ANALYSIS OF GLYCANS AND GLYCOPROTEINS BY MASS SPECTROMETRY

By

Dilusha S. Dalpathado

PhD., University of Kansas 2007
Submitted to the Department of Chemistry and the
Faculty of the Graduate School of the University of Kansas
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy

_____
Heather Desaire (Chairperson)

_____
Cindy L. Berrie

_____
Michael A. Johnson

_____
Julian Limburg

_____
Jennifer S. Laurence

Date defended _____

The Dissertation Committee for Dilusha S. Dalpathado certifies that this is the approved version of the following dissertation:

ANALYSIS OF GLYCANS AND GLYCOPROTEINS BY MASS SPECTROMETRY

Committee:

_____
Heather Desaire (Chairperson)

_____
Cindy L. Berrie

_____
Michael A. Johnson

_____
Julian Limburg

_____
Jennifer S. Laurence

Date Accepted _____

Abstract

Dilusha S. Dalpathado, PhD
Department of Chemistry, December 2007
University of Kansas

Glycosylation is among the important post-translational modifications that occurs in proteins. In particular the heterogeneity of the glycan moiety can affect a range of physio-chemical properties of a glycoprotein. Thus the increasing interest in glycan function has created a need for sensitive detection and analysis.

Mass spectrometric methods are described herein to obtain structural information from glycans. There are two common ways of characterizing glycans by mass spectrometry. The first method involves releasing glycans from protein, purification followed by mass spectrometric analysis. The second approach involves the analysis of glyopeptides after subjecting the glycoprotein to enzymatic digestion.

The first approach is widely used as a method of getting structural information. In this approach, glycans are usually derivatized using reductive amination to facilitate detection. An improved reductive amination procedure is introduced herein with less toxic reagents. In these experiments $NaBH(OAc)_3$ was used as the reducing agent which is an alternative to the toxic, but extensively used, reducing agent, $NaBH_3CN$.

The glycan release method is very useful in getting structural information about glycans. However if one needs to obtain an understanding about how these glycan structures affect the function of glycoproteins on a

molecular level, it is important to characterize the glycans in a glycosylation site-specific manner. The feasibility of site-specific analysis is demonstrated here by characterizing glycans of follicle stimulating hormone in two mammalian species.

Finally, a statistical approach (STEP – statistical test of equivalent pathways), which was developed on the ion trap mass spectrometry is extended to a routinely used instrument, the triple quad mass spectrometer. This method is capable of giving structural information for glycans and peptides. The STEP analysis is based on differentiating first generation product ions from the second-generation product ions in a tandem mass spectrum. This information is useful in getting information about ion genealogy.

Together, the three analytical methods developed herein for glycan and glycoprotein analysis serve as useful tools for future researchers who wish to characterize glycosylation on proteins. Additionally, this thesis provides the first direct comparison of site-specific glycosylation profiles of follicle stimulating hormone, isolated from two different species.

Acknowledgement

First, I would like to thank my parents for all their efforts in nurturing me to be the individual I am today. I am very privileged to have parents like you. Thank you very much for all the love and support you have given me. I need to thank both of you for sending me to one of the best schools in Sri Lanka for my early education because that helped me obtain a good foundation. Thaththi, I wish you were here to share this moment with me and make you proud. I miss you. Ammi, I am amazed at how strong you have been and I wish to be as strong as you are. You have been there for me for everything in my life and I thank you for all of your valuable advice and support. I was able to share with you all the ups and downs in my life and you were there to understand and listen to me. Thank you very much for making me believe in myself. Malli, I would also like to thank you for all the encouraging words that you have given me when I needed it. You have grown up to be an individual that I am so very proud of.

My host parents, Dad and Mom, thank you very much for treating me like one of your own kids. You have shown equal love and support that my own parents have shown me and I am very grateful for that. You made my transition from Sri Lanka to the United States very easy for me. I was very lucky and privileged to have you both as my parents in Lawrence. Thank you very much for listening to me and giving me valuable advice and trying to keep my spirits up when ever I felt low.

There are many people that I would like to thank at KU, especially my advisor, Dr. Heather Desaire. It was indeed a challenging choice that I had to make when I joined you as one of your first students. I saw the potential and the capabilities in you and you have gone far beyond my expectations within a very short time. I have been very happy to be a part of your group. Thank you very much for getting me involved in all of the important work and for being a wonderful mentor. Thank you for directing my training to fit the pharmaceutical industry since that's what I told you that I wanted to do after my PhD on the first day when I joined your group. I remember you trying to instill in us that good is the enemy of great, and that we need to always try to do great without settling for good. There were times that I felt that I was under too much pressure from your high expectations but I have come to realize that it was what was best for me. Also, I really appreciate the fact that you send us for conferences because that helped me in building confidence and making interactions with people doing similar work. Thank you very much for getting us to write our own papers because it helped me to transform one of my weaknesses to one of my strengths. Enjoying writing is one of my biggest achievements that I will be taking with me. Thank you very much for appreciating my work and making me a stronger individual.

Hui, you were one of my inspirational role models when I initially joined the group. I always admired your humble and calm nature and your willingness to help at anytime. Thank you very much for all the advice and

support that you have given me. I am very fortunate to have been associated with an individual such as you at that early stage of my career. Eden, thank you for coming to our group after Hui left and for sharing all of your expertise and knowledge. Your high standards in your work and training have helped me do things in a more perfect way. Thank you for all your advice and support and sharing with me your experiences because that has certainly helped me to become a better individual.

I wish to extend a big thank you to all group members, Mary, Janet, Ying, Lin, Melinda, Katie, Qing and Sushma for supporting me in different ways. Mary, I really admired how you came to work so early and got a lot done in the morning. It has been quite a struggle instilling this habit in me but I have managed to do it now also, which I am very happy about. Thank you for giving me helpful suggestions about my research. Janet, I was very fortunate to work closely with you because there was a lot that we were able to share. I enjoyed doing experiments with you and as a team we were able to achieve a lot. Apart from chemistry, our discussions about life made it a very enjoyable and a pleasant experience for me. Thank you for all the support and suggestions that you have given me. Ying, I consider myself fortunate to work with you too. I have always admired your hard working nature and wished that I could work like you. I enjoyed our conversations and I have learned a lot from you. Lin, I enjoyed doing experiments with you and have learned a lot from you, Thank you. Melinda and Katie, I appreciate you

both trying to help me out even with your busy schedules. Also our discussions on preparing for teaching helped me out a lot since it saved a lot of time. Qiang, I enjoyed mentoring you and thank you for working with me since I was able to develop my mentoring skills. Sushma, even though I haven't had to chance to work with you, thank you very much for all of your support and encouraging words. I also would like to thank Marcus and Colleen for working under me because I learned a lot from teaching you.

In addition I would also like to thank Dr. Todd Williams for helping me out in many ways. Thank you very much for teaching me various things on different mass spectrometers. I was able to learn a lot from the conversations that we had. Also I would like to thank Dr. Wynn Esch for all of your time spent with me discussing the fundamentals of mass spectrometry. This information was very useful for me especially when I was studying for my orals. Dr. George Bousfield is also another person that I would like to thank for giving me excellent suggestions about my research.

I cannot fail to gratefully thank Hemamali and Luke for their support in helping me through the graduate coursework. Our little study group helped me survive the cumes and course work successfully. Hemamali, I knew you from the time I arrived at KU and thank you for being such a wonderful friend and for all of the advice and support you have given me. Also thank you for offering your delicious Sri Lankan meals whenever I felt hungry for them. In addition I need to thank all my room mates Nan, Gauher, Anne, Naomi,

Priyanka and Laura for making my environment at home enjoyable and pleasant. I also want to thank Mauricio, Valentino, Albert and all my Sri Lankan colleagues at KU for all of their support and encouraging words.

Dr. Rohan Wimalasena has been one of my inspirational models with whom I have shared very important decisions since coming to the United States. Thank you very much for introducing me to KU and all the advice you have given me. I also want to thank Prof. George Wilson for corresponding with me while I was in Sri Lanka and trusting me with my credentials. I would like to offer a big thank you to the Department of Chemistry at KU for the training and education given to me. I will always be proud to be a Jayhawk. I also wish to extend my gratitude to several of my Sri Lankan professors such as Dr. Veranja Karunaratne, Prof. J. N.O. Fernando, Dr. Jayanthalal Ratnasekara, and Dr. G. Bandarage all for whom I have great respect. It has been a pleasure working with each of you.

Last but by no means the least, I would like to thank Mahen for supporting me in numerous ways. Thank you for being available to listen to me when I was stressed and for giving me valuable advice. You have motivated me on things that seem very challenging and it has been a pleasure sharing with you all of the ups and downs in my life. Thank you for making me laugh and helping to take the stress away from me. You have definitely inspired me to be more productive.

Table of Contents

CHAPTER 1

Introduction to the Analysis of Glycosylated Proteins

1.1. Introduction

Glycomic analysis is an emerging technology that promises to contribute to diverse areas of science, including early disease detection, vaccine development, and control of the reproductive and immune systems. Glycomics studies focus on glycosylation which is one of the most common and important post translational modifications that occurs in proteins, and it has a direct impact on protein function.[1, 2] Abnormal glycosylation is correlated with several disease states such as cancer, inflammatory diseases, and congenial disorders etc.[3-5] Glycans are also being used as targets for designing vaccine candidates for AIDS [6-8] and cancer.[9-11] While the importance of studying glycosylation has clearly been demonstrated, the glycosylation state of most proteins, along with the role of the glycans on those proteins, is still unknown. For instance, although 50% of the total number of proteins entered into the SWISS-PROT database are expected to be glycosylated, only ~ 10% of these proteins' glycans have been characterized [12], which indicates that significantly more research in glycosylation analysis is necessary, in order to better understand the structural and functional roles of glycans. There is no doubt that contribution from glycomics will facilitate promising new developments that could be used for disease diagnosis and management.

1.2. Types of glycans

Depending on the glycan attachment sites in the polypeptide chain, glycosylation can be divided into N- and O- linkage types.[13] In N-linked glycosylation, glycans are attached to asparagines in the sequence containing Asn-Xxx-Ser/Thr, where Xxx corresponds to any amino acid except for proline.[14] However, in a limited number of glycoproteins, presence of a cysteine has been identified instead of Ser/Thr in the amino acid sequence.[15] O-linked glycosylation, consists of attaching the glycans to a serine (Ser) or a threonine (Thr), and this glycosylation does not require a unique sequence such as in the case of N-linked glycans.[14]

Glycosylation takes place in the endoplasmic reticulum (ER) and golgi apparatus and is catalyzed by a series of enzymes.[1, 13] All N-linked glycans are derived from the precursor $Glc_3Man_9GlcNAc_2$, which is attached to the protein during translation. While many of the pendant residues on this initial building block are eventually removed, a well-defined trimannosyl pentasaccharide core ($Man_3GlcNAc_2$) always remains (Figure 1.1), even after glycan modification.[13]

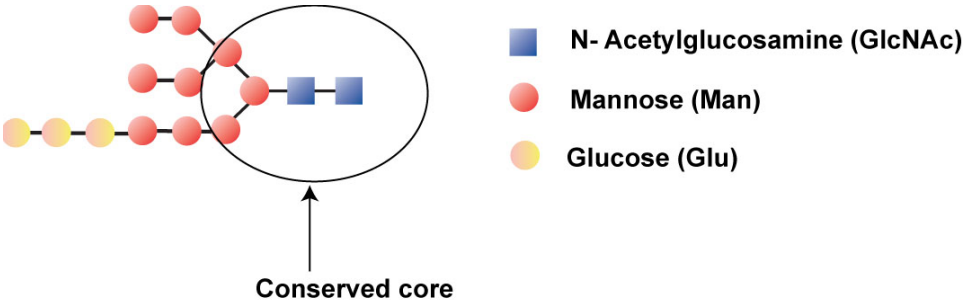


Figure 1.1: N- gycan precursor for all N-linked glycans. Adapted from reference 13.

This precursor is subjected to various modifications in the ER and Golgi apparatus, which results in three types of glycan structures; high mannose, complex and hybrid.[16] High mannose structures are formed from the precursor by cleaving the mannoses with α-mannosidases, without the addition of any new sugars.[13] On the other hand, complex type glycans are formed both by removal and addition of sugars and are characterized according to the number of branches attached to the terminal mannose residues on the core structure.[13,16] Depending on the number of branches attached, complex type glycans could be subdivided to bi-, tri- and tetra-antennary structures.[16] Hybrid type glycans shares structural features both from high mannose and complex type glycans.[13] The a pictoral representation of these glycans is shown Figure 1.2.

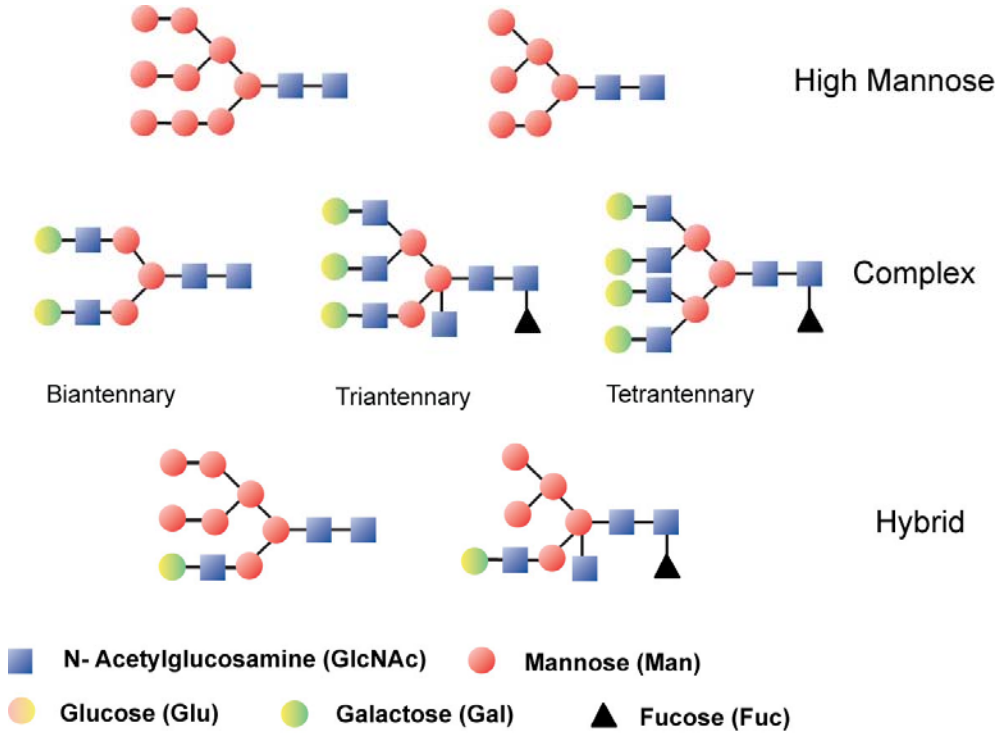


Figure 1.2: Three groups of N-linked glycans structures

1.3. Challenges in analyzing glycans

Glycan analysis can be challenging due to several reasons. For instance, unlike in the case of proteins, glycan biosynthesis is governed by a series of enzymes present in the cell, and it does not rely on an underlying template.[17] As a result, even though the cell can reproducibly generate a single amino-acid sequence each time it synthesizes a specified protein, the glycosylation does not conform to a single structure. It is typically very heterogeneous. Additionally, the enzymes responsible for glycosylation can be affected by the physiological state of the cell, thereby adding in even more variability to the resulting glycan profile and also allowing for the possibility of

modified profiles in certain disease states, where the cellular environment affects the glycosylation.[13] Furthermore, analysis of glycans has been very challenging due to the limited quantities that are released from glycoproteins.

1.4. Methods of characterizing glycans and glycopeptides by mass spectrometry

Mass spectrometry has proven to be an important tool for glycan analysis over the past few years, due to its high sensitivity, selectivity, and throughput. Early work in this area was performed using GC-MS, and these techniques are still being used today for detection of monosaccharide compositions.[18, 19] Due to the nonvolatility of underivatized oligosaccharides, a significant expansion in glycosylation analysis by mass spectrometry occurred with the advent of soft ionization techniques such as fast atom bombardment (FAB), electrospray ionization (ESI) and matrix assisted laser desorption/ionization (MALDI). Additionally, since soft ionization methods typically do not fragment the precursor ions, they are well suited for analyzing complex mixtures. Therefore mass spectrometry generally overcomes the limitations of other methods of glycan analysis, which require substantial purification.

There are two approaches of characterizing glycans using mass spectrometry. In the first approach, glycans are cleaved from the glycoprotein enzymatically or chemically and are subjected to purification prior to mass

spectral analysis.[14] There are several review papers that describe this analytical approach.[14, 20, 21] However, the utility of this method is limited when more than one glycosylation site is present on the glycoprotein, because it fails to correlate glycan structure with the different attachment sites. This information can be retained by using another approach, which is to characterize glycopeptides after subjecting the glycoprotein to proteolysis.[22-24] This method of analyzing glycans is rapidly developing and less established, compared to analysis of released carbohydrates. Using glycopeptide analysis, the glycan structure can be readily correlated to the attachment site, due to the presence of the peptide portion on each glycopeptide. Because the glycosylation information is connected to the glycosylation site in which the glycan is attached, this approach is often referred to as “a glycosylation site-specific analysis.” The information obtained in a site-specific manner can be important in correlating glycosylation profiles with individual glycosylation sites, which is useful in developing an understanding of the protein structure on a molecular level. This type of molecular-scale information is critical in understanding structure-function relationships.

1.5. Steps associated in characterizing glycopeptides by mass spectrometry

Glycopeptide analysis is usually performed in several steps. The schematic diagram illustrating these steps is shown in Figure 1.3. The first step involves isolating glycoprotein from complex biological mixtures.[25-28] The glycoprotein is then subjected to a variety of chemical and biochemical reactions, such as reduction and alkylation (for glycoproteins containing disulfide bonds) and proteolysis; this is followed by purification (glycopeptide enrichment) and mass spectral analysis. [26]

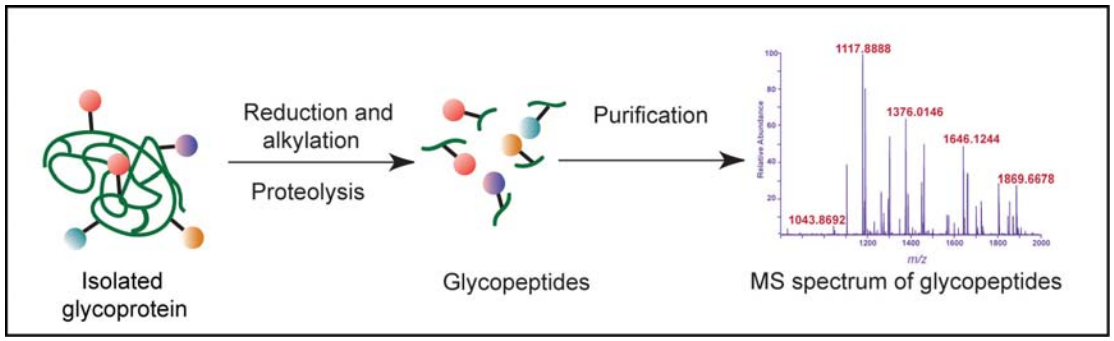


Figure 1.3: Schematic diagram of the steps involved in glycopeptide analysis

1.5.1. Isolating glycoproteins

Traditionally, gel electrophoresis has been employed as a method capable of isolating glycoproteins of interest.[29, 30] However due to glycan heterogeneity, multiple spots are observed on the gel for proteins having identical primary structure, which can be rather problematic.[26, 31] Affinity chromatography is another method that could be used to isolate glycoproteins

from biological extracts like serum, human urine, cell culture lysate etc.[28, 32, 33] For a protein-specific analysis, an antibody can be used. To isolate a broad range of glycosylated proteins, columns having immobilized lectins are employed. Lectins are proteins that have an affinity to bind to different types of carbohydrates.[28] Some examples of lectins include concanavalin A (Con A), wheat germ agglutinin (WGA) and jacaline etc.[33] In order to get more coverage, more than one type of lectin can be used.[27] These lectins can be arranged in different columns in series [34] or they can be combined in a single affinity column.[33]

1.5.2. Reduction and alkylation

Before proteolytic digestion is performed on a glycoprotein, it has to be unfolded so that the protease can act efficiently. One way to denature the protein is by adding denaturating agents such as surfactants to the initial buffer solution.[35] Examples of such surfactants include sodium dodecylsulfate (SDS), urea, guanidine hydrochloride (GHCl) and RapiGest SF (trade name for sodium-3-[2-methyl-2-undecyl-1,3-dioxolan-4-yl)-methoxy]-1-propanesulfonate) [35]. In addition, denaturation can also be achieved by subjecting the glycoprotein to reduction and alkylation. Usually DTT (dithiothreitol) and IAA (iodoacetamide) are used as reducing and alkylating reagents respectively.[28, 35] In this step, disulfide bonds that hold the three

dimensional structure intact are broken, and the sulfhydryl groups are capped by alkylating agents, to prevent unwanted reformation of the bonds.[36]

1.5.3. Enzymes used in glycopeptide analysis

Proteolysis of the glycoprotein is achieved using a variety of enzymes. Trypsin is a widely used enzyme that cleaves the protein at well-defined sites (on the C-terminal side of Lys and Arg). The advantage of using an enzyme with this high level of specificity is that the peptide portion of the glycopeptide can readily be predicted, if the amino acid sequence of the protein is available. However, there are certain limitations when using trypsin. For instance, missed cleavages and the presence of multiple glycosylation sites on the same glycopeptide can complicate the data analysis. [37] In addition, the large non glycosylated peptides that exist in tryptic digests can suppress the signal of glycopeptides, which have much poorer ionization efficiency.[38]

An alternative approach to proteolysis with trypsin is using non-specific enzymes like proteinase K [22, 37] and pronase [39], which can overcome some of the limitations of trypsin. However, predicting the peptide sequences that are produced in these digests is quite challenging. Non-specific enzymes cleave the glycoprotein in an undefined manner to produce shorter glycopeptides and peptides. On the positive side, reduction in the size of the peptide portion, results in obtaining smaller glycopeptides that could facilitate obtaining higher mass accuracy measurements.[38] In addition, reducing the size of the peptide

results in inducing fragmentation on specifically the glycan portion of the glycopeptide.[38] It is also is possible to use step-wise digestion with specific and non-specific enzymes.[38]

1.5.4. Glycopeptide purification methods

Reverse phase purification is one of the common methods of purifying glycopeptides.[40-42] The retention mechanism is governed by the hydrophobic character of the peptide portion, and as a result, glycans with different peptide portions generally elute at different times.[42] This technique can be used on its own[40-42] or can be combined with another purification technique. Combining multiple purification methods helps to avoid signal suppression of glycopeptides from unmodified peptides in the mass spectrometer, and helps to enrich the glycopeptides, when they are present in a complex mixture containing nonglycosylated peptides.[26]

The techniques that are combined with RP-HPLC include a variety of purification strategies such as lectin affinity chromatography,[25, 26, 28] carbohydrate based gels such as cellulose or sepharose [43, 44] and gel filtration, or size exclusion chromatography.[45] Lectin affinity chromatography can be used to isolate glycopeptides with specific glycan structures, and the methodology used is similar to that described in the glycoprotein isolation section. Sepharose separation utilizes the formation of hydrogen-bonding between the carbohydrates on substrates as well as on the glycopeptides.[44]

Hydrophilic interaction chromatography (HILIC) or normal phase chromatography[39, 46] and porous graphitized carbon which is an affinity purification method[38, 47] are some other purification technique that have been implemented for purifying glycopeptides. HILIC makes use of the polar interactions between the hydroxyl groups of glycans and stationary phase.[39, 46] In this method the retention mechanism is governed by the size of the glycan, resulting in separating glycoforms with the same peptide moiety.[39] Porous graphitized carbon is well suited for enriching glycopeptides with smaller peptide portions but sufficient selectivity is not achieved with larger tryptic glycopeptides.[44]

1.5.5. Steps involved in assigning compositions for glycopeptides

Fully characterizing the glycopeptides involves identifying the peptide containing the glycosylation site as well as characterizing the glycan structure. These two objectives are most easily obtainable if the analysis occurs in a step-wise fashion, as shown in Figure 1.4. First, it is important to identify with certainty whether or not the peaks in the mass spectrum correspond to glycopeptides. After verification, the peptide portions of the glycopeptides need to be identified, and this can be done using several methods, as described in more detail below. Finally the glycan portion of the glycopeptide needs to be characterized, and the assignment should be validated with MS/MS data.

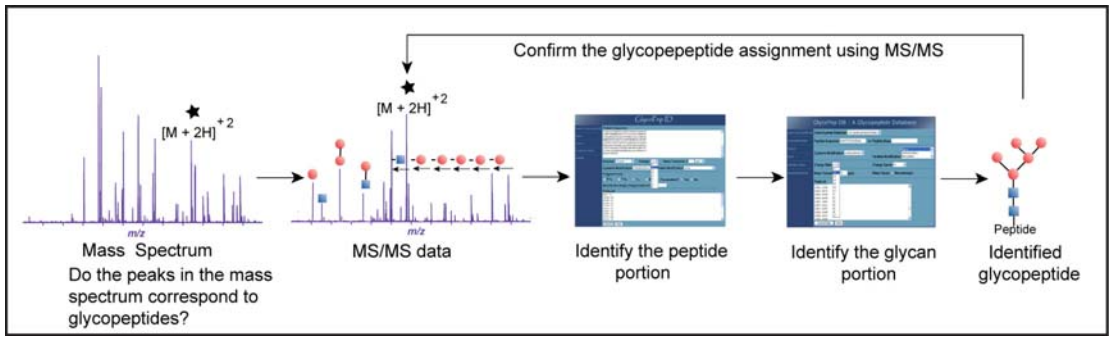


Figure 1.4: Steps involved in assigning compositions for glycopeptides

1.5.5.1. Identification of glycopeptide ions

When glycopeptides are subjected to collision induced dissociation, marker ions corresponding to low molecular weight oxonium ions such as *m/z* 163 (Hex^+), *m/z* 204 ($HexNAc^+$), *m/z* 366 ($Hex\text{-}HexNAc^+$) and *m/z* 292 ($NeuNAc^+$) are produced [23, 40], provided the instrument has the ability to scan down to these low masses. These ions can be used as diagnostic ions to detect the presence of glycopeptides in the mass spectrum. Marker ions can also be generated at the skimmer region of the mass spectrometer. This is done by placing a high potential on the ion optics next to the skimmer causing ion fragmentation, due to acceleration.[41] This process could be coupled with single ion monitoring (SIM) for the detection of marker ions with high sensitivity.[40, 48, 49] Precursor and product ion scanning experiments can be performed to identify the glycopeptides using the marker ions mentioned above.[23, 40, 50]

In general, glycopeptides fragment to produce glycosidic cleavage ions of the Y and B type[49] (nomenclature by Demon and Costello[51]). As a result,

consecutive losses of 162 Da (Hexose), 203 Da (HexNAc), and 291 Da (NeuNAc) are frequently observed in CID spectra. The presence of these ions also can be used as a tool for confirming the presence of a glycopeptide.[43] In addition, neutral loss scan experiments can be performed making use of this occurrence to detect glycopeptides.[39] Moreover, when MS/MS experiments are performed on glycopeptide ions generated by a MALDI-Qq-TOF, usually there are four peaks that appear on the TOF spectrum that can be used as diagnostic ions. These peaks include $[Peptide + H]^+$, $[peptide +H -17]^+$, $[Peptide + GlcNAc +H]^+$, and $[Peptide + C_2H_2NHAc]^+$.[52, 53] Similar results had also been observed when the experiments were performed on a MALDI-TOF/TOF-MS instrument.[54] In this case, in addition to the ions mentioned above, there were two other ions that were observed at [Peptide + GlcNAc Fucose + H]$^+$ and [Peptide + GlcNAc + 2Fucose + H]$^+$ due to the core structure being fucosylated (Figure 1.5). When the peptide is unknown, the best way to identify these ions is to search for differences in peaks that correspond to 120 Da, 266 Da or 412 Da. These differences correspond to $[Peptide + GlcNAc + H]^+$ and $[Peptide + C_2H_2NHAc]^+$, [Peptide + GlcNAc Fucose + H]$^+$ and $[Peptide + C_2H_2NHAc]^+$ and [Peptide + GlcNAc + 2Fucose + H]$^+$ and Peptide + $C_2H_2NHAc]^+$ respectively.

