Site-Specific *N*-Glycosylation and Oligosaccharide Structures of Recombinant HIV-1 gp120 Derived from a Baculovirus Expression System†

Jiunn-chern Yeh,‡ Jonathan R. Seals,⊥ Cheryl I. Murphy,⊥ Herman van Halbeek,§ and Richard D. Cummings*,‡

Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, P.O. Box 26901, BSEB-325, Oklahoma City, Oklahoma 73190, Complex Carbohydrate Research Center, Department of Biochemistry, University of Georgia, 220 Riverbend Road, Athens, Georgia 30602, and Department of Protein Chemistry, Cambridge Biotech Corporation, 365 Plantation Street, Worcester, Maryland 01605

Received July 1, 1993; Revised Manuscript Received August 5, 1993®

ABSTRACT: We report the complete structures of the *N*-linked oligosaccharides and the site-specificity of the *N*-glycosylation of recombinant gp120 (rgp120) of the HIV-1 BH8 isolate produced by a baculovirus expression system. Glycopeptides derived from the tryptic digests of intact rgp120 or of cyanogen bromide-generated fragments of rgp120 were isolated by their binding to concanavalin A–Sepharose and were purified by reversed-phase HPLC. The isolated glycopeptides were treated with PNGase F, releasing the carbohydrate moiety while converting Asn to Asp, and identified by amino acid analysis and/or peptide sequencing. Our results indicate that all 22 potential *N*-glycosylation sites in the rgp120 sequence are utilized. We did not detect *N*-acetylgalactosamine in rgp120, indicating that the glycoprotein lacks typical *O*-linked oligosaccharides. To investigate the oligosaccharide structures at the sites of glycosylation, we determined the carbohydrate composition for each site and characterized the oligosaccharides by ^{1}H-NMR spectroscopy and by oligosaccharide mapping using high pH anion-exchange chromatography. Mannose and *N*-acetylglucosamine were the only sugars observed in the intact rgp120 and likewise in individual glycopeptides. All glycopeptides derived from rgp120 contained high mannose-type *N*-linked oligosaccharides, ranging from $GlcNAc_2Man_5$ to $GlcNAc_2Man_9$. However, different glycosylation sites showed varied degrees of processing of the high mannose-type oligosaccharides, as characterized by the ratio of $GlcNAc_2Man_{8-9}$ to $GlcNAc_2Man_{5-7}$. These results demonstrate that *N*-glycosylation of rgp120 in the baculovirus expression system occurs at all potential sites and is site specific in terms of oligosaccharide structures.

Human immunodeficiency virus (HIV),[1] the causative agent of acquired immunodeficiency syndrome, contains glycoproteins gp120 and gp41 within its envelope. These two glycoproteins are produced by the proteolytic cleavage of a precursor protein, gp160, and remain associated through noncovalent association (Robey et al., 1985). Gp120 plays an important role in the first step of viral infection, because it serves as a high-affinity ligand for the T-cell receptor CD4 (Dalgleish et al., 1984; Klatzman et al., 1984). After binding to CD4, gp41 is then able to mediate virus–cell fusion via its *N*-terminal hydrophobic domain, and subsequently the virus particle is released into the cells.

Gp120 is a heavily glycosylated protein with about 50% of its molecular weight contributed by carbohydrate (Allan et al., 1985), and in all characterized strains of the virus, gp120 contains between 20 and 26 consensus Asn-glycosylation sites. Because of the lack of proofreading function of the viral reverse transcriptase, the genome of the virus shows a high degree of heterogeneity, especially on the *env* gene, which encodes proteins facing the host immune surveillance. The gp120s from different isolates sequenced to date show an extraordinarily high-degree of variability in their amino acid sequences. This sequence variability has been mapped primarily to five hypervariable regions (Modrow et al., 1987; Willey et al., 1986) and can contribute up to 30% of the heterogeneity between two isolates. Eighteen cysteine residues from different isolates have been conserved in all isolates examined; however, only 13 of the approximately 24 potential *N*-glycosylation sites were conserved among different isolates. Even though many of these glycosylation sites are located in the hypervariable regions, the total number of *N*-glycosylation sites appears to remain constant (generally 20–26), which indicates the potential functional importance of glycosylation.

Many studies have investigated the potential roles of carbohydrates in the virus life cycle, infectivity, and cytopathic effects. Several mannose-binding lectins and monoclonal antibodies to specific carbohydrate structures can block syncytia formation and reduce viral infectivity *in vitro* (Lifson et al., 1986; Robinson et al., 1987; Hansen et al., 1990; Muller et al., 1991). Divergent inhibitors of *N*-linked oligosaccharide biosynthesis have also been shown to have dramatic effects on the virus life cycle (Gruters et al., 1987; Walker et al., 1987; Montefiori et al., 1988; Pal et al., 1989a). In other approaches to define the potential importance of carbohydrate for virus infectivity, potential *N*-linked glycosylation sites

† This work was supported by NIH Program Grant AI27135 to R.D.C. and H.H.

* Address correspondence to Dr. Richard D. Cummings at the University of Oklahoma Health Sciences Center. Phone: 405-271-2546. Fax: 405-271-3910.

‡ University of Oklahoma Health Sciences Center.

§ University of Georgia.

⊥ Cambridge Biotech Corporation.

® Abstract published in *Advance ACS Abstracts*, October 1, 1993.

[1] Abbreviations: HIV, human immunodeficiency virus; gp, glycoprotein; rgp, recombinant glycoprotein; Con A, concanavalin A agglutinin; TPCK, L-1-(*p*-tosylamido)-2-phenylethyl chloromethyl ketone; PNGase F, peptide *N*-glycosidase F; HPLC, high-performance liquid chromatography; TFA, trifluoroacetic acid; HPAEC, high-pH anion-exchange chromatography; PAD, pulsed amperometric detection; DTT, dithiothreitol; RCM, reduced and *S*-carboxymethylated; CNBr, cyanogen bromide; TBS, Tris-buffered saline; PTC, phenylthiocarbamyl; PTH, phenylthiohydantoin; NMR, nuclear magnetic resonance; Sf, *Spodoptera frugiperda*; GlcNAc, *N*-acetylglucosamine; Man, mannose; CHO, chinese hamster ovary cells; Endo H, *endo*-β-*N*-acetylglucosaminidase H; GalNAc, *N*-acetylgalactosamine.

0006-2960/93/0432-11087$04.00/0 © 1993 American Chemical Society

within gp120 were deleted, and some deletions blocked or impaired viral infectivity (Willey et al., 1988; Lee et al., 1992).

Data from these studies aimed at understanding the role of the carbohydrates on the gp120 molecule, however, have been difficult to evaluate because of a lack of specific information about the glycosylation of the protein, such as the utilization of individual glycosylation sites, the structures of oligosaccharides at particular glycosylation sites, and the effects of deletion of *N*-glycosylation sites on the glycosylation of nearby sites. A major reason this information has been difficult to obtain is the limited availability of gp120 from natural sources. Some specific information about the glycosylation of gp120 was derived from studies of a recombinant form of the glycoprotein produced in Chinese hamster ovary cells (Leonard et al., 1990).

The baculovirus expression system has been used to express a variety of proteins in large quantities in insect cells under the control of a strong polyhedrin promoter (Luckow & Summers, 1988). Most of the recombinant proteins produced in this system have been shown to be antigenically, immunogenically, and functionally similar to their authentic counterparts, because the insect cells are able to carry out many types of posttranslational modifications, such as addition of *N*- and *O*-linked oligosaccharides (Luckow & Summers, 1988). The envelope glycoproteins of HIV-1 have been expressed in insect cells using baculovirus vectors and shown to retain the ability to bind to the CD4 receptor (Wells & Compans, 1990; Murphy et al., 1990). In addition, antibodies from goats immunized with baculovirus recombinant gp160 have been shown to neutralize the viral infections *in vitro* (Rusche et al., 1987). Furthermore, the recombinant HIV-1 envelope expressed using baculovirus vectors received FDA approval for clinical evaluation as a candidate vaccine for AIDS (Luckow & Summers, 1988).

In this paper, we have characterized the sites of glycosylation in recombinant gp120 produced in the baculovirus system. Our results indicate that in this system the gp120 of the HIV-1 BH8 isolate contains high mannose-type *N*-linked oligosaccharides at each of the 22 potential *N*-glycosylation sites and that the degree of processing of these oligosaccharides is highly site specific.

EXPERIMENTAL PROCEDURES

Materials. Dithiothreitol, iodoacetic acid, cyanogen bromide, and PITC were purchased from Sigma Chemical Co. (St. Louis, MO). Con A–Sepharose and Superose 12 (pre grade) were purchased from Pharmacia LKB Biotechnology (Picataway, NJ). TPCK-treated trypsin was purchased from Worthington Biomedical Corp. (Freehold, NJ). *Staphylococcus aureus* V8 protease was obtained from Boehringer Mannheim Biochemicals (Indianapolis, IN). Human pancreatic chymotrypsin was obtained from Calbiochem (La Jolla, CA). The PNGase F (recombinant *N*-glycanase) was obtained from Genzyme Corp. (Cambridge, MA). HPLC/spectro grade trifluoroacetic acid, BCA protein assay reagent, Extracti-Gel D detergent-removing gel, and constant-boiling 6 N HCl were obtained from Pierce Chemical Co. (Rockford, IL). Monosaccharide standards were obtained from Dionex Corp. (Houston, TX). High mannose-type oligosaccharide standards were purchased from Oxford GlycoSystems (Rosedale, NY).

Purification of Baculovirus Recombinant HIV-1 gp120. The baculovirus recombinant HIV-1 gp120 was produced according to the procedures described by Murphy et al. (1993). The recombinant gp120 produced by plasmid pVLBSegt120 was purified from culture medium by immunoaffinity chromatograpy (Murphy et al., 1993), followed by SDS–PAGE using a Bio-Rad Model 491 Prep Cell to remove any smaller, incomplete fragments of the protein from the final preparation. Samples purified by immunoaffinity chromatography, containing 15–25 mg of gp120, were subjected to SDS–PAGE through 6% acrylamide gels and fractionated by continuous elution with 0.05 M ammonium bicarbonate. Fractions containing the full-length protein were identified by analytical SDS–PAGE, pooled, lyophilized, and stored dry at −40 °C prior to analysis.

Carbohydrate Compositional Analysis. Samples (intact protein or glycopeptides) were dried down in 13- × 100-mm screw-capped tubes with Teflon-coated caps and were hydrolyzed in 200 μL of 2 N TFA at 100 °C for 4 h. After hydrolysis, the samples were dried and resuspended in deionized water. The carbohydrate composition analysis was done by high-pH anion-exchange chromatography (HPAEC) using a Carbopac PA1 column (4 × 250 mm). The samples were separated with an isocratic gradient of 12 mM sodium hydroxide for 25 min. The flow rate was 1 mL/min, and the separation was monitored with a pulsed amperometric detector. Postcolumn addition of 300 mM sodium hydroxide at 0.6 mL/min was used to increase the detection. Quantitation was done by calculating the peak area of monosaccharides on the basis of the peak area of the monosaccharide standards treated under the same conditions.

