



## Methods development for analysis of partially deglycosylated proteins and application to an HIV envelope protein vaccine candidate

Eden P. Go<sup>a</sup>, Geetha S. Hewawasam<sup>a</sup>, Ben J. Ma<sup>b</sup>, Hua-Xin Liao<sup>b</sup>, Barton F. Haynes<sup>b</sup>, Heather Desaire<sup>a,\*</sup>

<sup>a</sup> Department of Chemistry, University of Kansas, Lawrence, KS, United States

<sup>b</sup> Duke Human Vaccine Institute, Duke University Medical Center, Durham, NC, United States

### ARTICLE INFO

#### Article history:

Received 2 August 2010

Received in revised form 5 November 2010

Accepted 8 November 2010

Available online 18 November 2010

#### Keywords:

Glycosylation

PNGase F

Endoglycosidase

HIV

Envelope protein

Glycosylation site

### ABSTRACT

The work presented herein describes the first comprehensive analysis of a partially deglycosylated HIV vaccine candidate envelope protein (Env). The Env, JRFL gp140 ΔCF, with 27 potential glycosylation sites, was partially deglycosylated with PNGase F as part of a strategy to generate a more immunogenic HIV vaccine, and the resulting protein glycosylation was characterized in a unique workflow using two different glycosidases, Endo H and Endo F3. This unique analysis protocol provided for coverage on 26 of the 27 glycosylation sites, and the data showed that the biochemical treatment with PNGase F resulted in a highly heterogeneous glycoprotein product that had been partially deglycosylated at most of the glycosylation sites. The protocols described in this work could be useful for characterizing the glycosylation site occupancy of other native or biochemically deglycosylated proteins.

© 2010 Elsevier B.V. All rights reserved.

### 1. Introduction

Glycoproteins are an important and growing class of biopharmaceuticals. Currently, glycoprotein-based pharmaceuticals are used or under development for solving a broad variety of health problems, including low fertility [1], renal failure [2], or to prevent the transmission of HIV [3,4]. The glycosylation on proteins is often necessary to assure proper protein folding, especially during protein expression. However, the glycosylation can also have detrimental effects, such as targeting the protein for excretion [5,6], or suppressing the protein's intended function [7].

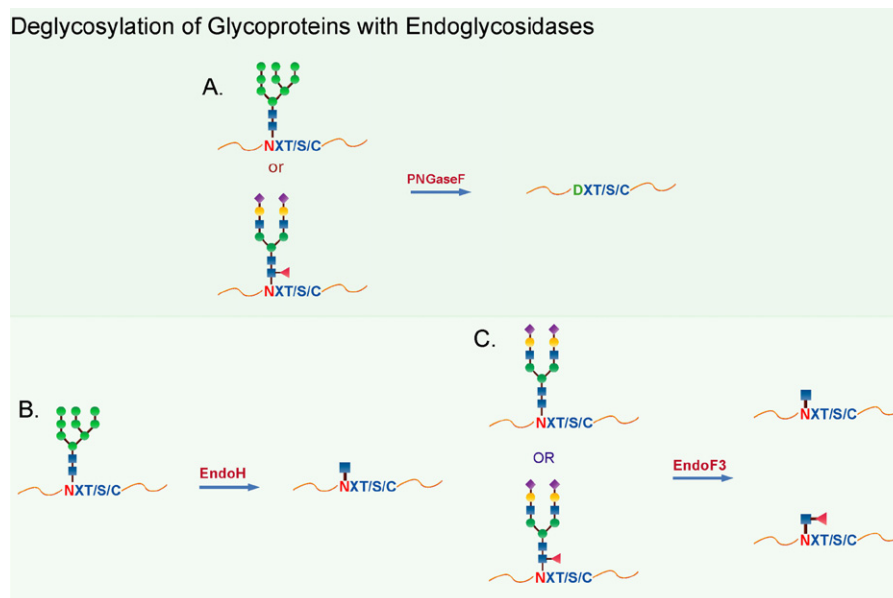
One poignant example of the deleterious effects of glycosylation is found in the hunt for an HIV vaccine. The glycans on Env, the protein on the surface of the HIV virus which is expressed in mammalian cells and used in vaccine trials, cover immunogenic epitopes of the protein [8] and may also suppress the immune system by binding to carbohydrate receptors on T cells [9–11]. Unfortunately, Env, like many proteins, requires the glycans to be present during protein folding [12,13]. Otherwise, unfolded, inactive protein is produced.

One possible strategy to generating better biopharmaceuticals, such as a more effective HIV vaccine, would be to express a well-folded glycoprotein of interest with its glycosylation intact, then

either fully or partially remove the glycans under non-denaturing conditions, using an enzyme such as peptide N-glycosidase F (PNGase F). This strategy is currently being explored as a means to develop an effective HIV vaccine [14]. A key limitation of this strategy is that significant efforts must be undertaken to characterize the partially deglycosylated species, so that the proteins' change in function can be related back to its modified composition. Characterizing the remaining glycosylation on a protein such as HIV Env is a quite burdensome process. Fully glycosylated Env proteins require at least 200 μg of protein at >4 mg/mL for full glycosylation site-specific analysis [15,16], and heterogeneous mixtures of glycosylated and deglycosylated proteins would likely result in an increase in this sample requirement. In addition to sample consumption issues, analyzing glycosylation in a glycosylation site-specific manner is quite a labor and skill intensive process [17], where a single protein could take months to analyze. Therefore, instead of characterizing the remaining glycosylation on Env, which would be quite burdensome, we introduce a new analytical platform to rapidly characterize these species using a complementary approach. In this platform, partially deglycosylated proteins are treated with endoglycosidases (Endo H and Endo F3) to reduce the complexity of the remaining N-linked glycopeptides, and the products are detected by LC/MS and MS/MS. The glycosidases cleave the glycans between the first and second N-acetylhexosamine in the N-linked core, leaving a single N-acetylhexosamine attached to the peptide (Fig. 1) [18–20]. The advantages of the endoglycosidase treatment at least two-fold: (1) The resulting glycopeptides

\* Corresponding author. Tel.: +1 785 864 3015.

E-mail address: [hdesaire@ku.edu](mailto:hdesaire@ku.edu) (H. Desaire).



**Fig. 1.** (A) Deglycosylation reaction with PNGase F. Both high mannose and complex glycans are cleaved by the enzyme. Concomitant with glycan cleavage, the N at the glycosylation site is converted into D. (B) Glycosidic cleavage reaction with Endo H. (C) Glycosidic cleavage reaction with Endo F3.

(with the single HexNAc attached) are quite simple to characterize because the glycan mass is known; therefore, the exact mass of the glycopeptide is known, and the glycosylated peptides can be searched for and identified in an automated fashion. (2) Since all heterogeneity of the glycan component is clipped off, the MS signal is stronger because all the heterogeneous glycans with different  $m/z$ 's now produce one product, with a single  $m/z$ . This results in lower detection limits and lower sample requirements for the analysis. It is important to note however, that if glycoproteins are partially core-fucosylated, two peaks, instead of one, may be detected.