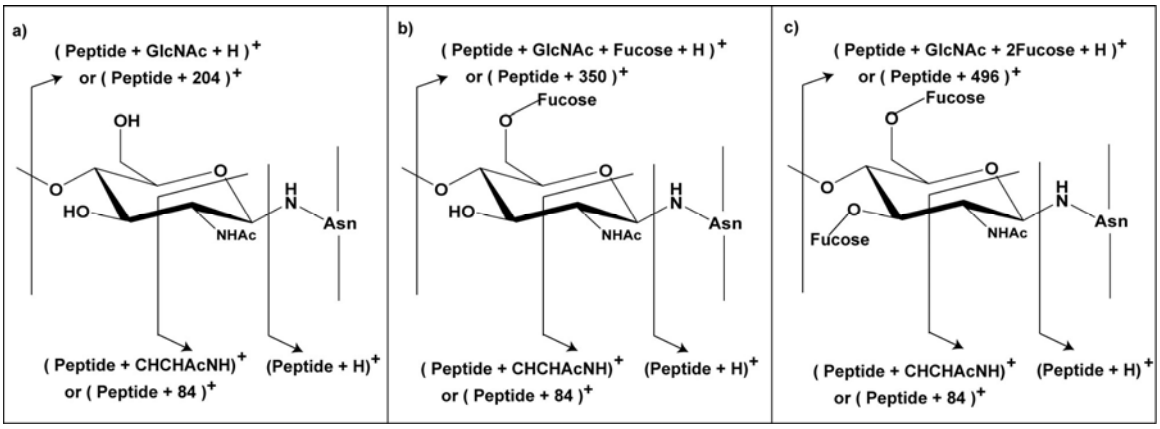


Figure 1.5: The first N-acetylglucosamine residue is shown attached to Asparagine (Asn). Fragmentation patterns correspond to the ions observed in the MS/MS spectra from glycopeptides generated by MALDI having differences in core fucosylation. a) no fucose b) one fucose c) two fucoses. Adapted from reference 53.

1.5.5.2. Identification of the peptide portion

After determining which ions in the mass spectrum are likely to be glycopeptides, the compositional assignment of these ions is most readily achieved if the peptide portion is identified separately from the glycan portion. Characterizing the peptide can be challenging when using a non-specific enzymes for proteolysis, since a robust method of predicting the possible peptide candidates is not available. When a non-specific enzyme is used, one way to deduce the peptide is by identifying the glycan structure from MS/MS experiments first, and subtracting this mass from the precursor ions so that the remaining mass can be used to obtain a peptide mass. This can be done by scanning the peptide masses against the protein sequence obtained from Swiss-Prot using the program FindPep tool (http://ca.expasy.org/tools/findpept.html)[39] The tool can also can be used to

obtain the mass of the peptide by making use of the Peptide + GlcNAc ion that appears on both the ESI and MALDI spectra.[23, 35] The peptide mass can be determined by subtracting the mass of the GlcNAc from the ion corresponding to [Peptide + GlcNAc+ H]$^{+}$.

GlycoPep ID (http://hexosechem.ku.edu/suagr.php)[37] is another promising web-based tool can also be used to identify the peptide portion of glycoforms. This method makes use of the presence of cross ring cleavage ion products (^{0,2}X and ^{0,2}A) in the MS/MS data.[37] These two ions are complementary ions that result in fragmenting the GlcNAc residue connected to the peptide (Figure 1.6). The mass corresponding to the ^{0,2}X (peptide + 83 Da) can be used to infer the mass of the peptide portion. GlycoPep ID identifies the ^{0,2}X or ^{0,2}A ions in the MS/MS data in an automated fashion, so manual interpretation of the data is obviated.

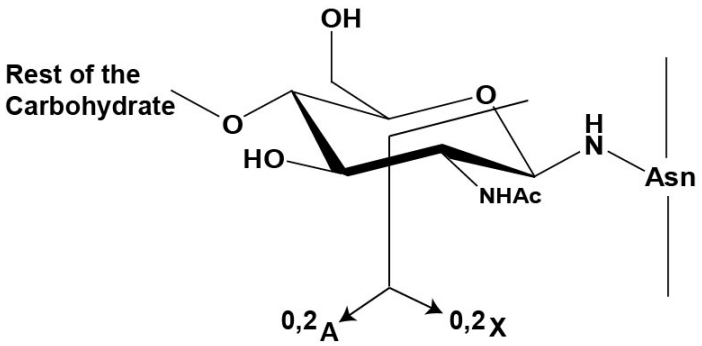


Figure 1.6: ^{0,2}X and ^{0,2}A cross ring fragmentations which occurs duing MS/MS on the first N-acetylglucosamine residue attached to Asparagine (Asn).

In addition to obtaining the peptide mass in MS/MS experiments, several researchers have determined that additional information about the peptide sequence can be obtained, under certain conditions. For instance, when CID MS/MS experiments are performed on glycopeptides using MALDI-Qq-TOF,[52, 53] MALDI-TOF/TOF,[54] and MALDI-QIT-rTOF[55] instruments, fragments from both the peptide and glycan can be detected. In contrast, peptide fragmentation will occur only to a very small extent on an ESI-QIT[27] or a ESI-QqQ mass spectrometer[23] during the first collisional activation event. However, this is only a minor disadvantage for ion traps, because a wealth of information regarding the peptide can be obtained if another stage of tandem mass spectrometry is performed on the ion containing the peptide and the GlcNAc residue.[27, 55, 56]

Usually MS/MS data obtained by Collision Induced Dissociation (CID) of glycopeptides are dominated by glycosidic cleavage of the sugar moeities and y and b type ions from the peptide portion. In addition to MS/MS by CID, ECD (electron-capture dissociation)[57-59] and ETD (electron transfer dissociation)[27, 60] are two other MS/MS techniques that could be performed on glycopeptides, to obtain information about the peptide fragmentation. In both these techniques, the glycan portion remains intact while the peptide fragments to give both z and c type ions [27, 59, 60]. ECD experiments are usually performed on FT-ICR instruments while ETD can be performed in an ion trap mass spectrometers.[27]

Two techniques that do not require MS/MS data for identifying the peptide portion include Edman sequencing and PNGase F treatment. In Edman sequencing,[22, 24] all the amino acids in the glycopeptide sequence except for the glycosylated amino acid are identified as PhNCS derivatives. The peptide sequence can be interpreted from this information.[22] Another method to get the peptide mass is to cleave the glycans from the fractions that are collected during purification. This process is done by using the enzyme, PNGase F. The enzyme cleaves the N-glycosydic bond resulting in converting the asparagine at the N-glycosylated site to aspartic acid. The peptide mass is calculated by taking the difference before and after treating the glycopeptide fractions with PNGase.[42, 61] Alternatively, the de-glycosylated peptide ions can be identified and sequenced using mass spectrometry.

1.5.5.3. Composition assignment of glycopeptides

Once the peptide is identified, the glycopeptide compositions can be characterized using data bases such as Glycopep DB (http://hexose.chem.ku.edu)[22, 62] or GlycoMod (http://us.expasy.org/cgi-bin/glycomod.pl).[38, 52] To use these tools, the mass spectrum is converted to a peak list and searched against these databases for possible matches. The former site has the advantage over the latter site in that the peak list of all multiply charged species can be used to search for matches, without having to convert the peak list to singly charged species. In addition Gycopep DB

contains only biologically relevant glycans, resulting in much more expiditous data interpretation. Although there are several websites that are capable of identifying glycans,[63-65] web-enabled glycopeptide analysis is mainly accomplished using the two sites given above. Since the output of these websites depends on the mass error that is specified by the user, the search is more effective if the data is obtained on a high-resolution mass spectrometer.[22, 24] After the compositions have been obtained, they can be further verified by MS/MS experiments.

1.6. Summary of Glycopeptide Analysis

Glycopeptide analysis is challenging work that requires expertise in sample preparation, mass spectral data acquisition, and computer-assisted data analysis. Many recent research advances have facilitated this work, and glycopeptide analysis is now becoming a more common approach for analyzing glycoproteins. While these experiments are more challenging than studying released glycans, they provide critical information about which carbohydrates are attached to which glycosylation sites, for multiply glycosylated proteins. The studies allow researchers to obtain molecular-scale information about glycoprotein structures, and they open the doorway to understanding the complexities of how glycosylation structures affect protein function.

1.7. Other methods of characterizing glycans on glycoproteins

Structural analysis of glycans is quite a complicated process because it involves determining a number of parameters such as nature and order of monosaccharides, branching points, anomericity of the glycosidic bonds (α or β), stereochemistry of residues (D or L), and nature of substituents (sulfate, phosphate etc).[66] There is no single analytical method that is capable of routinely giving all this information.[66] However this not a major drawback, since often total characterization of carbohydrates is not necessary. Apart from mass spectrometric analysis of glycopeptides,, which was discussed in detail above, nuclear magnetic resonance spectroscopy (NMR) and enzymatic analysis are two major techniques that could be used to characterize glycans on glycoproteins.[66]

1.7.1. Nuclear magnetic resonance spectroscopy (NMR)

Out of these methods, nuclear magnetic resonance spectroscopy (NMR) can be considered as one of the single most powerful techniques that could provide detailed information about the carbohydrates on glycoproteins.[66, 67] In this technique, interactions of radio-frequency electromagnetic radiation with magnetically active nuclei (^{1}H and ^{13}C) are studied in the presence of an external magnetic field.[67] In ^{1}H-NMR only the protons that are attached directly to the carbon atom are taken into consideration, and the samples are usually dissolved in D_2O to replace fast

exchanging protons that are present in OH groups with deuterium atoms, to avoid broad lines in the spectrum.[67, 68] ^{1}H-NMR spectroscopy is suitable for analyzing oligosaccharides, reduced oligosaccharide and glycopeptides.[67] However, a disadvantage of this method is that it requires large amounts of sample that needs to be purified well before the analysis.[66, 67] Oligosaccharides are much easier to analyze using this method than glycopeptides, due to the lack of spectral overlap between the glycans and the peptides and the ease of purification.[67] Data analysis is usually performed using a fingerprint matching technique whereby the 1D ^{1}H-NMR spectrum of a unknown carbohydrate is matched against a database.[66, 67, 69, 70] However, if a spectrum does not match a spectrum in the database more complex experiments such as two dimensional total correlation spectroscopy (TOCSY) and rotating-frame (NOE) spectroscopy (ROESY) should be performed.[66, 67, 71]

1.7.2. Enzymatic analysis

Enzymatic analysis is a biochemical technique that has been widely used to obtain information regarding the nature, stereochemistry and anomericity of monosaccharides in glycoproteins.[66] In this method, glycan sample is divided into several aliquots and each aliquot is incubated with one or more sequence specific exoglycosydases.[1] If the appropriate linkage is present in the glycan structure, a cleavage product is observed. This product

can be detected by various techniques such as chromatography [72], mass spectrometry [73, 74] and gel electrophoresis.[75] This method has usually been performed on glycans that have been released from the peptide backbone using an enzymatic method (PNGase-F)[76] or chemical methods (anhydrous hydrazine).[77, 78] The released glycans are usually labeled prior to performing the enzymatic analysis to facilitate the detection. One method of labeling a glycan is to attach a fluorescent compound such as 2- aminobenzamide.[79] Radio labeling is another method of labeling glycans and it is performed by reducing the glycan with $NaBT_4$ [80] or $NaB[^3H]_4$.[72] However, MALDI-MS has been used as a method of detection by several investigators to detect the products of an enzymatic reaction without prior labeling of glycans after they have been released.[73, 74] One of the disadvantages of enzymatic analysis is that specific enzymes are still not available for some linkages.[1]

1.7.3. High performance anion exchange chromatography (HPAEC)

High performance anion exchange chromatography (HPAEC) with pulsed amperometric detection (PAD) has been a technique that had been used to separate oligosaccharides according to their molecular size, linkage and composition [81, 82]. The presence of large number of hydroxyl groups in glycans makes it amendable to electrochemical detection techniques. This method has the advantage of detecting underivatized ologosaccharides at a relatively low concentration levels.[81, 83] However, since this is essentially a

separation technique, it is important to calibrate the instrument with standards that have been characterized by another reliable method. Therefore this method cannot be used on its own.

While NMR, enzymatic analysis, and HPLC-PAD all have utility in analyzing glycans from glycoproteins, each of these methods suffers from a need for a purified starting compound. Since biologically relevant glycans are always present in complex mixtures, analytical methods that can accommodate their innate heterogeneity are generally more efficient. Mass spectrometry is well-suited for mixture analysis, so it is the analytical method of choices for most of the work presented in this thesis. Accordingly, the fundamentals of mass spectrometry are described in detail below.

1.8. Mass spectrometry

Mass spectrometry (MS) is a powerful analytical technique that is widely used to obtain both quantitative and qualitative information from various molecules. This technique has undergone tremendous developments from the time the initial studies were performed by J.J Thompson in 1912.[18] MS is capable of detecting mass-to-charge ratios (*m/z*) from samples that form gas phase ions.[84] The output from this method is recorded in the form of a spectrum consisting of relative abundances (or intensities) verses the *m/z* values.[84] A typical mass spectrometer consists of three basic components (Figure 1.7), the ion source, analyzer and the detector. In addition, sample inlet system and a data recording system are present as supplementary components. Low pressure conditions are usually maintained in analyzer and detector regions in order to avoid collisions between background gas molecules and ions.[84]

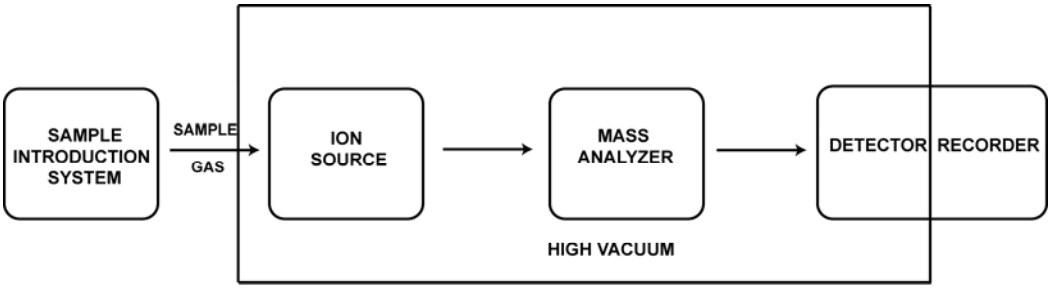


Figure 1.7: Basic components of a mass spectrometer. Adapted from reference 84.

1.9. Ionization sources

There are two methods of introducing the sample into a mass spectrometer. One method is to use direct insertion where by the sample is introduced in a probe or on a plate.[85] Both vaporization and ionization can be facilitated by heating the sample probe or by subjecting the sample to a laser treatment. The other method of sample introduction is by direct infusion. This method is usually performed by either liquid or gas chromatography techniques.[85]

There are many different ionization sources that are commercially available that fall into the two categories mentioned above. They include electron impact (EI),[86] chemical ionization (CI),[87] fast atom bombardment (FAB),[88] matrix-assisted later desorption/ionization (MALDI),[89] and electrospray ionization (ESI).[90] These different sources ionize samples by different methods. For instance in EI, ionization is usually achieved by electron capture or by electron ejection to form a radical anion or a radical cation.[85, 91] Some of the common ionization mechanisms when using other ionization sources include protonation /deprotonation, cationization and direct transfer of a charged molecule to gas phase.[85] The internal energy deposited on the analyte molecule will vary between the different ionization sources.[18] For instance extensive fragmentation can be observed in some ionization techniques, like EI, that are very energetic, and only the molecular ion can be observed with some of the softer ionization techniques, such as FAB, ESI and

MALDI.[18] EI and CI ionization sources are more suitable for analytes that are thermally stable and volatile.[18] In addition, these sources are well suited for analyzing molecules much smaller than the common biological compounds.[85] However the softer ionization techniques are capable of analyzing thermally labile compounds having very high molecular masses.[18, 85] The development of new generation softer ionization techniques has indeed been a major breakthrough in bio-molecular analysis.

1.9.1. Electrospray ionization

Electrospray ionization has undergone remarkable growth in LC/MS applications over the last fifteen years. ESI produces gaseous ions from compounds that are extracted directly from the solution phase.[18, 92] The electrospray is created by the formation of charged droplets in the presence of an electric field.[92-96] This technique had been used in the creation of aerosols and has existed long before it was used in mass spectrometry applications.[97] When there is a liquid flow in the capillary in the absence of an electric field, the drops start to form at the end of the capillary and when the gravitational force exceeds their surface tension they detach from the surface. However if a potential is applied between the capillary and the counter electrode (3-6 kV), the droplets tend to have a horizontal trajectory which prevents it from falling.[93] In addition, this field results in the partial separation of positive and negative ions in the solution inside the capillary.[98] If the

analysis is performed in the positive ion mode, the negative ions tend to be driven inside the capillary whereas the positive ions stay on the surface of the liquid.[98] As a result of the repulsions between the positive ions and the pull of the electric field the liquid at the end of the tip overcomes the surface tension and expands into a cone (Taylor cone). Charged droplets are formed from breaking this cone.[98] The solvent molecules are removed from the charged droplets by going through a sheath of inert gas and/or heated capillary.[18 98] A schematic diagram of the electrospray process is shown in Figure 1.8.

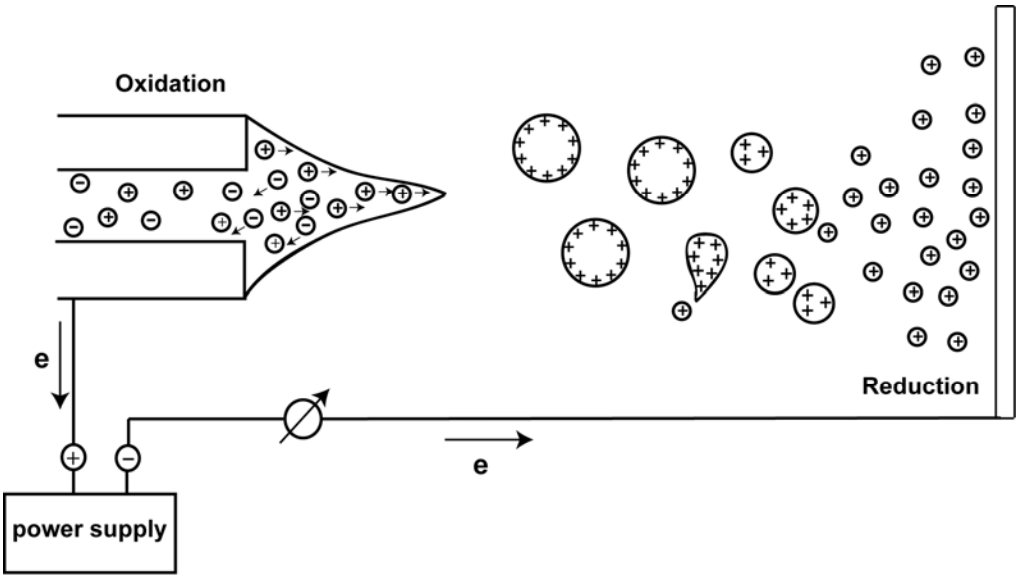


Figure 1.8: A schematic diagram of an electrospray process. Adapted from reference 97.

There are two mechanisms that explain the formation of droplets. However some parts of these mechanisms are still controversial.[98, 99] The two mechanisms include charge residue model (CRM) and ion evaporation model

(IEM) which were originally proposed by Malcolm Dole *et al* [100] and Iribarne *et al* [101] respectively. In both mechanisms the droplet decreases its size by evaporating the solvent, due to the gas which is injected coaxially.[98] Since this decrease takes place at a constant charge (q), a point at which the surface tension exceeds the Coulombic repulsions is reached. This point is known as the Rayleigh limit.[98] Coulombic fissions take place at this stage where smaller droplets are produced. These droplets also undergo the same process to produce even smaller droplets.[18, 98] According to the CRM model, ions are produced by having successive Coulombic fission steps until there is no further solvent evaporation. On the other hand, the IEM theory proposes that Coulombic fission takes place only after reaching a certain radius after solvent evaporation. Ions carrying a single charge (M^+) are then directly emitted from this droplet to the gas phase.[98]

A striking feature of electrospray mass spectra is that it is capable of producing multiple charged species.[18, 96] This phenomenon is due to the presence of several ionizable groups in a molecule. For instance, usually for approximately every 1000 Da of a given protein, there would be an additional charge acquired by the molecule. In addition, electrospray ionization could also be used in the absence of ionizable sites, through the formation of various adducts of sodium, ammonium and acetate etc.[18]

1.10. Analyzers

The ions that are produced from the source enter the analyzer. The analyzer represents the heart of a mass spectrometer and is responsible for separating the ions according to their mass to charge ratios. There can be two types of analyzers depending on where the separation takes place in the mass spectrometer.[85] For instance, ions can be separated in time or in space. In the former the separation occurs at different physical locations in the mass spectrometer. Triple quadrupole and (QqQ) and quadrupole time of flight are two analyzers that separate ions according to time. Trapping instruments such as Fourier transform ion cyclotron resonance (FTICR) and two or three dimensional quadrupole ion traps are two examples of analyzers that can separate ions in space.[85] In these instruments the separation occurs within the same trapping volume but at different times.

1.10.1. Triple quadrupole mass spectrometer

A triple quadrupole mass spectrometer consists of three sets of quadrupoles, and it is represented by QqQ. The first and the last set of quadrupoles (Q) are typically two single quadrupole mass spectrometers having both the RF and direct current (DC) potentials while the central quadrupole (q) consists of only a RF potential, and it can contain collision gas.[18]

A quadrupole mass analyzer consists of four parallel, symmetrical rods that are mounted in a square arrangement. The field within the analyzer is created by applying a RF and direct current (DC) potential to the opposite pairs of rods.[18, 102, 103] The potentials applied on the rods can be given by the following equations.[18]

$$\Phi_0 = +(u + v\cos\omega) \quad \text{and} \quad -\Phi_0 = -(u + v\cos\omega)$$

Where Φ_0 is the potential applied, ω represents the angular frequency = $2\pi v$ where v is the frequency of RF voltage, V is the amplitude of the RF potential and U is the direct potential. A schematic diagram of a quadrupole is shown in Figure 1.9.

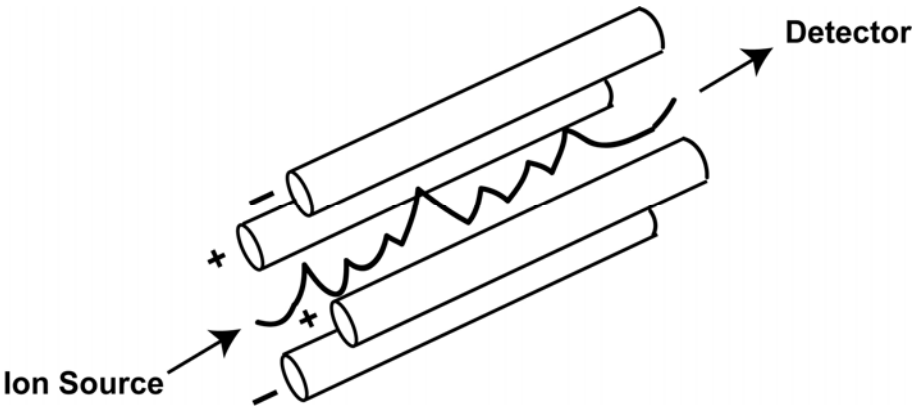


Figure 1.9: A schematic diagram of a quadrupole. Adapted from reference 104.

When ions created by the source enter the quadrupole, they experience trajectories under the influence of the field that is created. Only the ions having stable trajectories are successful at passing through the quadrupole towards the detector. The ions having unstable trajectories are

lost by colliding with the electrode.[102, 103] The stability for a given *m/z* will depend upon the RF potential, magnitude of the frequency associated and the ratio of RF and DC (U) potential amplitudes.[102] The resolution becomes zero when there is no direct current applied. However, the resolution can be increased by increasing the ratio of V/U.[102] This is because at a given time, *m/z's* corresponding to a single value can be made to transmit to the detector by making only those ions to have a stable trajectory.[102] A mass spectrum can be generated using two methods. One method is to scan the values of V and U at a constant U/V ratio keeping the frequency constant. The other method is to scan the frequency while holding the V and the U constant.[102]

1.10.1.1. MS/MS in the triple quadrupole mass spectrometer

Tandem mass spectrometry or MS/MS experiments can be performed in the triple quadrupole instrument due to the presence of the second quadrupole containing the neutral gas.[18] There are various types of scanning events that could be performed to acquire different types of MS/MS experiments on a triple quadrupole mass spectrometer.[18] They are product ion scan, parent ion or precursor ion scan, and neutral loss scan[18] The product ion scan is performed by fixing the first quadrupole to transmit only the ion of interest. This ion then enters the second quadrupole where it is subjected to collisions with inert gas molecules to smaller fragments. The product ions are scanned using the third quadrupole.[18] The precursor ion

scan is performed by fixing the third quadrupole to transmit only a selected ion, while at the same time scanning the first quadrupole. All precursor ions that give the selected fragment as a product are detected. In the neutral loss scan, both first and the third mass analyzers are scanned for a constant loss. The only precursor ions that give the chosen loss are detected.[18]

1.10.2. The 3D quadrupole Ion trap mass spectrometer (QIT-MS)

A 3D quadrupole ion trap can be considered as a three dimensional analogue of the linear quadrupole.[102, 104] As mentioned before in a quadrupole mass analyzer, all ions with a stable trajectory pass through the filter and strike the detector. On the other hand, in a quadrupole ion trap, ions that possess a stable trajectory are trapped in a volume composed of electrodes.[105] The 3D ion trap consists of three electrodes, consisting of two end-cap electrodes and one ring electrode. The ring electrode replaces two rods in a quadrupole, while the end cap electrode in an ion trap mass spectrometer replaces the other two rods.[105] A cross section of an ion trap is shown in Figure 1.10.

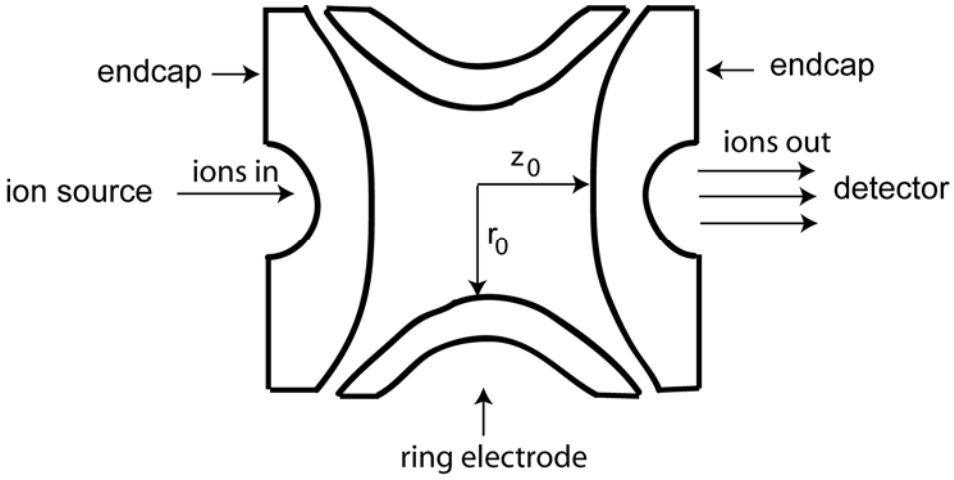


Figure 1.10: Cross-sectional view of the quadrupole ion trap. Adapted from reference 104.

The ions enter the trap through the first end-cap electrode and are detected from an electron multiplier located outside the trap, near the second end-cap electrode. A quadrupole field is generated by applying a AC voltage (RF voltage) to either the end-caps or ring electrode or both.[105] There could also be an optional DC voltage applied to the ring electrode or the end-caps.

The stability of an ion in the axial (z) and radial (r) direction can be explained in terms of Mathieu's parameters, a_z and q_z.[104, 105]

$$a_z = -2a_r = \frac{-16\,eU}{m(r_0^2 + 2z_0^2)\Omega^2}$$

$$q_z = -2q_r = \frac{8\,eV}{m(r_0^2 + 2z_0^2)\Omega^2}$$

Where U and V are DC and RF potentials, ω is the angular frequency, trap dimensions are given by r_o and z_o in the axial and radial direction and m and e are mass and charge of the ion. The stability of ions, as defined by a_z and q_z, is controlled in commercial instruments by controlling V, the RF potential on the ring electrode. Since ions are generally stable with a q_z value between 0 and 0.908 in a quadrupole ion trap, ramping the voltage on the ring electrode increases q_z for each ion and changes the stability of the ions. Controlling ion stability is important for a number of reasons, including ion storage and ejection. Once the ions are unstable – and thus, ejected – they can be detected by the electron multiplier that is present after the endcap electrode, as shown in the previous figure. This is described in more detail in section 1.10.2.2.

