNAc_2Man_9$
T3 TTFGFTNDSS FEPIK	1689.8			1055.3	$GlcNAc_2Man_{13}$
				1014.7	$GlcNAc_2Man_{12}$
				974.2	$GlcNAc_2Man_{11}$
T7 NTQ…GN**NIT** ….SVGGK	3999.4		2118.1	1588.8	$GlcNAc_2Man_{12}$
			2064.2	1548.4	$GlcNAc_2Man_{11}$
			2010.1	1507.7	$GlcNAc_2Man_{10}$
			1956.1	1467.3	$GlcNAc_2Man_9$

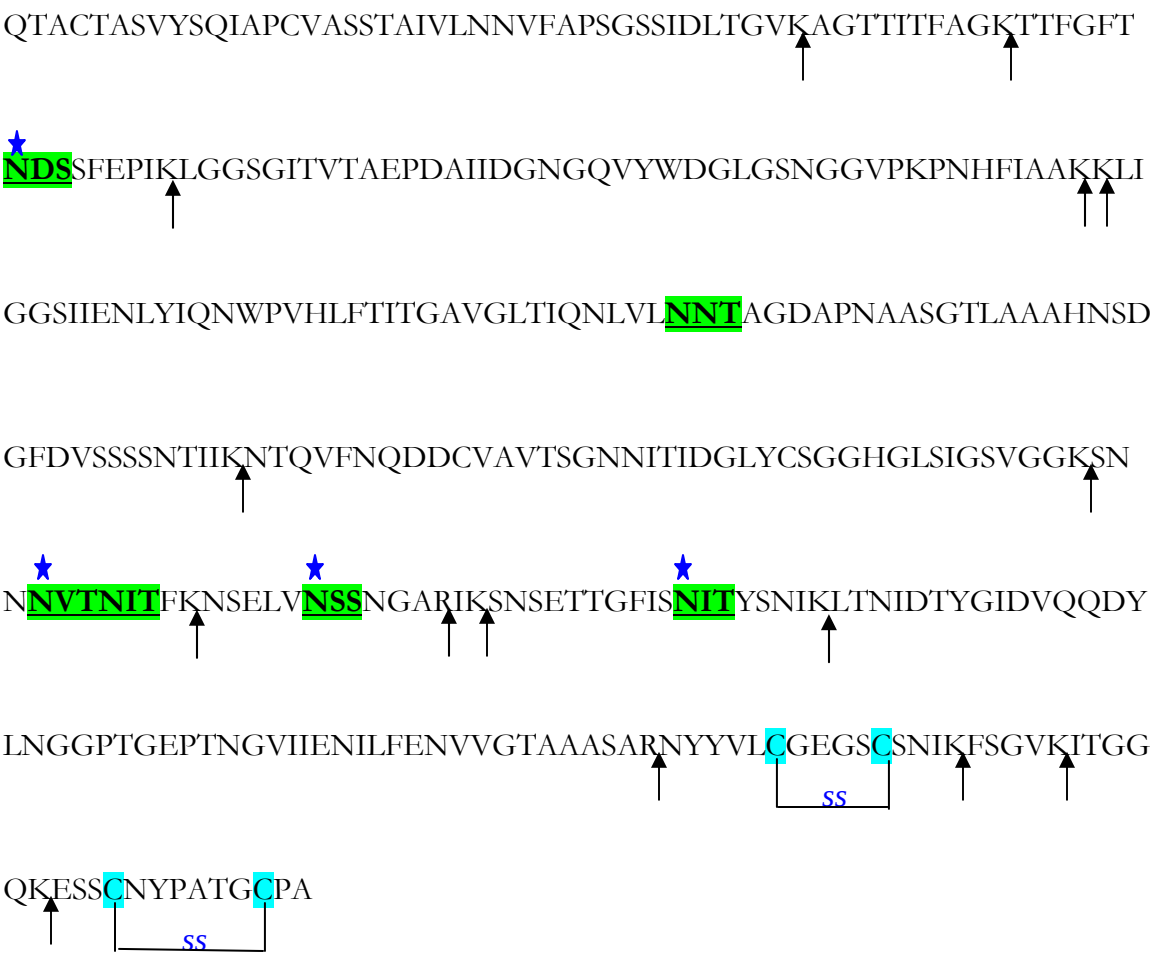


Figure 3.10: Amino acid sequence of BcPG6 protein

(arrows indicating the trypsin cleavage sites)

Seven potential N-linked glycosylation sites are indicated by shadeing, and the sites being occupied are labeled by★; 2 disulfide bonds are also indicated

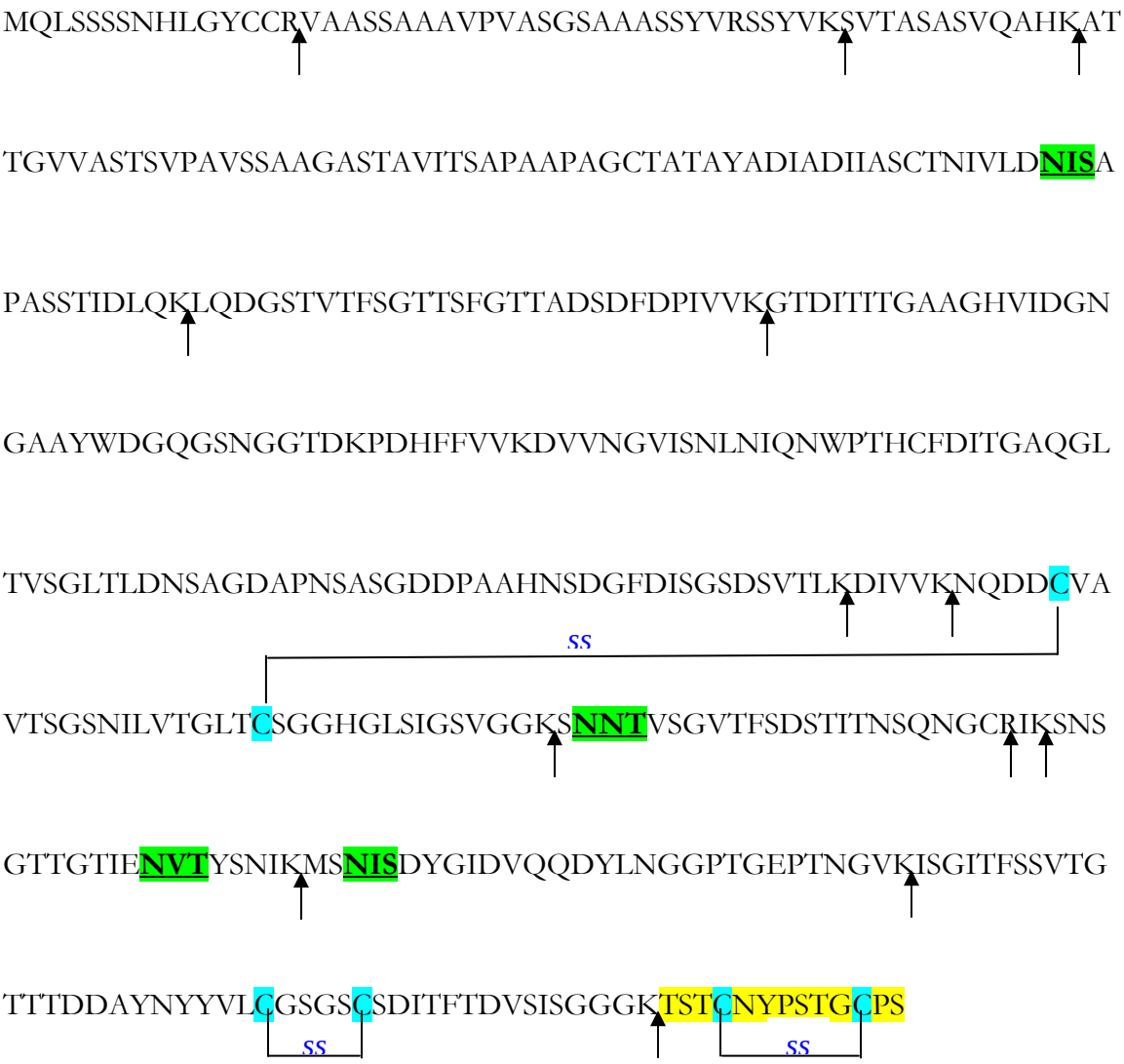


Figure 3.11: Amino acid sequence of BcPG3 protein

Four potential N-linked glycosylation sites are indicated by shading,
and the arrows indicate the trypsin cleavage sites.

Table 3.2: The glycopeptides present in BcPG3 and the composition of the attached oligosaccharides

Tryptic Peptide	peptide mass (Da)	Observed glycoform (m/z)			oligosaccharides composition
		M^{1+}	M^{2+}	M^{3+}	
T11 S**N**^{308}**NT**V SGVTFS… NGCR	2388.1			1094.8 1149.2 1203.3 1257.5 1311.6 1365.8	$GlcNAc_2Man_3$ $GlcNAc_2Man_4$ $GlcNAc_2Man_5$ $GlcNAc_2Man_6$ $GlcNAc_2Man_7$ $GlcNAc_2Man_8$
T13 SNSGTTG TIE**N**^{341}**VT** YSNIK	1884.9		1876.3 1957.3 2038.4 2119.4	1251.2 1305.2 1359.3 1413.3 1467.3 1521.4	$GlcNAc_2Man_9$ $GlcNAc_2Man_{10}$ $GlcNAc_2Man_{11}$ $GlcNAc_2Man_{12}$ $GlcNAc_2Man_{13}$ $GlcNAc_2Man_{14}$
T14 MS**N**^{351}**IS** DYG…… NGVK	3068.4			1645.8 1699.9 1726.8(+ 1 H_2PO_3) 1754.0 1780.8(+ 1 H_2PO_3) 1808.1 1835.0(+ 1 H_2PO_3) 1862.1 1881.9(+ 1 H_2PO_3) 1916.2 1943.1(+ 1 H_2PO_3) 1970.1 1997.1(+ 1 H_2PO_3) 2024.1 2051.0(+ 1 H_2PO_3)	$GlcNAc_2Man_9$ $GlcNAc_2Man_{10}$ $GlcNAc_2Man_{11}$ $GlcNAc_2Man_{12}$ $GlcNAc_2Man_{13}$ $GlcNAc_2Man_{14}$ $GlcNAc_2Man_{15}$ $GlcNAc_2Man_{16}$
T16 (O-linked) **TST**CN**Y**P **ST**GCP**S**	1316.5	1315.6 1477.7 1639.8 1801.8	658.3 739.3 820.4 901.4 982.5		T1 + 1 S-S T1+ 1 S-S + 1 Hexose T1+ 1 S-S + 2 Hexose T1+ 1 S-S + 3 Hexose T1+ 1 S-S + 4 Hexose

