GLYCOPEPTIDE ANALYSIS OF HIV-1 ENVELOPE PROTEIN USING HPLC/ESI-FTICR MS AND MALDI-TOF MS WITH ASIALOFETUIN ENRICHMENT AND SEPARATION TECHNIQUES

By

©2009

Qing Chang

M.S., University of Kansas 2009

Submitted to the Department of Chemistry and the
Faculty of the Graduate School of the University of Kansas
In partial fulfillment of the requirements for the degree of
Master of Science

_____
Heather Desaire (Chairperson)

_____
Michael Johnson

_____
Minae Mure

Date defended _____

The Thesis Committee for Qing Chang certifies that this is the approved Version of the following thesis:

GLYCOPEPTIDE ANALYSIS OF HIV-1 ENVELOPE PROTEIN USING HPLC/ESI-FTICR MS AND MALDI-TOF MS WITH ASIALOFETUIN ENRICHMENT AND SEPARATION TECHNIQUES

Committee:

_____
Heather Desaire (Chairperson)

_____
Michael Johnson

_____
Minae Mure

Date Approved _____

Abstract

Qing Chang, M.S.
Department of Chemistry, April 2009
University of Kansas

Glycosylation on proteins serve many biological roles and their heterogeneity can affect various biological functions. The site occupancy of the glycans as well as their various glycoforms makes glycans unique to analyze due to the fact that the variation in structure is not based on a genomic code. Glycosylation is a post-translational modification that provides diverse glycoforms.

Mass spectrometry has become a standard technique for characterizing the site occupancy and glycoforms that a protein exhibits. Methodologies in characterization of glycopeptides often include a proteolytic digest followed by purification of the digested protein with liquid chromatography. Mass spectrometers are sensitive to salts and buffer contents, thus a purification of the sample before analysis is vital for the detection of the sample.

The research presented herein studies the glycosylation coverage of two HIV-1 envelope proteins. The intention of this study is to determine the specific sites on the HIV-1 protein that are heavily masked/protected by glycans or are free of glycans, in order for future studies to correlate if exposed regions on the protein are favorable regions for vaccine design.

Another study was performed to aid in the glycosylation analysis of the HIV-1 proteins. Asialofetuin was used as a model protein to scrutinize the separation efficiencies of three various capillary columns. Specifically the separation of glycopeptides from a peptide mixture was conducted on the three columns in order to

determine which platform would optimize the separation of N-linked HIV glycopeptides from a peptide mixture.

Table of Contents

Introduction into HIV-1 and Glycopeptide Analysis

1.1 Introduction in Glycomics

Glycomics is defined as the analysis of a whole set of glycans produced in a single organism.[1] The study of glycomics is crucial because all living cells of all organisms bear glycans that are involved in many biological activities, such as cell development, differentiation or markers of cells types, morphogenesis, etc.[2] Glycans have also shown to be important in viral infections.[3]

There are no known viruses that encode within their genome the ability to biosynthesize glycans on their own.[4] Viruses are dependent on the host cell's glycosylation machinery for the glycans that are attached to their proteins.[4] The study of glycans present on viruses will lead to better understanding of viral infection and may lead to better treatment.

1.2 Types of Glycans

There are three types of glycosylation found on eukaryotic proteins (N-glycosylation, O-glycosylation, and GPI-anchoring).[1] This study focuses on N-glycosylation because its predominance in mammalian cells,[4] and particularly, its prevalence in HIV-1 Env proteins. N-glycosylation has a glycan attached to a asparagine, and this is called asparagine-linked or N-linked.[3] N-linked glycosylation occurs when glycans are attached to asparagine in a peptide chain containing a sequence Asn-X-Ser/Thr, where X is any amino acid except for proline.[3, 5] Glycosylation occurs in the endoplasmic reticulum (ER) and Golgi compartment.[5] Glycans that are not modified/processed through the ER and Golgi are called high mannose glycans.[5] Glycans that are processed in the ER and Golgi are called hybrid and complex glycans.[5] High mannose glycans have glycoforms that range from 5-9 mannose sugars.[7] Hybrid and complex glycans have glycoforms that have either three or more hexose (Hex) and four or more *N*-acetylglucosamine (HexNAc) or they have four or more hexose (Hex) and three or more *N*-acetylglucosamine (HexNAc).[7] Figure 1.1 shows a possible glycoform of a processed and unprocessed glycan.

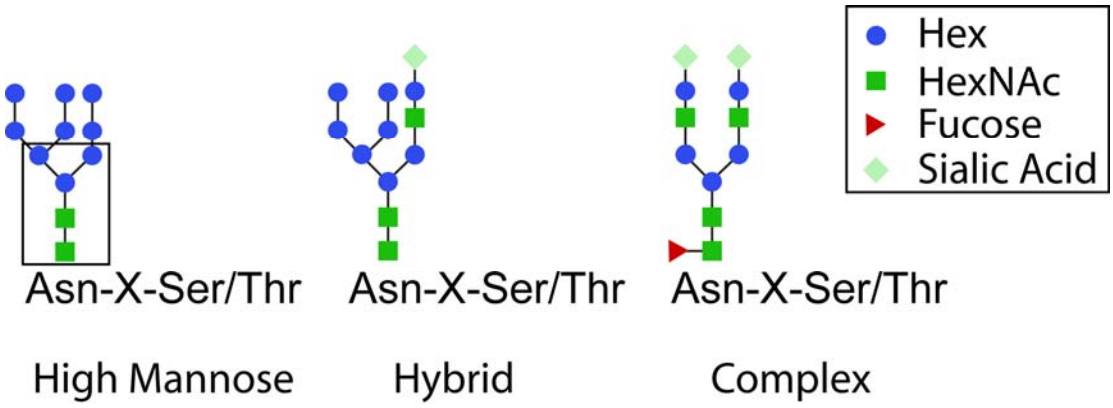

Fig. 1.1: N-linked glycoforms of both processed and unprocessed glycans

1.3 Introduction into HIV and HIV glycoproteins

Human immunodeficiency virus (HIV) infection is a pandemic that continues to increase at a rate of 14,000 new infections per day.[6] It is transmitted through blood, sexual contact, and from mother-to-infant.[6] HIV attacks a target cell by fusing the viral membrane with the plasma membrane of the target cell.[7] The viral membrane contains viral envelope spikes composed of a 120,000 dalton exterior envelope glycoprotein (gp120) which was first synthesized as an 845 to 870 amino acid precursor in the rough endoplasmic reticulum (rough ER)[7] and a 41,000 dalton transmembrane glycoprotein (gp41).[7, 8] High-mannose sugar chains are then added to form the glycoprotein gp160.[7] This polyprotein is transported to the Golgi apparatus to undergo glycosylation modification for proper folding and conformational stability.[7, 9] Within the Golgi apparatus, gp160 is cleaved by furin or equivalent endoproteases to produce gp120 (the exterior envelope glycoprotein) and gp41 (the transmembrane glycoprotein).[7, 9]

HIV initiates infection by targeting the CD4 molecule found on the surface of some CD4 T lymphocytes.[10] Through direct fusion between the virus envelope and the cell membrane[10], the gp120 first binds to CD4 to induce a conformational change that facilitates the binding of the co-receptor (CCR5 or CXCR4). When gp120 binds to the co-receptor, it induces a change in gp41 that leads to the virus fusing with the host cell.[11]

1.3.1 Synthesis of HIV proteins in mammalian cells

HIV uses the host's ER and Golgi to synthesize the glycans that attach to the Env protein. The ER synthesizes proteins and allows for the transport of proteins through the cell.[14] The majority of the synthesized proteins are glycoproteins.[15] When the N-linked glycans are first added to the polypeptide chains, they display a consensus "core glycan," which is shown in Figure 1.2.[12]

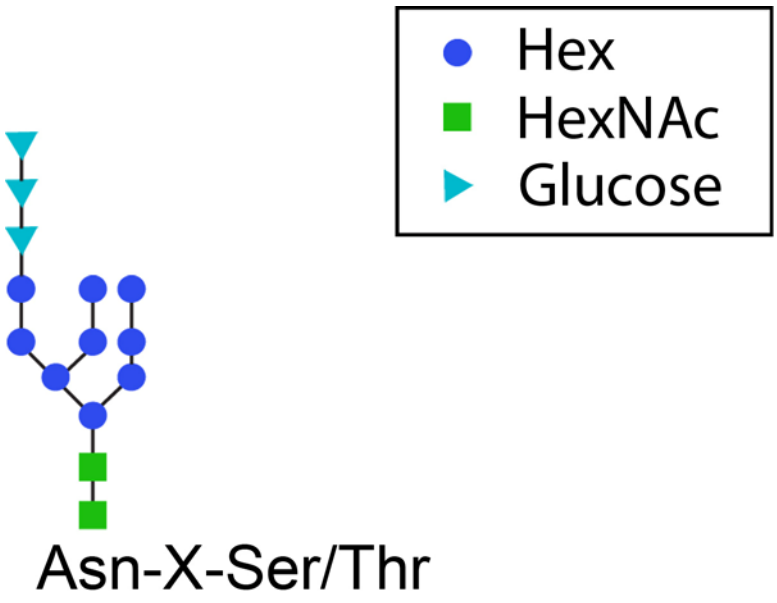

Fig. 1.2: The consensus "core glycans" of mammalian species within the ER.

The core glycans were first synthesized within the cytosol of the ER membrane.[12] Within mammalian cells, the core glycans were constructed as a membrane-bound dolichylpyrophosphate precursor by enzymes on both sides of the ER.[13] Sugars are added one at a time within the cytosol.[12] Two N-acetylglucosamines and five mannoses are added to the dolichylphosphate.[14] After

the addition of two N-acetyglucosamines and five mannoses, the oligosaccharide is placed on the lumen side of the ER membrane.[13] Within the lumen, seven sugars are then added onto to form the "core glycans".[12]

The core glycans synthesized in the ER consist of 14 saccharides.[12] Once the core oligosaccharides are assembled, the oligosaccharyltransferase (OST) transfers the core glycans onto a growing polypeptide chain by adding the N-linked glycans to the side chain nitrogen of the Asn residue with an N-glycosidic bond.[15] Every polypeptide that departs the translocon complex to the ER lumen has to be examined by the OST for the consensus sequence of Asn-X-Ser/Thr.[16] Three glucoses are cleaved by glucosidase I and glucosidase II and the terminal mannose are cleaved by mannosidases, shown in Fig. 1.3.[13]

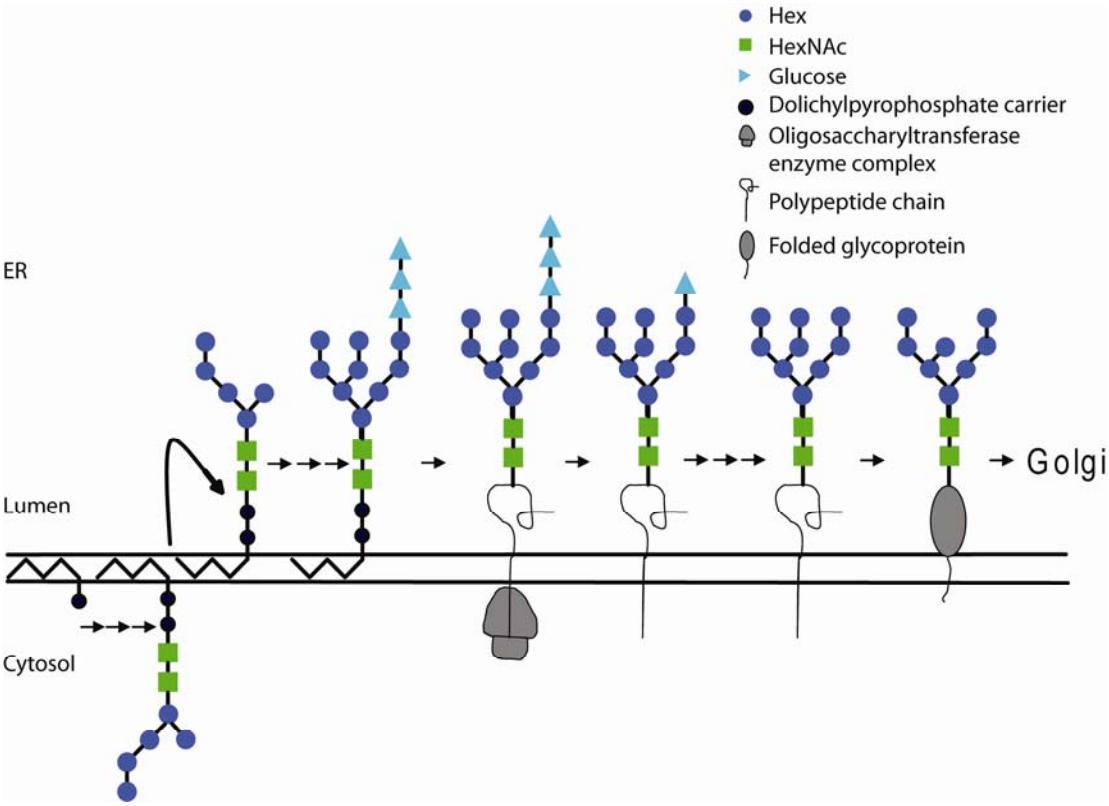

Fig. 1.3: Modification of glycans within the ER.

Glucosidase I cleaves the α-1,2 linkage of the outermost glucose linked to the middle glucose.[17] Glucosidase I in conjunction with glucosidase II prevent any binding of the protein-bound glycan to OST.[17] The glycoprotein then folds and is transported to the Golgi complex. The majority of proteins that fold in the ER gain disulfide bonds through oxidation process catalyzed by thiol-disulfide oxidoreductases.[18] If the glycoproteins undergo incomplete or improper folding, they are degraded within the ER.[18] The synthesized glycoproteins in the ER pass through the Golgi complex before leaving to their targeted functions.[18]

Glycans are also altered thru the calnexin/calreticulin cycle. Calnexin is a transmembrane protein and calreticulin is a soluble luminal ER protein.[13] Both are monomeric, calcium-binding proteins, and have ER localization signals.[19] Glucosidase I removes the first glucose and glucosidase II removes the second glucose.[13] The glycans that remain bind to calnexin or calreticulin.[13] This prevents the glycoproteins being exported before proper modifications.[13] The calnexin and calreticulin also protect any intermediates from premature degradation and also exposes them to a thiol-disulfide oxidoreductase, ERp57.[19] The ERp57 is a cofactor that aids in proper disulfide bond formation.[20] Glucosidase II releases the glycoprotein from calnexin and calreticulin.[20] This allows for the glycoprotein to leave the ER.[20]

After the glycoprotein were released from the calnexin and calreticulin, they are free to leave the ER unless recognized by a soluble enzyme, UDP-Glc:glycoprotein glucosyltransferase (GT).[20, 21] GT is a large soluble protein that is located near the exit of the ER and in the ERGolgi intermediate compartment (ERGIC).[21] GT targets incomplete folded glycoproteins for proper folding and oligomerization.[19] It is not clear how GT senses which glycoprotein is incompletely folded. One possibility is that protein may be able to sense if another protein is incompletely folded through exposed hydrophobic peptides and surface dynamics.[21] Studies have also shown that only glycopeptides with more than twelve amino acid residues and have hydrophobic residues on either the N-terminal side or C-terminal side will be recognized by GT if that glycoprotein was incompletely folded.[19] Studies

have also shown that GT does not require interactions and will recognize incomplete folding with the absence of glycans.[19]