1.10.2.1. Trapping ions

There are two methods of trapping ions in an ion trap mass spectrometer. One method is to raise the trapping field strength and the other method is to lower the ions' kinetic energies once the ions have entered the trap.[105] The field strength can be increased by changing the RF either as a pulse[106, 107] or by ramping[108] it up. This method is well suited with pulsed ionization techniques like MALDI.[105] However, the most common method to trap ions is by lowering the kinetic energy of ions so that the ions are efficiently trapped in the trapping field.[109] The kinetic energy can be lowered

by collisions with inert gas molecules in the trapping volume. For this method, the RF level has to be properly maintained such a way that it can both admit ions having high kinetic energy and also trap ions after they have lost energy.[105]

1.10.2.2. Ion detection

Mass selective instability mode was the first method of detection that was utilized in commercial ion trap mass spectrometers.[105] Using this method, mass analysis is performed by increasing the RF amplitude of the ion, resulting in increasing the ion's q_z value. Ions are ejected from the trap when the q_z value becomes just greater than 0.908 due to having an unstable trajectory.[105]

Resonance ejection has been another mode of ion detection method and this mode has made QIT-MS a very practical instrument.[105] An ion possesses a unique motion in the trap when it is subjected to external voltages and this motion has a frequency known as the ions's secular frequency. In the resonance ejection method, kinetic energy is provided to the ions in the form of a supplementary voltage that matches the ion's secular frequency. This excites the ion in the axial direction, and the ions exit the trap after gaining sufficient energy. Usually during ejection, an RF potential (on either the ring or endcap electrodes) is ramped to modulate the ions' secular frequencies. Simultaneously, a supplementary voltage is applied at a fixed

frequency on the endcaps. The secular frequencies of ions change with the RF amplitude, and at one point, the supplementary voltage comes into resonance with the ions' secular frequency and this results in increasing the ions' kinetic energy, which causes the ion to get ejected.[105, 110] The resolution and the sensitivity can be increased from this mode of analysis compared to the mass selective instability mode. Hence this method is used as the detection system in modern QIT instruments.[105]

1.10.2.3. MS/MS experiments in the ion trap mass spectrometer

The MS/MS analysis starts by isolating the ion of interest in the trap followed by activation and detection.[105] Isolation of a parent ion involves ejecting out ions that have both higher and the lower *m/z* values than the parent ion. These ions are ejected from the trap by the two ion ejection methods, mass selective instability mode and resonance ejection, respectively.[105] After the ion of interest has been isolated, it is activated. Collision induced dissociation is one of the common methods of ion activation.[111-113] In this process, the internal energy of a precursor ion is increased by collisions with a neutral gas, which results in decomposing it into smaller fragments.[113] Resonance excitation is one of the common methods by which an ion can be excited to induce fragmentation.[114-116] This process increases the probability of collisions with the neutral gas molecules. The theory behind resonance excitation is similar to the resonance ejection

process that was previously described. However in resonance excitation, the supplementary rf voltage amplitude and the time it is applied should be controlled so that enough energy is applied without ejecting the ions [105].

Other investigators have also used a variety of other activation methods.[105] These include boundary activated dissociation (BAD),[117, 118] red shifted off resonance large amplitude excitation (RSORLAE),[119, 120] and non resonance excitation.[121] Activation methods that do not involve collision gas are surface induce dissociation (SID),[122] infrared multiphoton dissociation (IRMPD),[123, 124] electron capture dissociation (ECD),[125] electron detachment dissociation (EDD)[126] and electron transfer dissociation (ETD)[127]. In SID the ions gain kinetic energy by colliding with a surface such as an electrode [122, 128] and energy deposition involves photons in IRMPD [123, 124]. In ECD and EDD ions gained energy by electron-ion interactions [129] while in ETD it is due to ion-ion interactions[127].

The ability to perform MS^n experiments is a unique feature of the ion trap mass spectrometer.[104] This arises due to the ability to perform a sequence of isolating and fragmenting events. For instance a compound can be fragmented and the resulting fragments can further be isolated and fragmented.

1.10.3. Linear ion trap mass spectrometer

One of the limitations of the 3D ion trap is the space charge effect which arises as a result of having too many ions the in the trapping volume.[102, 104, 105] Space charging occurs as a result of Coulombic repulsions between ions and causes the resolution to decrease.[102, 105] In order to obtain good resolution, the ions having the same *m/z* have to be ejected as one packet to the detector. However, due to the space charging effect, the ions spread into a larger cloud and hit the detector over a longer time period giving broader peaks.[105] In order to overcome this phenomenon, a new linear ion trap (LIT) has been built with increased ion storage capacity by different vendors, Thermo Finnigan [130] and Applied Biosystems.[131] All the operating characteristics present in a three dimensional quadrupole ion trap have been retained in the linear ion trap (LIT).[102, 130]

The Thermo Finnigan linear ion trap consists of a quadrupole similar to the quadrupole present in a triple quadrupole.[130] The schematic diagram of a Thermo Finnigan linear ion trap is shown in Figure 1.11.

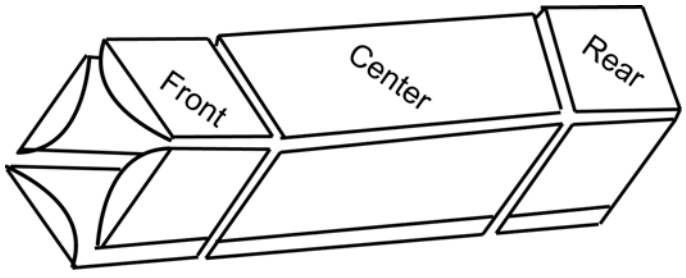


Figure 1.11: Angled section of the two-dimensional LIT. Adapted from reference 102.

There are three axial sections for each rod. The axial confinement of ions is achieved by applying a DC potentials to front and rear lenses where as the radial confinement is achieved by applying a RF potential to the center section.[102, 130]

1.10.4. Fourier Transform-ion Cyclotron Resonance (FT-ICR)

FT-ICR is one of the most powerful mass analyzers, capable of having very high mass resolving power, which is critical in obtaining accurate mass measurements.[105, 132] The resolving power depends on the magnetic field strength, and at present superconducting magnets with very high magnetic fields are being used for FT-ICR instruments.[105, 132, 133] The most common geometry of this analyzer is a six sided cubic box located in a magnetic field.[18, 105] There are three pairs of plates in each cell. They are trapping plates, detection plates and excitation plates (Figure 1.12).

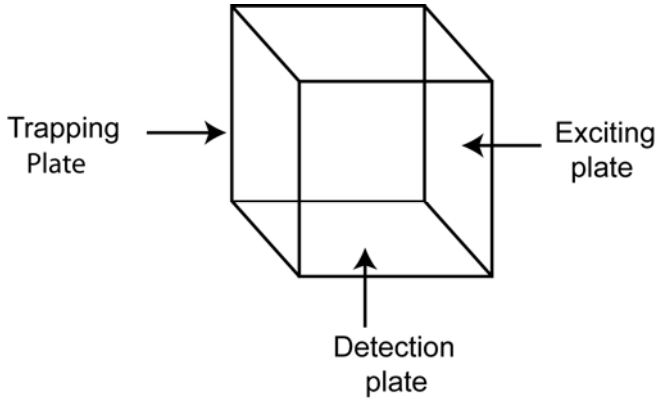


Figure 1.12: Cross-sectional view of the FT-ICR cell. Adapted from reference 105.

The open ended cylindrical cell is another type of geometry that has been used for these instruments.[133] The two ends function as the trapping electrodes and the center has four electrodes to do the same functions as mentioned before.[133] The principle behind this instrument is based on monitoring the circular motion of a charged particle in a magnetic field.[85] The magnetic field traps ions only in two dimensions.[105] The motion of an ion in a magnetic field can be described by the following equations[18, 105, 132, 133] and this is referred to as the cyclotron motion.

$$\text{Centripetal force} : F = evB$$

$$\text{Centrifugal force} : F' = \frac{mv}{r}$$

In order for the ion to stabilize, the two forces have to be equal.

$$evB = \frac{mv}{r}$$

Where m is the mass of the ion, r is the radius of the ion motion, e is the charge on the ion, B is the magnetic field strength and v is the velocity

$$\frac{m}{e} = \frac{Br}{v} = B\omega_C$$

ω_C is the angular velocity of the ion and is related to the cyclotron frequency (f_c) by

$$f_C = 2\pi\omega_C$$

Each mass-to-charge ratio will have a unique frequency because the magnetic field is usually fixed.

In addition to the cyclotron motion, two other motions are described in ICR instruments.[133] The first one is called the trapping motion which is arises as a result of having an electric field in the third dimension to help trap ions. The other motion is called magnetron motion which results from a combination of magnetic and electric fields.[105, 133]

1.10.4.1. Trapping ions

Injecting ions into the analyzer is quite a challenging task because the ions have to penetrate a large magnetic field and also must slow down to get trapped in the cell.[105] In order to facilitate the ions entering the cell, the front trapping plate is kept at a lower potential than the kinetic energy of ions. There are two things that could be done to prevent the ions from going past the rear trapping plate. One method is to have a higher voltage on the rear trapping plate and the other method is to decrease the kinetic energy after the ions have entered the cell.[105] When a higher potential is placed on the rear trapping plate the ions that have kinetic energies lower than the rear trapping plate will turn around and go towards the front trapping plate and at this point this is kept at higher voltage than the kinetic energy of ions. More ions will stop from entering the cell when this happens. The time it takes for an ion to go from the front plate to the rear plate and back to the front plate will depend

on the mass of the ion. For instance the larger ions will have smaller velocities. The other method of trapping ions is to decrease the kinetic energy of the ions after entering the cell. This is done by colliding with a background gas.[105]

1.10.4.2. Ion detection

Ion detection in a Fourier Transform mass analyzer is performed inside the cell using the detection plates, and this is a two step process which involves ion excitation and detection.[105] The ions are all simultaneously excited by applying a rapid scan of a larger frequency range for a short duration of time (1 μs).[18, 105, 133] Ions absorb energy by resonance absorption when a frequency from this range matches with its characteristic frequency.[18] This causes the ions to gain kinetic energy.[18]

It can be shown that all ions will have the same radius when excited by the same potential V_0 applied over the same time (T_{exc}) by the following equation.[18]

$$r = \frac{V_0 T_{exc}}{B_0}$$

As can be seen from this equation the radius in independent of the mass-to-charge ratio.[18] However, each ion will have a characteristic orbiting frequency and all ions that have the same mass-to charge ratio will move together as a packet.[105] This movement of ions induces an image current on the detection

plates, which can be measured. The Fourier equations can be used to convert the time domain signal to time domain spectra.[105]

1.10.4.3. MS/MS experiments in the FT-ICR mass spectrometer

After an ion has been isolated in the ICR cell there are number of different methods that could be used to increase the kinetic energy of ions. Most of these techniques that are used are similar to the methods that were used with QIT-MS. Here too, collisional activation remains the most common method of activating ions.[105] However, in order to get high resolution, an ultra low pressure environment has to be maintained, but it becomes problematic with the use of collision gas. However, a few approaches have been developed to help to overcome this problem.[105, 134] Briefly one method involves pulsing the collision gas only during the collisional activation and the delaying the detection process so that the pressure can return to low values after pumping out the gas.[134]

Another method of ion activation involves dipolar resonance excitation, which is similar to resonant excitation in QIT-MS.[105] A limitation of this method is that the cyclotron radii of most ions that are produced are quite large which results in complicating the detection process. To overcome this several alternative resonance excitation processes have been developed.[105] They are sustained off-resonance irradiation (SORI),[135] multiple excitation collisional activation (MECA) [136], very low energy excitation (VLE).[137]

In addition surface induce dissociation (SID),[128] UV photodissociation, [138] infrared multiphoton dissociation (IRMPD),[139] black body infrared radiative dissociation (BIRD),[140] electron capture dissociation (ECD)[129] and electron detachment dissociation (EDD)[141] are some of the other activation techniques that are used with no collision gas involved.

In summary, a wide range of mass analyzers are available for mass spectrometric detection of glycans. The construction of the mass analyzer dictates its unique uses and applications. Triple quadrupole instruments, for example, are uniquely suited to do precursor ion scanning and neutral loss scanning experiments, because of their geometry, containing three sets of quadrupole rods. They are one of the most commonly implemented instruments in the pharmaceutical industry because of these capabilities. Quadrupole ion trap instruments are also routinely used because they can be manufactured affordably and they have the ability to do MS^n experiments. Finally, FT-ICR-MS analysis allows for high-resolution analysis, which is particularly important in analyzing glycopeptides.

1.11. Overview and summary of the following chapters

In the work presented herein, mass spectrometric experiments are designed to extract information about glycans' structures on glycoproteins. This research is significant for two reasons. First, glycoproteins are an important class of macromolecules that are released in very minute quantities. Characterizing glycans in glycoproteins usually involves analyzing complex mixtures of either glycans or glycopeptides. Therefore developing methods on an instrument such as a mass spectrometer is important because it is very sensitive and can handle complex mixtures. Second, developing methods on a mass spectrometer can be useful because this technique is a high throughput analytical method that can provide data faster than other currently available analytical techniques for analyzing glycoproteins, reducing the analysis time.

There were three goals that were met in this research. The first goal is to improve the current methods for derivatizing and detecting released glycans, and this goal is met in the second chapter, which describes a new derivatization technique for glycans. Glycans are usually derivatized by adding a UV or a fluorescent label after they have been released from the peptide backbone. The advantages in the improved derivatization conditions described in Chapter 2 are as follows. 1) This reaction avoided the use of hazardous chemicals and therefore can be used for routine studies. 2) The derivative helped in enhancing the ionization efficiency in the mass

spectrometer. 3) The derivatized product can readily be detected by UV /fluorescent detectors. 4) More structural information could be obtained.

The second goal of this research was to identify glycopeptides from glycoproteins, and a detailed description is given in chapter 3. Derivatizing glycans after they have been released from the peptide backbone is an excellent method of getting structural information about glycans on glycoproteins. However, if one needs to obtain an understanding about how these glycan structures affect the function of these glycoproteins on a molecular level, it is advantageous to analyze the glycans in a glycosylation site specific manner. This is because when glycans are cleaved from the protein, the attachment site information is lost which is important in determining structure function relationships. However, this information can be retained if glycopeptide analysis is performed instead of glycan analysis. A glycopeptide consists of a peptide and a glycan portion on the same structure. The peptide not only provides information about the glycan attachment site on the protein but also acts as a derivatizing agent and helps in the ionization process. The third chapter shows that site specific glycopeptide analysis is feasible by characterizing the glycans of follicle stimulating hormone (FSH) on two mammalian species (Human and equine).

The third goal for my research is to extend a method of obtaining structural information for glycans and peptides to a routinely used instrument, the triple quad mass spectrometer. The details for this analysis are given in

chapter 4. This method was performed by adapting a statistical test (STEP – statistical test of equivalent pathways) that was previously developed on an ion trap mass spectrometer.[142] The STEP test is based on differentiating first generation product ions from the second generation product ions in a tandem mass spectrum. This information can be useful in getting information about ion genealogy.

Chapter 2

Six different carbohydrates were derivatized using 4-amino-N-[2-diethylamino) ethyl] benzamide (DEAEAB) by reductive amination. $NaBH(OAc)_3$ was used as the reducing agent for this reaction which is an alternative to hazardous $NaBH_3CN$ which is commonly used. When the two reducing agents were compared by a side-by-side analysis both seemed equally efficient. Previously when $NaBH(OAc)_3$ had been used for derivatizing glycans, it had taken 16 hrs to complete the reaction. We were able to cut down the time to 2.5 hrs, which is a clear advantage. The final product was purified using either HPLC or the zip-tip. When the two purification strategies were compared, ~1000 fold lower detection limit was observed for the product purified by HPLC. In addition MS/MS data of derivatized and underivatized glycans were compared. In the derivatized glycan, the reducing end can be easily identified from the nonreducing end, which helps in identifying the

product ions more easily in the MS/MS data than in the data for the underivatized glycan.

Chapter 3

Reproductive functions in the body are regulated by hormones and follicle stimulating hormone (FSH) is an important hormone that falls under this category. This hormone has two subunits that are noncovalently attached, and each subunit contains of two glycosylation sites. Several investigators have revealed that the glycans are responsible for functional differences at different glycosylation sites.[143-145] Therefore one goal of this research was to characterize glycans in a site specific manner. Another goal was to study glycan variability between two mammalian species (human and equine) having similar peptide sequences. Glycoproteins from both equine and human FSH were subjected to reduction and alkylation followed by proteolysis. Proteinase K, which is a non specific enzyme, was used for the proteolytic digestion. The glycopeptides were separated from the peptides using gel filtration. The purified glycopeptides were subjected to Edman sequencing and FT-ICR-MS analysis. Edman sequencing was performed to identify the peptide sequences, and these were used to identify the glycopeptide compositions in the MS analysis. The glycopeptides were assigned using a computer program, and the compositions were verified using MS/MS experiments. Glycans were successfully identified in site

specific manner from both equine and human FSH. There were many site-specific differences that were observed between glycosylation sites of the same species and between different species.

Chapter 4

Two carbohydrates and two peptides were subjected to the STEP analysis in the triple quadrupole mass spectrometer by acquiring two MS/MS spectra for each analyte at high and low energy conditions. STEP ratios were calculated for each product ion in the MS/MS spectrum and these ratios were used to distinguish whether each ion was a first generation (primary) or a second generation (secondary) ion. When the conditions that were used on the ion trap were directly transferred to the triple quadrupole mass spectrometer, it was found that secondary product ions were incorrectly identified. Therefore, new conditions had to be developed in order to accommodate the energetic differences between the two instruments. Since fragmentation is affected by the collision gas pressure, this was one of the important parameters that was taken into consideration when developing new conditions. After modifications, the method was successfully implemented on the triple quadrupole mass spectrometer and out of a total of 108 ions, this method accurately predicted 104 ions (96%) to belong to the category of primary or secondary ions.

1.12. References

(1) Dwek, R. A. *Chem. Rev*.**1996**, *96*, 683-720.
(2) Rudd, P. M.; Dwek, R. A. *Crit. Rev. Biochem. Mol. Biol.***1997**, *32*, 1-100.
(3) Daniels, M. A.; Hogquist, K. A.; Jameson, S. C. *Nat. Immunol*. **2002**, *3*, 903-910.
(4) Dwek, M. V.; Ross, H. A.; Leathem, A. J. C. *Proteomics* **2001**, *1*, 756-762.
(5) Freeze, H. H. *Glycobiology* **2001**, *11*, 129R-143R.
(6) Pantophlet, R.; Burton, D. R. *Annu. Rev. Immunol.* **2006**, *24*, 739-769.
(7) Wang, L.-X. Curr Opin Drug Discov Devel **2006**, *9*, 194-206.
(8) Zhou, T.; Xu, L.; Dey, B.; Hessell, A. J.; Van Ryk, D.; Xiang, S.-H.; Yang, X.; Zhang, M.-Y.; Zwick, M. B.; Arthos, J.; Burton, D. R.; Dimitrov, D. S.; Sodroski, J.; Wyatt, R.; Nabel, G. J.; Kwong, P. D. *Nature* **2007**, *445*, 732-737.
(9) Slovin, S. F.; Keding, S. J.; Ragupathi, G. *Immunol. Cell Biol*. **2005**, *83*, 418-428.
(10) Musselli, C.; Livingston, P. O.; Ragupathi, G. *J. Cancer Res. Clin. Oncol*. **2001**, *127 Suppl 2*, R20-26.
(11) Vlad, A. M.; Finn, O. J. *Breast Dis* **2004**, *20*, 73-79.
(12) Apweiler, R.; Hermjakob, H.; Sharon, N. *Biochim. Biophys. Acta.* **1999**, *1473*, 4-8.
(13) Dell, A.; Morris, H. R. *Science* **2001**, *291*, 2351-2356.
(14) Harvey, D. J. *Expert Rev Proteomics* **2005**, *2*, 87-101.
(15) Satomi, Y.; Shimonishi, Y.; Takao, T. *FEBS Lett.* **2004**, *576*, 51-56.
(16) Nilsson, B. *Mol. Biotechno*l.**1994**, *2*, 243-280.
(17) Hagglund, P.; Bunkenborg, J.; Elortza, F.; Jensen Ole, N.; Roepstorff, P. *J. Proteome Res*. **2004**, *3*, 556-566.
(18) Mass spectrometry: Principles and applications, DeHoffman, E.; Stroobant, V. Second Ed, John Wiley & Sons Ltd: England, **2001**,
(19) Montilla, A.; van de Lagemaat, J.; Olano, A.; del Castillo, M. D. *Chromatographia* **2006**, *63*, 453-458.
(20) Mechref, Y.; Novotny, M. V. *Chem. Rev.* **2002**, *102*, 321-369.
(21) Zaia, J. *Mass Spectrom Rev* **2004**, *23*, 161-227.
(22) Dalpathado, D. S.; Irungu, J.; Go, E. P.; Butnev, V. Y.; Norton, K.; Bousfield, G. R.; Desaire, H. *Biochemistry* **2006**, *45*, 8665-8673.
(23) Conboy, J. J.; Henion, J. D. *J. Am. Soc. Mass Spectrom.***1992**, *3*, 804-814.
(24) Irungu, J.; Dalpathado, D. S.; Go, E. P.; Jiang, H.; Ha, H.-V.; Bousfield, G. R.; Desaire, H. *Anal. Chem*. **2006**, *78*, 1181-1190.
(25) Hirabayashi, J. *Glycoconj. J.* **2004**, *21*, 35-40.
(26) Geng, M.; Zhang, X.; Bina, M.; Regnier, F. *J. Chromatogr. B Biomed. Sci. Appl*. **2001**, *752*, 293-306.

(27) Wuhrer, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H. *J. Chromatogr. B Analyt. Technol. Biomed.* Life Sci. **2007**, *849*, 115-128.
(28) Bunkenborg, J.; Pilch, B. J.; Podteleinikov, A. V.; Wisniewski, J. R. *Proteomics* **2004**, *4*, 454-465.
(29) Kuster, B.; Krogh, T. N.; Mortz, E.; Harvey, D. J. *Proteomics* **2001**, *1*, 350-361.
(30) Gooley, A. A.; Packer, N. H. *Proteome Res* **1997**, 65-91.
(31) Packer, N. H.; Pawlak, A.; Kett, W. C.; Gooley, A. A.; Redmond, J. W.; Williams, K. L. *Electrophoresis* **1997**, *18*, 452-460.
(32) Wang, L.; Li, F.; Sun, W.; Wu, S.; Wang, X.; Zhang, L.; Zheng, D.; Wang, J.; Gao, Y. *Mol. Cell Proteomics* **2006**, *5*, 560-562.
(33) Wang, Y.; Wu, S.-l.; Hancock, W. S. *Biotechnol. Prog.* **2006**, *22*, 873-880.
(34) Yamamoto, K.; Tsuji, T.; Osawa, T. *Mol. Biotechnol.* **1995**, *3*, 25-36.
(35) Imre, T.; Schlosser, G.; Pocsfalvi, G.; Siciliano, R.; Molnar-Szollosi, E.; Kremmer, T.; Malorni, A.; Vekey, K. *J Mass Spectrom* **2005**, *40*, 1472-1483.
(36) Hale, J. E.; Butler, J. P.; Gelfanova, V.; You, J.-S.; Knierman, M. D. *Anal. Biochem.* **2004**, *333*, 174-181.
(37) Irungu, J.; Go, E. P.; Dalpathado, D. S.; Desaire, H. *Anal. Chem*. **2007**, *79*, 3065-3074.
(38) Larsen, M. R.; Hojrup, P.; Roepstorff, P. *Mol. Cell Proteomics*. **2005**, *4*, 107-119.
(39) Wuhrer, M.; Koeleman, C. A. M.; Hokke, C. H.; Deelder, A. M. *Anal. Chem*. **2005**, *77*, 886-894.
(40) Huddleston, M. J.; Bean, M. F.; Carr, S. A. *Anal. Chem*. **1993**, *65*, 877-884.
(41) Peterman, S. M.; Mulholland, J. J. *J. Am. Soc. Mass Spectrom,* **2006**, *17*, 168-179.
(42) Harmon, B. J.; Gu, X.; Wang, D. I. *Anal. Chem*. **1996**, *68*, 1465-1473.
(43) Wada, Y.; Tajiri, M.; Yoshida, S. *Anal. Chem*. **2004**, *76*, 6560-6565.
(44) Tajiri, M.; Yoshida, S.; Wada, Y. *Glycobiology* **2005**, *15*, 1332-1340.
(45) Alvarez-Manilla, G.; Atwood, J., III; Guo, Y.; Warren, N. L.; Orlando, R.; Pierce, M. *J. Proteome Res*. **2006**, *5*, 701-708.
(46) Hemstroem, P.; Irgum, K. *J. Sep. Sci.* **2006**, *29*, 1784-1821.
(47) Davies, M. J.; Smith, K. D.; Carruthers, R. A.; Chai, W.; Lawson, A. M.; Hounsell, E. F. *Journal of Chromatography* **1993**, *646*, 317-326.
(48) Sullivan, B.; Addona, T. A.; Carr, S. A. *Anal. Chem*. **2004**, *76*, 3112-3118.
(49) Ritchie, M. A.; Gill, A. C.; Deery, M. J.; Lilley, K. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 1065-1077.
(50) Jiang, H.; Desaire, H.; Butnev, V. Y.; Bousfield, G. R. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 750-758.
(51) Domon, B.; Costello, C. E. *Glycoconj. J*. **1988**, *5*, 397-409.

(52) Bykova, N. V.; Rampitsch, C.; Krokhin, O.; Standing, K. G.; Ens, W. *Anal. Chem*. **2006**, *78*, 1093-1103.
(53) Krokhin, O.; Ens, W.; Standing, K. G.; Wilkins, J.; Perreault, H. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2020-2030.
(54) Wuhrer, M.; Balog, C. I. A.; Catalina, M. I.; Jones, F. M.; Schramm, G.; Haas, H.; Doenhoff, M. J.; Dunne, D. W.; Deelder, A. M.; Hokke, C. H. *FEBS J.* **2006**, *273*, 2276-2292.
(55) Demelbauer, U. M.; Zehl, M.; Plematl, A.; Allmaier, G.; Rizzi, A. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 1575-1582.
(56) Wuhrer, M.; Balog Crina, I. A.; Koeleman Carolien, A. M.; Deelder Andre, M.; Hokke Cornelis, H. *Biochim. Biophys. Acta.* **2005**, *1723*, 229-239.
(57) Kjeldsen, F.; Haselmann, K. F.; Budnik, B. A.; Sorensen, E. S.; Zubarev, R. A. *Anal. Chem*. **2003**, *75*, 2355-2361.
(58) Hakansson, K.; Cooper, H. J.; Emmett, M. R.; Costello, C. E.; Marshall, A. G.; Nilsson, C. L. *Anal. Chem*. **2001**, *73*, 4530-4536.
(59) Hkansson, K.; Chalmers, M. J.; Quinn, J. P.; McFarland, M. A.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem*. **2003**, *75*, 3256-3262.
(60) Hogan, J. M.; Pitteri, S. J.; Chrisman, P. A.; McLuckey, S. A. *J. Proteome Res*. **2005**, *4*, 628-632.
(61) An, H. J.; Peavy, T. R.; Hedrick, J. L.; Lebrilla, C. B. *Anal. Chem*. **2003**, *75*, 5628-5637.
(62) Go, E. P.; Rebecchi, K. R.; Dalpathado, D. S.; Bandu, M. L.; Zhang, Y.; Desaire, H. *Anal. Chem*. **2007**, *79*, 1708-1713.
(63) Cooper, C. A.; Joshi, H. J.; Harrison, M. J.; Wilkins, M. R.; Packer, N. H. *Nucleic Acids Res*. **2003**, *31*, 511-513.
(64) Lohmann, K. K.; von der Lieth, C.-W. *Proteomics* **2003**, *3*, 2028-2035.
(65) Lapadula, A. J.; Hatcher, P. J.; Hanneman, A. J.; Ashline, D. J.; Zhang, H.; Reinhold, V. N. *Anal. Chem.* **2005**, *77*, 6271-6279.
(66) Dwek, R. A.; Edge, C. J.; Harvey, D. J.; Wormald, M. R. *Annu. Rev. Biochem.***1993**, *62*, 65-100.
(67) Van Halbeek, H. *Meth. Enzymol.***1994**, *230*, 132-168.
(68) Poppe, L.; Van Halbeek, H. *J. Am. Chem. Soc.***1991**, *113*, 363-365.
(69) Vliegenthart, J. F. G.; Dorland, L.; Van Halbeek, H. *Adv Carbohydr Chem Biochem* 1983*, 41,* 209-374.
(70) Brockhausen, I.; Grey, A. A.; Pang, H.; Schachter, H.; Carver, J. P. *Glycoconj. J.* **1988**, *5*, 419-448.
(71) Bush, C. A. *Bull.Magn. Reson* **1988**, *10*, 73-95.
(72) Green, E. D.; Baenziger, J. U. J. *Biol. Chem.***1988**, *263*, 25-35.
(73) Sutton, C. W.; O'Neill, J. A.; Cottrell, J. S. *Anal. Biochem.* **1994**, *218*, 34-46.
(74) Colangelo, J.; Orlando, R. *Anal. Chem.* **1999**, *71*, 1479-1482.
(75) Guttman, A. *Electrophoresis* **1997**, *18*, 1136-1141.

