

BIOANALYTICAL APPLICATIONS OF MASS SPECTROMETRY TO IDENTIFY
PROTEINS AND GLYCANS FROM COMPLEX BIOLOGICAL SAMPLES

by

JULIANNE COOK BOTELHO

(Under the Direction of Ron Orlando)

ABSTRACT

Herein we explore the applications of mass spectrometry on three different biological samples, mosquito cuticles, human blood serum, and porcine mucin.

A proteomics analysis of the cast cuticles of *Anopheles gambiae* was performed. From this analysis 259 peptides from putative cuticular proteins were matched to 119 proteins in 69 protein families. Authentic cuticular proteins previously reported were identified following extraction, isolation and sequencing of individual proteins isolated from manually cleaned cuticles. The proteomics approach allowed for the identification of more proteins, in a high throughput manner, than previously reported. Furthermore, several of the identified proteins are reported as authentic *Anopheles gambiae* cuticle proteins for the first time.

A glycoproteomics approach using *Phaseolus Vulgaris* Leukoagglutinin lectin affinity chromatography is described. This analysis allows for identification of low abundance complex type glycoproteins that carry a $\beta 6$ -N-Acetylglucosamine (GlcNAc) branch from the blood serum of patients with benign and malignant ovarian tumors. From this study 53 proteins were identified in the serum of patients with malignant tumors, 44 in patients with benign tumors, and 34 in a control sample of pooled male serum. This accounted for a total of 63 unique proteins

from the 9 serum samples. From this analysis, 10 proteins were identified that with further investigation could serve as potential biomarker candidates. Nine of these proteins were identified only in the serum of patients with malignant tumors while an additional protein was observed to have a higher abundance in malignant serum based on relative quantification by normalized spectral counts.

A quantitative isobaric labeling (QUIBL) technique for glycomics was validated for O-linked glycans. QUIBL, which is based on permethylation of glycans with isobaric labels, proved to be applicable for O-linked glycans based on the analysis of fetuin on a high resolution mass spectrometer. The reproducibility and accuracy of QUIBL for O-linked glycans was confirmed over 2 orders of magnitude and its application to a more complex sample, porcine mucin, was verified in this study

INDEX WORDS: Mass Spectrometry, Proteomics, Mosquito Cuticular Proteins, Glycoproteomics, Ovarian Cancer, Lectins, Glycomics, Isobaric labeling, Stable isotope labeling, Quantitation O-linked Glycans

BIOANALYTICAL APPLICATIONS OF MASS SPECTROMETRY TO IDENTIFY
PROTEINS AND GLYCANS FROM COMPLEX BIOLOGICAL SAMPLES

by

JULIANNE COOK BOTELHO

B.S. University of Florida, 2002

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2007

© 2007

Julianne Cook Botelho

All Rights Reserved

BIOANALYTICAL APPLICATIONS OF MASS SPECTROMETRY TO IDENTIFY
PROTEINS AND GLYCANS FROM COMPLEX BIOLOGICAL SAMPLES

by

JULIANNE COOK BOTELHO

Major Professor: Ron Orlando

Committee: Jonathan Amster
Carl Bergmann

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2007

DEDICATION

To all the strong, educated women that have passed through my life, my grandmother, my mother, my sister and all the female science teachers and professors I have had through out the years. You lead by example and proved to be an inspiration.

And to my husband, David Botelho, who despite it all makes me feel like the strongest woman of them all. I can only hope to express how grateful I am for your support and love through this venture.

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to several people at the University of Georgia who have been of so much help during my graduate career. I would like to thank my major professor, Ron Orlando, for his guidance and support throughout this process. His help was greatly appreciated and his constant reminder to 'not stress about it' brought me back to earth every once in a while. He allowed me to be independent and find my own way during this process to which I am eternally grateful. I would also like to thank my committee members, Carl Bergmann and Jon Amster. Your insight was invaluable and your patience very much appreciated. Last, I would like to acknowledge specific people: Caroline Watson, Judy Willis, Ningia He, Lei Cheng, James Atwood, Lin Lin, and Joshua Sharp. Each of you has been instrumental in this process and I would like to thank each of you.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
CHAPTER	
1 INTRODUCTION	1
2 LITERATURE REVIEW	5
3 PROTEOMICS - ANALYSIS OF PROTEINS FROM CAST CUTICLE OF <i>ANOPHELES GAMBIAE</i>	37
4 GLYCOPROTEOMICS - A PROOF OF PRINCLPLE STUDY: BIOSPECIFIC ISOLATION AND LABEL FREE COMPARISON OF COMPLEX N-LINKED GLYCOPROTEINS IN SERUM OF PATIENTS WITH A MALIGNANT AND BENIGN OVARIAN TUMORS.....	59
5 GLYCOMICS - QUANTIFICATION BY ISOBARIC LABELING (QUIBL) FOR THE COMPARATIVE GLYCOMICS STUDY OF O-LINKED GLYCANS	107
6 CONCLUSIONS.....	133

CHAPTER 1

INTRODUCTION

For years, researchers have worked to understand the cell down to its smallest components. Today, scientists seek to understand the cell as a whole with all of its complexities. As result, the field of *-omics* sciences has emerged. This discipline includes the fields of proteomics, glycoproteomics, and glycomics. Within the *-omics* revolution, mass spectrometry has emerged as the platform of choice and has shown extensive progress in high-throughput identification of proteins and carbohydrates. Herein we explore the applications of mass spectrometry with *-omic* analysis on three different biological samples, mosquito cuticles, human blood serum, and porcine mucin. As Dr. John N. Weinstein at the National Cancer Institute in Bethesda, MD stated, “With 35,000 genes and hundreds of thousands of protein states to identify, correlate, and understand; it no longer suffices to rely on studies of one gene, gene product, or process at a time. We have entered the “omic” era”¹.

In chapter 3, we begin our investigation into the *-omics* field by exploring the cast cuticle proteins of *Anopheles gambiae* by mass spectrometry. This proteomics study identified proteins as authentic *Anopheles gambiae* cuticle proteins for the first time. In contrast, authentic cuticular proteins previously reported were identified via extraction, isolation and sequencing of individual proteins isolated from manually cleaned cuticles, proving to be more time consuming than the proteomics approach applied here.

In chapter 4, we continue our analysis of proteins using mass spectrometry. Here we discuss a glycoproteomics technique where glycoproteins were isolated using lectin affinity chromatography in a proof of principle study on human blood serum of patients with malignant and benign ovarian tumors. Here, our studies led us to the identification of proteins with an affinity for PHA-L from which we can infer that the proteins contain complex N-linked glycans with a β 6GlcNAc branching. This type of glycosylation has been shown to increase with

malignancy in cancer. Both the methodology associated with this analysis and the biological implications of the glycoprotein identifications that could serve as potential biomarkers for ovarian cancer are discussed.

In chapter 5, our *-omics* studies shifts to detection of glycans. Here, we validate quantitative isobaric labeling (QUIBL) as an applicable quantitative glycomics approach for O-linked glycans using a high resolution mass spectrometer. The O-linked glycans of fetuin and mucin were released by β -elimination and permethylated with isobaric labels, $^{13}\text{CH}_3\text{I}$ and $^{12}\text{CH}_2\text{DI}$. The potential of QUIBL was analyzed by comparing the signal intensities of isobaric labeled O-linked glycans of fetuin at varying ratios. The quantification capabilities of QUIBL were further tested on a glycoprotein with a more complex mixture of O-linked glycans, porcine stomach mucin.

REFERENCES

1. Weinstein, J. N., Searching for pharmacogenomic markers: The synergy between omic and hypothesis-driven research. *Disease Markers* **2001**, 17, (2), 77-88.

CHAPTER 2

LITERATURE REVIEW

PROTEOMICS

Mass spectrometry (MS) coupled with a variety of separation techniques provides the principal methodology for proteomics. The two primary strategies used in protein identification by MS are referred to as bottom-up or top-down¹. In bottom-up proteomics purified proteins or complex protein mixtures undergo proteolytic digestion forming peptides to be analyzed by MS; this method was first introduced as gel-based proteomics. In comparison top-down proteomics uses gas-phase fragmentation for MS analysis of intact protein ions or large protein fragments. In this section methods and technologies available for separation, identification, and quantification of proteins based on the bottom-up approach will be discussed.

Gel-based Proteomics

The field of proteomics evolved from the development of two-dimensional gel electrophoresis (2DGE). 2DGE was first introduced by Klose and O'Farrell in the mid-1970s where complex protein mixtures are separated across a gel by two dimensions, first by isoelectric point (pI) and second by relative molecular weight (M_r)²⁻⁵. pI separation is achieved by loading the sample onto an isoelectric focusing (IEF) or immobilized pH gradient (IPG) strip. Once proteins reach their isoelectric point a loss of net charge halts electrophoretic mobility resulting in precise, very concentrated protein bands that can be loaded onto an SDS-PAGE. The bands are separated in the second dimension by a polyacrylamide matrix of an SDS-PAGE. The protein spots formed by the 2D separation are stained for visualization, excised, and digested for further analysis. 2DGE approach has been used extensively for separation of proteins and provides additional information about the pI and M_r of the protein, aiding in the validity of the protein match.

2DGE methods can be problematic because of moderate reproducibility and the robustness of the technique ³. It has been reported that 3,000-10,000 protein spots can be visualized with a standard 2DGE method however, this represents approximately 25% of the total proteins present ⁶. High abundance proteins occupy a significant portion of the available analytical space making observation of low abundance proteins difficult resulting in minimal coverage of an entire proteome by 2DGE separation. Furthermore, proteins with pIs outside the range 2-10, heavier than 100 kDa, lighter than 10 kDa or membrane proteins typically are excluded or under represented in 2DGE analysis^{3, 7, 8}. Techniques to overcome these shortcomings, such as narrowing the pH range of the IPG strips and increasing the load capacity of the SDS-PAGE, to improve the dynamic range have been applied but limitations still remain ⁹. Gel-based methods come at a high cost, based on time, labor, and consumables required to run a 2DGE ¹⁰.

Edman degradation was initially used to identify digested proteins that were separated by gel electrophoresis. Edman's reagent (phenylisothiocyanate) is used to sequentially remove N-terminal amino acids to form a cyclic phenylthiocarbamyl (PTC) derivative that is identified by its retention time using liquid chromatography¹¹. This method is very robust and can take upwards of 10-20 hours to generate enough sequence data to identify one protein. Post-translational modifications (PTMs) of amino acids effect the retention times of the PTC derivatives effecting residue assignment and the blocking of N-terminal due to modifications restriction the chemistry required to perform Edman degradation.

Peptide Mass Fingerprinting

Peptide mass fingerprinting (PMF) compares the peptide mass measurements made by MS from a digested protein to predicted masses using a protein databases. Typically PMF is applied to a single protein or simple protein mixtures which requires significant protein separation and purification prior to analysis. PMF was originally introduced by Henzel et. al. using 2DGE as an alternative to Edman degradation to sequence *E. coli* proteins¹¹. The protein gels spots obtained from the 2DGE analysis were excised and trypsin digested for analysis by matrix assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF). With this approach Henzel et. al. was able to match proteins based on peptide masses from a database search algorithm constructed in-house. PMF demonstrated a more rapid approach to matching proteins that was not limited by modification.

Tandem MS experiments

Tandem mass spectrometry (MS/MS) acquires information from the induced peptide fragmentation. The amino acid sequence of the original precursor ion can be inferred from the masses of the fragment ions. The most common method of fragmentation is collision induced dissociation (CID). The precursor ion enters the collision cell or in the case of ion traps and FT-MS, the precursor ion is isolated in the trap. A high pressure of an energized, chemically inert collision gas, such as Ar, He, or N₂, repeatedly collides with the precursor ion. The collisions convert part of the translational energy of the ion to internal vibrational energy which results in bond breakage and the fragmentation of the precursor ion into smaller fragments.

Typically, fragmentation of the peptide chain occurs at three locations on the backbone, C α -C, C-N, or N-C α bond. If the charge is retained on the N-terminal fragment of the peptide,

the ion is either an *a*, *b* or *c* fragment ion and if the charge is retained on the C terminal, the ion is either *x*, *y* or *z* fragment ion (FIGURE 2.1). The types of fragment ions produced in an MS/MS analysis is controlled in part by the pressure of inert gas used. At low energy CID (1-100 eV range) typically used in triple quadrupole, ion trap, and orthogonal quadrupole time-of-flight instruments the observed fragmentation produces predominantly *b* and *y* ions. High energy CID (keV range) requires MS instruments that can function at high energy, such as a magnetic sector MS¹². High energy CID allows for an increase in types of fragmentation ions formed due to the increase in collision frequency used. While high energy CID has its advantages the typically low energy CID is used in MS/MS proteomics studies.

Some common examples of tandem MS instruments include triple-quadrupoles (QqQ), quadrupole-time-of-flight (Q-TOF), and TOF/TOFs. These instruments use a tandem-in-space configuration where the first mass analyzer is used to isolate the precursor ion and the second mass analyzer to detect the fragment ions. Other MS/MS instrumentation relies on a tandem-in-time configuration where a single mass analyzer is used sequentially and include instruments such as ion-traps (ITMS) and Fourier transform ion cyclotron resonance mass spectrometers (FTICRMS or FTMS). Tandem-in-space mass spectrometers can only perform the tandem process once for an isolated precursor ion, but tandem-in-time instruments can perform repetitive tandem mass spectrometry, MSⁿ. A more detailed description of the analyzers is included below.

Shotgun Proteomics

Shotgun proteomics applies a combination of MS/MS and protein sequence databases to identify proteolytic digested proteins from a complex mixture^{4, 13}. A common method used to match the data from an MS/MS analysis to proteins is by correlating each the experimental

fragmentation spectrum to the calculated spectra for all peptides in a database. A score is given to determine how much the MS/MS spectrum agrees with the correlating calculated sequence spectrum. In a typical shotgun proteomics experiment 50,000–100,000 MS/MS spectra can be generated therefore analysis would be impossible without the software tools to automate this correlation¹⁴. The major search algorithms used currently include, MASCOT, SEQUEST, PeptideSearch, PROWL, and Protein Prospector^{1, 15}.

A common approach to validate peptide matches is to accept all assignments above a certain threshold¹. These thresholds can be determined by searching the same dataset against a decoy database, often created by inverting each protein sequence in the original database. This will provide an estimate of false discovery rate (FDR) seeing as any match made via decoy database is a false positive. As a result, threshold values can be set to maintain a certain FDR.

Separation Techniques for Shotgun Proteomics

Proteome analysis by shotgun proteomics relies on the simplification of peptide mixtures before they are analyzed by MS. Typically the combination of high performance liquid chromatography (HPLC) and mass spectrometry are used. Both single dimensional HPLC and multidimensional LC can be directly interfaced with the mass spectrometer. The most common LC based shotgun proteomics method is reverse phase liquid chromatography mass spectrometry (RPLC-MS) where peptides are separated by hydrophobicity¹⁶.

Typically, nanoRPLC columns (50-100 μm) and the lower flow rates (<500 nl/min) are used for separation to achieve higher sensitivity^{13, 17}. NanoRPLC requires a specially designed LC system or modification of a conventional LC and smaller columns than are available commercially or can be packed by the individual. An in-house nanocolumn is typically

fabricated from fused-silica capillary tubing that has an internal diameter between 50-100 μm and is pulled to form at tip of approximately 5 μm ¹⁷. The nanocolumn is packed using pneumatic displacement via a pressurized cell with polymeric or silica-based, C18 coated, stationary phase with a particle sizes of 3-10 μm ¹³. To maximize sensitivity of peptide detection in the low femtomole range and to minimize sample loss, a peptide mixture can be loaded directly onto the nanocolumn offline ^{8, 13}. The nanocolumn can be interfaced to the MS and the peptides eluted into a mass spectrometer by ESI ⁸.

RPLC-MS is typically combined with 2 or more orthogonal chromatographic approaches, such as size exclusion and ion exchanged which is called multidimensional liquid chromatography (MDLC) or multidimensional protein identification technology (MudPIT)^{17, 18}. MDLC separates peptides by 2D liquid chromatography and a common MDLC combination begins with ion exchange chromatography, typically in the form of strong cation exchange (SCX), followed by RP into the MS ^{8, 18, 16}. This was first shown in the work of the Yates et. al. where MDLC, combining SCX and RP, was used to identify the proteome of *Saccharomyces cerevisiae* ¹⁹.

SCX chromatography is ideal for the first dimension of a MDLC analysis since is directly compatible with RP and has a high loading capacity. Moreover, SCX is compatible with the preceding trypsin digestion given that peptides bind to the stationary phase in the presence of molecules commonly used for digestion, urea and DTT, and intact trypsin binds tightly to the stationary phase removing it from the analysis ⁸. Peptides are eluted from the SCX column primarily based on peptide solution charge ^{8, 19}. Tryptic peptides generally have a solution charge state of 2⁺ since they terminate in lysine or arginine and have a free N-terminus. A charge state of 3⁺ is seen for tryptic peptides that contain a histidine residue in addition to the

terminal lysine or arginine and the N-terminus, a 1^+ charge state can represent C-terminal peptides, peptides whose N-terminus is blocked, or phosphorylated peptides and a 4^+ or higher charge state are more specialized and can possibly contain disulfide linked tryptic peptides or missed cleavages⁸. Peptides that have the same charge are then eluted with SCX by decreasing hydrophobicity.

SCX separation combined with RP can be done off line sequentially or on line with the two chromatography steps interfaced in one fused silica capillary. If done off line multiple fractions are collected from the first dimension of separation and are analyzed independently by RPLC-MS. For example, 80 fractions were collected at one minute intervals from the SCX separation of *Saccharomyces cerevisiae* and then analyzed by RPLC-MS/MS by Yates¹⁹. The off line approach is common in a MDLC method carried where upwards of 100 fractions can be produced. While some methods can be automated, the time required to complete such an analysis is quite lengthy, spanning days to months¹⁸.

Electrophoretic separation combined with chromatographic separation is another type of multidimensional separation used in a shotgun proteomic analysis. Whole proteins are separated by molecular mass with single-dimension polyacrylamide gel electrophoresis (1DGE) and the gel is cut into slices or bands that contain a mixture of proteins. The band is digested and separated by hydrophobicity using RPLC-MS. This method provides a less elaborate intermediate between analyzing 2DGE on a single protein spots, used in PMF compared to the large numbers of fractions produced by MDLC analysis¹⁷.

Instrumentation used for Proteomic Analyses: Ionization Sources

One of the most common ionization methods used to ionize non volatile peptides is matrix-assisted laser desorption ionization (MALDI) developed by both Hillenkamp and Tanaka in the late 1980s^{13, 20, 21}. MALDI produces mostly singly charged ions, $[M+H]^+$, using laser pulses on a sample co-crystallized with a molecular matrix. The matrix has the same absorbance wavelength of the laser and as a result absorbs most of the energy. Laser pulses on co-crystallized samples induce rapid heating from the excitation of the matrix molecules. The sublimation and expansion of the matrix as a result of the excitation produces an explosion like event and the sample and matrix molecules are ejected into the gas phase. The most widely accepted ion formation mechanism involves gas-phase photon transfer with photoionized matrix molecules¹². Different lasers can be used in MALDI, however ultraviolet lasers are the most common found in the form of a Nitrogen lasers. The common matrices for proteomic analysis with a UV laser include α -cyano-4-hydroxycinnamic acid and 2-(4-Hydroxyphenylazo)-benzoic acid.

Another common non volatile peptide ionization method is electrospray ionization (ESI), developed by Fenn²². An electrospray is formed when a strong electric field is applied to a liquid passed through a capillary tube at atmospheric pressure. The electric field generates a charge accumulation on the liquid surface which dissociates to form highly charged droplets. Desolvation by a heated capillary or heated inert gas triggers a cascade of smaller droplets to form alleviating charge repulsion due to the proximity of the charged solvent molecules. Eventually, the charge on the solvent droplet is transferred to the analyte ion and several high-capacity pumps are used to link the atmospheric pressure of the source to the low pressure mass analyzers where the ions are detected¹². Large molecules with several ionizable sites generate

multiple charged ions, $[M+nH]^{n+}$. The multiple charges produced with ESI allows for the analysis of very large molecules even if the mass analyzer has a very small m/z range¹². ESI is commonly employed for shotgun proteomics due to its ability to bring labile biomolecules into the gas phase without decomposition, its capability to interface with the liquid phase of HPLC, and its compatibility with MS/MS analysis from continuous ionization.

Instrumentation used for Proteomic Analyses: Mass Analyzers

Mass analyzers separate ions according to their m/z ratios by either an electric or magnetic field. The most common types of analyzers used in proteomic research include time-of-flight (TOF), quadrupole, ion trap (IT) and Fourier transform ion cyclotron resonance (FTICR). In this section the general principles of each analyzer are briefly described.

TOF mass analyzer

TOF mass analyzers, in basic terms, measure the time it takes an ion to travel from the source to the detector in a field free drift path through a flight tube. The m/z of an ion can be determined given that the kinetic energy (KE) of an ion is associated with its velocity (v) and mass (m) by the following equation, $KE = \frac{1}{2} m v^2$. Ions have approximately the same kinetic energy, 20 to 30 keV, as a result of the acceleration into the flight tube of a TOF; consequently, their velocities differ as a function of their mass. The velocity of an ion is inversely proportional to the square-root of its mass-to-charge. Therefore, the time it takes an ion to travel to the detector through a flight tube will differ as a function of mass. In theory ions with the same mass, that start at the identical point and time, will arrive concurrently at the detector. Some

peptide ions will decay because of the energy departed in the desorption process, post source decay (PSD), but this can be used to obtain structural information.

Mass resolution in a TOF is in part proportional to the flight time; however, variations in the initial kinetic energy can prove to be problematic. By correcting the energy difference the velocity of ions with same m/z will narrow so that an increase in resolution and sensitivity is achieved. Methods to correct for this discrepancy include the use of delayed pulse extraction and reflectors. Delayed pulse extraction employs a short time lag (hundreds of nanoseconds to several milliseconds) after ionization that allows for ions to distribute themselves based on the amount of kinetic energy conveyed during the ionization process. After the delay an accelerating voltage is applied over a potential gradient focusing ions and correcting for the energy discrepancy before sending the ions into the flight tube. Another method used to improve the resolving power of a linear TOF is to incorporate the addition of one or more ion reflectors. Reflectors typically involve a series of grids and ring electrodes that correct the energy dispersion of the ions by creating a field that acts as an ion mirror at the end of a flight tube. Ions with a higher kinetic energy penetrate further into the reflector and as a result require more time to exit. This time delay focuses ions of the same m/z with varying kinetic energies. Modern TOF or TOF/TOF that incorporate these improvements can achieve resolving power of 10,000+, mass accuracy in the range of a few parts per million, and require only about a femtomole of peptide material.

Quadrupole mass analyzers

Quadrupole mass analyzers are made up of four parallel rods with one pair being at the positive side of a variable DC source and other at the negative. Variable RF AC potentials are applied to each rod 180° out of phase and as ions enter the space between the rods they are subjected to the electric field created by the four rods. By systematically adjusting the AC and DC voltages, while maintaining a constant ratio, ions with varying m/z values will reach the detector while other ions will destabilize and collide into the quadrupole rods. The m/z of the ions that can successfully reach the detector are proportional to the voltage applied to the rods. A quadrupole mass analyzer is essentially a mass filter that is capable of transmitting only the ion of choice. Quadrupoles are capable of unit mass resolution, mass accuracy of 0.1-1 Dalton, and moderate sensitivity¹⁵.

IT mass analyzer

IT mass analyzers separate ions by selective ejection from a three-dimensional trapping field. An IT consists of ring shaped electrode that has two hemi-spherical electrodes caps above and below it, essentially a quadrupole bent in on itself. Comparable to the operation of a quadrupole, the RF and DC voltages are applied to the electrodes and ramped at a constant RF/DC ratio for storage of ions within the ion trap. Ions are ejected for detection when their resonance frequency matches the resonance applied to the trap. 3D ion traps generally have poor trapping efficiencies primarily due to their small volume of the trap. The linear ion traps address this problem with a two dimensional quadrupole field instead of a 3D field as described above in which the trapping volume would be roughly a cylinder so that ions can be stored in the entire length of the linear trap giving it a much higher capacity. The ability for an IT mass analyzer to

accumulate ions within the ring electrode prior to ejections aids in the high sensitivity of the analyzer.

FTICR Mass Analyzer

FTICR mass analyzers utilize the motion of an ion located within a high magnetic field under ultra-high vacuum. FTICR utilizes the ion cyclotron motion induced from the interaction of an ion with a unidirectional magnetic field. Ion cyclotron motion is characterized by a cyclotron frequency (f) that is determined based on the strength of the magnetic field (B), the mass (m) and charge (z) of the ion shown by the following equation $f = (z B) / (2\pi m)$. Since the magnetic field is held constant the m/z ratio can be determined by measuring the cyclotron frequency of a given ion. The ICR cell is where ions are stored, mass analyzed and detected by six plates (two excitation, two detector, and two trapping plates). The ions trapped in the cell are excited by applying a sinusoidal voltage to the excitation plates. Ions whose cyclotron frequency is in resonance with the frequency of the applied RF electric field will begin to spiral outwards with ions of the same m/z moving together. As the excited ions come in close proximity to the detection plate current is measured as a function of time to yield a free inductive decay (FID) signal. Fourier Transform (FT) methods are used to deconvolute this complex signal into the m/z versus intensity spectrum. The frequency parameter can be measured with high precision and accuracy so FTICR mass analyzers have resolving power of up to 10^6 or more and ppb mass accuracy. Moreover FTICR provides increased sensitivity due to the trapping of ions as well as a non-destructive method of ion detection. However, FTICR mass analyzers are very cumbersome and expensive.

Quantification Strategies for Proteomic Analyses

Quantifying changes in protein abundance between samples is an important goal of proteomics. This provides a snapshot of the protein expression in response to biological perturbations such as disease, morphogenesis, or treatment ²³. While all the quantification methods discussed here are currently being used in one form or another, the proteomics field has no agreement on a standardized strategy for quantification of identified proteins.

Gel Based Quantification

Attempts at proteomic based quantification started with 2DGE, where samples are run in parallel on separate gels, stained, and intensities for relative quantification compared. Unfortunately 2DGE experiments, even when run in parallel, are only mildly reproducible. As a result, inconsistencies produced incomplete quantitation results. Differential imaging gel electrophoresis (DIGE) improved gel based quantification by using fluorescent dyes to differentially label proteins so that a single 2DGE analysis can be performed. DIGE dyes covalently attach to the lysine residues and create a negligible perturbation so each tagged protein will migrate to the same point in the gel ²⁴. Each fluorescent dye is excited at the appropriate wavelength and relative quantitation is performed using an image analysis. Proteins with differential expression are identified, excised from the gel and analyzed by MS. DIGE continues to be demonstrated for a number of proteomic applications; however, this method is prone to errors associated with the 2DGE separation, as discussed earlier, and can rather be labor intensive ^{6, 25-27}.

Stable Isotopic/Mass Tag Labeling for Quantification

Stable isotopic labels can be added by either enzymatic or chemical reactions of a light or heavy stable isotope tag to proteomes to be compared. The proteomes to be compared are labeled, combined in one-to-one ratio, and analyzed simultaneously by MS. The mass difference of the tags can be used for quantification by comparing peak areas or heights. Stable isotopes or mass tag labeling includes such isotopes as, ^2H , ^{13}C , ^{15}N , and ^{18}O ^{1, 7, 8, 15, 18, 28}. These stable isotopes can be added at either the protein or peptide level by different incorporation techniques.

