9032 *Biochemistry* **2010,** *49,* 9032–9045
DOI: 10.1021/bi1011332

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Structural Analysis of a Highly Glycosylated and Unliganded gp120-Based Antigen Using Mass Spectrometry[†]

Liwen Wang,[‡] Yali Qin,[§] Serguei Ilchenko,[‡] Jen Bohon,[‡,∥] Wuxian Shi,[‡,∥] Michael W. Cho,[§] Keiji Takamoto,[⊥] and Mark R. Chance*[,‡,∥]

[‡]*Center for Proteomics and Bioinformatics, School of Medicine, Case Western Reserve University, Cleveland, Ohio 44106,* [§]*Department of Biomedical Sciences, College of Veterinary Medicine, Iowa State University, Ames, Iowa 50011,* [∥]*Center for Synchrotron Biosciences, Brookhaven National Laboratory, Upton, New York 11973, and* [⊥]*Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York 10461*

Received July 16, 2010; Revised Manuscript Received September 3, 2010

ABSTRACT: Structural characterization of the HIV-1 envelope protein gp120 is very important for providing an understanding of the protein's immunogenicity and its binding to cell receptors. So far, the crystallographic structure of gp120 with an intact V3 loop (in the absence of a CD4 coreceptor or antibody) has not been determined. The third variable region (V3) of the gp120 is immunodominant and contains glycosylation signatures that are essential for coreceptor binding and entry of the virus into T-cells. In this study, we characterized the structure of the outer domain of gp120 with an intact V3 loop (gp120-OD8) purified from *Drosophila* S2 cells utilizing mass spectrometry-based approaches. We mapped the glycosylation sites and calculated the glycosylation occupancy of gp120-OD8; 11 sites from 15 glycosylation motifs were determined as having high-mannose or hybrid glycosylation structures. The specific glycan moieties of nine glycosylation sites from eight unique glycopeptides were determined by a combination of ECD and CID MS approaches. Hydroxyl radical-mediated protein footprinting coupled with mass spectrometry analysis was employed to provide detailed information about protein structure of gp120-OD8 by directly identifying accessible and hydroxyl radical-reactive side chain residues. Comparison of gp120-OD8 experimental footprinting data with a homology model derived from the ligated CD4−gp120-OD8 crystal structure revealed a flexible V3 loop structure in which the V3 tip may provide contacts with the rest of the protein while residues in the V3 base remain solvent accessible. In addition, the data illustrate interactions between specific sugar moieties and amino acid side chains potentially important to the gp120-OD8 structure.

HIV-1[1] (human immunodeficiency virus) envelope protein gp120 can recognize the CD4 cell receptor and initiate the entry of the virus into cells. Epitopes from gp120 can generate neutralizing antibodies such that antigens derived from gp120 have been examined as potential candidates for the development of an HIV-1 vaccine. Structural characterization of gp120 is very important in understanding its immunogenicity and antigenicity. However, crystallographic structures for free glycosylated gp120 that include all its variable loops have been extremely difficult to achieve. Removal and/or restriction of flexible parts of proteins (including envelope glycoproteins) can enhance the overall probability of crystallization (*1*, *2*). Therefore, several strategies have been used to facilitate crystallization of human and simian gp120, including binding to CD4 or gp120 specific antibodies (*2*−*5*), deglycosylation of gp120 (*2*, *3*), or truncation of gp120 (v1-v2 or v3) loops (*4*−*6*). However, the extensive glycosylation (∼50%) and the variable polypeptide chain loops of gp120 are critical factors for defining molecular recognition in the immune response (*7*); thus, it is very important to obtain structural information relevant to glycosylated HIV-1 gp120 protein forms than include variable loops of interest.

Glycosylation is a very common post-translational modification in higher eukaryotes, and glycoproteins usually include multiple glycoforms carrying between one and dozens of different glycans with varying degrees of site occupancy. Glycosylation linked to asparagine residues in the consensus motif Asn-X-S/T (X can be any amino acid except proline) via an *N*-acetylglucosamine (GlcNac) linkage are called N-linked glycans. N-Linked glycans are distributed among three subtypes on the basis of their nature and location. High-mannose (oligomannose) structures are composed of mannose residues attached to the core structure with a $Man_{5-9}GlcNac_2$ composition (*8*). A second type contains *N*-acetyllacosamine (Galb1-3/4GlcNac) in its antenna region, while a third type is represented by hybrid structures that are composed of mannose, fucose, and *N*-acetyllacosamine attached to a trimannosyl chitobiose core. The first two categories are the most common structures for *N*-glycans. Glycosylated envelope glycoproteins are mainly composed of N-linked high-mannose type oligosaccharides, which control the proper folding and conformational stability of the protein (*9*). Precise characterization of sugar occupancy via glycoproteomics analysis is challenging because of the complex structures, labile nature, and microheterogeneity inherent in glycoproteins. A comprehensive analysis

[†]The study was funded by the National Institutes of Health (P01-AI-074286).

*To whom correspondence should be addressed: Case Western Reserve University, Center for Proteomics and Bioinformatics, Department of Physiology and Biophysics, Cleveland, OH 44106. Phone: (216) 368-4406. Fax: (216) 368-3812. E-mail: mark.chance@case.edu.

[1]Abbreviations: HIV-1, human immunodeficiency virus; V3, third variable region; OD, outer domain protein; GlcNac, *N*-acetylglucosamine; PNGase F, peptide *N*-glycosidase F; Endo H, endo-*β*-*N*-acetylglucosaminidase H.

 Published on Web 09/09/2010

typically involves three key tasks: identification of peptides and/or proteins, glycosylation site mapping, and evaluation of glycosylation types. Mass spectrometry is a powerful and efficient technique for completing all three tasks. Glycoprotein identification and site localization are simplified by enzymatic removal of glycans, such as peptide *N*-glycosidase F (PNGase F) deglycosylation, followed by mass spectrometry (*10*−*16*), while stable isotopic labeling can be combined with enzymatic treatment to quantitate the glycosylation occupancy (*14*). Endo-β-*N*-acetylglucosaminidase H (Endo H) treatment before mass spectrometry analysis can be used to evaluate the glycan types (*17*, *18*) and quantitate high-mannose glycosylation occupancy (*15*) as high-mannose and hybrid structures are sensitive to Endo H treatment. Because of the labile nature of glycans, peptide identification can be challenging using traditional methods of fragmentation, such as collision-induced dissociation (CID). Therefore, alternative dissociation techniques, electron capture dissociation (ECD) (*19*, *20*) or electron transfer dissociation in conjunction with CID (*21*), have been used to provide complementary information about both peptide sequence and glycan structure.

To develop approaches for the structural analysis of gp120 constructs intended as antigens for evaluating immune responses, we characterized the glycosylation of a recombinant construct composed of the outer domain of gp120, including amino acid residues 260−485 (gp120-OD8), a region that contains the structurally important V3 loop using mass spectrometry-based approaches. We used PNGase F and Endo H to deglycosylate the protein. CID and ECD MS were combined to identify peptide sequences generated by digestion using specific proteases as well as glycan structures of gp120-OD8. A relative quantitation method combining Endo H, PNGase F treatment, and LC−MS/MS was employed to estimate the degree of occupancy at each glycosylation site for high-mannose and/or hybrid glycans of gp120-OD8 (*15*).

In addition, we employed synchrotron protein footprinting, a mass spectrometry-based protein structure analysis technique, to provide detailed information about the solvent accessibility of the side chains in the gp120-OD8 structure. Hydroxyl radicals generated from radiolysis of water using millisecond pulses of a synchrotron X-ray beam directly attack accessible and reactive side chain residues of proteins in solution (*22*−*27*). After synchrotron radiolysis, oxidized side chain products can be quantified and detected by LC−MS/MS, providing measures of surface accessibility for specific side chains in the protein structure. This technique has been successfully applied to structural elucidation of proteins and protein interactions (*22*−*30*). Apparent differences in surface accessibility between this construct and surface accessibility predicted from a homology model generated from a gp120 crystal structure bound to CD4 were used to define ligand-dependent variations in structure in the V3 loop region.

EXPERIMENTAL PROCEDURES

Construction, Expression, and Purification of gp120-OD8. A gene segment encoding the outer domain of gp120 (amino acids 260−485) was PCR-amplified from the codon-optimized M-group consensus envelope sequence (MCON6) (*31*). The amplified DNA fragment was cloned into the pMT/Bip/V5-his vector (Invitrogen, Carlsbad, CA) between the BglII and PmeI sites to yield pMT/Bip/gp120-OD8. The recombinant protein contains the Bip signal sequence at the N-terminus to allow secretion and a six-His tag at the C-terminus to facilitate purification.

The S2 cell line stably expressing gp120-OD8 protein was generated according to the manufacturer's protocol. Briefly, exponentially growing S2 cells were cotransfected with the pMT/Bip/gp120-OD8 and pCoBlast plasmids using the calcium phosphate method. Transfected cells were maintained at 27 °C in Schneider's *Drosophila* medium (Invitrogen) supplemented with 10% fetal bovine serum (Hyclone, Logan, UT), penicillin and streptomycin (50 units/mL and 50 μg/mL, respectively), and blasticidin (25 μg/mL) for selection. Protein expression was induced using 10 mM cadmium chloride ($CdCl_2$).

Expressed gp120-OD8 secreted into cell culture medium was purified using tandem affinity chromatography. First, gp120-OD8 was enriched from the culture supernatant using Con A Sepharose (GE Health, Inc.). gp120-OD8 was eluted from the Con A Sepharose column using 500 mM methyl α-D-mannopyranoside, Tris-HCl (pH 7.4), and 500 mM NaCl. gp120-OD8 was further purified using a Ni-NTA column (Qiagen, Inc., Valencia, CA). After the column had been washed with 20 mM Tris-HCl (pH 8.0), 500 mM NaCl, and 5 mM imidazole, gp120-OD8 was eluted using the same buffer containing 250 mM imidazole. The eluted protein was dialyzed in Tris-HCl (pH 8.0) and 50 mM NaCl. Finally, the protein was loaded into a Q-Sepharose column. The protein, which was in the flow-through fraction, was concentrated with an Amicon Ultra concentrator (Millipore, Billerica, MA) and was stored frozen at −80 °C.

Evaluation of Interactions between gp120-OD8 and IgG1 b12 by Surface Plasmon Resonance (*SPR*). SPR analyses were performed with Biacore 3000 (Biacore, Columbia, MD) at room temperature in HBS-P running buffer [0.01 M HEPES (pH 7.4), 0.15 M NaCl, and 0.005% (v/v) surfactant P20]. IgG1 b12 was covalently bound to the CM5 sensor chip (Biacore) via carboxyl moieties on the dextran by the standard primary amine coupling method, and the resonance signal reached ∼1300 resonance units (RUs). IgG1 b12 was diluted in 10 mM glycine-HCl buffer (pH 3.0) to a final concentration of 10 μg/mL. We prepared a reference surface by activating and blocking a flow cell in the absence of IgG1 b12. For kinetic measurements of gp120-OD8 binding to immobilized IgG1 b12, sensorgrams were obtained by passing various concentrations of gp120-OD8 (1−100 nM) over the sensor surface at a flow rate of 10 μL/min using a 2 min association phase and a 2 min dissociation phase. The profiles of specific binding of gp120-OD8 to immobilized IgG1 b12 were obtained after subtraction of the response signal from the reference flow cell. Binding kinetics of each protein was evaluated using BiaEvaluation (Biacore) based on 1:1 Langmuir binding model.