While endoglycosidases have been used previously to characterize the natural site occupancy of a protein [21–23], this is the first example of using these enzymes in a rapid and reliable method to determine which glycosylation sites on a glycoprotein have been deglycosylated by another enzyme, PNGase F, and detailed information is provided about every glycosylation site in the protein of interest. The analysis of one HIV envelope vaccine candidate, HIV Env JRFL gp140  $\Delta$ CF, with 27 potential glycosylation sites, is used to demonstrate the utility of the method. The method developed herein could be applied broadly to other glycosylated HIV vaccine candidates or to any other protein whose glycosylation is fully or partially removed.

## 2. Experimental

### 2.1. Reagents

Trizma<sup>®</sup> hydrochloride, Trizma<sup>®</sup> base, ethylenediaminetetraacetic acid (EDTA), glacial acetic acid, HPLC grade acetonitrile ( $\text{CH}_3\text{CN}$ ) and methanol ( $\text{CH}_3\text{OH}$ ), hydrochloric acid, ammonium acetate, sodium acetate, ammonium hydroxide, insulin from bovine pancreas, iodoacetamide (IAA), dithiothreitol (DTT), 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 endoproteinase Glu-C from *Staphylococcus aureus* V8 were obtained from Promega (Madison, WI) and Sigma (St. Louis, MO), respectively. Glycerol-free peptide N-glycosidase F (PNGase F) from *Flavobacterium*

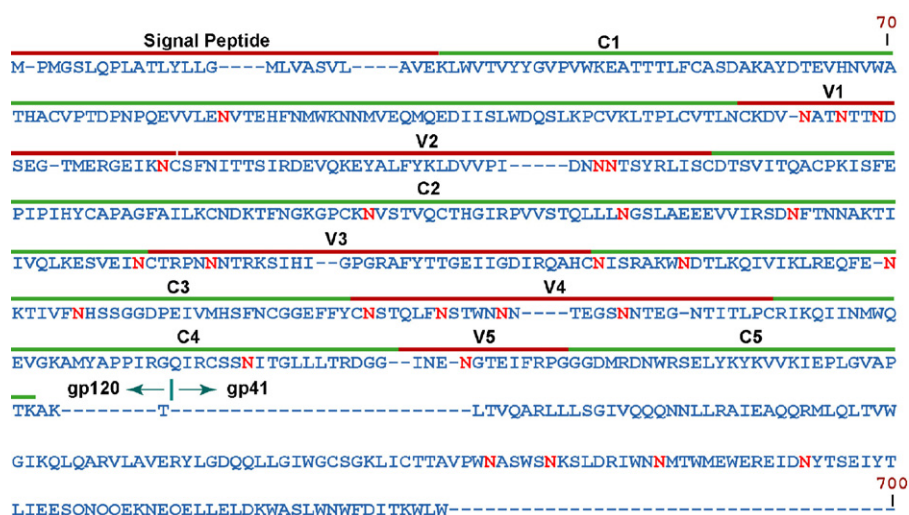
*meningosepticum*, endo- $\beta$ -N-acetylglucosaminidase H (Endo H), and endo- $\beta$ -N-acetylglucosaminidase F3 (Endo F3) from *Elizabethkingia meningosepticum* from *Streptomyces plicatus*, were obtained from New England BioLabs (Ipswich, MA), and EMD Biosciences (Gibbstown, NJ).

### 2.2. Expression and purification of HIV-1 subtype B envelope proteins

Recombinant JRFL gp140  $\Delta$ CF envelope protein was obtained from the Duke Human Vaccine Research Institute in Durham N.C. The envelope protein was constructed with internal deletions and was expressed and purified as described previously [24,25]. Briefly, recombinant vaccinia viruses (rVVs) expressing JR-FL gp140  $\Delta$ CF were used for production of soluble oligomeric envelope protein. Batch production was carried out by infecting 293T cells with rVVs. The 293T cells were cultured in Dulbecco modified Eagle medium (DMEM; Invitrogen Corp., Carlsbad, CA) supplemented with 10% fetal calf serum in T150 tissue culture flasks and were grown to confluence at a multiplicity of infection (MOI) of about one before infecting with rVVs. At 2 h post-infection, the cell culture was washed with serum free DMEM and the infection was allowed to proceed for 72 h. Recombinant JR-FL gp140  $\Delta$ CF was purified from supernatants of rVV-infected 293T cell cultures using *Galanthus nivalis* lectin-agarose (Vector Labs, Burlingame, CA) column chromatography and stored at  $-70^\circ\text{C}$  until use [16,24–27]. A typical batch production using 30 T-150 TC flasks would yield 1–3 mg of purified recombinant HIV-1 envelope protein. Protein concentration was determined by absorbance. Purified recombinant envelope proteins were concentrated for MS-based glycosylation analysis.

### 2.3. Partial N-deglycosylation of the envelope protein

Glycans were removed partially by incubating 100  $\mu\text{g}$  of the envelope protein with 4.0  $\mu\text{L}$  PNGase F (2000 U) under non-denaturing conditions. The reaction mixture (200  $\mu\text{L}$ ) was incubated at  $37^\circ\text{C}$  for 24 h and was stopped by freezing at  $-70^\circ\text{C}$  and was stored at this temperature prior to analysis.



**Fig. 2.** Sequence of the HIV Envelope protein, JRF1 gp140ΔCF, used in these studies. All the potential N-linked glycosylation sites are in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

#### 2.4. Endo H and Endo F3 deglycosylation of envelope proteins

Deglycosylated envelope proteins were treated either with Endo H or with Endo F3. These enzymes cleave between the N-acetylhexosamine residues in the chitobiose core of N-linked glycans leaving an N-acetylhexosamine or N-acetylhexosamine with fucose on the asparagine residue. Endo H is specific for high mannose and hybrid glycans while Endo F3 is specific for complex glycans. Typical deglycosylation with these endoglycosidases was performed by incubating ~7 fmol of the envelope protein (protein concentration of ~7 mg/mL) with the enzyme solutions ( $\geq 30$  units/mg protein). For the Endo H deglycosylation experiment, the sample was denatured with 2 M urea in 100 mM Tris buffer (pH 5.5) followed by the addition of 2  $\mu$ L of Endo H. After thorough mixing, the reaction was incubated for 48 h at 37 °C. For the Endo F3 deglycosylation experiment, samples were incubated with ~6  $\mu$ L of 250  $\mu$ M NH<sub>4</sub>C<sub>2</sub>H<sub>2</sub>O<sub>2</sub> (pH 4.5) and 2  $\mu$ L of Endo F3 for two weeks at 37 °C. Endo H and Endo F3 treated deglycosylated envelope protein were digested either with trypsin or with Glu-C as described below. As a control, native envelope protein was treated with Endo H and Endo F3 as described above and then subsequently digested with trypsin or endoproteinase Glu-C.

