



Published in final edited form as:

J Am Soc Mass Spectrom. 2008 August ; 19(8): 1209–1220. doi:10.1016/j.jasms.2008.05.010.

Comparison of HPLC/ESI-FTICR MS Versus MALDI-TOF/TOF MS for Glycopeptide Analysis of a Highly Glycosylated HIV Envelope Glycoprotein

Janet Irungu^a, Eden P. Go^a, Ying Zhang^a, Dilusha S. Dalpathado^a, Hua-Xin Liao^b, Barton F. Haynes^b, and Heather Desaire^a

^aDepartment of Chemistry, University of Kansas, Lawrence, Kansas, USA

^bDuke University Medical Center, Durham, North Carolina, USA

Abstract

Defining the structures and locations of the glycans attached on secreted proteins and virus envelope proteins is important in understanding how glycosylation affects their biological properties. Glycopeptide mass spectrometry (MS)-based analysis is a very powerful, emerging approach to characterize glycoproteins, in which glycosylation sites and the corresponding glycan structures are elucidated in a single MS experiment. However, to date there is not a consensus regarding which mass spectrometric platform provides the best glycosylation coverage information. Herein, we employ two of the most widely used MS approaches, online high performance liquid chromatography-electrospray ionization mass spectrometry (HPLC/ESI-MS) and offline HPLC followed by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS), to determine which of the two approaches provides the best glycosylation coverage information of a complex glycoprotein, the group M consensus HIV-1 envelope, CON-S gp140ΔCFI, which has 31 potential glycosylation sites. Our results highlight differences in the informational content obtained between the two methods such as the overall number of glycosylation sites detected, the numbers of N-linked glycans present at each site, and the type of confirmatory information obtained about the glycopeptide using MS/MS experiments. The two approaches are quite complementary, both in their coverage of glycopeptides and in the information they provide in MS/MS experiments. The information in this study contributes to the field of mass spectrometry by demonstrating the strengths and limitations of two widely used MS platforms in glycoprotein analysis.

Glycoproteomics is a newly emerging field focusing on analysis of protein glycosylation. These studies are important because glycosylation is the most common post-translational modification present in both eukaryotic and prokaryotic proteins [1, 2], and this modification affects proteins' biological and cellular processes and influences their physiochemical properties [3-5]. Additionally, glycans have been shown to play a vital role in various parasitic, bacterial, and viral disease infections [6]. For instance, interaction and fusion of the human immunodeficiency virus (HIV) with its target host cells is mediated by its envelope protein, gp160, which has over 50% of its mass comprising of glycans [7-10]. The high population and diverse range of glycans on this protein acts as a shield for the virus against the immune system; the glycans also mask epitopes that impact HIV disease progression [8, 11-17]. Consequently, conducting glycoproteomics studies on this target,

defining the structures and locations of glycans in the HIV envelope protein, is important in understanding how variation in glycosylation affects the functions of this protein, and the studies may also provide valuable information that can be useful in identifying new vaccine candidates. To acquire this information, sensitive, rapid, and reliable glycoproteomics methods for mapping and profiling protein glycosylation are of paramount importance.

Analytical methods in the glycoproteomics field are still evolving. Mass spectrometry (MS) has gained a widespread use in protein glycosylation analysis and has become an indispensable, powerful analytical technique in this research area. Analysis of protein glycosylation by mass spectrometry is typically achieved by two main approaches: The glycans can be released from the peptide backbone either enzymatically or chemically, or the glycoprotein can be subjected to a protease digestion, producing a mixture of peptides and glycopeptides. The latter approach is advantageous to releasing glycans from the protein, since it does not require extra sample manipulation and allows for site-specific glycosylation profiling [18]. However, there are several obstacles encountered when using a glycopeptide-based MS analysis. For example, glycopeptides exhibit poor ionization efficiency and their signal is usually suppressed by nonglycosylated peptides. In addition, most glycosylation sites contain various glycoforms and each glycoform may exist at low concentration in the total glycopeptide mixture [19, 20]. To obviate these obstacles, it is often necessary to perform an enrichment or chromatographic separation before MS analysis. Several studies have addressed this issue and proposed effective enrichment or/and chromatographic methods that can be utilized before MS analysis of glycopeptides [3, 20].

Although glycopeptide-based MS approaches are often used for glycoprotein analysis, so far there is no consensus as to which MS approach would provide the most glycosylation information, especially for a complex glycoprotein. Recent advances in glycopeptide-based MS analysis have been achieved by two emerging platforms, online high performance liquid chromatography electrospray ionization Fourier transform ion resonance (HPLC/ESI-FTICR)-MS and offline HPLC/MALDI-TOF/TOF. These methods are known for their unique high resolution and high mass accuracy capabilities, along with their ability to accommodate tandem mass spectrometry (MS/MS) experiments. MALDI tandem time of flight (MALDI-TOF/TOF) is widely used partly because it has a higher dynamic range and has a high tolerance to salts and other contaminants. Besides, the complexity of data obtained in ESI-FTICR-MS, due to the presence of multiply charged ions and formation of salt adducts, greatly complicate data interpretation of complex glycopeptide mixtures [21]. However, unlike offline HPLC/MALDI-MS, online HPLC-ESI-FTICR-MS is a powerful technique for analyzing glycopeptide mixtures rapidly, efficiently, and in great detail in a single experiment [22]. Furthermore, glycan-specific ions can be selectively identified from full MS¹ scans and used to trigger subsequent MSⁿ scans during chromatographic separation, thereby providing a plethora of information about the glycopeptides in question [23]. On the contrary, MSⁿ experiments cannot be performed in MALDI-TOF/TOF, which limits the amount of information that can be acquired using this platform. In addition, MALDI analyses suffer from matrix-dependent ionization and fragmentation processes [24–27]. For instance, the “hot” matrix, *α*-cyano 4-hydroxycinnamic acid (CHCA), commonly used with peptides and small neutral glycans, is known to promote formation of glycosidic cleavages, whereas 2,5-dihydroxybenzoic acid (DHB), a “cooler” matrix that is typically used for glycopeptides and glycans, is said to suppress MALDI-induced fragmentation [24]. As a result, the type of matrix used for glycopeptide analysis largely influences the extent and type of fragmentation ions produced during MALDI-MS/MS experiments [26]. Since neither of these two platforms stands out as a clearly superior approach, we performed a head-to-head comparison on both platforms, using a highly complex glycoprotein sample, to investigate the merits and limitations of each method.

Herein, we present a detailed study to investigate the merits of offline HPLC/MALDI-TOF/TOF and online HPLC-ESI-FTICR when used to provide glycosylation information of a recently characterized glycoprotein containing 31 potential glycosylation sites [28]. Specifically, we employed the two platforms to analyze the number of glycopeptides and the quality of MS data obtained from the analysis of the glycoprotein, the group M consensus HIV-1 envelope protein, CON-S gp140 Δ CFI [29]. To ensure that the intrinsic worth of each platform was fully exploited, we determined how well each platform could answer several specific research questions that will eventually contribute in understanding how glycosylation affects the function and immunogenicity of the Env protein. These questions included: How many of the 31 potential glycosylation sites, if glycosylated, could be detected by each technique; what is the extent of glycosylation coverage provided by each platform, for each glycosylation site; what type of confirmatory information can be obtained on both the peptide and glycan portions of the glycopeptides identified using collision induced dissociation (CID) experiments. Our results revealed significant differences in the glycosylation sites detected by using each method, differences in the population of glycoforms identified and the type of structural information obtained on either the peptide or glycan portion of the identified glycopeptides. These results suggest that the two techniques are highly complementary, and when possible, the glycosylation information is maximized by combining the two platforms.

Experimental

Materials and Reagents

Purified CON-S gp140 Δ CFI protein was produced as recombinant vaccinia virus expressed protein from Duke Human Vaccine Research Institute in Durham, as described previously [29]. Urea, ethylenediaminetetraacetic acid (EDTA), dithiothreitol (DTT), iodoacetamide (IAA), HPLC grade acetonitrile (ACN), ammonium bicarbonate, trizma hydrochloride and base, 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 obtained from Promega (Madison, WI). N-Glycosidase F (PNGase F) from *Elizabethkingia meningosepticum* was obtained from CalBioChem (San Diego, CA). Water used for these studies was purified using a Millipore Direct-Q3 Water Purification System (Billerica, MA).