Stable isotope labeling with amino acids in cell culture (SILAC) can be used for *in vivo* incorporation of a label into proteins ²⁹. Proteins can undergo metabolic incorporation of labels by growing cell populations in normal or enriched isotope containing culture media. A common example involves the use of media labeled with ^{15}N enriched and normal ^{14}N . Cells grown in this manner can undergo typical proteomics MS analysis for quantification of the proteins. This method of isotope incorporation can only be applied to cells grown in culture which is problematic for direct proteomic studies of tissue.

Proteins can also be labeled with isotope-coded affinity tag (ICAT) reagents. The reagent bind to cysteine residues and are internally labeled with either 8 deuterium atoms (D8) or 8 regular hydrogen atoms (D0) and an affinity tag ⁶. After labeling, the samples are pooled for fractionation and digestion. The cysteine labeled peptides are purified by affinity chromatography and can be analyzed by LC-MS/MS. The ICAT tag increase the molecular weight of the peptide by about 500 Da and issues with tagged peptides not co-eluting making relative quantification difficult. A new version of the ICAT reagent use an acid-cleavable bond allowing for the removal of the biotin moiety prior to MS which addresses the large mass increase induced by the tag³⁰. Additionally ICAT reagents that incorporated 9 carbon-13 atoms

or 9 carbon-12 atoms rather than ^2H or ^1H atoms which eliminated the slight delay in retention time observed by the light peptide as found in the regular ICAT reagent ³⁰.

Stable isotopic labels can also be introduced at the peptide level. Typically, the labels are incorporated at either the C- or N- termini since this is universal within the entire peptide population. Such is the case with trypsin induced ^{18}O labeling, where two ^{18}O atoms can be incorporated into the C-terminus of the peptide by digestion in H_2^{18}O ⁷. Another labeling method for peptides is iTRAQ where the N-terminus is chemically tagged by isobaric iTRAQ reagents. In single MS mode iTRAQ produces labeled peptides identical in mass, but produce strong, diagnostic, low-mass MS/MS signature ions for quantitation. Currently, four iTRAQ reagents are commercially available with the promise of a total of 8 becoming available in the near future. This will allow for quantification of up to 8 different samples simultaneously.

Label Free Approaches of Quantification

Label free methods for quantification have recently been introduced as an alternative labeling method. Label free methods require less sample handling and experimental manipulation prior to MS analysis. For label free quantification each sample set is run in parallel and measurements made during MS analysis, such as spectral counting and ion intensities, are compared ^{1,31-35}.

One of the most promising label free approaches involves the use of spectral counts. A single spectrum is acquired from a matched peptide and combined with other peptide hits for a protein group. The sum of spectral hits per proteins match is referred to as the total spectral count for the given protein. Initial studies by Liu et. al. reported that spectral count quantification of spiked standard proteins added to yeast extracts showed a reproducibly high

correlation to the relative protein concentration as well as a linearity over two orders of magnitude^{23, 36}. Xia et. al. also demonstrated the use of spectral counts as the superior method for label free quantification of *Porphyromonas gingivalis*, a Gram-negative intracellular pathogen, when compared to other label free approaches such as signal intensity at the peptide and protein level³⁷. In addition to these examples spectral counting has been employed in several other quantitative proteomic studies making it the most developed method for label free comparative proteomics^{23, 38, 39}.

GLYCOPROTEOMICS

Large scale proteomic studies rarely include the study of post-translational modification (PTM) such as glycosylation. Glycosylation involves the covalent attachment of carbohydrates through either an O-glycosidic or N-glycosidic bond to a protein. The type and number of carbohydrates involved in glycosylation varies greatly making it very difficult to analyze with a typical proteomics platform. However, recent attempts have been made to define these modifications on a proteome scale. While no satisfactory methods for a global, proteome-wide analysis of all glycoproteins exists to date, it is possible to characterize a “sub-glycoproteome” with specific glycogroups by lectin affinity, to chemically capture glycoproteins and/or to map glycosylation sites on a protein⁴⁰⁻⁴⁴. After enrichment and/or removal of glycans a typical proteomics approach, as described previously, can be applied to identify glycosylated proteins and sites glycosylation.

Mapping Glycosylation Sites

A consensus sequence, asparagine-X-serine/threonine (N-X-S/T), where X is any amino acid other than proline, is required for N-linked glycosylation. The N-X-S/T sequence does not guarantee glycosylation and to determine which asparagines are glycosylated a isotope coded glycosylation-site-specific tagging methods, IGOT, is used ⁴⁴. Typically N-glycanase F (PNGaseF) is used to cleave N-linked glycans in the presence of H₂¹⁸O. Upon removal of the glycan the asparagines that were glycosylated are converted to ¹⁸O-labeled aspartic acids. The mass tags created by the incorporation of ¹⁸O are discernible by MS analysis to identify the sites of N-linked glycosylation. The mass tag can be matched during a typical database search by looking for the modification of Asp residues. Glycan mapping by PNGaseF digestion in H₂¹⁸O can identify sites of N-linked glycosylation on a peptide; however, the glycans that were attached can not be matched back to the site of glycosylation after removal.

O-linked glycans are found on serine or threonine residues and do not require a consensus sequence. To determine which serine or threonine contain glycans BEMAD, a mild β -elimination reaction followed by Michael addition with dithiothreitol (DTT) or biotin pentylamine (BAP) is used ⁴⁵. O-linked oligosaccharides are typically released chemically from the protein backbone by a beta elimination reaction ⁴⁶⁻⁴⁹. This chemical reaction takes advantage of the slight acidity of Ser and Thr. In an alkaline environment the proton on alpha carbon of Ser or Thr are eliminated and the attached O-linked carbohydrate is released. DTT or BAP are added to the peptide at the site of O-linked glycosylation by a Michael addition reaction and are used as unique mass tags. DTT addition will show a mass shift of 136 Da and BAP a 310 Da shift that can be matched during a typical database search by looking for the modification

of Ser/Thr residues. As with N-linked glycan mapping, BEMAD can identify sites of O-linked glycosylation on a peptide; however, the glycans that were attached can not be determined.

Lectin Affinity

Affinity chromatography is used for the purification of biological molecules such as proteins and other macromolecules and operates on the principle that ligands, attached to a matrix made up of an inert substance, bind to the desired molecule within a solution to be analyzed⁵⁰. Ideally, the ligand will only interact with the desired molecule and other compounds in the solution will elute, leaving the target molecules on the column. The molecules attached to the ligand can then be removed by using a wash that lowers the binding constant and allows for recovery of the enriched compounds.

For glycoproteomics lectins typically serve as the ligand molecule for affinity chromatography. Lectins are carbohydrate binding proteins found in nature that recognize with high affinity particular structural characteristics of the glycans^{51, 52}. Lectin affinity chromatography (LAC) is a technique for separating glycoconjugates in complex mixtures and can provide a notable degree of separation and purification^{53, 54}. LAC separation technique, also referred to as a “glyco-catch”, is relatively inexpensive to run and maintain while providing a way for glycoproteins or glycopeptides to be isolated simply in a rather reproducible, rapid manner⁴⁰. As a result, it is possible to draw inference about the structural characteristics of glycans based on the interactions with specific lectins.

The separation of glycoconjugates by LAC is based on the specific interaction of glycans with the immobilized lectin. There are several important factors to consider when optimizing the use of these interactions for glycoproteomics analysis. The binding affinity of a lectin for

oligosaccharides are usually on the order of 10^{-5} to 10^{-6} M requiring a concentration of approximately 2-3mg/mL of conjugated lectin per mL^{53, 55}. The bed volume of the column is also important since in general, small amounts of sample, both in volume and concentration, are used in LAC. Therefore a column that contain a bed with less than 1 mL of immobilized lectin is typically used^{53, 55}. A variety of buffers can be used in LAC, the two most common being: tris-buffered saline (TBS) and phosphate-buffered saline (PBS). However, any buffer used should be fortified with metals and salts to prevent inactivation of the lectin⁵⁵. The rate of binding of oligosaccharides to a lectin should also be considered. Plant lectins bind comparatively slower than mammalian lectins and under solution conditions can take minutes or hours for equilibrium to be reached^{53, 55}. This requires an incubation time for the sample to interact with the lectin and a slow flow rate to maximize interactions, typically performed by gravity flow. The use of haptenic sugars to which lectins preferentially bind are used to elute bound glycoconjugate; however, it should be noted that some lectins isolate glycoconjugates by retardation chromatography and it may be necessary to collect multiple column volumes of eluent⁵⁵. Optimization of LAC needs to be considered for individual glycoproteomic analysis and altered based on the lectin used in the analysis.

After enrichment for a specific glycogroup by LAC the proteins can be analyzed in a similar manner as described in the proteomics section of this paper. If glycan mapping is used with LAC the site of glycosylation can be determined but the structure of the glycan that was removed can only be inferred based on the lectin that was utilized in the LAC analysis. The use of LAC, as a glycoproteomics tool, is wide spread in the research community^{56, 40-43, 57-61}

Chemical Capture Method

A chemical capture method has also been reported by Zhang et. al. where glycoproteins undergoes periodate oxidation of the *cis*-diol groups typically found on the non-reducing terminal of N- and O-linked glycoprotein⁶². The sugar ring is cleaved and the *cis*-diol groups of the sugar are converted to aldehydes by sodium periodate oxidation. The modified glycoproteins are passed through a solid support of hydrazide beads and the aldehydes on the glycoproteins covalently couple to hydrazide groups on the solid support. The nonglycosylated proteins do not bind to the hydrazide beads and are removed. The immobilized glycoproteins are digested on the solid support so that only glycosylated peptides remain bound. The isolated N-linked glycopeptides can then be released with PNGaseF and identified using a typical proteomics analysis. There is no enzyme comparable to PNGaseF for O-linked glycans therefore the release of O-linked glycopeptides involves beta-elimination or the sequential removal of monosaccharides by a panel of exoglycosidases until only a Gal β 1,3GalNAc core remains. This core can be released by O-glycosidase. While chemical capture provides a more complete approach to glycoproteomics no inference can be made about the type of glycans that were attached as can be done with lectins.

Quantification Strategies for Glycoproteomics

Glycoproteomics focuses on the enrichment and/or site mapping of proteins that are glycosylated, therefore quantification strategies are performed at the peptide or proteins level. A variety of quantification strategies similar to those applied in proteomics have been utilized in the glycoproteomics field including, quantification on a 2D gel after enrichment for glycopeptides⁴³, addition of stable isotopic labels on the glycopeptides^{41, 59, 62}, and a label free

approach using peak intensities⁶³. Additionally, stable isotope labels can be incorporated during deglycosylation of N-linked glycans with ¹⁸O and ¹⁶O water or via BEMAD on O-linked using deuterated DTT^{61, 64}.

GLYCOMICS

Glycoproteomics methods as described above allow for the analysis of proteins that are glycosylated; however, information about the attached glycans can only be inferred. The aim of a glycomic analysis is to identify the sequences of the released glycans. The enrichment and removal techniques used to remove glycans are similar to those described in the glycoproteomics section however; glycomics deals with analysis of the released glycans rather than the proteins.

Carbohydrate Fragmentation

Typically fragment ions of carbohydrates are based on the successive loss of sugar residues by glycosidic and cross-ring cleavage starting at the non-reducing end of the glycan^{66,67}. Carbohydrate fragmentation ion nomenclature describe *A*, *B* and *C* ions where the charge is retained on the non-reducing end of the carbohydrate and *X*, *Y*, and *Z* ions where the charge is retained on the reducing end⁶⁵. *B*, *C* and *Y*, *Z* fragment ions are products of glycosidic cleavage between the sugars where the glycosidic oxygen atom is retained in the *B* and *Y* fragment. The product ions from cross ring C-C cleavages are designated as *A* and *X* fragment ions and can occur at different locations within the ring. Fragmentation of $[M+H]^+$ carbohydrates by CID predominantly produces *B* and *Y* ions; however, with the addition of an adduct, $[M+Na]^+$ or $[M+Li]^+$, *A* and *X* fragment ions are also formed⁶⁷. An increase in *A* and *X* fragment ions could result from alkali metal ions coordinating with oxygen atoms at various positions on the sugar

ring. The use of fragmentation ions from a standard MS/MS analysis will only provide partial sequence information since many of the constituent monosaccharides, such as mannose, galactose and glucose, are isobaric. Several methods can be used to fragment a glycan such as CID, fast atom bombardment (FAB), electron capture dissociation (ECD), infrared multiphoton dissociation (IRMPD), and post source decay (PSD). While these methods differ on how the fragments are formed the nomenclature used to describe the fragments is universal.

A glycomic analysis has an inherent complexity not seen in other areas of applied biology due to the extensive number of monosaccharides that form in both a linear and branched manner. There have been considerable advances in development of computational methods that can assign glycans based on MS/MS fragmentation patterns; however, such search algorithms are not as widely applied limiting large scale glycomic studies⁶⁸⁻⁷¹.

Quantification Strategies

Stable isotopic labeling has widespread application in the proteomics field, but limited use in the glycomics field. Yuan et. al. and Hitchcock et. al. introduced isotopic varied reductive amination tags for quantification of glycans^{72, 73}. Bowman et. al. synthesized tags that could modify oligosaccharides with four isotope-enriched variants for quantification⁷⁴. In these cases derivatives were added to the reducing termini of N-linked glycans which limits quantification of O-linked glycans since the reducing end of the glycans are reduced during β -elimination. Xie et. al. reported a method for quantification of O-linked glycans using NaBH_4 or NaBD_4 during β -elimination to incorporate deuterium⁷⁵. While successful for O-linked glycans this method would require additional chemistry to be applied to N-linked glycans which is not ideal.

Alvarez-Manilla et. al. developed a method for comparative isotopic labeling of N-linked oligosaccharides with ^{13}C labeled and ^{12}C methyl iodide during standard permethylation conditions ⁷⁶. Permethylation to introduce stable isotopic label is also applicable to O-linked analysis making stable isotopic labeling by permethylation a more global glycomics technique. While the quantification study was successful for N-linked glycans and holds great potential for O-linked analysis, this method induces large m/z shifts between heavy and light labeled glycans that increase as the detected glycans becomes larger. The large mass difference could prove to be problematic for quantification of complex mixtures. A method that minimizes the mass shift was introduced by Atwood et. al., as a novel strategy for quantitative glycomics based on the use of quantitative isobaric labeling (QUIBL)⁷⁷. In this approach, N-linked glycans were permethylated with either $^{13}\text{CH}_3\text{I}$ or $^{12}\text{CH}_2\text{DI}$. This isobaric labeling keeps the nominal masses of the glycans the same but varies the exact mass. The variation in exact mass can be observed using a high resolution mass spectrometer ⁷⁷. This method holds great potential for application to O-linked glycans as well.

REFERENCES

1. Resing, K. A.; Ahn, N. G., Proteomics strategies for protein identification. *FEBS Lett* **2005**, 579, (4), 885–9.
2. O'Farrell, P. H., High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry* **1975**, 250, (10), 4007-4021.
3. Lottspeich, F., Proteome Analysis: A Pathway to the Functional Analysis of Proteins. *Angew Chem Int Ed Engl* **1999**, 38, (17), 2476-2492.
4. Anderson, N. L.; Anderson, N. G., The Human Plasma Proteome History, Character, and Diagnostic Prospects*. In ASBMB: 2002.
5. Klose, J., Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **1975**, 26, (3), 231-43.
6. Issaq, H. J., The role of separation science in proteomics research. *Electrophoresis* **2001**, 22, (17), 3629.
7. Zybaylov, B. L.; Washburn, M. P., Mass Spectrometry-based Methods of Proteome Analysis. *Methods* 1, 3.
8. Peng, J.; Gygi, S. P., Proteomics: the move to mixtures. *Journal of Mass Spectrometry* **2001**, 36, (10), 1083-1091.
9. Adkins, J. N.; Varnum, S. M.; Auberry, K. J.; Moore, R. J.; Angell, N. H.; Smith, R. D.; Springer, D. L.; Pounds, J. G., Toward a Human Blood Serum Proteome Analysis By Multidimensional Separation Coupled With Mass Spectrometry* S. *Molecular & Cellular Proteomics* **2002**, 1, (12), 947-955.
10. Li, X.; Gong, Y.; Wang, Y.; Wu, S.; Cai, Y.; He, P.; Lu, Z.; Ying, W.; Zhang, Y.; Jiao, L., Comparison of alternative analytical techniques for the characterisation of the human serum proteome in HUPO Plasma Proteome Project. *Proteomics* **2005**, 5, (13), 3423-41.
11. Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C., Identifying Proteins from Two-Dimensional Gels by Molecular Mass Searching of Peptide Fragments in Protein Sequence Databases. *Proceedings of the National Academy of Sciences* **1993**, 90, (11), 5011-5015.
12. de Hoffmann, E.; Stroobant, V., *Mass spectrometry: principles and applications*. Wiley: 2001.
13. Mann, M.; Hendrickson, R. C.; Pandey, A., ANALYSIS OF PROTEINS AND PROTEOMES BY MASS SPECTROMETRY. *Annual Review of Biochemistry* **2001**, 70, 437-473.

14. Marcotte, E. M., How do shotgun proteomics algorithms identify proteins. *Nature Biotechnology* **2007**, 25, 755 - 757
15. Gygi, S. P.; Aebersold, R., Mass spectrometry and proteomics. *Curr Opin Chem Biol* **2000**, 4, (5), 489-94.
16. Motoyama, A.; Venable, J. D.; Ruse, C. I.; Yates 3rd, J. R., Automated Ultra-High-Pressure Multidimensional Protein Identification Technology (UHP-MudPIT) for Improved Peptide Identification of Proteomic Samples. *Analytical chemistry(Washington, DC)* **2006**, 78, (14), 5109-5118.
17. Romijn, E. P.; Krijgsveld, J.; Heck, A. J. R., Recent liquid chromatographic-(tandem) mass spectrometric applications in proteomics. *Journal of Chromatography A* **2003**, 1000, (1-2), 589-608.
18. Nair, K. S.; Jaleel, A.; Asmann, Y. W.; Short, K. R.; Raghavakaimal, S., Proteomic research: potential opportunities for clinical and physiological investigators. In *Am Physiological Soc*: 2004.
19. Washburn, M. P.; Wolters, D.; Yates Iii, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology* **2001**, 19, 242-247.
20. Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry* **1988**, 60, (20), 2299-2301.
21. Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T., Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom* **1988**, 2, (8), 151-153.
22. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization-principles and practice. *Mass Spectrom. Rev* **1990**, 9, (1), 37-70.
23. 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* S. *Molecular & Cellular Proteomics* **2005**, 4, (10), 1487-1502.
24. Lilley, K. S.; Friedman, D. B., All about DIGE: quantification technology for differential-display 2D-gel proteomics. *Expert Rev Proteomics* **2004**, 1, (4), 401-409.
25. Hoorn, E. J.; Hoffert, J. D.; Knepper, M. A., The Application of DIGE-Based Proteomics to Renal Physiology. *Nephron Physiol* **2006**, 104, (1), p61-p72.
26. Okano, T.; Kondo, T.; Kakisaka, T.; Fujii, K.; Yamada, M.; Kato, H.; Nishimura, T.; Gemma, A.; Kudoh, S.; Hirohashi, S., Plasma proteomics of lung cancer by a linkage of multi-dimensional liquid chromatography and two-dimensional difference gel electrophoresis. *Proteomics* **2006**, 6, (13), 3938-3948.

27. Kakhniashvili, D. G.; Griko, N. B.; Bulla, L. A.; Goodman, S. R., The Proteomics of Sick Cell Disease: Profiling of Erythrocyte Membrane Proteins by 2D-DIGE and Tandem Mass Spectrometry. In SEBM: 2005.
28. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003**, 422, 198-207.
29. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics*. *Molecular & Cellular Proteomics* **2002**, 1, (5), 376-386.
30. Yi, E. C.; Li, X.; Cooke, K.; Lee, H.; Raught, B.; Page, A.; Aneliunas, V.; Hieter, P.; Goodlett, D. R.; Aebersold, R., Increased quantitative proteome coverage with 13 C/12 C-based, acid-cleavable isotope-coded affinity tag reagent and modified data acquisition scheme. *PROTEOMICS* **2005**, 5, (2), 380-387.
31. Radulovic, D.; Jelveh, S.; Ryu, S.; Hamilton, T. G.; Foss, E.; Mao, Y.; Emili, A., Informatics Platform for Global Proteomic Profiling and Biomarker Discovery Using Liquid Chromatography-Tandem Mass Spectrometry* S. *Molecular & Cellular Proteomics* **2004**, 3, (10), 984-997.
32. Wang, E.; Miller, L. D.; Ohnmacht, G. A.; Liu, E. T.; Marincola, F. M., High-fidelity mRNA amplification for gene profiling. *Nature Biotechnology* **2000**, 18, 457-459.
33. Elrick, M. M.; Walgren, J. L.; Mitchell, M. D.; Thompson, D. C., Proteomics: Recent Applications and New Technologies. *BASIC AND CLINICAL PHARMACOLOGY AND TOXICOLOGY* **2006**, 98, (5), 432.
34. Bondarenko, P. V.; Shaler, T. A.; Chelius, D. H., Quantitation of biological molecules. **2006**.
35. 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. *Nat. Biotechnol* **1999**, 17, 994-999.
36. Liu, H.; Sadygov, R. G.; Yates Iii, J. R., A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem* **2004**, 76, (14), 4193-4201.
37. Xia, Q.; Wang, T.; Park, Y.; Lamont, R. J.; Hackett, M., Differential quantitative proteomics of *Porphyromonas gingivalis* by linear ion trap mass spectrometry: Non-label methods comparison, q-values and LOWESS curve fitting. *International Journal of Mass Spectrometry* **2007**, 259, (1-3), 105-116.

38. Zybaylov, B.; Mosley, A. L.; Sardu, M. E.; Coleman, M. K.; Florens, L.; Washburn, M. P., Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J Proteome Res* **2006**, 5, 2339-47.
39. Kislinger, T.; Cox, B.; Kannan, A.; Chung, C.; Hu, P.; Ignatchenko, A.; Scott, M. S.; Gramolini, A. O.; Morris, Q.; Hallett, M. T., Global Survey of Organ and Organelle Protein Expression in Mouse: Combined Proteomic and Transcriptomic Profiling. *Cell* **2006**, 125, (1), 173-186.
40. Zhang, H.; Yi, E. C.; Li, X.; Mallick, P.; Kelly-Spratt, K. S.; Masselon, C. D.; Camp, D. G.; Smith, R. D.; Kemp, C. J.; Aebersold, R., High Throughput Quantitative Analysis of Serum Proteins Using Glycopeptide Capture and Liquid Chromatography Mass Spectrometry*. *Molecular & Cellular Proteomics* **2005**, 4, (2), 144-155.
41. Xiong, L.; Andrews, D.; Regnier, F., Comparative proteomics of glycoproteins based on lectin selection and isotope coding. *J. Proteome Res* **2003**, 2, 618-625.
42. Xiong, L.; Regnier, F. E., Use of a lectin affinity selector in the search for unusual glycosylation in proteomics. *Journal of Chromatography B* **2002**, 782, (1-2), 405-418.
43. Block, T. M.; Comunale, M. A.; Lowman, M.; Steel, L. F.; Romano, P. R.; Fimmel, C.; Tennant, B. C.; London, W. T.; Evans, A. A.; Blumberg, B. S., Use of targeted glycoproteomics to identify serum glycoproteins that correlate with liver cancer in woodchucks and humans. *Proceedings of the National Academy of Sciences* **2005**, 102, (3), 779-784.
44. Hirabayashi, J., Lectin-based structural glycomics: Glycoproteomics and glycan profiling. *Glycoconjugate Journal* **2004**, 21, (1), 35-40.
45. Wells, L.; Vosseller, K.; Cole, R. N.; Cronshaw, J. M.; Matunis, M. J.; Hart, G. W., Mapping Sites of O-GlcNAc Modification Using Affinity Tags for Serine and Threonine Post-translational Modifications*. *Molecular & Cellular Proteomics* **2002**, 1, (10), 791-804.
46. Carlson, D. M., Oligosaccharides Isolated from Pig Submaxillary Mucin. *Journal of Biological Chemistry* **1966**, 241, (12), 2984-2986.
47. Karlsson, N. G.; Packer, N. H., Analysis of O-Linked Reducing Oligosaccharides Released by an In-line Flow System. *Analytical Biochemistry* **2002**, 305, (2), 173-185.
48. Varki, A., *The Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press: 1999.
49. Dell, A.; Reason, A. J.; Khoo, K. H.; Panico, M.; McDowell, R. A.; Morris, H. R., Mass spectrometry of carbohydrate-containing biopolymers. *Methods Enzymol* **1994**, 230, 108-32.
50. Felton, M. a. M. S. L., To Affinity and Beyond: Analytical Alphabet Soup. In *Chromatography*, 2001.

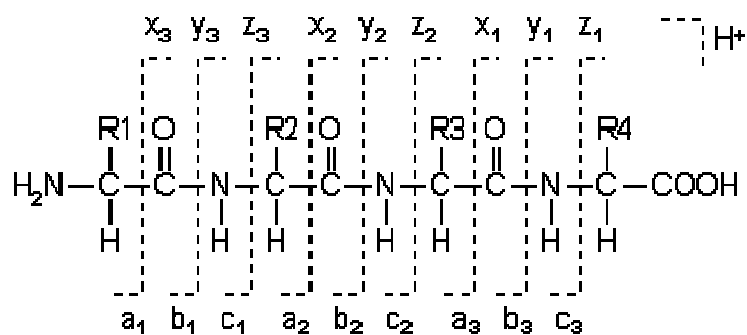
51. Hamelryck, T. W.; Dao-Thi, M. H.; Poortmans, F.; Chrispeels, M. J.; Wyns, L.; Loris, R., The Crystallographic Structure of Phytohemagglutinin-L. *Journal of Biological Chemistry* **1996**, 271, (34), 20479.
52. Cummings, R. D., Affinity Chromatography of Oligosaccharides and Glycopeptides. In *Affinity Separations: A Practical Approach*, 1997; pp 122-139.
53. Cummings, R. D., Roberta K. Merkle, Nancy L. Stults, Separation and Analysis of Glycoprotein Oligosaccharides. In 1989; pp 329-371.
54. Merkle, R. K., Richard D. Cummings, Lectin Affinity Chromatography of Glycopeptides. In *Analytical Methods*, 1987; pp 232-259.
55. Cummings, R. D., Guide to Techniques in Glycobiology. In 1994; Vol. 230, pp 66-88.
56. Bunkenborg, J.; Pilch, B. J.; Podtelejnikov, A. V.; Wisniewski, J. R., Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics* **2004**, 4, (2), 454-65.
57. Madera, M.; Mechref, Y.; Klouckova, I.; Novotny, M. V., High-sensitivity profiling of glycoproteins from human blood serum through multiple-lectin affinity chromatography and liquid chromatography/tandem mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* **2006**.
58. Wang, Y.; Wu, S.; Hancock, W. S., Approaches to the study of N-linked glycoproteins in human plasma using lectin affinity chromatography and nano-HPLC coupled to electrospray linear ion trap--Fourier transform mass spectrometry. *Glycobiology* **2006**, 16, (6), 514.
59. Qiu, R.; Regnier, F. E., Use of multidimensional lectin affinity chromatography in differential glycoproteomics. *Anal Chem* **2005**, 77, (9), 2802-2809.
60. Geng, M.; Ji, J.; Regnier, F. E., Signature-peptide approach to detecting proteins in complex mixtures. *J. Chromatogr. A* **2000**, 870, 295–313.
61. Kaji, H.; Saito, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Hirabayashi, J.; Kasai, K.; Takahashi, N.; Isobe, T., Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nature Biotechnology* **2003**, 21, (6), 667-672.
62. Zhang, H.; Li, X.; Martin, D. B.; Aebersold, R., Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nature Biotechnology* **2003**, 21, (6), 660-666.
63. He, P.; He, H. Z.; Dai, J.; Wang, Y.; Sheng, Q. H.; Zhou, L. P.; Zhang, Z. S.; Sun, Y. L.; Liu, F.; Wang, K., The human plasma proteome: Analysis of Chinese serum using shotgun strategy. *Proteomics* **2005**, 5, (13), 3442–3453.