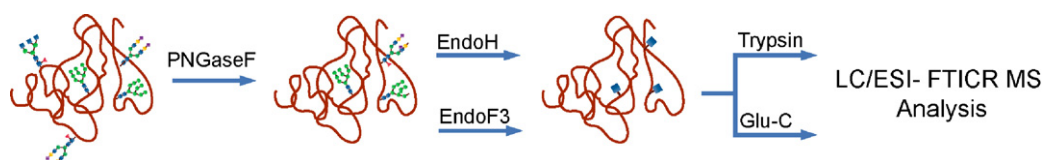
#### 2.5. Digestion of envelope proteins

Samples containing 75  $\mu$ g of the HIV-1 envelope protein, with protein concentration ~7 mg/mL, were denatured with 6 M urea and digested with trypsin or endoproteinase Glu-C. The proteins were reduced and alkylated with 10 mM DTT and 15 mM IAA at RT, respectively. Protein digestion with trypsin was performed at a protein:enzyme ratio of 30:1 (w/w) in 100 mM Tris buffer (pH 8.5) containing 3 mM EDTA while protein digestion with endoproteinase Glu-C was performed at a protein:enzyme ratio of 20:1

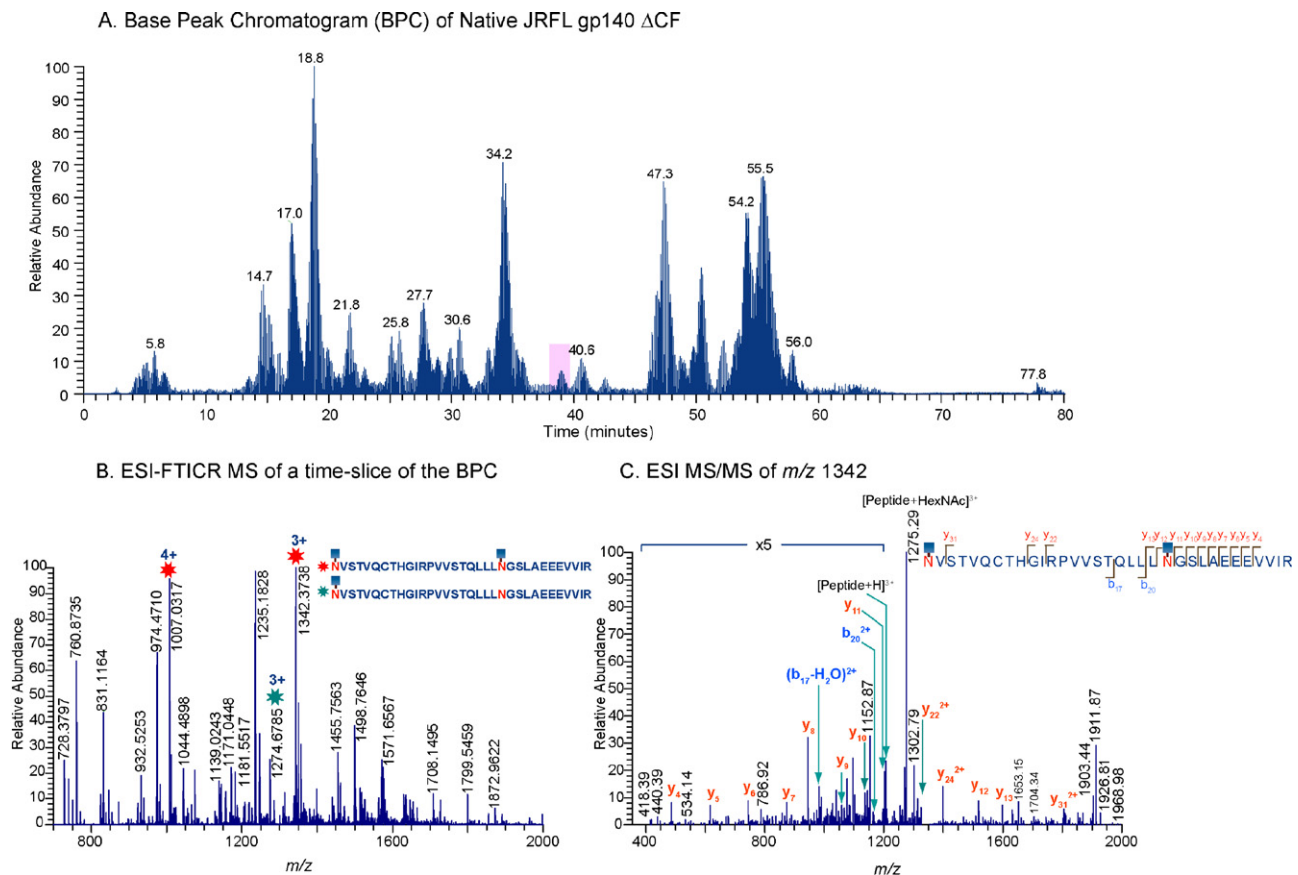
(w/w) in 100 mM Tris buffer (pH 7.8) containing 3 mM EDTA. Digestion was allowed to proceed overnight at 37 °C and was stopped with the addition of 1  $\mu$ L of glacial acetic acid. The resulting HIV envelope glycoprotein digest was analyzed by RP-HPLC/ESI-FTICR MS. Protein digestion was performed three times on different days with samples obtained from the same batch to ensure reproducibility and reliability of the method.

#### 2.6. Mass spectrometry

LC/ESI-FTICR MS experiments were performed using a hybrid linear ion-trap Fourier transform ion cyclotron resonance mass spectrometer (LTQ-FT, ThermoScientific, San Jose, CA) directly coupled to a Dionex UltiMate capillary LC system (Sunnyvale, CA) equipped with a FAMOS well plate autosampler. Mobile phases utilized for the experiment consisted of solvent A: 99.9% deionized H<sub>2</sub>O + 0.1% formic acid and solvent B: 99.9% CH<sub>3</sub>CN + 0.1% formic acid. Five microliters of the digested sample was injected onto C18 PepMap™ 300 column (300  $\mu$ m i.d.  $\times$  15 cm, 300 Å, LC Packings, Sunnyvale, CA) at a flow rate of 5  $\mu$ L/min. The following CH<sub>3</sub>CN/H<sub>2</sub>O multistep gradient was used: 5% mobile phase B for 5 min, followed a linear increase to 40% B in 50 min a linear increase to 90% B in 10 min. The column was held at 95% B for 10 min before re-equilibration. A short wash and blank run were performed between every sample to ensure no sample carry-over. The ESI source was operated in the following conditions: source voltage of 2.8 kV, capillary temperature of 200 °C, and capillary voltage of 46 V. ESI source parameters were optimized for maximum sensitivity using 100 nM insulin at *m/z* 1147.70. Data were collected in a data-dependent mode in which the five most intense ions in an FT scan were sequentially and dynamically selected for subsequent collision-induced dissociation (CID) in the LTQ linear ion trap using a normalized collision energy of 30% and a 3 min dynamic exclusion window.