Trypsin Digestion of CON-S gp140 Δ CFI Protein

Approximately 300 μ g of protein (about 2 nmol) was prepared in 100 mM Tris-HCl buffer, containing 6 M urea and 3 mM EDTA, pH 7.5. The protein was reduced for 1 h with 15 mM DTT and alkylated for another hour at room-temperature with 25 mM IAA. The excess amount of IAA was neutralized by adding DTT to a final concentration of 40 mM. Extra buffer solution was added to reduce the concentration of Urea. Trypsin was added at a protein:enzyme ratio of 30:1 (wt/wt) to generate glycopeptides. The protein solution was incubated overnight at 37 °C. The reaction was quenched the following day by adding 1 μ L of concentrated acetic acid. The final protein concentration was ~10 pmol/ μ L. Two aliquots were removed from the total digest and each aliquot was subjected to either online HPLC/ESI-FTICR or offline HPLC fractionation before MALDI-TOF/TOF MS. For each HPLC injection, a protein solution of 1.4 μ g/ μ L was utilized.

Reverse Phase HPLC Fractionation

The tryptic glycopeptides/peptides mixture was purified and separated on a Shimadzu model HPLC system (Shimadzu, Columbia, MD). For each run, 20 μ L of the tryptic digest was injected onto a C18 column (150 \times 4.6 mm, 5 μ M; Alltech, Deerfield, IL) at a flow rate of 1 mL/min. Purified water and HPLC grade ACN each containing 0.1% formic acid were used

as mobile phase A and B, respectively, with a linear gradient from 5% to 40% B over 50 min, followed by a ramp to 95% B in 10 min [30]. Fractions were manually collected every 1 min for 60 min. Each fraction was evaporated to dryness on a CentriVap (Labconco Corporation, Kansas City, MO) before reconstituting with 10 μL of water. The reconstituted fractions were first screened and analyzed by MALDI-TOF/TOF and all fractions containing glycopeptides were then deglycosylated and reanalyzed by MALDI-TOF/TOF MS.

Deglycosylation

Reconstituted HPLC glycopeptide fractions were enzymatically deglycosylated using PNGase F (CalBioChem) by applying the protocol recommended by the manufacturer. Briefly, each enriched glycopeptide fraction was deglycosylated by adding 4 μL of 500 units/mL of PNGase F and 25 μL of 20 mM NH_4HCO_3 (pH = 8), and then incubated overnight at 37 °C. The reaction was stopped by boiling and analyzed by MALDI-MS.

MALDI-TOF/TOF MS Analysis

Equal volumes of saturated solutions of DHB and CHCA matrixes prepared in 50% ACN with 0.1% TFA were used and mixed with each sample (1:1 by volume). Approximately 0.75 μL of the mixture was spotted on a stainless steel MALDI target plate (Applied Biosystems, Foster City, CA) and air-dried. All MALDI MS and MS/MS data were acquired in the positive reflectron ion mode on an Applied Biosystems 4700 proteomics analyzer mass spectrometer. The samples were irradiated by a 355 nm Nd-YAG laser (355 nm) at 200 Hz. The acceleration voltage was 25 kV. Each mass spectrum was generated by averaging 3200 laser shots. The laser intensity was optimized to give the best signal-to-noise ratio and resolution for each sample. All the data were processed in Data Explorer version 4.5 (Applied Biosystems). Glycopeptide analysis was performed by using the high resolution MALDI-TOF/TOF MS in conjunction with our previously described web-based tool (GlycoPep DB), [31] to assign glycopeptide compositions. The assigned compositions were then confirmed by using MALDI-MS/MS experiments.

Capillary HPLC/ESI- FTICR MS Analysis

Analysis of the tryptic glycopeptides on HPLC/ESI-FTICR-MS was performed by using a Dionex Ultimate capillary HPLC system (Sunnyvale, CA) equipped with a Famous well plate autosampler coupled to a high resolution Thermo Finnigan (San Jose, CA) linear ion trap-Fourier transform ion cyclotron resonance (LTQ-FTICR) mass spectrometer equipped with a 7 telsa actively shielded magnet. Samples were loaded onto a Famous well plate autosampler and 5 μL of the tryptic digest was injected onto an LC Packings C18 PeMap 300 column (LC Packings, Sunnyvale, CA) (300 μm i.d. \times 15 cm, 5 μm , 300 \AA). Water and HPLC grade ACN each containing 0.06% formic acid were used as mobile phase A and B, respectively, with a linear gradient starting from 5% to 40% B over 50 min, followed by a ramp to 95% B in 10 min. The eluting solution was directly infused into the mass spectrometer at a flow rate of 5 $\mu\text{L}/\text{min}$.

High resolution data were acquired on FTICR MS by maintaining resolution at 50,000 for m/z 400. The instrument was externally calibrated before the analysis over the entire mass range of interest. The data were acquired in the mass range of m/z 800–2000 using a spray voltage of \sim 4.0 kV. N_2 was used as a nebulizing gas at 20 psi, and the capillary temperature was maintained between 200 and 230 °C. Data were acquired and processed using Xcalibur 1.4 SR1 software (Thermo Finnigan, San Jose, CA). The glycopeptide compositions were assigned using the high resolution data together with GlycoPep DB as described previously [31].

CID Experiments in HPLC/ESI-FTICR MS

All MS/MS data were acquired in the linear ion trap of the hybrid LTQ-FTICR in a data-dependent scanning fashion in the positive ion mode. Data-dependent MS/MS data were acquired for the first three most intense ions observed in full MS¹ scan using a dynamic exclusion window. To maximize the amount of data-dependent MS/MS scans collected for the glycopeptide ions observed in full MS¹ scan data, three more scan events were set with each subsequent scan event selecting the second, third, and fourth most intense ions from MS¹ data. If a neutral loss of a hexose or a HexNAc was detected in these scans, an MS³ scan event was triggered. Each selected precursor ion was activated for 30 ms with q_z value of 0.25 and an isolation width of 3 Da. Activation amplitudes were in the range of 22%–25% as defined by the instrument software.

Data Analysis

To interpret the high resolution data acquired from HPLC/ESI-FTICR MS, several steps were undertaken. The first step was to determine if the peaks observed in MS¹ were glycopeptides or not. To verify this, the lower mass range region of MS/MS data of those peaks were examined for the presence of glycan characteristic product ions like m/z 528 [HexNAc + 2Hex+H]⁺, m/z 690 [HexNAc + 3Hex + H]⁺, m/z 893 [2HexNAc + 3Hex +H]⁺, or m/z 657 [HexNAc + Hex + Sialic Acid + H]⁺. If any of these ions was observed, the next step was to input the MS/MS peak list of the glycopeptides in question into our newly developed web-based tool, GlycoPep ID. A complete description of how this tool operates was provided previously [32]. Briefly, GlycoPep ID uses characteristic fragment ions, such as ^{0,2}X ion [Peptide + 83 + H]⁺ or Y₁ ion [Peptide + 203 + H]⁺, observed in MS/MS to predict the potential peptide portion of the glycopeptides in question. From HPLC/ESI-MS/MS data in the positive ion mode, each of the scan events provided a specific characteristic fragmentation ion, Y₁ glycosidic bond cleavage that occurs at the inner core of *N*-acetyl glucosamine (HexNAc) attached to the peptide, and this ion was automatically predicted by GlycoPep ID, thus identifying the peptide portion of the glycopeptide in question. The identified peptide portion was then imported into GlycoPep DB, described previously [31], which utilizes the high resolution MS¹ peak list to generate all the plausible glycan compositions attached to that specific peptide. All the glycopeptide compositions generated were then inspected manually and verified using MS¹ and MS/MS data. MS/MS data were also used to confirm the assigned carbohydrate compositions.

For MALDI-TOF/TOF data analysis, MALDI-MS/MS data obtained from each glycopeptide fraction was first utilized to identify the ^{0,2}X ion [Peptide + 83 + H]⁺, a characteristic product ion that is typically observed in MALDI-MS/MS corresponding to the peptide plus a portion of the innermost HexNAc residue, which remains after the cross ring cleavage [33]. The identified peptide portion for each fraction was then input into GlycoPep DB and using the high resolution MS¹ peak-list of that fraction, all the plausible glycopeptide compositions could be identified. These glycopeptide compositions were then verified manually using MS¹ and MS/MS data.