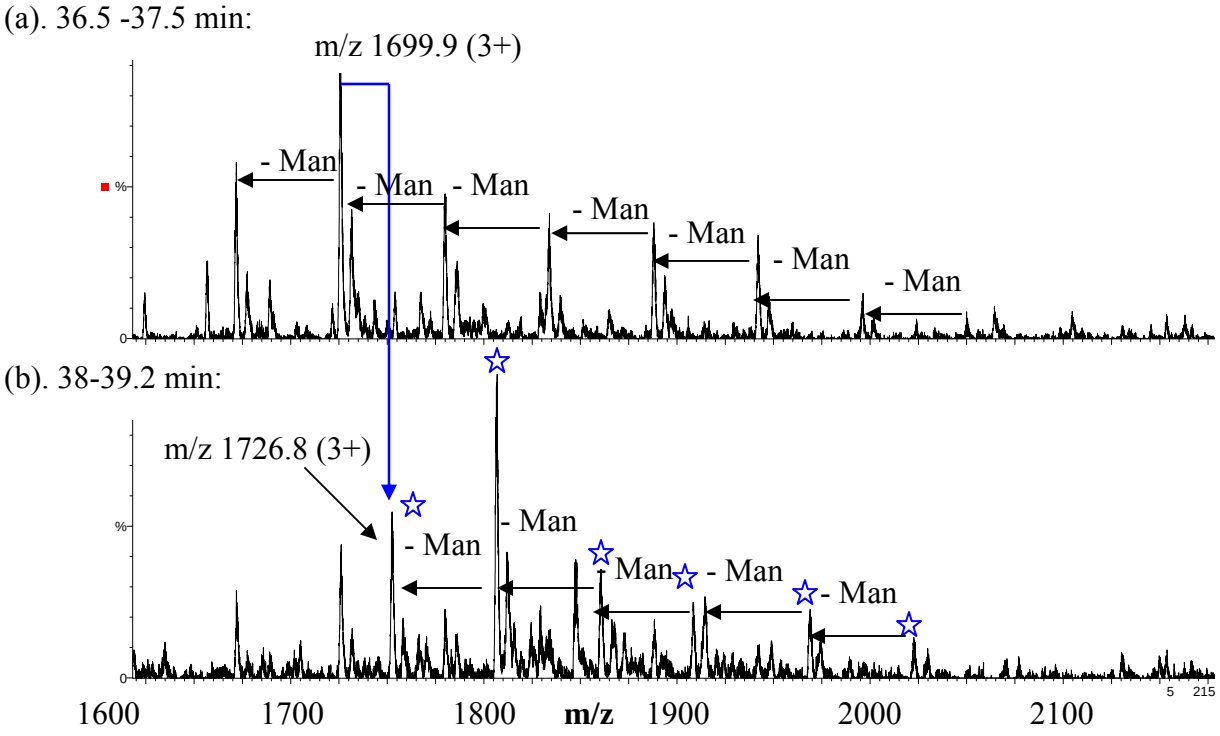


Figure 3.12: LC/MS spectra of glycopeptides corresponding to ion peaks at (a) 36.5-37.5 min and (b) 38-39.2 min
(the observed m/z value of each ion is described in Table 2)

A new serial of peaks appeared at (b) labeled by ☆. The 80.6 Da mass difference between m/z 1700.4 and m/z 1727.3 indicated one phosphrylation occurred on this peptide.

m/z 1699.9 = T14 + $GlcNAc_2Man_{11}$
(T14: MSN[331]ISDYGIDVQQDYLNGGPTGEPTNGVK, with 6 potential phosphorylation sites present)

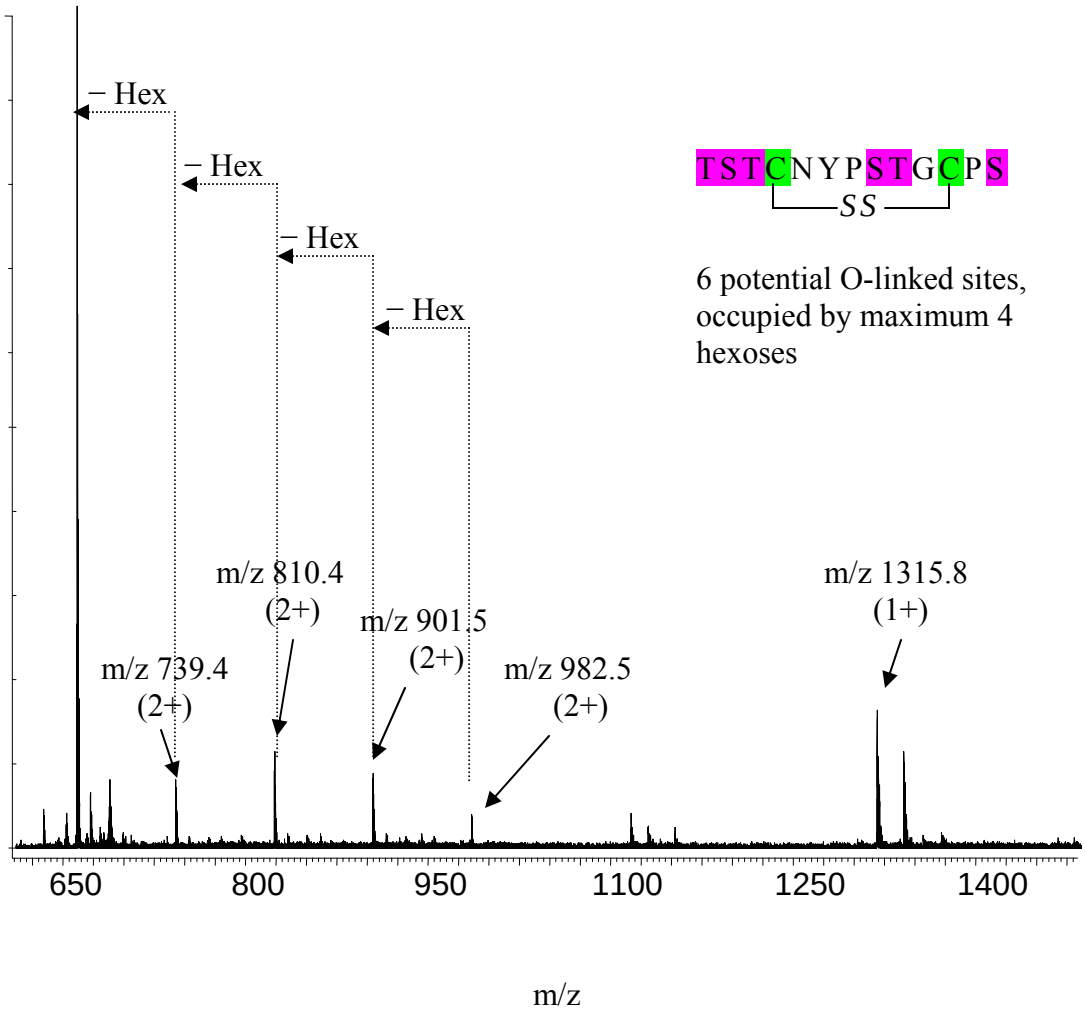


Figure 3.13: O-linked glycopeptides identified from LC-MS with conventional ESI source. m/z 1315.8 (2+) and m/z 658.4 (2+) is 2 Da less than T16 peptide backbone, indicating disulfide bond exist. A maximum of 4 hexoses (Hex) are attached to the peptide.

Reference List

1. Dept of crop sciences, University of illinois at Urbana-Champaign. Gray-mold rot or Botrytis blight of vegetables . 2000.
Ref Type: Report

2. Javis, W. R. Botrytinia nad Biotrytis species: taxonomy, physiology, and pathogenicity: a guid to the literature. 1977. Ottawa, Ontario, Canada, Canada Department of Agriculture.
Ref Type: Data File

3. Elad, Y. *Crop Protect.* **1988,** *7*, 361-66.

4. A.Berrie *J.Agric.Sci.* **1994,** *S17*, 383-89.

5. R.C.Staples , A. M. Mayer *FEMS Microbiology Letters* **1995,** *134*, 1-7.

6. W.Wdlich; G.Lorenz; H.Lyr; E.Nega; E.H.Pommer *Pers.Neth.J.Plant Pathol.* **1989,** *95*, 53-62.

7. Pashkoulov, D.; Giannetti, I.; Benvenuto, E.; De Martinis, D. *Mycological Research* **2002,** *106*, 827-31.

8. G.Leone; E.A.M.Schoffelmeer; J.van den Heuvel *Can.J.Bot.* **1990,** *68*, 1921-30.

9. G.van der Cruyssen; E.de Meester; O.Kamoen *Med.Fac.Landbouwwet.Rijksuniv.Gent.* **1994,** *59*, 895-905.

10. P.H.Reignault; C.Kunz; N.Nelage; E.Moreau; R.Vedel; W.Hamada; G.Bompeix; M.Boccara *Mycol.Res.* **2000,** *104*, 428.

11. A.ten Have; W.Mulder; J.Visser; J.A.van Kan *Plant-Microbe Interact.* **1998,** *11*, 1009-16.

12. B.Poinssot; E.Vandelle; M.Bentejac; M.Adrian; C.Levis; Y.Brygoo; J.Garin; F.Sicilia; P.Coutos-Thevenot; A.Pugin *Molecular Plant-Microbe Interaction* **2003,** *16*, 553-64.

13. A.ten Have; W.O.Breuil; J.P.Wubben; J.Visser; J.A.van Kan *Fungal Genetics and Biology* **2001,** *33*, 97-105.

14. Nigel Jenkins *Handbook of Industrial Cell Culture Mannalian, Microbial and Plant Cells*, V.A.Vinci; S.R.Parekh, Eds.; Humana Press: Totowa, NJ, 2003; Chapter 1.

15. Yang, C. Bergmann J. Benen R. Orlando. *Rapid Commun.Mass Spectrom* **1997,** *11*, 1257-62.

16. J.Colangelo, V. Licon J. Benen C. Bergmann R. Orlando. *Rapid Commun.Mass Spectrom.* **1999,** *13*, 1448-53.

17. M.Xie; K.V.S.Kolli; J.A.E.Benen; J.Visser; C.Bergmann; R.Orlando *2003 ASMS conference*, 2003 317.

18. Tetsuya Shimizu, Kazuo Miyairi and Toshikatsu Okuno *Eur.J.Biochem.* **2000,** *267*, 2380-88.

19. S.A.Carr; M.J.Huddleston; M.F.Bean *Protein Sci.* **1993,** *2*, 183-96.

20. B.Domon; C.E.Costello *Biochemistry* **1988,** *27*, 1534-43.

21. R.Pickersgill, M. Scott D. Smith K. Worboys & J. Jenkins *Acta Crystallogr.* **1999,** *D55*, 320-22.

22. R.Pickersgill, D. Smith K. Worboys & J. Jenkins *J.Biol.Chem.* **1998,** *273*, 24660-64.

Chapter 4. Glycosylation Site Mapping and Carbohydrate Structure Elucidation of Recombinant *Aspergillus niger* Endopolygalacturonase III [1]

[1] Xie, M.; Kolli, K.V.S.; Bergmann, C.; Benen, J.A.E.; Visser, J. and Orlando, R. Submit later to Analytical Biochemistry

Background

The fungus *Aspergillus niger* (*A. niger*) secretes a wide variety of plant polysaccharide-modifying enzymes, which are capable of degrading plant cell wall polysaccharides such as pectin and xylan, as well as plant storage polysaccharides such as starch and inulin. Pectin is the most complex of all these carbohydrates. The pectic polysaccharides, rhamnogalacturonan I and II (RGI and RGII), are heavily branched molecules composed of different saccharides connected by various types of linkages. The homogalacturonan is unbranched, but is decorated with substitutes such as methylesters. For a complete degradation of the pectin polysaccharide network, a vast array of interacting enzymes is necessary[1;2]. However, only a fraction of the pectin degrading enzymes have been fully characterized to date.

A. niger is able to utilize pectin as a sole carbon source for growth and produces a broad spectrum of polygalacturonases with a variety of patterns of action on homogalacturonan. Presently, a family of seven endopolygalacturonase encoding genes has been cloned and individually overexpressed including PGI, PGII, PGA, PGC, PGC, PGD an PGE[3]. The role of each enzyme in pectin degradation was investigated by analyzing the basic biochemical properties such as the pH optimum and the kinetic parameters of each EPG isozyme [4]. A great deal of study has been performed to analyze the specific activities of these *A. niger* EPGs on partially methyl esterified pectins[5-7]. It has been suggested that PGA and PGB are the scouting enzymes that make the fungus sense the presence of pectin by generating low molecular weight inducers. Next, the full array of EPGs are expressed in which PGII, with the highest specific activity, is the major force generating oligoGal*p*A which are quickly converted to monomers by the action of PGI. PGC and PGE have low activities and most likely attack parts of the homogalacturonan which have not yet been identified[8].

The destructive nature of *A.niger* EPGs on cell walls has gained them substantial commercial value in the food processing, fruit juice and wine industries[9]. The fundamental understanding of the functionalities of these enzymes will be of great value in improving their industrial applications, developing new commercial applications and fully exploiting the potential of *A. niger*. Thus knowledge about present carbohydrate-modifying enzymes will be helpful during the development of the new and improved products.