Removal of SDS from Electroeluted Recombinant gp120. Lyophilized sample was reconstituted to 10 mg/mL in degassed 0.1 M ammonium bicarbonate buffer. Extracti-Gel D detergent-removing gel was degassed under vacuum and equilibrated in degassed 0.1 M ammonium bicarbonate buffer. The sample was applied to the gel (generally, 1 mL of gel was used for 50 mg of sample) and eluted with the same buffer for 10 bed volumes. The sample was collected, and the fractions were assayed by BCA protein assay and lyophilized.

Reduction and S-Carboxymethylation. Baculovirus recombinant gp120 (50 mg) was dialyzed against 0.36 M Tris buffer, pH 8.6, containing 8 M urea and 3 mM EDTA. DTT was added to a concentration of 10 mM, and the sample was incubated for 4 h at room temperature. The sample was then treated with 25 mM iodoacetic acid in the dark for 30 min at room temperature. The reaction was quenched with excess DTT. The sample was dialyzed against 0.1 M ammonium bicarbonate at 4 °C and then lyophilized and was designated reduced and carboxymethylated rgp120 (RCM-rgp120).

Cyanogen Bromide (CNBr) Digestion of RCM-rgp120. The RCM-rgp120 (50 mg) was resuspended in 70% formic acid to a concentration of 10 mg/mL. A small pinch of CNBr (about 5 mg) was added. The reaction mixture was incubated at room temperature in an argon atmosphere in the dark for 24 h. The reaction was stopped by adding 10 volumes of water and was lyophilized.

Separation of CNBr Fragments on FPLC. CNBr-digested RCM-rgp120 (50 mg) was resuspended in 1 mL of 0.1 M ammonium bicarbonate buffer and separated on Superose 12 (prep grade) in an HR 16/50 column. The column was preequilibrated in 0.1 M ammonium bicarbonate buffer (pH 7.8), and the column flow rate was 1 mL/min. Fractions were monitored by absorbance at 280 nm, and 1-mL fractions were collected. Peaks were pooled and rerun on Superose 12 under the same conditions.

Trypsin Digestion. Samples were suspended in 0.1 M ammonium bicarbonate buffer to a final concentration of about 10 mg/mL. TPCK-treated trypsin was added to a ratio of

1:100 of the protein sample weight at 0 h, and an additional similar portion was added 6 h later. The mixtures were incubated at 37 °C for a total of 24 h. The samples were boiled at 100 °C to inactivate trypsin and then lyophilized.

Treatment of Peptides with S. aureus V8 Protease. For specific cleavage at glutamic acid residues, the samples were reconstituted in 0.1 M ammonium bicarbonate buffer (pH 7.8); for cleavage at glutamic acid and aspartic acid, the samples were reconstituted in 50 mM sodium phosphate buffer, pH 7.8. V8 protease was added to a ratio of 1:100 of the estimated sample weight, and the samples were incubated at 37 °C overnight. The digested samples were injected directly onto reversed-phase HPLC to separate the cleaved peptides.

Treatment of Peptides with Chymotrypsin. The peptides isolated from reversed-phase HPLC were dried and reconstituted in 0.1 M ammonium bicarbonate buffer (pH 7.8). Chymotrypsin was added at a ratio of 1:100 by weight to the sample, according to the estimated sample weight. The digestion was conducted overnight at 37 °C. The samples were injected directly onto reversed-phase HPLC to separate the digested products.

Isolation of Glycopeptides by Concanavalin A–Sepharose. The amount of concanavalin A (Con A)–Sepharose used depended on the amount of the sample. Generally, 1 mL of Con A–Sepharose was used for 20 nmol of starting material. Lyophilized tryptic digests were resuspended in 1 mL of Tris-buffered saline (TBS/NaN_3, 0.01 M Tris, pH 8.0, 0.15 M NaCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$, and 0.02% NaN_3) and applied to a Con A–Sepharose column preequilibrated in the same buffer. The column was washed with TBS/NaN_3, and unbound material eluted. The bound material was eluted by application onto the column about one-half the column bed volume of 200 mM α-methylmannoside in TBS/NaN_3. After the bottom of the column was capped, the column was immersed in a boiling water bath for 10 min. This step was necessary to inactivate the Con A; hapten sugars alone were unable to quantitatively elute the bound glycopeptides. After boiling, the column was again connected to the fraction collector and was eluted with addition 200 mM α-methylmannoside preequilibrated to 60 °C. The fractions were monitored for absorbance at 280 nm. The bound and unbound materials were pooled and desalted on a Bio-Gel P2 column (1.5 × 50 cm) in deionized water. The void fractions were collected and lyophilized.

Reversed-Phase HPLC. Separation of tryptic peptides was performed by reversed-phase HPLC with a 5-μm Vydac C_{18} endcapped column (4.6 × 250 mm). The column was first equilibrated with 0.1% trifluoroacetic acid, and the separation was carried out by a linear gradient from 0 to 45% acetonitrile containing 0.08% trifluoroacetic acid in 90 min. The system used was a Shimadzu low-pressure mixing liquid chromatograph system with a single LC-10AD pump and a FCV-10AL low-pressure flow channel selection valve. The detector used was a Shimadzu SPD-10A UV–vis detector set at 214 nm.

Treatment of Glycopeptides or Intact gp120 with PNGase F. The glycopeptides (ranging from 1 to 5 nmol) purified from reversed-phase HPLC were dried down and resuspended in 50 μL of 50 mM sodium phosphate buffer containing 0.2% sodium azide, pH 7.0. Two milliunits of PNGase F was added, and the sample was incubated at 37 °C for 16 h. For deglycosylation of intact rgp120, 10 mg of RCM-rgp120 was used. The lyophilized sample was reconstituted in 1 mL of 50 mM sodium phosphate buffer containing 0.2% sodium azide, pH 7.0. Fifty milliunits of PNGase F was added, and the solution was incubated at 37 °C for 24 h. The reaction mixture was run on a Bio-Gel P6 column (1 × 110 cm) equilibrated in 0.1 M ammonium bicarbonate buffer to separate protein and released carbohydrate. For protein, the fractions were monitored by the BCA protein assay; for sugar determination, the fractions were monitored by the phenol–sulfuric acid assay (Dubois et al., 1956).

Peptide Identification. Peptides collected from reversed-phase HPLC were identified by PTC amino acid analysis and/or *N*-terminal sequence analysis. Peptides for amino acid analyses were hydrolyzed by gas-phase hydrolysis. The samples were dried down in 15- × 60-mm Pyrex tubes and were then set inside a larger 50-mL Pyrex tube (2.5 × 10 cm). About 0.5 mL of 6 N constant-boiling HCl containing 2% phenol was added to the bottom of the larger tube, and the tube was flushed with argon before being capped. The tube was then placed in a 165 °C oven for 1 h. The PTC amino acid analysis was performed as described by Gimenez-Gallego and Thomas (1987). *N*-Terminal sequencing was carried out at the Molecular Genetics Instrumentation Facility at the University of Georgia or Oklahoma Molecular Biology Facility on an Applied Biosystems 470A protein sequencer utilizing Edman degradation with an on-line Applied Biosystems PTH analyzer 120A.

Oligosaccharide Mapping on HPAEC. Glycopeptides (about 5 nmol) were treated with PNGase F as described earlier. After treatment, the oligosaccharide mapping was done on Dionex HPAEC. The samples were directly injected into a Carbopac PA1 column (4 × 250 mm) preequilibrated in 24 mM sodium acetate in 100 mM sodium hydroxide. The same solvent was run for 25 min to separate the high mannose-type mixtures. Postcolumn addition of 300 mM sodium hydroxide at 0.3 mL/min was used to increase the detection. The flow rate was 1 mL/min at room temperature, and the separation was monitored with a pulsed amperometric detector.

^{1}H-NMR Spectroscopic Analysis of Con A-Bound Glycopeptides. Tryptic glycopeptides, obtained after CNBr proteolytic cleavage of recombinant gp120, isolated by Con A chromatography, and purified by reversed-phase HPLC, were analyzed by ^{1}H-NMR spectroscopy at 500 MHz. The samples (typically 15–20 nmol) were repeatedly dissolved in D_2O (Cambridge Isotope Laboratories; 99.96 atom % D) at room temperature and pD 7, with intermediate lyophilization. Prior to ^{1}H-NMR spectroscopic analysis, the samples were redissolved in 0.5 mL of D_2O (Cambridge Isotope Laboratories; 99.99 atom % ^{2}H) and transferred into 5-mm NMR tubes (Wilmad; 535-PP). ^{1}H-NMR spectroscopy was performed on a Bruker AM-500 spectrometer interfaced with an Aspect-3000 computer. The residual HOD resonance was suppressed by low-power presaturation. The probe temperature was kept at 23 °C. Further experimental details have been described (van Halbeek, 1993). Data were processed on an IBM PC with a 486 processor, using Felix for Windows, version 1.01 (BioSym). Chemical shifts (δ) are calibrated by reference to internal acetone (δ 2.225).

RESULTS

Identification of Glycosylation Sites in rgp120. Analysis of Trypsin-Derived Glycopeptides. Recombinant gp120 of the HIV-1 BH8 isolate used in this study was expressed in *Spodoptera frugiperda* (Sf9) cells using the baculovirus vector pVLBSlegt, which contained an ecdysterid UDPglucosyltransferase signal peptide in place of the HIV-1 signal peptide. The vector allows recombinant gp120 to be efficiently secreted from the cells into the medium, where it accumulates to much higher levels than is possible without the signal peptide

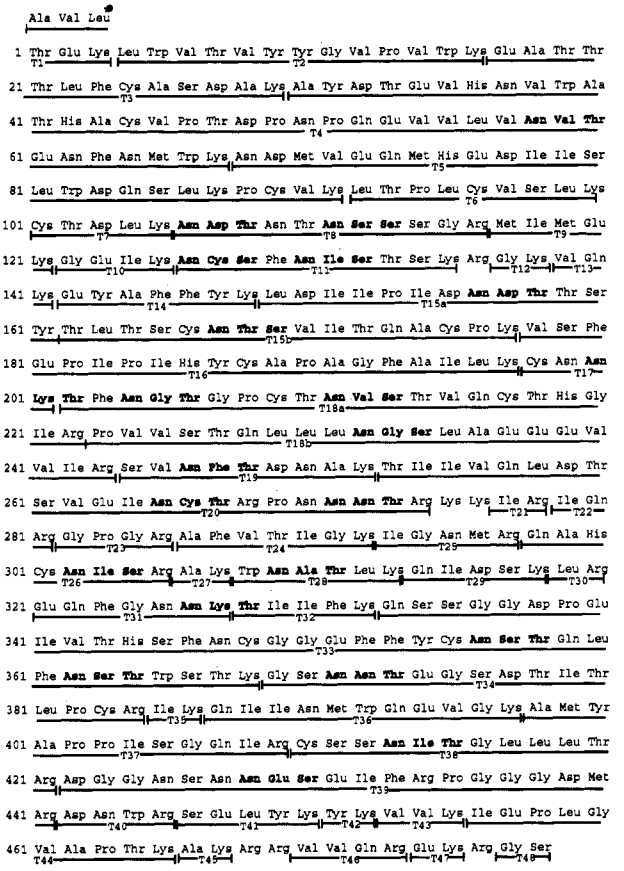


FIGURE 1: Amino acid sequence of mature gp120 from the BH8 isolate of HIV-1 produced in Sf9 cells using a baculovirus vector. The * represents the *N*-terminal segment after cleavage of the signal peptide. Recombinant gp120 was produced by plasmid pVLBSegt120 encoding the signal peptide of baculovirus ecdysteroid UDPglycosyltransferase. The letter T refers to the peptides predicted to be generated by trypsin digestion, and the tryptic peptides are numbered sequentially from the *N*-terminal end of the molecule. The lowercase letters following the T number indicate unexpected proteolytic cleavage. The potential *N*-glycosylation sequences (Asn-Xaa-Ser/Thr) are highlighted in bold type.