(76) Tarentino, A. L.; Gomez, C. M.; Plummer, T. H., Jr. *Biochemistry* **1985**, *24*, 4665-4671.
(77) Patel, T.; Bruce, J.; Merry, A.; Bigge, C.; Wormald, M.; Parekh, R.; Jaques, A. *Biochemistry* **1993**, *32*, 679-693.
(78) Takasaki, S.; Mizuochi, T.; Kobata, A. *Meth. Enzymol.* **1982**, *83*, 263-268.
(79) Bigge, J. C.; Patel, T. P.; Bruce, J. A.; Goulding, P. N.; Charles, S. M.; Parekh, R. B. *Anal. Biochem.* **1995**, *230*, 229-238.
(80) Yoshima, H.; Mizuochi, T.; Ishii, M.; Kobata, A. *Cancer Res.***1980**, *40*, 4276-4281.
(81) Hardy, M. R.; Townsend, R. R. *Proc. Natl. Acad. Sci. U.S.A.***1988**, *85*, 3289-3293.
(82) Hardy, M. R.; Townsend, R. R. *Carbohydr. Res.* **1989**, *188*, 1-7.
(83) Townsend, R. R.; Hardy, M. R.; Lee, Y. C. *Meth. Enzymol.***1989**, *179*, 65-76.
(84) Practical Aspects Of Ion Trap Mass Spectrometry; March, R. E., Todd, J.F.J. vol 3, CRC Press Inc, Boco Raton, FL, **1995**
(85) The Expanding Role of Mass Spectrometry in Biotechnology; Siuzdak, G. MCC Press, San Diago, CA, **2003**.
(86) Mariner, T.; Bleakney, W. *Phys Rev.***1947**, *72*, 792-806.
(87) Harrison, A. G. *Adv Mass Spectrom* **1989**, *11A*, 582-595.
(88) Barber, M.; Bordoli, R. S.; Sedgwick, R. D.; Tyler, A. N. *Nature* **1981**, *293*, 270-275.
(89) Karas, M.; Bachmann, D.; Bahr, U.; Hillenkamp, F. *Int. J. Mass Spectrom. Ion Process.* **1987**, *78*, 53-68.
(90) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64-71.
(91) Interpretation of Mass Spectra*;* McLafferty, F. W.; Turecek, F. Fourth Ed. University of Science Books*:* Mill Valley*, CA,***1993**.
(92) Kebarle, P.; Tang, L. *Anal. Chem.* **1993**, *65*, 972A-986A.
(93) Yamashita, M.; Fenn, J. B. *J Phys Chem* **1984**, *88*, 4451-4459.
(94) Yamashita, M.; Fenn, J. B. *J Phys Chem* **1984**, *88*, 4671-4675.
(95) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Mass Spectrom Rev* **1990**, *9*, 37-70.
(96) Smith, R. D.; Loo, J. A.; Edmonds, C. G.; Barinaga, C. J.; Udseth, H. R. *Anal. Chem.***1990**, *62*, 882-899.
(97) Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentations & Applications, Cole, R. B. John Wiley and Sons, Inc, New York, NY,**1997**.
(98) Kebarle, P. *J Mass Spectrom* **2000**, *35*, 804-817.
(99) Nguyen, S.; Fenn, J. B. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1111-1117.
(100) Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. *J Chem Phys* **1968**, *49*, 2240-2249.

(101) Iribarne, J. V.; Thomson, B. A. *J Chem Phys* **1976**, *64*, 2287-2294.
(102) Quadrupole Ion Trap Mass Spectrometry; March, R. E., Todd, J.F.J. Second Ed, John Wiley and Sons, Inc, Hoboken, NJ, **2005**.
(103) Introduction to Mass Spectrometry: Biomedical, Environmental and Forensic Applications; Watson, J., T. Raven Press, New York, NY **1976**.
(104) Wong, P. S. H.; Cooks, R. G. *Curr Sep* **1997**, *16*, 85-92.
(105) Payne, A. H.; Glish, G. L. *Meth. Enzymol.* **2005**, *402*, 109-148.
(106) O, C. S.; Schuessler, H. A. *Int. J. Mass Spec. Ion Physics* **1981**, *40*, 67-75.
(107) O, C. S.; Schuessler, H. A. *Int. J. Mass Spec. Ion Physics* **1981**, *40*, 53-66.
(108) Doroshenko, V. M.; Cotter, R. *J. Mass Spectrom* **1997**, *32*, 602-615.
(109) Quarmby, S. T.; Yost, R. A. *Int J Mass Spectrom* **1999**, *190/191*, 81-102.
(110) Louris, J. N.; Cooks, R. G.; Syka, J. E. P.; Kelley, P. E.; Stafford, G. C., Jr.; Todd, J. F. J. *Anal. Chem.***1987**, *59*, 1677-1685.
(111) McLuckey, S. *A. J. Am. Soc. Mass Spectrom.* **1992**, *3*, 599-614.
(112) Shukla, A. K.; Futrell, J. H. *J Mass Spectrom* **2000**, *35*, 1069-1090.
(113) Sleno, L.; Volmer, D. A. *J Mass Spectrom* **2004**, *39*, 1091-1112.
(114) Charles, M. J.; McLuckey, S. A.; Glish, G. L. *J. Am. Soc. Mass Spectrom.***1994**, *5*, 1031-1041.
(115) March, R. E.; McMahon, A. W.; Allinson, E. T.; Londry, F. A.; Alfred, R. L.; Todd, J. F. J.; Vedel, F. *Int. J. Mass Spectrom. Ion Process.* **1990**, *99*, 109-124.
(116) Splendore, M.; Londry, F. A.; March, R. E.; Morrison, R. J.; Perrier, P.; Andre, J. *Int. J. Mass Spectrom. Ion Process.* **1996**, *156*, 11-29.
(117) Vachet, R. W.; Glish, G. L. *Anal. Chem.***1998**, *70*, 340-346.
(118) March, R. E.; Weir, M. R.; Londry, F. A.; Catinella, S.; Traldi, P.; Stone, J. A.; Jacobs, W. B. *Can J Chem* **1994**, *72*, 966-976.
(119) Qin, J.; Chait, B. T. *Anal. Chem* **1996**, *68*, 2108-2112.
(120) Qin, J.; Chait, B. T. *Anal. Chem* **1996**, *68*, 2102-2107.
(121) Wang, M.; Schachterle, S.; Wells, G. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 668-676.
(122) Lammert, S. A.; Cooks, R. G. J. *Am. Soc. Mass Spectrom.* **1991**, *2*, 487-491.
(123) Colorado, A.; Shen, J. X.; Vartanian, V. H.; Brodbelt, J. *Anal. Chem.* **1996**, *68*, 4033-4043.
(124) Boue, S. M.; Stephenson, J. L., Jr.; Yost, R. A. Rapid Commun. Mass Spectrom. **2000**, *14*, 1391-1397.
(125) Baba, T.; Hashimoto, Y.; Hasegawa, H.; Hirabayashi, A.; Waki, I. *Anal. Chem.* **2004**, *76*, 4263-4266.
(126) Kjeldsen, F.; Silivra, O. A.; Ivonin, I. A.; Haselmann, K. F.; Gorshkov, M.; Zubarev, R. A. *Chem. Eur.J.* **2005**, *11*, 1803-1812.

(127) Pitteri, S. J.; Chrisman, P. A.; Hogan, J. M.; McLuckey, S. A. *Anal. Chem.* **2005**, *77*, 1831-1839.
(128) Williams, E. R.; Henry, K. D.; McLafferty, F. W.; Shabanowitz, J.; Hunt, D. F. *J. Am. Soc. Mass Spectrom.* **1990**, *1*, 413-416.
(129) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120*, 3265-3266.
(130) Schwartz, J. C.; Senko, M. W.; Syka, J. E. P. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 659-669.
(131) Hager, J. W. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 512-526.
(132) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Mass Spectrom Rev **1998**, *17*, 1-35.
(133) Amster, I. J. *J Mass Spectrom* **1996**, *31*, 1325-1337.
(134) Carlin, T. J.; Freiser, B. S. *Anal. Chem.***1983**, *55*, 571-574.
(135) Gauthier, J. W.; Trautman, T. R.; Jacobson, D. B. *Anal. Chim. Acta* **1991**, *246*, 211-225.
(136) Lee, S. A.; Jiao, C. Q.; Huang, Y.; Freiser, B. S. *Rapid Commun. Mass Spectrom.***1993**, *7*, 819-821.
(137) Boering, K. A.; Rolfe, J.; Brauman, J. I. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 303-305.
(138) Hunt, D. F.; Shabanowitz, J.; Yates, J. R., III. *J. Chem. Soc. Chem. Commun.***1987**, 548-550.
(139) Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W. *Anal. Chem.* **1994**, *66*, 2809-2815.
(140) Price, W. D.; Schnier, P. D.; Williams, E. R. *Anal. Chem.***1996**, *68*, 859-866.
(141) Yang, J.; Mo, J.; Adamson, J. T.; Haakansson, K. *Anal. Chem.* **2005**, *77*, 1876-1882.
(142) Bandu, M.L.; Wilson, J.; Vachet, R.W.; Dalpathado, D.S.; Desaire, H. *Anal Chem.* **2005**, *77***,** 5886-5893.
(143) Flack, M.R.; Froehlich, J.; Bennet, A, P.; Anasti, J.; Nisula, B.C. *J. Biol. Chem*. **1994**, *269*, 14015-14020
(144) Valove, F. M.; Finch, C.; Anasti, J. N.; Froehlich, J.; Flack, M. R. *Endocrinology* **1994**, *135*, 2657-2661
(145) Saneyoshi, T.; Min, K.; Ma, X.J.; Nambo, Y.; Hiyama, T.; Tanaka, S.; Shiota, K. *Biol. Reprod.* **2001**, *65*, 1686-1690

CHAPTER 2

Reductive Amination of Carbohydrates Using $NaBH(OAc)_3$

2.1. Introduction

Glycoproteins comprise several important classes of macromolecules and perform diverse biological functions. In particular the heterogeneity of the oligosaccharide moiety can affect both the structure and function of a glycoprotein.[1] For example, specific sets of oligosaccharides are associated with different stages of cell differentiation, and alterations to these oligosaccharides can lead to certain disease states, like cancer.[2] Many scientists now believe that study of oligosaccharides may offer opportunities of therapeutic treatment for life-threatening diseases such as cancer, AIDS, and diabetes.[3] Thus the increasing interest in oligosaccharide function has created the need for sensitive detection and analysis.

Mass spectrometry is an important tool for the structural analysis of oligosaccharides, because it offers precise results with high sensitivity.[4] High sensitivity is especially important in carbohydrate analysis since oligosaccharides are liberated in minute amounts from naturally occurring glycoproteins. To enhance sensitivity, oligosaccharides may be derivatized, either by methylation, acetylation or reductive amination[5] prior to mass spectrometric detection. These derivatives enhance the hydrophobicity and

help to increase the signal strength, regardless of the ionization technique used.[5-8] Derivatization not only improves the signal but it may provide more structural information, especially with MS/MS experiments [9]. Unfortunately permethylation and acetylation become increasingly difficult with small-scale sample sizes.[10]

The most common derivatization technique for small-scale samples is reductive amination. This reaction is displayed in Figure 2.1.

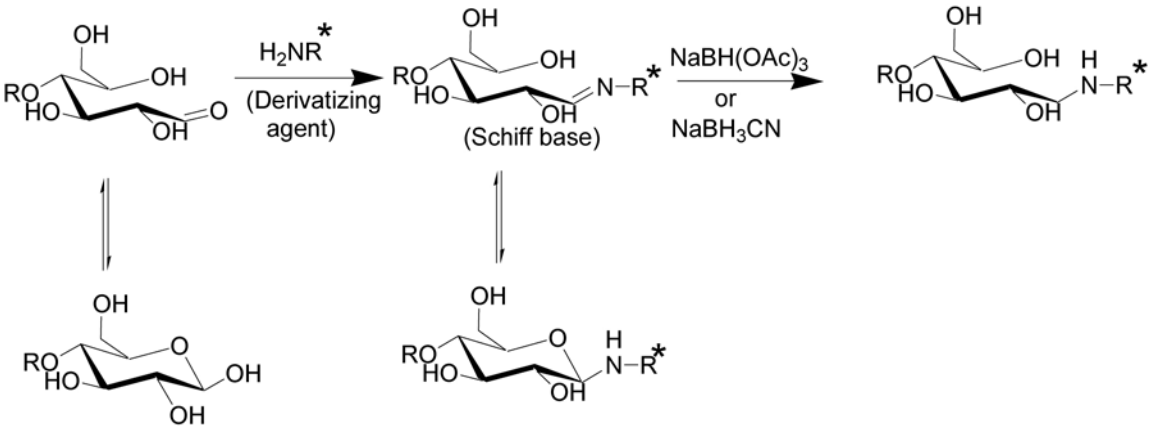


Figure 2.1: Reductive amination of oligosaccharides.

The ring opened (carbonyl) form of the carbohydrate reacts with an amine, that is usually aromatic. The resulting Schiff base is comparatively unstable and is reduced to the corresponding secondary amine when the reducing agent is added. Since these derivatives usually contain a chromophore, this derivatization not only enhances sensitivity for MS analyses but it also allows for chromatographic detectability.[11,12] Several

amines have been used in this context, and a number of these have been investigated as derivatives for improving mass spectrometric detection limits and MS/MS experiments. Some of the common reducing-end derivatizing agents used in carbohydrate analysis are 2-amino pyridine (2-AP),[13,14] 4-aminobenzoic acid 2-(diethylamino)ethyl ester (ABDEAE),[8,12,15] 4-amino-N-[2-(diethylamino)ethyl] benzamide (DEAEAB),[7,10] 2- aminobenzamide (2-AB),[16,17] trimethyl-(4-aminophenyl)aminonium (TMAPA),[9] 9-aminofluorene (9AmFL),[18] 8-aminonaphthalene-1,3,6-trisulfonic acid (ANTS),[19] 1-phenyl-3-methyl-5-pyrazolone,[20] aminopyrazine[11] and 2-aminoacridone (2-AMAC).[21,22]

$NaBH_3CN$ is the most extensively used reducing agent in these procedures. This reducing agent works remarkably well for carbohydrates, and the analysis can be achieved by mass spectrometry[8,10,16] and HPLC.[12,21] Although used infrequently, dimethylamine-borane ($(CH_3)_2NHBH_3$), has been utilized as the reducing agent when separation and detection was completed with HPLC-UV.[11,23] $NaBH(OAc)_3$ had been used for reductive amination of carbohydrates in one instance; but in this case, the derivatization procedure was optimized for detection with gel electrophoresis.[24] No reports comparing the efficacy of these alternative reducing agents to $NaBH_3CN$ (for reductive amination of glycans) are available, and no protocols using alternative reducing agents have been optimized for mass spectral detection.

We sought an alternative to the toxic $NaBH_3CN$ to eliminate the risk of residual cyanide in the product and residual waste. $NaBH(OAc)_3$ is an

attractive alternative reducing agent, since it exhibits remarkable selectivity and is less toxic. In addition to identifying a better reducing agent than $NaBH_3CN$, a reductive amination procedure that did not require the long (16 h) derivatization times[24] was sought. In this report, the results of a comprehensive investigation of the utility and limitations of $NaBH(OAc)_3$ as a reducing agent for the reductive amination of a variety of carbohydrates is reported. We have used DEAEAB as a derivatization agent (the amine) in this work because it has shown promising results in sensitive detection using ESI-MS.[7,10] Comparative results of side-by-side analysis of the extensively used reducing agent $NaBH_3CN$ and $NaBH(OAc)_3$ are reported. Final purification was done either by zip-tip or by HPLC and the limit of detection of each case was found. MS/MS spectra of derivatized and underivatized oligosaccharides were compared as well.

2.2. Experimental Section

N-linked glycans were obtained from Prozyme (San Leandro, CA). Maltopentaose, $NaBH(OAc)_3$, $NaBH_3CN$, DMSO , LNFP I & II and 4-amino-N-[2-(diethylamino)ethyl] benzamide monohydrochloride (DEAEAB) were purchased from Sigma- Aldrich (St. Louis MO). All solvents used were of HPLC grade.

2.2.1. Preparation of derivatives- purification by zip-tip

Derivatives using $NaBH_3CN$ were prepared by the method described by Prime *et al* [25] and purification was followed as described by Harvey.[7] Briefly, the dried glycan (500 pmole-30 nmole) was dissolved in 8 μL of dry DMSO containing 0.3 M DEAEAB in DMSO and 1 μL of acetic acid. To the mixture an excess of $NaBH_3CN$ (0.6 mg) was added. The mixture was heated at 60 ^{0}C for 2 hrs, cooled, and applied to a strip (100 × 30mm) of Whatman 3MM Chromatography paper (Fisher Scientific Co, Pittsburg, PA) and allowed to dry. The paper was placed in a chromatography tank containing acetonitrile and the solvent was allowed to rise to the top of the paper. The spot at the origin (about 10 mm diameter) was cut out, and the carbohydrate derivatives were extracted with 150 μL of water. Purification was achieved using a C-18 Millipore "zip-tip" (Millipore, Billerica, MA). Initial wetting (50%ACN) and equilibration (0.1%HOAc) of the zip-tip was achieved according to the instructions provided by the manufacturer. The sample was aspirated and dispensed ~30 times through the zip-tip in order to assure maximum binding. The wash solution (0.1%HOAc) was aspirated once through the zip-tip. The derivatized carbohydrate was finally eluted with 15 μL of 50/50 MeOH/H_2O, 0.1% HOAc. The elution solution was aspirated and dispensed 30 times to maximize recovery of the glycans.

Derivatives using $NaBH(OAc)_3$ were prepared by adding 1 μL of HOAc, and 3 μL of 0.3 M DEAEAB in DMSO to the dried glycan (500 pmol to 30

nmol). The solution was heated at 60 ^{0}C for one hour and 5 μL of 6.2 M $NaBH(OAc)_3$ in DMSO was added and reheated at 60 ^{0}C for 11/2 hrs. The purification was done in the same manner as described above.

2.2.2. Preparation of derivatives – purification by HPLC

Preparation of derivatives for HPLC was done using the same synthetic procedure described above, using $NaBH(OAc)_3$, including the preliminary purification using the Whatman 3MM chromatography paper and eluting with 150 μL of water. Samples were then lyophilized on a Labconco cetrivap cold trap (Kansas City, MO) and reconstituted to 250 pmole/μL-500 pmole/μL in 10% ACN with 0.1% TFA before applying to the HPLC column (C18, 150mm× 4.6mm). The derivatized oligosaccharides were separated and desalted by RP-HPLC using Shimadzu model LC-10ATvp. The mobile phases used were H_2O containing 0.1% TFA (A) and 95%ACN/H_2O (B). The product was eluted using a linear gradient (5%-20% B in 15 min) at flow rate of 1 mL/min, which was a slight modification to the method described in reference 12. The absorbance was monitored at 289 nm using a diode array detector of model SPD-M10Avp (Shimadzu). HPLC fractions were collected manually. These were concentrated to dryness and reconstituted with 50% MeOH/H_2O, 0.1% HOAc before mass spectrometric analysis.

2.2.3. Mass Spectrometry

Electrospray mass spectra were recorded on a Thermo Finnigan LCQ Advantage mass spectrometer (San Jose, CA). The final eluent obtained after purification using the zip-tip was diluted with 50% MeOH/H_2O containing 0.1%HOAc prior to injection into the mass spectrometer. The fractions collected from HPLC were lyophilized and reconstituted in the same solvent. Each sample was injected into a mobile phase of 50% methanol in water (v/v) containing 0.5% acetic acid, at a flow rate of 10 μL/min. Spray voltage was maintained at approximately 3.8 kV and N_2 was used as nebulizing gas at a flow rate of 10 psi. Ions were desolvated in a heated ion transfer tube, maintained at 230 ^{0}C. MS data were acquired in the positive ion mode, and spectra were processed using Xcaliber software, version 1.3. For MS/MS experiments, activation time and activation q were set at 30 ms and 0.250 respectively. Isolation width of 3 Da was used, and the activation energies between 28-32%, as defined by the Xcalibur software, were used.

2.3. Results and Discussion

The reductive amination reaction, as shown in Figure 2.1, produces amino-alcohol products. Figure 2.2 shows the chemical structure of the reaction product, when glucose is used as the carbohydrate and DEAEAB is the amine.

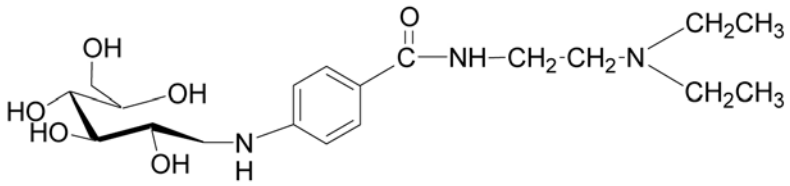
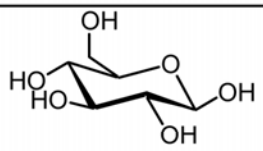
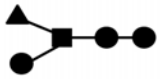
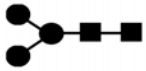
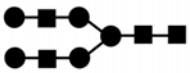


Figure 2.2: Glucose derivatized with DEAEAB. The molecular weight is 399.2 Da, while underivatized glucose weighs 180.1 Da.

The reaction conditions reported here have been optimized to require minimal reaction time and maximal flexibility in the conditions. The best reaction conditions were identified by using different heating times, temperatures and adding different concentrations of DEAEAB and $NaBH(OAc)_3$ to the carbohydrate. After optimal conditions were developed, the reaction was performed on six different carbohydrates, which are depicted in Table 2.1.

Table 2.1: Derivatized Carbohydrates

Name	Structure	Ions expected for derivatized products			
		Cal'd $[M+H]^+$	Cal'd $[M+Na]^+$	Cal'd $[M+H+Na]^+$	Cal'd $[M+2H]^{2+}$
Glucose Monoisotopic mass -180.1		400.3	422.3	211.6	200.6
Maltopentaose Monoisotopic mass- 828.3		1048.5	1070.4	535.7	524.7
LNFP I Monoisotopic mass- 853.3		1073.5	1095.5	548.2	537.2
LNFP II Monoisotopic mass -853.3		1073.5	1095.5	548.2	537.2
$(Man)_3$ $(GlcNAc)_2$ Monoisotopic mass-910.3		1130.5	1152.5	576.7	565.8
$(Gal)_2$ $(GlcNAc)_4$ $(Man)_3$ Monoisotopic mass-1640.6		1860.8	1882.7	941.9	930.9

● - Mannose or galactose ■ - *N*-acetylhexosamine, ▲ - Fucose
The neutral derivatized products weigh 219.2 Da more than the underivatized carbohydrates.
Exact masses for the ions were calculated and the values in the table are rounded to one decimal place.

The table provides the expected *m/z* values for the derivatized products, including values for the singly charged species (protonated, sodiated) and doubly charged molecular ions that could be detected by mass

spectrometry. [As described in Figure 2.1, the derivatized products have a mass that is 219.2 Da greater than the original carbohydrate.]

The ESI-MS spectra for each of the reaction products are shown in Figure 2.3. These spectra were acquired after minimal clean-up, using a "zip-tip" purification strategy. Generally, the derivatized product is the only major ion detected in the mass spectrum. In two cases, Figure 2.3d and 2.3f, an additional ion is readily abundant, namely *m/z* 927 in Figure 2.3d and *m/z* 1495 in Figure 2.3f. In both these cases, the additional ion corresponds to a fragment of the product – loss of fucose in the case of 2.3d, and loss of a disaccharide (Gal-GlcNAc) in the case of 2.3f. These ions are most likely the result of in-source fragmentation, occurring during the ionization of the sample, and not side-products of the reaction. This has been verified previously by Shimonishi and co-workers, who described this phenomenon for similarly derivatized carbohydrates.[8] For the first four derivatized carbohydrates in Figure 2.3, the singly charged ion, $[M+H]^+$ was readily observed. The most prominent ion for *N*-linked glycans was a doubly charged species, $[M+H+Na]^{+2}$ or $[M+2H]^{+2}$. This finding, that the doubly charged ions are more prominent for *N*-linked glycans, is in agreement with previous MS data obtained by Harvey.[7] In those studies, the same amine was used, DEAEAB, but the reducing agent was $NaBH_3CN$.

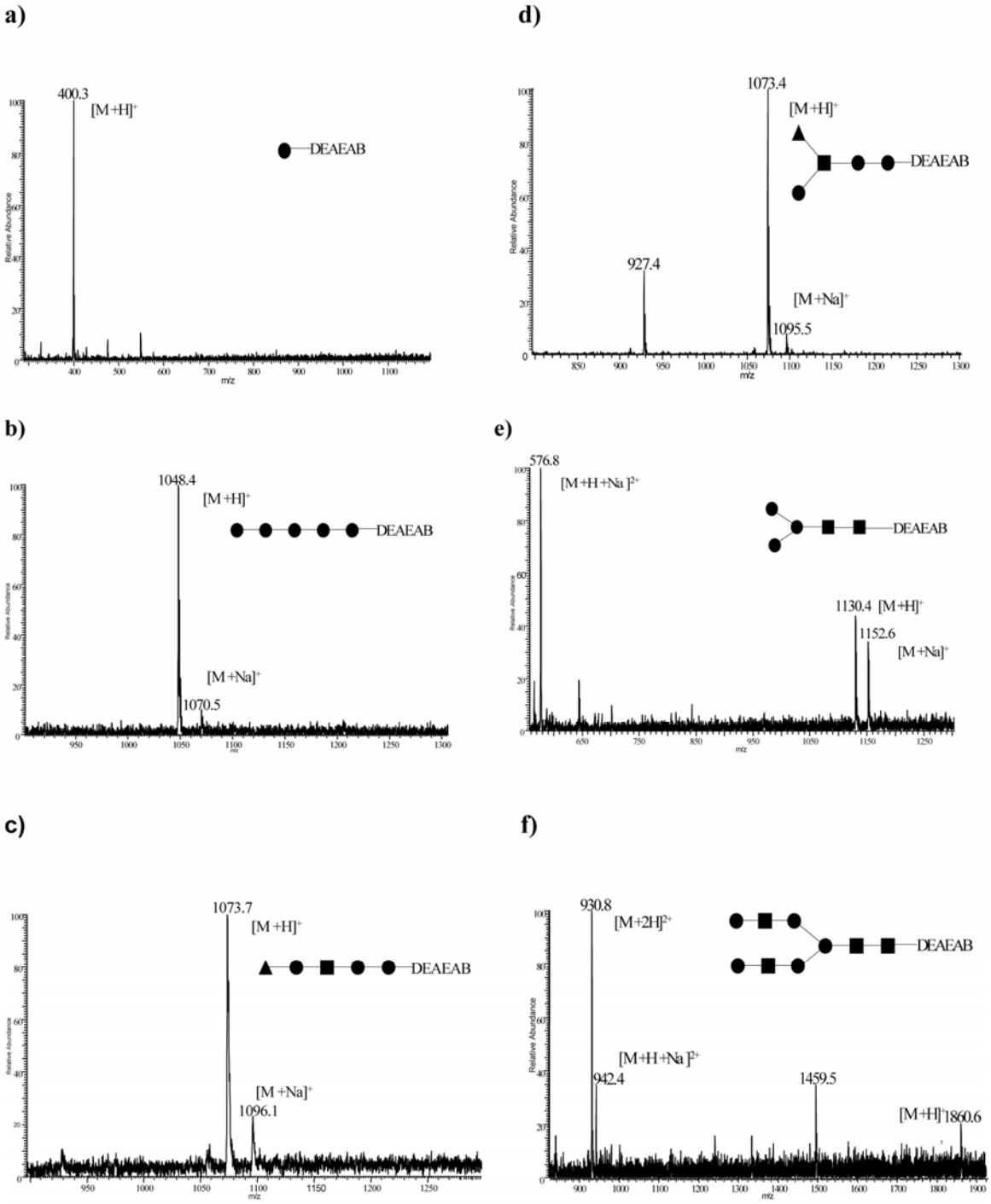


Figure 2.3: ESI-MS spectra of DEAEAB derivatized carbohydrates. a) Glucose b) Maltopentaose c) LNFP I d) LNFP II e) $(GlcNAc)_2(Man)_3$ f) $(Gal)_2(GlcNAc)_4$ Man_3. Clean up was done using zip-tip. The spectrum has been background subtracted.

To determine whether $NaBH(OAc)_3$ would be as effective as $NaBH_3CN$ at the reductive amination reaction, results obtained from this protocol using $NaBH(OAc)_3$ were compared with those obtained with the extensively used reducing agent, $NaBH_3CN$.[7,25] In side-by-side analysis of the two reducing agents, both protocols gave similar results, as shown in Figure 2.4.

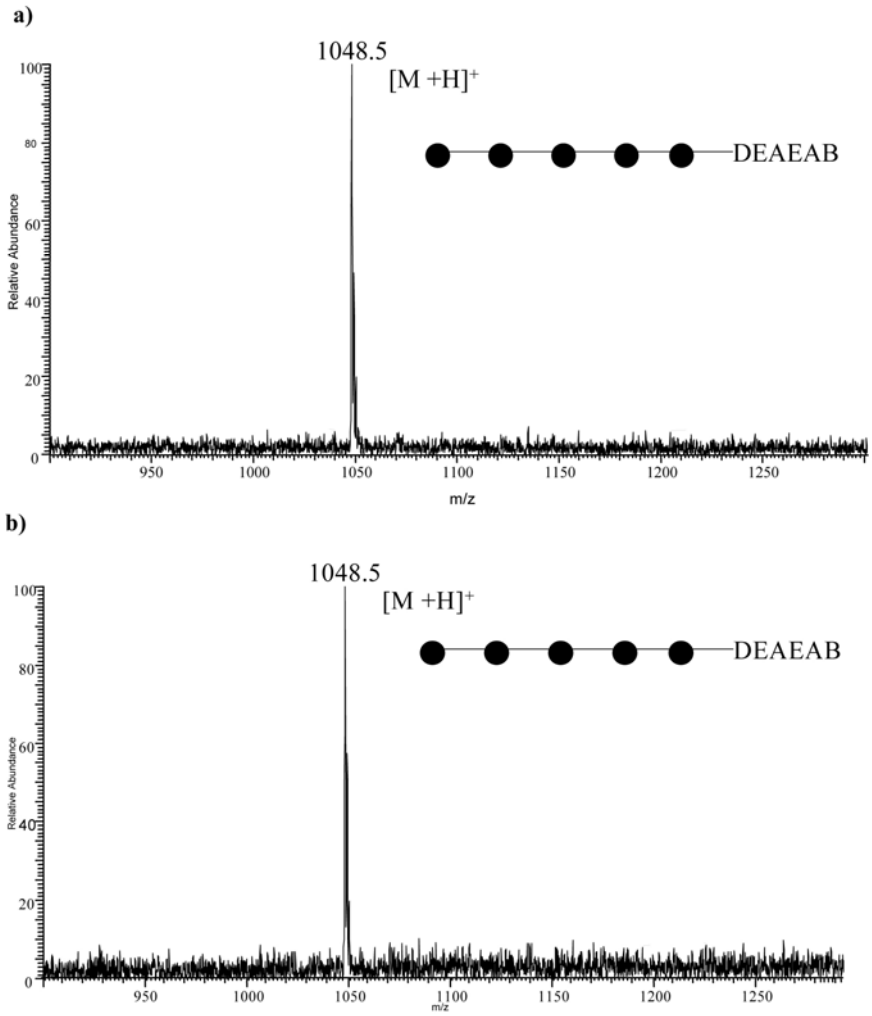


Figure 2.4: ESI-MS spectra of DEAEAB derivatized maltopentaose using a) $NaBH_3CN$. 1.9 nmole consumed from 5 µg of starting material. b) $NaBH(OAc)_3$. 1.9 nmole consumed from 5µg of starting material. The spectra are background subtracted.