64. Vosseller, K.; Hansen, K. C.; Chalkley, R. J.; Trinidad, J. C.; Wells, L.; Hart, G. W.; Burlingame, A. L., Quantitative analysis of both protein expression and serine/threonine post-translational modifications through stable isotope labeling with dithiothreitol. *PROTEOMICS* **2005**, 5, (2), 388-398.
65. Domon, B.; Costello, C. E., A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate Journal* **1988**, 5, (4), 397-409.
66. Harvey, D. J., Identification of protein-bound carbohydrates by mass spectrometry. *Proteomics* **2001**, 1, (2), 311-28.
67. Harvey, D. J., Collision-induced fragmentation of underivatized N-linked carbohydrates ionized by electrospray. *Journal of Mass Spectrometry* **2000**, 35, (10), 1178-1190.
68. Ethier, M.; Saba, J. A.; Ens, W.; Standing, K. G.; Perreault, H., Automated structural assignment of derivatized complex N-linked oligosaccharides from tandem mass spectra. *Rapid Communications in Mass Spectrometry* **2002**, 16, (18), 1743-1754.
69. Joshi, H. J.; Harrison, M. J.; Schulz, B. L.; Cooper, C. A.; Packer, N. H.; Karlsson, N. G., Development of a mass fingerprinting tool for automated interpretation of oligosaccharide fragmentation data. *PROTEOMICS* **2004**, 4, (6), 1650-1664.
70. Lohmann, K. K.; von der Lieth, C. W., GlycoFragment and GlycoSearchMS: web tools to support the interpretation of mass spectra of complex carbohydrates. *Nucleic Acids Research*.
71. Tang, H.; Mechref, Y.; Novotny, M. V., Automated interpretation of MS/MS spectra of oligosaccharides. *Bioinformatics* **2005**, 21, (1), i431-i439.
72. Yuan, J.; Hashii, N.; Kawasaki, N.; Itoh, S.; Kawanishi, T.; Hayakawa, T., Isotope tag method for quantitative analysis of carbohydrates by liquid chromatography-mass spectrometry. *J Chromatogr A* **2005**, 1067, (1-2), 145-52.
73. Hitchcock, A. M.; Costello, C. E.; Zaia, J., Glycoform quantification of chondroitin/dermatan sulfate using a liquid chromatography-tandem mass spectrometry platform. *Biochemistry* **2006**, 45, (7), 2350-61.
74. Bowman, M. J.; Zaia, J., Novel Tags for the Stable Isotopic Labeling of Carbohydrates and Quantitative Analysis by Mass Spectrometry.
75. Xie, Y.; Liu, J.; Zhang, J.; Hedrick, J. L.; Lebrilla, C. B., Method for the Comparative Glycomic Analyses of O-Linked, Mucin-Type Oligosaccharides. *Nat. Biotechnol* **1999**, 17, 994-999.
76. Alvarez-Manilla, G.; Warren, N. L.; Abney, T.; Atwood Iii, J.; Azadi, P.; York, W. S.; Pierce, M.; Orlando, R., Tools for glycomics: relative quantitation of glycans by isotopic permethylation using ¹³CH₃I. *Glycobiology* **2007**.

77. Atwood, J., Cheng, L., Alvarez-Manilla, G., Warren, N., York, W., Orlando, R., Quantitation by Isobaric Labelling: Application for Glycomics *Submitted* **2007**.

FIGURE 2.1:

The ions produced as a result of fragmentation of a peptide, with x , y and z fragment ions containing the C-terminus with one or more additional amino acid residues and a , b , and c ions containing the N-terminus with one or more additional amino acid residues.



CHAPTER 3

PROTEOMICS-

ANALYSIS OF PROTEINS FROM CAST CUTICLE OF

*ANOPHELES GAMBIAE**

* Botelho, J. C., Orlando, R.
To be submitted to *Journal of Proteome Research*

ABSTRACT

The aim of this work was identify multiple authentic cuticular proteins from *Anopheles gambiae* cast cuticles using a proteomics approach. Cuticles were collected from cast larval head capsules (CLHC) and cast pupal cuticles (CPC) of *Anopheles gambiae* during molting. The proteins were extracted from the cast cuticles and separated based on molecular weight by 1D SDS gel electrophoresis. Sections of the gels were excised for reduction, carbamidomethylation and in gel trypsin digestion. Additionally, the pellet that remained after protein extraction was also digested for analysis. The peptides from the gel slices and the pellet were analyzed using reverse phase liquid chromatography coupled with tandem mass spectrometry (RPLC MS/MS). The spectra were matched to peptides using the Matrix Science search engine, MASCOT, and Thermo Finnegan search engine, SEQUEST with a database constructed from known, novel, and ab initio predicted cuticular peptides of *Anopheles*.

From this analysis 259 peptides from putative cuticular proteins which matched to a minimum of 69 and a maximum of 119 different proteins were identified from the CLCH and CPC. The identified proteins are reported as authentic *Anopheles gambiae* cuticle proteins for the first time. In contrast, authentic cuticular proteins previously reported were identified via extraction, isolation and sequencing of individual proteins isolated from manually cleaned cuticles, proving to be more time consuming than the proteomics approach applied here.

INTRODUCTION

Anopheles gambiae

Anopheles gambiae is one of the primary malaria vectors of the *Plasmodium* parasite which transmits malaria to humans from the female *Anopheles* via a blood meal. According to the Center for Disease Control (CDC), malaria infects 300 to 500 million people worldwide and is fatal for more than a million and a half, mainly African children, each year. In addition to AIDS, malaria is one of the primary sources of mortality in Africa, Southeast Asia and Latin America and contributes to the continual abject poverty of these populations^{2,3}.

Drug resistance and lack of vaccine availability makes insecticides an important alternative to control major vectors of diseases. However, dwindling availability of insecticides, as a result of resistance, is a continuous problem in the mosquito vector²⁻⁴. As a result, there has been a disbandment of many vector control programs in the past three decades². It is believed that advances in genomics, including the completed genome sequence of *Anopheles gambiae*, will allow for a better understanding of the *Anopheles* mosquito which can lead to superior means of controlling it.

Anopheles gambiae has four stages in its life cycle: egg, larva, pupa, and adult. Adult females lay 50-200 eggs per oviposition, these eggs hatch within 2-3 days and emerge as larvae that develop through four stages, or instars. At the end of each instar, the larvae molt and shed their exoskeleton to allow for further growth. After the fourth instar the larvae transform into pupae. Pupae have a curved shaped where the head and thorax are combined and the abdomen curves underneath. After less than two days as a pupa, the dorsal cuticle of the head and thorax splits, leaving behind a cast exoskeleton, allowing the adult mosquito to emerge. Under the conditions used for to raise the *Anopheles* mosquito for this experiment the pupae emerge as

adults in less than 24 hours. It is the adult female *Anopheles* mosquito that acts as a malaria vector.

Cuticles

The cuticle is the outer layer, or exoskeleton, of the body surface of an insect and has been defined as an external compartment ⁵. Originally, in the 1940s, it was thought that there was a single protein, arthropodin, found in the cuticle ^{6,7}. However, as superior methods became available for separating proteins, the complexity of the multiple protein components of cuticles was revealed. The cuticle is defined by three distinct zones, the envelope, epicuticle, and procuticle (Figure 3.1) that shape and protect the insect from physical injury and dehydration ^{6,8-10}. Each zone is deposited on the surface of the insect by different processes ⁸.

The envelope, the 10-30nm outermost layer of the cuticle, is laid down at the plasma membrane surface, at the tips of microvilli ⁵. This is the first layer of new cuticle to be formed after molting. This coating serves as the first face of protection for the insect ^{5,8}. The next zone to form after molting is the epicuticle. This segment is approximately 1 µm thick and is believed to be chitin-free ⁸. Formation of the epicuticle entails discharge from the secretory vesicles into the cuticular region where assembly is completed on the inner face of the envelope. Due to the affinity the epicuticle components have for one another a homogenous layer is formed and is stabilized by quinones. The procuticle, a grouping of what has been called the *exo*- and *endo*-cuticle, the 3rd zone to form, is usually hundreds of times wider and much more complicated than the envelope and epicuticle combined. The procuticle is normally formed as stacks of laminae, with each lamina being a single layer of chitin fibers that oriented parallel to each other. Chitin

and associated proteins are added at the cell surface by secretion from the epidermis of the insect into the procuticle.

It was unreasonable to attempt to manually collect whole *Anopheles* cuticles free of cellular contamination, so analysis on cast cuticles was performed. Between each developmental stage, insects shed their old cuticle and a cast cuticle remains with the envelope intact along with portions of procuticle that were not digested by molting fluid ¹⁰. Most of the cuticular proteins are degraded by molting fluid and resorbed before the cuticle is cast; however, the various proteins that are left behind in the cuticle can be used in a preliminary evaluation of the first authentic cuticular proteome of *Anopheles*. Previously, the identification of authentic cuticular proteomes of other species has been based on isolation and sequencing of individual proteins extracted from manually cleaned cuticles.

Cuticular Proteins

There is recurring sequence motifs found in cuticular protein, the most common is a region, first reported by Rebers and Riddiford in 1988, referred to as the R&R Consensus ^{5, 6, 11}. The original R&R Consensus contained the following sequence: G-x(8)-G-x(6)-Y-x-A-x-E-x-G-Y-x(7)-P-x-P, where x is any amino acid and the number in parentheses represents the number of amino acids. There have been some modifications to this original sequence, where the last Y can be replaced by an F and up to 6 amino acids can be found between the first Y and GY instead of the original five as shown above ⁵. Proteins with the R&R Consensus can be split into three groups: RR-1, RR-2, and RR-3 ⁵.

The RR-1 proteins were first isolated from flexible or soft cuticles and contain the short sequence of the original consensus listed above –Y-x-A-x-E-x-G-(FY)-x(7)-P, where the

multiple letters in parentheses indicates that either amino acid may be present. Additionally, a glutamic acid is found in the conserved portion in bold and the N-terminus of this region contains an aromatic triad, such as (FY)-x-(FY)-x-(FY). The RR-2 proteins have been associated with hard cuticles and were first recognized by Bouhin et. al. and Charles et. al. in 1992^{5, 12}. RR-2 proteins contain an extension of the original sequence, shown here in italics, G-x(8)-G-x(6)-Y-x-A-x-E-x-**G-F-x-A-x-V**-x(3)-P, additionally histidine and lysine residues are abundant in this region as compared to RR-1^{5, 13}. The class of RR-3 proteins was introduced in 2000 by Anderson et. al., where RR-3 proteins contain a C-terminal sequence similar to RR-1 and RR-2 as described above but has conserved regions in the N-terminal region¹⁴.

Other common motifs found in cuticular proteins, with and without the R&R consensus, are repeated sequence of A-A-P-(A/V), long stretches of glycine, leucine, and tyrosine designated as a G(Y/L) motif, an 18 amino acid repeat, and a 51 amino acid repeat⁵. The 18 amino acid repeat contains the basis sequence (PV)-x-D-T-P-E-V-A-A-A-(KR)-A-A-(HF)-x-A-A-(HY) and presumably is a structurally important part of cuticles since it occurs in the cuticle of various insects and crustaceae^{5, 14}. The 51 residue motif identified in cuticles of *Locusta* and *Tenebrio* are not found in proteins with the R&R consensus and contain the sequence x(6)-A-x(9)-R-S-x-G-x(4)-V-S-x-Y-x-K-x(2)-D-x(3)-S-S-V-x-K-x-D-x-R-x(2)-N-x(3)^{5, 15}.

As of May 2007, 519 cuticular protein sequences were available through the cuticle database, cuticleDB, containing all structural proteins of arthropod cuticles identified to date (<http://bioinformatics2.biol.uoa.gr/cuticleDB/index.jsp>). Most of the known authentic cuticular proteins in the database were identified by Anderson et. al. through analyses of extracted proteins of cuticles from diverse species of arthropods or by N-terminal sequences of proteins extracted

from cuticles. However, a majority of the sequences available are identified *in silico* from conceptual translation of cDNAs or genomics sequences that share common characteristics with authentic cuticular proteins.

To classify annotated protein sequences derived from whole genome a protein families database, Pfam, can be used. Pfam is a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains and families (<http://www.sanger.ac.uk/Software/Pfam>)¹⁶. An annotated protein can be examined using a Basic Local Alignment Search Tool, BLAST, against all known and predicted proteins; information obtained correlates to Pfam entries. For classification of putative cuticular proteins with an R&R consensus sequence a match to pfam00379 is required. Pfam0037 is a 68 amino acid sequence that is based on proteins in both RR-1 and RR-2 classes and is identified as chitin-bind-4, insect cuticle protein. A majority of the proteins identified in the cuticleDB, approximately 70%, contain the R&R consensus⁵.

EXPERIMENTAL PROCEDURE

Sample Collection and Preparation

Anopheles gambiae colonies maintained by Mark Brown at the University of Georgia were used for this analysis. The cast cuticles of fourth instar larval head capsules and pupae were collected (Figure 3.2), lyophilized, weighed, and pulverized in liquid nitrogen with a mortar and pestle. Cuticle proteins were extracted from 20mg of cast larvae head capsules (CLHC) and cast pupae cuticles (CPC) at 100°C for 15 minutes with SDS-sample buffer and 50mM dithiothreitol (DTT) (Invitrogen, Carlsbad, CA). The supernatant and the pellet were used for further analysis. Collection and preparation of samples were done by Dr. Ningjia He in the

Willis Group at the University of Georgia. The supernatant from the protein extraction was loaded onto a 4-20% Tris-HCl pre-cast gradient gel (Bio-Rad, Hercules, CA) and electrophoresis was performed to separate the cuticle proteins by size. The gel was cut into three unequal horizontal regions (Figure 3.3) and washed with water.

LC-MS/MS analysis of tryptic peptides

The gel slices were rinsed twice in 25mM NH_4HCO_3 / 50% acetonitrile (ACN) and once with 25mM NH_4HCO_3 at a pH of 8.5. The slices were then incubated at 55°C for 1 hour with 10mM DTT in 25mM NH_4HCO_3 and for 45min in the dark at room temperature with 55mM iodoacetamide. The gel pieces were rinsed again with 25mM NH_4HCO_3 and dried down completely by a SpeedVac. Sequencing Grade Modified Trypsin (Promega, Madison, WI) was added to the dried sample at approximate 1:100 ratio (w/w) and incubated at 37°C for a 16 hour digestion. The peptides obtained from this digestion were extracted from the gel pieces with 25mM NH_4HCO_3 followed by 50%ACN/5% formic acid and concentrated on a SpeedVac to a final volume of 100 μL . The peptide solution was cleaned up through a 10 kDa MWCO centrifugal filter device; Microcon YM-10 regenerated cellulose (Millipore, Billerica, MA). The pellet, remaining after protein extraction with SDS from the cast cuticles, was also digested. The pellet was washed five times with 25mM NH_4HCO_3 and digested with Sequencing Grade Modified Trypsin (Promega, Madison, WI) at approximate 1:100 ratio (w/w) and incubated at 37°C for a 16 hour digestion.

A column was prepared from a silica capillary with a pulled tip (75 $\mu\text{m} \times 10\text{cm}$; tip 15 ± 1 μm , PicoTip EMITTER, New Objective, Woburn, MA) by packing silica C18 resins (Rainin Microsorb MV, 5 μm , 300 Å pore size) using a pressurized stainless steel cell and nitrogen gas at 1,000 psi. The trypsin digested sample was loaded directly onto the column for 30 min using the

pressurized cell. The loaded spray column was interfaced directly with a Finnegan LTQ Linear Ion Trap Mass Spectrometer (Thermo Finnegan, San Jose, CA) coupled to a reversed-phase capillary HPLC system. The spray voltage was held at 1.62 kV with a capillary temperature of 220°C. The peptides were eluted into the MS at approximately 400 nanoL/min with a 65-minute linear gradient of 5 to 60% ACN in 0.1% formic acid. The MS was operated in a data-dependent mode in which the top 9 precursor ions identified during each MS scan were further fragmented and analyzed by MS/MS.

Data evaluation

A database was assembled by Augustine Dunn in the Willis laboratory for this analysis from 35,641 *Anopheles* protein sequences obtained February, 2005 from Ensembl (www.ensembl.org). This database contained three types of peptides: known, novel, and ab initio predicted. In addition, sequences generated by manual annotation of genomic sequences that appeared to code for cuticular proteins based on the presence of known motifs or other sequence similarities were added. It is also important to note that the DNA used for the whole genome sequencing came from the PEST strain of *Anopheles gambiae*, this strain no longer exists, the peptides came from G3 strain.

MS/MS data obtained from the LTQ runs were searched using both MASCOT (Matrix Science, Boston, MA) and SEQUEST (Thermo Finnegan, San Jose, CA) search engines. Both search engines were employed to insure that the maximum numbers of peptides were matched. Searches were limited to tryptic peptides, allowing up to 3 missed cleavages. Peptide mass tolerance was set to ± 1.2 Da and MS/MS tolerance to 0.2 Da which corresponded to the instrument used for this analysis. A variable modification of carbamidomethylation on cysteine

residues (+57.0293 Da) was allowed because of the possible reduction and alkylation of disulfide bonds. Peptide matches obtained from both search engines were filter using peptide scores deemed statistically significant. A MASCOT score of 36 or greater for a matched peptide and a SEQUEST cross-correlation value greater than 2.0 for a singly charged, 2.5 for a doubly charged, and 3.0 for a triply charged ion with ion coverage of 60% or more was accepted.

RESULTS AND DISCUSSION

Cuticular proteins were successfully matched from the CLHC and CPC in both the gel slices and the final pellet, allowing for the identification of the first authentic *Anopheles gambiae* cast cuticle proteome. The combined search results of MASCOT and SEQUEST resulted in a total of 436 unique peptides from the CLHC and 409 unique peptides from the CPC composed of five classes of proteins. These classes include structural cuticular proteins, secretory proteins associated with the cuticle, remnants of muscle proteins, proteins from molting fluids, and proteins that can not be confidently assigned to any of the four previously described groups.

Novel Approach to Cuticular Protein Matching

As stated, the identification of authentic cuticular proteins previously has been based on isolation and sequencing of individual proteins extracted from manually cleaned cuticles. However, this proteomics method, using mass spectrometry, allowed for the identification of multiple proteins without extraction of individual proteins. MS/MS data dependent analysis of the top nine precursor ions of each MS scan in combination with database searching allowed multiple cuticular proteins to be matched from a single sample.

Preliminary studies of cuticular proteins, by our laboratory, included MS/MS analysis of cast cuticle using a typical online LC-MS/MS proteomics set up. The samples were injected into the HPLC via an auto sampler and flowed through an online commercial C18 capillary column. The eluents from the column was sprayed into the MS instrument for fragmentation. Prior attempts to analyze the cast cuticles of *Anopheles gambiae* using this method proved to be somewhat problematic as a result a minimal number of proteins were identified in the preliminary studies. It was believed this was due, in part, to dilution and loss of the sample in the plumbing of the HPLC prior to introduction into the MS. An additional challenge that occurred during sample analysis was the column had a tendency to clog due to the make up of the sample. Some of these issues were addressed in the sample collection and preparation steps; however, the analysis could be further optimized by altering the typical LC-MS/MS set up.

To address the problem of sensitivity, a silica capillary C18 column fabricated using a pressurized cell was utilized. The pressurized cell allowed for direct application of the sample onto the fabricated column off line. Not only was the sample directly loaded onto the column but also a more concentrated sample could be added for an extended time period. After application of the peptides, the column was interfaced directly with the MS and a voltage applied so that the column also serves as the spray tip into the instrument. As the change in gradient was applied the appropriate peptides are eluted from the C18 column directly into the MS. It should be noted that this off line method was more time consuming and did not allow for automation. However, using this approach allowed for the identification of a greater number of proteins which will be discussed further in this section. Additionally, any issues that were previously encountered by the column clogging were addressed off line prior to analysis by the MS and therefore did not directly interfere as it was previously.

Structural Cuticular Proteins

No obvious differences in the cuticular peptides identified between the 2 sources of cast cuticles, CLHC or CPC, was observed in this analysis. Therefore, the results from these 2 anatomical sources will be grouped together during discussion of structural cuticular proteins identified. There were, on the other hand, noticeable differences between the two peptide preparation methods, SDS extraction run out on a gel compared to the final pellet left after extraction. Additional peptides were identified in the final pellet versus the gel slices with a larger amount of peptides from structural and secreted cuticular proteins in the final pellet. It is probable that a greater number of peptides that are found in the final pellet came from proteins that were insoluble to the SDS extraction.

Structural cuticular proteins were identified based on their similarity to authentic cuticular proteins in the database which included R&R Consensus and the 51-amino acid repeat proteins. This proteomic analysis resulted in a total of 295 peptides matches to putative cuticular proteins. These peptides correspond to at least a minimum of 69 and a maximum of 119 different authentic cuticular proteins.

A total of 77 structural cuticular peptides were found in the 6 gel slice samples and 276 peptides were found in the 2 trypsin treated pellet. Only 19 peptides were unique to the gel slice preparation method however, none of these peptides identified unique proteins. In comparison, the final trypsin digested pellet allowed for the identification of an additional 23 proteins.

A majority of identified cuticular proteins contained the R&R Consensus sequence with 199 different peptides. These peptides are represented here as CPR#. A total of 87 cuticular proteins were matched with a maximum of 16 proteins from the RR-1 family, 69 from RR-2 and 2 from RR-3 as shown in Table 3.1. It should be noted that 50 of these proteins contained only

peptides that were common in other proteins therefore these proteins can not be uniquely identified in this analysis. This results in a minimum of 37 RR Family cuticular proteins that were matched in this analysis. The proteins that do not contain unique peptides are printed in *italics* in Table 3.1. The RR-Family cuticular proteins identified in this analysis have an average length of 176 amino acid residues and an average 26% coverage of the sequence was achieved.

Additional structural proteins were identified in this proteomic analysis. These include 4 proteins with the 51 amino acid-repeat domain represented as CPF1-4. Peptides from all four annotated members of this 51 repeat domain family were matched, although it should be noted that in *Anopheles* this conserved domain is only 40-41 amino acids. The CPF Family of proteins contained an average of 195 amino acids and an average 24% sequence coverage was achieved. Peptides were also found in 7 groups of proteins represented in ENSEMBL as ENSF0000001867. These proteins have a carboxyl terminal region that is similar to the CPF family and are represented as CPFL1-7. The CPFL proteins have an average of 197 amino acids found in their sequence and an average of 19% coverage was attained in this investigation. Expression analyses were performed by Togawa et. al., their results coupled with this proteomics analyses provided evidence that all of the 4 CPFs and 7 CPFLs identified are authentic cuticular proteins¹⁷.

We also identified peptides unique to 16 cuticular proteins with low sequence complexity named CPLC1-16. This family includes proteins with approximately 255 amino acids. An average sequence coverage of 25% was achieved for this protein family. Additionally, a new family of cuticular proteins that contains 2 conserved cysteines, named CPTC, was identified. 4 proteins in the CPTC family (CPTC1-4) were matched in this proteomic analysis with an average 48% sequence coverage of it 103 amino acid sequence.

Secretory Proteins Associated with the Cuticle

The Ensembl database was used to obtain sequences of secretory proteins associated with the cuticle. In our proteomic study we matched 44 peptides from 14 secreted proteins likely to be present in the cuticle in the gel slice samples and 85 peptides from 14 proteins in the final pellet. These proteins included prophenoloxidasases, arylphorin-like, C-type lectins, peritrophins, and “yellow” proteins.

Nine prophenoloxidasases (ProPO) have been annotated from *Anopheles* genome, this analysis allowed for identification of peptides from 5, ProPO-1, 2, 3, 7, and 8. In addition, 2 prophenoloxidasase activating factors were also identified in this proteomic investigation. These results are not surprising given that prophenoloxidasases has been established as part of the cuticle with suggested roles in sclerotization, wound healing and defense against microbes ¹⁸.

In this study, 4 arylphorins-like proteins were matched. Arylphorins are proteins with high content of aromatic amino acids which are known to be synthesized in the hemocytes and epidermis but are found in the cuticle ⁵. Their role is generally assumed to be involved in sclerotization due to its high tyrosine content. Four C-type lectins were matched. One of them, C-type lectin A16, was previously identified in *Anopheles* by Dimopoulos et. al. ¹⁹. Their function in the cuticle has not been determined. Eight proteins with chitin binding domains, classified as peritrophins, were identified. Peritrophins are associated with the peritrophic matrix which is a film-like structure that separates food from midgut tissue. The peritrophin membrane protects the epithelium against food abrasion and microorganisms ²⁰. Whether these proteins contribute to the cuticle in addition to the peritrophic membranes is yet to be established. In our analysis 4 “yellow” or major royal jelly proteins (MRJP) proteins were found. This family of proteins has been extensively studied in *Drosophila* by Drapeau et. al. ²¹. For this project the

“yellow” proteins were annotated for *Anopheles* resulting in 13 punitive “yellow” proteins when the MRJP domain, pf03022, was searched. “Yellow” proteins are localized to the cuticle as shown in immunocytochemistry analyses and have been portrayed as a structural component of the cuticle that aids in the deposition of melanin or black cuticle formation ⁵.

Other Proteins

The other proteins identified in this analysis include muscle-derived proteins, molting fluid-derived proteins and miscellaneous proteins. Muscle attachment fibers are known to be implanted in the cuticle, therefore identification of such muscle derived proteins such as actin, myosin, and tubulins is not surprising ¹⁸. 68 peptides which associated to 23 muscle derived-proteins were identified in the gel slice samples and only 4 peptides from 4 proteins in the final pellet analysis. Additionally enzymes that are associated with muscle were also found.

Molting fluid-derived proteins were identified in this analysis even with the extensive washing of the cast cuticles. Prominent among them were proteases and chitinases. 37 peptides from 16 molting fluid derived proteins were matched in the gel slices samples as compared to 14 peptides from only 9 proteins in the final pellet.

Approximately 9% of our proteome, in both sample preparation methods, was comprised of proteins with no description or were based on a single MS/MS spectrum. As a result, these 81 proteins are grouped as miscellaneous from this study. Whether these proteins are false positives or authentic cuticle components cannot be determined at this time.

CONCLUSIONS

This study of cast cuticles of *Anopheles gambiae* allowed for the identification of the first authentic cast cuticular proteome. From this analysis 259 peptides from putative cuticular proteins that matched to a minimum of 69 and a maximum of 119 different proteins were identified from the CLCH and CPC. At least 33 and possibly as many as 87 authentic cuticular proteins were identified from the 149 annotated *Anopheles* RR proteins. This quantity is especially notable given that cuticles came from only 2 sources, head capsules and pupa, and from cast cuticles rather than whole. The use of cast cuticles allowed for a more convenient sample collection method but limited the protein pool since the endocuticle was likely digested by molting fluid and primarily exocuticle was analyzed.

