

Qualitative and quantitative analysis of proteolytically digested glycoproteins by mass spectrometry

By

Kathryn R. Rebecchi

Submitted to the graduate degree program in Chemistry and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

---

Chairperson: Dr. Heather Desaire

---

Dr. Susan Lunte

---

Dr. Robert Dunn

---

Dr. Jon Tunge

---

Dr. Roberto De Guzman

Date Defended: April 25, 2011

The Dissertation Committee for Kathryn R. Rebecchi  
certifies that this is the approved version of the following dissertation:

Qualitative and quantitative analysis of proteolytically digested glycoproteins by mass  
spectrometry

---

Chairperson: Dr. Heather Desaire

Date approved: April 25, 2011

## ABSTRACT

Glycoproteins are a very large and biologically relevant class of proteins that comprise more than 50 % of proteins in the human body. The glycosylation present on proteins, specifically *N*-linked glycosylation has been shown to be important for a variety of processes including protein folding, protein stability, and cell-cell interactions. Many glycoproteins are currently being considered as therapeutic drug targets. Glycosylation on proteins has also been shown to be altered with the onset of diseases, such as cancer, which has opened up the field of glycoproteomics, which aims to detect glycosylation changes for earlier detection of disease states. Mass spectrometry is a versatile technique that is frequently utilized for the analysis of glycoproteins, and it is particularly useful in the detection of glycosylation present on proteins. Most glycoproteins are prepared for mass spectrometric analysis by performing a protease digestion, followed by either a separation by HPLC or some other technique for enrichment of glycopeptides. In this work, the protease digestion procedure was optimized for maximized protein sequence coverage and detection of *N*-linked glycopeptides and other post-translational modifications. This method was applied to a recombinant glycoprotein that had never before been fully characterized by mass spectrometry and is a potential protein therapeutic as well as known to play a role in different types of cancer. Furthermore, a mass spectrometric relative quantitation method was developed by creating glycosylation profiles from glycopeptides detected at individual glycosylation sites on different glycoproteins. This method allowed for distinguishing between changes in protein concentration from changes in glycosylation. Lastly, glycoprotein structure and stability was probed by circular dichroism

spectroscopy before and after glycan removal on glycoproteins containing high mannose type glycans with the enzyme peptide-N-glycosidase F. Protease digestion and mass spectrometry was performed to ensure that the deglycosylation reaction went to completion.

## ACKNOWLEDGEMENTS

I would like to thank all my family and friends for your support over the years. I would not have been able to complete graduate school without you. Additionally, I would like to thank all the teachers and professors who taught me science. I want to especially thank Dr. Bellamy, my analytical professor at Northwest Missouri State University. Without Dr. Bellamy's direction and guidance, I may not have chosen analytical chemistry for a career path.

To everyone in the Desaire group, past and present, thank you for all your help and support. I have learned so much from all of you. Heather Desaire, I could not have asked for a better advisor. I could always count on you to answer my questions anytime your door was open and for always challenging me to do my best. Mary Bandu, thank you for teaching me how to use a mass spectrometer and always checking on my progress. Dilusha Dalpathado, you were always so patient with my questions. Janet Irungu and Ying Zhang, you both were so kind to help me out anytime I asked. Daniel Clark, you always kept me on my toes and I enjoy talking with you, whether discussing science or hunting or whatever else. Carrie Woodin, thank you for teaching me biology and being such a great friend. I hope we never lose contact. Eden Go, you are an amazing mentor. I learned more from you than anyone else in graduate school.

I am so blessed to have an amazing family support system. Mom, thank you for instilling in me the need to succeed and to keep me grounded in my thoughts and dreams. Jenny Saunders, I have always looked up to you and trust in your judgement wholeheartedly. Thank you for listening when I need you. David Smith, thank you for being my dad even though you did not have to and for always having faith in my

abilities. Even though you went to Heaven while I was still in high school, Grandma Cain, you helped me get to where I am now. I still think of you often and hope that I have made you proud. David Rebecchi, I am so thankful for your patience with me throughout graduate school. Margo, you always make me smile even through the midst of stress. I love you so much, my precious daughter!

Billie Sue Miller, thank you for being my best friend for longer than anyone else. Sometimes I do not know how you have put up with me so long. I know I can call you anytime day or night and you will be there. Merci Decker, thank you for all of our wonderful philosophical conversations and all the times we shared. It was you that taught me about graduate school and the importance of getting a PhD. Christina Munson, thank you for all your help and support during graduate school. All your advice about parenting and how to handle being a parent while going to school has been invaluable. Also, I greatly appreciate you sharing with me the new technologies that you came across in the chemical education field. Stacey Wiley, I absolutely could not have finished graduate school without you. I will never be able to fully repay you fully for taking such good care of my daughter.

Lastly, I want to thank everyone at my church for your support and prayers for my family and me. These last 3 years have been especially tough and I am not sure I would have finished school without all of you believing in me. I am so blessed to know each and every one of you. I especially want to thank Margi Colerick, Julie Roush, Debbie Lohmann, and Carol Everley for always being there for me when I need someone to talk to.

<b>TABLE OF CONTENTS</b>	<b>PAGE</b>
<b>1. Introduction</b>	<b>1</b>
1.1 Overview of glycoprotein analysis	1
1.1.1 <i>N</i> -linked glycoproteomics	2
1.1.2 Glycoprotein therapeutics	4
1.2 Preparation for mass spectrometric analysis	6
1.2.1 Enrichment of the glycoproteome	6
1.2.1.1 Lectin affinity chromatography	7
1.2.1.2 Glycoprotein capture	8
1.2.1.3 Isolating a single glycoprotein	9
1.2.2 Protease digestion	9
1.2.2.1 Glycopeptide enrichment strategies	11
1.3 Mass spectrometric analysis of glycoproteins	13
1.3.1 Electrospray Ionization	15
1.3.2 FT-ICR mass spectrometry	17
1.3.3 Ion trap mass spectrometry	19
1.3.4 Identification of peptides and glycopeptides	25
1.3.5 Quantitation Strategies	27
1.3.5.1 Quantitation of glycoproteins after glycan release	28
1.3.5.1.1 Quantitation methods for deglycosylated peptides using isotopic labeling	30
1.3.5.1.2 Label-free quantitation methods for deglycosylated peptides	34
1.3.5.2 Quantitation of glycopeptides	37
1.3.5.2.1 Label-free quantitation methods for glycopeptides	38
1.4 Structural/stability analysis of glycoproteins	41
1.4.1 Circular dichroism spectroscopy	42
1.5 Concluding Remarks	43
1.6 Summary of subsequent chapters	44
1.7 References	48
<b>2. Determination of an Ideal Protease Digestion Procedure for LC/MS Analysis of Soluble Proteins and Glycoproteins</b>	<b>60</b>
2.1 Introduction	60
2.2 Experimental	63

2.2.1 Materials and reagents	63
2.2.2 Glycoprotein protease digestion denatured with RapiGest™ SF	63
2.2.3 Glycoprotein protease digestion denatured with urea	64
2.2.4 Liquid chromatography/mass spectrometry	65
2.2.5 Data analysis	66
2.3 Results and discussion	67
2.3.1 Survey of various protease digestion conditions	67
2.3.2 MS data analysis of peptides and criteria chosen to compare the different digestion conditions	69
2.3.3 Comparison of denaturing and reducing agents	72
2.3.4 Under alkylation of Cys residues	75
2.3.5 Over alkylation	75
2.3.6 Evaluation of sequence coverage	75
2.3.7 Determination of the optimal conditions	76
2.3.8 Detection and analysis of glycopeptides	78
2.4 Concluding remarks	79
2.5 References	81
<b>3. Mass spectrometric analysis of sequence coverage and post-translational modifications of human lysyl oxidase-like 2 glycoprotein expressed in a <i>Drosophila</i> cell line</b>	<b>85</b>
3.1 Introduction	85
3.2 Experimental Procedures	87
3.2.1 Materials and reagents	87
3.2.2 Protease digestion	87
3.2.3 Liquid chromatography/mass spectrometry	88
3.2.3.1 Mass spectrometry on an ESI-LTQ-FTICRMS	88
3.2.3.2 Mass spectrometry on an ESI-LTQ Velos MS	89
3.2.4 Data analysis	89
3.3 Results and discussion	91
3.3.1 Protein sequence coverage	92
3.3.2 Glycopeptide data analysis	94
3.3.3 Detection of the lysyl – tyrosyl cross-link	99
3.4 Concluding remarks	102
3.5 References	103
<b>4. Label-free quantitation: A new glycoproteomics approach</b>	<b>106</b>

4.1 Introduction	106
4.2 Experimental	111
4.2.1 Materials and reagents	111
4.2.2 Enzymatic glycan trimming with $\alpha$ -mannosidase	111
4.2.3 Glycoprotein protease digestion	112
4.2.4 Glycopeptide enrichment	112
4.2.5 Mass spectrometry	113
4.2.6 Data analysis	114
4.2.7 Data treatment for quantitative method	115
4.3 Results and discussion	116
4.3.1 Developing a quantitative method for glycopeptides	116
4.3.2 Quality control experiment 1 – robustness	121
4.3.3 Quality control experiment 2 – applicability to different glycoproteins	123
4.3.4 Quality control experiment 3 – Instrument precision	125
4.3.5 Mixture analysis	127
4.3.6 Monitoring changes in glycosylation	129
4.4 Concluding remarks	131
4.5 References	133
<b>5. Assessing secondary structure of glycoproteins in the presence and absence of high mannose <i>N</i>-linked glycans</b>	<b>138</b>
5.1 Introduction	138
5.2 Experimental	141
5.2.1 Materials and reagents	141
5.2.2 Preparation of glycoprotein stock solutions	141
5.2.3 Circular dichroism (CD) spectroscopy	141
5.2.4 Protease digestion	143
5.2.5 Liquid chromatography/mass spectrometry	144
5.3 Results and discussion	146
5.3.1 Validation of glycan removal	149
5.3.2 Circular dichroism analysis on glycosylated and deglycosylated proteins	155
5.3.2.1 CD Secondary Structure Results	155
5.3.2.2 Thermal melt studies using CD spectroscopy	158
5.4 Concluding remarks	161
5.5 References	162

# CHAPTER 1

## INTRODUCTION

### 1.1 Overview of glycoprotein analysis

Glycoproteins are a large class of proteins involved in a wide variety of cellular processes and regulatory mechanisms. In fact, more than 50% of proteins found in serum are thought to be glycosylated.<sup>1-6</sup> The oligosaccharides present on glycoproteins impart unique functionality onto a nascent protein chain, and this added functionality helps ensure that protein folding, transport, and signaling events are properly carried out.<sup>3, 5</sup> Vital events such as protein degradation and modification of many cell to cell interactions are regulated by the glycan moieties comprising these glycoproteins.<sup>7-9</sup>

Glycoproteins have been increasingly associated with biomarkers for many different types of cancer and other disease states. Specifically, these protein concentrations have been shown to be up or down regulated,<sup>10, 11</sup> and changes in the typical glycosylation have been shown to occur with the onset of numerous pathologies.<sup>11, 12</sup> Monitoring changes in glycosylation or glycoprotein concentration is a critical step in identifying new biomarkers to improve early detection of adverse pathological states.

Glycoproteins are also an important class of pharmaceuticals. The manipulation of glycosylation on proteins used in pharmaceutical development is becoming increasingly common, since the glycan moieties found on therapeutic glycoproteins have been shown to increase the protein's efficacy and circulation half-life.<sup>13-15</sup> Thus, additional care must be taken to ensure that the glycosylation engineered during drug design is metabolically compatible with the human body. Therefore, the ability to profile

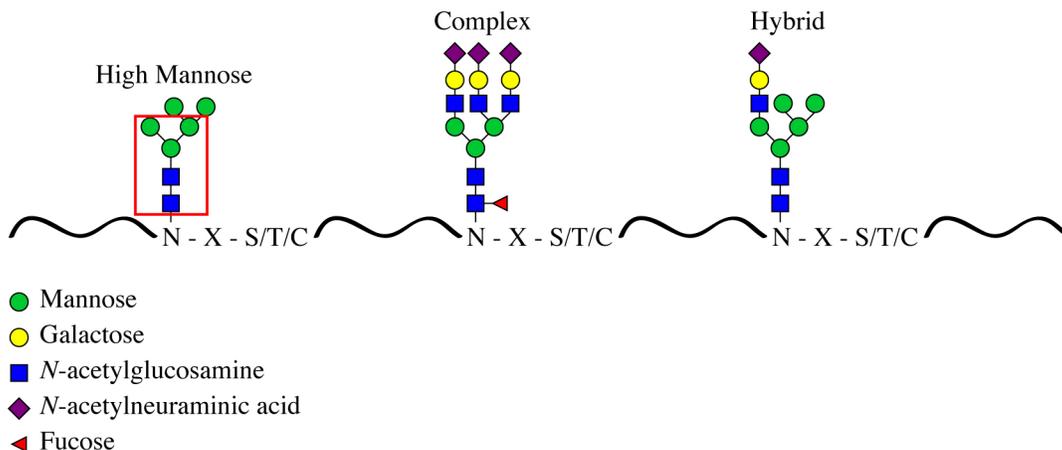
glycosylation on proteins is important to a variety of fields, including biomarker discovery and glycoprotein drug development.

Mass spectrometry has been shown to be an effective tool in glycoprotein analysis, as tandem mass spectrometry experiments, along with high resolution mass spectral data, work together to allow unambiguous identification and, at times, quantification of glycosylated proteins.<sup>2, 3, 5</sup> Mass spectrometry is a technique that is capable of glycoprotein analysis in both large scale studies of multiple glycoproteins that have been extracted from tissues, serum, or other types of bodily fluids (glycoproteomics) and for analysis of a single glycoprotein, such as a purified recombinant glycoprotein with potential as a pharmaceutical candidate (therapeutics). The work described herein has applications that can be important for both glycoproteomics and therapeutic approaches.

### **1.1.1 N-linked glycoproteomics**

Glycoproteomics is an analytical approach for studying glycoproteins; it relies heavily on the use of mass spectrometry. This field is a subset of the larger field of proteomics, which focuses on the characterization, identification, and quantitation of proteins.<sup>3, 16</sup> Glycoproteomic studies typically focus on the two most common types of glycosylation, O-linked and N-linked. O-linked glycans are oligosaccharides that are covalently attached to a serine or threonine amino acid residue on some proteins. This type of glycosylation has several core structures and no consensus amino acid sequence by which to determine where the glycan will be attached.<sup>17, 18</sup> Oligosaccharides undergoing N-linked glycosylation are attached through an asparagine residue containing the amino acid sequence Asn-Xxx-Ser/Thr and rarely

Asn-Xxx-Cys<sup>19, 20</sup> where Xxx can be any amino acid except proline. The core structure for *N*-linked glycosylation is a pentasaccharide consisting of two *N*-acetylglucosamine and three mannose residues.<sup>5, 17, 21</sup> Figure 1 illustrates the different types of *N*-linked glycosylation.



**Figure 1.** Representative types of *N*-linked glycosylation. The conserved pentasaccharide core is boxed in red on the high mannose type glycan.

One aspect of glycoproteomics is to measure relative changes in glycoproteins by either measuring changes in glycoprotein concentration or changes in the glycosylation present on the protein. These studies are necessary for a variety of reasons, including monitoring alterations in the glycosylation and/or the protein abundance of glycoproteins during progression of certain diseases, such as cancer and congenital disorders of glycosylation (CDG),<sup>6, 18</sup> where the overall goal is for early disease detection. Thus, sensitive and quantitative analyses are extremely important for detection of disease states, as levels of glycoproteins in body fluids vary widely.<sup>22-24</sup>

The example above is a situation where it is generally necessary to retain the glycans on the protein for analysis; however, there are other applications for studying the glycoproteome that do not require retention of the glycans, once enrichment of the

glycoproteome is completed. For example, if the goal of the study is to determine protein expression level changes in glycosylated proteins,<sup>25-27</sup> then retention of the glycans after glycoprotein enrichment may hinder the analysis. Additionally, studies aimed at determining glycosylation site occupancy, typically cleave off the glycans during the sample preparation steps.<sup>28, 29</sup>

### **1.1.2 Glycoprotein therapeutics**

As the pharmaceutical industry continues to progress in the expansion of protein drugs, more and more glycosylated proteins are becoming targets for therapeutic development. In fact, there are several glycoprotein drugs already on the market, such as erythropoietin,<sup>30</sup> glycoprotein hormones (i.e. thyroid stimulating hormone, luteinizing hormone, etc.),<sup>31</sup> and various human antibodies.<sup>32</sup> Potential areas of growth include vaccine candidates like the envelope glycoprotein on the surface of the HIV virus, glycoproteins shown to be deficient in the body, and those glycoproteins that are found to be down regulated during disease progression, such as cancer. Most protein drugs available today are expressed in a variety of cell lines including bacteria, fungi, insects, and mammalian cells.<sup>13, 14</sup> Production of nonglycosylated proteins is achievable in a wider variety of expression systems compared to the cellular production of glycoproteins, because most nonglycosylated proteins are only dependent on transfecting the desired DNA into a particular cell line and stimulating the cell to produce the protein. Thus, recombinant (nonglycosylated) proteins can be expressed in any of the cell lines mentioned above. Glycoproteins, on the other hand depend heavily on the glycosylation machinery present in the Golgi apparatus of a cell. Therefore, in addition to transfecting and producing glycoproteins in a cell, the cell line used is of utmost

importance because resulting glycosylation is cell dependent and not encoded into the DNA.<sup>13</sup> Therefore, glycoproteins are commonly expressed in eukaryotic cell lines where glycosylation machinery is available as explained in more detail below.

In many glycoproteins, one role of the glycans is to aid in protein folding; therefore, glycosylation is often necessary to produce properly folded and active glycoproteins.<sup>9, 33</sup> Since bacteria cell lines, such as *Escherichia coli*, do not contain glycosylation machinery, proteins requiring glycosylation cannot be produced in these types of cell lines.<sup>13</sup> Fungi and insect cell lines can produce glycosylation on proteins; however, the type of glycans formed from recombinant glycoproteins in these systems are very different from the glycans seen in human glycoproteins.<sup>13, 34</sup> For example, humans produce *N*-linked glycoproteins with sialylated complex type glycans and fungi/insect cells produce *N*-linked glycans of the high mannose type (see Figure 1).<sup>9</sup> Therapeutic glycoproteins expressed in fungi and insect cell lines will inevitably have a much shorter circulation half-life in the body compared to the same glycoprotein with sialylated complex type *N*-linked glycans because mannose binding lectins recognize high mannose glycoproteins as non-self, and will remove these glycoproteins from the body.<sup>32, 35-37</sup> Unfortunately, fungi and insect cell lines typically have much higher production yields compared to mammalian cell systems making these cell lines more amenable to the mass production required for recombinant therapeutic drugs.<sup>14, 37-39</sup> Mammalian cell systems produce glycoproteins with glycosylation most similar to human glycans.<sup>13, 14</sup> One of the most commonly used mammalian cell expression systems for therapeutic glycoproteins is Chinese hamster ovary (CHO) cells.<sup>13, 14, 32</sup> A problem with using CHO cells for producing recombinant glycoproteins, aside from the

lower production yields when compared to yeast or insect cells, is that the resultant *N*-linked glycans contain two different types of sialic acids, *N*-acetylneuraminic acid and *N*-glycolylneuraminic acid, but human *N*-linked glycans only comprise one type of sialic acid, *N*-acetylneuraminic acid.<sup>40, 41</sup> There have been a few reports where glycoprotein drugs expressed in CHO cells have caused an immunogenic response.<sup>15, 31, 42</sup> Therefore, as interest in glycoprotein biopharmaceuticals increases, and a greater number of glycoproteins are designed for therapeutic use, new methods to evaluate the efficacy, metabolism, and *in vivo* circulation of these drugs will need to be developed, approved, and regulated by the Food and Drug Administration.

## **1.2 Preparation for mass spectrometric analysis**

Many glycoproteins for mass spectrometric analysis come from very crude biological matrices. For example, samples for glycoproteomics studies typically originate from different bodily fluids, such as serum.<sup>8</sup> Recombinant glycoproteins, such as therapeutic drug candidates, are present either within the cells used for protein production or excreted into the cellular extract.<sup>43</sup> In both of these situations, many purification steps are often required prior to mass spectrometry experiments.

### **1.2.1 Enrichment of the glycoproteome**

While approximately 50% of the proteins in the body are glycosylated, the remaining non-glycosylated proteins are also abundant, but not important for analysis of the glycoproteome. To obtain optimal mass spectrometry (MS) data, efforts need to be made to remove as many of these interferents as possible. Some research aimed at quantitative analysis of glycoproteins has a sample preparation component to separate out interfering proteins. The two most common strategies for enriching glycoproteins

are lectin affinity chromatography and glycoprotein capture. Other enrichment techniques can be found in the literature, as well. Tables 1 describes the different types of glycoproteome enrichment strategies at the protein level, as well as a summary of advantages and concerns for each type.

**Table 1.** Description of different types of enrichment for glycoproteins

Type of Enrichment	Ideal Application	Key Considerations	References
<b>Lectin Affinity Chromatography</b>	Analysis of a specific glycan type	Lectins are somewhat promiscuous in their affinity for glycans. Multiple lectins are required to enrich multiple glycan types	[26-32]
<b>Glycoprotein Capture</b>	Quantitation of glycosylated proteins	Must remove glycans prior to MS, thus all glycosylation information is lost	[21, 33-35]
<b>Affinity Chromatography</b>	Analysis of 1 glycoprotein	An antibody is not always available for the glycoprotein of interest	[36, 37]

### 1.2.1.1 Lectin affinity chromatography

Lectins are a class of mammalian and plant proteins that have highly specific binding sites for monosaccharide moieties or particular glycan chains containing certain branching patterns.<sup>56</sup> Thus, lectins are often exploited for glycoprotein enrichment. Due to their high specificity, lectins are best utilized for glycoproteomics experiments when a specific type of glycosylated protein is to be purified from a complex mixture.<sup>57</sup> For example, wheat germ agglutinin (WGA) has a specificity for sialic acid residues<sup>58</sup> and *N*-acetylglucosamine residues containing a  $\beta$  linkage, which is present on the pentasaccharide core of *N*-linked glycans.<sup>58, 59</sup> Hill *et. al.* used WGA to enrich for glycoproteins in their glycoproteomic analysis.<sup>60</sup> Many lectins, while very useful for enrichment of specific types of glycosylated proteins, are often too specific for glycoproteomics studies, when the goal is to profile all the glycosylation diversity present on the glycoprotein(s). As an alternative to using the WGA lectin, a series of

lectins can be employed to enrich a large variety of glycoproteins.<sup>61, 62</sup> A variety of lectin combinations could be used. The supplementary data from Tao *et. al.* lists many known lectins and their specificities,<sup>59</sup> which could help researchers in choosing lectins for their experiments. Since the glycoforms are not cleaved during lectin affinity chromatography, this method is especially useful when the end-goal is to characterize the glycosylation on the proteins or peptides.

### **1.2.1.2 Glycoprotein capture**

When the goal of the experiment is to quantify the glycoprotein, and not to characterize glycosylation, many researchers opt for glycoprotein capture systems to facilitate sample preparation.<sup>63-65</sup> Glycoprotein capture strategies utilize hydrazide chemistry. A bead or resin containing a hydrazide functional group is covalently attached to a glycan through cis-diol groups present on some sugar moieties.<sup>34</sup> The advantage to this strategy is that any accessible glycan on the glycoproteins that contains a cis-diol group will be attached to the bead, and all non-glycosylated proteins/peptides can then be removed. Compared to lectin affinity chromatography, glycoprotein capture strategies are much less specific, allowing for a more general enrichment of the glycoproteome.<sup>25</sup> Once the non-glycosylated proteins are successfully removed, an enzyme is added to remove the glycan from the protein.<sup>64</sup> For glycoproteomics studies that seek to determine protein concentration changes, this method has proven to be excellent. However, when information about the glycosylation is required, the capture method is not adequate. Removal of glycoproteins/glycopeptides from the bead/resin necessitates loss of some or all of the glycan.

### 1.2.1.3 Isolating a single glycoprotein

When only one glycoprotein or one class of glycoproteins needs to be enriched, techniques that are specific to the glycoprotein of interest can be used. For example, Wuhrer *et. al.* used Protein A immobilized onto Sepharose beads to enrich for Immunoglobulin G (IgG) proteins in human serum because of Protein A's high affinity for IgGs.<sup>66, 67</sup>

### 1.2.1 Protease digestion

For analysis of proteins containing post-translational modifications (PTMs) by mass spectrometry, most proteins are cleaved by proteases to create several peptides, some of which contain PTMs.<sup>5, 44</sup> MS analysis on intact proteins may prove difficult when trying to unambiguously identify the individual PTMs present, especially in cases where the specific site of attachment must be ascertained as well.<sup>45</sup> This is especially true for glycosylated proteins, as glycoforms are typically heterogeneous for a given glycosylation site and often multiple glycosylation sites are present within the primary protein sequence as well.<sup>5, 45</sup> Even when only one glycosylation site is present, individual glycoforms may not be resolved well enough within the MS data to allow for compositional assignment of the glycans. This further necessitates the need for protease digestion in glycoproteins.

There are many different options available for choosing proteases to cleave proteins. Some proteases cleave before or after specific amino acid residues including trypsin,<sup>5, 46</sup> endoproteinase GluC,<sup>46</sup> and endoproteinase AspN,<sup>46</sup> while other proteases cleave relatively non-specifically, such as proteinase K and pronase.<sup>5</sup> Trypsin is one of the most commonly used proteases for mass spectrometric analysis of proteins.<sup>47</sup> The

specificity of trypsin allows for the prediction of expected  $m/z$  values for resultant peptides, allowing for targeted MS data searches which help to speed up the data analysis process.<sup>5, 47</sup> Under appropriate conditions, trypsin cleaves after Arg and Lys amino acid residues, which are fairly common in most proteins.<sup>48</sup> A key advantage in the MS analysis of tryptic peptides is that an amine group is present due to the retainment of Arg and Lys residues at the C-terminal end of every peptide.<sup>48</sup> The presence of these side chains help to ensure an addition of at least one proton to peptides during the ionization process, which helps to improve the ionization efficiency of tryptic peptides over peptides that do not contain amine groups at the C-terminal ends.

Most proteolytic enzymes are not effective at cleaving proteins in their native conformations; therefore, proteins must be completely unfolded prior to the addition of the protease in order to achieve efficient digestion.<sup>49-51</sup> Denaturants, such as chaotropes and detergents are among the most common reagents to unfold a protein. Chaotropes work by hydrogen bonding to a protein; competing with the intramolecular hydrogen bonding that keeps the protein structure intact.<sup>52, 53</sup> Detergents allow the hydrophobic regions, which are typically buried in the center of the protein, to interact with the solvent, thereby unfolding the protein.<sup>52, 53</sup> Detergents are excellent for unfolding proteins that are difficult to solubilize, e. g. membrane proteins, and new mass spectrometry friendly detergents, such as Waters corporation's Rapigest<sup>TM</sup> SF allow for the use of detergents in the denaturing step of the protease digestion process.<sup>54</sup>

Many proteins contain disulfide bonds. Although denaturing agents may effectively unfold a protein, which will allow for better access of the protein to the

protease's cleavage site, the presence of disulfide bonds will impede protease efficiency.<sup>51</sup> Therefore, most digestion procedures (except those where disulfide bonded peptides are sought) include a step to break (or reduce) the disulfide bonds.<sup>55</sup> After disulfide bonds are reduced, the free Cys residues that result are typically derivatized with an alkylating agent to ensure that reformation of disulfide bonds cannot occur.<sup>51</sup> At this point, the protein should be completely unfolded, thereby maximizing the efficiency of the protease. After protease digestion, the protein can either be directly injected onto a separation platform, such as reversed phase high performance liquid chromatography (HPLC) with data dependent mass spectrometry analysis or further enriched for PTM containing peptides, such as glycopeptides.

#### **1.2.2.1 Glycopeptide enrichment strategies**

When the goal of the experiment is to quantify glycosylation on individual peptides, enrichment strategies are often needed at the peptide level as well. To separate interferences at the peptide level (separating glycopeptides from their nonglycosylated counterparts) a sepharose-based enrichment, or chromatographic separation, is often implemented.<sup>68-70</sup> Table 2 summarizes different glycopeptide enrichment strategies. These enrichment methods exploit the chemical differences between glycan moieties and peptides, including differences in hydrophilicity, hydrophobicity, size, and net charge. Wada *et. al.*, for example, developed an in-solution extraction method using Sepharose CL-4B, a beaded hydrophilic polysaccharide polymer. When a mixture of peptides and glycopeptides is added to the Sepharose CL-4B solution, the glycopeptides remain in the hydrophilic Sepharose CL-4B fraction and the peptides are extracted out of solution with organic solvents before

elution of the glycopeptide fraction with aqueous solvents.<sup>68-70</sup> Normal phase high performance liquid chromatography (HPLC) or hydrophilic interaction liquid chromatography (HILIC) can also be performed to separate glycopeptides from peptides. Glycopeptides will be retained on a normal phase column longer than peptides, due to the hydrophilic glycan portion; thus, a separation of the two components will occur.<sup>71</sup> An alternative to HILIC methods is to separate glycopeptides from peptides by hydrophobicity. In these circumstances, reversed phase HPLC may be performed, and glycopeptides will tend to elute off the column before peptides.<sup>5, 70</sup> Alvarez-Manilla *et. al.*<sup>72</sup> and Joenväärä *et. al.*<sup>73</sup> used a glycopeptide enrichment method with size exclusion chromatography. Glycopeptides tend to be larger species compared to their peptide counterparts. Thus, glycopeptides will elute earlier from a size exclusion column.<sup>72, 73</sup> Strong cation exchange chromatography has also been shown to be useful for separating glycopeptides from peptides when there are sialic acids or other negatively charged species present on the glycopeptides. The negative charge of the sialic acids causes the glycopeptides to elute quickly off the column, compared to the peptides.<sup>74</sup>

**Table 2.** Description of different types of enrichment for glycopeptides

Type of Enrichment	Separation Mechanism	Key Considerations	References
<b>Sepharose Extraction</b>	Method has specific affinity for glycans	Further separations may be necessary for analysis of complex mixtures	[38-40]
<b>Normal Phase HPLC</b>	Glycopeptides tend to elute <i>later</i> than peptides	Glycopeptides with high peptide contribution may not elute in the glycopeptide fraction	[41]
<b>Reversed Phase HPLC</b>	Glycopeptides tend to elute <i>earlier</i> than peptides	Glycopeptides with high peptide contribution may not elute in the glycopeptide fraction	[12, 40]
<b>Size Exclusion Chromatography</b>	Glycopeptides are typically larger than peptides	Not all glycopeptides are larger than peptides	[42, 43]
<b>Strong Cation Exchange Chromatography</b>	Many glycans on glycopeptides have negatively charged components, whereas peptides do not	Not all glycopeptides contain negatively charged components	[44]

### 1.3 Mass spectrometric analysis of glycoproteins

Mass spectrometry (MS) is one of the most common techniques used for analysis of proteins, especially those containing post-translational modifications. The high throughput, sensitivity, and selectivity keep mass spectrometry as the workhorse instrument for protein identification and quantitation.<sup>48</sup> The instrumental design of a mass spectrometer consists of three main components: An ionization source, a mass analyzer, and detector. Most mass spectrometers used in protein analysis utilize either electrospray ionization (ESI) or matrix assisted laser desorption ionization (MALDI) for formation of gaseous ions.<sup>8</sup> These techniques are considered “soft” ionization techniques, because analyte molecules are not broken apart during the ionization process.<sup>75</sup> This ionization enables the MS detection of multiple analytes simultaneously. Examples of mass analyzers include ion traps (3-D or 2-D), triple quadrupole, time of flight (TOF), Fourier transform – ion cyclotron resonance (FT-ICR),

and, most recently, the orbitrap.<sup>76, 77</sup> Each of these analyzers has their own merits and limitations.<sup>77</sup>

All mass analyzers are capable of producing MS<sup>1</sup> data, which gives  $m/z$  values in a mass spectrum representative of all ions detected. A mass spectrum helps to identify proteins and peptides by their  $m/z$ ; however, proteins and peptides that have similar  $m/z$  values cannot be distinguished by their mass alone. Tandem mass spectrometry (or MS/MS) experiments allow for isolation of a specific ion, which is then fragmented into pieces.<sup>78</sup> The fragment ions detected help uniquely identify an ion present in a mass spectrum. Collision induced dissociation (CID) is one type of tandem mass spectrometry experiment that involves the introduction of inert gas, such as helium, into the mass analyzer or collision cell.<sup>78</sup> The inert gas collides with analyte ions that have been activated by an electric potential, where these collisions cause fragmentation of the analyte to occur.<sup>78</sup>

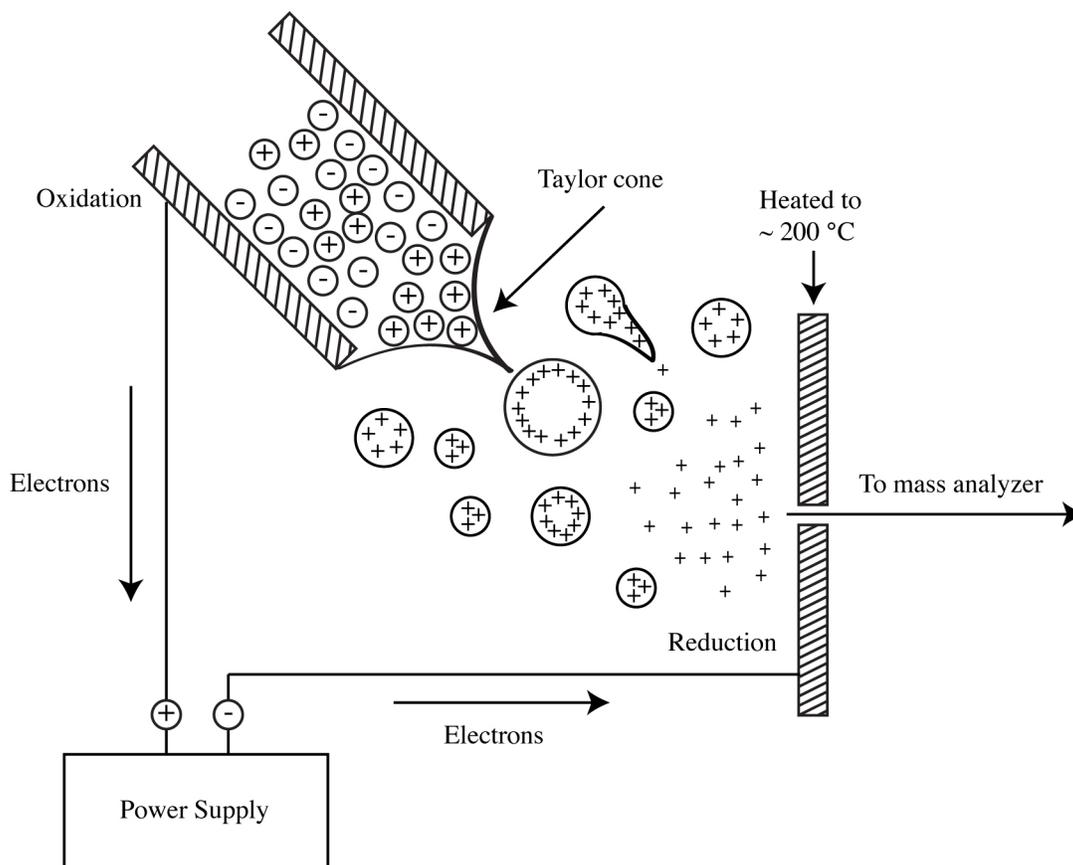
Ion traps and triple quadrupoles are readily capable of performing tandem mass spectrometry experiments; however, the MS<sup>1</sup> data is only available at unit resolution.<sup>77-78</sup> Although time of flight mass analyzers have the largest mass range, they are not amenable to tandem mass spectrometry experiments without the addition of some sort of collision cell (such as that found in the Qq-TOF).<sup>77</sup> The resolution capable on a TOF instrument is greatly improved compared to the ion trap or triple quadrupole; yet, it is not as high as the FT-ICR or orbitrap mass analyzers.<sup>76, 77</sup> The FT-ICR is capable of high resolution mass spectra, but it is costly to purchase, and the most difficult type of mass spectrometer to maintain, as it requires liquid helium and nitrogen for its super cooled magnet.<sup>79-81</sup> The orbitrap is one of the most recent mass analyzers developed, and it is

capable of high resolution MS<sup>1</sup> data similar to the FT-ICR, without the need for liquid nitrogen or helium.<sup>76</sup> The work described in this dissertation focuses on utilizing electrospray as the ionization source, FT-ICR MS for high resolution mass spectra, and the linear ion trap for tandem mass spectrometry experiments as well as low resolution MS<sup>1</sup> scans.

### **1.3.1 Electrospray ionization**

Electrospray ionization (ESI) is one of the most common ionization methods used to analyze peptides and glycopeptides, since it can be readily coupled to a high performance liquid chromatography (HPLC) system for online LC-MS and MS/MS analysis.<sup>82</sup> Figure 2 is a schematic of the electrospray ionization process.<sup>83</sup> A sample is introduced to the source through a capillary tube (ESI needle) with an applied potential between 2.5 to 5 kV, causing the analyte to become charged.<sup>82</sup> When working in positive ion mode, negatively charged species are attracted to the inside of the capillary while positively charged species (containing the analyte) form a Taylor cone at the end of the needle, as shown in Figure 2.<sup>83, 84</sup> Once the repulsive forces in the Taylor cone become greater than the surface tension, a spray of charged droplets will be released. The electrospray source is heated to approximately 200 °C, causing the solvent in the charged droplets to evaporate. A combination of solvent evaporation and Coulombic fission, due to increased space charge effects of the ions in the droplets as evaporation is occurring, allows for the formation of gaseous ions that enter the mass spectrometer.<sup>84, 85</sup> The ions generated by ESI may be present in more than one charge state,<sup>84</sup> and the presence of multiple charge states is often useful in the MS data, because it allows for the detection of proteins and peptides that would normally be well

beyond the mass range of most mass analyzers.<sup>82</sup> At times, however, multiple charge states are problematic as their presence complicates MS data analysis, especially in cases where there are mixtures of proteins or peptides present.<sup>82</sup>



**Figure 2.** Schematic of electrospray ionization illustrating ion formation and solvent de-solvation. Adapted from Kobarle.<sup>83</sup>

Lastly, some considerations are necessary for obtaining sufficient ionization efficiencies with ESI. Positive ion mode is typically utilized for ESI on many peptides and glycopeptides because of amine groups present on peptides that are favorable for protonation. To enhance the ionization of peptides, small amounts of acid (typically 0.1 to 0.5 % formic or acetic acid, respectively) are added to the solvent to provide excess protons for protonation of peptides and glycopeptides.<sup>82</sup> Solvent conditions need to be

fairly volatile to aid in the de-solvation process. Typical solvents include mixtures of water and organic solvents, such as acetonitrile and methanol.<sup>82</sup> The presence of salts inevitably decreases ionization efficiency as salts are nonvolatile and alter droplet formation from the Taylor cone.<sup>86</sup> Consequently, salts need to be minimized whenever possible. This is accomplished by using buffers that are ESI-friendly, such as Tris or ammonium bicarbonate for protein preparation.

### 1.3.2 FT-ICR mass spectrometry

Fourier transform – ion cyclotron resonance mass spectrometry (FT-ICR MS) is based on the principle that ions in a magnetic field will move in a direction perpendicular to that field, thereby forcing ions to move in a circular motion.<sup>79-81</sup> Equations 1.1 and 1.2 describe the different forces in play when ions are in the presence of a magnetic field; where  $F$  is force,  $e$  is the charge of an electron,  $v$  is velocity,  $B$  is magnetic field strength,  $m$  is mass and  $r$  is the radius of ion motion.<sup>81</sup> For stable movement of ions within the magnetic fields, a balance between Equations 1.1 and 1.2 needs to be met as is shown in Equation 1.3.<sup>81</sup>

$$F = evB$$

Centripetal Force  
Equation 1.1

$$F = \frac{mv^2}{r}$$

Centrifugal Force  
Equation 1.2

$$\frac{mv^2}{r} = evB$$

Equation 1.3

Because ions move in a circular path an angular frequency ( $\omega$ ) is considered and its relationship to velocity is shown in Equation 1.4.<sup>81</sup> Rearranging Equation 1.3 in terms of magnetic field strength ( $B$ ) and incorporating in the angular frequency term, Equation 1.5 clearly shows that under the same magnetic field strength, ions of different  $m/z$  will travel at different frequencies.