Deglycosylation. Complete deglycosylation of gp120-OD8 was achieved using PNGase F (New England Biolabs, Ipswich, MA) treatment, while Endo H (New England Biolabs) treatment yields partially deglycosylated protein. Briefly, 1−2 μL of 1 μg/μL gp120-OD8, 1 μL of glycoprotein denaturing buffer (5% SDS and 0.4 M DTT), and HPLC water were mixed to yield a total of 10 μL of reaction buffer. The gp120-OD8 reaction buffer was heated at 95 °C for 10 min followed by cooling on ice. After the addition of 2 μL of 0.5 M sodium phosphate (pH 7.5), 2 μL of 10% NP-40, 1 μL of PNGase F, and 5 μL of HPLC water, the gp120-OD8 reaction buffer was then subjected to a deglycosylation reaction at 37 °C for 3 h. Similarly, gp120-OD8 was denatured with glycoprotein denaturing buffer and heated at 95 °C for the partial deglycosylation reaction using Endo H.

After the addition of 2 μL of 0.5 M sodium citrate (pH 5.5), 1 μL of Endo H, and 7 μL of HPLC water, the gp120-OD8 reaction buffer was then subjected to partial deglycosylation at 37 °C for 3 h. Prior to proteolysis, deglycosylated gp120-OD8 was purified by TCA precipitation to remove detergent, sugars, and small molecules; 12.5 μL of the TCA solution was added to 20 μL of gp120-OD8 reaction buffer, and the mixture was incubated on ice for 30 min followed by two washes with 200 μL of ice-cold acetones. Deglycosylated gp120-OD8 was then vacuum-dried and reconstituted in 10 μL of 100 mM ammonium biocarbonate buffer (pH 7.5).

Determination of Deglycosylation Efficiency by Electrophoresis. After deglycosylation by Endo H and PNGase F, gel electrophoresis was used to monitor overall deglycosylation efficiency; 100 ng each of gp120-OD8, Endo H-treated gp120-OD8, and PNGase F-treated gp120-OD8 were loaded onto a 4 to 20% SDS−PAGE gradient gel (Invitrogen). After electrophoresis, the gel was fixed with 50% methanol and 10% acetic acid for 10 min, washed with distilled water for 1 h, rocked in dithiothreitol (DTT, 32.5 μL of 0.1 M DTT in 100 mL of water) for 10 min, stained with silver nitrate for 15 min, and distained in developer (6 g of sodium carbonate and 100 μL of formaldehyde in 200 mL of water) for 5−10 min, which was quenched by addition of 10 mL of 2.3 M citric acid.

Proteolysis of Glycoproteins. gp120-OD8 and deglycosylated gp120-OD8 were subjected to proteolysis using sequencing-grade modified trypsin (Promega Biosciences, Madison, WI), Asp-N, and GluC (Roche Diagnostics, Penzberg, Germany). Endo H-deglycosylated gp120-OD8 was digested by three different enzymes: trypsin, GluC, and chymotrypsin (Roche Diagnostics). Before digestion, 1−2 μg gp120-OD8 was denatured by adding 10 μL of 50% ACN and heating at 95 °C for 10 min. Samples were reduced in 2.5 mM DTT at 56 °C for 15 min and subsequently alkylated in 10 mM iodoacetamide in the dark at 37 °C for 30 min. DTT concentration was adjusted to 11 mM to reduce excessive iodoacetamide at 37 °C for 30 min gp120-OD8 or deglycosylated gp120-OD8 samples were then diluted 10 times by 100 mM ammonium biocarbonate and subjected to proteolysis at a protease-to-protein ratio of 1:20 (w:w) at 37 °C for 12 h. Digestions where GluC and Trypsin were used in conjunction, the protein was initially digested with GluC for 2 h at room temperature with gentle shaking. Next, GluC was deactivated by heating the sample to 95 °C for 10 min and cooled on ice. Finally, Trypsin was added to the sample for overnight digestion at 37 °C. Digestion was terminated by addition of 0.1% (v: v) trifloroacetic acid (TFA). The digests were vacuum-dried and desalted by C18 Ziptip before mass spectrometry analysis.

Synchrotron X-ray Radiolysis. Synchrotron X-ray radiolysis was performed at beamline X28C of the National Synchrotron Light Source at the Brookhaven National Laboratory (Upton, NY). Prior to radiolysis, protein samples were dialyzed against 10 mM sodium cacodylate buffer (Electron Microscopy Science, Hatfield, PA) at pH 7.4 and 4 °C. Exposure conditions were optimized by following the dose-dependent degradation of the fluorescent compound Alexa 488 (Invitrogen, Carlsbad, CA) in the presence of sample buffer (*26*). An initial set of gp120-OD8 samples (3.5 μM, 5 μL volume) was exposed to mirror-focused X-rays (*32*) by use of a multiple-sample holder for 0−30 ms (0, 10, 20, and 30) at a beam current of 276 mA. To access shorter exposure times, all subsequent samples (1.7 μM, 50 μL injection volume) were exposed using the KinTek (Austin, TX) stopped-flow apparatus. Two additional sets of gp120-OD8 were then exposed to X-rays for 0−20 ms (0, 5, 10, 15, and 20 ms) at beam currents of 281 and 258 mA. After radiolysis, Met-NH_2 (pH 7.0) was immediately added to a final concentration of 10 mM to prevent secondary oxidation (*24*). The samples were then flash-frozen in liquid nitrogen and stored at −80 °C until they were digested.

Liquid Chromatography−Tandem Mass Spectrometry (LC−MS/MS). MS experiments were conducted using an LTQ-FT mass spectrometer (Thermo Finnigan, San Jose, CA). Nano-reverse phase liquid chromatography (RPLC) separations were performed on a Dionex Ultimate U3000 HPLC system (Dionex, Sunnyvale, CA) with a 5 cm × 75 μm Pico Frit C18 column (New Objectives, Woburn MA) directly connected to a New Objectives nanospray emitter (10 μm). Chromatography was performed using mobile phases A (0.1% formic acid in water) and B (80% acetonitrile and 0.04% formic acid in water) with a linear gradient of 1%/min, starting with 100% solution A at a flow rate of 0.3 μL/min. All data were acquired in positive ion mode. CID (collision-induced dissociation) and/or ECD (electron capture dissociation) was used to fragment peptides in FT-MS. In these experiments, full MS scans (m/z 300−2000) were followed by eight subsequent MS^2 scans of the top eight most abundant peptide ions using a normalized collision energy of 35%. In CID or ECD mode, full MS spectra were acquired from the ICR cell at a resolution of 100000. In ECD and CID mode, full MS spectra were acquired from the ICR cell at a resolution of 25000 to reduce the time required per scan. For each precursor ion, MS/MS spectra were acquired by both CID and ECD scans. MS/MS data-dependent CID scans were obtained from the LTQ, while the ECD spectra were produced in the ICR cell at an electron energy of 1.7−2.7 eV.

Mass Spectrometry Data Search. Tandem MS data were analyzed by use of MassMatrix (*33*, *34*) and Mascot (*35*). gp120-OD8 tandem MS data were searched against a protein database containing the consensus sequence, MCON6, for gp120-OD8, sequences of human keratins, proteolytic enzymes used such as trypsin or GluC, and decoy sequences comprised of reversed sequences of all the proteins listed in the database. Carbamidomethylation of cysteine was included as a fixed modification. For glycosylation site identification, acetylation of the N-terminus and lysine, deamidation of asparagine and glutamine (after PNGase F treatment), and glycosylation that has a 203 Da mass shift (after Endo H treatment) were included as variable modifications. For gp120-OD8 footprinting, acetylation of the N-terminus and lysine, deamidation of asparagine and glutamine, and all possible oxidation modifications (*25*) were included as variable modifications. The enzyme was defined to cleave the protein after lysine, arginine, asparatic acid, glutamic acid, and asparagine when trypsin and GluC were used for digestion. When Asp-N was used for proteolysis, asparagine was included as one of the cleavage sites. Four missed cleavages were allowed for searching the data. The mass tolerance was set to 10 ppm for the precursor ion search and 1 Da for the product ion.

Calculation of Glycosylation Fraction and Oxidation Rates. The fraction of glycosylated peptides was calculated from the ratio of the chromatography area under the ion signals for the modified peptides to the sum of the glycosylated peptides and unmodified peptides. The fraction of unoxidized peptides was calculated from the ratio of the chromatography area under the ion signals for the unoxidized peptides to the sum of the unoxidized peptides and their radiolytic products (*25*, *30*, *36*). The fraction of unmodified peptide at each time point was

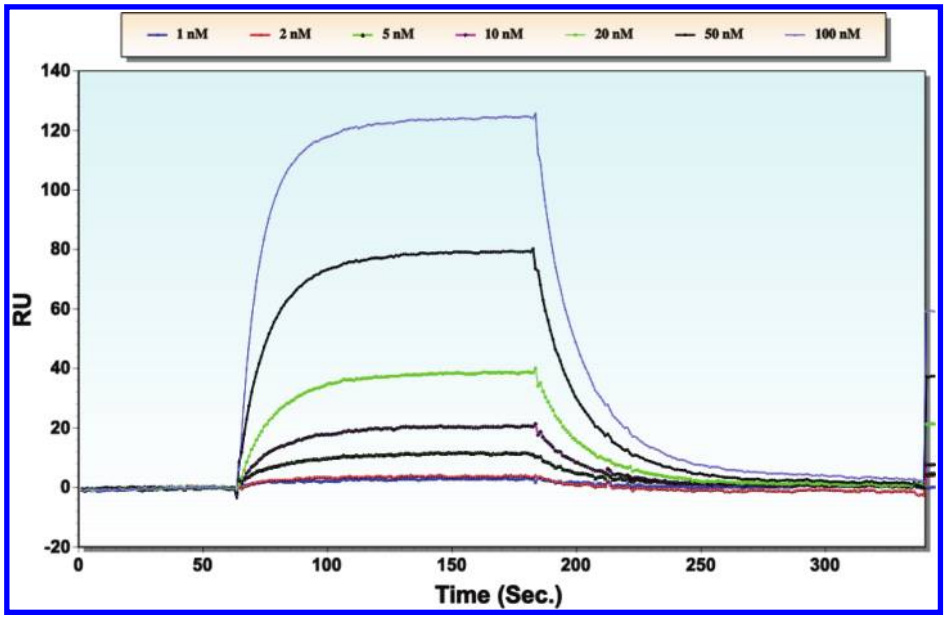

FIGURE 1: SPR kinetic analysis of binding of mAb IgG1 b12 to gp120-OD8. mAb IgG1 b12 was immobilized on a CM5 chip, and gp120-OD8 at various concentrations was passed over the chip surface.

normalized as the fraction unmodified at 0 ms is 1.0. The fraction unmodified peptide was fit to the equation $Y = Y_0e^{-kt}$. Y and Y_0 are the fractions of unmodified peptides at time t and time zero, respectively, and k is a first-order rate constant. Dose−response curves are presented as unmodified fraction versus exposure time points. The oxidation constant rate k and fitted dose−response curves were calculated with ProtMapMS (*36*) and confirmed by manual calculation.

Structure Simulation of gp120-OD8 from Homology. The homology model of gp120-OD8 was built using the Swiss-Model server (*37*−*39*), an automated comparative protein modeling server using residues 260−485 translated from the MCON6 sequence. The homology model was generated using a template of the HIV-1 JR-FL gp120 core protein containing the V3 loop in complex with human CD4 and the X5 antibody [Protein Data Bank (PDB) entry 2B4C] (*3*), which is 76.4% identical in sequence with MCON6 gp120-OD8. The ligands were removed in the modeling. The homology model generated from this template included all but the first residue.