All of the cuticular proteins in this study could be identified from the material in the final pellet. As a result, future studies could be done solely on the final pellet without running the SDS extractions on a 1D gel. This would require a reduced amount of time for analysis since 1D electrophoresis would not have to be preformed and only 1 sample, the final pellet, would need to be analyzed. Extensive SDS extractions would still be required since it may help in the removal of some of the non cuticular proteins such as molting fluid and muscle remnants. However, analysis of these extractions would not necessary.

ACKNOWLEDGEMENTS

This work was supported by a grant from the National Institutes of Health (AI55624) to JHW and by the National Institutes of Health/NCRR funded Integrated Technology Resource for Biomedical Glycomics (P41RR018502).

REFERENCES

1. He, N.; Botelho, J. M.; McNall, R. J.; Belozarov, V.; Augustine Dunn, W.; Mize, T.; Orlando, R.; Willis, J. H., Proteomic analysis of cast cuticles from *Anopheles gambiae* by tandem mass spectrometry. *Insect Biochem Mol Biol* **2007**, 37, (2), 135-46.
2. Guerin, P. J.; Olliaro, P.; Nosten, F.; Druilhe, P.; Laxminarayan, R.; Binka, F.; Kilama, W. L.; Ford, N.; White, N. J., Malaria: current status of control, diagnosis, treatment, and a proposed agenda for research and development. *Lancet Infectious Diseases* **2002**, 2, (9), 564-573.
3. Sachs, J.; Malaney, P., The economic and social burden of malaria. *Nature* **2002**, 415, (6872), 680-685.
4. Sachs, J. D., A New Global Effort to Control Malaria. In 2002.
5. Willis, J. H., V. A. Iconomidou, R. F. Smith, S. J. Hamodrakas *Cuticular Proteins* Elsevier 2005; Vol. 4, p 79-104.
6. Willis, J. H., Cuticular Proteins in Insects and Crustaceans. *American Zoology* **1999**, 39, 600-609.
7. Fraenkel, G.; Rudall, K. M., A Study of the Physical and Chemical Properties of the Insect Cuticle. *Proceedings of the Royal Society of London. Series B, Biological Sciences* **1940**, 129, (854), 1-35.
8. Locke, M., The Wigglesworth Lecture: Insects for studying fundamental problems in biology. *J Insect Physiol* **2001**, 47, (4-5), 495-507.
9. Vincent, J. F. V.; Wegst, U. G. K., Design and mechanical properties of insect cuticle. *Arthropod Structure & Development* **2004**, 33, (3), 187-199.
10. Vincent, J. F., Insect cuticle: a paradigm for natural composites. *Symp Soc Exp Biol* **1980**, 34, 183-210.
11. Rebers, J. E.; Riddiford, L. M., Structure and expression of a *Manduca sexta* larval cuticle gene homologous to *Drosophila* cuticle genes. *J Mol Biol* **1988**, 203, (2), 411-23.
12. Charles, J. P.; Bouhin, H.; Quennedey, B.; Courrent, A.; Delachambre, J.; Laboratoire de Zoologie, U. A.; de Bourgogne, U.; Dijon, F., cDNA cloning and deduced amino acid sequence of a major, glycine-rich cuticular protein from the coleopteran *Tenebrio molitor*. *Eur. J. Biochem* **1992**, 206, 813-819.

13. Iconomidou, V. A.; Willis, J. H.; Hamodrakas, S. J., Unique features of the structural model of hard cuticle proteins: implications for chitin-protein interactions and cross-linking in cuticle. *Insect Biochem Mol Biol* **2005**, 35, (6), 553-60.
14. Andersen, S. O., Studies on proteins in post-ecdysial nymphal cuticle of locust, *Locusta migratoria*, and cockroach, *Blaberus craniifer*. *Insect Biochem Mol Biol* **2000**, 30, (7), 569-77.
15. Andersen, S. O.; Rafn, K.; Roepstorff, P., Sequence studies of proteins from larval and pupal cuticle of the yellow meal worm, *Tenebrio molitor*. *Insect Biochem Mol Biol* **1997**, 27, (2), 121-31.
16. Bateman, A.; Birney, E.; Cerruti, L.; Durbin, R.; Eddy, S. R.; Griffiths-Jones, S.; Howe, K. L.; Marshall, M.; Sonnhammer, E. L. L., The Pfam Protein Families Database. *Nucleic Acids Research* **2002**, 30, (1), 276-280.
17. Togawa, T.; Augustine Dunn, W.; Emmons, A. C.; Willis, J. H., CPF and CPFL, two related gene families encoding cuticular proteins of *Anopheles gambiae* and other insects. *Insect Biochemistry and Molecular Biology* **2007**, 37, (7), 675-688.
18. Asano, T.; Ashida, M., Cuticular Pro-phenoloxidase of the Silkworm, *Bombyx mori* PURIFICATION AND DEMONSTRATION OF ITS TRANSPORT FROM HEMOLYMPH. *Journal of Biological Chemistry* **2001**, 276, (14), 11100-11112.
19. Dimopoulos, G.; Richman, A.; della Torre, A.; Kafatos, F. C.; Louis, C., Identification and characterization of differentially expressed cDNAs of the vector mosquito, *Anopheles gambiae*. In National Acad Sciences: 1996.
20. Wijffels, G.; Hughes, S.; Gough, J.; Allen, J.; Don, A.; Marshall, K.; Kay, B.; Kemp, D., Peritrophins of adult dipteran ectoparasites and their evaluation as vaccine antigens. *Int J Parasitol* **1999**, 29, (9), 1363-77.
21. Drapeau, M. D., The Family of Yellow-Related *Drosophila melanogaster* Proteins. *Biochemical and Biophysical Research Communications* **2001**, 281, (3), 611-613.

FIGURE 3.1:

Artist drawn depiction of the 3 layers of an insect cuticle⁸

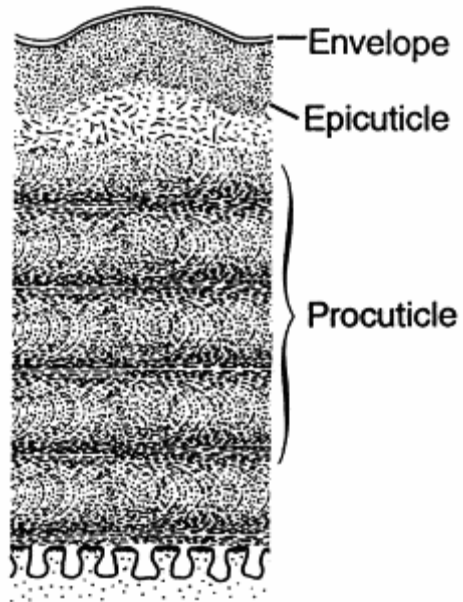


FIGURE 3.2:

Microphotographs of two cast cuticles, cast pupal cuticle (CPC) and cast larvae head capsule (CLHC) used in this study ¹

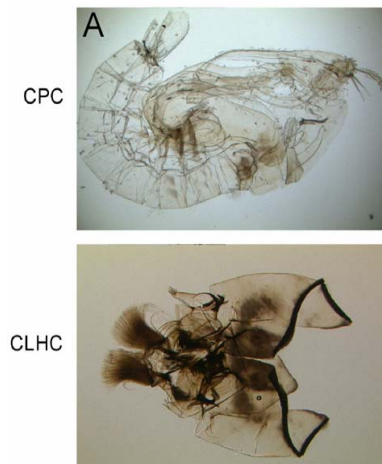


FIGURE 3.3:

A photograph of the comassie-stained 1D SDS gel on the extracted proteins from CLHC and CPC with lines added to show the approximate location of where gel slices were cut for further analysis ¹

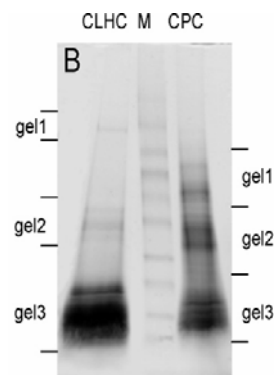


TABLE 3.1:

Distribution of the cuticular proteins of the RR-family matched in this proteomic analysis.

Proteins listed in ***italics*** contained only peptides that were common in other proteins therefore these proteins can not be uniquely identified.

RR-1	CPR 9, 11, <i>12, 13</i> , 15, 16, 21, 23, 62, 79, 106, 113, 125, 126, 126, 134
RR-2	CPR 1, 2, 3, 4, 5, 6, <i>34, 35, 36, 39, 40, 41, 42, 43, 44, 45, 46, 48, 49, 50, 51, 52, 53, 54, 58, 59, 60, 65, 66, 70, 72, 82, 83, 84, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 107, 108, 109</i> , 110, <i>115, 116, 117, 118, 119</i> , 120, <i>121, 123, 124, 131, 135, 136, 140, 146, 148, 149</i>
RR-3	CPR 111, 130

CHAPTER 4

GLYCOPROTEOMICS –

A PROOF OF PRINCIPLE STUDY: BIOSPECIFIC ISOLATION AND LABEL FREE
COMPARISON OF COMPLEX N-LINKED GLYCOPROTEINS IN SERUM OF PATIENTS
WITH MALIGNANT AND BENIGN OVARIAN TUMORS *

* Botelho, J. C., Orlando, R.
To be submitted to *Journal of Proteome Research*

ABSTRACT

The aim of this work was to compare low abundance complex type glycoproteins that carry a β 6 -N-Acetylglucosamine (GlcNAc) branch from the blood serum of patients with benign and malignant ovarian tumors using a glycoproteomics approach. The identified glycoproteins could serve as potential biomarker candidates for early detection of ovarian cancer. Lectin affinity chromatography (LAC) was used to enrich for complex glycoproteins with tri and/or tetra - antennary N-linked glycans from depleted serum using a *Phaseolus Vuglaris* Leukoagglutinin (PHA-L) agarose column. The isolated glycoproteins were digested with trypsin and analyzed in triplicate using reverse phase liquid chromatography coupled with tandem mass spectrometry (RPLC MS/MS). The spectra were matched to peptides using the Matrix Science search engine, MASCOT, with the National Center for Biotechnology Information (NCBI) human database. The matched peptides were grouped into proteins with a 1% false discovery rate using ProValT, protein validation software program. From this analysis 53 proteins were identified in the serum of patients with malignant tumors, 44 in patients with benign tumors, and 34 in a control sample of pooled male serum. This accounted for a total of 63 unique proteins from the 9 serum samples. From this analysis, 11 proteins were identified that with further investigation could serve as potential biomarker candidates.

INTRODUCTION

Cancer is the second leading cause of death in the United States with half of all men and one third of all women developing cancer during their lifetime¹. Today, millions of people are living with cancer or have had cancer at some point in their lives. Ovarian cancer is the ninth most common cancer and the fifth leading cause of cancer death among women in the US. According to the American Cancer Society there will be approximately 23,000 new cases of ovarian cancer in the US this year (2007) and half of the women diagnosed will not survive more than 5 years.

The survival rate among women with malignant ovarian cancer is correlated with the stage at which it is diagnosed¹. When the diagnosis is made in Stage 1, where the cancer is found only in the ovaries, there is a 93% 5 year survival rate. A diagnosis made in stage 2, where the cancer has spread to other organs in the pelvic region, results in a 70% 5 year survival rate. Diagnosis made in stage 3 and 4, where the cancer has spread beyond the pelvic region, results in a 5 year survival rate of only 15-20%. Currently, over 80% of ovarian cancer diagnoses are made in stage 3 or 4 and in spite of the increase in cancer research these statistics have remained almost stagnant for the last 50 years².

Late detection, often in what is the final stage of the disease, is in large part a result of the inadequate detection methods presently available. Current diagnostic tools including pelvic exams, ultrasounds, and the CA125 blood test do not provide a reliable diagnosis for early detection due to a lack of sensitivity. It is obvious that new reliable and cost effective detection methods for ovarian cancer has a potential to dramatically reduce mortality^{1,2}. One such method could be the use of a protein based molecular biomarker to indicate the early onset of the disease state. Since malignant transformation typically involves a change in protein expression with

subsequent proliferation of the altered cells changes in these proteins could be monitored, both qualitatively and quantitatively and provide valuable information that may be an aid to more effective diagnosis of the disease.

One type of malignant transformation that can be observed is changes in glycosylation. Altered glycosylation occurs in proteins produced by tumor cells and has been observed in essentially all types of experimental and human cancers^{4, 9, 10} and is considered a universal feature of cancer cells^{8, 11-16}. Because the glycosylation of proteins is known to change in tumor cells during the development of cancer, a glycoproteomics approach could be useful to find relevant biomarkers of ovarian cancer. One of the most widely occurring changes in glycosylation observed in malignant cancer cells is the increase in the number of branches on complex N-linked glycan structures. The addition of GlcNAc β 6 branch has been directly associated with invasive and metastatic potential^{4, 10, 17-19}. An increase in glycan branching was first reported in 1969 by Meezan *et al.* in a comparative study of the plasma membrane glycoproteins of normal mouse 3T3 cells and cells with the tumor Simian virus (SV)-transformants, a virus known to cause tumors in rodents^{9, 13, 20, 21}. This analysis showed that the transformed cells were physically larger than those from non-transformed cells^{9, 20, 21}. Later, Buck *et al.* reported that the same phenomenon was detected by the comparative study of glycopeptides of various fibroblast cell lines and their malignant counterparts produced virally and chemically^{9, 21-23}. From these studies it was confirmed that larger glycopeptides are found in malignant cells and it became known as the “Warren-Glick” phenomenon^{9, 21}. It was estimated and later established by structural study of released oligosaccharides that the molecular basis of the “Warren-Glick” phenomenon was an increase in the tri- and tetra- antennary complex N-

linked glycans, more specifically the GlcNAc β 6 branch, in malignant cells, shown in figure 4.1^{13, 21, 24, 25}.

The source of increased GlcNAc β 6 branching in transformed cells was reported to be the result of an increase in the specific activity of N-acetylglucosaminyltransferase-V (GnT-V), the glycosyltransferase that initiates the synthesis of this branch^{21, 24, 26}. Recent studies, employing manipulation of glycosyltransferase genes, show the activity of GnT-V is elevated in malignant cancer cells^{10, 17-19, 26, 27}. It is believed that elevated GnT-V activity in tumors might commonly occur by a direct effect on GnT-V's promoter region caused by the up regulation of proto-oncogenes such as the Ets gene family of transcriptional activators, *src*, and *erbB2*^{8, 10, 21, 28, 29}. The cascade of events responsible for affecting the GnT-V promoter is thought to be initiated by activation of a growth factor receptor in the cell membrane through a cellular proliferation signal pathway^{8, 10, 21, 28}. The increased branching of glycoproteins, as well as the increased activity of GnT-V, has been observed in several types of cancer cells including: breast, colon, esophagus, blood, and skin^{4, 9, 12, 13, 17, 18, 21, 26, 27, 30, 31}.

In this paper a proof of principle study for high throughput quantitative proteomic analysis on blood serum is described, for the identification of potential ovarian cancer biomarkers. This method, described as a glycoproteomics approach, is based on the selective isolation of tri- and tetra- antennary complex N-glycoproteins with a GlcNAc β 6 branch from depleted serum using LAC and the analysis of the isolated complex glycoproteins using a conventional one dimensional proteomic approach with LC/MS/MS. Additionally, normalized spectral counts (NSC) were utilized to compare the relative abundance of proteins identified in each population.

EXPERIMENTAL PROCEDURE

Sample collection and depletion

Human blood serum samples were acquired from 2 sources: The Ovarian Cancer Institute (OCI) in Atlanta, GA and Sigma Aldrich (St. Louis, MO). Four OCI serum samples were obtained from patients diagnosed with malignant ovarian cancer tumors (C1-C4). In addition, four OCI serum samples were obtained from patients diagnosed with benign ovarian cancer tumors (A1-A4). These samples were collected intravenously by a registered nurse while patients were prepped for a surgical procedure. The clinical diagnosis of these patients is shown in Tables 4.1a and 4.1b. The serum sample purchased from Sigma Aldrich consisted of pooled serum from human male plasma.

All samples were depleted of six proteins found in high abundance in serum (Albumin, IgG, IgA, Transferrin, Haptoglobin, and Antitrypsin) by a Multiple Affinity Removal System, MARS, 4.6 x 50mm column (Agilent, Wilmington, DE). The depleted serum collected from the MARS column was concentrated with Vivaspin 20 concentrators with a 5k MWCO (VivaScience, Hannover, Germany).

Lectin Affinity Chromatography (LAC)

GlcNAc β 6 branched tri and/or tetra – antennary complex N-linked glycoproteins were isolated using a lectin column with agarose bound PHA-L at a concentration of 3mg/mL (Vector Labs, Burlingame, CA). 40 μ L of this lectin was packed into a Micro Bio-Spin Chromatography Column (Bio-Rad Laboratories, Inc., Richmond, CA). The column was equilibrated with a binding buffer solution of phosphate buffered saline (1x PBS) before use, as described previously³⁸. Depleted serum was loaded onto the lectin column and allowed to incubate for at

least 2 hours. The unbound proteins were collected by passing 1 mL of binding buffer through the lectin column. An elution buffer containing the haptenic sugar N-Acetyl-D-galactosamine (GalNAc) at a concentration of 0.3M in PBS (Sigma, St. Louis, MO) was then added to the lectin column and allowed to incorporate for 30 minutes. The bound glycoproteins were collected by passing the elution buffer through the lectin column in addition to 1 mL of binding buffer in 100 μ L increments to insure complete collection. All of the bound fractions were pooled and desalted using 10 kDa MWCO centrifugal filter device, Microcon YM-10 regenerated cellulose (Millipore, Billerica, MA).

LC-MS/MS analysis of digested glycoproteins

The serum glycoproteins with an affinity for PHA-L were subject to tryptic digestion in the presence of RapiGest (Waters, Milford, MA) which was used as a mild denaturant to help solubilize and unfold the glycoproteins, making them more amenable to cleavage. The proteins were reduced by incubation with 5mM dithiothreitol (DTT) at 50°C for 45 minutes and then alkylated by incubation with 15mM iodoacetamide (IAM) at room temperature in the dark for 45 minutes. Sequencing Grade Modified Trypsin (Promega, Madison, WI) was added to the reduced and carbamidomethylated proteins at approximately a 1:100 ratio (w/w) and incubated at 37°C overnight. The peptide mixture was concentrated on a SpeedVac and brought up in 0.1% formic acid to a final volume of 50 μ L.

A column was prepared from a silica capillary (75 μ m $\times$ 10cm; tip 15 $\pm$ 1 μ m, PicoTip EMITTER, New Objective, Woburn, MA) by packing silica C18 resins (Rainin Microsorb MV, 5 μ m, 300 Å pore size) using a pressurized stainless steel cell and nitrogen gas at 1,000 psi. The trypsin digested sample was loaded directly onto the column for 30 min using the pressurized

cell. The loaded spray column was interfaced directly with a Finnegan LTQ Linear Ion Trap Mass Spectrometer (Thermo Finnegan, San Jose, CA) coupled to a reversed-phase capillary HPLC system. The spray voltage was held at 1.9 kV with a capillary temperature of 220°C. The peptides were eluted into the MS at approximately 400 nanoL/min with a 65-minute linear gradient of 5 to 60% ACN in 0.1% formic acid. The MS was operated in a data-dependent mode in which the top 9 precursor ions during each MS scan were further fragmented and analyzed by MS/MS. The raw data obtained from the LTQ was converted into mzXML format and into peak lists for further investigation. Analyses were repeated in triplicate for each serum sample.

Data evaluation

Two sequence databases were assembled for this analysis, normal and decoy. A representative database (normal) was acquired from the National Center for Biotechnology Information (NCBI) on October 15, 2005 and utilized for the final protein identifications. A decoy database, created by inverting the sequences of the normal database, was used to establish accurate scoring thresholds for protein identification in the normal database by ProValT.

The peak lists constructed from the LTQ raw data were submitted for database searching with Mascot (Matrix Science, Boston, MA) on both the normal and decoy databases. Mascot searches were limited to tryptic peptides allowing up to 2 missed cleavages. Peptide mass tolerance was set to +/- 500ppm and the MS/MS tolerance to 0.6 Daltons, these values corresponded to the instrument used for this analysis. A variable modification of carbamidomethylation on cysteine residues was allowed.

ProValT, a protein validation software developed at the University of Georgia, was used to group non redundant peptides to protein groups with a false-discovery rate (FDR) of $\leq 1\%$ ³⁹.

These proteins were clustered according to homology of the identified peptides. As a result, the protein with the most unique peptides was cataloged as “TOP” and the remainder as “OTHER”. This grouping, by homology, gave a maximum protein list that includes the TOP and OTHER proteins and a minimum proteins list which contained only proteins cataloged as TOP in the homology group. A manual filtering of proteins identified by ProVault was performed to remove keratin and any proteins without an N-linked motif in its amino acid sequence (N-X-S/T). In addition, proteins that did not appear in at least 2 of the 3 triplicate runs or proteins that were only identified by 1 spectrum were not considered in this analysis.

Normalized spectral counts (NSC) were used as a label free method to compare the relative abundance of proteins identified in this proof of principle study. Identification of a peptide requires the generation of a tandem mass spectrum which is called a spectral count. A peptide may be identified and re-identified many times during a MS/MS analysis which in turn causes an accumulation of spectra. The total number of MS/MS spectra associated with an identified protein can be easily obtained from the instrument output. In general, it has been reported that as protein abundance increases, so does the spectral count for that protein⁷⁹⁻⁸¹. Therefore normalized spectral counts (NSC) were used to determine the relative abundance of proteins in each population.

The total number of spectral counts obtained from each technical replicate of a sample was standardized to establish a correction factor. This correction factor was applied to the spectral counts of the individual proteins and averaged to give an individual spectral count ($SC_{\text{individual}}$). The NSC for a protein identified in a sample was calculated from the $SC_{\text{individual}}$ and the number of amino acid residues ($\#AA_{\text{individual}}$) in the identified protein. Additionally, the total spectral count obtained (SC_{total}) and the total number of amino acids residues found in all of the

proteins identified (#AA_{total}) were used in the following equation as previously described by Zybaylov et al ⁴⁰.

$$NSC = \frac{SC_{individual} / \#AA_{individual}}{\Sigma (SC_{Total} / \#AA_{Total})}$$

These calculations were repeated and averaged for each technical triplicate of the nine serum samples. Comparative quantification was performed after normalization. The NSC of pooled male serum was used as a control and was compared to the NSC of the malignant and benign serum samples to detect up regulated glycoproteins. Moreover, evaluation of glycoproteins not identified in the male serum compared the NSC of the benign serum samples to the NSC of the malignant serum samples.

RESULTS AND DISCUSSION

Lectin Selection and Verification

Phaseolus vulgaris Leukoagglutinin (PHA-L), a lectin from the red kidney bean, exhibits attraction for the GlcNAc β 6 branched tri and/or tetra – antennary complex glycans and is not affected by other glycans on the outer branches, i.e. fucose and sialic acids ^{17, 18, 34}. PHA-L is a tetramer comprised of four structurally similar subunits that form a central channel containing the binding site for the complex glycoproteins ¹⁰. High levels of PHA-L reactive oligosaccharides have been observed in several types of malignant tumor cells by lectin histochemistry, as well as, in lysates of malignant tumor cells by lectin blot analysis ^{4, 6, 17, 18, 29-31, 35}. Lectins have been long established as an effective analytical tool for glycoprotein isolation and glycan structure analysis and have lately been applied to proteomic studies in the form of lectin affinity chromatography (LAC) ^{36, 37}.

The selectivity of the LAC was evaluated by analyzing a mixture of standard protein/glycoproteins on the PHA-L column prior to serum analysis. These standards include myoglobin, transferrin, and thyroglobulin. Myoglobin, a 17 kDa protein, has no N-linked sites and transferrin, an 80 kDa protein, has 2 N-linked sites that are bi-antennary were used as negative controls since they should not show any affinity for PHA-L. Thyroglobulin, a 670 kDa protein, contains N-linked sites with β -6 GlcNAc branches served as a positive control. Both the bound and unbound proteins collected from LAC column were analyzed by 1D SDS gel electrophoresis. From this analysis we were able to conclude that PHA-L had a higher affinity for thyroglobulin then myoglobin and transeferrin verifying it as a useful lectin for this proof of principle study. Additionally the conditions used for this analysis showed minimal non-specific binding of the negative controls, myoglobin and transferrin, to the PHA-L column (data not shown for brevity). As a result non-specific binding of proteins in serum during the glycoproteomics analysis should be negligible.

Use of Male Serum

Complex type N-linked glycoproteins with β -6GlcNAc branching does occur naturally. As a result, not all β -6GlcNAc branches are associated with malignancy. In order to determine which glycoproteins are normally present a pooled sample of male serum was used. Male serum was used in this analysis due to its commercial accessibility and because the male population is unaffected by ovarian cancer. Approximately 15-20% of the proteins identified in the male serum for this analysis were not found in any of the malignant or benign samples, shown in Table 4.2.

Reasons for the identification of additional proteins in the male serum could be the result in sample collection methods and the sample itself. Serum is the liquid left after whole blood is allowed to clot either naturally, by addition of clotting agents, and/or centrifugation. The malignant and benign samples were collected and stored by protocols contingent with the Ovarian Cancer Institute, where as, the control was collected and stored by the vender, Sigma, according to their standard protocols. These methods could vary resulting in a difference in proteins identified. This is especially notable with coagulation factor II precursor which is a protein in blood plasma that is the inactive precursor of thrombin, a key clot promoter. Differences in the proteins identified in only the OCI samples as shown in Table 4.3 where approximately 13-18% of the proteins are not found in the male serum. Examples include heparin cofactor, plasminogen, fibulin, fibronectin, and von Willebrand factor. These proteins are involved in the process of blood clotting and could be found as a result of sample collection differences however, it should be noted that cancer patients have evidence of increased intravascular coagulation which could explain the increased number of coagulation proteins matched in the serum of patients with ovarian tumors ^{44, 45}. This would require further study to verify if this is a result of sample collection.

Differences in the male and female proteome should be considered ^{46, 47}. In this analysis gender specific proteins were not identified in either sample set nonetheless differences in protein expression levels need to be contemplated. For example, ceruloplasmin (ferroxidase) is shown to be unregulated in female cerebrospinal fluid (CSF) when compared to male CSF ⁴⁸. While a clear-cut difference in expression levels of ceruloplasmin was not observed in this analysis a difference in protein expression as a result of gender should be considered.

Description of Important Matched Proteins

Human blood serum was depleted and proteins with a high affinity for PHA-L were isolated using LAC. The isolated proteins were digested into peptides and the glycopeptides were left intact. Peptides generated by digestion of serum proteins were analyzed in triplicate by LC-MS/MS on a Finnegan LTQ for peptide identification. Intact glycopeptides were not matched during the database search due to the change in mass associated with their glycosylation attachment. This differs from other glycoproteomic studies where the glycans are typically cleaved from the proteins or peptide with the use of enzymes, such as PNGaseF, or other chemical treatments^{5, 15, 16, 36, 37, 42, 43}. Such treatments add experimental steps that not only increase experiment time but could result in sample loss or manipulation. This could have been problematic given the low abundance of glycoproteins with a β -6 GlcNAc branch, therefore these steps were excluded to avoid sample loss.

The peptides matched using the NCBI human database were grouped into proteins with a $\leq 1\%$ FDR by ProValT and screened for the appropriate consensus sequence for N-linked glycans (N-x-S/T) showing that they have the potential to be glycosylated. This method infers that the proteins matched contain β -6GlcNAc branching based on the affinity the samples had for PHA-L. Even with the selectivity analysis described above the potential for PHA-L to have non specific binding should be considered. Further analysis would be needed in order to conclude that the proteins matched actually contained β -6GlcNAc branching. It should be noted that both malignant tumor cells and cell lysates have shown a higher affinity for PHA-L when compared to their benign counter parts^{4, 6, 17, 18, 29-31, 35}. Therefore identification of the proteins that have a high affinity for PHA-L, whether they truly possess the β -6GlcNAc branch or not, could prove to be useful for potential biomarkers discovery.

