Assignment of Intrachain Disulfide Bonds and Characterization of Potential Glycosylation Sites of the Type 1 Recombinant Human Immunodeficiency Virus Envelope Glycoprotein (gp120) Expressed in Chinese Hamster Ovary Cells*

(Cordelia K. Leonard, Michael W. Spellman, Lavon Riddle‡, Roed J. Harris, James N. Thomas§, and Timothy J. Gregory¶)

From the Departments of Medicinal and Analytical Chemistry, ‡Recovery Process Research and Development and §Cell Culture Research and Development, Genentech, Inc., South San Francisco, California 94080

(Received for publication, January 25, 1990)

The major envelope glycoprotein (gp120) of HIV-1 has been the object of intensive investigation since the initial identification of HIV-1 as the etiological agent of acquired immunodeficiency syndrome (Barre-Sinoussi et al., 1983). The gp120 molecule is of interest as a vaccine candidate (Berman et al., 1988; Arthur et al., 1987), as the mediator of viral attachment via the virus receptor CD4 (Dalgleish et al., 1984; Klatzman et al., 1984) and as an agent with immunosuppressive effects of its own (Shalaby et al., 1987; Diamond et al., 1988). It is also a potential mediator of the pathogenesis of HIV-1 in acquired immunodeficiency syndrome (Siliciano et al., 1988; Sodroski et al., 1986). The gp120 molecule is synthesized as part of a membrane-bound glycoprotein, gp160 (Allan et al., 1985). Via a host-cell-mediated process, gp160 is cleaved to form gp120 and the integral membrane protein gp41 (Robey et al., 1985). Together gp120 and gp41 form the spikes observed on the surface of newly released HIV-1 virions (Gelderblom et al., 1987). As there is no covalent attachment between gp120 and gp41, free gp120 is released from the surface of virions and infected cells (Gelderblom et al., 1986).

The gp120 molecule consists of a polyepitope core of 60,000 daltons; extensive modification by N-linked glycosylation increases the apparent molecular weight of the molecule to 120,000 (Lasky et al., 1986). The amino acid sequence of gp120 contains five relatively conserved domains interspersed with five hypervariable domains (Modrow et al., 1987; Willey et al., 1986). The hypervariable domains contain extensive amino acid substitutions, insertions, and deletions. Sequence variations in these domains result in up to 25% overall sequence variability between gp120 molecules from the various viral isolates. Despite this variation, several structural and functional elements of gp120 are highly conserved. Among these are the ability of gp120 to bind to the viral receptor CD4, the ability of gp120 to interact with gp41 to induce fusion of the viral and host cell membranes, the positions of the 18 cysteine residues in the gp120 primary sequence, and the positions of 13 of the approximately 24 N-linked glycosylation sites on the gp120 molecule (Lasky et al., 1986). The gp120 molecule is of interest as a vaccine candidate (Berman et al., 1988; Arthur et al., 1987), as the mediator of viral attachment via the virus receptor CD4 (Dalgleish et al., 1984; Klatzman et al., 1984) and as an agent with immunosuppressive effects of its own (Shalaby et al., 1987; Diamond et al., 1988). It is also a potential mediator of the pathogenesis of HIV-1 in acquired immunodeficiency syndrome (Siliciano et al., 1988; Sodroski et al., 1986). The gp120 molecule is synthesized as part of a membrane-bound glycoprotein, gp160 (Allan et al., 1985). Via a host-cell-mediated process, gp160 is cleaved to form gp120 and the integral membrane protein gp41 (Robey et al., 1985). Together gp120 and gp41 form the spikes observed on the surface of newly released HIV-1 virions (Gelderblom et al., 1987). As there is no covalent attachment between gp120 and gp41, free gp120 is released from the surface of virions and infected cells (Gelderblom et al., 1986).

The gp120 molecule consists of a polyepitope core of 60,000 daltons; extensive modification by N-linked glycosylation increases the apparent molecular weight of the molecule to 120,000 (Lasky et al., 1986). The amino acid sequence of gp120 contains five relatively conserved domains interspersed with five hypervariable domains (Modrow et al., 1987; Willey et al., 1986). The hypervariable domains contain extensive amino acid substitutions, insertions, and deletions. Sequence variations in these domains result in up to 25% overall sequence variability between gp120 molecules from the various viral isolates. Despite this variation, several structural and functional elements of gp120 are highly conserved. Among these are the ability of gp120 to bind to the viral receptor CD4, the ability of gp120 to interact with gp41 to induce fusion of the viral and host cell membranes, the positions of the 18 cysteine residues in the gp120 primary sequence, and the positions of 13 of the approximately 24 N-linked glycosylation sites in the gp120 sequence.

