Published in final edited form as:
J Proteome Res. 2008 April ; 7(4): 1660–1674. doi:10.1021/pr7006957.

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

Eden P. Go[†], **Janet Irungu**[†], **Ying Zhang**[†], **Dilusha S. Dalpathado**[†], **Hua-Xin Liao**[‡], **Laura L. Sutherland**[‡], **S. Munir Alam**[‡], **Barton F. Haynes**[‡], and **Heather Desaire**[†,*]

[†]Department of Chemistry, University of Kansas, Lawrence, Kansas 66045

[‡]Duke Human Vaccine Institute, Duke University Medical Center, Durham, North Carolina 22710

Abstract

The HIV-1 envelope (Env) is a key determinant in mediating viral entry and fusion to host cells and is a major target for HIV vaccine development. While Env is typically about 50% glycan by mass, glycosylation sites are known to evolve, with some glycosylation profiles presumably being more effective at facilitating neutralization escape than others.[1] Thus, characterizing glycosylation patterns of Env and native virions and correlating glycosylation profiles with infectivity and Env immunogenicity are necessary first steps in designing effective immunogens. Herein, we describe a mass spectrometrybased strategy to determine HIV-1 Env glycosylation patterns and have compared two mammalian cell expressed recombinant Env immunogens, one a limited immunogen and one that induces crossclade neutralizing antibodies. We have used a glycopeptide-based mass mapping approach to identify and characterize Env's glycosylation patterns by elucidating which sites are utilized and what type of glycan motif is present at each glycosylation site. Our results show that the immunogens displayed different degrees of glycosylation as well as a different characteristic set of glycan motifs. Thus, these techniques can be used to (1) define glycosylation profiles of recombinant Env proteins and Env on mature virions, (2) define specific carbohydrate moieties at each glycosylation site, and (3) determine the role of certain carbohydrates in HIV-1 infectivity and in modulation of Env immunogenicity.

Keywords

HIV; envelope glycoprotein; glycosylation; vaccine; mass spectrometry

Introduction

The viral defense mechanisms against anti-HIV-1 neutralizing antibodies and the structural complexity of HIV envelope (Env) proteins present an unprecedented challenge in the development of an effective vaccine against AIDS.[2–5] The evolution rate of HIV-1 quasispecies coupled with the evolution of Env escape mutants prevents the anti-HIV-1 antibody response from neutralizing HIV-1.[6,7] While CTL responses are crucial in limiting

[*]To whom correspondence should be addressed. hdesaire@ku.edu.

Supporting Information Available: A complete list of the glycopeptide compositions for JR-FL and CON-S envelope proteins (Supplementary Tables 1A, 1B, 2A, and 2B). This material is available free of charge via the Internet at http://pubs.acs.org.

virus replication, the induction of broadly neutralizing antibodies against a broad spectrum of HIV isolates remains to be the best option for protective immunity.[8–11] Despite recent advances in vaccine research, an immunogen that could elicit high levels of broadly neutralizing antibodies has not yet been developed.[4,5,11,12]

The HIV-1 envelope trimer that mediates viral entry is extensively glycosylated with at least 24 potential N-linked glycosylation sites spread throughout the conserved and the variable regions.[13,14] The transmembrane subunit gp41, which anchors gp120, promotes the insertion of the viral RNA to the host cell and is known to have 4–5 potential N-linked glycosylation sites located at the extraviral domain.[15] Glycosylation plays a pivotal role in HIV pathogenesis by impacting the proper folding and functional conformation of the envelope spike, thereby influencing the protein's antigenicity and immunogenicity.[16–20] In addition, glycosylation is involved in the HIV immune evasive mechanisms that include conformational masking of epitopes and glycan shielding.[1,21–25] Thus, the loss or gain of potential glycosylation sites can significantly alter the biological activity of the envelope spike.[23,26–30] One of the most promising targets for induction of anti-HIV-1 neutralizing antibodies is Env carbohydrates, as they comprise ~50% of the Env mass.[13,31] However, HIV-1 carbohydrates are not immunogenic and may vary with the cell type in which the Env is produced.

To date, the glycosylation pattern of gp120 of several HIV strains has been defined.[13,14,32–37] Seminal studies on the gp120 glycosylation pattern of the first prototype HIV strain, IIIB, derived from Chinese hamster ovary (CHO) cells show that the protein is comprised of high mannose, complex, and hybrid type glycans.[13] The study revealed that all of the 24 glycosylation sites were utilized with 13 sites occupied by complex type glycans and 11 sites with high mannose and/or hybrid type glycans. A similar glycosylation pattern within the 24 consensus sites was also observed in another CHO cell derived strain (SF2) with 26 glycosylation sites.[36,37] Identified complex type structures were diverse spanning from bi- to tetraantennary structures with varying degrees of sialylation and fucosylation. Notably, SF2 glycoforms were found to exist as two distinct clusters, one with mostly high mannose and the other complex type localized on distinct domains of the envelope surface.[36]

While these studies provided the benchmark for elucidating gp120 glycosylation profiles, the CHO cell derived monomeric form of gp120 used in VAXGEN phase III HIV vaccine trial was not effective.[38,39] The glycosylation on the failed vaccine candidate is well-characterized, but it is highly probable that different vaccine candidates, constructed from different sequences, produced in different cell lines, and/or present in oligomeric forms, would display different glycosylation profiles. This notion is supported by the fact that different cell types display different types of glycosylation patterns.[35,40–42] In addition, glycan attachment is not always 100% efficient resulting in variable site occupancy.[43–45] Thus, host cell specific variations that include processing of complex glycans[35,41] and the variation in the distribution of glycosylation sites can modulate the antigenic and immunogenic presentation of the protein to the immune system. Therefore, much work remains in the characterization of glycosylation profiles of Env immunogens to reveal the correlation between glycosylation and immunological response.

Most efforts on the effect glycosylation on virus' susceptibility to neutralization have focused on the deletion of glycosylation sites flanking the receptor binding site, the V1/V2 loop, and the V3 loop.[24,46–50] Variation in global glycosylation pattern of the Env proteins retrieved from the Los Alamos National Laboratory (LANL) HIV database has been reported,[51] but the variation in glycosylation contributing to Env protein immunogenicity

has not been fully explored. Such assessment is beneficial in the design of an efficacious Env-based immunogen.

To correlate glycosylation profiles with immunological response, one must identify a form of the Env protein with immunogenic properties. An immunogen design approach using synthetic Env genes generated from a centralized consensus gene sequences in the LANL HIV sequence database used in combination with DNA/recombinant protein vaccines has shown some success in this area.[52–57] These synthetic Env genes express the soluble secreted oligomeric forms of the Env protein that have the same antigenic, immunogenic, and functional properties as the native HIV-1 Env proteins when used in DNA vaccines. Notably, out of the six synthetic immunogens generated so far, three Env genes generated high titers of neutralizing antibodies. One of these Env genes was CON-S which elicited neutralizing antibody response spanning three M clades when used in DNA/recombinant protein vaccine in animal models.[54]

While the use of a centralized gene approach has provided valuable insights regarding the antigenic and immunogenic properties of the synthetic Env proteins, the role of glycosylation and its contribution to the antigenic/immunogenic activity have not been defined. To begin to prove the relevance of the pattern of glycosylation of Env to Env antigenicity and immunogenicity, we have developed the strategy to profile Env glycosylation patterns, and determined glycosylation species of two recombinant HIV-1 Env oligomers, one a poor immunogen for neutralizing antibodies, and one that induces a more robust cross-clade neutralizing antibody response. We characterized the glycosylation pattern of an rVV derived synthetic Env immunogen, CON-S, representing the centralized sequence of the group M consensus, by mass spectrometry and performed a comparative glycosylation study with the wild-type clade B Env protein, JR-FL. This comparison provides an avenue to understand how changes in glycosylation influence the immunogenicity of the Env proteins. We used glycopeptidebased mass spectral analysis to elucidate the glycosylation pattern of the recombinant glycoproteins, CON-S and JR-FL. Proteolytic digests of the glycoproteins were examined directly with matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS) and liquid chromatography electrospray ionization Fourier Transform ion cyclotron resonance mass spectrometry (LC/ESI-FTICR MS). Comparison of the glycosylation profile of CON-S and JR-FL at each glycosylation site reveals characteristic glycan patterns with processed (complex and hybrid) type and more high mannose structures for CON-S and predominantly processed (complex and hybrid) type structures for JR-FL. One key glycosylation feature of these Env proteins is that not all glycosylation sites are utilized. The information gained from this study can be used to correlate variation in glycosylation patterns with respect to the breadth of neutralizing antibody response induced by each immunogen. Identification of the immunogen's specific glycosylation footprint that elicits broadly neutralizing antibody response will be an informative addition to current paradigms used in the design of effective immunogens.

Materials and Methods

Materials and Reagents

All reagents were obtained in high purity from Sigma-Aldrich except when noted otherwise. Ammonium bicarbonate (NH_4HCO_3), trizma hydrochloride, trizma base, ethylenediaminetetraacetic acid (EDTA), phosphate buffered saline solution (PBS), acetic acid, HPLC grade acetonitrile (CH_3CN), methanol (CH_3OH), 2,5-dihydroxybenzoic acid (DHB), urea, α-cyano-4-hydroxycinnamic acid (CHCA), iodoacetamide (IAA), Sepharose CL-4B, butanol, ethanol, dithiothreitol (DTT), trifluoroacetic acid (TFA), and formic acid were purchased from Sigma (St. Louis, MO). Water was purified using a Millipore Direct-

Q3 Water Purification System (Billerica, MA). Sequencing grade trypsin (Tp) and *N*-glycosidase F (PNGase F) from *Elizabethkingia meningosepticum* were obtained from Promega (Madison, WI) and Calbiochem (San Diego, CA), respectively.

Expression and Purification of Envelope Glycoproteins (JR-FL and CON-S)

All Env proteins were expressed and purified from the Duke Human Vaccine Research Institute in Durham, NC. The Env proteins were constructed, expressed, and purified using the method described in literature.[53,54]

Tryptic Digestion

Samples containing 300 μg aliquots of the HIV-1 Env proteins, JR-FL gp140ΔCF and CON-S gp140ΔCFI, were denatured with 6 M urea in 100 mM Tris buffer (pH 7.5) containing 3 mM EDTA. The proteins were reduced and alkylated with 15 mM DTT at room temperature for 1 h and 40 mM IAA at room temperature in the dark for another hour, respectively. The samples were brought to a final concentration of 50 mM DTT to neutralize excess IAA. Proteins were then digested at 37 °C with trypsin at a protein/enzyme ratio of 30:1 (w/w) overnight, followed by a second trypsin digestion under the same conditions. The resulting HIV Env protein digest mixture was either subjected to off-line, reversed-phase, high-performance liquid chromatography (RP-HPLC) fractionation for MALDI analysis or RP-HPLC ESI-FT MS analysis.