Overall, from this analysis 63 unique proteins in 48 protein groups, that have an affinity for PHA-L, were matched in the 9 serum samples. Fifty-three proteins in 43 proteins groups were identified in the serum of patients with malignant tumors, 44 proteins in 34 proteins groups in patients with benign tumors, and 34 proteins in 28 protein groups in pooled male serum. Approximately 40% of all proteins were common between the 3 populations. The overlap of these proteins is shown in Figure 4.2. Common along these populations are the top scoring proteins typically found in proteomic analysis such as alpha-2-macroglobulin precursor, apolipoprotein B precursor, and complement component 3 precursor. A complete breakdown of proteins matched for each of the individual 9 samples is listed in the appendix (Tables A-I)

A total of 11,530 total MS/MS spectra were matched to peptides using MASCOT. The distribution of the number of spectra for each sample is shown in Table 4.4. The variety in the number of spectra for each sample demonstrates a difference in protein recovery. Reasons for this discrepancy could come from a variety of sources including recovery from the patient as well as protein or peptide recovery at any stage in the protocol. Additionally it should be noted that the OCI samples came from individual patients and the male serum from a pooled sample set. This could enrich or add diversity of the proteins identified in the pooled serum versus individual patient samples providing a possible explanation for the much greater number of spectra obtained in the LC-MS/MS analysis from the male serum when compared to the other samples. However, in general it should be noted that proteomes by nature are dynamic and the diversity could be a result of such.

Further investigation of the proteins matched based on the diagnosis of the patients allowed for the identification of proteins that could be potential biomarkers for ovarian cancer. The first group of proteins to consider is those only matched in malignant serum (Table 4.5).

Based on these results 10 proteins were matched in malignant serum only. Complement component 5 and the Fc fragment of IgG binding proteins were found in multiple malignant serum samples while other proteins were identified solely in a single malignant serum sample. The matching of a protein to a single sample should not be overlooked due to the diversity of diagnosis in the serum provided by OCI (Table 4.1A). These proteins could provide information that is diagnosis specific with further investigation.

The 10 proteins were examined using a literature search so that their biological roles could be determined and potential as a biomarker investigated. In order to determine their clinical application a much more in depth analysis of these proteins will need to be preformed.

Complement factor H (FH) is a 150-kDa glycosylated plasma protein that is the main inhibitor of the alternative pathway of the complement cascade activation. FH binds to some cancer cell surfaces and appears to protect the structures from complement attack⁴⁹. Increased expression of FH has been confirmed in cell lines from several types of cancers including cervix, bladder, lung, kidney, and ovary cells⁵⁰⁻⁵². More specifically, ovarian tumors produce FH as shown by the immunohistochemical analysis using antibodies specific for FH, 196X and VIG8⁵⁰. In addition BTA TRAKTM and BTAsatTM are a clinically approved immunoassay for the detection of bladder cancer in urine which is based on the quantification of FH⁵².

Protein S (alpha) is a vitamin K-dependent plasma glycoprotein that exists in two forms; free or bound to complement protein C4b. *In vivo* Protein S forms a complex with Protein C and thrombomodulin to slow down the process of coagulation. Cancer patients have evidence of increased intravascular coagulation and have shown an increase in coagulants^{44, 45}. This was demonstrated in a study of 81 patients with gastrointestinal or pelvic cancer and 58 healthy subjects who were evaluated for free and total Protein S plasma levels. When compared,

the total Protein S levels were significantly higher in the cancer group with no significant difference in the concentrations of free and the functional Protein S⁵³.

Histidine-rich glycoprotein precursor (HRG), also known as histidine-proline-rich glycoprotein, is a relatively abundant plasma protein whose functions include defense, coagulation, and to innate immune responses⁵⁴. HRG is also thought to be involved in angiogenesis. Angiogenesis, the development of new blood vessels, is involved in normal development, wound healing, and may lead to the progression and metastasis of tumors⁵⁵. This glycoprotein was shown by an immunohistochemical study of human breast cancer where HRP to be localized in the tumor matrix⁵⁵

Serum amyloid P component precursor (SAP) is a 25kDa pentameric protein that is considered to be an acute phase protein. Acute phase proteins are non-antibody proteins whose concentrations are increased in serum as a response to inflammation and tissue injury. The function of SAP is still unknown but it is believed that it acts as an opsonising protein and is involved in the removal of apoptotic cells. This conclusion is drawn because SAP, like the other pentraxins, binds to apoptotic cells⁵⁶. Patients with breast cancer show a significant increase in SAP concentrations in serum, and the concentrations of SAP were found to correlated to the severity of the diagnosis of the disease⁵⁷.

Complement component 5 (C5) is a 190 kDa complement protein that consists of two polypeptide chains connected by disulfide bonds. Typically, C5 is cleaved into 2 fragments, C5a and C5b, by C5 convertase which allows the continuation of the complement cascade. Once cleaved C5a acts as the most important inflammatory mediator of the complement system⁴⁹. It has been shown that C5a acts in parallel with the up-regulation of the SAP, a protein also found in the malignant serum of this analysis⁵⁸. C5b initiates the membrane attack complex (MAC)

with C6-9 to aid in the destruction of extracellular pathogens⁴⁹. As a result, C5a and C5b play significant roles in vital immunological pathways.

Fc fragment of IgG binding protein (FcGBP) binds to monomeric and aggregated IgG molecules and it is believed that this receptor is responsible for the movement of antibodies. FcGBP shares some structural features with mucin proteins suggesting that it plays an important role in the mucosal immune system and inflammation⁵⁹. FcGBP potential as a biomarker has been reported recently from a meta-review of thyroid cancer⁶⁰.

Centrosome-associated protein 350 (CAP 350) is a component of cell-cycle control with unknown function. Perturbation of centrosome-proteins has been linked to cell-cycle misregulation and cancer⁶¹. A proteomics study using two-dimensional protein gel electrophoresis and protein sequencing specifically reported an increased expression of CAP 350 in primary glioblastoma cells^{62, 63}. It should be noted that this protein is typically found in the cell cytoplasm and nucleus but where matched to serum proteins in this analysis⁶⁴.

The biological function of hypothetical protein FLJ13646 currently remains unknown⁶⁵⁻⁶⁷. Although its function is unknown it has been shown to be over expressed in patients with colon cancer tumor reoccurrence based on a study of 12 colon cancer patients by microarray analysis where⁶⁷.

A disintegrin and metalloprotease with thrombospondin motifs-8 (ADAMTS 8 or METH-2) is a secreted cartilage-degrading metalloproteases^{68, 69}. ADAMTS proteins are similar to matrix metalloproteinases (MMPs) which have long been associated with cancer-cell invasion and metastasis⁷⁰. The relevance of ADAMTS with cancer progression is unclear but it is expected that it is equally as important as MMPs, as was shown in a recent study where high

expression of ADAMTS8 was associated with poor survival rates of patients with breast cancer^{71, 72}

TAF4b RNA polymerase II, also known as a TATA box binding protein (TBP), assembles with transcription factors to aid in the transcription of protein-encoding genes in eukaryotes^{73, 74}. TBP is typically selected for microarray analysis as an internal reference due to its relative low abundance in normal tissue^{75, 76}. As a result, clinical studies to investigate TBP's role, if any, in cancer has yet to be completed.

Description of Differentially Expressed Proteins

To determine if an up regulation of glycoproteins with an affinity for PHA-L is associated with malignancy an evaluation of the changes in protein abundance between the samples by relative quantification was used. Current methods to quantify proteins are based on isotope or mass tag labeling which require experimental manipulation as a results there is a possibility that low abundance proteins will not react to completion or become lost through the labeling process^{77, 78}. In order to minimize sample handling and to insure proteins of low abundance are not lost due to labeling methods, a label free approach using spectral counts was applied.

The NSC of pooled male serum was compared to the NSC of the malignant and benign serum samples to detect up regulated glycoproteins in the presence of a tumor (not specified as malignant or benign). Previous reports using spectral counts report variations of approximately 2 to 2.5 fold changes as biologically significant^{40, 81, 82}. However, for this analysis we set a more conservative ≥ 3 fold change limit for reporting potentially up regulated proteins. The proteins shown in Table 4.6 showed a 3 fold or greater up regulation using NSC. These proteins showed

an approximately equal up regulation malignant and benign serum. As a result, this up regulation could be an inflammatory response towards a tumor in general not specific for malignant or benign.

The malignant glycoproteins were compared with the benign sample set serving as the control. Again, an up regulation of 3 or greater was considered significant for this analysis. One protein was found to be particularly up regulated in the malignant serum, NP_000604.1 hemopexin. Hemopexin (Hpx) a 60 kD glycoprotein found in plasma is an acute phase reactant that is synthesized after inflammation is induced⁸³. Hpx is a member of the matrix metalloproteinases (MMPs) family that participates in extracellular matrix (ECM) degradation. MMPs have been associated with tumor growth, invasion, metastasis and are frequently overexpressed in malignant tumors⁸⁴. More specifically increased Hpx concentrations have been found in some malignancies, especially malignant melanoma and breast cancer but not in benign breast fibroadenoma^{83, 85, 86}. With these findings hemopexin should be added to the list of potential biomarkers requiring further investigation.

CONCLUSIONS

This proof of principle glycoproteomics study allowed for the identification of 63 proteins in human serum that have an affinity for PHA-L. From this we can infer that these proteins are glycosylated and contain a β 6GlcNAc branch. Additionally, by comparing the matched proteins, based on the diagnosis of the patients, 11 potential biomarkers for ovarian cancer were identified. These proteins will require further investigation on a much larger sample set to determine if they are clinically applicable biomarkers.

While this research was specific for ovarian cancer the methodology used in this study can be applied to other sample sets to identify proteins that have an affinity for PHA-L. This could include serum from patients with different types of cancer or diseases. Also different sample sources such as urine, saliva, or cerebrospinal fluid (CSF), which are compatible with the MARS column and LAC, could be investigated in this glycoproteomics approach using this lectin to isolate glycoproteins with increased branching.

REFERENCES

1. American Cancer Society, I. Detailed Guide: Ovarian Cancer. www.cancer.org
2. National Ovarian Cancer Coalition. www.ovarian.org
3. Bunkenborg, J.; Pilch, B. J.; Podtelejnikov, A. V.; Wisniewski, J. R., Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics* **2004**, 4, (2), 454-65.
4. Dall'olio, F., Protein glycosylation in cancer biology: an overview. *Clin Mol Pathol* **1996**, 49, (3), M126-M135.
5. Wang, Y.; Wu, S.; Hancock, W. S., Approaches to the study of N-linked glycoproteins in human plasma using lectin affinity chromatography and nano-HPLC coupled to electrospray linear ion trap--Fourier transform mass spectrometry. *Glycobiology* **2006**, 16, (6), 514.
6. Taniguchi, N.; Ihara, S.; Saito, T.; Miyoshi, E.; Ikeda, Y.; Honke, K., Implication of GnT-V in cancer metastasis: A glycomic approach for identification of a target protein and its unique function as an angiogenic cofactor. *Glycoconjugate Journal* **2001**, 18, (11), 859-865.
7. Block, T. M.; Comunale, M. A.; Lowman, M.; Steel, L. F.; Romano, P. R.; Fimmel, C.; Tennant, B. C.; London, W. T.; Evans, A. A.; Blumberg, B. S., Use of targeted glycoproteomics to identify serum glycoproteins that correlate with liver cancer in woodchucks and humans. *Proceedings of the National Academy of Sciences* **2005**, 102, (3), 779-784.
8. Orntoft, T. F.; Vestergaard, E. M., Clinical aspects of altered glycosylation of glycoproteins in cancer. *Electrophoresis* **1999**, 20, (2), 362-71.
9. Kobata, A., Altered glycosylation of surface glycoproteins in tumor cells and its clinical application. *Pigment Cell Res* **1989**, 2, (4), 304-8.
10. Hakomori, S., Glycosylation defining cancer malignancy: New wine in an old bottle. In National Acad Sciences: 2002.

11. An, H. J.; Miyamoto, S.; Lancaster, K. S.; Kirmiz, C.; Li, B.; Lam, K. S.; Leiserowitz, G. S.; Lebrilla, C. B., Profiling of Glycans in Serum for the Discovery of Potential Biomarkers for Ovarian Cancer. *Journal of Proteome Research* **2006**.
12. Kim, Y. J.; Varki, A., Perspectives on the significance of altered glycosylation of glycoproteins in cancer. *Glycoconjugate Journal* **1997**, 14, (5), 569-576.
13. Kobata, A., A retrospective and prospective view of glycopathology. *Glycoconjugate Journal* **1998**, 15, (4), 323-331.
14. Ørntoft, T. F.; Bech, E., Circulating blood group related carbohydrate antigens as tumour markers. *Glycoconjugate Journal* **1995**, 12, (3), 200-205.
15. Xiong, L.; Andrews, D.; Regnier, F., Comparative proteomics of glycoproteins based on lectin selection and isotope coding. *J. Proteome Res* **2003**, 2, 618–625.
16. Zhang, H.; Li, X.; Martin, D. B.; Aebersold, R., Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nature Biotechnology* **2003**, 21, (6), 660-666.
17. Fernandes, B., Beta 1-6 branched oligosaccharides as a marker of tumor progression in human breast and colon neoplasia. *Cancer Research* **1991**, 51, (2), 718-723.
18. Seelentag, W. K., Prognostic value of beta1, 6-branched oligosaccharides in human colorectal carcinoma. *Cancer Research* **1998**, 58, (23), 5559-5564.
19. Przybylo, M.; Hoja-Lukowicz, D.; Litynska, A.; Laidler, P., Different glycosylation of cadherins from human bladder non-malignant and cancer cell lines. *feedback* **2004**.
20. Meezan, E.; Wu, H. C.; Black, P. H.; Robbins, P. W., Comparative Studies on the Carbohydrate-containing membrane components of normal and virus-transformed mouse fibroblasts. II. Separation of glycoproteins and glycopeptides by Sephadex chromatography. *Biochemistry* **1969**, 8, (6), 2518-2524.
21. Pierce, M.; Buckhaults, P.; Chen, L.; Fregien, N., Regulation of N-acetylglucosaminyltransferase V and Asn-linked oligosaccharide β (1, 6) branching by a growth factor signaling pathway and effects on cell adhesion and metastatic potential. *Glycoconjugate Journal* **1997**, 14, (5), 623-630.

22. Buck, C. A.; Fuhrer, J. P.; Soslau, G.; Warren, L., Membrane Glycopeptides from Subcellular Fractions of Control and Virus-transformed Cells. *Journal of Biological Chemistry* **1974**, 249, (5), 1541-1550.
23. Warren, L.; Cossut, G., The bound carbohydrates of glycoproteins in normal and pathological states. *Ann NY Acad Sci* **1982**, 401, 85-91.
24. Yamashita, K.; Ohkura, T.; Tachibana, Y.; Takasaki, S.; Kobata, A., Comparative study of the oligosaccharides released from baby hamster kidney cells and their polyoma transformant by hydrazinolysis. *Journal of Biological Chemistry* **1984**, 259, (17), 10834-10840.
25. Ogata, S.; Muramatsu, T.; Kobata, A., New structural characteristic of the large glycopeptides from transformed cells. *Nature* **1976**, 259, (5544), 580-582.
26. Taniguchi, N.; Miyoshi, E.; Ko, J. H.; Ikeda, Y.; Ihara, Y., Implication of N-acetylglucosaminyltransferases III and V in cancer: gene regulation and signaling mechanism. *Biochim Biophys Acta* **1999**, 1455, (2-3), 287-300.
27. Saito, T.; Miyoshi, E.; Sasai, K.; Nakano, N.; Eguchi, H.; Honke, K.; Taniguchi, N., A Secreted Type of β 1, 6-N-Acetylglucosaminyltransferase V (GnT-V) Induces Tumor Angiogenesis without Mediation of Glycosylation A Novel Function of GnT-V Distinct from the Original Glycosyltransferase Activity. *Journal of Biological Chemistry* **2002**, 277, (19), 17002-17008.
28. Ito, Y.; Miyoshi, E.; Sakon, M.; Takeda, T.; Noda, K.; Tsujimoto, M.; Ito, S.; Honda, H.; Takemura, F.; Wakasa, K., Elevated expression of UDP-N-acetylglucosamine: α -mannoside β 1, 6 N-acetylglucosaminyltransferase is an early event in hepatocarcinogenesis. *Int J Cancer* **2001**, 91, (5), 631-7.
29. Ihara, S.; Miyoshi, E.; Ko, J. H.; Murata, K.; Nakahara, S.; Honke, K.; Dickson, R. B.; Lin, C. Y.; Taniguchi, N., Prometastatic Effect of N-Acetylglucosaminyltransferase V Is Due to Modification and Stabilization of Active Matrilysin by Adding β 1-6 GlcNAc Branching. *Journal of Biological Chemistry* **2002**, 277, (19), 16960-16967.
30. Dennis, J. W.; Granovsky, M.; Warren, C. E., Glycoprotein glycosylation and cancer progression. *Biochim Biophys Acta* **1999**, 1473, (1), 21-34.
31. Schwarz, R. E.; Wojciechowicz, D. C.; Park, P. Y.; Paty, P. B., Phytohemagglutinin-L (PHA-L) lectin surface binding of N-linked β 1-6 carbohydrate and its relationship to

- activated mutant ras in human pancreatic cancer cell lines. *Cancer Lett* **1996**, 107, (2), 285-91.
32. Adkins, J. N.; Varnum, S. M.; Auberry, K. J.; Moore, R. J.; Angell, N. H.; Smith, R. D.; Springer, D. L.; Pounds, J. G., Toward a Human Blood Serum Proteome Analysis By Multidimensional Separation Coupled With Mass Spectrometry* S. *Molecular & Cellular Proteomics* **2002**, 1, (12), 947-955.
 33. Echan, L. A.; Tang, H. Y.; Ali-Khan, N.; Lee, K.; Speicher, D. W., Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. *Proteomics* **2005**, 5, (13), 3292-303.
 34. Merkle, R. K., Richard D. Cummings, Lectin Affinity Chromatography of Glycopeptides. In *Analytical Methods*, 1987; pp 232-259.
 35. Dosaka-Akita, H.; Miyoshi, E.; Suzuki, O.; Itoh, T.; Katoh, H.; Taniguchi, N., Expression of N-Acetylglucosaminyltransferase V Is Associated with Prognosis and Histology in Non-Small Cell Lung Cancers. In AACR: 2004.
 36. Qiu, R.; Regnier, F. E., Use of multidimensional lectin affinity chromatography in differential glycoproteomics. *Anal Chem* **2005**, 77, (9), 2802-2809.
 37. Xiong, L.; Regnier, F. E., Use of a lectin affinity selector in the search for unusual glycosylation in proteomics. *Journal of Chromatography B* **2002**, 782, (1-2), 405-418.
 38. Cummings, R. D., Guide to Techniques in Glycobiology. In 1994; Vol. 230, pp 66-88.
 39. Weatherly, D. B.; Atwood, J. A.; Minning, T. A.; Cavola, C.; Tarleton, R. L.; Orlando, R., A Heuristic Method for Assigning a False-discovery Rate for Protein Identifications from Mascot Database Search Results*. *Molecular & Cellular Proteomics* **2005**, 4, (6), 762-772.
 40. Zybaylov, B.; Mosley, A. L.; Sardi, M. E.; Coleman, M. K.; Florens, L.; Washburn, M. P., Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J Proteome Res* **2006**, 5, 2339-47.
 41. Yocum, A. K.; Yu, K.; Oe, T.; Blair, I. A., Effect of immunoaffinity depletion of human serum during proteomic investigations. *J. Proteome Res* **2005**, 4, (5), 1722-1731.

42. Kaji, H.; Saito, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Hirabayashi, J.; Kasai, K.; Takahashi, N.; Isobe, T., Proteomics: the first decade and beyond. *Nature Biotechnology* **2003**, 21, 667-672.
43. Morelle, W.; Flahaut, C.; Michalski, J. C.; Louvet, A.; Mathurin, P.; Klein, A., Mass spectrometric approach for screening modifications of total serum N-glycome in human diseases: application to cirrhosis. *Glycobiology* **2006**, 16, (4), 281.
44. Green, D.; Maliekal, K.; Sushko, E.; Akhtar, R.; Soff, G. A., Activated-protein-C resistance in cancer patients. *Haemostasis* **1997**, 27, (3), 112-8.
45. Zecchina, G.; Ghio, P.; Bosio, S.; Cravino, M.; Camaschella, C.; Scagliotti, G. V., Reactive thrombocytosis might contribute to chemotherapy-related thrombophilia in patients with lung cancer. *Clin Lung Cancer* **2007**, 8, (4), 264-7.
46. Zhan, X.; Desiderio, D. M., Heterogeneity Analysis of the Human Pituitary Proteome. In *Am Assoc Clin Chem*: 2003.
47. He, P.; He, H. Z.; Dai, J.; Wang, Y.; Sheng, Q. H.; Zhou, L. P.; Zhang, Z. S.; Sun, Y. L.; Liu, F.; Wang, K., The human plasma proteome: Analysis of Chinese serum using shotgun strategy. *Proteomics* **2005**, 5, (13), 3442–3453.
48. Ogata, Y.; Muddiman, D. C.; Higgins, L. A.; Lomas, L.; Keegan, M. B.; Vernino, S. A.; Bergen Iii, H. R.; Ligand, H., Differential Protein Expression Between Male and Female Human Lumbar Cerebrospinal Fluid after Abundant Protein Depletion. *Brain* 20, 80.
49. Gelderman, K. A.; Tomlinson, S.; Ross, G. D.; Gorter, A., Complement function in mAb-mediated cancer immunotherapy. *Trends Immunol* **2004**, 25, (3), 158-164.
50. Junnikkala, S.; Hakulinen, J.; Jarva, H.; Manuelian, T.; Bjørge, L.; Bützow, R.; Zipfel, P. F.; Meri, S., Molecular and Cellular Pathology. *British Journal of Cancer* **2002**, 87, 1119-1127.
51. Ajona, D.; Castano, Z.; Garayoa, M.; Zudaire, E.; Pajares, M. J.; Martinez, A.; Cuttitta, F.; Montuenga, L. M.; Pio, R., Expression of Complement Factor H by Lung Cancer Cells Effects on the Activation of the Alternative Pathway of Complement. In *AACR*: 2004.

52. Cheng, Z. Z.; Corey, M. J.; Parepalo, M.; Majno, S.; Hellwage, J.; Zipfel, P. F.; Kinders, R. J.; Raitanen, M.; Meri, S.; Jokiranta, T. S., Complement Factor H as a Marker for Detection of Bladder Cancer. *Clinical Chemistry* **2005**, 51, (5), 856-863.
53. Battistelli, S.; Vittoria, A.; Cappelli, R.; Stefanoni, M.; Roviello, F., Protein S in cancer patients with non-metastatic solid tumours. *Eur J Surg Oncol* **2005**, 31, (7), 798-802.
54. Borza, D. B., Life without histidine-rich glycoprotein: modulation of the hemostatic balance revisited. *Journal of Thrombosis and Haemostasis* **2005**, 3, (5), 863-864.
55. Simantov, R.; Febbraio, M.; Crombie, R.; Asch, A. S.; Nachman, R. L.; Silverstein, R. L., Histidine-rich glycoprotein inhibits the antiangiogenic effect of thrombospondin-1. *Journal of Clinical Investigation* **2001**, 107, (1), 45-52.
56. Bijl, M.; Bootsma, H.; van der Geld, Y.; Limburg, P. C.; Kallenberg, C. G. M.; van Rijswijk, M. H., Serum amyloid P component levels are not decreased in patients with systemic lupus erythematosus and do not rise during an acute phase reaction. In *BMJ*: 2004.
57. Levo, Y.; Wollner, S.; Treves, A. J., Serum amyloid P-component levels in patients with malignancy. *Scand. J. Immunol* **1986**, 24, 151.
58. Szalai, A. J.; van Ginkel, F. W.; Wang, Y.; McGhee, J. R.; Volanakis, J. E., Complement-Dependent Acute-Phase Expression of C-Reactive Protein and Serum Amyloid P-Component 1. *The Journal of Immunology* **2000**, 165, (2), 1030-1035.
59. Harada, N.; Iijima, S.; Kobayashi, K.; Yoshida, T.; Brown, W. R.; Hibi, T.; Oshima, A.; Morikawa, M., Human IgGFc Binding Protein (Fc γ BP) in Colonic Epithelial Cells Exhibits Mucin-like Structure. *Journal of Biological Chemistry* **1997**, 272, (24), 15232-15241.
60. Griffith, O. L.; Melck, A.; Jones, S. J. M.; Wiseman, S. M., Meta-Analysis and Meta-Review of Thyroid Cancer Gene Expression Profiling Studies Identifies Important Diagnostic Biomarkers. *Journal of Clinical Oncology* **2006**, 24, (31), 5043.
61. Nigg, E. A., Centrosome aberrations: cause or consequence of cancer progression? *Nature Reviews Cancer* **2002**, 2, (11), 815-825.

62. Ohgaki, H.; Kleihues, P., Genetic Pathways to Primary and Secondary Glioblastoma. *American Journal of Pathology* **2007**, 170, (5), 1445.
63. Furuta, M.; Weil, R. J.; Vortmeyer, A. O.; Huang, S.; Lei, J.; Huang, T. N.; Lee, Y. S.; Bhowmick, D. A.; Lubensky, I. A.; Oldfield, E. H., Protein patterns and proteins that identify subtypes of glioblastoma multiforme. *Oncogene* **2004**, 23, 6806-6814.
64. Patel, H.; Truant, R.; Rachubinski, R. A.; Capone, J. P., Activity and subcellular compartmentalization of peroxisome proliferator-activated receptor α are altered by the centrosome-associated protein CAP350. *Journal of Cell Science* **2005**, 118, (1), 175-186.
65. Hardwidge, P. R.; Rodriguez-Escudero, I.; Goode, D.; Donohoe, S.; Eng, J.; Goodlett, D. R.; Aebersold, R.; Finlay, B. B., Proteomic Analysis of the Intestinal Epithelial Cell Response to Enteropathogenic Escherichia coli. *J Biol Chem* **2004**, 279, (19), 20127-20136.
66. Fox, M. S.; Ares, V. X.; Turek, P. J.; Haqq, C.; Reijo Pera, R. A., Feasibility of global gene expression analysis in testicular biopsies from infertile men. *Molecular Reproduction and Development* **2003**, 66, (4), 403-421.
67. Barrier, A.; Boelle, P. Y.; Lemoine, A.; Tse, C.; Brault, D.; Chiappini, F.; Lacaine, F.; Houry, S.; Huguier, M.; Flahault, A., Gene Expression Profiling of Nonneoplastic Mucosa May Predict Clinical Outcome of Colon Cancer Patients. *Diseases of the Colon & Rectum* **2005**, 48, (12), 2238-2248.
68. Collins-Racie, L. A.; Flannery, C. R.; Zeng, W.; Corcoran, C.; Annis-Freeman, B.; Agostino, M. J.; Arai, M.; DiBlasio-Smith, E.; Dorner, A. J.; Georgiadis, K. E., ADAMTS-8 exhibits aggrecanase activity and is expressed in human articular cartilage. *Matrix Biol* **2004**, 23, (4), 219-30.
69. Overall, C. M.; Lopez-Otin, C., Strategies for mmp inhibition in cancer: innovations for the post-trial era. *Nature Reviews Cancer* **2002**, 2, (9), 657-672.
70. Egeblad, M.; Werb, Z., New Functions for the Matrix Metalloproteinases in Cancer Progression. *Nature Reviews Cancer* **2002**, 2, (3), 163-176. all caps title
71. Porter, S.; Span, P. N.; Sweep, F.; Tjan-Heijnen, V. C. G.; Pennington, C. J.; Pedersen, T. X.; Johnsen, M.; Lund, L. R.; Romer, J.; Edwards, D. R., ADAMTS8 and ADAMTS15 expression predicts survival in human breast carcinoma. *International Journal of Cancer* **2006**, 118, (5), 1241.