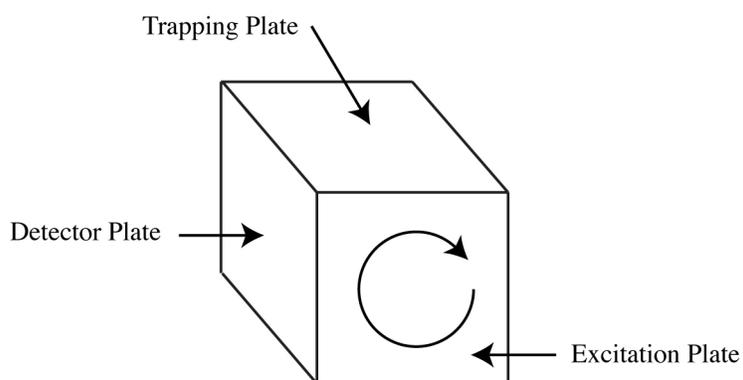
$$\omega = \frac{v}{2\pi r}$$

Equation 1.4

$$B = \frac{m}{e} \omega$$

Equation 1.5

One arrangement for an FT-ICR mass analyzer is composed of 6 plates placed together as a 6-sided cubic box that is located in the presence of an external magnetic field.<sup>79-81</sup> Figure 3 illustrates the FT-ICR cell and an ion traveling within the cell. Plates facing opposite one another have similar roles, where two plates keep ions trapped in the mass analyzer (trapping plates), two plates are for ion excitation (excitation plates), and two plates for ion detection (detector plates).<sup>79-81</sup>



**Figure 3.** Illustration of an FT-ICR cell showing the different plates where there is a second plate of the same type opposite to the one labeled. The circular arrow describes an ion in motion in the FT-ICR cell.

The trapping plates (one of which is the entrance to the FT-ICR cell) are set perpendicular to the magnetic field direction so that after ions enter the cell, a small positive potential can be applied (if detecting positive ions) to the trapping plates to keep ions trapped within that cell.<sup>79, 81</sup> Once ions are trapped, they will begin to move in a circular path, as shown in Figure 3, due to the external magnetic field. For detection of ions, a sinusoidal potential is scanned and applied to the excitation plates over a frequency range large enough to excite all the ions present in the cell.<sup>79, 81</sup> This results

in an increase in kinetic energy of the ions causing the orbit of ions to increase and packets of ions to form by  $m/z$ .<sup>80, 81</sup> As the ions pass by one detector plate, a small image current is induced and measured.<sup>79-81</sup> Since there are two detector plates opposite one another, an alternating current is induced as the ions pass by each plate.<sup>79</sup> The sinusoidal wave that results from the alternating current of ions passing the detector plates can be transformed into a frequency through the use of Fourier transform equations.<sup>79</sup> Using equation 1.5, the frequency of ion motion in the FT-ICR cell is converted to  $m/z$  to construct a mass spectrum. This method for the detecting ions allows for very accurate frequencies to be obtained for the ion packets, resulting in high resolution mass spectra.<sup>79-81</sup>

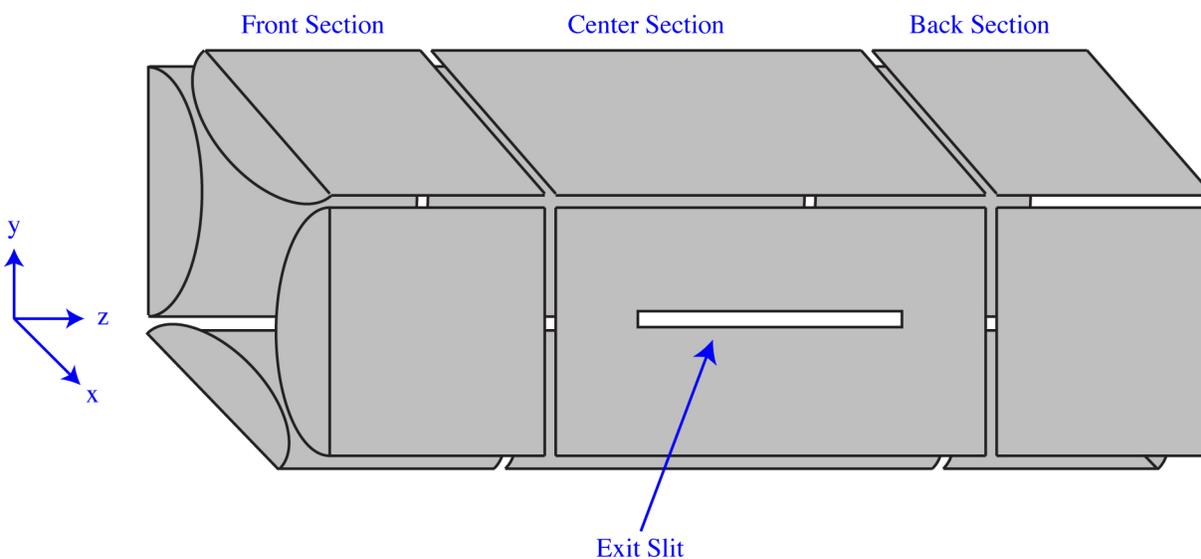
### **1.3.3 Ion trap mass spectrometry**

There are two different types of ion trap mass spectrometers commercially available; 3-D and 2-D (or linear) ion traps.<sup>87</sup> Both instruments work in the same manner to separate ions by  $m/z$  where ions are first stored, or trapped, within the instrument and then ejected sequentially by  $m/z$  to construct a mass spectrum.<sup>87, 88</sup> The 3-D ion trap uses only RF frequencies to store ions within the trap, whereas the linear ion trap utilizes a DC potential to trap ions in the axial direction and RF potentials for trapping ions in the radial direction.<sup>87</sup>

There are several advantages in using linear ion traps over 3-D ion traps, including the ability to trap more ions due to an increased analyzer volume, lower space charge effects, (as ions are focused onto a line across the center of the linear ion trap rather than a point in the center of the 3-D ion trap), and the ability to eject ions in a perpendicular direction to entrance of the trap, enabling the use of two detectors as

opposed to only one for ion detection.<sup>87-89</sup> The design changes employed in the development of linear ion traps allow for greater sensitivity and dynamic range in the mass spectral data.<sup>87-89</sup>

Linear ion traps look most similar to a quadrupole where either 4 circular or 4 hyperbolic rods are positioned in a circular fashion, where opposite rods are parallel to one another, as shown in Figure 4, which represents a schematic of the linear ion trap. In contrast to the quadrupole, rods in linear ion traps are separated into 3 sections; the front, center, and back, where the front and back sections are typically 12 mm in length and the center section is 37 mm in length.<sup>88-89</sup> The separation between sections allows for different DC potentials to be applied in each individual section when needed.<sup>88</sup> For example, to keep ions trapped in the mass analyzer, a higher DC potential will be applied to the front and back sections compared to the center section to repel ions away from the edges of the trap in the z direction (see Figure 4).<sup>88</sup> To contain ions in the x and y directions (see Figure 4) an RF potential is applied to all the rods where rods opposite one another are in phase, which creates a restoring force that keeps the ions focused along the center line of the trap.<sup>88</sup>



**Figure 4.** Schematic of a linear ion trap mass analyzer showing the quadrupole hyperbolic rods in three sections and the exit slit where ions are ejected from the trap in the x direction. This figure was adapted from Schwartz *et. al.*<sup>89</sup>

For ion isolation, activation, and detection, resonance excitation is used. Each ion present in a quadrupolar field moves at a specific frequency within the field, where ions of different  $m/z$  move at different frequencies.<sup>89</sup> To keep ions in the trap, the movement of the ions must be less than the dimensions of the trap, otherwise ions will hit the rods and not be detected or will leave the trap through the exit slits.<sup>89</sup> Ion stability within the linear ion trap at different RF potentials can be predicted through the use of Mathieu's a- and q-parameter equations:

$$a = \frac{8zU}{m\Omega^2 r_0^2}$$

Equation 1.6

$$q = \frac{4zV_{RF}}{m\Omega^2 r_0^2}$$

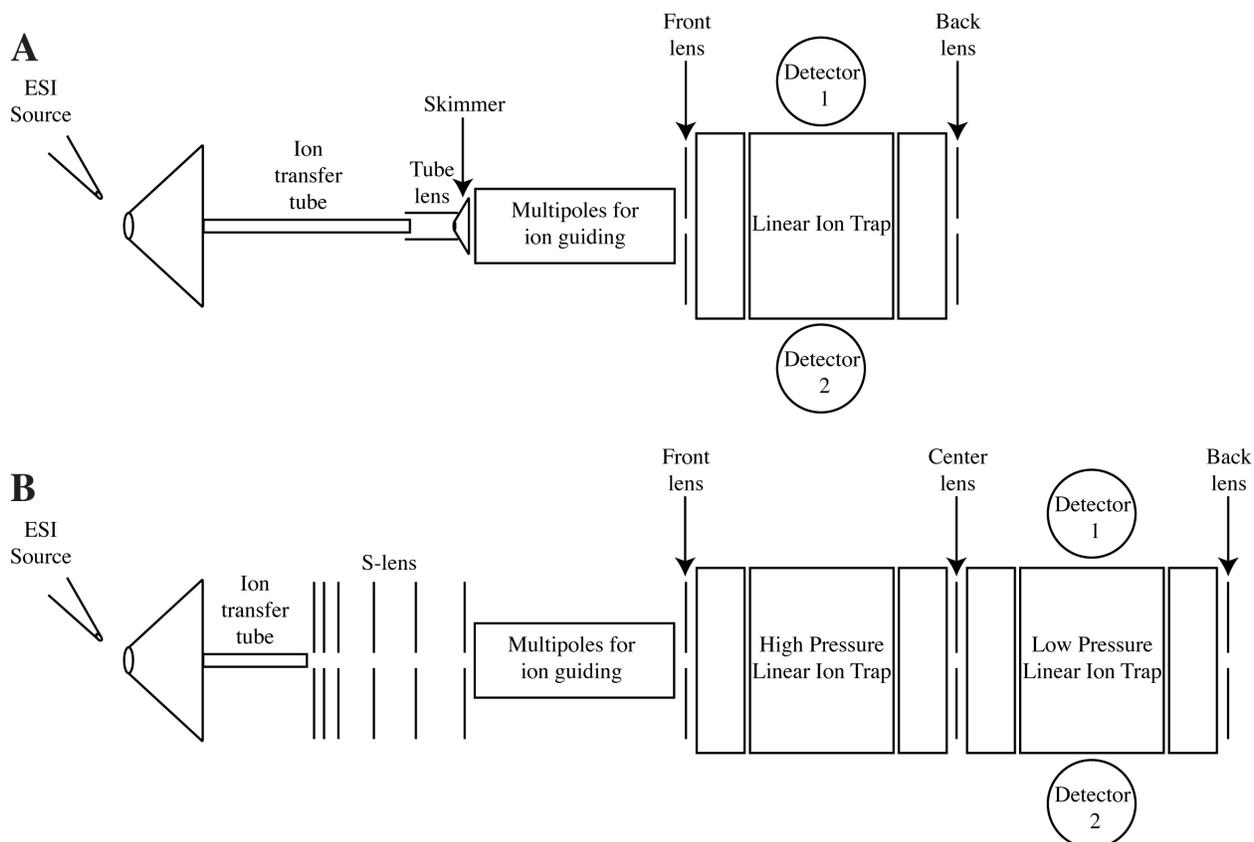
Equation 1.7

where  $U$  is the direct potential,  $V_{RF}$  is the RF potential,  $\Omega$  is the angular frequency,  $r_0$  is the distance from the center of the trap to one of the rods, and  $m/z$  is the mass to charge ratio.<sup>89</sup> When these two parameters are plotted against one another the regions

of overlap in the plot are representative of regions where ions are stable (i. e. ions will not hit the rods, but move within the confines of the trap).<sup>87-89</sup> Douglas *et. al.* show a representative Mathieu stability plot illustrating the areas where an ion will be stable in a linear ion trap.<sup>89</sup> Once an ion in a quadrupole ion trap system reaches a  $q$  value of 0.908, it will become unstable and leave the trap.<sup>89</sup> For ion ejection and detection, in the radial direction, the RF potential is scanned such that all the ions in the trap become unstable (or reach a  $q$  value of 0.908) and are removed from the trap and detected by their  $m/z$  values.<sup>89</sup>

Ion traps are particularly well suited for tandem mass spectrometry experiments because of the ability to isolate ions of one  $m/z$  in the trap. Ion isolation is achieved by removing all ions except the  $m/z$  of interest (or precursor ion). Two different mechanisms are used to remove unwanted ions. The first method is to scan the RF potentials to cause the ions below the precursor  $m/z$  to reach a  $q$  of 0.908 and become unstable.<sup>89</sup> In order to eject ions above the precursor  $m/z$  a method called resonance ejection is used.<sup>88, 89</sup> This method takes advantage of the fact that each ion has a different frequency of motion.<sup>89</sup> When the RF frequency matches an ion's frequency of motion it causes the amplitude of the ion trajectory to increase and eventually the ion will either hit a rod or leave the trap.<sup>78, 89</sup> A shorter RF frequency pulse is used to activate ions away from the center of the trap, but not far enough away from the center to remove the ions from the trap. This ion activation in combination with the addition of inert gas causes fragmentation of the precursor ion for CID MS/MS experiments.<sup>78, 89</sup> The RF potential is scanned after fragmentation to detect the fragment ions.<sup>89</sup>

The experiments herein utilize two linear ion trap (LTQ and LTQ Velos) mass analyzers, both manufactured by ThermoScientific, to achieve MS analysis of peptides and glycopeptides. The LTQ used is part of a hybrid mass spectrometer where it is coupled to an FT-ICR MS. This instrumental set-up was used to obtain high resolution mass spectra with the FT-ICR MS, while MS/MS experiments were performed within the LTQ component. The LTQ Velos mass analyzer was used as a stand-alone mass spectrometer, and though the MS<sup>1</sup> data from the LTQ Velos cannot match the resolution possible on the FT-ICR MS, the MS/MS data acquired on the LTQ Velos are improved compared to the LTQ by higher sensitivity and faster scan rates.<sup>90, 91</sup> Figure 5 illustrates schematic of the LTQ (Figure 5A) and the LTQ Velos (Figure 5B) where key differences can be elucidated.



**Figure 5.** (A) Schematic of a ThermoScientific LTQ mass spectrometer and (B) schematic of a ThermoScientific LTQ Velos mass spectrometer. Note that the ion transfer tube is about half the length in the LTQ Velos (B) compared to the LTQ (A), the skimmer and tube lens from the LTQ (A) were replaced with an S-lens (B), and there are 2 linear ion traps in tandem in the LTQ Velos (B) with differential pressures and only 1 linear ion trap in the LTQ (A). The schematics are adapted from Second *et. al.*<sup>90</sup> and Olsen *et. al.*<sup>91</sup>

One change that was implemented into the LTQ Velos design includes a different mechanism for focusing ions into the mass analyzer. The LTQ uses a tube lens and skimmer system for ion focusing whereas the LTQ Velos uses an S-lens (or an Stacked Ring Ion Guide),<sup>91</sup> which consists of a set of flat ring electrodes that are spaced further and further apart from one another, as shown in Figure 5B. Instead of a DC potential gradient being applied to focus ions toward the mass analyzer, the S-lens uses RF potentials where the odd and even ring electrodes are of opposite phases, which keeps ions focused towards the center of the S-lens lens as ions are guided into the linear ion

trap.<sup>90, 91</sup> The use of RF potentials for ion focusing in the S-lens system has been shown to increase the efficiency of ions that make it to the trap.<sup>92</sup>

The second innovation incorporated into the LTQ Velos is the use of two linear ion traps, rather than just one (see Figure 5). The first trap is maintained at a higher pressure, thereby allowing better collisional cooling and faster ion activation in collision induced dissociation (CID) MS/MS experiments, while the second trap is maintained at a lower pressure to be utilized solely for ion ejection toward the detectors. This dual pressure system has been shown to increase sensitivity and lower scan rates compared to the LTQ.<sup>90, 91</sup>

The last major innovation in the LTQ Velos is the use of predictive automatic gain control (AGC). Some ion trap mass spectrometers (including the LTQ) take a pre-scan of mass spectrum prior to the analytical scan that the user sees in the MS software. This pre-scan allows the instrument to assess the ion intensities and dynamically adjust injection time. A pre-scan takes ~ 30 ms to perform and is repeated for every spectrum taken. The predictive AGC that is utilized in the LTQ Velos removes the pre-scan, and uses prior MS<sup>1</sup> data that has already been collected to assess ion intensities and adjust injection times. This combination of improvements made in the LTQ Velos mass analyzer contribute to the increased sensitivity, high dynamic range, higher resolution, and dramatically shorter scan rates for MS/MS data compared to the LTQ.<sup>90, 91</sup>

#### **1.3.4 Identification of peptides and glycopeptides**

For identification proteins and their post-translation modifications, MS analysis is performed after proteolytic digestion of a protein(s) into peptides. When possible, mass spectra are obtained on a high resolution instrument, and tandem mass spectra are

utilized to aid in compositional elucidation of peptides.<sup>93</sup> Because protease digestions typically yield several peptides, data analysis of peptides and PTM-containing peptides have complicated MS data where large proteins or mixture of proteins can take weeks or even months for complete manual data analysis.

To ease the data analysis bottleneck, computer programs are available that can help analyze tandem mass spectra. Databases such as Mascot,<sup>94</sup> SEQUEST,<sup>95</sup> and X!Tandem<sup>96</sup> can be used to help identify peptides in the MS/MS data. Parameters can be set to search for some PTMs, such as phosphorylation and methionine oxidation.<sup>94</sup> Additionally, artifacts of the protease digestion can be taken into account, such as alkylation of Cys residues, alkylation of the N-terminus, and carbamylation of the N-terminus when urea is used for denaturation.<sup>94</sup> After a program examines the MS/MS spectra, a list of “hits” or peptide matches is obtained. The matches identified in the database programs can be compared to the MS/MS data for confirmation of identified peptides. The use of database programs does indeed speed up MS analysis time, however, manual data analysis is still important.<sup>97</sup> For manual analysis of peptides, characteristic ions (b and y ions in CID MS/MS data) are searched for in the MS/MS data. When the monoisotopic mass is near the calculated monoisotopic mass (mass accuracy depends on the mass analyzer used for MS analysis) for MS<sup>1</sup> data and MS/MS data contains appropriate fragment ions, a peptide can be considered identified.<sup>46</sup>

There are certain modifications that database programs do not take into account. Two examples include partial alkylation of Cys residues on peptides containing more than one cysteine and glycosylation of peptides. Glycopeptides are not searchable in

peptide database due to the high heterogeneity of glycans present on proteins, making the prediction of  $m/z$  values for potential glycopeptides quite complex. Thus, glycopeptide MS data is typically still analyzed manually. MS/MS spectra that are indicative of glycopeptides contain certain marker ions such as  $m/z$  366, 528, and 657 when CID is used for MS/MS data collection.<sup>5</sup> The presence of marker ions and losses of sugar residues (observed by subtracting the mass of different monosacchrides from the precursor ion mass) help to identify the glycan composition of glycopeptides. The peptide portion is identified from the difference in the mass of the precursor ion and the mass of the glycan portion.<sup>5</sup> There are a few software programs and databases available for aiding in MS analysis of glycopeptides, such as GlycoPep DB,<sup>98</sup> GlycoPep ID,<sup>99</sup> and GlycoMiner,<sup>100</sup> however, none of the available programs can analyze data as completely as the peptide database programs like Mascot. Until that time, interpretation of glycopeptides will be primarily completed manually.

### **1.3.5 Quantitation Strategies**

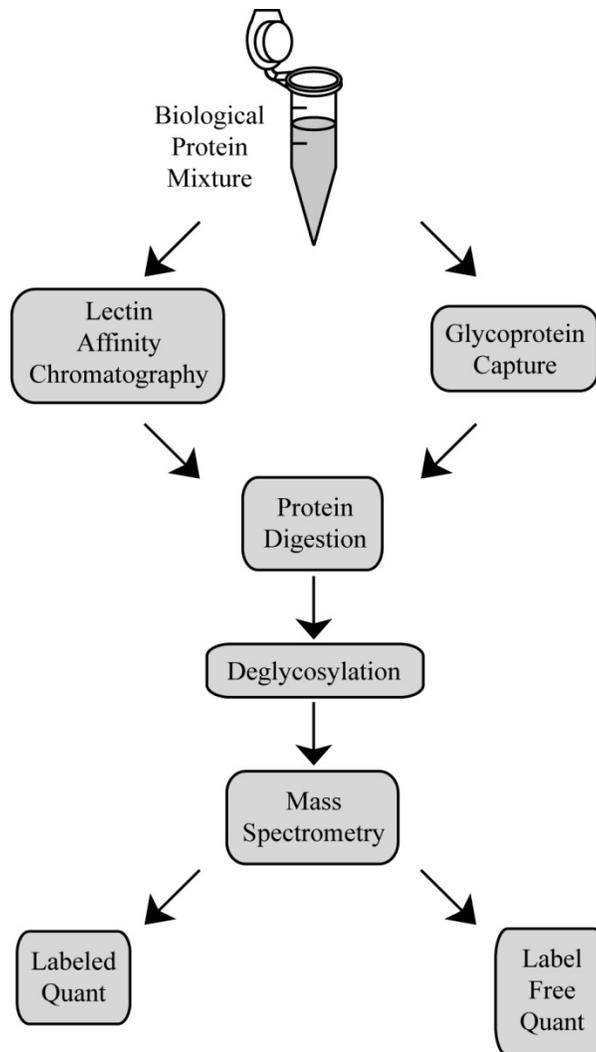
There are two basic strategies applied for relative quantitation of the glycoproteome; differential labeling and label-free analysis. Both strategies have their associated advantages and disadvantages. Many biomarker discovery experiments seeking to determine protein expression level changes employ labeling techniques to achieve relative quantitation between healthy and diseased samples.<sup>26, 101</sup> On the other hand, glycopeptide-based strategies typically encompass label-free approaches for construction of a glycosylation profile showing the distribution of glycoproteins present within a biological sample.<sup>67, 68</sup> These and other applications of quantitative glycoproteomics studies are described in more detail below.

### 1.3.5.1 Quantitation of glycoproteins after glycan release

Many quantitative glycoproteomics studies can be conducted by quantifying peptides after glycan removal. These studies could be done, for example, to calculate protein expression levels relevant to glycoprotein produced, to evaluate the amount of glycoprotein bound to a lectin column that has defined glycan specificity, or to measure and define glycosylation site occupancy, as cleaving the glycans introduces a 1 Da mass shift in the peptide mass. [It should be appreciated, however, that direct quantification of co-eluting peptides with a 1 Da mass difference can be problematic. Therefore, when the goal is to measure glycosylation site occupancy, labeling strategies are generally incorporated to increase the mass difference between the glycosylated and nonglycosylated compounds.] Regardless of whether the goal is to monitor site occupancy, glycosylation expression level, or quantify glycoprotein binding to lectins, all three types of experiments generally involve cleaving the glycans after the initial glycoprotein selection steps and quantifying the bare peptides.

Removal of glycans is often advantageous for many reasons. 1) Glycosylated peptides have a high degree of heterogeneity, due to the multiple glycoforms that can be attached to the peptide portion. Thus, when glycopeptides are identified in mass spectra, a peptide containing just one glycosylation site produces many spectral peaks, which correspond to those different glycoforms present along the peptide backbone. This heterogeneity contributes to difficulty in obtaining quantitative data with respect to the peptide sequence.<sup>102</sup> 2) The glycopeptide ion signal is spread among all the different glycoforms, lowering the abundance of the MS signal of each peak. 3) Glycopeptides are more difficult to analyze, due to the need to deduce both the peptide

composition and the glycan composition.<sup>5</sup> 4) Additionally, negatively charged glycans commonly present in most mammalian glycoproteins, such as or sialic acid residues, also negatively impact the MS signal of glycopeptides detected in positive ion mode.<sup>103</sup> The following sections describe differential labeling and label-free quantitation techniques that achieve quantitative glycoproteomics by analyzing (de-glycosylated) peptides from the glycoproteome. Typical work flows of glycoproteomics studies where the glycans are cleaved prior to mass spectral analysis are described in Figure 6. Biological samples are first enriched for glycoproteins primarily by lectin affinity chromatography or glycoprotein capture, followed by protein digestion with a protease such as trypsin. After protease digestion, the glycans are cleaved from the protein for mass spectrometry analysis, where finally the peptide data is evaluated for quantitative changes using either isotopic labeling or a label-free approach.



**Figure 6.** Flow chart summarizing experimental/data analysis protocols of the quantitative analyses from glycoprotein samples that are eventually de-glycosylated and analyzed as peptides by mass spectrometry.

### 1.3.5.1.1 Quantitation methods for de-glycosylated peptides using isotopic labeling

These methods are best used when only two types of samples are being compared, such as in biomarker studies in which healthy (control) samples are profiled and evaluated against diseased-state samples.<sup>26, 62, 101</sup> There are a variety of options available for differential labeling of peptides, including those that utilize the incorporation

of stable isotopes into the amino acid sequence, as well as methods that instead require the derivatization of the peptides with isotope-enriched tags. The greatest advantage of these strategies, with respect to label-free methods, is that samples being compared are mixed and analyzed simultaneously. Simultaneous analysis of the two samples is not only more efficient; it removes much of the signal variability associated with run-to-run inconsistencies. Mixing the samples together also eliminates slight differences in retention times for peptide peaks, which is a common problem for samples not run simultaneously. Moreover, labeling can be very useful in detecting subtle changes within samples.<sup>104</sup>

Stable isotope labeling is a powerful tool for comparing glycosylation changes across a variety of biological samples. Quantitation using this strategy is accomplished by calculating the ratio of intensities between the “light” and “heavy” sample populations. This ratio is then used to determine if protein concentration is up- or down-regulated. Stable isotope labeling is particularly helpful when identifying cell-surface glycosylation changes between complex biological sample types, as demonstrated by Wollscheid *et. al.* in the comparison of cell surface glycosylation of T and B cells during neuronal activation.<sup>25</sup> In this work, the glycoproteins were “pre-labeled”, using the SILAC (Stable Isotope Labeling of Amino acids in Cell culture) method. The cell-surface glycoproteins were captured using a bi-functional linker (one side with a hydrazide functional group to affix covalently to the glycans of cell surface glycoproteins, the other, a biotin tag that attaches to streptavidin beads for glycoprotein purification). After tryptic digestion, the glycopeptides were released with Peptide N-

Glycosidase F (PNGase F). Protein concentration was inferred based on the differences in abundance of the various isotopically labeled peptides.

Stable isotopic labels were also used to quantify protein expression of the serum glycoproteome in lung cancer patients by Ueda *et. al.*<sup>26</sup> In this analysis, glycoproteins were enriched using serial lectin affinity fractionation, and the isotopic label was incorporated by <sup>13</sup>C<sub>6</sub>-2-Nitrobenzenesulfonyl labeling of tryptophan residues. This method has been reported to be highly sensitive, as the screening was focused toward low-abundance proteins.

Numerous other research groups have focused on novel quantification strategies making use of H<sub>2</sub><sup>18</sup>O as the isotopic label.<sup>28, 29, 74, 101</sup> For example, Hülsmeier *et. al.* introduced a mass difference of 2 Da between the control and experimental samples by diluting PNGase F, the deglycosylation enzyme, in heavy water (H<sub>2</sub><sup>18</sup>O) or non-labeled water (H<sub>2</sub><sup>16</sup>O). In these experiments, the isotopic label is introduced at the Asn that was formerly glycosylated. As the N converts to D, one <sup>18</sup>O is incorporated. This strategy is particularly useful when the research goal is to probe site occupancy of proteins. Hülsmeier's work is also notable in that the glycoproteins were initially captured using a sepharose affinity resin.<sup>28</sup>

A manuscript published by Liu *et. al.* reported the monitoring of changes in glycosylation from a patient with ovarian stage II b cancer vs. a healthy control, where three heavy <sup>18</sup>O atoms were incorporated into one of the samples by labeling the 2 O atoms at the C-terminus (during tryptic digest) and the site of deglycosylation, after PNGase F treatment. This gave a 6 Da mass shift between the two samples, allowing

for complete resolution of healthy and diseased peptide peaks in the mass spectral data.<sup>29</sup>

Additionally, other glycoproteomics studies have used labeling approaches where the stable isotope is incorporated at the N terminus. For example, Qiu *et. al.* used acetic anhydride to differentially label glycoproteins from serum that were retained on lectin columns. The lectin encoded for the glycoform type (bi-, tri-, or tetra-antennary) and the resulting glycans were cleaved prior to MS analysis.<sup>61</sup> These studies demonstrate that, through the use of serial lectin affinity chromatography, tri- and tetra-antennary *N*-linked glycopeptides were separated from biantennary glycopeptides. After the separation, the isotopically labeled samples allowed for a quantitative comparison of branching patterns in serum glycoproteins.<sup>61</sup>

Finally, quantitative glycoproteomics can be completed using iTRAQ (isobaric Tag for Relative and Absolute Quantitation) reagents.<sup>57, 62, 65</sup> Deglycosylated peptides or glycoproteome enriched samples are tagged with iTRAQ reagents and mixed prior to mass spectrometric analysis. One advantage of this technique over other labeling strategies is that up to four<sup>105</sup> or, more recently, eight<sup>106</sup> samples can be analyzed at a time. Additionally, as a result of the isobaric tag, MS<sup>1</sup> data is not diluted over multiple peaks, which is the case for traditional isotopically labeled samples. To quantify using iTRAQ, tandem mass spectrometry (MS/MS) data is analyzed for reporter ions (ions of the naked iTRAQ reagents) and the ratio in the MS signal among reporter ions is used to determine up- or down-regulation of glycoproteins.<sup>65, 105, 106</sup>

Furthermore, the use of iTRAQ has become a very useful quantitative tool in biomarker studies.<sup>62, 65</sup> For example, Zhou, *et. al.* captured glycoproteins from tear fluid

using hydrazide chemistry and used iTRAQ to compare glycoprotein concentrations between control and climatic droplet keratopathy patients.<sup>65</sup> Lee, *et. al.* used multiple lectins to enrich glycoproteins from plasma and made use of iTRAQ for biomarker analyses on a comparative study between healthy and hepatocellular carcinoma specimens.<sup>62</sup>

#### **1.3.5.1.2 Label-free quantitation methods for deglycosylated peptides**

These methods described herein are another attractive option for glycoproteomics studies. One advantage to label-free studies is that there is no limitation to the number of samples for comparison, as is the case with labeling strategies. Additional reagents and extra sample preparation steps are also unnecessary, yielding a simpler work-flow.<sup>63</sup> The key disadvantage of label-free methods, however, is that the data analysis can be more challenging. This is due to a variety of reasons. 1) Since all the samples in a label-free analysis are not analyzed simultaneously on a mass spectrometer, run-to-run variability in signal intensity must be taken into account. Thus, label-free approaches can use normalization techniques to overcome the instrumental variability.<sup>107, 108</sup> 2) Label-free studies are often repeated multiple times to ensure reproducibility among samples.<sup>27, 107</sup> 3) LC-MS experiments require alignment, as the peptide peaks of interest have slight shifts in retention time among sample runs.<sup>27, 60</sup> Quantification applied to label-free approaches is done by comparing either the relative intensity of peptide peaks in mass spectra or the peak area under the curve of the extracted ion chromatograms for a given peptide. Often, software programs such as SuperHirn<sup>27</sup> and ProteinQuant<sup>107, 108</sup> are utilized to

accomplish the complex spectral alignment and data processing required for label-free studies.

Zhang *et. al.* demonstrated that it is possible to do a very simple quantitative analysis on serum glycoproteins by using glycoprotein capture for enrichment and label free quantitation (without normalization).<sup>109</sup> The capture strategy greatly reduced sample complexity and allowed for direct comparison of relative abundances from relevant peptide peaks between healthy and cancerous mouse serum. To validate this non-normalized label-free analysis strategy, N-termini of captured peptides were labeled with heavy and light succinic anhydride. Results from the labeled component were similar to the label-free analysis. It was indicated in this work, however, that variability was detected among relative abundances of replicate samples (i.e. the 3 healthy serum samples that were analyzed); however, differences were evident between healthy and cancerous peptides.<sup>109</sup>

Another method commonly used to quantify changes among samples is to calculate the area under the curve of a particular peptide peak from extracted ion chromatograms.<sup>27, 60, 63, 107, 108</sup> Schiess *et. al.* analyzed *drisophila melanogaster* cells to quantify changes in the cell-surface glycoproteome before and after perturbation of the cell with chemical reagents that are known to affect intracellular signaling events.<sup>27</sup> As described above, a bi-functional linker was used to capture cell-surface glycoproteins for analysis of the deglycosylated *N*-linked peptides. To quantify changes in the cell-surface glycoproteome, a ratio between perturbed peptide peak areas and control peptide peak areas was calculated. In order to maximize reproducibility, 3 cells were

analyzed in triplicate, yielding 9 replicate sample sets that were averaged together for comparison between the perturbed and non-perturbed cells.<sup>27</sup>

The practice of calculating a ratio of peak areas to describe changes between two different sample sets has also been done by Hill *et. al.* in two separate cancer-related studies.<sup>60, 63</sup> Glioblastoma (cancerous) cells were exposed to a drug believed to decrease the proliferation of tumors. Glycoprotein capture was utilized to enrich for glycoproteins on cells before and after drug treatment.<sup>63</sup> A second study by Hill *et. al.* was done on model tumor cells to compare changes in glycoprotein expression levels before and after the addition of protein transforming growth factor beta, which is known to affect tumor cells. For this particular study, lectin affinity chromatography was used to enrich for glycoproteins.<sup>60</sup> In both sets of experiments, each sample was run in triplicate to improve reproducibility,<sup>60, 63</sup> and a ratio between treated and untreated peptides peak areas was calculated to determine relative changes in glycoprotein abundance.

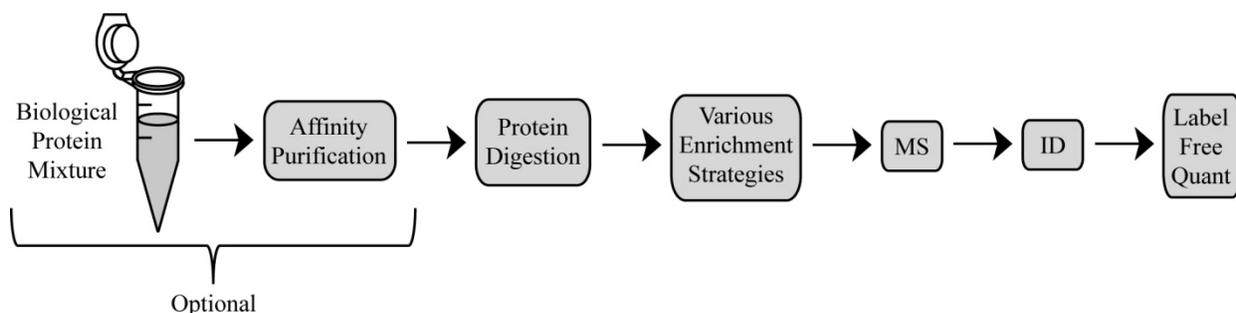
Other researchers have sought to remove the run-to-run variability that is common for label-free quantitative analysis by applying a normalization strategy during data analysis. Mann *et. al.* developed software tools, entitled ProteinQuant Suite, for automated proteomic quantitative analysis.<sup>108</sup> This program aids in removing run-to-run variability by either dividing the peak area (in SIM chromatogram) of a peptide peak by the sum of the peak areas of all peptide peaks in a given sample, or by dividing the peak area of a peptide peak by the peak area of an internal standard.<sup>108</sup> Madera *et. al.* used ProteinQuant to compare changes in glycoproteins due to differences in lectin matrix preparations.<sup>107</sup> Normalization was completed by summing the peak areas of all

peptide peaks present in a sample. It should be noted that the glycans were not removed in this study; however, the quantitative work-up was for the non-glycosylated peptides only.<sup>107</sup>

### 1.3.5.2 Quantitation of glycopeptides

At times, it is necessary to retain the glycans on glycoproteins during analysis of glycopeptides. Changes in glycosylation may occur without protein expression level changes as the glycosylation pattern is very much dependent on changes in the localized cellular environment.<sup>5, 17, 18, 110</sup> It is also known that different glycosylation patterns may occur on different glycosylation sites of glycoproteins.<sup>111</sup> Thus, it is sometimes necessary to profile glycosylation changes in a site-specific manner by analyzing glycopeptides directly.<sup>68, 111</sup> In addition to naturally occurring changes in glycosylation at specific glycosylation sites, disease progression of certain cancers and CDG are well known to change glycosylation.<sup>102, 112</sup> Tajiri *et. al.*, for example, developed a method to site-specifically monitor changes in fucose levels in known fucosylated proteins, as fucose is a biomarker for certain cancers.<sup>112</sup> In glycopeptide-based strategies, quantitation is carried out in a label-free fashion, by using either the relative intensity (peak height) or peak areas from extracted ion chromatograms. The intensities can be used directly to compare samples<sup>67</sup> or glycopeptide peaks can be normalized by dividing the intensity of a given glycopeptide peak by the sum of all intensities from all glycopeptide peaks present in either a mass spectrum,<sup>68</sup> window of retention time,<sup>113</sup> or the sum of peaks for a specific glycosylation site.<sup>114</sup> Biomarker analyses and other experiments aimed at comparing two samples to one another typically calculate a ratio for comparison between samples, as Uematsu *et. al.* did to

compare murine dermal and epidermal glycopeptides.<sup>115</sup> Glycopeptide-based methodologies are illustrated in Figure 7. When working with complex biological mixtures, such as serum or cell culture, an affinity purification step may be employed for initial enrichment for glycoproteins. Glycoproteins are proteolytically digested, followed by glycopeptide enrichment. Exploitation of the differences between peptides and glycopeptides are utilized to enrich for glycopeptides. Potential differences include changes in hydrophilicity, hydrophobicity, and charge (in the case of sialylated, sulfated, and phosphorylated *N*-linked glycans). Mass spectrometry is often used for glycopeptide detection, and identification of the glycopeptides is typically completed with the aid of web-based software such as GlycoMod,<sup>116</sup> GlycoPep DB,<sup>98</sup> GlycoPep ID,<sup>99</sup> and GlycoMiner.<sup>100</sup>



**Figure 7.** Illustration of the most common protocols for quantitative analyses of glycopeptides by mass spectrometry.

### 1.3.5.2.1 Label-free quantitation methods for glycopeptides

The label-free approach is the only currently adopted approach for glycopeptide-based glycoproteomics strategies. In these types of studies, researchers may be keenly interested in comparing glycopeptide profiles from one sample to another, or the focus may be to compare the quantities of the individual glycoform abundances on a particular glycoprotein. For instance, Kondo *et. al.* was interested in comparing the

glycopeptide profiles of sialylated glycopeptides between healthy and antiphospholipid syndrome patients in  $\beta$ 2-glycoprotein 1.<sup>113</sup> These researchers purified their protein with an antibody, digested it, conducted an additional purification at the glycopeptide level using solid phase extraction, and probed the glycopeptides using MALDI- (Matrix Assisted Laser Desorption Ionization) and ESI- (Electrospray Ionization) MS.<sup>113</sup>

Thaysen-Anderson *et. al.* was interested in quantifying all the different glycoform abundances at key glycosylation sites on several example proteins, not comparing how glycosylation a given protein can change during a disease state.<sup>111</sup> The label-free studies he carried out were done on model proteins such as RNase B, IgG, and fetuin. While the proteins were available in purified form, the glycopeptides still needed to be enriched from nonglycosylated species after tryptic digestion. To achieve efficient sample preparation, the authors compared a series of purification methods and analyzed data by MALDI-MS, and validation of their results was accomplished by performing label-free quantitation on the released glycans.<sup>111</sup>

Additional examples of researchers profiling the glycosylation on a key protein or set of proteins are also available. Ivancic *et. al.* analyzed  $\alpha_1$ -acid glycoprotein for differences in *N*-linked glycosylation branching (ie. biantennary, triantennary, or tetraantennary) by summing the extracted ion chromatographic peak areas for each branching type and comparing the percent of each branching type between two gene products of  $\alpha_1$ -acid glycoprotein.<sup>114</sup> In this work, the protein was first isolated by affinity chromatography and the de-sialylated protein was proteolyzed and subjected to reverse phase HPLC separation, prior to ESI-MS analysis.