Offline RP-HPLC Fractionation

A 20 μL aliquot of the HIV Env protein digest mixture was injected onto a C18 column (150 mm × 4.6 mm, 5 μm size column particle, Alltech, Deerfield, IL). Mobile phases utilized for the fractionation were A, 99.9% deionized H_2O and 0.1% formic acid, and B, 99.9% CH_3CN and 0.1% formic acid. The following CH_3CN/H_2O multistep gradient was used to elute the glycopeptides from the column at a flow rate of 1 mL/min: 5% mobile phase B for 3 min, followed by a linear increase to 40% B in 15 min, a 15 min hold at 40% B, then a linear increase to 85% B in 20 min. The column was then held at 85% B for 7 min before re-equilibration. Fractions were collected every minute for 60 min. The HPLC fractions were dried in a centrivap (Labconco Corporation, Kansas City, MO) and reconstituted with 10 μL of H_2O. A total of 60 fractions were analyzed by MALDI MS.

N-Deglycosylation

HIV glycopeptide enriched fractions were deglycosylated using *N*-glycosidase F (CalBioChem) utilizing the protocol recommended by the manufacturer. Briefly, a solution containing 500 units/mL of *N*-glycosidase F was prepared by adding 100 μL of deionized H_2O to 50 units of lyophilized enzyme in a vial. Glycans were released in each glycopeptide enriched fraction by adding 25 μL of 20 mM NH_4HCO_3 (pH) 8) and 4 μL of *N*-glycosidase F solution. The reaction was incubated overnight at 37 °C and was stopped the following day by heating the sample to 100 °C. The resulting solution was subsequently analyzed by MALDI MS.

Mass Spectrometry and Liquid Chromatography

MALDI MS and MS/MS experiments were performed on an Applied Biosystems 4700 Proteomics Analyzer mass spectrometer (Foster City, CA) operated in the positive ion mode. Samples were prepared by mixing equal volumes of the analyte and matrix solutions in a microcentrifuge tube, then immediately deposited on a MALDI plate, and allowed to dry in air. The matrix solution was prepared by mixing equal volumes of saturated solutions of CHCA in 50% CH_3CN in H_2O with 0.1% TFA and DHB in 50% CH_3CN in H_2O. Samples were irradiated with an ND:YAG laser (355 nm) operated at 200 Hz. This

instrument was equipped with automated and multisampling capabilities for rapid sample analysis. Mass spectra were acquired in both reflectron and linear ion modes and were generated by averaging 3200 individual laser shots into a single spectrum. Each spectrum was accumulated from 80 shots at 40 different locations within the MALDI spot. The laser intensity was optimized to obtain adequate signal-to-noise (S/N) ratio and resolution for each sample. MALDI MS/MS data were acquired using a collision energy of 1 kV with nitrogen as collision gas.

LC/ESI-FTICR MS and MS/MS experiments were performed using a hybrid linear ion-trap (LIT) Fourier Transform ion cyclotron resonance mass spectrometer (LTQ-FT, ThermoElectron, San Jose, CA) directly coupled with a Dionex UltiMate capillary LC system (Sunnyvale, CA) equipped with FAMOS well plate autosampler. Mobile phases utilized for the experiment consisted of solvent A, 99.9% deionized H_2O and 0.1% formic acid, and solvent B, 99.9% CH_3CN and 0.1% formic acid, which were pumped at a flow rate of 5 μL/min. Samples were injected into a C18 PepMap 300 column (300 μm i.d. × 15 cm, 300 Å, LC Packings, Sunnyvale, CA). After loading 5 μL of sample, glycopeptides were eluted at a flow rate of 5 μL/min using the same gradient described above. A short wash and blank run were performed between every sample to ensure that there was no sample carryover. The hybrid LIT-FT-MS was operated in a data-dependent scanning mode with scan event details as follows: A full FT-MS scan within the mass range *m/z* 800–2000 followed by three data-dependent MS/MS scans of the three most intense glycopeptide molecular ions from the full MS scan were sequentially and dynamically selected for subsequent collision-induced dissociation (CID) in the LTQ linear ion trap using a normalized collision energy of 35% and a 3 min dynamic exclusion window. When the data-dependent MS/MS scan detects a neutral loss corresponding to monosaccharide units (hexose or HexNAc), an MS^3 scan event is triggered for the parent ion. The temperature for the capillary of the ion source was set at 200 °C and an ESI voltage of 4.0 kV was used, scanning in the positive ion mode.

Glycopeptide Identification

Tandem mass spectra generated from MALDI MS and ESI-MS analyses were analyzed using GlycoPep DB[58] and GlycoPep ID[59] to elucidate the HIV Env protein tryptic glycopeptide compositions. The details of the glycopeptide compositional analysis have been described previously. Briefly, the peptide portion is determined using the collision induced dissociation (CID) data. For MALDI MS analysis, CID data from glycopeptide enriched fractions were inspected manually to identify the characteristic signature fragment ion of the cross-ring cleavage, $^{0,2}X$ ion. This ion was used to verify the peptide portion of the glycopeptide. Once the peptide was identified, the peak list of the MS data from the fraction in which the peptide sequence was ascertained was searched against all carbohydrate entries in the GlycoPep DB database. The query result generated all plausible glycopeptide compositions which were subsequently refined and verified by evaluating the MS and MS/MS data. The compositional assignment for ESI FTICR-MS data is realized using GlycoPep ID to determine the peptide portion of the glycopeptide and GlycoPep DB for glycopeptide composition. Identification of peptide portion is facilitated using GlycoPep ID, in which a peptide prediction table is generated from a theoretical digest of the glycoprotein of interest with their corresponding sequence and *m/z* values as well as a list of predicted *m/z* values of the cross-ring cleavages, $^{0,2}X$ or $^{0}Y_1$ ions.

Results and Discussion

CON-S and JR-FL Envelope Glycoproteins

Modified forms of the synthetic Env immunogen, CON-S, representing the group M and the wild-type clade B Env protein, JR-FL were used in this study due to the marked improvement in immunogenicity and the protein's ability to express soluble oligomeric protein.[53,54,60] On the basis of the full-length sequence of gp160 of the reference HIV strain, HXB2, the gene construct of the modified form of CON-S (gp140ΔCFI) was constructed with three internal deletions that included the gp120/gp41 proteolytic cleavage site (C, residues 510–511), fusion domain of gp41 (F, residues 512–527), and the region between two heptad repeats (residues 546–579, 628–655) in the immunodominant region (I) and shortened variable loops, whereas the modified form of JR-FL (gp140ΔCF) lacked the gp120/gp41 proteolytic cleavage site (C), and the fusion domain (F) of gp41.[60] Both of the Env immunogens were lectin-purified and were propagated in rVV.[53,54] The full sequence alignment of CON-S gp140ΔCFI and JR-FL gp140ΔCF is shown in Figure 1 with the potential glycosylation sites shown in red. The protein sequence of CON-S gp140ΔCFI and JR-FL gp140ΔCF is 81% identical. As can be seen in Figure 1, the Env proteins differ in nine potential glycosylation sites. Missing potential glycosylation sites in JR-FL gp140ΔCF but present in CON-S gp140ΔCFI and vice versa are shown in boxes. Glycosylation analysis using N-glycosite (http://www.hiv.lanl.gov/) reveals 31 and 27 potential glycosylation sites for CON-S gp140ΔCFI and JR-FL gp140ΔCF, respectively. Throughout the text, CON-S gp140ΔCFI and JR-FL gp140ΔCF are referred as CON-S and JR-FL, respectively.

Glycosylation Mapping by Mass Spectrometry

The global mapping and comprehensive identification of glycosylation using glycopeptide-based MS analysis is perhaps one of the most challenging tasks in glycoproteomics. This difficulty mainly stems from the low ionization efficiency of glycopeptides, the extensive glycan heterogeneity at each glycosylation site, the relatively lower glycopeptide concentration compared to peptides, and the complexity of data analysis. Central to any successful MS analysis is the design and choice of sample preparative aides. Recently, we performed a thorough evaluation of several chromatographic and enrichment methods used for glycopeptide-based MS analysis and optimized them to markedly improve the glycopeptide coverage.[61] Accordingly, we tailored an optimized sample preparation method that is well-suited for the analysis of the extensively glycosylated recombinant HIV Env proteins, CON-S and JR-FL. Figure 2 provides a schematic representation of the experimental template used in this study. Our approach was to integrate both MALDI MS and LC/ESI FTICR MS analyses to obtain a global profile of the glycosylation and distinguish differences in the glycosylation profile between the two Env immunogens.

CON-S and JR-FL both have ~600 amino acid residues. The reduced Env proteins have 15 cysteine residues, which were alkylated at the protein level by reacting the cysteine residues with IAA producing carbamidomethylated Env proteins. The carbamidomethylated Env proteins were subjected to all operations typical of an in-solution trypsin digestion generating about 100 possible tryptic fragments when allowing for up to one internal trypsin cleavage site (single missed cleavage). After digestion with trypsin, two separate aliquots of the glycoprotein digest were either subjected to HPLC fractionation for MALDI MS analysis or LC/ESI FTICR MS analysis (Figure 2). For both MS experiments, glycopeptides were separated within an 80-min gradient. Out of the 60 HPLC fractions collected for each sample for MALDI analysis, there were 36 and 38 fractions that contained glycopeptides for JR-FL and CON-S, respectively. These fractions were subjected to MS/MS analysis to deduce the glycopeptide composition present in each fraction. A portion of each

glycopeptide fraction was treated further with PNGase F to validate the peptide assignment and determine which of the potential glycosylation sites were occupied (Figure 2). Glycopeptide analysis using LC/ESI-FTICR-MS and MS/MS analysis was performed using data-dependent acquisition mode with the hybrid LTQ linear ion trap. Glycopeptides were identified by locating clusters of peaks whose characteristic mass differences correspond to the masses of the monosaccharide units (hexose, HexNAc, etc.) in the ESI FTICR-MS data. We have identified all of the 31 and 27 potential glycosylation sites for CON-S and JR-FL from our analysis (Table 1).

Glycosylation Profiles of CON-S and JR-FL

A total of 19 and 16 tryptic glycopeptides with both single and multiple glycosylation sites were identified for CON-S and JR-FL, respectively (see Figure 1). The glycosylation profiles, which include the glycan motif and site occupancy, were compared for CON-S and JR-FL. A partial list of the glycopeptide compositions is shown in Table 2 and the complete list is found in Supplementary Tables 1A, 1B, 2A and 2B in Supporting Information. Representative MALDI MS and LC/ESI-FTICR MS and the corresponding tandem MS spectra of the glycopeptide fractions shown in Figure 3 depict the tryptic glycopeptides in the V2 and V3 regions for JR-FL (Figure 3A) and CON-S (Figure 3B), respectively. Glycopeptide compositions were deduced after identifying the peptide portion of the glycopeptide from MS/MS data (Figure 3 inset),and the glycan portion was verified using the mass list from each spectrum. Compositional analysis generated ~500 putative identifications for each immunogen per MS technique used. The query results were refined by evaluating the MS spectra manually and further confirming the assignments with all the available MS/MS data. Overall, we have identified over 300 unique glycopeptide compositions per immunogen consisting of high mannose, hybrid, and complex type N-linked glycans. This high level of coverage of glycosylation heterogeneity is unprecedented.