Side Chain Solvent Accessibility Calculation. MSMS version 2.5 developed by M. F. Sanner et al. was used to calculate the solvent accessible surface areas of all side chains (square angstroms) (*40*). The homology model of gp120-OD8 (*3*) was used for this analysis.

RESULTS AND DISCUSSION

Evaluation of the Binding Efficiency of gp120-OD8 to Antibody b12. To examine interactions of gp120-OD8 with IgG1 b12, which is an antibody specific for the CD4 binding site of gp-120, we analyzed the kinetics of binding of b12 to gp120-OD8 using SPR. IgG1 b12 was immobilized on a CM5 chip, and gp120-OD8 solutions at various concentrations were allowed to flow over the surface. The kinetic curves (Figure 1) of gp120-OD8 were relatively homogeneous, indicating a single type of binding site. The equilibrium dissociation constant (K_D) of gp120-OD8 inferred from these data was 113 nM, which is ~3.9 times lower than that (438 nM) of an outer domain protein of gp120 (OD1, amino acids 252−482) (*41*). The data for binding of gp120-OD8 to the b12 antibody indicate an intact and nativelike tertiary

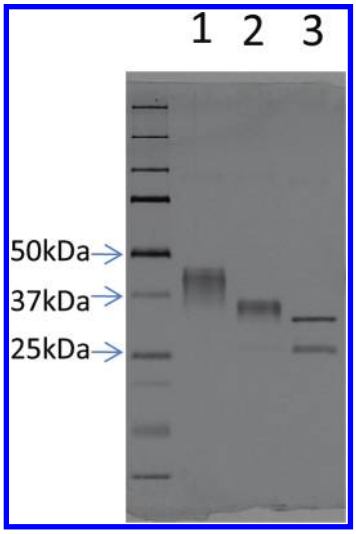

FIGURE 2: SDS−PAGE gel (4 to 20%) with gp120-OD8 (lane 1), Endo H-treated gp120-OD8 (lane 2), and PNGase F-treated gp120-OD8 (bottom band in lane 3). The top band in lane 3 is PNGase F (M_w = 36 kDa).

structure, at least with respect to b12 binding. The observation that OD1 was "unable" to bind to CD4 but can bind to dodecameric CD4 while the gp120 core can bind to CD4 suggested that binding of CD4 to gp120 needs the avidity of the cell wall or the inner domain to enhance the stability of the outer domain gp120−CD4 complex (*41*). Therefore, CD4 binding was not used to test the native structure of gp120-OD8 here.

Identification of N-Linked Glycans of gp120-OD8. gp120-OD8 was treated with PNGase F and Endo H to examine the glycosylation of the purified protein. Deglycosylation efficiency was examined using gel electrophoresis (Figure 2). The apparent molecular mass of gp120-OD8 ranges from 48 to 35 kDa (lane 1), which results from the heterogeneous glycoforms of gp120-OD8, while the apparent molecular mass of gp120-OD8 decreased to 30 kDa (lane 2) after partial deglycosylation via Endo H treatment. After complete deglycosylation by PNGase F treatment, the apparent molecular mass of gp120-OD8 decreased

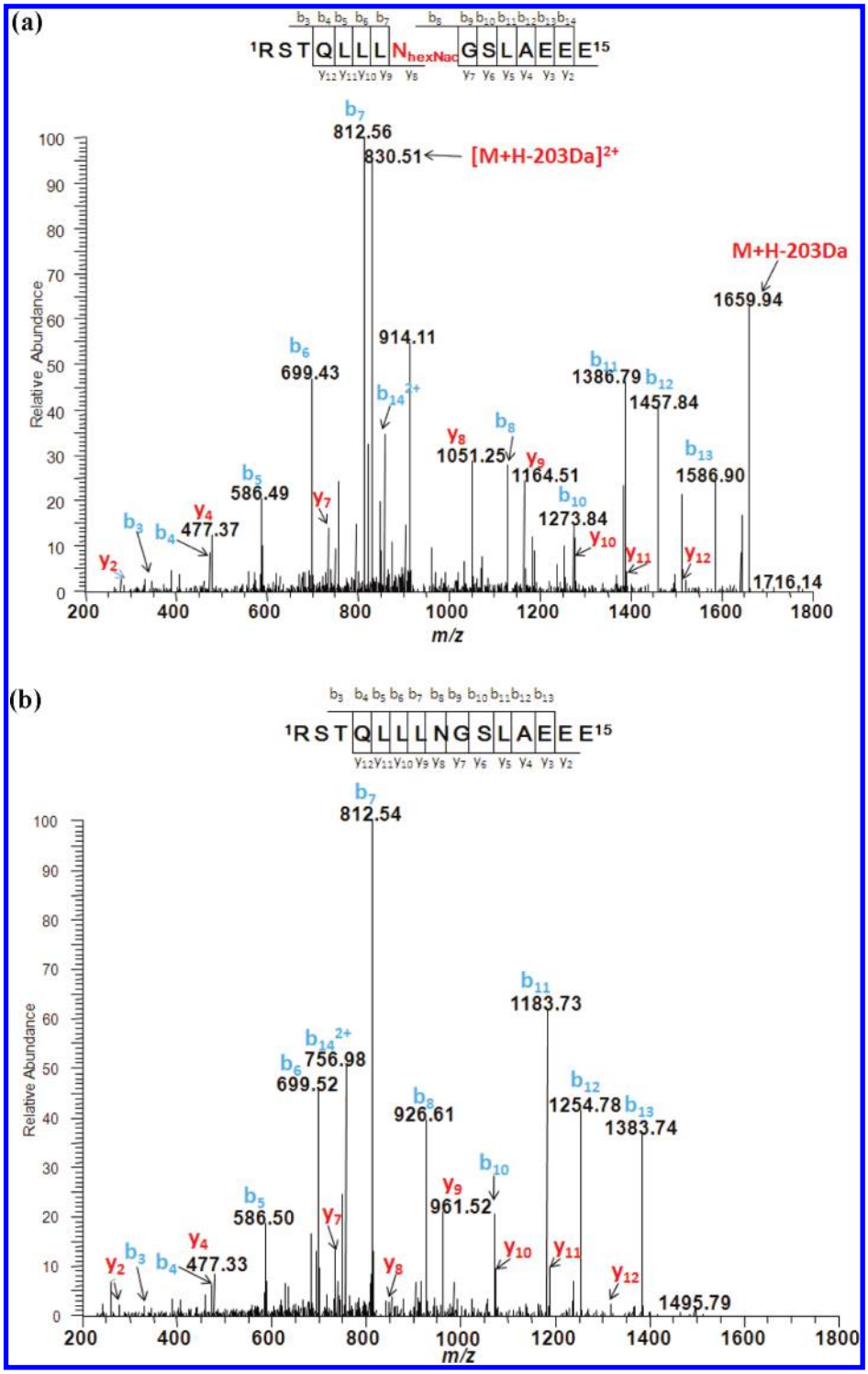

FIGURE 3: Identification of glycosylation sites using Endo H treatment. (a) MS/MS spectrum of the glycopeptide precursor ion at m/z 931.9750 Da. (b) MS/MS spectrum of the unmodified peptide of the same sequence [RSTQLLLNGSLAEEE (residues 1−15)].

to 25 kDa, reflected by the bottom band in lane 3 [the top band in lane 3 is PNGase F (M_w = 36 kDa)].

Trypsin, GluC, or Asp-N was used to digest the treated and untreated gp120-OD8. Protein digests were separated by nano-LC and detected by a high-mass accuracy FT ICR/ion-trap hybrid mass spectrometer. The MS data set was subjected to MASCOT and MassMatrix MS database searches against consensus gp120-OD8 sequence MCON6. Thirty-two unique peptides were detected from digestion with trypsin, resulting in 78% sequence coverage, while 40 unique peptides were detected from digestion with Asp-N, resulting in 87% sequence coverage. Trypsin and GluC revealed 35 unique peptides and 85% sequence coverage, with the additional identification of peptides 102−114, 115−125, 112−130, 126−136, and 137−151 due to digestion by GluC. The total sequence coverage using all the digestion methods is 100%.

After PNGase F treatment, the glycosylasparagine residues in the protein sequence lose their sugar moieties and the side chain is converted to aspartate (*42*, *43*). Therefore, the mass difference between asparagine and aspartate (Δm = 0.9847 Da) was used to

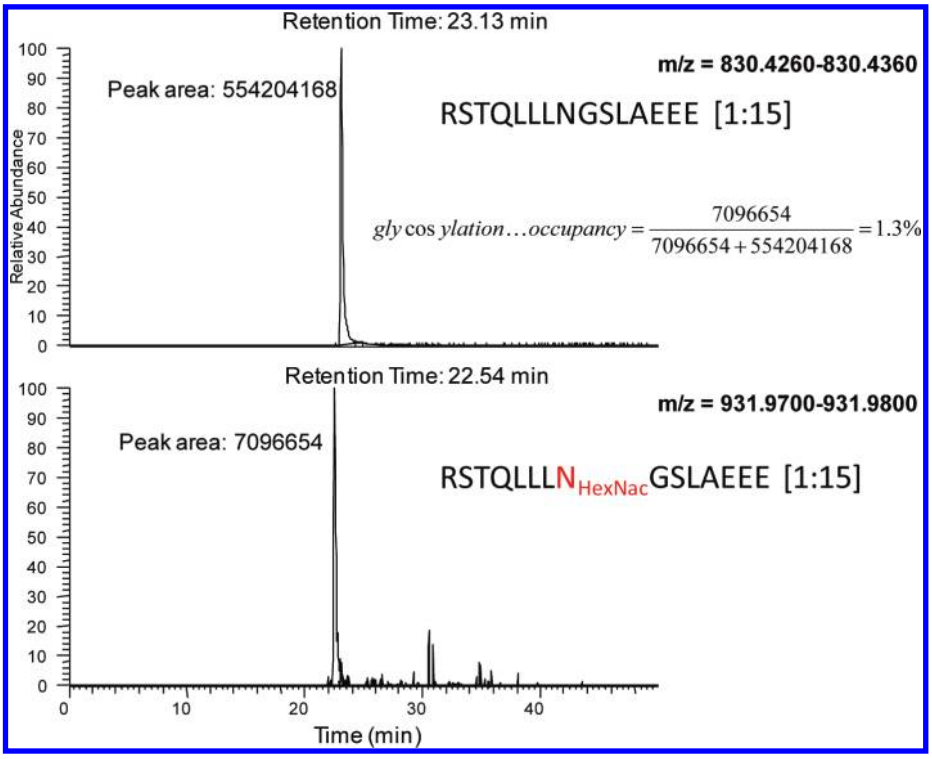

FIGURE 4: Quantiation of the glycosylation occupany of Asn 8 in peptide 1−15 (RSTQLLLNGSLAEEE) by use of Endo H treatment. The top panel shows the chromtography peak of unmodified peptide 1−15. The peak area was calculated as m/z values ranging from 830.4260 to 830.4360. The bottom panel shows the chromtography peak of glycopeptide 1−15. The peak area was calculated as m/z values ranging from 931.9700 to 931.9800.

identify modification of asparagine in the database search and to localize the glycosylation sites. Twenty unique Asp-N digests from gp120-OD8 were identified as possible glycosylated peptides after PNGase F treatment. It was determined that all 15 asparagines with a glycosylation motif were converted to aspartate. However, deamidation of asparagine in vivo and in vitro results in the loss of the amino ammonia group and formation of aspartate; thus, deamidation can result in false positive identification of a glycosylation site (*44*−*46*). Five asparagines with a glycosylation motif were determined to be deamidated before PNGase F treatment based on analysis of untreated gp120-OD8. The nonenzymatic deamidation rate of Asn and Gln is dependent on the peptide sequence, buffer type, pH, temperature, and ionic strength (*46*). Peptide sequences containing an "AsnGly" were found to be especially unstable (*47*). In this study, peptide 1−15 (RSTQLLLNGSLAEEE) containing "NG" sequence in the untreated sample exhibited a deamidation ratio of >70%. Therefore, identification of the glycosylation sites and use of PNGase F treatment are used as only one point of reference in the analysis of MS data of glycosylated gp120-OD8.