The starting materials and the amount injected to the mass spectrometer were identical in the two cases and the purification procedures were the same (zip-tip purification). In both cases the reaction had gone to completion and there were no starting material remaining.

An additional benefit of using $NaBH(OAc)_3$ as the reducing agent is that the products can be analyzed by MS without *any* purification. The derivatization bi-products in the protocol developed herein are not toxic, so MS analysis can also be performed directly, if the initial amount of carbohydrate is in the 30 µg range. A typical spectrum, which was obtained without purification, is shown in Figure 2.5.

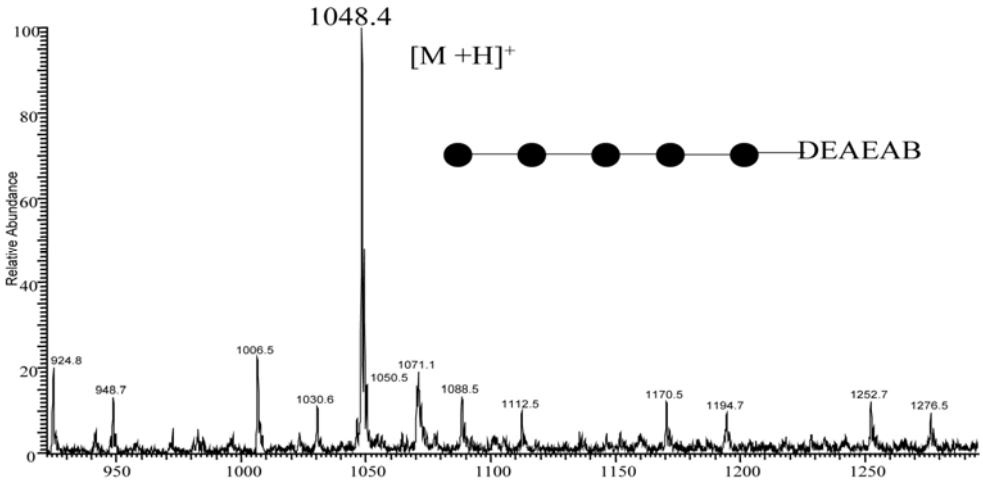


Figure 2.5: ESI-MS spectra of DEAEAB derivatized maltopentaose without any purification. 74 pmole injected. The spectrum has been background subtracted.

When using $NaBH_3CN$ as the reducing agent, one cannot inject the reaction mixture directly into the mass spectrometer, because that could release HCN into the air. Being able to inject samples directly into the mass spectrometer prior to purification is a distinct advantage of using $NaBH(OAc)_3$, because such experiments allow for rapid detection of the derivatized carbohydrates.

2.4. Validation Data

To determine the robustness of the method, the quantity of the carbohydrate was varied, while all other parameters were kept constant. The synthetic conditions described herein are applicable for a range of initial carbohydrate concentrations, from 500 pmol to 30 nmol. Using starting quantities anywhere in this range provides very reproducible derivatization results. This flexibility is quite useful for derivatizing carbohydrates isolated from biological sources, where the exact amount of carbohydrate present is generally unknown.

Sensitivity studies were performed on the derivatized maltopentaose to determine the instrumental detection limits. For these studies two samples of maltopentaose, each containing 30 μg, were derivatized. One sample was purified using the zip-tip and the other was purified using HPLC. After purification, by zip-tip, the eluent was further diluted with 50% $MeOH/H_2O$, 0.1% HOAc. The subsequent dilutions were injected to the mass

spectrometer until a signal to noise ratio of ~3 was obtained, providing the instrumental limit of detection, which was 150 pmole for the sample purified by zip-tip. Note: Others have reported detection limits as low as 100 fmole for similar products,[10] when a Q-TOF MS is used instead of a QIT-MS. The differences in the detection limits are because a different amine was used and the Q-TOF is a more sensitive mass analyzer. Previous research has demonstrated that, for glycopeptide analysis, the Q-TOF is approximately 200 times more sensitive than the quadrupole ion trap mass spectrometer.[26]

When the derivatized maltopentaose is purified by HPLC, a chromatogram like the one in Figure 2.6 is typically obtained.

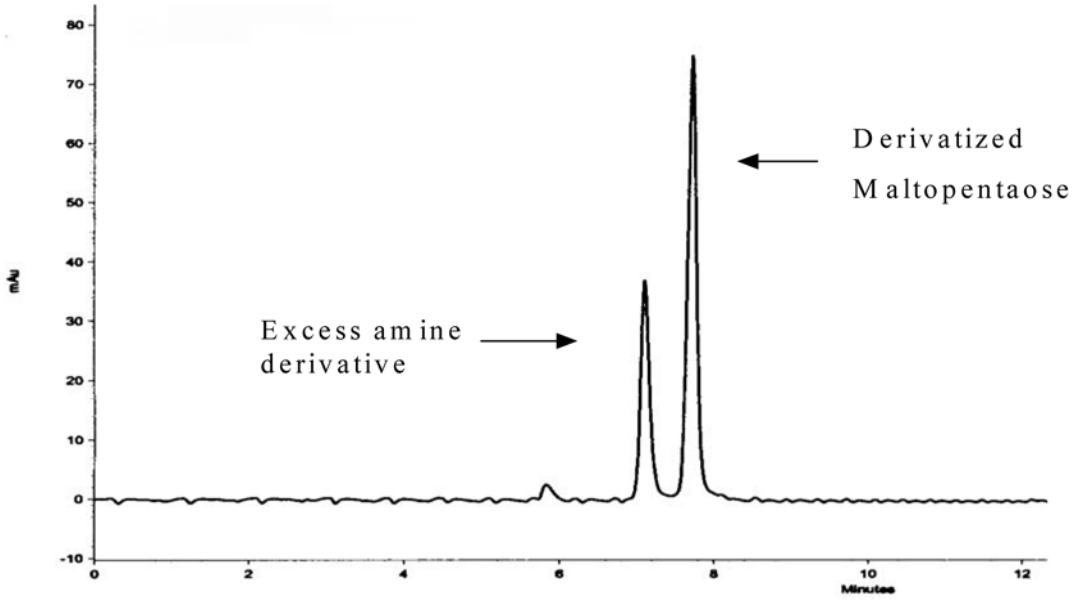


Figure 2.6: Purification of DEAEAB derivatized maltopentaose by RP-HPLC.

To compare the zip-tip purification to HPLC purification, the fractions collected from the HPLC were evaporated to dryness and reconstituted in

50% MeOH/H_2O and H_2O, 0.1% HOAc and the limit of detection was found in the same manner as above. The instrumental detection limit of the product purified by HPLC is 100 fmol. This detection limit is ~1000 times lower than the LOD that was determined when the zip-tip is used for purification. The zip-tip, while useful at removing some of the salts, depletes product recoveries. All the derivatized carbohydrates in Figure 3 were purified by the zip-tip protocol. This is the more rapid, but less effective, purification strategy. The low product recoveries of the zip-tip purification likely contribute to the moderate background observed in Figure 2.3.

2.5. MS/MS Spectral interpretation

In addition to providing enhanced sensitivity for MS or HPLC detection, derivatization of carbohydrates also facilitates structural analysis by MS/MS. This has been demonstrated previously by MS/MS methods[16,20] for the analysis of neutral oligosaccharides and N-glycans, and it is also illustrated here, in Figure 2.7, where LNFP I is the carbohydrate analyzed. The derivatized oligosaccharides produced more easily interpretable product ions, because the reducing end and nonreducing end are easily differentiable, whereas the native compound gives rise to product ions that could have originated from either end of the molecule. The carbohydrate fragmentation nomenclature is described by Domon and Costello.[27] Sequence information could be readily obtained from MS/MS data of derivatized $[M+H]^+$ ion, due to

the presence of only Y-fragment ions in the spectrum. This is an agreement with Franz et al.[18]

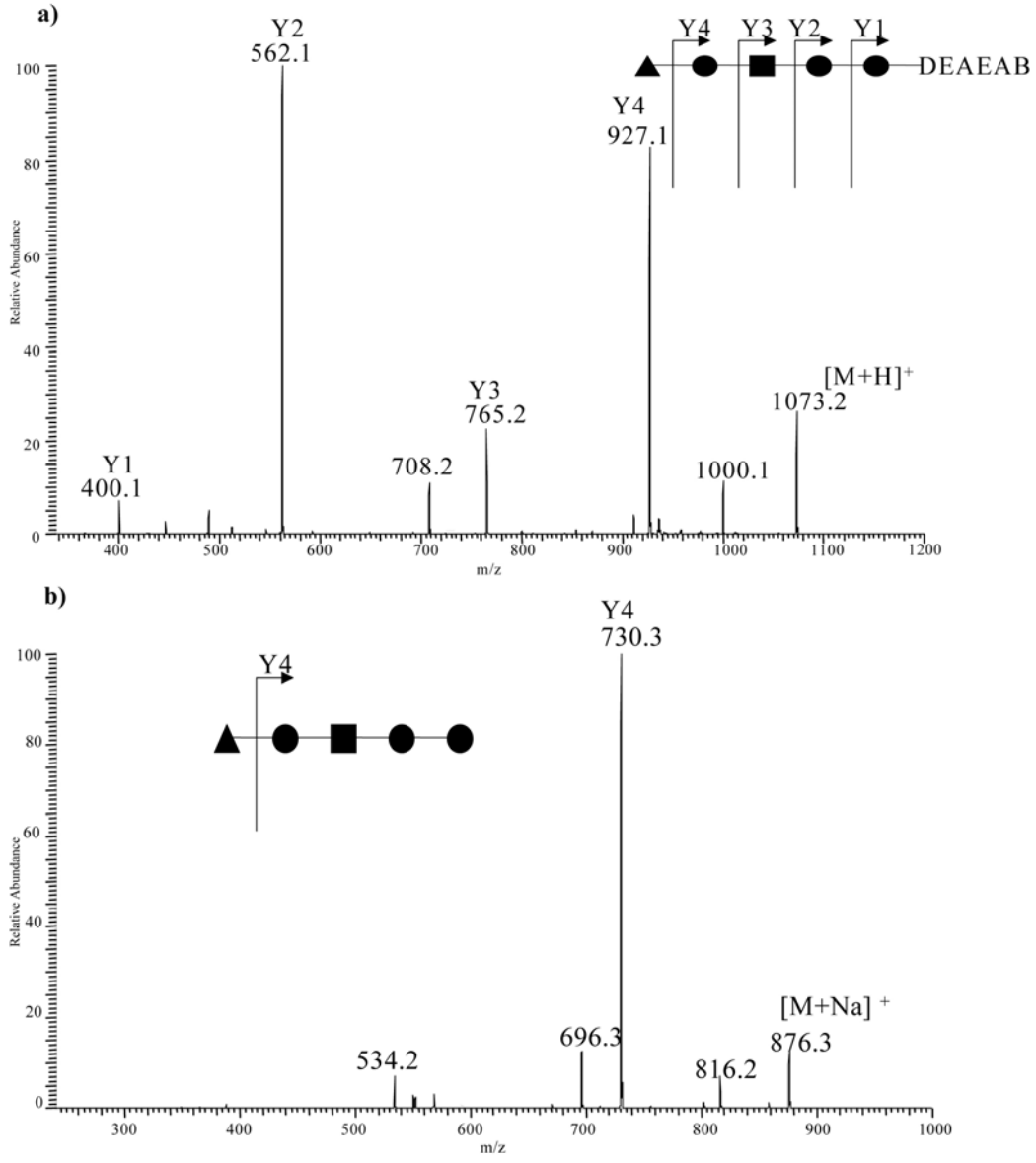


Figure 2.7: Positive ion ESI –MS/MS spectra of a) Derivatized LNFP I- 1.9 nmole sample consumed. b) Underivatized LNFP II –1.9 nmole sample consumed. Clean up was done using zip-tip. The spectra have been background subtracted.

2.6. Conclusion

In this paper, a novel derivatization reaction for oligosaccharides using the non-toxic reagent, $NaBH(OAc)_3$, is developed. While a previous report requires 16 hours for reduction with $NaBH(OAc)_3$,[24] the method reported herein requires a *total* of 2.5 hours for the derivatization – this includes both the amination and reduction reactions. These reaction conditions afford minimal reaction time and significant flexibility. The reaction yield was comparable to the yield obtained with the extensively used reducing agent $NaBH_3CN$. Two methods of purification were employed, the zip-tip and HPLC. There was ~1000 fold increase in the recovery of the sample when the purification was done using HPLC, compared to using the zip-tip. MS/MS spectra of derivatized and underivatized LNFP I were compared, more structural information could be obtained by derivatized product, as described previously.

2.7. References

(1) Dwek R.A. *Chem. Rev.***1996**, *96*,683-720.
(2) Essentials of carbohydrate chemistry; Lindhorst, T.K. Second Ed, Wiley-VCH verlag GmbH & Co. KGaA, Weinheim, **2003**.
(3) Oligosaccharides: Their synthesis and biological roles; Osborn, H.M.I.; Khan, T.H. Oxford University press Inc, New York, **2000**.
(4) Zaia, J. *Mass Spectrom Rev* **2004**, *23*, 161-227.
(5) Mass Spectrometry: principles and Applications; De Hoffman, E.; Stroobant V. Second Ed, John Wiley & Sons Ltd, England, **2001**.
(6) Dell, A.; Carman, N.H.; Tiller, P.R.; Thomas-Oates, J.E. *Biomed. Environ. Mass Spectrom.* **1988**,*16*, 19-24.
(7) Harvey, D.J. *J. Am. Soc. Mass Spectrom.* **2000**,*11*, 900-915.
(8) Takao, T.; Tambara, Y.; Nakamura, A.; Yoshino, K.; Fukuda, H.; Fukuda, M.; Shimonishi, Y. *Rapid Commun. Mass Spectrom.* **1996**,*10*, 677-640.
(9) Okamoto, M.; Takahashi, K.; Doi, T. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 641-643.
(10) Harvey, D.J. *Rapid Commun. Mass Spectrom.* **2000**, *14,* 862-871.
(11) Wu, W.; Hamase, K.; Kiguchi, M.; Yamamoto, K.; Zaitsu, K. *Anal Sc*i, **2000** *16*, 919-922.
(12) Yoshino, K.; Takao, T.; Murata, H.; Shimonishi, Y. *Anal. Chem*. **1995**, 67 4028 - 4031.
(13) Hase. S.; Ibuki, T.; Ikenaka, T. *J. Biochem.* **1984**, *95*, 197-203.
(14) Okamoto, M.; Takahashi, K.; Doi, T.; Takimoto, Y. *Anal.Chem.* **1997**, *69*, 2919-2926.
(15) Mo, W.; Takao, T.; Sakamoto, H. Shimonishi, Y. *Anal. Chem.* **1998**, *70*, 4520 -4526.
(16) Harvey, D.J. Analyst, **2000**, 125, 609 - 617.
(17) Bigge, J.C.; Patel, T. P.; Bruce, J.A.; Goulding, P.N.; Charles, S.M.; Parekh, R.B. *Anal. Biochem.* **1995**, 230, 229-238.
(18) Franz, A.H.; Molinski, T.F.; Lebrilla, C.B. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 1254-1261.
(19) Che, F.Y.; Song, J.F.; Zeng, R.; Wang, K.Y.; Xia, Q.C. J*. Chromatogr A* **1999**, 858: 229-238.
(20) Shen, X.; Perreault, H. *J Mass Spectrom* **1999**, *34*, 502-510.
(21) Camilleri P, Tolson, D, Birrell, H. *Rapid Commun. Mass Spectrom.***1998**, *12*, 144-148.
(22) Okafo, G.; Burrow, L.; Carr, S.A.; Roberts, G.D.; Johnson, W.; Camilleri, P. *Anal.Chem.* **1996**, 68, 4424-4430.
(23) Kiguchi, M.; Hamase, K.; Wu, W.; Yamamoto, K.; Zaitsu, K. *Anal Sci* **1999**, 15, 903-905.

(24) Drummond, K.J.; Yates, E.A.; Turnbull, J.E. *Proteomics* **2001**, *1*, 304-310.
(25) Prime, S.B.; Shipston, N.F.; Merry, T.H. *BioMethods* **1997**, 9, 235-260.
(26) Jiang, H.; Butnev, V.Y.; Bousfield, G.R.; Desaire. H. *J. Am. Soc. Mass Spectrom.* **2004**, 15, 750-758.
(27) Domon, B.; Costello, C.E. *Glycoconj. J.* **1988**, 5, 397-409.

CHAPTER 3

"Comparative glycomics of the glycoprotein follicle-stimulating hormone (FSH): Glycopeptide analysis of isolates from two mammalian species."

3.1. Introduction

Follicle stimulating hormone (FSH) is a heterodimeric glycoprotein consisting of two non-covalently linked α- and β-subunits found in the pituitary of all mammals.[1,2] It is one of the hormones that regulate reproductive functions of the body. Specifically, it is responsible for stimulating follicle maturation in females and for supporting spermatogenesis in males.[3,4] While the amino acid sequence of this glycoprotein is highly homologous between species[5] the glycan structures not only differ significantly between species,[5] but also between the glycosylation sites for any given species.[6] It is known that the glycans greatly influence the biological properties of this hormone.[6] For example, signal transduction,[7] receptor binding,[8] and half life in the circulation[9] are all controlled by glycans, and the final biological response of the target cell (granulosa and Sertoli cells)[10] in vivo is determined by these factors.[11] Previous studies focused on a single critical glycosylation site in the common α subunit have been unable to account for differences in biological activity[1,2] due to compensatory effects of β subunit glycans.[3] Hence, in order to relate their biological functions to their structures, it is essential to

determine the variability of the glycans in a glycosylation site-specific manner. The knowledge gained from this approach can be applied in the development and production of analogues that can be used to regulate gonadal functions.

FSH initiates its action by binding to G protein-coupled receptors, which are on the surface of the target cells.[3,12] This binding causes adenylate cyclase activation, resulting in hormonal responses. Previous studies have reported that this process is greatly affected by glycans at particular sites. For instance when site- directed mutagenesis was performed on Asn^{52} of hFSH to selectively eliminate the glycan at this position, it resulted in either increased[13,14] or unaltered[15] receptor binding activity with a reduction of receptor activation and signal transduction. Similar results have also been observed for eFSH when deglycosylation was performed at αAsn^{56} site using either PNGase digestion[11] or site-directed mutagenesis.[16] It is still unclear about which glycans are responsible for this function even though there is a clear indication that Asn^{52}/Asn^{56} glycans play an essential role.

Studies done on recombinant human variants of hFSH reveal that glycans from the β subunit play a more important role than from those of the α subunit in metabolic clearance/ half life circulation.[9] It is evident that the glycans mediate different functions at different glycosylation sites. Therefore in order to accurately study glycan structure and relate it back to its functional significance, one needs to characterize glycans in a glycosylation site-specific manner.

The glycans from the pituitary and recombinant FSH have been previously characterized by various methods. Conventional approaches employed for this type of analysis involve cleaving the glycans enzymatically from the peptide backbone[2,17,18-21] followed by a combination of techniques such as gel filtration, high performance liquid chromatography (HPLC) and lectin affinity chromatography with endo- and exoglycosidase digestion.[2,17] While these techniques are clearly useful, purifying carbohydrates by chromatographic methods is quite a daunting task. ^{1}H-nuclear magnetic resonance (NMR) has also been used in this type of analysis, however this analysis also requires purified samples (and large sample quantities).[19] Despite these limitations, several abundant glycoprotein hormones have been characterized in a site specific manner. Alternatively, mass spectrometry has the advantage of small sample requirements and the ability to profile complex mixtures of glycans. But unlike the other methods, MS does not provide structural information for isomeric structures.[20,21] In addition to all pros and cons already discussed, most conventional methods of glycosylation analysis suffer from the fact that no glycosylation site-specific information is obtained, because the carbohydrates are cleaved from the protein prior to analysis.

In order to obtain glycosylation information in a site-specific fashion, glycoproteins can be subjected to proteolysis, followed by MS analysis of glycopeptides.[22-25] We implement this method herein to characterize FSH glycopeptides in two species (human and equine). A nonspecific protease is

used for proteolysis, to produce small glycopeptide fragments, and a combination of Edman sequencing and mass spectrometry are used to analyze the digestion products. The information obtained from this method is used to compare the glycans at different glycosylation sites of human and equine FSH preparations.

3.2. Experimental procedures

The horse pituitary glycoprotein hormone preparation, eFSH, was prepared as previously described.[26] Human FSH (AFP-4161B) was obtained from the NIDDK National Hormone and Pituitary Program and Dr. A.F. Parlow.[27,28]

3.2.1. Protease digestion

Glycoprotein hormone preparations, which typically contain approximately 2-5 mg of protein were reduced with dithiothreitol and alkylated with iodoacetic acid as described previously.[26] The reaction mixtures were desalted using 4-ml Amicon (Millipore Corporation, Bedford, MA) Ultra-4 centrifugal ultrafiltration cartridges (MW cutoff 10,000) in a Sorvall (Kendro Laboratory Products, Newtown, CT) RC-3B Plus centrifuge (5,000 rpm, LA/S-400 rotor). After washing 3 times with 0.2 M ammonium bicarbonate, pH 8.5, the protein solutions were transferred to 2-ml polypropylene screw cap tubes and 10% (w/w) proteinase K (Boerhinger Mannheim, Indianapolis, IN) was added. The tubes were incubated overnight at 37°C with continuous shaking.

The next day, samples were dried in a Thermo Savant (Marietta, OH) Speed Vac centrifugal evaporator.

3.2.2. Superdex gel filtration

A 10/30 Amersham Biosciences (Piscataway, NJ) Superdex peptide column was connected to a Waters (Milford, MA) model 600 HPLC system equipped with a model 484 variable wavelength detector controlled by Empower Pro (build 1154) control/acquisition software. The column was equilibrated with 0.2 M ammonium bicarbonate buffer at a flow rate of 0.4 mL/min. Glycopeptide samples, (typically 1-2.5 mg) were dissolved in 200 µL 30% acetonitrile in 0.2 M ammonium bicarbonate and injected into the HPLC. The chromatogram was developed at a flow rate of 0.4 mL/min. Fractions were collected by hand in 12 x 75 mm polypropylene tubes and dried in the Speed Vac. The dried residues were dissolved in 500-1000 µL water, 25 µL aliquots were removed and dried in 2-mL screw cap polypropylene tubes. Dried samples were hydrolyzed in 200 µL 4 N trifluoroacetic acid at 100°C for 4 hr. TFA was removed by evaporation, and the hydrolysates were dissolved in 100 µL water containing deoxyglucose. Monosaccharide composition was determined using a Dionex (Sunnyvale, CA) carbohydrate analyzer consisting of a model GP50 gradient pump and LC20 chromatography enclosure equipped with a 4 x 50 mm BorateTrap column, a 4 x 50 mm AminoTrap column, which functioned as a guard column, a 4 X 250 mm CarboPak PA-10 column, model ED50 electrochemical detector, and controlled by Chromeleon control/acquisition software. A Waters model 717 refrigerated autosampler maintained the hydrolysates at 4°C and injected 20 µL samples from each

vial. Carbohydrate-positive samples were subsequently analyzed in duplicate. The fraction containing the largest amount of carbohydrates (90% of the glycopeptides in eFSH and 75% of the glycopeptides in hFSH) were subjected to mass spectral analysis.

3.2.3. Peptide sequencing

Automated Edman degradation was carried out using an Applied Biosystems model 492 Procise sequencer.[24] Samples were applied to glass fiber membranes coated with Biobrene that had been precycled in the sequencer. Typical sequencer experiments consisted of 7 Edman degradation cycles, sufficient to sequence the entire length of 3-5 residue peptides that were the most common products of proteinase K digestion.

3.2.4. Mass Spectrometry

MS data was acquired on a high-resolution Thermo Finnigan linear ion trap-Fourier Transform Ion Cyclotron Resonance mass spectrometer, LTQ-FTICR-MS, (San Jose, CA) equipped with a 7 Tesla actively shielded magnet. The dried glycopeptides were first dissolved in water and diluted with MeOH: H_2O (4:1) containing 0.3% acetic acid, to a final concentration of 0.03 µg/µL. Samples were directly infused into the mass spectrometer using a syringe pump at a flow rate of 5 µL/min. High resolution data was acquired by maintaining resolution at 50,000, for *m/z* 400. The instrument was externally calibrated over the mass range of interest (*m/z* 800-2000) prior to the analysis. ESI spray voltage was maintained between -3.0kV to -4.0 kV for

data acquisition in the negative mode. Capillary temperature was held between 200-230^{o}C and N_2 was used as a nebulizing gas at 20 psi. Data was acquired and processed using Xcalibur 1.4 SR1 software (Thermo Finnigan San Jose, CA).

3.2.5. Glycopeptide composition assignment

The glycopeptide compositions were assigned for the peaks in the mass spectrum using the aid of two computer algorithms written in VBA and Perl. The VBA algorithm was described previously.[24] The perl program follows a mass fingerprinting algorithm that matches theoretical and experimental masses at a given tolerance. Whenever possible, MS/MS analysis was performed on the glycopeptides, to verify the peak assignments. This provided supporting information but could not be used to unequivacably assign branching.

3.3. Results

The approach utilized for characterizing the glycosylation in a site-specific fashion is as follows. The hormones were digested with a non-specific serine protease, proteinase K. Typically, smaller peptides are generated from the proteinase K digestion than from a tryptic digest. (Pronase, another nonspecific enzyme was also tested in preliminary experiments, but it produced less optimal signal). The small peptides produced by proteinase K facilitate separation of nonglycosylated peptides

from glycopeptides by gel filtration. Other strategies, including reverse phase HPLC and separations using porous graphite were also investigated, but showed less optimal results. Reverse-phase HPLC irreversibaly adsorbed eFSH glycopeptides. Adsorption to Carbograph graphitized charcoal cartridges removed proteinase-K, but not the small peptides. Some glycopeptides were lost during the Carbograph adsorption process, and resulting mass spectrometry signals were weak, perhaps due to residual TFA. Gel filtration of the total digest removed both protease and small peptides.

An example of the gel filtration fractionation data appears in Figure 3.1. The eFSH and hFSH proteinase K digest chromatograms included a large protease peak followed by a small peak that was eluted before the peptide peaks. As low UV absorbance is characteristic of glycopeptides, we surveyed hydrolyzed samples of all the column fractions for carbohydrate and found it was largely restricted to the putative glycopeptide peak (Fig. 3.1, solid bar). Most of the eFSH carbohydrate was restricted to a single fraction (Figure 3.1A). However, 13% of the hFSH carbohydrate co-eluted with the protease fraction, 13% emerged in the next fraction, and 67% was recovered from the 27-33 min fraction [hFSH glycopeptide yield is increased to 90% by altering collection times to 25-31 min] (Figure 3.1B). This was consistent with the larger size of hFSH oligosaccharides as compared with those derived from eFSH. The size fractionation strategy enhances the mass spectral signal of the glycopeptides by removing the nonglycosylated peptide interference.