72. Int, J., Gene Expression Tied to Breast Cancer Outlook. *Int J Cancer* **2006**, 118, 1241-1247.
73. Shao, H.; Revach, M.; Moshonov, S.; Tzuman, Y.; Gazit, K.; Albeck, S.; Unger, T.; Dikstein, R., Core Promoter Binding by Histone-Like TAF Complexes. *Molecular and Cellular Biology* **2004**, 25, (1), 206-219.
74. Hernandez, N., TBP, a universal eukaryotic transcription factor? *Genes Dev* **1993**, 7, (7B), 1291-308.
75. Di Nicolantonio, F.; Mercer, S. J.; Knight, L. A.; Gabriel, F. G.; Whitehouse, P. A.; Sharma, S.; Fernando, A.; Glaysher, S.; Di Palma, S.; Johnson, P., Cancer cell adaptation to chemotherapy. *Feedback* **2006**.
76. Helenius, M. A.; Saramaki, O. R.; Linja, M. J.; Tammela, T. L. J.; Visakorpi, T., Amplification of Urokinase Gene in Prostate Cancer 1. In AACR: 2001.
77. Elrick, M. M.; Walgren, J. L.; Mitchell, M. D.; Thompson, D. C., Proteomics: Recent Applications and New Technologies. *Basic and Clinical Pharmacology and Toxicology* **2006**, 98, (5), 432.
78. Resing, K. A.; Ahn, N. G., Proteomics strategies for protein identification. *FEBS Lett* **2005**, 579, (4), 885–9.
79. Bonner, A. J.; Liu, H., Towards the prediction of protein abundance from tandem mass spectrometry data. *Proceedings of the SIAM International Conference on Data Mining* **2006**, 6, 599–603.
80. Liu, H.; Sadygov, R. G.; Yates Iii, J. R., A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem* **2004**, 76, (14), 4193–4201.
81. 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* S. *Molecular & Cellular Proteomics* **2005**, 4, (10), 1487-1502.

82. Hendrickson, E. L.; Xia, Q.; Wang, T.; Leigh, J. A.; Hackett, M., Comparison of spectral counting and metabolic stable isotope labeling for use with quantitative microbial proteomics. *Analyst* **2006**, 131, 1335-1341.
83. Delanghe, J. R.; Langlois, M. R., Hemopexin: a review of biological aspects and the role in laboratory medicine. *Clin Chim Acta* **2001**, 312, (1-2), 13-23.
84. Hidalgo, M.; Eckhardt, S. G., Development of Matrix Metalloproteinase Inhibitors in Cancer Therapy. *JNCI* **2001**, 93, (3), 178-193.
85. Manuel, Y.; Defontaine, M. C.; Bourgoin, J. J.; Dargent, M.; Sonneck, J. M., Serum haemopexin levels in patients with malignant melanoma. *Clin Chim Acta* **1971**, 31, (2), 485-6.
86. Coombes, R. C.; Powles, T. J.; Gazet, J. C.; Ford, H. T.; Sloane, J. P.; Laurence, D. J.; Neville, A. M., Biochemical markers in human breast cancer. *Lancet* **1977**, 1, (8003), 132-4.

FIGURE 4.1:

A figure of a complex type N-linked glycan with the $\beta 6$ GlcNA branch circled. This type of branching can be found in a 2,2,6 tri-antennary and a 2,2,4,6 tetra-antennary N-linked glycan.

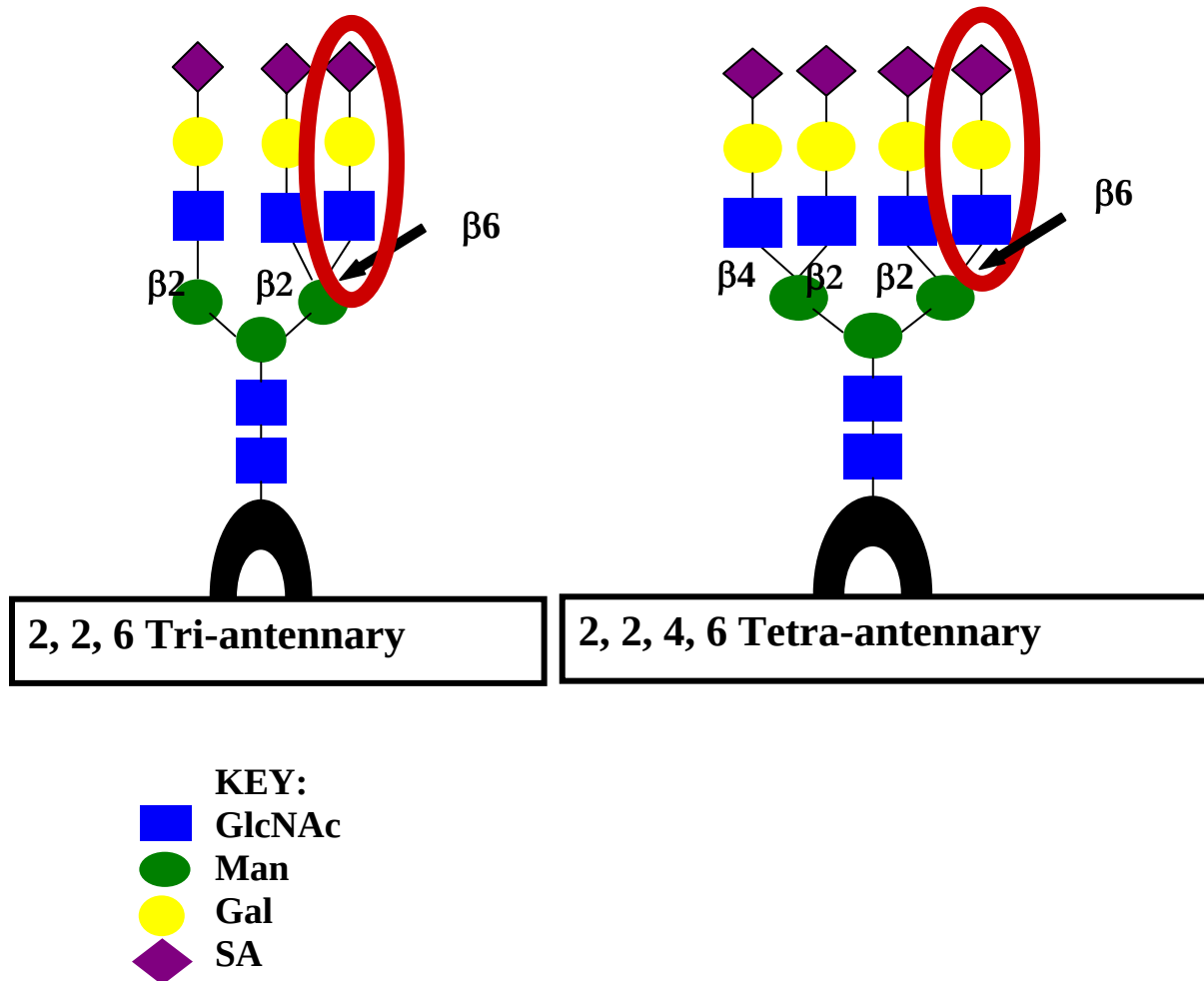


TABLE 4.1a:

The clinical diagnosis of the malignant serum samples obtained from the Ovarian Cancer Institute.

C1	Poorly differentiated adenocarcinoma, features of serous papillary Stage IIIa, grade 3
C2	Poorly differentiated papillary serous adenocarcinoma; likely primary fallopian Stage III, grade 3
C3	High grade endometrioid adenocarcinoma Stage Ic, grade high
C4	Poorly differentiated serous papillary carcinoma Stage IIIC, grade 3

TABLE 4.1b:

The clinical diagnosis of the benign serum samples obtained from the Ovarian Cancer Institute

A1	Multiloculated serous cystadenoma; cystadenofibroma
A2	Benign serous cystadenofibroma
A3	Mucinous cystadenoma
A4	Mucinous cystadenoma

TABLE 4.2:

Proteins matched only in sample of pooled male serum. There are 4 unique proteins identified as TOP proteins and an additional 3 labeled as OTHER to give a maximum of 7 proteins identified only in the pooled male serum sample set.

<u>Sample</u>	<u>TOP/OTHER</u>	<u>Accession #</u>	<u>Protein Description</u>
SS	TOP	NP_000479.1	serine (or cysteine) proteinase inhibitor, cladeC
SS	TOP	NP_000497.1	coagulation factor II precursor
SS	TOP	NP_000599.1	orosomucoid 2
SS	TOP	NP_002208.1	inter-alpha (globulin) inhibitor H3
SS	OTHER	NP_002420.1	matrix metalloproteinase 19 isoform rasi-1, preproprotein
SS	OTHER	NP_073629.1	matrix metalloproteinase 19 isoform rasi-9
SS	OTHER	XP_370616.3	similar to hypothetical protein

TABLE 4.3:

Proteins matched only in the malignant and benign samples. There are 10 unique proteins identified as TOP proteins and an additional 7 labeled as OTHER to give a maximum of 17 proteins identified only in the malignant and benign sample sets.

<u>Sample</u>	<u>TOP/OTHER</u>	<u>Accession #</u>	<u>Protein Description</u>
A3-4/C3	TOP	NP_000176.1	heparin cofactor II
A3-4/C3-4	TOP	NP_000292.1	Plasminogen
A2-4/C3-4	TOP	NP_000604.1	hemopexin
A2,4/C4	TOP	NP_001701.1	complement factor B preproprotein
A3/C3	TOP	NP_001987.2	fibulin 1 isoform C precursor
			fibronectin 1 isoform 3
A2-4/C2-4	TOP	NP_002017.1	preproprotein
A3/C1,3	TOP	NP_002046.1	glial fibrillary acidic protein
A3/C3	TOP	NP_006477.2	fibulin 1 isoform D
A3-4/C1,3	TOP	XP_291977.3	similar to Olfactory receptor 52L1
A3/C3	TOP	NP_000543.1	von Willebrand factor precursor
A3/C3	OTHER	NP_006476.2	fibulin 1 isoform B precursor
A3/C3	OTHER	NP_006478.2	fibulin 1 isoform A precursor
			fibronectin 1 isoform 6
A2-3/C2-4	OTHER	NP_997639.1	preproprotein
			fibronectin 1 isoform 2
A2-3/C2-4	OTHER	NP_997640.1	preproprotein
			fibronectin 1 isoform 5
A2-3/C2-4	OTHER	NP_997641.1	preproprotein
			fibronectin 1 isoform 4
A2-3/C2-4	OTHER	NP_997643.1	preproprotein
			fibronectin 1 isoform 1
A2-3/C2-4	OTHER	NP_997647.1	preproprotein

FIGURE 4.2:

Distribution by patients' diagnosis of the 63 maximum unique proteins matched after triplicate analysis and stringent filtering of results.

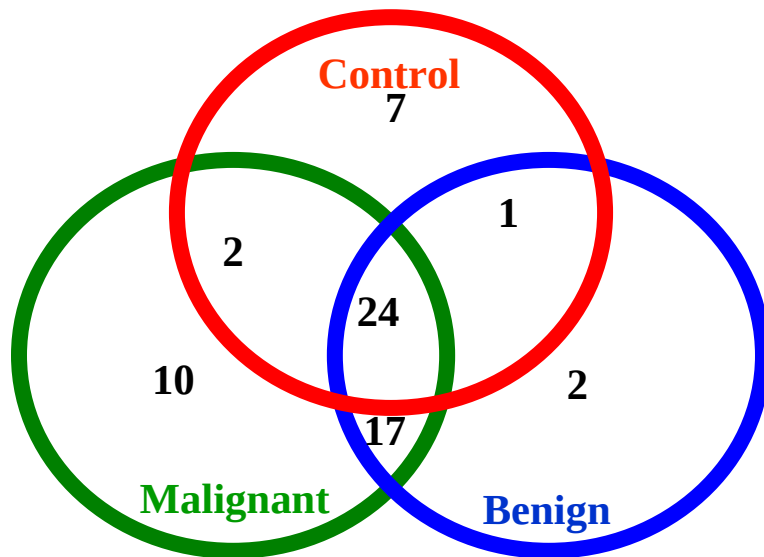


TABLE 4.5:

Distribution of the total number of MS/MS spectra matched to peptides using MASCOT for each sample.

<u>Sample</u>	<u>Total # of Spectra</u>
A1	625
A2	822
A3	1589
A4	1079
C1	394
C2	342
C3	1648
C4	1389
SS	3642

TABLE 4.5:

Proteins matched only in the malignant samples. There are 8 unique proteins identified as TOP proteins and an additional 2 labeled as OTHER to give a maximum of 10 proteins identified only in the malignant sample set.

<u>Sample</u>	<u>TOP/OTHER</u>	<u>Accession #</u>	<u>Protein Description</u>
C4	TOP	NP_000177.1	complement factor H
C3	TOP	NP_000304.1	protein S (alpha)
C4	TOP	NP_000403.1	histidine-rich glycoprotein precursor
C1	TOP	NP_001630.1	serum amyloid P component precursor
C3-4	TOP	NP_001726.2	complement component 5
C1-3	TOP	NP_003881.1	Fc fragment of IgG binding protein
C4	TOP	NP_055625.2	centrosome-associated protein 350
C4	TOP	NP_078860.2	hypothetical protein FLJ13646
C3	OTHER	NP_008968.2	a disintegrin & metalloprotease w/ thrombospondin motifs-8
C3	OTHER	XP_290809.3	TAF4b RNA polymerase II

TABLE 4.6:

Proteins matched in malignant and benign sample sets that are up regulated ≥ 3 fold when NSC are compared to the NSC of the control set

<u>Sample</u>	<u>TOP/OTHER</u>	<u>Accession #</u>	<u>Protein Description</u>
C1-4/A1-4/SS	TOP	NP_001076.1	alpha-1-antichymotrypsin
C1/A1,4/SS	TOP	NP_000020.1	angiotensinogen precursor
C3/A3-4/SS	TOP	NP_000053.1	complement component 1 inhibitor precursor
C1-3/A1-4/SS	TOP	NP_005558.1	galectin 3 binding protein
C4/A4/SS	TOP	NP_000574.2	vitamin D-binding protein precursor

Supplementary Appendix

The distribution of the individual samples (Tables A- I)

TABLE A: *Proteins Matched in A1 (12 minimum/ 13 maximum)*

TOP	NP_000005.1	alpha-2-macroglobulin precursor
TOP	NP_000020.1	angiotensinogen precursor
TOP	NP_000055.1	complement component 3 precursor
TOP	NP_000087.1	ceruloplasmin (ferroxidase)
TOP	NP_000375.1	apolipoprotein B precursor
TOP	NP_000437.3	paraoxonase 1
TOP	NP_000583.1	complement component 4B proprotein
TOP	NP_001076.1	alpha-1-antichymotrypsin, precursor
TOP	NP_001822.2	clusterin isoform 1
TOP	NP_002207.1	inter-alpha (globulin) inhibitor H2
TOP	NP_002855.1	pregnancy-zone protein
TOP	NP_005558.1	galectin 3 binding protein
OTHER	NP_976084.1	clusterin isoform 2

TABLE B: *Proteins Matched in A2 (17 minimum/ 26 maximum)*

TOP	NP_000005.1	alpha-2-macroglobulin precursor
TOP	NP_000055.1	complement component 3 precursor
TOP	NP_000087.1	ceruloplasmin (ferroxidase)
TOP	NP_000375.1	apolipoprotein B precursor
TOP	NP_000437.3	paraoxonase 1
TOP	NP_000583.1	complement component 4B proprotein
TOP	NP_000604.1	hemopexin
TOP	NP_000629.2	vitronectin precursor
TOP	NP_000884.1	kininogen 1
TOP	NP_001076.1	alpha-1-antichymotrypsin, precursor
TOP	NP_001701.1	complement factor B preproprotein
TOP	NP_001822.2	clusterin isoform 1
TOP	NP_002017.1	fibronectin 1 isoform 3 preproprotein
TOP	NP_002207.1	inter-alpha (globulin) inhibitor H2
TOP	NP_003656.2	ficolin 3 isoform 1 precursor
TOP	NP_005558.1	galectin 3 binding protein
TOP	NP_570602.2	alpha 1B-glycoprotein
OTHER	NP_001002029.1	complement component 4B preproprotein
OTHER	NP_002855.1	pregnancy-zone protein
OTHER	NP_775628.1	ficolin 3 isoform 2 precursor
OTHER	NP_976084.1	clusterin isoform 2
OTHER	NP_997639.1	fibronectin 1 isoform 6 preproprotein
OTHER	NP_997640.1	fibronectin 1 isoform 2 preproprotein
OTHER	NP_997641.1	fibronectin 1 isoform 5 preproprotein
OTHER	NP_997643.1	fibronectin 1 isoform 4 preproprotein
OTHER	NP_997647.1	fibronectin 1 isoform 1 preproprotein

TABLE C: *Proteins Matched in A3 (24 minimum/ 34 maximum)*

TOP	NP_000005.1	alpha-2-macroglobulin precursor
TOP	NP_000053.1	complement component 1 inhibitor precursor
TOP	NP_000055.1	complement component 3 precursor
TOP	NP_000176.1	heparin cofactor II
TOP	NP_000286.2	serine (or cysteine) proteinase inhibitor, clade A
TOP	NP_000292.1	plasminogen
TOP	NP_000375.1	apolipoprotein B precursor
TOP	NP_000437.3	paraoxonase 1
TOP	NP_000583.1	complement component 4B proprotein
TOP	NP_000598.1	orosomucoid 1 precursor
TOP	NP_000604.1	hemopexin
TOP	NP_000629.2	vitronectin precursor
TOP	NP_000706.1	complement component 4 binding protein, alpha
TOP	NP_000884.1	kininogen 1;
TOP	NP_001076.1	alpha-1-antichymotrypsin, precursor
TOP	NP_001822.2	clusterin isoform 1
TOP	NP_001987.2	fibulin 1 isoform C precursor
TOP	NP_002017.1	fibronectin 1 isoform 3 preproprotein
TOP	NP_002046.1	glial fibrillary acidic protein
TOP	NP_002855.1	pregnancy-zone protein
TOP	NP_005558.1	galectin 3 binding protein
TOP	NP_006477.2	fibulin 1 isoform D
TOP	NP_079474.2	transcription termination factor-like protein
TOP	XP_291977.3	similar to Olfactory receptor 52L1
OTHER	NP_000543.1	von Willebrand factor precursor
OTHER	NP_001002029.1	complement component 4B preproprotein
OTHER	NP_006476.2	fibulin 1 isoform B precursor
OTHER	NP_006478.2	fibulin 1 isoform A precursor
OTHER	NP_976084.1	clusterin isoform 2
OTHER	NP_997639.1	fibronectin 1 isoform 6 preproprotein
OTHER	NP_997640.1	fibronectin 1 isoform 2 preproprotein
OTHER	NP_997641.1	fibronectin 1 isoform 5 preproprotein
OTHER	NP_997643.1	fibronectin 1 isoform 4 preproprotein
OTHER	NP_997647.1	fibronectin 1 isoform 1 preproprotein

TABLE D: *Proteins Matched in A4 (20 minimum/25 maximum)*

TOP	NP_000005.1	alpha-2-macroglobulin precursor
TOP	NP_000020.1	angiotensinogen precursor
TOP	NP_000053.1	complement component 1 inhibitor precursor
TOP	NP_000055.1	complement component 3 precursor
TOP	NP_000087.1	ceruloplasmin (ferroxidase)
TOP	NP_000176.1	heparin cofactor II
TOP	NP_000292.1	plasminogen
TOP	NP_000375.1	apolipoprotein B precursor
TOP	NP_000437.3	paraoxonase 1
TOP	NP_000574.2	vitamin D-binding protein precursor
TOP	NP_000583.1	complement component 4B proprotein
TOP	NP_000604.1	hemopexin
TOP	NP_001076.1	alpha-1-antichymotrypsin, precursor
TOP	NP_001701.1	complement factor B preproprotein
TOP	NP_001822.2	clusterin isoform 1
TOP	NP_002017.1	fibronectin 1 isoform 3 preproprotein
TOP	NP_002206.1	inter-alpha (globulin) inhibitor H1
TOP	NP_002207.1	inter-alpha (globulin) inhibitor H2
TOP	NP_570602.2	alpha 1B-glycoprotein
TOP	XP_209550.4	similar to Carboxypeptidase N 83 kDa chain
OTHER	NP_001002029.1	complement component 4B preproprotein
OTHER	NP_002855.1	pregnancy-zone protein
OTHER	NP_976084.1	clusterin isoform 2
		serine (or cysteine) proteinase inhibitor, clade A
OTHER	NP_997261.1	
OTHER	XP_291977.3	similar to Olfactory receptor 52L1

TABLE E: *Proteins Matched in C1 (11 minimum/ 15 maximum)*

TOP	NP_000005.1	alpha-2-macroglobulin precursor
TOP	NP_000020.1	angiotensinogen precursor
TOP	NP_000055.1	complement component 3 precursor
TOP	NP_000087.1	ceruloplasmin (ferroxidase)
TOP	NP_000375.1	apolipoprotein B precursor
TOP	NP_000583.1	complement component 4B proprotein
TOP	NP_001076.1	alpha-1-antichymotrypsin, precursor
TOP	NP_001630.1	serum amyloid P component precursor
TOP	NP_001822.2	clusterin isoform 1
TOP	NP_003881.1	Fc fragment of IgG binding protein
TOP	NP_005558.1	galectin 3 binding protein
OTHER	NP_001002029.1	complement component 4B preproprotein
OTHER	NP_002046.1	glial fibrillary acidic protein
OTHER	NP_976084.1	clusterin isoform 2
OTHER	XP_291977.3	similar to Olfactory receptor 52L1

TABLE F: *Proteins Matched in C2 (11 minimum/ 17 maximum)*

TOP	NP_000005.1	alpha-2-macroglobulin precursor
TOP	NP_000055.1	complement component 3 precursor
TOP	NP_000087.1	ceruloplasmin (ferroxidase)
TOP	NP_000375.1	apolipoprotein B precursor
TOP	NP_000437.3	paraoxonase 1
TOP	NP_000583.1	complement component 4B proprotein
TOP	NP_000884.1	kininogen 1
TOP	NP_001076.1	alpha-1-antichymotrypsin, precursor
TOP	NP_002017.1	fibronectin 1 isoform 3 preproprotein
TOP	NP_003881.1	Fc fragment of IgG binding protein
TOP	NP_005558.1	galectin 3 binding protein
OTHER	NP_001002029.1	complement component 4B preproprotein
OTHER	NP_997639.1	fibronectin 1 isoform 6 preproprotein
OTHER	NP_997640.1	fibronectin 1 isoform 2 preproprotein
OTHER	NP_997641.1	fibronectin 1 isoform 5 preproprotein
OTHER	NP_997643.1	fibronectin 1 isoform 4 preproprotein
OTHER	NP_997647.1	fibronectin 1 isoform 1 preproprotein

TABLE G: *Proteins Matched in C3 (31 minimum/ 34 maximum)*

TOP	NP_000005.1	alpha-2-macroglobulin precursor complement component 1 inhibitor
TOP	NP_000053.1	precursor
TOP	NP_000055.1	complement component 3 precursor
TOP	NP_000087.1	ceruloplasmin (ferroxidase)
TOP	NP_000176.1	heparin cofactor II serine (or cysteine) proteinase inhibitor,
TOP	NP_000286.2	clade A
TOP	NP_000292.1	plasminogen
TOP	NP_000304.1	protein S (alpha)
TOP	NP_000375.1	apolipoprotein B precursor
TOP	NP_000437.3	paraoxonase 1
TOP	NP_000543.1	von Willebrand factor precursor
TOP	NP_000583.1	complement component 4B proprotein
TOP	NP_000598.1	orosomucoid 1 precursor
TOP	NP_000604.1	hemopexin
TOP	NP_000629.2	vitronectin precursor complement component 4 binding protein,
TOP	NP_000706.1	alpha
TOP	NP_001076.1	alpha-1-antichymotrypsin, precursor
TOP	NP_001613.1	alpha-2-HS-glycoprotein
TOP	NP_001726.2	complement component 5
TOP	NP_001822.2	clusterin isoform 1
TOP	NP_001987.2	fibulin 1 isoform C precursor
TOP	NP_002017.1	fibronectin 1 isoform 3 preproprotein
TOP	NP_002206.1	inter-alpha (globulin) inhibitor H1
TOP	NP_002207.1	inter-alpha (globulin) inhibitor H2
TOP	NP_002209.2	inter-alpha (globulin) inhibitor H4
TOP	NP_002855.1	pregnancy-zone protein
TOP	NP_003656.2	ficolin 3 isoform 1 precursor
TOP	NP_003881.1	Fc fragment of IgG binding protein
TOP	NP_005558.1	galectin 3 binding protein
TOP	NP_006477.2	fibulin 1 isoform D
TOP	XP_291977.3	similar to Olfactory receptor 52L1
OTHER	NP_001002029.1	complement component 4B preproprotein
OTHER	NP_002046.1	glial fibrillary acidic protein
OTHER	NP_006476.2	fibulin 1 isoform B precursor
OTHER	NP_006478.2	fibulin 1 isoform A precursor a disintegrin and metalloprotease with
OTHER	NP_008968.2	thrombospondin motifs-8
OTHER	NP_775628.1	ficolin 3 isoform 2 precursor+C36
OTHER	NP_976084.1	clusterin isoform 2
OTHER	NP_997639.1	fibronectin 1 isoform 6 preproprotein
OTHER	NP_997640.1	fibronectin 1 isoform 2 preproprotein

OTHER	NP_997641.1	fibronectin 1 isoform 5 preproprotein
OTHER	NP_997643.1	fibronectin 1 isoform 4 preproprotein
OTHER	NP_997647.1	fibronectin 1 isoform 1 preproprotein
OTHER	XP_290809.3	TAF4b RNA polymerase II