Results and Discussion

The group M consensus HIV-1 envelope protein, CON-S gp140ΔCFI, is a very heavily *N*-glycosylated protein with 31 potential glycosylation sites [29]. Figure 1 shows the CON-S gp140ΔCFI protein sequence with all the potential glycosylation sites highlighted in red. The peptides boxed in green represent all the possible tryptic peptides containing one or more potential glycosylation sites produced from this protein, with no missed cleavages. The glycosylation on this protein has recently been described [28], and in that work, glycosylation analysis was demonstrated to be an effective technique in correlating

glycosylation profiles with vaccine efficacy. The work presented here uses this same protein in a case study detailing the relative merits of off-line HPLC, followed by MALDI TOF/TOF MS and on-line HPLC-ESI-LTQ-FTICR-MS for glycopeptide analysis. The protein was subjected to typical sample preparation conditions (reduction/alkylation and tryptic digest) and analyzed using two of the most powerful MS techniques, HPLC/ESI-FTICR-MS and MALDI-TOF/TOF.

Figure 2 illustrates the analytical protocol employed in this study. After the glycoprotein was digested with trypsin, the total digest was divided into two portions. Each portion was subjected to either capillary HPLC/ESI-FTICR MS or HPLC fractionation followed by MALDI-TOF/TOF MS analysis. In addition, the reconstituted HPLC fractions collected for MALDI-TOF/TOF analysis were also deglycosylated and reanalyzed by MALDI-TOF/TOF MS. The glycosylation information content (sequence coverage, number, and type of glycans) obtained from each MS approach was compared to determine the strengths and weaknesses of the two methods.

Assigning Glycopeptide Compositions

One of the key challenges in glycopeptide-based MS analysis is assigning compositions to the masses observed in MS¹ data with a high confidence level. This is because it is often possible to assign different glycopeptide compositions to the same mass, even when mass accuracy is less than 5 ppm [34]. The process used to assign the glycopeptide masses is detailed in the data analysis portion of the Experimental section. Briefly, the steps include: (1) determining whether or not an MS peak is a glycopeptide, by looking for “marker ions” produced when the peak is subjected to MS/MS analysis; (2) determining the peptide composition of the glycopeptide, using MS/MS data and our data analysis tool, GlycoPep ID; (3) determining the remaining portion of the glycopeptide (the glycan mass) by using the high resolution mass data and our analysis tool, GlycoPep DB; (4) confirming the composition with MS/MS data. At the onset of this project, it was not apparent which instrument platform would provide more readily interpretable data; thus, a comparison was undertaken to determine whether HPLC/ESI-FTICR-MS or MALDI-TOF/TOF-MS had advantages in terms of providing the most confirmatory information about the glycopeptide compositions assigned. Figure 3a and b represent MS/MS data of an example glycopeptide, acquired via CID in the HPLC/ESI-LTQ-FTICR and MALDI-TOF/TOF MS platforms, respectively. This glycopeptide demonstrates the relative merits of MS/MS analysis from each technique in providing high confidence assignments for the peptide and glycan compositions.

MS/MS Data from HPLC/ESI-FTICR MS

Figure 3a illustrates MS/MS data acquired in the linear ion trap of the LTQ-FTICR mass spectrometer. In this figure, the precursor ion, m/z 1477.15, was first identified to be a glycopeptide based on the marker ions identified in the MS/MS data as described in the Experimental section. The glycan composition for this ion is confirmed by product ions resulting from glycosidic bond cleavage of this glycopeptide. The glycosidic cleavages are indicated in Figure 3a, and they include sequential losses of nine hexoses (mannose residues), confirming the presence of Man₉, a high-mannose type of N-linked glycan. The glycosidic cleavage resulting from a loss of a HexNAc is represented by a square in Figure 3a. As indicated in this spectrum, the cleavage of all the glycosidic bonds present in this glycopeptide are observed up to the innermost N-acetyl-glucosamine residue (HexNAc), which is attached to the peptide moiety of the glycopeptide. This corresponds to the Y₁ ion [Peptide + 203 + H]⁺, or both [Peptide + 203 + H]⁺ and [Peptide + 349 + H]⁺, when core fucosylation is present. The Y₁ ion is a useful characteristic production, [35] and it is observed in all MS/MS data of the glycopeptides subjected to ESI MS/MS experiments;

since the Y_1 ion corresponds to the peptide attached to the innermost N-acetyl glucosamine (HexNAc), subtracting the mass of this sugar (203 Da) from the observed mass of the Y_1 ion provides the mass of the glycosylated peptide moiety. The product ion corresponding to the Y_1 ion can be identified either manually or by simply inputting the MS/MS peaklist for this glycopeptide (m/z 1477.15) into GlycoPep ID, available at <http://hexose.chem.ku.edu/predictiontable2.php>, which automatically outputs the potential peptide and its corresponding Y_1 ion (see the Experimental section). In this case, GlycoPep ID was used to identify the Y_1 ion, which was identified as m/z 1290.74 (singly charged) and m/z 646.08 (doubly charged ion) and its corresponding peptide moiety, SN⁴⁵³ITGLLLTR. Taken together, the glycosidic cleavage product ions explicitly confirm the glycan portion of this glycopeptide, and the Y_1 ion verifies the peptide composition. However, there were no other product ions resulting from cleavage along the peptide backbone; thus, further confirmation of the peptide sequence was not feasible.

MS/MS Data from MALDI-TOF/TOF MS

Figure 3b represents MS/MS data obtained from MALDI-TOF/TOF of the same glycopeptide shown in Figure 3a. As indicated in Figure 3b, fewer product ions are observed, compared with the ones observed in Figure 3a. These ions include two sets of product ions at or near the innermost HexNAc residues. The two sets of product ions represent the Y_1 ion ($[\text{Peptide} + 203 + \text{H}]^+$) and the $^{0,2}\text{X}$ ion ($[\text{Peptide} + 83 + \text{H}]^+$) [35]. These pairs of ions were always observed in all glycopeptides subjected to MALDI-MS/MS experiments regardless of the type of N-linked glycans (high-mannose, complex, or hybrid) present. Like in MS/MS on the linear ion trap, when core fucosylation is present, the Y_1 ion, corresponding to $[\text{Peptide} + 349 + \text{H}]^+$, is observed, along with the $^{0,2}\text{X}$ ion. In Figure 3b, Y_1 and $^{0,2}\text{X}$ ions are observed at m/z 1290.9 and 1170.8, respectively. Besides these two product ions, there were no other glycan related cleavage ions observed in the MALDI MS/MS experiments [33]. This is because, unlike the low-energy CID in the linear ion trap, MALDI-MS/MS is a high-energy process that yields predominantly fragmentation ions originating from peptide bond cleavages [27]. As a result, MS/MS of the glycopeptide at m/z 2952.55 yields several y and b ions resulting from peptide bond cleavage. Thus, this technique provides detailed sequence and site attachment data for the glycosylated peptide but provides minimal information about the glycan moiety.

Overall, MS/MS acquired from the two high resolution MS techniques, HPLC/ESI-LTQ-FTICR and MALDI-TOF/TOF, provided confirmatory information in that in both methods, the Y_1 ion was always observed. This ion was used as a characteristic ion to identify the peptide moiety of the glycopeptide in question. The remaining mass of the glycopeptide after subtracting the mass of the Y_1 ion can be used to indirectly determine the glycan moiety for that glycopeptide. In MS/MS in the linear ion trap, the characteristic Y_1 ion was always observed as the base peak for glycopeptides containing high-mannose glycan compositions, but was not the base peak for glycopeptides containing complex or hybrid glycans. This is probably because loss of fucose and loss of 1-4 linked N-acetylhexosamines (HexNAc) are both facile cleavages, compared with glycosidic cleavages of high-mannose glycans; therefore, glycosidic cleavages are often observed as the base peak in MS/MS analysis of glycopeptides containing hybrid and complex glycans. An example of MS/MS data of a complex glycopeptide is shown in Supplemental Figure 1.