Tajiri *et al.* was interested in calculating a ratio of fucosylated to de-fucosylated glycopeptides to determine the extent of fucosylation at specific glycosylation sites in glycoproteins.<sup>112</sup> In those studies, different antibody-purified proteins were digested, and the glycopeptides were subjected to a glycopeptide enrichment protocol using sepharose. Rebecchi *et al* also used the sepharose enrichment method in developing a label-free glycopeptide-based quantitation method that can differentiate between glycosylation changes and changes in protein abundance in ESI-MS data.<sup>68</sup> This method bridged the gap between studies that determine protein up and down regulation vs. analyses to quantify glycosylation changes.<sup>68</sup> Selman *et. al.* analyzed the glycopeptide profiles of their glycoprotein of interest (IgG) for a multitude of clinical samples.<sup>67</sup> In this case, isolated IgG glycopeptides were subjected to a rapid SPE purification, and MALDI-MS was used to profile IgG from 62 human serum samples. Glycopeptides were quantified based on relative intensity of each of the peaks.<sup>37</sup>

Other research in glycoproteomics is to use glycopeptide strategies in conjunction with other techniques. Wagner-Rousset *et. al.* developed a three-tiered approach to studying glycoproteins; they analyzed intact protein, released glycans, and glycopeptides.<sup>117</sup> To quantify glycopeptides, the relative intensities of the glycoforms present on the glycopeptides were compared among samples.<sup>117</sup> Wada *et. al.* conducted a multi-laboratory study to compare reproducibility in glycoproteomics methods across different instrumentation, researchers, and sample preparation procedures. In addition to quantifying glycans released from glycopeptides, a glycopeptide-based relative quantification approach was adopted by two laboratories; in these cases, the quantitative results were consistent with the glycan released studies.<sup>102</sup>

Finally, while the main thrust of glycoproteomic quantitation at the glycopeptide level is focused on the analysis of glycoforms comprising a single, purified protein; a few reports demonstrate the intriguing possibility of performing the same analysis on a larger sample set. For example, Uematsu and coworkers were able to compare high-mannose-containing glycopeptides from a variety of proteins isolated from murine dermis and epidermis.<sup>115</sup> A lectin affinity selection using a ConA-agarose column (which selects for high mannose glycans) simplified the sample complexity, and peak intensities from MALDI TOF/TOF were used to quantify differences between the dermal and epidermal samples. In the MALDI-TOF-TOF experiments, high resolution MS/MS data were used to identify the peptide.<sup>115</sup>

Similarly, glycopeptides from complex protein mixtures were quantified by LC-ESI-MS, using peak areas of the ions of interest by Ding and coworkers.<sup>71</sup> When working with even larger protein mixtures, significant effort must be placed on identifying the glycopeptide compositions. With ESI-MS data, identification of the protein from which the glycopeptide originates is possible, if MS<sup>3</sup> data of the unglycosylated peptide is obtainable.<sup>71</sup>

#### **1.4 Structural/stability analysis of glycoproteins**

While mass spectrometry is very useful for protein identification and relative quantification, especially for proteins containing post-translational modifications, it is difficult to use mass spectrometry to assess protein structure and stability. There has been some success in using chemical cross-linking,<sup>118</sup> hydrogen-deuterium exchange,<sup>119, 120</sup> and limited proteolysis<sup>121</sup> for determining proximity of certain protein regions and exposed residues by mass spectrometry; however, these approaches

cannot match the structural detail determined by techniques like nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography.<sup>122, 123</sup> Unfortunately, both NMR and X-ray crystallography require large amounts of protein in a very pure state.<sup>119-121, 124</sup> An alternative approach for structural analysis of glycoproteins, which has an added benefit of being able to also assess protein thermal stability is circular dichroism (CD) spectroscopy.<sup>124</sup> CD spectroscopy would be classified as somewhere between mass spectrometry and NMR for structural analysis of proteins because CD spectroscopy determines the level of secondary structure present rather than either primary amino acid sequence (mass spectrometry) or tertiary structure (NMR).

#### **1.4.1 Circular dichroism spectroscopy**

CD spectroscopy works by shining left and right circularly polarized light through a protein sample, where changes in the absorbance of the light are assessed, in terms of ellipticity, for the secondary structural elements of the protein, such as alpha helices and beta sheets.<sup>124, 125</sup> A protein's structural elements absorb the left and right circularly polarized light differently.<sup>125</sup> Measuring the changes in the absorbance of the light at different wavelengths in the far UV region helps to determine the percentages of alpha helices and beta sheets present in a protein sample.<sup>125, 126</sup> Thermal stability studies can also be performed on a CD instrument, as well. The change in ellipticity is monitored at one wavelength as the temperature of the protein solution is linearly increased, causing the protein to slowly unfold.<sup>127</sup> Melt plots typically can be fitted to a sigmoidal curve where the inflection point, or melt temperature ( $T_m$ ), is considered the point where half the protein is folded and the other half unfolded.<sup>128</sup> The higher the  $T_m$ , the more stable (or resistant to unfolding) the protein is considered.

## 1.5 Concluding remarks

Mass spectrometry is a versatile tool for identification and quantitation of glycoproteins where both large-scale glycoproteomics studies containing thousands of proteins and glycoproteins can be analyzed, as well as the analysis of a single glycoprotein, such as in the development of pharmaceuticals. High performance liquid chromatography mass spectrometry utilizing high resolution mass spectrometers (FT-ICR and orbitrap MS) for MS analysis combined with powerful tandem mass spectrometry instruments (performed in linear ion traps) make it possible to separate and unambiguously detect many peptides present in proteins, including those containing post-translational modifications, such as glycosylated peptides, in a single LC-MS run. The biological functions of glycoproteins, as well as their increasing importance in disease progression and pharmaceutical development, make identification and quantitation of glycoproteins necessary for disease prevention and overall health. Developing methods aimed at improving detection of glycoproteins are needed, as well as innovative strategies for relative quantitation of glycoproteins, for detection of glycosylation and glycoprotein concentration changes. The work described in this dissertation sought to fill both voids. Chapters 2 and 3 focus on improvement of the protease digestion procedure, which can be a bottleneck in the analysis of glycoproteins. Chapter 2 describes the method development steps and Chapter 3 applies the optimized digestion protocol for MS analysis on a novel glycoprotein. Chapter 4 illustrates a new mass spectrometric method for relative quantitation of glycoproteins. Chapter 5 uses circular dichroism spectroscopy as a pre-screening method to assess changes in glycoprotein structure and stability upon deglycosylation.

## 1.6 Summary of subsequent chapters

**Chapter 2** discusses a systematic approach for optimizing the protease digestion protocol on soluble proteins as a preparatory step before mass spectrometric analysis. As recombinant proteins are increasingly utilized for vaccine, pharmaceutical, and research development, improved methodologies ensuring the characterization of post-translational modifications (PTMs) are needed. Typically, proteins prepared for PTM analysis are proteolytically digested and analyzed by mass spectrometry. To assure full coverage of the PTMs on a protein, one must obtain complete sequence coverage of the protein, which is often quite challenging. The objective of the research described here is to design a protocol that maximizes protein sequence coverage and enables detection of post-translational modifications, specifically *N*-linked glycosylation. To achieve this objective, a high efficiency proteolytic digest protocol using trypsin was designed by comparing the relative merits of denaturing agents (urea and Rapigest<sup>TM</sup> SF), reducing agents (dithiothreitol or DTT, and tris(2-carboxylethyl)phosphine or TCEP), and various concentrations of alkylating agent (iodoacetamide or IAA). After analysis of human apo-transferrin using various protease digestion protocols, ideal conditions were determined to contain 6 M urea for denaturation, 5 mM TCEP for reduction, 10 mM IAA for alkylation, and 10 mM DTT, to quench excess IAA before the addition of trypsin. These digestion conditions were specifically designed for PTM analysis of recombinant protein therapeutics and can be widely implemented in biopharmaceutical analysis.

**Chapter 3** applies the protease digestion approach described in Chapter 2 to a biologically relevant protein, human lysyl oxidase-like 2 (hLOXL2) glycoprotein, which

has been expressed for the first time in *Drosophila* S2 cells. As hLOXL2 has been shown to be involved in the metastasis of several different types of cancer, there is strong interest in the development of hLOXL2 as a therapeutic drug for those who have a deficiency of the protein. Experimental evidence shows that hLOXL2 contains 2 *N*-linked glycosylation sites, 8 disulfide bonds, and a lysyl-tyrosol cross-link that is conserved across all LOX and LOX-like proteins. This class of proteins, though studied extensively, is difficult to isolate as well as to express. Thus, there are no prior studies where the LOX and LOX-like proteins have been completely mapped by mass spectrometry. Through mass spectrometric analysis, high protein sequence coverage was achieved (> 90 %); the glycosylation was detected and was determined to be consistent with the typical glycosylation profile for the cell line used. Finally, the cross-link was identified as partially intact. In order to obtain all the needed information on hLOXL2 by mass spectrometry, two different mass spectrometers were used. The first was an ESI-LTQ-FTICRMS, where a combination of high resolution MS and MS/MS spectra yielded high protein sequence coverage, but only glycopeptides from one of the two glycosylation sites were detected and the intact lysyl-tyrosol cross link was not found in the MS or MS/MS data. The second instrument utilized was an ESI-LTQ Velos MS, which provides low resolution MS data, but improved MS/MS sensitivity compared to the LTQ MS. By using the ESI-LTQ Velos MS, glycopeptides were detected at both glycosylation sites; additionally, the intact lysyl-tyrosol cross link was detected as well. The high success rate on the analysis of hLOXL2 validated the protease digestion procedure that was optimized in Chapter 2. These experiments illustrate the

importance for novel method development in protease digestion protocols that aim to prepare soluble glycoproteins for MS analysis.

**Chapter 4** describes a method for quantifying glycosylation changes on glycoproteins. This novel method uses MS data of glycopeptides to analyze glycosylation profiles, and several quality control tests were done to demonstrate the method is reproducible, robust, applicable to different types of glycoproteins, and tolerant of instrumental variability during ionization of the analytes. This method is unique in that it is the first label-free quantitative method specifically designed for glycopeptide analysis. It can be used to monitor changes in glycosylation in a glycosylation site-specific manner on a single glycoprotein, or it can be used to quantify glycosylation in a glycoprotein mixture. During mixture analysis, the method can discriminate between changes in glycosylation of a given protein, and changes in the glycoprotein's concentration in the mixture. This method is useful for quantitative analyses in biochemical studies of glycoproteins, where changes in glycosylation composition can be linked to functional differences; it could also be implemented in the pharmaceutical industry, where glycosylation profiles of glycoprotein-based therapeutics must be quantified. Finally, quantification of glycopeptides is an important aspect of glycopeptide-based biomarker discovery, and our quantitative approach could be a valuable asset to this field as well.

**Chapter 5** describes an approach for using circular dichroism (CD) spectroscopy to detect changes in structure and thermal stability on high mannose containing *N*-linked glycoproteins after glycan removal. This work is potentially useful as a pre-screening technique for identifying glycoproteins that do not require glycosylation once

a protein is properly folded, thereby allowing protein expression to be performed in cell lines, such as yeast and insect cells. These cell lines produce proteins with the yields necessary for mass production of therapeutic protein drugs, but do not produce the human-like *N*-linked glycosylation necessary for therapeutic effects in the body. Glycoproteins that “pass” the CD spectroscopy test (i. e. no measurable changes in structure and thermal stability are detected upon deglycosylation) can then be targeted for further functional assays, to confirm that deglycosylation does not affect protein function. This approach for screening the effects of glycosylation was first implemented on a model glycoprotein (ribonuclease B) containing one *N*-linked glycosylation site with high mannose type glycans. CD structure and thermal stability scans were taken before and after deglycosylation with the enzyme Peptide-N-glycosidase F. Mass spectrometric analysis was also performed to ensure that the deglycosylation reaction went to completion. CD secondary structural scans on ribonuclease B resulted in identical percentages for  $\alpha$ -helical and  $\beta$ -sheet content and the calculated  $T_m$  for thermal stability was within 0.2 °C for the glycosylated and deglycosylated forms of ribonuclease B. In summary, these results indicate that little or no change was observed in the structure and stability upon deglycosylation. This procedure was then applied to a potential therapeutic glycoprotein, human lysyl oxidase-like 2 (hLOXL2). Mass spectrometric analysis of hLOXL2 indicated that the deglycosylation reaction with PNGase F went to completion, and similarly to ribonuclease B, there was essentially no change in structure detected by CD spectroscopy upon deglycosylation of hLOXL2. Thermal stability analysis was inconclusive, however, because rather than unfolding, hLOXL2 precipitated out of solution as the temperature was increased.

## 1.7 References

- (1) Apweiler, R.; Hermjakob, H.; Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* **1999**, *1473*, 4-8.
- (2) Häggglund, P.; Bunkenborg, J.; Elortza, F.; Jensen, O. N.; Roepstorff, P. J. A new strategy for identification of *N*-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *Proteome Res.* **2004**, *3*, 556-566.
- (3) Budnik, B. A.; Lee, R. S. and Steen, J. A. J. Global methods for protein glycosylation analysis by mass spectrometry. *Biochim. Biophys. Acta*, **2006**, *1764*, 1870-1880.
- (4) Pan, S.; Chen, R.; Aebersold, R.; Brentnall, T. A. Mass spectrometry based glycoproteomics— from a proteomics perspective. *Mol. Cell. Proteomics* **2011**, *10*, Epub ahead of print.
- (5) Dalpathado, D. S. and Desaire, H. Glycopeptide analysis by mass spectrometry. *Analyst*, **2008**, *133*, 731-738.
- (6) Haltiwanger, R. S. and Lowe, J. B. Role of glycosylation in development. *Annu. Rev. Biochem.*, **2004**, *73*, 491-537.
- (7) Varki, A. Biological roles of oligosaccharides – all of the theories are correct. *Glycobiology*, **1993**, *3*, 97-130.
- (8) Morelle, W.; Canis, K.; Chirat, F.; Faid, V.; Michalski, J. –C. The use of mass spectrometry for the proteomic analysis of glycosylation. *Proteomics* **2006**, *6*, 3993-4015.
- (9) Helenius, A.; Aebi, M. Intracellular functions of *N*-linked glycans. *Science* **2001**, *291*, 2364-2369.
- (10) Fong, S. F. T.; Dietzsch, E.; Fong, K. S. K.; Hollosi, P.; Asuncion, L.; He, Q.; Parker, M. I.; Csiszar, K. Lysyl oxidase-like 2 expression is increased in colon and esophageal tumors and associated with less differentiated colon tumors. *Genes, Chromosomes & Cancer* **2007**, *46*, 644-655.
- (11) Saldova, R.; Wormald, M. R.; Dwek, R. A.; Rudd, P. M. Glycosylation changes on serum glycoproteins in ovarian cancer may contribute to disease pathogenesis. *Disease Markers* **2008**, *25*, 219-232.
- (12) Arnold, J. N.; Saldova, R.; Hamid, U. M. A.; Rudd, P. M. Evaluation of the serum *N*-linked glycome for the diagnosis of cancer and chronic inflammation. *Proteomics* **2008**, *8*, 3284-3293.

- (13) Dingermann, T. Recombinant therapeutic proteins: production platforms and challenges. *Biotechnol. J.* **2008**, *3*, 90-97.
- (14) Verma, R.; Boleti, E.; George, A. J. T. Antibody engineering: comparison of bacterial, yeast, insect, and mammalian expression systems. *J. Immunol. Methods* **1998**, *216*, 165-181.
- (15) Walsh, G.; Jefferis, R. Post-translational modifications in the context of therapeutic proteins. *Nature Biotechnol.* **2006**, *24*, 1241-1252.
- (16) Tissot, B.; North, S. J.; Ceroni, A.; Pang, P. -C.; Panico, M.; Rosati, F.; Capone, A.; Haslam, S. M.; Dell, A. and Morris, H. R. Glycoproteomics: past, present, and future. *FEBS Lett.*, **2009**, *583*, 1728-1735.
- (17) Dell, A. and Morris, H. R. Glycoprotein structure determination mass spectrometry. *Science*, **2001**, *291*, 2351-2356.
- (18) Dennis, J. W.; Granovsky, M. and Warren, C. E. Protein glycosylation in development and disease. *BioEssays*, **1999**, *21*, 412-421.
- (19) Vance, B. A.; Wu, W.; Ribaud, R. K.; Segal, D. M.; Kearse, K. P. Multiple dimeric forms of human CD69 result from differential addition of *N*-glycans to typical (Asn-X-Ser/Thr) and Atypical (Asn-X-Cys) glycosylation motifs. *J. Biol. Chem.*, **1997**, *272*, 23117-23122.
- (20) Spiro, R. G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*, **2002**, *12*, 43R-56R.
- (21) North, S. J.; Hitchen, P. G.; Haslam, S. M. and Dell, A. Mass spectrometry in the analysis of *N*-linked and *O*-linked glycans. *Curr. Opin. Struct. Biol.*, **2009**, *19*, 498-506.
- (22) Novotny, M. V. and Mechref, Y. New hyphenated methodologies in high-sensitivity glycoprotein analysis. *J. Sep. Sci.*, **2005**, *28*, 1956-1968.
- (23) Ramachandran, P.; Boontheung, P.; Pang, E.; Yan, W.; Wong, D. T. and Loo, J. A. Comparison of *N*-linked glycoproteins in human whole saliva, parotid, submandibular, and sublingual glandular secretions identified using hydrazide chemistry and mass spectrometry. *Clin. Proteomics*, **2008**, *4*, 80-104.
- (24) Sun, B.; Ranish, J. A.; Utleg, A. G.; White, J. T.; Yan, X.; Lin, B. and Hood, L. Shotgun glycopeptide capture approach coupled with mass spectrometry for comprehensive glycoproteomics. *Mol. Cell. Proteomics*, **2007**, *6*, 141-149.

- (25) Wollscheid, B.; Bausch-Fluck, D.; Henderson, C.; O'Brien, R.; Bibel, M.; Schiess, R.; Aebersold, R. and Watts, J. D. Mass-spectrometric identification and relative quantification of *N*-linked cell surface glycoproteins. *Nat. Biotechnol.*, **2009**, *27*, 378-386.
- (26) Ueda, K.; Katagiri, T.; Shimada, T.; Irie, S.; Sato, T. -A.; Nakamura, Y. and Daigo, Y. Comparative profiling of serum glycoproteome by sequential purification of glycoproteins and 2-nitrobenzensulfonyl (NBS) stable isotope labeling: a new approach for the novel biomarker discovery for cancer. *J. Proteome Res.*, **2007**, *6*, 3475-3483.
- (27) Schiess, R.; Mueller, L. N.; Schmidt, A.; Mueller, M.; Wollscheid, B. and Aebersold, R. Analysis of cell surface proteome changes via label-free, quantitative mass spectrometry. *Mol. Cell. Proteomics*, **2009**, *8*, 624-638.
- (28) Hulsmeier, A. J.; Paesold-Burda, P. and Hennet, T. *N*-glycosylation site occupancy in serum glycoproteins using multiple reaction monitoring liquid chromatography-mass spectrometry. *Mol. Cell. Proteomics*, **2007**, *6*, 2132-2138.
- (29) Liu, Z.; Cao, J.; He, Y.; Qiao, L.; Xu, C.; Lu, H. and Yang, P. Tandem <sup>18</sup>O stable isotope labeling for quantification of *N*-glycoproteome. *J. Proteome Res.*, **2010**, *9*, 227-236.
- (30) Balaguer, E.; Demelbauer, U.; Pelzing, M.; Sanz-Nebot, V.; Barbosa, J. and Neusüß, C. Glycoform characterization of erythropoietin combining glycan and intact protein analysis by capillary electrophoresis electrospray time-of-flight mass spectrometry. *Electrophoresis*, **2006**, *27*, 2638-2650.
- (31) Higgins, E. Carbohydrate analysis throughout the development of a protein therapeutic. *Glycoconjugate J.*, **2010**, *27*, 211-225.
- (32) Beck, A.; Wagner-Rousset, E.; Bussat, M. -C.; Lokteff, M.; Klinguer-Hamour, C.; Haeuw, J. -F.; Goetsch, L.; Wurch, T.; van Dorsselaer, A. and Corvaia, N. Trends in glycosylation, glycoanalysis and glycoengineering of therapeutic antibodies and Fc-fusion proteins. *Curr. Pharm. Biotechnol.*, **2008**, *9*, 482-501.
- (33) Mitra, N.; Sinha, S.; Ramya, T. N. C.; Surolia, A. *N*-linked oligosaccharides as outfitters for glycoprotein folding, form and function. *Trends Biochem. Sci.* **2006**, *31*, 156-163.
- (34) Gemmill, T. R.; Trimble, R. B. Overview of *N*- and *O*-linked oligosaccharide structures found in various yeast species. *Biochim. Biophys. Acta, Gen. Subj.* **1999**, *1426*, 227-237.
- (35) Ip, W. K. E.; Takahashi, K.; Ezekowitz, R. A.; Stuart, L. M. Mannose-binding lectin and innate immunity. *Immunol. Rev.* **2009**, *230*, 9-21.

- (36) Degn, S. E.; Jensenius, J. C.; Bjerre, M. The lectin pathway and its implications in coagulation, infections and auto-immunity. *Curr. Opin. Organ Transplant* **2011**, *16*, 21-27.
- (37) De Pourcq, K.; De Schutter, K.; Callewaert, N. Engineering of glycosylation in yeast and other fungi: current state and perspectives. *Appl. Microbiol. Biotechnol.* **2010**, *87*, 1617-1631.
- (38) Liu, C.; Cashion, L. M.; Pu, H. Protein expression both in mammalian cell lines and in yeast *Pichia pastoris* using a single expression plasmid. *Biotechniques* **1998**, *24*, 266-268.
- (39) Gerngross, T. U. Advances in the production of human therapeutic proteins in yeast and filamentous fungi. *Nat. Biotechnol.* **2004**, *22*, 1409-1414.
- (40) Cointe, D; Beliard, R; Jorieux, S; Leroy, Y; Glacet, A; Verbert, A; Bourel, D Chirat, F. Unusual *N*-glycosylation of a recombinant human erythropoietin expressed in a human lymphoblastoid cell line does not alter its biological properties. *Glycobiology* **2000**, *10*, 511-519.
- (41) Hokke, C. H.; Bergwerff, A. A.; Vandedem, G. W. K.; Vanoostrum, J.; Kamerling, J. P.; Vliegenthart, J. F. G. Sialylated carbohydrate chains of recombinant human glycoproteins expressed in Chinese-hamster ovary cells contain traces of *N*-glycolylneuraminic acid. *FEBS Lett.* **1990**, *275*, 9-14.
- (42) Baker, K. N.; Rendall, M. H.; Hills, A. E.; Hoare, M.; Freedman, R. B.; James, D. C. Metabolic control of recombinant protein *N*-glycan processing in NS0 and CHO cells. *Biotechnol. Bioeng.* **2001**, *73*, 188-202.
- (43) Pantazaki, A.; Taverna, M.; Vidal-Madjar, C. Recent advances in the capillary electrophoresis of recombinant glycoproteins. *Anal. Chim. Acta* **1999**, *383*, 137-156.
- (44) Wührer, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J. Chromatography B* **2007**, *849*, 115-128.
- (45) Ongay, S.; Neusüß, C. Isoform differentiation of intact AGP from human serum by capillary electrophoresis-mass spectrometry. *Anal. Bioanal. Chem.* **2010**, *398*, 845-855.
- (46) Steen, H.; Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 699-711.

- (47) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198-207.
- (48) Shevchenko, A.; Jensen, O. N.; Podtelejnikov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Shevchenko, A.; Boucherie, H.; Mann, M. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci.* **1996**, *93*, 14440-14445.
- (49) Meyer, B.; Papatotiriou, D. G.; Karas, M. 100 % protein sequence coverage: a modern form of surrealism in proteomics. *Amino Acids* **2010**, Epub ahead of print.
- (50) Barnes, C. A. S.; Lim, A. Applications of mass spectrometry for the structural characterization of recombinant protein pharmaceuticals. *Mass Spectrometry Reviews* **2007**, *26*, 370-388.
- (51) Ren, D.; Julka, S.; Inerowicz, H. D.; Regnier, F. E. Enrichment of cysteine-containing peptides from tryptic digests using a quaternary amine tag. *Anal. Chem.* **2004**, *76*, 4522-4530.
- (52) Rabilloud, T. Solubilization of proteins for electrophoretic analyses. *Electrophoresis* **1996**, *17*, 813-829.
- (53) Tiefenbach, K. -J.; Durchschlag, H.; Jaenicke, R. Hydrodynamic and spectroscopic analysis of the denaturation of serum albumin induced by guanidinium chloride and sodium dodecyl sulfate. *Prog. Colloid Polym. Sci.* **2004**, *127*, 136-147.
- (54) Yu, Y. -Q.; Gilar, M.; Lee, P. J.; Bouvier, E. S. P.; Gebler, J. C. Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Anal. Chem.* **2003**, *75*, 6023-6028.
- (55) Getz, E. B.; Xiao, M.; Chakrabarty, T.; Cooke, R.; Selvin, P. R. A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. *Anal. Biochem.* **1999**, *273*, 73-80.
- (56) Madera, M.; Mechref, Y.; Klouckova, I. and Novotny, M. V. Semiautomated high-sensitivity profiling of human blood serum glycoproteins through lectin preconcentration and multidimensional chromatography/tandem mass spectrometry. *J. Proteome Res.*, **2006**, *5*, 2348-2363.
- (57) Jung, K., Cho, W. and Regnier, F. E. Glycoproteomics of plasma based on narrow selectivity lectin affinity chromatography. *J. Proteome Res.*, **2009**, *8*, 643-650.
- (58) Kaji, H. and Isobe, T. Liquid chromatography/mass spectrometry (LC/MS)-based glycoproteomics technologies for cancer biomarker discovery. *Clin. Proteomics*, **2008**, *4*, 14-24.

- (59) Tao, S. -C.; Li, Y.; Zhou, J.; Qian, J.; Schnaar, R. L.; Zhang, Y.; Goldstein, I. J.; Zhu, H. and Schneck, J. P. Lectin microarrays identify cell-specific and functionally significant cell surface glycan markers. *Glycobiology*, **2008**, *18*, 761-769.
- (60) Hill, J. J.; Tremblay, T. -L.; Cantin, C.; O'Connor-McCourt, M.; Kelly, J. F. and Lenferink, A. E. G. Glycoproteomic analysis of two mouse mammary cell lines during transforming growth factor (TGF)-beta induced epithelial to mesenchymal transition. *Proteome Sci.*, **2009**, *7*, 1-17.
- (61) Qiu, R. and Regnier, F. E. Comparative glycoproteomics of *N*-linked complex-type glycoforms containing sialic acid in human serum. *Anal. Chem.*, **2005**, *77*, 7225-7231.
- (62) Lee, H. -J.; Na, K.; Choi, E. -Y.; Kim, K. S.; Kim, H. and Paik, Y. -K. Simple method for quantitative analysis of *N*-linked glycoproteins in hepatocellular carcinoma specimens. *J. Proteome Res.*, **2010**, *9*, 308-318.
- (63) Hill, J. J.; Moreno, M. J.; Lam, J. C. Y.; Haqqani, A. S. and Kelly, J. F. Identification of secreted proteins regulated by cAMP in glioblastoma cells using glycopeptide capture and label-free quantification. *Proteomics*, **2009**, *9*, 535-549.
- (64) Zhang, H.; Li, X.; Martin, D. B. and Aebersold, R. Identification and quantification of *N*-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.*, **2003**, *21*, 660-666.
- (65) Zhou, L.; Beuerman, R. W.; Chew, A. P.; Koh, S. K.; Cafaro, T. A.; Urrets-Zavalía, E.A.; Urrets-Zavalía, J.A.; Li, S.F.Y. and Serra, H.M. Quantitative analysis of *N*-linked glycoproteins in tear fluid of climatic droplet keratopathy by glycopeptide capture and iTRAQ. *J. Proteome Res.*, **2009**, *8*, 1992-2003.
- (66) Wührer, M.; Stam, J. C.; van de Geijn, F. E.; Koeleman, C. A. M.; Verrips, C. T.; Dolhain, R. J. E. M., Hokke, C. and Deelder, A. M. Glycosylation profiling of immunoglobulin G (IgG) subclasses from human serum. *Proteomics*, **2007**, *7*, 4070-4081.
- (67) Selman, M. H. J.; McDonnell, L. A.; Palmblad, M.; Ruhaak, R.; Deelder, A. M. and Wührer, M. Immunoglobulin G glycopeptide profiling by matrix-assisted laser desorption ionization Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.*, **2010**, *82*, 1073-1081.
- (68) Rebecchi, K. R.; Wenke, J. L.; Go, E. P. and Desaire, H. Label-free quantitation: a new glycoproteomics approach. *J. Am. Soc. Mass Spectrom.*, **2009**, *20*, 1048-1059.

- (69) Wada, Y., Tajiri, M. and Yoshida, S. Hydrophilic affinity isolation and MALDI multiple-stage tandem mass spectrometry of glycopeptides for glycoproteomics. *Anal. Chem.*, **2004**, *76*, 6560-6565.
- (70) Zhang, Y.; Go, E. P. and Desaire, H. Maximizing coverage of glycosylation heterogeneity in MALDI-MS analysis of glycoproteins with up to 27 glycosylation sites. *Anal. Chem.*, **2008**, *80*, 3144-3158.
- (71) Ding, W.; Nothhaft, H.; Szymanski, C. M. and Kelly, J. Identification and quantification of glycoproteins using ion-pairing normal-phase liquid chromatography and mass spectrometry. *Mol. Cell. Proteomics*, **2009**, *8*, 2170-2185.
- (72) Alvarez-Manilla, G.; Atwood, J. A., III; Guo, Y.; Warren, N. L.; Orlando, R. and Pierce, M. Tools for glycoproteomic analysis: size exclusion chromatography facilitates identification of tryptic glycopeptides with *N*-linked glycosylation sites. *J. Proteome Res.*, **2006**, *5*, 701-708.
- (73) Joenväärä, S; Ritamo, I; Peltoniemi, H and Renkonen, R. *N*-glycoproteomics – an automated workflow. *Glycobiology*, **2008**, *18*, 339-349.
- (74) Lewandrowski, U.; Zahedi, R. P.; Moebius, J.; Walter, U. and Sickmann, A. Enhanced *N*-glycosylation site analysis of sialoglycopeptides by strong cation exchange prefractionation applied to platelet plasma membranes. *Mol. Cell. Proteomics*, **2007**, *6*, 1933-1941.
- (75) El-Aneed, A.; Cohen, A.; Banoub, J. Mass spectrometry, review of the basics: electrospray, MALDI, and commonly used mass analyzers. *Appl. Spectrosc. Rev.* **2009**, *44*, 210-230.
- (76) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Cooks, R. G. The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **2005**, *40*, 430-443.
- (77) Domon, B.; Aebersold, R. Mass spectrometry and protein analysis. *Science* **2006**, *312*, 212-217.
- (78) March, R. E. An introduction to quadrupole ion trap mass spectrometry. *J. Mass Spectrom.* **1997**, *32*, 351-369.
- (79) Amster, I. J. Fourier transform mass spectrometry. *J. Mass Spectrom.* **1996**, *31*, 1325-1337.
- (80) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev.* **1998**, *17*, 1-35.

- (81) Payne, A. H.; Glish, G. L. Tandem mass spectrometry in quadrupole ion trap and ion cyclotron resonance mass spectrometers. *Meth. Enzymol.* **2005**, *402*, 109-148.
- (82) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray ionization – principles and practice. *Mass Spectrom. Rev.* **1990**, *9*, 37-70.
- (83) Kebarle, P. A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J. Mass Spectrom.* **2000**, *35*, 804-817.
- (84) Cole, R. B. Some tenets pertaining to electrospray ionization mass spectrometry. *J. Mass Spectrom.* **2000**, *35*, 763-772.
- (85) Constantopoulos, T. L.; Jackson, G. S.; Enke, C. G. Effects of salt concentration on analyte response using electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 625-634.
- (86) Jackson, A. U.; Talaty, N.; Cooks, R. G.; Van Berkel, G. J. Salt tolerance of desorption electrospray ionization (DESI). *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 2218-2225.
- (87) Hager, J. W. A new linear ion trap mass spectrometer. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 512-526.
- (88) Schwartz, J. C.; Senko, M. W.; Syka, J. E. P. A two-dimensional quadrupole ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 659-669.
- (89) Douglas, D. J.; Frank, A. J.; Mao, D. Linear ion traps in mass spectrometry. *Mass Spectrom. Rev.* **2005**, *24*, 1-29.
- (90) Second, T. P.; Blethrow, J. D.; Schwartz, J. C.; Merrihew, G. E.; MacCoss, M. J.; Swaney, D. L.; Russell, J. D.; Coon, J. J. Zabrouskov, V. Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures. *Anal. Chem.* **2009**, *81*, 7757-7765.
- (91) Olsen, J. V.; Schwartz, J. C.; Griep-Raming, J.; Nielsen, M. L.; Damoc, E.; Denisov, E.; Lange, O.; Remes, P.; Taylor, D.; Splendore, M.; Wouters, E. R.; Senko, M.; Makarov, A.; Mann, M.; Horning, S. A dual pressure linear ion trap orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* **2009**, *8*, 2759-2769.
- (92) Kim, T.; Tolmachev, A. V.; Harkewicz, R.; Prior, D. C.; Anderson, G.; Udseth, H. R.; Smith, R. D. Design and implementation of a new electrodynamic ion funnel. *Anal. Chem.* **2000**, *72*, 2247-2255.

- (93) Mann, M.; Kelleher, N. L. Precision proteomics: the case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 18132-18138.
- (94) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551-3567.
- (95) Eng, J. K.; McCormack, A. L.; Yates, J. R. III An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976-989.
- (96) Craig, R.; Beavis, R. C. Tandem: matching proteins with tandem mass spectra. *Bioinformatics* **2004**, *20*, 1466-1467.
- (97) Chen, Y.; Zhang, J.; Xing, G.; Zhao, Y. Mascot-derived false positive peptide identifications revealed by manual analysis of tandem mass spectra. *J. Proteome Res.* **2009**, *8*, 3141-3147.
- (98) Go, E. P.; Rebecchi, K. R.; Dalpathado, D. S.; Bandu, M. L.; Zhang, Y. and Desaire, H. GlycoPep DB: a tool for glycopeptide analysis using a "smart search". *Anal. Chem.*, **2007**, *79*, 1708-1713.
- (99) Irungu, J.; Go, E. P.; Dalpathado, D. S. and Desaire, H. Simplification of mass spectral analysis of acidic glycopeptides using GlycoPep ID. *Anal. Chem.*, **2007**, *79*, 3065-3074.
- (100) Ozohanics, O.; Krenyacz, J.; Ludányi, K.; Pollreisz, F.; Vékey, K. and Drahos, L. GlycoMiner: a new software tool to elucidate glycopeptide composition. *Rapid Commun. Mass Spectrom.*, **2008**, *22*, 3245-3254.
- (101) Chaerkady, R.; Thuluvath, P. J.; Kim, M. -S.; Nalli, A.; Vivekanandan, P.; Simmers, J.; Torbenson, M. and Pendey, A. 18O Labeling for a quantitative proteomics analysis of glycoproteins in hepatocellular carcinoma. *Clin. Proteomics*, **2008**, *4*, 136-155.
- (102) Wada, Y.; Azadi, P.; Costello, C. E.; Dell, A.; Dwek, R. A.; Geyer, H.; Geyer, R.; Kakehi, K.; Karlsson, N. G.; Kato, K.; Kawasaki, N.; Khoo, K. H.; Kim, S.; Kondo, A.; Lattova, E.; Mechref, Y.; Miyoshi, E.; Nakamura, K.; Narimatsu, H.; Novotny, M. V.; Packer, N. H.; Perreault, H.; Peter-Katalinic, J.; Pohlentz, G.; Reinhold, V. N.; Rudd, P. M.; Suzuki, A. and Taniguchi, N. Comparison of the methods for profiling glycoprotein glycans - HUPO human disease glycomics/proteome initiative multi-institutional study. *Glycobiology*, **2007**, *17*, 411-422.

- (103) Irungu, J.; Go, E. P.; Zhang, Y.; Dalpathado, D. S.; Liao, H. -X.; Haynes, B. F. and Desaire, H. Comparison of HPLC/ESI-FTICR MS versus MALDI-TOF/TOF MS for glycopeptide analysis of a highly glycosylated HIV envelope glycoprotein. *J. Am. Soc. Mass Spectrom.*, **2008**, *19*, 1209-1220.
- (104) Atwood, J. A., III; Cheng, L.; Alvarez-Manilla, G.; Warren, N. L.; York, W. S. and Orlando, R. Quantitation by isobaric labeling: applications to glycomics. *J. Proteome Res.*, **2008**, *7*, 367-374.
- (105) Wiese, S.; Reidegeld, K. A.; Meyer, H. E. and Warscheid, B. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics*, **2007**, *7*, 340-350.
- (106) Ow, S. Y.; Salim, M.; Noirel, J.; Evans, C.; Rehman, I. and Wright, P. C. iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". *J. Proteome Res.*, **2009**, *8*, 5347-5355.
- (107) Madera, M.; Mann, B.; Mechref, Y. and Novotny, M. V. Efficacy of glycoprotein enrichment by microscale lectin affinity chromatography. *J. Sep. Sci.*, **2008**, *31*, 2722-2732.
- (108) Mann, B.; Madera, M.; Sheng, Q.; Tang, H.; Mechref, Y. and Novotny, M. ProteinQuant Suite: a bundle of automated software tools for label-free quantitative proteomics. *Rapid Commun. Mass Spectrom.*, **2008**, *22*, 3823-3834.
- (109) Zhang, H.; Yi, E. C.; Li, X.; Mallick, P.; Kelly-Spratt, K. S.; Masselon, C. D.; Camp, D. G., II; Smith, R. D.; Kemp, C. J. and Aebersold, R. High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. *Mol. Cell. Proteomics*, **2005**, *4*, 144-155.
- (110) Nana, K.; Satsuki, I.; Noritaka, H.; Akira, H.; Daisuke, T. and Teruhida, Y. Mass spectrometric analysis of carbohydrate heterogeneity for the characterization of glycoprotein-based products. *Trends Glycosci. Glycotechnol.*, **2008**, *20*, 97-116.
- (111) Thaysen-Andersen, M.; Mysling, S. and Hojrup, P. Site-specific glycoprofiling of N-linked glycopeptides using MALDI-TOF MS: strong correlation between signal strength and glycoform quantities. *Anal. Chem.*, **2009**, *81*, 3933-3943.
- (112) Tajiri, M., Kadoya, M. and Wada, Y. Dissociation profile of protonated fucosyl glycopeptides and quantitation of fucosylation levels of glycoproteins by mass spectrometry. *J. Proteome Res.*, **2009**, *8*, 688-693.
- (113) Kondo, A.; Miyamoto, T.; Yonekawa, O.; Giessing, A. M.; Østerlund, E. C. and Jensen, O. N. Glycopeptide profiling of beta-2-glycoprotein I by mass spectrometry reveals attenuated sialylation in patients with antiphospholipid syndrome. *J. Proteomics*, **2009**, *73*, 123-133.

- (114) Ivancic, M. M.; Gadgil, H. S.; Halsall, H. B. and Treuheit, M. J. LC/MS analysis of complex multiglycosylated human alpha(1)-acid glycoprotein as a model for developing identification and quantitation methods for intact glycopeptide analysis. *Anal. Biochem.*, **2010**, *400*, 25-32.
- (115) Uematsu, R.; Furukawa, J.; Nakagawa, H.; Shinohara, Y.; Deguchi, K.; Monde, K. and Nishimura, S. High throughput quantitative glycomics and glycoform-focused proteomics of murine dermis and epidermis. *Mol. Cell. Proteomics*, **2005**, *4*, 1977-1989.
- (116) Cooper, C. A., Gasteiger, E. and Packer, N. H. GlycoMod - A software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics*, **2001**, *1*, 340-349.
- (117) Wagner-Rousset, E.; Bednarczyk, A.; Bussat, M. -C.; Colas, O.; Corvaia, N.; Schaeffer, C.; Dorsselaer, A. V. and Beck, A. The way forward, enhanced characterization of therapeutic antibody glycosylation: comparison of three level mass spectrometry-based strategies. *J. Chromatogr. B*, **2008**, *872*, 23-37.
- (118) Sinz, A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom. Rev.* **2006**, *25*, 663-682.
- (119) Sours, K. M.; Kwok, S. C.; Rachidi, T.; Lee, T.; Ring, A.; Hoofnagle, A. N.; Resing, K. A.; Ahn, N. G. Hydrogen-exchange mass spectrometry reveals activation-induced changes in conformational mobility of p38a MAP kinase. *J. Mol. Biol.* **2008**, *379*, 1075-1093.
- (120) Konermann, L.; Tong, X.; Pan, Y. Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling and related approaches. *J. Mass Spectrom.* **2008**, *43*, 1021-1036.
- (121) Sajnani, G.; Pastrana, M. A.; Dynin, I.; Onisko, B.; Requena, J. R. Scrapie prion protein structural constraints obtained by limited proteolysis and mass spectrometry. *J. Mol. Biol.* **2008**, *382*, 88-98.
- (122) Wishart, D. NMR Spectroscopy and protein structure determination: applications to drug discovery and development. *Curr. Pharm. Biotechnol.* **2005**, *6*, 105-120.
- (123) Manjasetty, B. A.; Turnbull, A. P.; Panjikar, S.; Bussow, K.; Chance, M. R. Automated technologies and novel techniques to accelerate protein crystallography for structural genomics. *Proteomics* **2008**, *8*, 612-625.
- (124) Kelly, S. M.; Price, N. C. The use of circular dichroism in the investigation of protein structure and function. *Curr. Protein Pept. Sci.* **2000**, *1*, 349-384.

- (125) Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **2006**, *1*, 2876-2890.
- (126) Whitmore, L.; Wallace, B. A. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **2004**, *32*, W668-W673.
- (127) Benjwal, S.; Verma, S.; Rohm, K. -H.; Gursky, O. Monitoring protein aggregation during thermal unfolding in circular dichroism experiments. *Protein Sci.* **2006**, *15*, 635-639.
- (128) Chiti, F.; van Nuland, N. A. J.; Taddei, N.; Magherini, F.; Stefani, M.; Ramponi, G.; Dobson, C. M. Conformational stability of muscle acylphosphatase: the role of temperature, denaturant concentration, pH. *Biochemistry* **1998**, *37*, 1447-1455.