(Murphy et al., 1993). The recombinant gp120 produced using this vector contains 481 amino acid residues, with the first three amino acid residues derived from the signal sequence of UDPglucosyltransferase and the last two residues derived from the vector. The predicted sequence of the recombinant glycoprotein is shown in Figure 1, with the 22 potential *N*-linked glycosylation sites (Asn-Xaa-Ser/Thr) (Kornfeld & Kornfeld, 1985) marked in bold.

It has been reported that baculovirus recombinant human plasminogen produced in *S. frugiperda* cells contained complex-type *N*-linked oligosaccharides composed of mannose, fucose, *N*-acetylglucosamine, galactose, and sialic acid (Davidson et al., 1990). However, carbohydrate compositional analysis of rgp120 (Table I) revealed that *N*-acetylglucosamine and mannose were the major monosaccharides present in baculovirus recombinant gp120. These results suggested that rgp120 contained only high mannose-type oligosaccharides. On the basis of this prediction, we decided to use concanavalin A agglutinin, a plant lectin which binds to high mannose-type and hybrid-type oligosaccharides with high affinity (Kornfeld et al., 1981), to isolate the glycopeptides from the total tryptic digests. The experimental approach we took to study the site-specific glycosylation of the baculovirus-derived HIV gp120 is shown in Figure 2.

The baculovirus recombinant gp120 (50 mg) was first reduced and *S*-carboxymethylated with iodoacetic acid and

Table I: Carbohydrate Composition of Baculovirus Recombinant gp120

sugar residues	nmol/100 μg gp120[a]	sugar residues	nmol/100 μg gp120[a]
fucose	t[b]	galactose	n.d.
galactosamine	n.d.[c]	glucose	n.d.
glucosamine	9.0	mannose	37.9

[a] Recombinant gp120 (125 μg of protein based on the BCA protein assay) was hydrolyzed with 2 N TFA at 100 °C for 4 h, and the released monosaccharides were determined by Dionex-HPAEC as described in Experimental Procedures. [b] Trace amount. [c] Not detectable.

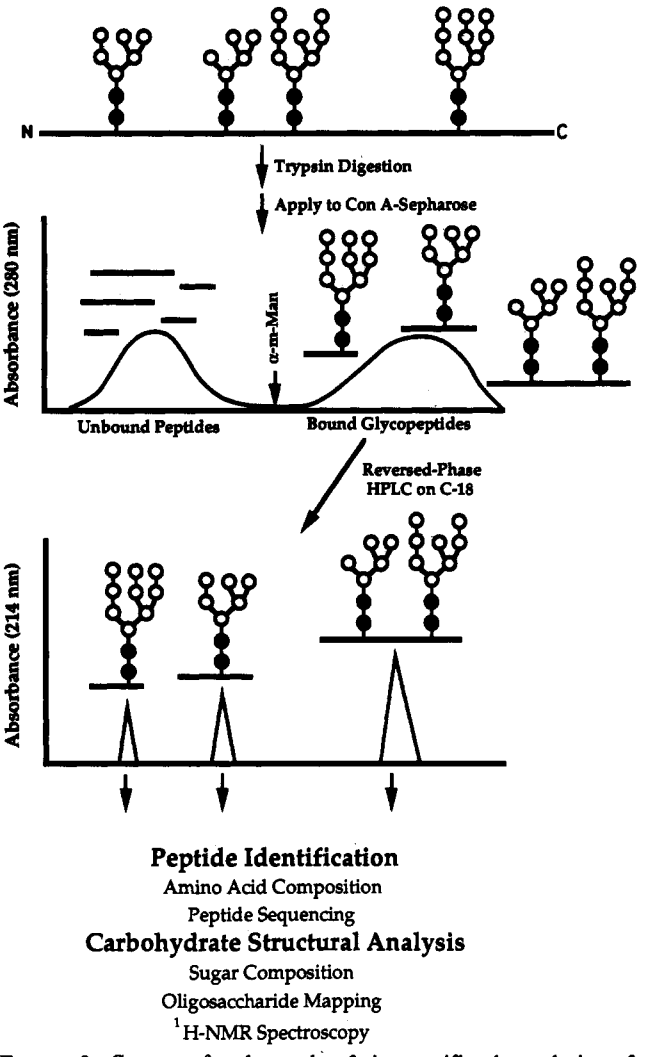


FIGURE 2: Strategy for the study of site-specific glycosylation of recombinant baculovirus-derived HIV-1 gp120. Con A–Sepharose was used to isolate the glycopeptides from the total tryptic digests. The Con A-bound glycopeptides were separated on reversed-phase HPLC and identified by amino acid analysis and/or peptide sequencing. The carbohydrate structures of the glycopeptides were analyzed by sugar compositional analysis and oligosaccharide mapping on HPAEC and by ^{1}H-NMR spectroscopy. Detailed procedures are described in Experimental Procedures.

then digested with trypsin. The glycopeptides were separated from nonglycosylated peptides by chromatography over a column of Con A–Sepharose (Figure 3). Because of the high-affinity interaction between Con A and the glycopeptides, we were unable to elute the glycopeptides at room temperature even with 500 mM α-methylmannoside or with organic solvents such as acetonitrile or ethanol. However, the glycopeptides were eluted when the column was boiled in the presence of 200 mM α-methylmannoside for 10 min. The Con A–Sepharose-bound and -unbound fractions were pooled sepa-

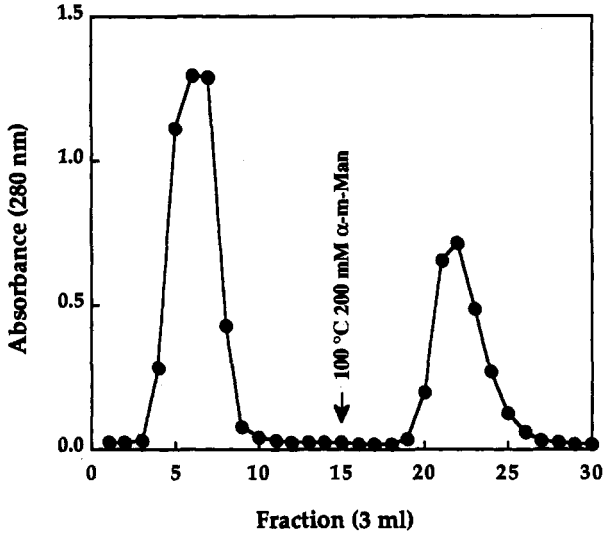


FIGURE 3: Con A–Sepharose profile of trypsin-digested RCM-rgp120. The total tryptic digests of RCM-rgp120 were applied onto a column of Con A-Sepharose. The bound glycopeptides were eluted by boiling the column in the presence of 200 mM α-methylmannoside. The fractions were monitored for absorbance at 280 nm. The bound and unbound fractions were pooled and desalted by a Bio-Gel P2 column.

rately and desalted on a Bio-Gel P2 column. By assaying the desalted material bound and not bound by Con A–Sepharose with the phenol–sulfuric acid assay, we found that 98% of the hexose-containing oligosaccharides were recovered in fractions bound by Con A–Sepharose (data not shown). The Con A–Sepharose-bound and -unbound fractions were then subfractionated on reversed-phase HPLC. As shown in Figure 4, the HPLC pattern of the total tryptic glycopeptides and peptides prior to chromatography on Con A–Sepharose was complex. However, the material bound by Con A–Sepharose had a much simpler pattern. The peaks in the HPLC profile of the Con A-bound fraction were generally broad and appeared as multiplets; this behavior is typical of all the glycopeptides on reversed-phase HPLC (Figure 4C). By contrast, the peaks in the profile of the Con A-unbound fraction were relatively sharp (Figure 4B).

The Con A-bound glycopeptides separated by reversed-phase HPLC were collected as pools of peaks with similar retention times, as indicated in Figure 4C. Each glycopeptide pool was treated with PNGase F, an enzyme that specifically cleaves *N*-linked oligosaccharides from their peptides and results in those asparagines in sugar linkage being converted to aspartic acid residues (Tarentino et al., 1985). The PNGase F-treated glycopeptides were rerun on reversed-phase HPLC under the same conditions. Treatment of glycopeptides with PNGase F resulted in sharpened peaks and increased the retention times on HPLC as a result of the removal of the carbohydrate moiety from the glycopeptides. An example of HPLC analysis after PNGase F treatment is shown in Figure 5 for the glycopeptides designated T19, T15, and T38. For some glycopeptides, it was necessary to use a shallower gradient to separate the released peptide from the glycopeptide. In all cases, the purified peptides were then identified by PTC amino acid analysis. Because of the distinct amino acid composition of each tryptic peptide, we were able to identify most of the peptides by amino acid analysis (Table II).

Products of some unexpected proteolytic cleavages (T15a and T18a) were also observed in the Con A-bound fraction. However, these unexpected cleavage products represented less than 5% of their original tryptic peptides, as calculated from the amino acid composition analyses. *N*-Terminal sequence

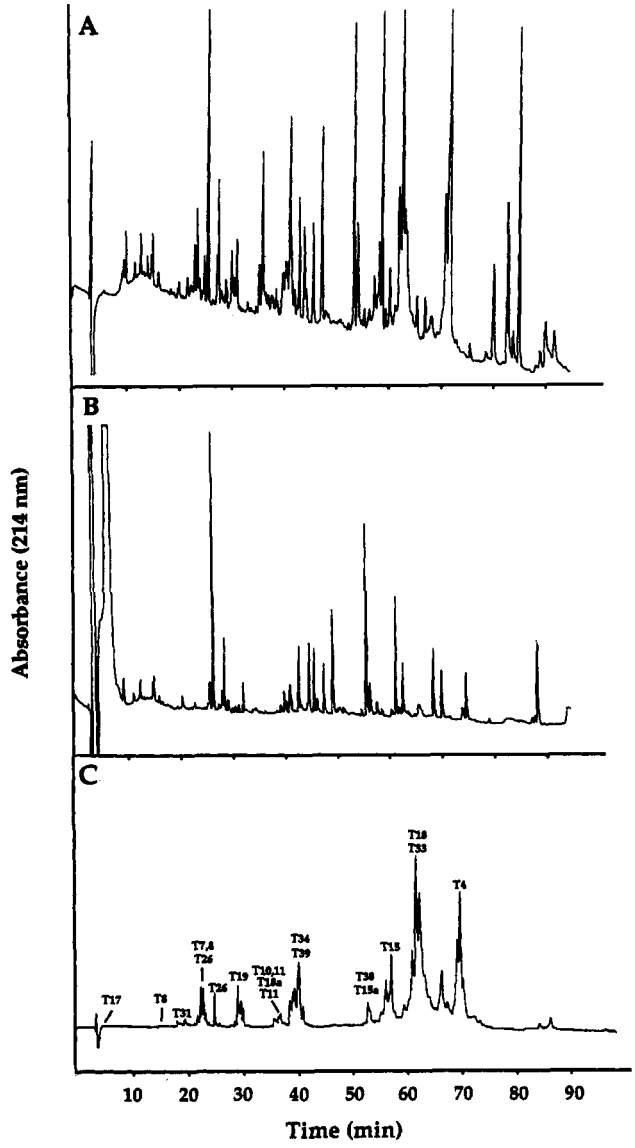


FIGURE 4: Reversed-phase HPLC profiles of trypsin-digested RCM-rgp120. (A) Total tryptic digests of RCM-rgp120; (B) tryptic peptides not bound by Con A–Sepharose; (C) tryptic glycopeptides bound by Con A–Sepharose. The tryptic glycopeptides in C were isolated and identified as described in the text and in Experimental Procedures by amino acid analysis and/or *N*-terminal sequencing following deglycosylation with PNGase F. Identified peptides are labeled according to the nomenclature given in Figure 1. Two T numbers separated by a comma indicate peptides containing uncleaved tryptic sites.

analysis was done on some of the peptides containing more than one potential *N*-glycosylation site.