In MALDI-MS/MS, in addition to the Y_1 ion, the $^{0,2}\text{X}$ ion was also always observed, and either of these two ions formed the base peak. It is worth noting that although these ions identify the peptide moiety of the glycopeptide in question, by providing the mass of the peptide, deglycosylation experiments generally identified the same peptides as identified by the Y_1 ion and the $^{0,2}\text{X}$ ions. This increased the confidence level of the identified peptides from both ESI- and MALDI-MS/MS data. Furthermore, for smaller mass ions ($< m/z$ 5000)

and strongly ionizing peptides, like arginine-containing tryptic peptides, the peptide sequence could easily be obtained from MALDI MS/MS data without deglycosylation. Therefore, MALDI-TOF/TOF provided a higher confidence level for identifying the peptide moiety than the HPLC/ESI-FTICR-MS data. However, in terms of the glycan moiety identified by both MS/MS techniques, HPLC/ESI-FTICR provided a higher confidence level than MALDI-TOF/TOF. When the two MS techniques are used together, extensive information can be obtained about both the peptide sequence and the monosaccharide units contained in the glycan.

Number of Glycoforms Identified

Table 1 shows the number of glycoforms identified at each glycosylation site, detected from both HPLC/ESI-FTICR and MALDI-TOF/TOF MS. This list is reflective of the number of unique glycan *compositions* present at each site only, as no attempt was made to infer glycan *structures* or resolve isomers. The confidence level in these glycan assignments is high, as the monosaccharide units can usually be verified in the ESI-MS/MS data, as described above. A complete list of all the assigned glycan compositions can be found in Supplemental Table 1. The purpose of this table is to show which glycopeptides were identified in MALDI-TOF/TOF versus ESI-LTQ-FT MS analyses. From Table 1, it is quite evident that the number of glycans obtained from each glycosylation site differed greatly between the two instruments. For instance, from HPLC/ESI-FTICR MS data in Table 1, we identified 27 different glycan compositions attached to EANTTLFCASDAK peptide whereas, from the same glycosylation site, only four glycan compositions were identified using MALDI-TOF/TOF MS. However, when another glycosylation site is examined, for example, from the peptide LREHFNN³⁶¹K/EHFNN³⁶¹K, 35 different glycan compositions attached to this site were identified using MALDI-TOF/TOF (Table 1) whereas, from HPLC/ESI-FTICR MS, only eight different glycan compositions were identified from the same site (Table 1). As a result, the number of glycan compositions identified varied from one glycosylation site to the other between the two instruments.

Figure 4 shows a Venn diagram that demonstrates glycan population coverage for MALDI-TOF/TOF and HPLC/ESI-FTICR MS. As indicated in this figure, about 130 unique glycan compositions were identified using each of the two MS techniques alone. About 90 identical glycan compositions were identified by both methods. Overall, ~350 different glycan compositions were identified from all detected glycosylation sites by using HPLC/ESI-FTICR and MALDI-TOF/TOF MS, and this provided the best profile for the glycan population present in CON-S gp140ΔCFI.

Identification of the Most Abundant Type of N-Linked Glycan Present

Table 1 also shows the most abundant type of N-linked glycans identified from each glycosylation site using both HPLC/ESI-FTICR MS and MALDI-TOF/TOF MS. (Isomeric structures of the ones shown in Table 1 are also possible). All three types of N-linked glycans, high-mannose, hybrid, and complex type, were detected from all the identified glycosylation sites; see Supplemental Table 1. Although the number of glycans detected at each site using the two methods differed, in most cases they both provided similar results about the most abundant glycan species present at each site. For example, the total number of glycans found attached to EHFNN³⁶¹K/LREHFNN³⁶¹K using MALDI-TOF/TOF was 35 while from HPLC/ESI-FTICR MS, only eight were detected. However, regardless of the significant difference in number of glycans detected, the same glycan structure ([Hex₉HexNAc₂]) was identified as the most abundant species in both cases. Additionally, seven of the nine glycosylated tryptic peptides identified by both MS methods produced the exact same glycan composition as the most abundant species. The remaining two glycosylated tryptic peptides both contained high-mannose glycans, although the exact

composition varied slightly between the two instruments (see Table 1). From this table, it can also be seen that out of all the glycosylation sites detected by both methods, about 80% of them contained high-mannose N-linked glycans as the most abundant species. As a result, it can be inferred that CON-S gp140 Δ CFI has a high degree of high-mannose N-linked glycan structures.

Glycosylation Site Coverage

Theoretically, digestion of CON-S gp140 Δ CFI with trypsin would produce a total of 19 unique tryptic peptides containing one or more potential glycosylation site(s) (assuming no missed cleavages), which would account for the 31 potential glycosylation sites present in this protein. To determine if any of the two MS techniques could identify all the 31 potential glycosylation sites (19 tryptic peptides), we examined the number of tryptic peptides and their corresponding number of glycosylation sites detected by each MS technique and then compared the results.

Table 2 shows the glycosylated tryptic peptides and the corresponding number of peptides detected from both MS techniques. As shown in this table, from HPLC/ESI-FTICR MS, a total of 14/19 tryptic peptides each containing one or more potential glycosylation sites were detected, which accounted for 18/31 potential glycosylation sites present in this protein. Figure 5a is a representative example of MS¹ data containing glycoforms from two co-eluting tryptic peptides obtained from HPLC/ESI-FTICR MS. As can be seen from this figure, each tryptic peptide contained various glycoforms. A complete list of glycoforms from each of these tryptic peptides can be found in Supplemental Table 1.

In MALDI-TOF/TOF, two analyses were performed in parallel. The first analysis was performed by subjecting each of the reconstituted HPLC fractions to MALDI-TOF/TOF analysis. Figure 5b illustrates an example of MS¹ data from MALDI-TOF/TOF; it contains the same tryptic peptide as that shown in Figure 5a, from the HPLC/ESI-FTICR MS data. The compositions of the glycoforms shown in this figure can be found in Supplemental Table 1.

In the second analysis, PNGase F was used to deglycosylate each of the reconstituted HPLC fractions analyzed in the first experiment. This enzyme releases N-linked glycans from the protein, converting the asparagine residues (N) from which the glycans are removed into aspartic acid (D). As a result, a mass shift of 1 Da is expected to occur for every utilized glycosylation site that is deglycosylated [36]. This experiment was used to determine glycosylation site occupancy [28] (reported in Table 2) and also to confirm glycopeptides whose abundance was low in the high resolution MALDI-TOF/TOF data. For instance, glycosylated tryptic peptides LINCN²⁰¹TSITQACPK and EAN⁴⁸TTLFCASDAK were detected in low abundance in MALDI-TOF/TOF in the first analysis, making it difficult to verify them using MALDI-MS/MS before deglycosylation. However, after deglycosylation (second analysis), these tryptic peptides could be confirmed; more importantly, new tryptic peptides were also detected, such as the two tryptic peptides shown in Table 2c, LTPLCVTLN¹²⁹CTNVN¹³⁵VTN¹³⁸TTN¹⁴¹NTEEK and GEFFYCN³⁹¹TSGLN³⁹⁷STWIG N⁴⁰³GTK, which contain four and three potential glycosylation sites, respectively. These tryptic peptides were also not detected in HPLC/ESI-FTICR MS, probably because of their high masses, when glycosylated.

Another issue with these two peptides that hinders their ionization by MALDI-TOF/TOF is that they are terminated in lysine. The lysine-containing tryptic peptides are known to ionize less efficiently during MALDI analysis than tryptic peptides containing arginine residues [37], and their ionization efficiency is even more compromised since they are multiply glycosylated (and thus large and heterogeneous). Furthermore, with the high resolution of

MALDI-TOF/TOF in the reflectron mode, sensitivity for higher masses is lower than for smaller masses, making it more difficult to detect these tryptic glycopeptides in the first MALDI-TOF/TOF MS analysis. As a result, it is possible to observe these multiply glycosylated tryptic peptides after deglycosylation (second analysis) but not when glycosylated (first analysis). Overall, from MALDI-TOF/TOF analyses, a total of 14/19 potentially glycosylated tryptic peptides, corresponding to 21/31 potential glycosylation sites, were identified.