To differentiate the glycosylation motifs between the synthetic and wild-type Env immunogen and facilitate analysis, the glycoforms were broadly categorized into two groups, namely, high mannose and processed glycans. High mannose glycans consist of mannose-containing structures with 5–9 mannose sugars, whereas processed glycans include all hybrid and complex type structures. The processed glycans were counted according to the following criteria: hexose (Hex) ≥ 3 and *N*-acetylglucosamine (HexNAc) ≥ 4 or Hex ≥ 4 and HexNAc ≥ 3. Figure 4A shows a summary of the glycosylation data populating each site. A closer look indicates that the CON-S and JR-FL differ in their site occupancy and the glycosylation pattern particularly surrounding the immunodominant V3 loop. To elucidate how these two differences in glycosylation profiles can affect the immunogenicity of the Env proteins, a detailed comparison of the glycosylation of the Env proteins was made, and its implications to vaccine design are discussed below.

A. Open Glycosylation on CON-S and JR-FL

The identified glycopeptides isolated from the tryptic digests of CON-S and JR-FL accounted for 31 and 27 potential glycosylation sites, respectively. While the glycopeptide mass mapping experiment allowed for the identification of glycopeptides, it cannot directly predict the site occupancy. Several identified glycopeptides contain more than one glycosylation site (see Table 1). To identify which sites were occupied on these glycopeptides, the glycopeptides were enzymatically deglycosylated to determine the site occupancy.[62–65] The enzymatic release of glycans converts asparagine in the NXT/S to aspartic acid resulting in a mass shift of 1 Da per site. Figure 5A contains representative MS data of a deglycosylated glycopeptide fraction for CON-S. Four deglycosylated glycopeptides exhibiting a mass increment of 1 Da each were detected, indicating that each of these glycopeptides contains one utilized glycosylation site. Tandem MS analysis of these

peptides was performed to validate the peptide sequence as well as to determine site utilization for peptides with multiple glycosylation sites. MS/MS data (Figure 5B) of the tryptic peptide, NNN^{413}NTN^{416}DTITLPCR, in the V4 loop shows which of the two potential glycosylation sites is occupied. The N^{413}NT site is occupied, as evidenced by the fact that this asparagine has been converted to aspartic acid. The second site, N^{416}DT, must be unutilized, since the MS/ MS data clearly indicate that this residue is still an asparagine, after PNGase F treatment. Analysis of another glycopeptide fraction containing this peptide shows that both N^{413}NT and N^{416}DT are utilized. Thus, N^{416}DT is variably occupied, as both glycosylated and nonglycosylated asparagines (N^{416}) were identified by MALDI MS from PNGase F treated glycopeptide fractions. Overall, our results revealed that out of the 31 potential glycosylation sites detected for CON-S (Table 1), glycosylation sites at N141 in the C1–V1 region, N191 in the V1/V2 region, and N631 and N643 in the transmembrane region were not utilized at any time, and nine sites that included N135 in C1–V1 region, N159 and N201 in the V1/V2 region, N245 and N293 in the conserved region 2 (C2), N305 in the V3 region, N344 in the conserved region 3 (C3), N416 in the V4 loop, and N466 in the V5 loop were variably utilized (Figure 4B, top). For JR-FL, the sites at N138 in the V1 region, N192 in the V1/V2 region, and two sites at N617 (or 622) N643 in the transmembrane region were not utilized and the site at N159 was variably utilized (Figure 4B, bottom). It should be noted that, for the long peptide portion in the V4 region, two sites were occupied out of five potential glycosylation sites. We were not able to determine which sites were occupied by MS/MS experiments due to the inherent low ionization efficiency of this high mass species. Table 1 shows the list of identified tryptic glycopeptides with open and variable glycosylation sites for both CON-S and JR-FL.

The observed variation in the degree of glycosylation occupancy for both immunogens can be attributed to several factors that commonly affect the efficiency of protein glycosylation. It is known that the extent of protein glycosylation is governed by the primary structure of X in the consensus glycosylation site, NXT/ S, the amino acid residue flanking the NXT/S, the structural conformation of the local environment surrounding NXT/S, and the availability of the array of key glycosylating enzymes during glycan biosynthesis/processing.[43,45,66] Close examination of the amino acid sequence surrounding NXT/S for each glycopeptide in Table 1 indicates that the heterogeneity in site occupancy can be attributed to the conformation of the local environment surrounding NXT/S and the amino acid at position X. For instance, the absence of glycosylation on the second asparagine residue for the glycopeptide, LDVVPIDDNNN^{190}N^{191}SSNYR, in the V2 region for CON-S is most likely due to steric occlusion. The same trend is observed for a similar glycopeptide on JR-FL. After the transfer of N-linked glycans on N190 (CON-S) or N191 (JR-FL), the glycan addition to N191 or N192 is not favorable due to steric hindrance. Inefficient glycosylation at N135, N141, N159, N245, N293, N305, N397, N416, N466, and N631 for CON-S and N138 and N159 for JR-FL may also be due to steric hindrance of occupied glycosylation sites in close proximity. Glycosylation efficiency is also regulated by the presence of large hydrophobic residues (W, L, F, and Y) and negatively charged residues (E and D) at position X in the consensus, NXT/S.[67,68] Thus, the lack of glycosylation at N643, for both CON-S and JR-FL, located at the transmembrane region is likely influenced by the presence of tyrosine (Y) at the X position. Inefficient glycosylation at the variably occupied site at N293 before the V3 region and N466 in the V5 region for CON-S is most likely influenced by the presence of glutamic acid (E) at the X position.

In addition to the primary sequence effecting glycosylation site occupancy, it is also quite possible that the cell line used to express the protein contributes to the number of open glycosylation sites in CON-S and JR-FL. Previous analyses, which have shown that all the potential glycosylation sites on Env are occupied, were conducted on proteins expressed in CHO cells.[13,36] In contrast, JR-FL and CON-S were both expressed in 293-T cell line. Other

glycoproteins have been described that contain dramatically different glycosylation patterns, when the proteins are obtained from different mammalian species.[41,42] If the differences in cell line used contribute to the differences in glycosylation site occupancy in Env, this would imply that changing cell lines for an immunogen could have an effect in its glycosylation profile and, as a result, its immunogenicity.

The existence of open and variable glycosylation can directly affect the immunogenicity of the Env proteins. Indeed, the susceptibility of the virus toward antibody neutralization is dependent on whether the unutilized sites could either effectively or ineffectively expose key protein epitopes. An increase in antibody neutralization was observed when the glycosylation sites in the V3 regions and near the receptor binding region were deglycosylated.[47,50] Our results show that more open sites surround the V1/V2, C2, and V3 regions for CON-S, compared to the less effective immunogen, JR-FL.

B. Glycosylation and Protein Structure

In addition to determining whether the glycosylation sites are occupied, glycosylation analysis also provides some insight into the three-dimensional (3D) conformation of the protein, and its structural heterogeneity *in vivo*. Given that the appended carbohydrates could either be heavily processed or not, depending on the site's accessibility to key glycosylating enzymes, the variability in glycosylation processing provides a probe of the protein's local structure at a particular glycosylation site. High mannose glycans indicate minimal processing and a more protected local 3D structure, whereas fully processed glycans indicate that the glycosylation site was readily accessible to glycosyltransferase enzymes in the Golgi. In fact, high mannose glycans have been previously shown to help stabilize the local 3D conformation of proteins.[45,69,70] Of course, Env immunogens must be able to adapt stable 3D conformations so they can provide conserved structural epitopes that elicit strong immunological response. To probe differences in protein structure between JR-FL, a poor immunogen, and CON-S, a more effective immunogen, we highlight two key regions/segments where glycosylation is different between the two proteins.

The first region where the difference in glycosylation was observed is the region before the V1 loop. The glycosylation site is located in the first conserved (C1) region. The sequence alignment of the 62-residue segment of the glycopeptide identified in the C1 region of CON-S and JR-FL is 95% identical (Figure 6A). Since the two sites have identical sequence, one would reasonably expect that these sites are equally exposed and accessible to glycosylation enzymes. However, our glycosylation data clearly shows different glycan motifs on CON-S and JR-FL (Figure 6B). CON-S glycans consist of high mannose structures with 5–9 mannose (Man) sugars and processed glycans which are mostly complex type structures with some degree of sialylation and a minimal amount of fucosylation. In contrast, JR-FL glycans consist of a single high mannose structure (Man5) and a greater number of processed glycoforms, most of which have been more fucosylated than CONS. It should be noted that both immunogens are rVV expressed proteins and have undergone the same sample processing prior to MS analysis, so the expression and analysis conditions cannot be used to explain the glycosylation differences detected, and neither can the primary sequence. The results suggest that the difference in glycosylation is dictated by the structural conformation of the local environment surrounding the glycosylation site. Certain conformations would be more favorable for processing; in this case, JR-FL's higher degree of processing indicates that it must have a different 3D structure at this site, compared to CON-S. The relatively low accessibility of the same glycosylation site for CON-S evidenced by the presence of more high mannose structures implies that this site is partially buried in the folded protein. Thus, it is likely that this particular region for CON-S has lower protein flexibility allowing more stable 3D conformation in this region. This hypothesis is supported

by the precedent that the extent of glycosylation of a particular site reflects the particular local conformation of the protein surrounding the glycosylation site.[44,71]

The next region where CON-S and JR-FL differ in glycosylation is the region surrounding the V3 loop (Figure 4). The V3 loop spans a 35-residue segment connected by a disulfide bond and this region is known to be a determinant for HIV tropism and receptor binding. This region is glycosylated before and after the loop. Our data shows that the sites at N245, N266, N293, and N299 in the C2 region and N305 in the V3 loop for CON-S are variably unoccupied as discussed in the preceding section. When glycosylated, these sites were generally occupied by high mannose, hybrid type glycans, and smaller complex type structures (Figure 7). In contrast, corresponding sites in JR-FL are fully glycosylated with larger processed glycoforms. The mere presence of a large population of high mannose glycans on these sites for CON-S gp140ΔCFI suggests that these sites went through very minimal processing (i.e., glucose trimming) before the protein exited the Golgi. These sites are either completely or partially buried within the protein backbone after folding and thereby less accessible to glycosyltransferase enzymes for processing in the Golgi. With more high mannose structures on these regions, the protein flexibility is also reduced.[44,71] While the V3 region of CON-S is less accessible to glycosyltransferase enzymes, this degree of inaccessibility could ultimately be a reason why the V3 region of CON-S elicits more neutralizing antibodies, compared to JR-FL. Since JR-FL is more accessible to glycosyltransferase enzymes, its glycosylation is substantially larger than the glycosylation on CON-S. As a result of JR-FL's heavier glycosylation both in the number of sites occupied and in the size of the glycans present, the protein sequence in this local area is covered more effectively by glycans, masking key epitopes from antibodies. In contrast, the protein sequence in the V3 loop of CON-S is generally less shielded and key epitopes are more accessible to antibodies.