Endo H cleaves the chitobiose core of high-mannose sugars and some hybrid oligosaccharides from N-linked glycoproteins (*43*). Because the recombinant proteins expressed in insect cells mainly contain high-mannose sugars, removal of glycosylation using Endo H is reasonable (*48*−*50*). An *N*-acetylglycosamine group remains in the asparagine after Endo H treatment, which results in a 203.0794 Da shift in the mass spectrum compared to the mass of asparagine. Therefore, 203.0794 Da was used as an indicator of asparagine modification after Endo H treatment. Glycosylasparagine sites were localized by matching mass spectra of unmodified and modified peptides unambiguously as shown in Figure 3a,b and Figure 1 of the Supporting Information. The mass spectra of modified and unmodified peptides showed similar y ion and b ion distribution patterns with corresponding mass shifts resulting from the modification in the spectra of modified peptides. Eleven glycosylated sites were identified with a mass shift of 203.0794 Da as listed in Table 2. Neutral loss peaks of 203 Da from precursor ions in the MS/MS spectrum were observed (Figure 3a). The intensity of a neutral loss peak is usually greater than that of the product ion that contains glycans because the labile glycosidic bonds are more easily fragmented than the peptide backbone in mass spectrometry (Figure 3a). Because of inefficient cleavage caused by the glycan groups remaining after Endo H treatment, some long peptides were observed with two missed cleavages. To obtain better sequence coverage and more accurate identification, we used four different digestion methods (trypsin, chymotrypsin, GluC-trypsin, and GluC). Multiple charge states of mass spectra were detected from GluC protelytic digests with several lysine residues in the sequence. The occupancies of high-mannose or hybrid glycosylation were assessed on the basis of the chromatographic area under the MS signal of modified species and unmodified species. Quantitative glycosylation mapping was generated by calculating the fraction of the chromatographic peak areas of modified peptides versus total peptides (*15*) as shown in Figure 4. In summary, 11 glycosylation sites of 15 predicted (possible) sites were identified. Replicates or triplicates were analyzed by mass spectrometry. The data were highly reproducible. The glycosylation occupancy of different sites in gp120-OD8 varies from 0.8 to 99.2% as listed in Table 1 and Figure 5. In some cases, we were unable to deconvolute the occupancy of the individual sites from Endo H-treated samples. For example, the total glycosylation occupancies of two asparagine sites, Asn 77 and Asn 84 on peptide 67−96, were 77.2%. Therefore, data for PNGase F-treated gp120-OD8 were used to isolate the occupancy of the individual sites. As discussed in the previous paragraph, in vivo or in vitro

Table 1: Overview of Glycopeptides Containing High-Mannose or Hybrid Structure Identified by Endo H and PNGase F Treatment and Mass Spectrometry

peptide sequence (residue range)	glycosylation sites	occupancy
	Endo H Deglycosylation	
RSTQLLLNGSLAEEE (1−15)	N8	0.8%, 0.7%, 1.3%
IIIRSENITNNAKTIIVQLNE (16−36)	N22	98.3%, 100%
IIIRSENITNNAKTIIVQLNE (16−36)	N22 and N35	18.8%, 23.1%
TIIVQLNE (29−36)	N35	23.5%, 21.0%
SVEINCTRPNNNTR (37−50)	N41	97.8%
SVEINCTRPNNNTR (37−50)	N41 and N47	10.5%
SVEINCTRPNNNTRKSIHIGPGQAFYATGE (37−66)	N41 or N47	~100%
SVEINCTRPNNNTRKSIHIGPGQAFYATGE (37−66)	N41 and N47	10.8%
IIGDIRQAHCNISRTKWNKTLQQVAKKLRE (67−96)	N77 or N84	77.2%, 74.4%
HFNNKTIIFKPSSGGDLE (97−114)	N100	15.3%, 12.7%
FYCNTSGLF (127−135)	N130	94%, 100%
SNITGLLLTR (190−199)	N191	5.2%, 5.6%, 6.8%
DGGNNSNKNKTETFRPGGGDMR (200−221)	N208	67.7%
GKITCKSNITGLLLTRDGGNNSNKNKTE (184−211)	N208	75%, 61.2%
	PNGase F Deglycosylation	
SVEINCTRPNNNTR (37−50)	N41 and N47	10.1%, 17.2%
QAHCNISR (73−80)	N77	42.1%, 55.5%
GTYMFNGTR (143−151)	N148	92%, 100%
STWMFNGTYMFNGTR (137−151)	N142 and N148	82.7%
SNITGLLLTR (190−199)	N191	39.0%, 43.3%

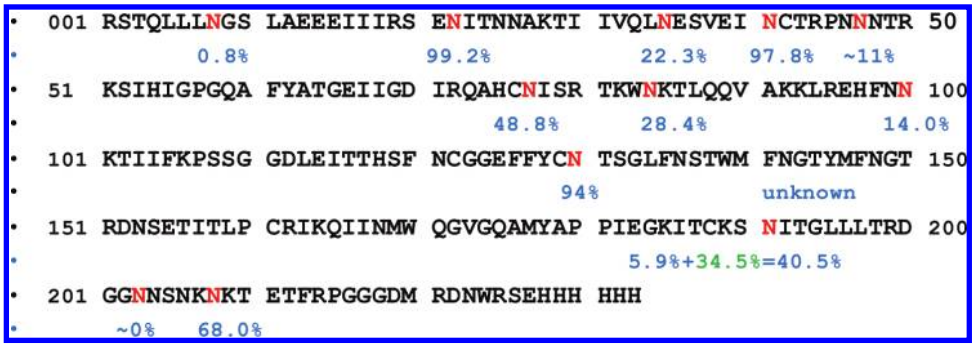

FIGURE 5: Quantitative glycosylation occupancy of gp120-OD8. The level of glycosylation of 11 sites varies from 0.8 to 99.2%. N191 might contain two types of glycosylation; N136, N142, and N148 were potential glycosylation sites but not identified in Endo H treatment. No glycosylation of N203 was detected.

deamidation were detected in fully glycosylated gp120-OD8 before PNGase F treatment. Therefore, data of fully glycosylated gp120 were used to verify the accuracy of the data for PNGase F-treated gp120-OD8. Because no deamidation was observed for Asn 77 in fully glycosylated samples, the glycosylation occupancy of Asn 77 on peptide 73−80 was determined to be 48.8% based on PNGase F-treated samples. Asn 84 was deamidated in fully glycosylated gp120-OD8. Therefore, the glycosylation occupancy on Asn 84 was calculated to be 28.4% via subtraction of the glycosylation occupancy of Asn 77 (48.8% for data for PNGase F-treated gp120-OD8) from the total glycosylation occupancies of Asn 77 and Asn 84 (77.2% for data for Endo H-treated gp120-OD8) (Table 1).

Peptide 37−50 in the trypsin-digested Endo H-treated sample and peptide 37−66 in the GluC-digested Endo H-treated sample were detected. Glycosylation occupancies of both peptide 37−50 and peptide 37−66 containing one glycosylation site were determined to be greater than 97.8% (97.8 and ~100%, respectively). Glycosylation occupancies of both peptide 37−50 and peptide 37−66 containing glycosyl Asn 41 and glycosyl Asn 47 were ~11% (10.5 and 10.8%, respectively). Asn 41 was localized as the glycosylated site in glycosylated peptide 37−50 containing one glycosylation site. Therefore, Asn 41 was determined to be highly glycosylated (97.8%), and another site Asn 47 is partially glycosylated (~11%). Additionally, these data are consistent with those of PNGase F-treated samples. Peptide 190−199 was determined to be occupied with 5.9% high-mannose or hybrid glycosylation in Endo H-treated samples and 41.2% in PNGase F-treated samples. Only 0.7% deamidation was observed in fully glycosylated gp120-OD8 samples. Therefore, this site might be occupied with two different types of glycosylation: 35.3% (=41.2% − 5.9%) complex glycosylation and 5.9% high-mannose/or hybrid glycosylation. Peptides containing Asn 136, Asn 142, and Asn 148 were not detected in the Endo H-treated sample or fully glycosylated samples. Heavy glycosylation and different glycosylation types on the three Asn sites may result in the undetectability of the peptides containing these Asn sites in fully glycosylated samples and Endo H-treated samples. Among them, the Asn 142- and Asn 148-containing NG motif might undergo a high degree of deamidation in the sample preparation process. The sequence (amino acids 126−156) was not detected in fully glycosylated gp120-OD8, which makes the determination of

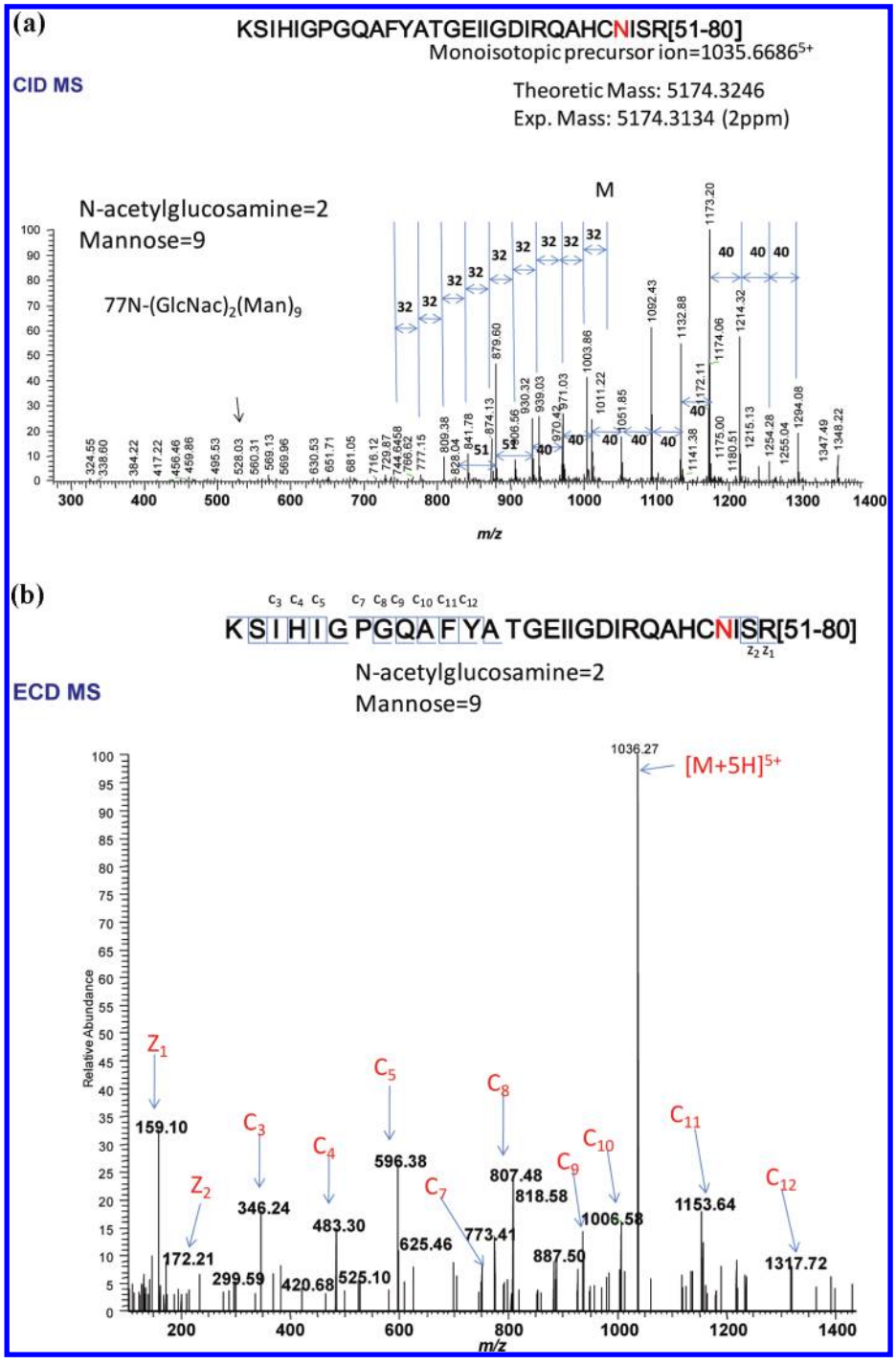

FIGURE 6: Identification of glycopeptides by a combination of CID and ECD MS. The peptide sequence was identified as [51]KSIHIGPGQAFYATGEIIGDIRQAHCNISR[80]; Asn 77 was identified as the glycosylation site containing two *N*-acetylglucosamines and nine mannoses. (a) Typical CID MS/MS spectrum of the high-mannose glycopeptides with many abundant neutral loss peaks and the 528 Da (two mannoses + one GlcNac) signature peak. (b) ECD MS/MS spectrum of the glycopeptide providing information about peptide sequence by fragmenting peptides to c ions and z ions.