Since using either one of the two high resolution MS techniques may be inadequate in detecting all the potential glycosylation sites present in a heavily glycosylated protein like CON-S gp140 Δ CFI, a comparison was performed to determine if there was any benefit derived from combining the two high resolution MS techniques in terms of the number of glycosylation sites detected. A closer look at these results obtained from both techniques revealed that only 11/19 potentially glycosylated tryptic peptides, accounting for 15/31 potential glycosylation sites, were identified from both methods (see Table 2a). The remaining three and six glycosylation sites were uniquely identified from either high resolution HPLC/ESI-FTICR or MALDI-TOF/TOF MS, respectively, and are listed in Table 2b. When the numbers of glycosylation sites identified from the two MS techniques are combined, 17/19 tryptic peptides bearing one or more glycosylation sites are identified, resulting in a total of 24 of the 31 potential glycosylation sites. This coverage is higher than the 18 or 21 glycosylation sites obtained from either HPLC/ESI-FTICR or MALDI-TOF/TOF MS alone, thus increasing the probability of detecting as many potential glycosylation sites as possible. Approximately 80% glycosylation coverage was obtained when the two high resolution MS techniques were used together.

In cases where full coverage of all glycosylation sites is desired, lower resolution MS techniques like MALDI MS analysis in the linear mode can be employed. This is because at lower resolution, sensitivity is not as compromised as in high resolution analyses. However, the low resolution analysis results in glycopeptides mass assignments of lower confidence levels since the assignments are based on average masses rather than monoisotopic masses. Furthermore, the assigned glycopeptide compositions cannot be confirmed since MS/MS capabilities are not available in the linear MALDI MS mode. However, sometimes full coverage is the ultimate goal, and low resolution MALDI-MS methods were necessary to afford 100% coverage in this case.

Conclusions

This study demonstrates the use of two high resolution MS techniques; MALDI-TOF/TOF and HPLC/ESI-FTICR MS, to provide glycosylation information of a highly glycosylated protein, CON-S gp140 Δ CFI. CID experiments acquired on both instruments indicated that ESI-MS/MS in a linear ion trap provided the most complete information set for confirming the glycan moieties present, including the acidic monosaccharide units, while MS/MS on a MALDI-TOF/TOF provided higher confidence assignments for confirming the peptide portion of the same glycopeptide. When used together, the two instruments provided a high degree of information about the glycopeptides, affording high confidence assignments. From the high resolution data, 14/19 tryptic peptides were obtained from each MS technique accounting for 18/31 and 21/31 potential glycosylation sites in this protein from HPLC/ESI-FTICR and MALDI-TOF/TOF, respectively. When the two instruments were used to complement each other, 24/31 tryptic peptides accounting for about 80% glycosylation sites coverage was obtained, providing the best glycosylation site coverage. However, to achieve 100% glycosylation coverage, deglycosylation experiments and lower resolution MALDI MS were required.

In terms of glycosylation data, different populations of N-linked glycans comprising a wide range of high-mannose, hybrid, and complex type N-linked glycans were identified and characterized in a glycosylation site-specific manner. Overall, the high-mannose glycans were identified as the most abundant glycoforms from both MS techniques. Approximately 350 glycopeptide compositions were identified, when data from the two techniques were combined. The information presented in this study provides other researchers with useful insight about what MS methods are most appropriate for glycopeptide analysis, and how those methods can be used synergistically to provide optimal glycosylation coverage and high confidence assignments.

Acknowledgments

The authors acknowledge the National Institutes of Health for funding (project numbers RO1GM077266 and PO1AI61734). This work was also supported by collaboration from the AIDS Vaccine Development Grant from the Bill and Melinda Gates Foundation to B.F.H. The authors also thank Dr. Todd Williams of the University of Kansas for his support and for allowing them to use mass spectrometry facilities in KU.

References

1. Apweiler R, Hermjakob H, Sharon N. On the Frequency of Protein Glycosylation, as Deduced from Analysis of the SWISS-PROT Database. *Biochim Biophys Acta*. 1999; 1473:4–8. [PubMed: 10580125]
2. Moens S, Vanderleyden J. Glycoproteins in Prokaryotes. *Arch Microbiol*. 1997; 168:169–175. [PubMed: 9382700]
3. Morelle W, Michalski J-C. Glycomics and Mass Spectrometry. *Curr Pharm Des*. 2005; 11:2615–2645. [PubMed: 16101462]
4. Baenziger JU. The Role of Glycosylation in Protein Recognition. *Am J Pathol*. 1985; 121:382–391. [PubMed: 3934981]
5. Rudd PM, Wormald MR, Dwek RA. Sugar-Mediated Ligand-Receptor Interactions in the Immune System. *Trends Biotechnol*. 2004; 22:524–530. [PubMed: 15450746]
6. Wang L-X. Toward Oligosaccharide- and Glycopeptide-Based HIV Vaccines. *Curr Opin Drug Discov Devel*. 2006; 9:194–206.
7. Geyer H, Holschbach C, Hunsmann G, Schneider J. Carbohydrates of Human Immunodeficiency Virus. Structures of Oligosaccharides Linked to the Envelope Glycoprotein 120. *J Biol Chem*. 1988; 263:11760–11767. [PubMed: 2841333]
8. Koch M, Pancera M, Kwong PD, Kolchinsky P, Grundner C, Wang L, Hendrickson WA, Sodroski J, Wyatt R. Structure-Based, Targeted Deglycosylation of HIV-1 gp120 and Effects on Neutralization Sensitivity and Antibody Recognition. *Virology*. 2003; 313:387–400. [PubMed: 12954207]
9. Leonard CK, Spellman MW, Riddle L, Harris RJ, Thomas JN, Gregory TJ. Assignment of Intrachain Disulfide Bonds and Characterization of Potential Glycosylation Sites of the Type 1 Recombinant Human Immunodeficiency Virus Envelope Glycoprotein (gp120) Expressed in Chinese Hamster Ovary Cells. *J Biol Chem*. 1990; 265:10373–10382. [PubMed: 2355006]
10. Mizuochi T, Matthews TJ, Kato M, Hamako J, Titani K, Solomon J, Feizi T. Diversity of Oligosaccharide Structures on the Envelope Glycoprotein gp120 of Human Immunodeficiency Virus 1 from the Lymphoblastoid Cell Line H9. Presence of Complex-Type Oligosaccharides with Bisecting N-Acetylglucosamine Residues. *J Biol Chem*. 1990; 265:8519–8524. [PubMed: 2341393]
11. Back NK, Smit L, De Jong JJ, Keulen W, Schutten M, Goudsmit J, Tersmette M. An N-Glycan Within the Human Immunodeficiency Virus Type 1 gp120 V3 Loop Affects Virus Neutralization. *Virology*. 1994; 199:431–438. [PubMed: 8122371]
12. Cheng-Mayer C, Brown A, Harouse J, Luciw PA, Mayer AJ. Selection for Neutralization Resistance of the Simian/Human Immunodeficiency Virus SHIV_{SF33A} Variant in Vivo by Virtue of Sequence Changes in the Extracellular Envelope Glycoprotein that Modify N-Linked Glycosylation. *J Virol*. 1999; 73:5294–5300. [PubMed: 10364275]