Analysis of CNBr-Derived Glycopeptides from rgp120. The trypsin-derived glycopeptides identified from the Con A-bound fractions (Table II) accounted for 19 of the 22 potential glycosylation sites. The sites we were unable to identify were located on two predicted tryptic glycopeptides, T20 (Asn265 and Asn271) and T28 (Asn309). We considered the possibility that these sites were glycosylated but were missed in our analysis because of the complexity of the HPLC profile. To determine whether these three potential glycosylation sites were utilized, we took a different approach. As indicated in Figure 1, there are nine methionine residues in gp120 from HIV-1 BH8 isolate (Figure 1). Treatment of rgp120 with cyanogen bromide, which cleaves specifically at the C-terminus of methionine residues, should generate fragments containing various numbers of potential glycosylation sites (Figure 6). The largest predicted fragment

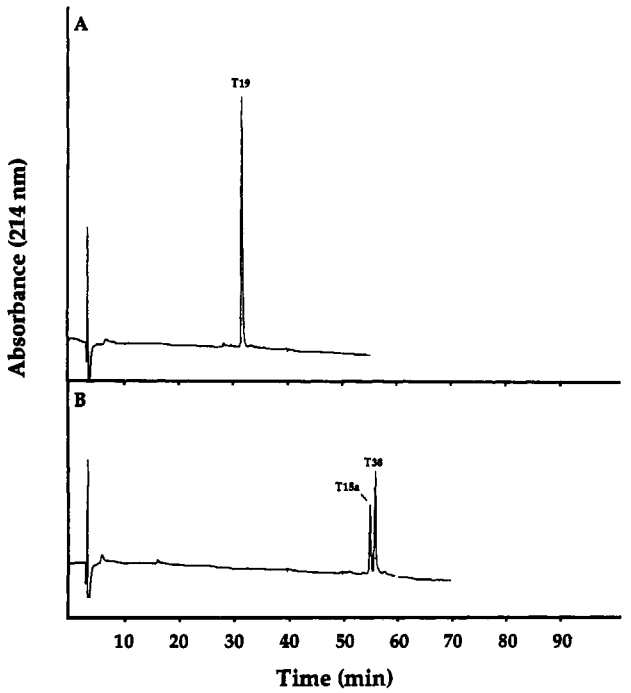


FIGURE 5: PNGase F treatment of tryptic glycopeptides bound by Con A–Sepharose, monitored by reversed-phase HPLC. The glycopeptide pools from Con A-bound glycopeptides shown in Figure 4C were treated with PNGase F and run on reversed-phase HPLC under conditions that allowed separation of the released peptide from the untreated glycopeptide. The sharpened peaks after PNGase F treatment were identified by amino acid analysis and/or peptide sequencing. (A) Peak generated after PNGase F treatment of the glycopeptide pool, with retention time around 30 min in Figure 4C. The peak was identified as T19. (B) Peaks generated after PNGase F treatment of the glycopeptide pool, with retention time around 53 min in Figure 4C. These two peaks were identified as T38 and T15a, respectively.

contained 11 potential sites, including two of the sites, Asn265 and Asn271, that we were unable to identify from analysis of tryptic glycopeptides. The second largest fragment contained six sites, including the third unidentified site as Asn309 (Figure 6).

The reduced and *S*-carboxymethylated rgp120 was treated with CNBr, as described in the Experimental Procedures, and the CNBr fragments were separated by FPLC on a Superose 12 (prep grade) column. The chromatography separated the CNBr fragments into five peaks according to size (Figure 7). To remove cross-contamination from nearby peaks, each peak was rerun over the same column under the same conditions and analyzed on SDS–PAGE to check the purity. The first peak, which was eluted off the column in the void volume, appeared to be a contaminant generated during the reaction; we were unable to detect any peptide in this fraction on SDS–PAGE (data not shown). The second, third, and fourth peaks appeared as broad bands on SDS–PAGE with sizes close to 40, 30, and 15 kDa, respectively. The fifth peak appeared on SDS–PAGE as a broad, faint band with molecular weight less than 10 kDa and probably represented the predicted small peptide products lacking carbohydrate (Figure 6).

We designated the second, third, and fourth peaks as CNBr I, II, and III, respectively (Figure 7). From the apparent molecular weights, we predicted that CNBr I represented the largest fragment and contained 11 *N*-glycosylation sites; CNBr II was the second largest fragment and contained six glycosylation sites; CNBr III appeared to be a combination of three fragments which covered the remaining four sites.

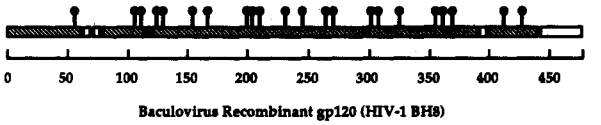


FIGURE 6: Predicted fragments generated by CNBr treatment. Hatched bars represent the CNBr-generated fragments which contain potential *N*-glycosylation sites. The Asn residues in the potential *N*-glycosylation sites are represented across the top of the figure.

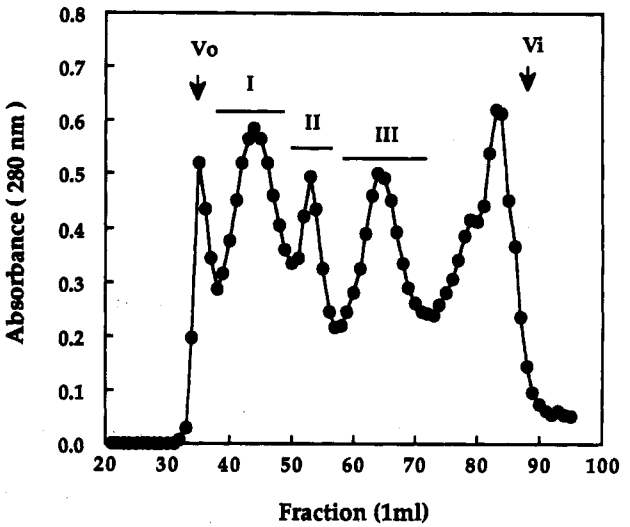


FIGURE 7: Superose 12 profile of CNBr-treated RCM-rgp120. CNBr-treated RCN-rgp120 (50 mg) was separated by FPLC on a Superose 12 (prep grade) column as described in the Experimental Procedures. The fractions were monitored for absorbance at 280 nm. The second, third, and fourth peaks were labeled as CNBr I, II, and III, respectively.

The same procedure described above for analysis of the direct trypsin-derived glycopeptides was used to analyze the CNBr-derived fragments purified from FPLC. CNBr I, II, and III were digested with trypsin, and the resultant glycopeptides were isolated by chromatography on Con A–Sepharose. The bound glycopeptides were fractionated on reversed-phase HPLC (Figure 8). The peaks from each HPLC profile were directly analyzed by PTC amino acid analysis. The Con A-bound glycopeptides identified from each tryptic fragment were separated on reversed-phase HPLC with retention times similar to those from the HPLC profile of the total Con A-bound glycopeptides. T4 and T39, both containing a methionine within their sequences, were identified in CNBr III and were designated as T4-CB and T39-CB to represent the proteolytic products of both cyanogen bromide and trypsin. Homoserine residues were observed in the composition of these two glycopeptides as expected, since cyanogen bromide treatment converts the methionine residue to a homoserine residue as it cleaves the peptide bond. T20, which contained two of the unidentified glycosylation sites, was identified in the Con A–Sepharose-bound tryptic CNBr I, as we predicted from the amino acid sequence (Table II). This glycopeptide appeared as a very small peak on the HPLC profile, probably because the peptide lacked aromatic amino acid. T28 was identified in the Con A-bound tryptic CNBr II, as expected (Table II).

From the analysis of the Con A-bound total tryptic digests and Con A-bound tryptic CNBr fragments, we identified glycopeptides covering all the potential glycosylation sites. Nine of these glycopeptides (T4, T17, T19, T26, T28, T31, T34, T38, and T39), identified in the Con A-bound fraction, contained only a single glycosylation site and thus could be characterized unambiguously with regard to their binding to the Con A–Sepharose.

Different approaches were taken to characterize the peptides containing more than one glycosylation site. To determine whether both potential glycosylation sites were utilized in those peptides containing multiple predicted glycosylation sites, following treatment with PNGase F the peptides were subjected to *N*-terminal Edman sequencing. As noted above, successful cleavage of an *N*-glycosylated asparagine with PNGase F converts the asparagine residue to aspartic acid (Tarentino et al., 1985). Upon sequencing the peptide, aspartic acid residues are observed on the cycles where the *N*-glycosylated asparagine residues used to be located. Using this approach, we found that sites Asn106 and Asn111 in T7,8; Asn126 and Asn130 in T11; and Asn265 and Asn271 in T20 were glycosylated. For Asn126, which was located at the *N*-terminal end of T11, we found a blank instead of aspartic acid residue because PNGase F will not cleave the oligosaccharides attached to either the amino- or the carboxyl-terminal end of peptides (Tarentino et al., 1985).

T15 contains two glycosylation sites, Asn156 and As167. Since there was a tyrosine residue located between these two sites, we treated T15 with chymotrypsin and separated the products on reversed-phase HPLC. We were able to identify the chymotrypsin-generated glycopeptides and confirmed that both Asn156 and Asn 167 were glycosylated on the basis of their susceptibility to PNGase F (data not shown). Residues Asn356 and Asn362, located near the carboxyl-terminal end of T33, were difficult to determine directly by Edman degradation, since there were 23 amino acids preceding the first glycosylation site and 29 amino acids preceding the second sites. We therefore treated T33 with V8 protease at conditions which cleave only at glutamic acid residues to generate a peptide with the *N*-glycosylation sites closer to the amino-terminal end. The V8-treated glycopeptide was isolated by reversed-phase HPLC. By *N*-terminal sequencing of the PNGase F-treated glycopeptide (data not shown), both glycosylation sites were confirmed to be utilized.

