Glycosylation and Disulfide Bond Analysis of Transiently and Stably Expressed Clade C HIV-1 gp140 Trimers in 293T Cells Identifies Disulfide Heterogeneity Present in Both Proteins and Differences in *O*-Linked Glycosylation

Eden P. Go, David Hua, and Heather Desaire*

Department of Chemistry, University of Kansas, Lawrence, Kansas 66047, United States

S *Supporting Information*

ABSTRACT: The HIV-1 envelope protein (Env) mediates viral entry into host cells to initiate infection and is the sole target of antibody-based vaccine development. Significant efforts have been made toward the design, engineering, and expression of various soluble forms of HIV Env immunogen, yet a highly effective immunogen remains elusive. One of the key challenges in the development of an effective HIV vaccine is the presence of the complex set of post-translational modifications (PTMs) on Env, namely, glycosylation and disulfide bonds, that affect protein folding, epitope accessibility, and immunogenecity. Although these PTMs vary with expression systems, variations in Env's PTMs due to changes in the expression method are not yet well established. In this study, we compared the disulfide bond network and glycosylation profiles of clade C recombinant HIV-1 Env trimers, C97ZA012 gp140, expressed by stable and transient transfections using an integrated mass mapping workflow that combines collision induced dissociation (CID) and electron transfer dissociation (ETD). Site-specific analysis of the *N*- and *O*-glycosylation profiles revealed that C97ZA012 gp140 produced by both transfection methods displayed a high degree of similarity in *N*-glycosylation profiles and site occupancy except for one site. By contrast, different *O*-glycosylation profiles were detected. Analysis of the disulfide bond networks of the Env revealed that both transfection methods yielded C97ZA012 gp140 adopting the expected disulfide bond pattern identified for the monomeric gp120 and gp41 as well as alternative disulfide bond patterns in the C1, V1/V2, and C2 regions. The finding that disulfide bonding is consistently heterogeneous in these proteins is perhaps the most significant outcome of these studies; this disulfide heterogeneity has been reported for multiple other recombinant gp140s, and it is likely present in most recombinantly expressed Env immunogens.

KEYWORDS: *Envelope protein, mass spectrometry, HIV-1, disulfide bond, glycosylation, immunogens, protein expression, glycopeptide, glycosylation site occupancy*
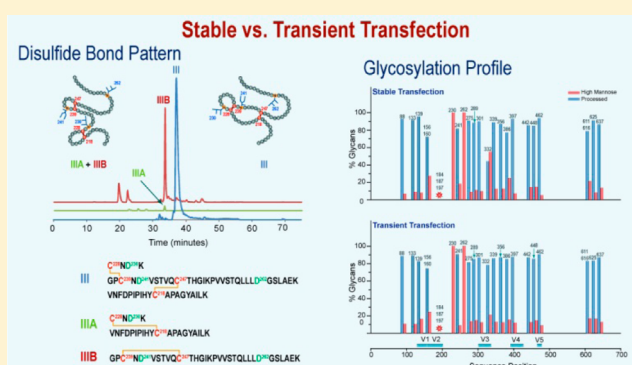
INTRODUCTION

Research toward the development of an HIV vaccine has made considerable progress over the last 30 years, but a highly efficacious vaccine product still remains elusive. Many vaccine strategies currently under development include administration of recombinant versions of the HIV-1 envelope protein (Env), a >100 kDa glycoprotein, that resides in trimeric form on the viral surface.[1] To date, the oligomeric gp140, which consists of the surface glycoprotein, gp120, and the truncated form of the transmembrane glycoprotein, gp41, has been one of the most studied forms of Env-based vaccines. Although the ideal preparation of this antigen is not yet identified, recent clinical findings show that a vaccine including the recombinantly expressed Env antigen has demonstrated the best performance to date of any potential HIV vaccine product.[1,2]

One unique feature of the Env antigen is the remarkably complex set of post-translational modifications present on the protein. It contains >20 *N*-linked glycosylation sites, at least one *O*-linked site, and 10 disulfide bonds. Among the range of PTMs that occur naturally in proteins, those that are reversible, like disulfide bond formation,[3] and diverse, such as glycosylation,[4,5] can significantly impact the structural integrity of proteins and therefore affect their biological functions and activities.[6−8]

Indeed, the importance of glycosylation on Env is well-documented.[9−16] Yet, a few recent publications provide particular insight into the potential impact of glycosylation on HIV vaccine development. Novak and co-workers reported the first comprehensive study of how glycosylation on recombinantly expressed gp120 varies with the cellular expression system used.[13] They further showed that, concomitant with

glycan modifications, antibody binding properties were also substantially impacted when the same protein is expressed in different cell lines.[13] These studies provide the strongest evidence to date that controlling glycosylation on Env is essential for controlling antibody binding. Furthermore, two recent studies have shown that the glycosylation profile on virion-derived Envs is unique and conserved irrespective of the production system and clade compared to the glycosylation profile typically observed on the recombinantly expressed Env.[16,17] Taken together, these studies make a very strong case that Env-based HIV vaccine development necessitates understanding and controlling Env's glycosylation.

Glycosylation on a protein may be modulated by changing the protein sequence or changing the expression conditions. We have characterized the glycosylation from four Envs that were generated using transient transfection[18,19] and four Envs generated using stable cell lines.[20,21] The glycosylation profiles were dramatically different in these cases. The Envs produced from stable cell lines had significantly lower degree of glycosylation, and they contained more high-mannose glycans at key sites.[20,21] Unfortunately, in addition to varying the transfection conditions, the protein sequences were also all unique, so the glycosylation profiles may have been different either because of the changes in protein sequence or because of the change from transient transfection to stable cell lines. Determining causality in this case is important because most preliminary vaccine studies are done using transient transfections, but large quantities of protein for vaccine trials must be produced in stable cell lines. Therefore, identifying potentially confounding differences in the post-translational modifications of Env (or identifying whether no difference exists) is valuable information that will help speed vaccine scale-up.

In addition to glycosylation, disulfide bonding is another important post-translational modification on Env that can impact vaccine antigenicity and immunogenicity. Growing evidence suggests that disulfide bonding patterns in recombinant Env may not always mimic the canonical pattern believed to be on viral, trimeric Env. We have obtained the first disulfide map of an oligomeric gp140 protein and shown that the disulfide bonding in the V1/V2 region was different than the expected profile, mapped almost 20 years ago on a monomeric Env.[22] More recently, we have shown that the recombinant gp120 protein, 1086.C gp120, also has some disulfide heterogeneity in the V1/V2 region.[23] In addition to this direct evidence, other researchers have also recently contributed indirect evidence of disulfide bonding variability in the V1/V2 region. Finzi and co-workers have shown, through a variety of sequence deletion studies, that this part of the protein is involved in forming dimeric Env.[24] Jobes et al. showed that the cysteine variability in the V1/V2 region is high, and they further suggest that this could lead to disulfide bonding differences.[25] Clearly, a better understanding of the disulfide bonding patterns in recombinant Env proteins is necessary, especially when the goal of generating those proteins is to develop a protein that is an effective mimic of trimeric, viral Env.

Here, we compared the disulfide bond topology and glycosylation profiles of a recombinant HIV-1 envelope protein, C97ZA012 gp140, that was produced by both stable and transient transfections in 293T cells with the goal of determining the extent to which the transfection conditions impact the protein's PTM profile. These studies were completed using an integrated MS-based experimental workflow that entails proteolytic digestion of the protein in question followed by a chromatographic separation and then mass analysis, combining collision induced dissociation (CID) and electron transfer dissociation (ETD).[21,23,26] Our results show that both stably and transiently expressed C97ZA012 gp140 displayed similar N-glycosylation profiles and glycosylation site occupancy except for one site. By contrast, O-glycans observed from transiently expressed C97ZA012 gp140 were more heterogeneous than those of stably expressed C97ZA012 gp140. Regarding disulfide bonds, both transfection methods yielded C97ZA012 gp140 with heterogeneous disulfide bond patterns consisting of the expected disulfide bond pattern identified for the monomeric gp120 and gp41[27−31] and alternate disulfide patterns in the C1, V1/V2, and C2 regions. These studies comprehensively address the following question: What changes do the PTM's on HIV-1 envelope proteins undergo when the expression conditions change from transient transfection to production in stable cell lines? The results, that relatively few changes are detected, are quite promising to vaccine developers who rely on these two techniques to produce their proteins. However, these data also show another example of a significant, but typically overlooked, aspect of the heterogeneity of these proteins, namely, a complex disulfide bonding profile.

■ EXPERIMENTAL SECTION

Reagents

Trizma hydrochloride, Trizma base, ethylenediaminetetraacetic acid (EDTA), iodoacetamide (IAM), tris(2-carboxyethyl)-phosphine (TCEP) hydrochloride, dithiothreitol (DTT), urea, HPLC grade acetonitrile (CH_3CN), methanol (CH_3OH), ammonium acetate ($NH_4C_2H_3O_2$), 4-vinylpyridine, glacial acetic acid, 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) was obtained from Promega (Madison, WI). Glycerol-free peptidyl-N-glycosidase F (PNGase F) cloned from *Flavobacterium meningosepticum* was purchased from New England BioLabs (Ipswich, MA), and endo-β-N-acetylglucosaminidase H (Endo H) cloned from *Streptomyces plicatus* was purchased from EMD Millipore (Billerica, MA).

Expression and Purification of Envelope Proteins

The C97ZA012 gp140 trimers used for this study were generous gifts of Bing Chen (Division of Molecular Medicine, Children's Hospital and Department of Pediatrics, Harvard Medical School, Boston, MA). The expression and purification of the C97ZA012 gp140 trimer in a stably transfected 293T cell line were described previously.[32,33] Briefly, a 293T cell line stably transfected with the C97ZA012 gp140 construct was generated by Codex Biosolutions (Gaithersburg, MD). The stable cell lines were grown in DMEM with 10% FBS to confluence and then were changed to Freestyle 293 expression medium (Invitrogen, Life Technologies, Grand Island, NY). The cell supernatants were harvested at 96−108 h after medium change. The His-tagged gp140 protein was purified by Ni-NTA (Qiagen, Inc., Valencia, CA) followed by gel-filtration chromatography as described.[33] The purified proteins were concentrated, frozen in liquid nitrogen, and stored at −80 °C. For expression by transient transfection, 293T cells were grown in DMEM with 10% FBS and transfected with the C97ZA012 gp140 expression construct using calcium phosphate. Transfected cells were then changed to Freestyle 293 expression

medium, and the cell supernatants were harvested at 96−144 h after medium change. The gp140 protein was purified by Ni-NTA and gel-filtration chromatography and stored at −80 °C.

Deglycosylation and Proteolytic Digestion of C97ZA012 gp140 for Disulfide Analysis

Approximately 75 μg samples were alkylated with a 10-fold molar excess of 4-vinylpyridine for 1 h at room temperature in the dark to cap free cysteine residues prior to deglycosylation to prevent disulfide bond shuffling. Deglycosylation was performed by incubating the alkylated Env sample (protein concentration of 12 mg/mL) with 1 μL of PNGase F solution (500 000 units/mL) in 100 mM Tris buffer (pH 7) at 37 °C for 1 week.