occupancy for these three potential glycosylation sites difficult, although they were shown to be deamidated in PNGase F-treated samples. No glycosylation of Asn 203 having the NxS motif was identified in Endo H-treated samples. A quantitative glycosylation map based on the data and analysis described above is shown in Figure 5.

Composition Analysis of Glycosylation by ECD and CID MS. Glycopeptides fragment in a characteristic manner when subjected to CID. The labile glycosidic bonds of the carbohydrate moiety are much more easily fragmented than the peptide backbone (Figure 6a). Thus, it is difficult to obtain backbone sequence information about the glycopeptide CID MS/MS spectrum dominated by glycan neutral loss signals. Fragmentation using ECD is lower in energy and better preserves labile post-translational modifications. Therefore, ECD MS and CID MS data were used to complement each other in identifying the glycopeptides as well as interpreting the glycan moieties accurately.

We used two software programs to analyze mass spectral data of glycosyl peptides. One is software developed in house (*51*), which was used to locate glycosyl peptide mass spectra.

Table 2: List of Eight Unique Glycopeptides Containing Nine Glycosylation Sites Identified by CID MS and ECD MS

peptide sequence (residue range)	glycosylation sites	glycan moieties	theoretical M_w (Da)	observed M_w (Da)
IIIRSENITNNAK (16−28)	N22	$(GlcNac)_2(Man)_3(Fuc)_1$	2524.2132	2524.2077
TIIVQLNE (29−36)	N35	$(GlcNac)_2(Man)_3$	1821.8474	1821.8445
INCTRPNNNTR (40−50)	N41 and N47	$(GlcNac)_2(Man)_x$ and $(GlcNac)_2(Man)_{9-x}$	3766.5055	3766.4984
KSIHIGPGQAFYATGEIIGDIRQAHCNISR (51−80) with alkylation of C76	N77	$(GlcNac)_2(Man)_9$	5174.3246	5174.3134
TKWNKTLQQVAKK (81−93)	N84	$(GlcNac)_2(Man)_7$	3113.4504	3113.4368
$_{Ace}$KLREHFNNKTIIFKPSSGGDLE (93−114)	N99	$(GlcNac)_2(Man)_3(Fuc)_1$	3610.7333	3610.7422
SNITGLLLTR (190−199)	N191	$(GlcNac)_2(Man)_3(Fuc)_1$	2126.0221	2126.0184
DGGNNSNKNKTETFRPGGGDMRDNWR (200−225)	N208	$(GlcNac)_2(Man)_3(Fuc)_1$	3961.6993	3961.6908

Glycopeptide spectra generated using CID typically contain abundant neutral loss peaks, such as the loss of 146 Da for fucose, 162 Da for hexose (Hex) or mannose (Man), and 203 Da for *N*-acetylglucosamine (GlcNac). Sugar fragment masses such as 366 Da $[(GlcNac)_1(Hex)_1]$, 528 Da $[(GlcNac)_2(Hex)_1]$, and 690 Da $[(GlcNac)_3(Hex)_1]$ are also typical signatures of glycopeptide CID MS spectra. We also used ProteinProspector version 5.3.2 (*52*, *53*), a proteomics tool for mining sequence databases, in conjunction with mass spectrometry experiments. It was used to calculate theoretical gp120-OD8 proteolytic digests based on the MCON6 consensus sequence and to calculate theoretical CID and ECD MS product ions based on glycopeptide sequences. The type and number ofglycans were determined by calculating the mass and number of neutral losses of product ions in the CID mass spectra of glycosyl peptides as shown in Figure 6a and Figure 2 of the Supporting Information. The c ions and z ions in the ECD mass spectrum were used to confirm the peptide sequence as shown in Figure 6b and Figure 2 of the Supporting Information. For instance, glycopeptides with pentacharged (+5) precursor ions (m/z 1035.6686) were detected as shown in Figure 6a. Sugar moieties were identified by multiple neutral losses of glycan signals. For example, a pattern of nine continuous "32" or "40" (m/z) differences between product ions indicated that the glycopeptides lost nine hexose moieties (m/z = 162 Da/5 = 32.4 ± 1, or m/z = 162 Da/4 = 40.5 ± 1). Losses of *N*-acetylglucosamines resulted in a difference of "51" for product ions when the product ion were tetracharged (m/z = 203 Da/4 = 50.8 ± 1). It can be determined that Asp 77 first connected with two *N*-acetylglucosamines because they were fragmented (two differences of "51" in the MS spectrum) following the fragmentation of nine hexoses (nine differences of "40" in the MS spectrum), but from these data, we are not able to unambiguously determine the connections between all the glycans. The variance is 2 ppm upon comparison of the experimental mass with the theoretical mass of the glycopeptide calculated by addition of masses of nine hexoses and two *N*-acetylglucosamines to the mass of peptide 51−80, so the assignment is relatively good. This was confirmed by ECD MS (Figure 6b) as the sequence "KSIHIGP-QAFYA...NISR" (peptide 51−80) through detection of z ions and c ions from the same precursor ion. In summary, eight unique glycosyl peptides containing nine glycosylation sites were identified as listed in Table 2. The glycosylation sites listed in Table 2 were confirmed to contain high-mannose or hybrid glycosylation structure. These results are also consistent with the fact that the recombinant glycoproteins expressed in insect cells usually contain simple glycan types (high-mannose and trimannose), and a lack of complex glycosylation (*48*−*50*). This is one feature of using insect cells to produce glycoproteins; although the insect cell system is efficient for the expression of recombinant proteins with glycosylation, the occupancy may not be native.

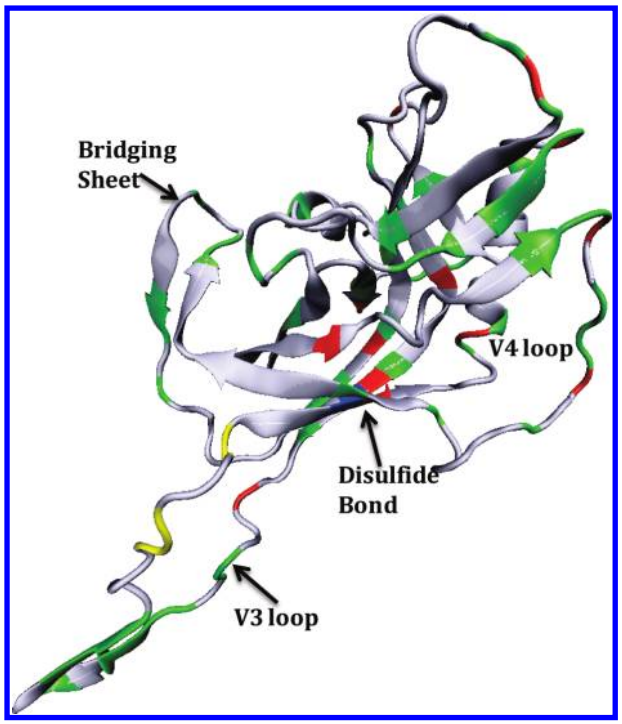

FIGURE 7: Homology model of gp120-OD8, including V3 and V4 loops and the bridging sheet. Glycosylation sites are identified experimentally by mass spectrometry. Red denotes high-mannose glycosylation sites. Green denotes oxidized residues in synchrotron radiolysis. Yellow denotes oxidized residues in N-terminus of the V3 loop in synchrotron radiolysis that should be more exposed than other residues, including those on the V3 tip according to footprinting data. Blue denotes the disulfide bond in the V3 loop.

Structural Analysis of gp120-OD8. Two regions have been considered important for gp120 coreceptor binding: (i) the V3 tip and (ii) the gp120 core around the bridging sheet, the V3 base, and neighboring residues (*54*−*56*). The entire V3 loop of gp120, which includes peptides 40−50, 51−66, 52−72, 67−72, and 73−80, is an important neutralizing antibody (Nab) target for the virus (*57*−*59*). The binding site for CD4 on the ligated gp120 structure is formed by the interface among the inner domain, bridging sheet, and outer domain (*4*). The bridging sheet of gp120 includes peptide 163−183. Another loop structure V4 region includes peptides 137−151, 143−151, and 152−162. The peptides that include these three regions are shown in Figure 7.

Purified and glycosylated gp120-OD8 was exposed to X-rays for intervals of 0, 10, 20, and 30 ms or 0, 5, 10, 15, and 20 ms, then deglycosylated with PNGase F, and then digested withy Asp-N

Table 3: Oxidation Rate Constants of gp120-OD8 Asp Digests

peptide sequence (residue range)	modified residues and their mass shifts (Da)	SASA (Å^2) in homology	rate (s^{-1})
RSTQLLLNGSLAEEEIIIRSE (1−21)	L6 (+16/+14), L11 (+16), R19 (+16)	R19 (104.64)	8.5 ± 1.2
NNTRKSIHIGPGQA-FYATGEIIG (47−69) (V3 loop)	K51, I55, F61, I67, H54 (+16)	K51 (134.9), I55 (129.9), F61 (85.8), H54 (81.7), I67 (137.8)	7.4 ± 1.7
NISRTKW (77−83)	W83 (+32)	W83 (5.1)	7.0 ± 1.4
NKTIIFKPSSGG (100−111)	K101, F104, K106 (+16)	K101 (56.7), F104 (33.3), K106 (87.0)	1.1 ± 0.06
NSTWMF (136−141) (+1)	W139 (+32/+4), M140, F141 (+16)	W139 (28.3), M140 (76.2), F141 (135.9)	29 ± 8.4
NTSGLFNSTWMF (130−141)	W139 (+32), M140 (+16), F141 (+16)	W139 (28.29), M140 (76.2), F141 (135.9)	37 ± 7.3
NGTYMF (142−147) (+1)	F141, M146 (+16)	F141 (135.9), M146 (198.0)	23 ± 3.4
NSTWMFNGTYMF (136−147) (+2)	W139 (+32), M140 (+32)	W139 (28.29), M140 (76.2)	36 ± 4.0
NGTYMFNGTR (142−151) (+2)	M146, F147, R151 (+16)	M146 (198.0), F147 (144.0), R151 (190.6)	31 ± 1.7
NITGLLLTR (191−199)	L196 or L197 (+16)	L196 (0), L197 (0)	0.36 ± 0.13
DGGNNSNKNKTETFRPGGGD (200−218)	K207/K209 (+16)	K207 (136.6), K209 (100.2)	1.5 ± 0.05
NKTETFRPGGGDMR (208−221)	P215, M220, R221 (+16), F213 (−30)	P215 (0), M220 (24.5), R221 (156.0), F213 (0)	87 ± 11

or GluC and trypsin. Deglycosylation of gp120-OD8 by PNGase F treatment was performed to make the samples easier to analyze. The peptide digests were separated by LC−MS and analyzed by FT-ICR mass spectrometry. The tandem mass spectra of oxidized peptides are shown in Figures 3 and 4 of the Supporting Information. Dose−response curves were calculated by plotting the fraction of unmodified peptides (on a log scale) as a function of exposure time as shown in Figure 5 of the Supporting Information. First-order rate constants for all oxidized peptides were derived from this analysis and are listed in Tables 3 and 4.