Implications to Vaccine Design

The first step in understanding how glycosylation influences Env's immunogenicity is to define its global glycosylation profile. Distinguishing the differences in glycosylation provide insights on how the glycan profiles affect the functional conformation of the protein, necessary for eliciting potent immune response. While the overall degree of glycosylation within isolates and across clades is conserved to maintain the glycan shield, glycosylation continues to evolve to effectively mask underlying epitopes and perhaps eliminate nonself glycosylation patterns generated by the host cell glycosylation machinery to evade immune recognition.[71,72] Therefore, the design of an effective Env protein-based immunogen as a vaccine component should require a detailed analysis of the differences in glycosylation profile between immunogens to improve vaccine development.

In an effort to accomplish this goal, we have characterized the glycosylation of the synthetic Env protein CON-S representing group M with shortened variable loops derived from clade C and the wild-type clade B JR-FL and performed a comparative study of their glycosylation profile. In the context of amino acid sequence of these two Env proteins, they differ in nine potential glycosylation sites. Such differences could affect glycosylation efficiency and how proteins fold in general. While the differences in amino acid sequence between Env's provides an avenue to understand protein structure and glycosylation, defining the specific glycosylation footprint at each site provides additional insights as to why one vaccine candidate is more immunogenic than the other. Our analysis shows a substantial difference in glycosylation in terms of the degree and the type of glycosylation pattern between CON-S and JR-FL. This difference can be correlated to differences in protein structure and ultimately immunogenicity. Our results indicated that the more immunogenic Env protein has more unutilized sites surrounding the V1, V2, C2, andV3

regions and has more high mannose structures as well as smaller processed type glycoforms in the C2 and the immunodominant V3 and V4 regions.

An effective Env immunogen must have a low degree of structural heterogeneity to allow better neutralization of underlying structural epitopes and the glycosylation of CON-S suggests that its structure is more highly conserved than JR-FL. Specifically, CON-S contains a higher degree of high mannose glycans in the C2 domain and V4 region, along with minimally processed glycoforms and high mannose structures in the V3 loop. This observation is reflective of the presence of more occluded glycosylation sites surrounding the C2, V3, and V4 regions. Since the high mannose glycoforms are known to reduce protein flexibility, these glycans are likely to promote protein stability and preserve specific protein configuration in these regions. In addition, the presence of more unutilized glycosylation sites surrounding these regions indicates that the key protein epitopes are more exposed, which would assist in eliciting antibody response. From the data obtained thus far, the glycosylation features that appear to add to CON-S's enhanced immunogenicity include the number of open glycosylation sites,and the regions containing high-mannose glycans in the early part of the sequence, which correlate to a more well-conserved protein structure. These key findings are consistent with recent immunology data on CON-S gp140ΔCFI where high titers of antibodies were elicited when used in DNA vaccine in small animal models.[54] Further study and refinement of the correlation between glycosylation and immunogenicity will provide the opportunity to enable identification of certain glycosylation footprints of Env proteins that will promote the induction of antibodies to a broad spectrum of HIV-1 isolates. Such study is a step toward improving HIV vaccine design/development. We are currently characterizing other synthetic as well as wild-type Env immunogens and correlating the glycosylation pattern with immunologic activity in small animal models.

Conclusions

A glycopeptide-based mass mapping approach was used to characterize the glycosylation of two Env protein vaccine candidates in a glycosylation site-specific fashion. Our results show that the two Env proteins have different glycosylation site occupancy and different characteristic sets of glycan motifs populating each glycosylation site. CON-S and JR-FL are the first two proteins shown to contain unoccupied potential glycosylation sites in the Env, and CON-S has a particularly high level of unoccupied sites: 19/31 are unoccupied at least some of the time. The open sites could be present in these proteins in part because a 293-T cell line was used as the expression system for both CON-S and JRFL. Additionally, the higher level of unoccupied sites on CON-S, compared to JR-FL, could be due, in part, to its unique primary sequence.

Unraveling the differences in glycosylation provided important biological insights as to why CON-S may be more immunogenic than JR-FL. The characteristic features of CON-S include more unoccupied sites and glycosylation sites that are populated with smaller glycoforms and/or high mannose structures. Such a glycosylation pattern would render better accessibility of antigenic epitopes to neutralizing antibodies. Together with the immunological data, glycosylation sitespecific analysis is one avenue of research that could provide information in directing antibody response and perhaps in the future hold the key in designing effective immunogen against the HIV virus.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH grant RO1GM077226, PO1AI61734, and a collaboration from the AIDS Vaccine Development Grant from the Bill and Melinda Gates Foundation to Barton F. Haynes.

References

1. Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, Salazar-Gonzalez JF, Salazar MG, Kilby JM, Saag MS, Komarova NL, Nowak MA, Hahn BH, Kwong PD, Shaw GM. Nature. 2003; 422(6929):307–312. [PubMed: 12646921]
2. Letvin NL, Barouch DH, Montefiori DC. Annu. Rev. Immunol. 2002; 20:73–99. [PubMed: 11861598]
3. Letvin NL. Annu. Rev. Med. 2005; 56:213–223. [PubMed: 15660510]
4. McMichael AJ. Annu. Rev. Immunol. 2006; 24:227–255. [PubMed: 16551249]
5. Singh M. Virol. J. 2006; 3:60. [PubMed: 16939652]
6. Evans DT, Desrosiers RC. Immunol. Rev. 2001; 183:141–158. [PubMed: 11782254]
7. Humbert M, Dietrich U. AIDS Rev. 2006; 8(2):51–59. [PubMed: 16848273]
8. Burton DR, Desrosiers RC, Doms RW, Koff WC, Kwong PD, Moore JP, Nabel GJ, Sodroski J, Wilson IA, Wyatt RT. Nat. Immunol. 2004; 5(3):233–236. [PubMed: 14985706]
9. Zolla-Pazner S. Nat. Rev. Immunol. 2004; 4(3):199–210. [PubMed: 15039757]
10. Burton DR, Stanfield RL, Wilson IA. Proc. Natl. Acad. Sci. U.S.A. 2005; 102(42):14943–14948. [PubMed: 16219699]
11. Haynes BF, Montefiori DC. Expert Rev. Vaccines. 2006; 5(4):579–595. [PubMed: 16989638]
12. Letvin NL. Nat. Rev. Immunol. 2006; 6(12):930–939. [PubMed: 17124514]
13. Leonard CK, Spellman MW, Riddle L, Harris RJ, Thomas JN, Gregory TJ. J. Biol. Chem. 1990; 265(18):10373–10382. [PubMed: 2355006]
14. Mizuochi T, Matthews TJ, Kato M, Hamako J, Titani K, Solomon J, Feizi T. J. Biol. Chem. 1990; 265(15):8519–8524. [PubMed: 2341393]
15. Johnson WE, Sauvron JM, Desrosiers RC. J. Virol. 2001; 75(23):11426–11436. [PubMed: 11689624]
16. Li Y, Luo LZ, Rasool N, Kang CY. J. Virol. 1993; 67(1):584–588. [PubMed: 8416385]
17. Jeffs SA, McKeating J, Lewis S, Craft H, Biram D, Stephens PE, Brady RL. J. Gen. Virol. 1996; 77:1403–1410. [PubMed: 8757980]
18. Hebert DN, Zhang JX, Chen W, Foellmer B, Helenius A. J. Cell Biol. 1997; 139(3):613–623. [PubMed: 9348279]
19. Land A, Braakman I. Biochimie. 2001; 83(8):783–790. [PubMed: 11530211]
20. Saunders CJ, McCaffrey RA, Zharkikh I, Kraft Z, Malenbaum SE, Burke B, Cheng-Mayer C, Stamatatos L. J. Virol. 2005; 79(14):9069–9080. [PubMed: 15994801]
21. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Nature. 1998; 393(6686):648–659. [PubMed: 9641677]
22. Reitter JN, Means RE, Desrosiers RC. Nat. Med. 1998; 4(6):679–684. [PubMed: 9623976]
23. Poignard P, Saphire EO, Parren PWHI, Burton DR. Annu. Rev. Immunol. 2001; 19:253–274. [PubMed: 11244037]
24. Kwong PD, Doyle ML, Casper DJ, Cicala C, Leavitt SA, Majeed S, Steenbeke TD, Venturi M, Chaiken I, Fung M, Katinger H, Parren PW, Robinson J, Van RD, Wang L, Burton DR, Freire E, Wyatt R, Sodroski J, Hendrickson W, Arthos J. Nature. 2002; 420(6916):678–682. [PubMed: 12478295]
25. Kwong PD. Nature. 2005; 433(7028):815–816. [PubMed: 15729327]
26. Lee W, Syu W, Du B, Matsuda M, Tan S, Wolf A, Essex M, Lee T. Proc. Natl. Acad. Sci. U.S.A. 1992; 89(6):2213–2217. [PubMed: 1549584]
27. Fenouillet E, Gluckman JC, Jones IM. Trends Biochem. Sci. 1994; 19(2):65–70. [PubMed: 8160267]