**Fig. 3.** Workflow for these studies. After treatment with PNGase F to deglycosylate the protein, the sample is split and treated with Endo H and Endo F3. Each of these samples are then treated with Trypsin or Glu-C, then analyzed by LC/MS.



**Fig. 4.** (A) Base peak chromatogram of the tryptic digest from the native JRFL gp140  $\Delta$ CF. A peak at around 39 min is highlighted, and the MS data for this portion of the chromatogram is shown in (B). The spectrum in (C) is the MS/MS data of the peak at  $m/z$  1342, highlighted in B.

### 2.7. Peptide identification

Glycopeptides and deglycosylated peptides were identified by searching raw MS/MS data acquired on the hybrid LTQ FTICR mass spectrometer against a custom HIV database with 127 protein entries, obtained from the Los Alamos HIV sequence database (<http://www.hiv.lanl.gov/content>), using Mascot (Matrix Science, London, UK, version 2.2.04). The peak list was extracted from raw files and converted to mgf files using DTASupercharge v. 1.18. Mgf files were searched against the custom HIV database specifying the following parameters: (a) enzyme: trypsin or V8-E, (b) missed cleavage: 2, (c) fixed modification: carbamidomethyl, (d) variable modification: methionine oxidation, carbamyl, pyro-carbamidomethyl, Gln  $\rightarrow$  pyro-Glu, HexNAc, and HexNAc(1)dHex(1)(N) (e) peptide tolerance of 0.8 Da, and (f) MS/MS tolerance of 0.4 Da. Deglycosylated peptides identified from Mascot search were manually validated from MS/MS spectra to ensure major fragmentation ions (b and y ions) were observed especially for deglycosylated peptides generated from PNGase F treated envelope proteins containing potential N to D conversions.

## 3. Results and discussion

### 3.1. Experimental workflow

The glycoprotein used in these studies, JRFL gp140  $\Delta$ CF, is an Env immunogen expressed in 293T cells (Fig. 2). The protein has 27 unique potential N-linked glycosylation sites (highlighted in red), and we have previously shown that the glycans at these sites are a heterogeneous mix of various types of processed and high mannose glycans [15].

The ultimate goal of this study was to develop methods to characterize the partially deglycosylated form of this Env using the enzyme PNGase F. Therefore, the primary objective is to identify when glycosylation is present (and when it has been removed) at each glycosylation site on the protein. The experimental protocol in Fig. 3 describes how glycosylation removal is detected.

### 3.2. Fully glycosylated sample: each site is detected

Prior to analyzing deglycosylated proteins, control experiments on the fully glycosylated protein were conducted. The purpose of these experiments was to verify that full coverage of the glycosylation sites was obtained on a protein that had not been deglycosylated. (These experiments were designed to prove that if a peptide was glycosylated, the method described herein would detect it.)

In these experiments, the protein, JRFL gp140  $\Delta$ CF, was subjected to two separate glycosidase reactions, one with Endo H and the other with Endo F3, as displayed in Fig. 3, and the deglycosylated products were subjected to tryptic digestion and Glu-C digestion and LC/MS/MS analysis. The expected outcome of these experiments are as follows: after treatment with Endo H, each glycosylation sites known to have high mannose glycans present should be detected as a glycopeptide containing a single N-acetylhexosamine at the glycosylation site. Similarly, after treatment with Endo F3, each glycopeptide that is known to have complex glycans present at the glycosylation site should be detected as a peptide with an N-acetylhexosamine (with or without an additional attached fucose).

An example of the LC/MS data for the glycopeptides produced after one of the endoglycosidase experiments on the fully glyco-



sylated JRFL Env protein is shown in Fig. 4. The top panel (4A) is the HPLC chromatogram, with one segment of the chromatogram highlighted. Panel 4B shows the MS data for this time window. The peak at  $m/z$  1342 corresponds to a glycopeptide that contains the peptide sequence NVSTVQCTHGIRPVVSTQLLNGLAEEVVIR plus a single N-acetylhexosamine at each of the two glycosylation sites (in red). The composition of this glycopeptide is verified by MS/MS experiments, as shown in Fig. 4C.

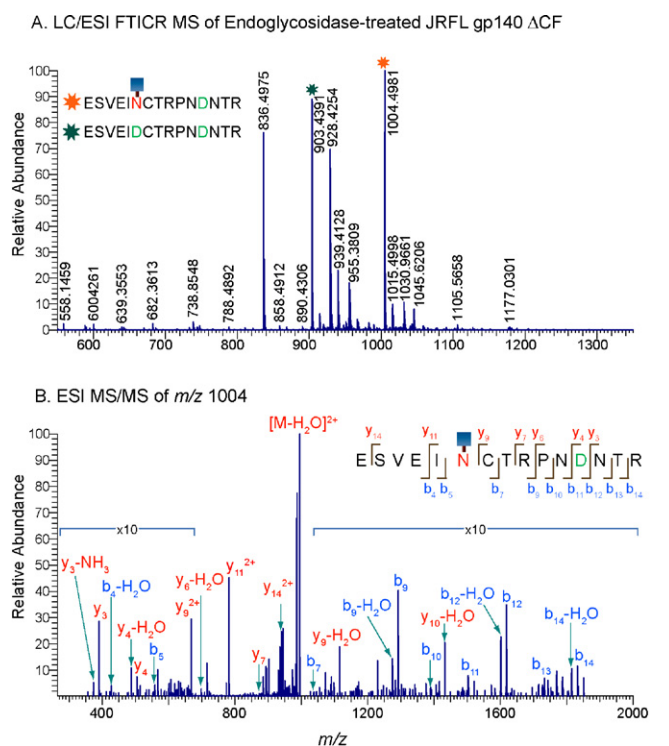
After LC/MS/MS analysis of the protein digested with endoglycosidase and trypsin, a high level of sequence coverage was obtained. The characterized fully glycosylated Env JRFL peptides are listed in Table 1. Every glycosylation site on the protein is represented in the data as a tryptic peptide, with the exception of a segment of the protein in the V4 loop. The tryptic peptide corresponding to this segment is quite large. With no missed cleavages, the peptide would be: TIVFNHSSGGDPEIVMHSFNCGGEFYFCNSTQLFNSTWNNNTEGSNTEGNTITLPCR. Since this peptide is not detected as a tryptic peptide, we digested the deglycosylated protein with Glu-C, in a separate experiment. This enzyme produces slightly smaller peptides for this region: NKTIVFNHSSGGDPE and IVMHSFNCGGEFYFCNSTQLFNSTWNNNTEGSNTE. Both the smaller peptides were detected in the Glu-C digestion, affording 100% sequence coverage of the glycosylation sites on the protein. Both of these peptides appear in Table 1.