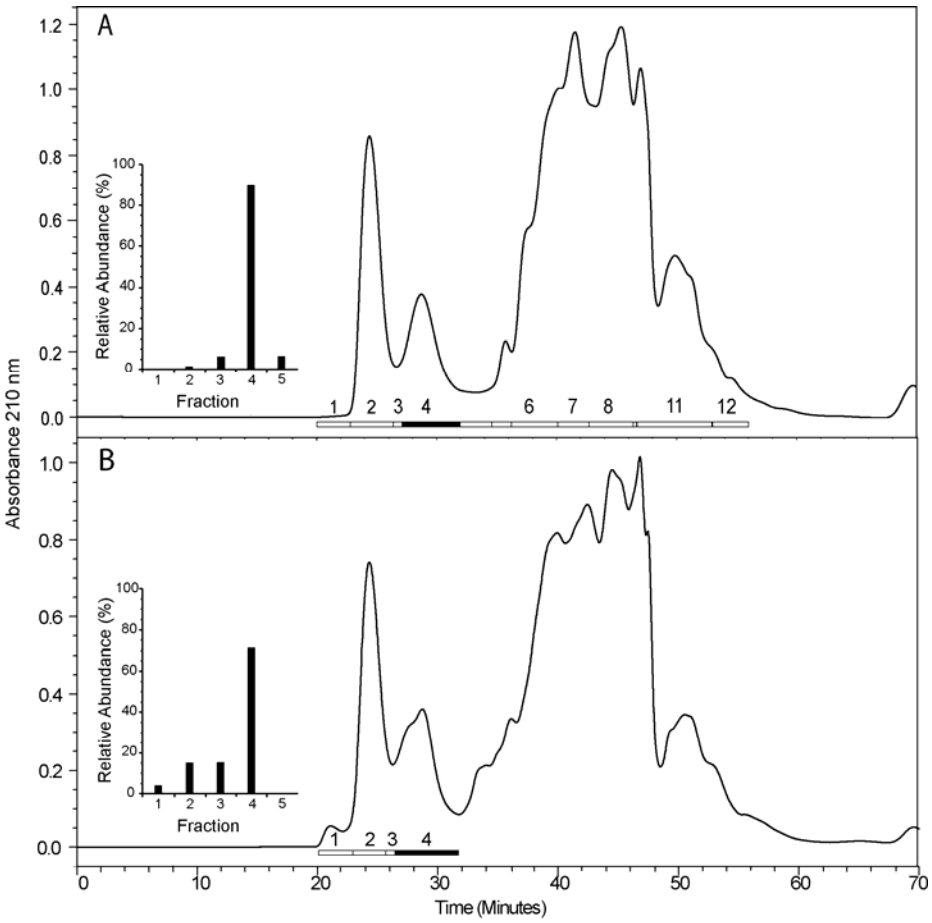


Figure 3.1 : Fractionation of 200 µL samples containing 2500 µg of A) eFSH or B) hFSH proteinase K digest on a Superdex peptide column (absorbance at 210 nm) equilibrated with 0.2M NH_4HCO_3 at a flow rate of 0.4 mL/min. The protease peak elutes first, followed by the glycopeptide peak and then the peptide peaks. *Inset*: The solid bars represent monosaccharides quantified from each of the different fractions, collected manually in 12 x 75 millimeter tubes. For each fraction, carbohydrate content was determined by hydrolyzing any glycopeptides present with TFA to release monosaccharides. These were quantified against an internal standard, deoxyglucose, using a Dionex HPAEC system, as described in the experimental section. Quantitative results from five fractions of the dozen collected are shown here, because carbohydrate composition analysis indicated that glycans were present only in the first 4-5 fractions

Subsequently, the glycopeptide fractions were subjected to both Edman sequencing and FTICR- MS for analysis. Edman data is useful for identifying the amino terminus and predicting the peptide sequence for each of the glycopeptides released from the digest. Quantitative output from the Edman degradations is shown in Figure 3.2 and interpretation of Edman data is illustrated below. This study focuses on characterizing glycopeptides in a glycosylation site specific manner, and special emphasis is placed on determining the unique properties of glycans that are present and absent from each site in human and equine FSH respectively.

a) **Cycle 1 -** K (185pmole) ; I (182pmole)
Cycle 2 - V (46pmole) ; H (127pmole) ; I (65pmole)
Cycle 3 - T (318pmole) ; V (182pmole)
Cycle 4 - T (234pmole)

Possible sequences in hFSH

KNVT, NVT, NHT, NIT, INTT

b) **Cycle 1 -** L (361pmole)
Cycle 2 - E (310pmole) ; H (70pmole) ; I (477pmole)
Cycle 3 - T (388pmole)
Cycle 4 - H (176pmole) ; Q (63pmole)
Cycle 5 - T (207pmole)

Possible sequences from eFSH alpha subunit

LENHTQ, NHTQ, NIT

c) **Cycle 1 -** S (112pmole) ; I (171pmole) ; T (41pmole)
Cycle 2 - I (299pmole)
Cycle 3 - I (40pmole) ; T (266pmole)
Cycle 4 - T (281pmole)
Cycle 5 - T (152pmole)
Cycle 6 - T (66pmole)

Possible sequences from eFSH beta subunit

SINTT, INTT, TNIT, NIT

Figure 3.2: Quantitative output of Edman data from different cycles of hormone isolates. a) hFSH b) eFSH α subunit c) eFSH β subunit. The absolute amount of Aspargine (N) is absent in these cycles because glycosylated amino acid derivatives are not extracted from the Edman chemistry cartridge.

3.4. Peptide compositions from Edman sequencing

To interpret the Edman data for a mixture of glycopeptides, the protein sequence and the position of glycosylation sites must be known. The amino acid sequences of eFSH and hFSH in Figure 3.3 show that α subunit glycosylation takes place at Asn^{52} and Asn^{78} in human FSH[29] while the corresponding glycosylation sites in eFSH are Asn^{56} and Asn^{82}. Both

hormones are glycosylated on the β subunit, at Asn^{7} and Asn^{24}.[29] From this information, the peptide sequences from Edman data can be inferred.

a) eFSH

Amino acid sequence of eFSH α subunit

FPDGEFTTQDCPECKLRENKYFFKLGVPIYQCKGCCFSRAYPTPARSRKTMLV
PK**N** ^{56}**IT**SESTCCVAKAFIRVTVMG NIK**LEN**^{82}**HTQ**CYCSTCYHHKI

Amino acid sequence of eFSH β subunit

NSCEL**TN**^{7}**IT**IAVEKEECGFCI**SIN**^{24}**TT**WCAGYCYTRDLVYKDPARPNIQKTCTFK
ELVYETVK VPGCAHHADSLYTYPVATACHCGKCNSDSTDCTVRG LGPSYCSF
GDMKE

b) hFSH

Amino acid sequence of hFSH α subunit

APDVQDCPECTLQENP FFSQPGAPILQCMGCCFSRAYPTPLRSKKT MLVQK
N^{52}**VT**SE STCCVAKSYNRVTVMGGFKVE**N**^{78}**HT**ACHCSTCYYHKS

Amino acid sequence of hFSH β subunit

NSCELT**N**^{7}**IT**IAIEKEECRFCIS**IN**^{24}**TT**WCAGYCYTRDLVYKDPARPKIQKTCTFK
ELVYETVRVPGCAHHADSLYTYPVATQCHCGKCDS DSTDCTVR GLGPSYCS
F GEMKE

Figure 3.3: Amino acid sequences for both a) eFSH and b) hFSH α and β subunits showing the glycosylation sites.

The Edman data are interpreted by matching the sequences of amino acids near the known glycosylation sites of the protein (Figure 3.3) with the amino acid derivatives observed in the Edman sequencing experiments (Figure 3.2). For example, the glycopeptide KNVT is presumed to be present in the hFSH digestion, since >PhNCS-Lys is abundant in cycle one, >PhNCS -Val is in cycle three and >PhNCS -Thr is present in cycle four. No >PhNCS -

Asn (N) was detected in cycle two, because glycosylated amino acid derivatives are not extracted from the Edman chemistry cartridge by the adsorptive protocol employed in these studies. In addition to KNVT, several other glycopeptides are likely to be present in the glycopeptide mixture, based on the Edman data. All the reasonable compositions obtained for hFSH are shown in Figure 3.2a.

Even though the Edman data identify a list of possible sequences, the mass spectral data cannot be used to distinguish among sequences that are isomeric. For instance, glycans that originate from $\alpha N^{56}IT$ and $\beta N^{7}IT$ of eFSH cannot be distinctively identified. Therefore there was a need to separate the alpha and the beta subunits of eFSH, in order to uniquely identify glycans present on these two glycopeptides. Figures 3.2b and 3.2c show the Edman data for the separated subunits of eFSH alpha and beta subunits.

As described in the experimental section, Edman data were combined with the mass spectral data to obtain glycopeptide compositions for both equine and human FSH. Figures 3.4, 3.5 and 3.6 represent the mass spectra obtained for both hormones in the negative ion mode. The peak list corresponding to the numbers on the spectra can be found in supplemental Tables 1, 2 and 3. More than 35 glycopeptides were identified for each of these two hormones.

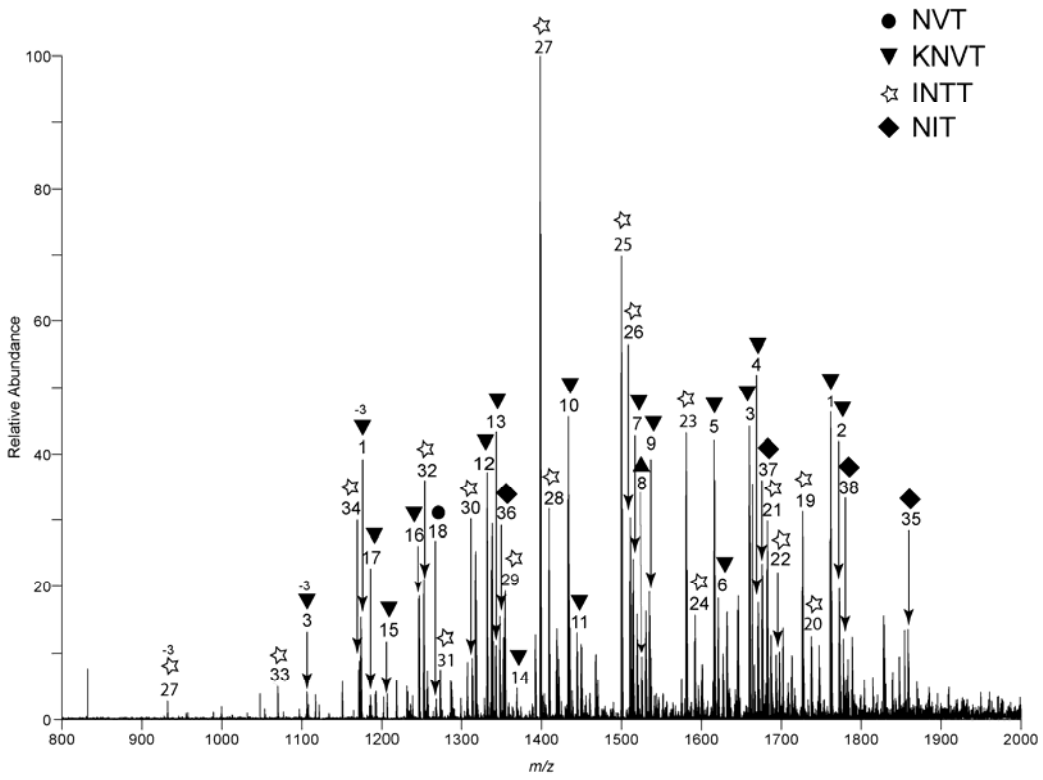


Figure 3.4: FT-ICR-MS spectrum obtained for hFSH in negative mode. Different symbols have been used identify glycopeptides from different glycosylation sites from α and β subunits. Symbols : ◆ = NIT, ✩ = INTT, ● = NVT, ▼ = KNVT. All peaks are numbered according to supplement Table 1. Peaks without a charge are all doubly charged. The same number is used to identify the composition having different charge states on the spectrum.

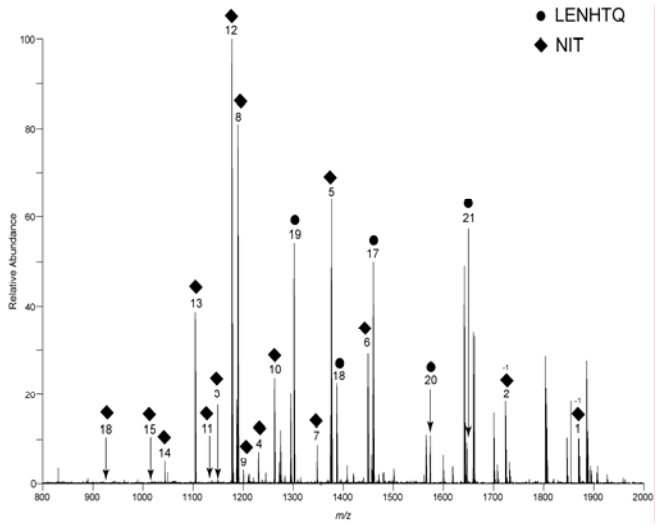


Figure 3.5: FT-ICR-MS spectrum obtained for eFSH α subunit in the negative mode. Different symbols have been used identify glycopeptides associated from different glycosylation sites from the α subunit. Symbols : ◆ = NIT, ● = LENHTQ. All peaks are numbered according to supplement Table 2. Peaks without a charge are all doubly charged.

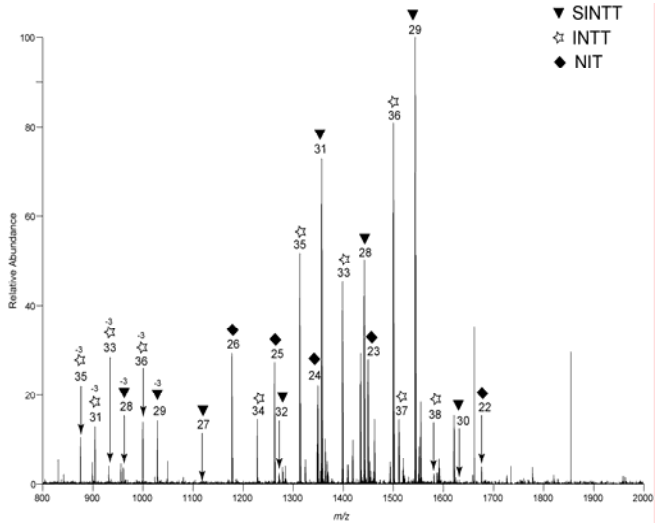
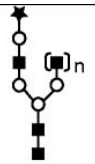
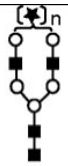
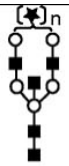
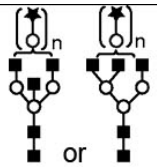
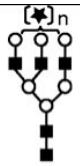
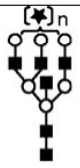
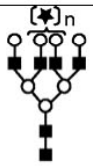
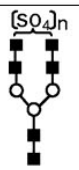
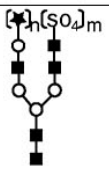
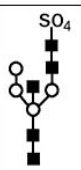
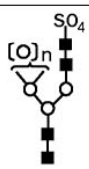


Figure 3.6: FT-ICR-MS spectrum obtained for eFSH β subunit in the negative mode. Different symbols have been used identify glycopeptides associated with different glycosylation sites from β subunit. Symbols : ☆ = INTT, ▼ = SINTT, ◆ = NIT. All peaks are numbered according to supplement Table 3. Peaks without a charge are all doubly charged.

Pictorial description of the glycans can be found in Table 3.1.This is further illustrated in Figure 3.7, which provides a comparative description of the glycans present on each of the glycosylation sites in eFSH and hFSH. This study exclusively compares the anionic species found in FSH since it had been found previously that 90% of the total glycans in human FSH are sialylated or sulfated,[2] and the heterogeneity of these hormones is primarily determined by the variability of the anionic species.[30]

Table 3.1 : Glycan structures of equine and human pituitary FSH.[a]

Peptide	Site on protein						
		A	B	C	D or E	F	G
		n=0 n=1	n=1 n=2	n=1	n=1 n=2 n=1n=2	n=1 n=2 n=3	n=1
hFSH							
KNVT	α^{52}		12 13		15 10 11	14 7 3 8 4	9
NVT	α^{52}		18				
INTT	β^{24}	33* 34*	32* 27* 28*	29*	31* 25* 26*	23* 19* 24* 20*	
NIT	β^{7}		36*			37*	
eFSH							
NIT	α^{56}		7*	4	5,6*		
	β^{7}		24*		23*	22*	
LENHTQ	α^{82}				20 21*		
SINTT	β^{24}	27*	28*		29*	30*	
INTT	β^{24}/β^{7}		33*		36* 37*	38*	

Peptide	Site on protein						
		H	I	J	K	L	M
		n=1 n=2 n=3	n=3	n=1 n=2	n=1 n=0 n=1 m=1 m=1 m=0		n=0 n=1 n=2
hFSH							
KNVT	α^{52}	5,6 1,2			16 17		
NVT	α^{52}						
INTT	β^{24}	21* 22*			30*		
NIT	β^{7}	38*	35*				
eFSH							
NIT	α^{56}			11* 12* 13	8,9 3 10*	14	1*,2 16 15
	β^{7}			26*	25*		
LENHTQ	α^{82}			19	17*,18		
SINTT	β^{24}			32*	31*		
INTT	β^{24}/β^{7}			34*	35*		

[a]Numbers appearing on the table correspond to the peaks on the mass spectra, in Figures 5, 6, and 7. hFSH glycans correspond to data in Figure 5, while eFSH glycans correspond to data from Figures 6 and 7. Structures having an asterisk are all fucosylated.
Symbols : ■ = HexNAc, ○ = hex, ★ = Sialic acid

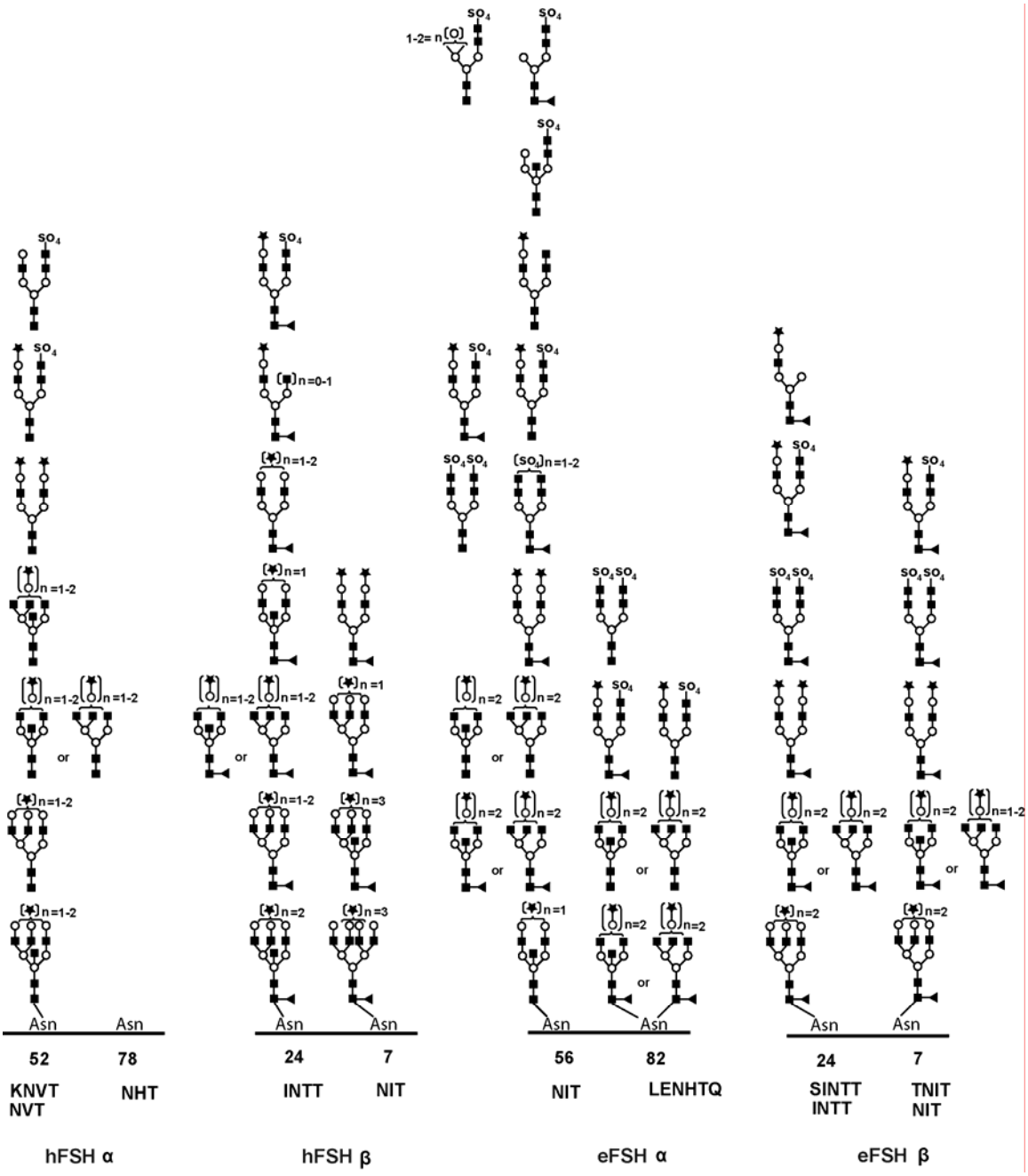


Figure 3.7: A comparative description of the glycans present on each of the glycosylation sites in eFSH and hFSH. Symbols: ■ = HexNAc, ○ = Hexose, ▲ = Fucose, ★ = Sialic acid. Where bisecting, biantennary and triantennary glycans are possible, both structures are shown. Tri antennary structures could be branched at the 3-linked or 6-linked mannose, but only one structure was depicted for simplicity.

3.5. Pictorial representation of glycans

While others have previously reported the glycan compositions and structures for FSH[2,18,20] this is the first report that provides information about which glycans are present at all four eFSH glycosylation sites and three of four hFSH sites. While MS data only provide compositions for each of the glycopeptides, these compositions can be converted to glycan structures, by comparing the data in this study to previous carbohydrate analyses.[2,18] Table 1 contains pictorial descriptions of each of the observed glycans, and it shows which glycans were found at which glycosylation site, for each of the hormones. Baenziger and coworkers[2] had identified all but seven of the structures that we observed in pituitary FSH from their glycosidase digestion and glycan chromatography data. [Those seven are: A(n=1), C, G, I, K(n=0, m=1), K(n=1, m=0), and L]. Four of these species, A(n=1), C, G, and L, are virtually identical to four other glycans in the table, A(n=0), B, E(n=1) and M (n=1), with the exception that the former glycans contain an additional GlcNAc moiety. Two other structures, K(n=0, m=1) and K(n=1, m=0) are identical to K (n=1, m=1), except they contain only a sulfate or a sialic acid group instead of both groups. The composition of one of the structures, I, matches a tetraantennary glycan previously reported by Renwick et al,[18] which was identified from another FSH isolate.

3.6. Discussion

3.6.1. Characteristic patterns of glycans in eFSH at different sites

It can be seen from Table 3.1 and Figure 3.7 that a majority of the glycans associated with the peptide, NIT, are attached to α Asn^{56}. This is consistent with the fact that the glycans attached to NIT in the α subunit exhibited a large degree of variation.[20] Bi-antennary sialylated and sulfated glycans were present. These were predominantly of the complex type, but some hybrid forms were also observed. The glycans attached to $N^{7}IT$ in the beta subunit consisted of fewer structures, suggesting a lesser degree of heterogeneity. The extent of sulfation was less than at α Asn^{56}, and the structures were all complex type and mostly biantennary.

Glycopeptides derived from the α subunit Asn^{82} site of eFSH can be definitively identified from the unique peptide sequence $LEN^{82}HTQ$. These possess mainly biantennary glycans having di- sialylated or sulfated structures as well as glycans containing both a sulfate and a sialic acid group. Since $SIN^{24}TT$ is uniquely mapped to the β Asn^{24} glycosylation site, glycans at β Asn^{24} can also be identified definitively, and these belong to either bi- or triantennary classes of oligosaccharide structures (A, B, D or E, F, J (n=2) , and K(n=1, m=1)), which are mainly sialylated. Except for a single glycan attached to the peptide sequence $SIN^{24}TT$, all other glycans associated with SINTT are also present as glycans attached to the related peptide INTT. When pairs of ions corresponding to the glycopeptides SINTT and INTT are

present, one can infer that these ions come from the same glycosylation site in the original protein, in this case the β Asn^{24} site. However, since $TN^{7}IT$ was also identified as a potential peptide from Edman sequencing, and mass spectrometry could not discriminate between $TN^{7}IT$ and $IN^{24}TT$, it is possible that some or all of the glycans that were identified from Asn^{24}, with the sequence $IN^{24}TT$, were also present at Asn^{7}.

All glycans attached to both glycosylation sites in the β subunit were fucosylated, in contrast to the α subunit where only a fraction of glycans showed fucosylation. Similar observations were reported by Renwick et al. for ovine and human lutropin,[31,32] human thyrotropin,[33] and human choriogonadotropin.[34] The function of a fucose residue in glycoprotein hormone glycans is not very well understood[35] but according to some investigators, addition of fucose can be as important as the addition of sulfate or sialic acid, since it ensures the proper binding to the appropriate receptor.[36] While fucosylated glycans enabled hTSH to activate more than one intracellular signaling pathway,[29,37] creating hybrid hormone preparations based on fucosylation of αAsn^{56} glycans had no effect on LH activation of the PLC pathway in cultured ovarian luteal cells.[1,38] The differences in fucosylation correlate with the fact that glycans on the α subunit are buried within the dimer while those on the β subunit are solvent exposed.[9]

The eFSH data we present herein represent a major advance in site-specific glycosylation analysis. Only a single eFSH glycosylation site had

been characterized previously. Neutral and negatively charged glycan structures released by selective PNGase hydrolysis were identified by MALDI-MS.[20] The present study identified 16 of these negatively charged glycans, which represented most of the major negatively charged glycans found in the earlier study. Phosphorylated high mannose glycans were the only forms not identified as glycopeptide ions. While the low-abundance glycans were not detected herein, this work is highly complementary to previous analyses because it also presents glycosylation site-specific information about the other three eFSH glycosylation sites.

3.6.2. Characteristic patterns of glycans in hFSH at different sites

Glycans attached to both KNVT and INTT can be easily mapped to their specific glycosylation sites because of the unique peptide sequences. The KNVT glycopeptides were derived from the α Asn^{52} site, while INTT containing glycopeptides mapped to the β Asn^{24} site. Both sites possessed almost the same glycan population, which were exclusively sialylated, mainly bi- and triantennary structures. The second glycosylation site on the β subunit (β Asn^{24}), had a significantly different oligosaccharide population than what was observed at the second site, β Asn^{7}. The glycans at the former site were largely biantennary with only 2-3 triantennary glycans and a single monoantennary glycan. While most glycans were terminated with sialic acid, a sulfated glycan was present in this population. The glycans at the latter site

were mostly triantennary, with only a single biantennary glycan structure. The only tetrantennary structure identified in the present study came from this site. One commonality between glycosylation at β Asn^{24} and β Asn^{7} was that all glycans were fucosylated, whereas the glycans from the α subunit had no fucose. Therefore in general, the hFSHβ subunit contains fucoslylated, bi- ,tri- and tetraantennary glycans terminated with sialic acid.

There were no mono- or disulfated complex glycans or sulfated hybrid type glycans observed for hFSH in either subunit. The few sulfated glycans present also contained sialic acid. However Green et al.[2] had identified mono-sulfated, hybrid type glycans that constituted 2% of the total FSH carbohydrate population. This difference may be attributed to variation between two different hFSH preparations.

The present study represents a significant advance in characterizing hFSH glycosylation. We were able to characterize three of four N-glycosylation sites and account for all but two glycan structures reported in two separate studies by other investigators.[2,18] The site that this analysis was unable to detect Asn^{78}, must possess glycans that are also present at the other three glycosylation sites, since the carbohydrate structures detected in two total oligosaccharide analyses of hFSH were essentially the same as those that we detected from glycopeptides derived from all sites, excluding those at Asn^{78}. The fact that the Asn^{78} glycopeptides were not detected is not a significant limitation of this study, because glycans associated with the *other*

three sites, Asn^{52}, Asn^{7}, and Asn^{24} appear to regulate hFSH biological activity.[13-15] Therefore, the data presented here represent a major step toward to our goal of correlating glycan structure with hFSH function.

3.6.3. Comparison of glycans on equine and human FSH

From Table 1 it is evident that hFSH α subunit contains bi- and triantennary, complex type glycans, which are almost all sialylated, whereas the α subunit of eFSH possesses only biantennary glycans, some of which are sulfated and others of which are sialylated. (While some triantennary structures might exist, in every case an alternative biantennary glycan with a bisecting HexNAc is also possible.) Hybrid type glycans are only present in the α subunit of eFSH. The absence of larger glycans specifically at the Asn^{56} of the α subunit could imply that eFSH α subunit is less sterically hindered by smaller glycans present at this glycosylation site compared to larger glycans encountered at the corresponding site of the hFSH α subunit; therefore the former could bind more tightly to the receptor.

While all glycans in the β subunit are fucosylated in both eFSH and hFSH, several other structural differences are apparent. The β subunit of eFSH contains mostly biantennary glycans that are mainly sialylated, whereas in hFSH bi-, tri-, and tetra-antennary glycans are present. In eFSH, both sites in the β subunit contain sulfated glycans, whereas only the β Asn^{24} site contains sulfated glycans in hFSH. This comparison between human and

equine FSH glycosylation on the β subunit demonstrates that glycosylation can vary dramatically, even when comparing proteins from two mammalian sources, whose amino acid sequence is ~94% identical.