## CHAPTER 2

# DETERMINATION OF AN IDEAL PROTEASE DIGESTION PROCEDURE FOR LC/MS ANALYSIS OF SOLUBLE PROTEINS AND GLYCOPROTEINS

### 2.1 Introduction

Recombinant proteins are designed and produced for a variety of reasons, most notably for use as therapeutic agents<sup>1-3</sup> and vaccine candidates.<sup>4-8</sup> Utilizing recombinant DNA technology and genetic engineering to produce a wide-range of proteins has been shown to be beneficial in the development of various pharmaceuticals, as demonstrated by pharmacological studies involving interferons,<sup>2</sup> reproductive hormones,<sup>9</sup> and monoclonal antibodies.<sup>1, 10</sup> More recently, the use of recombinant proteins has focused on the development of potential biopharmaceutical protein drugs that contain novel post-translational modifications (PTMs), in an effort to alter the solubility, efficacy, half-life, and in-vivo clearance rate in comparison to the characteristics of the corresponding native protein sequences.<sup>1, 10, 11</sup> For potential protein pharmaceuticals, full characterization including PTMs is critical in the drug development process.

Mass spectrometry (MS) is an important tool for protein identification<sup>10, 12</sup> and quantification,<sup>11, 13-15</sup> and is especially powerful in the analysis of proteins. MS is perhaps the most commonly utilized technique for primary sequence characterization of recombinant proteins, as well as for the detection of PTMs.<sup>2, 16</sup> Common preparatory steps leading to MS involving proteins (recombinant or otherwise) include a protease digestion procedure, where the primary protein sequence is cleaved into peptides prior to MS analysis. The peptides formed from digestion typically retain their PTMs, thereby

allowing MS and tandem MS experiments to detect these modifications to be used to detect these modifications, and determine the PTM location on the protein.<sup>17-18</sup> For detection of all the different PTMs present in proteins, it is advantageous to achieve full protein sequence coverage.<sup>19</sup> Therefore, efficient protease digestions are crucial in order to achieve accurate characterization and full detection for peptides containing PTMs.<sup>3, 20</sup>

In order to develop optimized methods for mass spectrometric analysis of peptides and PTMs on proteins, previous work has focused on several different stages of the protein preparation process ranging from evaluation of different types of mass spectrometers<sup>21, 22</sup> or separation methods,<sup>20</sup> to comparing specific aspects of a protease digestion procedure.<sup>23-26</sup> Inefficient protease digestion procedures will inevitably result in poor mass spectrometry data, no matter how efficient the separation methods or mass spectrometer parameters are.<sup>23, 25-28</sup> In many instances, if a protein is not properly unfolded prior to addition of protease, the protease will not efficiently cleave the protein, and therefore, MS data interpretation suffers because several peptides, consisting of different degrees of enzymatic mis-cleavage, will be diluted over multiple *m/z* values, and those peptides that are difficult to ionize will not be detected, leading to lower protein sequence coverage.<sup>10, 19, 23, 27</sup>

Proteolytic digestion methods consist of several procedural steps prior to the addition of an enzyme to cleave a protein into peptides, including: denaturation, or unfolding of protein(s), reduction of disulfide bonds, and subsequent alkylation, or “capping,” of reduced cysteine residues. Previous studies have focused on each of the individual steps in the protease digestion process, and hence much of the previous

optimization research has concentrated on the individual parameters, such as denaturation.<sup>24, 25, 27, 29, 30</sup> Additionally, most of the recent emphasis has centered on the MS analysis of membrane proteins<sup>27, 29</sup> and cellular proteome elucidation.<sup>24, 25</sup> Analysis of cellular proteomes incorporates numerous membrane-bound proteins; thus, it was no surprise that published experimental findings were favorable toward MS friendly detergents, such as Rapigest<sup>TM</sup> SF, as detergents are known to aid in the solubilization of membrane-bound proteins.<sup>29</sup> It is unknown whether these conditions would be optimal for recombinantly expressed proteins, which are typically secreted, post-translationally modified, and not membrane-bound.

Due to the common presence of disulfide bonds present in many proteins analyzed by MS, research also has focused on optimizing conditions for reduction<sup>30, 31</sup> and alkylation.<sup>23, 32, 33</sup> When reduction and alkylation are incomplete, a lower signal to noise ratio is often seen, as peaks may be present in the MS data that correspond to both derivatized and underivatized peptides.<sup>15</sup> Thus, peptides with already low ionization efficiencies, such as glycosylated peptides, may not be detected because splitting peptide ions over multiple  $m/z$  values can result in ion abundance too low for MS detection.<sup>15</sup> Moreover, over alkylation of peptides, or alkylation of the N-terminus or other amino acid side chains besides cysteine, can also occur when alkylation is allowed to incubate with the sample long periods of time, such as when alkylating agent is not removed during the protease procedural step.<sup>34</sup> Therefore, optimization of reduction and alkylation conditions is also essential to maximize protein sequence coverage by MS.

The work described herein focuses on designing an ideal protease digestion protocol for readily soluble proteins containing both disulfide bonds and *N*-linked glycosylation, with a specific goal of identifying reaction conditions yielding the highest protein sequence coverage, as well as effective detection of *N*-linked glycosylation. Multiple parameters in the protease digestion process were assessed by developing several different reaction conditions on a model protein for determination of the most optimal digestion strategy.

## **2.2 Experimental**

### **2.2.1 Materials and Reagents**

Human apo-transferrin (transferrin), urea, tris(2-carboxyethyl)phosphine (TCEP), dithiothreitol (DTT), iodoacetamide (IAA), hydrochloric acid, acetic acid, and formic acid were purchased from Sigma (St. Louis, MO). Ammonium bicarbonate ( $\text{NH}_4\text{HCO}_3$ ) was purchased from Fluka (Milwaukee, WI). HPLC grade acetonitrile ( $\text{CH}_3\text{CN}$ ) was purchased from Fisher Scientific (Fair Lawn, NJ). RapiGest™ SF was purchased from Waters Corporation (Milford, MA). PNGase F from *Flavobacterium meningosepticum* was purchased from New England BioLabs (Ipswich, MA). Sequencing Grade Modified Trypsin was purchased from Promega (Madison, WI). Ultrapure water was obtained from an in-house Millipore Direct-Q® UV 3 system (Billerica, MA) with a resistance greater than 18 M $\Omega$ .

### **2.2.2 Glycoprotein Protease Digestion Denatured with RapiGest™ SF**

Human apo-transferrin (~ 10 mg/mL) was dissolved in 0.1% RapiGest™ SF containing 50 mM  $\text{NH}_4\text{HCO}_3$ , pH 7.8 buffer. For reduction of disulfide bonds, either dithiothreitol (DTT) or tris(2-carboxyethyl)phosphine (TCEP) was added. Table 1 shows

the concentrations and type of reducing agent added for each of the 7 different reaction conditions. Samples were incubated for 45 min. at 60 °C. Iodoacetamide (IAA) was added as the alkylating agent for 60 min. at room temperature in the dark. As shown in Table 1, reaction condition 3 contained a step where DTT was added to quench the alkylation reaction after IAA had incubated with the protein samples for 1 hr. in the dark. Trypsin was added at a 1:30 (w/w) enzyme:protein ratio, and all samples were incubated at 37 °C for 18 hrs. HCl was added to a final concentration of 50 mM to stop the tryptic digestion, as well as to provide an acidic solution for RapiGest™ SF precipitation. Samples were re-incubated at 37 °C for an additional 45 min., then centrifuged to pellet out RapiGest™ SF. The supernatant was removed and stored at -20 °C until analysis with mass spectrometry.

**Table 1.** Protease digestion preparation conditions for human apo-transferrin.

Reaction Conditions	Denaturing Agent	Reducing Agent	Alkylating Agent	Quenching Agent
#1	0.1% RapiGest™ SF	5 mM DTT	15 mM IAA	
#2	6 M Urea	5 mM DTT	15 mM IAA	
#3	0.1% RapiGest™ SF	10 mM DTT	15 mM IAA	20 mM DTT
#4	0.1% RapiGest™ SF	5 mM TCEP	10 mM IAA	
#5	6 M Urea	10 mM DTT	15 mM IAA	20 mM DTT
#6	6 M Urea	5 mM TCEP	10 mM IAA	
#7	6 M Urea	5 mM TCEP	10 mM IAA	10 mM DTT

### 2.2.3 Glycoprotein Protease Digestion Denatured with urea

Urea (6 M) was added to transferrin (~ 10 mg/mL), which had been dissolved in 50 mM NH<sub>4</sub>HCO<sub>3</sub>, pH 7.8. DTT or TCEP was added to reduce the disulfide bonds (see Table 1), and samples were incubated at room temperature for 1 hr before IAA was

added to alkylate Cys residues by incubation in the dark for 1 hr. Reaction conditions 5 and 7 contained an additional step where DTT was added to quench the alkylation reaction after IAA had been allowed to incubate in the dark for 1 hr. The  $\text{NH}_4\text{HCO}_3$  buffer was added to dilute the urea concentration to 1 M before the addition of trypsin, at a 1:30 (w/w) enzyme:protein ratio. Samples were then incubated at 37 °C for 18 hrs. To stop the trypsin reaction, 1  $\mu\text{L}$  acetic acid was added per 100  $\mu\text{L}$  of solution before storing samples at -20 °C until ready for mass spectrometry.

#### **2.2.4 Liquid chromatography/mass spectrometry**

The tryptic digest samples (5  $\mu\text{L}$ , ~ 15  $\mu\text{g}$ ) were injected onto a reversed phase ( $\text{C}_{18}$ ) column {300  $\mu\text{m}$  i.d. x 5 cm, 3  $\mu\text{m}$  particle size, CVC MicroTech, (Fontana, CA)} using a Dionex UltiMate capillary HPLC system (Sunnyvale, CA) containing a FAMOS well plate autosampler. The HPLC system was connected to an ESI-LIT-FTICRMS {electrospray ionization – linear ion trap – Fourier transform ion cyclotron resonance mass spectrometer, ThermoScientific (San Jose, CA)} containing a 7 Tesla actively shielded magnet. Mobile phase solvents A and B were composed of  $\text{H}_2\text{O}$  and  $\text{CH}_3\text{CN}$ , respectively, where both solvents contained 0.1 % formic acid. To elute peptides from the column, the solvent conditions were held initially for 5 min at 5 % B, linearly increased to 40 % B over 50 min, further increased to 90 % B in 10 min, then held at 90 % B for 10 min, and lastly the column was re-equilibrated before the next injection. To ensure that no sample carryover was detected in the MS data, a 30 min wash followed by a blank injection was implemented between each sample run. The mass spectrometer was set to the following parameters for all samples: ESI source voltage was 2.8 kV, capillary voltage offset was 47 V, capillary temperature was 200 °C, FT-ICR

resolution was set to 25,000 for  $m/z$  400, and MS/MS data were collected in a data dependent manner by selecting the 5 most intense ions in an FT-ICR MS<sup>1</sup> scan for collision induced dissociation (CID), where a collision energy of 30 % and a dynamic exclusion window of 3 min was utilized.

### **2.2.5 Data analysis**

Peptides were analyzed with Mascot (Matrix Science, London, U.K., version 2.2.04), as well as confirmed manually. Peak lists were extracted from raw files using BioWorks Browser (ThermoScientific, Version 3.5). The following parameters were searched in the DTA files: 1) enzyme, trypsin; 2) up to 2 missed cleavages; 3) fixed modification, Cysteine carbamidomethyl; 4) variable modifications: methionine oxidation, carbamyl, N-terminal carbamidomethyl; 5) peptides tolerance, 0.8 Da; and 6) MS/MS tolerance, 0.4 Da. If searching for under, or incomplete, alkylation of Cys residues, then the fixed modification of carbamidomethyl was not added to the parameters searched. The database searched for both transferrin and hLOXL2 was SwissProt 2010, taxonomy Mammalia. All peptides identified from Mascot were manually checked to corroborate the presence of b and y fragmentation ions in the MS/MS data. Some peptides, and all glycopeptides, could not be detected using Mascot. In these cases, manual interpretation of the MS<sup>1</sup> and MS/MS data was done to identify peptides and glycopeptides by their characteristic fragmentation patterns (b and y ions for peptides, and losses of sugar residues for glycopeptides). To be considered as a detected peptide through manual inspection, two criteria must be met: MS/MS data illustrating appropriate fragmentation was required, and the monoisotopic peak present in the mass spectrum had to be within 20 ppm mass error.

## **2.3 Results and discussion**

The goal of the study described herein is to determine ideally suited reaction conditions for proteolytic digestion of recombinant proteins and glycoproteins. Protease digestion protocols include denaturation (unfolding) of a protein, reduction and alkylation of Cys residues (when present), and addition of a proteolytic enzyme to cleave a protein(s) into peptides. It is imperative to have an effective digestion method when performing mass spectrometry experiments on peptides, especially in those cases when analyzing for post-translational modifications and/or characterizing novel proteins, as inefficient protein digests will inevitably lead to poor MS data. To achieve the objectives, various reaction conditions are tested on a model protein, so that the optimal protein digest condition could be identified. Human apo-transferrin (transferrin) was chosen as a model glycoprotein for the initial testing of different digestion procedures because transferrin possesses a high number of co- and post-translational modifications, including 19 disulfide bonds and two *N*-linked glycosylation sites. Figure 1 shows the amino acid sequence for transferrin.

### **2.3.1 Survey of various protease digestion conditions**

Six different reaction conditions (Table 1) were initially tested on transferrin and the resulting MS and MS/MS data was assessed for the detection of peptides. The different reaction conditions chosen allowed for the comparison of two different denaturants, RapiGest<sup>TM</sup> SF and urea, as well as the reducing agents, DTT and TCEP, and various concentrations of the alkylating agent (IAA). Additionally, the necessity of adding an extra procedural step where DTT was added to quench the excess IAA, preventing unwanted side reactions and complications in MS data analysis, was

investigated. In order to systematically compare 2 different denaturants concomitantly with comparing 2 different reductants, certain sets of conditions only differ in the denaturant used, such as Conditions 1 and 2 (see Table 1) where the rest of the digestion parameters (reducing agent type and concentration and alkylating agent concentration) were kept the same.

```

1          70
MRLAVGALLV CAVLGLCLAV PDKTVRWCAV SEHEATKCQS FRDHMKSVIP SDGPSVACVK KASYLDCIRA
71          140
IAANEADAVT LDAGLVYDAY LAPNNLKPVV AEFYGSKEDP QTFYYAVAVV KKDSGFQMNQ LRGKKSCHTG
141        210
LGRSAGWNIP IGLLYCDLPE PRKPLEKAVA NFFSGSCAPC ADGTDFPQLC QLCPCGCGCST LNQYFGYSGA
211        280
FKCLKDGAGD VAFVKHSTIF ENLANKADR D QYELLCLDNT RKPVDEYKDC HLAQVPSHTV VARSMSGGKED
281        350
LIWELLNQAQ EHFGGDKSKE FQLFSSPHGK DLLFKDSAHG FLKVPRPRMDA KMYLGYEYVT AIRNLREGTC
351        420
PEAPTDECKP VKWCALSHHE RLKCDEWSVN SVGKIECVSA ETTEDCIAKI MNGEADAMSL DGGFVYIAGK
421        490
CGLVPVLAEN YNKSDNCEDT PEAGYFAVAV VKKSASDLTW DNLKGGKKSCH TAVGRTAGWN IPMGLLYNKI
491        560
NHCRFDEFFS EGCAPGSKKD SSLCKLCMGS GLNLCEPNNK EGYGYTGAF RCLVEKGDVA FVKHQTVPQN
561        630
TGGKNPDPWA KNLNEKDYEL LCLDGTRKPV EEYANCHLAR APNHAVVTRK DKEACVHKIL RQQOHLFGSN
631        698
VTDCSGNFCL FRSETKDLLF RDDTVCLAKL HDRNTYEKYL GEEYKAVGN LRKCSTSSLL EACTFRRRP

```

**Figure 1.** Amino acid sequence of human apo-transferrin. Blue text indicates the signal peptide. Green text signifies amino acids that were detected using the seventh set of digestion conditions (See Table 1). Black text illustrates amino acids that were not detected using these conditions. Red text shows Cys residues that were not detected. Brown text highlights N-linked glycosylation sites.

In addition to simply detecting transferrin peptides in the MS and MS/MS data, other factors governing digestion efficiency were also assessed, including: 1) complete alkylation (or under alkylation) of the Cys residues, 2) alkylation on the N-terminus of peptides (over alkylation), and 3) incomplete detection of glycans. MS<sup>1</sup> and MS/MS analysis was performed to evaluate these factors for each of the seven different

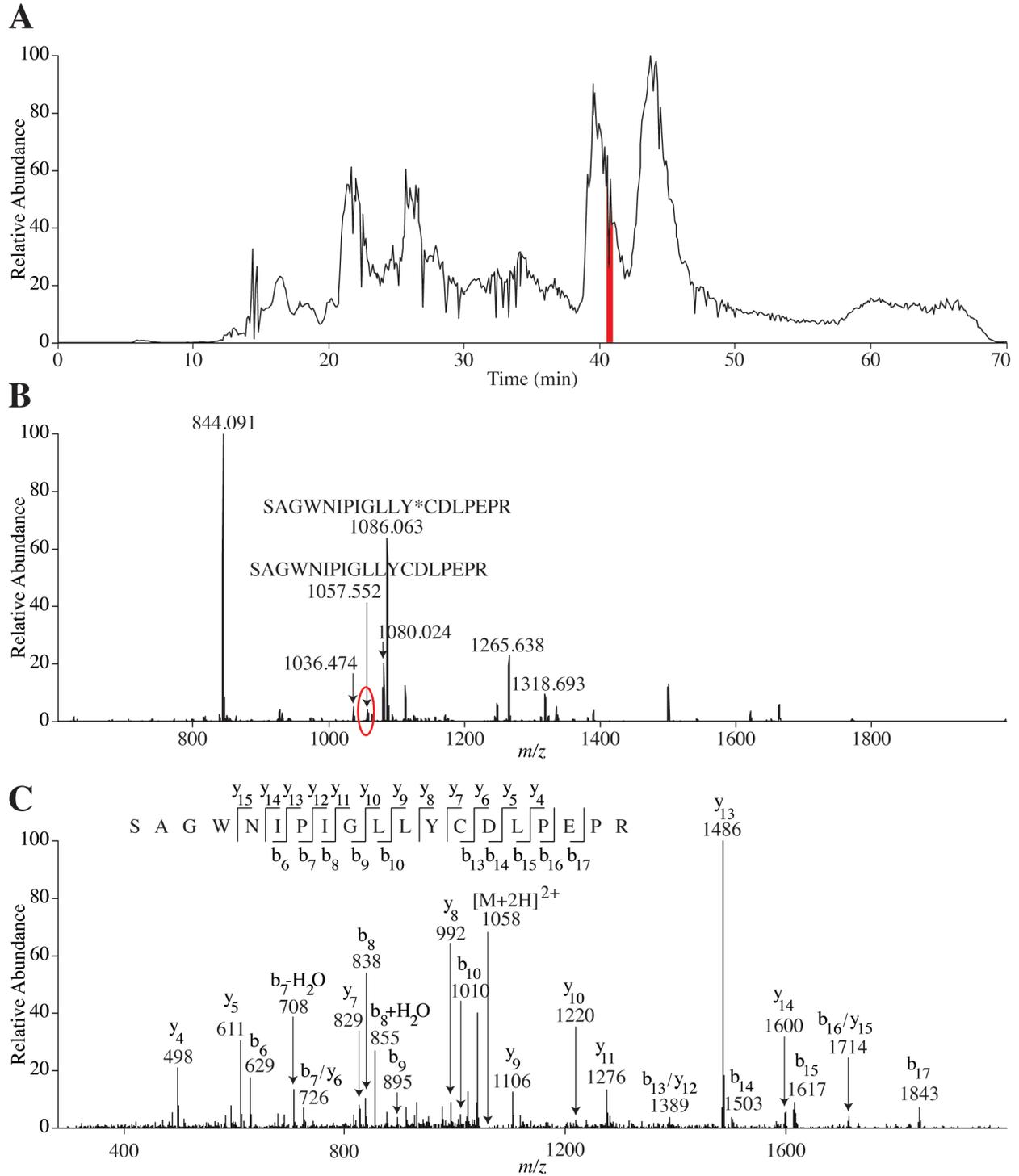
reaction conditions (see Table 1) in an effort to determine optimal protease digestion conditions for transferrin.

### **2.3.2 MS data analysis of peptides and criteria chosen to compare the different digestion conditions**

One of the key goals in this work was to obtain high sequence coverage on the proteins that were analyzed, because the greater the sequence coverage of a protein, the more optimal the digestion conditions. Additionally, because reduction and alkylation were also being evaluated, it was critical to determine the extent of alkylation in the peptides that were analyzed, because the chosen optimal conditions should have complete alkylation of Cys residues without detection of over alkylation on the N-terminus of peptides. Thus, each peptide in the MS data needed to be searched for a fully alkylated peptide peak (when Cys residues were present), peaks corresponding to peptides with incomplete alkylation of Cys residues, and peaks containing N-terminal alkylation for all peptides. The proteins being analyzed in this study are quite large, especially transferrin (~80 kDa); so there were numerous predicted peptides that needed to be searched for in the data. Mascot was chosen to aid in MS data analysis to help reduce the amount of time needed to search for all the predicted peptide peaks. While Mascot was useful for detecting the peptides present in high abundance and minimizing the analysis time in identifying peptides from each of the different reaction conditions tested, it was not an all-inclusive solution to the data analysis problem. Upon manual inspection of the Mascot results, both false positives and false negatives were detected, especially for peptides that were present in low abundance in the MS data. False positives consisted of peptides that Mascot considered a “hit” for the MS/MS data,

however, the MS<sup>1</sup> peak from the high resolution data was not sufficient to be considered a correct match (i.e. the MS<sup>1</sup> monoisotopic peak was > 20 ppm from the calculated *m/z* or an insufficient isotopic distribution was present and no monoisotopic peak could be confirmed). False negatives were peptides that Mascot did not detect, however, upon manual inspection of the MS and MS/MS data the peptides were identified. In other words, the monoisotopic peak was present at < 20 ppm mass error; the charge state was correct, and MS/MS data clearly supported the assignment. Therefore, all the Mascot results were manually validated.

Additionally, there were cases where Mascot could not be used to search for predicted peptides, specifically when searching the MS data for glycosylated peptides as well as peptides with more than one Cys, where one of the two Cys were alkylated. For detection of peptides not in the Mascot results, the LC-MS chromatogram was searched in 1 min increments for MS<sup>1</sup> peaks within 20 ppm mass error of a given peptide. If there was a match, then the MS/MS data was further searched to confirm the identity of the MS<sup>1</sup> peak. Figure 2a shows an example LC chromatogram from condition 3 in Table 1. The highlighted region from 41-42 min corresponds to the retention time averaged for the high resolution mass spectrum shown in Figure 2b. The MS/MS spectrum in Figure 2c resulted from the circled peak labeled in the high resolution spectrum (Figure 2b).



**Figure 2.** Representative human apo-transferrin data from condition 3 (See Table 1). (A) Total ion chromatogram. Highlighted in red is the region (41-42 min.) where the MS<sup>1</sup> in (B) is shown. (B) The 2 labeled peptides illustrate both the fully and non-alkylated peptide SAGWNIPIGLLYCDLPEPR. The peak that is circled in red is the *m/z* where MS/MS was acquired. Additional peaks labeled with numbers alone indicate other detected human apo-transferrin peptides. The \* indicates the alkylated Cys residue. (C) MS/MS of the non-alkylated SAGWNIPIGLLYCDLPEPR with a retention time of 41.55 min.

These particular MS and MS/MS spectra correspond to a peptide that was not alkylated, indicating that the alkylation reaction was incomplete in this case. Thus, as shown in Figure 2b, the peptide is diluted between the fully alkylated and the non-alkylated form. Fortunately, this particular peptide happened to have a strong enough ionization efficiency to detect both the alkylated and non-alkylated forms. However, if the MS signal for the fully alkylated peptide were near the limit of detection, neither of these peaks would be detected in the MS/MS data, leading to a lower protein sequence coverage.

### **2.3.3 Comparison of denaturing and reducing agents**

The most common denaturing agents available, urea (a chaotropic agent) and Rapigest™ SF (a detergent), were chosen to compare for protease digestion optimization, because they have different chemical properties and unique characteristics, and thus have different mechanisms by which proteins are unfolded.<sup>35, 36</sup> Chaotropic agents have a high capacity for forming hydrogen bonds; thus these species compete with the protein's intramolecular forces, ultimately disrupting those forces and causing the protein to unfold. Due to the strong denaturing behavior of chaotropic agents, it is necessary to dilute or remove them prior to adding protease, as proteolytic activity could be affected.<sup>24, 35, 36</sup> Detergents aid in solubilizing and denaturing proteins by adding a degree of hydrophobicity to the solution that allows the hydrophobic regions of the protein to interact with the solution, and thus unfold.<sup>35, 36</sup> Unfortunately, most detergents are incompatible with mass spectrometry. Even “compatible” detergents like the acid labile RapiGest™ SF must be removed from the digestion solution prior to mass spectrometry by precipitating out the surfactant, which can in turn cause

hydrophobic peptides to precipitate out as well.<sup>24</sup> As shown in Table 2, urea is clearly the optimal denaturing agent for the model protein used in this study. In three out of the four reaction conditions where urea was used to denature transferrin, high sequence coverage was obtained. As shown in Table 2, the only reaction condition where a urea denatured sample did not achieve greater sequence coverage than samples utilizing Rapigest™ SF for denaturation was condition 2. The analogous reaction condition for condition 2 was condition 1, where the type and concentration of reducing agent, as well as concentration of alkylating agent were the same, and only the type of denaturing agent was different. Both of these analogous reaction conditions performed poorly and had the lowest sequence coverage compared to all other the reaction conditions tested. Therefore, it was not necessarily that urea did not perform well in condition 2 (it still outperformed condition 1 where Rapigest™ SF was utilized for denaturation), but most likely a poor combination of reducing and alkylating agents led to the poor sequence coverage result in condition 2.

**Table 2.** MS/MS analysis on transferrin peptides for seven different protein digestion reaction conditions.

Reaction Conditions	#1	#2	#3	#4	#5	#6	#7
<b>Protein Coverage</b>	49.8%	66.4%	69.4%	72.5%	72.9%	80.4%	83.5%
<b>Cys Containing Peptides Detected</b>	50.0%	71.1%	52.6%	73.7%	63.2%	78.9%	81.6%
<b>Under Alkylation Detected</b>	0.0%	0.0%	30.0%	0.0%	0.0%	0.0%	0.0%
<b>Over Alkylation Detected</b>	69.6%	48.4%	0.0%	41.7%	0.0%	17.5%	0.0%
<b>Type of Glycans Detected at N<sup>432</sup></b>	Bi	Bi	Bi	Bi	<sup>a</sup> Bi	Bi	Bi
<b>Type of Glycans Detected at N<sup>630</sup></b>	Bi	Bi	Bi	Bi	Bi	Bi	Bi/ <sup>b</sup> Tri

<sup>a</sup>Bi = Biantennary N-linked Glycans. <sup>b</sup>Tri = Triantennary N-linked Glycans.

DTT and TCEP were chosen as competing reducing agents for the optimization study. When disulfide bonds are present in a protein and not reduced prior to the addition of protease, the resulting MS data contains disulfide-bonded peptides, which in

turn complicates the data analysis compared to a solution of peptides in which all disulfide bonds have been broken. Therefore, unless disulfide-bonded peptides are required for a particular analysis, proteins are typically reduced prior to protease digestion. DTT is an inexpensive, common, and readily available reducing agent; however, TCEP has been cited in the literature in recent years to be a potentially better alternative to DTT, because it does not interfere with the common alkylating agent IAA.<sup>30, 31</sup> DTT competes with Cys residues for IAA alkylation due to the presence of thiol functional groups in the structure of DTT. TCEP also has an improved pH range, compared to DTT, and it is stable over a longer period of time, as DTT is known to degrade rapidly within a few hours when in solution.<sup>30, 31</sup> Conversely, TCEP is an acidic compound and thus, more care needs to be taken in the preparation of solutions to which it is added. TCEP also has been cited to produce protein backbone cleavage around Cys residues.<sup>37</sup> Therefore, careful evaluation of both reducing agents was necessary for protease digestion optimization. No cleavages around Cys residues (unless a Cys residue was next to an Arg or Lys where trypsin would be expected to cleave it) were observed in the MS data presented here. As shown in Table 2, two out of the three reaction conditions where TCEP was used as the reducing agent, TCEP yielded greater sequence coverage than the reaction conditions where DTT was the reducing agent. The one time where TCEP did not outperform a DTT reduced sample was in condition 4, where Rapigest<sup>TM</sup> SF was used to denature transferrin instead of urea. As already discussed above, Rapigest<sup>TM</sup> SF was not determined to be an optimal denaturing agent. Therefore, that particular sample of transferrin was probably not

completely unfolded prior to addition of TCEP, which led to TCEP ineffectively reducing the disulfide bonds.

### **2.3.4 Under alkylation of Cys residues**

Incomplete alkylation of Cys residues may occur when a protein is not entirely unfolded or when disulfide bonds are inefficiently reduced, thereby rendering the alkylating agent inaccessible to those residues.<sup>15</sup> The only protease digestion procedure where under alkylation was detected was condition 3, as shown in Table 1. Aside from condition 3, which is obviously not an optimal protocol, under alkylation was not a significant issue.

### **2.3.5 Over alkylation**

Over alkylation occurs when alkylation is detected on the N-terminus of peptides, as opposed to being limited to Cys residues. Although alkylating agents are most selective to thiol groups, amines can also become reactive when given enough incubation time.<sup>34</sup> As no reagents were removed from the reaction mixtures during the protease digestion, there was ample time for IAA to react with the N-terminus of newly-formed peptides that formed after cleavage by trypsin. Indeed, the results from Table 2 indicate that over alkylation was detected in all reaction conditions that lacked an additional step of adding DTT to quench alkylation of IAA, regardless of the identity of reducing or denaturing agent used. Therefore, the optimal protease digestion condition must incorporate a step to quench IAA after alkylation.

### **2.3.6 Evaluation of sequence coverage**

As described above, the main criterion used to determine the optimal protease digestion conditions was protein sequence coverage. The reaction conditions listed in

Table 2 are shown from left to right in increasing sequence coverage. Sequence coverage, however, was not the only criterion used to determine optimal reaction conditions. Due to the many peptides that contained one or more Cys residues, and because under and over alkylation was being evaluated, the percentage of Cys-containing peptides was also evaluated and compared to total sequence coverage. An efficient trypsin digestion with proper reduction and alkylation should contain similar results for Cys-containing peptides compared to total sequence coverage. As shown in Table 2, most of the digestion conditions did have similar values for sequence coverage and Cys-containing peptides, except for conditions 3 and 5, where there is an approximate 10 % drop in detection of Cys-containing peptides compared to sequence coverage, which is probably due to inefficient reduction and alkylation reactions in these reaction procedures.

### **2.3.7 Determination of the optimal conditions**

As described above, transferrin was proteolytically digested utilizing the first 6 reaction conditions from Table 1 and subsequently analyzed by LC-MS. High resolution MS<sup>1</sup> and MS/MS data were searched for transferrin peptides. The results from this study were compiled in Table 2 where the reaction conditions are listed in order of lowest sequence coverage to highest sequence coverage. As described above, it was determined that urea outperformed Rapigest<sup>TM</sup> SF as a denaturant and TCEP outperformed DTT as a reducing agent. Condition 6, which utilized both urea and TCEP for denaturation and reduction, respectively, was nearly optimal, except for the presence of several over alkylated peptides. Upon reviewing these results, a seventh reaction condition was developed with the same protease digestion procedure from

condition 6, but with an additional step after alkylation where DTT was added to quench unreacted IAA (see Table 1 for full digestion details). The transferrin peptides identified in the MS data from the condition 7 digestion protocol are highlighted in Figure 1, with a complete list, including the mass error of each of the identified peptides and glycopeptides, compiled in Table 3. As described in Table 2, condition 7 had the greatest sequence coverage, no under or over alkylation was detected. Thus, condition 7 was expected to be the optimal protease digestion condition.

**Table 3.** List of human apo-transferrin peptides detected under the optimal conditions (Conditions 4 in Table 1). <sup>a</sup>Hex = Hexose, HexNAc = *N*-acetyl hexosamine, NeuNAc = *N*-acetylneuraminic acid.

Peptide	Charge State	Theoretical <i>m/z</i>	Experimental <i>m/z</i>	Mass Error (ppm)
WCAVSEHEATK	2	659.2983	659.2985	0.38
SVIPSDGPSVACVK	1	1415.7197	1415.7278	5.72
ASYLDCIR	1	997.4771	997.4789	1.80
AIAANEADAVTLDAAGLVYDAYLAPN				
NLKPVVAEFYGSK	3	1318.6772	1318.6864	7.00
EDPQTFYYAVAVVK	2	815.4116	815.4159	5.33
KDSGFQMNQLR	2	662.3274	662.3269	0.75
DSGFQMNQLR	1	1195.5525	1195.5559	2.84
SAGWNIPIGLLYCDLPEPR	2	1086.0513	1086.0589	7.00
AVANFFSGSCAPCADGTDFPQLCQ				
LCPGCGCSTLNQYFGYSGAFK	3	1663.3706	1663.3918	12.75
DGAGDVAFVK	1	978.4891	978.4930	3.99
HSTIFENLANK	1	1273.6536	1273.6555	1.49
ADRDQYELLCLDNTR	2	941.4416	941.4424	0.85
DQYELLCLDNTR	2	770.3591	770.3610	2.53
KPVDEYKDCHLAQVPSHTVVAR	3	850.4358	850.4371	1.57
DCHLAQVPSHTVVAR	2	845.4281	845.4292	1.36
SMGGKEDLIWELLNQAQEHFGK	2	1265.6236	1265.6341	8.34
SMGGKEDLIWELLNQAQEHFGKDK	2	1387.1845	1387.1968	8.87
EDLIWELLNQAQEHFGK	2	1035.5183	1035.5227	4.25
SKEFQLFSSPHGK	2	746.3832	746.3837	0.67
EFQLFSSPHGK	2	638.8197	638.8220	3.60
DLLFK	1	635.3764	635.3776	1.89
DSAHGFLK	1	874.4418	874.4449	3.55
MYLGYEYVTAIR	2	739.8711	739.8723	1.62

EGTCPEAPTDECKPVK	2	909.4058	909.4064	0.66
WCALSHHER	2	598.2749	598.2750	0.17
LKCDEWSVNSVGK	2	761.3720	761.3740	2.69
CDEWSVNSVGK	1	1280.5575	1280.5590	1.17
IECVSAETTEDCIAK	2	863.3872	863.3891	2.26
IMNGEADAMSLDGGFVYIAGK	2	1080.0111	1080.0165	5.00
CGLVPVLAENYNK + <sup>3</sup> [Hex]5[HexNAc]4[NeuNAc]2	2	1841.2656	1841.2623	1.79
SDNCEDTPEAGYFAVAVVK	2	1036.4674	1036.4744	6.75
KSASDLTWDNLK	2	689.3541	689.3543	0.29
TAGWNIPMGLLYNK	2	789.4110	789.4151	5.26
FDEFFSEGCAPGSK	2	789.3325	789.3357	4.12
LCMGSGLNLCPEPNNK	2	853.8866	853.8923	6.68
EGYYGYTGAFR	2	642.2883	642.2887	0.70
CLVEK	1	648.3386	648.3375	1.70
GDVAFVK	1	735.4036	735.4005	4.22
HQTVPQNTGGKNPDPWAK	2	987.9927	987.9954	2.78
NPDPWAK	1	827.4045	827.4043	0.24
NLNEKDYELLCLDGTR	2	976.9728	976.9775	4.81
DYELLCLDGTR	1	1354.6308	1354.6392	6.20
KPVEEYANCHLAR	2	793.8908	793.8914	0.76
QQQHFLFGSNVTDCSGNFCLFR + <sup>3</sup> [Hex]5[HexNAc]4[NeuNAc]2	4	1180.7296	1180.7332	3.05
QQQHFLFGSNVTDCSGNFCLFR + <sup>3</sup> [Hex]6[HexNAc]5[NeuNAc]3	3	1792.7042	1792.7129	4.85
DLLFR	1	663.3825	663.3842	2.56
DLLFRDDTVCLAK	2	783.4033	783.4047	1.85
DDTVCLAK	1	921.4346	921.4362	1.74
YLGEEYVK	1	1000.4987	1000.5039	5.20
KCSTSSLLEACTFR	2	830.3951	830.3975	2.89
CSTSSLLEACTFR	2	766.3476	766.3476	0.00

### 2.3.8 Detection and analysis of glycopeptides

In addition to the criteria above, one key feature of an ideal digestion protocol is that it produces high coverage of the post-translational modifications on the protein being analyzed. Therefore, in addition to checking for sequence coverage and alkylation state, the seven data sets were also searched for the known PTMs on transferrin, which contains two *N*-linked glycosylation sites. For a set of reaction

conditions to be considered optimal, the glycopeptides detected in the transferrin MS data needed to encompass all the glycoforms described in the literature for this protein sample. The major glycoform known to be present on transferrin is an *N*-linked biantennary sialylated complex type glycan.<sup>38</sup> However, transferrin also has an *N*-linked triantennary sialylated complex type glycan present in lower abundance.<sup>38</sup> Therefore, the MS data was searched for both biantennary and triantennary sialylated complex type glycopeptides in the data sets from all 7 reaction conditions. As shown in Table 2, biantennary *N*-linked sialylated glycopeptides were detected at both glycosylation sites in transferrin in data sets from all 7 reaction conditions. This was expected, since the biantennary glycans are the most abundant glycoforms in transferrin.<sup>38</sup> The data set where the lesser abundant triantennary glycans were detected was from condition 7. The triantennary *N*-linked glycan details are listed in Table 3 with the QQQHFLFGSNVTDCSGNFCLFR peptide portion. The presence of the less common glycoform detected only in the MS data for condition 7 was further confirmation that condition 7 was indeed the most optimal reaction protocol.

## **2.4 Concluding remarks**

Properly preparing proteins for mass spectrometric analysis through protease digestion is a critical process for protein identification, especially on those proteins that containing PTMs because many PTMs, such as glycosylation, have poor ionization efficiencies compared to peptides that do not contain PTMs. Inefficient digestions will lead to poor MS data, thus the goal of this research was to employ a systematic strategy for developing an ideal protease digestion protocol for recombinantly expressed proteins containing post-translational modifications. Additionally, we

illustrated the general applicability of our findings by using the identified conditions in the successful analysis of a protein that had never before been analyzed by mass spectrometry; therefore the developed protocol should be an efficient method for other soluble proteins.

The ideal protease digestion reaction contained 6 M urea for denaturation, 5 mM TCEP for reduction, 10 mM IAA for alkylation, and 10 mM DTT for quenching the alkylation reaction. As described in Table 2, these conditions illustrated the highest protein sequence coverage with no under or over alkylation detected in the MS data. These conditions were also the only reaction conditions where triantennary *N*-linked glycopeptides were detected in the transferrin data. This is significant because other researchers have been able to detect the triantennary *N*-linked glycans in deglycosylated transferrin data.<sup>38</sup> Therefore, these glycoforms are expected to be detected as glycopeptides as well.

### **Acknowledgements**

The authors acknowledge financial support from an NSF Career award (project number 0645120) to HD, and an NSF Fellowship to KR and CW (DGE-0742523).