13. Kwong PD, Doyle ML, Casper DJ, Cicala C, Leavitt SA, Majeed S, Steenbeke TD, Venturi M, Chaiken I, Fung M, Katinger H, Parren PWIH, Robinson J, Van Ryk D, Wang L, Burton DR, Freire E, Wyatt R, Sodroski J, Hendrickson WA, Arthos J. HIV-1 Evades Antibody-Mediated Neutralization Through Conformational Masking of Receptor-Binding Sites. *Nature*. 2002; 420:678–682. [PubMed: 12478295]
14. Reitter JN, Means RE, Desrosiers RC. A Role for Carbohydrates in Immune Evasion in AIDS. *Nat Med*. 1998; 4:679–684. [PubMed: 9623976]
15. Wyatt R, Kwong PD, Desjardins E, Sweet RG, Robinson J, Hendrickson WA, Sodroski JG. The Antigenic Structure of the HIV gp120 Envelope Glycoprotein. *Nature*. 1998; 393:705–711. [PubMed: 9641684]
16. Zhu X, Borchers C, Bienstock RJ, Tomer KB. Mass Spectrometric Characterization of the Glycosylation Pattern of HIV-gp120 Expressed in CHO Cells. *Biochemistry*. 2000; 39:11194–11204. [PubMed: 10985765]
17. Yeh J-C, Seals JR, Murphy CI, van Halbeek H, Cummings RD. Site-Specific N-Glycosylation and Oligosaccharide Structures of Recombinant HIV-1 gp120 Derived from a Baculovirus Expression system. *Biochemistry*. 1993; 32:11087–11099. [PubMed: 8218172]
18. Morelle W, Canis K, Chirat F, Faïd V, Michalski J-C. The Use of Mass Spectrometry for the Proteomic Analysis of Glycosylation. *Proteomics*. 2006; 6:3993–4015. [PubMed: 16786490]
19. Wada Y, Azadi P, Costello CE, Dell A, Dwek RA, Geyer H, Geyer R, Kakehi K, Karlsson NG, Kato K, Kawasaki N, Khoo K-H, Kim S, Kondo A, Lattova E, Mechref Y, Miyoshi E, Nakamura K, Narimatsu H, Novotny MV, Packer NH, Perreault H, Peter-Katalinic J, Pohlentz G, Reinhold VN, Rudd PM, Suzuki A, Taniguchi N. Comparison of the Methods for Profiling Glycoprotein Glycans—HUPO Human Disease Glycomics/Proteome Initiative Multi-Institutional Study. *Glycobiology*. 2007; 17:411–422. [PubMed: 17223647]
20. Budnik BA, Lee RS, Steen JAJ. Global Methods for Protein Glycosylation Analysis by Mass Spectrometry. *Biochim Biophys Acta*. 2006; 1764:1870–1880. [PubMed: 17118724]
21. Mechref Y, Novotny MV. Structural Investigations of Glycoconjugates at High Sensitivity. *Chem Rev*. 2002; 102:321–369. [PubMed: 11841246]
22. Wührer M, Koeleman CAM, Hokke CH, Deelder AM. Protein Glycosylation Analyzed by Normal-Phase Nano-Liquid Chromatography Mass Spectrometry of Glycopeptides. *Anal Chem*. 2005; 77:886–894. [PubMed: 15679358]
23. Carr SA, Huddleston MJ, Bean MF. Selective Identification and Differentiation of N- and O-Linked Oligosaccharides in Glycoproteins by Liquid Chromatography-Mass Spectrometry. *Protein Sci*. 1993; 2:183–196. [PubMed: 7680267]
24. Stephens E, Maslen SL, Green LG, Williams DH. Fragmentation Characteristics of Neutral N-Linked Glycans Using a MALDI-TOF/TOF Tandem Mass Spectrometer. *Anal Chem*. 2004; 76:2343–2354. [PubMed: 15080747]
25. Burlingame AL. Characterization of Protein Glycosylation by Mass Spectrometry. *Curr Opin Biotechnol*. 1996; 7:4–10. [PubMed: 8742373]
26. Kuroguchi M, Nishimura SI. Structural Characterization of N-Glycopeptides by Matrix-Dependent Selective Fragmentation of MALDI-TOF/TOF Tandem Mass Spectrometry. *Anal Chem*. 2004; 76:6097–6101. [PubMed: 15481958]
27. Wührer M, Catalina MI, Deelder AM, Hokke CH. Glycoproteomics Based on Tandem Mass Spectrometry of Glycopeptides. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2007; 849:115–128.
28. Go EP, Irungu J, Zhang Y, Dalpathado DS, Liao H-X, Sutherland LL, Alam SM, Haynes BF, Desaire H. Glycosylation Site-Specific Analysis of HIV Envelope Proteins (JR-FL and CON-S) Reveals Major Differences in Glycosylation Site Occupancy, Glycoform Profiles, and Antigenic Epitopes' Accessibility. *J Proteome*. 2008; 7(4):1660–1674.
29. Liao H-X, Sutherland LL, Xia S-M, Brock ME, Scearce RM, Vanleeuwen S, Alam SM, McAdams M, Weaver EA, Camacho ZT, Ma B-J, Li Y, Decker JM, Nabel GJ, Montefiori DC, Hahn BH, Korber BT, Gao F, Haynes BF. A Group M Consensus Envelope Glycoprotein Induces Antibodies that Neutralize Subsets of Subtype B and C HIV-1 Primary Viruses. *Virology*. 2006; 353:268–282. [PubMed: 17039602]

30. Cutalo JM, Deterding LJ, Tomer KB. Characterization of Glycopeptides from HIV-1_{gp120} by Liquid Chromatography Mass Spectrometry. *J Am Soc Mass Spectrom.* 2004; 15:1545–1555. [PubMed: 15519221]
31. Go EP, Rebecchi KR, Dalpathado DS, Bandu ML, Zhang Y, Desaire H. GlycoPep DB: A Tool for Glycopeptide Analysis Using a “Smart Search”. *Anal Chem.* 2007; 79:1708–1713. [PubMed: 17297977]
32. Irungu J, Go EP, Dalpathado DS, Desaire H. Simplification of Mass Spectral Analysis of Acidic Glycopeptides Using GlycoPep ID. *Anal Chem.* 2007; 79:3065–3074. [PubMed: 17348632]
33. Wuhler M, Hokke CH, Deelder AM. Glycopeptide Analysis by Matrix-Assisted Laser Desorption/Ionization Tandem Time-of-Flight Mass Spectrometry Reveals Novel Features of Horseradish Peroxidase Glycosylation. *Rapid Commun Mass Spectrom.* 2004; 18:1741–1748. [PubMed: 15282773]
34. Irungu J, Dalpathado DS, Go EP, Jiang H, Ha H-V, Bousfield GR, Desaire H. Method for Characterizing Sulfated Glycoproteins in a Glycosylation Site-Specific Fashion, Using Ion Pairing and Tandem Mass Spectrometry. *Anal Chem.* 2006; 78:1181–1190. [PubMed: 16478110]
35. Dalpathado DS, Desaire H. Glycopeptide Analysis by Mass Spectrometry. *Analyst.* 2008; 133:731–738. [PubMed: 18493671]
36. Carr SA, Roberts GD. Carbohydrate Mapping by Mass Spectrometry: A Novel Method for Identifying Attachment Sites of Asn-Linked Sugars in Glycoproteins. *Anal Biochem.* 1986; 157:396–406. [PubMed: 3096166]
37. Brancia FL, Oliver SG, Gaskell SJ. Improved Matrix-Assisted Laser Desorption/Ionization Mass Spectrometric Analysis of Tryptic Hydrolysates of Proteins Following Guanidination of Lysine-Containing Peptides. *Rapid Commun Mass Spectrom.* 2000; 14:2070–2073. [PubMed: 11085420]

MRV R G I Q R N C Q H L W R W G T L I L G M L M I C S A A E N L W V T V Y Y G V P V W K E A N T T
 L F C A S D A K A Y D T E V H N V W A T H A C V P T D P N P Q E I V L E N V T E N F N M W K N N M V
 E Q M H E D I I S L W D Q S L K P C V K L T P L C V T L N C T N V N V T N T T N N T E E K G E I K N
 C S F N I T T E I R D K K Q K V Y A L F Y R L D V V P I D D N N N S S N Y R L I N C N T S A I T Q
 A C P K V S F E P I P I H Y C A P A G F A I L K C N D K K F N G T G P C K N V S T V Q C T H G I K P
 V V S T Q L L L N G S L A E E E I I I R S E N I T N N A K T I I V Q L N E S V E I N C T R P N N N T
 R K S I R I G P G Q A F Y A T G D I I G D I R Q A H C N I S G T K W N K T L Q Q V A K K L R E H F N
 N K T I I F K P S S G G D L E I T T H S F N C R G E F F Y C N T S G L F N S T W I G N G T K N N N N
 T N D T I T L P C R I K Q I I N M W Q G V G Q A M Y A P P I E G K I T C K S N I T G L L L T R D G G
 N N N T N E T E I F R P G G G D M R D N W R S E L Y K Y K V V K I E P L G V A P T K A K L T V Q A R
 Q L L S G I V Q Q Q S N L L R A I E A Q Q H L L Q L T V W G I K Q L Q A R V L A V E R Y L K D Q Q L
 E I W D N M T W M E W E R E I N N Y T D I I Y S L I E E S Q N Q Q E K N E Q E L L A L D K W A S L W
 N W F D I T N W L W

Figure 1.