This work also complements other emerging evidence that demonstrates structural differences between human and horse FSH. Human and equine FSH preparations analyzed in this study represent two patterns of mammalian FSH glycosylation. In the former, which may be representative for primate FSH,[39] the glycans are larger, consist of 2-4 complex branches, are terminated with sialic acid, and contain very little sulfation. In contrast, smaller, 1-2 branch, sulfated glycans are identified at every glycosylation site of eFSH. This places the horse FSH at the upper end of the continuum for glycan sulfation,[2] and this structural difference is important because higher degrees of sulfated are associated with more rapid hormone clearance.[6]

While we have demonstrated that the individual glycans on the β subunit of human and equine FSH are different, other studies have also demonstrated that the global pattern of glycosylation on this subunit also varies, for these two hormones. It has been recently shown that hFSHβ subunit is glycosylated in an all-or-none manner,[39] whereas eFSHβ is glycosylated either exclusively at Asn^{24} or at both Asn^{7} and Asn^{24}.[1] In humans, the relative abundance of glycosylated and non-glycosylated FSHβ subunit appears to be physiologically regulated.[39] Both the information presented here, which provides information about glycan composition at

specific glycosylation sites, and earlier work, which documents differences in the presence and absence of glycosylation on the β subunit, are important studies needed to fully understand the glycoprotein structure. However, to clearly understand how the structure of these hormones induces specific biological effects, thouroughly characterized isoforms, with defined glycan composition, will be needed for use in functional assays. These functional studies are a future emphasis of our research, and.the site-specific glycolsyation analysis method presented in this report is a key component to making such structure-function studies possible.

3.7. Supporting information

Table 1: Glycopeptides compositions for hFSH hormone

Peak Number	C 12 Exp	C 12 Cal	Mass error	Compositions	Charge
1	1761.1704	1761.1647	3.236495	[KNVT + 6HexNAc + 6Hex + 3sialic acid]	2#
2	1772.1629	1772.1559	3.949991	[KNVT + 6HexNAc + 6Hex + 3sialic acid+Na]	2
3	1659.6329	1659.6250	4.760111	[KNVT + 5HexNAc + 6Hex + 3sialic acid]	2#
4	1670.6210	1670.6162	2.873191	[KNVT + 5HexNAc + 6Hex + 3sialic acid+Na]	2
5	1615.6238	1615.6170	4.208918	[KNVT + 6HexNAc + 6Hex + 2sialic acid]	2
6	1626.6109	1626.6082	1.659896	[KNVT + 6HexNAc + 6Hex + 2sialic acid +Na]	2
7	1514.0821	1514.0773	3.170248	[KNVT + 5HexNAc + 6Hex + 2sialic acid]	2
8	1525.0717	1525.0685	2.098266	[KNVT + 5HexNAc + 6Hex + 2sialic acid +Na]	2
9	1534.5962	1534.5906	3.649182	[KNVT + 6HexNAc + 5Hex + 2sialic acid]	2
10	1433.0538	1433.0509	2.023655	[KNVT + 5HexNAc + 5Hex + 2sialic acid]	2
11	1444.0449	1444.0421	1.939002	[KNVT + 5HexNAc + 5Hex + 2sialic acid + Na]	2
12	1331.5149	1331.5112	2.778798	[KNVT + 4HexNAc + 5Hex + 2sialic acid]	2
13	1342.5036	1342.5024	0.893853	[KNVT + 4HexNAc + 5Hex + 2sialic acid +Na]	2
14	1368.5309	1368.5296	0.949925	[KNVT + 5HexNAc + 6Hex + 1sialic acid]	2
15	1206.4798	1206.4768	2.486579	[KNVT + 5HexNAc + 4Hex + 1sialic acid]	2
16	1246.4549	1246.4552	0.240683	[KNVT + 5HexNAc + 4Hex + 1sialic acid +SO3	2
17	1185.9620	1185.9635	1.264794	[KNVT + 4HexNAc + 5Hex + 1sialic acid]	2
18	1267.4622	1267.4637	1.264794	[NVT + 4HexNAc + 5Hex + 2sialic acid]	2
19	1726.1460	1726.1381	4.576691	[INTT + 5HexNAc + 6Hex + 3Sialic acid + Fucose]	2
20	1737.1376	1737.1294	4.720431	[INTT + 5HexNAc + 6Hex + 3Sialic acid + Fucose + Na]	2
21	1682.1368	1682.1301	3.983045	[INTT + 6HexNAc + 6Hex + 2Sialic acid + Fucose]	2
22	1693.1278	1693.1214	3.780001	[INTT + 6HexNAc + 6Hex + 2Sialic acid + Fucose + Na]	2
23	1580.5947	1580.5904	2.720502	[INTT + 5HexNAc + 6Hex + 2Sialic acid + Fucose]	2
24	1591.5859	1591.5817	2.638884	[INTT + 5HexNAc + 6Hex + 2Sialic acid + Fucose + Na]	2
25	1499.5675	1499.5640	2.334012	[INTT + 5HexNAc + 5Hex + 2Sialic acid + Fucose]	2
26	1510.5603	1510.5553	3.310041	[INTT + 5HexNAc + 5Hex + 2Sialic acid + Fucose + Na]	2
27	1398.0274	1398.0243	2.217415	[INTT + 4HexNAc + 5Hex + 2Sialic acid + Fucose]	2#
28	1409.0188	1409.0156	2.271089	[INTT + 4HexNAc + 5Hex + 2Sialic acid + Fucose + Na]	2
29	1354.0183	1354.0163	1.477087	[INTT + 5HexNAc + 5Hex + 1Sialic acid + Fucose]	2
30	1312.9706	1312.9683	1.751756	[INTT +5HexNAc + 4Hex + 1Sialic acid + Fuc + SO3]	2
31	1272.9876	1272.9899	1.80677	[INTT + 5HexNAc + 4Hex + 1Sialic acid + Fucose]	2
32	1252.4790	1252.4766	1.916203	[INTT + 4HexNAc + 5Hex + 1Sialic acid + Fucose]	2
33	1069.9114	1069.9105	0.841192	[INTT + 3HexNAc + 4Hex + 1Sialic acid + Fucose]	2
34	1171.4509	1171.4502	0.59755	[INTT + 4HexNAc + 4Hex + 1Sialic acid + Fucose]	2
35	1858.1835	1858.1804	1.668299	[NIT + 6HexNAc + 7Hex + 3Sialic acid + 1Fucose]	2
36	1347.5020	1347.5005	1.113172	[NIT + 4HexNAc + 5Hex + 2Sialic acid + 1Fucose]	2
37	1675.6231	1675.6143	5.251805	[NIT + 5HexNAc + 6Hex + 3Sialic acid + 1Fucose]	2
38	1777.1622	1777.1540	4.614119	[NIT + 6HexNAc + 6Hex + 3Sialic acid + 1Fucose]	2

#Apart from the doubly charged ions, triply charged ions also exist for the same composition. The same number is used to indicate both on the spectrum.

Table 2: Glycopeptide compositions for eFSH α subunit

Peak Number	C 12 Exp	C 12 Cal	Mass error	Compositions	Charge
1	1869.6678	1869.6686	0.427883	[NIT + 4HexNAc + 3Hex + 1SO3 + Fucose]	1
2	1723.6042	1723.6107	0.427883	[NIT + 4HexNAc + 3Hex + 1SO3]	1
3	1149.4392	1149.4371	1.826981	[NIT + 5HexNAc + 4Hex + 1Sialic acid]	2
4	1230.4647	1230.4635	0.975242	[NIT + 5HexNAc + 5Hex + 1Sialic acid]	2
5	1376.0146	1376.0112	2.47091	[NIT + 5HexNAc + 5Hex + 2Sialic acid]	2
6	1449.0415	1449.0402	0.897146	[NIT + 5HexNAc + 5Hex + 2Sialic acid + Fucose]	2
7	1347.5019	1347.5005	1.038961	[NIT + 4HexNAc + 5Hex + 2Sialic acid + Fucose]	2
8	1189.4176	1189.4155	1.765573	[NIT + 5HexNAc + 4Hex + Sialic acid + 1SO3]	2
9	1200.4063	1200.4068	0.416525	[NIT + 5HexNAc + 4Hex + Sialic acid + SO3 + Na]	2
10	1262.4451	1262.4445	0.475268	[NIT + 5HexNAc + 4Hex + Sialic acid + 1SO3 + Fucose]	2
11	1137.4104	1137.4100	0.351676	[NIT + 6HexNAc + 3Hex + 1SO3 + Fucose]	2
12	1177.3888	1177.3884	0.339735	[NIT + 6HexNAc + 3Hex + 2SO3 + Fucose]	2
13	1104.3599	1104.3595	0.362201	[NIT + 6HexNAc + 3Hex + 2SO3]	2
14	1043.8692	1043.8678	1.341166	[NIT + 5HexNAc + 4Hex + 1SO3]	2
15	1023.3539	1023.3545	0.586307	[NIT + 4HexNAc + 5Hex + 1SO3]	2
16	942.3269	942.3281	1.273442	[NIT + 4HexNAc + 4Hex + 1SO3]	2
17	1459.5273	1459.5245	1.918433	[LENHTQ + 5HexNAc + 4Hex + 1Sialic acid + 1SO3 + Fucose]	2
18	1386.4975	1386.4956	1.370361	[LENHTQ + 5HexNAc + 4Hex + 1Sialic acid + 1SO3]	2
19	1301.4396	1301.4395	0.076838	[LENHTQ + 6HexNAc + 3Hex + 2SO3]	2
20	1573.0938	1573.0913	1.589228	[LENHTQ + 5HexNAc + 5Hex + 2Sialic acid]	2
21	1646.1244	1646.1202	2.551454	[LENHTQ + 5HexNAc + 5Hex + 2Sialic acid + Fucose]	2

Table 3: Glycopeptide compositions for eFSH β subunit

Peak Number	C 12 Exp	C 12 Cal	Mass error	Compositions	Charge
22	1675.6179	1675.6143	2.1484	[NIT + 5HexNAc + 6Hex + 3Sialic acid + Fucose]	2
23	1449.0415	1449.0402	0.8971	[NIT + 5HexNAc + 5Hex + 2Sialic acid + Fucose]	2
24	1347.5019	1347.5005	1.0389	[NIT + 4HexNAc + 5Hex + 2Sialic acid + Fucose]	2
25	1262.4451	1262.4445	0.4752	[NIT + 5HexNAc + 4Hex + Sialic acid + 1SO3 + Fucose]	2
26	1177.3888	1177.3884	0.3397	[NIT + 6HexNAc + 3Hex + 2SO3 + Fucose]	2
27	1113.4273	1113.4260	1.1675	[SINTT + 3HexNAc + 4Hex + 1Sialic acid + Fucose]	2
28	1441.5410	1441.5404	0.4162	[SINTT + 4HexNAc + 5Hex + 2Sialic acid + Fucose]	2#
29	1543.0806	1543.0800	0.3888	[SINTT + 5HexNAc + 5Hex + 2Sialic acid + Fucose]	2#
30	1624.1108	1624.1065	2.6476	[SINTT + 5HexNAc + 6Hex + 2Sialic acid + Fucose]	2
31	1356.4852	1356.4843	0.6634	[SINTT + 5HexNAc + 4Hex + Sialic acid + 1SO3 + Fucose]	2#
32	1271.4314	1271.4283	2.4382	[SINTT + 6HexNAc + 3Hex + 2SO3 + Fucose]	2
33	1398.0295	1398.0243	3.7195	[INTT + 4HexNAc + 5Hex + 2Sialic acid + Fucose]	2#
34	1227.9126	1227.9123	0.2443	[INTT + 6HexNAc + 3Hex + 2SO3 + Fucose]	2
35	1312.9709	1312.9683	1.9802	[INTT + 5HexNAc + 4Hex + Sialic acid + 1SO3 + Fucose]	2#
36	1499.5650	1499.5640	0.6668	[INTT + 5HexNAc + 5Hex + 2Sialic acid + Fucose]	2#
37	1510.5572	1510.5553	1.2578	[INTT + 5HexNAc + 5Hex + 2Sialic acid + Fucose +Na]	2
38	1580.5928	1580.5904	1.5184	[INTT + 5HexNAc + 6Hex + 2Sialic acid + Fucose]	2

#Apart from the doubly charged ions, triply charged ions also exist for the same composition. The same number is used to indicate both on the spectrum.

3.8 References

1. Bousfield, G.R.; Butnev, V.Y.; Gotschall, R.R.; Baker, V. L.; Moore, W.T. *Mol. Cell. Endocrinol.* **1996**,*125*, 3-19.
2. Green, E. D.; Baenziger, J. U. *J. Biol. Chem.* **1988**, *263*, 36-44.
3. Fan, Q.R.; Hendrickson, W. A. *Nature* **2005**, *433*, 269-277.
4. Shoham, Z.; Insler, V. *Fertil. Steril.* **1996**, *66*,187-201.
5. Baenziger, J.U.; Green, E. D. *Biochim. Biophys. Acta* **1988**, *947*, 287-306.
6. Thotakura, N.R.; Blithe, D.L. *Glycobiology* **1995**, *5*, 3-10.
7. Sairam, M. R.; Bhargavi, G. N. *Science,***1985**, *229*, 65-67.
8. Liu, C.; Dias, J. A. *Arch. Biochem. Biophys.***1996**, *329*, 127-135.
9. Bishop, L. A.; Nguyen, T. V.; Schofield, P.R. *Endocrinology* **1995**,*136*, 2635-2640.
10. Ulloa-Aguirre, A.; Timossi, C. *Hum. Reprod. Update* 1998, *4*, 260-283.
11. Butney, V. Y.; Gotschall, R. R.; Butnev, V.Y.; Baker, V. L.; Moore, W. T.; Bousfield, G. R. *Biol. Reprod.* **1998**, *58*, 458-469.
12. Sairam, M.R. *FASEB J.* **1989**, *3,* 1915-1926.
13. Flack, M.R.; Froehlich, J.; Bennet, A, P.; Anasti, J.; Nisula, B.C. *J. Biol. Chem.* **1994**, *269*, 14015-14020.
14. Valove, F. M.; Finch, C.; Anasti, J. N.; Froehlich, J.; Flack, M. R. *Endocrinology* **1994**, *135*, 2657-2661.
15. Bishop, L.A.; Robertson, D.M.; Cahir, N.; Scholfield, P.R. *Mol. Endocrinol.* **1994**, *8,* 722-731.
16. Saneyoshi, T.; Min, K.; Ma, X.J.; Nambo, Y.; Hiyama, T.; Tanaka, S.; Shiota, K. *Biol. Reprod.* **2001**, *65*, 1686-1690.
17. Green, E.D.; Baenziger, J. U. *J. Biol. Chem.* **1988**, *263*, 25-35.
18. Renwick, A.G.C.; Mizuochi, T.; Kochibe, N.; Kobata, A. *J. Biochem.* 1987,*101*, 1209-1221.
19. Hard, K.; Mekking, A.; Damm, J. B. L.; Kamerling, J.P.; De Boer, W.; Wijnands, R. A.; Vliegenthart, J. F.G. 1990, *Eur. J. Biochem. 193*, 263-271.
20. Bousfield, G. R.; Butnev, V. Y.; Butnev, V. Y.; Nguyen, V. T.; Gray, C.M.; Dias, J.A.; MacColl, R.; Eisele, L.; Harvey, D.J. *Biochemistry* 2004, *43*, 10817-10833.
21. Gervais, A.; Hammel, Y.; Pelloux, S.; Lepage, P.; Baer, G.; Carte, N.; Sorokine, O.; Strub, J.; Koerner, R., Leize, E.; Dorsselaer, A. V. *Glycobiology* **2003**, *13*, 179-189.
22. An, H.J.; Peavy, T.R.; Hedrick, J.L.; Lebrilla, C.B. *Anal. Chem*. **2003,** *75*, 5628-5637.
23. Jiang, H., Irungu, J., and Desaire, H. *J. Am. Soc. Mass Spectrom.* 2005, *16*, 340-348.

24. Jiang, H.; Desaire, H.; Butnev, V.Y.; Bousfield, G.R *J. Am. Soc. Mass Spectrom*. **2004,** *15*, 750-758.
25. Conboy, J. J., and Henion, J. D. *J. Am. Soc. Mass Spectrom.* 1992, *3*, 804-814.
26. Bousfield, G. R.,and Ward, D. N. *J. Biol.Chem.* **1984**, *259*, 1911-1921.
27. Ward, D. N.; Glenn, S. D.; Nahm, H. S.; Wen, T. *Int. J. Pept. Protein Res.* 1986, *27*, 70-78.
28. Gordon, W. L.; Bousfield, G. R.; Ward, D. N. *J. Endocrinol. Invest.*1989, *12*, 383-392.
29. Gonadotropins: chemistry and biosynthesis, in *Knobil and Neill: Physiology of Reproduction* (Neill, J. D., Ed.) Bousfield, G. R.; Jia, L.;Ward, D. N. pp 1581-1634, Elsevier, San Diego, 2006.
30. Green, E.D.; Boime, I.; Baenziger, J.U. *Mol. Cell. Biochem.*1986, *72*, 81-100.
31. Weisshaar, G.; Hiyama, J.; Renwick, A.G.C. *Eur. J. Biochem*, **1990**, *192*, 741-751.
32. Weisshaar, G.; Hiyama, J.; Renwick, A. G. C.; Nimtz, M. *Eur. J. Biochem.* **1991**, *195*, 257-268.
33. Hiyama, J.; Weisshaar, G.; Renwick, A. G. C. *Glycobiology* 1992, *2*, 401-409.
34. Weisshaar, G.; Hiyama, J., Renwick, A. G. C. *Glycobiology* **1991**,*1*, 393-404.
35. Witt, M.; Klessen, C. *Folia Histochem. Cytobiol.* **1987**, *25*, 115-118.
36. Baenziger, J.U. *Biochem. Soc. Trans.* **2003**, *31*,326-330.
37. Schaaf, L.; Leiprecht, A.; Saji, M.; Huebner, U.; Usadel, K. H.; Kohn, L. D. *Mol. Cell. Endocrinol.* **1997**,*132*, 185-194.
38. Nguyen, V. T.; Singh, V.; Butnev, V. Y.; Gray, C. M.; Westfall, S.; Davis, J. S.; Dias, J. A.; Bousfield, G. R. *Mol. Cell. Endocrinol.* **2003**,*199*, 73-86.
39. Bousfield, G.R.; Butnev, V.Y.; Walton, W.J.; Nguyen, V.T.; Huneidi, J.; Singh, V.; Kolli, K.; Harvey, D.J.; Rance, N.E. *Mol. Cell. Endocrinol*, in press **2006**

CHAPTER 4

Application of the STEP Method to the Triple Quadrupole Mass Spectrometer

(Submitted to *Rapid Commun. Mass Spectrom. for publication)*

4.1. Introduction

The study of the nature of characteristic fragmentations of gaseous ions by mass spectrometry is important because understanding basic fragmentation information is helpful in deducing precursor ion structures for unknown compounds.[1-3] Fragmentation information can be used to establish genealogical relationships, which are important in discerning information about the connectivity of groups that fragment during MS/MS experiments.[4] Since first generation product ions (primary product ions) directly originate from the precursor ion while the second generation of ions (secondary product ions) originate from other product ions in the MS/MS experiment, genealogy information can be useful in piecing together a precursor ion's structure, based on the product ions detected in the MS/MS data.[5 6, 7]

The information about the connectivity of ions from MS/MS data can be determined in a number of ways. For example, previous researchers have shown that some information can be inferred by comparing results in MS^3 experiments to MS^2 experiments on selected ions.[5] The product ions that appear on both the MS^2 and MS^3 data were assigned as a result of secondary fragmentation. However, since product ions can be formed from more than

one source, this method has limited applicability in obtaining genealogy information with certainty. Additionally, MS^n capability is only present in trapping instruments like the quadrupole ion trap MS and is not present in instruments like the triple quadrupole mass spectrometer.

Double resonance experiments in conjunction with SWIFT (stored wave form inverse Fourier transform) techniques are another avenue for obtaining genealogy information in quadruple ion traps and Fourier transform ion cyclotron resonance mass spectrometers.[2, 3, 8 9] Application of the SWIFT technique allows product ions to be selectively ejected as they are formed.[3] The disappearance of fragment ions after the ejection of an ion can be used to confirm the origin of the ions that was ejected. SWIFT technique is well suited for instruments that have the capability of ion storage and trapping.[10] Since the triple quadruple MS is not a trapping instrument, the applicability of this technique to gain genealogy information from the triple quadruple MS is not possible.

Energy resolved mass spectrometry is another method that can be used to distinguish primary ions from secondary ions in the mass spectrometer, [7, 8, 11-14] and this method can be applied to the triple quadrupole instruments.[7] This method involves graphing the product ion intensity with varying collision energy to yield breakdown curves.[15] When IRMPD (Infrared multiphoton dissociation) is used for the activation of ions, exposure time is plotted against product ion abundances, to yield breakdown

curves useful in diagnosing ion origins.[3, 13] Identification of ions as primary and secondary is based on the shapes of the curves that are obtained with increasing energy. For instance as the energy increases the intensity of primary product ions decrease and the intensity of secondary product ions start to increase. The breakdown curves help to obtain information about the major reaction pathways, but detailed information about the minor ions seems to be unobtainable.[14] Breakdown curves can be constructed in trapping instruments as well as in triple quadrupole mass spectrometers, but this technique can be time consuming because numerous collision energies or exposure times must be utilized to collect the data necessary to identify product ion origin.

The STEP (statistical test of equivalent pathways) analysis, which was originally developed on the ion trap mass spectrometer, is another tool that differentiates primary product ions from secondary product ions.[4,16] Applicability of this method was previously tested with peptides and carbohydrates.[16] This technique utilizes ion ratios obtained from two MS/MS experiments in order to obtain genealogy information. Application of the STEP method to the triple quadrupole MS can be very useful because of the limited availability of techniques that could differentiate primary ions from secondary ions in this instrument. Since the triple quadrupole mass spectrometer is widely used, application of this method to this instrument is expected to have broad applicability among the mass spectrometry community.

4.2. Methods

4.2.1. Sample preparation

Peptides (VEPIPY and YGGFL) were purchased from Sigma-Aldrich (St. Louis, MO). Samples were dissolved in 50:50, MeOH, H_2O, 0.5% HOAc to a final concentration of 0.01mM.

The carbohydrate M3N2 was purchased from Prozyme (San Leandro, CA). Derivatized maltopentaose was synthesized in the lab using a method described previously.[17] 100 µL of H_2O was added to each sample and samples were diluted to 0.01mM using 50:50, MeOH, H_2O, 0.5% HOAc.

4.2.2. Data acquisition on a triple quadrupole mass spectrometer

The MS/MS data for the STEP analysis was acquired on a Micromass Quattro Ultima triple quadrupole mass spectrometer equipped with an electrospray ionization source. The samples were directly infused using a syringe pump at a flow rate of 10 µL/min. Data was collected at four different collision gas pressures: 0.2 mTorr, 0.4 mTorr, 0.5 mTorr and 1.4 mTorr. Two MS/MS spectra were acquired at high and low dissociation energy conditions at each of these collision gas pressures. For each experiment, data was acquired for 2 min with a scan time of 10 s and interscan time of 0.1 s. The low energy spectrum was obtained when any product ion in spectrum was at 20% abundance. The high energy spectrum was obtained at an energy that

was 14eV higher than the energy needed for the low energy spectrum. Each analysis was performed in triplicate in positive ion mode.

4.2.3. Calculating the STEP Ratios

STEP calculations were performed as previously described.[16] Briefly, the STEP ratio is calculated using ion ratios in MS/MS data. For each ion in the MS/MS spectrum the ratio is calculated by dividing the product ion area by the total product ion area.

$$\text{Ratio of product ion area} = \frac{\text{Product ion area}}{\text{Total product ion area}}$$

This ratio is calculated for all ions in the high and low energy spectra and a STEP ratio is assigned for each ion.

$$\text{STEP ratio} = \frac{\text{Ratio of product ion area in high energy spectrum}}{\text{Ratio of product ion area in low energy spectrum}}$$

The STEP ratio can be used to distinguish primary ions from secondary ions. Primary ions that have a STEP ratio ≤ 1 are then utilized as a control group for a Q test. All product ions with a STEP ratio greater than one are subjected to a Q test of statistical significance to determine whether they are significantly different from the primary ions. A confidence interval of 90% is used to classify ions that are significantly different from primary ions. A STEP ratio is calculated for all ions that have a relative abundance above 3%.

4.3. Results and Discussion

The STEP method is very useful in identifying primary product ions that originate directly from the precursor ion and secondary product ions that originate from primary product ions. Classifying product ions as primary or secondary is very useful when characterizing the structures of carbohydrates and peptides, as previously demonstrated.[16]

The theoretical basis of the STEP analysis can be explained in terms of breakdown curves.[16] Figure 4.1 a and b are representative examples of breakdown curves obtained for a peptide, VEPIPY in the triple quadrupole.

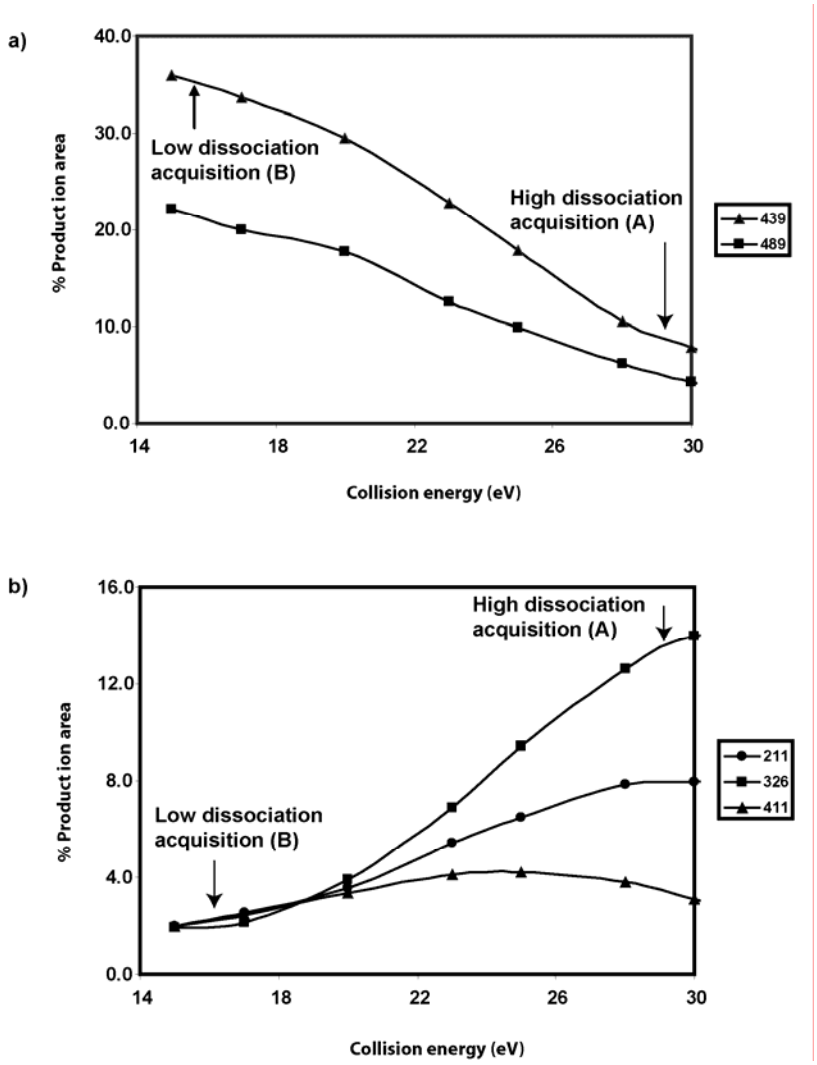


Figure 4.1: Breakdown curves obtained for a) primary ions b) secondary ions of VEPIPY.

The ions that are used to construct these breakdown curves were previously validated by double resonance experiments[16] to be primary or secondary ions. As can be seen from the figure, as collision energy increases, the percent product ion area for primary product ions decreases whereas secondary product ion areas increase. There is a marked difference

between the behavior of primary and secondary ions with increasing collision energy. As a result one can expect that primary product ions to be differentiated from secondary ions using a ratio obtained when MS/MS data is collected under high and low dissociation conditions. This ratio also known as the STEP ratio, has been well established in previous publications.[4, 16] The two data collection points for calculating this ratio is indicated by A and B on the graphs, where A represents the high energy product ion area and B represents the low energy product ion area. Primary product ions have low STEP ratios, because the STEP ratio is obtained by dividing the high dissociation product ion area (A) by the low dissociation product ion area (B).

4.4. STEP analysis on triple quadrupole mass spectrometer

In a first attempt at adapting the STEP method to the triple quadrupole mass spectrometer, the criteria that distinguished primary and secondary ions on the quadrupole ion trap were directly implemented in the analysis. In that case, the high energy CID spectrum was obtained when the precursor ion was depleted to 0 – 2% relative abundance, as described previously.[4, 16] Additionally, the low energy spectrum was obtained by choosing the most abundant product ion peak from the high energy spectrum and depleting it to only 30% abundance, relative to the precursor ion. In these experiments, product ions with STEP ratios less than one were assigned as primary product ions. These ions were used as a basis set of data to perform a Q-test.

Product ions whose STEP ratios were significantly different from the STEP ratios in the basis set were defined as secondary product ions. This direct application of the method to the triple quadrupole mass analyzer was unsuccessful.

Since the fragmentation is effected by the collision gas pressure,[10, 18 19 20] the analysis was also performed at four different gas pressures (0.2 mTorr, 0.4 mTorr, 0.5 mTorr and 1.4 mTorr), to determine whether the pressure had an effect on correctly identifying primary ions from secondary ions. At each pressure four compounds (two carbohydrates and two peptides) were subjected to the STEP analysis. The results indicated that the majority of primary and secondary ions were incorrectly identified at all pressures. A representative example of the results obtained for the peptide, VEPIPY at 0.4mTorr is shown in Table 4.1 a.