28. Hemming A, Bolmstedt A, Jansson B, Hansen JE, Travis B, Hu SL, Olofsson S. Arch. Virol. 1994; 134(3–4):335–344. [PubMed: 8129620]
29. Hemming A, Gram GJ, Bolmstedt A, Losman B, Hansen JE, Ricksten A, Olofsson S. Arch. Virol. 1996; 141(11):2139–2151. [PubMed: 8973529]
30. Perrin C, Fenouillet E, Jones IM. Virology. 1998; 242(2):338–345. [PubMed: 9514971]
31. Pantophlet R, Burton DR. Annu. Rev. Immunol. 2006; 24:739–769. [PubMed: 16551265]
32. Geyer H, Holschbach C, Hunsmann G, Schneider J. J. Biol. Chem. 1988; 263(24):11760–11767. [PubMed: 2841333]
33. Mizuochi T, Spellman MW, Larkin M, Solomon J, Basa LJ, Feizi T. Biochem. J. 1988; 254(2): 599–603. [PubMed: 2845957]
34. Yeh JC, Seals JR, Murphy CI, van HH, Cummings RD. Biochemistry. 1993; 32(41):11087–11099. [PubMed: 8218172]
35. Liedtke S, Adamski M, Geyer R, Pfutzner A, Rubsamen-Waigmann H, Geyer H. Glycobiology. 1994; 4(4):477–484. [PubMed: 7827409]
36. Zhu X, Borchers C, Bienstock RJ, Tomer KB. Biochemistry. 2000; 39(37):11194–11204. [PubMed: 10985765]
37. Cutalo JM, Deterding LJ, Tomer KB. J. Am. Soc. Mass Spectrom. 2004; 15(11):1545–1555. [PubMed: 15519221]
38. Flynn NM, Forthal DN, Harro CD, Judson FN, Mayer KH, Para MF. J. Infect. Dis. 2005; 191(5): 654–665. [PubMed: 15688278]
39. Gilbert PB, Peterson ML, Follmann D, Hudgens MG, Francis DP, Gurwith M, Heyward WL, Jobes DV, Popovic V, Self SG, Sinangil F, Burke D, Berman PW. J. Infect. Dis. 2005; 191(5): 666–677. [PubMed: 15688279]
40. Geyer H, Kempf R, Schott HH, Geyer R. Eur. J. Biochem. 1990; 193(3):855–862. [PubMed: 2174368]
41. Liedtke S, Geyer R, Geyer H. Glycoconjugate J. 1997; 14(7):785–793.
42. Dalpathado DS, Irungu J, Go EP, Butnev VY, Norton K, Bousfield GR, Desaire H. Biochemistry. 2006; 45(28):8665–8673. [PubMed: 16834341]
43. Kornfeld R, Kornfeld S. Annu. Rev. Biochem. 1985; 54:631–664. [PubMed: 3896128]
44. Rudd PM, Dwek RA. Crit Rev. Biochem. Mol. Biol. 1997; 32(1):1–100. [PubMed: 9063619]
45. Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR. Glycobiology. 2004; 14(2):103–114. [PubMed: 14514716]
46. Wyatt R, Sullivan N, Thali M, Repke H, Ho D, Robinson J, Posner M, Sodroski J. J. Virol. 1993; 67(8):4557–4565. [PubMed: 8331723]
47. Kolchinsky P, Kiprilov E, Bartley P, Rubinstein R, Sodroski J. J. Virol. 2001; 75(7):3435–3443. [PubMed: 11238869]
48. Hartley O, Klasse PJ, Sattentau QJ, Moore JP. AIDS Res. Hum. Retroviruses. 2005; 21(2):171–189. [PubMed: 15725757]
49. Wolk T, Schreiber M. Med. Microbiol. Immunol. 2006; 195(3):165–172. [PubMed: 16547752]
50. Koch M, Pancera M, Kwong PD, Kolchinsky P, Grundner C, Wang L, Hendrickson WA, Sodroski J, Wyatt R. Virology. 2003; 313(2):387–400. [PubMed: 12954207]
51. Zhang M, Gaschen B, Blay W, Foley B, Haigwood N, Kuiken C, Korber B. Glycobiology. 2004; 14(12):1229–1246. [PubMed: 15175256]
52. Gaschen B, Taylor J, Yusim K, Foley B, Gao F, Lang D, Novitsky V, Haynes B, Hahn BH, Bhattacharya T, Korber B. Science. 2002; 296(5577):2354–2360. [PubMed: 12089434]
53. Gao F, Weaver EA, Lu Z, Li Y, Liao HX, Ma B, Alam SM, Scearce RM, Sutherland LL, Yu JS, Decker JM, Shaw GM, Montefiori DC, Korber BT, Hahn BH, Haynes BF. J. Virol. 2005; 79(2): 1154–1163. [PubMed: 15613343]
54. Liao HX, Sutherland LL, Xia SM, Brock ME, Scearce RM, Vanleeuwen S, Alam SM, McAdams M, Weaver EA, Camacho Z, Ma BJ, Li Y, Decker JM, Nabel GJ, Montefiori DC, Hahn BH, Korber BT, Gao F, Haynes BF. Virology. 2006; 353(2):268–282. [PubMed: 17039602]

55. Weaver EA, Lu Z, Camacho ZT, Moukdar F, Liao HX, Ma BJ, Muldoon M, Theiler J, Nabel GJ, Letvin NL, Korber BT, Hahn BH, Haynes BF, Gao F. J. Virol. 2006; 80(14):6745–6756. [PubMed: 16809280]
56. Kothe DL, Li Y, Decker JM, Bibollet-Ruche F, Zammit KP, Salazar MG, Chen Y, Weng Z, Weaver EA, Gao F, Haynes BF, Shaw GM, Korber BT, Hahn BH. Virology. 2006; 352(2):438–449. [PubMed: 16780913]
57. Kothe DL, Decker JM, Li Y, Weng Z, Bibollet-Ruche F, Zammit KP, Salazar MG, Chen Y, Salazar-Gonzalez JF, Moldoveanu Z, Mestecky J, Gao F, Haynes BF, Shaw GM, Muldoon M, Korber BT, Hahn BH. Virology. 2007; 360(1):218–234. [PubMed: 17097711]
58. Go EP, Rebecchi KR, Dalpathado DS, Bandu ML, Zhang Y, Desaire H. Anal. Chem. 2007; 79(4): 1708–1713. [PubMed: 17297977]
59. Irungu J, Go EP, Dalpathado DS, Desaire H. Anal. Chem. 2007; 79(8):3065–3074. [PubMed: 17348632]
60. Chakrabarti BK, Kong WP, Wu BY, Yang ZY, Friborg J, Ling X, King SR, Montefiori DC, Nabel GJ. J. Virol. 2002; 76(11):5357–5368. [PubMed: 11991964]
61. Zhang Y, Go EP, Desaire H. Anal. Chem. 2008 in press.
62. Tarentino AL, Gomez CM, Plummer TH. Biochemistry. 1985; 24(17):4665–4671. [PubMed: 4063349]
63. Tarentino AL, Plummer TH Jr. Methods Enzymol. 1994; 230:44–57. [PubMed: 8139511]
64. Morelle W, Canis K, Chirat F, Faid V, Michalski JC. Proteomics. 2006; 6(14):3993–4015. [PubMed: 16786490]
65. Morelle W, Michalski JC. Nat. Protocols. 2007; 2(7):1585–1602.
66. Jones J, Krag SS, Betenbaugh MJ. Biochim. Biophys. Acta. 2005; 1726(2):121–137. [PubMed: 16126345]
67. Mellquist JL, Kasturi L, Spitalnik SL, Shakin-Eshleman SH. Biochemistry. 1998; 37(19):6833–6837. [PubMed: 9578569]
68. Shakin-Eshleman SH, Spitalnik SL, Kasturi L. J. Biol. Chem. 1996; 271(11):6363–6366. [PubMed: 8626433]
69. Nishimura I, Uchida M, Inohana Y, Setoh K, Daba K, Nishimura S, Yamaguchi H. J. Biochem. (Tokyo). 1998; 123(3):516–520. [PubMed: 9538236]
70. Jitsuhara Y, Toyoda T, Itai T, Yamaguchi H. J. Biochem. (Tokyo). 2002; 132(5):803–811. [PubMed: 12417032]
71. Scanlan CN, Offer J, Zitzmann N, Dwek RA. Nature. 2007; 446(7139):1038–1045. [PubMed: 17460665]
72. Pashov A, Perry M, Dyar M, Chow M, Kieber-Emmons T. Curr. Pharm. Des. 2007; 13(2):185–201. [PubMed: 17269927]

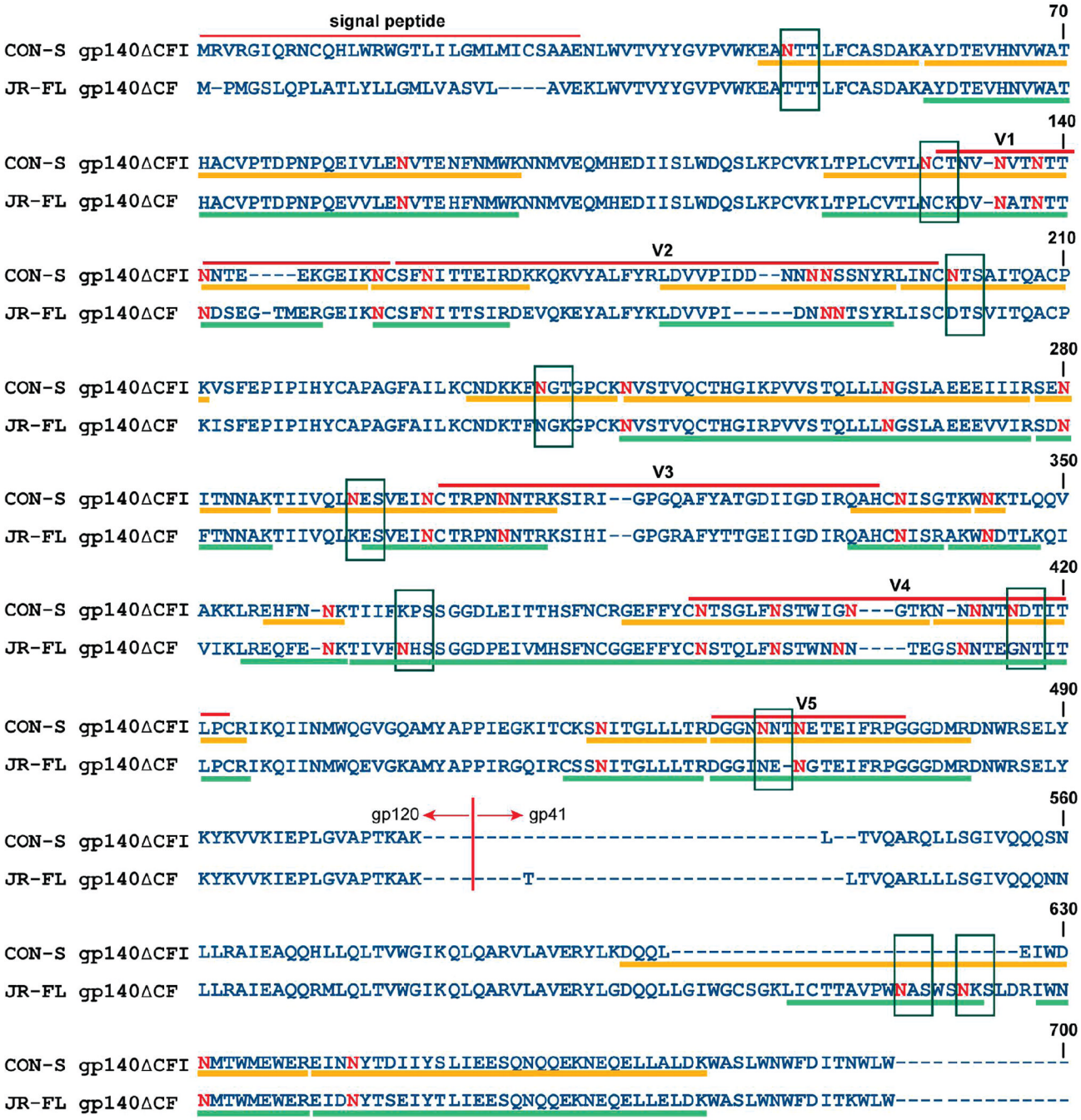

Figure 1.
Sequence alignment of CON-S gp140 ΔCFI and JR-FL gp140ΔCF. Dashes indicate gaps in amino acid sequence and the location of the variable regions (V1–V5) are shown. Potential glycosylation sites are in red. and differences in potential glycosylation sites are boxed. Identified tryptic fragments are underlined.