Forty unique peptides were detected from digestion of Asp-N (resulting in sequence coverage of >81%). The enzyme will cleave at Asp residues and Asn residues that were converted to Asp after PNGase F treatment. Twelve of the 40 unique peptides were seen to be oxidized (Table 3). Oxidative modification of these peptides mainly resulted in product peptides with 16 or 14 Da mass increases. Changes in retention times of oxidized peptides relative to those of unmodified peptides were observed via chromatography because of the change in hydrophobicity after oxidation (*23*).

The Asp-N peptides examined had rates of oxidation ranging from 0.36 ± 0.13 s^{-1} for peptide 191−199 to 87 ± 11 s^{-1} for peptide 208−221. Asp-N can cleave at the N-terminus of aspartate generated from deamidation of asparagine. Of these 12 oxidized Asp-N peptides, one was the N-terminal peptide (1−21) that contains partial glycosylated asparagines, one peptide's (200−218) N-terminal amino acid was aspartate, five peptides' (47−69, 77−83, 100−111, 130−141, and 208−221) N-termini were deamidated asparagine generated by deglycosylation, one peptide's (191−199) N-terminus was deamidated asparagine that contains both in vivo deamidation (0.7%) and deglycosylation deamidation (40.5%), and four peptides' (136−141, 142−147, 136−147, and 142−151) N-termini were deamidated asparagine, but it was not determined whether these deamidations were from deglycosylation. Thus, oxidation data of Asp-N peptides (47−69, 77−83, 100−111, 130−141, and 208−221) reflect solely the oxidation behavior of glycosylated peptides, while for the other peptides, the population of peptides is mixed.

Trypsin can cleave proteins after lysine and arginine, which can generate unglycosylated peptides as well as glycosylated peptides of the same overall sequence if glycosylation is incomplete. Thirty-five unique peptides were detected from digestion of trypsin and GluC (resulting in sequence coverage of >85%). Eighteen of the 35 GluC-tryptic peptides were seen to be oxidized. Oxidation of His 54 included formation of the 16, −22, and −23 Da species. The −23 Da oxidation product of His has not been reported in previous protein radiolysis experiments (*22*−*25*, *30*, *34*, *60*−*65*). All the oxidized GluC-tryptic peptide mass spectra are shown in Figure 4 of the Supporting Information. The oxidation rates of the 18 peptides are listed in Table 4, and representative dose−response curves are shown in Figure 5 of the Supporting Information. The overall oxidation rates for different peptides vary significantly from 0.20 ± 0.16 s^{-1} for peptide 190−199 to 59 ± 8.6 s^{-1} for peptide 52−72. Comparison of the oxidation rates of the two sets of data from Tables 3 and 4 revealed that the same sites of oxidation detected in different peptide sequences exhibited similar oxidation rates except peptide 208−221. Secondary oxidation of methionine (*25*) may cause the oxidation rate of peptide 208−221 in Asp-N digests (87 ± 11 s^{-1}) to be greater than in GluC-tryptic digests (39 ± 6.1 s^{-1}) as higher oxidation rates (164 ± 53 s^{-1}) were observed when samples were deglycosylated, digested, and detected by mass spectrometry in experiments at different times. Although 10 mM Met-NH_2 was added to suppress the secondary oxidation typically observed in sulfur-containing residues, Met-NH_2 was removed from the solution after TCA precipitation, which may have contributed to the irreproducibility of the oxidation rate for this sequence.

Correlation of Footprinting and Crystallographic Data for gp120-OD8. The order of relative reactivity of amino acid side chains under aerobic conditions in mass spectrometry-based footprinting experiments is as follows: Cys > Met > Trp > Tyr > Phe > cystine > His > Leu ~ Ile > Arg ~ Lys ~ Val > Ser ~ Thr ~ Pro > Gln ~ Glu > Asp ~ Asn > Ala > Gly (*24*). Lower-reactivity residues such as Asp, Asn, Ala, Gly, and residues whose oxidation products are difficult to detect such as Ser and Thr often do not provide much information for protein structure analysis. There are 6 Cys, 13 Arg, 4 Trp, 4 Tyr, 10 Phe, 4 His (neglecting the 6-His tag), 13 Leu, 23 Ile, 13 Lys, 4 Val, and 7 Pro residues in gp120-OD8. Additionally, we detected 2 oxidized Ala residues (of 8 Ala residues in total) by mass spectrometry after synchrotron radiolysis. To provide a framework for interpreting the observed sites of oxidation, we employed the solvent accessibility of a model of our construct, based on the structure of the core protein of gp120 with the V3 loop from clade B isolate (JR-FL) (*3*). The level of sequence identity between the consensus sequence of recombinant gp120-OD8 (MCON6) and its counterpart in the CD4−gp120 complex from clade B isolate (JP-FL) is 76.4%, thus providing an excellent template for modeling from this perspective; however, the effects

Table 4: Oxidation Rate Constants of gp120-OD8 GluC and Tryptic Digests

peptide sequence (residue range)	modified residues and their mass shifts (Da)	SASA (Å^2) in homology	rate (s^{-1})
RSTQLLLNGSLAEEE (1−15)	R1 (+16/−43), L5/L6 (+14/+16), L11 (+14/+16), E13 (−30), E14 (−30)	L5 (0), L6 (12.8), L11 (49.7), E13 (127.3), E14 (132.4)	7.7 ± 1.4
IIIRSENITNNAK (16−28)	I16 (+14/+16), I18 (+14/+16), R19 (+16/−43), I23 (+16)	I16 (7.6), I18 (21.6), R19 (104.6), I23 (24.5)	3.7 ± 0.15
TIIVQLNE (29−36)	I31 (+14/+16)	I31 (19.8)	3.9 ± 0.37
INCTRPNNNTR (40−50)	I40 (+16), C42 (+32/+16), R50 (+16)	I40 (1.2), C42 (4.0), R150 (145.4)	4.8 ± 1.6
KSIHIGPGQAFYATGE (51−66) (V3)	K51 (+16), I53 (+16), I55 (+16), P57 (+16), A60 (+16), F61 (+16), Y62 (+16), H54 (−23/−22/+16)	K51 (134.9), I53 (91.8), I55 (129.9), P57 (107.2), A60 (43.6), F61 (85.8), Y62 (64.7), H54 (81.7)	14 ± 1.3
SIHIGPGQAFYATGE (52−66)	I53 (+14), I55 (+16), A60 (+16), F61 (+16), P57 (+16), Y62 (+16), H54 (−23/−22/+16)	I53 (91.8), I55 (129.9), P57 (107.2), A60 (43.6), F61 (85.8), Y62 (64.7), H54 (81.7)	7.8 ± 0.50
SIHIGPGQAFYATGEIIGDIR (52−72) (V3)	I67 or I68 (+16), I71 (+16/+14), R72 (+16)	I67 (137.8) or I68 (143.9), I71 (98.9), R72 (61.2)	59 ± 8.6
TKWNKTLQQVAK (81−92)	W83 (+16/+32), K85 (+16)	W83 (5.1), K85 (61.6)	8.9 ± 0.61
LREHFNNK (94−101)	H97 (+16/−22)	H97 (65.1)	14 ± 1.9
TIIFKPSSGGDLE (102−114)	I103 (+16), F105 (+16), K106 (+16), P107 (+16), L113 (+16)	I103 (4.4), F105 (0), K106 (87.0), P107 (19.6), L113 (91.4)	2.2 ± 0.13
ITTHSFNCGGE (115−125)	I115 (+16), F120 (+16), C122 (+32)	I115 (55.2), F120 (16.8), C122 (21.3)	2.0 ± 0.19
GTYMFNGTR (143−151)	M146 (+16), Y145 (+16), F147 (+16), R151 (+16)	M146 (198.0), Y145 (152.7), F147 (144.0), R151 (190.6)	28 ± 5.7
TITLPCR (156−162)	I157 (+16), L159 (+16), P160 (+16) [or R162 (+16)]	I157 (37.9), L159 (2.4), P160 (73.3), R162 (102.1)	1.8 ± 0.39
ITCKSNITGLLLTR (186−199)	C188 (+48), K189 (+16)	C188 (17.1), K189 (111.8)	41 ± 2.5
SNITGLLLTR (190−199)	I192 (+16), L196 (+16), L197 (+16)	I192 (0), L196 (0), L197 (0)	0.20 ± 0.16
QIINMWQGVGQAMYAPPIEGK (165−185)	M169, W170, V173, M177, Y178, P180 (+16)	M169 (56.79), W170 (127.27), V173 (127.23), M177 (116.98), Y178 (89.4), P180 (76.26)	41 ± 6.4
NKTETFRPGGGDMR (208−221)	P215 (+16), M220 (+16), R221 (+16)	P215 (0), M220 (24.5), R221 (156.0)	41 ± 15

of CD4 and antibody binding on the template relative to the target must be kept in mind in the analysis.

The observed oxidations were consistent overall with the surface accessibility calculations (SASA) (Table 4). For example, the solvent accessibility and likely flexibility of the V3 and V4 loop structures were shown to be consistent with the oxidation of peptides 40−50, 51−66, 51−72, and 67−72 (V3 loop) and peptides 137−151, 143−151, and 152−162 (V4 loop), including multiple oxidized probe sites (Tables 3 and 4). However, variations in the oxidation rates for different V3 loop peptides argued for variations in the structure across the V3 loop. For example, V3 loop peptide 52−72 (59 ± 8.6 s^{-1}) had a much higher oxidation rate than V3 tip peptides 51−66 (13.7 ± 1.3 s^{-1}) and 52−66 (7.8 ± 0.5 s^{-1}). V3 loop peptide 47−69 resulting from Asp-N digestion also had an oxidation rate of 7.4 ± 1.7 s^{-1}. As the oxidation behavior of this peptide was influenced by the presence of sugar residues (N47), we mainly use GluC-tryptic peptide data for the sake of comparison in this case. The predicted SASA values of residues close to the C-terminus of the V3 loop (Ile 67, 137.8 Å^2; Ile 68, 143.9 Å^2; Ile 71, 98.88 Å^2; Arg 72, 61.2 Å^2) are very close to those of residues in the V3 tip (Lys 51, 134.9 Å^2; Ile 53, 91.8 Å^2; His 54, 81.7 Å^2; Ile 55, 129.9 Å^2; Pro 57, 107.2 Å^2; Phe 61, 85.8 Å^2; Tyr 62, 64.7 Å^2) in the homology model (*3*). Because the observed reactivitites are very different, while types of residues in terms of absolute reactivity are similar, it suggests that residues close to the C-terminus of the V3 loop, such as Ile 67, Ile 68, Ile 71, and Arg 72 (yellow area in Figure 7), exhibit much higher accessibility than those at the tip of the V3 loop.