The protein sequence for CON-S gp140 Δ CFI with all the 31 potential glycosylation sites highlighted in red. The peptides boxed in green represent all the potentially glycosylated tryptic peptides present in this protein with no missed cleavages.

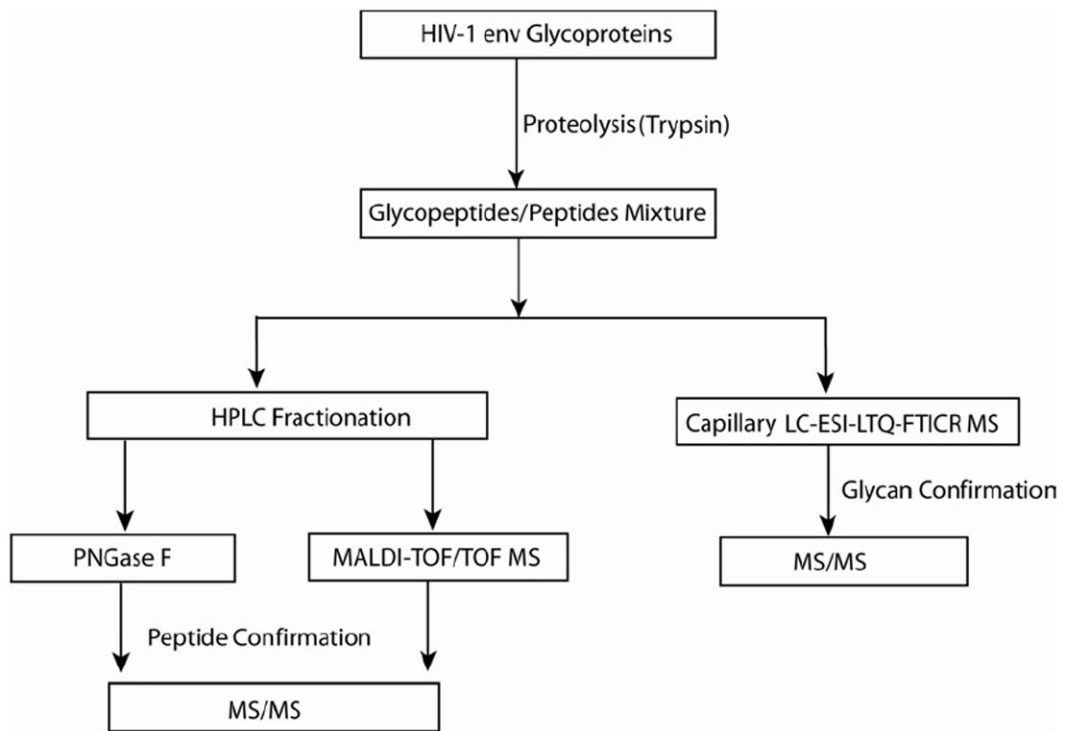


Figure 2. Work flow to illustrate the protocol used to analyze CON-S gp140 Δ CFI glycopeptides.

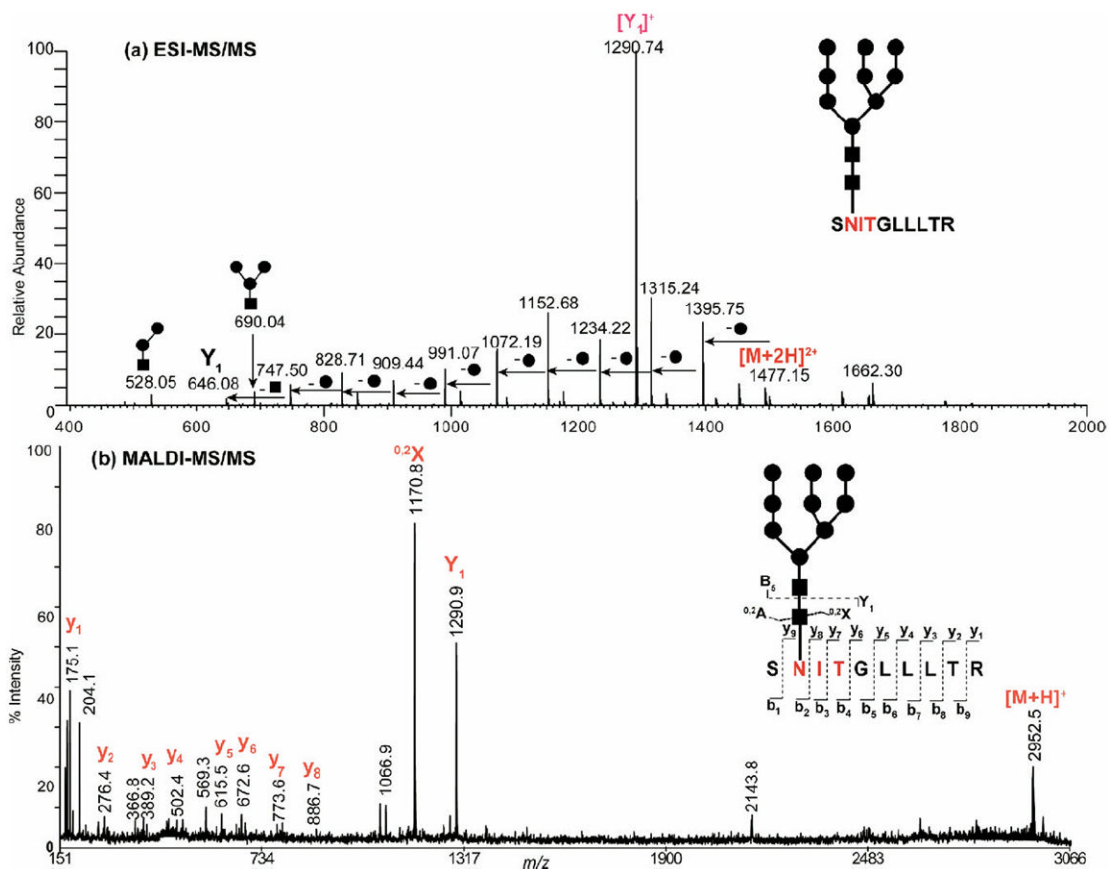


Figure 3. A representative example of MS/MS data used to confirm the assigned glycopeptide compositions. **(a)** ESI-MS/MS data for a doubly charged glycopeptide ion at m/z 1477.15. **(b)** MALDI-MS/MS data of the singly charged form of the same glycopeptide (m/z 2952.55) as in **(a)**.

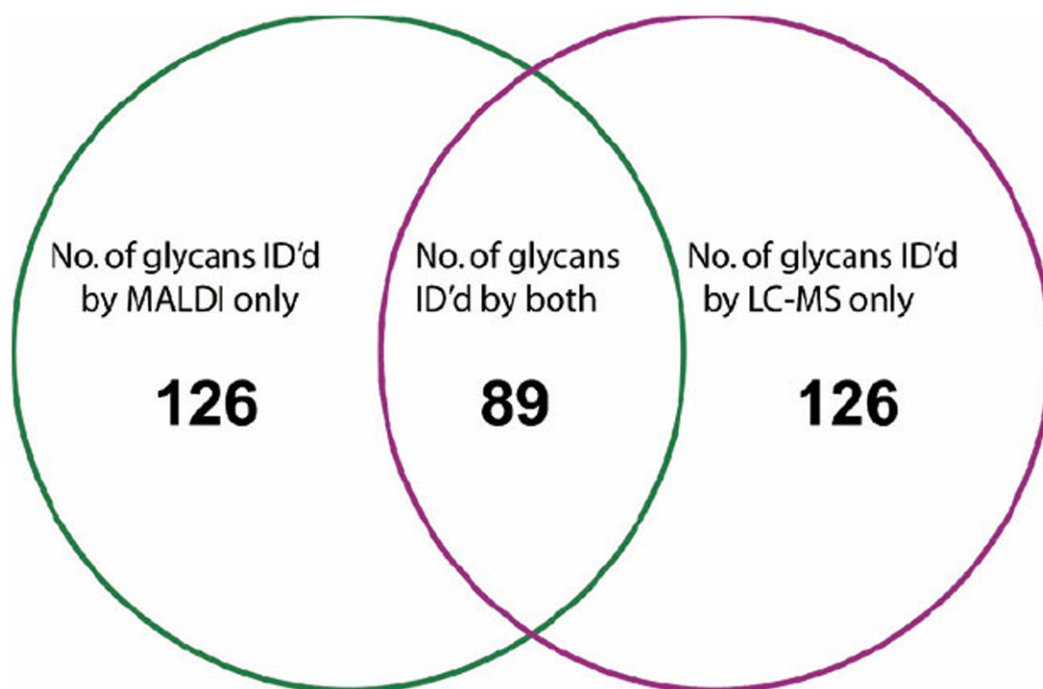


Figure 4. Venn diagram indicating the number of glycans detected by either high resolution HPLC/ESI-FTICR-MS, MALDI-TOF/TOF MS, or both.