The disulfide bonding pattern within gp120 and the positions of actual oligosaccharide moieties on the molecule would be useful information for directing mutagenesis and fragmentation studies aimed at defining the functional domains of gp120 and sites for potential pharmacological interuption of its functions (e.g. type-common neutralizing epitopes). This information has been difficult to obtain due to the small amounts of gp120 available from natural sources, the complexity of the disulfide bonding and oligosaccharide structures in gp120, and uncertainty regarding the functionality or structural relevance (Moore et al., 1990) of rgp120 produced in non-mammalian systems. We have been able to produce large amounts of two different rgp120 fusion proteins in a mammalian cell system (Lasky et al., 1986). This has allowed us...
to elucidate all nine of the disulfide bonds, the positions of the glycosylation sites that are utilized, and the type of oligosaccharide moieties present at each site in rgp120 from the IIb isolate of HIV-1 produced in CHO cells.

**EXPERIMENTAL PROCEDURES**

**Materials**—Recombinant gp120 proteins were produced in CHO cells and purified by immunoaffinity chromatography as previously described (Lasky et al., 1986). DTT, iodoacetic acid, and 2-iodoacetyl-5-L-(aspartamido)-1,2-dideoxy-D-glucose (GlcNAc-Asn) were obtained from Sigma. HPLC/Gradient grade trifluoroacetic acid (Pierce Chemical Co.), Acetoniure virus (American B&D), and Milli-Q water (Millipore) were used for reversed-phase HPLC. The enzymes used were TPCK trypsin from Worthington Biomedical Corp., endoproteinase Asp-N (*"sequencing grade") obtained from Boehringer Mannheim GmbH, Staphylococcus aureus V8 protease from ICN ImmunoBiologicals, and PNGase F (N-Glycanase") and endo H from Genzyme.

**Reduction and S-Carboxymethylation**—Recombinant gp120 (2.0 mg of CL44) was dialyzed against 0.36 mM 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 ambient temperature. The reaction was quenched with excess DTT, the sample was diaлизed against 0.1 M ammonium bicarbonate, and then lyophilized.

**Treatment of RCA4 rgp120 with PNGase F—**RCM rgp120 (0.5 mg) was reconstituted in 0.1 ml of 0.25 M sodium phosphate, pH 8.6, containing 10 mM EDTA and 0.02% NaCN. PNGase F was added to the sample in the ratio of 0.1 units/mg of protein, and the sample was incubated overnight at 37°C. RCM rgp120 treated with PNGase F was diaлизed against 0.1 M ammonium bicarbonate.

**Treatment of RCA4 rgp120 with Endo H—**RCM rgp120 (0.5 mg) was reconstituted in 0.1 ml of 0.05 M sodium phosphate, pH 6.0, containing 0.02% NaCN. Endo H (2 units/ml) was added to the sample in the ratio of 0.1 unit/mg of protein and the sample was incubated overnight at 37°C. RCA4 rgp120 treated with endo H was diaлизed against 0.1 M ammonium bicarbonate.

**Treatment with TPCK Trypsin—**Samples of untreated, PNGase F-treated, and endo H-treated RCA4 rgp120 (0.5 mg aliquots of CL44) in 0.1 M ammonium bicarbonate were treated at ambient temperature with TPCK-trypsin by the addition of aliquots of enzyme (enzyme to substrate ratio of 1:100 w/w) at 37°C. The digestion was stopped after 24 h by freezing the samples. For disulfide determinations, a sample of rgp120 (0.5 mg of 9AA) was treated with TPCK-trypsin using the same conditions.

**Treatment of Tryptic Peptides with PNGase F Followed by Endoproteinase Asp-N—**Peptides (ranging from 0.5 nmol to 3.7 nmol) purified by reversed-phase HPLC of a 9AA tryptic digest were reconstituted in 0.05 M sodium phosphate, pH 7.0, containing 0.02% NaCN (0.06 ml). PNGase F was added to the 0.05 M sodium phosphate buffer, pH 7.0, containing 0.02% NaCN, and the samples were incubated for 20 h at 37°C. Endoproteinase Asp-N (2 μg) was then added, and the samples were incubated for 20 h at 37°C.