TABLE H: *Proteins Matching in C4 (24 minimum/ 32 maximum)*

TOP	NP_000005.1	alpha-2-macroglobulin precursor
TOP	NP_000055.1	complement component 3 precursor
TOP	NP_000087.1	ceruloplasmin (ferroxidase)
TOP	NP_000177.1	complement factor H
TOP	NP_000292.1	plasminogen
TOP	NP_000375.1	apolipoprotein B precursor
TOP	NP_000403.1	histidine-rich glycoprotein precursor
TOP	NP_000437.3	paraoxonase 1
TOP	NP_000574.2	vitamin D-binding protein precursor
TOP	NP_000583.1	complement component 4B preprotein
TOP	NP_000604.1	hemopexin
TOP	NP_000629.2	vitronectin precursor
TOP	NP_000706.1	complement component 4 binding protein, alpha
TOP	NP_001076.1	alpha-1-antichymotrypsin, precursor
TOP	NP_001613.1	alpha-2-HS-glycoprotein
TOP	NP_001701.1	complement factor B preproprotein
TOP	NP_001726.2	complement component 5
TOP	NP_001822.2	clusterin isoform 1
TOP	NP_002017.1	fibronectin 1 isoform 3 preproprotein
TOP	NP_002206.1	inter-alpha (globulin) inhibitor H1
TOP	NP_002207.1	inter-alpha (globulin) inhibitor H2
TOP	NP_002209.2	inter-alpha (globulin) inhibitor H4
TOP	NP_055625.2	centrosome-associated protein 350
TOP	NP_078860.2	hypothetical protein FLJ13646
OTHER	NP_001002029.1	complement component 4B preproprotein
OTHER	NP_002855.1	pregnancy-zone protein
OTHER	NP_976084.1	clusterin isoform 2
OTHER	NP_997639.1	fibronectin 1 isoform 6 preproprotein
OTHER	NP_997640.1	fibronectin 1 isoform 2 preproprotein
OTHER	NP_997641.1	fibronectin 1 isoform 5 preproprotein
OTHER	NP_997643.1	fibronectin 1 isoform 4 preproprotein
OTHER	NP_997647.1	fibronectin 1 isoform 1 preproprotein

TABLE I: Proteins Matching to Pooled Male Serum (28 minimum/ 34 maximum)

TOP	NP_000005.1	alpha-2-macroglobulin precursor
TOP	NP_000020.1	angiotensinogen precursor
TOP	NP_000053.1	complement component 1 inhibitor precursor
TOP	NP_000055.1	complement component 3 precursor
TOP	NP_000087.1	ceruloplasmin (ferroxidase)
TOP	NP_000286.2	serine (or cysteine) proteinase inhibitor, clade A
TOP	NP_000375.1	apolipoprotein B precursor
TOP	NP_000437.3	paraoxonase 1
TOP	NP_000479.1	serine (or cysteine) proteinase inhibitor, clade C
TOP	NP_000497.1	coagulation factor II precursor
TOP	NP_000574.2	vitamin D-binding protein precursor
TOP	NP_000583.1	complement component 4B proprotein
TOP	NP_000598.1	orosomucoid 1 precursor
TOP	NP_000599.1	orosomucoid 2
TOP	NP_000629.2	vitronectin precursor
TOP	NP_000706.1	complement component 4 binding protein, alpha
TOP	NP_000884.1	kininogen 1
TOP	NP_001076.1	alpha-1-antichymotrypsin, precursor
TOP	NP_001613.1	alpha-2-HS-glycoprotein
TOP	NP_001822.2	clusterin isoform 1
TOP	NP_002206.1	inter-alpha (globulin) inhibitor H1
TOP	NP_002207.1	inter-alpha (globulin) inhibitor H2
TOP	NP_002208.1	inter-alpha (globulin) inhibitor H3
TOP	NP_002209.2	inter-alpha (globulin) inhibitor H4
TOP	NP_003656.2	ficolin 3 isoform 1 precursor
TOP	NP_005558.1	galectin 3 binding protein
TOP	XP_209550.4	similar to Carboxypeptidase N 83 kDa chain
OTHER	NP_001002029.1	complement component 4B preproprotein
		matrix metalloproteinase 19 isoform rasi-1,
OTHER	NP_002420.1	preproprotein
OTHER	NP_002855.1	pregnancy-zone protein
OTHER	NP_073629.1	matrix metalloproteinase 19 isoform rasi-9
OTHER	NP_775628.1	ficolin 3 isoform 2 precursor
OTHER	NP_976084.1	clusterin isoform 2
OTHER	XP_370616.3	similar to hypothetical protein

CHAPTER 5

GLYCOMICS –

QUANTIFICATION BY ISOBARIC LABELING (QUIBL) FOR THE COMPARATIVE GLYCOMICS STUDY OF O-LINKED GLYCANS *

*Botelho, J. C., Atwood III, J. A., Cheng, L., Alvarez-Manilla, G., York, W., Orlando, R.
To be submitted to *International Journal of Mass Spectrometry*

ABSTRACT

Quantification by isobaric labeling (QUIBL) has recently been described as a tool for the relative quantification of N-linked glycans in glycomic studies. In this paper we expand the application of QUIBL to the relative quantification of O-linked glycans. Fetuin, which contains two O-linked glycans, NeuAc-Hex-HexNAc and NeuAc₂-Hex-HexNAc, was used to validate this study. The glycans were released by β -elimination and permethylated with the isobaric labeling reagents, $^{13}\text{CH}_3\text{I}$ and $^{12}\text{CH}_2\text{DI}$. The exact mass difference between the isobaric labels is minimal, (0.00292 Da), but since glycans contain multiple permethylation sites, the mass difference can be resolved with a high resolution mass spectrometer, such as an FTMS. Quantitative data were obtained by comparing the signal intensities from various mixtures of isobarically labeled O-linked glycans of fetuin. Five different ratios of $^{13}\text{CH}_3$ to $^{12}\text{CH}_2\text{D}$ ranging from 10:1 to 1:10 were analyzed for linearity and reproducibility of the QUIBL method on the standard glycan. QUIBL was also evaluated using porcine mucin, a sample with a larger variety of more complex O-linked glycans. These results indicate that QUIBL allows for the relative quantification of O-linked glycans over a linear dynamic range of at least two orders of magnitude, validating QUIBL as an applicable quantitative glycomics approach for both N-linked and O-linked glycans.

INTRODUCTION

Glycosylation is one of the most common and complex forms of post-translational modifications found on proteins [1-3]. Protein glycosylation is typically referred to as N-linked or O-linked based on its origin of attachment to the peptide backbone. The study of this type of post-translation modification has become of increasing relevance given the functions that glycans have been demonstrated to exhibit in development, differentiation, cell adhesion and growth [1-4]. The potential complexity of glycosylation types requires that each site be structurally characterized. This has led to the emerging field called glycomics, which involves the characterization of glycans at each stage of a biological event, i.e., differentiation, tumor progression, etc. Mass spectrometry plays a key role in glycomics by providing a highly sensitive analytical technique that can offer a rapid and reliable method to distinguish and identify individual glycans in mixtures.

O-linked glycans are a prevalent form of glycosylation found on cell surfaces and secreted proteins. They are highly enriched in epithelial cells due to their role in mucus production. This type of glycan is most commonly attached to the peptide chain through the oxygen residue on the side chains of serine (Ser) or threonine (Thr) residues. There does not appear to be an amino-acid consensus sequence for O- glycosylation, nor is there a single characteristic O-glycan core structure, as observed in N-glycans. Mucin-type glycans (the most common type of O-linked glycans) are attached to the peptide backbone via a GalNAc residue. Mucin-type O-linked glycans are thought to have many functions, including blocking the accessibility of the polypeptide backbone to proteases, regulating the serum-half life of chemokines or hormones to attenuate their activity *in vivo*, and modulating the intracellular trafficking of proteins [5].

The first step in the analysis of O-linked glycans is typically to chemically release them from the protein backbone by a beta elimination reaction [6-9]. This chemical reaction takes advantage of the slight acidity of the proton on the alpha carbon of Ser and Thr. In an alkaline environment, abstraction of the alpha proton of the glycosylated Ser or Thr residue leads to release of the O-linked carbohydrate that is attached to the beta-carbon via a β -elimination mechanism. The free oligosaccharide contains a carbonyl group as a result of the release, which is very susceptible to further elimination, called “peeling”. Several methods have been used to prevent peeling, with the most common approach involving the chemical derivatization of the aldose into an alditol with the assistance of a reducing agent such as sodium borohydride [10]. Released glycans can be further derivatized by permethylation [9, 11]. In this reaction, the hydrogen on both oxygen and nitrogen atoms in the oligosaccharide are replaced by methyl groups making the oligosaccharide less polar and more hydrophobic [12]. This chemical modification yields more predictable ion products, stabilizes sialic acids residues, and allows oligosaccharides to be separated from salts and other contaminants due to the hydrophobic nature of methylated glycans. In addition, methylated carbohydrates ionize more efficiently, which aids in MS analysis [9, 11, 13].

While MS offers a powerful tool for qualitative glycomics, quantitation can be problematic for a multitude of reasons. For instance, the ionization efficiency of each analyte depends on factors such as its molecular mass, proton/cation affinity, surface activity, the presence of compounds which compete with or interfere with the ionization of the analyte, etc., and thus the intensities of the ions does not directly correlate with concentration. In addition, the instrument’s response can vary over time, so that the direct comparison of two or more spectra obtained at different times yields a qualitative, rather than quantitative, indication of the glycan

content of the two analyte samples. To compensate for these factors, quantitative measurements are typically performed by adding an internal standard and measuring the analyte's response relative to this standard. One of the best internal standards is an isotopically labeled (^{13}C , D, ^{15}N , etc.) form of the analyte itself, as this often retains the physical-chemical characteristics of the analyte while allowing the mass analyzer to resolve the isotopomers permitting their relative abundances to be determined by comparing the intensity of ions from the analyte to that of the isotopically labeled standard. For comparative studies of complex samples where isotopically labeled standards are not available for all species, an isotopic labeling approach where one of the samples is modified with a “light” tag while the other is derivatized with a “heavy” tag can be used [14].

Stable isotopic labeling has shown widespread application in the proteomics field,^{14, 15} whereas it has shown limited use in the glycomics field. Some advances in the quantitative glycomics field have been shown by Yuan *et al.* [16] and Hitchcock *et al.* [17], who used an isotopically varied reductive amination tag for quantification. Bowman *et al.* synthesized tags that could modify oligosaccharides with four isotope-enriched variants [18]. In these cases, derivatives were added to the reducing terminus of N-linked glycans. This method would become problematic for quantification for O-linked glycans since the reducing end of the glycans are usually reduced as a result of β -elimination. Xie *et al.* reported a method for quantification of O-linked glycans using sodium borohydride or sodium borodeuteride during β -elimination to incorporate deuterium [19]. While successful for O-linked glycans, this method would require additional chemistry to be applied to N-linked glycans. Another proposed method for comparative isotopic labeling of oligosaccharides uses heavy methyl iodide ($^{13}\text{CH}_3$ or $^{12}\text{CD}_3$) vs. light methyl iodide ($^{12}\text{CH}_3$), which employs standard permethylation conditions, which are

routinely used for both N-linked and O-linked glycan analysis [11, 20]. While this study was successful for N-linked glycans and shows great potential for O-linked glycans, the large m/z shift between heavy and light labeled glycans can become problematic in the analysis of complex unknown mixtures where the number of modified sites is not known making it difficult to identify differentially labeled mass pairs.

Here, we evaluate the use of a new approach, known as quantitative isobaric labeling (QUBIL), for relative quantification of O-linked glycans in glycomic studies. We have previously demonstrated the utility of QUIBL for comparative N-linked glycomics [21]. In this previous report, released N-linked glycans were permethylated with either $^{13}\text{CH}_3\text{I}$ or $^{12}\text{CH}_2\text{DI}$ and the small mass shift between these two isobaric labels permitted the ions from the two samples to be resolved in the MS, which in turn allowed relative quantification of glycans. Since the isobaric labels were introduced during the permethylation step, QUIBL should be applicable to the analysis of O-linked glycans without additional sample manipulation. This possibility was evaluated here by the analysis of the two O-linked glycans released from fetal calf fetuin and with porcine stomach mucin. This analysis demonstrates the extension of QUIBL to O-linked glycans, and thus further supports that this approach is useful for comparative glycomic investigations.

EXPERIMENTAL PROCEDURE

Materials

Bovine fetuin, type III partially purified porcine stomach mucin, and Dowex ion exchange resin was purchased from Sigma Aldrich (St. Louis, MO). The 99% $^{13}\text{CH}_3\text{I}$ and 98% $^{12}\text{CH}_2\text{DI}$ used during permethylation were purchased from Cambridge Isotopes Inc. (Andover, MA).

Beta-Elimination of Glycoproteins

Beta-elimination was carried on fetuin from fetal calf serum, as previously described by Dell *et al.* [9]. Briefly, a 1.0 M solution of sodium borohydride in 50mM NaOH was added to 400 μg of fetuin and incubated for 16 hours at 45°C. After the allotted time, the sample was cooled to room temperature and neutralized by the addition of 10% acetic acid. Sodium ions were removed using an ion exchange resin (Dowex 50 W x 8-100) and the sample was lyophilized to dryness. Next, borate was removed by at least repeated additions of a methanol/acetic acid solution, which was dried under a stream of nitrogen.

Permethylation of Released O-Glycans

Permethylation was performed on the O-linked glycans according to standard procedures [22]. Briefly, dried glycans were suspended in DMSO (200 μL) and stirred until completely dissolved. Approximately 200 μL of 50% sodium hydroxide in DMSO was added and the solution was mixed vigorously. Each sample was divided into two equal portions, which were permethylated separately by the addition 100 μL of either 99% $^{13}\text{CH}_3\text{I}$ or 98% $^{12}\text{CH}_2\text{DI}$. After at least 15 minutes, the methylation reaction was terminated with 2 mL of water. Excess

iodomethane was purged from the resulting mixture with N₂ gas. The permethylated glycans were then extracted with dichloromethane and dried for further analysis by mass spectrometry.

MS Analysis of Released O-Glycans

The glycans were resuspended in 1 mM NaOH in 50% MeOH and analyzed on a hybrid linear ion trap Fourier transform mass spectrometer, LTQ-FT (Thermo Finnegan, San Jose, CA). Each glycan mixture was directly infused into the LTQ-FT by syringe injection at a flow rate of 0.3 μ L/min. The sample was nano electrosprayed through a pulled silica capillary tip (75 μ m $\times$ 10cm Polymicro Technologies, L.L.C, Phoenix, AZ) into the MS where the spray voltage was held at 2.0 kV and the capillary temperature at 250°C.

The ¹³CH₃ and ¹²CH₂D labeled glycans from fetuin were analyzed independently to validate that the proper glycans were present. Quasimolecular ions at 909 and 1275 *m/z* were assigned as permethylated fetuin glycans. MS/MS experiments were carried out in the ion trap using normalized collision energy of 29% after isolation of the precursor masses in a 5.0 *m/z* window to fragment the sample. The fragmentation pattern was used to validate the glycan structures assigned to these two quasimolecular ions.

Mixtures of permethylated fetuin glycans containing the isobaric labels were prepared for ¹³CH₃:CH₂D in the following ratios: 10:1, 8:3, 1:1, 3:8, and 1:10. Analysis was carried out in the FTMS on the ions at 909 *m/z* and 1275 *m/z* with an isolation width of 25.0 *m/z*. The resolving power was set to 100,000 *m*/ Δ *m*, which provide ample resolution to separate the ¹³C and D labeled glycans. The study was carried out in triplicate for each ratio of ¹³CH₃:CH₂D. Quantification was performed by comparing the sum of the peak intensities for all isotopes from each permethylated glycan pairs.

O-linked glycans from 400 µg of type III partially purified porcine stomach mucin were released and permethylated as described above. A mixture of $^{13}\text{CH}_3:\text{CH}_2\text{D}$ was prepared at a ratio of 1:1 and analyzed on the LTQ-FTMS in triplicate. Fragmentation patterns obtained by MS/MS analysis were used to identify structures of the O-linked glycans of mucin. Relative quantification for each differentially labeled glycan pair was accomplished by comparing the sum of all abundances for its ^{13}C -labeled ions to the sum of all abundances for its D-labeled ions.

RESULTS AND DISCUSSION

Principles of Quantification by QUIBL

The International Union of Pure and Applied Chemist (IUPAC) definition of isobars is “nuclides of equal mass number” [23]. Hence, isobars contain the same mass number (sum of all protons and neutrons) but have a different atomic number (sum of all protons), which in turn leads to isobars possessing different exact masses but the same nominal mass. Isobars are clearly different than isotopes, which have the same atomic number with different mass number, and isotones, which have equal number of neutrons but a different number of protons [23].

The IUPAC definition of an isobar is written for atomic species; however it can be extended to molecular species as well. Commonly encountered molecular isobars in the field of biochemistry are the two amino acids lysine (K) and glutamine (Q), whose chemical formulae are $\text{C}_6\text{H}_{14}\text{N}_2\text{O}_2$ and $\text{C}_5\text{H}_{10}\text{N}_2\text{O}_3$, respectively. These two species clearly fit the definition of isobars as they both have a molecular mass number of 146, but possess different numbers of protons, K has 80 while Q has 78. Consequently, these two species share the same nominal mass (146 Da), but have different exact masses, (146.1882 Da for K vs. 146.1451 Da for Q). Leucine (L) and isoleucine (I) are another pair of amino acids that share the same nominal mass, but in

this instance they have identical elemental compositions ($C_6H_{13}NO_2$), molecular mass number (131), and number of protons (72), consequently these two amino acids have the same exact molecular mass (131.174 Da). In this instance, L and I are structural isomers, not isobars, although they are often referred to as such in the literature [24-27]. Leucine and isoleucine, however, are isobars of hydroxyproline ($C_5H_9NO_3$), which has a molecular mass number of 131, a total of 70 protons, and an exact mass of 131.058 Da -- 0.115 Da below that of L and I. Similar logic has led to the extension of isobars to include two or more compounds possessing the same nominal mass (i.e., total number of protons and neutrons) with different elemental or isotopic compositions [28] and thus different exact masses.

The characteristic of isobars having nearly the same exact mass offers a unique advantage for the introduction of an isotopic label for various quantitative studies. For instance if two samples were labeled with isobaric reagents, the molecular ion species from these samples would appear as the same ion under low resolution conditions and therefore lead to increased ion intensity as ions from both samples would not be distributed between isotopic species having different m/z values. This reduction in spectral complexity clearly is beneficial for reducing the limit of detection of the compound being analyzed. Similarly, the small mass difference between these isobars allows the molecular ions from the two samples to be simultaneously selected for MS^n analysis, permitting qualitative confirmation of the analyte. Quantification would be achieved at either the MS or the MS^n level by performing the analysis in a high resolution instrument. Consequently, isobaric labeling offers the ability to rapidly switch between high sensitivity qualitative analysis at low resolution and quantitative studies at high resolution using an appropriate spectrometer such as a hybrid ion trap-FTMS system.

The primary obstacle to using isobars for quantitative labeling is their small mass difference. For example the isobaric tags we have selected for this study, $^{13}\text{CH}_3\text{I}$ and $^{12}\text{CH}_2\text{DI}$, have exact masses of 142.931 Da and 142.934 Da, respectively. The mass difference (ΔM) between these isobars is 0.003 Da, and would require a mass resolution ($M/\Delta M$) of over 700,000 in the case of a single label on an analyte with a molecular mass of 2,000 Da. This would be difficult to attain routinely on all but the most expensive mass spectrometers, and thus is impractical for general use. These two reagents were selected because methyl iodide are the common reagent for per-methylating glycans, which contain multiple methylation sites (*i.e.*, –OH and NH_2 groups). As the number of methylation sites increases so does the mass difference between a pair of differentially labeled isobaric species. For example, an O-linked glycan with a NeuAcHexHexNAc monosaccharide composition has 14 permethylation sites, resulting in a mass of 886.489 Da when labeled with $^{13}\text{CH}_3$ and a mass of 886.530 Da when labeled with $^{12}\text{CH}_2\text{D}$. The mass difference of 0.042 Da between these two species requires a mass resolution of only $\sim 22,000$ $M/\Delta M$ to resolve. Furthermore, the mass difference between oligosaccharide isobaric species becomes larger as the size of the structure increases due to the increasing number of permethylation sites, and in this situation, the resolution required to distinguish between the isobaric pairs becomes relatively independent of the size of glycan.

Validation of QUIBL on O-Glycans with Fetuin

The O-linked glycans released from the fetal calf serum protein fetuin were used to demonstrate the principles of QUIBL. The fetuin samples were β -eliminated to release the O-linked glycans, permethylated with isobaric labels, $^{13}\text{CH}_3$ and $^{12}\text{CH}_2\text{D}$, and analyzed with a hybrid linear ion trap-FTMS. The high resolution spectrum of a 1:1 mixture containing both

($^{13}\text{CH}_3$ and $^{12}\text{CH}_2\text{D}$) labeled forms of HexNAc-Hex-NeuAc (glycan 1) contains intense ions at 909.510 and 909.550 m/z , respectively, which correspond to (monoisotopic) sodium adducts of the differentially labeled glycan (FIGURE 5.1A). It is noteworthy that m/z for the quasimolecular ion of the $^{12}\text{CH}_2\text{D}$ labeled glycan is shifted by 0.040 units compared to its $^{13}\text{CH}_3$ labeled counterpart, in good agreement with the shift predicted for the presence of 14 methyl groups on a singly charged ion ($[0.003 \times 14] = 0.042$). Similarly, the 0.055 Da mass shift observed for HexNAc-Hex-NeuAc₂ (glycan 2, FIGURE 5.1B) agrees well with the calculated shift of 0.057 Da based on the 19 permethylation sites on this glycan. The excellent agreement between the calculated and experimental mass shifts suggests that these can be used to calculate the number of methylation sites in a molecule thereby aiding in the structural elucidation of unknown glycans.

The spectra of the isobarically labeled species (FIGURE 5.1A) also contain a doublet of ions at 908.507 and 908.544 m/z that is exactly 1 m/z unit below of that observed from the fully labeled species. These ions appear to result from incomplete isotope enrichment of the labeling reagents used, i.e., 99% $^{13}\text{CH}_3$ and 98% $^{12}\text{CH}_2\text{D}$. This spectrum demonstrates that in a QUIBL experiment, it is possible to resolve the incompletely labeled ions. This is possible because when one of the fourteen ^{13}C atoms is replaced with a ^{12}C atom or one of the fourteen D (or ^2H) atoms is replaced with an ^1H atom, the analyte's mass is consequently reduced by approximately 1 Da, however, the resulting ion is detected in the appropriate ($^{13}\text{CH}_3$ -labeled or $^{12}\text{CH}_2\text{D}$ -labeled) ion series because it still contains thirteen of the isotopic labels. This greatly simplifies quantitation, which is accomplished by summing the ion abundances for the $^{13}\text{CH}_3$ -labeled and $^{12}\text{CH}_2\text{D}$ -labeled series and comparing these two values. This characteristic of QUIBL provides an advantage over traditional isotopic labeling procedures, where the use of incompletely labeled

reagents results in overlapping isotopic peaks, i.e., the ion produced by the under incorporated “heavy” species appears at an m/z value that is indistinguishable from an ion produced by the “light” species [11, 13, 29, 30]. The ability of QUIBL to differentiate species from under isotope incorporation greatly simplifies quantitation calculation which simply involves summing the ion intensities for each isobaric distribution.

The quantification capabilities of QUIBL on O-linked glycans were then evaluated by analyzing 5 different ratios of $^{13}\text{CH}_3$ to $^{12}\text{CH}_2\text{D}$ ranging from 10:1 to 1:10. The glycan mixtures were analyzed by FTMS in technical triplicates. The experimental ratios for the two O-linked glycans were calculated based on the ion intensities and compared to the theoretical ratio. This approach showed good reproducibility with the calculated ratios for each triplicate run being in relatively good agreement. The theoretical and experimentally averaged ratios were compared demonstrating a linear response over two orders of magnitude, as shown in FIGURE 5.2.

Isotope Effects in QUIBL

The method of permethylation used in this analysis involves the deprotonation of the oxygens and nitrogens on the glycan prior to the addition of methyl iodide under strong basic conditions [31]. The addition of methyl iodide is established through an $\text{S}_{\text{N}}2$ reaction where the deprotonated oxygens and nitrogens of the glycan act as the nucleophile (Nu) to displace iodine so that the C-I bond and the formation of the new C-Nu bond occurs simultaneously. The use of isotopes in this reaction could induce a kinetic isotope effect.

A kinetic isotope effect (KIE) is a variation in the reaction rate of a chemical reaction as a result of atoms of a reactant that is replaced by its isotope. When an isotope is introduced, it lowers the minimum vibrational energy of the bond, the zero point energy. As a result, more

energy is required to raise the vibrational state of the bond to the point of breakage. The isotope effect is a function of the difference in the vibrational frequencies related to a bond being formed or broken which in turn affects the reaction rate. An isotopic substitution can modify the reaction rate, especially if the isotope is directly involved in the chemical bond that is broken and/or formed. The change of rate from the direct isotopic involvement in the reaction is referred to as a primary isotope effect. A secondary isotope effect occurs if the substituted isotope is not directly involved in the bond that is breaking or forming. The relative change in rate associated with the incorporation of these isotopes is a function of the inverse square root of the ratios of the reduced masses of the atoms involved in the bonds based the calculation of the vibrational energy required to break or form a bond.

The use of $^{13}\text{CH}_3\text{-I}$ as a labeling reagent would induce a primary kinetic isotope effect during the $\text{S}_{\text{N}}2$ reaction since the ^{13}C is directly involved in the chemical bond breaking and forming. Based on the inverse square root of the reduced mass ratios, the theoretical decrease in reaction rate is approximately a 2%, a minimal effect and well within the experimental precision of these measurements. In general, isotope effects involving ^{13}C tend to be much smaller than those involving deuterium. However, the use of $^{12}\text{CH}_2\text{D-I}$ as an isobaric labeling reagent would induce a secondary alpha kinetic isotope effect since the D is not directly involved in the reaction. This secondary alpha effect is also reported [32] to introduce about a 2% decrease in the reaction rate, providing an effect that is similar to the primary KIE due to the presence of ^{13}C . As the effects of these isotopes are small and comparable, any difference in reaction rates for incorporation of the two different isobaric labels is theoretically insignificant. An additional isotope effect to consider could be caused by the steric hindrance from the deuterium during the

nucleophilic backside attack of the S_N2 reaction. This hindrance could slow the reaction rate when compared to the S_N2 reaction involving all hydrogens as with the $^{13}\text{CH}_3\text{-I}$ labeling reagent.

If a rate change based on a kinetic isotope effect or steric hindrance occurs during permethylation using these isobaric labeling agents, under-labeling could occur at a greater rate for one sample over the other, affecting quantification. Although the kinetic differential isotope effects are theoretically small for the isobaric labels used in this analysis, avoiding incomplete labeling is important for reliable quantification. Therefore optimization of the reaction conditions for permethylation is a prerequisite for successful application of the QUIBL method.

Application of QUIBL for the Quantification of O-Glycans in Mucin

To evaluate QUIBL on a sample with a larger variety of more complex O-linked glycans, porcine mucin was subject to β -elimination and permethylated with the isobaric labels. The $^{13}\text{CH}_3$ and $^{12}\text{CH}_2\text{D}$ labeled glycans were then mixed in a 1:1 ratio and analyzed in technical triplicate. After identification of glycans by LTQ analysis (FIGURE 5.3A), quantification was preformed using the FTMS. For each glycan, pairs of differentially labeled ions were generated within the same nominal mass; a major advantage of the QUIBL method. This differs from traditional labeling techniques where large mass shifts between labeled species make identification difficult. Additionally, errors from discrepancies in detection efficiency between ions of very different m/z ratios were avoided using the QUIBL approach. The QUIBL method accurately quantified a broad range of O-linked glycans structures found in mucin from 709 to 1670 m/z . This demonstrates that this method is independent of glycan size or composition of the mucin glycans as illustrated in FIGURE 5.3B and 5.3C.

The average calculated ratios were determined as shown in TABLE 5.1 for each glycan. The maximum observed error for the experimental ratio when compared to the theoretical ratio obtained in the fetuin analysis was 26.43% and the average error for all glycans analyzed by QUIBL was 10.62%. This does not insure that the analysis of the mucin glycans were not without additional errors associated with the sample preparation as well as the sample itself. For example, the differences in the extent of glycan permethylation for the two reagents would introduce quantitative errors, which can be minimized by optimizing the methylation conditions (*i.e.*, by adjusting the reaction time and reagent amounts.)