When glycoproteins fail to fold or oligomerize, they are retained in the ER.[22] The ER-associated degradation (ERAD) prevents accumulation of misfolded proteins in the ER.[22] There are three steps in removal of misfolded proteins: recognition of a misfolded glycoprotein, retrotranslocation to the cytoplasm, and the ubiquitin-dependent degradation by the proteasome.[22] Recognition of misfolded proteins is based on a timer mechanism.[23] Degradation of misfolded proteins starts after a lag period of 30 to 90 minutes.[23] This allows for folding intermediates to have enough time to properly fold inside the ER.[23] This process is now well know, there may be other factors besides a timer delay.[23] The delay in degradation of glycoproteins by ERAD is linked to the trimming of mannoses.[24] The α-1,2 exomanosidase that removes the terminal mannose from the B branch is key in the timer mechanism.[25] Studies have show that if the mannosidase is inhibited by kifunensin, the degradation process is dramatically slower.[26] If the mannosidae I is overexpressed, then the degradation is accelerated.[26] The removal of mannose varies from 10 minutes in yeast cells to an hour in mammalian cells.[26] This lag time varies for different proteins.[26] The mannosidase is slow-acting to protect the newly synthesized proteins.[26]

Further modifications occur when the glycoproteins are transported to the Golgi complex. The Golgi apparatus modifies proteins by adding and trimming saccharides to form diverse structures of N-linked glycans shown in Fig. 1.4.[14]

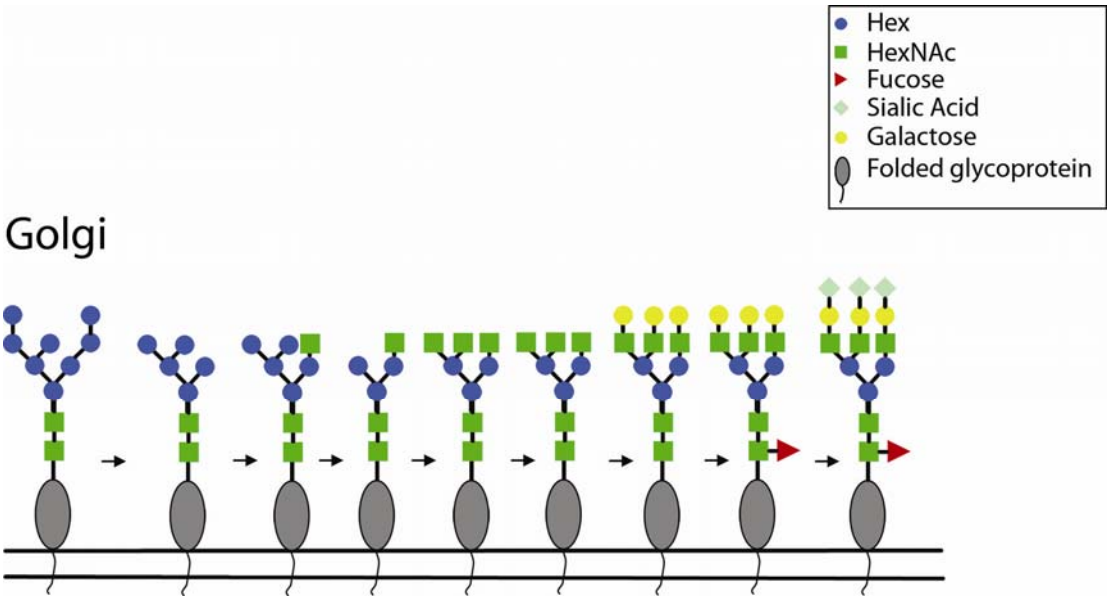

Fig. 1.4: Modifications of glycans within the Golgi Apparatus.

Sugars can be added during the terminal glycosylation stage to form complex N-linked glycans.[14] The glycoproteins are modified as they move from the entry side (cis) to the exit side (trans).[25] The average time that glycoproteins spent within the Golgi is usually between five to fifteen minutes.[25]

1.3.2 Variation of HIV Env protein sequence

The process of HIV infection is well known, yet developing a vaccine to prevent this infection is difficult, due to the protein's high diversity in genetic sequence. The diversity of HIV is characterized based on its many clades or subtypes. These clades differ in their genetic sequence and are categorized by their geographic prevalence.[27] HIV is broken down into human immunodeficiency virus type 1 (HIV-1) and human immunodeficiency virus type 2 (HIV-2). HIV-1 and HIV-2 are differentiated based on their genetic similarities to known simian immunodeficiency virus (SIV).[28] HIV-1 has the closest genetic similarity with SIVcpz, isolated from the chimpanzee subspecies *Pantroglodytes troglodytes* in Central Africa.[28, 29] HIV-2 has the closest genetic similarity with SIVsm, from sooty mangabeys in West Africa.[28, 29] These diverse set of viruses that dominate the HIV pandemic belong to HIV-1 M or "main" group.[17, 19] There are nine major clades within HIV-1 group M (A to D, F to H, J, and K).[29] Clade C is the most prevalent virus that causes more than half of all HIV infections worldwide.[30] Clade B virus is predominant in the United States and in Western Europe.[27]

Variation in sequence between different clades leads to immunological ambiguity during vaccine development. Strains of the same subtype can differ up to 20% in their envelope protein sequence, while strains of different subtypes can differ up to 35%.[31] There are many differences between clade B and clade C. Clade C is more variable, especially in the α2-helix in the C3 region, whereas clade B is more conserved in this region.[32] Studies show that the α2-helix region (composed of 18

residues in gp120 numbered 335 to 352) is more exposed in clade C.[32] The α2-helix region may be more antigenic in clade C because it is more exposed and therefore more susceptible to antibody attacks.[32]

An HIV vaccine must be capable of neutralizing different subtypes of the virus. Currently, most studies have focused on HIV-1 clade B viruses. Little is known about clade C viruses, yet they are the most prevalent subtype in the world. Few vaccine strategies focuses on the HIV-1 genetic variation.[30] These strategies predominately utilizes HIV-1 immunogens that come from viruses selected based on availability and geographic representation.[30]

1.3.3 Methods for addressing HIV sequence variation

There are currently two methods for selecting vaccine strains that address high levels of HIV sequence variation. The first method uses an isolate from one subtype selected based on their prevalence within a geographical region.[31] Essentially, a vaccine would be designed using an Indian strain to be used in India and a South African strain would be used in South Africa.[31] However, this method is limited by the scale of the epidemic within that region.[31] Strains from one region that match closely with a different region would be viewed as outliers. For instance, a strain such as IN101 from India matches more to African subtype C strains than African strains are to each other would not be used for developing vaccines in Africa since it is outside the geographical region.[33] The second method uses a consensus sequence or an ancestral sequence to minimize HIV diversity.[33] A consensus sequence is designed based on the most common amino acid in each position of the sequence.[31] An ancestral sequence is designed based on a phylogenetic tree that shows the genetic distances between potential vaccine strains.[33] This method has advantages in producing vaccines that have the potential to elicit cross-reactive response and is more economically feasible than producing multiple vaccines based on country-specific strains.[31]

1.4 Challenges for glycan analysis

While HIV sequence variation has its own challenges, analysis of glycans itself also has its unique set of challenges. Glycans are synthesized in a different manner than genes and proteins.[1] Glycoproteins of the same type may have small variations in the glycan units from each other.[3] This variation is termed microheterogeneity.[3] Studies on ovalbumin obtained from the eggs of the same hen have shown variation of glycan units from 6:5 to 5:2 ratios of mannose to *N*-acetylglucosamine.[22]

1.5 Experimental design using mass spectrometry for glycopeptide analysis

Glycan composition after modifications within the ER and Golgi apparatus can be characterized using mass spectrometry. There are two strategies for analyzing glycopeptides using mass spectrometry. One strategy is to cleave the glycans off the protein using enzymatic or chemical methods and analyzing the glycans directly.[34] Another strategy is to obtain glycopeptides in a peptide mixture using a protease digestion.[34] The second strategy is beneficial in gaining site-specific information of the glycan coverage of the protein.[34]

1.5.1 Reduction and alkylation

Glycosylated proteins are more susceptible to proteolytic enzymes when the protein's tertiary structure is distrupted.[35] Surfactants such as: urea, guanidine hydrochloride, and sodium dodecylsulfate are often used as denaturing agents to improve the unfolding and solubility of the proteins.[25] Dithiothreitol (DTT) and iodoacetamide (IAA) are often used as the reducing and alkylating reagents respectively.[36] DTT is used to promote reductive unfolding by breaking the protein's disulfide bonds.[36] IAA is used to prevent oxidative folding by preventing the protein from binding to it's native disulfide bonds.[36]

1.5.2 Enzymes used for digestion of glycopeptides

Trypsin was used for the protease digestion. Trypsin is well-known for cleaving the C-terminal side of lysine (Lys) and arginine (Arg) unless these residues are followed by a proline. Trypsin allows for the digestion of a protein into mixture of glycopeptides and peptides.

1.5.3 N-Deglycosylation with PNGase F

Site-occupancy of glycans on N-linked asparagine residue provides useful information on the glycan coverage of a protein. Knowing the amino acid sequences of the protein that were open, or lacking glycans, may aid in vaccine development. It is well-known that Protein N-glycosidase F (PNGase F) enzymatically releases N-linked glycans and converts the asparagines residue to aspartic acid causing, an observed mass shift of 1 Da as shown in Fig. 1.5.

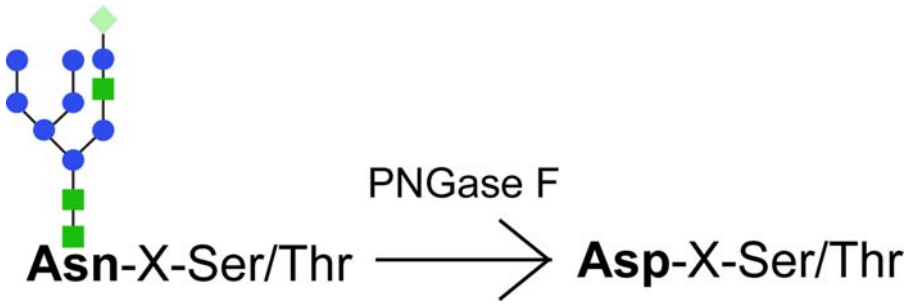

Fig. 1.5: N-Deglycosylation of N-linked glycans using PNGase F

1.5.4 Purification methods for glycopeptide analysis

Glycopeptide analysis from a proteolytic digest is problematic due to the presence of peptides along with glycopeptides in the sample. Peptides ionize better than glycopeptides and would suppress the signal intensity of the glycopeptides if both are present in the sample.[34]

A variety of methods have been proposed to isolate the glycopeptides within a peptide mixture. Lectin-affinity chromatography can isolate glycoproteins with different carbohydrate types.[1] Lectins are specific for a particular carbohydrate. For

example, Con A is specific for mannose and wheat germ agglutinin (WGA) is specific for N-Acetylglucosamine.[1] This method is useful if the glycoforms are known on the glycopeptides. Analysis of glycopeptides with an unknown variety of glycoforms would be problematic for this method. Affinity separation using cellulose or sepharose on glycopeptides from a peptide mixture is another possible approach. The affinity of the cellulose or sepharose is dependent on the hydrogen bonding between the hydroxyl group of carbohydrates on cellulose or sepharose and the glycopeptides.[37] The hydrogen bonding is not capable of completely separating the glycopeptides from the peptides.[28] High-performance anion-exchange (HPAE) chromatography with pulsed amperometric detection (PAD) has been used to separate oligosaccharides according to molecular size, sugar composition, and linkage of the monosaccharide units.[38] This technique requires high concentrations of sodium hydroxide and sodium acetate and these salt solvents have to be removed for further analysis.[39]

1.5.4.1 Reverse-phase chromatographic separation

Reverse-Phase High Performance Liquid Chromatography (rp-HPLC) is the conventional method for separation of glycopeptides after enzymatic digestion.[28] Silanes that possess alky chains (mainly C_{18}) are the standard stationary phase for reverse-phase separation.[40] The stationary phase may also contain polar alky groups to increase the retention of polar compounds.[40] Polar embedded or polar end-capped groups have been used to improve separation efficiencies of highly polar compounds.[41] However, the presence of polar functional group on the polar-embedded stationary phase reduces the hydrophobicity of the stationary phase.[41] The polar end-capped phase displays similar hydrophobic retention as the conventional C_{18} columns.[41]

1.5.4.2 Reverse-phase chromatographic separation with porous graphite carbon

Porous graphitized carbon (PGC) exhibit properties similar to reverse-phased columns, but are able to retain more hydrophilic analytes and provide separation of structurally similar oligosaccharides.[33]

PGC is composed of flat sheets of hexagonally-arranged carbon atoms.[42] London dispersion interactions hold the individual sheets of carbon atoms together.[35] More specifically, the dipole-induced-dipole interactions between the carbon atoms in adjacent sheets hold the flat carbon sheets together.[35]

The retention mechanism for PGC is due to a combination of dispersive interactions or hydrophobic interactions and charge-induced interactions.[36, 37] London dispersive interactions (as well as dipole-dipole and hydrogen bonding) between the analytes and the eluents decrease the analytes' retention on the graphite surface.[35] Similar London dispersive interactions also occur between the analytes and the graphite surface, and the net effect between the analyte-eluent dispersive interaction and analyte-graphite surface interaction determines the retention of the analyte for the graphite surface.[35, 38] Hydrophobic interactions between the analyte and the eluent will affect the retention for the graphite surface. Repulsions between the hydrophilic eluent and non-polar segments of the analytes will cause retention of the analyte on the graphite surface.[35, 37]

Charge-induced interactions force polar groups on the analyte to be closer to the graphite surface, as shown in Fig 1.6.[36, 38, 39] This is called PREG (polar retention effect on graphite).[36, 38] The flat surface reduces the retention of highly structural and

rigid analytes, which can only come in contact with the graphite surface with a portion of their own surface.[36, 38] The molecular area of an analyte that is in contact with the graphite surface largely determines the strength of the interaction.[43] Hence, planar molecules tend to come in contact with the graphite surface better than non-planar molecules, and they have a higher retention for the graphite surface than non-polar molecules.[42]

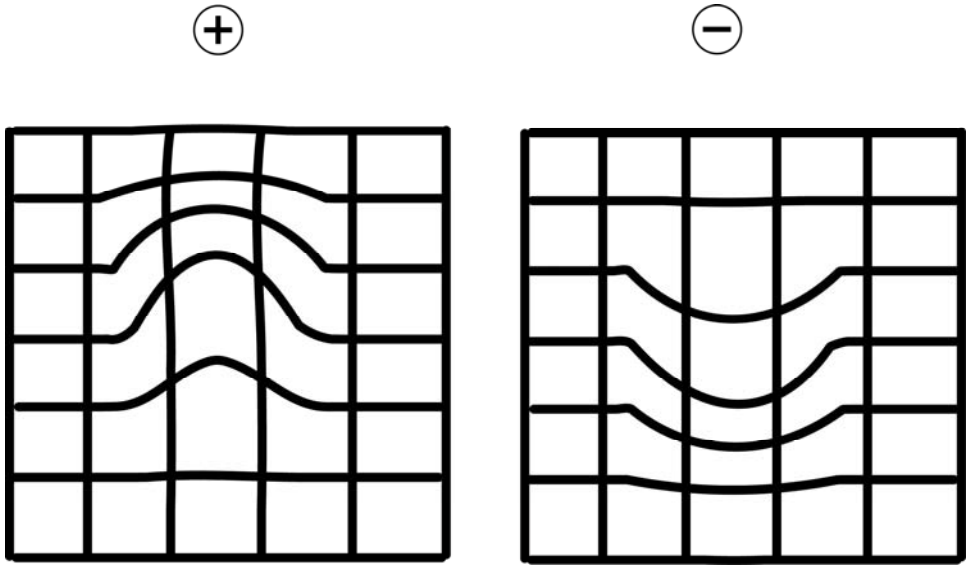

Fig. 1.6: Charge-induced interactions from polar groups of the analyte

The PGC is preferred over reverse-phase columns for highly hydrophilic compounds. In reverse-phase chromatography, the hydrophobic part of the analyte interacts with the hydrophobic stationary phase.[41] With increasing polarity of the analytes, the interaction between the analyte and solvents dominates and reduces the retention of the stationary phase.[34, 41]

The PGC can be used from pH 1 to 14 and are unaffected by acidic or alkaline conditions.[41] This allows for a wider range of solvents than reverse-phase chromatography.[34, 41]