These data demonstrate that the endoglycosidase conditions described here successfully identify when glycosylation is present on the glycopeptide. Additionally, the use of both glycosidase enzymes (Endo H and Endo F3) is important if one does not know the type of glycosylation on the protein in advance, since the enzymes are selective for only certain classes of glycans. While enzyme selectivity can be a disadvantage in that two separate experiments must be conducted, the selectivity is also an advantage because these experiments provide some information about the types of glycans that are present at each of the glycosylation site, without having to do an intense glycopeptide analysis of the fully glycosylated protein.

### 3.3. Partially deglycosylated HIV JRFL envelope

After successfully demonstrating that the experiments outlined in Fig. 3 provided 100% coverage of the glycosylation sites, we applied this same methodology to the same protein, that first had been partially deglycosylated with PNGase F, to show that the method could be used to track that sites had been deglycosylated under non-denaturing conditions. For this deglycosylated protein, the exact level of remaining glycosylation was unknown, however, the protein's dramatic decrease in molecular weight (as observed by SDS-PAGE) suggested that a significant fraction of the glycans had been removed (not shown) [14].

To analyze this partially deglycosylated Env, the same experimental protocol as described in Fig. 3 was used, and all of the species that had been identified in Table 1 were searched for in the new data set. When the ions in Table 1 are also present in the new data set, this indicates that the glycosylation sites corresponding to these ions still have glycans attached. In addition to searching for these species, deglycosylated peptides were also searched for and identified. The fully deglycosylated peptides, generated by PNGase F treatment, convert the N (Asn) at the glycosylation site to D (Asp), concomitant with glycan removal, as described in Fig. 1. Therefore, glycosylation sites that have been deglycosylated are identified by detecting the peptide sequences with N to D conversion at the glycosylation sites. It is important to note that while it is possible for proteins to naturally undergo an N to D conversion at potential glycosylation sites when the asparagines are not glycosylated, this problem is not relevant here because we probed



**Fig. 5.** (A) MS data showing a peptide with two glycosylation sites. This peptide is detected as a fully deglycosylated species ( $m/z$  903) and a partially deglycosylated species ( $m/z$  1004.) The location of the remaining glycosylation site on  $m/z$  1004 is determined using the MS/MS data in (B). This spectrum shows abundant b/y cleavages and verifies that the first of the two glycosylation sites is occupied with a glycan.

for such unusual modifications in the fully glycosylated protein, and found no such instances of this modification. Therefore, any peptides undergoing N to D conversion at a glycosylation site can be confidently identified as a site on the protein that has been deglycosylated.

For peptides with a single glycosylation site, the process of identifying whether or not PNGase F cleaved some or all the glycans is straightforward. For these peptides, one simply searches for the peptide undergoing N to D conversion, (to identify deglycosylated peptides) and the peptide containing one HexNAc at the glycosylation site (to identify cases where glycan is still attached). The analysis is slightly more complicated when a single tryptic peptide contains multiple glycosylation sites, since each site could be glycosylated or deglycosylated. As an example, the peptide ESVEINCTRPNNTR could be present in 4 different forms, either as ESVEIN(HexNAc)CTRPN(HexNAc)NTR, ESVEIDCTRPN(HexNAc)NTR, ESVEIN(HexNAc)CTRPNDNTR, or ESVEIDCTRPNDNTR. While two of these peptides are uniquely identifiable based on their mass, the remaining two glycoforms are only distinguishable by MS/MS analysis. The LC/MS/MS data was searched for each possible form of each peptide, and MS/MS data was used to confirm each of the assignments. Fig. 5 shows an example of the MS and MS/MS data for one of the glycopeptides with two glycosylation sites. Tandem mass spectrometry experiments must be used in this case to distinguish which site is deglycosylated and which site still contains a HexNAc.

After a thorough analysis of the partially deglycosylated Env protein, the results were compiled in Table 2. By coupling data from the tryptic digests with data from the Glu-C digest, all of the glycosylation sites except one were detected. The missing site is the final glycosylated Asn in the V4 loop. In the sample which was not treated with PNGase F, this site was