Fully deglycosylated alkylated Env samples were digested overnight with trypsin (protein/enzyme ratio of 30) at 37 °C. To generate consistent and reproducible Env digests for disulfide analysis, full deglycosylation and subsequent proteolytic digestion were performed at least two times on different days with gp140 samples obtained from the same expression batches and analyzed with the same experimental procedure. In addition, biological replicates consisting of gp140 samples that were expressed on different days were also analyzed using a buffer with pH 6.5 and the same experimental procedures. To preclude disulfide shuffling artifacts resulting from sample preparation, a control experiment was performed to determine the optimal pH that prevents formation of alternative disulfide bond linkages while maintaining sufficient enzyme activity during deglycosylation and proteolytic digestion. This control experiment was conducted utilizing three separate aliquots of a standard glycoprotein, bovine fetuin. Control samples containing ∼75 μg of bovine fetuin (protein concentration of 12 mg/mL) were fully deglycosylated using the same procedure described in the previous paragraph. In-solution proteolytic digestion were performed in three separate buffers with pH ≤ 7, specifically, pH 5.5, 6.5, and 7.0. A 100 mM Tris buffer, pH 7, was used for in-solution digestion at neutral pH, whereas 50 mM NH$_4$C$_2$H$_3$O$_2$ was used for control samples digested at pH 5.5 or 6.5. Samples digested at pH 5.5 and 6.5 were buffer-exchanged in 50 mM NH$_4$C$_2$H$_3$O$_2$ (pH 5.5 or 6.5) using a 10 kDa molecular weight cutoff filter after deglycosylation. Control samples were incubated overnight at 37 °C with trypsin at a protein/enzyme ratio of 30.

Complete and Partial Deglycosylation of C97ZA012 gp140 for Glycosylation Analysis

Samples containing ∼75 μg (protein concentration of 12 mg/mL) of C97ZA012 gp140 were treated with PNGase F for complete deglycosylation or Endo H for partial deglycosylation. For PNGase F treatment, Env samples were incubated with 1 μL of PNGase F solution (500 000 units/mL) at 37 °C for 1 week at pH 8.0. For Endo H treatment, the pH of the sample solution was adjusted to pH 5.5 with 200 mM HCl. Samples were then denatured with 2 M urea followed by the addition of 2 μL of Endo H (≥5 units/mL). After thorough mixing, the reaction mixture was incubated for 48 h at 37 °C. The pH of the Endo H-treated samples was adjusted to 8.0 with 300 mM NH$_4$OH prior to tryptic digestion. Deglycosylated samples were digested with trypsin as described below.

Proteolytic Digestion of C97ZA012 gp140 for Glycosylation Analysis

Deglycosylated and nondeglycosylated C97ZA012 gp140 samples (75 μg at a concentration of ∼12 mg/mL) were

denatured with 6 M urea in 100 mM tris buffer (pH 8.0) containing 3 mM EDTA and were fully reduced using 5 mM TCEP at room temperature for 1 h. Following reduction, samples were alkylated with 20 mM IAM at room temperature for another hour in the dark. Excess IAM in both samples was quenched by adding DTT to a final concentration of 25 mM and incubating the samples for 20 min at room temperature. The reduced and alkylated Env samples were buffer-exchanged and concentrated using a 30 kDa MWCO filter (Millipore, Billerica, MA) prior to trypsin digestion. Samples were subsequently digested with trypsin (50:1 protein/enzyme ratio) at 37 °C and incubated overnight, followed by a second trypsin addition under the same conditions. The resulting HIV-1 glycoprotein digest was either directly analyzed or stored at −20 °C until further analysis. To ensure reproducibility of the method, protein digestion was performed at least three times on different days with samples obtained from the same batch and analyzed with the same experimental procedure.

Chromatography and Mass Spectrometry

High- and low-resolution LC−MS experiments were performed using two different platforms. The first was a hybrid linear ion-trap Fourier transform ion cyclotron resonance mass spectrometer (LTQ-FT, ThermoScientific, San Jose, CA) coupled to a nanoAcquity Ultra performance liquid chromatography (UPLC) system (Waters, Milford, MA). The second was an LTQ Velos mass spectrometer with ETD (ThermoScientific, San Jose, CA) coupled to Acquity UPLC system (Waters, Milford, MA). In both the disulfide and glycosylation experiments, mobile phases consisted of solvent A (99.9% deionized H$_2$O + 0.1% formic acid) and solvent B (99.9% CH$_3$CN + 0.1% formic acid). Five microliters of the sample (∼7 μM) was injected onto C18 PepMap 300 column (300 μm i.d. × 15 cm, 300 Å; ThermoScientific Dionex, Sunnyvale, CA) at a flow rate of 5 μL/min. The following CH$_3$CN/H$_2$O multistep gradient was used: 5% B for 5 min, followed a linear increase to 40% B in 50 min, and then 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 that there was no sample carry-over. All mass spectrometric analysis was performed in data-dependent mode as described below. For the high-resolution experiments using the LTQ-FT, the ESI source was operated under the following conditions: source voltage of 2.8 kV, capillary temperature of 200 °C, and capillary offset voltage of 44 V. Data were collected in the positive ion mode in a data-dependent fashion, where 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 35% and a 3 min dynamic exclusion window. The FTICR mass spectrometer was set at a resolution (R) of 25 000 at m/z 400. Under these conditions, the measured R (fwhm) at m/z 1000 is 10 000 and at m/z 1500 is 6700.

The LTQ Velos mass spectrometer was set up to perform experiments by alternating CID and ETD acquisition. DDA was set up to acquire 10 scan events: for every one full MS scan in the mass range 300−2000 m/z, each selected m/z in the MS scan was subjected to three MS/MS events, (a) CID, (b) ETD, and (c) CID, of the charge reduced precursor in the previous ETD event. The mass spectrometric parameters used for the experiment were as follows: spray voltage of 3.0 kV, S-lens RF value between 45 and 55%, capillary temperature of 250 °C,
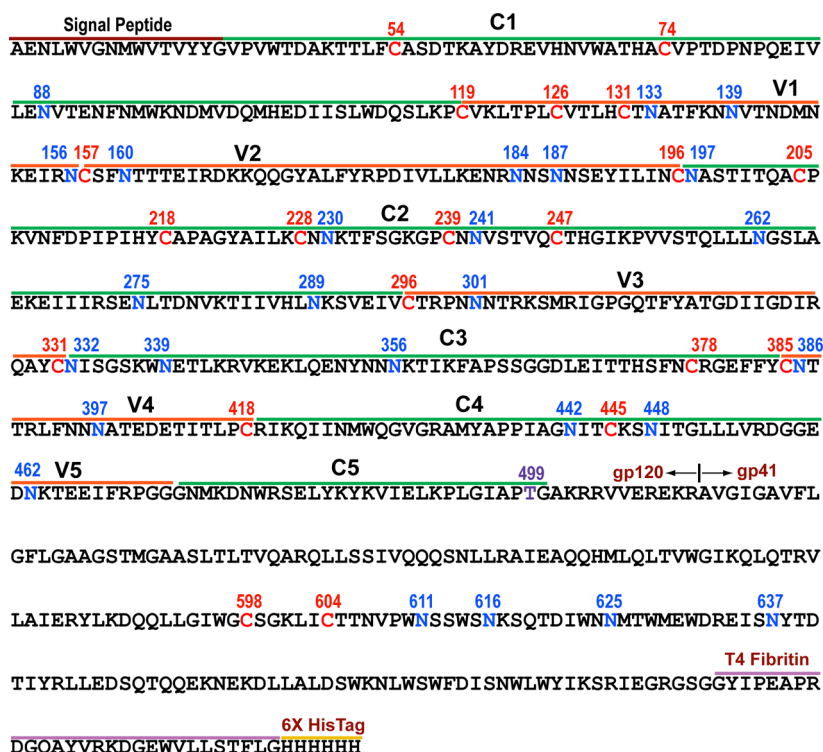
Figure 1. C97ZA012 gp140 sequence showing locations of conserved (C1−C5) and variable (V1−V5) regions, potential N- and O-linked glycosylation sites, and cysteine residues. Sequence positions were standardized using the reference HIV-1 strain, HXB2.

normalized collision energy of 35% for CID, the ion−ion reaction for ETD between the precursor ion and the radical anion, fluoranthane, was set at AGC target value of 2×10^5, and 100 ms ion−ion reaction time. To improve ETD efficiency, supplemental activation was turned on.

Glycopeptide Identification

Data were analyzed using a combination of freely available web-based informatics tools: GlycoPep DB, GlycoPep ID, GlycoMod, and Protein Prospector (http://prospector.ucsf.edu). Details of the compositional analysis have been described previously.[34,35] Briefly, glycopeptide compositional analysis of glycopeptides with one glycosylation site was carried out by first identifying the peptide portion from CID data. The peptide portion was inferred manually or by Glycopep ID after identification of the Y_1 ion, a glycosidic bond cleavage between the two N-acetyl glucosamine at the pentasaccharide core. Once the peptide sequence was determined, plausible glycopeptide compositions were obtained using the high-resolution MS data and GlycoPep DB, and the putative glycan candidate was confirmed manually by identifying the Y_1 ion and inspecting the glycan fragmentation pattern from the tandem MS data. For glycopeptides with multiple glycosylation sites, experimental masses of glycopeptide ions from the high-resolution MS data were converted to the corresponding m/z values of the singly-charged ions and submitted to GlycoMod. This program calculates plausible glycopeptide compositions from the set of experimental m/z values entered by the user, compares these m/z values with theoretical m/z values, and then generated a list of plausible glycopeptide compositions within a specified mass error. Plausible glycopeptide compositions in GlycoMod were deduced by providing the m/z of the singly charged glycopeptide ion, enzyme, protein sequence, cysteine modification, mass tolerance, and the

possible types of glycans present in the glycopeptide. Plausible glycopeptide compositions obtained from the analysis were manually confirmed and validated from MS^2 data.

Raw data containing mixed CID and ETD spectra acquired using an alternating CID/ETD scans on the LTQ-Velos were analyzed manually. ETD spectra of glycopeptides identified from a preceding CID scan were manually assessed for peptide fragment ions using Protein Prospector. Matched fragment ions that were within 0.5 Da of the theoretical value were accepted.

Peptide Identification

Deglycosylated peptides and peptides containing free cysteines were identified by searching raw data acquired on the hybrid LTQ FTICR mass spectrometer against a custom HIV database with 148 gp120/gp41 sequences, obtained from the Los Alamos HIV sequence database (http://www.hiv.lanl.gov/content), using Mascot (Matrix Science, London, UK, version 2.4.1). The peak list was extracted from raw files using Mass Matrix conversion tool. The mgf files were searched specifying the following parameters: (a) enzyme, trypsin; (b) missed cleavage, 2; (c) fixed modification, carbamidomethyl (or pyridlyethyl for vinylpyridine-alkylated samples); (d) variable modification, methionine oxidation, carbamyl, Pyro-glu, Pyro-cmC, HexNAc, and dHexNAc; (e) peptide tolerance, 0.8 Da; and (f) MS/MS tolerance, 0.4 Da. Peptides identified from the Mascot search with 95% confidence were manually validated from MS^2 data to ensure major fragmentation ions (b and y ions) were observed, especially for peptides generated from PNGase F-treated samples that contain N to D conversion.