Due to energetic differences between the ion trap and the triple quadrupole mass spectrometer[10, 21] the experimental conditions previously defined on an ion trap produced different results, when the method was transferred directly to the triple quadrupole mass analyzer. One example of the problem associated with directly applying the method, without modification, is the genealogy assignment of *m/z* 411 from the compound, VEPIPY. The genealogy assignments (1^o and 2^o) for this compound (in Table 4.1a) were validated based on previously acquired double- resonance data.[16] In the validation, the ion, *m/z* 411, (a_4) is determined to be a secondary ion

that is ~100% from the b_4 ion. This ion is expected to have a higher STEP ratio because it is a secondary ion. The average STEP ratio for this ion is 0.83 ± 0.05, which would classify this product ion as primary. This demonstrates that the experimental parameters for the STEP analysis utilizing the triple quadrupole mass spectrometer required adjustment.

Table 4.1: STEP analysis versus validation data for the peptide VEPIPY acquired using old and new experimental conditions

Peptides	STEP Method*					Validation**	
a) old method	*m/z*	trial 1	trial 2	trial 3	1° or 2°	1° or 2°	Method**
	489(y_4)	0.1	0.11	0.1	1°	1°	ST
	439 (b_4)	0.11	0.1	0.13	1°	1°	ST
Val-Glu-Pro-Ile-Pro-Tyr	411(a_4)	0.85	0.8	0.84	1°	2°	ref(16)/DR
	279(y_2)	1.87	1.98	1.97	1°	1° / 2°	ref(16)/DR
		3.29	**3.07**	**3.24**			
	211(PI/IP)	4.46	4.07	3.92	2°	2°	ST
	326(b_3)	6.14	5.84	5.72	2°	2°	ref(16)/DR
	229(b_2)	21.92	21.92	18.15	2°	?	?
b) new method	439(b_4)	0.25	0.25	0.24	1°	1°	ST
	489(y_4)	0.23	0.26	0.27	1°	1°	ST
		0.57	**0.42**	**0.74**			
Val-Glu-Pro-Ile-Pro-Tyr	411(a_4)	1.47	1.41	1.66	2°	2°	ref(16)/DR
	279(y_2)	2.07	1.89	1.88	2°	1° / 2	ref(16)/DR
	211(PI/IP)	3.23	3.25	3.24	2°	2°	ST
	326(b_3)	7.59	6.95	6.84	2°	2°	ref(16)/DR
	229(b_2)	7.28	12.55	9.01	2°	?	?

For the STEP analysis, primary and secondary ions are differentiated by using a Q-test (at the 90%CI). The cut off between primary and secondary is represented by the horizontal line. The STEP values that defined the boundary are incorporated.
**The primary and secondary ions were validated by double resonance experiments (DR), literature (ref) and reasonable arguments based on the structure (ST)

Table 4.1 b represents the STEP values obtained after some modification to the method. These values have been obtained when the collision gas pressure was 0.4mTorr. The corresponding MS/MS spectrum for the peptide analyte, VEPIPY, acquired at high dissociation conditions is shown in Figure 4.2.

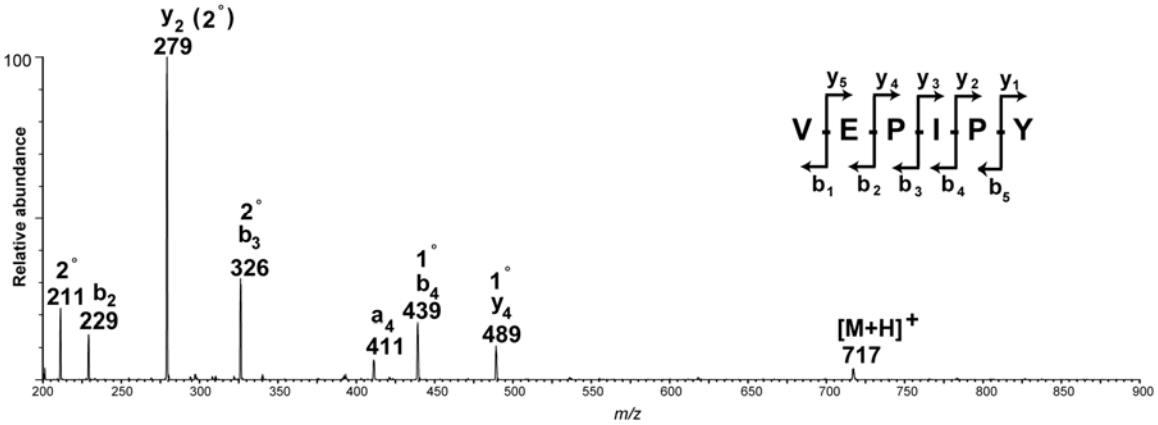


Figure 4.2: MS/MS data for the peptide, VEPIPY. All peaks are labeled as primary (1^0) or secondary (2^0) based on the results that were obtained from the STEP analysis.

After optimizing the conditions of the method, it was determined that differentiation between primary and secondary ions is optimal when the low energy spectrum is acquired using conditions that keep all product ions below ~20% abundance. The high energy spectrum was obtained at an energy that was 14eV higher than the energy required for the low energy spectrum. With the new experimental conditions, *m/z* 411, which was predicted to be secondary from double resonance experiments, has a relatively high STEP ratio.

In addition to *m/z* 411, there were four ions (*m/z* 326, *m/z* 279, *m/z* 211, *m/z* 229) that were categorized as secondary in this analysis. The ion *m/z* 326 was confirmed by double resonance experiments to be a secondary ion in our previous publication.[16] The ion *m/z* 279, which was previously confirmed to be a result of both primary and secondary fragmentation, always occurred as a secondary fragment in the triple quadrupole. Ions *m/z* 211 and *m/z* 229 were two new ions that were not observed in the ion trap data that was acquired previously. The former ion corresponds to a mass of an internal ion, which can have the sequence IP or PI. Therefore two bonds need to be cleaved in order to form this ion, which must be a secondary ion. The ion, *m/z* 229 corresponds to a b_2 cleavage in the peptide, and there is no validated data available on this ion. Relatively high STEP ratios were obtained for this ion, which lead us to believe that this is a secondary fragment originating from *m/z* 439 (b_4) or a tertiary fragment from *m/z* 326 (b_3). The two ions (*m/z* 439 (b_4) and *m/z* 489 (y_4)) previously identified as primary product ions had very low STEP values in this analysis and were categorized as primary.

In general, primary ions in the triple quadrupole have lower STEP ratios than the STEP values that were obtained previously in the ion trap. For example, *m/z* 439 and *m/z* 489 had an average STEP ratio of 0.80 and 0.81 when the experiments were previously performed on the ion trap mass spectrometer; whereas their ratios were 0.24 and 0.25 when the experiments were performed on the triple quadrupole mass spectrometer. This

phenomenon can be related to the high internal energy of activated ions and the continued activation of ions in the triple quadupole mass spectrometer.[22] The high internal energy is due to the higher efficiency in energy deposition in the triple quadrupole where Ar is the collision gas, rather than in the ion trap, which typically employs He.[21, 23] Therefore the precursor ions have higher internal energy to impart to the product ions. The product ions not only receive more energy from the precursor ions, they also can undergo collisions with the neutral gas and become further activated to produce more fragments. Therefore in the triple quadrupole there is a very high chance of depleting the primary product ions after they are formed. By contrast, in the ion trap, only the precursor ions are activated, so the secondary ions are only formed as a way of releasing excess energy from the primary product ions. Product ions, themselves, are not activated. These differences in activation in the two instruments could certainly affect the STEP ratios. Since the product ion area in the high energy spectrum is in the numerator of the STEP equation, primary product ions, which have relatively lower abundance in the triple quadrupole experiments, would be expected to have smaller overall STEP values.

In addition to optimizing the collision energy applied, the effect of collision gas pressure was also investigated. All of the data described above were collected using a gas pressure of 0.4 mTorr Ar. When the modifications to the STEP analysis protocol were applied using other collision gas

pressures, including 0.2 mTorr, 0.5 mTorr, and 1.4 mTorr, fewer correct assignments for the product ions were obtained. Therefore, we verified that 0.4mTorr was the best collision gas pressure to use, based on the fact that it gave the maximum number of correctly identified primary and secondary ions for all four compounds analyzed in this study. The STEP values for the rest of the compounds described below were obtained when the collision gas pressure was maintained at 0.4 mTorr.

The trends for STEP ratios for the second peptide in this study, YGGFL were similar to the trends observed for the peptide VEPIPY, described above. The spectrum and STEP values corresponding to the peptide YGGFL can be found in Figure 4.3 and Table 4.2 respectively.

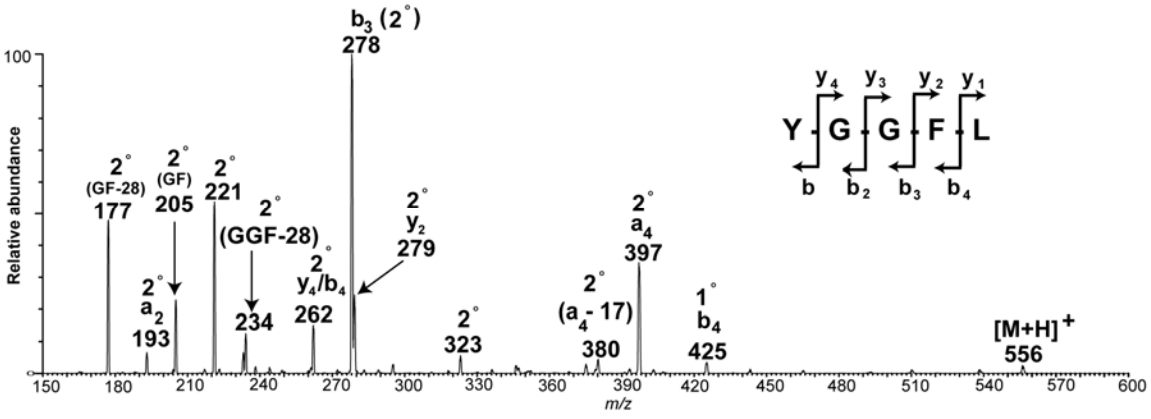


Figure 4.3 : MS/MS data for the peptide, YGGFL. All peaks are labeled as primary (1^0) or secondary (2^0) based on the results that were obtained from the STEP analysis

Table 4.2: STEP analysis versus validation data for three compounds

Peptides		STEP Method*				Validation**	
	m/z	trial 1	trial 2	trial 3	1° or 2°	1° or 2°	Method**
	425(b_4)	0.04	0.04	0.03	1°	1°	ref(16)/ST
	397(a_4)	0.43	0.44	0.43	1°	2°	ref (2,16)
	336(y_3)	0.06	0.05	0.05	1°	1°	ref(16)/ST
Tyr-Gly-Gly-Phe-Leu	538(-H_20)	0.06	0.06	0.06	1°	1°	ref(16)/ST
	279(y_2)	0.76	0.73	0.73	1°	1°	DR/ref(16)
		1.68	**1.61**	**1.62**			
	205(GF)	1.88	1.80	1.91	2°	2°	ST
	380(a_4-17)	2.09	2.33	2.19	2°	2°	ref (5)
	278(b_3)	4.40	4.58	4.30	2°	2°	ref (2,16)
	262(GGF)	2.58	2.63	2.57	2°	2°	ST
	323	10.91	11.42	10.11	2°	2°	ref (5)
	221(b_2)	14.78	15.06	14.73	2°	?	?
	234(GGF-28)	26.34	25.16	25.45	2°	2°	ST
	177(GF-28)	27.57	28.62	26.47	2°	2°	ST
	193(a_2)	36.50	32.81	24.59	2°	2°	ref (24)
Carbohydrates							
	749(y_3)	0.44	0.47	0.31	1°	1°	ref(16)/ST
M3N2	893(-H_2O)	0.61	0.41	0.40	1°	1°	ref(16)/ST
	690(b_3)	1.61	1.22	1.22	1°	1°	ref(16)/ST
		3.32	**1.43**	**1.83**			
	587($y_{3\alpha}$/$y_{3\beta}$)	3.46	3.34	3.11	2°	2°	ref(16)/ST
	731(y_3 -H_20)	5.89	2.82	3.38	2°	2°	ref(16)/ST
	407	6.87	5.70	3.91	2°	?	?
	528(b_3/$y_{3\alpha}$)	7.26	3.79	4.80	2°	2°	ref(16)/ST
	425(many poss)	8.19	5.86	7.00	2°	2°	ref(16)/ST
	325(many poss)	8.02	12.80	8.61	2°	2°	ref(16)/ST
	366(b_3/$y_{3\alpha}$/y_3)	16.86	19.39	9.75	2°	2°	ref(16)/ST
	886(y_4)	0.31	0.40	0.21	1°	1°	ref(16)/ST
Derivatized maltopentaose	724(y_3)	0.39	0.49	0.31	1°	1°	ref(16)/DR
	562(y_2)	0.49	0.54	0.44	1°	1°	ref(16)/DR
	975(-$C_4H_{11}N$)	0.48	0.51	0.47	1°	1°	ref(16)/ST
DEAEAB	400(y_1)	0.52	0.74	0.54	1°	1°	ref(16)/DR
		0.79					
	813(y_4-$C_4H_{11}N$)	1.22	1.00	0.68	1° & 2°	2°	ref(16)/DR/ST
			1.60	**1.16**			
	489(y_2-$C_4H_{11}N$)	3.09	3.52	2.50	2°	2°	ref(16)/DR/ST
	327(y_1-$C_4H_{11}N$)	2.65	3.31	3.73	2°	2°	ref(16)/DR/ST
	608(y_3-$C_6H_{16}N_2$)	1.53	1.78	1.62	2°	?	?
DEAEAB	446(y_2-$C_6H_{16}N_2$)	1.96	2.25	2.00	2°	?	?
	302	3.28	5.29	4.53	2°	?	?
	428	2.47	2.11	2.85	2°	?	?

*For the STEP analysis, primary and secondary ions are differentiated by using a Q-test (at the 90%CI). The cut off between primary and secondary is represented by the horizontal line. The STEP values that defined the boundary are incorporated.
**The primary and secondary ions were validated by double resonance experiments (DR), literature (ref) and reasonable arguments based on the structure (ST)

There were five ions, *m/z* 425 (b_4), *m/z* 538 (loss of water), *m/z* 336 (y_3), *m/z* 279 (y_2) *m/z* 397 (a_4), which were identified as primary ions, having STEP ratios less than 1. Except for *m/z* 397, the rest of the ions have been previously confirmed to result from primary fragmentation.[16] The ion, *m/z* 397, which had previously been classified as a secondary ion,[16] had quite low STEP values in this analysis. This ion had even lower STEP values when the experiments were performed using the conditions that were previously developed on the ion trap MS (data not shown). Therefore even with modifications to the analysis, as described above, this ion could not be correctly identified. However, rest of the ions were correctly identified. In addition to *m/z* 397 two other ions *m/z* 510 (loss of H_2O and CO) and *m/z* 278 (b_3) were identified from previous experiments as secondary ions. The ion at *m/z* 510 was not included in the analysis because its abundance was less than 3% in the triple quadrupole MS/MS data. The STEP analysis correctly identified *m/z* 278 to be a secondary ion.

In addition to these ions, there were some newly identified ions that also originate from secondary fragmentation. For example, *m/z* 177, *m/z* 205, *m/z* 234 and *m/z* 262 match the internal fragments which correspond to GF-28 Da, GF, GGF-28 Da, and GGF. Since two bonds must break for these ions to form, they must result from secondary fragmentation, and the STEP values confirm this scenario.

Finally, the ions *m/z* 193, 380, and 323 also appear in this analysis, and their origins can also be validated. The ion at *m/z* 193 (a_2) most likely comes from the b_2 ion (*m/z* 221) after the loss of CO, since this is common fragmentation pathway that forms 'a' ions.[24] Vachet *et al.* had previously found *m/z* 380 (a_4 -17) to be a product ion of *m/z* 397 when MS^3 experiments were performed on *m/z* 397, revealing that it is a secondary ion.[5] It has also been confirmed from double resonance experiments that *m/z* 323 is formed from both *m/z* 380 and from *m/z* 397.[5] Therefore *m/z* 323 has to be a result of secondary fragmentation and the STEP values are in agreement with this assignment. In general, the STEP analysis was very effective at identifying whether the ions originated from primary or secondary fragmentation.

In addition to the peptides, the STEP analysis was performed on two carbohydrates. The high energy MS/MS data corresponding to the first carbohydrate under study, M3N2 can be found in Figure 4.4.

As can be seen from Table 4.2, all ion genealogies except for one were correctly identified by the STEP analysis. There were three ions that were identified from M3N2 from previous studies as primary fragmentation products.[16] All three ions (*m/z* 893 (-H_2O), *m/z* 749 (y_3) and *m/z* 690 (b_3)) were identified as primary ions in this analysis as well. There were five ions that were identified as secondary from previous experiments. These ions (*m/z* 587, *m/z* 528, *m/z* 425, *m/z* 731 and *m/z* 366) were also identified as secondary in this study. There were two ions that were not detected

previously, *m/z* 325 and *m/z* 407. The former ion corresponds to the disaccharide ion, $[Hex\text{-}Hex]H^+$, and the only way to produce this ion from the precursor is by breaking at least two bonds. Therefore this ion has to result from a secondary fragmentation. The latter ion corresponds to an ion containing two HexNAc's, and this ion could have originated by cleaving one bond (z_2) or as a secondary fragment from another ion, like *m/z* 749. The STEP analysis identified this ion as a secondary fragment, but that assignment cannot be verified at this time.

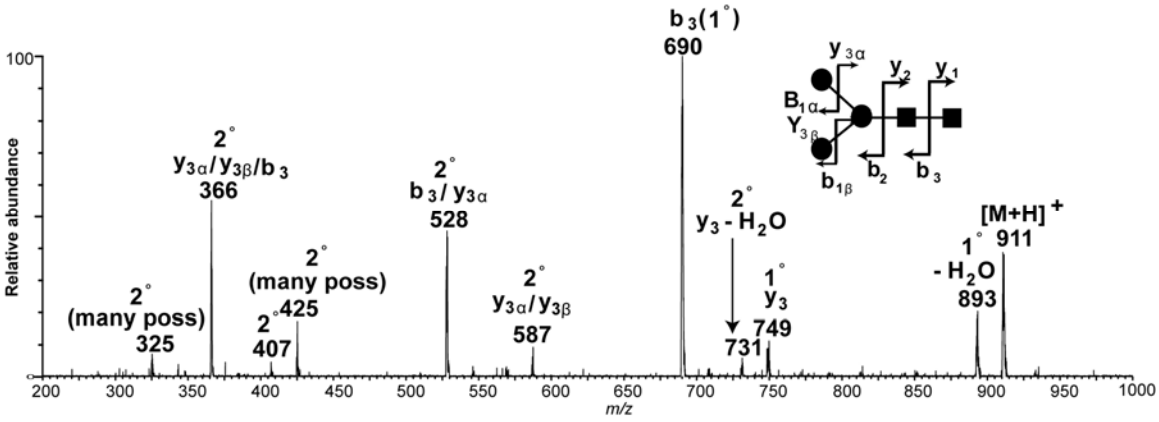


Figure 4.4: MS/MS data for the carbohydrate M3N2. All peaks are labeled as primary (1^0) or secondary (2^0) based on the results that were obtained from the STEP analysis

The MS/MS data and the STEP data for the other carbohydrate, derivatized maltopentaose, can be found in Figure 4.5 and Table 4.2 respectively. There were five ions (*m/z* 975, *m/z* 562, *m/z* 724, *m/z* 400 and *m/z* 886) that were confirmed to be from primary fragmentation from previous studies.[16] These ions also had low STEP ratios when the analysis was

performed in the triple quadrupole, which indicates that they are primary ions. In addition to correctly confirming the primary ions, the STEP analysis was effective at identifying secondary ions. Three secondary ions, (*m/z* 327, *m/z* 489, *m/z* 651) that were previously classified as secondary, were identified correctly by the STEP analysis conducted on the triple quadrulope MS, each of the three times the analysis was conducted. The final secondary ion, m/z 813 was also correctly identified two times, during the three replicate analyses.

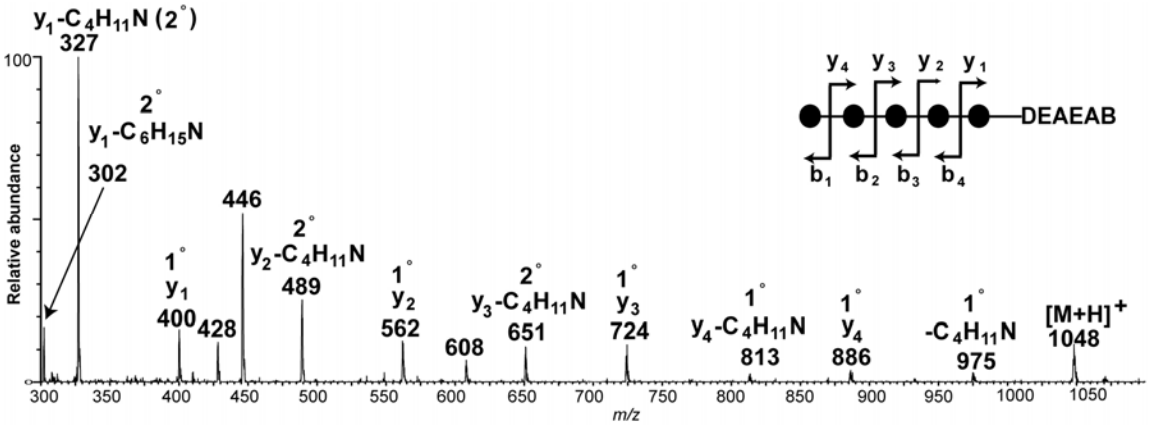


Figure 4.5: MS/MS data for the carbohydrate derivatized maltopentaose M3N2. All peaks are labeled as primary (1^0) or secondary (2^0) based on the results that were obtained from the STEP analysis

In summary, the STEP analysis method was validated on the triple quadrupole mass spectrometer using four compounds whose fragmentation properties have been well characterized previously. These four compounds produce a total of 108 product ions of known (primary or secondary) origin

when the analysis is conducted in triplicate and these ions can be used to test the effectiveness of the STEP analysis on the triple quadrupole mass spectrometer. Of these 108 ions, 104 of them (96%) were correctly identified as primary or secondary, indicating that this analysis is effective for carbohydrates and peptides undergoing CID on a triple quadrupole mass spectrometer.

4.5. Conclusion

This study demonstrated that the STEP method can be transferred from the quadrupole ion trap, to the triple quadrupole mass spectrometer, to differentiate primary and secondary ions in the triple quadrupole MS. The method transfer required overcoming some obstacles, due to the fundamental differences in the ion activation conditions between these two instruments. Initially, low STEP ratios were obtained for secondary ions, when the same experimental condition used in the ion trap were used to perform the experiments in the triple quadrupole MS. As a result, the secondary ions were not identified correctly. Therefore new ion activation conditions were adapted to accommodate for the fragmentation differences between the two instruments. These modifications afforded the successful implementation of the method on the triple quadrupole MS.

4.6. References

(1) Payne, A.H.; Chelf, J.H.; Glish G.L. *Analyst* **2000**, *125,* 635-640
(2) Vachet, R.W.; Ray, K.L.; Glish, G.L. *J. Am. Soc. Mass Spectrom.* **1998,** *9,* 341-344.
(3) Colorado, A.; Shen, J.X.; Vartanian, V.H.; Brodbelt, J. *Anal Chem.* **1996,** *68,* 4033-4043.
(4) Bandu, M.L.; Desaire, H. *Analyst* **2006**, *131***,** 268-274.
(5) Vachet, R.W.; Bishop, B.M.; Erickson, B.W.; Glish, G.L. *J. Am. Chem. Soc.* **1997,** *119***,** 5481-5488.
(6) Glish, G. *J. Am. Soc. Mass Spectrom.* **1991**, *2***,** 349.
(7) Wensing, M.W.; Snyder, A.P.; Harden, C.S. *J Mass Spectrom*. **1995**, *30,* 1539-1545.
(8) Crowe, M.C.; Brodbelt, J.S.; Goolsby, B.J.; Hergenrother, P. *J. Am. Soc. Mass Spectrom.* **2002**, *13***,** 630-649.
(9) Vachet, R.W.; McElvany, S.W. *J. Am. Soc. Mass Spectrom.* **1999**, *10,* 355-359.
(10) Sleno, L.; Volmer, D.A. *J Mass Spectrom*. **2004,** *39***,** 1091-1112.
(11) Wensing, M.W.; Snyder, A.P.; Harden, C.S. *Rapid Commun. Mass Spectrom.* **1996**, *10***,** 1259-1265.
(12) Speir, J.P.; Amster, I.J. *J. Am. Soc. Mass Spectrom.* **1995**, *6***,** 1069-1078.
(13) Goolsby, B.J.; Brodbelt, J.S. *J Mass Spectrom*. **1998***, 33,* 705-712.
(14) Harrison A.G. *Rapid Commun. Mass Spectrom.* **1999**, *13***,** 1663-1670.
(15) Fetterolf, D.D.; Yost, R.A. *Int. J. Mass Spectrom. Ion Phys.* **1982**, *44,* 37-50.
(16) Bandu, M.L.; Wilson, J.; Vachet, R.W.; Dalpathado, D.S.; Desaire, H. *Anal Chem.* **2005**, *77***,** 5886-5893.
(17) Dalpathado, D.S.; Jiang, H.; Kater, M.A.; Desaire, H. *Anal Bioanal Chem*, **2005**, *381***,** 1130-1137.
(18) Charles, M.J.; McLuckey, S.A.; Glish, G.L. *J. Am. Soc. Mass Spectrom.* **1994**, 5,1031-1041.
(19) Xia, Y.; Liang, X.; McLuckey, S.A. *Anal Chem*, **2006**, *7*, 1218-1227.
(20) Charles, M.J.; Marbury, G.D. *Anal Chem*, **1991**, *63*, 713-721.
(21) Brodbelt ,J.S.; Kenttamaa, H.I.; Cooks, R.G. *Org. Mass Spectrom.* **1988,** *23,* 6-9.
(22) Shukla, A.K.; Futrell, J.H. *J Mass Spectrom*, **2000**, *35*, 1069-1090.
(23) Bursey, M.M.; Nystrom, J.A.; Hass, J.R. *Anal Chim Acta.* **1984,** *159***,** 265-274.
(24) Yalcin, T.; Csizmadia, I.G.; Peterson, M.R.; Harrison, A. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 233.

5.1. Conclusion and future work

Glycoproteins are an important class of bio-molecules that have a significant effect on the proper functioning of a protein. Specifically, the glycan structures on glycoproteins have been found to be responsible for this behavior. Hence the focus of the work presented herein has been to research methods that enhance the analysis of glycan structures on glycoproteins. All methods have been developed on a mass spectrometer, which is becoming a promising tool for bio-molecular analysis. Developing methods on this instrument is useful because it is a high throughput method capable of handling complex mixtures which is a significant improvement from other methods capable of doing the same analysis.

There were two methods that were developed to obtain more structural information from glycans. The first method was performed by improving a derivatization reaction for glycans with less toxic reagents. This protocol not only provided more structural information but also enhanced the ionization efficiency in the mass spectrometer, and the reaction product could also be detected by optical detector because of the presence of a chromophore. The product was purified using a zip-tip or HPLC and there was a ~1000 fold improvement in the limit of detection when the purification was performed using HPLC. As for future work, it could be beneficial to use the method for glycans that are released from a complex glycoprotein.

The second method of obtaining structural information from glycans involved using a statistical method (statistical test of equivalent pathways-STEP), which was originally developed on an ion trap mass spectrometer. This method was further extended to a triple quadrupole mass spectrometer, which is an instrument that is routinely being used. After performing modifications to the analysis, this method could differentiate primary product ions from secondary product ions, and this differentiation helps to decipher genealogy information. This method not only works for glycans but it also works for peptides as well. The STEP method can be further improved by trying to find conditions that could differentiate tertiary product ions from secondary product ions.

The experimental conditions used for the first and the second methods are suited for glycans that are cleaved from glycoproteins. This is an excellent method of getting structural information, if the focus is only on glycans and not on the attachments site on the proteins. However, glycan structure alone cannot provide answers to questions regarding the functional significance of glycans structure. This is because glycans mediate different functions at different glycosylation sites. Therefore it is important to characterize glycans in a site specific manner so that glycan structure can be correlated to the function.

This information could be obtained by characterizing glycopeptides. This analysis is a relatively new method for glycan analysis, and for the first

time it was used for characterizing glycopeptides on a reproductive hormone, follicle stimulating hormone (FSH), obtained from the pituitaries of two mammalian species, human and horse. This hormone is a heterodimer consisting of two sub units having two glycosylation sites in each subunit. In addition, we wanted to see the differences of glycan structures between two mammalian species having similar amino acid sequences in their protein sequences. There were many site specific differences that were observed between species as well as between different sites on the same species. Currently, FSH derived from recombinant technology are being used for infertility treatment. Studies done so far have compared the differences between the pituitary glycans and recombinant glycans after they have been cleaved from the glycoprotein. Since different glycosylation sites play different functional roles it would be important to compare the glycans between the two forms in a site specific manner. These studies will facilitate the development of a better treatment for infertility.