**Treatment of Tryptic Peptides with S. aureus V8 Protease—**Peptides (3.0 nmol) purified by reversed-phase HPLC of a 9AA tryptic digest were reconstituted in 0.05 M sodium phosphate, pH 7.0, containing 0.02% NaCN (0.06 ml). V8 protease (5 μg) was added at 0 and 7 h, and the sample was incubated for 24 h at 37°C.

**Treatment of CL44 Peptides with Endo H Followed by PNGase F—**Peptides (typically 3 nmol) purified by reversed-phase HPLC were reconstituted in 0.05 M sodium phosphate, pH 6.0, containing 0.02% NaCN (0.1 ml). Endo H (0.05 unit in 0.025 ml of 0.05 M sodium phosphate, pH 6.0, containing 0.02% NaCN) was added, and the sample was incubated for 20 h at 37°C. PNGase F (0.25 units) and 0.05 M sodium phosphate, pH 10.3, containing 0.02% EDTA and 0.02% NaCN (0.125 ml) were then added, and the sample was incubated for 20 h at 37°C.

**Reversed-phase HPLC—**Tryptic digests were fractionated by reversed-phase HPLC on a 5-μm Vydac C18 endcapped column (4.6 × 250 mm). After equilibration with 0.1% aqueous trifluoroacetic acid, the elution of tryptic peptides was carried out at 1 ml/min with a linear gradient from 0 to 45% acetonitrile containing 0.08% trifluoroacetic acid in 90 min. The system used was a Waters gradient liquid chromatograph consisting of two 600A pumps, a 720 controller, a WISP 710B injector, and a Perkin-Elmer LC75 single wavelength 1 UV detector set at 214 nm.

**Peptide Identification—**Peptides collected from reversed-phase HPLC were identified by AAA and/or N-terminal sequence analysis. Samples for AAA were treated with constant boiling HCl at 110°C in vacuo for either 24 or 72 h, depending upon extent of glycosylation. The extended hydrolysis degrades glucosamine, which would otherwise interfere with quantitation of Ile and Leu. Analysis was performed on a Beckman model 6300 amino acid analyzer with ninhydrin detection.

**N-terminal sequence analysis** was performed on an Applied Biosystems model 477A/120A. The amino acid concentration in the equilibration buffer of the PTH analysis system was decreased from 10 to 9% to resolve the PTH derivative of GlcNAc-Asn from DTT. **FAB MS—**FAB mass spectra were acquired on a JEOL HX110HF/HX110HP tandem mass spectrometer operated in a negative ion reflector mode. H4-MS was performed with 6 keV xenon atoms (10 mA emission current). Data were acquired over a mass range of 380–4000 atomic mass units.

**RESULTS**

Lasky et al. (1986) expressed gp120 in CHO cells as a fusion protein using the signal peptide of the herpes simplex gD1. Two such fusion proteins were used in this study. The recombinant glycopolypeptide used in most of this study (CL44) was expressed as a 498-amino acid fusion protein containing the first 27 residues of gD1 fused to residues 31-501 of gp120 (Lasky et al., 1986). This construction lacks the first cysteine residue of mature gp120. Disulfide assignments were carried out on another recombinant fusion protein (9AA) which contains the first 9 residues of gD1 fused to residues 4-501 of gp120. This restores the first cysteine residue, Cys-24. Carboxy-terminal analysis of CL44 using carboxypeptidase digestion indicated that glutamic acid residue 479 is the carboxyl terminus of the fully processed molecule secreted by CHO cells (data not shown). The amino acid sequences of these two constructions are given in Fig. 1.

**RCM CL44 Tryptic Map—**Reversed-phase HPLC tryptic mapping was used to confirm the primary structure of the molecule, to assign intrachain disulfide bonds, and to characterize potential sites for N-glycosylation. In experiments not intended to give information about disulfides, the protein was RCM prior to digestion with trypsin. This treatment unfolds the protein and disrupts disulfide bonds, thereby resulting in smaller tryptic fragments than would be obtained with the native molecule.

The reversed-phase HPLC tryptic map of RCM CL44 is shown in Fig. 2. Tryptic peptides were separated by reversed-phase HPLC using an acetoniure/water system with trifluoroacetic acid as the ionic modifier. As will be discussed below, much of the peak heterogeneity derives from the extremely high (approximately 50% of total mass) carbohydrate content of the molecule. Peaks were collected and subjected to AAA for identification (Table I). In some cases, N-terminal sequence analysis was used for confirmation (these peaks are indicated in Table I). The peaks not assigned a label in Fig. 2 were not identified.