Disulfide Analysis

Raw data acquired from LTQ-FT and LTQ-Velos experiments were analyzed manually. To facilitate the analysis, peptides containing S-pyridylethyl-labeled cysteines were first identified using the Mascot search engine as described in the previous
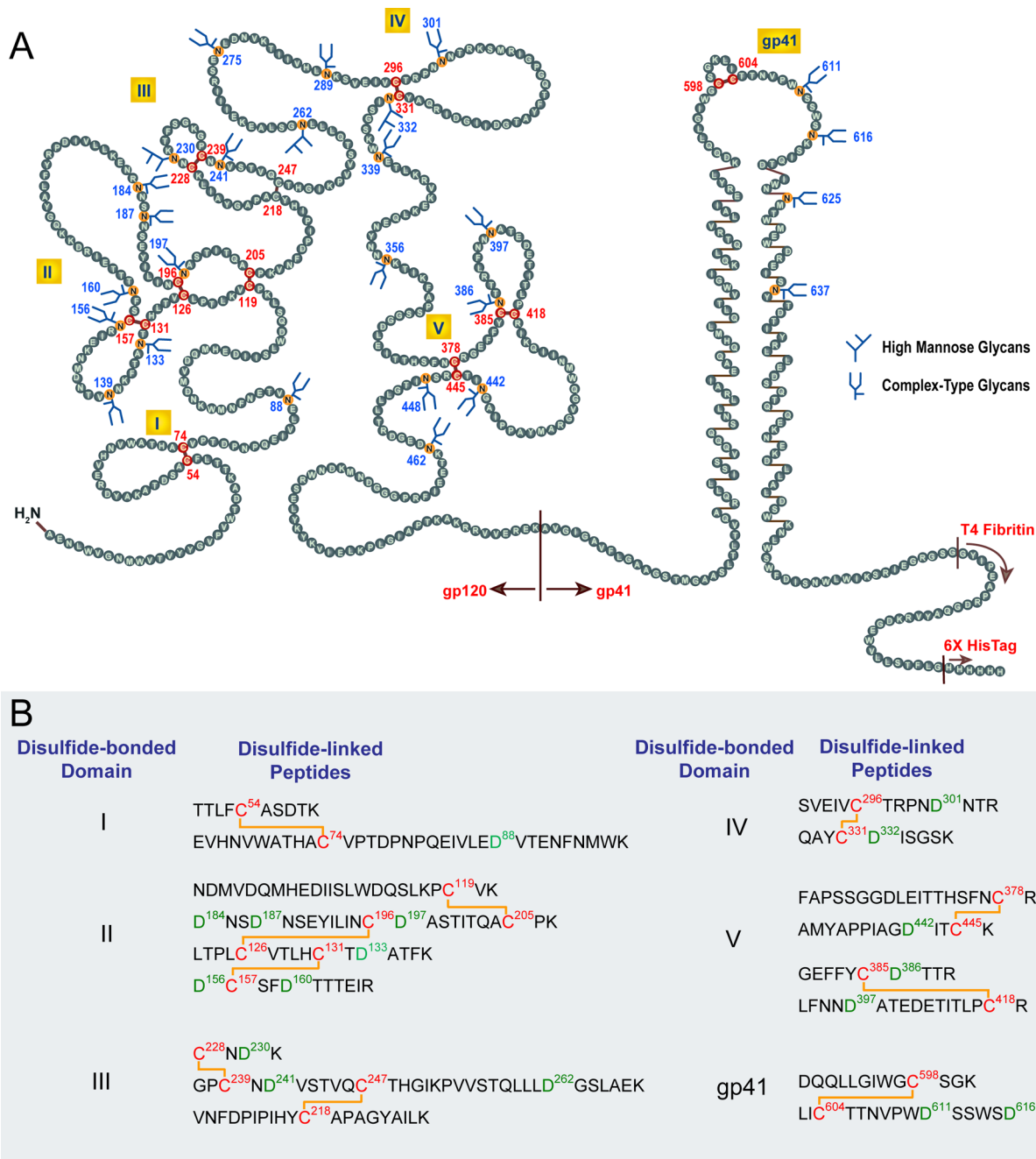
Figure 2. (A) Pictorial representation of C97ZA012 gp140 showing the five disulfide-bonded domains in gp120 and the single disulfide bond in the gp41 region. Pictorial representation was adapted from known disulfide bond arrangements of monomeric gp120 and gp41.[27,30] Locations of the potential glycosylation sites (red) and cysteine residues (in red) are also shown. (B) Expected disulfide-linked peptides for each disulfide-bonded domain generated from tryptic digest of C97ZA012 gp140.

paragraph. Once the free cysteines were identified, a prediction table containing the sum of the masses of all possible disulfide-linked peptide pairs minus 2 Da at different charge states was generated. From the calculated m/z's of the plausible candidates, the elution time profile (or extracted ion chromatogram, XIC) was constructed. For each plausible candidate, the MS[1] data and corresponding CID and/or ETD spectra of the targeted disulfide-linked peptide that are needed for identification were extracted from the constructed XIC. Observed peptide fragment ions from CID (b/y ions) and ETD (c/z ions) data for each identified ion in the MS[1] scan, were manually matched to the theoretical peptide fragment ions

generated using MSProduct from Protein Prospector. Matched fragment ions that were within 0.5 Da of the theoretical value were accepted. In addition to the characteristic c and z ion fragment ions observed in ETD, ions corresponding to the peptide fragments resulting from cleavage of the disulfide bond during ETD were required to be present in the ETD spectra for unambiguous assignment.

■ RESULTS AND DISCUSSION

C97ZA012 gp140 Trimers

The C97ZA012 gp140s were transiently and stably expressed in 293T cells using a gp140 construct containing a C-terminal T4
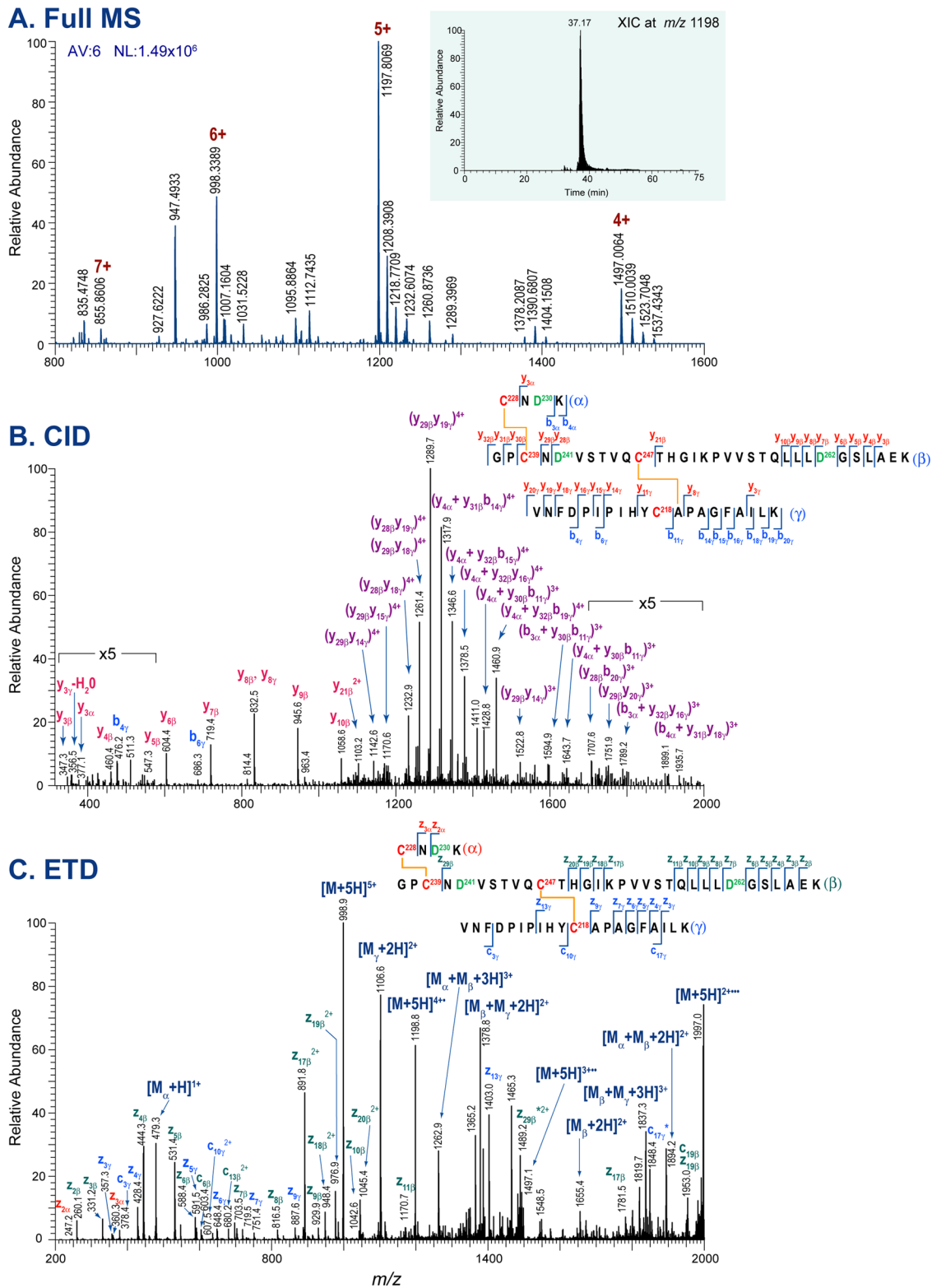
Figure 3. (A) Representative ESI−FTICR−MS data of the three-peptide disulfide-linked chain located in disulfide-bonded domain III. (B) CID and (C) ETD data of the three-peptide disulfide-linked chain supporting the assignment. (Inset) Extracted ion chromatogram (XIC) of the three-peptide disulfide-linked chain.

fibritin foldon trimerization domain and terminated with a histidine tag. The addition of this trimerization motif has been demonstrated to produce stable and homogeneous gp140 trimers at high expression levels with antigenicity that resembles that of native Env.[32,33,36] This protein contains 26