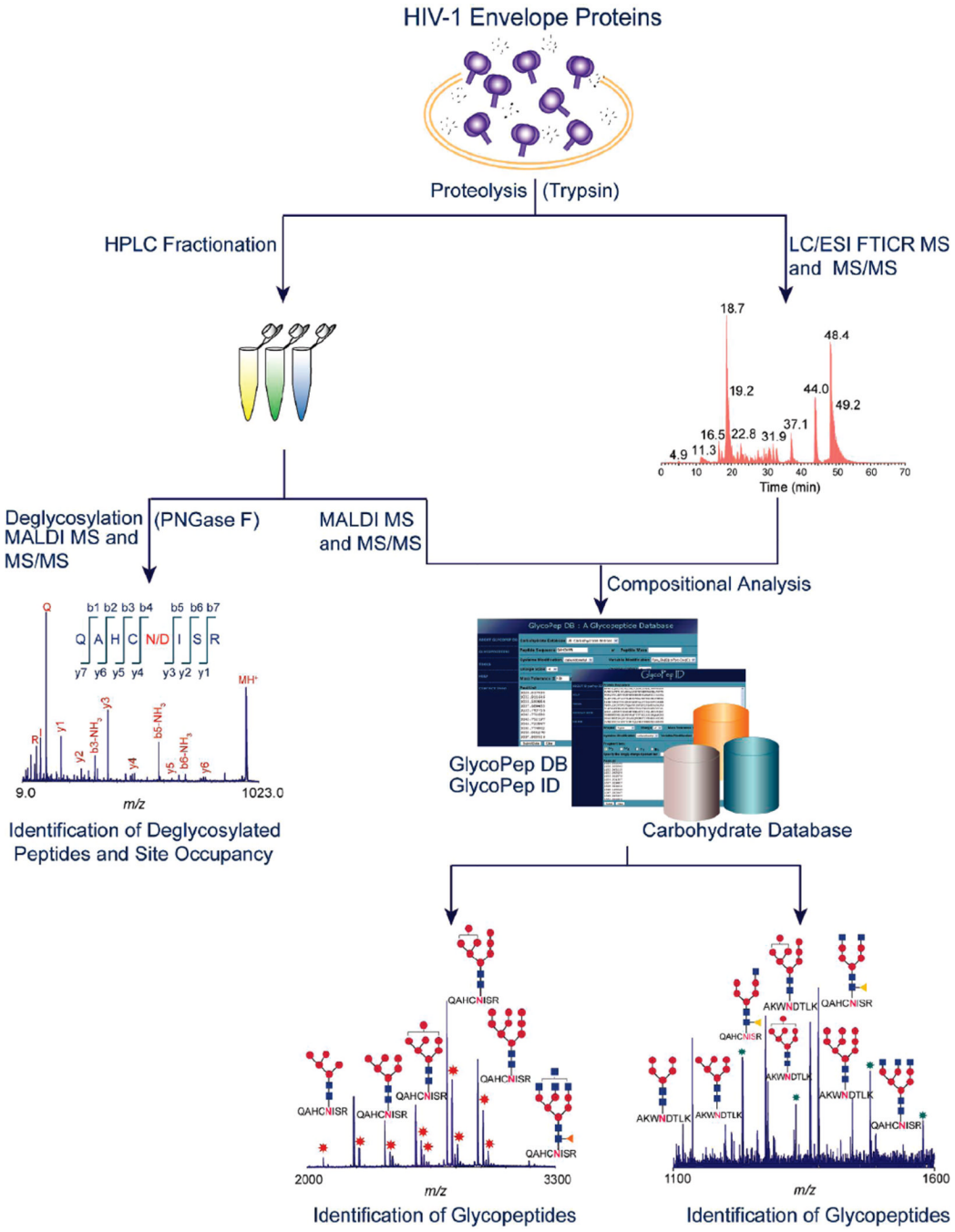

Figure 2.
Schematic representation of the experimental approach of the glycosylation mapping and profiling.

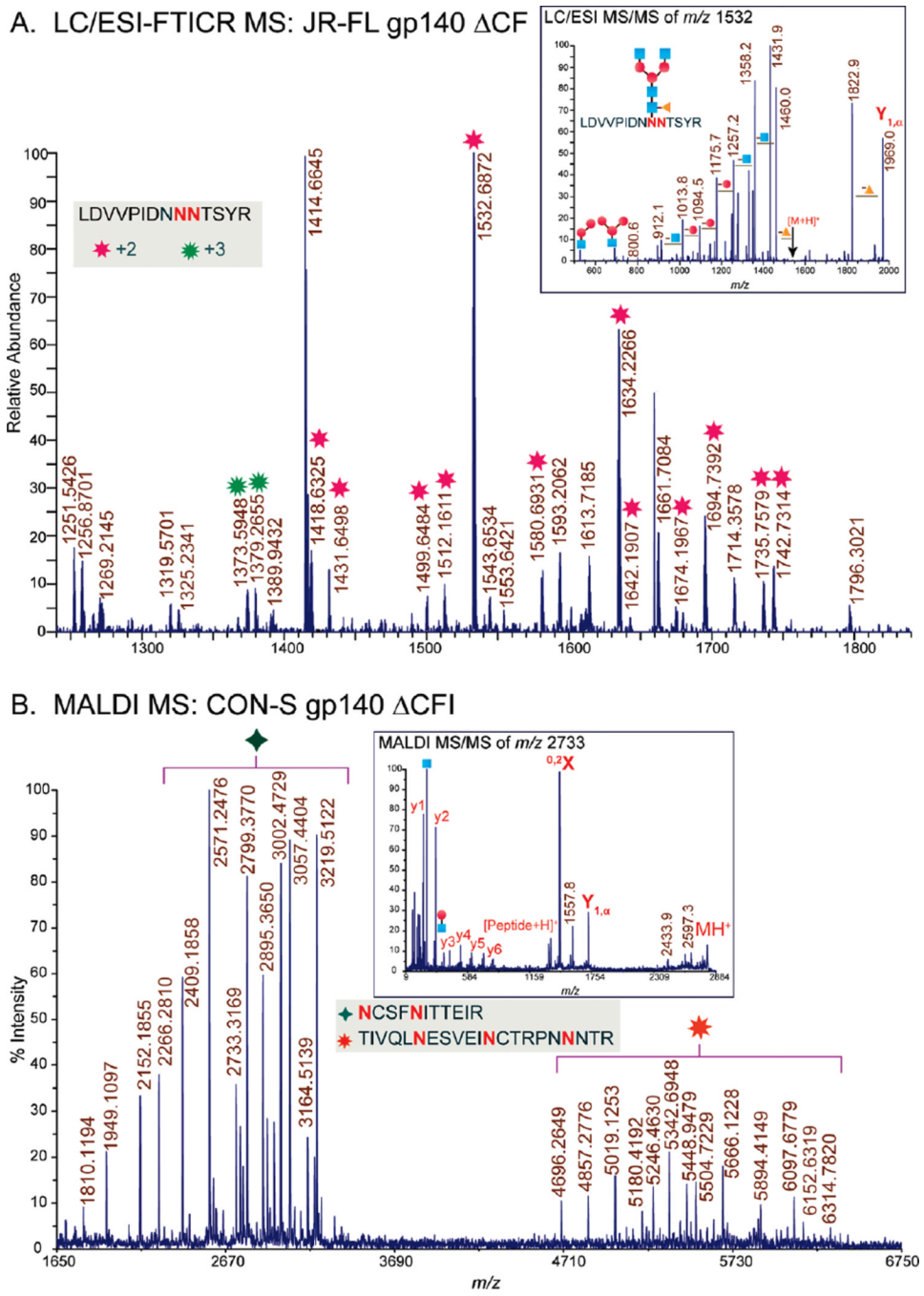

Figure 3.

Representative (A) LC/ESI-FTICR MS and (B) MALDI MS spectra of two separate glycopeptide fractions generated from the proteolytic digest of the envelope proteins. Inset: MS/MS spectra of the identified tryptic glycopeptides. Peptide portion was determined from the characteristic cross-ring cleavages, $^{0,2}X$ (in MALDI MS/MS) and $^{0}Y_{1}$ (in LC/ESI MS/MS) ions.

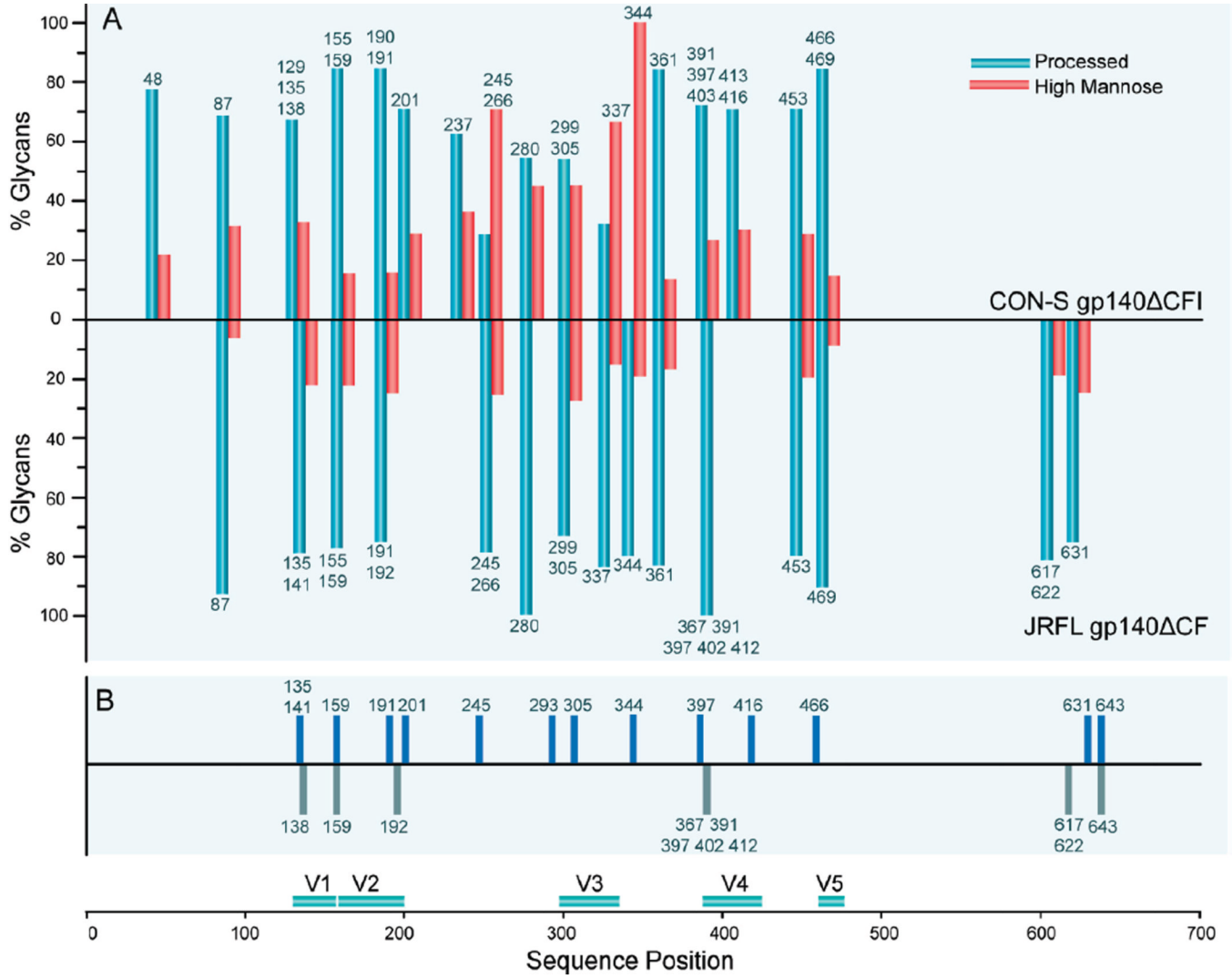

Figure 4.