**Table 1**  
Native JRFL gp140  $\Delta$ CF treated with Endo H and Endo F3.

Env Region	CS	Experimental <i>m/z</i>	Theoretical <i>m/z</i>	Mass Error (ppm)	Peptide Sequence	Glycan
C1	4+	1171.0467	1171.0404	5	AYDTEVHNWVWATHACVPTDPNPQEVVLENVTEHFNMWVK	[HexNAc]
	4+	1207.5623	1207.5549	6	AYDTEVHNWVWATHACVPTDPNPQEVVLENVTEHFNMWVK	[HexNAc][Fuc]
V1	2+	1029.4347	1029.4318	3	DVNATNTTNDSEGTMER	[HexNAc]
	2+	1130.9380	1130.9715	30	DVNATNTTNDSEGTMER	2[HexNAc]
	2+	1277.0503	1277.0294	16	DVNATNTTNDSEGTMER	2[HexNAc][Fuc]
V1-V2	2+	1058.0042	1058.0048	1	NCSFNITTSIRDEVQK	[HexNAc]
	2+	1131.0363	1131.0337	2	NCSFNITTSIRDEVQK	[HexNAc][Fuc]
	3+	754.3633	754.3582	7	NCSFNITTSIRDEVQK	[HexNAc][Fuc]
	2+	1159.5481	1159.5444	3	NCSFNITTSIRDEVQK	2[HexNAc]
	3+	870.7317	870.7373	6	NCSFNITTSIRDEVQK	2[HexNAc][Fuc]
V2	2+	911.9491	911.9445	5	LDVVPIDNNNTSYR	[HexNAc]
	2+	984.9748	984.9735	1	LDVVPIDNNNTSYR	[HexNAc][Fuc]
C2	3+	1274.6783	1274.6737	4	NVSTVQCTHGIRPVVSTQLLNGLSLAEEEEVIR	[HexNAc]
	3+	1342.3722	1342.3668	4	NVSTVQCTHGIRPVVSTQLLNGLSLAEEEEVIR	2[HexNAc]
	4+	1007.0289	1007.0269	2	NVSTVQCTHGIRPVVSTQLLNGLSLAEEEEVIR	2[HexNAc]
	4+	1080.0561	1080.0559	0.2	NVSTVQCTHGIRPVVSTQLLNGLSLAEEEEVIR	2[HexNAc][Fuc]
	1+	1213.5675	1213.5332	28	SDNFTNNAK	[HexNAc]
	1+	1359.5945	1359.5911	3	SDNFTNNAK	[HexNAc][Fuc]
C2-V3	2+	1105.5034	1105.5031	0.3	ESVEINCTRPNNTR	2[HexNAc]
	2+	1251.5450	1251.5610	13	ESVEINCTRPNNTR	2[HexNAc][Fuc]
V3-C3	2+	594.7764	594.7749	3	QAHCNISR	[HexNAc]
C3	2+	589.8063	589.8062	0.2	AKWNTLTK	[HexNAc]
	2+	662.8360	662.8352	1	AKWNTLTK	[HexNAc][Fuc]
	2+	633.8275	633.8199	12	LREQFENK	[HexNAc]
	2+	706.8502	706.8488	2	LREQFENK	[HexNAc][Fuc]
	2+	975.9664	975.9500	17	NKTIVFNHSSGGDPE <sup>a</sup>	[HexNAc][Fuc]
	2+	1004.4618	1004.4607	1	NKTIVFNHSSGGDPE <sup>a</sup>	2[HexNAc]
2+	1077.5077	1077.4897	17	NKTIVFNHSSGGDPE <sup>a</sup>	[HexNAc][HexNAc][Fuc]	
C3-V4	3+	1591.6204	1591.6585	24	IVMHSFNCGGEFFYCNSTQLFNSTWNNTEGSNNTE <sup>a</sup>	[HexNAc][HexNAc][Fuc]
C4	2+	769.4014	769.3981	4	CSSNITGLLLTR	[HexNAc]
	2+	842.4297	842.4271	3	CSSNITGLLLTR	[HexNAc][Fuc]
V5-C5	2+	1148.0197	1148.0189	1	DGGINENGTEIFRPGGGDMR	[HexNAc]
	2+	1221.0534	1221.0479	5	DGGINENGTEIFRPGGGDMR	[HexNAc][Fuc]
TM	2+	1025.9995	1025.9987	1	LICTTAVPWNASWSNK	[HexNAc]
	2+	1099.0324	1099.0277	4	LICTTAVPWNASWSNK	[HexNAc][Fuc]
	2+	1127.5409	1127.5384	2	LICTTAVPWNASWSNK	2[HexNAc]
	2+	1273.6085	1273.5963	10	LICTTAVPWNASWSNK	2[HexNAc][Fuc]
	2+	949.9131	949.9137	1	IWNNTWMEWER	[HexNAc]
	2+	1022.9459	1022.9426	3	IWNNTWMEWER	[HexNAc][Fuc]
	3+	1295.9773	1295.9515	20	EIDNYTSEIYTLIEESQNNQEQKNEQELLELDK	non glycosylated

<sup>a</sup>Peptides generated from Glu-C digestion of the envelope protein, JR-FL gp140  $\Delta$ CF.

only detected on a peptide with a missed cleavage, containing the sequence, VMHSFNCGGEFFYCNSTQLFNSTWNNTEGSNNTE (see Table 1). In the data for the PNGase F treated sample, this peptide was not detected. Instead, the shorter peptide, containing no missed cleavages, FFYCNSTQLFNSTWNNTE, was detected. This shorter peptide lacks the fourth glycosylation site, GSNNTTE.

For the 26 of 27 glycosylation sites that were detected, virtually all of them were detected as either fully deglycosylated peptides or mixtures of fully and partially deglycosylated species. The large

peptide in the C3-V4 region was the only species that was not detected at some level, as a fully deglycosylated glycoform. This peptide was the most difficult to detect species in the entire Env protein, so it is possible that a fully deglycosylated form of this species was present, but below our method's detection limit. Interestingly, the glycosylation sites at the N- and C-terminus of the protein were exclusively deglycosylated; these were likely most accessible to PNGase F. All other sites in the protein were detected as at least partially glycosylated (and therefore, only partially deglycosylated).