potential *N*-linked glycosylation (PNG) sites, at least one potential *O*-linked glycosylation site, and 20 cysteine residues that could form 10 disulfide bonds. Figure 1 shows the location of the 26 PNG sites (blue), a potential *O*-linked glycosylation site (purple) near the end of the gp120 region, and the 20 Cys
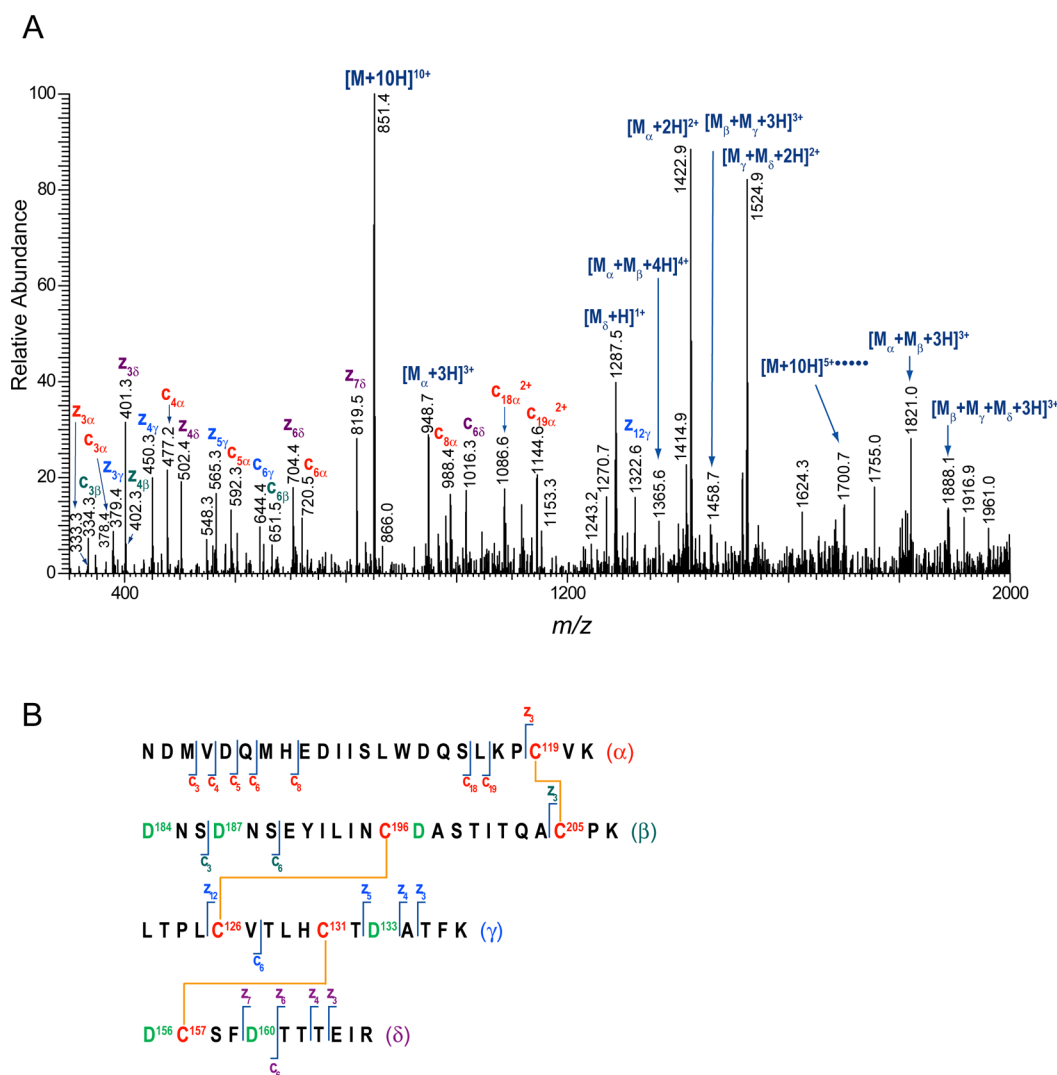
Figure 4. (A) ETD spectrum of the four-peptide disulfide-bonded chain located in the disulfide-bonded domain II obtained from the tryptic digest of stably expressed C97ZA012 gp140. (B) Peptide sequence with the corresponding observed c and z ions.

residues (red) in the conserved regions, C1–C5, in the variable regions, V1–V5, and in the gp41 region. Characterization of disulfide bond topology and the *N*- and *O*-glycosylation profiles of transiently and stably expressed C97ZA012 gp140 was accomplished using a typical in-solution proteolytic digestion followed by LC–MS and MS/MS analysis using both CID and ETD, as described in the following sections.

Disulfide Bond Mapping Approach

Because of the extensive glycosylation of HIV-1 Env, disulfide bond mass mapping necessitates complete deglycosylation prior to proteolytic digestion using trypsin. In an effort to maintain sufficient glycosidase and trypsin activity during deglycosylation and proteolytic digestion as well as to preserve the disulfide bond topology that reflects the native conformation of C97ZA012 gp140 trimers, both deglycosylation and trypsin digestion were performed at pH 7. In addition, potential free cysteines were capped with 4-vinylpyridine prior to deglycosylation at the same pH. Tryptic digests of the deglycosylated Env samples were analyzed by LC–MS using a hybrid LTQ-FTICR mass spectrometer for high-accuracy and high-resolution measurements and an LTQ-Velos mass spectrometer for ETD experiments using the same gradient and column. Data

analysis was done by first identifying peptides containing free cysteines using the Mascot search engine, as described in the Experimental Section. Results from the Mascot search indicated the absence of free cysteines in C97ZA012 gp140 trimers produced by stable and transient transfection.

To address the issue of disulfide shuffling during sample preparation specifically at neutral pH, a control experiment was conducted using three separate aliquots bovine fetuin samples, which were digested in buffers with pH $\leq$ 7, as described in the Experimental Section. MS analysis of the control samples digested in buffers with pH 5.5, 6.5, and 7.0 indicated the absence of alternate disulfide bond patterns (data not shown). Once we established that the disulfides in the control samples did not undergo shuffling at pH $\leq$ 7, a separate set of experiments was conducted for stably expressed C97ZA012 gp140 trimers under the same experimental conditions utilized for bovine fetuin. In addition to these experiments, we also analyzed Env samples expressed on different days under similar experimental conditions (with a buffer pH of 6.5), as described in the Experimental Section, to ensure the reproducibility between biological replicates. Results from these experiments show that disulfide bond patterns observed on the C97ZA012 gp140 trimers were highly reproducible.

Table 1. Mass Assignments of Expected Disulfide Bond in Stably and Transiently Expressed C97ZA012 gp140

Disulfide Loop Domain	Disulfide-linked Peptides	CS	Theoretical m/z	Stably Expressed C97ZA012 gp140		Transiently Expressed C97ZA012 gp140	
				Experimental m/z	Mass Error (ppm)	Experimental m/z	Mass Error (ppm)
I	TTLFC^{54}ASDTK / EVHNVWATHAC^{74}VPTDPNPQEIVLED^{88}VTENFNMWK	4+	1262.3400	1262.3470	6	1262.3376	2
		5+	1010.0735	1010.0709	3	1010.0723	1
		6+	841.8958	841.8997	5	841.8947	1
II	NDMVDQMHEDIISLWDQSLKPC^{119}VK / D^{184}NSD^{187}NSEYILINC^{196}D^{197}ASTITQAC^{205}PK / LTPLC^{126}VTLHC^{131}TD^{133}ATFK / D^{156}C^{157}SFD^{160}TTTEIR	6+	1417.6505	1417.6435	5	1417.6438	5
		7+	1215.2729	1215.2820	8	1215.2626	9
		8+	1063.4897	1063.4972	4	1063.4887	1
		9+	945.4361	945.4431	7	945.4379	2
III	C^{228}ND^{230}K / GPC^{239}ND^{241}VSTVQC^{247}THGIKPVVSTQLLLD^{262}GSLAEK / VNFDPIPIHYC^{218}APAGYAILK	4+	1496.9978	1497.0064	6	1496.9909	5
		5+	1197.7997	1197.8069	6	1197.7980	1
		6+	998.3343	998.3389	5	998.3331	1
		7+	855.8590	855.8606	2	855.8552	4
IV	SVEIVC^{296}TRPND^{301}NTR / QAYC^{331}D^{332}ISGSK	2+	1336.6238	1336.6267	2	1336.6209	2
		3+	891.4183	891.4204	2	891.4174	1
		4+	668.8155	668.8187	5	668.8152	0.4
V	FAPSSGGDLEITTHSFNC^{378}R / AMYAPPIAGD^{442}ITC^{445}K	3+	1162.8757	1162.8850	8	1162.8747	1
		4+	872.4086	872.4121	4	872.4090	0.5
		5+	698.1283	698.1317	5	698.1292	1
	GEFFYC^{385}D^{386}TR / LFNND^{397}ATEDETITLPC^{418}R	2+	1594.2028	1594.2114	5	1594.1978	3
		3+	1063.1468	1062.1422	4	1063.1363	10
		4+	797.6050	797.6053	0.4	797.6033	2
gp41	DQQLLGIWGC^{598}SGK / LIC^{604}TTNVPWD^{611}SSWSD^{616}K	3+	1085.1813	1085.1870	5	1085.1802	1
		4+	814.1378	814.1387	1	814.1405	3

The disulfide bond topology of C97ZA012 gp140 was determined by identifying all plausible disulfide-linked peptides from the tryptic digests of C97ZA012 gp140. To efficiently map the disulfide linkages, disulfide-linked peptides with linkage patterns identical to the established disulfide bond topology for the monomeric gp120 and gp41 were first identified.[27−31] For simplicity, these disulfide linkages will be referred to as expected disulfide bonds. The disulfide bond map that reflects the expected disulfide bond pattern in C97ZA012 gp140 is shown in Figure 2A. Tryptic digests of C97ZA012 gp140 generate three sets of disulfide-linked peptides consisting of five two-peptide disulfide-linked chains, a three-peptide disulfide-linked chain, and a four-peptide disulfide-linked chain (Figure 2B), which were all identified and assigned manually using MS and tandem MS data.

In a typical analysis, a prediction table was created containing the theoretical m/z's of the disulfide-linked peptides in different charge states. These theoretical m/z's were then matched to the observed m/z's in the total ion chromatogram (TIC). Once a match was found, the extracted ion chromatogram (XIC) was constructed using the m/z of the most intense charge state of the targeted disulfide-linked peptides. From the XIC, the MS[1] data containing the ions that correspond to the targeted disulfide-linked peptides along with the MS[2] data needed to assign the targeted disulfide-linked peptides were obtained within the elution time window. As an example, Figure 3A−C

illustrates the assignment of the three-peptide disulfide-linked chain generated from the tryptic digest of stably expressed C97ZA012 gp140. These disulfide-linked peptides consist of two fully deglycosylated peptides, C^{228}ND^{230}K (α) and GPC^{239}ND^{241}VSTVQC^{247}THGIKPVVSTQLLLD^{262}GSLAEK (β), and a nonglycosylated peptide, VNFDPIPIHYC^{218}APA-GYAILK (γ), connected by two disulfide bonds. Figure 3A shows a representative MS[1] data averaged over the constructed XIC window (inset) at full-width-half-maximum (fwhm). In this spectrum, the four ions corresponding to the three-peptide disulfide-linked chain were observed at m/z 1497.0064 (4+), 1197.8069 (5+), 998.3389 (6+), and 855.8606 (7+). The identity of these ions was determined using the CID (Figure 3B) and ETD (Figure 3C) data within the reconstructed XIC window. The CID spectrum of m/z 1198 (5+) in Figure 3B shows the b and y ions from each peptide as well as peptide fragment ions with the disulfide bond intact. Similarly, the ETD spectrum of m/z 998 (6+) in Figure 3C shows contiguous c and z ion series resulting from the peptide backbone fragmentation of each peptide, the peptide fragment ions resulting in the cleavage of the disulfide bond at m/z 479 (α, 1+), 1655 (β, 2+), and 1101 (γ, 2+), and the disulfide bonded peptide pairs, $\alpha + \beta$ at m/z 1261 (3+) and 1893 (2+) and $\beta + \gamma$ at m/z 1837 (2+). From the sequence information obtained from both CID and ETD spectra, the disulfide bond pattern of

Table 2. Mass Assignments of Alternative Disulfide Bond in Stably and Transiently Expressed C97ZA012 gp140