All of the peptides identified were consistent with the primary structure predicted from the cDNA sequence. Of the 38 predicted peptides with three or more amino acids, 36 were identified in the tryptic map of RCM CL44. In addition, four predicted peptides consisting of two amino acids each were also identified (H3, H4, T23, and T35). The tripeptide com-
Characterization of Recombinant gp120

FIG. 1. Amino acid sequences of the mature envelope glycoprotein (gp120) from the 111s isolate of HIV-1 (a) and the N-terminal sequence portion of the recombinant fusion glycoproteins (SAA or CL44) from the herpes simplex gD1 (b). Fusion sites between the gD1 and gp120 segments in the 9AA and CL44 constructions are marked with (*) and (**), respectively. The letter 2' refers to observed tryptic cleavage of the gp120 segment, and the peptides are ordered sequentially starting at the N terminus of the molecule. Lower case letters following the T number indicate other unexpected proteolytic cleavages. The letter H refers to the observed tryptic cleavage of the herpes simplex gD1 protein portion of CL44. Peptide T2' contains the fusion site in CL44. The cysteine residues of gp120 are enclosed in boxes, and potential N-glycosylation sites are indicated with a dot above the corresponding asparagine residue. Posed of residues 139-141 (VQK) was not identified in the map and was not given a label in Fig. 1. The only other peptide not identified was T13 (CNNK). Asparagine residue 200 of peptide T13 is a potential glycosylation site and the peptide lacks hydrophobic amino acids. Therefore, this glycopeptide is likely to be extremely hydrophilic and poorly resolved from the salt fraction on the reversed-phase column.

Tryptic cleavage did not occur between peptides T5 and T6 and between peptides T8 and T9. These are designated in Fig. 2 as two T-numbers separated by a comma (T5,6 and T8,9). The absence of cleavage was confirmed by N-terminal sequence analysis of the peptides. In both of these cases, the asparagine residue to the C-terminal side of the cleavage site is a potential N-glycosylation site, and it is likely that the carbohydrate moiety interferes with the action of trypsin. Incomplete tryptic cleavage was also observed between peptides H4 and T1' and between peptides T23 and T24 (H4,T1' and T23,24).

Several peptides arising from non-tryptic cleavages were observed in the tryptic map of RCM CL44. Two of the predicted tryptic peptides were further cleaved by "chymotryptic-like" cleavages. Peptide T12 was completely cleaved after tyrosine residue 187 and phenylalanine residue 193 to yield peptides T12a, T12b, and T12c. Peptide T4 was partially hydrolyzed after leucine residue 95 to yield peptides T4a and T4b. Intact peptide T4 was also present.

FIG. 2. Reversed-phase HPLC tryptic map of RCM CL44. This chromatogram was generated with 7.5 nmol of trypsin-digested RCM CL44. Chromatography conditions were as described under "Experimental Procedures." Peaks were collected and identified by AAA and in some cases confirmed by N-terminal sequence analysis (Table I). Identified peaks are labeled according to the nomenclature given in Fig. 1. Peptides containing potential tryptic sites that were not hydrolyzed are designated by two T numbers separated by a comma.

Disulfide Assignments in gp120—Mature gp120 contains 18 cysteine residues (enclosed in boxes in Fig. 1) and therefore could contain nine intrachain disulfide bonds. The CL44 construction lacks Cys-24, the first cysteine residue of gp120 (Lasky et al., 1986); therefore, a different construction (9AA), in which the first cysteine residue was restored, was purified to approximately the same degree as CL44. Ellman's reagent (Ellman, 1959) was used to demonstrate the absence of free sulfhydryl groups in 9AA (data not shown). Therefore, disulfide assignments were determined for the 9AA construction.

Disulfide assignments were determined for the 9AA construction. Tryptic mapping studies performed without reduction and S-carboxymethylation of cysteine residues allowed partial assignment of disulfides. The tryptic map of 9AA is shown in Fig. 3. Peaks were identified by N-terminal sequence analysis (Table II). These identifications allowed unequivocal assignment of three of the nine disulfide bonds: between Cys-101 and Cys-127 (peak A, Table II), between Cys-266 and Cys-301 (peak B, Table II), and between Cys-24 and Cys-44 (peak E, Table II).