1.5.4.3 Normal-phase chromatographic separation

Zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC) has been proposed as a tool to enrich N-linked glycopeptides in a peptide mixture.[42] A ZIC-HILIC stationary phase is based on a grafted polymeric layer with sulfoalkylbetaine zwitterionic moieties of 3-sulfopropyldimethylalkylammonio inner salts as the functional groups.[43] A silica-based sulfoalklbetaine zwitterionic phase is the least affected packing material with respect to changing pH compared to other HILIC packing such as: bare silica, aminopropyl silica, and amide silica.[44] The ZIC-HILIC is based on a partitioning mechanism with electrostatic interactions from the sulfonic group and quaternary ammonium group shown in Fig. 1.7.[43, 45]

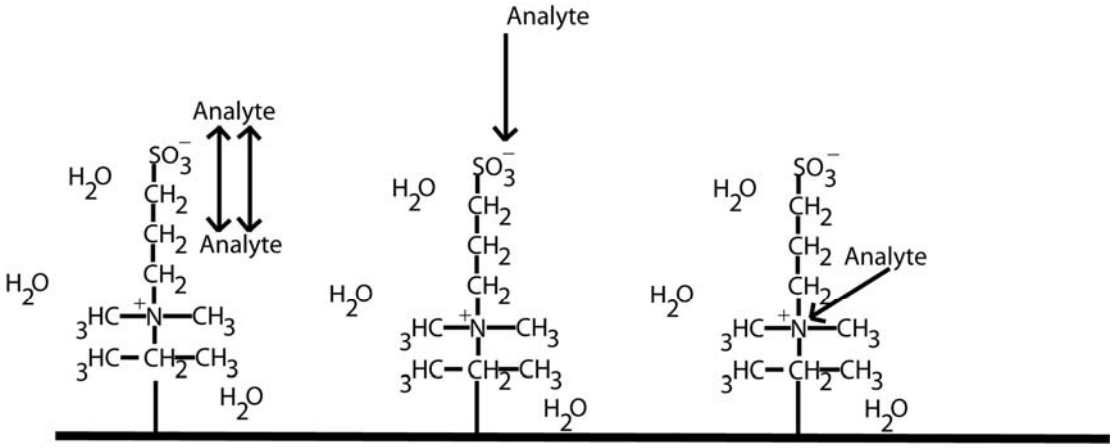

Fig. 1.7: Stationary phase of ZIC-HILIC with partitioning mechanism and electrostatic interactions with the analyte

The ZIC-HILIC stationary phase has a zero net charge due to the 1:1 ratio of quaternary ammonium and sulfonic acid group.[43] The stationary phases are very hydrophilic and easily absorb or imbibe water.[41] Having a very hydrophilic

stationary phase and a relatively hydrophobic mobile phase allows for the hydrophilic analytes to partition between the hydrophilic stationary phase and relatively hydrophobic bulk eluent.[41, 43, 46] The order of elution is the opposite of reverse-phase chromatography; the hydrophilic compounds will be retained longer than hydrophobic compounds.[43, 46] This makes HILIC an ideal tool for analysis of complex samples with poor retention towards reverse-phase chromatography.

Since the retention of compounds is based on polar interactions, the retention of glycopeptides tends to increase with an increase in the length of the glycans chains, and the elution position may provide a rough indication of the glycans size.[47]

1.5.5 Glycoform composition analysis in glycopeptides

Glycopeptide data analysis consists of identifying the peptide and carbohydrate portions of the glycopeptides. FTICR-MS and MALDI-TOF were the instrumentation used for the glycopeptide data analysis in this thesis. These techniques produced MS^1 and MS^2 data, and the CID experiments resulted in characteristic product ions of the glycan compositions.

Characteristic ions in the CID data, such as the Y_1 ion $[Peptide + 203 + H]^+$, which represents the peptide portion and an N-acetylhexosamine (HexNAc) attached at the N-linked asparagine. The Y_1 ion is shown in Fig. 1.8.

Fig. 1.8: Characteristic Y_1 ion found in CID spectrum

The Y_1 ion that resulted from the fragmentation of the glycopeptides are used to assign glycopeptide peaks from the MS data.[44] Each peak in the MS^1 spectrum that may potentially correspond to a glycopeptide is the combination of the peptide and carbohydrate portions. The *m/z* of each peak is the sum of the peptide and carbohydrates. This leads to challenges of knowing how much the peptide mass and the carbohydrate mass contribute to each *m/z* of each peak. It becomes time consuming to manually assign each peak, without knowing how much contribution the peptide mass or the carbohydrate mass has on the *m/z* of that peak. A need for

software to account for these challenges greatly facilitates assigning the glycopeptide composition.

The glycopeptide compositions in this work were assigned using two web-based programs: GlycoPep ID[44] and GlycoPep DB[45]. GlycoPep ID aids in identifying the peptide portion of the glycopeptides.[44] GlycoPep DB aids in identifying the glycan portion of the glycopeptides.[45] Both web-based programs were implemented based on their advantage over other methods. GlycoPep ID is capable of assigning the peptide portions when the proteolytic digest cleaves in a nonspecific manner.[44] GlycoPep DB only assigns carbohydrates that have been found in nature/literature.[45]

The methodology for the data analysis of glycopeptides using GlycoPep ID and GlycoPep DB is shown in Fig. 1.9.[48, 49] To use GlycoPep ID, the user inputs several parameters required to identify the peptide portion. The user inputs the protein sequence, the specific enzyme used during the digestion experiment, cysteine modification, charge state, and mass tolerance.[44]

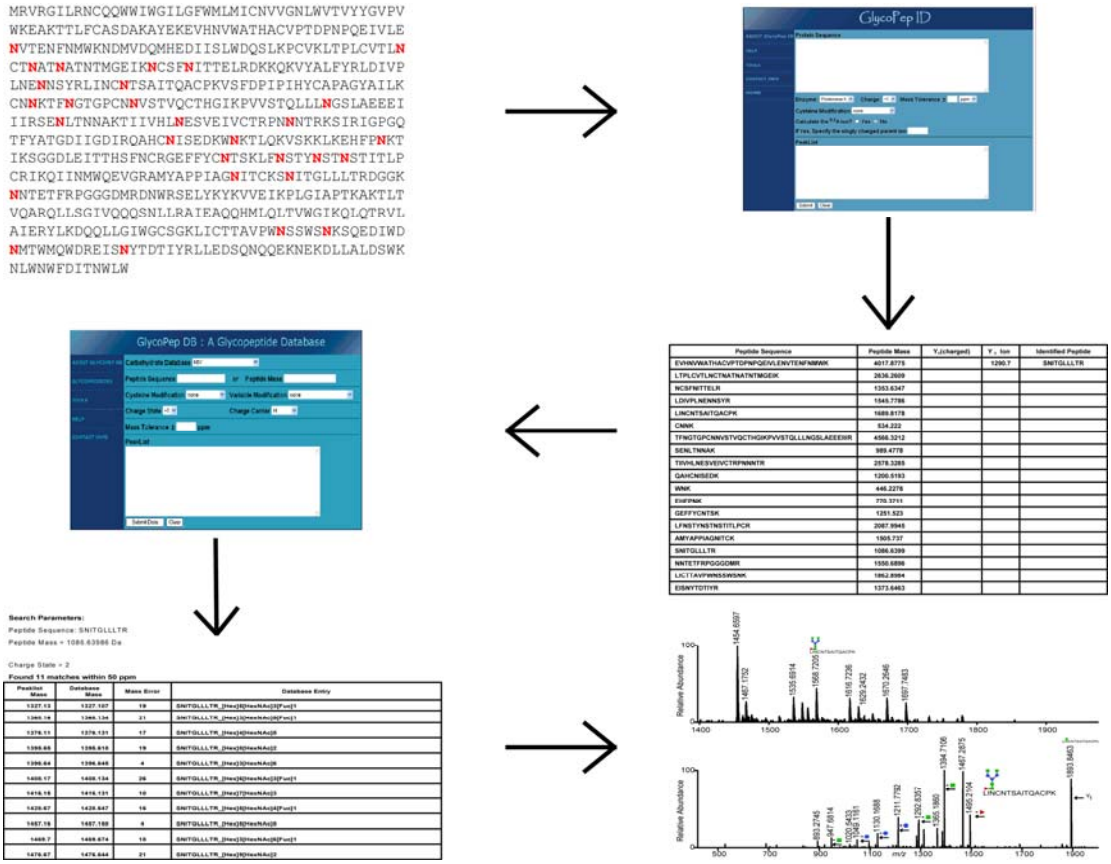

Fig. 1.9: Overview of glycopeptide analysis with GlycoPep ID and GlycoPep DB

The user also inputs the peak list from the MS^2 data.[44] GlycoPep ID is able to calculate for the characteristic Y_1 ion that occurs in MS^2 experiments.[44]

In order to determine the peptide portion of the glycopeptide, GlycoPep ID first uses the input data to determine all the possible peptides with N-linked glycosylation sites.[44] A table of all possible peptides is generated, with the glycosylation sites. This table lists all possible theoretical peptides that can occur with the specific digestion methods selected; this list of peptide masses is then compared to the experimental peak list from the MS^2 data.[44] The program matches

characteristic ions that fall within the mass tolerance specified by the user.[44] If the program matches more than one peptide, the user can repeat the search with a different characteristic ion or look manually at the spectrum to confirm which peptide portion is correct. Alternatively, both peptide possibilities could be considered in the next step, which is to determine the glycan portion using GlycoPep DB.

GlycoPep DB contains a database of 319 N-linked glycans that have been found in the literature.[45] GlycoPep DB only uses biologically relevant glycans in assigning the carbohydrate portion of the glycopeptides.[45] The user inputs the possible peptide sequence that is generated from GlycoPep ID, the charge state, the cysteine modification, any variable modifications from other amino acid residues, the charge carriers, the mass tolerance, and the experimental peak list.[45] The user also selects for a particular database of biologically relevant glycans to match their experimental peak list from the high resolution MS^1 experiment.[45] GlycoPep DB generates a list of possible glycopeptide compositions that fall within the range of the mass tolerance.[45]

The following is a summary of the data analysis process: First, the user can look at the lower mass range of an MS^2 spectrum for characteristic fragmentation product ions shown in Fig. 1.10. *m/z* 528 $[HexNAc + 2Hex + H]^+$, *m/z* 690 $[HexNAc +3Hex + H]^+$, *m/z* 893 $[2HexNAc + 3Hex + H]^+$, and *m/z* 657 [HexNAc + Hex + Sialic Acid + $H]^+$.[34]

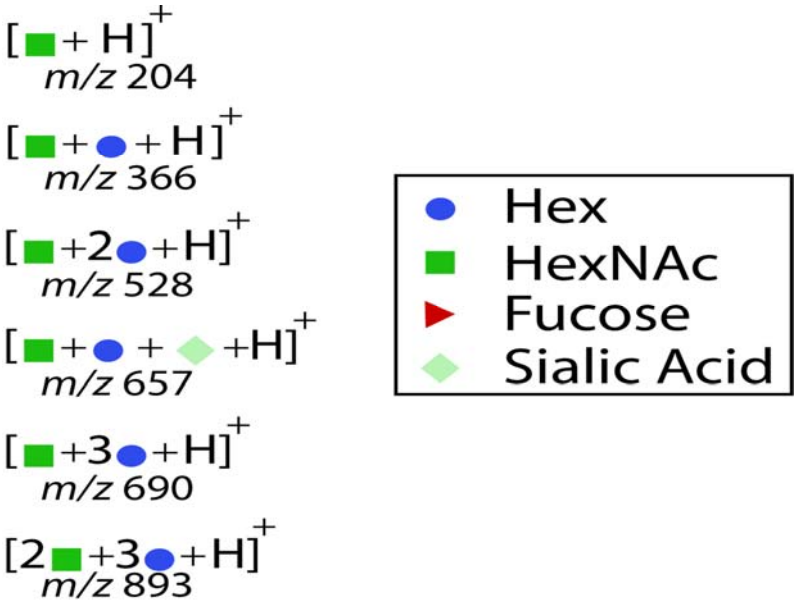

Fig. 1.10: Characteristic fragmentation product ions of glycopeptides

If these are present, this indicates the spectrum is a glycopeptide. The user can then input the MS^2 peak list that contains any of those characteristic fragmentation product ions into GlycoPep ID to assign the peptide portion of the glycopeptides.[34] The peptide that resulted from the GlycoPep ID assignment is then imputed into GlycoPep DB as well as the MS^1 peak list to generate a list of possible glycan compositions.[34] Once the plausible glycopeptide compositions have been generated using both GlycoPep ID and GlycoPep DB, the user can confirm those assignments by manually checking the MS^2 spectra for all of the peaks that have MS^2 data by looking for cleavages of neutral losses of individual sugars and cleavages along the peptide backbone.

1.6 Mass spectrometry

Since glycan composition after modifications within the ER and Golgi apparatus is characterized using mass spectrometry, a short review of mass spectrometry is provided here. Mass spectrometry (MS) is composed of three fundamental parts: ion source, mass analyzer, and detector. The ion source allows for the conversion of analytes to ions.[46] The ions are separated based on their *m/z* in the mass analyzer.[46] The detector converts the ions into electrical signals.[46]

1.7 Electrospray ionization

Electrospray ionization (ESI) was used on the HIV Env and alpha(1)-acid glycoprotein (AGP). ESI converts a liquid sample into a gaseous sample.[46] It generates charged droplets within an electrical field.[46] Solvent from the charged droplets start to evaporate when heat, dry gas, or both are applied to the charged droplets in atmospheric pressure.[46] Ions are created from the charged droplets when the Coulombic repulsion force reaches a point where the charge density on the surface exceeds the decreasing surface tension the charged droplets (Fig. 1.11).[46] The charged molecules on the surface of the ions under a strong electric field causes the desorption of the ions.[47] The sensitivity increases with higher concentration of charged molecules on the surface of the droplets.[47]

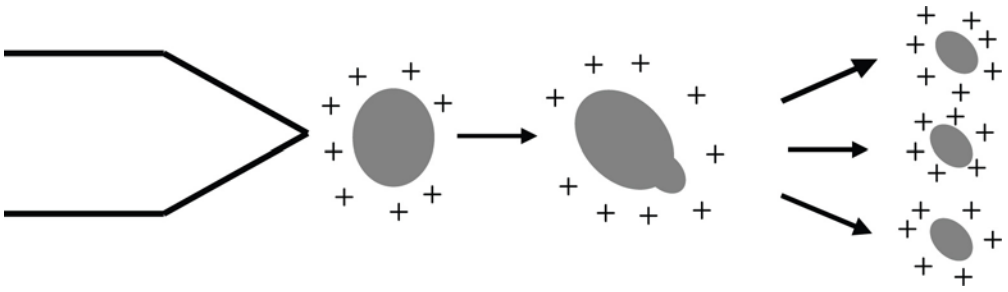

Fig. 1.11: Coulombic repulsion forces of ions created from charged droplets. When the charge density on the surface exceeds the size of the charged droplets, ions are created from the charged droplets.