T18, which contained three potential *N*-glycosylation sites (Asn204, Asn211, and Asn232), was treated with PNGase F, and the released peptide was isolated by HPLC. Direct *N*-terminal sequencing of this resultant peptide demonstrated that the first two sites (Asn204 and Asn211) were glycosylated, and aspartic acid residues appeared instead of asparagine residues at Edman sequencing cycles 3 and 10. To access whether glycosylation occurred also at the third site, the deglycosylated peptide was treated with V8 protease in phosphate buffer, which is a condition permitting cleavage at positions *C*-terminal of both aspartic acid and glutamic acid residues. We reasoned that if the third site (Asn232) were glycosylated, PNGase F treatment would convert the sugar-attached asparagine to aspartic acid, and thus this site would become susceptible to V8 protease. Upon analysis of the released peptides from reversed-phase HPLC of the V8-digested, deglycosylated T18, we were able to identify peptides corresponding to amino acid residues Val212 through Asp232

Table II: Amino Acid Composition of Identified Peptides from the PNGase F-Treated Con A–Sepharose-Bound Tryptic Glycopeptides[a]

peptides	T17	T8	T31	T7,8[b]	T26	T19	T11[b]	T10,11	T18a[b]	T34
Asx	1.9(2)	3.9(4)	1.6(2)	5.3(5)	1.2(1)	3.0(3)	3.7(2)	1.8(2)	2.0(2)	2.6(3)
Glx	0.0	0.0	2.3(2)	0.0	1.2(1)	0.0	0.0	0.9(1)	1.2(1)	1.1(1)
CMC	0.5(1)	0.0	0.0	0.6(1)	0.8(1)	0.0	1.2(1)	0.3(1)	0.8(2)	0.8(1)
Ser	0.0	2.9(3)	0.0	3.3(3)	1.1(1)	1.0(1)	4.2(3)	3.0(3)	1.6(1)	1.8(2)
Gly	0.0	1.4(1)	1.2(1)	1.1(1)	0.2	0.0	0.4	1.5(1)	1.9(3)	1.9(2)
His	0.0	0.0	0.0	0.0	0.8(1)	0.0	0.0	0.0	0.2(1)	0.0
Arg	0.0	1.0(1)	0.0	1.2(1)	1.0(1)	0.0	0.0	0.0	0.2(1)	1.0(1)
Thr	0.0	2.7(2)	0.0	3.3(3)	0.2	0.9(1)	1.1(1)	1.0(1)	3.1(6)	2.1(3)
Ala	0.0	0.0	0.0	0.0	0.9(1)	1.2(1)	0.0	0.0	0.0	0.0
Pro	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3(1)	0.8(1)
Tyr	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Val	0.0	0.0	0.0	0.0	0.0	1.0(1)	0.0	0.0	1.6(1)	0.0
Met	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ile	0.0	0.0	0.0	0.0	1.0(1)	0.0	1.2(1)	1.5(2)	1.1(1)	0.9(1)
Leu	0.0	0.0	1.5(1)	0.0	0.0	0.0	0.0	0.0	1.0	0.8(1)
Phe	0.0	0.0	1.1(1)	0.0	0.0	0.9(1)	1.8(1)	1.9(1)	0.9(1)	0.0
Lys	1.0(1)	0.0	1.0(1)	1.0(1)	0.0	1.0(1)	1.0(1)	2.8(2)	0.5	0.0
Trp	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

peptides	T39	T15a[b]	T38	T15	T18	T33	T4	T20[b,c]	T28[d]
Asx	5.0(5)	4.0(4)	2.0(1)	3.5(5)	3.1(3)	3.6(4)	7.0(7)	4.0(5)	0.7(1)
Glx	2.5(2)	0.0	0.0	1.5(1)	5.6(5)	4.5(4)	4.8(4)	2.5(2)	0.2
CMC	0.4	0.0	0.6(1)	0.9(2)	0.1(2)	0.7(2)	1.0(1)	0.3(1)	0.0
Ser	3.0(2)	2.1(1)	2.0(2)	3.6(3)	3.2(3)	4.0(6)	0.0	2.2(1)	0.2
Gly	5.6(5)	1.0	1.3(1)	0.3	4.2(4)	4.0(4)	0.0	0.9	0.2
His	0.0	0.0	0.0	0.1	0.9(1)	0.9(1)	1.6(2)	0.0	0.0
Arg	2.0(2)	0.6	1.0(1)	0.7	2.0(2)	0.9	1.7	2.0(2)	0.0
Thr	0.7	3.4(2)	2.4(2)	6.2(6)	5.3(6)	4.4(4)	3.8(4)	4.2(4)	0.8(1)
Ala	0.5	0.0	0.0	1.3(1)	1.1(1)	0.0	2.5(3)	0.6	0.9(1)
Pro	1.0(1)	0.4(1)	0.0	2.5(2)	1.7(2)	1.1(1)	4.0(3)	1.4(1)	0.1
Tyr	0.0	0.8(1)	0.0	1.1(1)	0.0	3.1(1)	1.0(1)	0.0	0.0
Val	0.0	0.0	0.0	1.4(1)	4.9(6)	2.5(1)	6.3(7)	2.3(2)	0.0
Met	0.3(1)	0.0	0.0	0.0	0.0	0.0	0.8(1)	0.0	0.0
Ile	1.6(1)	2.7(3)	2.3(1)	3.9(4)	2.9(2)	2.1(1)	1.0	3.5(3)	0.0
Leu	0.0	2.6(1)	3.8(3)	2.4(2)	2.4(4)	2.0(1)	1.1(1)	1.6(1)	0.7(1)
Phe	1.6(1)	0.3	0.0	1.1(1)	2.0(1)	1.9(4)	1.1(1)	0.4	0.0
Lys	0.0	0.8	0.0	1.0(1)	1.0(1)	0.6(1)	0.5(1)	0.0	1.0(1)
Trp	0.0	0.0	0.0	0.0	0.0	0.0(1)	0.0(2)	0.0	0.0(1)

[a] Peptides are listed according to elution position in the HPLC profile of the total mixture of tryptic Con A-bound glycoproteins except where noted. Theoretical values predicted for peptides on the basis of Figure 1 are in parentheses. [b] Confirmed by peptide sequencing. [c] Identified in Con A-bound tryptic CNBr I. [d] Identified in Con A-bound tryptic CNBR II.

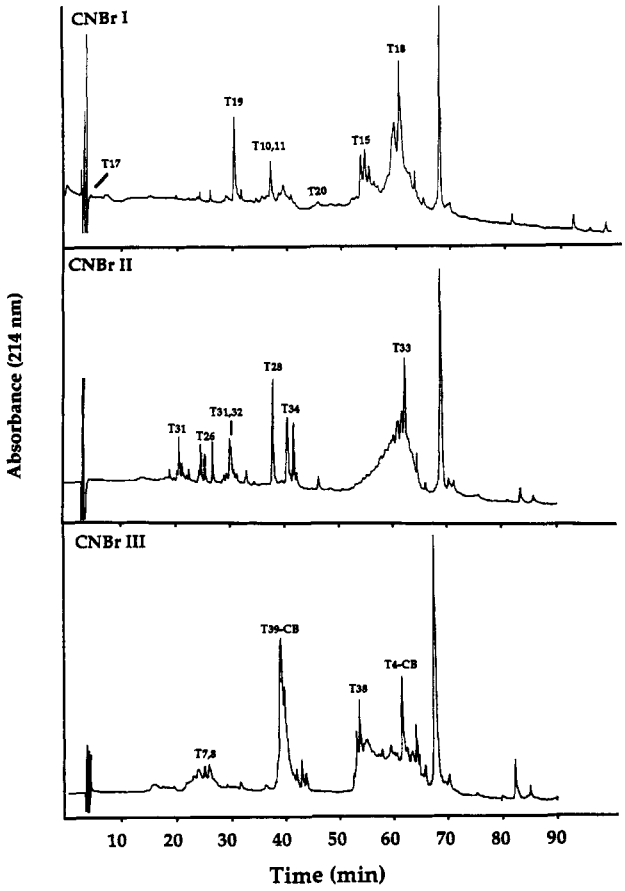


FIGURE 8: Reversed-phase HPLC profiles of Con A-bound glycopeptides produced by trypsin treatment of CNBr-derived glycopeptides I, II, and III. CNBr-derived glycopeptides I, II, and III (from Figure 7) were treated with trypsin, and the glycopeptides were isolated by their binding to Con A–Sepharose. (A) CNBr I; (B) CNBr II; and (C) CNBr III. Glycopeptides were separated on reversed-phase HPLC, as described in the Experimental Procedures.

and Gly233 through Glu237. In addition, all other possible peptides were identified. The cleavage between Asp232 and Gly233, which occurred following PNGase F treatment, indicates that Asn232 is also *N*-glycosylated. These results demonstrate that all three sites in T18 were glycosylated.

Taken together, the results of analysis of trypsin-derived and CNBr-derived glycopeptides indicate that all 22 potential *N*-glycosylation sites in rgp120 are glycosylated with PNGase F-sensitive oligosaccharides.

Structural Analysis of Oligosaccharides in the Glycopeptides. We performed a series of analyses to characterize the oligosaccharide structures in each glycopeptide. These analyses included monosaccharide compositional analysis, oligosaccharide mapping on the Dionex-HPAEC on all of the glycopeptides, and ^{1}H-NMR spectroscopy on some representative glycopeptides. The Con A-Sepharose-bound glycopeptides from tryptic CNBr I, II, and III were used for oligosaccharide structural analysis because these glycopeptides were well separated from each other on reversed-phase HPLC. T15 was treated with chymotrypsin to obtain T15a (Asn156) and T15b (Asn167) for the analysis. For compositional analysis, each glycopeptide was hydrolyzed with 2 N TFA at 100 °C for 4 h, and the hydrolysate was analyzed on a Dionex HPAEC with a CarboPac PA1 column, and eluting monosaccharides were detected with a pulsed-amperometric detector system (PAD). Soybean lectin, which has high mannose-type *N*-linked oligosaccharides with the structure of $GlcNAc_2Man_9$ (Dorland et al., 1981), served as a standard to normalize the average number of mannose residues per two *N*-acetylglucosamine residues for each glycopeptide. Most of the samples were analyzed in duplicate, depending on the availability of the samples.

As we observed from the carbohydrate composition of the intact rgp120, *N*-acetylglucosamine and mannose were the major monosaccharides found in all the glycopeptides analyzed. The average number of mannose residues per two GlcNAc residues on the glycopeptides we analyzed ranged from 6 to 9, after normalization based on the compositional results obtained for purified soybean lectin (Table III). These results suggest that each glycopeptide contains a high mannose-type *N*-linked oligosaccharide, and the approximate predicted size range was $GlcNAc_2Man_5$ to $GlcNAc_2Man_9$.

Analysis of individual glycopeptides by ^{1}H-NMR spectroscopy gave results consistent with this prediction. No structure smaller than $GlcNAc_2Man_5$ was found from the analysis of eight glycopeptides by ^{1}H-NMR spectroscopy. In addition, oligosaccharide mapping of PNGase F-released oligosaccharides from each glycopeptide was performed, and the results demonstrated that $GlcNAc_2Man_5$ to $GlcNAc_2Man_9$ represented the major structures on these glycopeptides (Table III).