## 2.5 References

- (1) Beck, A.; Wagner-Rousset, E.; Bussat, M. –C.; Lokteff, M.; Klinguer-Hamour, C.; Haeuw, J. –F.; Goetsch, L.; Wurch, T.; Dorsselaer, A. V.; Corvaia, N. Trends in glycosylation, glycoanalysis, and glycoengineering of therapeutic antibodies and Fc-fusion proteins. *Current Pharmaceutical Biotechnology* **2008**, *9*, 482-501.
- (2) Liu, Y. –H.; Wylie, D.; Zhao, J.; Cure, R.; Cutler, C.; Cannon-Carlson, S.; Yang, X.; Nagabhushan, T. L.; Pramanik, B. N. Mass spectrometric characterization of the isoforms in *Escherichia coli* recombinant DNA-derived interferon alpha-2b *Anal. Biochem.* **2011**, *408*, 105-117.
- (3) Walsh, G.; Jefferis, R. Post-translational modifications in the context of therapeutic proteins. *Nature Biotechnol.* **2006**, *24*, 1241-1252.
- (4) Koff, R. J. Hepatitis B vaccines: recent advances. *Int. J. Parasitol.* **2003**, *33*, 517-523.
- (5) Madrid-Marina, V.; Torres-Poveda, K.; Lopez-Toledo, G.; Garcia-Carranca, G. Advantages and disadvantages of current prophylactic vaccines against HPV. *Archives of Medical Research* **2009**, *40*, 471-477.
- (6) Go, E. P.; Irungu, J.; Zhang, Y.; Dalpathado, D. S.; Liao, H. –X.; Sutherland, L. L.; Alam, S. M.; Haynes, B. F.; Desaire, H. Glycosylation site-specific analysis of HIV envelope proteins (JR-FL and CON-S) reveals major differences in glycosylation site occupancy, glycoform profiles, and antigenic epitopes' accessibility. *J. Proteome Res.* **2008**, *7*, 1660-1674.
- (7) Go, E. P.; Chang, Q.; Liao, H. –L.; Sutherland, L. L.; Alam, S. M.; Baynes, B. F.; Desaire, H. Glycosylation site-specific analysis of clade C HIV-1 envelope proteins. *J. Proteome Res.* **2009**, *8*, 4231-4242.
- (8) Barouch, D. H. Challenges in the development of an HIV-1 vaccine. *Nature*, **2008**, *455*, 613-619.
- (9) Thakur, D.; Rejtar, T.; Karger, B.; Washburn, N. J.; Bosques, C. J.; Gunay, N. S.; Shriver, Z.; Venkataraman, G. Profiling the glycoforms of the intact  $\alpha$  subunit of recombinant human chorionic gonadotropin by high-resolution *Anal. Chem.* **2009**, *81*, 8900-8907.
- (10) Barnes, C. A. S.; Lim, A. Applications of mass spectrometry for the structural characterization of recombinant protein pharmaceuticals. *Mass Spectrometry Reviews* **2007**, *26*, 370-388.
- (11) Ezan, E.; Dubois, M.; Becher, F. Bioanalysis of recombinant proteins and antibodies by mass spectrometry. *Analyst* **2009**, *134*, 825-834.

- (12) Chen, G.; Mirza, U. A.; Pramanik, B. N. Macromolecules in drug discovery: Mass spectrometry of recombinant proteins and proteomics. *Advances in Chromatography* **2009**, *47*, 1-29.
- (13) Rebecchi, K. R.; Wenke, J. L.; Go, E. P.; Desaire, H. Label-free quantitation: a new glycoproteomics approach. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1048-1059.
- (14) Norrgran, J.; Williams, T. L.; Woolfitt, A. R.; Solano, M. I.; Pirkle, J. L.; Barr, J. R. Optimization of digestion parameters for protein quantification. *Anal. Biochem.* **2009**, *393*, 48-55.
- (15) Hamdan, M.; Righetti, P. G. Modern strategies for protein quantification in proteome analysis: advantages and limitations. *Mass Spectrom. Reviews* **2002**, *21*, 287-302.
- (16) Itoh, S.; Kawasaki, N.; Ohta, M.; Hayakawa, T. Structural analysis of a glycoprotein by liquid chromatography-mass spectrometry and liquid chromatography with tandem mass spectrometry – application to human thrombomodulin. *J. Chromatography A* **2002**, *978*, 141-152.
- (17) Dalpathado, D.; Desaire H. Glycopeptide analysis by mass spectrometry. *Analyst* **2008**, *133*, 731-738.
- (18) Wuhrer, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J. Chromatography B* **2007**, *849*, 115-128.
- (19) Meyer, B.; Papatotiriou, D. G.; Karas, M. 100 % protein sequence coverage: a modern form of surrealism in proteomics. *Amino Acids* **2010**, Epub ahead of print.
- (20) Zhang, Y.; Go, E. P.; Desaire, H. Maximizing coverage of glycosylation heterogeneity in MALDI-MS analysis of glycoproteins with up to 27 glycosylation sites. *Anal. Chem.* **2008**, *80*, 3144-3158.
- (21) Second, T. P.; Blethrow, J. D.; Schwartz, J. C.; Merrihew, G. E.; MacCoss, M. J.; Swaney, D. L.; Russell, J. D.; Coon, J. J.; Zabrouskov, V. Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures. *Anal. Chem.* **2009**, *81*, 7757-7765.
- (22) Garza, S.; Moini, M. Analysis of complex protein mixtures with improved sequence coverage using (CE-MS/MS)<sup>n</sup>. *Anal. Chem.* **2006**, *78*, 7309-7316.

- (23) Ren, D.; Julka, S.; Inerowicz, H. D.; Regnier, F. E. Enrichment of cysteine-containing peptides from tryptic digests using a quaternary amine tag. *Anal. Chem.* **2004**, *76*, 4522-4530.
- (24) Chen, E. I.; Cociorva, D.; Norris, J. L.; Yates, J. R., III Optimization of mass spectrometry-compatible surfactants for shotgun proteomics. *J. Proteome Res.* **2007**, *6*, 2529-2538.
- (25) Arnold, R. J.; Hrnčirova, P.; Annaiah, K.; Novotny, M. V. Fast proteolytic digestion coupled with organelle enrichment for proteomic analysis of rat liver. *J. Proteome Res.* **2004**, *3*, 653-657.
- (26) Strader, M. B.; Tabb, D. L.; Hervey, W. J.; Pan, C.; Hurst, G. B. Efficient and specific trypsin digestion of microgram to nanogram quantities of proteins in organic-aqueous solvent systems. *Anal. Chem.* **2006**, *78*, 125-134.
- (27) Yu, Y. -Q.; Gilar, M.; Lee, P. J.; Bouvier, E. S. P.; Gebler, J. C. Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Anal. Chem.* **2003**, *75*, 6023-6028.
- (28) Hervey, W. J., IV; Strader, M. B.; Hurst, G. B. Comparison of digestion protocols for microgram quantities of enriched protein samples. *J. Proteome Res.* **2007**, *6*, 3054-3061.
- (29) Masuda, T.; Tomita, M.; Ishihama, Y. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J. Proteome Res.* **2008**, *7*, 731-740.
- (30) Getz, E. B.; Xiao, M.; Chakrabarty, T.; Cooke, R.; Selvin, P. R. A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. *Anal. Biochem.* **1999**, *273*, 73-80.
- (31) Cline, D. J.; Redding, S. E.; Brohawn, S. G.; Psathas, J. N.; Schneider, J. P.; Thorpe, C. New Water-Soluble Phosphines as Reductants of Peptide and Protein Disulfide Bonds: Reactivity and Membrane permeability. *Biochemistry* **2004**, *43*, 15195-15203.
- (32) Jacobs, J. M.; Mottaz, H. M.; Yu, L. -R.; Anderson, D. J.; Moore, R. J.; Chen, W. -N. U.; Auberry, K. J.; Strittmatter, E. F.; Monroe, M. E.; Thrall, B. D.; Camp, D. G., II; Smith, R. D. Multidimensional proteome analysis of human mammary epithelial cells. *J. Proteome Res.* **2004**, *3*, 68-75.
- (33) Sechi, S.; Chait, B. T. Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Anal. Chem.* **1998**, *70*, 5150-5158.
- (34) Boja, E. S.; Fales, H. M. Overalkylation of a protein digest with iodoacetamide. *Anal. Chem.* **2001**, *73*, 3576-3582.

- (35) Tiefenbach, K. –J.; Durchschlag, H.; Jaenicke, R. Hydrodynamic and spectroscopic analysis of the denaturation of serum albumin induced by guanidinium chloride and sodium dodecyl sulfate. *Prog. Colloid Polym. Sci.* **2004**, *127*, 136-147.
- (36) Rabilloud, T. Solubilization of proteins for electrophoretic analyses. *Electrophoresis* **1996**, *17*, 813-829.
- (37) Liu, P.; O'Mara, B. W.; Warrack, B. M.; Wu, W.; Huang, Y.; Zhang, Y.; Zhao, R.; Lin, M.; Ackerman, M. S.; Hocknell, P. K.; Chen, G.; Tao, L.; Rieble, S.; Wang, J.; Wang-Iverson, D. B.; Tymiak, A. A.; Grace, M. J.; Russell, R. J. A Tris (2-carboxyethyl) phosphine (TCEP) related cleavage on cysteine-containing proteins. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 837-844.
- (38) Wada, Y.; Azadi, P.; Costello, C. E.; Dell, A.; Dwek, R. A.; Geyer, H.; Geyer, R.; Kakehi, K.; Karlsson, N. G.; Kato, K.; Kawasaki, N.; Khoo, K. –H.; Kim, S.; Kondo, A.; Lattova, E.; Mechref, Y.; Miyoshi, E.; Nakamura, K.; Narimatsu, H.; Novotny, M. V.; Packer, N. H.; Perreault, H.; Peter-Katalinic, J.; Pohlentz, G.; Reinhold, V. N.; Rudd, P. M.; Suzuki, A.; Taniguchi, N. Comparison of the methods for profiling glycoprotein glycans - HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. *Glycobiology* **2007**, *17*, 411-422.

## CHAPTER 3

# MASS SPECTROMETRIC ANALYSIS OF SEQUENCE COVERAGE AND POST-TRANSLATIONAL MODIFICATIONS OF HUMAN LYSYL OXIDASE-LIKE 2 GLYCOPROTEIN EXPRESSED IN A *DROSOPHILA* CELL LINE

### 3.1 Introduction

Lysyl oxidase (LOX) is a secreted copper-containing amine oxidase that forms reactive aldehydes by oxidating the  $\epsilon$ -amino group of lysine side chains in collagen and elastin.<sup>1</sup> LOX contains a cross-linked quinone cofactor arising from a post-translational modification (PTM) of its lysyl and tyrosyl residues, which are conserved across all LOX and lysyl oxidase-like (LOXL) proteins.<sup>1-7</sup> Enzymatically, LOX has been shown to promote stability in the extracellular matrix (ECM) through catalysis of intra- and intermolecular cross-linkages that act to determine mechanical properties.<sup>1, 8-10</sup> Research involving lysyl oxidase and lysyl oxidase-like proteins (LOXL, LOXL2, LOXL3 and LOXL4) has implicated that these enzymes participate in a variety of biological processes, including extracellular matrix stabilization, cellular growth, and homeostasis.<sup>1, 8-9</sup>

Each of these enzymes share a conserved C-terminal amino acid sequence containing residues that form a carbonyl cofactor, copper binding site, and a cytokine receptor-like domain that are crucial for enzymatic activity,<sup>5-6, 11</sup> but vary in their N-terminal domain content.<sup>1</sup> Much less is known about the function of the individual enzymes<sup>9</sup> and their post-translational modifications (PTMs); therefore characterization of the protein by mass spectrometry can contribute new knowledge to the understanding of LOX and LOXL proteins.

Moreover, developing MS-based methods to characterize these proteins is also important because LOX family participants are attractive pharmacological targets as dysregulation of LOX has been found to correlate to numerous diseases and adverse physiological states, including cancer formation and metastasis, connective tissue disorders, neurodegenerative pathologies, and cardiovascular abnormalities.<sup>9, 12-15</sup>

Specifically, LOXL2 has been shown to be involved in abnormal collagen deposition, tumor invasion, lymph node metastasis, and cancer progression in breast and ovarian cancers.<sup>4, 12-14, 16-17</sup> As such, characterization of LOXL2 is an essential step in assessing its viability as a pharmacological candidate of interest as regulatory roles for which the quinone cofactor become further elucidated and better understood.

The work described herein utilizes mass spectrometry for the complete characterization of hLOXL2 that was expressed for the first time in a *Drosophila* cell line. For protein characterization by mass spectrometry, a protease digestion was performed so that the protein sequence and PTMs present in hLOXL2 could be detected. To ensure an efficient protease digestion, the optimal digestion conditions identified from the transferrin study were used. The full details of the transferrin study are described in Chapter 2. Because hLOXL2 has not been fully analyzed by mass spectrometry before, full protein sequence coverage was desired, in addition to detection of the PTMs including glycosylation suspected at the 2 *N*-linked glycosylation consensus sequences and the intact lysyl-tyrosol cross-link. The analysis of hLOXL2 by mass spectrometry should prove valuable in assessing the structural details for this potential pharmaceutical target.

## **3.2 Experimental**

### **3.2.1 Materials and reagents**

Trizma® HCl, Trizma® base, tris(2-carboxyethyl)phosphine (TCEP), urea, ethylenediaminetetraacetic acid (EDTA), dithiothreitol (DTT), iodoacetamide (IAA), acetic acid and formic acid were purchased from Sigma (St. Louis, MO). HPLC grade acetonitrile was purchased from Fisher (Fair Lawn, NJ). Sequencing grade modified trypsin was purchased from Promega (Madison, WI). Water was purified in-house with a Millipore Direct-Q® UV-3 system (Billerica, MA) and was only used when the resistance was > 18 MΩ. A recombinant form of human lysyl oxidase-like 2 (hLOXL2) was prepared using an insect expression system prepared by the Mure lab at the University of Kansas.

### **3.2.2 Protease digestion**

Human lysyl oxidase-like 2 (hLOXL2) protein was supplied at 2 mg/mL in 100 mM Tris, pH 8.5 in 1 mM EDTA and 150 mM NaCl. Solid urea was added to the sample to a final concentration of 6 M, for protein denaturation, followed by the addition of TCEP, to a concentration of 5 mM, for reduction of disulfide bonds. The sample was then allowed to incubate at room temperature for 1 hr. IAA (10 mM) was added to alkylate the Cys residues so that refolding of the protein could not occur. This reaction was allowed to proceed for 1 hr at room temperature in the dark. The alkylation reaction was quenched by the addition of 10 mM DTT. Since the urea content was too high for efficient trypsin digestion, the sample was diluted until the final concentration of urea was 1 M. At this point trypsin was added at a 1:30 enzyme:protein ratio for protease digestion. The sample was placed in a 37 °C oven for 18 hrs and stopped by

the addition of 1  $\mu\text{L}$  acetic acid for every 100  $\mu\text{L}$  of solution. After digestion was complete, the sample was placed in a Labconco centrivap cold trap (Kansas City, MO) to concentrate the sample to  $\sim 3$  mg/mL. Samples were stored at  $-20$   $^{\circ}\text{C}$  until LC-MS could be performed.

### **3.2.3 Liquid chromatography/mass spectrometry**

#### **3.2.3.1 Mass spectrometry on an ESI-LTQ-FTICR MS**

The protease digestion sample, hLOXL2 (5  $\mu\text{L}$ ,  $\sim 15$   $\mu\text{g}$ ), was injected onto a  $\text{C}_{18}$  column (300  $\mu\text{m}$  i.d., 5 cm length, and 3  $\mu\text{m}$  particle size) produced by CVC Microtech (Fontana, CA) that was connected to a Dionex UltiMate capillary HPLC system (Sunnyvale, CA) containing a FAMOS well plate autosampler with an electrospray – linear ion trap – Fourier transform ion cyclotron resonance mass spectrometer (ESI-LTQ-FTICRMS), ThermoScientific (San Jose, CA) containing a 7 Tesla actively shielded magnet for detection. Aqueous mobile phase consisted of 99.9 % water and 0.1 % formic acid (solvent A) and organic mobile phase was composed of 99.9 % acetonitrile and 0.1 % formic acid (solvent B). For reversed phase separation of the peptides in hLOXL2, the following gradient was used: Solvent conditions were held at 5% B for 5 min, a linear increase to 40% B in 50 min, another linear increase to 90% B in 10 min, held at 90% B for 10 min, and re-equilibration of the column. Between each sample run, a short 30 min wash cycle and blank were run to ensure no carry over was detected between samples. The ESI source was set to 2.8 kV, capillary temperature was  $200$   $^{\circ}\text{C}$ , capillary voltage offset was 47 V, and FT-ICR resolution was set to 25,000 for  $m/z$  400. MS/MS data was collected in data-dependent mode where the 5 most intense ions from the high resolution FT-ICR MS<sup>1</sup> scan were selected for collision

induced dissociation (CID) at 30% collision energy and a 3 min dynamic exclusion window.

### **3.2.3.2 Mass spectrometry on an ESI-LTQ Velos MS**

hLOXL2 (5  $\mu$ L,  $\sim$ 15  $\mu$ g) was injected onto the same C<sub>18</sub> column described in the paragraph above. For these experiments, however, the column was connected to a Waters Acquity UPLC system (Milford, MA), which was directly coupled to an electrospray – linear ion trap mass spectrometer (ESI-LTQ Velos MS) from ThermoScientific (San Jose, CA). Mobile phase A and B are the same as in the above paragraph. The following gradient was used for reversed phase separation of hLOXL2 peptides: Initial conditions were 5% B with a linear increase to 10% B in 5 min, a linear increase to 40% B in 45 min, a linear increase to 90 % B in 10 min, held at 90% B for 10 min before re-equilibration of the column. Since sample carryover is sometimes an issue in LC-MS analysis, a short (30 min) wash cycle and a blank run were performed between each sample. For mass spectrometry, the electrospray source voltage was 3 kV, capillary temperature was 250 °C. For MS/MS analysis of peptides by CID (collision induced dissociation), a 30% collision energy was used. LC-MS/MS was set up in data dependent scan mode where the 5 most intense ions were chosen for MS/MS analysis with a 3 min dynamic exclusion window.

### **3.2.4 Data analysis**

For Mascot analysis of the MS data from the hLOXL2 sample that was run on the ThermoScientific ESI-LTQ-FTICRMS, the peak lists from XCalibur (Version 2.1.0.1139) raw files were exported into BioWorks browser (ThermoScientific, Version 3.5), which exported the data as a .DTA file. The .DTA file was searched for the following

parameters: 1) enzyme, trypsin; 2) up to 2 missed cleavages; 3) Cysteine carbamidomethyl as a fixed modification (when searching for under alkylation of Cys residues, this parameter was removed); 4) methionine oxidation, carbamyl, and N-terminal carbamidomethyl as variable modifications; 5) 0.8 Da peptide tolerance; and 6) 0.4 Da MS/MS tolerance. The MS data was searched in SwissProt 2010, taxonomy Mammalia, database. All peptides detected from Mascot were also manually verified to ensure they met the criteria established for a detected peptide and to search for peptides that Mascot did not identify. To be considered a manually detected peptide, two criteria must be met: MS/MS data illustrating appropriate b and y ions was required, and the monoisotopic peak present in the mass spectrum had to be within 20 ppm mass error.

For analysis of glycopeptides in the MS data, a manual interpretation was completed by two strategies. The first strategy was to create a prediction table where the peptide mass from an *N*-linked glycosylation site was added to probable glycoforms specific for *Drosophila* cells,<sup>18</sup> and the predicted *m/z* values were searched for in the MS/MS data. Secondly, the MS/MS data was scanned characteristic ions indicative of glycopeptide spectra. For example, *m/z* 366 (a hexose plus an *N*-acetylhexosamine residue) is often found in glycopeptide spectra and considered a marker ion for glycopeptides. Upon detection of *m/z* 366 and other marker ions, the spectrum was examined for appropriate glycan losses for compositional assignment of a glycopeptide. These strategies were applied for data analysis of hLOXL2 for both instruments used, the ESI-LTQ-FTICRMS and the ESI-LTQ Velos MS.

The lysyl-tyrosol cross-link was determined manually by calculating the mass of the cross-linked peptide with and without certain modifications, including the presence and absence of the phenylhydrazine tag that was added to ensure the cross-link would stay intact through protease digestion, up to 2 trypsin missed cleavages, and methionine oxidation. The predicted  $m/z$  values calculated were searched for in the data for an MS/MS spectrum that contained appropriate b and y ions corresponding to the cross-link.

### **3.3 Results and discussion**

hLOXL2 is a biologically important protein known to play a role in the development of several cancers,<sup>12, 14, 19</sup> as well as a potential therapeutic.<sup>9, 17</sup> This is the first example of a LOX or LOXL protein to be fully characterized by mass spectrometry. The goal of this work was to use MS to obtain complete protein sequence coverage, as well as detect the known PTMs that occur in hLOXL2, specifically analysis of glycosylation present at the two potential *N*-linked glycosylation sites and detection of the lysyl-tyrosyl cross-link known to be present in all LOX and LOXL proteins.<sup>5-7, 20</sup> Figure 1 illustrates the amino acid sequence for this recombinant hLOXL2 protein.

```

1
RSPWPGVPTS MRLNGGRNPY EGRVEVLVER NGSLVWGMVC GQNWGIVEAM VVCRQLGLGF ASNAFQETWY 70
71
WHGDVNSNKV VMSGVKCSGT ELSLAHCRHD GEDVACPQGG VQYGAGVACS ETAPDLVLNA EMVQQTTYLE 140
141
DRPMFMLQCA MEENCLSASA AQTDPTTGYR RLLRFSSQIH NNGQSDFRPK NGRHAWIWHD CHRHYSMEV 210
211
FTHYDLLNLN GTKVAEGHKA SFCLEDTECE GDIQKNYECA NFGDQGITMG CWDMYRHDID CQWVDITDVP 280
281
PGDYLFQVVI NPNFEVAESD YSNNIMCRS RYDGHRIMY NCHIGGSFSE ETEKKFEHFS GLLNNQLSPQ 350
351 359
SAWSHPQFE

```

**Figure 1.** Amino acid sequence of hLOXL2 expressed in a *Drosophila* cell line. Green text signifies amino acids that were detected by LC-MS and MS/MS analysis, as described below. Black text illustrates amino acids that were not detected. Red text shows Cys residues that were not detected. Brown text highlights *N*-linked glycosylation sites. Purple text highlights the Lys and Tyr residues involved in the ortho quinone cross-link. The peptide shown in blue was not detected in the high resolution ESI-LTQ-FTICRMS data, however, it was detected as part of the lysyl-tyrosol cross-link in the low resolution ESI-LTQ Velos MS data.

### 3.3.1 Protein sequence coverage

In order to obtain high protein sequence coverage for recombinant hLOXL2, a proteolytic digestion was performed followed by mass spectrometry. The digestion conditions used for hLOXL2 were previously optimized on a model glycoprotein, transferrin, and described in detail in Chapter 2. This digestion procedure utilized urea for denaturation, TCEP for reduction, IAA for alkylation, and DTT to quench alkylation. After protease digestion of hLOXL2, LC-MS and MS/MS analysis of the rendered peptides was performed. The high resolution LC-MS data collected on the ESI-LTQ-FTICR MS for the hLOXL2 digested sample was first analyzed using Mascot to aid in the identification of hLOXL2 peptides. This was followed by manual validation of the MS and MS/MS data, as described in the experimental section. A summary of the hLOXL2 amino acid residues detected is shown in Figure 1. The overall results from the analysis of hLOXL2 peptides that met both the high resolution MS<sup>1</sup> and MS/MS criteria are described in Table 1. The percent protein sequence coverage was 90.8%

and the percent Cys-containing peptides were 93.8%. In fact, only 1 Cys residue was not detected in the analysis of hLOXL2. The non-detected Cys residue is highlighted in red in Figure 1. This Cys residue was not expected to be detected by MS, because its tryptic peptide is only CR, which would have an  $m/z$  of 335, and this is outside the scan range used for this experiment. Thus, a missed trypsin cleavage is necessary for CR to be detected.

**Table 1.** Summary of results from hLOXL2 peptides detected for protein sequence coverage using the high resolution ESI-LTQ-FTICRMS for analysis.

	<b>hLOXL2</b>
<b>Protein Coverage</b>	90.8 %
<b>Cys Containing Peptides Detected</b>	93.8 %
<b>Under Alkylation Detected</b>	0.0 %
<b>Over Alkylation Detected</b>	0.0 %
<b>Major Glycoform Detected at N<sup>31</sup></b>	<sup>a</sup> [Hex]3[HexNAc]2[Fuc]1
<b>Major Glycoform Detected at N<sup>220</sup></b>	<sup>a</sup> [Hex]3[HexNAc]2[Fuc]1

<sup>a</sup>Hex = Hexose, HexNAc = *N*-acetyl hexosamine, Fuc = Fucose.

All undetected peptides in the hLOXL2 data consisted of short peptides, with 6 amino acid residues or less, found in various regions throughout the protein sequence, as highlighted in black in Figure 1. Of the 9 tryptic peptides not detected by MS (see Figure 1), the calculated  $m/z$  for 5 of the 9 peptides fall below scan range (ie. < 500 Da); thus these peptides would never be detected using the scanning parameters implemented. The other 4 peptides are 5 to 6 amino acid residues in length and could potentially be detected by MS with the scan range utilized. However, we expect that these peptides have lower ionization efficiencies because of the presence of acidic amino acid residues (Asp and Glu) in 3 of the 4 undetected peptides, which could lower the ionization efficiency. In summary, this work illustrates that high sequence coverage was obtained, with both MS<sup>1</sup> and MS/MS data being available for all the detected

peptides. Table 2 lists the theoretical and experimental masses of all the peptides that were detected in the analysis of hLOXL2.

**Table 2.** List of hLOXL2 peptides detected in the ESI-LTQ-FTICR mass spectrometer where both MS/MS and MS<sup>1</sup> < 20 ppm mass error were used as parameters for a detected peptide or glycopeptide. <sup>3</sup>Hex = Hexose, HexNAc = *N*-acetyl hexosamine, Fuc = Fucose, Na = Sodium.

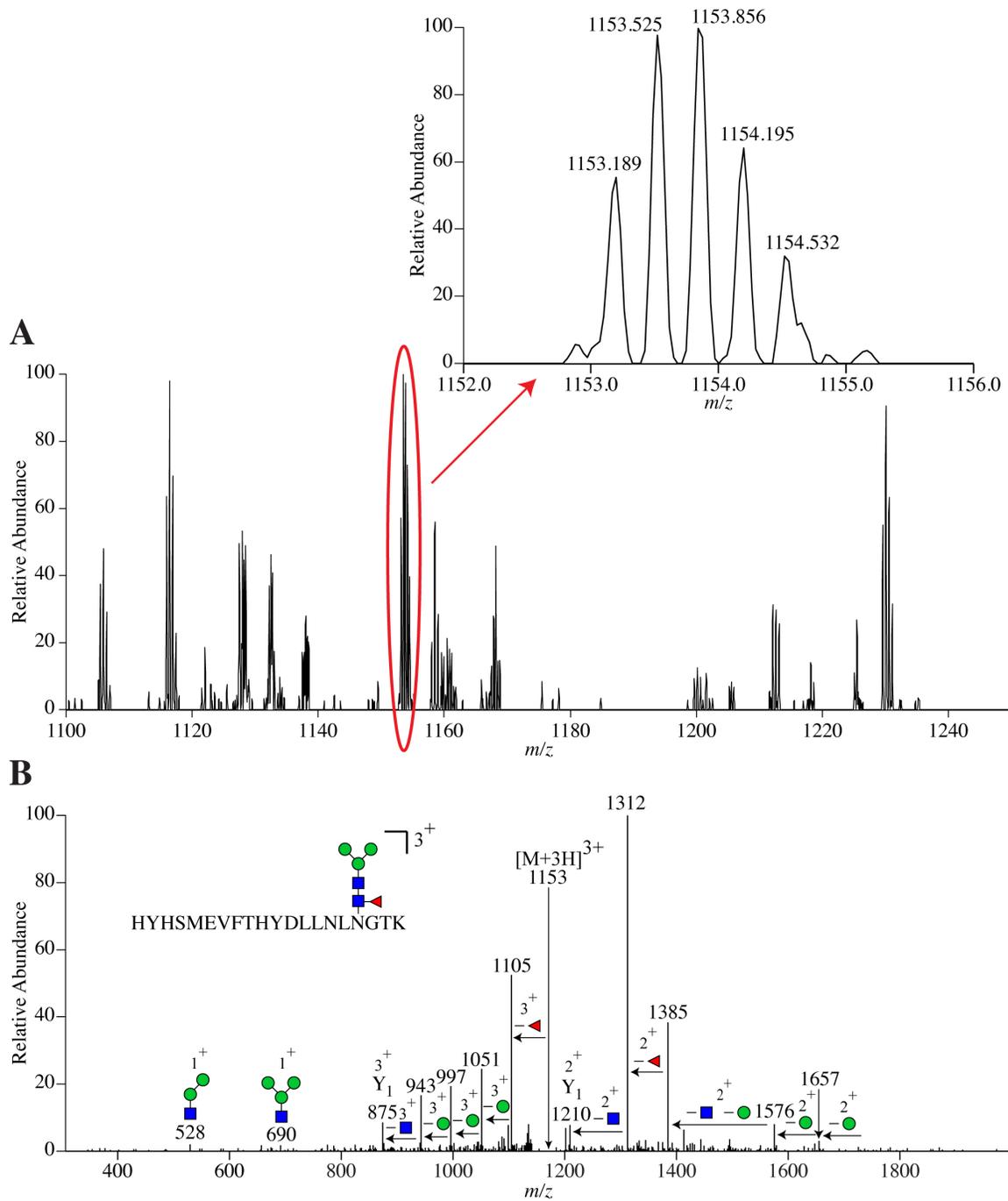
Peptide	Charge State	Calculated <i>m/z</i>	Experimental <i>m/z</i>	Mass error (ppm)
SPWPGVPTSMR	2	607.8029	607.8071	6.9
VEVLVER	1	843.4935	843.4980	5.3
NGSLVWGMVCGQNWGIVEAMVVCR + <sup>3</sup> [Hex]3[HexNAc]2[Fuc]1 + [Na]1	3	1261.8839	1261.8988	11.8
QLGLGFASNAFQETWYWHGDPVNSNK	2	1435.1701	1435.1881	12.5
VVMMSGVK	1	719.4121	719.4133	1.7
CSGTELSLAHCR	2	695.8032	695.8062	4.3
HDGEDVACPQGGVQYGAGVACSETA				
PDLVLNAEMVQQTTYLEDPRPMFMLQC				
AMEENCLSASAAQTDPTTGYR	5	1583.2994	1583.3253	16.4
FSSQIHNNQGSDFRPK	2	931.4506	931.4561	5.9
HAWIWHDCHR	1	1417.6330	1417.6391	4.3
HYHSMEVFTHYDLLNLNGTK + <sup>3</sup> [Hex]3[HexNAc]2[Fuc]1	3	1153.1799	1153.1894	8.2
ASFLEDTECEGDIQK	2	951.3983	951.3851	13.8
NYECANFGDQGITMGCWDMYR	2	1294.5085	1294.5273	14.5
HDIDCQWVDITDVPPGDYLFQVWINPN				
FEVAESDYSNNIMK	3	1599.7422	1599.7617	12.2
IWMYNCHIGGSFSEETEK	2	1094.4774	1094.4901	11.6
FEHFSGLLNNQLSPQSAWSHPQFE	2	1400.6593	1400.6693	7.1

### 3.3.2 Glycopeptide data analysis

The CID MS/MS data of glycopeptides, including the glycopeptides from hLOXL2, are distinct from peptides in that, unlike peptides, glycopeptides cannot be identified in an automated fashion using a Mascot search, but instead must be characterized using other search tools. The two strategies used to detect the glycopeptides present in the hLOXL2 MS data are described in the Experimental

section. A glycopeptide was considered identified when two criteria were met: An MS/MS spectrum was needed to support the assignment, and the monoisotopic peak must have an  $m/z$  value of less than 20 ppm mass error compared to the calculated  $m/z$  in the high resolution (FTICR) MS data.

Figure 2 is an example of a glycopeptide spectrum from one of the two glycosylation sites in hLOXL2. Figure 2A shows a high resolution mass spectrum present where the circled peak corresponded to a glycopeptide from hLOXL2. By zooming in, as shown in Figure 2A, the isotopic distribution can be seen and the monoisotopic  $m/z$  is determined and compared to the calculated  $m/z$ . In this case, the experimental mass error is 8.2 ppm, as shown in Table 2. Figure 2B shows the MS/MS data from the precursor ion circled in Figure 2A. As can be elucidated in Figure 2B, there are losses of monosaccharide sugar residues present in the glycopeptide data. These sugar losses help to identify the *N*-linked glycan present on a glycopeptide, a fucosylated *N*-linked glycan core ([Hex]3[HexNAc]2[Fuc]1, where Hex = hexose, HexNAc = *N*-acetylhexosamine, and Fuc = fucose). See Figure 2B. To confirm this assignment, the MS/MS data is searched for a peak corresponding to the potential peptide plus one hexNAc residue, called the  $Y_1$  ion. This ion is common to many glycopeptide MS/MS data taken in positive ion mode.<sup>21</sup> When a peak corresponding to the  $Y_1$  ion is detected in the data, and it correlates to the monosaccharide sugar losses, a glycopeptide is said to be identified, as shown in Figure 2B.



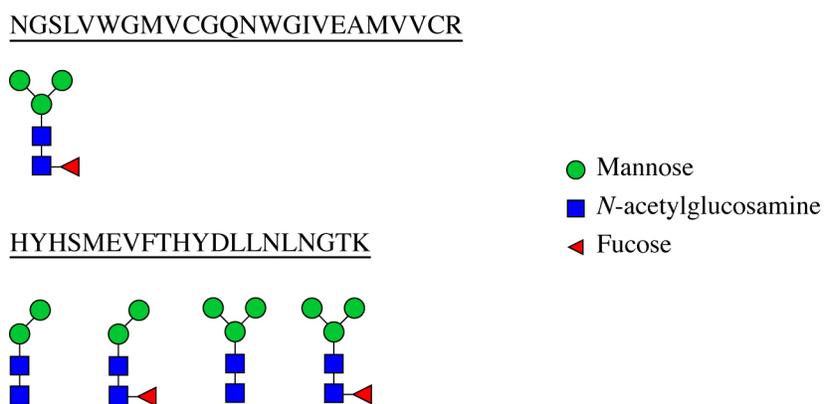
**Figure 2.** (A) High resolution MS<sup>1</sup> spectrum from hLOXL2 from the retention time 38-39 min. Circled in red illustrates where the hLOXL2 glycopeptide ion in (B) is located in the spectrum and the zoomed in region shows the isotopic distribution for the hLOXL2 glycopeptide ion where mass error can be calculated from the monoisotopic  $m/z$  value. (B) hLOXL2 glycopeptide MS/MS data at  $m/z$  1153. The blue squares are *N*-acetylhexosamines; green circles are hexoses, and the red triangle is fucose. This MS/MS data shows losses of glycan residues that aid in determining the glycan composition.

To aid in the manual analysis of the glycopeptide MS/MS data, GlycoPep DB<sup>22</sup> and GlycoPep ID<sup>23</sup> were also utilized. The predominant glycoform detected in the MS/MS data of hLOXL2 at both glycosylation sites was a fucosylated *N*-linked glycan core. This assignment is consistent with the fact that a fucosylated *N*-linked glycan core is known to be one of the most common glycoforms in insect cells and for proteins expressed in insect cell lines.<sup>18</sup>

Analysis of the hLOXL2 protease digestion sample run on a ThermoScientific ESI-LTQ-FTICRMS instrument allowed for the detection of several glycoforms for the HYHSMEVFTHYDLLNLNGTK glycosylation site. However, only the most common fucosylated *N*-linked glycan core was detected at the other glycosylation site (NGSLVWGMVCGQNWGIVEAMVVCR). Figure 3 shows the glycopeptides detected by ESI-LTQ-FTICRMS that met the criteria for glycopeptide identification: The MS/MS data contained appropriate glycan losses and the MS<sup>1</sup> monoisotopic peaks were within 20 ppm mass error. It was concerning, however, to detect only 1 glycoform for the NGSLVWGMVCGQNWGIVEAMVVCR glycosylation site. The reason for low coverage of heterogeneity for glycopeptides detected was not likely from the protease digestion process, since protein sequence coverage was very high at 90.8% (see Table 1). Instead, we hypothesized that the low coverage was due to the known problem of glycopeptides being more difficult to detect than peptides, due to their heterogeneity and, possibly, their lower ionization efficiency.<sup>24</sup>

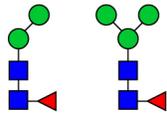
Therefore, a different mass spectrometer was chosen for analysis of hLOXL2 glycopeptides, a ThermoScientific ESI-LTQ Velos MS. The LTQ Velos mass analyzer has been shown to have advantages over the ThermoScientific LTQ regarding duty

cycle (~ 2x decrease in cycle time) and increased sensitivity.<sup>25</sup> Because the LTQ Velos MS was a stand alone instrument and not connected to a high resolution mass spectrometer, such as the FTICR MS, the requirement of less than 20 ppm mass error in the MS<sup>1</sup> data was removed as a criterion for glycopeptide detection. For ESI-LTQ Velos MS glycopeptide detection, the MS/MS spectrum with appropriate glycan losses must be observed, as well as the top of MS<sup>1</sup> peak had to be within 1 Da of the calculated *m/z*. The glycopeptides detected from hLOXL2 with the LTQ Velos MS are summarized in Figure 4, which shows 1 new glycoform detected for the NGSLVWGMVCGQNWGIVEAMVVCR glycosylation site and 3 new glycoforms for the HYHSMEVFTHYDLLNLNGTK glycosylation site.

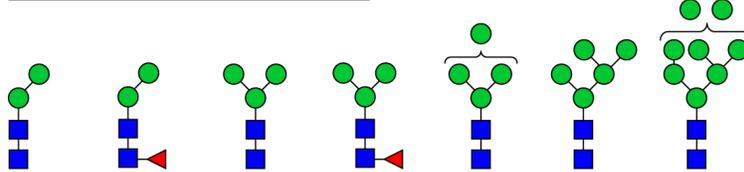


**Figure 3.** Compositions of the glycans detected in the hLOXL2 MS data at both *N*-linked glycosylation sites with the LTQ-FTICRMS.

NGSLVWGMVCGQNWGIVEAMVVCR



HYHSMEVFTHYDLLNLNGTK

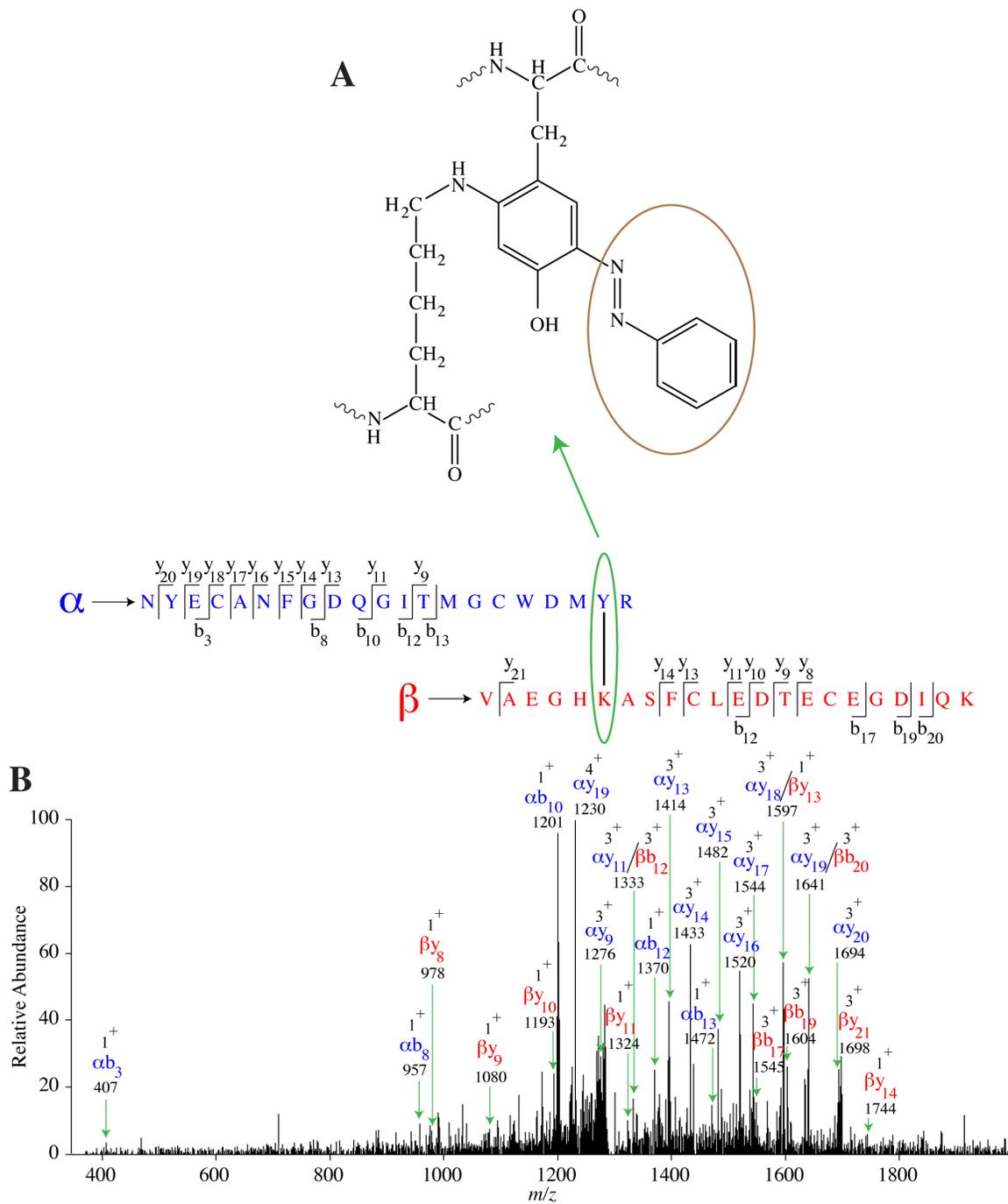


**Figure 4.** Compositions of the glycans detected in the hLOXL2 MS data at both *N*-linked glycosylation sites with the LTQ Velos MS.