1.8 Fourier Transform Ion Cyclotron Resonance

Fourier transform mass spectrometry (FTMS) was first introduced by Comisarow and Marshall in 1974.[47] FTMS is an ion trapping instrument that employs an analyzer cell to store and detect ions.[48] The cubic analyzer cell is a commonly used analyzer cell and is shown in Fig. 1.12.[49] Ions are excited by applying sinusoidal voltage to the emitter plate.[48] The sinsoidal voltage or radiofrequency voltage creates an electric field where ions that are in resonance with the radiofrequency electric field will spiral outwards until they strike the detection plate or receiver plate.[48, 49] Ions that are not in resonance with the radiofrequency electric field will remain at the center of the cubic cell.[48, 49]

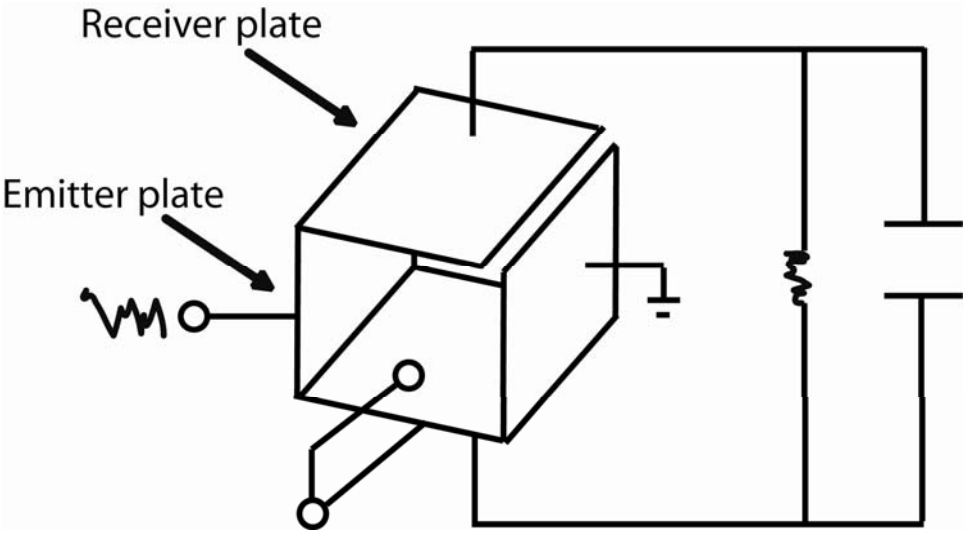

Fig. 1.12: Depiction of a typical FT-ICR cell. Ions within the cell are excited by the emitter plate and have different characteristic cyclotronic frequencies, depending on their mass. This gives them a unique image current, which is detected by the receiver plate.

Ions with the same mass-to-charge ratio undergo a cyclotron motion and are excited coherently.[48] The cyclotron motion is described by Newton's Second Law of Motion, which states that the force of an object is equal to the mass of the object times its acceleration. Trajectories of ions in a spatially uniform magnetic field (B) are curved by the magnetic field.[47] The moving ions are subject to a force with m as the ionic mass, q as the charge, and V as the velocity described in Eq. (1).[49]

$$\text{Force} = \text{mass} \cdot \text{acceleration} = m \cdot dV / dt = qV \times B \quad (1)$$

The trajectory of the ion is curved due to the magnetic field perpendicular to its original trajectory. Hence, the force is equal to the charge multiplied by the velocity times the vector cross product of the magnetic field.[49] Angular acceleration dV / dt are affected by the magnetic field bending the ion path into a circle with a radius r and changes with the ion's velocity v which is described in Eq. (2)[49]

$$\left| dV / dt \right| = V^2 / r \quad (2)$$

The force is described by replacing Eq. (2) into Eq. (1) shown in Eq. (3).

$$(mV^2) / r = qVB \text{ or } (mV) / r = qB \quad (3)$$

The angular velocity ω is described in Eq. (4) with a $2\pi \cdot r$ as its circular trajectory with a frequency v.

$$\omega = 2\pi v = V / r \quad (4)$$

Plugging Eq. (4) into Eq. (3) gives Eq. (5), which shows the angular velocity is independent of the velocity of the ion.[49]

$$m\omega^2 r = qB\omega r \quad (5)$$

$$\omega = (q/m)B \quad (5a)$$

A pulsed radio frequency is used to excite all of the ions orbiting in a magnet field.[46] Ions with different masses have different characteristic cyclotronic frequencies.[47] All ions are simultaneously accelerated giving each *m/z* a unique image current.[46] The image current is Fourier-transformed to a frequency-dependent intensity corresponding to the ion's *m/z*.[46, 47]

1.9 Matrix-Assisted Laser Desorption/Ionization with Time-of-Flight

Matrix-assisted laser desorption/ionization (MALDI) was created by Tanaka, Karas, and Hillenkamp in 1988.[46] An organic molecule in solution such as 2,5-Dihydroxybenzoic acid (DHB) is used as the matrix and the analyte is mixed with the matrix.[47] An UV laser pulse causes the ionization of the matrix along with ionization of the analyte and transfer of it into the gas phase.[46] The matrix absorbs most of the energy from the laser pulse to reduce sample damage.[46] The matrix has an absorption that matches the laser frequency.[47] Analytes with a variety of molecular weights including those exceeding 100,000 Da can be ionized using MALDI.[47] Typically, MALDI observes singly charged molecular species.[47]

The time-of-flight mass spectrometer uses a reflectron to increase the resolution, by increasing the amount of time for ions to reach the detector.[46] The reflectron creates a retarding field and deflects the ions back through the flight tube.[47] Ions that have the same *m/z* but different kinetic energies will reach the detector at the same time.[46, 47]

1.10 Detectors

Electron multipliers are the most common type of detectors. They generate the emission of multiple secondary particles, as shown in Fig. 1.13.[47] Electron multipliers are a series of dynodes with increasing potential.[46] Ions that strike the first dynode emit electrons.[46] The electrons are attracted to the next higher potential dynode and more secondary electrons are produced.[46]

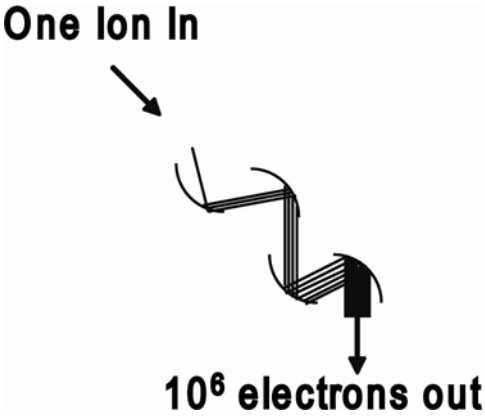

Fig.1.13 Depiction of a typical electron multiplier. A cascade of electrons is produced from a series of dynodes at increasing potentials.

1.11 Other methods for glycopeptide analysis

There are three alternative approaches to determine glycan composition on a glycopeptide. The first approach is an instrumental method, to use mass spectrometry to characterize glycans.[1] This approach has already been covered in this chapter.

The second approach is centers around differences in chemical interactions, using 2-D/3-D mapping to characterize the chemical properties such as: charge, molecular size, and hydrophobicity.[1] Takahashi and coworkers have developed a 2-D/3-D mapping using anion-exchange, normal-phase, and reverse-phase chromatography of pyridylaminated derivatives-oligosaccharides.[50] Reducing ends of glycans are labeled with 2-aminopyridine as a fluorophore.[50] The glycan elution data from the three different columns are places as separate axes to produce a three-dimensional glycan mapping.[39] Disadvantages of this method include time-consuming experiments (over 80 mins per analysis) and lack of reproducible data.[1]

The third approach is a biochemical approach, using frontal affinity chromatography (FAC) to determine dissociation constants (K_d) or association constants (K_a) between two biomolecules.[1] FAC works by placing excess volume of an analyte into a column with an immobilized affinity ligand.[1] The initial concentration of the analyte is known and if the analyte has any affinity for the immobilized ligand, the elution of the analyte would be retarded and the difference would reflect the affinity of the analyte to the immobilized ligand.[1] Disadvantages of this technique include requiring a large amount of the analyte for this analysis, and this method does not detect any glycans that are not recognized by any lectins.[1]

1.12 Overview and summary of each chapter

There were two main studies presented in this research. The overall purpose of these studies was to characterize the glycoforms present on two different HIV-1 envelope proteins and to determine their glycosylation coverage. The first study was to characterize the glycoforms and site occupancy of N-linked glycopeptides on two clade C HIV-1 envelope proteins. The second study was to identify a separation method corresponding to improving the identification of glycoforms for the HIV study. Both studies used chromatography and mass spectrometry as the analytical techniques to procure the data needed for the glycosylation analysis.

Chapter 2

Glycosylation on proteins impacts a variety of biological and cellular processes. They are involved in processes such as: cell development, morphogenesis, tumor metastasis, infection, etc.[2] Glycans on a viral glycoprotein assist in various viral infections.[51] The location of glycans on a viral glycoprotein can also affect the efficiency of a viral infection.[4] Roughly fifty percent mass of the HIV-1 envelope protein is contributed by the glycans.[4] It has been well established that glycans play a significant role in concealing and assisting HIV in escaping neutralizing antibodies.[4]

The design of this study was to determine the glycosylation coverage on the two clade C HIV-1 envelope proteins. Having a better understanding of the types of glycoforms and the locations of those glycoforms on the HIV-1 envelope protein may

assist future research in developing a neutralizing antibody to target the susceptible glycosylation regions.

Proteolytic digest of Con C and VRC_C were analyzed with MALDI-TOF MS and HPLC/ESI-FTICR MS to ascertain the glycosylation coverage on the two clade C HIV-1 envelope proteins. The determination from both mass spectrometers revealed predominately high-mannose glycoforms throughout the glycosylation coverage of both clade C envelope proteins. Complex and hybrid glycoforms were rarely seen throughout the glycosylation coverage. This may be due to the separation efficiencies of the C18 column to elute glycopeptides from a peptide mixture.

Chapter 3

Reverse-phased chromatography is the standard separation technique used for carbohydrate analysis. A model experiment was designed to inspect the separation efficiencies of glycopeptides from a peptide mixture with various columns (Aquasil C18, Hypercarb, and ZIC-HILIC) each with its unique retention mechanisms. Asialofetuin was used as the model protein because its glycosylation sites have been well characterized with mass spectrometry.[52] The intent of this study was to analyze the N-linked glycosylation sites on asialofetuin and differentiate which unique retention mechanism was the most desirable in separating glycopeptides from a peptide mixture, in order to improve the glycopeptides coverage for HIV-1 glycopeptides.

A proteolytic digest, similar to the HIV-1 digest, was accomplished on asialofetuin. The proteolyic glycopeptides were detected using HPLC/ESI-FTICR

MS, this approach does not require a rp-HPLC fractionation step, which would be necessary if MALDI-TOF MS was selected as the mass spectrometer. The ZIC-HILIC column was the only column to separate all three N-linked glycoforms on asialofetuin.

1.13 References

1. Hirabayashi, J.; Kasai, K., Separation technologies for glycomics. *J. Chromatogr. B* **2002,** 771, (1-2), 67-87.
2. Varki, A., Biological Roles Of Oligosaccharides - All Of The Theories Are Correct. *Glycobiology* **1993,** 3, (2), 97-130.
3. Spiro, R. G., Glycoproteins. *Annu. Rev. Biochem.* **1970,** 39, 599-&.
4. Rademacher, T. W.; Parekh, R. B.; Dwek, R. A., Glycobiology. *Annu. Rev. Biochem.* **1988,** 57, 785-838.
5. Dell, A.; Morris, H. R., Glycoprotein structure determination mass spectrometry. *Science* **2001,** 291, (5512), 2351-2356.
6. Takebe, Y.; Kusagawa, S.; Motomura, K., Molecular Epidemiology Of HIV: Tracking AIDS Pandemic. *Pediatr Int.* **2004,** 46, (2), 236-244.
7. Wyatt, R.; Sodroski, J., The HIV-1 Envelope Glycoproteins: Fusogens, Antigens, And Immunogens. *Science* **1998,** 280, (5371), 1884-1888.
8. Arthur, L. O.; Pyle, S. W.; Nara, P. L.; Bess, J. W.; Gonda, M. A.; Kelliher, J. C.; Gilden, R. V.; Robey, W. G.; Bolognesi, D. P.; Gallo, R. C.; Fischinger, P. J., Serological Responses In Chimpanzees Inoculated With Human-Immunodeficiency-Virus Glycoprotien (GP120) Subunit Vaccine. *Proc. Natl. Acad. Sci. U.S.A.* **1987,** 84, (23), 8583-8587.
9. Pantophlet, R.; Burton, D. R., gp120: Target For Neutralizing HIV-1 Antibodies. *Annu Rev Immunol.* **2006,** 24, 739-769.
10. Sattentau, Q. J.; Moore, J. P., Conformational-Changes Induced In The Human-Immunodeficiency-Virus Envelope Glycoprotein By Soluble CD4 Binding. *J Exp Med.* **1991,** 174, (2), 407-415.
11. Zhang, J.; Rao, E.; Dioszegi, M.; Kondru, R.; DeRosier, A.; Chan, E.; Schwoerer, S.; Cammack, N.; Brandt, M.; Sankuratri, S.; Ji, C. H., The Second Extracellular Loop Of CCR5 Contains The Dominant Epitopes For Highly Potent Anti-Human Immunodeficiency Virus Monoclonal Antibodies. *Antimicrob Agents Chemother.* **2007,** 51, (4), 1386-1397.
12. Kornfeld, R.; Kornfeld, S., Assembly Of Asparagine-Linked Oligosaccharides. *Annu. Rev. Biochem.* **1985,** 54, 631-664.
13. Helenius, A.; Aebi, M., Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **2004,** 73, 1019-1049.
14. Helenius, A.; Aebi, M., Intracellular functions of N-linked glycans. *Science* **2001,** 291, (5512), 2364-2369.
15. Burda, P.; Aebi, M., The dolichol pathway of N-linked glycosylation. *Biochim. Biophys. Acta-Gen. Subj.* **1999,** 1426, (2), 239-257.
16. Imperiali, B.; Rickert, K. W., Conformational Implications Of Asparagine-Linked Glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **1995,** 92, (1), 97-101.
17. Elbein, A. D., Inhibitors Of Glycoprotein-Synthesis. *Methods Enzymol.* **1983,** 98, 135-154.
18. De Praeter, C. M.; Gerwig, G. J.; Bause, E.; Nuytinck, L. K.; Vliegenthart, J. F. G.; Breuer, W.; Kamerling, J. P.; Espeel, M. F.; Martin, J. J. R.; De Paepe, A. M.;

Chan, N. W. C.; Dacremont, G. A.; Van Coster, R. N., A novel disorder caused by defective biosynthesis of N-linked oligosaccharides due to glucosidase I deficiency. *Am. J. Hum. Genet.* **2000,** 66, (6), 1744-1756.
19. Moremen, K. W.; Trimble, R. B.; Herscovics, A., Glycosidases Of The Asparagine-Linked Oligosaccharide Processing Pathway. *Glycobiology* **1994,** 4, (2), 113-125.
20. Hammond, C.; Braakman, I.; Helenius, A., Role Of N-Linked Oligosaccharide Recognition, Glucose Trimming, And Calnexin In Glycoprotein Folding And Quality-Control. *Proc. Natl. Acad. Sci. U. S. A.* **1994,** 91, (3), 913-917.
21. D'Alessio, C.; Fernandez, F.; Trombetta, E. S.; Parodi, A. J., Genetic evidence for the heterodimeric structure of glucosidase II - The effect of disrupting the subunit-encoding genes on glycoprotein folding. *J. Biol. Chem.* **1999,** 274, (36), 25899-25905.
22. Klausner, R. D.; Sitia, R., Protein-Degradation In The Endoplasmic-Reticulum. *Cell* **1990,** 62, (4), 611-614.
23. Jakob, C. A.; Burda, P.; Roth, J.; Aebi, M., Degradation of misfolded endoplasmic reticulum glycoproteins in Saccharomyces cerevisiae is determined by a specific oligosaccharide structure. *J. Cell Biol.* **1998,** 142, (5), 1223-1233.
24. Herscovics, A., Structure and function of Class I alpha 1,2-mannosidases involved in glycoprotein synthesis and endoplasmic reticulum quality control. *Biochimie* **2001,** 83, (8), 757-762.
25. Bause, E., Structural Requirements Of N-Glycosylation Of Proteins - Studies With Proline Peptides As Conformational Probes. *Biochem. J.* **1983,** 209, (2), 331-336.
26. Jaenicke, R., Protein Folding - Local Structures, Domains, Subunits, And Assemblies. *Biochemistry* **1991,** 30, (13), 3147-3161.
27. Grossman, Z.; Vardinon, N.; Chemtob, D.; Alkan, M. L.; Bentwich, Z.; Burke, M.; Gottesman, G.; Istomin, V.; Levi, I.; Maayan, S.; Shahar, E.; Schapiro, J. M., Genotypic Variation Of HIV-1 Reverse Transcriptase And Protease: Comparative Analysis Of Clade C And Clade B. *Aids* **2001,** 15, (12), 1453-1460.
28. Korber, B.; Muldoon, M.; Theiler, J.; Gao, F.; Gupta, R.; Lapedes, A.; Hahn, B. H.; Wolinsky, S.; Bhattacharya, T., Timing The Ancestor Of The HIV-1 Pandemic Strains. *Science* **2000,** 288, (5472), 1789-1796.
29. Choisy, M.; Woelk, C. H.; Guegan, J. F.; Robertson, D. L., Comparative Study Of Adaptive Molecular Evolution In Different Human Immunodeficiency Virus Groups And Subtypes. *J Virol.* **2004,** 78, (4), 1962-1970.
30. Kothe, D. L.; Li, Y. Y.; Decker, J. M.; Bibollet-Ruche, F.; Zammit, K. P.; Salazar, M. G.; Chen, Y. L.; Weng, Z. P.; Weaver, E. A.; Gao, F.; Haynes, B. F.; Shaw, G. M.; Korber, B. T. M.; Hahn, B. H., Ancestral And Consensus Envelope Immunogens For HIV-1 Subtype C. *Virology* **2006,** 352, (2), 438-449.
31. Gaschen, B.; Taylor, J.; Yusim, K.; Foley, B.; Gao, F.; Lang, D.; Novitsky, V.; Haynes, B.; Hahn, B. H.; Bhattacharya, T.; Korber, B., AIDS - Diversity Considerations In HIV-1 Vaccine Selection. *Science* **2002,** 296, (5577), 2354-2360.