(A) Summary of glycan compositions (in percent) present on the identified glycosylation site. Glycan compositions were broadly categorized into two classes (see text). Processed glycans include hybrid and complex type structures. (B) Variably utilized and unutilized glycosylation sites for CON-S gp140 ΔCFI (top, blue) and JR-FL gp140 ΔCF (bottom, gray). Note that numbers on top or bottom of the bars in A and B represent the glycosylation site/s. Multiple numbers at the bottom of the bar in B denote either of the glycosylation sites are utilized.

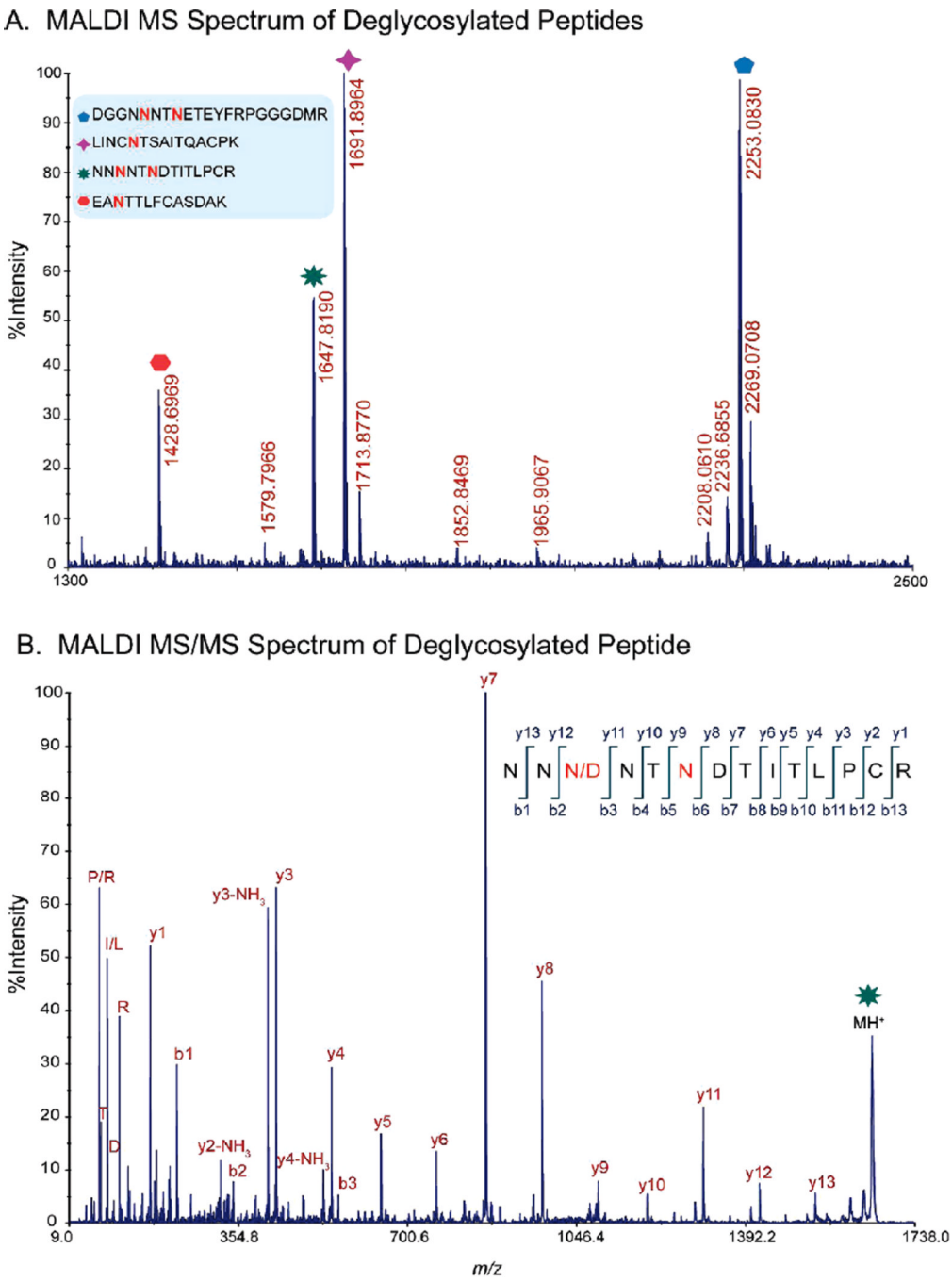

Figure 5.

(A) Representative MALDI MS spectrum of the deglycosylated glycopeptide fraction of CON-S gp140ΔCFI. Four tryptic peptides bearing potential glycosylation sites (see legend) were identified. (B) MS/MS spectrum of one of the identified peptides in A bearing two potential glycosylation sites. Only one site is utilized as shown.

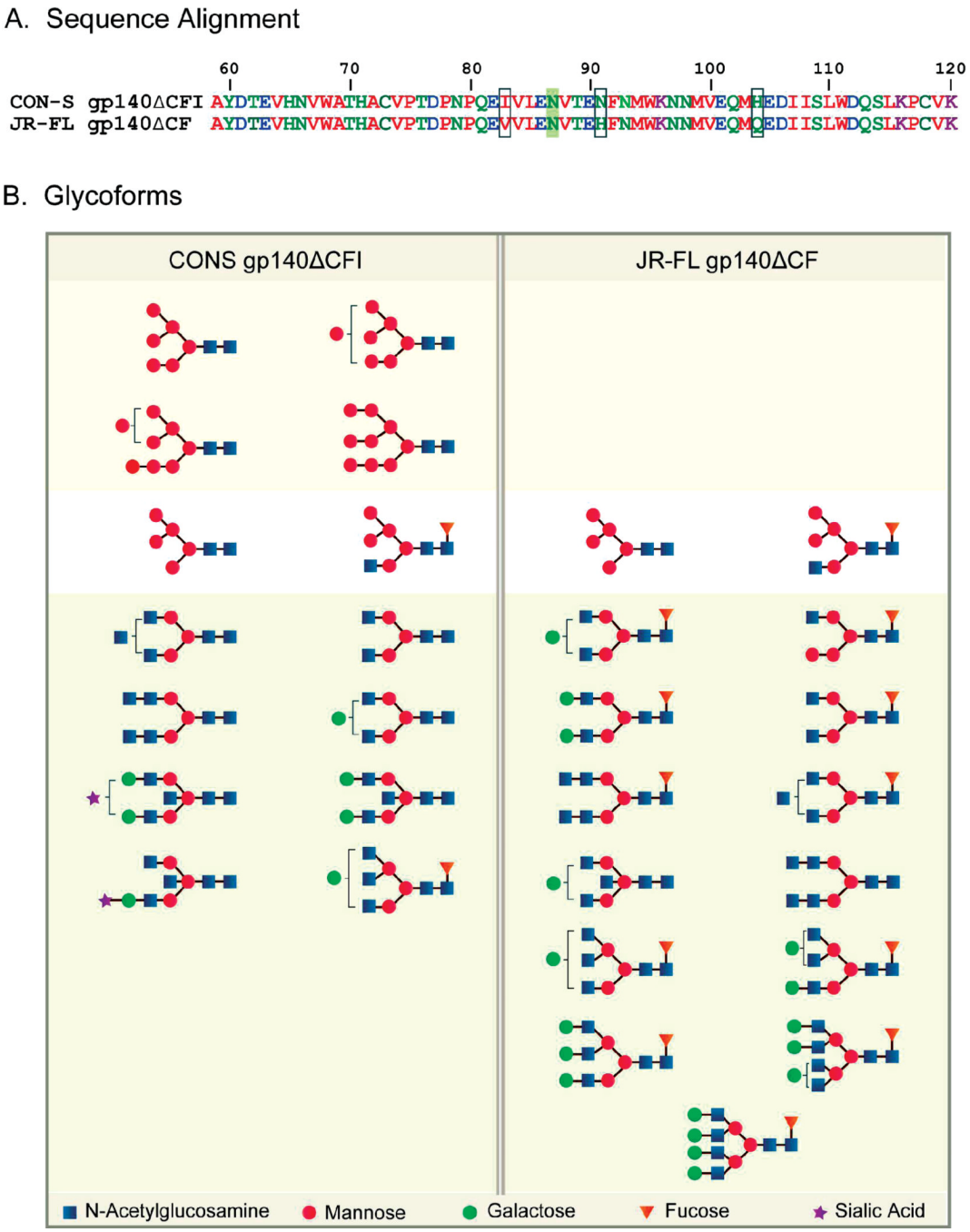

Figure 6.

(A) Sequence alignment of 62-residue segment of the glycopeptide identified in the C1 region of the envelope proteins. The amino acid residues are colored according to their properties. The sequence is 95% identical. Difference in amino acid residue is boxed. Glycosylation site is highlighted in green. (B). Pictorial representation of the identified glycoforms. The structures shown are biologically relevant structures. These are glycan structures commonly found in biological systems and note that isoforms exist.