### 3.3.3 Detection of the lysyl – tyrosyl cross-link

A unique and conserved cross-link is observed among the different types of LOX and LOXL proteins consisting of a lysine residue cross-linked to a tyrosine residue to form an ortho quinone cofactor.<sup>5-7, 20</sup> Because this cross-link is known to occur in active and properly folded LOX and LOXL proteins, detecting the in-tact cross-link was one of the goals sought after for complete characterization of hLOXL2. To aid in keeping the cross-link in-tact throughout the protease digestion procedure, the protein was derivatized at the cross-link site with phenylhydrazine, as circled in brown in Figure 5A.<sup>2, 6</sup> [This derivatization occurred as part of the sample preparation in the Mure lab and was not part of the experimental work described here.] Upon successful derivatization of the hydrazine group, the protein solution was expected to change from a clear color to pale yellow, which was observed upon receiving the hLOXL2 sample. The lysine and tyrosine residues involved in the cross-link are highlighted in purple in Figure 1. Because this lysine-tyrosine cross-link is known only to occur in LOX and LOXL proteins, it was not possible to utilize databases such as Mascot to aid detection of the

cross-linked peptide in the MS data. After close manual inspection of the ESI-LTQ-FTICRMS data, no evidence for the cross-linked peptide was found. However, we determined that the cross-link must not be 100 % in-tact, because one of the peptides that should have been involved in the cross-link was detected individually (see Figure 1 and Table 2). As in the analysis of glycopeptides, it was determined that the abundance of the intact cross-linked peptide may have been too low for detection with the ESI-LTQ-FTICRMS, so the ESI-LTQ Velos MS was employed for re-analysis of the hLOXL2 protease digestion sample. After inspection of the ESI-LTQ Velos MS data for the intact cross-link and its associated peptides, an MS/MS spectrum was found that corresponds to this species, as shown in Figure 5B. Several b and y ions were detected on both peptides involved in the cross-link, confirming that the cross-link is indeed in-tact, at least partially, in the recombinant hLOXL2 sample.



**Figure 5.** MS/MS data at  $m/z$  1299 corresponding to the lysyl-tyrosol crosslink common in all LOX and LOXL proteins. (A) Illustrates the chemical structure of the cross-link. Circled in brown corresponds to a phenyl hydrazine group that was attached to the protein before protease digestion to ensure the cross-link would stay intact throughout digestion. (B) MS/MS data for the cross-link.

### **3.4 Concluding remarks**

A recombinant form of hLOXL2 expressed in insect cells was successfully analyzed by mass spectrometry, with near complete characterization of the protein sequence. Protein sequence coverage was greater than 90% with all non-detected peptides having a sequence of 5 amino acids or less. In addition, 9 different glycopeptides were identified, with glycoforms present at both *N*-linked glycosylation sites. The glycosylation that was detected was consistent with the glycoforms known to be present in *Drosophila* cells. Lastly, the in-tact lysyl tyrosol cross-link was detected in the LTQ Velos MS data, completing the analysis of hLOXL2 including its post-translational modifications.

### **Acknowledgements**

The authors acknowledge financial support from an NSF CAREER award (project number 0645120) to HD, and an NSF Fellowship to KR (DGE-0742523), and NSF CAREER award (MCB-0747377) and NIH (5R01GM079446-02) to MM.

### 3.5 References

- (1) Molnar, J.; Fong, K. S. K.; He, Q. P.; Hayashi, K.; Kim, Y.; Fong, S. F. T.; Fogelgren, B.; Szauter, K. M.; Mink, M.; Csiszar, K. Structural and functional diversity of lysyl oxidase and the LOX-like proteins. *Biochim et Biophys Acta* **2003**, *1647*, 220-224.
- (2) Wang, S. X.; Mure, M.; Medzihradzsky, K. F.; Burlingame, A. L.; Brown, D. E.; Dooley, D. M.; Smith, A. J.; Kagan, H. M.; Klinman, J. P. A crosslinked cofactor in lysyl oxidase: Redox function for amino acid side chain. *Science* **1996**, *273*, 1078-1084.
- (3) Seve, S.; Decitre, M.; Gleyzal, C.; Farjanel, J.; Sergeant, A.; Ricard-Blum, S.; Sommer, P. Expression analysis of recombinant lysyl oxidase (LOX) in myofibroblastlike cells. *Conn. Tiss. Res.* **2002**, *43*, 613-619.
- (4) Vadasz, Z.; Kessler, O.; Akiri, G.; Gengrinovitch, S.; Kagan, H. M.; Baruch, Y.; Izhak, O. B.; Neufeld, G. Abnormal deposition of collagen around hepatocytes in Wilson's disease is associated with hepatocyte specific expression of lysyl oxidase and lysyl oxidase like protein-2. *J. Hepatology* **2005**, *43*, 499-507.
- (5) Rucker, R. B.; Kosonen, T.; Clegg, M. S.; Mitchell, A. E.; Rucker, B. R.; Uriu-Hare, J. Y.; Keen C. L. Copper, lysyl oxidase, and extracellular matrix protein cross-linking. *Am. J. Clin. Nutr.* **1998**, *67*, 996S-1002S.
- (6) Mure, M. Tyrosine-derived quinone cofactors. *Acc. Chem. Res.* **2004**, *37*, 131-139.
- (7) Anthony, C. Quinoprotein-catalysed reactions. *Biochem. J.* **1996**, *320*, 697-711.
- (8) Lucero, H. A.; Kagan, H. M. Lysyl oxidase: an oxidative enzyme and effector of cell function. *Cell. Mol. Life Sci.* **2006**, *63*, 2304-2316.
- (9) Rodriguez, C.; Rodriguez-Sinovas, A.; Martinez-Gonzalez, J. Lysyl oxidase as a potential therapeutic target. *Drug News Perspect.* **2008**, *21*, 218-224.
- (10) Kagan, H. M.; Li, W. Lysyl oxidase: properties, specificity, and biological roles inside and outside of the cell. *J. Cellular Biochemistry* **2003**, *88*, 660-672.
- (11) Tang, C.; Klinman, J. P. Inhibition of p53 transcriptional activity by the S100B calcium-binding protein. *J. Biol. Chem.* **2001**, *276*, 30575-30578.
- (12) Kirschmann, D. A.; Seftor, E. A.; Fong, S. F. T.; Nieva, D. R. C.; Sullivan, C. M.; Edwards, E. M.; Sommer P.; Csiszar, K.; Hendrix, M. J. C. A molecular role for lysyl oxidase in breast cancer invasion. *Cancer Res.* **2002**, *62*, 4478-4483.

- (13) Barker, H. E.; Chang, J.; Cox, T. R.; Lang, G.; Bird, D.; Nicolau, M.; Evans, H. R.; Gartland, A.; Erler, J. T. LOXL2-mediated matrix remodeling in metastasis and mammary gland involution. *Cancer Res.* **2011**, *71*, 1561-1572.
- (14) Fong, S. F. T.; Dietzsch, E.; Fong, K. S. K.; Hollosi, P.; Asuncion, L.; He, Q. P.; Parker, M. I.; Csiszar, K. Lysyl oxidase-like 2 expression is increased in colon and esophageal tumors and associated with less differentiated colon tumors. *Genes, Chromosomes & Cancer* **2007**, *46*, 644-655.
- (15) Erler, J. T.; Bennewith, K. L.; Nicolau, M.; Kong, C.; Le, Q. -T.; C, J. -T. A; Jeffrey, S. S.; Giaccia, A. J. Lysyl oxidase is essential for hypoxia-induced metastasis. *Nature* **2006**, *440*, 1222-1226.
- (16) Peinado, H.; del Carmen, M.; Iglesias-de la Cruz, M. C.; Olmeda, D.; Csiszar, K.; Fong, K. S. K.; Vega, S.; Nieto, M. A.; Cano, A.; Portillo, F. A molecular role for lysyl oxidase-like 2 enzyme in Snail regulation and tumor progression. *EMBO J.* **2005**, *24*, 3446-3458.
- (17) Peng, L.; Ran, Y. -L.; Hu, H.; Yu, L.; Liu, Q.; Zhou, Z.; Sun, Y. -M.; Sun, L. -C.; Pan, J.; Sun, L. -X.; Zhao, P.; Yang, Z. -H. Secreted LOXL2 is a novel therapeutic target that promotes gastric cancer metastasis via the Src/FAK pathway. *Carcinogenesis* **2009**, *30*, 1660-1669.
- (18) Fabini, G.; Freilinger, A.; Altmann, F.; Wilson, I. B. H. Identification of core alpha 1,3-fucosylated glycans and cloning of the requisite fucosyltransferase cDNA from *Drosophila melanogaster* - Potential basis of the neural anti-horseradish peroxidase epitope. *J. Biol. Chem.* **2001**, *276*, 28058-28067.
- (19) Payne, S. L.; Hendrix, M. J. C.; Kirschmann, D. A. Paradoxical roles for lysyl oxidases in cancer - a prospect. *J. Cellular Biochemistry* **2007**, *101*, 1338-1354.
- (20) Dooley, D. M. Structure and biogenesis of topaquinone and related cofactors. *J. Biol. Inorg. Chem.* **1999**, *4*, 1-11.
- (21) Segu, Z. M.; Mechref, Y. Characterizing protein glycosylation sites through higher-energy C-trap dissociation. *Rapid Comm. Mass Spectrom.* **2010**, *24*, 1217-1225.
- (22) Go, E. P.; Rebecchi, K. R.; Dalpathado, D. S.; Bandu, M. L.; Zhang, Y.; Desaire, H. GlycoPep DB: a tool for glycopeptide analysis using a "smart search". *Anal. Chem.* **2007**, *79*, 1708-1713.
- (23) Irungu, J.; Go, E. P.; Dalpathado, D. S.; Desaire, H. Simplification of mass spectral analysis of acidic glycopeptides using GlycoPep ID. *Anal. Chem.* **2007**, *79*, 3065-3074.

- (24) Zhang, Y.; Go, E. P.; Desaire, H. Maximizing coverage of glycosylation heterogeneity in MALDI-MS analysis of glycoproteins with up to 27 glycosylation sites. *Anal. Chem.* **2008**, *80*, 3144-3158.
- (25) Second, T. P.; Blethrow, J. D.; Schwartz, J. C.; Merrihew, G. E.; MacCoss, M. J.; Swaney, D. L.; Russell, J. D.; Coon, J. J.; Zabrouskov, V. Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures. *Anal. Chem.* **2009**, *81*, 7757-7765.

## CHAPTER 4

# LABEL-FREE QUANTITATION: A NEW GLYCOPROTEOMICS APPROACH

### 4.1 Introduction

Glycoproteomics, the study of the glycome attached to proteins, is a vital research field because as many as 50% of proteins in the human body are glycosylated.<sup>1</sup> Glycoproteins are not only common, but the glycans on the proteins are significant because they are known to play important biological roles in the body, including cell-cell interaction, cell recognition, and protein regulation.<sup>2</sup> While proteins are genetically encoded, the glycosylation on proteins depends on the glycosylating enzymes that are present and the local cell environment. Therefore, the amount of enzymes and cofactors involved in glycosylation affect the extent of glycosylation on the glycoprotein.<sup>3</sup> Changes in glycosylation are known to occur with the onset of certain diseases such as cancer.<sup>4-6</sup> Thus, detection methods to monitor changes in glycosylation of glycoproteins are essential to determine possible biomarkers for cancer and other diseases.

Methods to monitor changes in glycosylation of proteins are not just important for biomarker studies. These methods are also important for pharmaceutical development, since glycoproteins have become increasingly desirable targets as therapeutic agents. Some example glycoprotein-based pharmaceuticals include erythropoietin,<sup>7</sup> follicle stimulating hormone,<sup>8</sup> thyroid stimulating hormone,<sup>9</sup> and vaccine candidates, such as the heavily glycosylated envelope glycoprotein on the surface of the HIV virus.<sup>10-12</sup> Profiling and quantifying the glycosylation on these products is important because

studies indicate that glycosylation in cell expression systems can differ from the glycosylation that occurs in the human body,<sup>10</sup> and researchers are currently striving to overcome this problem by modifying the glycosylation, or humanizing it, during protein production.<sup>13, 14</sup> Therefore, as methods are developed for humanizing glycosylation, a quantitative method that distinguishes between glycosylation profiles on native and modified glycoproteins is imperative.

There are two options for quantifying glycosylation on proteins, quantifying the glycans after enzymatic or chemical cleavage from the protein, or quantifying glycopeptides. While glycan analysis is clearly a more established technique,<sup>15-18</sup> this approach restricts the amount of information one can obtain about the glycosylation profile. For example, in purified, multiply glycosylated proteins, the study of released glycans would only provide aggregate information about the glycosylation on a protein, and it would not provide information about the glycosylation profile at a specific glycosylation site. Yet, it is well established that monitoring glycosylation profiles at individual glycosylation sites is important because the glycosylation at particular sites in a protein can modulate the protein's structure, function, or metabolic clearance.<sup>19</sup> Since many therapeutic proteins are multiply glycosylated, including all the examples mentioned above, the analysis of released glycans is problematic in that it does not provide glycosylation site-specific information about the glycosylation profiles of the protein.

In biomarker analysis quantifying the released glycans has an additional disadvantage in that all the information about the proteins from which the glycans originated is lost. This introduces many problems in biomarker discovery, such as not

being able to identify whether or not the glycan's concentration increased because a protein containing that glycan was overexpressed or if one or several proteins' glycosylation profile changed, causing the glycan to be more abundant, even though the protein level(s) are not altered.<sup>4</sup> In contrast to glycan analysis, glycopeptide analysis provides glycosylation site-specific information for purified proteins<sup>17, 20, 21</sup> and it could potentially be useful in distinguishing between glycosylation changes and protein expression changes, since the protein information is encoded in the glycopeptide. For the reasons described herein, we are pursuing quantitative methods for glycopeptides.

There are two strategies to quantify changes in the glycosylation of proteins, either differentially labeling sets of samples or using label-free approaches. Several labeling techniques exist, including those with detection by optical methods<sup>22, 23</sup> and mass spectrometry.<sup>24-26</sup> One common quantitation strategy using labeling and optical detection involves the use of lectins to bind glycoproteins in complex samples, and detection of different types of glycosylation due to differential binding of the lectins. Because different lectins have different specificities for classes of *N*-linked glycans, it is possible to use lectins to distinguish between high mannose and complex type glycosylation, for example. Detecting the binding of lectins to glycans, glycopeptides, or glycoproteins is done by either tagging the lectins<sup>22, 27</sup> or tagging the analyte<sup>23</sup> with a fluorophore, followed by the monitoring of a change in fluorescence upon binding. If the ultimate quantitative goal is to detect changes in classes of *N*-linked glycans, optical methods that detect differences in lectin binding are ideal and have very low detection limits. However, the use of lectin microarrays is incapable of observing subtle

glycosylation changes, such as a change in the number of mannoses present on a high mannose glycan.<sup>23</sup>

If detecting subtle changes in glycosylation is required, such as distinguishing between the addition or subtraction of one monosaccharide unit between glycoprotein samples, other detection strategies, for example mass spectrometry, must be employed. Quantitative MS analysis of glycosylated species using isotopic labels is a growing field. Currently, methods are available to analyze glycans directly,<sup>25, 26, 28-31</sup> and the strategy used in some of these approaches could potentially be applied to glycopeptide analysis. The biggest draw-back is that many of these labeling methods are also limited to two sample sets, a control group and a test group; therefore, if more samples need to be compared to one another, such as analyzing the glycosylation differences between five different vaccine candidates, binary labeling approaches become difficult to implement.

An alternative strategy for accomplishing a quantitative analysis is to use label-free approaches. These methods have the potential to compare multiple samples with ease. Changes in intensities of mass spectral peaks have been assessed by comparing different sample sets through either glycan analysis<sup>4</sup> or glycopeptide analysis.<sup>32</sup> Because signal intensities can vary between mass spectrometric samples, label-free approaches are not as commonly used for quantitative analysis.<sup>26, 33</sup>

To alleviate much of the variation due to changes in MS response among samples, normalization of data has been applied in proteomics studies.<sup>34-36</sup> Old, *et. al.* employed a normalization technique that is easily adaptable to direct injection mass spectrometry by dividing intensities of individual peaks by the total intensity from all

peaks in the spectrum.<sup>33</sup> In the work presented here, this concept of normalization is built-upon to produce a label-free quantitative method for glycopeptide analysis. While in the proteomic field, one can obtain reliably quantitative data by normalizing the data to the total ion abundance,<sup>33</sup> this method is potentially problematic for glycopeptide analysis, because the glycopeptides ionize weakly, compared to the non-glycosylated peptides that may also be present. Therefore, small changes in the presence of a non-glycosylated interferent could impart large variability in a quantitative assay, when the total ion current is used to normalize the ion abundances of the analytes. To remedy this problem, a new normalization method is described herein, where the ion abundance from each glycopeptide is divided by the total intensity of all glycopeptide peaks present in a given spectrum (excluding all ions that are not assigned as glycopeptides). As demonstrated herein, this normalization produces reproducibly quantitative, label-free data.

The second major innovation of this work is using a two-tiered quantitative analysis. In the first tier of the analysis, the abundances of glycopeptide ions within a given sample are compared to each other. This internal analysis is used to generate a glycosylation profile for the sample, where the abundance of each glycoform is rank-ordered (from smallest to largest) within the sample. The second tier of the analysis involves comparing this generated glycopeptide profile from one sample to the profile of another sample. By comparing whole profiles, and not just the abundance of a given glycoform, one can readily discriminate between changes in the overall glycosylation profile of a protein and changes in the abundance of a given glycoprotein, present in a mixture of other species. The described method, which presents a new normalization

method custom-designed for the challenges of glycopeptide analysis, and a new approach to glycosylation profiling, where internal and external analyses are completed in parallel, is useful for those interested in glycosylation profiling of biopharmaceuticals as well as those quantifying mixtures of glycoproteins for various applications, including biomarker analysis.

## **4.2 Experimental**

### **4.2.1 Materials and Reagents**

All reagents were obtained in high purity from Sigma Aldrich except when noted otherwise. Ribonuclease B (RNase B) >80% pure, asialofetuin, urea,  $\alpha$ -mannosidase from *Canavalia ensiformis*, dithiothreitol (DTT), iodoacetamide (IAA), acetic acid, Sepharose<sup>®</sup> CL-4B, HPLC grade 1-butanol, and HPLC grade ethanol were all purchased from Sigma Aldrich (St. Louis). Sequencing grade modified trypsin was purchased from Promega (Madison, WI). Ammonium bicarbonate ( $\text{NH}_4\text{HCO}_3$ ) was purchased from Fluka (Milwaukee, WI). HPLC grade methanol was purchased from ThermoFisher Scientific (Fairlawn, NJ). Water was purified by a Millipore Direct Q-3 water purification system (Billerica, MA).

### **4.2.2 Enzymatic Glycan Trimming with $\alpha$ -mannosidase**

To analyze glycosylation change, one RNase B sample was subjected to cleavage by the enzyme  $\alpha$ -mannosidase as described by Toumi *et al.*<sup>37</sup> Briefly, approximately 300  $\mu\text{g}$  of RNase B was dissolved in enough 10 mM  $\text{NH}_4\text{HCO}_3$  (pH 5.0) to make 2 mg/mL.  $\alpha$ -mannosidase was added in an enzyme:protein ratio of 1:1000 (mol/mol). The sample was allowed to incubate for 24 hours in a 37 °C oven. Enough NaOH was added to raise the pH of the sample to approximately pH 8.0. The sample

was then treated as described below for protease digestion, glycopeptide enrichment, mass spectrometry, and data analysis.

#### **4.2.3 Glycoprotein Protease Digestion**

Approximately 300 µg of glycoprotein was dissolved in 25 mM NH<sub>4</sub>HCO<sub>3</sub> (pH 7.5 – 8.0) containing 4 M urea, to a glycoprotein concentration of 1 mg/mL (asialofetuin) or 2 mg/mL (RNase B). To this solution, dithiothreitol (DTT) was added to a final concentration of 15 mM, and it was incubated at room temperature for 1 hr. Following the incubation, iodoacetamide (IAA) was added to a final concentration of 25 mM, and the reaction was stored at room temperature, in the dark, for 1 hr. Additional DTT was added to a final concentration of 40 mM to neutralize excess IAA. The solution was diluted with 25 mM NH<sub>4</sub>HCO<sub>3</sub> until the urea concentration was less than 1 M. Trypsin was added at a 1:50 (w/w) protease/glycoprotein ratio. The solution was allowed to incubate at 37 °C for 18 hr and stopped by the addition of 1 µL acetic acid per 100 µL solution.

#### **4.2.4 Glycopeptide Enrichment**

To remove the nonglycosylated peptides from samples, a method adapted from Wada *et al.*<sup>38-40</sup> was used. Briefly, the digest solution was added to 800 µL of 5:1:1 (v/v) 1-butanol/ethanol/water and 25 µL Sepharose<sup>®</sup> CL-4B, and shaken gently for 45 minutes before centrifugation and extraction of the solution layer. Samples were washed twice with the addition of 1 mL 5:1:1 (v/v) 1-butanol/ethanol/water with gentle shaking for 5 minutes followed by the same centrifugation and extraction. After washing, 1 mL of 1:1 (v/v) ethanol/water was added, and samples were allowed to stand for 30 minutes, followed by gentle shaking for 30 minutes. The samples were

centrifuged, and the solution layer was extracted and collected. The ethanol/water extraction step was repeated a second time. The combined samples were dried using a Labconco centrivap cold trap (Kansas City, MO) and stored at -20 °C until use.

For the quality control studies, the samples were reconstituted prior to MS analysis in 1:1 (v/v) water/methanol containing 0.5% acetic acid, to a final glycopeptide concentration of 10  $\mu$ M. After the initial MS analysis, the remaining RNase B was stored in the reconstituted solvent at -20 °C for 8 weeks for a second MS analysis testing the method's robustness.

For the mixture analysis experiments, RNase B and asialofetuin glycopeptide digest samples were each reconstituted in 100  $\mu$ L 1:1 (v/v) water/methanol containing 0.5% acetic acid. The reconstituted samples of asialofetuin and RNase B glycopeptides were combined into four separate vials in varying concentrations. A total of 4 samples were prepared where the asialofetuin glycopeptides retained a fixed concentration of 1  $\mu$ M, and the RNase B glycopeptide concentration prepared at 1  $\mu$ M, 3  $\mu$ M, 5  $\mu$ M, and 10  $\mu$ M, respectively, in each mixture vial (See Table 1).

**Table 1.** Glycoprotein Mixture Concentrations

Sample #	1	2	3	4
RNase B	1 $\mu$ M	3 $\mu$ M	5 $\mu$ M	10 $\mu$ M
Asialofetuin	1 $\mu$ M	1 $\mu$ M	1 $\mu$ M	1 $\mu$ M

#### 4.2.5 Mass Spectrometry

MS and MS/MS data were acquired on a Thermo electrospray ionization - linear ion trap - Fourier transform ion cyclotron resonance - mass spectrometer, ESI-LTQ-FTICR-MS (San Jose, CA), containing a 7 Tesla actively shielded magnet. The samples were injected by direct infusion at a flow rate of 1  $\mu$ L/min in positive ion mode.

The spray voltage was optimized to maximize ion signal, and the value ranged between 2.8 and 4.0 kV. The nebulizing gas, N<sub>2</sub>, was set to 10 psi, and the capillary temperature was 200 °C. MS data were acquired with 100,000 resolution for *m/z* 400, over a mass range of *m/z* 500 – 2000 for RNase B and a mass range of *m/z* 800 – 2000 for asialofetuin. For all MS<sup>1</sup> data, 50 scans (with each containing 10 microscans) were averaged. For MS/MS data, the precursor ion was isolated with a 2 Da isolation range; the activation time was set to 30 ms, the activation *q<sub>z</sub>* was 0.250, and the activation energy was 30%, as defined by the instrument software. There were 20 to 30 scans (each containing 10 microscans) averaged during acquisition of MS/MS data. The instrument software used was Xcalibur version 1.4 SR1 (ThermoFisher Scientific San Jose, CA).

#### **4.2.6 Data Analysis**

The glycopeptide ions were assigned by matching theoretical masses to the actual masses acquired in the MS data. Prediction tables of possible theoretical glycopeptide masses were constructed for each glycoprotein studied. The prediction table was generated by the following steps: The amino acid sequence of each protein was obtained from Uniprot (<http://beta.uniprot.org>), and a theoretical tryptic digest of the given glycoprotein was completed by importing that sequence into Protein Prospector (<http://prospector.ucsf.edu/>), which calculates the possible tryptic fragments. The mass of the resulting tryptic fragments were adjusted to account for the alkylation of cysteine residues by iodoacetamide. The peptide masses that contained glycosylation sites were added to possible *N*-linked glycan masses to give predicted glycopeptide masses. The glycan masses used in this case were the known glycans that are appended to

these proteins, as described in references 41 and 42, for asialofetuin and RNase B, respectively. After combining the peptide masses with the known glycan masses, the calculated glycopeptide masses were converted to  $m/z$ 's (for the +1, +2, +3, +4, and +5 charge states) for comparison to the MS<sup>1</sup> data. A maximum of two missed tryptic cleavages was considered, as well as the presence of protonated and sodiated glycopeptide peaks. Possible peak identities from the MS data were confirmed through analysis of MS/MS data taken on each peak in the spectrum, as described previously.<sup>43</sup>

#### 4.2.7 Data Treatment for Quantitative Method

After identifying all the glycopeptide peaks in the spectrum, six steps were taken to process the data for quantitative analysis. The steps were: 1) Peak lists of  $m/z$  and intensity were generated in Xcalibur, and transferred to Microsoft Excel. 2) The first four isotopic peaks of each glycopeptide ion were summed to obtain each *glycopeptide's peak intensity*. 3) *Glycopeptide peak intensities* from all glycopeptide peaks in a spectrum were summed to calculate the *total glycopeptide intensity*. 4) The *glycopeptide peak percentage* was computed by Equation 4.1, below. This percentage reports how large each glycopeptide peak is, relative to the *total glycopeptide intensity*.

$$\text{GlycopeptidePeak\%} = \frac{\text{GlycopeptidePeakIntensity}}{\text{TotalGlycopeptideIntensity}} \times 100$$

Equation 4.1

5) *Glycopeptide peak percentages* corresponding to the same glycopeptide composition, but containing different charge states and/or charge carriers, were summed to give the *glycopeptide percentage*. If glycopeptides from a given glycosylation site also were generated with differing levels of missed tryptic cleavages,

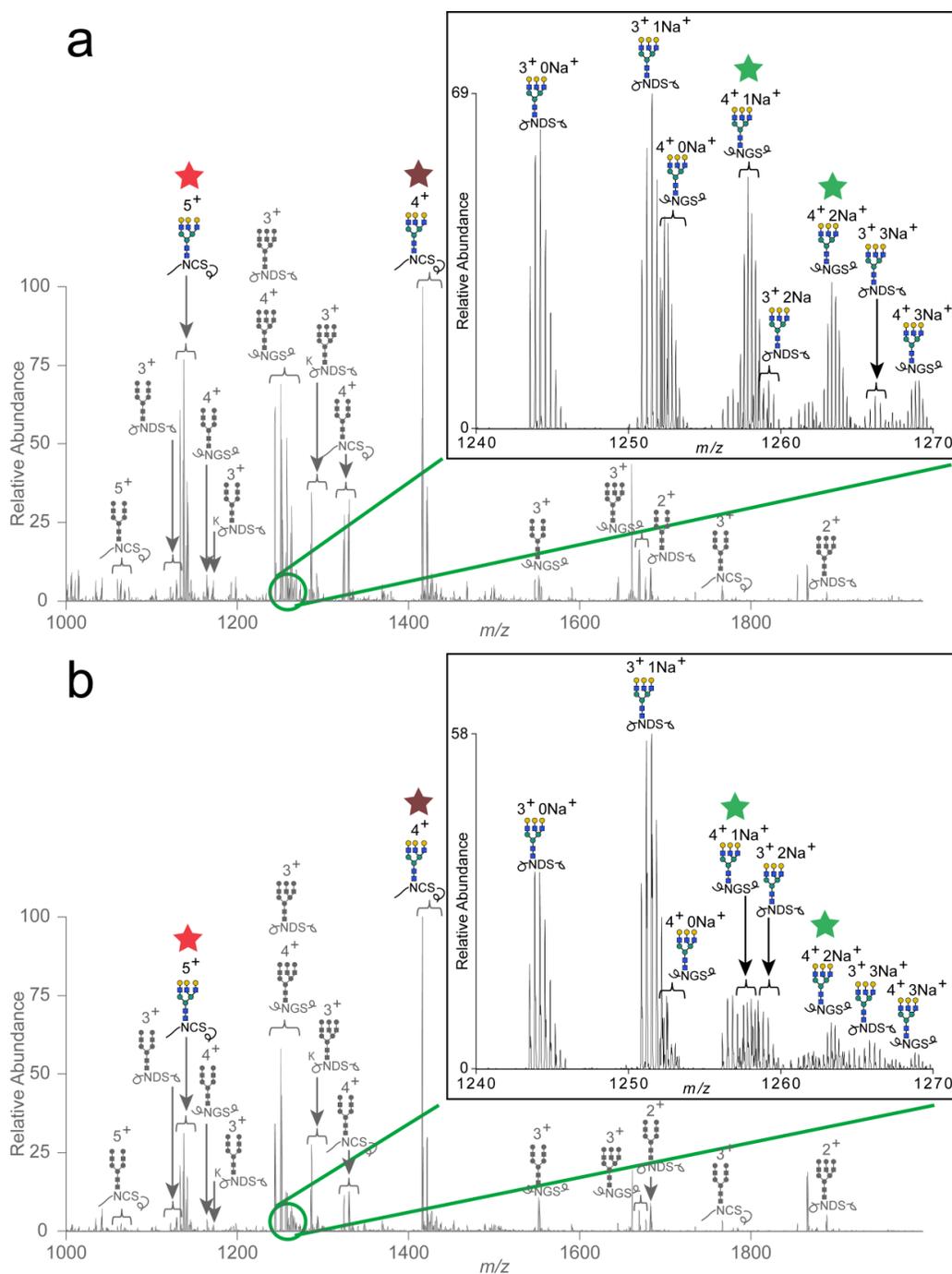
these species were also combined into one glycopeptide percentage. This percentage is a measure of how abundant a given glycopeptide composition is in the sample, regardless of whether or not it ionizes as a single peak, or as several peaks corresponding to different charge states, different numbers of sodiated adducts, or different lengths of peptide, due to missed tryptic cleavage. 6) *Glycopeptide percentages* were rank ordered to determine which glycopeptides were most abundant in the spectrum, and the rank order between different samples was compared.

### **4.3 Results and discussion**

#### **4.3.1 Developing a quantitative method for glycopeptides**

Since the goal is to develop a label-free quantitative approach for glycopeptides, the first problem that must be overcome is the fact that ion abundances in mass spectra are not very reproducible in run-to-run analyses. Figure 1 demonstrates this by showing MS data of glycopeptides generated from two replicate samples of glycopeptides generated from digesting asialofetuin with trypsin. The brown and red stars in Figure 1 label peaks that correspond to two different charge states of a single glycopeptide composition, the 4<sup>+</sup> and 5<sup>+</sup> charge states, respectively. The intensity of the peaks labeled with the brown stars do not change between Figure 1a and b, because this is the base peak in both spectra. The peaks labeled with red stars, on the other hand, have a much higher intensity in Figure 1a compared to Figure 1b, demonstrating that the ions partitioned differently into different charge states in the two analyses. The insets in Figure 1a and b show the intensities of sodiated adducts are also different, when the two spectra are compared. The peaks labeled with green stars, which

correspond to sodiated adducts of some of the glycopeptides, have much lower intensities in Figure 1b compared to Figure 1a.



**Figure 1.** (a) and (b) Mass spectra from two different glycopeptide samples of asialofetuin. Red and brown stars show the same glycopeptide in different charge states; the intensities of each of the charge states vary from sample to sample. Insets: a zoomed-in region of the spectrum containing multiple sodiated adducts of the glycopeptides. The green stars indicate sodiated adducts of glycopeptides whose intensity changes in spectrum a and b.

Consequently, the distribution between charge states and sodiated adducts of the glycopeptide peaks in Figure 1a is different than the distribution of the glycopeptide peaks in Figure 1b. Therefore, it makes sense that the peaks from Figure 1a and b would have poor reproducibility when compared by their relative abundance alone.

To alleviate the problem that changes in the form in which the peaks ionize leads to irreproducible peak intensity data, the relative abundances of the peaks containing the same glycopeptide composition but differing in the charge state and charge carrier were combined. By combining these values, the problem that the peaks partition differently into different charge states is mitigated. Once all the charge states and charge carriers for a particular glycopeptide composition are combined, we report the results as a percent of the total glycopeptide ion signal, and this value is henceforth referred to as the *glycopeptide percentage*. [Reporting the values as percentages of all the glycopeptides present in the spectrum is a normalization method used to mitigate the run-to-run ionization discrepancies that are known to be problematic in label-free quantitative approaches, as described in the introduction.] After obtaining the *glycopeptide percentages* for each of the different glycopeptide compositions, the compositions are ordered based on their percentage from smallest to largest (rank ordered), so an internal comparison of the glycan profile can be made. A list of all the identified glycopeptide ions from asialofetuin are shown in Table 2, and the chart is color-coded to show which species' ion abundances were combined. For example, all the blue entries correspond to glycopeptides with the peptide sequence RPTGEVYDIEIDTLETTCHVLDPTPLANCSVR. The darker blue indicates the biantennary glycans containing that amino acid sequence, and the lighter blue indicates

the triantennary glycans. The intensities of peaks with the same color are combined to produce a single glycopeptide percentage.

**Table 2.** Glycopeptide Peaks Identified in Asialofetuin FT-MS Data

Theoretical <i>m/z</i>	Observed <i>m/z</i>	Mass Error <sup>1</sup>	Charge State	# of Na	Peptide	Carbohydrate
1059.6757	1059.6767	0.9	5 <sup>+</sup>	0	RPTGEVYDIEIDTLETTCHVLDP TPLANCSVR	[Hex]5[HexNAc]4
1064.0722	1064.0783	5.7	5 <sup>+</sup>	1	"	"
1068.4687	1068.4746	5.5	5 <sup>+</sup>	2	"	"
1121.8122	1121.8183	5.4	3 <sup>+</sup>	0	LCPDCPLLAPLNSDR	"
1129.1397	1129.1438	3.6	3 <sup>+</sup>	1	"	"
1132.7022	1132.7059	3.3	5 <sup>+</sup>	0	RPTGEVYDIEIDTLETTCHVLDP TPLANCSVR	[Hex]6[HexNAc]5
1137.0987	1137.1055	6.0	5 <sup>+</sup>	1	"	"
1141.4952	1141.5009	5.0	5 <sup>+</sup>	2	"	"
1145.8917	1145.8944	2.4	5 <sup>+</sup>	3	"	"
1150.2882	1150.2762	10.4	5 <sup>+</sup>	4	"	"
1160.5443	1160.5529	7.4	4 <sup>+</sup>	0	VVHAVEVALATFNAESNGSYLQ LVEISR	[Hex]5[HexNAc]4
1164.5106	1164.5156	4.3	3 <sup>+</sup>	0	KLCPDCPLLAPLNSDR	"
1166.0400	1166.0479	6.8	4 <sup>+</sup>	1	VVHAVEVALATFNAESNGSYLQ LVEISR	"
1171.5356	1171.5389	2.8	4 <sup>+</sup>	2	"	"
1171.8381	1171.8426	3.8	3 <sup>+</sup>	1	KLCPDCPLLAPLNSDR	"
1243.5230	1243.5300	5.6	3 <sup>+</sup>	0	LCPDCPLLAPLNSDR	[Hex]6[HexNAc]5
1250.8505	1250.8559	4.3	3 <sup>+</sup>	1	"	"
1251.8274	1251.8252	1.8	4 <sup>+</sup>	0	VVHAVEVALATFNAESNGSYLQ LVEISR	"
1257.3230	1257.3286	4.5	4 <sup>+</sup>	1	"	"
1258.1780	1258.1814	2.7	3 <sup>+</sup>	2	LCPDCPLLAPLNSDR	"
1262.8186	1262.8220	2.7	4 <sup>+</sup>	2	VVHAVEVALATFNAESNGSYLQ LVEISR	"
1265.5055	1265.5081	2.1	3 <sup>+</sup>	3	LCPDCPLLAPLNSDR	"
1268.3143	1268.3221	6.1	4 <sup>+</sup>	3	VVHAVEVALATFNAESNGSYLQ LVEISR	"
1273.8099	1273.8070	2.3	4 <sup>+</sup>	4	"	"
1286.2213	1286.2286	5.7	3 <sup>+</sup>	0	KLCPDCPLLAPLNSDR	"
1293.5488	1293.5582	7.3	3 <sup>+</sup>	1	"	"
1300.8763	1300.8730	2.5	3 <sup>+</sup>	2	"	"
1308.2038	1308.2108	5.4	3 <sup>+</sup>	3	"	"
1324.3428	1324.3526	7.4	4 <sup>+</sup>	0	RPTGEVYDIEIDTLETTCHVLDP TPLANCSVR	[Hex]5[HexNAc]4
1329.8385	1329.8471	6.5	4 <sup>+</sup>	1	"	"
1335.3341	1335.3426	6.4	4 <sup>+</sup>	2	"	"
1415.6259	1415.6355	6.8	4 <sup>+</sup>	0	"	[Hex]6[HexNAc]5
1421.1215	1421.1297	5.8	4 <sup>+</sup>	1	"	"
1426.6171	1426.6263	6.4	4 <sup>+</sup>	2	"	"
1432.1128	1432.1185	4.0	4 <sup>+</sup>	3	"	"
1437.6084	1437.6084	0.0	4 <sup>+</sup>	4	"	"
1547.0587	1547.0695	7.0	3 <sup>+</sup>	0	VVHAVEVALATFNAESNGSYLQ LVEISR	[Hex]5[HexNAc]4
1554.3842	1554.3688	9.9	3 <sup>+</sup>	1	"	"
1668.7674	1668.7771	5.8	3 <sup>+</sup>	0	"	[Hex]6[HexNAc]5
1676.0949	1676.1073	7.4	3 <sup>+</sup>	1	"	"

1682.2147	1682.2204	3.4	2 <sup>+</sup>	0	LCPDCPLLAPLNSDR	[Hex]5[HexNAc]4
1683.4224	1683.4358	8.0	3 <sup>+</sup>	2	VVHAVEVALATFNAESNGSYLQ LVEISR	[Hex]6[HexNAc]5
1765.4547	1765.4575	1.6	3 <sup>+</sup>	0	RPTGEVYDIEIDTLETTCHVLDP TPLANCSVR	[Hex]5[HexNAc]4
1864.7808	1864.7876	3.6	2 <sup>+</sup>	0	LCPDCPLLAPLNSDR	[Hex]6[HexNAc]5

<sup>1</sup>Mass error is reported in ppm.

The color of the rows (blue, red, or green) represent the three glycosylation sites of asialofetuin, whereas the darker and lighter shades of a given color correspond to biantennary and triantennary glycans, respectively. Ion abundances corresponding to the same color and shade are added together to generate the “glycopeptide percentage” for each species.

Figure 2 illustrates the benefit of combining the different charge states and charge carriers for the asialofetuin glycopeptide data. Figure 2a shows a portion of the rank order of ion abundances between the two asialofetuin samples from Figure 1 when the ion abundances are *not* combined, prior to ranking the glycopeptides from smallest to largest. The rank order is clearly very different between the two samples. However, once the different charge states and charge carriers for each of the glycopeptide compositions are combined, the rank order is highly reproducible between the two samples (Figure 2b).

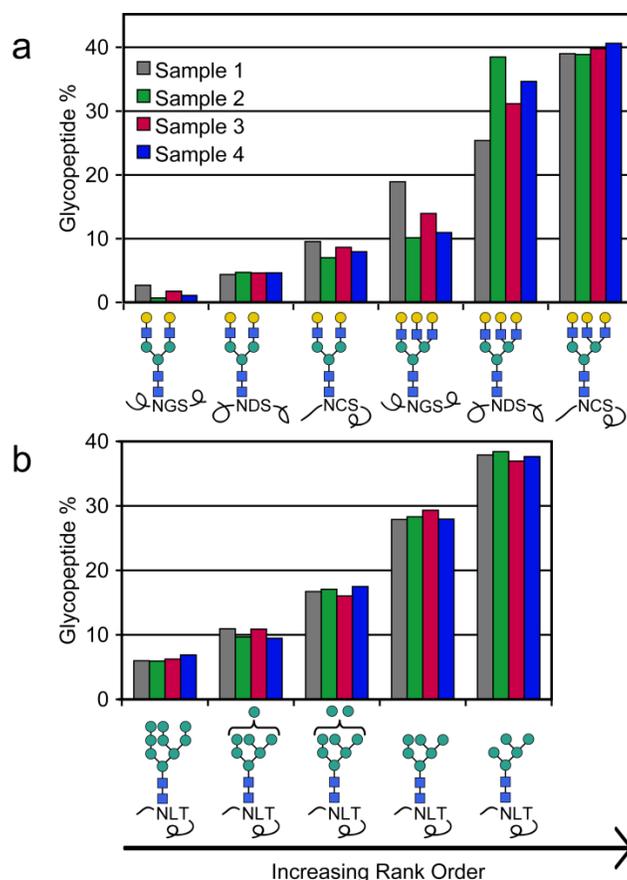
Glycopeptide Composition	Sample 1		Sample 2		After Combining By Species	Sample 1		Sample 2	
	Rank Order	Rank Order	Rank Order	Rank Order		Rank Order	Rank Order	Rank Order	Rank Order
	38	43				1	1		
	39	44				2	2		
	40	39				3	3		
	41	33				4	4		
	42	38				5	5		
	43	40				6	6		
	44	45							
	45	47							
	46	42							
	47	46							

**Figure 2.** (a) Rank order of glycopeptide ions from the data in Figure one. (b) Rank order of glycopeptide compositions, after combining all charge states and sodium adducts of ions with the same composition, prior to rank-ordering the components. Reproducible glycosylation profiles are achieved only in 2(b), when charge states and sodium adducts of the same species are combined.