32. Gnanakaran, S.; Lang, D.; Daniels, M.; Bhattacharya, T.; Derdeyn, C. A.; Korber, B., Clade-Specific Differences Between Human Immunodeficiency Virus Type 1 Clades B And C: Diversity And Correlations In C3-V4 Regions Of gp120. *J Virol.* **2007,** 81, (9), 4886-4891.
33. Schultz, A. M.; Bradac, J. A., The HIV vaccine pipeline, from preclinical to phase III. *Aids* **2001,** 15, S147-S158.
34. Irungu, J.; Go, E. P.; Zhang, Y.; Dalpathado, D. S.; Liao, H. X.; Haynes, B. F.; Desaire, H., Comparison of HPLC/ESI-FTICR MS versus MALDI-TOF/TOF MS for glycopeptide analysis of a highly glycosylated HIV envelope glycoprotein. *J. Am. Soc. Mass. Spectrom.* **2008,** 19, (8), 1209-1220.
35. Hale, J. E.; Butler, J. P.; Gelfanova, V.; You, J. S.; Knierman, M. D., A simplified procedure for the reduction and alkylation of cysteine residues in proteins prior to proteolytic digestion and mass spectral analysis. *Anal. Biochem.* **2004,** 333, (1), 174-181.
36. Wedemeyer, W. J.; Welker, E.; Narayan, M.; Scheraga, H. A., Disulfide bonds and protein folding. *Biochemistry* **2000,** 39, (23), 7032-7032.
37. Wada, Y.; Tajiri, M.; Yoshida, S., Hydrophilic affinity isolation and MALDI multiple-stage tandem mass spectrometry of glycopeptides for glycoproteomics. *Anal. Chem.* **2004,** 76, (22), 6560-6565.
38. Lee, J.; Kim, J.; Hyeon, T., Recent progress in the synthesis of porous carbon materials. *Adv. Mater.* **2006,** 18, (16), 2073-2094.
39. Davies, M.; Smith, K. D.; Harbin, A. M.; Hounsell, E. F., High-Performance Liquid-Chromatography Of Oligosaccharide Alditols And Glycopeptides On A Graphitized Carbon Column. *J. Chromatogr.* **1992,** 609, (1-2), 125-131.
40. Ikegami, T.; Hara, T.; Kimura, H.; Kobayashi, H.; Hosoya, K.; Cabrera, K.; Tanaka, N., Two-dimensional reversed-phase liquid chromatography using two monolithic silica C18 columns and different mobile phase modifiers in the two dimensions. *J. Chromatogr. A* **2006,** 1106, (1-2), 112-117.
41. Layne, J., Characterization and comparison of the chromatographic performance of conventional, polar-embedded, and polar-endcapped reversed-phase liquid chromatography stationary phases. *J. Chromatogr. A* **2002,** 957, (2), 149-164.
42. Koizumi, K., High-performance liquid chromatographic separation of carbohydrates on graphitized carbon columns. *J. Chromatogr. A* **1996,** 720, (1-2), 119-126.
43. Ross, P., The role of porous graphitic carbon in HPLC. *Lc Gc N. Am.* **2000,** 18, (1), 14-+.
44. Irungu, J.; Go, E. P.; Dalpathado, D. S.; Desaire, H., Simplification of mass spectral analysis of acidic glycopeptides using GlycoPep ID. *Anal. Chem.* **2007,** 79, (8), 3065-3074.
45. Go, E. P.; Rebecchi, K. R.; Dalpathado, D. S.; Bandu, M. L.; Zhang, Y.; Desaire, H., GlycoPep DB: A Tool For Glycopeptide Analysis Using A "Smart Search". *Anal. Chem.* **2007,** 79, (4), 1708-1713.
46. Siuzdak, G., *The Expanding Role of Mass Spectrometry in Biotechnology*. 2 ed.; MCC Press: 2006.

47. Stroobant, E. d. H. a. V., *Mass Spectrometry Principles and Applications*. 2 ed.; John Wiley & Sons, LTD: 2002.
48. Amster, I. J., Fourier transform mass spectrometry. *J. Mass Spectrom.* **1996,** 31, (12), 1325-1337.
49. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S., Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* **1998,** 17, (1), 1-35.
50. Takahashi, N., Three-dimensional mapping of N-linked oligosaccharides using anion-exchange, hydrophobic and hydrophilic interaction modes of high-performance liquid chromatography. *J. Chromatogr. A* **1996,** 720, (1-2), 217-225.
51. Wang, X. D.; Li, W. Y.; Rasmussen, H. T., Orthogonal method development using hydrophilic interaction chromatography and reversed-phase high-performance liquid chromatography for the determination of pharmaceuticals and impurities. *J. Chromatogr. A* **2005,** 1083, (1-2), 58-62.
52. Yet, M. G.; Chin, C. C. Q.; Wold, F., The Covalent Structure Of Individual N-Linked Glycopeptides From Ovomucoid And Asialofetuin. *J. Biol. Chem.* **1988,** 263, (1), 111-117.

Chapter 2

Glycosylation Coverage of HIV-1 Envelope Proteins

2.1 Introduction

Carbohydrates play an important role in HIV's ability to resist neutralization.[1] In addition, different strains of HIV-Env likely have different carbohydrate profiles. Information about variations in the carbohydrates among different HIV strains is needed for vaccine development. Few studies have analyzed glycosylation on clade C strains. Here, we look at the glycosylation of two different clade C strains (Con C and VRC_C) in the hopes that this will improve immunogenicity studies, by investigating how HIV uses different carbohydrates to defend against neutralization.

The work presented herein focuses on comparing the glycosylation on a consensus (Con C) glycoprotein with a wild-type (VRC_C) glycoprotein and analyzing trends between their glycosylation patterns using mass spectrometry. By identifying similarly conserved glycosylation patterns between a consensus clade C strain and a wild-type clade C strain, vaccine strategies can be further improved upon.

Con C Env sequences were synthesized from the 2001 Los Alamos HIV-1 sequence database. Con C was constructed using the most common amino acid at each position in the protein alignment.[2, 3] For the hypervariable regions, only the minimal lengths of consensus were retained.[2, 3] The VRC_C Env sequence is a wild-type clade C strain that has retained the wild-type variable loop sequences and was used for comparison with Con C. Both proteins were expressed in 293T cells by collaborators at Duke University Human Vaccine Institute.

There are potentially 29 N-linked glycosylation sites on Con C and 24 potentially N-linked glycosylation sites on VRC_C, shown in Fig. 2.1 and Fig. 2.2 respectively.

MRVRGILRNCQQWWIWGILGFWMLMICNVVGNLWVTVYYGVPV
WKEAKTTLFCASDAKAYEKEVHNVWATHACVPTDPNPQEIVLE
NVTENFNMWKNDMVDQMHEDIISLWDQSLKPCVKLTPLCVTLN
CTNATNATNTMGEIKNCSFNITTELRDKKQKVYALFYRLDIVP
LNENNSYRLINCNTSAITQACPKVSFDPIPIHYCAPAGYAILK
CNNKTFNGTGPCNNVSTVQCTHGIKPVVSTQLLLNGSLAEEEI
IIRSENLTNNAKTIIVHLNESVEIVCTRPNNNTRKSIRIGPGQ
TFYATGDIIGDIRQAHCNISEDKWNKTLQKVSKKLKEHFPNKT
IKSGGDLEITTHSFNCRGEFFYCNTSKLFNSTYNSTNSTITLP
CRIKQIINMWQEVGRAMYAPPIAGNITCKSNITGLLLTRDGGK
NNTETFRPGGGDMRDNWRSELYKYKVVEIKPLGIAPTKAKTLT
VQARQLLSGIVQQQSNLLRAIEAQQHMLQLTVWGIKQLQTRVL
AIERYLKDQQLLGIWGCSGKLICTTAVPWNSSWSNKSQEDIWD
NMTWMQWDREISNYTDTIYRLLEDSQNQQEKNEKDLLALDSWK
NLWNWFDITNWLW

Fig. 2.1 Sequence of Con C. The N-linked glycosylation sites are in red.

MRDRGIPRNWPQWWMWGILGFWMIIICRVVGNMWVTVYYGVPV
WTDAKTTLFCASDTKAYDREVHNVWATHACVPTDPNPQEIVLE
NVTENFNMWKNDMVDQMHEDIISLWDQSLKPCVKLTPLCVTLH
CTNATFKNNVTNDMNKEIRNCSFNTTTEIRDKKQQGYALFYRP
DIVLLKENRNNSNNSEYILINCNASTITQACPKVNFDPIPIHY
CAPAGYAILKCNNKTFSGKGPCNNVSTVQCTHGIKPVVSTQLL
LNGSLAEKEIIIRSENLTDNVKTIIVHLNKSVEIVCTRPNNNT
RKSMRIGPGQTFYATGDIIGDIRQAYCNISGSKWNETLKRVKE
KLQENYNNNKTIKFAPSSGGDLEITTHSFNCRGEFFYCNTTRL
FNNNATEDETITLPCRIKQIINMWQGVGRAMYAPPIAGNITCK
SNITGLLLVRDGGEDNKTEEIFRPGGGNMKDNWRSELYKYKVI
ELKPLGIAPTGAKLTVQARQLLSSIVQQQSNLLRAIEAQQHML
QLTVWGIKQLQTRVLAIERYLKDQQLEIWNNMTWMEWDREISN
YTDTIYRLLEDSQTQQEKNEKDLLALDSWKNLWSWFDISNWLW

Fig. 2.2 Sequence of VRC_C. The N-linked glycosylation sites are in red.

2.2 Experimental Design

Con C and VRC_C proteins were expressed and purified from the Duke Human Vaccine Research Institute in Durham, NC.[4] Ammonium bicarbonate (NH_4HCO_3), trizma hydrochloride, trizma base, urea, ethylenediaminetetraacetic acid (EDTA), dithiothreitol (DTT), iodoacetamide (IAA), HPLC grade acetonitrile (ACN), formic acid, 2,5-dihydroxybenzoic acid (DHB), and α-cyano-4-hydroxycinnamic acid (CHCA) were all purchased from Sigma-Aldrich (St. Louis, MO). Proteomics grade trypsin was purchased from Promega (Madison, WI). N-Glycosidase F (PNGase F) was acquired from *Elizabethkingia meningosepticum* and purchased from CalBioChem (San Diego, CA). Purified water was obtained using a Millipore Direct-Q3 Water Purification System (Billerica, MA).

The overall experiment is shown in Fig. 2.3 with explanations in the proceeding paragraphs.

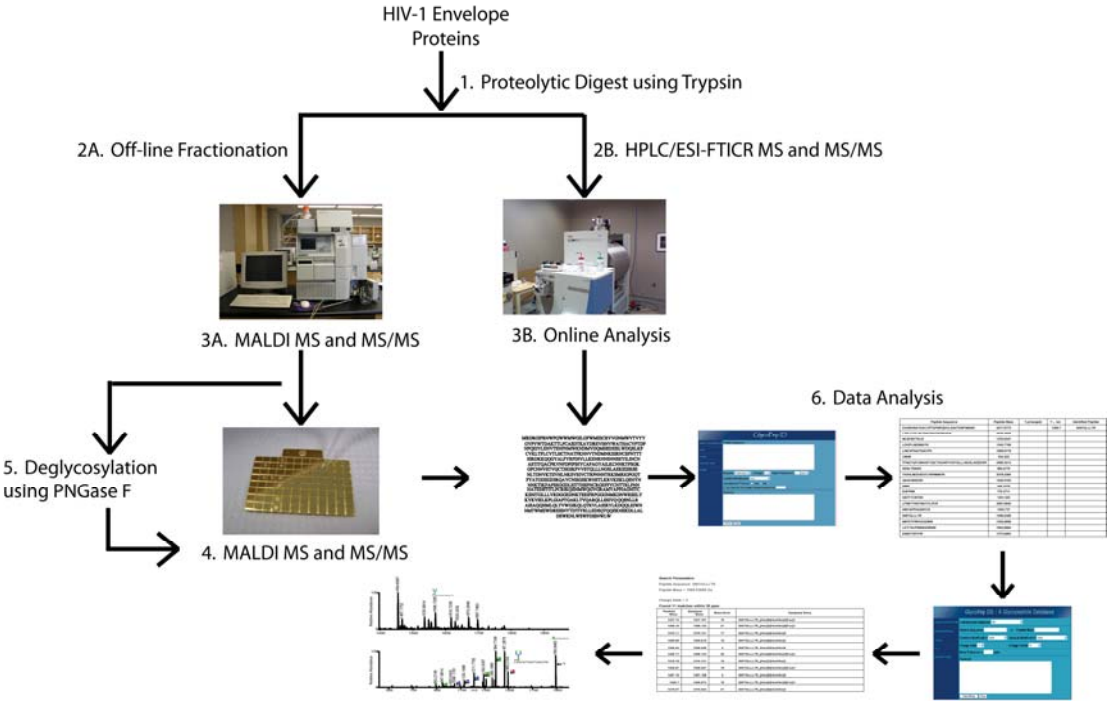

Fig. 2.3: Experimental design of the HIV-1 envelope glycopeptide analysis

2.2.1 Tryptic Digestion

The HIV-1 Env proteins (Con-C and VRC_C) were subjected to a proteolytic digest. Approximately 300 µg of the HIV-1 Env proteins were denatured with 6 M urea in 100 mM Tris-HCL buffer at a pH of 7.5 with 3 mM EDTA. The proteins were reduced with 15 mM DTT for 1 hour and alkylated in the dark with 40 mM IAA for 1 hour, both at room temperature. The sample was brought to a final concentration of 50 mM DTT to neutralize any excess IAA. The proteins were digested with trypsin at a protein/enzyme ratio of 30:1 (wt/wt) in 37 °C for 18 hours. This was followed by a further addition of trypsin at the same protein/enzyme ratio and temperature. The second trypsin addition was allowed to digest for 4 hours and was quenched by adding 1 µL of acetic acid for every 100 µL of total solution. The

digested proteins were analyzed either online using HPLC/ESI-FTICR-MS or offline, using HPLC fractionation with MALDI TOF/TOF MS.