As is clear from the results shown in Table III, some glycopeptides contained primarily small oligosaccharides of the size near $GlcNAc_2Man_5$, while other glycopeptides contained larger oligosaccharides of the size near $GlcNAc_2Man_9$. The data are presented as the percentage of Man_5–Man_9 structures of each glycopeptide, determined by the peak area of each structure following HPAE chromatography and detection with the PAD system. It should be pointed out that comparisons of the peak areas for each oligosaccharide might not directly reflect the authentic molar ratio of each high mannose-type structure. Defined molar amounts of these oligosaccharides are not available for standardization of the detector response. Nevertheless, these relative ratios are probably close to the actual ratios, based on the results predicted by both compositional and ^{1}H-NMR analyses (see Table III).

To aid in the interpretation of these data, we considered that the higher sized oligosaccharides of $GlcNAc_2Man_8$ and $GlcNAc_2Man_9$ represented relatively "unprocessed" oligosaccharides, whereas $GlcNAc_2Man_5$, $GlcNAc_2Man_6$, and $GlcNAc_2Man_7$ represented relatively more "processed" oligosaccharides. We therefore analyzed each glycopeptide for its so-called "processing index" by taking the ratio of the amount of processed oligosaccharides to the amount of unprocessed oligosaccharides. Comparison of the processing index allowed us to classify the *N*-glycosylation sites into three types: those with highly processed sites (H), having a processing index higher than 0.64; those with partially processed sites (P), having a processing index around 0.4; and those with slightly processed sites (S), having a processing index lower than 0.15 (Table III).

For the 11 tryptic glycopeptides that contained a single glycosylation site, the classification of oligosaccharide processing was unambiguous. However, for the tryptic glycopeptides that contained multiple glycosylation sites, determination of the processing index and classification required consideration of multiple lines of evidence. For example, T8, which contained two sites, had a processing index of 0.85 and a low mannose number in composition. These indicate that both sites in this glycopeptide contained highly processed oligosaccharides. However, T33, which also contained two sites, had a high mannose content and a low processing index.

Table III: Carbohydrate Analysis of Con A-Bound Glycopeptides from Tryptic CNBr I, II, and III

glycopeptides	site	composition analysis: no. of $GlcNH_2$ residues	composition analysis: no.[a] of Man residues	oligosaccharide mapping (%): Man_5	Man_6	Man_7	Man_8	Man_9	processing index: $(Man_5 + Man_6 + Man_7)/(Man_8 + Man_9)$	type[b]
T4-CB	58	2	7.5 ± 1.0	3	5	20	38	34	0.39	P
T7,8[c]	106, 111	2	6.4 ± 0.2	8	11	27	38	16	0.85	2H
T10,11	126, 130	2	6.5	4	7	21	36	32	0.47	2P
T15a	156	2	6.8	7	10	27	35	21	0.79	H
T15b	167	2	7.9	3	4	22	36	35	0.41	P
T17	200	2	8.4 ± 0.1	1	1	9	38	51	0.12	S
T18	204, 211, 232	2	8.1 ± 0.1	3	4	17	36	40	0.32	S + 2P
T19[c]	246	2	6.2 ± 0.4	4	11	24	36	35	0.64	H
T20	265, 271	2	7.7 ± 0.8	0	2	18	30	50	0.25	P + H
T26[c]	302	2	8.7 ± 0.1	0	2	6	36	56	0.09	S
T28[c]	309	2	8.9 ± 0.1	0	6	7	31	56	0.15	S
T31[c]	326	2	8.2	5	10	14	32	39	0.41	P
T33	356, 362	2	8.5 ± 0.3	1	2	10	39	48	0.15	2S
T34[c]	371	2	7.7 ± 0.7	2	11	16	45	26	0.41	P
T38[c]	413	2	7.6	1	3	19	42	35	0.30	P
T39-CB[c]	428	2	6.6 ± 0.8	4	13	29	39	15	0.85	H

[a] All data are averages of duplicate experiments (reported as ±), except for some where only a single analysis was performed. [b] Types of glycosylation site classified by processing index. S, slightly processed, P, partially processed, and H, highly processed. [c] Samples analyzed by ^{1}H-NMR spectroscopy.

We concluded that at both sites the oligosaccharides were slightly processed. For other glycopeptides containing multiple sites, such as T11, T18, and T20, we could deduce the types of oligosaccharides that were presented on each glycopeptides on the basis of the processing index. However, we have as yet been unable to define the oligosaccharide structures as individual sites of each of the multiply glycosylated peptides. The ^{1}H-NMR spectra of glycopeptides from each type are shown in Figure 9.

Examination of rgp120 for the Presence of O-Linked Oligosaccharides. We considered the possibility that in addition to the *N*-linked oligosaccharide defined above, rgp120 might also contain *O*-linked oligosaccharides. In a recent study, we demonstrated that gp120 derived from HIV-1-infected Molt-3 human T cells contained *N*-acetylgalactosamine (Merkle et al., 1991). The *O*-linked oligosaccharides are characterized by having a residue of *N*-acetylgalactosamine linked to either Ser or Thr. To assess whether the baculovirus-derived rgp120 contained *O*-linked oligosaccharides, we treated the *S*-carboxymethylated rgp120 with PNGase F to remove *N*-linked oligosaccharides. After treatment, the protein was separated from the released oligosaccharides on a Bio-Gel P6 column (Figure 10). PNGase F treatment was able to remove about 90% of the neutral sugars from the protein. We then performed carbohydrate analysis on the residual glycoprotein in the void volume of the Bio-Gel P6 column, which represented the PNGase F-resistant oligosaccharides on the rgp120. We considered the possibility that the rgp120 might contain only one *O*-linked oligosaccharide and consequently only a single mole of GalNAc per mole of protein. We than analyzed a sufficient amount of protein (2 mg) to allow detection of rgp120 having a single *O*-linked oligosaccharide. However, the carbohydrate composition indicated that the rgp120 recovered after PNGase F treatment contained only mannose and *N*-acetylglucosamine and no *N*-acetylgalactosamine (Table IV). These results demonstrate that rgp120 lacks any *O*-linked oligosaccharide.

DISCUSSION

In this study, we have characterized the sites of glycosylation and the carbohydrate structures of recombinant gp120 of HIV-1 BH8 isolate from the baculovirus expression system. All 22 potential *N*-glycosylation sites on this molecule were utilized and the recombinant glycoprotein lacks *O*-linked oligosaccharides. Each site of glycosylation contained high mannose-type *N*-linked oligosaccharides ranging in size from $GlcNAc_2Man_5$ to $GlcNAc_2Man_9$, but the degree of processing of these oligosaccharides was site specific.

Several previous studies have shown that the *N*-linked oligosaccharides in insect cell-derived glycoproteins are only high mannose-type (Luckow & Summers, 1988) and can be processed or trimmed to trimannosyl cores without further elongation (Butters et al., 1981; Hsieh & Robbins, 1984). However, in some recent studies it was shown that recombinant human plasminogen, expressed using the baculovirus expression vector in two lepidopteran insect cell lines *Spodoptera frugiperda* and *Mamestra brassicae*, contained complex-type oligosaccharides with fucose, galactose, and sialic acid (Davidson et al., 1990; Davidson & Castellino, 1991b). In addition, the amount of complex glycan found was correlated to infection time, and a considerable amount of complex-type oligosaccharides was present by 48 h postinfection (Davidson & Castellino, 1991a). The recombinant HIV-1 gp120 that we used in our study was purified from the culture medium of *S. frugiperda* (Sf9) cells after infection with the baculovirus vector for 48 h, but no complex-type oligosaccharides were found. It is not known whether this difference in results is due to differences in potential of different subclones, different culture conditions, or differences in expression vectors or levels of recombinant protein expression.

The presence of abundant amounts of $GlcNAc_2Man_8$ in the rgp120 in our study indicates that many of the glycosylation sites were acted upon by the ER α-mannosidase, which converts $GlcNAc_2Man_9$ to $GlcNAc_2Man_8$ (Bischoff et al., 1986). Presumably, further processing to $GlcNAc_2Man_5$ species requires the action of a Golgi α-mannosidase I. Our results suggest that different glycosylation sites are differentially acted upon by α-mannosidase I, possibly due to differences in accessibility of the oligosaccharides to the enzyme as a result of conformational factors in the glycoprotein.

Our results can be directly compared to those presented in the only other study on the site-specific glycosylation of a recombinant form of gp120, which was derived from a different isolate of HIV-1 expressed in Chinese hamster ovary cells (Leonard et al., 1990). These authors examined for sites of *N*-glycosylation and found that all 24 potential *N*-glycosylation sites were utilized. Although they did not provide a structural analysis of the oligosaccharides at each site, they did assess

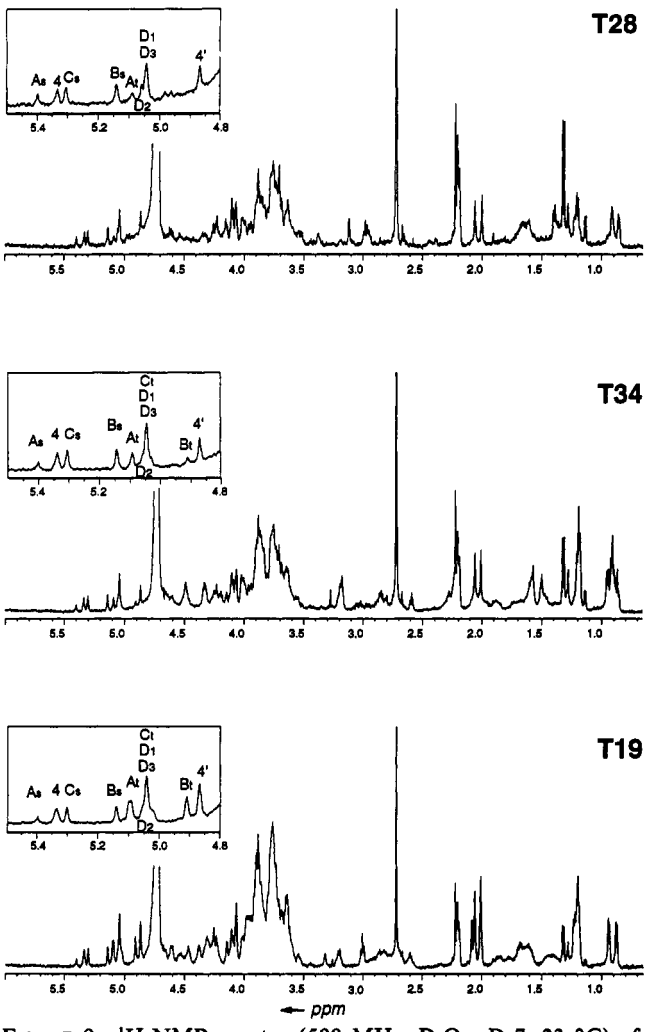
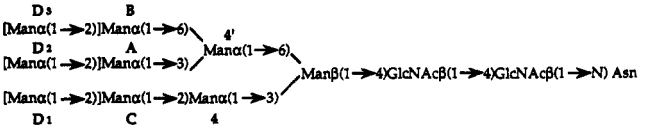


FIGURE 9: ^{1}H-NMR spectra (500 MHz, D_2O; pD 7; 23 °C) of glycopeptides with distinct types of *N*-linked oligosaccharide structures. (Top) T28, "slightly processed" type; (middle) T34, "partially processed" type; and (bottom) T19, "highly processed" type. Each panel shows the overall spectrum and (blown-up in the inset) the anomeric region. The numbers and letters in the spectra refer to the corresponding monosaccharide residues in the structure:



Subscripts t and s refer to terminal and substituted, respectively.

the types of oligosaccharides present by a combination of methods. Some sites were shown to be susceptible to *endo*-β-*N*-acetylglucosaminidase H (Endo H), which indicated the presence of high mannose-/hybrid-type *N*-linked oligosaccharides, while other sites were not susceptible to Endo H but were susceptible to PNGase F, which indicated the presence of complex-type *N*-linked oligosaccharides.