#### 4.3.2 Quality Control Experiment 1 – Robustness

Since the goal was to develop a robust quantitation method that can detect changes in glycopeptides' intensities for different samples, one important feature is that the method must produce the same results for the same glycoprotein batch. For

example, small differences that could be introduced during digestion or glycopeptide enrichment and dilution for MS analysis, should not cause changes in the quantitation results. Otherwise, the method would not be robust. To test robustness, a quality control experiment was developed where four replicate samples of asialofetuin were each digested and prepared separately. After each sample was subjected to MS analysis, the *glycopeptide percentages* were calculated as described above, and were subsequently rank ordered, by percentage, from smallest to largest. The results of this experiment are displayed in Figure 3a. It is evident from the figure that the same rank order was observed for all four samples. However, the percentages between each sample tended to vary slightly, albeit not enough to change the rank order. This variability in percentage of asialofetuin glycopeptides is attributed to change in the distribution of sodiated adducts between samples. Different numbers of sodiated adducts can have slightly different ionization efficiencies. Therefore, when samples produce spectra with different intensities of sodiated adducts, combining the percentages of the different sodiated adducts can introduce a small variability in the percentages of each glycopeptide composition. This is not a significant problem, however, because the changes are small enough that the rank order did not vary between the samples.



**Figure 3.** Similarity of the glycopeptide profiles for asialofetuin (a) and RNase B (b). The percentage of each glycopeptide composition is plotted for four replicate samples. The glycopeptide compositions are shown on the x-axis in order of their abundance in the spectrum. In each case, the rank order (smallest percentage to largest percentage) does not change; small fluctuations are observed in the actual percentage of each glycopeptide composition among the four replicate samples.

#### 4.3.3 Quality Control Experiment 2 – Applicability to Different Glycoproteins

Once the method was confirmed to produce reproducible results for asialofetuin, a second glycoprotein with very different properties than asialofetuin was analyzed to ensure the method is applicable for a wide variety of glycoproteins. RNase B was chosen for the analysis because it is much smaller than asialofetuin; it has only one glycosylation site; and it has a different type of glycosylation, high mannose type glycans. See Table 3. RNase B was subjected to the same sample preparation conditions as asialofetuin. Four replicate samples of RNase B were digested and

analyzed. Unlike the asialofetuin data, RNase B glycopeptides did not ionize as sodiated adducts. The RNase B glycopeptides did ionize in multiple charge states, and the spectra also contained peaks corresponding to missed tryptic cleavages of the glycopeptides (data not shown). The missed cleavages are likely the result of the glycosylation blocking the cleavage site, as described earlier,<sup>44</sup> since several arginine and lysine residues are located very near the glycosylation site. See the amino acid sequence in Table 3. These missed tryptic cleavages could potentially interfere with

**Table 3.** Comparison of Glycoproteins

<b>RNase B</b>	<b>Asialofetuin</b>
MALKSLVLLSLLVLLVLLVLRVQ	MKSFVLLFCLAQLWGCHSIPLDPVAGYKEPACDDPDTEQAALAADVYINKHLP
PSLGKETAAAKFERQHMSSTS	RGYKHTLNQIDSVKWPWRPTGEVYDIEIDTLETTCHVLDPTPLANCSVRQQT
AASSNYCNQMMKSRNLTKDRC	QHAVEGDCDIHVLKQDGFVSVLFTKCDSSPDSAEDVRKLCPCDPLLAPLNDLR
KPVNTFVHESLADVQAVCSQKN	VVHAVEVALATFNAESNGSYLQLVEISR
VACKNGQTNCYQSYSTMSITDC	AQFVPLPVSVSVEFAVAATDCIAKE
RETGSSKYPNCAYKTTQANKHI	VVDPTKCNLLAEKQYGFCKGSVIQKALGGEDVRVTCTLFQTQPVIPQPQPDGA
IVACEGNPYVPVHFDASV	EAEAPSAVPDAAGPTPSAAGPPVASVVVGPSVVAVPLPLHRAHYDLRHTFSGV
	ASVSSSGEAFHVGKTPIVGQPSIPGGPVRLCPGRIRYFKI
16.5 kDa	38.5 kDa
1 N-linked glycosylation site	3 N-linked glycosylation sites
High mannose glycans	Complex glycans

reproducible quantitative analysis, if the digestions do not generate identical proportions of peptides with missed tryptic cleavages near the glycosylation sites. To mitigate the potential for quantitative error due to differences in the digestion, glycopeptide peaks corresponding to a given glycosylation site and a given glycan composition were combined with species that contained the same glycan composition and glycosylation site, but different levels of missed tryptic cleavages. The results from the RNase B experiments are illustrated in Figure 3b. The RNase B glycopeptide data also exhibits a consistent rank order among the replicate samples.

#### 4.3.4 Quality Control Experiment 3 – Instrument precision

The use of a mass spectrometer over extended time frames can lead to reproducibility problems in label-free quantitative assays. Therefore, to ensure that minor changes in the instrument conditions do not lead to inaccuracies in assigning the rank order of glycopeptides, the four samples of glycopeptides from RNase B were run on two different dates, eight weeks apart. Between analyses, the samples were stored at -20 °C. Under these conditions, the RNase B glycopeptides do not degrade; therefore any changes in the rank order of the glycopeptide compositions would be attributed to instrument variability over time.<sup>45</sup> Prior to using our quantitative method on the two sets of data, we first determined whether or not the mass spectra showed deviations in peak intensities between weeks zero and eight, by comparing the raw ion abundances for each of the glycopeptides. Table 4 shows the results for the relative abundances of several peaks that were acquired from one RNase B glycopeptide sample, before and after the 8 week storage conditions. The relative abundance increases in the 3<sup>+</sup> charge state in week 8 compared to week 0, while the relative abundance in the 2<sup>+</sup> charge state decreases in week 8, compared to week 0. This is another example that shows the ions can partition themselves differently into different charge states. In this case, drastically different ion abundances were acquired for the exact same sample. Fortunately, the quantitative method described herein is designed to accommodate this variability by combining *glycopeptide peak percentages* for the same glycopeptide, partitioned into different charge states. After combining the

**Table 4.** Differences in relative abundance for glycopeptide ions from RNase B:

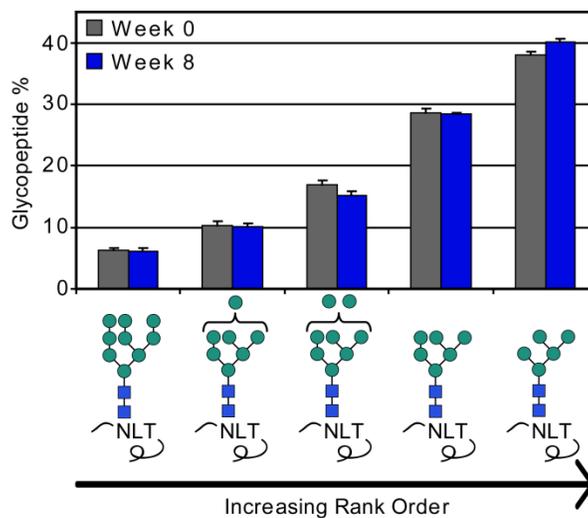
Charge State	Composition	<i>m/z</i>	Relative Abundance Week 0 <sup>a</sup>	Relative Abundance Week 8 <sup>a</sup>
3 <sup>+</sup>	[Hex]5[HexNAc]2 + SRNLTKDR	735.9958	32.7	77.6
3 <sup>+</sup>	[Hex]6[HexNAc]2 + SRNLTKDR	790.0134	34.2	66.8
3 <sup>+</sup>	[Hex]7[HexNAc]2 + SRNLTKDR	844.0310	14.8	31.1
3 <sup>+</sup>	[Hex]8[HexNAc]2 + SRNLTKDR	898.0486	36.1	50.8
3 <sup>+</sup>	[Hex]9[HexNAc]2 + SRNLTKDR	952.0662	16.1	21.0
2 <sup>+</sup>	[Hex]5[HexNAc]2 + SRNLTKDR	1103.489 9	67.4	40.0
2 <sup>+</sup>	[Hex]6[HexNAc]2 + SRNLTKDR	1184.516 3	30.1	15.7
2 <sup>+</sup>	[Hex]7[HexNAc]2 + SRNLTKDR	1265.542 7	6.4	3.3
2 <sup>+</sup>	[Hex]8[HexNAc]2 + SRNLTKDR	1346.569 1	5.7	3.8
2 <sup>+</sup>	[Hex]9[HexNAc]2 + SRNLTKDR	1427.595 5	0.8	0.6

<sup>a</sup>Two data sets using the same sample, analyzed eight weeks apart.

glycopeptide peak percentages as described above, the rank order of RNase B glycopeptides was obtained, and the data is shown in Figure 4 for the two different time points. The *glycopeptide percentages* for RNase B glycopeptides illustrated in Figure 4 show high reproducibility, small standard deviations, and the rank order is retained between the two runs.

The above experiments demonstrate that for a purified sample, the quantitative method described produces reproducible data, even under different instrumental conditions, and the reproducibility is unaltered after repeating the protease digestion and sample preparation conditions. Also, the method successfully analyzed two different glycoproteins that had a variety of different features, including varying numbers of glycosylation sites and different types of glycosylation. In summary, these studies demonstrate that the quantitative method described would be useful for classifying

glycosylation changes in purified proteins, which is useful in a variety of biopharmaceutical applications, as described in the introduction.



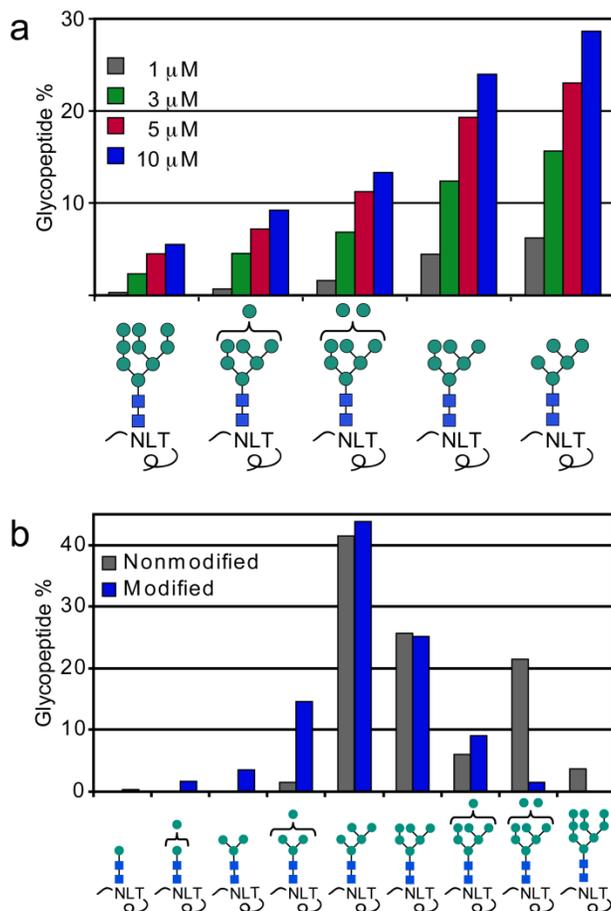
**Figure 4.** Graphical depiction of the quantified amounts of RNase B glycopeptides at two different time points. The mean from four digest samples was plotted. Error bars indicate the standard deviation. The glycopeptides are plotted from left to right in order of increasing abundance. The rank order does not change, even when the sample is re-analyzed eight weeks after the original analysis date.

#### 4.3.5 Mixture Analysis

In addition to characterizing the glycosylation on biopharmaceuticals, it would be ideal if this method of quantifying glycosylation profiles could also be used in other types of applications, for example in studies where a mixture of proteins is present. When a mixture of glycopeptides is analyzed, one current roadblock is being able to determine why the particular glycopeptide ion has changed in abundance, i.e. is it due to changes in glycosylation on a given protein or due to changes in the protein's net concentration, relative to the other species being analyzed?<sup>4</sup> Our method, which characterizes the entire glycosylation profile for a given glycoprotein as part of the quantitation process, is ideally suited to solving this problem.

To demonstrate that the quantitative method described herein can also distinguish between glycosylation changes and net protein abundance changes when proteins are present in a mixture, glycopeptides from the two proteins described above were combined in several different ratios, and the resulting samples were analyzed. In each case, glycopeptides from asialofetuin were present at a concentration of 1  $\mu\text{M}$ , while the glycopeptides from RNase B were present at varying concentrations, between 1 and 10  $\mu\text{M}$ . (See Table 1). MS data of the mixed samples was acquired, and the rank order for each of the glycopeptides present was obtained, as described previously. The quantitative results are shown in Figure 5. This Figure shows data for the RNase B glycopeptides for the four samples. Regardless of the relative concentrations of the two proteins, the rank order of the glycans for the given glycoprotein, RNase B, did not change. In every case, the Man9 glycopeptides from RNase B were present in lowest abundance and the Man5 glycopeptides were present in highest abundance, among the RNase B glycoforms. This demonstrates that varying the concentration of protein does not impact the rank order of the glycopeptides for a given glycoprotein. In addition, the order observed in Figure 5 is the same order as observed for RNase B alone, (in Figure 3b and 4), demonstrating that the presence of other proteins does not interfere with the method's ability to reproducibly rank order the glycopeptides. Most importantly, when one compares the data for a given glycopeptide, for example, Man5, the data in Figure 5 clearly demonstrate that the glycopeptide percentages increase as the concentration of the protein increases. Therefore, the data described herein can clearly distinguish between changes in a glycosylation profile and changes in a protein's concentration. For the four samples in Figure 5, the glycosylation profile was identical; yet the

concentrations of a given glycopeptide increased as the protein concentration increased. If the concentration of the protein remained the same but the glycan profile had been altered, the rank order of glycans would vary between the samples, but the overall glycopeptide concentration for the sum of all the glycopeptides of a given protein would remain approximately the same.



**Figure 5.** (a) Glycopeptide profiles for RNase B in mixtures of RNase B and asialofetuin. Glycopeptide percentages of all the RNase B glycoforms are increased as the concentration of RNase B is increased in the mixture; rank order of the glycoforms is conserved. (b) RNase B before and after cleavage with  $\alpha$ -mannosidase. Man8 and Man9 decreased in glycopeptide percentage, while Man1, Man2, Man3, and Man4 increased in glycopeptide percentage.

#### 4.3.6 Monitoring Changes in Glycosylation

The quantitative method proposed herein has been subjected to many experimental conditions described above to ensure that is reproducible, robust, and that

changes in glycoprotein concentration can be assessed. One additional experiment was completed to ensure that this method can indeed quantify changes in glycosylation. RNase B was subjected to a commercially available enzyme capable of cleaving mannose residues,  $\alpha$ -mannosidase from *Canavalia ensiformis*. For a complete digestion with this enzyme, a 72 hour incubation period was needed.<sup>37</sup> A partial digestion was more appropriate for analyzing changes in glycosylation, therefore the enzyme was only allowed to incubate with RNase B for 24 hours. Two experiments were performed simultaneously; one where RNase B was allowed to be digested with  $\alpha$ -mannosidase and a second experiment with RNase B unmodified. Figure 5b illustrates the results from this experiment. As can be deduced from the figure, the *glycopeptide percentage* from Man8 and Man9 were reduced, and the *glycopeptide percentage* from Man1, Man2, Man3, and Man4 were increased, as expected with an enzyme that cleavage glycan residues. This clearly shows that the quantitative method is capable of monitoring changes in glycosylation.

This quantitative method is similar to other analytical methods, in that there are expectations and limitations to this work. First of all, it is expected that other methods will be used to identify all the possible glycopeptides before using the quantitative method.<sup>12, 17, 20, 40, 43, 46, 47</sup> Obviously, without proper identification of the glycopeptides, the glycosylation profiles may be different when comparing multiple samples. This method is only applicable to glycoproteins containing neutral glycans, as mixtures of sialylated, other negatively charged glycan species, and neutral glycans typically require the use of both positive and negative mode to accurately identify such species. Using both modes would be extremely difficult for accurate glycosylation profiles to be

constructed. If certain glycoprotein samples were known to have varying levels of glycosylation site occupancy, the comparison between samples may produce similar results as is seen with glycoprotein concentration changes. Thus, this method cannot distinguish changes in site occupancy from glycoprotein concentration changes. Even though this method has some limitations, its utility for quantifying glycosylation profiles has many uses in glycoprotein analysis.

#### **4.4 Concluding Remarks**

This manuscript describes a new label-free quantitative method that can be applied to purified proteins as well as glycoprotein mixtures. In the mixture analysis, the method is able to distinguish between glycoprotein concentration changes and changes in glycosylation. The method was validated with several control experiments. The first control experiment analyzed replicate samples of glycopeptides from one glycoprotein, asialofetuin. The results from this experiment illustrate that the rank order of glycopeptides is consistent in all replicate samples, with slight variation in *glycopeptide percentages*. To be confident that the method would be applicable to a large set of glycoproteins, a second glycoprotein, RNase B, with very different properties was analyzed. RNase B results are also very consistent among four replicate samples, and the rank order of the glycoforms is retained among the replicate samples. Because the replicate samples in the two glycoproteins studied were digested in different vials before analysis, the results demonstrated that minor changes in digestion conditions did not alter the rank order. The third quality control experiment measured the ability of the method to tolerate small changes in the instrument conditions. The same RNase B glycopeptide samples were run on two different dates, 8 weeks apart, and subsequently

analyzed with the quantitative method. Similar RNase B glycopeptide percentages were observed, with no change in rank order; therefore similar samples can confidently be analyzed by this method at different times. This new quantitative method would be useful for anyone studying glycosylation profiles of proteins, either as purified proteins (as in the case of pharmaceutical development) or as glycoprotein mixtures, such as in the search for glycan-based biomarkers.

### **Acknowledgements**

The authors acknowledge the National Institutes of Health for funding (project number RO1GM077266) and the National Science Foundation (project number 0645120).

## 4.5 References

- (1) Budnik, B. A.; Lee, R. S.; Steen, J. A. Global Methods for Protein Glycosylation Analysis by Mass Spectrometry. *Biochim. Biophys. Acta* **2006**, *1764*, 1870-1880.
- (2) Morelle, W.; Canis, K.; Chirat, F.; Faid, V.; Michalski, J. -C. The Use of Mass Spectrometry for the Proteomic Analysis of Glycosylation. *Proteomics* **2006**, *6*, 3993-4015.
- (3) Hayes, B. K.; Freeze, H. H.; Varki, A. Biosynthesis of Oligosaccharides in Intact Golgi Preparations from Rat Liver. Analysis of *N*-linked Glycans by UDP-[6-<sup>3</sup>H]*N*-acetylglucosamine. *J. Biol. Chem.* **1993**, *268*, 16139-16154.
- (4) Zhao, J.; Qiu, W.; Simeone, D. M.; Lubman, D. M. *N*-linked Glycosylation Profiling of Pancreatic Cancer Serum Using Capillary Liquid Phase Separation Coupled with Mass Spectrometric Analysis. *J. Proteome Res.* **2007**, *6*, 1126-1138.
- (5) Orntoft, T. F.; Vestergaard, E. M. Clinical Aspects of Altered Glycosylation of Glycoproteins in Cancer. *Electrophoresis* **1999**, *20*, 362-371.
- (6) Kim, Y. J.; Varki, A. Perspectives on the Significance of Altered Glycosylation of Glycoproteins in Cancer. *Glycoconjugate J.* **1997**, *14*, 569-576.
- (7) Maiese, K.; Li, F.; Chong, Z. Z. Erythropoietin in the Brain: Can the Promise to Protect be Fulfilled? *Trends Pharmacol. Sci.* **2004**, *25*, 577-583.
- (8) Loumaye, E.; Dreano, M.; Galazka, A.; Howles, C.; Ham, L.; Munafo, A.; Eshkol, A.; Giudice, E.; De Luca, E.; Sirna, A.; Antonetti, F.; Giartosio, C. -E.; Scaglia, L.; Kelton, C.; Campbell, R.; Chappel, S. Recombinant Follicle Stimulating Hormone: Development of the First Biotechnology Product for the Treatment of Infertility. *Hum. Reprod. Update* **1998**, *4*, 862-881.
- (9) Szkudlinski, M. W.; Thotakura, N. R.; Bucci, I.; Joshi, L. R.; Tsai, A.; East-Palmer, J.; Shiloach, J.; Weintraub, B. D. Purification and Characterization of Recombinant Human Thyrotropin (TSH) Isoforms Produced by Chinese Hamster Ovary Cells: The Role of Sialylation and Sulfation in TSH Bioactivity. *Endocrinology* **1993**, *133*, 1490-1503.
- (10) Hamilton, S. R.; Davidson, R. C.; Sethuraman, N.; Nett, J. H.; Jiang, Y.; Rios, S.; Bobrowicz, P.; Stadheim, T. A.; Li, H.; Choi, B. -K.; Hopkins, D.; Wischnewski, H.; Roser, J.; Mitchell, T.; Strawbridge, R. R.; Hoopes, J.; Wildt, S.; Gerngross, T. U. Humanization of Yeast to Produce Complex Terminally Sialylated Glycoproteins. *Science* **2006**, *313*, 1441-1443.

- (11) Burton, D. R.; Desrosiers, R. C.; Doms, R. W.; Koff, W. C.; Kwong, P. D.; Moore, J. P.; Nabel, G. J.; Sodroski, J.; Wilson, I. A.; Wyatt, R. T. HIV Vaccine Design and the Neutralizing Antibody Problem. *Nat. Immunol.* **2004**, *5*, 233-236.
- (12) Go, E. P.; Irungu, J.; Zhang, Y.; Dalpathado, D. S.; Liao, H. -X.; Sutherland, L. L.; Alam, S. M.; Haynes, B. F.; Desaire, H. Glycosylation Site-Specific Analysis of HIV Envelope Proteins (JR-FL and CON-S) Reveals Major Differences in Glycosylation Site Occupancy, Glycoform Profiles, and Antigenic Epitopes' Accessibility. *J. Proteome Res.* **2008**, *7*, 1660-1674.
- (13) Gerngross, T. U. Advances in the Production of Human Therapeutic Proteins in Yeasts and Filamentous Fungi. *Nat. Biotechnol.* **2004**, *22*, 1409-1414.
- (14) Wong, C. -H. Protein Glycosylation: New Challenges and Opportunities. *J. Org. Chem.* **2005**, *70*, 4219-4225.
- (15) Wührer, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H. Glycoproteomics Based on Tandem Mass Spectrometry of Glycopeptides. *J. Chromatogr., B* **2007**, *849*, 115-128.
- (16) Mechref, Y.; Novotny, M. V. Structural Investigations of Glycoconjugates at High Sensitivity. *Chem. Rev.* **2002**, *102*, 321-369.
- (17) Dalpathado, D. S.; Desaire, H. Glycopeptide Analysis by Mass Spectrometry. *Analyst (Cambridge, U. K.)* **2008**, *133*, 731-738.
- (18) Joenväärä, S.; Ritamo, I.; Peltoniemi, H.; Renkonen, R. N-glycoproteomics – An Automated Workflow Approach. *Glycobiology* **2008**, *18*, 339-349.
- (19) Varki, A. Biological Roles of Oligosaccharides: All of the Theories Are Correct. *Glycobiology* **1993**, *3*, 97-130.
- (20) Dalpathado, D. S., Irungu, J., Go, E. P., Norton, K., Bousfield, G. R., Desaire, H. Comparative Glycomics of the Glycoprotein Follicle-Stimulating Hormone (FSH): Glycopeptide Analysis of Isolates from Two Mammalian Species *Biochemistry.* **2006**, *45*, 8665-8673.
- (21) Irungu, J., Dalpathado, D. S., Go, E. P., Jiang, H., Ha, H. V., Bousfield, G. R., Desaire, H. A Method for Characterizing Sulfated Glycoproteins in a Glycosylation Site-Specific Fashion, Using Ion-Pairing and Tandem Mass Spectrometry. *Anal Chem.* **2006**, *78*, 1181-1190.
- (22) Patwa, T. H.; Zhao, J.; Anderson, M. A.; Simeone, D. M.; Lubman, D. M. Screening of Glycosylation Patterns in Serum Using Natural Glycoprotein Microarrays and Multi-Lectin Fluorescence Detection. *Anal. Chem.* **2006**, *78*, 6411-6421.

- (23) Pilobello, K. T.; Krishnamoorthy, L.; Slawek, D.; Mahal, L. K. Development of a Lectin Microarray for the Rapid Analysis of Protein Glycopatterns. *ChemBiochem* **2005**, *6*, 985-989.
- (24) Xiong, L.; Andrews, D.; Regnier, F. Comparative Proteomics of Glycoproteins Based on Lectin Selection and Isotope Coding. *J. Proteome Res.* **2003**, *2*, 618-625.
- (25) Zhang, H.; Li, X. -J.; Martin, D. B.; Aebersold, R. Identification and Quantification of *N*-linked Glycoproteins Using Hydrazide Chemistry, Stable Isotope Labeling and Mass Spectrometry. *Nat. Biotechnol.* **2003**, *21*, 660-666.
- (26) Atwood, J. A., III; Cheng, L.; Alvarez-Manilla, G.; Warren, N. L.; York, W. S.; Orlando, R. Quantitation by Isobaric Labeling: Applications to Glycomics. *J. Proteome Res.* **2008**, *7*, 367-374.
- (27) Zhao, J.; Patwa, T. H.; Qiu, W.; Shedden, K.; Hinderer, R.; Misek, D. E.; Anderson, M. A.; Simeone, D. M.; Lubman, D. M. Glycoprotein Microarrays with Multi-Lectin Detection: Unique Lectin Binding Patterns as a Tool for Classifying Normal, Chronic Pancreatitis and Pancreatic Cancer Sera. *J. Proteome Res.* **2007**, *6*, 1864-1874.
- (28) Alvarez-Manilla, G.; Warren, N. L.; Abney, T.; Atwood, J., III; Azadi, P.; York, W. S.; Pierce, M.; Orlando, R. Tools for Glycomics: Relative Quantitation of Glycans by Isotopic Permethylation Using  $^{13}\text{CH}_3\text{I}$ . *Glycobiology* **2007**, *17*, 677-687.
- (29) Uematsu, R.; Furukawa, J. -I.; Nakagawa, H.; Shinohara, Y.; Deguchi, K.; Monde, K.; Nishimura, S. -I. High Throughput Quantitative Glycomics and Glycoform-focused Proteomics of Murine Dermis and Epidermis. *Mol. Cell. Proteomics* **2005**, *4*, 1977-1989.
- (30) Kita, Y.; Miura, Y.; Furukawa, J. -I.; Nakano, M.; Shinohara, Y.; Ohno, M.; Takimoto, A.; Nishimura, S. -I. Quantitative Glycomics of Human Whole Serum Glycoproteins Based on the Standardized Protocol for Liberating *N*-glycans. *Mol. Cell. Proteomics* **2007**, *6*, 1437-1445.
- (31) Furukawa, J. -I.; Shinohara, Y.; Kuramoto, H.; Miura, Y.; Shimaoka, H.; Kuroguchi, M.; Nakano, M.; Nishimura, S. -I. Comprehensive Approach to Structural and Functional Glycomics Based on Chemoselective Glycoblotting and Sequential Tag Conversion. *Anal. Chem.* **2008**, *80*, 1094-1101.
- (32) Wuhler, M.; Stam, J. C.; van de Geijn, F. E.; Koeleman, C. A. M.; Verrips, C. T.; Dolhain, R. J. E. M.; Hokke, C. H.; Deelder, A. M. Glycosylation Profiling of Immunoglobulin G (IgG) Subclasses from Human Serum. *Proteomics* **2007**, *7*, 4070-4081.

- (33) Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G. Comparison of Label-free Methods for Quantifying Human Proteins by Shotgun Proteomics. *Mol. Cell. Proteomics* **2005**, *4*, 1487-1502.
- (34) Wang, W.; Zhou, H.; Lin, H.; Roy, S.; Shaler, T. A.; Hill, L. R.; Norton, S.; Kumar, P.; Anderle, M.; Becker, C. H. Quantification of Proteins and Metabolites by Mass Spectrometry without Isotopic Labeling or Spiked Standards. *Anal. Chem.* **2003**, *75*, 4818-4826.
- (35) Callister, S. J.; Barry, R. C.; Adkins, J. N.; Johnson, E. T.; Qian, W. -J.; Webb-Robertson, B. -J.; Smith, R. D.; Lipton, M. S. Normalization Approaches for Removing Systematic Biases Associated with Mass Spectrometry and Label-Free Proteomics. *J. Proteome Res.* **2006**, *5*, 277-286.
- (36) Ono, M.; Shitashige, M.; Honda, K.; Isobe, T.; Kuwabara, H.; Matsuzuki, H.; Hirohashi, S.; Yamada, T. Label-Free Quantitative Proteomics Using Large Peptide Data Sets Generated by Nanoflow Liquid Chromatography and Mass Spectrometry. *Mol. Cell. Proteomics* **2006**, *5*, 1338-1347.
- (37) Toumi, M. L.; Go, E. P.; Desaire H. Development of fully functional proteins with novel glycosylation via enzymatic glycan trimming. *J. Pharm. Sci.* **2009**, *98*, 2581-2591.
- (38) Wada, Y.; Tajiri, M.; Yoshida, S. Hydrophilic Affinity Isolation and MALDI Multiple-Stage Tandem Mass Spectrometry of Glycopeptides for Glycoproteomics. *Anal. Chem.* **2004**, *76*, 6560-6565.
- (39) Tajiri, M.; Yoshida, S.; Wada, Y. Differential Analysis of Site-Specific Glycans on Plasma and Cellular Fibronectins: Application of a Hydrophilic Affinity Method for Glycopeptide Enrichment. *Glycobiology* **2005**, *15*, 1332-1340.
- (40) Zhang, Y.; Go, E. P.; Desaire, H. Maximizing Coverage of Glycosylation Heterogeneity in MALDI-MS Analysis of Glycoproteins with Up to 27 Glycosylation Sites. *Anal. Chem.* **2008**, *80*, 3144-3158.
- (41) Palm, A. K.; Novotny, M. V. A Monolithic PNGase F Enzyme Microreactor Enabling Glycan Mass Mapping of Glycoproteins by Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1730-1738.
- (42) Rudd, P. M.; Joao, H. C.; Coghill, E.; Fiten, P.; Saunders, M. R.; Opdenakker, G.; Dwek, R. A. Glycoforms Modify the Dynamic Stability and Functional Activity of an Enzyme. *Biochemistry* **1994**, *33*, 17-22.

- (43) Irungu, J.; Go, E. P.; Zhang, Y.; Dalpathado, D. S.; Liao, H. -X.; Haynes, B. F.; Desaire, H. Comparison of HPLC/ESI-FTICR MS Versus MALDI-TOF/TOF MS for Glycopeptide Analysis of a Highly Glycosylated HIV Envelope Glycoprotein. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1209-1220.
- (44) Imre, T.; Schlosser, G.; Pocsfalvi, G.; Siciliano, R.; Molnar-Szollosi, E.; Kremmer, T.; Malorni, A.; Vekey, K. Glycosylation Site Analysis of Human Apha-1-acid Glycoprotein (AGP) by Capillary Liquid Chromatography – Electrospray Mass Spectrometry. *J. Mass Spectrom.* **2005**, *40*, 1472-1483.
- (45) Dalpathado, D. S.; Irungu, J.; Zhang, Y.; Go, E. P.; Desaire, H. Unpublished data.
- (46) Go, E. P., Rebecchi, K. R.; Dalpathado, D. S.; Bandu, M. L.; Zhang, Y.; Desaire, H. GlycoPep DB: A Tool for Glycopeptide Analysis Using a “Smart Search.” *Anal. Chem.* **2007**, *79*, 1708-1713.
- (47) Irungu, J.; Go, E. P.; Dalpathado, D. S.; Desaire, H. Simplification of Mass Spectral Analysis of Acidic Glycopeptides Using GlycoPep ID. *Anal. Chem.* **2007**, *79*, 3065-3074.

## CHAPTER 5

# ASSESSING SECONDARY STRUCTURE OF GLYCOPROTEINS IN THE PRESENCE AND ABSENCE OF HIGH MANNOSE *N*-LINKED GLYCANS

### 5.1 Introduction

High mannose glycans are known to be the simplest type of *N*-linked glycans. They are considered the least processed because they have undergone the fewest modifications during post-translational processing.<sup>1-3</sup> Typically, in humans, these types of glycans are rare. Instead, *N*-linked glycans become extensively processed in the Golgi, after protein folding, to form complex-type *N*-linked glycans.<sup>1,3</sup> It is likely that when high mannose glycans are exposed on a human protein, the body recognizes these as being from a mis-folded or non-self protein. Most proteins in the human body that contain high mannose glycans on their surface will be removed quickly by mannose binding lectins.<sup>4, 5</sup> Rapid protein removal is to be avoided when administering glycoproteins as therapeutics; otherwise, dosing must be increased.<sup>6</sup> Thus, pharmaceutical companies interested in developing glycoprotein therapeutics must attempt to find an optimal glycan profile for producing their glycoproteins.

One option for producing proteins with an optimal glycan profile is to consider expressing the protein in mammalian cell lines, which typically generate proteins with more human-like glycosylation (complex glycans).<sup>7</sup> The disadvantage of this route is that much lower protein production yields are typical, compared to other cell lines, such as yeast or insect cells.<sup>8-10</sup> These production platforms produce the protein in much higher yields, but they contain non-human, high mannose type *N*-linked glycans.<sup>6</sup>

Even mammalian cell expression systems can cause problems for the safety and efficacy of glycoproteins. Erythropoietin, for example, is a common glycoprotein drug that is expressed in Chinese hamster ovary (CHO) cells.<sup>7, 11, 12</sup> The resulting glycosylation in erythropoietin is very similar to human glycans; however, there are two different types of sialic acids produced in CHO cells, whereas humans only produce one of the two sialic acid types, *N*-acetylneuraminic acid.<sup>11, 12</sup> This small change in the type of sialic acid residue can cause those who use this drug to experience adverse side effects and cause increased rates of clearance from the body.<sup>13, 14</sup>

One possible solution is to assess if the glycosylation is even necessary on glycoproteins, once the protein is expressed. In fact, several proteins that are glycosylated naturally in the human body, including interleukin-2, tumor necrosis factor- $\alpha$ , and some interferons have been successfully expressed in *Escherichia coli* and have been shown to be therapeutically active.<sup>7</sup> Two major roles of *N*-linked glycans are to aid in protein folding<sup>3</sup> and protein secretion from the cells.<sup>15</sup> In certain circumstances, the glycosylation may have no functional impact on the protein, after expression and secretion. If the glycans are not needed once a protein is folded and secreted, cell lines such as yeast and insect cells can be utilized for protein production; and, once produced, an enzyme such as Peptide-N-glycosidase F (PNGase F) may be used to cleave the glycans from the protein.

A first step in assessing the need for *N*-linked glycosylation present on a glycoprotein is expression of the protein in either yeast or insect cell line where greater protein yield can be obtained, followed by the use of circular dichroism (CD) spectroscopy to check for changes in structure and thermal stability due to

deglycosylation.<sup>16</sup> Recombinant glycoproteins that exhibit measurable changes in structure and/or thermal stability by CD spectroscopy after deglycosylation is a key clue in confirming that the glycosylation is critical for protein structure and stability. Therefore, these types of glycoproteins would need to be expressed in a mammalian cell line where human-like glycosylation will be present. On the other hand, if no measurable changes in the structure and thermal stability of a deglycosylated glycoprotein are detected, then further steps can be taken to ensure protein activity is not impacted after deglycosylation. This CD spectroscopic pre-screening technique is an easy approach requiring relatively small sample amounts for narrowing down appropriate cell lines for further development of recombinant therapeutic glycoproteins.

The work described herein utilizes CD spectroscopy to assess changes in protein structure and/or stability due to removal of glycans with PNGase F in glycoproteins. To further confirm that all the glycans were removed from the proteins after PNGase F treatment, a protease digestion was performed, followed by LC-MS and MS/MS analysis of the resulting peptides and glycopeptides. A model glycoprotein, ribonuclease B (RNase B) was first assessed by this method. Later, human lysyl oxidase-like 2 (hLOXL2) glycoprotein expressed in *Drosophila* Schneider 2 (S2) cells was analyzed. hLOXL2 is known to be an important protein in development of certain cancers,<sup>17-19</sup> as well as a potential pharmaceutical drug for those with a deficiency of this protein.<sup>20</sup> Additionally, both of these glycoproteins contained only high mannose type *N*-linked glycans.

## 5.2 Experimental

### 5.2.1 Materials and Methods

Ribonuclease B (RNase B), Trizma® HCl, Trizma® base, NaCl, ethylenediaminetetraacetic acid (EDTA), tris(2-carboxyethyl)phosphine (TCEP), iodoacetamide (IAA), acetic acid, and formic acid were purchased from Sigma (St. Louis, MO). Peptide-N-glycosidase F (PNGase F) from *Flavobacterium meningosepticum* was purchased from New England BioLabs (Ipswich, MA). Acetonitrile (CH<sub>3</sub>CN) was purchased from Fisher (Fair Lawn, NJ). Sequencing grade-modified trypsin was purchased from Promega (Madison, WI). Water was purified in house with a Millipore Direct-Q® UV 3 water purification system (Billerica, MA) that had a resistance > 18 MΩ. Human lysyl oxidase (hLOXL2) expressed in *Drosophila* Schneider 2 (S2) cells was obtained from the Mure lab at the University of Kansas (Lawrence, KS). See Chapter 3 for more details about the hLOXL2 glycoprotein.

### 5.2.2 Preparation of Glycoprotein Stock Solutions

RNase B was diluted to 2 mg/mL in 20 mM Tris buffer, pH 7.8 and 50 mM NaCl. hLOXL2 was supplied at 2 mg/mL in 100 mM Tris, pH 8.5, 150 mM NaCl, and 1 mM EDTA. Next, 75 µL aliquots of the 2 mg/mL stocks solutions were flash frozen in liquid nitrogen. The vials were then stored at -80 °C until ready for CD analysis.

### 5.2.3 Circular Dichroism (CD) Spectroscopy

Approximately 2 days prior to performing CD spectroscopy, vials containing the 2 mg/mL protein samples were slowly thawed by storing them at -20 °C for ~8 hrs, then storing at 4 °C overnight. Glycoprotein samples requiring deglycosylation were treated with 1 µL of 1:30 diluted PNGase F (a deglycosylating enzyme) for each N-linked

glycosylation site present in the proteins (1  $\mu$ L of PNGase F for RNase B and 2  $\mu$ L of PNGase F for hLOXL2). Samples containing PNGase F were incubated at 37 °C for 24 hrs, for glycan removal. Samples not being deglycosylated were also incubated at 37 °C oven for 24 hrs, to ensure any protein degradation caused by the incubation would be the same in all samples. Immediately prior to CD analysis, samples were diluted in the original buffer to 0.5 mg/mL.

For the CD experiments, the “blank” sample consisted of the buffer from a given glycoprotein solution. The same amount of PNGase F added to glycoprotein was also added to the blank. The blank was run at the beginning of each day and re-run any time the nitrogen tank was changed, or when the type of sample being analyzed was changed. The data was baseline adjusted for the blank run within the instrument software.

CD spectroscopy was completed on a Jasco J-815 spectrometer (Tokyo, Japan), and Jasco Spectra Manager version 1.54.03 (Build 1) software for CD analysis. All CD experiments were conducted under nitrogen flow. For completing a secondary structure scan, the wavelength range was set from 300 nm to 190 nm with a data pitch of 1 nm in continuous scan mode, a constant temperature of 25 °C, a scan speed of 100 nm/min, response time of 2 secs, bandwidth of 1 nm. An accumulation of 3 spectra were averaged before reporting. Parameters for the melt studies were as follows: 25 to 90 °C, data pitch of 0.2 °C, delay time was 10 sec, the temperature slope was 1 °C/min, sensitivity was standard (100 mdeg), response was 0.5 sec, bandwidth was set to 1 nm. The CD signal was measured at 222 nm.