CONCLUSIONS

This glycomics study allowed for the validation of quantitative isobaric labeling (QUIBL) on mucin-type O-linked glycans. In this analysis, fetuin was analyzed using the QUIBL approach for relative quantification. These results show that QUIBL allows for relative quantification of O-linked glycans over a linear dynamic range of two orders of magnitude. Additionally, QUIBL was effectively utilized on a highly O-glycosylated species, porcine mucin. In agreement with the results of Atwood *et al.* [21], QUIBL has proven to be a useful technique for quantitative glycomics of both N- and O-linked glycans without additional manipulation. Additionally, the concept of quantitative isobaric labeling is likely to be pertinent to other quantitative analyses in the *-omics* field.

REFERENCES

- [1] J. Bunkenborg, B.J. Pilch, A. V. Podtelejnikov, and J. R. Wisniewski. *Proteomics* **4** (2004), p. 454.
- [2] F. Dall'olio. *Clin Mol Pathol.* **49** (1996), p. M126.
- [3] Y. Wang, S. Wu, and W. S. Hancock. *Glycobiology.* **16** (2006) p. 514.
- [4] N. Taniguchi, E. Miyoshi, J. H. Ko, Y. Ikeda, and Y. Ihara. *Biochim Biophys Acta.* **1455** (1999), p. 287.
- [5] H. C. Hang and C. R. Bertozzi. *Bioorg Med Chem.* **13** (2005), p. 5021.
- [6] A. Varki. *The Essentials of Glycobiology.* Cold Spring Harbor Laboratory Press: **1999**.
- [7] M. D. Carlson. *Journal of Biological Chemistry* **241** (1966), p. 2984.
- [8] N. G. Karlsson and N. H. Packer. *Analytical Biochemistry.* **305** (2002), p. 173.
- [9] A. Dell, A. J. Reason, K. H. Khoo, M. Panico, R. A. McDowell, and H. R. Morris. *Methods Enzymol.* **230** (1994), p. 108.
- [10] L. G. Wade. *Organic chemistry.* Prentice Hall Englewood Cliffs, NJ: **1991**.
- [11] G. Alvarez-Manilla, N. L. Warren, T. Abney, J. A. Atwood III, P. Azadi, W. S. York, M. Pierce, and R. Orlando. *Glycobiology.* **17** (2007), p. 677.
- [12] J. Delaney and P. Vouros. *Rapid Communications in Mass Spectrometry.* **15** (2001), p. 325.
- [13] P. Kang, Y. Mechref, I. Klouckova, and M. V. Novotny. *Rapid Commun Mass Spectrom.* **19** (2005) p. 3421.
- [14] W. A. Tao and R. Aebersold. *Curr Opin Biotechnol.* **14** (2003), p. 110.
- [15] M. B. Goshe and R. D. Smith. *Curr Opin Biotechnol.* **14** (2003), p. 101.

- [16] J. Yuan, N. Hashii, N. Kawasaki, S. Itoh, T. Kawanishi and T. Hayakawa. *J Chromatogr A*. **1067** (2005), p. 145.
- [17] A. M. Hitchcock, C. E. Costello, and J. Zaia. *Biochemistry*. **45** (2006), p. 2350.
- [18] M. J. Bowman and J. Zaia. *Journal of Biomolecular Techniques*. **18** (2007) p. 268.
- [19] Y. Xie, J. Liu, J. Zhang, J. L. Hedrick, and C. B. Lebrilla. *Anal. Chem.* **76** (2004), p. 5186.
- [20] P. Kang, Y. Mechref, Z. Kyselova, K. Goetz, and M. V. Novotny. *Anal. Chem.* **79** (2007), p. 6064.
- [21] J. A. Atwood III, L. Cheng, G. Alvarez-Manilla, N. Warren, W. York, and R. Orlando *Submitted* 2007.
- [22] Ciucanu, I.; Costello, C. E., Elimination of oxidative degradation during the per-O-methylation of carbohydrates. *J Am Chem Soc* **2003**, 125, (52), 16213-9.
- [23] A. D. McNaught and A. Wilkinson. *IUPAC Compendium of Chemical Terminology: The Gold Book*. Blackwell Science: New York, **1997**.
- [24] R. S. Youngquist, G. R. Fuentes, M. P. Lacey, and T. Keough. *Journal of the American Chemical Society*. **117** (1995), p. 3900.
- [25] B. Casetta, D. Tagliacozzi, B. Shushan, and G. Federici. *Clin Chem Lab Med.* **38** (2000), p. 391.
- [26] B. C. Searle, S. Dasari, M. Turner, A. P. Reddy, D. Choi, P. A. Wilmarth, A. L. McCormack, L. L. David, and S. R. Nagalla. *Anal. Chem.* **76** (2004), p. 2220.
- [27] W. Kulik, L. Van Toledo-Eppinga, R. M. Kok, and W. S. Guérard. *Journal of Mass Spectrometry*. **30** (1995), p. 1260.
- [28] L. Aalberg, C. R. Clark, and J. DeRuiter. *Journal of Chromatographic Science*. **42** (2004), p. 464.
- [29] K. Aoki, M. Perlman, J. M. Lim, R. Cantu, L. Wells, and M. Tiemeyer. *J Biol Chem.* **282** (2007), p. 9127.
- [30] P. M. Angel and R. Orlando. *Anal Biochem.* **359** (2006), p. 26.
- [31] I. Ciucanu and F. Kerek. *Carbohydrate Research*. **131** (1984), p. 209.

[32] H. Simon and D. Palm. *Angewandte Chemie* 5 (2003), p. 920.

FIGURE 5.1A:

FTMS spectrum of $^{13}\text{CH}_3$ and $^{12}\text{CH}_2\text{D}$ labeled glycan 1 from fetuin at a 1:1 mixture. The following carbohydrate residues are represented: square: *N*-acetylhexosamine (HexNAc); circle: Hexose (Hex); and diamond: *N*-acetylneuraminic acid (NeuAc)

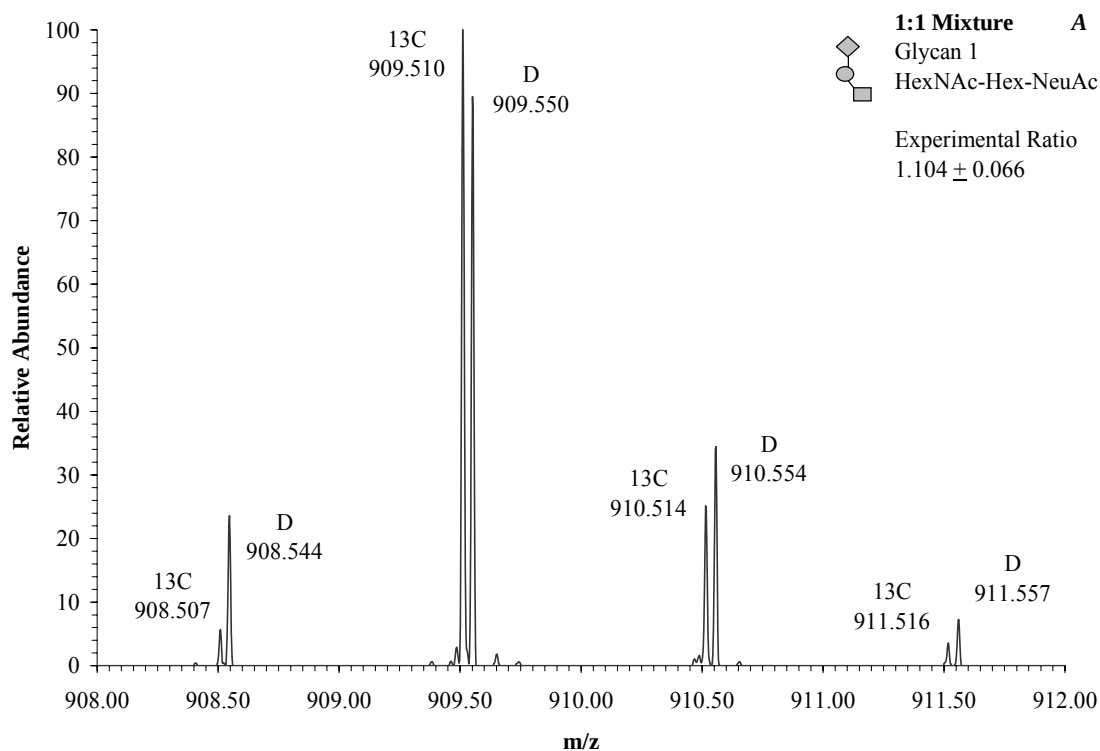


FIGURE 5.1B:

FTMS spectrum of $^{13}\text{CH}_3$ and $^{12}\text{CH}_2\text{D}$ labeled glycan 2 from fetuin at a 1:1 mixture. The following carbohydrate residues are represented: square: *N*-acetylhexosamine (HexNAc); circle: Hexose (Hex); and diamond: *N*-acetylneuraminic acid (NeuAc)

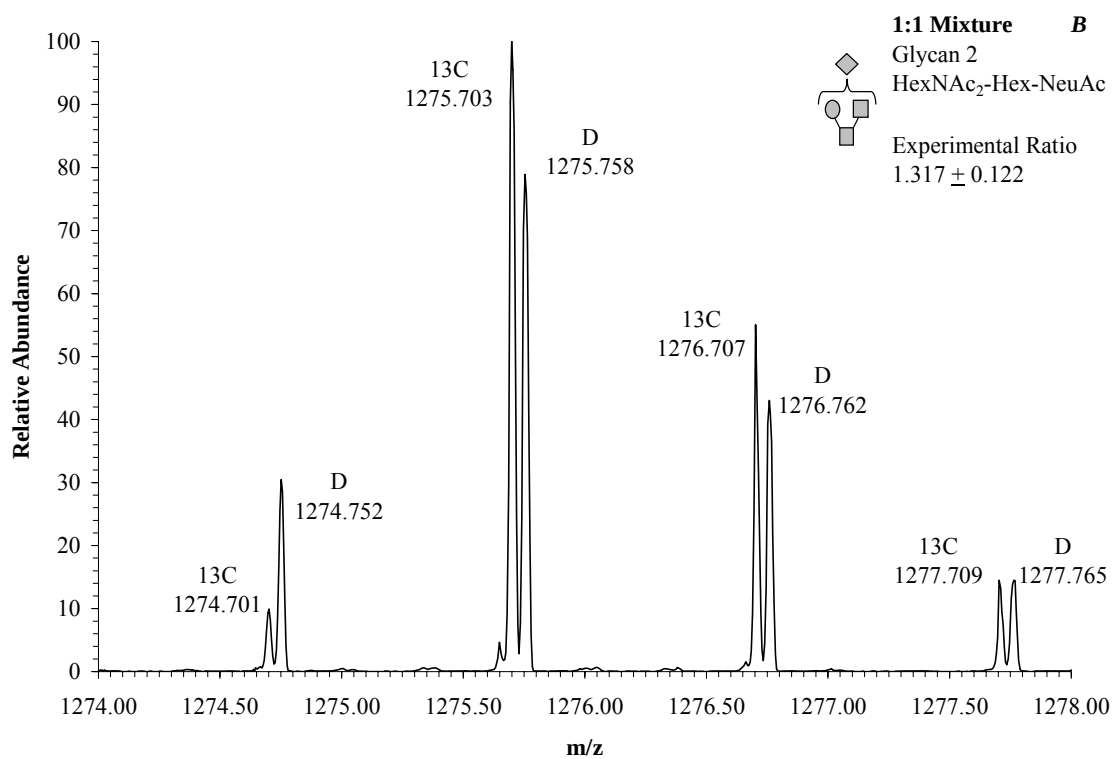


FIGURE 5.2:

This graph illustrates the correlation between the theoretical and experimental ratios for quantification of the two O-linked fetuin glycans by QUIBL. In each experiment the $^{13}\text{CH}_3$ and $^{12}\text{CH}_2\text{D}$ labeled glycans were mixed together at the ratios 10:1, 8:3, 1:1, 3:8, and 1:10 ($^{13}\text{CH}_3:\text{CH}_2\text{D}$) and analyzed by FTMS. The calculated expression ratios were determined by comparing the sum of the peak intensities for all isotopes between $^{13}\text{CH}_3$ and $^{12}\text{CH}_2\text{D}$ labeled precursor ions for each glycan. For both glycans a linear correlation was observed over two orders of magnitude with a minimum R^2 of 0.99.

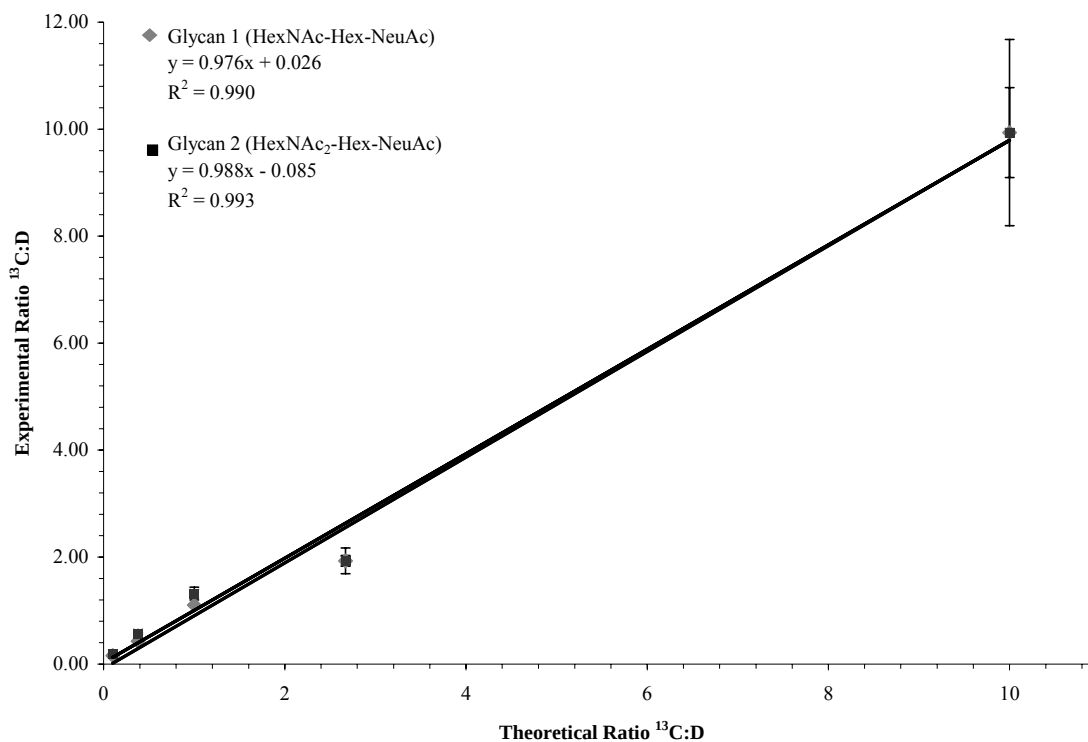


FIGURE 5.3A:

Full low resolution LTQ spectrum of $^{13}\text{CH}_3$ and $^{12}\text{CH}_2\text{D}$ labeled glycan from mucin at a 1:1 mixture. The following carbohydrate residues are represented: square: *N*-acetylhexosamine (HexNAc); circle: Hexose (Hex); diamond: *N*-acetylneuraminic acid (NeuAc); and triangle: Deoxyhexose (Deoxyhex).

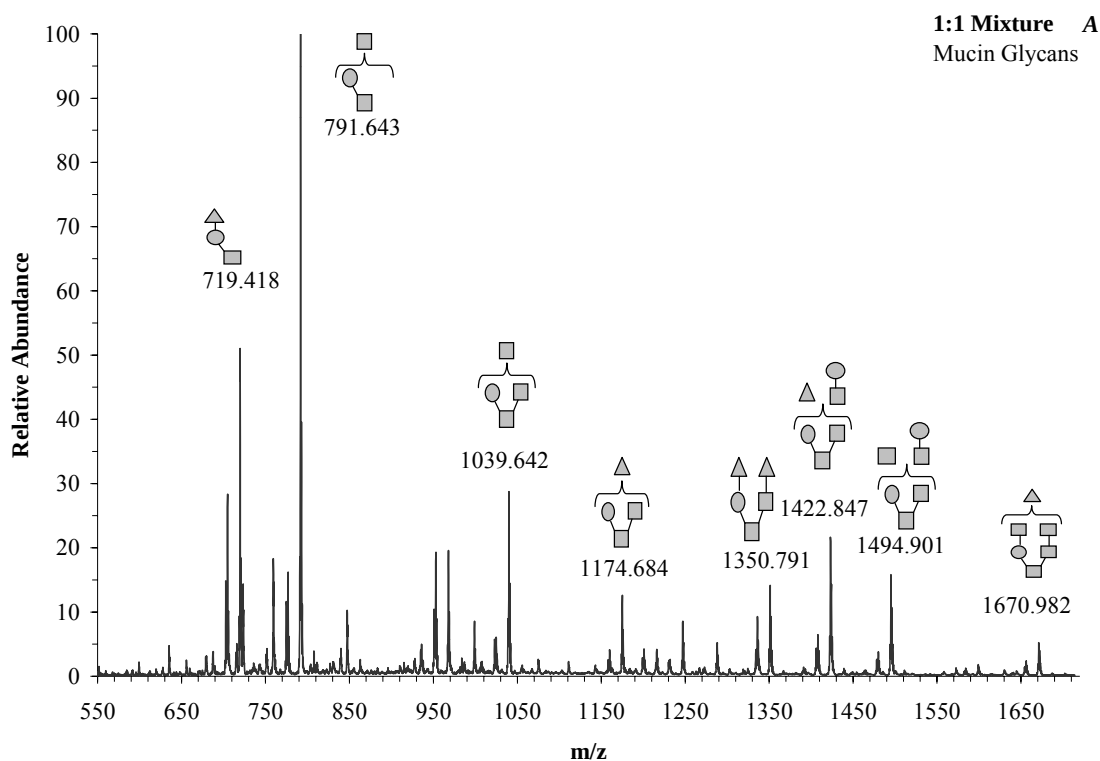


FIGURE 5.3B:

FTMS spectra of $^{13}\text{CH}_3$ and $^{12}\text{CH}_2\text{D}$ labeled of a small glycan from mucin at a 1:1 mixture. The following carbohydrate residues are represented: square: *N*-acetylhexosamine (HexNAc); circle: Hexose (Hex); and diamond: *N*-acetylneuraminic acid (NeuAc)

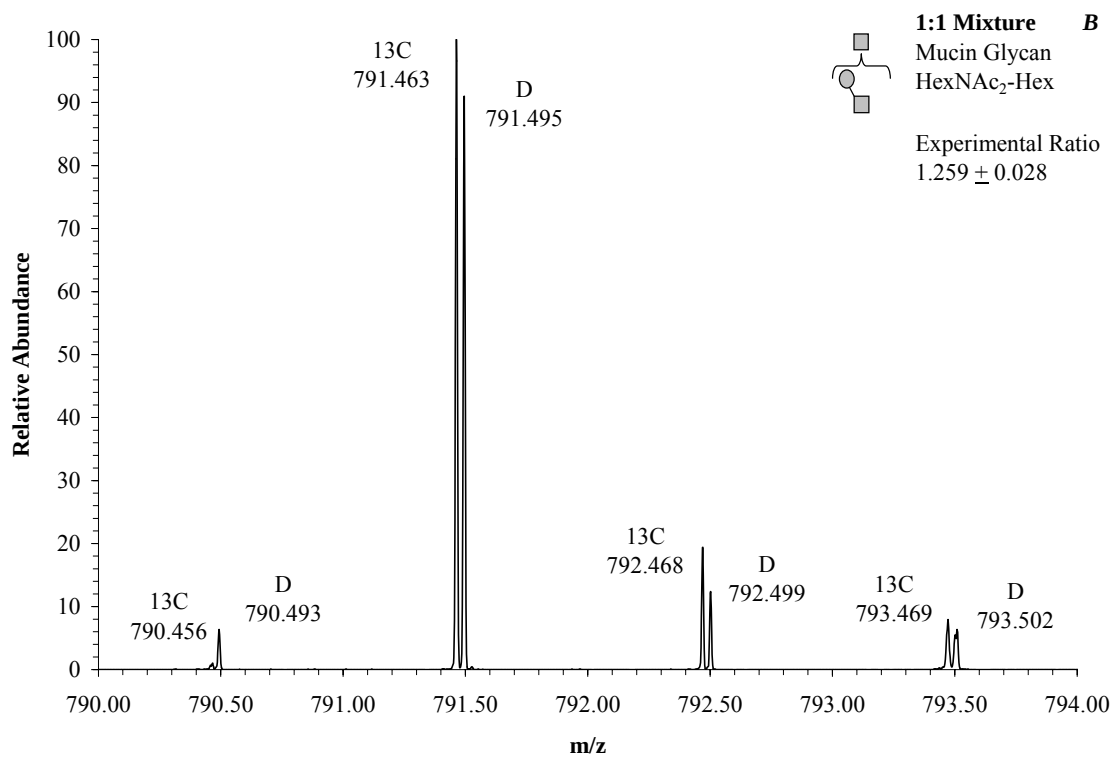


FIGURE 5.3C:

FTMS spectra of $^{13}\text{CH}_3$ and $^{12}\text{CH}_2\text{D}$ labeled of a large glycan from mucin at a 1:1 mixture. The following carbohydrate residues are represented: square: *N*-acetylhexosamine (HexNAc); circle: Hexose (Hex); and diamond: *N*-acetylneuraminic acid (NeuAc)

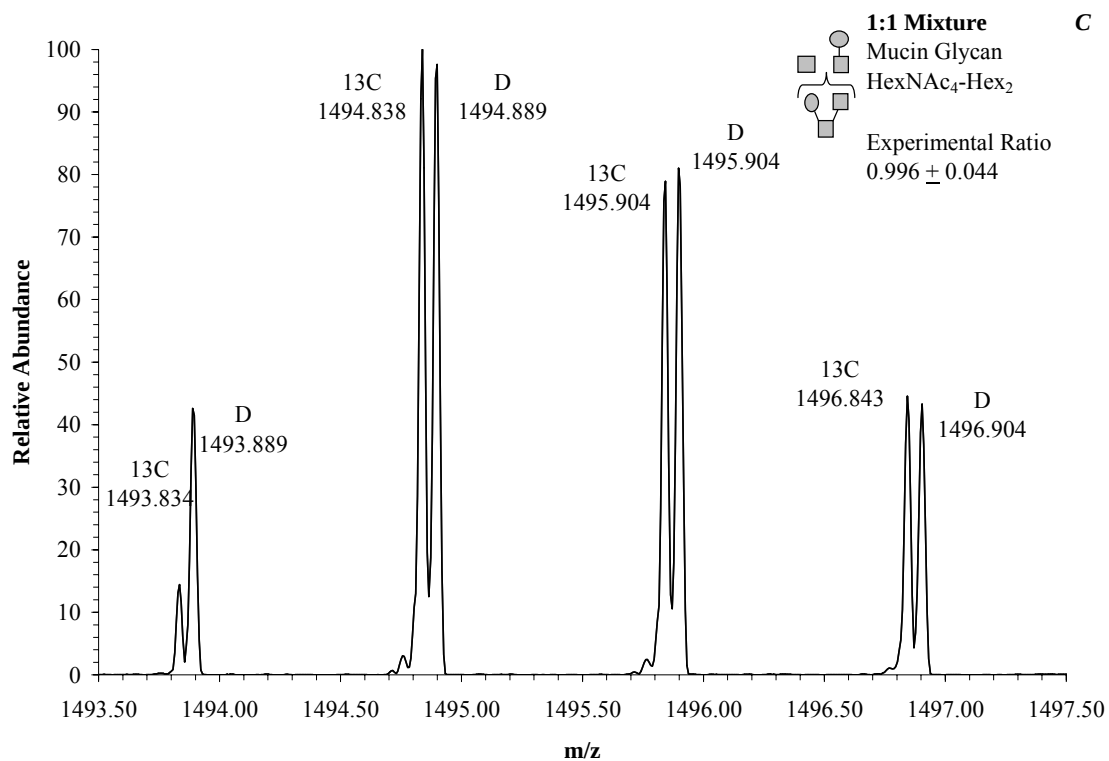


TABLE 5.1:

Quantification of various Porcine Mucin Glycans by QUIBL

<u>m/z</u>	<u>Glycan</u>	<u>Experimental ratio</u> (Ave $\pm$ 2SD)	<u>Theoretical ratio</u>	<u>% Error</u>
719.418	HexNAc-Hex-Deoxyhex	1.013 $\pm$ 0.084	1	1.37
791.643	HexNAc ₂ -Hex	1.259 $\pm$ 0.028	1	26.43
1039.684	HexNAc ₃ -Hex	1.155 $\pm$ 0.062	1	15.04
1174.684	HexNAc ₂ -Hex ₂ -Deoxyhex	0.971 $\pm$ 0.040	1	3.19
1350.791	HexNAc ₂ -Hex ₂ -Deoxyhex ₂	0.909 $\pm$ 0.026	1	10.02
1422.847	HexNAc ₃ -Hex ₂ -Deoxyhex	0.809 $\pm$ 0.034	1	20.27
1494.838	HexNAc ₄ -Hex ₂	0.996 $\pm$ 0.044	1	0.91
1670.982	HexNAc ₄ -Hex ₂ -Deoxyhex	0.920 $\pm$ 0.148	1	7.74

CHAPTER 6

CONCLUSIONS

The overall purpose of this work was to use mass spectrometry to identify and in some cases quantify proteins and glycans in a variety of complex biological samples. While the identifications of cuticular proteins of a mosquito, glycoproteins in human serum, and the quantification of O-linked glycans in porcine mucin seem to have no connection they all employed the use of mass spectrometry with *-omic* technologies.

Chapter 3

The proteome of cast cuticles of *Anopheles gambiae* was analyzed by LC-MS/MS. Cuticles from both the head capsule and pupal were studied after extensive washing, separation, and digestion. The resulting data were analyzed by the traditional methods of protein identification using MASCOT and SEQUEST on a database comprised of known, novel, and ab initio predicted peptides. Analysis of the cast cuticle proteome allowed for the identification of the first authentic cast cuticular proteome with 119 proteins being identified. Additionally, several annotated cuticular proteins were reported as authentic for the first time.

Chapter 4

The method described in this chapter allowed for the identification of the glycoproteome of blood serum proteins that have an affinity for PHA-L. Glycoproteins with a β 6GlcNAc branching were isolated from the depleted serum of patients with malignant and benign ovarian cancer by PHA-L lectin affinity. This glycoproteome is important since an increase in β 6GlcNAc branching has been associated with malignancy. The isolated glycoproteins were digested and analyzed using LC-MS/MS. Peptides were matched and stringently filtered which allowed for the identification of 63 unique glycoproteins in this study. Additionally, 9 proteins

were matched in malignant serum only or identified using quantification by spectral counting that could serve as potential biomarkers. Since this was a proof of principle study further studies will be need to verify the clinical potential of these glycoproteins as biomarkers.

Chapter 5

This chapter validated QUIBL (quantitative isobaric labeling) as a quantitative glycomics approach for O-linked glycans. QUIBL, which is based on permethylation of glycans with isobaric labels, proved to be applicable for O-linked glycans established by the analysis of O-linked fetuin glycans on a high resolution mass spectrometer. The reproducibility and accuracy of QUIBL for O-linked glycan was confirmed over 2 orders of magnitude and its application to a more complex sample, porcine mucin, was verified in this study.