Thus, the conformation of the V3 loop in free gp120-OD8 must be quite different from that derived from the homology model. Part of the V3 loop of gp120, including the V3 tip, might fold back in the unliganded state of gp120. Ligation of the antibody or CD4 to gp120-OD8 might cause distortion of the protein tertiary structure, resulting in the V3 tip being pushed out. There is a disulfide bond (peptide 42−76, blue area in Figure 7) in the V3 loop, but it is located inside the sheet, ∼8 Å (two pairs of residues) from other oxidized residues on the V3 base such as Arg 72. Therefore, the disulfide bond should not restrict the conformation of Arg 72, Ile 71, Ile 68, and Ile 67.

We performed solvent accessibility calculations for other V3 loop models, including one V3 loop model based on experimental NMR data (*66*). In this model, Ile 67, Ile 68, Ile 71, and Arg 72 from the V3 loop show relatively higher solvent accessibility values (Ile 67, 7.673 Å^2; Ile 68, 58.82 Å^2; Ile 71, 25.21 Å^2; Arg 72, 75.86 Å^2) than many residues on the V3 tip (Lys 51, 0.131 Å^2; Ile 53, 1.083 Å^2; Pro 57, 31.62 Å^2; Phe 61, 0.269 Å^2), consistent with our overall footprinting results (Figure 6 and Table 1 of the Supporting Information). However, there are also some discrepancies between this model and our results. The indicated SASA values of some residues differ from the observed experimental results. For example, some of the reactive residues in the V3 tip such as His 54, Ile 55, Ala 60, and Tyr 62 were shown to be oxidized in radiolysis, but SASA values of these residues in this structure are 0 Å^2. This may be because this V3 loop construct is an independent folding unit lacking potential interactions with other parts of gp120-OD8.

The V3 loop is known to be important in HIV-1 immunopathogenesis (*67*), and the loop protrudes when CD4 and other antibodies are bound. Our footprinting data indicate a quite different structure for the V3 loop in gp120-OD8 and are consistent with observations that addition of thrombin to virion-associated gp120 results in cleavage at the V3 loop when recombinant soluble CD4 (rsCD4) is added (*68*−*70*). Moreover, without

addition of an exogenous protease, proteolytic cleavage of gp120 was induced upon binding to rsCD4. This cleavage was observed most likely at the V3 loop, depending on the rsCD4 concentration and the incubation time (*71*).

In our construct, peptides containing CD4 binding regions had varying oxidation rates consistent with their solvent accessibility calculations; these included peptide 16−28 ($3.3 \pm 0.21\ s^{-1}$), peptide 102−114 ($2.1 \pm 0.13\ s^{-1}$), peptide 165−189 ($41 \pm 6.4\ s^{-1}$), and peptide 200−218 ($1.5 \pm 0.05\ s^{-1}$). The bridging sheet exhibited a high oxidation rate ($41 \pm 6.4\ s^{-1}$) with many reactive and accessible sites, which suggested that removal of the inner domain protein sequence leaves a large surface area of solvent accessible sites. However, there are some minor exceptions to this trend. For example, the SASA values of I192, L196, and L197 in peptide 190−199 are 0 $Å^2$, while this peptide exhibited detectable (but very low) reactivity with a rate of $0.1\ s^{-1}$. Because the oxidation rate is low, it may suggest a very modest solvent accessibility or dynamic flexibility in these residues.

Influence of Glycosylation on Solvent Accessibility. We examined the rates of oxidation for peptides containing a glycosylation motif for potential inconsistencies with results obtained from solvent accessibility calculations to explore potential attenuation of reactivity due to sugar structures. In most cases, the predicted solvent accessibility did not appear to be influenced by glycosylation. However, the specific oxidation products observed for the glycopeptide in some cases reflected a conformation influenced by glycosylation. In PNGase F-deamidated Asp-N peptides 77−83 and GluC-tryptic peptide 83−92, only 32 Da oxidation products of Trp 83 were detected; 16 Da oxidation products of tryptic peptide 83−92 were detected in unglycosylated GluC-trypsin peptides, but it was difficult to determine the exact oxidation sites (e.g., Trp 83 or Lys 85). Similarly, analysis of PNGase F-deamidated Asp-N peptides 137−141 and 130−141 showed that Trp 139 was oxidized to 32 and 4 Da products only (Figure 3 of the Supporting Information), while the oxidized products of Trp 139 in tryptic digests were not detected. Hydroxyl radicals attack the benzene ring in tryptophan side chains to generate 16 Da products, while when the pyrrole ring is oxidized, 16, 32, and/or 4 Da products are generated (*24*). These signatures of oxidation of specific structures within a side chain have been previously reported to reflect the specific conformation of the Trp side chain in protein−protein interactions (*22*). Thus, the sugar on residues Asn 84 and Asn 136 adjacent to Trp 83 and Trp 139, respectively, may preferentially protect the Trp benzene, resulting in oxidation primarily of the pyrrole ring. Because Asn 84 is partially occupied by glycosylation, it is possible to detect 16 Da oxidation products in nonglycosylated peptides on Trp 83 (undeamidated peptides after enzyme deglycosylation).

CONCLUSIONS

In this study, we characterized the structure of gp120-OD8 by identifying the glycosylation sites and structure and monitoring the accessible residues on the protein surface. We determined glycosylation sites of gp120-OD8 and estimated the occupancy of glycosylation by enzymatically removing the glycans before mass spectrometry analysis. Eleven of 15 glycosylation motifs were shown to have high-mannose structures or hybrid structures of glycosylation. The oligosaccharide occupancy at each site varied from 0.7 to 99%. Glycan structure for nine glycosylation sites in eight unique glycopeptides was assessed by combination of ECD and CID FT-MS data. Comparison of gp120-OD8 experimental footprinting data with structural data from a homology model derived from the ligated CD4−gp120-OD8 crystal structure revealed that the V3 loop of free gp120-OD8 is solvent accessible, like ligated gp120, but that the V3 tip may be bound to the surface of the protein leaving the base exposed. In addition, we detected specific associations between specific sugar moieties and adjacent Trp residues in the structure.

ACKNOWLEDGMENT

We acknowledge Janna Kiselar and Parminder Kaur for their discussion of this study. Use of the X28C beamline at the National Synchrotron Light Source, Brookhaven National Laboratory, was supported by the National Institute of Biomedical Imaging and Bioengineering (P30-EB-09998) and the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract DE-AC02-98CH10886.

SUPPORTING INFORMATION AVAILABLE

Supplementary figures and tables. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Ostermeier, C., Iwata, S., Ludwig, B., and Michel, H. (1995) Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome c oxidase. *Nat. Struct. Biol. 2* (10), 842–846.
2. Kwong, P. D., Wyatt, R., Desjardins, E., Robinson, J., Culp, J. S., Hellmig, B. D., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (1999) Probability analysis of variational crystallization and its application to gp120, the exterior envelope glycoprotein of type 1 human immunodeficiency virus (HIV-1). *J. Biol. Chem. 274* (7), 4115–4123.
3. Huang, C. C., Tang, M., Zhang, M. Y., Majeed, S., Montabana, E., Stanfield, R. L., Dimitrov, D. S., Korber, B., Sodroski, J., Wilson, I. A., Wyatt, R., and Kwong, P. D. (2005) Structure of a V3-containing HIV-1 gp120 core. *Science 310* (5750), 1025–1028.
4. Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature 393* (6686), 648–659.
5. Kwong, P. D., Wyatt, R., Majeed, S., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (2000) Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. *Structure 8* (12), 1329–1339.
6. Chen, B., Vogan, E. M., Gong, H., Skehel, J. J., Wiley, D. C., and Harrison, S. C. (2005) Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature 433* (7028), 834–841.
7. 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., and Shaw, G. M. (2003) Antibody neutralization and escape by HIV-1. *Nature 422* (6929), 307–312.
8. Morelle, W., Canis, K., Chirat, F., Faid, V., and Michalski, J. C. (2006) The use of mass spectrometry for the proteomic analysis of glycosylation. *Proteomics 6* (14), 3993–4015.
9. Fenouillet, E., Gluckman, J. C., and Jones, I. M. (1994) Functions of HIV envelope glycans. *Trends Biochem. Sci. 19* (2), 65–70.
10. Albach, C., Damoc, E., Denzinger, T., Schachner, M., Przybylski, M., and Schmitz, B. (2004) Identification of N-glycosylation sites of the murine neural cell adhesion molecule NCAM by MALDI-TOF and MALDI-FTICR mass spectrometry. *Anal. Bioanal. Chem. 378* (4), 1129–1135.
11. Morelle, W., Faid, V., Chirat, F., and Michalski, J. C. (2009) Analysis of N- and O-linked glycans from glycoproteins using MALDI-TOF mass spectrometry. *Methods Mol. Biol. 534*, 5–21.
12. Picariello, G., Ferranti, P., Mamone, G., Roepstorff, P., and Addeo, F. (2008) Identification of N-linked glycoproteins in human milk by hydrophilic interaction liquid chromatography and mass spectrometry. *Proteomics 8* (18), 3833–3847.

13. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Loo, J. A. (2006) Identification of N-linked glycoproteins in human saliva by glycoprotein capture and mass spectrometry. *J. Proteome Res. 5* (6), 1493–1503.
14. Zhang, H., Li, X. J., Martin, D. B., and Aebersold, R. (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol. 21* (6), 660–666.
15. Zeng, C., and Biemann, K. (1999) Determination of N-linked glycosylation of yeast external invertase by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Mass Spectrom. 34* (4), 311–329.
16. Khoshnoodi, J., Hill, S., Tryggvason, K., Hudson, B., and Friedman, D. B. (2007) Identification of N-linked glycosylation sites in human nephrin using mass spectrometry. *J. Mass Spectrom. 42* (3), 370–379.
17. Geyer, R., Geyer, H., Egge, H., Schott, H. H., and Stirm, S. (1984) Structure of the oligosaccharides sensitive to endo-β-N-acetylglucosaminidase H in the glycoprotein of Friend murine leukemia virus. *Eur. J. Biochem. 143* (3), 531–539.
18. Danielle, L. P., Paul, K., Theresa, M., and Alain, B. (2009) Characterization of Minor N-linked Glycans on Antibodies Using Endo H Release and MALDI-Mass Spectrometry. *Anal. Lett. 42* (11), 1711–1724.
19. Hakansson, K., Cooper, H. J., Emmett, M. R., Costello, C. E., Marshall, A. G., and Nilsson, C. L. (2001) Electron capture dissociation and infrared multiphoton dissociation MS/MS of an N-glycosylated tryptic peptic to yield complementary sequence information. *Anal. Chem. 73* (18), 4530–4536.
20. Hakansson, K., Chalmers, M. J., Quinn, J. P., McFarland, M. A., Hendrickson, C. L., and Marshall, A. G. (2003) Combined electron capture and infrared multiphoton dissociation for multistage MS/MS in a Fourier transform ion cyclotron resonance mass spectrometer. *Anal. Chem. 75* (13), 3256–3262.
21. Hogan, J. M., Pitteri, S. J., Chrisman, P. A., and McLuckey, S. A. (2005) Complementary structural information from a tryptic N-linked glycopeptide via electron transfer ion/ion reactions and collision-induced dissociation. *J. Proteome Res. 4* (2), 628–632.
22. Kiselar, J. G., Mahaffy, R., Pollard, T. D., Almo, S. C., and Chance, M. R. (2007) Visualizing Arp2/3 complex activation mediated by binding of ATP and WASp using structural mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A. 104* (5), 1552–1557.
23. Takamoto, K., and Chance, M. R. (2006) Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes. *Annu. Rev. Biophys. Biomol. Struct. 35*, 251–276.
24. Xu, G., and Chance, M. R. (2005) Radiolytic modification and reactivity of amino acid residues serving as structural probes for protein footprinting. *Anal. Chem. 77* (14), 4549–4955.
25. Xu, G., and Chance, M. R. (2007) Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem. Rev. 107* (8), 3514–3543.
26. Gupta, S., Sullivan, M., Toomey, J., Kiselar, J., and Chance, M. R. (2007) The Beamline X28C of the Center for Synchrotron Biosciences: A national resource for biomolecular structure and dynamics experiments using synchrotron footprinting. *J. Synchrotron Radiat. 14* (Part 3), 233–243.
27. Goldsmith, S. C., Guan, J. Q., Almo, S., and Chance, M. (2001) Synchrotron protein footprinting: A technique to investigate protein-protein interactions. *J. Biomol. Struct. Dyn. 19* (3), 405–418.
28. Zheng, X., Wintrode, P. L., and Chance, M. R. (2008) Complementary structural mass spectrometry techniques reveal local dynamics in functionally important regions of a metastable serpin. *Structure 16* (1), 38–51.
29. Kiselar, J. G., Maleknia, S. D., Sullivan, M., Downard, K. M., and Chance, M. R. (2002) Hydroxyl radical probe of protein surfaces using synchrotron X-ray radiolysis and mass spectrometry. *Int. J. Radiat. Biol. 78* (2), 101–114.
30. Kiselar, J. G., Janmey, P. A., Almo, S. C., and Chance, M. R. (2003) Structural analysis of gelsolin using synchrotron protein footprinting. *Mol. Cell. Proteomics 2* (10), 1120–1132.
31. Gao, F., Weaver, E. A., Lu, Z., Li, Y., Liao, H. X., Ma, B., Alam, S. M., Scearce, R. M., Sutherland, L. L., Yu, J. S., Decker, J. M., Shaw, G. M., Montefiori, D. C., Korber, B. T., Hahn, B. H., and Haynes, B. F. (2005) Antigenicity and immunogenicity of a synthetic human immunodeficiency virus type 1 group m consensus envelope glycoprotein. *J. Virol. 79* (2), 1154–1163.
32. Sullivan, M. R., Rekhi, S., Bohon, J., Gupta, S., Abel, D., Toomey, J., and Chance, M. R. (2008) Installation and testing of a focusing mirror at beamline X28C for high flux X-ray radiolysis of biological macromolecules. *Rev. Sci. Instrum. 79* (Part 1), 025101.
33. Xu, H., and Freitas, M. A. (2007) A mass accuracy sensitive probability based scoring algorithm for database searching of tandem mass spectrometry data. *BMC Bioinf. 8*, 133.
34. Xu, H., Zhang, L., and Freitas, M. A. (2008) Identification and characterization of disulfide bonds in proteins and peptides from tandem MS data by use of the MassMatrix MS/MS search engine. *J. Proteome Res. 7* (1), 138–144.
35. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis 20* (18), 3551–3567.
36. Kaur, P., Kiselar, J. G., and Chance, M. R. (2009) Integrated algorithms for high-throughput examination of covalently labeled biomolecules by structural mass spectrometry. *Anal. Chem. 81* (19), 8141–8149.
37. Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003) SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res. 31* (13), 3381–3385.
38. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics 22* (2), 195–201.
39. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis 18* (15), 2714–2723.
40. Sanner, M. F., Olson, A. J., and Spehner, J. C. (1996) Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers 38* (3), 305–320.
41. Zhou, T., Xu, L., Dey, B., Hessell, A. J., Van Ryk, D., Xiang, S. H., Yang, X., Zhang, M. Y., Zwick, M. B., Arthos, J., Burton, D. R., Dimitrov, D. S., Sodroski, J., Wyatt, R., Nabel, G. J., and Kwong, P. D. (2007) Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature 445* (7129), 732–737.
42. Plummer, T. H., Jr., and Tarentino, A. L. (1991) Purification of the oligosaccharide-cleaving enzymes of *Flavobacterium meningosepticum*. *Glycobiology 1* (3), 257–263.
43. Maley, F., Trimble, R. B., Tarentino, A. L., and Plummer, T. H., Jr. (1989) Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. *Anal. Biochem. 180* (2), 195–204.
44. Wold, F. (1981) In vivo chemical modification of proteins (post-translational modification). *Annu. Rev. Biochem. 50*, 783–814.
45. Robinson, N. E. (2002) Protein deamidation. *Proc. Natl. Acad. Sci. U.S.A. 99* (8), 5283–5288.
46. Robinson, N. E., and Robinson, A. B. (2004) Mocecular Clocks: Deamidation of Asparaginyl and Glutaminyl Residues in Peptides and Proteins, Althouse Press, London, ON.
47. Geiger, T., and Clarke, S. (1987) Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. *J. Biol. Chem. 262* (2), 785–794.
48. Kubelka, V., Altmann, F., Kornfeld, G., and Marz, L. (1994) Structures of the N-linked oligosaccharides of the membrane glycoproteins from three lepidopteran cell lines (Sf-21, IZD-Mb-0503, Bm-N). *Arch. Biochem. Biophys. 308* (1), 148–157.
49. Altmann, F., Staudacher, E., Wilson, I. B., and Marz, L. (1999) Insect cells as hosts for the expression of recombinant glycoproteins. *Glycoconjugate J. 16* (2), 109–123.
50. hang, G. D., Chen, C. J., Lin, C. Y., Chen, H. C., and Chen, H. (2003) Improvement of glycosylation in insect cells with mammalian glycosyltransferases. *J. Biotechnol. 102* (1), 61–71.
51. Ilchenko, S. A., Chance, M. R., Whittaker, L. J., and Whittaker, J. (2008) Determination of the Glycopeptide Structure of Insulin and IGF-I Receptors. 56th ASMS Conference on Mass Spectrometry and Allied Topics, Denver, CO.
52. Chalkley, R. J., Baker, P. R., Medzihradszky, K. F., Lynn, A. J., and Burlingame, A. L. (2008) In-depth analysis of tandem mass spectrometry data from disparate instrument types. *Mol. Cell. Proteomics 7* (12), 2386–2398.
53. Chalkley, R. J., Baker, P. R., Hansen, K. C., Medzihradszky, K. F., Allen, N. P., Rexach, M., and Burlingame, A. L. (2005) Comprehensive analysis of a multidimensional liquid chromatography mass spectrometry dataset acquired on a quadrupole selecting, quadrupole collision cell, time-of-flight mass spectrometer: I. How much of the data is theoretically interpretable by search engines? *Mol. Cell. Proteomics 4* (8), 1189–1193.
54. Rizzuto, C. D., Wyatt, R., Hernandez-Ramos, N., Sun, Y., Kwong, P. D., Hendrickson, W. A., and Sodroski, J. (1998) A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science 280* (5371), 1949–1953.

55. Rizzuto, C., and Sodroski, J. (2000) Fine definition of a conserved CCR5-binding region on the human immunodeficiency virus type 1 glycoprotein 120. *AIDS Res. Hum. Retroviruses 16* (8), 741–749.
56. Cormier, E. G., Tran, D. N., Yukhayeva, L., Olson, W. C., and Dragic, T. (2001) Mapping the determinants of the CCR5 amino-terminal sulfopeptide interaction with soluble human immunodeficiency virus type 1 gp120-CD4 complexes. *J. Virol. 75* (12), 5541–5549.
57. Burton, D. R., Desrosiers, R. C., Doms, R. W., Koff, W. C., Kwong, P. D., Moore, J. P., Nabel, G. J., Sodroski, J., Wilson, I. A., and Wyatt, R. T. (2004) HIV vaccine design and the neutralizing antibody problem. *Nat. Immunol. 5* (3), 233–236.
58. Javaherian, K., Langlois, A. J., McDanal, C., Ross, K. L., Eckler, L. I., Jellis, C. L., Profy, A. T., Rusche, J. R., Bolognesi, D. P., and Putney, S. D.; et al. (1989) Principal neutralizing domain of the human immunodeficiency virus type 1 envelope protein. *Proc. Natl. Acad. Sci. U.S.A. 86* (17), 6768–6772.
59. Putney, S. D., Matthews, T. J., Robey, W. G., Lynn, D. L., Robert-Guroff, M., Mueller, W. T., Langlois, A. J., Ghrayeb, J., Petteway, S. R., Jr., and Weinhold, K. J.; et al. (1986) HTLV-III/LAV-neutralizing antibodies to an *E. coli*-produced fragment of the virus envelope. *Science 234* (4782), 1392–1395.
60. Guan, J. Q., and Chance, M. R. (2005) Structural proteomics of macromolecular assemblies using oxidative footprinting and mass spectrometry. *Trends Biochem. Sci. 30* (10), 583–592.
61. Guan, J. Q., Takamoto, K., Almo, S. C., Reisler, E., and Chance, M. R. (2005) Structure and dynamics of the actin filament. *Biochemistry 44* (9), 3166–3175.
62. Kiselar, J. G., Janmey, P. A., Almo, S. C., and Chance, M. R. (2003) Visualizing the Ca^{2+}-dependent activation of gelsolin by using synchrotron footprinting. *Proc. Natl. Acad. Sci. U.S.A. 100* (7), 3942–3947.
63. Liu, R., Guan, J. Q., Zak, O., Aisen, P., and Chance, M. R. (2003) Structural reorganization of the transferrin C-lobe and transferrin receptor upon complex formation: The C-lobe binds to the receptor helical domain. *Biochemistry 42* (43), 12447–124454.
64. Takamoto, K., Das, R., He, Q., Doniach, S., Brenowitz, M., Herschlag, D., and Chance, M. R. (2004) Principles of RNA compaction: Insights from the equilibrium folding pathway of the P4-P6 RNA domain in monovalent cations. *J. Mol. Biol. 343* (5), 1195–1206.
65. Xu, G., Liu, R., Zak, O., Aisen, P., and Chance, M. R. (2005) Structural allostery and binding of the transferrin*receptor complex. *Mol. Cell. Proteomics 4* (12), 1959–1967.
66. Vranken, W. F., Fant, F., Budesinsky, M., and Borremans, F. A. (2001) Conformational model for the consensus V3 loop of the envelope protein gp120 of HIV-1 in a 20% trifluoroethanol/water solution. *Eur. J. Biochem. 268* (9), 2620–2628.
67. Moore, J. P., and Nara, P. L. (1991) The role of the V3 loop of gp120 in HIV infection. *AIDS 5* (Suppl. 2), S21–S33.
68. Clements, G. J., Price-Jones, M. J., Stephens, P. E., Sutton, C., Schulz, T. F., Clapham, P. R., McKeating, J. A., McClure, M. O., Thomson, S., and Marsh, M.; et al. (1991) The V3 loops of the HIV-1 and HIV-2 surface glycoproteins contain proteolytic cleavage sites: A possible function in viral fusion? *AIDS Res. Hum. Retroviruses 7* (1), 3–16.
69. Stephens, P. E., Clements, G., Yarranton, G. T., and Moore, J. (1990) A chink in HIV's armour? *Nature 343* (6255), 219.
70. Sattentau, Q. J., and Moore, J. P. (1991) Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. *J. Exp. Med. 174* (2), 407–415.
71. Werner, A., and Levy, J. A. (1993) Human immunodeficiency virus type 1 envelope gp120 is cleaved after incubation with recombinant soluble CD4. *J. Virol. 67* (5), 2566–2574.