Many EPGs produced by *A. niger* have been identified as glycoproteins[10-12]. As the most complex of the post-translational modifications of proteins, protein glycosylation has proven to be important inmaintaining protein structure and function, and is also important during protein-protein interactions. However, the roles of the carbohydrate substituents on EPGs during their interactions with the substrates are not yet well understood. In addition, the plant produces a class of self-protective proteins capable of inactiving EPGs, namely polygalacturonase-inhibiting proteins (PGIPs). Whether carbohydrate substituents affect the ability of PGIPs to inhibit EPGs is unknown. One proposed approach to study the roles of glycan substituents is using site-directed mutagenesis to create mutant forms of EPGs and PGIPs, which lack one or more glycosylation sites normally found in the native proteins. The specificity, selectivity, and activities of the mutants are then compared to native proteins to reveal any alternation in functionality. Therefore, a complete characterization of glycosylation of EPGs is required prior to the design of their mutants for EPG-PGIP interaction studies.

In this chapter, the glycosylation of one recombinant *A. niger* EPG isozyme, PGC, is the subject of study. Its carbohydrate structures and glycosylation site heterogeneities were analyzed by various mass spectrometric methods. MALDI-MS was used to analyze the molecular mass of the intact protein as well as its deglycosylated form. The *N*-linked glycopeptides were identified

from ESI-MS using precursor ion scan by the orifice potential stepping technique [13-15]. The presence of *O*-linked glycosylation was further detected based on the carbohydrate heterogeneity by another LC-MS method. The sites of *O*-linked glycosylation on PGC were identified by electron-capture dissociation (ECD) MS/MS experiments.

The common fragmentation techniques used in MS/MS experiments (such as CID) are based on vibrational excitation, in which energy is rapidly redistributed over many available degrees of freedom. This is known as intramolecular vibrational energy distribution. Under such condition, the fragmentation is most likely to occur at the weakest bonds. Due to thermal lability, glycosidic bonds are the ones most easily cleaved. Thus, it is difficult to obtain information about the glycosylation sites by CID. On the other hand, the ECD process is based on ion-electron recombination, and cleavage occurs prior to energy distribution. Cleavage of the backbone N-C_α is often observed, except at the N-terminal side of proline[16] (due to the presence of a tertiary nitrogen). The ability to preserve labile groups makes ECD an important tool for the studying of glycosylation as well as phosphorylation [17].

Experimental methods

Recombinant PGC sample

The gene encoding PGC was cloned, sequenced and expressed in *A. niger* as described before[18]. The purified recombinant PGC (concentration 2.68mg/mL) was stored in 50mM sodium acetate buffer (pH=5.5) and kept frozen till further usage. The sample was generously provided by Dr. Benen from The Netherlands.

ELISA detection of glycosylation and phosphorylation

DIG Glycan differentiation kit (Roche Molecular Biochemical Inc.) was used to detect the present of the glycosylation using the procedure described in Chapter 3. In a similar manner to that of the detection of carbohydrate, phosphorylation was detected by a modified ELISA method. Two different phosphoprotein antibodies (Zymed laboratories Inc., South San Francisco, CA) were used to probe for the presence of phosphorylation. The rabbit anti-phosphorylated protein is a polyclonal antibody purified from rabbit antiserum, which recognizes Ser, Thr and Tyr phosphorylation. Mouse anti-phosphotyrosine is a monoclonal antibody, which reacts specifically with tyrosine-phosphorylated proteins and has no cross-reactivity to phosphoserine or phosphothreonine. The α-casein was used as positive control and 1% BSA in TBS buffer was used as negative control.

MALDI-TOF MS of intact and deglycosylated recombinant PGC:

The MALDI-MS analysis was carried out on a Hewlett-Packard (Palo Alto, CA, USA) G2025A time-of-flight mass spectrometer. A 4ug/ul solution of sinapinic acid (Aldrich) was used as matrix and was made by dissolving the sinapinic acid in a solution of 50% (V/V) acetonitrile and 0.1% trifluoroacetic acid (TFA) in water. The recombinant PGC was mixed with sinapinic acid solution in a 1:1 ratio, and vacuum-dried on the MALDI target.

The deglycosylation was performed directly on the MALDI probe in order to reduce the sample and enzyme required. [19] Basically, 0.2µl of Endoglycosidase-H (Endo-H) (Prozyme Inc., San Leandro, CA) was mixed with 0.5µl of recombinant PGC on the MALDI target. The digestion proceeded at room temperature for 30min. MilliQ water (MQ-H_2O) was added during

the digestion to keep the reaction environment moist. After incubation, 0.5ul of sinapinc acid matrix solution was added to the spot and dried for MALDI analysis.

A nitrogen laser (λ=337nm) was used to ionize the sample. The instrument operated in the positive mode with an accelerating voltage of 28kV, extractor voltage of 7kV, and a pressure below 10^{-6} torr. Prior to sample analysis, a mixture of standard proteins was used to calibrate the instrument externally.

Trypsin digestion of recombinant PGC:

Recombinant PGC was dissolved in 0.2M Tris-HCl in 8M urea and the disulfide bonds were reduced with 40mM dithiothreitol at 55ºC for 30min, and the resulting free –SH groups were then subjected to carboxymethylation with 80mM iodoacetamide. This reaction was carried on in the dark at room temperature for 30min. Afterwards, MQ-H_2O was added to the mixture to reduce the concentration of urea and Tris-HCl down to 2M and 50mM respectively. The pH value of the resulting mixture was adjusted to around 8.0 with 10mM HCl, optimum for trypsin digestion, which was done by adding sequencing grade trypsin at a ratio of 1:25 (w/w, enzyme to substrate), and the digestion was carried out at 37ºC for 24h with gentle shaking.

LC-nanoESI-MS with stepped-orifice potential scanning technique:

Approximately 5pmol of a tryptic digest of recombinant PGC was injected into a 180mm×1mm C-18 column (LC Packing, PepMap™, 3µm particle size) interfaced to a QTOF instrument. Water's® CapLC™ system was used to deliver solvents at the rate of 1 µl/min. The mobile phases for gradient elution were 0.1% formic acid (v/v) in MQ-H_2O (buffer A) and 0.1% formic acid (v/v) in methanol (buffer B). The column was first desalted with 90% Buffer A

wash for 20min, and the separation was achieved by increasing buffer B concentration from 10% to 80% over 60min.

The mass spectrometer was a hybrid quadrupole-time of flight instrument equipped with nanospray (Q-TOF-2^{TM}, Micromass®). In order to selectively detect glycopeptides, the stepped orifice voltage technique was used. Instrument operation was described in detail in Chapter 3. The voltage on the electrospray source was set at 3000V, and the scan rate was 1.0s/scan with 0.1s inter-scan delay. The instrument was externally calibrated using 1.5pmole Glu^{1}-Fibrino-peptide B prior to sample analysis.

O-linked glycopeptide identification and purification

All N-linked glycan was first released by peptide N-glycanase-F (PNGase-F) digestion. Therefore any glycopeptide identified afterwards could be assumed to result from *O*-linked glycosylation. PNGase-F digestion was carried out by adding 5mU PNGase-F into ~50μg of digested PGC peptides at pH 7.5, and the reaction was continued for 18 hr at 37ºC. The resulting mixture was injected onto a C18 column (Phenomenex Luna 5u C18 (2), 150mm×2mm, 5μm particle size). The Hewlett-Packard 1100 series HPLC station was used to deliver the solvents at a flow rate of 20 μL/min. The *N*-linked glycans were removed with a 10 min wash of 5% buffer B. Then the concentration of buffer B was increased to 85% over 1 min and kept at 85% B for the next 10min to elute all peptides.

Detection of the *O*-linked glycopeptide was achieved in a different LC-MS setting from *N*-linked glycosylation detection. The collected peptides (after the removal of *N*-linked glycan) were injected onto the C18 column mentioned above and the separation was achieved by increasing buffer B from 10% to 80% over 60min. The solvents were delivered, again, by a

Hewlett-Packard 1100 series HPLC station which was coupled to QTOF2. To detect the potential *O*-linked glycopeptides, effluents were analyzed directly in the mass spectrometer without passing through the UV detector.

In order to collect the purified *O*-linked glycopeptide, an identical amount of PGC double digests were separated on the same C18 column, which was then disconnected from the QTOF2 after detection. The effluents again were not monitored by UV detection to avoid any unknown dead volume. The collected fractions was concentrated in a speed-vac, and analyzed by ECD-MS/MS for site mapping.

_O_-linked glycosylation site mapping by ECD-FTICR MS/MS

ECD-MS/MS was conducted at the National High Magnetic Field Laboratory (Florida State University). The experiments were performed in the positive ion mode with a laboratory-built mass spectrometer positioned in a passively shielded 9.4 T superconducting magnet[20]. The experimental event sequence was controlled by a modular ICR data acquisition system (MIDAS). The peptides were infused at a flow rate of 500 nL/min through an electrospray emitter consisting of a 50μm-i.d. fused-silica capillary which had been mechanically ground to a uniform thin-walled tip. The fomed ions are transported into the mass spectrometer through a Chait-style atmosphere-to-vacuum interface and externally accumulated for 2s in an octapole pretrap. After accumulation, the collected ions are transferred through multipole ion guides and captured by gated trapping in an open cylindrical cell.

Electrons for ECD were generated with a 10 mm diameter dispenser cathode (No. 1109; Heat Wave, Watsonville, CA, USA). The cathode was mounted on axis of the system, inside the passive shielding and 73 cm from the center of the ICR cell. A molybdenum grid was located 8

mm from the emitting surface. ECD was performed by application of −5.5 V to the cathode, and raising the grid potential to +100 V for 3–51 ms while holding the trap plates at 5–10 V. Excess electrons were scavenged by changing the potentials on the trap plates to 2 V, grid to +5 V and cathode to +10 V for 10 ms. At all other times, the cathode bias voltage was −0.1 V and the grid potential was −200 V.

Ions were subjected to chirp excitation [48–480 kHz (or between) at 150 Hz/μs] and direct-mode broadband detection (512 Kword or 1 Mword data points). Hanning apodization and one zero fill were applied to all data prior to fast Fourier transform (FFT) and subsequently magnitude transformed. Spectra were normally internally calibrated (Ledford equation) with three ions of known *m*/*z*.

_O_-linked carbohydrate composition analysis by High-Performance Anion Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD)[21]

Approximately 200ug of recombinant PGC was digested with PNGase-F at 37ºC overnight to release all N-linked carbohydrate. The resulting de-*N*-glycoslated protein was purified by passing the mixture through a 1mm C18 column (OASIS® HLB 30μm). The *N*-linked carbohydrate was discarded. The collected protein fraction was then subjected to β-elimination with 1.0M sodium borohydride in 50 mM sodium hydroxide at 50ºC overnight, in order to release the *O*-linked carbohydrates. The treated sample was neutralized with 5% acetic acid, and salts were removed by treatment with an ion exchange resin (DOWEX® 50WX8-100). The sample was dried, and the resulting borates were removed by repeated evaporation with methanol. Then, the sample was treated with 4M TFA for 4 hrs at 100ºC in order to hydrolyze the polysaccharides into monomers.

The HPAEC-PAD analysis was on a Dionex DX500 system equipped with a GP40 gradient pump, an ED40 electrochemical detector and a Thermo-Separations AS3500 autosampler. Analyses were performed using a Dionex CarboPac MA1 (4 × 250 mm) analytical column with an Amino Trap and Borate trap column, at a flow rate of 0.5 mL/min. The gradient program used eluents A, water and B, 1.0 M NaOH and was set at an initial concentration of 10% NaOH for 5 min, then increased to 70% NaOH over the next 10 min and kept for another 10 min. Injections were made every 50 minutes. Instrument control and data processing were accomplished using Dionex PeakNet software, version 5.01.

Results and Discussion

Among seven *A.Niger* EPGs expressed to date, PGI and PGII are the two most extensively characterized regarding the structure and functionality[9;22;23]. To understand the different properties of EPGs, other isozymes were also characterized. For PGC, the pH region for its activity is 3-5 (optimum at pH 4.1), similar to other EPGs[4]. The kinetic parameters of PGC activity were also determined, which showed much lower activity compared to PGI and PGII[4]. The specific activity study of PGC indicated that it prefers the non-methyl esterified substrate and the activity decreased when the degree of methyl esterification increased[4]. However, further study is need to understand the definite role of PGC during the pectin hydrolysis process.