Disulfide Loop Domain	Disulfide-linked Peptides	CS	Theoretical m/z	Stably Expressed C97ZA012 gp140		Transiently Expressed C97ZA012 gp140*	
				Experimental m/z	Mass Error (ppm)	Experimental m/z	Mass Error (ppm)
I-II	TTLFC54ASDTK / NDMVDQMHEDIISLWDQSLKPC119VK	3+	1309.9472	1309.9517	3	1309.9452	2
		4+	982.7122	982.7164	4	982.7119	0.3
		5+	786.3712	786.3736	3	786.3714	0.2
	EVHNVWATHAC74VPTDPNPQEIVLED88VTENFNMWK / D156C157SFD160TTTEIR	4+	1312.6106	1312.6079	2	1312.6010	7
		5+	1050.2828	1050.2825	0.3	1050.2839	1
		6+	875.4022	875.4030	1	875.3984	4
II	NDMVDQMHEDIISLWDQSLKPC119VK / D156C157SFD160TTTEIR	3+	1376.9601	1376.9686	6	1376.9619	0.1
		4+	1032.9719	1032.9800	8	1032.9698	2
		5+	826.5790	826.5834	5	826.5821	4
	D184NSD187NSEYILINC196D197ASTITQAC205PK	2+	1306.5838	1306.5882	3	1306.5837	0.1
		3+	871.3916	871.3939	3	871.3938	3
	LTPLC126VTLHC131TD133ATFK	2+	880.9391	880.9434	5	880.9347	5
		3+	587.6285	587.6310	4	587.6270	3
	LTPLC126VTLHC131TD133ATFK / D184NSD187NSEYILINC196D197ASTITQAC205PK	3+	1458.3408	1458.3452	3	ND	ND
		4+	1094.0075	1094.0117	4		
		5+	875.4074	875.4091	2		
III	VNFDPIPIHYC218APAGYAILK / C228ND230K	2+	1339.6589	1339.6696	8	1339.6575	1
		3+	893.4417	893.4446	3	893.4447	3
	GPC239ND241VSTVQC247THGIKPVVSTQLLLD262GSLAEK	2+	1654.3367	1654.3430	4	1654.3377	1
		3+	1103.2269	1103.2323	5	1103.2324	5
		4+	827.6720	827.6748	3	827.6742	3

*ND means not detected.

the three-peptide disulfide-linked chain located in disulfide-bonded domain III (C2 region) was deduced.

While the use of CID and ETD data have facilitated the assignment and allowed for the unambiguous identification of the disulfide-linked peptides, it is of interest to note the relative efficiencies of CID and ETD with respect to the assignment of the four-peptide disulfide-linked chain. The CID spectrum of the four-peptide disulfide-linked chain did not contain sufficient ions for the assignment (data not shown), but the corresponding ETD spectrum (Figure 4) of the four-peptide disulfide-linked chain at m/z 850 (10+) provided a series of characteristic c and z ions of each peptide in the low mass region, the peptide fragment ions resulting from the cleavage of the disulfide bond at m/z 948 (α, 3+) and 1287 (δ, 1+), and the disulfide-bonded peptide pairs, $\alpha + \beta$ at m/z 1365 (4+) and 1821 (3+), $\beta + \gamma$ at m/z 1458 (3+), $\gamma + \delta$ at m/z 1525 (3+), and $\beta + \gamma + \delta$ at m/z 1888 (3+). These ions were sufficient to ensure confident identification of the four-peptide disulfide-linked chain. The ability of ETD to provide sequence information for larger species with mass >3000 Da[37] and at the same time to efficiently cleave disulfide bonds[38] reflects the merits of ETD in the analysis of peptide chains held by multiple disulfide bonds. Overall, we have identified and assigned all seven disulfide-linked peptides shown in Figure 2B in both stably and transiently expressed C97ZA012 gp140 (Table 1). These results indicate that C97ZA012 gp140 produced from stable and transient transfections yielded C97ZA012 gp140

containing disulfide bond topology identical to the expected disulfide bond pattern consisting of three simple disulfide loops located in the disulfide-bonded domains I, IV, and gp41 and three complex disulfide loops located in disulfide-bonded domains II, III, and IV (Figure 2A). Tandem MS data supporting these assignments are shown in Supporting Information Figures S1−S5.

Identification of Alternative Disulfide Bonds

Having established that C97ZA012 gp140s produced in stable and transient transfections contain cysteine pairings consistent with the expected disulfide bond pattern of the monomeric gp120 and gp41, we explored the possibility of identifying alternative disulfide bonds. Considering the different sets of disulfide bonds that can be formed from the 20 cysteines, it would not be possible to form all linkages because of energetic and structural constraints. As a result, a disulfide bond would be favorably formed between two cysteines that are optimally located from each other in the Env structure. Indeed, variations in the disulfide bond arrangements in the V1/V2 regions have been observed in the oligomeric synthetic Env, CON-S gp140 ΔCFI, the recombinant clade C transmitted/founder monomeric Env, 1086.C gp120 expressed in 293T cells, and clade B gp120 sequences derived from early HIV-1 infections from phase 3 vaccine efficacy trials.[22,23,25] On the basis of these studies, we asked whether such heterogeneity is a common or an inherent feature among recombinant Envs. If so, then are there other possible alternative disulfide bonds from other

disulfide-bonded domains? We address these issues by systematically identifying alternative disulfide bonds within and between disulfide-bonded domains.

Disulfide-Bonded Domain II: V1/V2 Loop. This domain has six cysteines and is located in V1/V2 region. There are 15 possible ways in which three disulfide bonds can be formed from the six cysteines in this domain. Each plausible candidate was evaluated from the MS data that can be extracted from its corresponding XIC. Aside from expected disulfide bond arrangement (Figure 2A) defined by the cysteine pairs, $Cys^{119}-Cys^{205}$, $Cys^{126}-Cys^{196}$, and $Cys^{131}-Cys^{157}$, for both stably and transiently expressed C97ZA012 gp140, LC−MS and MS/MS analysis directly identified five alternate cysteine pairs, $Cys^{126}-Cys^{131}$, $Cys^{196}-Cys^{205}$, $Cys^{126}-Cys^{196}$, $Cys^{131}-Cys^{205}$, and $Cys^{119}-Cys^{157}$. All five cysteine pairs were observed in stably expressed C97ZA012 gp140, whereas only three alternate cysteine pairs were observed in transiently expressed C97ZA012 gp140 (Table 2). The corresponding mass assignments of the disulfide-linked tryptic peptides associated with these alternative cysteine pairings for both stably and transiently expressed C97ZA012 gp140 are shown in Table 2. All of these assignments were confirmed by both CID and ETD data, and representative tandem MS data that support these assignments are shown in Supporting Information Figures S8−S11. Comparison of the data revealed that C97ZA012 gp140 produced from stable and transient transfections differ in two alternate disulfide linkages defined by the cysteine pairings $Cys^{126}-Cys^{196}$ and $Cys^{131}-Cys^{205}$ (Table 2). These cysteine pairings were observed only in stably expressed C97ZA012 gp140 (Table 2). Interestingly, the disulfide bond configuration in this domain common to both stably and transiently expressed C97ZA012 gp140s (Table 2) is also similar to that observed in the synthetic oligomeric CON-S gp140[22] and CHO-derived 1086.C gp120,[23] suggesting that this disulfide bond pattern is an inherent feature of recombinantly expressed HIV-1 Envs

Disulfide-Bonded Domain III: C2 region. The expected disulfide bond pattern in this domain is a single double loop structure with cysteine pairings $Cys^{228}-Cys^{239}$ and $Cys^{247}-Cys^{218}$, identified from LC−MS and MS/MS analysis (Figure 2A). In addition to the expected disulfide linkages, two alternate disulfide linkages were identified, wherein Cys^{218} was linked to Cys^{228} and Cys^{239} was linked to Cys^{247} (Table 2). These alternative disulfide linkages were observed in both stably and transiently expressed C97ZA012 gp140 trimers. Mass assignments of the alternative disulfide-linked peptides identified from high-resolution measurements are shown in Table 2. The corresponding representative tandem MS data used to support these assignments are found in Supporting Information Figures S12 and S13.

Disulfide-Bonded Domain V: V4−C4 Region. Disulfide bonds derived from the four cysteines in this domain are defined by the linkage of the tryptic peptides FAPSSGGD-LEITTHSFNC[378]R with AMYAPPIAGD[442]ITC[445]K and GEF-FYC[385]D[386]TR with LFNND[397]ATEDETITLPC[418]R. Although two more alternative disulfide arrangements are theoretically possible, there were no alternate linkages observed experimentally in both stably and transiently expressed C97ZA012 gp140 trimers.

Alternate Disulfide Linkages between Domains. We also examined the cysteine pairings between the disulfide-bonded domains. Cysteines that form disulfide bonds are usually optimally located from each other in the protein's structure, where an optimal location would include cysteines that are either in close proximity to each other or located in protein regions with loops and no regular secondary structures. Although no disulfide linkages were identified between domains in previous studies,[22,27,28] we identified alternative disulfide linkages between peptides associated with disulfide domains I and II in both stably and transiently expressed C97ZA012 gp140 trimers. From different sets of LC−MS data and separate aliquots of Env samples digested at pH 5.5, 6.5, and 7.0 as well as those acquired from biological replicates, we consistently identified two alternate linkages consisting of the peptides TTLFC[54]ASDTK, in disulfide-bonded domain I linked to the peptide, NDMVDQMHEDIISLWDQSLKPC[119]VK, in disulfide-bonded domain II, and EVHNVWATHAC[74]VPTDPNPQ-EIVLED[88]VTENFNMWK, in disulfide-bonded domain I linked to D[156]C[157]SFD[160]TTTEIR in disulfide-bonded domain II. Representative tandem MS data supporting these assignments are found in Supporting Information Figures S6−S7, and the mass assignments are provided in Table 2.

Disulfide Bond Topology of Stably vs Transiently Expressed C97ZA012 gp140 Trimers

Tables 1 and 2 summarize all of the experimentally determined disulfide-linked peptides in C97ZA012 gp140 trimers. Our results indicate that both transfection methods yielded secretion of C97ZA012 gp140 with heterogeneous disulfide profiles that included the 10 disulfide bonds, consistent with the expected disulfide bond pattern of monomeric gp120 and gp41,[27−31] seven identical alternate disulfide bonds, and two alternate disulfide bonds present only in stably expressed C97ZA012 gp140 (Table 2). Examination of the data revealed that disulfide linkages in disulfide-bonded domains IV, V, and in the gp41 regions were intact, whereas the cysteines in disulfide-bonded domains I, II, and III formed alternate disulfide linkages in addition to the expected disulfide linkages in these domains.

Clearly, these proteins have heterogeneous disulfide bonding. However, it is important to note that these proteins' antigenicity profiles also very closely resemble that of native Env: they do not bind CD4i antibodies in the absence of CD4; they are reactive with the bNABs PG9 and PG16, which target quaternary epitopes present on trimers; and they do not interact with MPER-directed antibodies.[33] Therefore, it is possible that the disulfide heterogeneity detected on this recombinant, trimeric protein is, in fact, consistent with that found on viral Env. Further experiments must be done to determine the reason that these species are so reproducibly detected among different preparations of gp140.