**Table 2**  
Partially deglycosylated JRFL gp140 ΔCF treated with Endo H and Endo F3.

Env Region	CS	Experimental m/z	Theoretical m/z	Mass Error (ppm)	Peptide Sequence	Glycan
C1	3+	1493.6951	1493.6863	6	AYDTEVHNWVWATHACVPTDPNPQEVVLE <b>D</b> VTEHFNMWK	
V1	2+	928.3909	928.3842	7	DV <b>D</b> AT <b>N</b> TT <b>N</b> DSEGTMER	
	2+	928.8835	928.8762	8	DV <b>D</b> AT <b>D</b> TT <b>N</b> DSEGTMER	
V1-V2	2+	758.3779	758.3590	25	<b>N</b> CSF <b>N</b> ITTSIR <sup>b</sup>	[HexNAc]
	2+	831.3786	831.3880	11	<b>N</b> CSF <b>N</b> ITTSIR <sup>b</sup>	[HexNAc][Fuc]
	2+	1058.0029	1058.0048	2	<b>N</b> CSF <b>N</b> ITTSIRDEVQK <sup>b</sup>	[HexNAc]
	2+	1159.5515	1159.5444	6	<b>N</b> CSF <b>N</b> ITTSIRDEVQK	2[HexNAc]
	2+	956.9609	956.9571	4	<b>N</b> CSF <b>D</b> ITTSIRDEVQK	
V2	2+	911.9479	911.9445	4	LDVVPIDN <b>N</b> NTSYR	[HexNAc]
	2+	810.8977	810.8968	1	LDVVPIDN <b>D</b> NTSYR	
C2	3+	1275.0082	1275.0017	5	<b>N</b> VSTVQCTHGIRPVVSTQ <b>L</b> LL <b>D</b> GS <b>L</b> AEEEEVVIR	[HexNAc]
	3+	1342.3742	1342.3668	6	<b>N</b> VSTVQCTHGIRPVVSTQ <b>L</b> LL <b>N</b> GS <b>L</b> AEEEEVVIR	2[HexNAc]
	3+	1207.6556	1207.6366	16	<b>D</b> VSTVQCTHGIRPVVSTQ <b>L</b> LL <b>D</b> GS <b>L</b> AEEEEVVIR	
	1+	1213.5409	1213.5332	6	SD <b>N</b> FTNNAK <sup>b</sup>	[HexNAc]
C2-V3	1+	1011.4425	1011.4378	5	SD <b>D</b> FTNNAK	
	2+	1004.4555	1004.4554	0.1	ESVE <b>I</b> NCTRP <b>D</b> NTR	[HexNAc]
C2-V3	2+	1105.5063	1105.5031	3	ESVE <b>I</b> NCTRP <b>N</b> NTR	2[HexNAc]
	2+	902.9205	902.9157	5	ESVE <b>I</b> NCTRP <b>D</b> NTR	
	2+	903.4125	903.4077	5	ESVE <b>I</b> DCTRP <b>D</b> NTR	
	2+	903.4125	903.4077	5	ESVE <b>I</b> DCTRP <b>D</b> NTR	
V3-C3	2+	594.7767	594.7749	3	QAHC <b>N</b> ISR	[HexNAc]
	2+	493.7288	493.7272	3	QAHC <b>D</b> ISR	
C3	1+	777.3800	777.3777	3	W <b>D</b> DTLK <sup>a</sup>	
	1+	1178.6120	1178.6052	6	AKW <b>N</b> DTLK	[HexNAc]
	1+	976.5173	976.5098	8	AKW <b>D</b> DTLK	
	2+	633.8318	633.8199	19	LREQF <b>E</b> NK	[HexNAc]
	1+	1064.5444	1064.5371	7	LREQF <b>E</b> DK	
	2+	532.7743	532.7722	4	LREQF <b>E</b> DK	
	2+	903.4230	903.4130	11	<b>N</b> KTIV <b>F</b> DHSSGGDPE <sup>a</sup>	[HexNAc]
	2+	903.4195	903.4130	7	<b>D</b> KTIV <b>F</b> NHSSGGDPE <sup>a</sup>	[HexNAc]
C3-V4	2+	976.4559	976.4420	14	<b>N</b> KTIV <b>F</b> DHSSGGDPE <sup>a</sup>	[HexNAc][Fuc]
	2+	1004.4714	1004.4607	11	<b>N</b> KTIV <b>F</b> NHSSGGDPE <sup>a</sup>	2[HexNAc]
	2+	801.8792	801.8733	7	<b>D</b> KTIV <b>F</b> NHSSGGDPE <sup>a</sup>	
	2+	802.3737	802.3653	10	<b>D</b> KTIV <b>F</b> DHSSGGDPE <sup>a</sup>	
C3-V4	2+	1397.5998	1397.5766	17	FFY <b>C</b> NSTQL <b>F</b> N <b>S</b> TW <b>D</b> NTE <sup>a</sup>	[HexNAc]2
C4	2+	769.4005	769.3981	3.1	CSS <b>N</b> ITGLLLTR	[HexNAc]
	2+	842.4017	842.4271	30	CSS <b>N</b> ITGLLLTR	[HexNAc][Fuc]
	2+	668.3522	668.3504	2	CSS <b>D</b> ITGLLLTR	
V5-C5	2+	1148.0254	1148.0189	6	DGGIN <b>E</b> NGTEIFRPGGGDMR	[HexNAc]
	2+	1046.9871	1046.9712	15	DGGIN <b>E</b> DGTEIFRPGGGDMR	
TM	2+	1025.9916	1025.9987	7	LICTTAVPW <b>N</b> ASWS <b>N</b> K	[HexNAc]
	2+	1127.5482	1127.5384	9	LICTTAVPW <b>N</b> ASWS <b>N</b> K <sup>b</sup>	2[HexNAc]
	2+	924.9650	924.9510	15	LICTTAVPW <b>N</b> ASWS <b>D</b> K	
	8+	925.4501	925.4430	8	LICTTAVPW <b>D</b> ASWS <b>D</b> K <sup>b</sup>	
	2+	848.8742	848.8660	10	I <b>W</b> ND <b>M</b> T <b>W</b> MEWER	
	2+	1337.6458	1337.6219	18	E <b>I</b> D <b>N</b> Y <b>T</b> SEI <b>Y</b> L <b>I</b> EESQN <b>Q</b> Q <b>E</b> K	non-glycosylated

<sup>a</sup>Peptides generated from Glu-C digestion of the envelope protein, JR-FL gp140 ΔCF.

<sup>b</sup>Peptides with no MS/MS data but eluted within the elution window of either glycosylated or deglycosylated form of the same peptide.

While the apparent molecular weight (from SDS-PAGE) of the PNGase F treated protein suggests that this protein is heavily deglycosylated, the mass spectrometry data shows that deglycosylation for this protein does not occur in a one-site-at-a-time manner. Instead, all of these sites must be at least somewhat accessible to PNGase F, although the extent of accessibility likely varies. Future analyses of samples that have been deglycosylated to different extents could be useful in revealing which areas of the protein are most accessible to PNGase F and, therefore, most accessible to antibodies. Additionally, quantitative analyses, using label free methods or isotopically labeled internal standards, would allow additional insights into how much deglycosylation had occurred at each site. Testing of the efficacy of these vaccines is also ongoing, and those results will be published separately.

#### 4. Conclusion

The data presented herein show the following: the experimental protocol in Fig. 3 can be used to verify when glycosylation is present on glycoproteins. The protocol accurately identified the presence of high mannose and complex glycans present on all the occupied glycosylation sites for the protein studied here. This protocol can also be used to characterize proteins that have been partially deglycosylated with PNGase F. When the protein is partially deglycosylated, both the deglycosylated species and their glycosylated counterparts can be detected in the same experiment. These studies were used to chemically characterize a potential HIV vaccine candidate that has been partially deglycosylated, and these studies show that partial deglycosylation of Env proteins produces a highly heterogeneous product, with most sights being partially deglycosylated.

#### Acknowledgements

This work was supported by NIH grant RO1GM077226, P01AI61734, and a collaboration from the AIDS Vaccine Development Grant from the Bill and Melinda Gates Foundation to Barton F. Haynes. We would also like to acknowledge the Analytical Proteomics Laboratory at KU for instrument time.