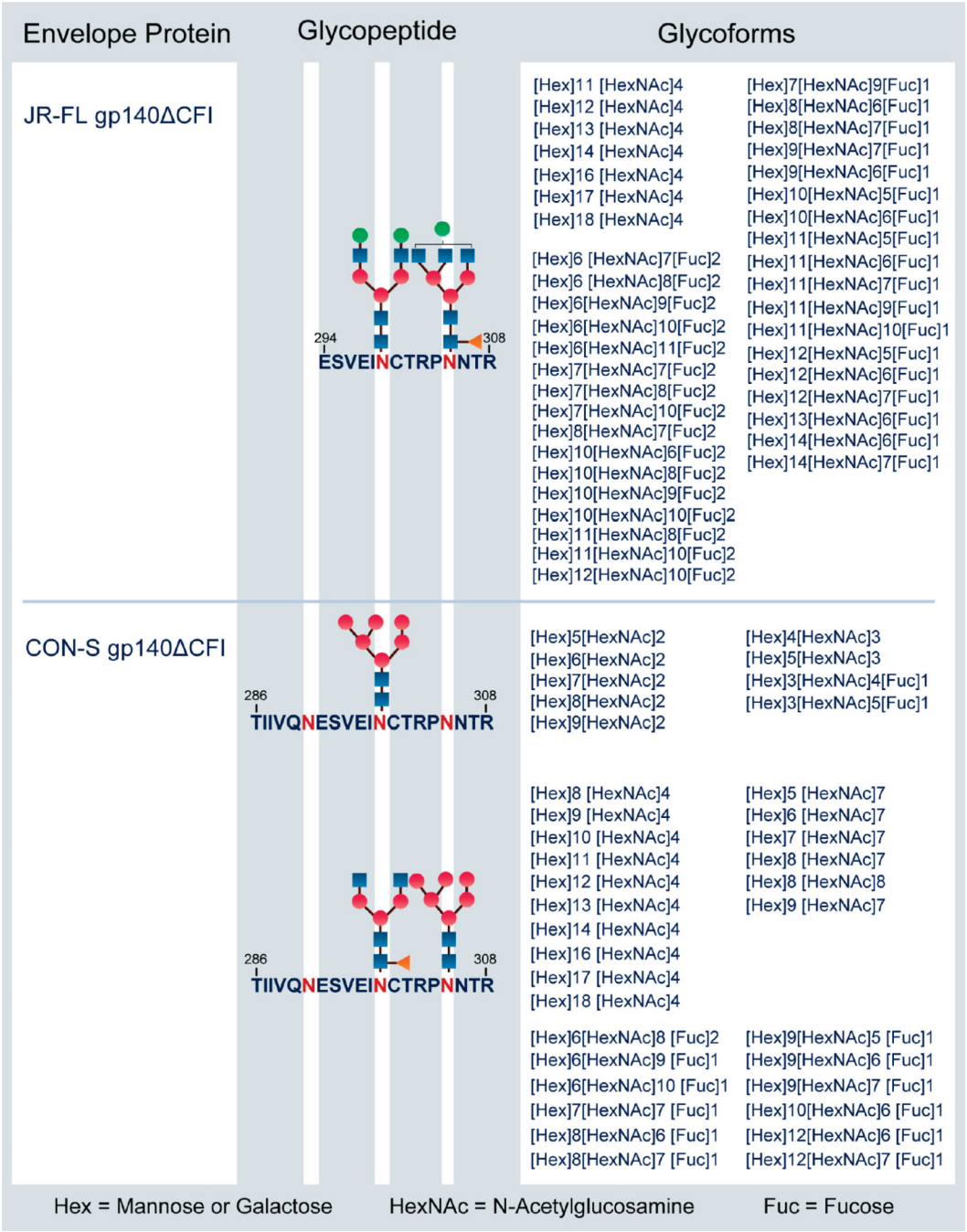

Figure 7.

Glycosylation of the tryptic glycopeptide at the beginning of the V3 region for JR-FL gp140 ΔCF and CON-S gp140 ΔCFI showing the utilized sites and identified glycan compositions.

Table 1

Tryptic Glycopeptides Detected by MALDI MS and LC ESI-FTICR MS

Glycopeptide	Number of Potential Glycosylation Sites	Number of Sites Occupied
CON-S gpl40 ΔCFI		
A. Glycopeptides with fully occupied Sites		
EAN[48]TTLFCASDAK	1	1
AYDTEVHNVWATHACVPTDPNPQEIVLEN[87]VTENFNMWK	1	1
CNDKKFN[237]GTGPCK	1	1
SEN[280]ITNNAK	1	1
QAHCN[337]ISGTK	1	1
LREHFNN[361]K	1	1
SN[453]ITGLLLTR	1	1
B. Glycopeptides with open and/or partially occupied sites		
LTPLCVTLN[129]CTNVN[135]VTN[138]TTN[141]NTEEK	4	2 and 3
N[155]CSFN[159]ITTEIR	2	1 and 2
LDVVPIDDNNN[190]N[191]SNYR	2	1
LINCN[201]TSAITQACPK	1	0 and 1
N[245]VSTVQCTHGIKPVVSTQLLLN[266]GSLAEEEIIIR	2	1 and 2
TIIVQLN[293]ESVEIN[299]CTRPNN[305]NTR	3	1 and 2
WN[344]K	1	0 and 1
GEFFYCN[391]TSGLFN[397]STWIGN[403]GTK	3	2 and 3
NNN[413]NTN[416]DTITLPCR	2	1 and 2
DGGNN[466]NTN[469]ETEIFRPGGGDMR	2	1 and 2
C. Nonglycosvlated Peptides		
DQQLEIWDN[631]MTMEWER	1	0
EINN[643]YTDIIYSLIEESQNQQEK	1	0
JR-FL gpl40 ΔCF		
A. Glycopeptides with fully occupied sites		
AYDTEVHNVWATHACVPTDPNPQEVVLEN[87]VTEHFNMWK	1	1
N[245]VSTQCTHGIRPVVSTQLLLN[266]GSLAEEEIIIR	2	2
SDN[280]FTNNAK	1	1
ESVEIN[299]CTRPNN[305]NTR	2	2
QAHCN[337]ISR	1	1
AKWN[344]DTLK	1	1
LREQFEN[361]K	1	1
CSSN[453]ITGLLLTR	1	1
DGGINEN[469]GTE1FRPGGGDMR	1	1
IWNN[631]MTWMEWER	1	1
B. Glycopeptides with open and partially occupied sites		

Glycopeptide	Number of Potential Glycosylation Sites	Number of Sites Occupied
LTPLCVTLNCKDVN^{135}ATN^{138}TTN^{141}DSEGTMER	3	2
N^{155}CSFN^{159}ITTSIRDEVQK	2	1 and 2
LDVVPIDNN^{191}N^{192}TSYR	2	1
TIVFN^{367}HSGGDPEIVMHSFNCGGEFFYCN^{391}STQLFN^{397}STWNN^{402}NTEGSN^{412}NTEGNTITLPCR*	5	2
LICTTAVPWN^{617}ASWSN^{622}K	2	1
C. Nonglycosylated Peptide		
EIDN^{643}YTSEIYTLIEESQNQQEK	1	0

Table 2

Representative Glycopeptide Compositions for JR-FL and CON-S[a]

env domain	Charge state	experimental mass, (*m/z*)	theoretical mass, (*m/z*)	mass error (ppm)	peptide sequence	mod*	carbohydrate composition
					JR-FL gp140 ΔCF		
V1–V2	1+	3721.7263	3721.5889	37	NCSFNITTSIRDEVQK		[Hex]4[HexNAc]5[Fuc]1
	3+	1295.2152	1295.2188	2.8	NCSFNITTSIRDEVQK		[Hex]5[HexNAc]5[Fuc]1
	3+	1392.2512	1392.2506	0.4	NCSFNITTSIRDEVQK		[Hex]5[HexNAc]5[Fuc]1[NeuNAc]1
	1+	3452.4651	3452.4514	4	NCSFNITTSIRDEVQK		[Hex]7[HexNAc]2
	1+	3614.5623	3614.5042	16	NCSFNITTSIRDEVQK		[Hex]8[HexNAc]2
	3+	1259.5146	1259.5239	7	NCSFNITTSIRDEVQK		[Hex]9[HexNAc]2
C2–V3	1+	4788.6450	4788.9026	11	ESVEINCTRPNNNTR		[Hex]10 [HexNAc]6[Fuc]1
	1+	5340.7336	5341.1192	72	ESVEINCTRPNNNTR		[Hex]10 [HexNAc]8[Fuc]2
	1+	5543.6627	5544.1986	97	ESVEINCTRPNNNTR		[Hex]10 [HexNAc]9[Fuc]2
	1+	5154.0062	5154.0347	6	ESVEINCTRPNNNTR		[Hex]11[HexNAc]7[Fuc]1
	1+	5503.9859	5503.1720	148	ESVEINCTRPNNNTR		[Hex]11[HexNAc]8[Fuc]2
	1+	5112.9775	5113.0082	6	ESVEINCTRPNNNTR		[Hex]12[HexNAc]6[Fuc]1
C4	2+	1491.6762	1491.6652	7	CSSNITGLLLTR		[Hex]3[HexNAc]5[Fuc]1
	2+	1369.6179	1369.6122	4	CSSNITGLLLTR		[Hex]4[HexNAc]3[Fuc]1
	3+	1078.1394	1078.1355	4	CSSNITGLLLTR		[Hex]4[HexNAc]4[Fuc]1[NeuNAc]1
	2+	1572.7100	1572.6916	12	CSSNITGLLLTR		[Hex]4[HexNAc]5[Fuc]1
					CON-S gp140 ΔCFI		
C1	2+	1525.6365	1525.6180	12	EANTTLFCASDAK		[Hex]5[HexNAc]4
	2+	1403.5756	1403.5651	7	EANTTLFCASDAK		[Hex]6[HexNAc]2
	2+	1505.1039	1505.1047	0.5	EANTTLFCASDAK		[Hex]6[HexNAc]3
	2+	1578.1262	1578.1337	5	EANTTLFCASDAK		[Hex]6[HexNAc]3[Fuc]1
	2+	1606.6709	1606.6444	17	EANTTLFCASDAK		[Hex]6[HexNAc]4
	4+	1509.4340	1509.4105	64	AYDTEVHNVWATHACVPTDPN-PQEVVLENVTEHFNMWK		[Hex]4[HexNAc]5[NeuNAc]1
	4+	1592.3395	1592.4410	16	AYDTEVHNVWATHACVPTDPN-PQEVVLENVTEHFNMWK		[Hex]5 [HexNAc]3 [Fuc]1
	4+	1422.1360	1422.1261	7	AYDTEVHNVWATHACVPTDPN-PQEVVLENVTEHFNMWK		[Hex]5[HexNAc]2
V2	2+	1603.6862	1603.6780	5	LDVVPIDDNNNNSSNYR		[Hex]4[HexNAc]3
	2+	1879.7935	1879.7863	4	LDVVPIDDNNNNSSNYR		[Hex]4[HexNAc]5[Fuc]1

J Proteome Res. Author manuscript; available in PMC 2013 May 20.

env domain	Charge state	experimental mass, (*m/z*)	theoretical mass, (*m/z*)	mass error (ppm)	peptide sequence	mod*	carbohydrate composition
	2+	1684.7036	1684.7044	0.4	LDVVPIDDNNNNSSNYR		[Hex]5[HexNAc]3
C2	1+	6213.3965	6213.8508	74	NVSTVQCTHGIKPVVSTQLL-LNGSLAEEEIIIR		[Hex]11 [HexNAc]4
	1+	6700.3182	6700.0093	46	NVSTVQCTHGIKPVVSTQLL-LNGSLAEEEIIIR		[Hex]14[HexNAc]4
	1+	6862.0877	6862.0621	4	NVSTVQCTHGIKPVVSTQLL-LNGSLAEEEIIIR		[Hex]15 [HexNAc]4
	1+	7024.3167	7024.1149	29	NVSTVQCTHGIKPVVSTQLL-LNGSLAEEEIIIR		[Hex]16[HexNAc]4
V5	1+	4222.7593	4222.6981	15	DGGNNNTNETEIFR-PGGGDMR		[Hex]5[HexNAc]5[Fuc]1
	1+	4425.8540	4425.7774	17	DGGNNNTNETEIFR-PGGGDMR		[Hex]5[HexNAc]6[Fuc]1
	3+	1326.8643	1326.8689	4	DGGNNNTNETEIFR-PGGGDMR		[Hex]6[HexNAc]3[Fuc]1
	3+	1375.2253	1375.2148	8	DGGNNNTNETEIFR-PGGGDMR		[Hex]6[HexNAc]3[NeuNAc]1

[a]The complete list is found in Supplementary Tables 1A, 1B, 2A and 2B in Supporting Information.