Site-Specific Glycosylation Analysis of Clade C gp140 Trimers

Having determined the disulfide bond map of C97ZA012 gp140 trimers produced from two different transfection methods, the glycan profile and glycosylation site occupancy of the 26 PNG sites were determined and compared using an integrated glycopeptide-based MS approach described in our previous studies.[20,21,39] In a typical experiment, glycopeptides generated from in-solution trypsin digest of untreated or glycosidase-treated gp140 samples were subjected to LC−MS and MS/MS analysis. Raw data acquired from the experiment were subsequently analyzed using software analysis tools Mascot, GlycoPep DB, GlycoMod, and GlycoPep ID[34,35,40] and by manual evaluation of tandem MS data to deduce the glycopeptide compositions. Tryptic digests of C97ZA012

Table 3. Glycosylation Site Occupancy of Stably and Transiently Expressed C97ZA012 gp140

Peptides Containing Potential Glycosylation Sites	No. of Potential Glycosylation sites	Occupancy	
		Stably Expressed C97ZA012 gp140	Transiently Expressed C97ZA012 gp140
EVHNVWATHACVPTDPNPQEIVLEN^{88}VTENFNMWK	1	0 and 1	0 and 1
LTPLCVTLHCTN^{133}ATFK	1	0 and 1	0 and 1
NN^{139}VTNDMNK	1	1	0 and 1
N^{156}CSFN^{160}TTTEIR	2	0, 1 and 2	0, 1 and 2
N^{184}NSN^{187}NSEYILINCN^{197}AST ITQACPK	3	2 and 3	2 and 3
CNN^{230}K	1	1	1
GPCNN^{241}VSTVQCTHGIKPVVSTQLLLN^{262}GSLAEK	2	0, 1, and 2	0, 1, and 2
SEN^{275}LTDNVK	1	0 and 1	0 and 1
TIIVHLN^{289}K	1	0 and 1	0 and 1
SVEIVCTRPNN^{301}NTR	1	0 and 1	0 and 1
QAYCN^{332}ISGSK	1	0 and 1	0 and 1
WN^{339}ETLK/ WN^{339}ETLKR	1	0 and 1	0 and 1
LQENYNNN^{356}K	1	1	1
GEFFYCN^{386}TTR	1	0 and 1	0 and 1
LFNNN^{397}ATEDETITLPCR	1	0 and 1	0 and 1
AMYAPPIAGN^{442}ITCK	1	0 and 1	0 and 1
SN^{448}ITGLLLVR	1	0 and 1	0 and 1
DGGEDN^{462}KTEEIFRPGGGNMK	1	0 and 1	0 and 1
LICTTNVPWN^{611}SSWSN^{616}K	2	0, 1, and 2	0, 1, and 2
SQTDIWNN^{625}MTWMEWDR	1	0 and 1	0 and 1
EISN^{637}YTDTIYR	1	0 and 1	0 and 1
VIELKPLGIAPT^{499}GAK	1	0 and 1	0 and 1

gp140 generated 21 tryptic peptides bearing PNG sites, of which 17 peptides had a single PNG site, three peptides had two PNG sites, and one peptide had three PNG sites (Table 3). In addition, we also identified an O-linked site (T499) near the end of the gp120 region. Glycosylation site occupancy of each of these sites was determined from LC−MS and MS/MS analysis of glycosidase treated Env samples.[20,21] The use of the glycosidase allowed for the unambiguous identification of glycosylated and nonglycosylated sequons. Raw data acquired from these experiments were analyzed using Mascot and by manual evaluation of tandem MS data. The degree of glycosylation site occupancy of stably and transiently expressed C97ZA012 gp140 trimers is summarized in Table 3. The data show that out of 26 PNG sites four sites (N139, N184, N230, and N356) were fully occupied in stably expressed C97ZA012 gp140, whereas three sites (N184, N230, and N356) were fully occupied in transiently expressed C97ZA012 gp140. Overall, C97ZA012 gp140 produced by both expression methods has essentially identical site occupancy.

For the glycosylation sites that are occupied, glycan compositions at each of the potential N- and O-linked glycosylation site were deduced from both MS^1 and tandem MS data originating from the same elution window in the TIC. Typical data are shown in Figure 5 (a TIC in shown in Figure 5A). Figure 5B shows the averaged mass spectrum from the retention time range specified on the TIC. Putative glycopeptide species were identified from peaks separated by the masses of monosaccharide units. Each glycopeptide peak in the spectrum (Figure 5B) was assigned using GlycoPep DB,[34] GlycoMod,[40] and GlycoPep ID[35] and was manually verified

from fragment ion information obtained from both CID and ETD data to elucidate the compositions.

As an example, Figure 6A,B shows representative CID and ETD data used to assign the high-mannose glycopeptide containing the peptide portion, LICTTNVPWN^{611}SSWSN^{616}K, with one site that is modified. CID analysis of the doubly charged high-mannose glycopeptide at m/z 1562 (Figure 6A) shows product ion spectrum typical of a glycopeptide fragmentation in an ion trap. Fragment ions observed in the CID spectrum consist mostly of ions resulting from the glycosidic bond cleavages from which the glycan component was deduced. The lower half of the mass range was dominated by the oxonium ions at m/z 528, 690, 852, and 1014, the Y_1 ion from which the peptide portion is inferred, a single peptide backbone cleavage at m/z 802 (b_7), and the m/z of the peptide portion at 954 (2+). While all of this information is sufficient to deduce the glycopeptide composition, additional information from ETD data is needed to unambiguously identify which of the two sites, N611 or N616, is modified with the glycan. The ETD spectrum of the triply charged species of the same glycopeptide in Figure 6B shows that out of the 16 available peptide backbone bond cleavages, 13 bond cleavages were observed. The contiguous fragment ion series, z_{3-6}, z_{13-15}, c_{5-6}, c_{8-9}, and c_{12-15}, from the ETD fragmentation not only confirmed the initial peptide assignment from CID data but also established that the high-mannose glycan is appended to N611. Using both CID and ETD, we identified the four sets of glycopeptides in Figure 5B, including an O-linked modified peptide eluting within the selected retention window of the TIC. CID and ETD analysis revealed that the potential O-linked glycopeptide contains the peptide portion, VIELKPL-
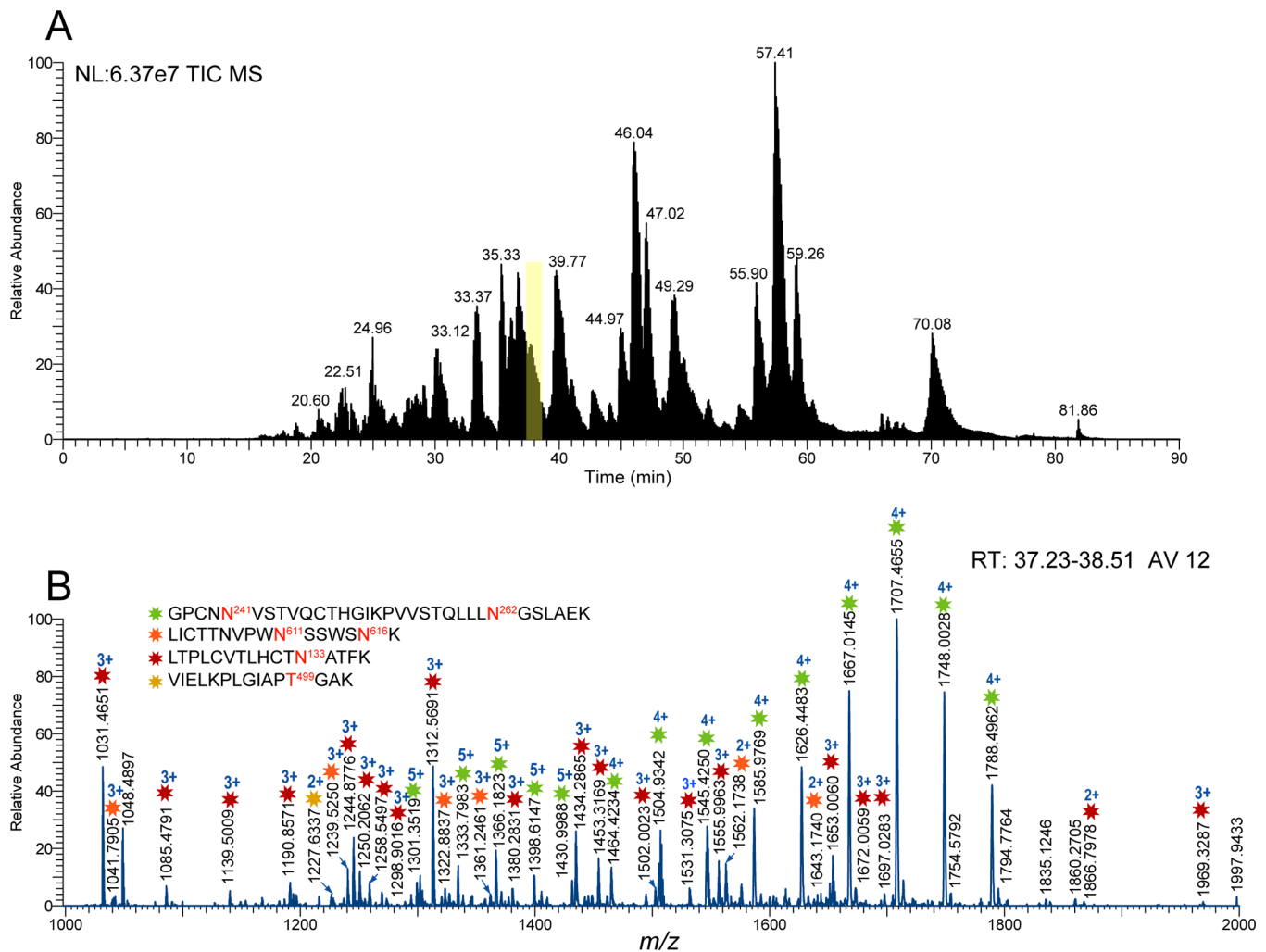
Figure 5. (A) Total ion chromatogram (TIC) and (B) ESI−FTICR−MS data of tryptic digest of stably expressed C97ZA012 gp140. The mass spectrum averaged within 37.23−38.51 min shows elution of four glycopeptide species. Asterisks (*) in the MS data denotes glycopeptide peaks. Glycan compositions of each observed glycopeptide peak in the mass spectra were verified from CID and ETD data.

GIAPT[499]GAK, modified with disialylated O-linked glycan (Figure 6C,D). The glycopeptides identified in this study for both stably and transiently expressed C97ZA012 gp140 are summarized in Supporting Information Tables S1A, S1B, S2A, and S2B. Overall, we identified ∼500 unique glycopeptides representing the 26 potential N-linked sites and the lone potential O-linked site for each of the gp140 trimers.

Glycosylation Profile of Stably vs Transiently Expressed C97ZA012 gp140 Trimers

One of the goals of this study was to determine whether there is any difference in the overall glycan profile as well as the glycosylation site occupancy when C97ZA012 gp140 trimers are produced by stable or transient transfection. Glycopeptide compositions summarized in Supporting Information Tables S1A, S1B, S2A, and S2B show that for N-linked glycosylation, glycopeptides were modified with a diverse array of glycans consisting of high-mannose, hybrid-type, complex-type glycans and an N-linked HexNAc that has been consistently observed with Envs that we analyzed.[18−21,39] Closer inspection of the data revealed that most of the sites are predominantly modified with complex-type glycans. To differentiate the glycan profile between the stably and transiently expressed C97ZA012 gp140, a bar graph was generated on the basis of the glycan profile of

each glycopeptide arranged according their sequence position. Glycan compositions of each glycopeptide were broadly sorted into two distinct groups, high-mannose and processed (hybrid- and complex-type) glycans, based on the criteria used in our previous studies.[18] Figure 7 shows that stably and transiently expressed C97ZA012 gp140 trimers displayed very similar glycan profiles except for one site, N332. This glycosylation site is populated with a higher level of processed glycans in transiently expressed C97ZA012 gp140 trimer, whereas the same glycosylation site is equally populated with high-mannose and processed glycans in stably expressed C97ZA012 gp140 trimer.