#### References

- [1] M. van Wely, N. Bayram, F. van der Veen, Recombinant FSH in alternative doses or versus urinary gonadotrophins for ovulation induction in subfertility associated with polycystic ovary syndrome: a systematic review based on a Cochrane review, *Hum. Reprod.* 18 (2003) 1143–1149.
- [2] S.B. Krantz, Erythropoietin, *Blood* 77 (1991) 419–434.
- [3] R. Wyatt, P.D. Kwong, E. Desjardins, R.W. Sweet, J. Robinson, W.A. Hendrickson, J.G. Sodroski, The antigenic structure of the HIV gp120 envelope glycoprotein, *Nature* 393 (1998) 705–711.
- [4] T. Zhou, L. Xu, B. Dey, A.J. Hessel, R.D. Van, S.H. Xiang, X. Yang, M.Y. Zhang, M.B. Zwick, J. Arthos, D.R. Burton, D.S. Dimitrov, J. Sodroski, R. Wyatt, G.J. Nabel, P.D. Kwong, Structural definition of a conserved neutralization epitope on HIV-1 gp120, *Nature* 445 (2007) 732–737.
- [5] M.N. Fukuda, H. Sasaki, L. Lopez, M. Fukuda, Survival of recombinant erythropoietin in the circulation: the role of carbohydrates, *Blood* 73 (1989) 84–89.
- [6] A.G. Morell, G. Gregoriadis, I.H. Scheinberg, J. Hickman, G. Ashwell, The role of sialic acid in determining the survival of glycoproteins in the circulation, *J. Biol. Chem.* 246 (1971) 1461–1467.
- [7] R.L. Shields, J. Lai, R. Keck, L.Y. O'Connell, K. Hong, Y.G. Meng, S.H. Weikert, L.G. Presta, Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fcγ<sub>3</sub> and antibody-dependent cellular toxicity, *J. Biol. Chem.* 277 (2002) 26733–26740.
- [8] X. Wei, J.M. Decker, S. Wang, H. Hui, J.C. Kappes, X. Wu, J.F. Salazar-Gonzalez, M.G. Salazar, J.M. Kilby, M.S. Saag, N.L. Komarova, M.A. Nowak, B.H. Hahn, P.D. Kwong, G.M. Shaw, Antibody neutralization and escape by HIV-1, *Nature* 422 (2003) 307–312.
- [9] G. Lin, G. Simmons, S. Pohlmann, F. Baribaud, H. Ni, G.J. Leslie, B.S. Haggarty, P. Bates, D. Weissman, J.A. Hoxie, R.W. Doms, Differential N-linked glycosylation of human immunodeficiency virus and Ebola virus envelope glycoproteins modulates interactions with DC-SIGN and DC-SIGNR, *J. Virol.* 77 (2003) 1337–1346.
- [10] M. Shan, P.J. Klasse, K. Banerjee, A.K. Dey, S.P. Iyer, R. Dionisio, D. Charles, L. Campbell-Gardener, W.C. Olson, R.W. Sanders, J.P. Moore, HIV-1 gp120 mannoses induce immunosuppressive responses from dendritic cells, *PLoS Pathog.* 3 (2007) e169.
- [11] S.K. Wang, P.H. Liang, R.D. Astronomo, T.L. Hsu, S.L. Hsieh, D.R. Burton, C.H. Wong, Targeting the carbohydrates on HIV-1: interaction of oligomannose dendrons with human monoclonal antibody 2G12 and DC-SIGN, *Proc. Natl. Acad. Sci. U.S.A.* 105 (2008) 3690–3695.
- [12] J.M. Binley, Y.E. Ban, E.T. Crooks, D. Eggink, K. Osawa, W.R. Schief, R.W. Sanders, Role of complex carbohydrates in human immunodeficiency virus type 1 infection and resistance to antibody neutralization, *J. Virol.* 84 (2010) 5637–5655.
- [13] A. Helenius, M. Aebi, Intracellular functions of N-linked glycans, *Science* 291 (2001) 2364–2369.
- [14] J.M. Ma, S.M. Alam, E.P. Go, H. Desaire, H.X. Liao, B. Haynes, Enhanced antigenicity of the HIV-1 envelope gp41 membrane proximal neutralizing epitope by selective deglycosylation envelope of gp140 oligomer protein, in preparation.
- [15] E.P. Go, J. Irungu, Y. Zhang, D.S. Dalpathado, H.X. Liao, L.L. Sutherland, S.M. Alam, B.F. Haynes, H. Desaire, Glycosylation site-specific analysis of HIV envelope proteins (JR-FL and CON-S) reveals major differences in glycosylation site occupancy, glycoform profiles, and antigenic epitopes' accessibility, *J. Proteome Res.* 7 (2008) 1660–1674.
- [16] E.P. Go, Q. Chang, H.X. Liao, L.L. Sutherland, S.M. Alam, B.F. Haynes, H. Desaire, Glycosylation site-specific analysis of clade C HIV-1 envelope proteins, *J. Proteome Res.* 8 (2009) 4231–4242.
- [17] D.S. Dalpathado, H. Desaire, Glycopeptide analysis by mass spectrometry, *Analyst* 133 (2008) 731–738.
- [18] F. Maley, R.B. Trimble, A.L. Tarentino, T.H. Plummer Jr., Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases, *Anal. Biochem.* 180 (1989) 195–204.
- [19] A.L. Tarentino, G. Quinones, L.M. Changchien, T.H. Plummer Jr., Multiple endoglycosidase F activities expressed by *Flavobacterium meningosepticum* endoglycosidases F2 and F3. Molecular cloning, primary sequence, and enzyme expression, *J. Biol. Chem.* 268 (1993) 9702–9708.
- [20] A.L. Tarentino, T.H. Plummer Jr., Enzymatic deglycosylation of asparagine-linked glycans: purification, properties, and specificity of oligosaccharide-cleaving enzymes from *Flavobacterium meningosepticum*, *Methods Enzymol.* 230 (1994) 44–57.
- [21] P. Hagglund, J. Bunkenborg, F. Elortza, O.N. Jensen, P. Roepstorff, A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation, *J. Proteome Res.* 3 (2004) 556–566.
- [22] P. Hagglund, R. Matthiesen, F. Elortza, P. Hojrup, P. Roepstorff, O.N. Jensen, J. Bunkenborg, An enzymatic deglycosylation scheme enabling identification of core fucosylated N-glycans and O-glycosylation site mapping of human plasma proteins, *J. Proteome Res.* 6 (2007) 3021–3031.
- [23] Z.M. Segu, A. Hussein, M.V. Novotny, Y. Mechref, Assigning N-glycosylation sites of glycoproteins using LC/MSMS in conjunction with Endo-M/exoglycosidase mixture, *J. Proteome Res.* 9 (2010) 3598–3607.
- [24] F. Gao, E.A. Weaver, Z. Lu, Y. Li, H.X. Liao, B. Ma, S.M. Alam, R.M. Scarsce, L.L. Sutherland, J.S. Yu, J.M. Decker, G.M. Shaw, D.C. Montefiori, B.T. Korber, B.H. Hahn, B.F. Haynes, Antigenicity and immunogenicity of a synthetic human immunodeficiency virus type 1 group m consensus envelope glycoprotein, *J. Virol.* 79 (2005) 1154–1163.
- [25] H.X. Liao, L.L. Sutherland, S.M. Xia, M.E. Brock, R.M. Scarsce, S. Vanleeuwen, S.M. Alam, M. McAdams, E.A. Weaver, Z. Camacho, B.J. Ma, Y. Li, J.M. Decker, G.J. Nabel, D.C. Montefiori, B.H. Hahn, B.T. Korber, F. Gao, B.F. Haynes, A group M consensus envelope glycoprotein induces antibodies that neutralize subsets of subtype B and C HIV-1 primary viruses, *Virology* 353 (2006) 268–282.
- [26] F. Gao, H.X. Liao, B.H. Hahn, N.L. Letvin, B.T. Korber, B.F. Haynes, Centralized HIV-1 envelope immunogens and neutralizing antibodies, *Curr. HIV Res.* 5 (2007) 572–577.
- [27] J. Irungu, E.P. Go, Y. Zhang, D.S. Dalpathado, H.X. Liao, B.F. Haynes, H. Desaire, Comparison of HPLC/ESI-FTICR MS versus MALDI-TOF/TOF MS for glycopeptide analysis of a highly glycosylated HIV envelope glycoprotein, *J. Am. Soc. Mass Spectrom.* 19 (2008) 1209–1220.