Though there is no direct evidence to indicate that glycosylation plays a role in EPG activities, study of their glycosylation could provide information valuable for mutagenesis studies as mentioned earlier. Such experiments had been performed on *A. niger* pectin methyl esterase (PME)[24], and the observed activities of mutants, which had one or more N-linked sites

removed by site directed mutagenesis, were similar to the native enzymes. The glycosylation on EPG II and EPG I had been studied before and both enzymes contain one *N*-linked glycosylation site with a high mannose structure attached to it, and neither *O*-linked glycosylations nor phosphorylation was present. The glycosylation of PGC is the subject of this work.

Lectin and ELISA result:

The presence of carbohydrate on PGC was detected by modified lectin blotting. Among five lectins used, which have different carbohydrate binding specificities, only GNA showed a strong positive reaction with recombinant PGC. This indicated the presence of terminally linked mannose resulting from either high mannose or hybrid type *N*-linked glycans. An examination of the amino acid sequence of PGC indicated one potential *N*-linked site at Asn220 exists.

Another common protein post-translational modification, phosphorylation was tested with antibodies specific for phosphorylations. Rabbit antibodies specific for phosphor serine/threonine showed a strong positive reaction, but not mouse anti-phosphotyrosine. This suggested that phosphoserine and/or phosphothreonine were present on PGC, but not phosphortyrosine. Among the homologously expressed *A.Niger* EPGs studied to date, only PGC showed evidence of phosphorylation. Possible phosphorylation on PGA was also indicated using this method and subsequently determined to be negative. The presence of phosphorylation on PGC, but not other *A. Niger* EPGs could be a very interesting point worthy of further investigation.

MALDI-MS of intact and deglycosylated protein:

The molecular mass of intact PGC was determined by MALDI-MS to be approximately 39600 Da, about 3300 Da larger than the calculated molecular mass from the amino acid

sequence (Figure 4.1a). This mass discrepancy is large enough to indicate the presence of post-translational modifications on PGC. The observed broad peak and poor mass resolution in the MALDI spectrum are further indications of heterogeneous glycoforms. The intact protein was deglycosylated by Endo-H directly on the MALDI target to confirm the present of *N*-linked glycosylation. After a 30 min digestion, a new peak appeared at a slightly lower mass range, approximately 1200 Da less than the intact protein peak (Figure 4.1b). The resulting mass migration following hydrolysis with Endo-H indicated that the recombinant PGC was indeed modified by *N*-linked glycans with either high-mannose or hybrid structure. This result was consistant with the lectin GNA test. Based on the fact that there is only one N-linked consensus sequence presented on the PGC amino acid sequence, Asn260 should be the site undergoing the *N*-linked glycosylation.

Theoretically, removal of *O*-linked glycosylation on the MALDI target with "Endo-*O*-glycosidase" could be done in the similar way. However, there are some practical difficulties that prohibited such an analysis. First, while *N*-linked glycans can be released from the protein by either Endo-H or peptide N-glycanase-F (PNGase F), no equivalent *O*-glycanase is available to release a variety of *O*-glycans. The available *O*-glycanases have much higher structure and linkage specificity. Thus, it is difficult to reveal the presence of *O*-linked glycosylation enzymatically with on target digestion experiments.

_N_-linked glycopeptide and _N_-linked glycan structure identification:

The trypsin digested PGC peptides were analyzed by MALDI-MS and LC/MS respectively. At the higher mass ranges of the MALDI-MS spectrum (Figure 4.2), a series of signals of moderate intensity were observed with a peak difference of 162. This is characteristic

of a peptide modified by a series of heterogeneous *N*-linked glycans, and the 162 mass differences indicated each peak in this series contained an additional mannose residue as found in high-mannose glycans. Since high mannose type contains only mannoses residues bound to the core structure, thus the 162 Da mass difference indicates the difference of one mannose residue. On the other hand, the monosaccharide residues presented at the other two types include sialic acids and GlcNAc, which will result in different peak spacing other than m/z 162 (or the m/z values corresponding to the multiply charged ions). The observed peptide masses were then compared with *in silico* digestion results (using Masslynx version 3.4, Micromass). The mass of the peptide containing the potential *N*-linked site Asn220 (T11) is determined to be 1310.7 Da, which is not observed in spectrum., and the mass difference between T11 and the observed glycopeptides are around 1500-1700 Da. The rest of the peaks observed in MALDI spectrum was identified to be T12, T6, T8, T5, T13 and T13-14.

The specific structure of the attached glycan was then determined by LC-MS with the stepped orifice potential technique on a Q-TOF2. The cone voltage settings were optimized based on the standard glycopeptide fragmentations under variable cone voltage conditions: while no carbohydrate fragments are being observed at a low cone voltage of 35, strong carbohydrate fragments can be generated when the cone voltage is increased to 80. Approximately 3 pmole of trypsin digested peptide mixture was analyzed using such settings and the total ion currents (TIC) corresponding to two cone voltage conditions are shown in figure 4.3a. The ion currents for three carbohydrate fragment ions extracted from high cone voltage TIC are showing in figure 4.3b. A single ion peak centered around 25 min contained all three marker ions, which suggests that only one peptide was detected as glycosylated, and its summed spectrum is showing in figure 4.3c. One series of doubly charged peaks is present from m/z

859.5 to 1752.0 with a peak spacing of m/z 81. The mass for each glycopeptide was calculated and compared to peptide masses calculated from *in silico* digestion, and the composition of the attached glycan was calculated accordingly. For example, the ion peak at m/z 1508.3 corresponds to a peptide with molecular mass of 3014.6 Da, and it is not a match with any calculated tryptic peptide masses. The mass difference between this ion and T11 is 1703.9 Da, which is exactly equal to the mass of $GlcNAc_2Man_8$. Therefore, this ion is determined to be T11 + $GlcNAc_2Man_8$. The rest of the glycopeptides were all analyzed in a similar manner and were determined to be T11 attached with high mannose *N*-linked glycans ($GlcNAc_2Man_{3\sim11}$). Another series of triply charged peaks appeared at slightly higher mass region with the peak spacing of m/z 54, which is corresponded to peptide T11-12 (one miscleavage) with $GlcNAc_2Man_{4\sim7}$ attached. The observed higher masses of this series peaks than the doubly charged series are resulted from the longer polypeptide backbone.

The molecular mass of $GlcNAc_2Man_{11}$ is calculated to be 2190 Da, which only compensates for part of the molecular mass discrepancy observed from intact protein analysis (~3300 Da). The remaining mass difference (~1110 Da) indicated there should be other post-translational modification present, and it seemed most likely to be glycosylation. However, the LC-MS using stepped orifice technique only revealed one *N*-linked glycopeptide. Another method had to be used in order to detect other potential glycopeptides. Meanwhile, peptide fingerprint mapping using both MALDI and LC-MS data located 11 out of 15 tryptic peptides, including the glycosylated T11. Thus, additional glycosylation could be located at any one of the unmatched peptides, T1, T4, T9 and T15.

O-linked glycopeptide identification and O-linked site mapping:

The doubly digested peptide mixtures (trypsin and PNGase-F) were separated on a microbore C18 column which was interfaced with a conventional ESI source. The QTOF instrument parameters were optimized to accommodate the high flow rate (20μL/min vs. 1 μL/min used in nanospray). Figure 4.4a shows the TIC obtained by this LC-MS setting and the mass spectra corresponding to each ion peak were obtained and compared to *in silico* digested masses. Two summed spectra, corresponding to the ion peaks at 15.2min and 15.8min respectively, showed the characteristic features of heterogeneous peaks (Fig 4.4b, 4.4c) similar to that seen for *N*-linked glycopeptides. The mass spectra obtained from the effluent at 15.2min show a dominant ion at m/z 1107.5, and a series of peaks with much weaker signals in the lower mass region. The rest of the ion peaks were also observed in the spectra obtained from the effluent at 15.8min. The ion peak at m/z 702.4, which is present in both spectra, corresponds to the T1 peptide backbone (ATTCTFSGSEGASK, 1403.6 Da after carboxymethylation of Cys residue). It indicated that the two series of glycopeptides actually originate from the same T1 peptide. The 15.2min fraction corresponds to the T1 peptide with a maximum of 5 hexoses attached, while the same peptide attached with a maximum of 4 hexoses eluted from the column at 15.8min. The extra hexose residue in the 15.2 min fraction increased the hydrophilic character of the peptide, thus resulting in a shorter retention time. Since O-linked glycosylation can occur at any Ser or Thr residues, any one of the 6 Ser/Thr residues present in the T1 amino acid sequence could be potential *O*-linked glycosylation sites.

The *O*-linked glycosylation site mapping was performed on the purified *O*-linked glycopeptide. The fractions from the C18 column in between 15 and 16min were collected based on the retention time of the two *O*-linked glycopeptides (15.2min and 15.8min,

respectively). The purified O-linked glycopeptide was first analyzed by MALDI-MS to ensure the presence of O-linked peptides (data not shown), and then analyzed by ECD-MS/MS (Figure 4.5) on a 9.4 Tesla FTICR mass spectrometer for site mapping.

The common fragmentation techniques used in MS/MS experiments (such as CID) are based on vibrational excitation, in which energy is rapidly redistributed over many available degrees of freedom. This is known as intramolecular vibrational energy distribution. Under such condition, the fragmentation is most likely to occur at the weakest bonds. Due to thermal lability, glycosidic bonds are the ones most easily cleaved. Thus, it is difficult to obtain information about the glycosylation sites by CID. On the other hand, the ECD process is based on ion-electron recombination, and cleavage occurs prior to energy distribution. Cleavage of the backbone N-C_α is often observed, except at the N-terminal side of proline[16] (due to the presence of a tertiary nitrogen). The ability to preserve labile groups makes ECD an important tool for the studying of glycosylation as well as phosphorylation [25].

Two precursor ions T1 + 4hexose and T1 + 3hexose were fragmented by ECD MS/MS. (Figure 4.6). Since the energy absorbed by the precursor ion to generate the fragmentation is fairly weak, the precursor ions remain the dominant peaks in the resulting MS/MS spectra. Several fragment ions such as the c13, y10, and z10 ions were detected. The masses of each fragment ion contain the *O*-glycan mass since the glycans remain on the peptide backbone during ECD MS/MS analysis. Thus, these fragment ions contain information about the location of the *O*-glycan location. For example, the y10 ion from the T1+ 4hexoses precursor ion corresponds to the peptide segment **T**F**S**G**S**EGA**S**K (970.0 Da), which contains 4 potential *O*-linkage sites (1 Thr and 3 Ser) (table 4.1). The mass difference between this fragment ion (1618.64 Da) and the calculated mass (970.0 Da) is 648.64 Da, which equals the combined mass from the 4 hexoses.

Therefore, based on this fragment ion, there are 4 hexoses attached to the segment **T**F**S**G**S**EGA**S**K. The other fragment ions were analyzed in a similar manner. For example, analysis of the precursor ion of T1 + 3 hexoses reveals that 3 hexoses are attached to the segment **S**G**S**EGA**S**K. Comparison of both precursor ion fragment ions indicates that the Thr[5] **(**ATTC**T**[5]FSGSEGASK) is one of the O-linked glycosylation sites and it is attached to a single hexose. The fragment efficiency of ECD is lower than the other dissociation techniques due to the nature of the charge neutralization after capture of the electron by the polycations (peptides), thus the fragment ion peaks in the ECD spectrum are very weak. The obtained data only permitted the direct identification of this one O-linked glycosylation site. However, the complete interpretation of the O-linked glycosylation is possible on the condition that fragment information can be obtained from other precursor ions, such as T1 + 5 hexose, T1 + 2 hexose and T1 + 1 hexose. On the other hand, there is evidence that fungal *O*-glycan sites are usually occupied by a single monosaccharide residue [26], a finding confirmed by the occupancy of Thr5 site by one hexose. It is possible that each of remaining 3 Ser/Thr residues present on TF**S**G**S**EGA**S**K segment (Ser7, Ser9 and Ser13) are also occupied by a single hexose. The ECD data could not provide any information regarding the other two potential O-linked sites (Thr 2 and Thr 3).