One important aspect of the glycosylation analysis of any HIV-1 envelope protein is whether the glycosylation profile matches the glycosylation pattern known to bind to broadly neutralizing antibodies. For example, the broadly neutralizing antibodies PG9 and PG16 bind to high-mannose glycans at the N160 site.[41] Typically, monomeric and even trimeric gp140 do not show a preponderance of high-mannose glycans at these sites, as we have shown previously.[18,19] To date, it is unknown whether the trimeric, viral envelope shows a preponderance of high-mannose glycans at these sites or if the PG9 and PG16 antibodies are simply binding favorably to a glycan population that is present in moderate abundance. Additional glycosylation
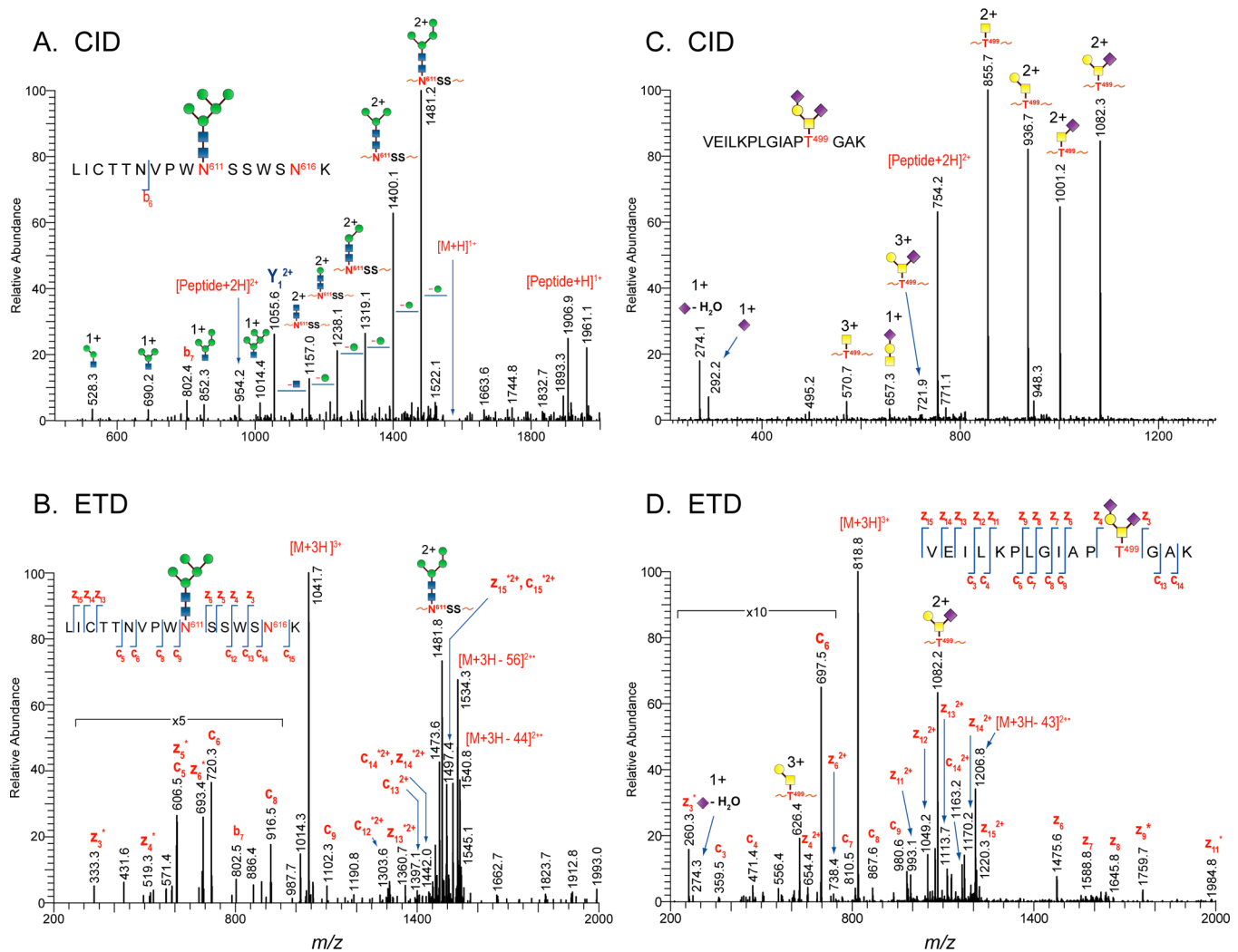
Figure 6. (A) CID data of glycopeptide with peptide portion, LICTTNVPWN^{611}SSWSN^{616}K, showing the glycan fragmentation of the high-mannose glycans. (B) ETD data of the same glycopeptide in panel A showing the peptide backbone cleavages. CID and ETD data indicate that the glycopeptide is occupied at N611. (C) CID and (D) ETD data of the *O*-linked glycopeptide, VIELKPLGIAPT^{499}GAK.

analysis on viral envelopes is required prior to determining whether a low degree of high-mannose glycans present at these sites is indicative of a protein being non-native and/or nonoptimal for eliciting these antibodies.

It should be noted that we were not able to fully characterize the glycan composition of the large tryptic glycopeptide bearing three PNG sites, N^{184}NSN^{187}NSEYILINCN^{197}ASTITQACPK, located in the V1/V2 region. These three sites are either fully occupied or one of the sites is partially occupied, based on the LC−MS and MS/MS analysis of the partially and fully deglycosylated C97ZA012 gp140 trimers. Large glycopeptides with multiple glycosylation sites usually do not ionize efficiently, especially when they are predominantly modified with complex-type glycans. These species are not usually detected, or when they are present in the spectrum, ions are usually observed with relatively low intensity. Overall, of the 23 glycosylation sites that were fully elucidated, 21 sites were modified with mostly processed glycans and two sites, N230 and N262, were exclusively modified with high-mannose glycans. Although that the glycan profiles of the three sites, N184, N187, and N197, located in the V1/V2 region were not fully elucidated, we predict that these sites are predominantly

modified with processed glycans based on our analysis of other gp140s.[18−21]

Analysis of *O*-linked glycosylation indicated the presence of *O*-linked glycans attached to T499, of the tryptic peptide, VIELKPLGIAPT^{499}GAK, near the end of the gp120 region. This *O*-linked site is partially occupied for both stably and transiently expressed C97ZA012 gp140 trimers. However, when glycosylated, only one *O*-linked glycopeptide containing disialylated core 1 type glycan was identified for stably expressed C97ZA012 gp140, and five *O*-linked glycopeptides containing both core 1 and core 2 type glycans were identified for transiently expressed C97ZA012 gp140, as shown in Table 4. All but one of the five *O*-linked glycans was sialylated. These data suggest that C97ZA012 gp140 produced from transient transfection displays more heterogeneous *O*-linked glycans compared to that from stable transfection.

■ CONCLUSIONS

We have used an integrated MS-based approach that combined CID and ETD to compare the disulfide bond topology and glycosylation profiles of C97ZA012 gp140 expressed by stable and transient transfections in 293T cells. Both transfection methods produced C97ZA012 gp140s with heterogeneous
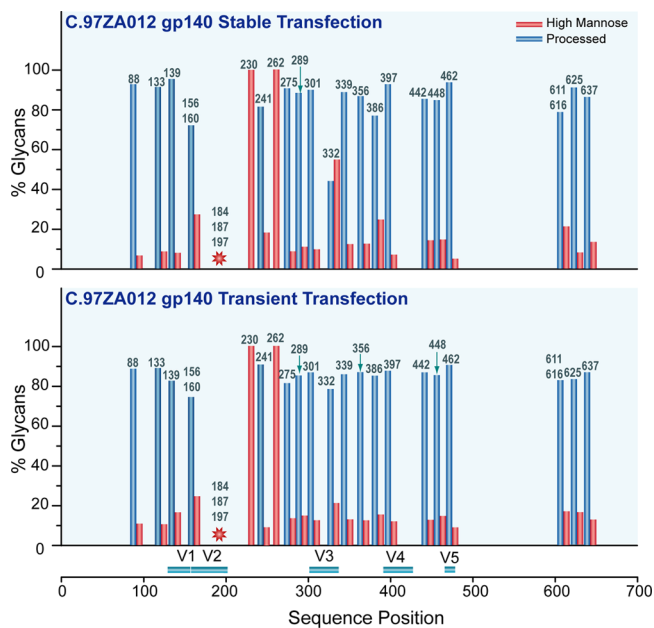
Figure 7. Bar graph showing the glycan profiles at each identified glycosylation site of the stably and transiently expressed C97ZA012 gp140. The glycan compositions (in percent) were broadly categorized into two classes: high-mannose (red bar) and processed glycans (blue bar). Red asterisk means that the glycan composition on these sites was not fully characterized.

disulfide bond patterns that included the expected disulfide bond pattern identified for the disulfide-bonded domains in the monomeric gp120 and gp41 and alternate disulfide bond patterns in disulfide-bonded domains I, II, and III. Although the disulfide bonding pattern of C97ZA012 gp140 did not vary significantly between the two different transfection methods, it is important to note that disulfide heterogeneity was present, as has been reported for similar proteins previously, and this heterogeneity appears to be a common feature of recombinantly expressed gp140.

Comparison of the glycosylation profiles revealed that C97ZA012 gp140 expressed by both transfection methods displayed similar N-glycosylation profiles and glycosylation site occupancy with very minor differences. Site-to-site comparison of the glycosylation profiles between stably and transiently

expressed C97ZA012 gp140 revealed that the glycan profile and site occupancy differ only at one site. By contrast, our results showed that C97ZA012 gp140 produced from transient transfection display more heterogeneous O-linked glycans compared to those generated from stable transfection. These data verify that only minor differences in the PTM profile of recombinant gp140s are observable when the protein expression conditions change from transient transfection to production in stable cell lines. This finding is important for vaccine developers who rely on data acquired from proteins expressed by transient transfection methods for designing vaccines that will ultimately be expressed in stable cell lines.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Tandem MS data of disulfide-linked peptides (Figures S1−S13) and complete list of the glycopeptide compositions of stably and transiently expressed C97ZA012 gp140 trimers (Tables S1A, S1B, S2A, and S2B). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*Phone: (785) 854-3015; Fax: (785) 864-5396; E-mail: hdesaire@ku.edu.

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Walker, L. M.; Burton, D. R. Rational antibody-based HIV-1 vaccine design: current approaches and future directions. *Curr. Opin. Immunol.* **2010**, *22*, 358−366.

(2) Rerks-Ngarm, S.; Pitisuttithum, P.; Nitayaphan, S.; Kaewkungwal, J.; Chiu, J.; Paris, R.; Premsri, N.; Namwat, C.; de Souza, S. M.; Adams, E.; Benenson, M.; Gurunathan, S.; Tartaglia, J.; McNeil, J. G.; Francis,

Table 4. O-Linked Glycans Observed from Stably and Transiently Expressed C97ZA012 gp140

Peptide	O-linked Glycans	Stably Expressed*	Transiently Expressed*
VIELKPLGIAPT⁴⁹⁹GAK		ND	✓
		ND	✓
	or	ND	✓
	or	✓	✓
		ND	✓

*Check mark means present; ND means not detected.

D. P.; Stablein, D.; Birx, D. L.; Chunsuttiwat, S.; Khamboonruang, C.; Thongcharoen, P.; Robb, M. L.; Michael, N. L.; Kunasol, P.; Kim, J. H. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N. Engl. J. Med.* **2009**, *361*, 2209−2220.