2.2.2 RP-HPLC Fractionation

A 20 μL of the tryptic digestion solution was injected onto a C18 column (150 mm x 4.6 mm, 5 μM size column particle, Alltech, Deerfield, IL) at a flow rate of 1 mL/min. The mobile phase A was purified water with 0.1% formic acid. The mobile phase B was HPLC grade acetonitrile with 0.1% formic acid. A linear gradient was used to elute the glycopeptides starting with a 5% B for 3 min. This was followed by a linear increase to 40% B in 15 min, and held at 40% B for 15 min. A linear increase to 95% B occurred for 20 minutes and held at 95% B for 3 min. Fractions were collected every 1 min for 60 min. The fractions were dried in a centrivap (Labconco Corporation, Kansas City, MO) before reconstituting with 10 μL of water for MALDI-TOF/TOF MS analysis.

2.2.3 N-Deglycosylation

The HPLC glycopeptide fractions were first analyzed by MALDI-TOF/TOF and all fractions that contained glycopeptides were then deglycosylated using N-glycosidase F or PNGase F (CalBioChem) with the protocol recommended by the manufacturer. A 500 units/mL of PNGase F solution was prepared by adding 100 μL of deionized water to 50 units of PNGase F. Approximately 3 μL of each glycopeptide fraction was deglycosyated by adding 3 μL of PNGase F with 9 μL of 20 mM NH_4HCO_3 (pH=8) and incubated overnight at 37 °C. The reaction was stopped the next day by heating the sample to 90 °C. Any fractions that contained

glycopeptides from the rp-HPLC fractionation with MALDI-TOF/TOF were reanalyzed again with MALDI-TOF/TOF after deglycosylation.

2.2.4 MALDI-TOF/TOF Analysis

The matrix was made with equal volumes of saturated solutions of DHB in 50% ACN in water and CHCA in 50% ACN in water with 0.1% TFA. The HPLC fractionated glycopeptides were spotted on a stainless steel MALDI plate and allowed to air-dry. All MALDI MS and MS/MS experiments were collected by an Applied Biosystems 4700 Proteomics Analyzer mass spectrometer (Foster City, CA) in the positive reflectron ion mode. The samples were irradiated with a 355 nm ND:YAG laser at 200 Hz. The mass spectra was generated by averaging 3200 laser shots and optimized to provide the best signal-to-noise (S/N) ratio and resolution for each sample.

2.2.5 HPLC/ESI-FTICR MS Analysis

The digested sample was analyzed using a hybrid linear ion-trap (LIT) Fourier Transform ion cyclotron mass spectrometer (LTQ-FTICR, ThermoElectron, San Jose, CA) that was coupled to a Dionex UltiMate capillary LC system (Sunnyvale, CA). Approximately 5 μL of the digested sample was injected into a C18 PepMap 300 column (15 cm x 300 μm I.D., particle size 5 μm, 300 Å) at a flow rate of 5 μL/min. The mobile phases used for the elution of the glycopeptides were purified water with 0.1% formic acid for mobile phase A and HPLC grade acetonitrile with 0.1% formic acid for mobile phase B. A linear gradient was used to elute the glycopeptides starting with a 5% B for 3 min. This was followed by a linear increase to 40% B in

15 min, and held at 40% B for 15 min. A linear increase to 95% B occurred for 20 minutes and held at 95% B for 3 min. Fractions were collected every 1 min for 60 min. A full LTQ-FTICR MS^1 scan within the mass range *m/z* 800-2000 were followed by three data-dependent MS/MS scans in the positive ion mode. The data-dependent scans for the three most intense ions from the MS^1 data. MS^3 data was also acquired if a neutral loss of hexose or a HexNAc was detected in the MS^2 scans. A spray voltage of 4.0 kV and the capillary temperature was set to 200 °C. All the mass spectrometry data for this study was collected in collaboration with Dr. Eden Go, another member of the Desaire group.

2.2.6 Data Analysis

These proteins were enzymatically cleaved using trypsin and separated using reverse phase high-performance liquid chromatography (rp-HPLC). FTICR-MS and MALDI-TOF were the instrumentations used for the glycopeptide data analysis. The analysis of the glycopeptide composition has been described previously.[5, 6] Briefly, the glycopeptide compositions were assigned using two web-based programs: GlycoPep ID[6] and GlycoPep DB[5]. GlycoPep ID identifies for characteristic ions such as the Y_1 ion $[Peptide + 203 + H]^+$ that occur after collision-induced dissociation (CID).[6]

GlycoPep ID determines the number of N-linked glycosylation sites using the protein sequence along with the digestion methods such as the protelytic enzyme.[6] A table of all possible peptides is generated with the glycosylation sites. This table lists all possible theoretical peptides that can occur with the specific digestion methods

selected and is then compared to the experimental peak list from the MS^2 data.[6] The program matches characteristic ions that fall within the mass tolerance specified by the user.[6] If the program matches more than one peptide, the user can repeat the search with a different characteristic ion or look manually at the spectrum to confirm the peptide portion. Alternatively, both peptide possibilities could be considered in the next step, which is to determine the glycan portion using GlycoPep DB.

GlycoPep DB contains a database of 319 N-linked glycans that have been found in the literature.[5] GlycoPep DB only uses biologically relevant glycans in assigning the carbohydrate portion of the glycopeptides.[5] GlycoPep DB generates a list of possible glycopeptide compositions that fall within the range of the mass tolerance.[5]

Once the plausible glycopeptide compositions have been generated using both GlycoPep ID and GlycoPep DB, the user can confirm those assignments by manually checking the MS^2 spectra for all of the peaks that have MS^2 data by looking for cleavages of neutral losses of individual sugars and cleavages along the peptide backbone.

2.3 Results and Discussion

When comparing the same sites in the two different proteins, many of the peptides within these have the same glycan composition. Out of the possible 29 N-linked glycosylation sites on Con C and 24 possible N-linked glycosylation sites on VRC_C, ten were assigned for Con C with various high-mannose glycoforms and six were assigned for VRC_C with various high-mannose glycoforms shown in Table 1 and Table 2 respectively. The current data presented shows there are some similarities within the glycosylation sites between a consensus sequence and a wild-type sequence. However, this data is not complete and further analysis needs to be done to compare these two sequences further.

Glycopeptides Results from Con C

Charge State	Experimental mass (*m/z*)	Theoretical Mass (*m/z*)	Mass Error (ppm)	Peptide Sequence	Carbohydrate Composition
2+	1454.1570	1454.1276	20	LINCN^{185}TSAITQACPK	[Hex]5[HexNAc]2
2+	1535.1853	1535.1540	20	LINCN^{185}TSAITQACPK	[Hex]6[HexNAc]2
2+	1568.7205	1568.7181	15	LINCN^{185}TSAITQACPK	[Hex]3[HexNAc]4 [Fuc]1
2+	1559.1140	1559.0858	18	GEFFYCN^{372}TSK	[Hex]9[HexNAc]2
2+	1478.0848	1478.0594	17	GEFFYCN^{372}TSK	[Hex]8[HexNAc]2
2+	1316.0281	1316.0066	16	GEFFYCN^{372}TSK	[Hex]6 [HexNAc]2
2+	1397.0546	1397.0330	15	GEFFYCN^{372}TSK	[Hex]7[HexNAc]2
2+	1532.1665	1532.1375	19	AMYAPPIAGN^{416}ITCK	[Hex]7[HexNAc]2
2+	1296.0548	1296.0418	10	EISN^{576}YTDTIYR	[Hex]5[HexNAc]2
2+	1605.1929	1605.1664	17	AMYAPPIAGN^{416}ITCK	[Hex]8[HexNAc]2
2+	1605.1974	1605.1664	19	AMYAPPIAGN^{416}ITCK	[Hex]8[HexNAc]2
2+	1686.2271	1686.1928	20	AMYAPPIAGN^{416}ITCK	[Hex]9[HexNAc]2
2+	1382.1437	1382.1080	26	LDIVPLNEN^{176}NSYR	[Hex]5[HexNAc]2
2+	1529.1542	1529.1153	25	N^{145}CSFN^{149}ITTELR	[Hex]8[HexNAc]2
2+	1476.6749	1476.6443	21	SN^{422}ITGLLLTR	[Hex]9[HexNAc]2
2+	1395.6450	1395.6179	19	SN^{422}ITGLLLTR	[Hex]8[HexNAc]2
2+	1540.7130	1540.6679	29	LICTTAVPWN^{550}SSWSN^{555}K	[Hex]5[HexNAc]2

Table 1. Glycopeptides of Con C with *m/z* from the peak list compared to the theoretical database mass and their mass errors.

Glycopeptides Results from VRC_C

Charge State	Experimental mass (*m/z*)	Theoretical Mass (*m/z*)	Mass Error (ppm)	Peptide Sequence	Carbohydrate Composition
2+	1158.5632	1158.5330	26	TIIVHLN^{248}K	[Hex]6 [HexNAc]2
2+	1239.5927	1239.5594	27	TIIVHLN^{248}K	[Hex]7[HexNAc]2
2+	1077.5352	1077.5066	27	TIIVHLN^{248}K	[Hex]5[HexNAc]2
2+	1158.5657	1158.5330	28	TIIVHLN^{248}K	[Hex]6 [HexNAc]2
2+	1280.0540	1280.0178	28	N^{110}CSFN^{114}TTTEIR	[Hex]5[HexNAc]2
2+	1418.0923	1418.0439	34	GEFFYCN^{344}TTR	[Hex]7[HexNAc]2
2+	1499.1262	1499.0703	37	GEFFYCN^{344}TTR	[Hex]8[HexNAc]2
2+	1580.1622	1580.0967	41	GEFFYCN^{344}TTR	[Hex]9[HexNAc]2
2+	1256.0236	1255.9911	26	GEFFYCN^{344}TTR	[Hex]5[HexNAc]2
2+	1337.0559	1337.0175	29	GEFFYCN^{344}TTR	[Hex]6 [HexNAc]2
2+	1174.9998	1174.9647	30	GEFFYCN^{344}TTR	[Hex]4[HexNAc]2
2+	1276.5404	1276.5044	28	GEFFYCN^{344}TTR	[Hex]4[HexNAc]3
2+	1524.1969	1524.1400	37	AMYAPPIAGN^{387}ITCK	[Hex]7[HexNAc]2
2+	1605.2278	1605.1664	38	AMYAPPIAGN^{387}ITCK	[Hex]8[HexNAc]2
2+	1686.2626	1686.1928	41	AMYAPPIAGN^{387}ITCK	[Hex]9[HexNAc]2
2+	1710.8176	1710.7143	60	AMYAPPIAGN^{387}ITCK	[Hex]5 [HexNAc]4 [NeuNAc]1
2+	1394.6696	1394.6282	30	SN^{393}ITGLLLVR	[Hex]8[HexNAc]2
2+	1313.6399	1313.6018	29	SN^{393}ITGLLLVR	[Hex]7[HexNAc]2
2+	1475.7002	1475.6546	31	SN^{393}ITGLLLVR	[Hex]9[HexNAc]2
2+	1151.5769	1151.5490	24	SN^{393}ITGLLLVR	[Hex]5[HexNAc]2
2+	1232.6073	1232.5754	26	SN^{393}ITGLLLVR	[Hex]6 [HexNAc]2

Table 2. Glycopeptides of VRC_C with *m/z* from the peak list compared to the theoretical database mass and their mass errors

While a significant number of glycoforms have been identified, it is likely that many more glycoforms are present that have not yet been characterized. One problem with the current data set is that high mannose glycans were predominately found between these two proteins, yet it is likely that several hybrid and complex glycans are also attached to these peptides. The hybrid and complex N-linked glycans are more difficult to analyze than the high-mannose N-linked glycans. This is due to the increased difficulty in finding the characteristic Y_1 ion in the MS^2 spectra of the hybrid and complex glycans. The characteristic Y_1 ion for a high-mannose glycans in a linear ion trap can be seen as the base peak whereas the characteristic Y_1 ion for a hybrid or complex glycan is not seen as the base peak.[7] This can be attributed to the ease at which fucose and the 1-4 linked HexNAc can be cleaved, compared to the glycosidic cleavages of high-mannose glycans.[7] The data analysis steps for assigning glycopeptides with unprocessed glycans, shown in Fig. 2.4, is accomplished by observing sequential losses of sugar moieties within the overall glycans profile. This data analysis method was not successful in assigning all of the possible N-linked sites. The sequence coverage is not complete and not all of the glycopeptides were assigned. It's possible that some glycopeptides co-elute and better separation techniques can solve this problem. Chapter three addresses the question of whether or not a better separation technique could provide better sequence coverage for glycopeptide analysis.

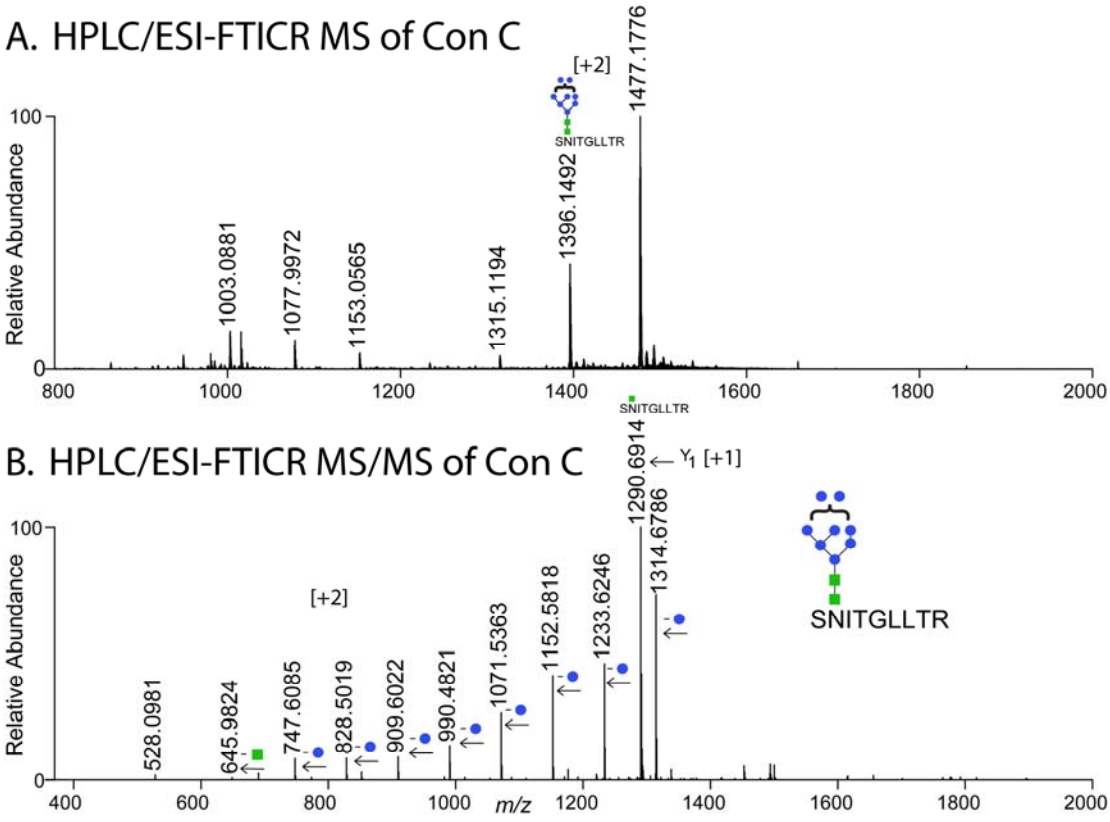

Fig. 2.4: Representative example MS data of an unprocessed glycoform of Con C (A) MS and (B) MS/MS spectra using HPLC/ESI-FTICR

2.4 Conclusion

Ten N-linked glycosylation sites were found on Con C and six N-linked glycosylation sites were found on VRC_C. The remainder of the potential glycosylation sites were not able to be determined by HPLC/ESI-FTICR-MS or MALDI TOF/TOF MS. Processed glycans were not seen in the spectra. Another technique to aid in assigning processed and unprocessed glycans can be performed to enhance the coverage. Also, by refining the chromatographic separation methods to further separate the unprocessed glycans from the processed glycans, more glycoforms may be detected.