In order to further confirm the ECD-MS/MS data, reductive alkali-catalyzed beta-elimination of the purified O-glycopeptide was performed. The reaction was carried out by dissolving the purified *O*-linked glycopeptide in 70% ethylamine and incubating at 50ºC for 18 hr. This method has been documented as an effective way to label *O*-linked sites with ethylamine following the release of the O-linked glycan without serious break down the peptide backbone structure[27]. The resulting products are stable during collisional induced dissociation

(CID), which provideed the possibility of determining the *O*-linked sites. However, our attempt at using this method to reveal the presence of *O*-linked sites on PGC T1 peptide was unsuccessful. It is worth noting that most modified beta-elimination reactions are designed for the analysis of *O*-GlcNAc linked samples. For other types of O-linked glycosylation, the reaction conditions may not be effective and require modification.

HPAEC-PAD is an established method for glycan composition analysis. Compared with GC-MS, another common method for analysis of glycan composition, HPAEC is relatively simple. The method used in this work is a modified version of the Carlson degradation[28], in which the PGC is heated for several hours in dilut NaOH with an excess of $NaBH_4$. The excess of $NaBH_4$ is used to ensure fast reduction of released oligosaccharides, thus protecting them from further degradation under alkaline conditions. After releasing the *O*-linked glycans, acid hydrolysis using TFA was performed to cleave any glycosidic bonds between monosaccharide residues. The HPAEC-PAD result is showing in Figure 4.7. Only one peak was observed for PGC, which suggested that it contains only one type of monosaccharide. To confirm the identity of this monosaccharide, a mannose standard was analyzed in the same manner. The peaks present in both chromatograms overlap with each other, thus, the monosaccharide in PGC is determined to be mannose. There is very little direct evidence about the composition of fungal *O*-linked glycan. The results presented here indicat that *O*-mannose is one type of fungal glycosylation.

The presence of *O*-linked glycosylation had not been found in recombinant *A. Niger* EPGI and EPGII, but is observed on *A. Niger* PGA and PGE. The *O*-linked glycosylation present on those EPGs are all located at the N-terminal, T1 peptide. LC-MS data for PGA *O*-linked glycopeptides suggests the glycan is most likely the mannose, as seen for PGC, since the

peak difference between each glycoforms is 162 Da[29]. The initial study of PGE glycosylation by LC-MS indicates the possible presence of *O*-GlcNAc (or *O*-GalNAc) as well as *O*-mannose. O-linked glycosylation was also observed from overexpressed *A. niger* pectate lyase, which was predicted to be located at either the N-terminal (T1) and/or T12–13 peptides[30].

Conclusion:

The *N*- and *O*-linked glycosylation sites on PGC and the structures of the carbohydrate chain were determined by various mass spectrometric methods. A single *N*-linked glycosylation site (N220) was found in the recombinant PGC, and the attached glycan was determined to have a high mannose structure ($GlcNAc_2$ Man_{3-11}). T1 peptide was identified as an *O*-linked glycopeptide from LC-MS data based on the heterogeneity of glycoforms, and the attached glycan was further identified as mannose by HPAEC-PAD experiment. The Thr 5 residue was determined to be attached with a single mannose from ECD MS/MS data, and it is highly possible that each one of the other three potential *O*-linked sites, Ser7, Ser9 and Ser13, is occupied by a single mannose respectively. Figure 4.8 shows the PGC amino acid sequence and *N*- and *O*-linked glycosylation sites.

The combined molecular masses from both *N*-linked and *O*-linked glycans was 3000 Da, which accounted for the majority of the mass discrepancy between the experimental and the calculated molecular mass for PGC (~3300 Da). There are several reasons may cause the remaining 300Da mass difference. The main reason is that the carbohydrate heterogeneity distributes the PGC molecular ion envelope over a wider m/z range when its molecular mass is detected by MALDI-MS, and this fairly broad peak (ΔM at 50% height approximately equal 1000) prohibits an accurate determination of the intact protein molecular mass. Secondly, it may

be explained by the possible presence of phosphorylation, which was detected by the ELISA but has not been detected by mass spectrometric analysis. The peptide fingerprint mapping was performed based on MALDI and LC-MS data and 14 out of 15 tryptic peptides were identified. Only the T9 peptide (9980.7 Da) with amino acid residues from 119 to 213 could not be detected. If phosphorylation is presenton PGC, the phosphate groups can only be attached to this peptide.

Acknowledgments:

This work is supported by the grants from NIH and NSF.

Special appreciation to Dr. Mark R. Emmett in the National High Magnetic Field Laboratory, Florida State University for the support of ECD-FTICR MS experiments.

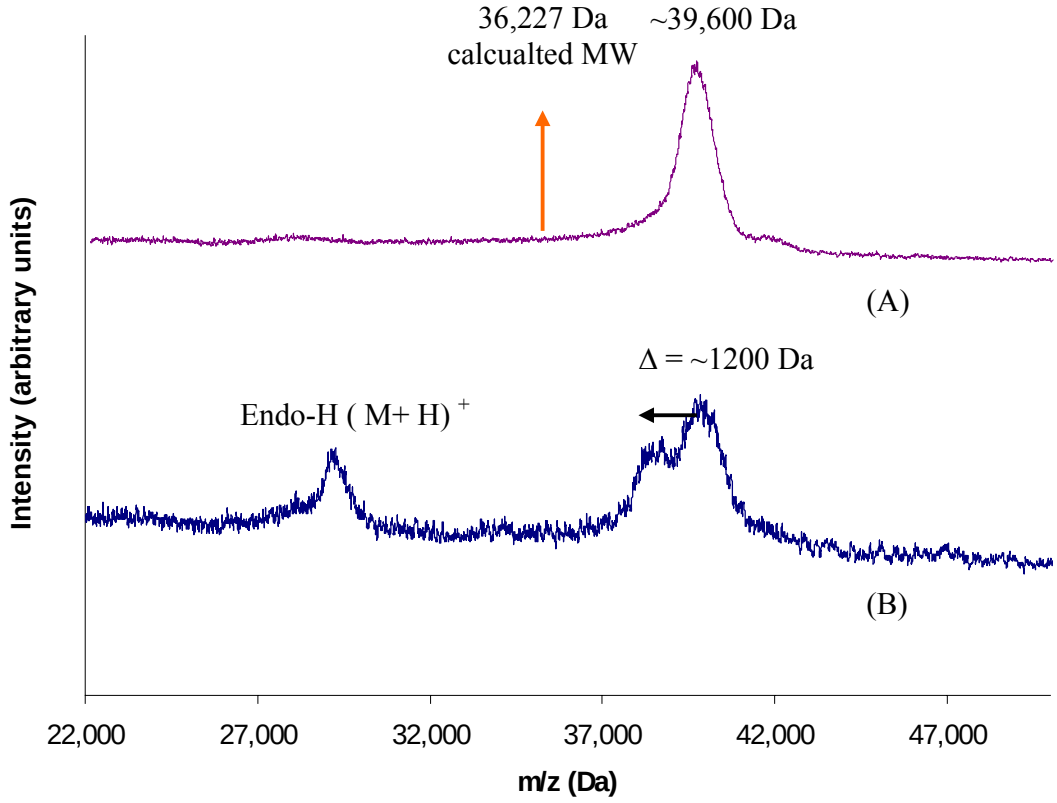


Figure 4.1: MALDI-MS spectra of (a): intact PGC protein and (b): the Endo-H digested PGC

The molecular weight of intact PGC was approximately 39600 Da, 3300 Da larger the molecular weight calculated from amino acid sequence. After Endo-H on target digestion for 30min, a new peak with approximately 1200 Da less appeared. But the intact protein still was the major peak which suggested the digestion was not complete. Another peak around 29000 Da was the Endo-H molecular ion peak.

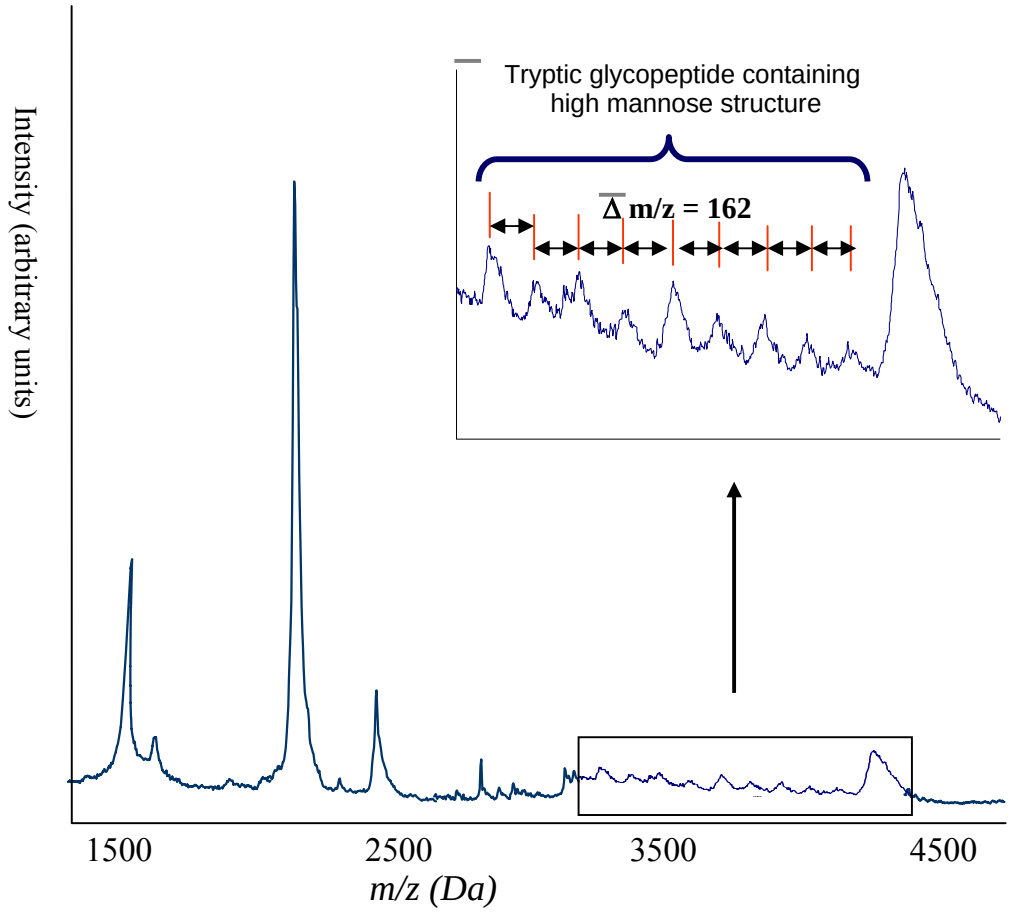


Figure 4.2: MALDI-MS spectrum of trypsin digested PGC peptide mixture

The series of peaks with average mass difference of 162 Da indicated this is a glycopeptide containing high mannose structure carbohydrate