We have compared in Figure 11 the results obtained by Leonard et al. (1990) with our own. In recombinant gp120 derived from CHO cells, most of the high mannose-/hybrid-type oligosaccharides were located in conserved regions. We have observed a similar distribution on the baculovirus-derived rgp120 and have classified the sites into three types according to the degree of oligosaccharide processing. Those sites in the baculovirus-derived rgp120 that contained slightly processed oligosaccharides contained high mannose-/hybrid-type oligosaccharides (Endo H-sensitive) in the recombinant gp120 from CHO cells. In addition, the sites in the baculovirus-

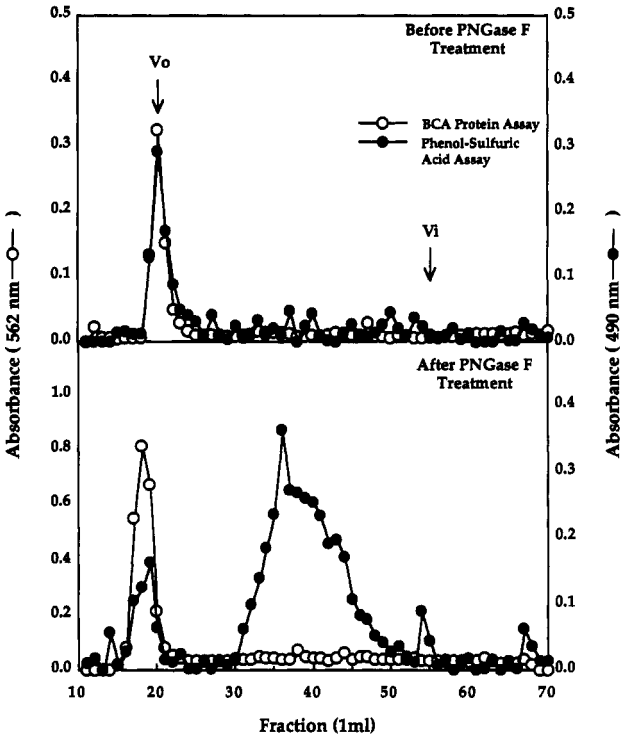


FIGURE 10: Bio-Gel P6 profiles of RCM-rgp120 before and after PNGase F treatment. RCM-rgp120 (2 mg) and PNGase F-treated RCM-rgp120 (10 mg) were applied to a column of Bio-Gel P6 (1 × 100 cm). Fractions were monitored by their protein and sugar content using the BCA protein assay and phenol–sulfuric acid assay, respectively.

Table IV: Carbohydrate Composition Analysis of Baculovirus Recombinant gp120 after PNGase F Treatment

sugar residues	nmol/100 μg gp120[a]	sugar residues	nmol/100 μg gp120[a]
fucose	n.d.[b]	galactose	n.d.
galactosamine	n.d.	glucose	n.d.
glucosamine	0.14	mannose	0.55

[a] Recombinant gp120 (2 mg of protein based on the BCA protein assay) was hydrolyzed with 2 N TFA at 100 °C for 4 h, and the released monosaccharides were determined by Dionex-HPAEC as described in Experimental Procedures. [b] Not detectable.

derived rgp120 which we identified as having highly processed oligosaccharides were found to containing complex-type oligosaccharides in the CHO-derived recombinant gp120. Finally, the sites that we found to contain partially processed oligosaccharides in the baculovirus-derived rgp120 were found to contain either Endo H resistant or Endo-H-susceptible oligosaccharides in their corresponding sites on the CHO recombinant gp120. The sites which we defined as partially processed in the baculovirus-derived gp120 might correspond to the sites in CHO-derived rgp120 that contained either hybrid-type oligosaccharides (Endo H-sensitive) or complex-type oligosaccharides with relatively simple structures (Endo H-resistant). Similar spatial distribution of oligosaccharides has been observed in the hemagglutinin of fowl plague virus derived from vertebrate hosts (Keil et al., 1985) and insect cells (Kuroda et al., 1990). Glycosylation sites found to contain complex-type oligosaccharides in hemagglutinin derived from chicken embryo cells were found to be of the $GlcNAc_2Man_3$ structure with fucose attached to the chitobiose core in hemagglutinin derived from Sf9 cells, while the sites containing high mannose-type structures in chicken-derived hemagglutinin were found to contain less-processed high mannose-type

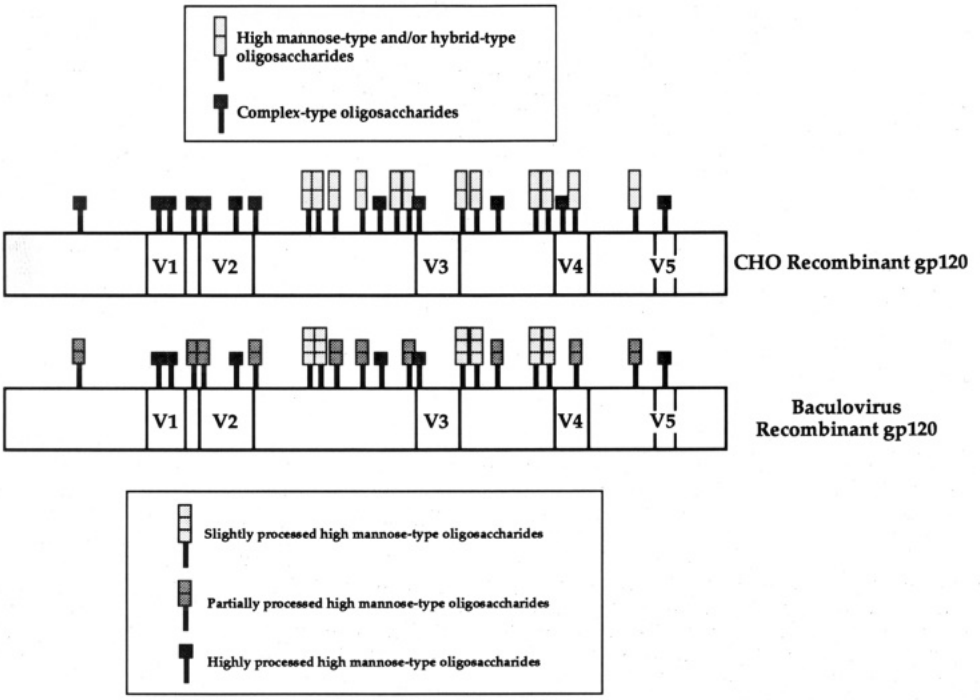


FIGURE 11: Comparison of glycosylation between CHO recombinant gp120 of HIV-III_B isolate and baculovirus recombinant gp120 of HIV-1 BH8 isolate. Glycosylation sites of CHO recombinant gp120 are classified as high mannose-type and/or hybrid-type oligosaccharides or complex-type oligosaccharides on the basis of their susceptibility to Endo H as reported by Leonard et al. (1990). Glycosylation sites of baculovirus-derived gp120 are classified into sites containing "slightly processed", "partially processed", and "highly processed" structures according to their processing index.

oligosaccharides ($GlcNAc_2Man_5$ to $GlcNAc_2Man_9$) in Sf9-derived hemagglutinin.

The envelope glycoprotein of HIV contain regions with pronounced amino acid variability, and these hypervariable domains are interspersed among five conserved regions (Willey et al., 1986; Starcich et al., 1986). For the functional integrity of the protein to be maintained, one might expect that the functional domains will be located in the conserved regions, and so far this appears to be the case. Studies aimed at defining the functional domains of gp120 have shown that several of these conserved regions mediate critical steps in the productive viral infection of human T-cells (Kowalski et al., 1987; Lasky et al., 1987). The "canyon theory" has been proposed concerning the receptor binding sites of the envelope proteins of the retroviruses, based on studies with the three-dimensional structure of the rhinovirus capsid. These studies revealed that receptor recognition occurs within a canyon lined with highly conserved amino acids and a narrow rim consisting of variable residues (Rossman et al., 1985). It has been suggested that by limiting the size of the receptor-binding pocket and exposing the highly variable residues at its surface, the virus can conceal its receptor-binding site from antibody recognition (Rossman et al., 1985). Although the three-dimensional structure of gp120 is not available, the canyon model has been proposed for the receptor-binding site of HIV envelope glycoprotein (Arthos et al., 1989). From the distribution of oligosaccharides in gp120 produced in CHO cells or baculovirus-infected insect cells, we found that the less-processed oligosaccharides were primarily located within the conserved regions (Figure 11). This observation might reflect the phenomenon that the conserved regions of the HIV envelope proteins are less exposed, and oligosaccharides located in these regions are also less susceptible to the carbohydrate processing enzymes. Interestingly, several mannose-binding lectins (Balzarini et al., 1992; Gattegno et al., 1992), but not galactose/*N*-acetylglactosamine-binding or sialic acid-binding lectins (Balzarini et al., 1992), can block viral infectivity *in vitro*. These results suggest the possibility that the oligosaccharides (high mannose-type) bound by these lectins are located at the functionally-important domains.

Several inhibitors of the *N*-linked carbohydrate biosynthetic pathway have been used to study their effects on the virus life cycle and the viral infectivity. Among the compounds tested, inhibitors of α-glucosidases I and II, which include castanospermine and 1-deoxynojirimycin or their alkyl derivatives and inhibitors of α-mannosidase-I, such as 1-deoxymannojirimycin, showed dramatic effects on the virus life cycle and were able to reduce viral infectivity (Gruters et al., 1987; Walker et al., 1987; Montefiori et al., 1988; Pal et al., 1989a). However, swainsonine, which inhibits Golgi α-mannosidase-II, had no effect on viral infectivity (Pal et al., 1989a). These observations indicate that the synthesis of hybrid-type *N*-linked oligosaccharides and/or of their intermediate $GlcNAc_2Man_{5-8}$ structures, is critical for HIV-1 infectivity, and the prevention of the synthesis of complex-type oligosaccharides may have little if any effect on viral infectivity in T-cells. Interestingly, treatments with inhibitors of *N*-linked oligosaccharide biosynthesis do not inhibit envelope protein binding to CD4 (Walker et al., 1987; Pal et al., 1989b); only tunicamycin has been found to result in an envelope protein, whose CD4 binding is abolished (Fennie and Lasky, 1989). The abolishment of receptor-binding by tunicamycin appears to be the result of dramatic alterations in protein conformation due to a complete lack of carbohydrates, rather than the lack of certain types of oligosaccharides.