2.5 References

1. Reitter, J. N.; Means, R. E.; Desrosiers, R. C., A Role For Carbohydrates In Immune Evasion In AIDS. *Nat Med.* **1998,** 4, (6), 679-684.
2. Kothe, D. L.; Li, Y. Y.; Decker, J. M.; Bibollet-Ruche, F.; Zammit, K. P.; Salazar, M. G.; Chen, Y. L.; Weng, Z. P.; Weaver, E. A.; Gao, F.; Haynes, B. F.; Shaw, G. M.; Korber, B. T. M.; Hahn, B. H., Ancestral And Consensus Envelope Immunogens For HIV-1 Subtype C. *Virology* **2006,** 352, (2), 438-449.
3. Gaschen, B.; Taylor, J.; Yusim, K.; Foley, B.; Gao, F.; Lang, D.; Novitsky, V.; Haynes, B.; Hahn, B. H.; Bhattacharya, T.; Korber, B., AIDS - Diversity Considerations In HIV-1 Vaccine Selection. *Science* **2002,** 296, (5577), 2354-2360.
4. Gao, F.; Liao, H. X.; Hahn, B. H.; Letvin, N. L.; Korber, B. T.; Haynes, B. F., Centralized HIV-1 envelope immunogens and neutralizing antibodies. *Current HIV research* **2007,** 5, (6), 572-577.
5. Go, E. P.; Rebecchi, K. R.; Dalpathado, D. S.; Bandu, M. L.; Zhang, Y.; Desaire, H., GlycoPep DB: A Tool For Glycopeptide Analysis Using A "Smart Search". *Anal. Chem.* **2007,** 79, (4), 1708-1713.
6. Irungu, J.; Go, E. P.; Dalpathado, D. S.; Desaire, H., Simplification of mass spectral analysis of acidic glycopeptides using GlycoPep ID. *Anal. Chem.* **2007,** 79, (8), 3065-3074.
7. Irungu, J.; Go, E. P.; Zhang, Y.; Dalpathado, D. S.; Liao, H. X.; Haynes, B. F.; Desaire, H., Comparison of HPLC/ESI-FTICR MS versus MALDI-TOF/TOF MS for glycopeptide analysis of a highly glycosylated HIV envelope glycoprotein. *J. Am. Soc. Mass. Spectrom.* **2008,** 19, (8), 1209-1220.

Chapter 3

Glycopeptide Separation Methods Using Asialofetuin

3.1 Introduction

The analysis of glycopeptides in a proteolytic digest is problematic due to the presence of peptides along with glycopeptides in the sample. Peptides ionize better than glycopeptides and would suppress the signal intensity of the glycopeptides if both are present in the sample.

A variety of methods have been proposed to isolate the glycopeptides within a peptide mixture. Lectin-affinity chromatography can isolate glycoproteins with different carbohydrate types.[1] This method is useful if the glycoforms are known on the glycopeptides. Analysis of glycopeptides with an unknown variety of glycoforms would be problematic for this method. Affinity separation using cellulose or Sepharose on glycopeptides from a peptide mixture is another method. The affinity of the cellulose or Sepharose is dependent on the hydrogen bonding between the hydroxyl group of carbohydrates on cellulose or Sepharose and the glycopeptides.[2] High-performance anion-exchange (HPAE) chromatography with pulsed amperometric detection (PAD) has been used to separate oligosaccharides according to molecular size, sugar composition, and linkage of the monosaccharide units.[3] This technique requires high concentrations of sodium hydroxide and sodium acetate, and these salts have to be removed before further analysis.[4]

Liquid chromatography has been extensively used in separation of glycans.[5] The study in this chapter investigates the separation efficiencies of three various

capillary columns (C18 column, porous graphite carbon column, and a zwitterionic hydrophilic interaction column) with three unique retention mechanisms, to selectively purify/enrich glycopeptides present in a peptide mixture.

Reverse-Phase High Performance Liquid Chromatography (rp-HPLC) is the conventional method for separation of glycopeptides after enzymatic digestion.[6] Silanes that possess alky chains (mainly C_{18}) are the standard stationary phase for reverse-phase separation. The stationary phase also contains polar alky groups to increase the retention of polar compounds.[7] Polar embedded or polar end-capped groups have been used to improve separation efficiencies of highly polar compounds.[8]

Porous graphitized carbon (PGC) exhibit properties similar to reverse-phased columns, but they are able to retain more hydrophilic analytes and provide separation of structurally similar oligosaccharides.[9] The retention mechanism for PGC is due to a combination of dispersive interactions or hydrophobic interactions and charge-induced interactions.[10-12] London dispersive interactions (as well as dipole-dipole and hydrogen bonding) between the analytes and the eluents increase the analytes' retention to the graphite surface.[13] Hydrophobic interactions between the analyte and the eluent will affect the retention for the graphite surface. Repulsions between the hydrophilic eluent and non-polar segments of the analytes will increase retention of the analyte for the graphite surface.[12, 13]

Charge-induced interactions force polar groups on the analyte to be closer to the graphite surface, as shown in Fig. 3.1.[10, 11, 14, 15] The flat surface reduces the

retention of highly structural and rigid analytes, which can only come in contact with the graphite surface with a portion of their own surface.[11, 16] The molecular area of an analyte that is in contact with the graphite surface largely determines the strength of the interaction.[10] Hence, planar molecules tend to come in contact with the graphite surface better than non-planar molecules and have a higher retention for the graphite surface than non-polar molecules.[17]

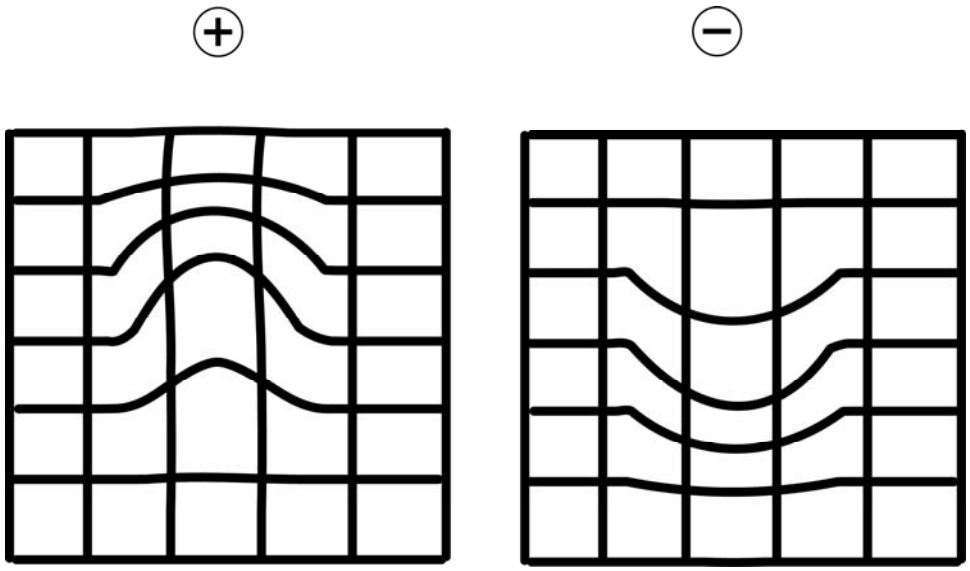

Fig. 3.1: Charge-induced interactions from polar groups of the analyte

Hydrophilic interaction liquid chromatography (HILIC) is similar to normal-phase chromatography but distinguishes itself from normal-phase in that the mobile phases contain water-miscible solvents, whereas normal-phase chromatography uses non-water-miscible solvent buffers.[18]

Interaction with the ZIC-HILIC stationary phase is based on a partitioning mechanism, with electrostatic interactions from the sulfonic group and quaternary ammonium group, as shown in Fig. 3.2.[19, 20] The ZIC-HILIC surface has a zero net charge due to the 1:1 ratio of quaternary ammonium and sulfonic acid group.[20] The stationary phases are very hydrophilic and easily absorb or imbibe water.[5] Having a very hydrophilic stationary phase and a relatively hydrophobic mobile phase allows for the hydrophilic analytes to partition between the hydrophilic stationary phase and relatively hydrophobic bulk eluent.[5, 18, 20] The order of elution is the opposite of reverse-phase chromatography; the hydrophilic compounds will be retained longer than hydrophobic compounds.[18, 20] This makes a HILIC column an ideal tool for analysis of complex samples with poor retention towards reverse-phase chromatography.

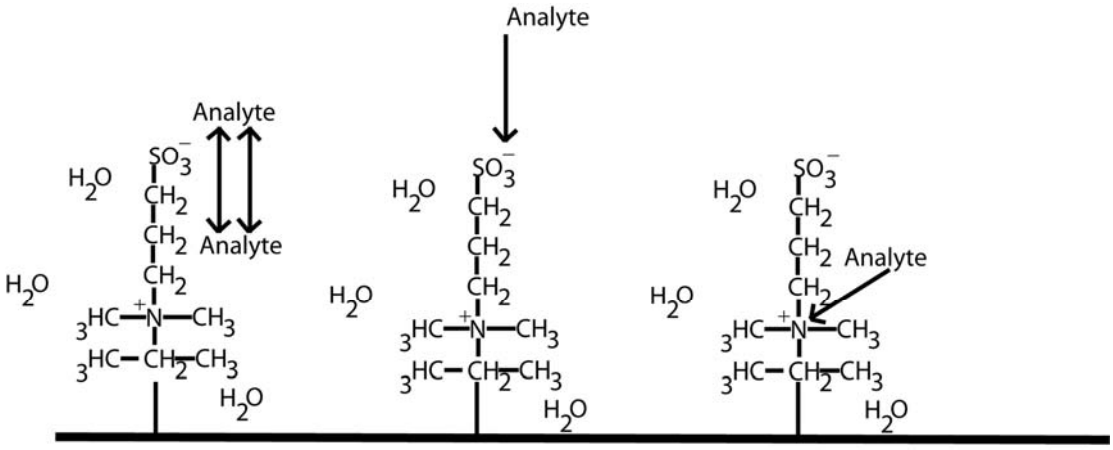

Fig. 3.2: Stationary phase of ZIC-HILIC with partitioning mechanism and electrostatic interactions with the analyte

Since the retention of compounds is based on polar interactions, the retention of glycopeptides tends to increase with an increase in the length of the glycans chains, and the elution position may provide a rough indication of the glycans size.[21]

Asialofetuin was chosen in this experiment because its complete structure has been well studied.[22-24] Fig. 3.3 shows three well characterized N-linked glycosylation sites in asialofetuin (residues at N^{99}CS, N^{156}DS, and N^{176}GS) which make it ideal for distinguishing the capability of each column in separating glycopeptides from the peptide mixture.

MKSFVLLFCLAQLWGCHSIPLDPVAGYKEPACDDPDTE
QAALAAVDYINKHLPRGYKHTLNQIDSVKVWPRRPTGE
VYDIEIDTLETTCHVLDPTPLANCSVRQQTQHAVEGDC
DIHVLKQDGQFSVLFTKCDSSPDSAEDVRKLCPDCPLL
APLNDSRVVHAVEVALATFNAESNGSYLQLVEISRAQF
VPLPVSVSVEFAVAATDCIAKEVVDPTKCNLLAEKQYG
FCKGSVIQKALGGEDVRVTCTLFQTQPVIPQPQPDGAE
AEAPSAVPDAAGPTPSAAGPPVASVVVGPSVVAVPLPL
HRAHYDLRHTFSGVASVESSSGEAFHVGKTPIVGQPSI
PGGPVRLCPGRIRYFKI

Fig. 3.3: Protein sequence of asialofetuin. N-linked glycosylation sites are in red.

3.2 Experimental Design

A capillary ZIC-HILIC column (150 mm x 0.3 mm I.D., particle size 5μm) was purchased from SeQuant (Umeå, Sweden). Hypercarb (150 mm x 0.32 mm I.D., particle size 5 μm) and Aquasil C18 (150 mm x 0.32 mm I.D., particle size 5 μm) were purchased from Thermo Scientific (San Jose, CA)

Asialofetuin was purchased from Sigma-Aldrich (St. Louis, MO). Ammonium bicarbonate (NH_4HCO_3), trizma hydrochloride, trizma base, urea, ethylenediaminetetraacetic acid (EDTA), dithiothreitol (DTT), iodoacetamide (IAA), HPLC grade acetonitrile (ACN), and formic acid were all purchased from Sigma-Aldrich (St. Louis, MO). Proteomics grade trypsin was purchased from Promega (Madison, WI). N-Glycosidase F (PNGase F) was acquired from Elizabethkingia meningosepticum and purchased from CalBioChem (San Diego, CA). Purified water was obtained using a Millipore Direct-Q3 Water Purification System (Billerica, MA).

The overall experiment is shown in Fig. 3.4 with explanations in the proceeding paragraphs.

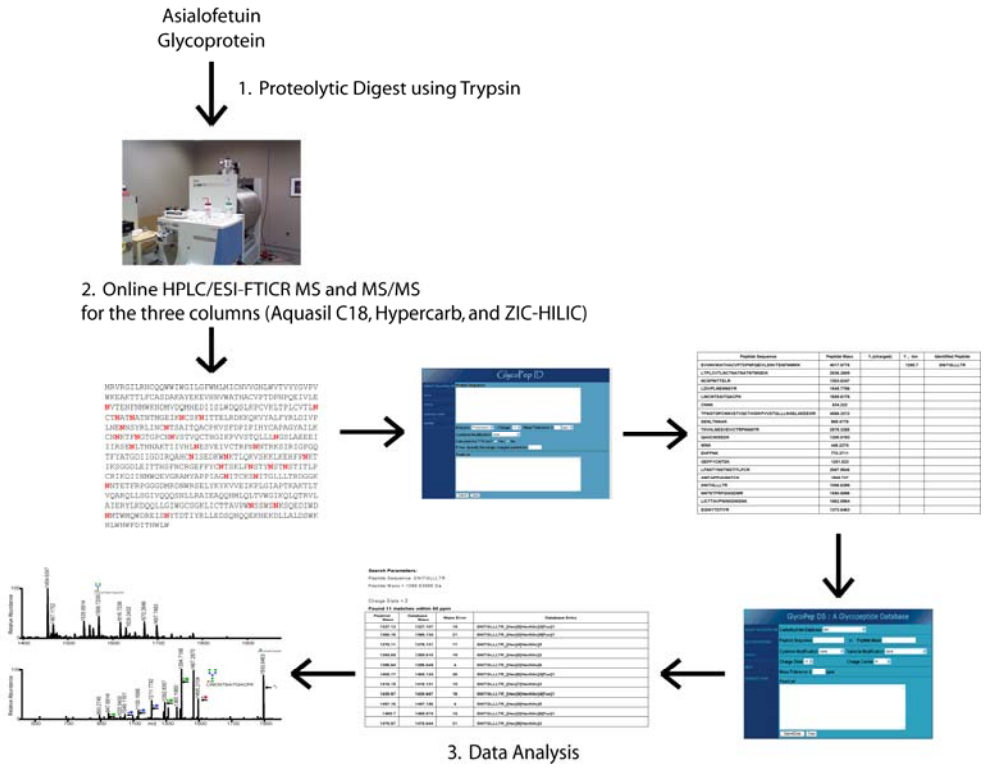

Fig. 3.4: Experimental design of the asialofetuin glycopeptide analysis

3.2.1 Protease digestion

Approximately 300 µg of asialofetuin protein was denatured in 25 mM ammonium hydrogen carbonate buffer with 4 M urea. The proteins were reduced with 15 mM DTT for 1 hour and alkylated in the dark with 25 mM IAA for 1 hour, both at room temperature. The sample was brought to a final concentration of 40 mM DTT to neutralize any excess IAA. Ammonium hydrogen carbonate buffer was added to reduce the Urea concentration to 1 M. Trypsin was added to digest the protein at a protein/enzyme ratio of 50:1 (wt/wt) in 37 °C for 18 hours. The reaction was quenched by adding 1 µL of acetic acid for every 100 µL of total solution. The final digested protein sample was analyzed by an online HPLC/ESI-FTICR-MS.