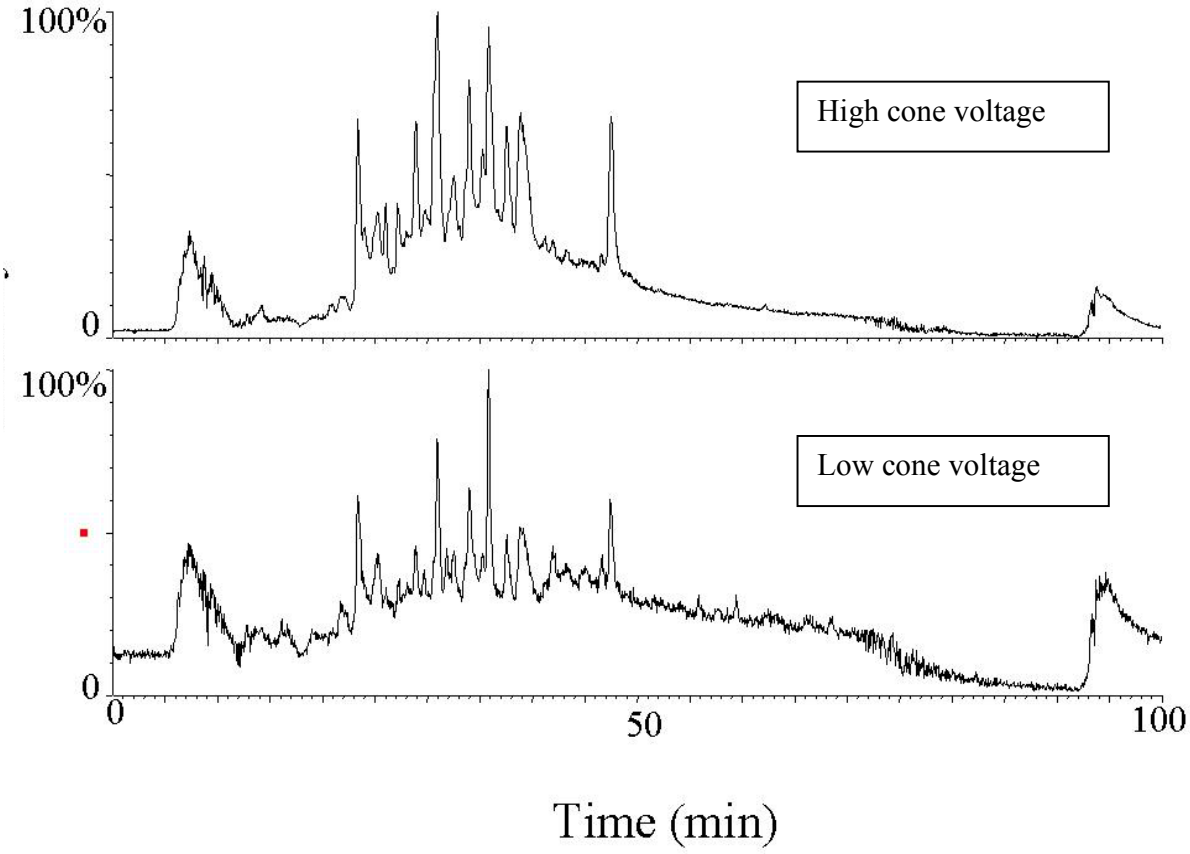


Figure 4.3a: Total ion current (TIC) for two MS acquisition functions to detect N-linked glycopeptide

High cone voltage = 80v, ion range from m/z 50-400
Low cone voltage = 30v, ion range from m/z 400-2500

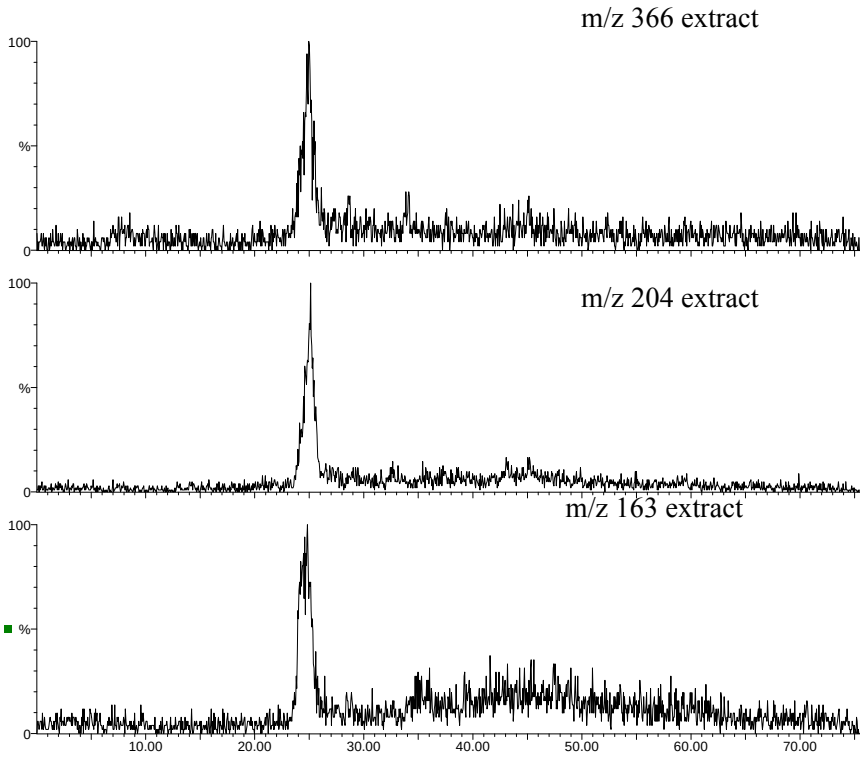


Figure 4.3b: Ion currents of three carbohydrate fragment ions extracted from TIC under high cone voltage acquisition

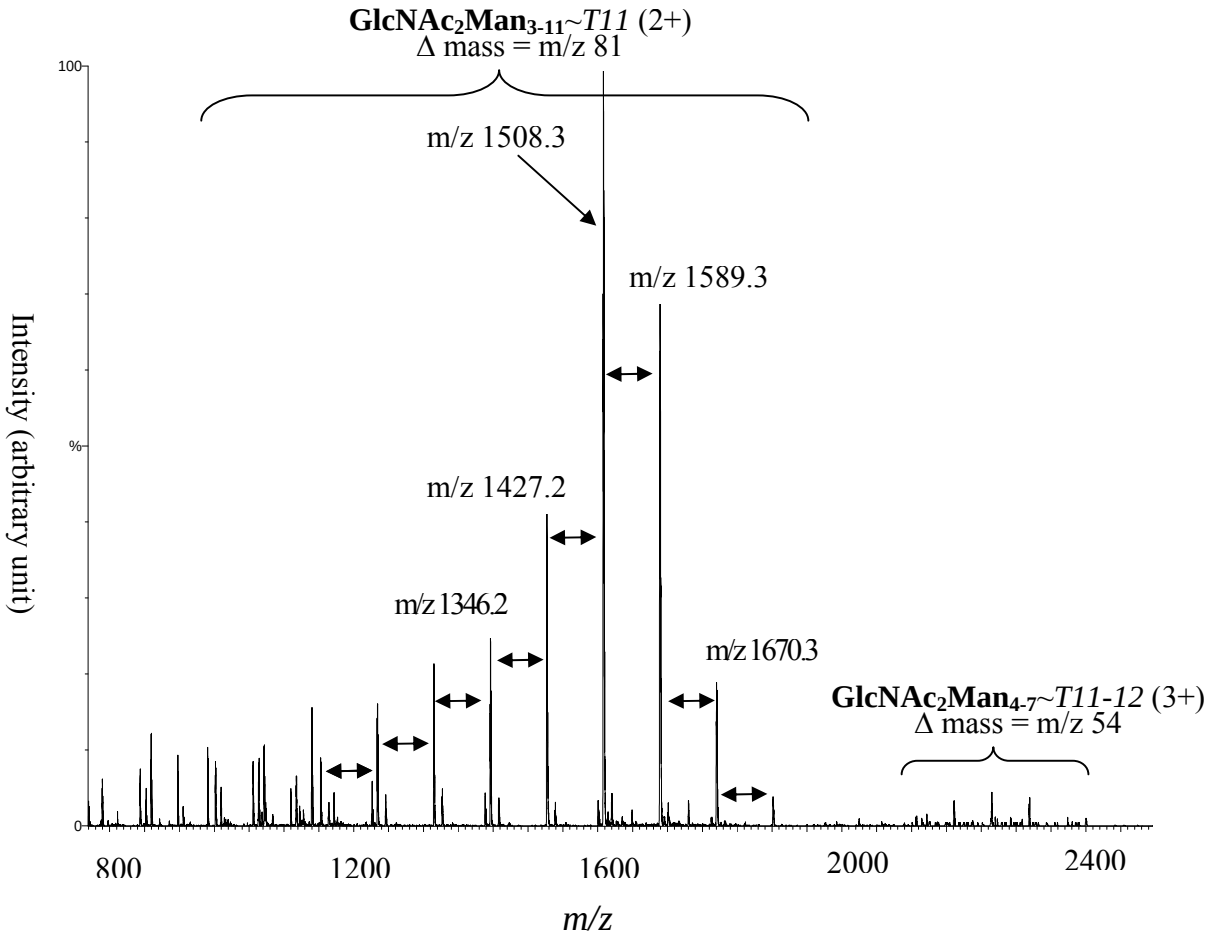


Figure 4.3c: The averaged spectrum of N-linked glycopeptide

The 2+ charged peaks had m/z 81 incremental between peaks, while 3+ charged peals has m/z 54 incremental. Each species within both series was separated by a single hexose residue

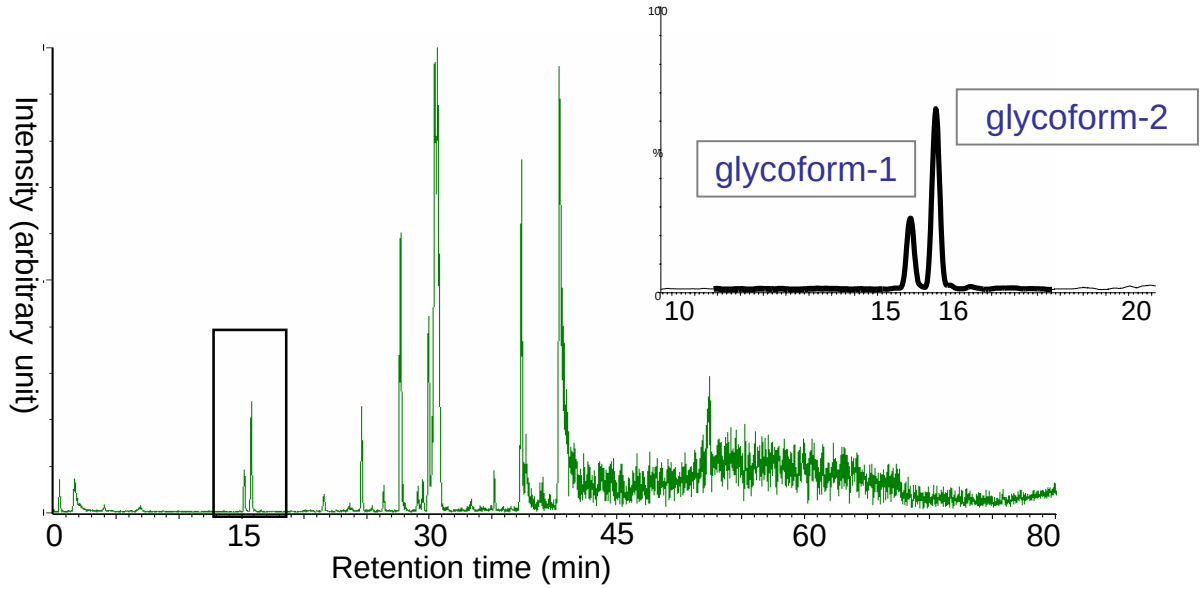


Figure 4.4a: Total ion current of PGC tryptic peptides obtained under conventional ESI LC-MS

Effluents at 15.3min and 15.8 min contain the glycopeptides. The purified O-linked peptide was collected in the time window of 15-16min.