Data collected was exported to .txt files, so plots could be re-constructed in Microsoft Excel. Secondary structural elements, or percentages of  $\alpha$ -helices and  $\beta$ -sheets, were calculated on the Dichroweb website (<http://dichroweb.cryst.bbk.ac.uk/html/home.shtml>) using the K2D analysis program. The melt temperature ( $T_m$ ) was calculated by importing the data from the melt plot into Microcal Origin 6.0 software and fitting the data to a sigmoidal curve. The inflection point in the sigmoidal curve (also calculated within the Microcal Origin software) was determined to be the melt temperature ( $T_m$ ).

#### **5.2.4 Protease digestion**

To ensure that the deglycosylating enzyme, PNGase F, completely removed the glycans from the glycoprotein samples, deglycosylated RNase B and hLOXL2 (as well as glycosylated RNase B and hLOXL2 control samples) were prepared for protease digestion, followed by LC-MS and MS/MS analysis. Urea was added to each of the samples, to a final concentration of 6 M, for protein denaturation. To reduce the disulfide bonds in the proteins, TCEP was added to a final concentration of 5 mM, and the samples were allowed to incubate at room temperature for 1 hr. Cysteine residues were then derivatized by the addition of IAA to 10 mM with incubation at room temperature in the dark for 1 hr, followed by the addition DTT to 10 mM to quench the alkylation reaction. Samples were then diluted with buffer to a final urea concentration of 1 M. Trypsin was added in a 1:30 enzyme:protein ratio, and samples were allowed to incubate for 18 hours at 37 °C. The protease digestion was stopped by the addition of 1  $\mu$ L acetic acid for every 100  $\mu$ L in solution. Digested samples were concentrated in a

Labconco centrivap cold trap (Kansas City, MO) until the final protein concentration was ~3 mg/mL, followed by storage at -20 °C until ready for LC-MS analysis.

### **5.2.5 Liquid Chromatography/Mass Spectrometry**

For LC-MS and LC-MS/MS analysis of glycoproteins, 5  $\mu$ L (~15  $\mu$ g) of the proteolytically digested samples were injected onto a CVC MicroTech (Fontana, CA) C18 column (300  $\mu$ m i.d., 5 cm length, and 3  $\mu$ m particle size) that was attached to a Dionex UltiMate capillary HPLC system (Sunnyvale, CA) equipped with a FAMOS well plate autosampler and directly connected to an electrospray ionization linear ion trap Fourier transform ion cyclotron resonance mass spectrometer (ESI-LTQ-FTICRMS). Mobile phase HPLC solvents consisted of 99.9 % water + 0.1 % formic acid for solvent A and 99.9 % acetonitrile + 0.1 % formic acid for solvent B. For separation of peptides and glycopeptides, solvent B was initially held at 5 % for 5 min, linearly increased to 40 % over 50 min, increased further to 90 % in 10 min, held at 90 % B for 10 min, and re-equilibrated before the next injection. Between each glycoprotein sample injection, a short wash cycle and blank run were performed to ensure that no peptides/glycopeptides from previous runs were detected. For mass spectrometry, the ESI source voltage was set to 2.8 kV; the capillary temperature was 200 °C; the capillary had an offset voltage of 47 V; the FT-ICR resolution was set to 25,000 for  $m/z$  400; and tandem mass spectrometry data were acquired in data-dependent mode. The five most intense ions from an FT-ICR scan were chosen for CID (collision induced dissociation) analysis with a three min dynamic exclusion window and collision energy of 30%.

The glycosylated hLOXL2 sample was re-run on the same HPLC column connected to a Waters Acquity ultra performance liquid chromatography (UPLC) system (Milford, MA) using an electrospray linear ion trap mass spectrometer (ESI-LTQ Velos MS), ThermoScientific (San Jose, CA) for detection, because no glycopeptides were detected at one of the two glycosylation sites for hLOXL2 on the ESI-LTQ-FTICRMS). The LTQ Velos is known to have distinct advantages, such as higher sensitivity and shorter duty cycle, compared to the LTQ.<sup>21</sup> To elute peptides and glycopeptides the solvent conditions began at 5% B and were linearly increased 10 % in 5 min, increased to 40 % B over 45 min, further increased to 90 % B in 10 min, held at 90 % B for 10 min, and re-equilibrated before the next injection. The LTQ Velos MS parameters are as follows: 3 kV for ESI source voltage, 250 °C for capillary temperature, and data dependent mode was used for collection of MS/MS data by selecting the five most intense ions for collision induced dissociation (CID). A 30% collision energy was used, along with a 3 min dynamic exclusion window.

MS and MS/MS data were manually assigned. MS data was collected for RNase B and hLOXL2 before and after glycan removal with PNGase F to confirm that PNGase F completely removed the glycans from the glycoproteins. Therefore, only the peptides from the potential glycosylation sites were searched in the MS and MS/MS data. For MS/MS analysis peptide and glycopeptides, the deglycosylated peptides needed to contain characteristic b and y ions, and glycosylated peptides needed to show losses of glycan residues. For high resolution MS data collected on the FTICRMS, the monoisotopic  $m/z$  had to be within 20 ppm mass error to be considered a detected peptide or glycopeptide. This high resolution mass spectrometric detection parameter

was relaxed for the hLOXL2 glycosylated sample that was run on the LTQ Velos MS because the instrument is a low resolution mass spectrometer. Therefore, to be considered a detected peptide or glycopeptide in the LTQ Velos data, the peak had to be within 1 Da of the calculated  $m/z$ .

### 5.3 Results and discussion

CD spectroscopy was used to assess changes in structure and/or stability among glycoproteins containing high mannose type *N*-linked glycans after glycan removal. Two proteins were chosen for analysis. The first was a model glycoprotein, ribonuclease B (RNase B), a fairly small glycoprotein (~15 kDa) with one *N*-linked glycosylation site containing high mannose type glycans. After RNase B was analyzed, the method was applied to a protein that has potential as a glycoprotein pharmaceutical, human lysyl oxidase-like 2 (hLOXL2) glycoprotein. hLOXL2 is ~40 kDa and contains two *N*-linked glycosylation sites. Chapter 3 describes the glycosylation present on hLOXL2 and confirmed that the glycans are *N*-linked and of the high mannose type.

Figure 1 illustrates a workflow of the experiments that were performed. Part of the native (glycosylated) proteins were deglycosylated by the enzyme PNGase F, which cleaves the *N*-linked glycosylation between the Asn residue and the first *N*-acetylglucosamine (HexNAc) glycan. In the process of removing the glycan, the Asn residue is converted to an Asp residue (as shown in Figure 1, steps 1 and 2). Before CD spectroscopy could be utilized for structural and stability analysis of the glycoproteins, the completeness of the glycan removal reaction was assessed by mass spectrometry. A glycosylated and a deglycosylated sample were subjected to protease digestion (step 3 in Figure 1). During this step, the proteins were cleaved into peptides

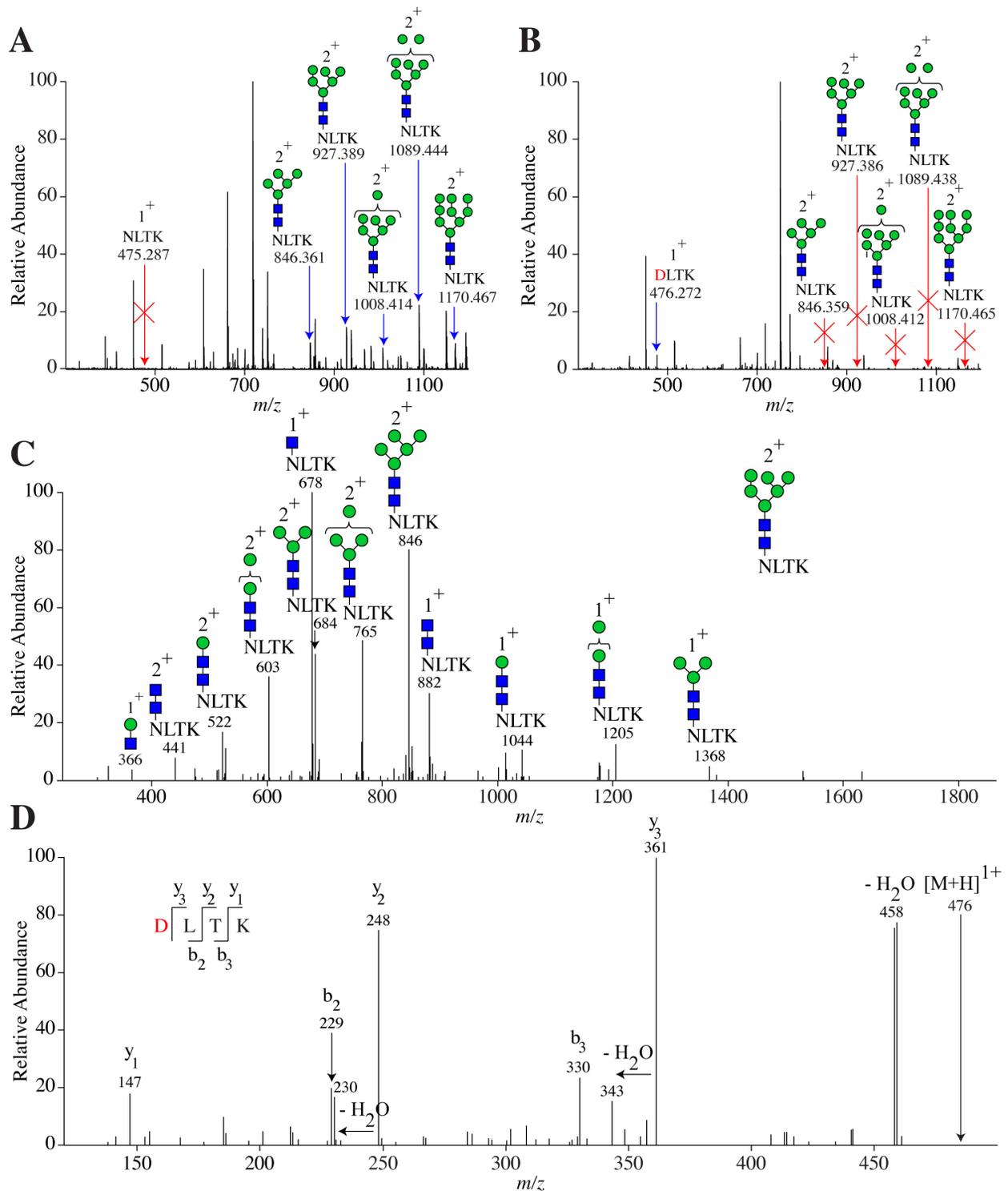
and glycopeptides. After digestion, LC-MS and data dependent MS/MS was performed on the proteins (step 4 in Figure 1). Once the deglycosylation conditions were validated to be effective at completely removing the glycan, the CD experiments were performed (step 5 in Figure 1). Two different types of CD experiments were utilized to determine changes in glycoprotein structure and stability; a wavelength scan, for structural analysis, and a melt study, for stability.



### 5.3.1 Validation of glycan removal

In order to ensure that PNGase F effectively removed the glycans from the proteins, deglycosylated samples were proteolytically digested and then subjected to LC-MS and MS/MS analysis. Because hLOXL2 contained two *N*-linked glycosylation sites, protease digestion before mass spectrometric analysis, as opposed to MS on the intact protein, would give the most complete compositional information. Figure 2 illustrates mass spectrometry data for the glycosylated and deglycosylated RNase B sample. Figure 2A shows the high resolution mass spectrum, where all the glycoforms from RNase B were present with no peak corresponding to a non-glycosylated peptide. Figure 2B is the high resolution mass spectrum after deglycosylation. As can be elucidated from Figure 2B, the glycans are not present in this spectrum, thereby illustrating that the deglycosylation reaction with PNGase F went to completion. Figure 2C and 2D are representative RNase B MS/MS data that were used to confirm the identities of the peaks present in Figure 2A and 2B. Figure 2C is tandem mass spectrum from one of the glycopeptides present in RNase B. The losses of glycan residues in Figure 2C are readily seen; they confirm the composition predicted from the high resolution mass spectrum in Figure 2A. Figure 2D displays the MS/MS data for the deglycosylated peptide from RNase B, where the peptide was easily deduced by the presence of b and y ions for all but one peptide backbone cleavage at b<sub>1</sub>, which was beyond the lower limit of the scan range. This spectrum, along with the mass spectrum

shown in Figure 2B, confirmed that the deglycosylated peptide was formed, and PNGase F successfully removed the glycans from RNase B.



**Figure 2.** Proteolytically digested RNase B mass spectra for glycosylated (A) and (C), as well as deglycosylated (B) and (D) forms. (A) Mass spectrum for glycosylated RNase B. Glycopeptides were

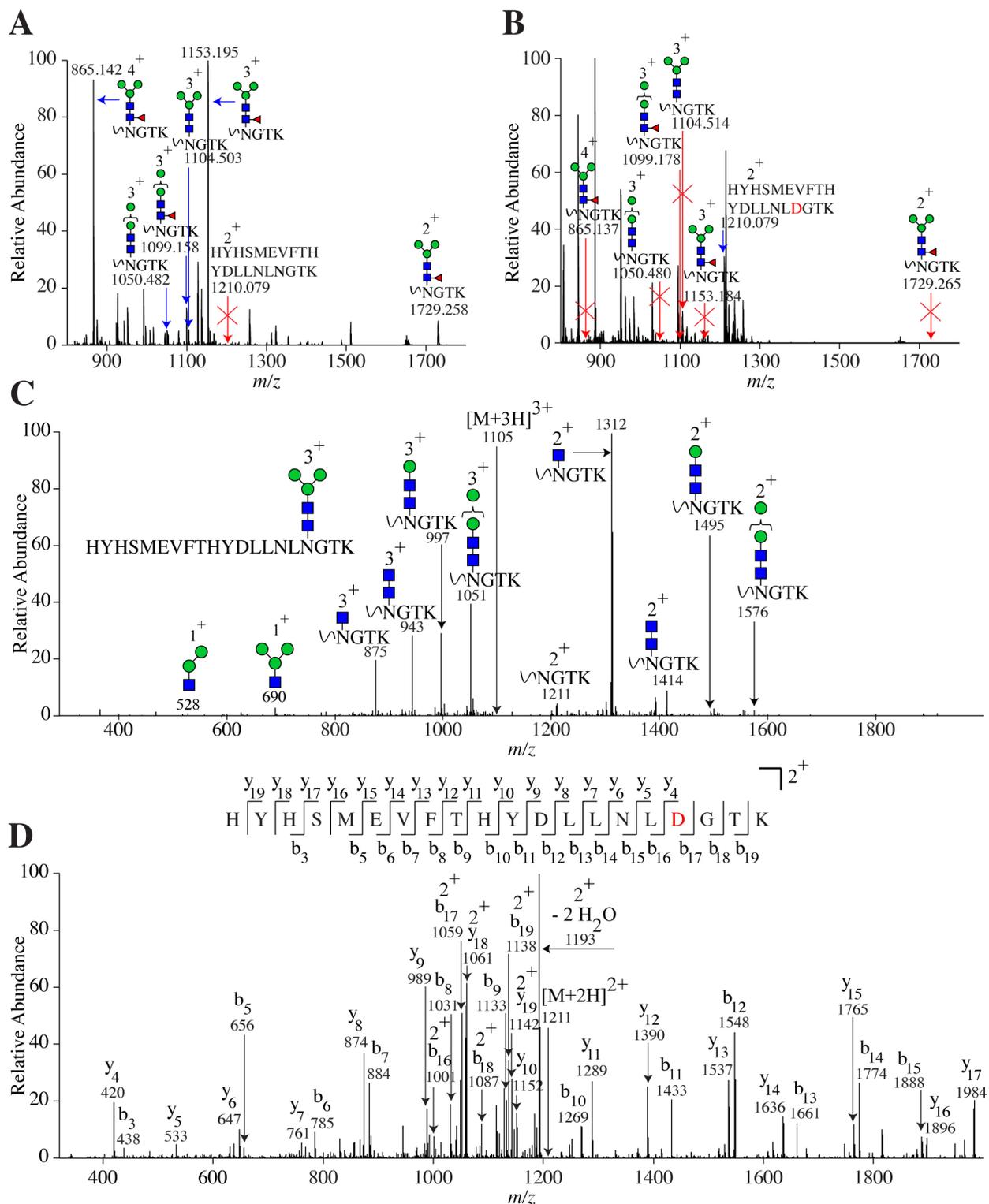
detected, but not a nonglycosylated peptide. (B) Mass spectrum for deglycosylated RNase B. The glycosylated peptide was detected, but no glycopeptides. (C) MS/MS data for a representative RNase B glycopeptide for validation of the presence of glycans at the glycosylation site. (D) MS/MS data for the deglycosylated RNase B peptide.

After successful deglycosylation of RNase B in the native state with PNGase F, this approach was applied to hLOXL2 using the same digestion procedure. The MS data was collected on both an ESI-LTQ-FTICRMS and an ESI-LTQ Velos MS with the same HPLC column for the glycosylated and deglycosylated forms of hLOXL2. Glycosylation was detected at both glycosylation sites with the ESI-LTQ Velos MS and no glycopeptides were detected in the deglycosylated samples (data not shown). hLOXL2 samples collected on the ESI-LTQ-FTICRMS contained glycopeptides at only one of the two *N*-linked glycosylation sites for the glycosylated samples. No glycopeptides were detected in the deglycosylated hLOXL2 sample, similar to the ESI-LTQ Velos MS data. Figures 3 and 4 are representative MS and MS/MS spectra for glycosylated and de-glycosylated peptides at both *N*-linked glycosylation sites in hLOXL2. Figure 3A and 3C show ESI-LTQ Velos MS data at the NGSLVWGMVCGQNWGIVEAMVVCR glycosylation site (because this glycosylation site was not detected in the ESI-LTQ-FTICRMS data). Figure 3B and 3D show the same glycosylation site after deglycosylation where the spectra are from the ESI-LTQ-FTICRMS data because the FT-MS data in Figure 3B has much higher resolution compared to the ESI-LTQ Velos MS and therefore a more confident assignment of the deglycosylated peptide is obtained. Figure 3C and 3D both show representative MS/MS spectra for the glycosylated and deglycosylated forms of hLOXL2, respectively.

The mass spectral data for the HYHSMEVFTHYDLLNLNGTK glycosylation site showed sufficient glycopeptide and deglycosylated peptide data with the ESI-LTQ-

FTICRMS run. Therefore, Figure 4 illustrates spectra from only the ESI-LTQ-FTICRMS because of the high mass accuracy obtained. Like the other glycosylation site from hLOXL2, the MS data shown in Figure 4A and 4B illustrate that only glycopeptides were detected in the glycosylated sample (Figure 4A) and only the deglycosylated peptide was detected in the deglycosylated sample (Figure 4B). Figure 4C and 4D show representative MS/MS data for the glycosylated and deglycosylated forms of this glycosylation site, respectively. The hLOXL2 mass spectrometry data, like RNase B, illustrated that no glycopeptides were detected in the deglycosylated samples, once again confirming that PNGase F successfully cleaved the glycans at both glycosylation sites in hLOXL2 without unfolding or reducing the protein.





**Figure 4.** hLOXL2 mass spectra at the other *N*-linked glycosylation site illustrating glycosylated (A) and (C), as well as deglycosylated (B) and (D) sample. (A) MS<sup>1</sup> spectrum for the glycosylated hLOXL2 sample where glycopeptides (but not a non-glycosylated peptide) were detected. (B) MS<sup>1</sup> spectrum of deglycosylated hLOXL2 where no glycopeptide peaks were detected. (C) Glycopeptide MS/MS spectrum for validation of the presence of glycans at the glycosylation site. (D) Peptide spectrum of the *N*-linked glycosylation site after deglycosylation with PNGase F.

### **5.3.2 Circular dichroism analysis on glycosylated and deglycosylated proteins**

Circular dichroism was performed on both RNase B and hLOXL2 to assess if changes could be detected in the secondary structure and in the stability of the proteins before and after deglycosylation with PNGase F. To ensure that PNGase F was not interfering in the CD analysis, since PNGase F is also a protein, the deglycosylating enzyme was added to the blank for the deglycosylated protein samples, whereas only buffer was added to the blank for the glycosylated samples. Two different types of CD analysis were performed on each of the proteins. The first type of CD analysis was a secondary structural scan where the wavelengths in the far UV region (260 nm – 190 nm) were analyzed at a constant temperature for changes in ellipticity indicative of the secondary structural elements of the proteins. The second CD scan type was a stability experiment, or melt study, where the change in ellipticity was assessed as the temperature was increased and the wavelength was kept constant at 222 nm (a minima observed for alpha helical content). A typical melt study results in a sigmoidal curve, where the inflection point, or melt temperature ( $T_m$ ), is where half the protein is folded and half is unfolded. The higher the  $T_m$ , the more stable the protein.

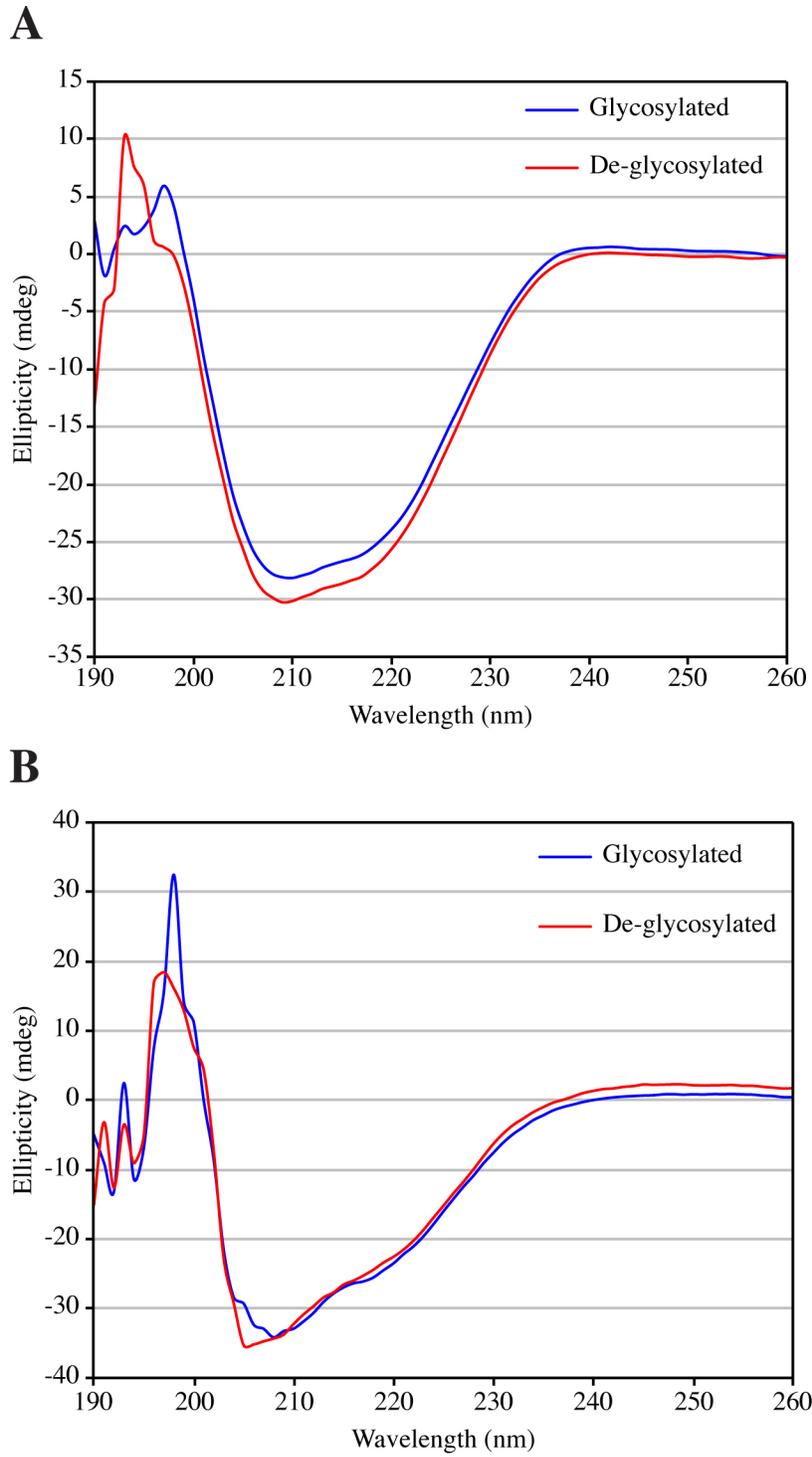
#### **5.3.2.1 CD Secondary Structure Results**

Figure 5A illustrates the CD secondary structural scans for glycosylated and deglycosylated RNase B samples. Four experiments were averaged together for each sample and plotted. As is clearly seen in the Figure, there is very little difference between the glycosylated and deglycosylated spectra. It is expected that RNase B would have a similar CD spectrum without glycosylated present, as RNase A is native nonglycosylated RNase B, with the same protein sequence and structure. Table 1

describes the secondary structural results from the glycosylated and deglycosylated forms of RNase B, where the secondary structure was calculated to be the same. The literature contains estimates for alpha helical content from RNase B X-ray crystallography data between 6 and 18 %, and the percentage of random coil between 46 and 58 %.<sup>22, 23</sup> The results obtained here fall within the ranges from the literature. Additionally, no changes in structure were detected by CD by deglycosylating RNase B.

**Table 1.** RNase B secondary structural elements as calculated in the Dichroweb K2D analysis program.

	$\alpha$ -helix	$\beta$ -sheet	Random coil
<b>Glycosylated</b>	0.16	0.33	0.51
<b>Deglycosylated</b>	0.16	0.33	0.51



**Figure 5.** (A) Secondary structure plot of the glycosylated and deglycosylated forms of RNase B. (B) Secondary structure plot of the glycosylated and deglycosylated forms of hLOXL2.

Figure 5B illustrates the CD secondary structural scans performed on hLOXL2 before and after deglycosylation with PNGase F after averaging 3 different CD runs. Like RNase B, the hLOXL2 results show very similar scans to one another indicating that removal of glycans most likely did not alter the protein structure. Table 2 describes the secondary structure results for hLOXL2 broken down by alpha helical, beta sheet, and random coil content. Because this is a novel recombinant protein, as well as a truncated version compared to endogenous hLOXL2, there is no other data for comparison of these results. Therefore, these data are the first to demonstrate that this protein's structure is unaltered when it undergoes deglycosylation under non-denaturing conditions.

**Table 2.** hLOXL2 secondary structural elements as calculated using Dichroweb's K2D analysis program.

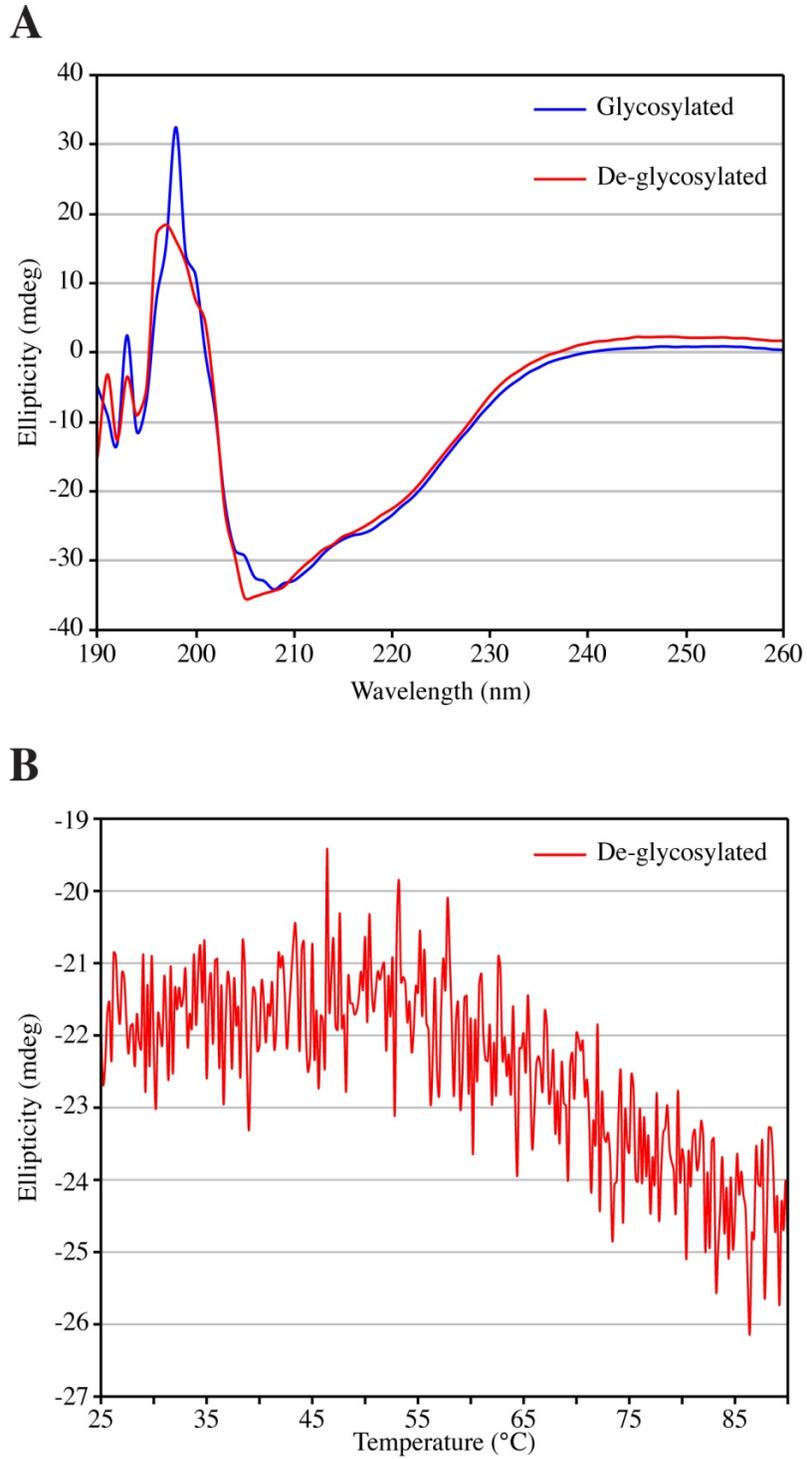
	$\alpha$ -helix	$\beta$ -sheet	Random coil
<b>Glycosylated</b>	0.19	0.32	0.49
<b>Deglycosylated</b>	0.19	0.33	0.49

### 5.3.2.2 Melt studies using circular dichroism spectroscopy

In addition to analyzing the secondary structure of RNase B and hLOXL2 by circular dichroism, a second type of CD experiment was performed to assess changes in protein stability due to glycosylation. It has been shown previously in the literature that there is a slight decrease in protein stability in RNase A (the naturally occurring deglycosylated form) compared to RNase B.<sup>24-27</sup> However, other research has indicated that the decrease in stability upon deglycosylation is either negligible<sup>27</sup> or due to steric hindrance of the glycan,<sup>29</sup> as most of the literature studies measure RNase B stability by how well it is cleaved by different proteases.<sup>29, 30</sup> In this work, a melt study

using CD spectroscopy was performed to assess changes in protein thermal stability in the glycosylated and deglycosylated forms. Figure 6A illustrates the results from the melt study on RNase B. The ellipticity was measured at 222 nm, which is one minimum observed in CD secondary structure plots corresponding to alpha helical content.<sup>16, 31</sup> As a protein is heated and unfolded, its alpha helical content decreases; thus, the ellipticity becomes less negative.<sup>16, 31</sup> Figure 6A also shows the characteristic sigmoidal curve that is observed in CD melt studies. The calculated  $T_m$  from the inflection point in the plot from Figure 6A for the glycosylated and deglycosylated RNase B was 65.6 °C and 65.8 °C, respectively. These results show that there is effectively no change in  $T_m$  due to removal of glycans using circular dichroism to determine protein stability.

After obtaining the results from RNase B that illustrate no difference between the glycosylated and deglycosylated forms' thermal stability, hLOXL2 was also subjected to a melt study using CD spectroscopy. The results from the melt study of hLOXL2 were inconclusive. Rather than slowly unfolding as the temperature was increased, hLOXL2



**Figure 6.** (A) Melt study plot of the glycosylated and deglycosylated forms of RNase B. (B) Melt study plot of the deglycosylated form of hLOXL2 illustrating protein aggregation as the temperature was increased. Similar results were detected in the glycosylated hLOXL2 melt plots.

(both the glycosylated and deglycosylated forms) stayed at the same relative ellipticity until about 55 °C. At higher temperatures, the ellipticity increased gradually indicating that the protein was becoming more structured. Unfortunately, rather than unfolding, like RNase B, hLOXL2 was aggregating and precipitating out of solution, as solid precipitate was observed in the cuvette immediately after the melt studies were completed. Figure 6B is an example from one melt study that was completed for hLOXL2 in the deglycosylated form. All hLOXL2 melt experiments resulted in similar plots. Therefore, a  $T_m$  could not be calculated for hLOXL2. Since the secondary structure of hLOXL2 was very similar in both the glycosylated and deglycosylated forms, and since both proteins behaved identically during the melt studies, it is probable that removing the glycans had minimal or no effect on the stability of the protein.

#### **5.4 Concluding remarks**

CD spectroscopy was used to analyze glycoproteins containing high mannose type *N*-linked glycosylation to determine if changes in structure or protein stability were detected upon deglycosylation of the proteins with the enzyme PNGase F. RNase B and hLOXL2 were successfully deglycosylated, as mass spectrometry experiments confirmed. For analysis of secondary structure by CD spectroscopy, both proteins exhibited no change in the secondary structural elements upon glycan removal. Stability analysis by CD spectroscopy was successful for RNase B. After removing the glycans, no effective change was detected in the  $T_m$  for RNase B, indicating that the protein stability was not changed. Stability analysis on hLOXL2 was inconclusive because protein aggregation was preferred over protein unfolding.

## Acknowledgements

The authors acknowledge Dr. De Guzman and Dr. Gamblin from the University of Kansas Molecular Bioscience department for the use of the Jasco J-815 Circular Dichroism spectrometer. These studies were financially supported from an NSF Career award (project number 0645120) to HD and an NSF Fellowship to KR (DGE-0742523).

## 5.5 References

- (1) Vagin, O.; Kraut, J. A.; Sachs, G. Role of *N*-glycosylation in trafficking of apical membrane proteins in epithelia. *Am. J. Physiol. Renal Physiol.* **2009**, *296*, F459-F469.
- (2) de Leoz, M. L. A.; Young, L. J. T.; An, H. J.; Kronewitter, S. R.; Kim, J. H.; Miyamoto, S.; Borowsky, A. D.; Chew, H. K.; Lebrilla, C.B. High-mannose glycans are elevated during breast cancer progression. *Mol. Cell. Proteomics* **2011**, *10*, 1-9.
- (3) Helenius, A.; Aebi, M. Intracellular functions of *N*-linked glycans. *Science* **2001**, *291*, 2364-2369.
- (4) Ip, W. K. E.; Takahashi, K.; Ezekowitz, R. A.; Stuart, L. M. Mannose-binding lectin and innate immunity. *Immunol. Rev.* **2009**, *230*, 9-21.
- (5) Degn, S. E.; Jensenius, J. C.; Bjerre, M. The lectin pathway and its implications in coagulation, infections and auto-immunity. *Curr. Opin. Organ Transplant* **2011**, *16*, 21-27.
- (6) Liu, C.; Cashion, L. M.; Pu, H. Protein expression both in mammalian cell lines and in yeast *Pichia pastoris* using a single expression plasmid. *Biotechniques* **1998**, *24*, 266-268.
- (7) Dingermann, T. Recombinant therapeutic proteins: production platforms and challenges. *Biotechnol. J.* **2008**, *3*, 90-97.
- (8) Pandhal, J.; Wright, P. C. *N*-Linked glycoengineering for human therapeutic proteins in bacteria. *Biotechnol. Lett.* **2010**, *32*, 1189-1198.
- (9) Verma, R.; Boleti, E.; George, A. J. T. Antibody engineering: comparison of bacterial, yeast, insect, and mammalian expression systems. *J. Immunol. Methods* **1998**, *216*, 165-181.
- (10) Gerngross, T. U. Advances in the production of human therapeutic proteins in yeast and filamentous fungi. *Nat. Biotechnol.* **2004**, *22*, 1409-1414.

- (11) Cointe, D; Beliard, R; Jorieux, S; Leroy, Y; Glacet, A; Verbert, A; Bourel, D Chirat, F. Unusual *N*-glycosylation of a recombinant human erythropoietin expressed in a human lymphoblastoid cell line does not alter its biological properties. *Glycobiology* **2000**, *10*, 511-519.
- (12) Hokke, C. H.; Bergwerff, A. A.; Vandedem, G. W. K.; Vanoostrum, J.; Kamerling, J. P.; Vliegthart, J. F. G. Sialylated carbohydrate chains of recombinant human glycoproteins expressed in Chinese-hamster ovary cells contain traces of *N*-glycolylneuraminic acid. *FEBS Lett.* **1990**, *275*, 9-14.
- (13) Walsh, G.; Jefferis, R. Post-translational modifications in the context of therapeutic proteins. *Nature Biotechnol.* **2006**, *24*, 1241-1252.
- (14) Baker, K. N.; Rendall, M. H.; Hills, A. E.; Hoare, M.; Freedman, R. B.; James, D. C. Metabolic control of recombinant protein *N*-glycan processing in NS0 and CHO cells. *Biotechnol. Bioeng.* **2001**, *73*, 188-202.
- (15) Dorner, A. J.; Bole, D. G.; Kaufman, R. J. The relationship of *N*-linked glycosylation and heavy chain-binding protein association with the secretion of glycoproteins. *J. Cell Biol.* **1987**, *105*, 2665-2674.
- (16) Kelly, S. M.; Price, N. C. The use of circular dichroism in the investigation of protein structure and function. *Curr. Protein Pept. Sci.* **2000**, *1*, 349-384.
- (17) Kirschmann, D. A.; Seftor, E. A.; Fong, S. F. T.; Nieva, D. R. C.; Sullivan, C. M.; Edwards, E. M.; Sommer P.; Csiszar, K.; Hendrix, M. J. C. A molecular role for lysyl oxidase in breast cancer invasion. *Cancer Res.* **2002**, *62*, 4478-4483.
- (18) Barker, H. E.; Chang, J.; Cox, T. R.; Lang, G.; Bird, D.; Nicolau, M.; Evans, H. R.; Gartland, A.; Erler, J. T. LOXL2-mediated matrix remodeling in metastasis and mammary gland involution. *Cancer Res.* **2011**, *71*, 1561-1572.
- (19) Fong, S. F. T.; Dietzsch, E.; Fong, K. S. K.; Hollosi, P.; Asuncion, L.; He, Q. P.; Parker, M. I.; Csiszar, K. Lysyl oxidase-like 2 expression is increased in colon and esophageal tumors and associated with less differentiated colon tumors. *Genes, Chromosomes & Cancer* **2007**, *46*, 644-655.
- (20) Rodriguez, C.; Rodriguez-Sinovas, A.; Martinez-Gonzalez, J. Lysyl oxidase as a potential therapeutic target. *Drug News Perspect.* **2008**, *21*, 218-224.
- (21) Second, T. P.; Blethrow, J. D.; Schwartz, J. C.; Merrihew, G. E.; MacCoss, M. J.; Swaney, D. L.; Russell, J. D.; Coon, J. J.; Zabrouskov, V. Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures. *Anal. Chem.* **2009**, *81*, 7757-7765.

- (22) Greenfield, N.; Fasman, G. D. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **1969**, *8*, 4108-4116.
- (23) Kartha, G.; Bello, J.; Harker, D. Tertiary structure of ribonuclease. *Nature* **1967**, *213*, 862-865.
- (24) Puett, D. Conformational studies on a glycosylated bovine pancreatic ribonuclease. *J. Biol. Chem.* **1973**, *248*, 3566-3572.
- (25) Joao, H. C.; Dwek, R. Effects of glycosylation on protein-structure and dynamics in ribonuclease-B and some of its individual glycoforms. *Eur. J. Biochem.* **1993**, *218*, 239-244.
- (26) Del Vecchio, P.; Catanzano, F.; Barone, G. Thermodynamic stability of ribonuclease B. *J. Therm. Anal. Calorim.* **2000**, *61*, 363-368.
- (27) Arnold, U.; Ulbrich-Hofmann, R. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **1997**, *36*, 2166-2172.
- (28) Xu, G.; Zhai, H.; Narayan, M.; McLafferty, F. W.; Scheraga, H. A. Simultaneous characterization of the reductive unfolding pathways of RNase B isoforms by top-down mass spectrometry. *Chem. Biol.* **2004**, *11*, 517-524.
- (29) Rudd, P. M.; Joao, H. C.; Coghill, E.; Fiten, P.; Saunders, M. R.; Opdenakker, G.; Dwek, R. A. Glycoforms modify the dynamic stability and functional activity of an enzyme. *Biochemistry* **1994**, *33*, 17-22.
- (30) Barnard, E. A. Ribonucleases. *Annu. Rev. Biochem.* **1969**, *38*, 677-732.
- (31) Mehl, A. F.; Crawford, M. A.; Zhang, L. Determination of myoglobin stability by circular dichroism spectroscopy: classic and modern data analysis. *J. Chem. Educ.* **2009**, *86*, 600-602.