3.2.2 Glycopeptide Enrichment using capillary columns

A hybrid linear ion-trap (LIT) Fourier Transform ion cyclotron resonance mass spectrometer (LTQ-FTICR, ThermoElectron, San Jose, CA) was coupled to a Dionex UltiMate capillary LC system (Sunnyvale, CA). The tryptic glycopeptide/peptide mixture was placed on a FAMOS well plate autosampler that is attached to the Dionex UltiMate capillary LC system. For each run, 5 μL of the tryptic digest was injected onto one of the three various columns at a flow rate of 5 μL/min. The Aquasil C18 and Hypercarb used the same mobile phases. The mobile phase A was purified water with 0.1% formic acid. The mobile phase B was HPLC grade acetonitrile with 0.1% formic acid. The HILIC column used 50% acetonitrile/10 mM ammonium acetate as the mobile phase A and 80% acetonitrile /10 mM ammonium acetate as the mobile phase B. A linear gradient elution was used on all columns. For the Aquasil C18 column, a linear gradient elution of 3% B for the first 3 minutes at a flow rate of 5 μL/min. This was followed by a linear increase to 25% B in 12 minutes, then increased to 50% B for 15 minutes, then increased to 75% B for an additional 20 minutes, and lastly to 95% B for 10 minutes. The total elution time was 60 minutes. For the Hypercarb column, a linear gradient of 3% B for the first 2 minutes at a flow rate of 5 μL/min. This was followed by a linear increase to 25% B for 18 minutes, then an increase to 40% B for 10 minutes, followed by 50% B for 15 minutes, then 70% B for 5 minutes, then 82% B for the last 10 minutes. The total elution time was 60 minutes. For the HILIC column, a linear gradient of 70% B for the first 5 minutes was used, then 55% B for 10 minutes, then

50% B for 15 minutes, then 40% B for 30 minutes, and lastly 32% B for 20 minutes. The total elution time was 80 minutes. A short wash and a blank were run prior to each sample run to prevent any sample carryover.

A full LTQ-FTICR MS^1 scan within the mass range *m/z* 800-2000 were followed by three data-dependent MS/MS scans in the positive ion mode. The data-dependent scans collected MS/MS data for the three most intense ions from the MS^1 data. MS^3 data was also acquired if a neutral loss of hexose or a HexNAc was detected in the MS^2 scans. A spray voltage of 4.0 kV and the capillary temperature was set to 200 °C.

3.2.3 Data Analysis

The analysis of the glycopeptide composition has been described previously.[25, 26] Briefly, the glycopeptide compositions were assigned using two web-based programs: GlycoPep ID[25] and GlycoPep DB[26]. GlycoPep ID identifies for characteristic ions such as the Y_1 ion $[Peptide + 203 + H]^+$ that occur after collision-induced dissociation (CID).[25]

GlycoPep ID determines the number of N-linked glycosylation sites using the protein sequence along with the digestion methods such as the protelytic enzyme.[25] A table of all possible peptides is generated with the glycosylation sites. This table lists all possible theoretical peptides that can occur with the specific digestion methods selected and is then compared to the experimental peak list from the MS^2 data.[25] The program matches characteristic ions that fall within the mass tolerance specified by the user.[25] If the program matches more than one peptide, the user can repeat the

search with a different characteristic ion or look manually at the spectrum to confirm the peptide portion. Alternatively, both peptide possibilities could be considered in the next step, which is to determine the glycan portion using GlycoPep DB.

GlycoPep DB contains a database of 319 N-linked glycans that have been found in the literature.[26] GlycoPep DB only uses biologically relevant glycans in assigning the carbohydrate portion of the glycopeptides.[26] GlycoPep DB generates a list of possible glycopeptide compositions that fall within the range of the mass tolerance.[26]

Once the plausible glycopeptide compositions have been generated using both GlycoPep ID and GlycoPep DB, the user can confirm those assignments by manually checking the MS^2 spectra for all of the peaks that have MS^2 data by looking for cleavages of neutral losses of individual sugars and cleavages along the peptide backbone. All of the plausible glycopeptide compositions generated by GlycoPep ID and GlycoPep DB were scrutinized manually by looking for neutral losses of glycans in the MS^2 data. MS^1 data was also used to confirm the plausible glycopeptides.

3.3 Results and Discussion

The purpose of this study was to determine which column was the most advantageous in separating N-linked glycopeptides of asialofetuin from a peptide mixture, in order to better the separation methods and improve the glycosylation coverage for the HIV analysis.

In order to determine the glycopeptide compositions in asialofetuin, a proteolytic digest was performed using trypsin. HPLC/ESI-FTICR MS and MS/MS experiments were chosen to investigate the separation capabilities of the three columns. HPLC/ESI-FTICR MS was also chosen because the experiment can be performed online and would not require a RP-HPLC fractionation step if MALDI-TOF MS and MS/MS was selected as the mass spectrometer.

Glycopeptide Identification

Hypercarb, Aquasil C18, and ZIC-HILIC columns were used to elucidate the glycoforms on asialofetuin. There are three N-linked glycosylation sites on asialofetuin (residues at N^{99}CS, N^{156}DS, and N^{176}GS). The Hypercarb column was able to separate the N^{156}DS glycoform shown in Table 1. The Aquasil C18 column was able to separate the N^{156}DS and N^{176}GS glycoforms shown in Table 2. The ZIC-HILIC column separated all three N-linked glycosylation sites on asialofetuin shown in Table 3.

Hypercarb Results of N-linked Asialofetuin

Charge State	Experimental mass (*m/z*)	Theoretical Mass (*m/z*)	Mass Error (ppm)	Peptide Sequence	Carbohydrate Composition
3+	1243.5982	1243.5230	60.5	LCPDCPLLAPLN^{156}DSR	[Hex]6 [HexNAc]5
3+	1121.8069	1121.8127	5.2	LCPDCPLLAPLN^{156}DSR	[Hex]5 [HexNAc]4
2+	1864.8670	1864.7813	46.0	LCPDCPLLAPLN^{156}DSR	[Hex]6 [HexNAc]5
2+	1682.2846	1682.2152	41.3	LCPDCPLLAPLN^{156}DSR	[Hex]6 [HexNAc]5

Table 1. Hypercarb separation of glycopeptides in asialofetuin with *m/z* from the peak list compared to the theoretical database mass and their mass errors

Aquasil C18 Results of N-linked Asialofetuin

Charge State	Experimental mass (*m/z*)	Theoretical Mass (*m/z*)	Mass Error (ppm)	Peptide Sequence	Carbohydrate Composition
4+	1160.6475	1160.5448	88.5	VVHAVEVALATFNAESN^{176}GSYLQLVEISR	[Hex]5 [HexNAc]4
3+	1286.3039	1286.2218	63.8	KLCPDCPLLAPLN^{156}DSR	[Hex]6 [HexNAc]5
3+	1243.5855	1243.5235	50.0	LCPDCPLLAPLN^{156}DSR	[Hex]6 [HexNAc]5
2+	1682.3695	1682.2152	91.7	LCPDCPLLAPLN^{156}DSR	[Hex]5 [HexNAc]4
3+	1121.8960	1121.8127	74.3	LCPDCPLLAPLN^{156}DSR	[Hex]5 [HexNAc]4

Table 2. Aquasil C18 separation of glycopeptides in asialofetuin with *m/z* from the peak list compared to the theoretical database mass and their mass errors

ZIC-HILIC Results of N-linked Asialofetuin

Charge State	Experimental mass (*m/z*)	Theoretical Mass (*m/z*)	Mass Error (ppm)	Peptide Sequence	Carbohydrate Composition
3+	1547.1085	1547.0572	33.2	VVHAVEVALATFNAESN^{176}GSYLQLVEISR	[Hex]5 [HexNAc]4
3+	1668.8517	1668.7679	50.2	VVHAVEVALATFNAESN^{176}GSYLQLVEISR	[Hex]6 [HexNAc]5
2+	1864.8464	1864.7813	34.9	LCPDCPLLAPLN^{156}DSR	[Hex]6 [HexNAc]5
3+	1243.5528	1243.5235	23.6	LCPDCPLLAPLN^{156}DSR	[Hex]6 [HexNAc]5
3+	1682.2750	1682.2152	35.5	LCPDCPLLAPLN^{156}DSR	[Hex]5 [HexNAc]4
3+	1121.8361	1121.8127	20.9	LCPDCPLLAPLN^{156}DSR	[Hex]5 [HexNAc]4
4+	1415.6661	1415.6264	28.0	RPTGEVYDIEIDTLETTC HVLDPTPLAN^{99}CSVR	[Hex]6 [HexNAc]5
3+	1765.5534	1765.4552	55.6	RPTGEVYDIEIDTLETTC HVLDPTPLAN^{99}CSVR	[Hex]5 [HexNAc]4
3+	1286.3161	1286.2218	73.3	KLCPDCPLLAPLN^{156}DSR	[Hex]6 [HexNAc]5
3+	1164.5226	1164.5111	9.9	KLCPDCPLLAPLN^{156}DSR	[Hex]5 [HexNAc]4

Table 3. ZIC-HILIC separation of glycopeptides in asialofetuin with *m/z* from the peak list compared to the theoretical database mass and their mass errors

The Aquasil C18 and the Hypercarb capillary columns are both reverse-phase chromatographic separation techniques, whereas the ZIC-HILIC capillary column is similar to a normal-phase chromatographic separation technique. The ZIC-HILIC uses sulfoalkylbetaine functional groups as the stationary phase, which gives it better separation of highly polar compounds than common reverse-phase stationary phases employed with alkyl chains.[7, 20] A charged polar stationary phase allows the ZIC-HILIC column to attract water molecules from the mobile phase to form water-enriched layers.[7]

3.4 Conclusion

Aquasil C18 and Hypercarb both used reverse-phase elution mechanism whereas the ZIC-HILIC is the opposite of reverse-phase chromatography. The ZIC-HILIC showed the best separation on asialofetuin glycopeptides and should be further studied in order to apply its partitioning mechanism to the study of HIV-1 envelope proteins.

3.5 References

1. Hirabayashi, J.; Kasai, K., Separation technologies for glycomics. *J. Chromatogr. B* **2002,** 771, (1-2), 67-87.
2. Wada, Y.; Tajiri, M.; Yoshida, S., Hydrophilic affinity isolation and MALDI multiple-stage tandem mass spectrometry of glycopeptides for glycoproteomics. *Anal. Chem.* **2004,** 76, (22), 6560-6565.
3. Lee, J.; Kim, J.; Hyeon, T., Recent progress in the synthesis of porous carbon materials. *Adv. Mater.* **2006,** 18, (16), 2073-2094.
4. Davies, M.; Smith, K. D.; Harbin, A. M.; Hounsell, E. F., High-Performance Liquid-Chromatography Of Oligosaccharide Alditols And Glycopeptides On A Graphitized Carbon Column. *J. Chromatogr.* **1992,** 609, (1-2), 125-131.
5. Neue, U. D.; O'Gara, J. E.; Mendez, A., Selectivity in reversed-phase separations - Influence of the stationary phase. *J. Chromatogr. A* **2006,** 1127, (1-2), 161-174.
6. Kanie, Y.; Enomoto, A.; Goto, S.; Kanie, O., Comparative RP-HPLC for rapid identification of glycopeptides and application in off-line LC-MALDI-MS analysis. *Carbohydr. Res.* **2008,** 343, (4), 758-768.
7. Ikegami, T.; Hara, T.; Kimura, H.; Kobayashi, H.; Hosoya, K.; Cabrera, K.; Tanaka, N., Two-dimensional reversed-phase liquid chromatography using two monolithic silica C18 columns and different mobile phase modifiers in the two dimensions. *J. Chromatogr. A* **2006,** 1106, (1-2), 112-117.
8. Layne, J., Characterization and comparison of the chromatographic performance of conventional, polar-embedded, and polar-endcapped reversed-phase liquid chromatography stationary phases. *J. Chromatogr. A* **2002,** 957, (2), 149-164.
9. Kawasaki, N.; Ohta, M.; Hyuga, S.; Hashimoto, O.; Hayakawa, T., Analysis of carbohydrate heterogeneity in a glycoprotein using liquid chromatography mass spectrometry and liquid chromatography with tandem mass spectrometry. *Anal. Biochem.* **1999,** 269, (2), 297-303.
10. Ross, P., The role of porous graphitic carbon in HPLC. *Lc Gc N. Am.* **2000,** 18, (1), 14-+.
11. Knox, J. H.; Kaur, B.; Millward, G. R., Structure And Performance Of Porous Graphitic Carbon In Liquid-Chromatography. *J. Chromatogr.* **1986,** 352, 3-25.
12. Jacquet, R.; Pennanec, R.; Elfakir, C.; Lafosse, M., Liquid chromatography analysis of monosubstituted sulfobutyl ether-beta-cyclodextrin isomers on porous graphitic carbon. *J. Sep. Sci.* **2004,** 27, (14), 1221-1228.
13. Pereira, L., Porous graphitic carbon as a stationary phase in HPLC: Theory and applications. *J. Liq. Chromatogr. Relat. Technol.* **2008,** 31, (11-12), 1687-1731.
14. Hanai, T., Separation of polar compounds using carbon columns. *J. Chromatogr. A* **2003,** 989, (2), 183-196.
15. Ross, P.; Knox, J. H., Carbon-based packing materials for liquid chromatography: Applications. In *Advances in Chromatography, Vol 37*, Marcel Dekker: New York, 1997; Vol. 37, pp 121-162.

16. Reitter, J. N.; Means, R. E.; Desrosiers, R. C., A Role For Carbohydrates In Immune Evasion In AIDS. *Nat Med.* **1998,** 4, (6), 679-684.
17. Koizumi, K., High-performance liquid chromatographic separation of carbohydrates on graphitized carbon columns. *J. Chromatogr. A* **1996,** 720, (1-2), 119-126.
18. Boersema, P. J.; Mohammed, S.; Heck, A. J. R., Hydrophilic interaction liquid chromatography (HILIC) in proteomics. *Anal. Bioanal. Chem.* **2008,** 391, (1), 151-159.
19. Takegawa, Y.; Deguchi, K.; Ito, H.; Keira, T.; Nakagawa, H.; Nishimura, S. I., Simple separation of isomeric sialylated N-glycopeptides by a zwitterionic type of hydrophilic interaction chromatography. *J. Sep. Sci.* **2006,** 29, (16), 2533-2540.
20. Hemstrom, P.; Irgum, K., Hydrophilic interaction chromatography. *J. Sep. Sci.* **2006,** 29, (12), 1784-1821.
21. Wuhrer, M.; Deelder, A. M.; Hokke, C. H., Protein glycosylation analysis by liquid chromatography-mass spectrometry. *J. Chromatogr. B* **2005,** 825, (2), 124-133.
22. Yet, M. G.; Chin, C. C. Q.; Wold, F., The Covalent Structure Of Individual N-Linked Glycopeptides From Ovomucoid And Asialofetuin. *J. Biol. Chem.* **1988,** 263, (1), 111-117.
23. Spiro, R. G., Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **2002,** 12, (4), 43R-56R.
24. Baenziger, J. U.; Fiete, D., Structure Of The Complex Oligosaccharides Of Fetuin. *J. Biol. Chem.* **1979,** 254, (3), 789-795.
25. Irungu, J.; Go, E. P.; Dalpathado, D. S.; Desaire, H., Simplification of mass spectral analysis of acidic glycopeptides using GlycoPep ID. *Anal. Chem.* **2007,** 79, (8), 3065-3074.
26. Go, E. P.; Rebecchi, K. R.; Dalpathado, D. S.; Bandu, M. L.; Zhang, Y.; Desaire, H., GlycoPep DB: A Tool For Glycopeptide Analysis Using A "Smart Search". *Anal. Chem.* **2007,** 79, (4), 1708-1713.

4.1 Conclusion and future work

The ZIC-HILIC capillary column should be used as the separation method for the glycosylation analysis of the HIV-1 envelope proteins. Another study should be performed on the ZIC-HILIC capillary column to separate glycopeptides from a peptide mixture with various buffer solvents and elution gradients in order to optimize the advantages of its unique partitioning mechanism.