Although *O*-linked oligosaccharides have been found on other retrovirus envelope glycoproteins (Pinter & Honnen, 1988), the existence of serine/threonine-linked sugar on the HIV-1 envelope glycoproteins has been controversial. Some studies showed the lack of *O*-linked oligosaccharides on either recombinant gp120 or gp160 produced in CHO cells (Kozarsky et al., 1989), while evidences suggesting the existence of *O*-linked oligosaccharides on either recombinant gp120 (Hansen et al., 1992) or gp120 from HIV-1-infected T-cell

lines (Hansen et al., 1990; Stein & Engleman, 1990; Merkle et al., 1991) have also been reported. It has been reported that recombinant proteins produced in Sf9 cells using baculovirus vectors contained simple *O*-linked chains (Luckow & Summers, 1988; Chen et al., 1991). However, our results demonstrate that the baculovirus-derived gp120 of HIV-1 BH8 isolate lacks *O*-linked oligosaccharide. It is conceivable that the soluble form of rgp120 expressed in the baculovirus system might not express a proper conformation for *O*-glycosylation. However, it should be noted that the baculovirus-derived gp120 was properly folded, as evidenced by its high affinity binding to CD4 (Wells & Compans, 1990; Murphy et al., 1990) and its expression of HIV-1 neutralizing epitopes (Rusche et al., 1987). It is possible that *O*-glycosylation of the envelope protein normally occurs at the precleavage stage when the protein is in the gp160 precursor form. Alternatively, differences in *O*-glycosylation might reflect the different glycosylation machineries of specific host cells used in different studies. More detailed and thorough studies need to be done to address this question.

In this paper, we provide a model system for studying the site-specific glycosylation of glycoproteins with multiple glycosylation sites. Using a combination of protein fragmentation, lectin affinity chromatography, and reversed-phase HPLC techniques, we characterized the sites of glycosylation of the recombinant gp120. The structural information we obtained from this study will be useful for the future study of gp120 from the same isolate or of gp120 from isolates with similar sequences. In conjunction with appropriate radiolabeling techniques, investigation of the site-specific glycosylation of gp120 derived from natural sources will also be possible.

ACKNOWLEDGMENT

The authors acknowledge Patricia Wilkins for critical reading of the manuscript and Judy Gaar for assistance in preparing the manuscript. We also thank Dr. Lihua Huang (CCRC, Athens, GA) for recording the ^{1}H-NMR spectra, Jim McIntire and Holly Hodgdon for their help in producing and purifying the rgp120, and Dr. William Canfield for technical assistance on FPLC and helpful discussions.

REFERENCES

Allan, J. S., Coligan, J. E., Barin, F., McLane, M. F., Sodrowski, J. G., Rosen, C. A., Haseltine, W. A., Lee, T. H., & Essex, M. (1985) *Science 228*, 1091–1094.

Arthos, J., Deen, K., Chaikin, M., Fornwald, J., Sathe, G., Sattentau, Q., Clapham, P., Weiss, R., McDougal, J., Pietropaolo, C., Axel, R., Truaneh, A., Maddon, P., & Sweet, R. (1989) *Cell 57*, 469–481.

Balzarini, J., Neyts, J., Schols, D., Hosoya, M., Van Damme, E., Peumans, W., & De Clercq, E. (1992) *Antiviral Res. 18*, 191–207.

Bischoff, J., Liscum, L., & Kornfeld, R. (1986) *J. Biol. Chem. 261*, 4766–4774.

Butters, T. D., Hughes, R. C., & Vischer, P. (1981) *Biochim. Biophys. Acta 640*, 672–686.

Chen, W., Shen, Q. X., & Bahl, O. P. (1991) *J. Biol. Chem. 266*, 4081–4087.

Dalgleish, A. G., Beverely, P. C. L., Clapham, P. R., Crawford, D. H., Greaves, M. F., & Weiss, R. A. (1984) *Nature 312*, 763–767.

Davidson, D. J., & Castellino, F. J. (1991a) *Biochemistry 30*, 6167–6174.

Davidson, D. J., & Castellino, F. J. (1991b) *Biochemistry 30*, 6689–6696.

Davidson, D. J., Fraser, M. J., & Castellino, F. J. (1990) *Biochemistry 29*, 5584–5590.

Dorland, L., van Halbeek, H., Vleigenthart, J. F., Lis, H., & Sharon, N. (1981) *J. Biol. Chem. 256*, 7708–7711.

Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A., & Smith, F. (1956) *Anal. Chem. 28*, 350–360.

Fennie, C., & Lasky, L. A. (1989) *J. Virol. 63*, 639–646.

Gattegno, L., Ramdani, A., Jouault, T., Saffar, L., & Gluckman, J. C. (1992) *AIDS Res. Hum. Retroviruses 8*, 27–37.

Gimenez-Gallego, G., & Thomas, K. A. (1987) *J. Chromatogr. 409*, 299–304.

Gruters, R. A., Neefjes, J. J., Termette, M., de Goede, R. E., Tulp, A., Huisman, H. G., Miedema, F., & Ploegh, H. L. (1987) *Nature 330*, 74–77.

Hansen, J. E. S., Clausen, H., Nielsen, C., Teglbjaerg, L. S., Hansen, L. L., Nielsen, C. M., Dabelsteen, E., Mathiesen, L., Hakomori, S. I., & Nielsen, J. O. (1990) *J. Virol. 64*, 2833–2840.

Hansen, J. E. S., Clausen, H., Hu, S. L., Nielsen, J. O., & Olofsson, S. (1992) *Arch. Virol. 126*, 11–20.

Hsieh, P., & Robbins, P. W. (1984) *J. Biol. Chem. 259*, 2375–2382.

Keil, W., Geyer, R., Dabrowski, J., Niemann, H., Stirm, S., and Klenk, H.-D. (1985) *EMBO J. 4*, 2711–2720.

Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J.-C., & Montagnier, L. (1984) *Nature 312*, 767–768.

Kornfeld, R., & Kornfeld, S. (1985) *Annu. Rev. Biochem. 54*, 631–664.

Kornfeld, K., Reitman, M., & Kornfeld, R. (1981) *J. Biol. Chem. 256*, 6633–6640.

Kowalski, M., Potz, J., Basiripour, L., Dorfman, T., Goh, W. C., Terwilliger, E., Dayton, A., Rosen, C., Haseltine, W., & Sodroski, J. (1987) *Science 237*, 1351–1355.

Kozarsky, K., Penman, M., Basiripour, L., Haseltine, W., Sodroski, J., & Krieger, M. (1989) *J. AIDS 2*, 163–169.

Kuroda, K., Geyer, H., Geyer, R., Doerfer, W., & Klenk, H.-D. (1990) *Virology 174*, 418–429.

Lasky, L. A., Nakamura, G., Smith, J. H., Fennie, L., Shimosaki, E., Patzer, P., Berman, P., Gregory, T., & Capon, D. J. (1987) *J. Virol. 63*, 3595–3600.

Lee, W.-R., Syu, W.-J., Du, B., Matsuda, M., Tan, S., Wolf, A., Essex, M., & Lee, T.-H., (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 2213–2217.

Leonard, C. K., Spellman, M. W., Riddle, L., Harris, R. F., Thomas, G. N., & Gregory, T. J. (1990) *J. Biol. Chem. 265*, 10373–10382.

Lifson, J., Coutre, S. Huang, E., & Engleman, E. (1986) *J. Exp. Med. 164*, 2101–2106.

Luckow, V. A., & Summers, M. D. (1988) *Bio/Technology 6*, 47–55.

Merkle, R. K., Helland, D. E., Welles, J. L., Shilatifard, A., Haseltine, W. A., & Cummings, R. D. (1991) *Arch. Biochem. Biophys. 290*, 248–257.

Modrow, S., Hahn, B. H., Shaw, G. M., Gallo, R. C., Wong-Stahl, F., & Wolf, H. (1987) *J. Virol. 61*, 570–578.

Montefiori, D. C., Robinson, W. E., Jr., & Mitchell, W. M. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 9248–9252.

Muller, W. E. G., Bachmann, M., Weiler, B. E., Schroder, H. C., Uhlenbruck, G., Shinoda, T., Shimizu, H., & Ushijima, H. (1991) *J. AIDS 4*, 694–703.

Murphy, C. I., Lennick, M., Lehar, S. M., Beltz, G. A., & Young, E. (1990) *Gen. Anal. Tech. Appl. 7*, 160–171.

Murphy, C. I., McIntire, J. R., Davis, D. R., Hodgdon, H., Seals, J. R., & Young, E. (1993) *Protein Expression and Purification* (in press).

Pal, R., Kalyanaraman, V. S., Hoke, G. M., & Sarngadharan, M. G. (1989a) *Intervirology 30*, 27–35.

Pal, R., Hoke, G. M., & Sarngadharan, M. G. (1989b) *Proc. Natl. Acad. Sci. U.S.A. 86*, 3384–3388.

Pinter, A., & Honnen, W. J. (1988) *J. Virol. 62*, 1016–1021.
Robey, W. G., Safai, B., Oroszlan, S., Arthur, L. O., Gonda, M. A., Gallo, R. C., & Fischinger, P. J. (1985) *Science 228*, 593–595.
Robinson, W. E., Jr., Montefiori, D. C., & Mitchell, W. M. (1987) *AIDS Res. Hum. Retroviruses 3*, 265–282.
Rossman, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., Hecht, H. J., Johnson, J. E., Kamer, G., Luo, M., Mosser, A. G., Ruekert, R. R., Sherry, B., & Vriend, G. (1985) *Nature 317*, 145–152.
Rusche, J. R., Lynn, D. L., Robert-Guroff, M., Langlois, A. J., Lyerly, H. K., Carson, H., Krohn, K., Ranki, A., Gallo, R. C., Bolognesi, D. P., Putney, S. D., & Matthews, T. J. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6924–6928.
Starcich, B. R., Hahn, B. H., Shaw, G. M., McNeely, P. D., Modrow, S., Wolf, H., Parks, E. S., Josephs, S. F., Gallo, R. C., & Wong-Staal, F. (1986) *Cell 45*, 637–648.
Stein, B. S., & Engleman, E. G. (1990) *J. Biol. Chem. 265*, 2640–2649.
Tarentino, A. L., Gomez, C. M., & Plummer, T. H., Jr. (1985) *Biochemistry 24*, 4665–4671.
van Halbeek, H. (1993) in *Methods in Molecular Biology*, (Jones, C., Mulloy, B., & Thomas, A. H., Eds.) Vol. 17, pp 115–148; Humana Press, Totowa, NJ.
Walker, B. D., Kowalski, M., Goh, W. C., Kozarsky, K., Krieger, M., Rosen, C., Rohrschneider, L., Haseltine, W. A., & Sodroski, J. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8120–8124.
Wells, D. E., & Compans, R. W. (1990) *Virology 176*, 575–586.
Willey, R. L., Rutledge, R. A., Dias, S., Folks, T., Theodore, T., Buckler, C. E., & Martin, M. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5038–5042.
Willey, R. L., Smith, D. H., Lasky, L. A., Theodore, T. S., Earl, P. L., Moss, B., Capon, D. J., & Martin, M. A. (1988) *J. Virol. 62*, 139–147.