(3) Sevier, C. S.; Kaiser, C. A. Formation and transfer of disulphide bonds in living cells. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 836−847.

(4) Lauc, G.; Zoldos, V. Protein glycosylation—an evolutionary crossroad between genes and environment. *Mol. BioSyst.* **2010**, *6*, 2373−2379.

(5) Marino, K.; Bones, J.; Kattla, J. J.; Rudd, P. M. A systematic approach to protein glycosylation analysis: a path through the maze. *Nat. Chem. Biol.* **2010**, *6*, 713−723.

(6) Baumann, M.; Meri, S. Techniques for studying protein heterogeneity and post-translational modifications. *Expert Rev. Proteomics* **2004**, *1*, 207−217.

(7) Walsh, G.; Jefferis, R. Post-translational modifications in the context of therapeutic proteins. *Nat. Biotechnol.* **2006**, *24*, 1241−1252.

(8) Walsh, G. Post-translational modifications of protein biopharmaceuticals. *Drug Discovery Today* **2010**, *15*, 773−780.

(9) Wei, X.; Decker, J. M.; Wang, S.; Hui, H.; Kappes, J. C.; Wu, X.; Salazar-Gonzalez, J. F.; Salazar, M. G.; Kilby, J. M.; Saag, M. S.; Komarova, N. L.; Nowak, M. A.; Hahn, B. H.; Kwong, P. D.; Shaw, G. M. Antibody neutralization and escape by HIV-1. *Nature* **2003**, *422*, 307−312.

(10) Reitter, J. N.; Means, R. E.; Desrosiers, R. C. A role for carbohydrates in immune evasion in AIDS. *Nat. Med.* **1998**, *4*, 679−684.

(11) Kwong, P. D.; Doyle, M. L.; Casper, D. J.; Cicala, C.; Leavitt, S. A.; Majeed, S.; Steenbeke, T. D.; Venturi, M.; Chaiken, I.; Fung, M.; Katinger, H.; Parren, P. W.; Robinson, J.; Van, R. D.; Wang, L.; Burton, D. R.; Freire, E.; Wyatt, R.; Sodroski, J.; Hendrickson, W. A.; Arthos, J. HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* **2002**, *420*, 678−682.

(12) Scanlan, C. N.; Offer, J.; Zitzmann, N.; Dwek, R. A. Exploiting the defensive sugars of HIV-1 for drug and vaccine design. *Nature* **2007**, *446*, 1038−1045.

(13) Raska, M.; Takahashi, K.; Czernekova, L.; Zachova, K.; Hall, S.; Moldoveanu, Z.; Elliott, M. C.; Wilson, L.; Brown, R.; Jancova, D.; Barnes, S.; Vrbkova, J.; Tomana, M.; Smith, P. D.; Mestecky, J.; Renfrow, M. B.; Novak, J. Glycosylation patterns of HIV-1 gp120 depend on the type of expressing cells and affect antibody recognition. *J. Biol. Chem.* **2010**, *285*, 20860−20869.

(14) Montefiori, D. C.; Robinson, W. E., Jr.; Mitchell, W. M. Role of protein *N*-glycosylation in pathogenesis of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 9248−9252.

(15) Fischer, P. B.; Karlsson, G. B.; Butters, T. D.; Dwek, R. A.; Platt, F. M. N-Butyldeoxynojirimycin-mediated inhibition of human immunodeficiency virus entry correlates with changes in antibody recognition of the V1/V2 region of gp120. *J. Virol.* **1996**, *70*, 7143−7152.

(16) Doores, K. J.; Bonomelli, C.; Harvey, D. J.; Vasiljevic, S.; Dwek, R. A.; Burton, D. R.; Crispin, M.; Scanlan, C. N. Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 13800−13805.

(17) Bonomelli, C.; Doores, K. J.; Dunlop, D. C.; Thaney, V.; Dwek, R. A.; Burton, D. R.; Crispin, M.; Scanlan, C. N. The glycan shield of HIV is predominantly oligomannose independently of production system or viral clade. *PLoS One* **2011**, *6*, e23521.

(18) Go, E. P.; Irungu, J.; Zhang, Y.; Dalpathado, D. S.; Liao, H. X.; Sutherland, L. L.; Alam, S. M.; Haynes, B. F.; Desaire, H. Glycosylation site-specific analysis of HIV envelope proteins (JR-FL and CON-S) reveals major differences in glycosylation site occupancy, glycoform profiles, and antigenic epitopes' accessibility. *J. Proteome Res.* **2008**, *7*, 1660−1674.

(19) Go, E. P.; Chang, Q.; Liao, H. X.; Sutherland, L. L.; Alam, S. M.; Haynes, B. F.; Desaire, H. Glycosylation site-specific analysis of clade C HIV-1 envelope proteins. *J. Proteome Res.* **2009**, *8*, 4231−4242.

(20) Go, E. P.; Hewawasam, G.; Liao, H. X.; Chen, H.; Ping, L. H.; Anderson, J. A.; Hua, D. C.; Haynes, B. F.; Desaire, H. Characterization of glycosylation profiles of HIV-1 transmitted/founder envelopes by mass spectrometry. *J. Virol.* **2011**, *85*, 8270−8284.

(21) Go, E. P.; Liao, H. X.; Alam, S. M.; Hua, D.; Haynes, B. F.; Desaire, H. Characterization of host-cell line specific glycosylation profiles of early transmitted/founder HIV-1 gp120 envelope proteins. *J. Proteome Res.* **2013**, *12*, 1223−1234.

(22) Go, E. P.; Zhang, Y.; Menon, S.; Desaire, H. Analysis of the disulfide bond arrangement of the HIV-1 envelope protein CON-S gp140 delta CFI shows variability in the V1 and V2 regions. *J. Proteome Res.* **2011**, *10*, 578−591.

(23) Clark, D. F.; Go, E. P.; Desaire, H. Simple approach to assign disulfide connectivity using extracted ion chromatograms of electron transfer dissociation spectra. *Anal. Chem.* **2013**, *85*, 1192−1199.

(24) Finzi, A.; Pacheco, B.; Zeng, X.; Kwon, Y. D.; Kwong, P. D.; Sodroski, J. Conformational characterization of aberrant disulfide-linked HIV-1 gp120 dimers secreted from overexpressing cells. *J. Virol. Methods* **2010**, *168*, 155−161.

(25) Jobes, D. V.; Daoust, M.; Nguyen, V.; Padua, A.; Michele, S.; Lock, M. D.; Chen, A.; Sinangil, F.; Berman, P. W. High incidence of unusual cysteine variants in gp120 envelope proteins from early HIV type 1 infections from a Phase 3 vaccine efficacy trial. *AIDS Res. Hum. Retroviruses* **2006**, *22*, 1014−1021.

(26) Go, E. P.; Rebecchi, K. R.; Desaire, H. In-solution digestion of glycoproteins for glycopeptide-based mass analysis. *Methods Mol. Biol.* **2013**, *951*, 103−111.

(27) Leonard, C. K.; Spellman, M. W.; Riddle, L.; Harris, R. J.; Thomas, J. N.; Gregory, T. J. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J. Biol. Chem.* **1990**, *265*, 10373−10382.

(28) Hoxie, J. A. Hypothetical assignment of intrachain disulfide bonds for HIV-2 and SIV envelope glycoproteins. *AIDS Res. Hum. Retroviruses* **1991**, *7*, 495−499.

(29) Kwong, P. D.; Wyatt, R.; Robinson, J.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. A. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **1998**, *393*, 648−659.

(30) Gallaher, W. R.; Ball, J. M.; Garry, R. F.; Griffin, M. C.; Montelaro, R. C. A general model for the transmembrane proteins of Hiv and other retroviruses. *AIDS Res. Hum. Retroviruses* **1989**, *5*, 431−440.

(31) Peisajovich, S. G.; Blank, L.; Epand, R. F.; Epand, R. M.; Shai, Y. On the interaction between gp41 and membranes: the immunodominant loop stabilizes gp41 helical hairpin conformation. *J. Mol. Biol.* **2003**, *326*, 1489−1501.

(32) Frey, G.; Peng, H.; Rits-Volloch, S.; Morelli, M.; Cheng, Y.; Chen, B. A fusion-intermediate state of HIV-1 gp41 targeted by broadly neutralizing antibodies. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 3739−3744.

(33) Kovacs, J. M.; Nkolola, J. P.; Peng, H.; Cheung, A.; Perry, J.; Miller, C. A.; Seaman, M. S.; Barouch, D. H.; Chen, B. HIV-1 envelope trimer elicits more potent neutralizing antibody responses than monomeric gp120. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 12111−12116.

(34) Go, E. P.; Rebecchi, K. R.; Dalpathado, D. S.; Bandu, M. L.; Zhang, Y.; Desaire, H. GlycoPep DB: a tool for glycopeptide analysis using a "Smart Search". *Anal. Chem.* **2007**, *79*, 1708−1713.

(35) Irungu, J.; Go, E. P.; Dalpathado, D. S.; Desaire, H. Simplification of mass spectral analysis of acidic glycopeptides using GlycoPep ID. *Anal. Chem.* **2007**, *79*, 3065−3074.

(36) Nkolola, J. P.; Peng, H.; Settembre, E. C.; Freeman, M.; Grandpre, L. E.; Devoy, C.; Lynch, D. M.; La Porte, A.; Simmons, N. L.; Bradley, R.; Montefiori, D. C.; Seaman, M. S.; Chen, B.; Barouch, D. H. Breadth of neutralizing antibodies elicited by stable, homogeneous clade A and clade C HIV-1 gp140 envelope trimers in guinea pigs. *J. Virol.* **2010**, *84*, 3270−3279.

(37) Coon, J. J.; Ueberheide, B.; Syka, J. E.; Dryhurst, D. D.; Ausio, J.; Shabanowitz, J.; Hunt, D. F. Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 9463−9468.

(38) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E.; Shabanowitz, J.; Hunt, D. F. The utility of ETD mass spectrometry in proteomic analysis. *Biochim. Biophys. Acta* **2006**, *1764*, 1811−1822.

(39) Go, E. P.; Hewawasam, G. S.; Ma, B. J.; Liao, H. X.; Haynes, B. F.; Desaire, H. Methods development for analysis of partially deglycosylated proteins and application to an HIV envelope protein vaccine candidate. *Int. J. Mass Spectrom.* **2011**, *305*, 209−216.

(40) Cooper, C. A.; Gasteiger, E.; Packer, N. H. GlycoMod—a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics* **2001**, *1*, 340−349.

(41) Pancera, M.; Shahzad-Ul-Hussan, S.; Doria-Rose, N. A.; McLellan, J. S.; Bailer, R. T.; Dai, K.; Loesgen, S.; Louder, M. K.; Staupe, R. P.; Yang, Y.; Zhang, B.; Parks, R.; Eudailey, J.; Lloyd, K. E.; Blinn, J.; Alam, S. M.; Haynes, B. F.; Amin, M. N.; Wang, L. X.; Burton, D. R.; Koff, W. C.; Nabel, G. J.; Mascola, J. R.; Bewley, C. A.; Kwong, P. D. Structural basis for diverse *N*-glycan recognition by HIV-1-neutralizing V1−V2-directed antibody PG16. *Nat. Struct. Mol. Biol.* **2013**, *20*, 804−813.