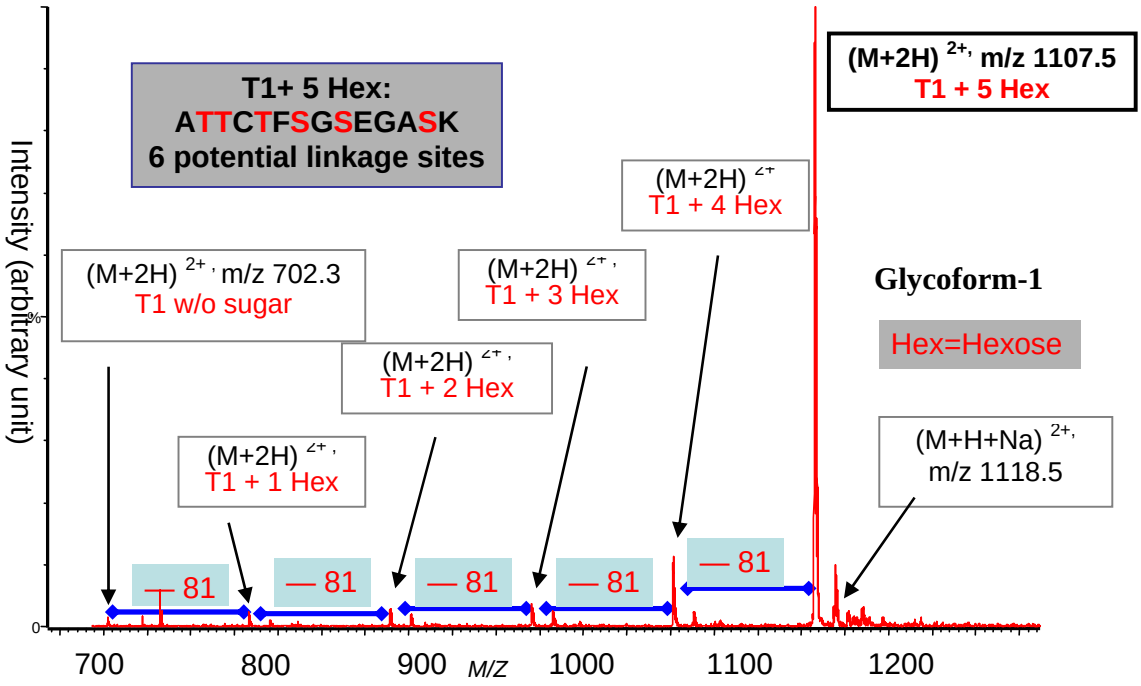


Figure 4.4b: Mass spectra of the O-glycopeptide with retention time at 15.3min

T1+ 5Hex as the domiant ion

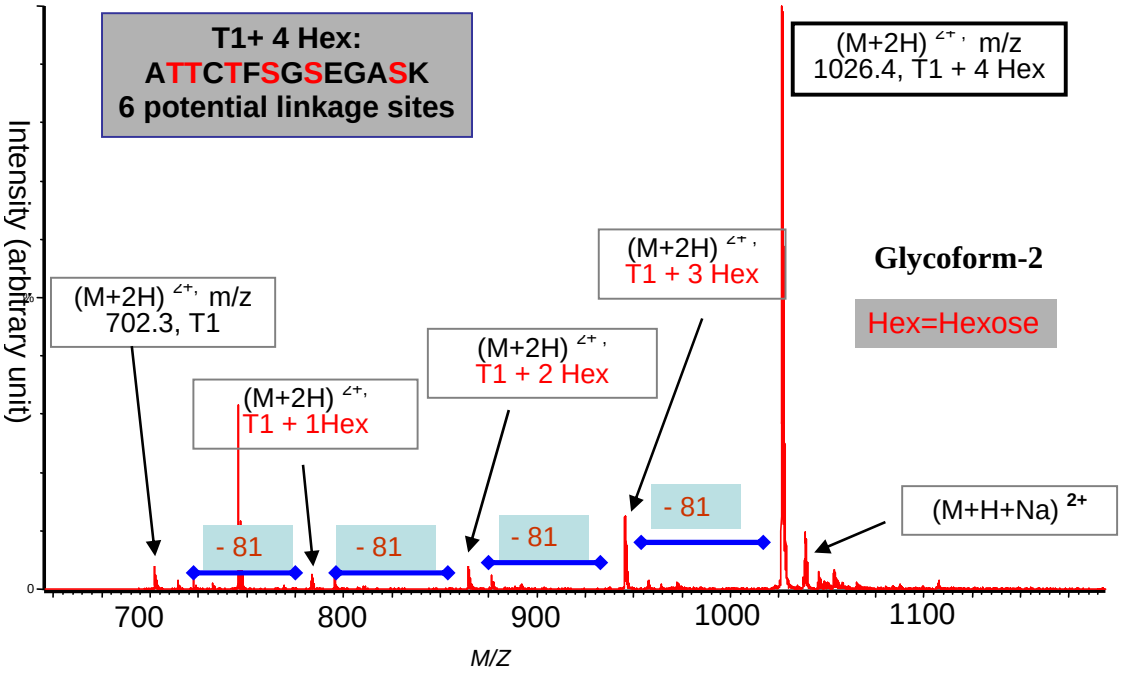


Figure 4.4c: Mass spectrum of the O-glycopeptide with retention time of 15.8min T1 + 4 Hex as the dominate ion

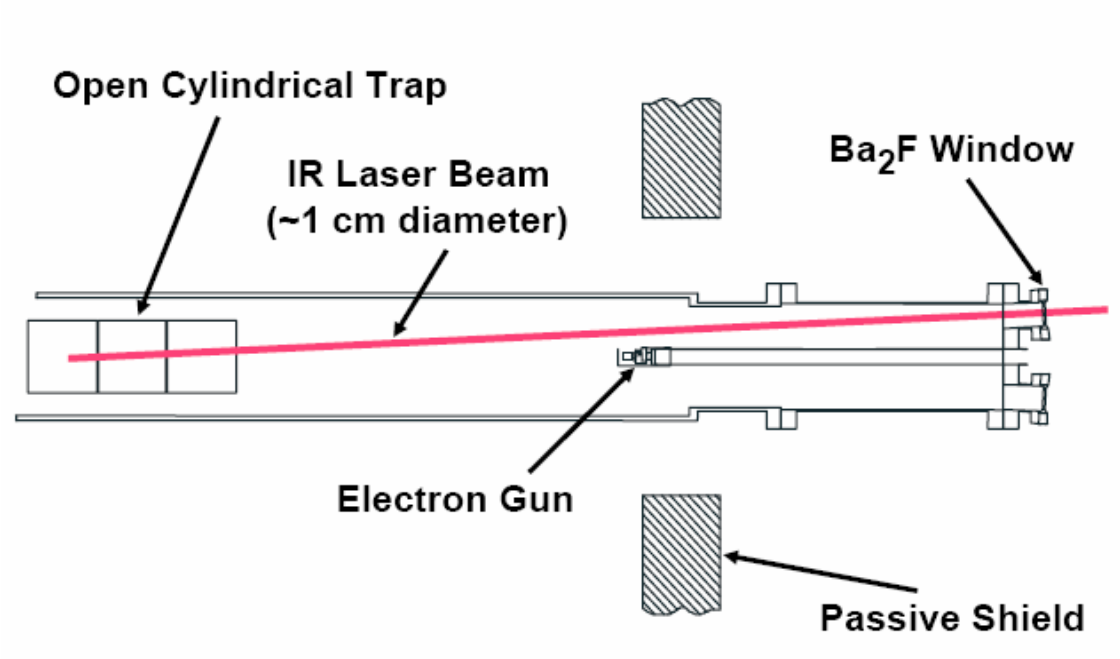


Figure 4.5 The ECD source for FTICR instrument (provided by FSU)

(Infrared Multiphoto Dissociation source can be used along with ECD source for MS/MS analysis to generate complimentary fragmentations)

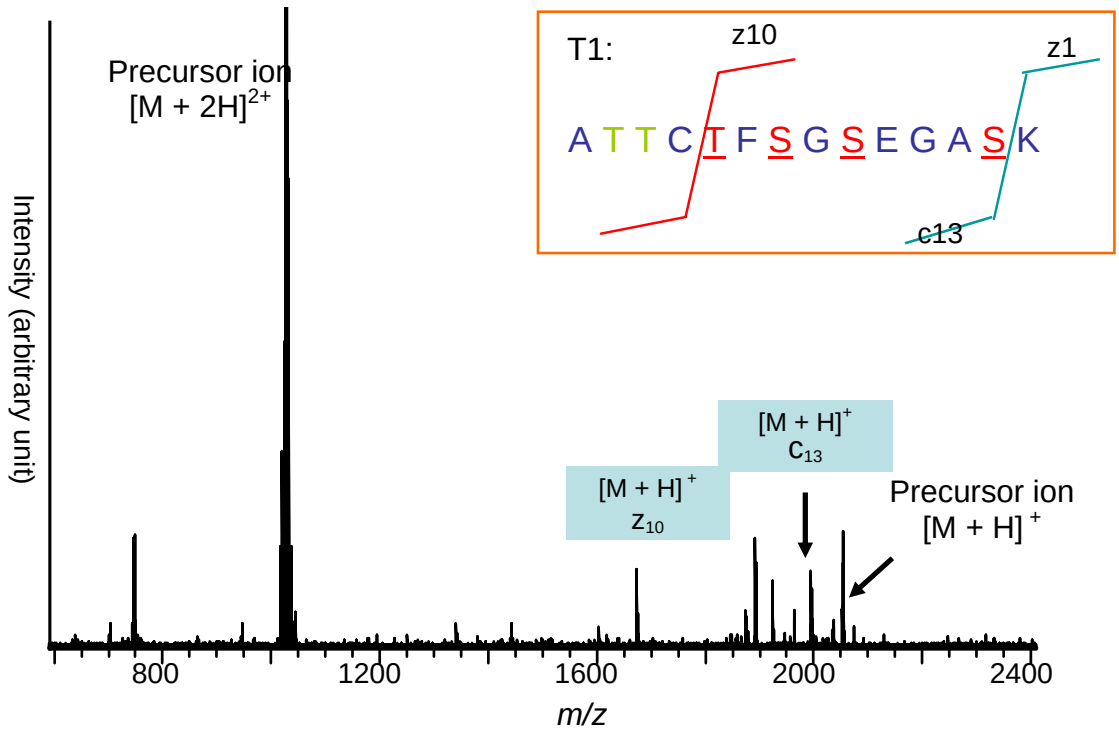


Figure 4.6: ECD-MS/MS spectrum of the precursor ion of T1 + 4 hexoses

The 1+ and 2+ charged precursor ion remained the major peaks with some c and z ions as fragment ions.

Table 4.1: The ECD MS/MS fragment ions and the predicted hexose linkage sites on the peptide backbone

Precursor	Selected fragment ions	Glycan attachment	Conclusion
T1 + 4 Hex	Y10 = 1618.66 Da	4 Hex on TFSGSEGASK	4 Hex located within animo acid 5-13 sequence: T F S G S E G A S which contains 4 potential linkage sites
	C13=1922.74 Da	4 Hex on ATTCTFSGSEGAS	
	z10=1602.64 Da	4 Hex on TFSGSEGASK	
T1 + 3 Hex	z10=1440.57 Da	3 Hex on TFSGSEGASK	3 Hex located within animo acid 5-11 sequence: T F S G S E G which contains 3 potential linkage sites
	z12=1407.22 Da	3 Hex on TCTFSGSEGASK	
	C11=1602.64 Da	3 Hex on ATTCTFSGSEG	
T1 sequence	**A T T C T F S G S E G A S K** Hex Hex Hex Hex Hex		