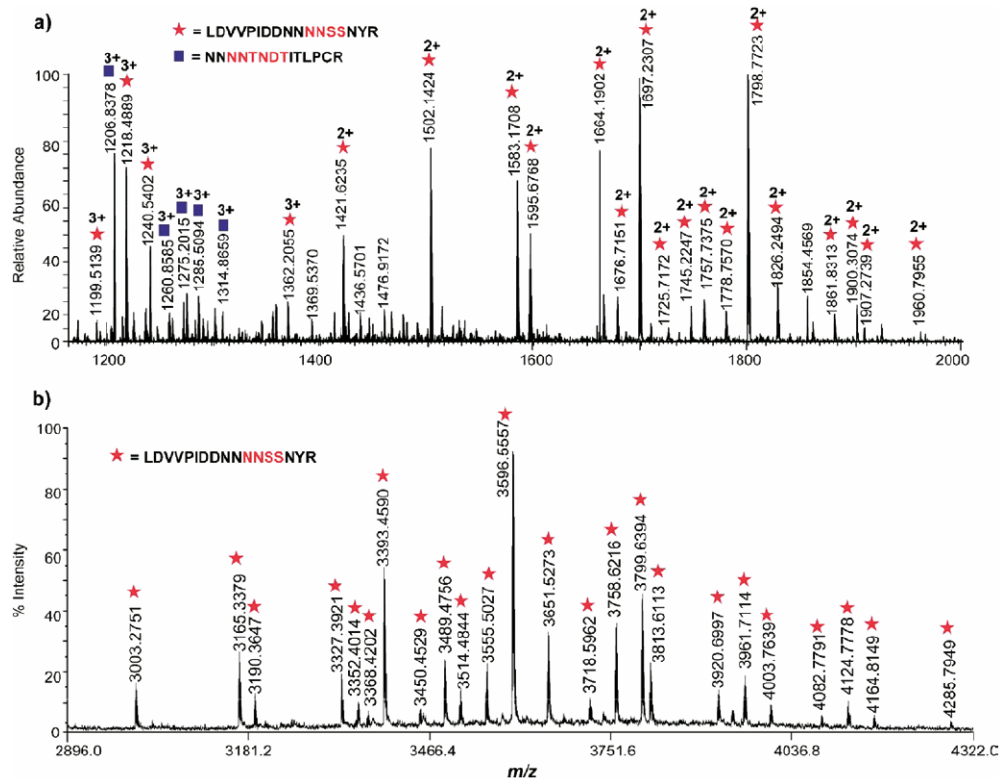





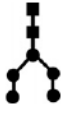


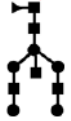
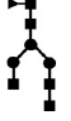












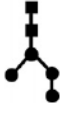



Figure 5. Representative examples of MS data containing the same glycopeptide peaks for CON-S gp140ΔCFI acquired on ESI-FTICR-MS and MALDI-TOF/TOF-MS. (a) MS data from ESI, (b) MALDI mass spectrum for the similar glycopeptide peaks as in (a).

Table 1

A summary of type and number of glycoforms identified from CON-S gp140ΔCFI

Identified peptide	LC-ESI/FTICR-MS			HPLC/MALDI-TOF/TOF		
	No of glycans	Most abundant structure	m/z	No of glycans	Most abundant structure	m/z
ENGTGPCK/CNDKKFNGTGPCK	4		1373.0222 (+2)	12		3373.3042 (+1)
EHFNK/ LREHFNK	8		1327.0049 (+2)	35		2922.1624 (+1)
QAHCNISGTK	5		1490.5868 (+2)	6		2315.0049 (+1)
SENTNNAK	11		1428.0684 (+2)	1		2206.8865 (+1)
NNNNINDTITLPCR	27		1206.8378 (+2)	23		3294.3725 (+1)
DGGNNNTNETEIPRGGGDMR	20		1300.2075 (+3)	45		3898.6094 (+1)
LDVVPIDDDNNNNSSNYR	20		1199.5139 (+3)	32		3596.5557 (+1)
NCSFNITTEIR	32		1610.1519 (+2)	26		3219.3145 (+1)
SNITGLLLTR	21		1476.6450 (+2)	12		2952.2661 (+1)
WNKTLQQVAKK/WNK	2		1156.4409 (+2)	-	-	-

Identified peptide	LC-ESI-FTICR-MS			HPLC/MALDI-TOF/TOF		
	No of glycans	Most abundant structure	m/z	No of glycans	Most abundant structure	m/z
EANTLFCASDAK	27		1646.6523 (+2)	4	N/A	-
LINCNTSAITQACPK	24		1778.2159 (+2)	3	N/A	-
AYDTEVHNVWATHACVPTDPN PQEIVLENVTFENFMWK	14		1584.1990 (+4)	-	-	-
TIIVQLNESVEINCTRPNNTNR	-	-	-	7		3639.7512 (+1)
NVSTVQCTHGKFPVSTQLLLN GSLAEHHIR	-	-	-	9		5483.6548 (+1)

N/A, Not Applicable, glycopeptide peaks were very low in abundance making it impossible to identify the most abundant type of glycoform present.

Table 2

Glycosylation sites coverage from high-resolution MS

	Potential sites/occupied sites ^a
Tryptic peptides identified by HPLC/ESI-FTICR and MALDI-TOF/TOF MS	
FN ²³⁷ GTGPCK/CNDKKFN ²³⁷ GTGPCK	1
EHFNN ³⁶¹ K/LREHFNN ³⁶¹ K	1
QAHCN ³³⁷ ISGTK	1
SEN ²⁸⁰ ITNNAK	1
NNN ⁴¹³ NTN ⁴¹⁶ DTITLPCR	2/(1 and 2)
DGGNN ⁴⁶⁶ NTN ⁴⁶⁹ ETEIFRPGGDMR	2/(1 and 2)
LDVVPIDNNN ¹⁹⁰ N ¹⁹¹ SSNYR	2/1
N ¹⁵⁵ CSFN ¹⁵⁹ ITTEIR	2/(1 and 2)
SN ⁴⁵³ ITGLLLTR	1
^b EAN ⁴⁸ TTLFCASDAK	1
^b LINCN ²⁰¹ TSAITQACPK	1
Unique tryptic peptides detected only in HPLC/ESI-FTICR MS	
WN ³⁴⁴ KTLQVAKK/ WN ³⁴⁴ K	1
AYDTEVHNVWATHACVPTDPNPQEIVLEN ⁸⁷ VTENFNMWK	1
EINN ⁶⁴³ YTDIIYSLIEESQNQKEK	1
Unique peptides detected only in MALDI-TOF/TOF MS	
TIIVQLN ²⁹³ ESVEIN ²⁹⁹ CTRPNN ³⁰⁵ NTR	3/(1 and 2)
N ²⁴⁵ VSTVQCTHGKIPVSTQLLN ²⁶⁶ GSLAEEIIIIR	2/(1 and 2)
DQQLIWDN ⁶³¹ MTWMEWER	1
Tryptic peptides undetected by both high resolution methods	
LTPLCVTLN ¹²⁹ CTNVN ¹³⁵ VTN ¹³⁸ TTN ¹⁴¹ NTEEK	4/(2 and 3)
GEFFYCN ³⁹¹ TSGLFN ³⁹⁷ STWIGN ⁴⁰³ GTK	3/(2 and 3)

^aFor peptides with more than one potential glycosylation site, the site occupancy was determined by deglycosylation with PNGase F and MS/MS on resulting peptides. Experimental details are described in reference [28].

^bPeptide sequences detected in low abundance in MALDI-TOF/TOF; verified by deglycosylation with PNGase F and MS/MS on resulting peptides.