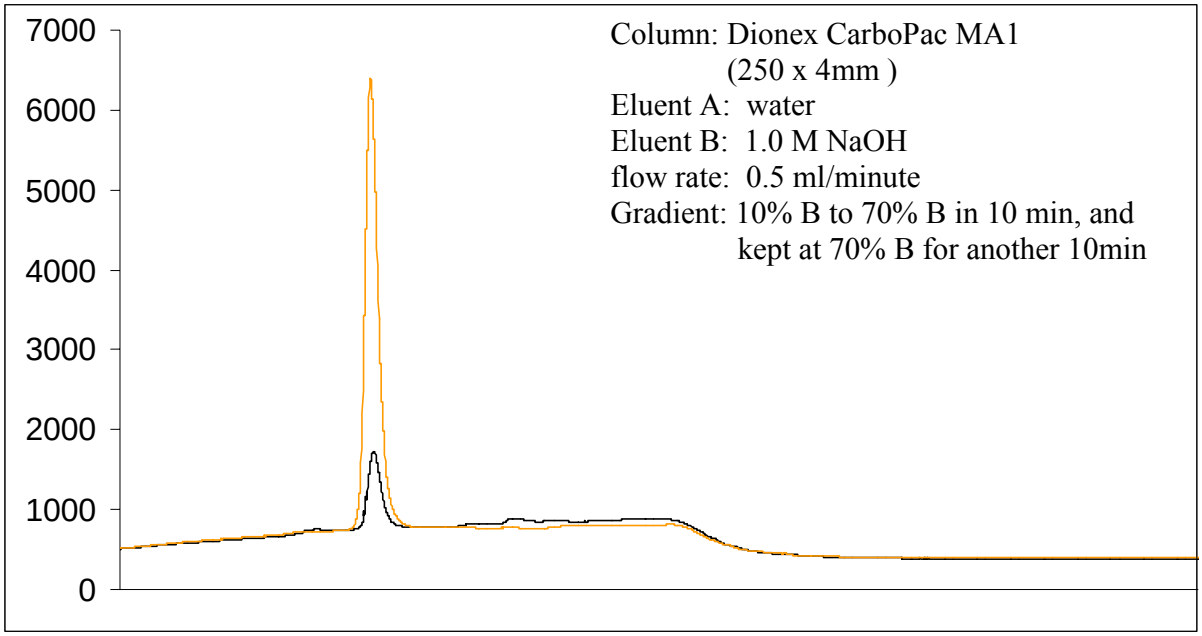


Figure 4.7: HPAEC-PAD of *O*-lined glycan composition analysis

(black line: PGC O-linked glycan; yellow line: mannose standard)

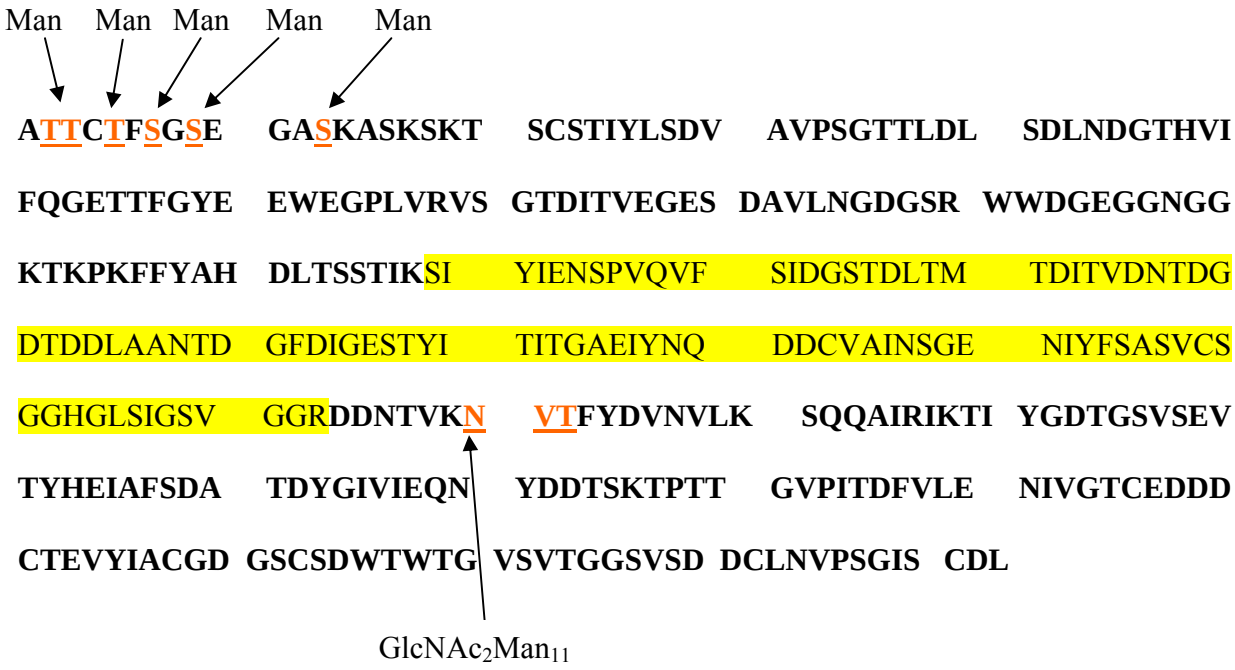


Figure 4.8: The amino acid sequence of PGC containing 1 *N*-linked glycosylation and five *O*-linked sites as indicated by the arrows. T9 (highlighted by yellow color) couldn't be detected by mass spectrometric method and was suspected to contain phosphorylation.

Reference List

1. A.G.J.Voragen, W. P. J. F. T. M. A. V. A. C. M. G. C. R. *Food Polysaccharides and their Application*, A.M.Stephen, Ed.; *Marcel Dekker*: *New York*, 1995.

2. M.A.O'Neill, P. A. A. G. D. *Methods in Plant Biochemistry* , P.M.Dey, Ed.; *Academic Press*: *London*, 1990.

3. H.J.Bussink; F.P.Buxton; B.A.Fraaye *Eur.J.Biochem,* **1991,** *208*, 83-90.

4. Jacques A.E.Benen, Harry C. M. Kester and Jaap Visser *Eur.J.Biochem.* **1999,** *259*, 577-85.

5. L.Parenicova; J.A.E.Benen; H.C.M.Kester; J.Visser *Eur.J.Biochem.,* **1998,** *251*, 72-80.

6. L.Parenicova; J.A.E.Benen; H.C.M.Kester; J.Visser *Eur.J.Biochem,* **2000,** *259*, 577-85.

7. L.Parenicova; H.C.M.Kester; J.A.E.Benen; J.Visser *FEBS* **2000,** *467*, 333-36.

8. J.A.E.Benen; G.J.W.M.van Alebeek; A.G.J.Voragen; J.Visser *Advances in pectin and pectinase research*, F.Voragen; H.ScholsR.Visser, Eds.; Kluwer Academic Punlishers: Dordrecht, The Netherlands, 2003; Chapter 4.

9. Gerrit Limberg, Roman Korner Hans Christian Buchholt etc. *Carbohydrate Research* **2000,** *327*, 293-307.

10. J.Colangelo, V. Licon J. Benen J. Visser C. Bergmann R. Orlando *Rapid Commun.Mass Spectrom.* **1999,** *13*, 2382-87.

11. Archer DB, Oeberdy JF. *Crit.Rev.biotechnol.* **1997,** ***17***, **273**.

12. Yang, C. Bergmann J. Benen R. Orlando. *Rapid Commun.Mass Spectrom* **1997,** *11*, 1257-62.

13. Huddleston, M. J.; Bean, M. F.; Carr, S. A. *Anal.Chem.* **1993,** *65*, 877-84.

14. Ritchie M.A.; Gill A.C.; Deery M.J.; Lilley K. *J.Am.Soc.Mass Spectrom.* **2002,** *13*, 1052-64.

15. Hunter A.P.; Games D.E *Rapid Commun Mass Spectrom.* **1995,** *9*, 42-56.

16. A.A.Zubarev; K.F.Haselmann; B.Budnik; F.Kjeldsen; F.Jensen *Eur.J.Mass Spectrom.* **2002,** *8*, 337-49.

17. Emmett, Mark R. *Journal of Chromatography A* **2003,** *1013*, 203-13.

18. Bussink, H. J. D. Kester H. C. M. and Visser J. *FEBS Lett.,* **1990,** *273*, 127-30.

19. Colangelo J and Orlando R *Anal.Chem.* **1999,** *71*, 1479-82.

20. M.W.Senko, C. L. Hendrickson L. Pasa-Tolic J. A. Marto F. M. White S. Guan and A. G. Marshall. *Rapid Commun.Mass Spectrom.* **1996,** *10*, 1824.

21. Hardy, M. R. and Townsend R. R *Methods Enzymol.* **1994,** *230*, 208-25.

22. Biely, P; Benen, J. A. E. Heineichova K; Kester, H. C. M; and Visser *J.FEBS lett.,* **1996,** *382*, 249-55.

23. Armand S., Wagemarker M. J. M. Sanchez-Torres P. Kester H. C. M. etc. *J.of Biological Chemistry* **2000,** *275*, 691-96.

24. M.E.Warren, H. C. M. Kester J. Benen J. Colangelo J. Visser C. Bergmann R. Orlando *Carbohydr.Res.* **2002,** *227*, 803-12.

25. Emmett, Mark R. *Journal of Chromatography A* **2003,** *1013*, 203-13.

26. P.E.Van Den Steen; P.M.RUdd; M.R.Wormald; R.A.Dwek; G.Opdenakker *Trends in Glycoscience and Glycotechnology* **2000,** *12*, 35-49.

27. F.Hanisch; M.Jovanovic; J.Peter-Katalinic *Anal.Biochem* **2001,** *290*, 47-59.

28. Carlson, D. M. *J.Biol.Chem.* **1968,** ***243***, 616-26.

29. M.Xie; K.V.S.Kolli; J.A.E.Benen; J.Visser; C.Bergmann; R.Orlando *2003 ASMS conference*, 2003 317.

30. J.Colangelo, V. Licon J. Benen. J. Visser C. Bergmann1 and R. Orlando *Rapid Commun.Mass Spectrom.* **1999,** *13*, 2382-87.

Chapter 5. Conclusions

The glycosylation of various endopolygalacturonases (EPGs) were successfully characterized by the implementation of several mass spectrometric techniques. Matrix-assisted laser desorption/ionization-mass spectrometry (MALDI-MS) was employed to measure the glycoprotein masses in order to determine the degree of glycosylation. Combined with on-target endoglycosidase digestion, MALDI-MS could also provide information about whether *N*-linked glycosylation is present, and such information can further be used to evaluate the possible presence of *O*-linked glycosylation. The glycoproteins were then digested with an endoproteinase and the resulting peptide mixtures were analyzed by MALDI-MS and LC-electrospray MS (LC-ESI/MS) using a stepped orifice voltage techniques to identify the glycopeptides ad to map the protein backbone. The glycopeptides were analyzed by collisional-induced dissociation (MS/MS) to determine the primary carbohydrate structure.

The *N*-linked glycosylations presented at various EPGs were readily analyzed by these techniques. PGC expressed from *A. niger* was found to have one *N*-linked glycosylation site, which was occupied by nine different high manse strucutures, each containing three to eleven mannose residues. BcPG3 expressed from *Botrytis cinerea* was found to have three out of four potential *N*-linked glycosylation sites, which were also occupied by a variety of high mannose glycans. Five out of seven potential *N*-linked glycosylation sites present on BcPG6 was found to be occupied, and each site was attached to several high mannose structures. Complete amino acid sequence of BcPG6 was identified from peptide fingerprint mapping, therefore, there is no *O*-linked glycosylation presence.

O-linked glycosylations were detected to present at BcPG3 and PGC by LC-convention ESI/MS technique using higher flow rate (20-50 uL/min) for peptide mixture separation. The identification of O-linked glycopeptides were achieved based on While N-terminal peptide T1 of

PGC was detected as O-linked glycopeptide, C-terminal peptide T16 of BcPG3 was identified as O-linked glycopeptide. Using the same LC-MS setup, O-linked glycosylatons were also detected from other EPGs including *A. niger* PGA and *B. cinerea* PGE. The composition of *O*-linked glycan present at PGC was determined by HPAEC-PAD to be mannose. To further mapping O-linked sites present at PGC T1 peptide, the purified glycopeptide was analyzed by electron-capture dissociation (ECD) MS/MS. The ECD fragmentation is relatively weak as compared to that of CID, but the preservation of glycosidic bonds during ECD makes it a useful technique for glycosylation site mapping. Two glycopeptide precursor ions were analyzed and their data revealed that one Thr residue was attached with a single mannose, and it was highly possible that each of the other three potential sites was also attached with a single mannose residue. However, the lack of the MS/MS data for several other glycopeptide precursor ions does not permit the identification of the remaining two potential sites.

Besides the N- and O-linked glycosylation, other post-translational modifications were also analyzed. Several disulfide bonds were detected form BcPG6 and BcPG3 samples based on LC-MS data. This observation is in agreement of other EPG expressed from *B. cinerea*. Phosporylation was found to present at PGC from phosphorylation antibodies screening, which was further determined to be located at the potential sites within T9 peptide based on the peptide fingerprint mapping.