Peptides containing the remaining cysteine residues were also identified (Table II). Peptide T28 contains 3 cysteine residues and coelutes with peptide T31, which contains 1 cysteine residue.
cysteine residue (peak D, Table II). Peptide T11 contains 2 cysteine residues and coelutes with peptides T3 and T4, each of which contains a single cysteine residue (peak F, Table II). Similarly, peptide T14 contains 2 cysteine residues and coelutes with peptides T12 and T13, each of which has a single cysteine residue (peaks C and E, Table II). In each of these cases more than one disulfide bond was present in the group of tryptic peptides, thereby preventing unambiguous assignment. These tryptic peptides were further manipulated as described below to introduce selective cleavage between cysteine residues located on a single peptide.

The procedure used to cleave between the cysteine residues of peptides T11 and T14 is summarized (Scheme 1). Each of the peptides has a potential N-linked glycosylation site located between the cysteine residues. The peptides were treated with PNGase F, which removes asparagine-linked carbohydrate while converting the attachment asparagine residue to aspartic acid (Tarentino et al., 1985). The resulting aspartic acid residue serves as a point for selective cleavage of the peptides with endoproteinase Asp-N (Drapeau, 1980). The peptides were separated by reversed-phase HPLC and identified by N-terminal sequence analysis. The HPLC chromatogram obtained after treatment of peptides T12, T13, and T14 (peak C, Fig. 3) with PNGase F followed by endoproteinase Asp-N is given in Fig. 4a, and the
Characterization of Recombinant gp120

Assignment of disulfides from peptides isolated in Fig. 4

The tryptic peptides that could not be assigned unambiguously in Table II were further manipulated as described in Fig. 4. Peaks were identified by N-terminal sequence analysis.

| Peak | Sequence |
|------|----------|
| 1    | DGN     |
| 2    | EPHGAGF  |
| 3    | DQSLKPB9VK |
| 4    | DTSVITC,A175PK |
| 5    | ITHGENDE  |
| 6    | FYVY6NSTLPPSTW5NSTWSTWE |

The tryptic mapping analysis and the further selective degradations permitted the assignment of all nine intrachain disulfide bonds of gp120. Parallel experiments performed on CL44 produced similar results for the eight disulfide bonds remaining in that construction (not shown). The disulfide bond assignments of gp120 are summarized in Fig. 6.

Glycosylation Sites of gp120—Mature gp120 contains 24 potential sites for N-glycosylation, as recognized by the sequence Asn-Xaa-Ser(Thr) (Kornfeld and Kornfeld, 1985). These sites are indicated by a dot above the corresponding asparagine residue in Fig. 1. In the present study, tryptic mapping of enzymatically deglycosylated CL44 was used in conjunction with Edman degradation and FAB-MS of individually treated peptides to determine which of the 24 potential N-glycosylation sites are glycosylated and which contain less fully processed (i.e. high mannose-type or hybrid-type) oligosaccharides.

The two enzymes used for deglycosylation were PNGase F and endo H. PNGase F releases all types of N-linked oligosaccharide structures by cleavage of the β-asparaginylglucosamine linkage (Tarentino et al., 1985). Endo H releases only high mannose-type and hybrid-type oligosaccharide structures by cleaving between the two core N-acetylglucosamine residues (Tai et al., 1977). Deglycosylation of a peptide can be monitored by the increase in retention time of the peak corresponding to the glycopeptide in the reversed-phase elution profile. Thus, it was possible to determine which peptides were glycosylated by treatment with PNGase F and, on the basis of susceptibility to endo H, to distinguish those with attached high mannose-type and/or hybrid-type oligosaccharides as the predominant structures.
The 24 potential glycosylation sites of CL44 are contained in 14 tryptic glycopeptides. Thirteen of these glycopeptides were identified in the tryptic map of RCM CL44 (Fig. 2). As mentioned above, T13 (CNNK) was not identified. The tryptic maps of PNGase F-treated RCM CL44 and endo H-treated RCM CL44 are compared with the RCM CL44 tryptic map in Fig. 5. The peaks corresponding to glycopeptides are labeled in each of the three tryptic maps.

As would be expected for a heavily glycosylated molecule, treatment of RCM CL44 with PNGase F (Fig. 5b) simplified the tryptic map significantly. Typically, the peaks corresponding to potential glycopeptides in the RCM CL44 tryptic map (Fig. 5a) were broad and often appeared as multiplets. Deglycosylation resulted in sharp, single peaks for each peptide, indicating that the glycopeptide peak multiplicity and broadness was due to carbohydrate heterogeneity.

Of the 13 potential glycopeptides that had been identified in the tryptic map of RCM CL44, all were shifted to later retention times in the tryptic map of PNGase F-treated material. This demonstrates that at least 13 of the 24 potential sites are glycosylated. Peptide T28 was not recovered after deglycosylation. This peptide contains a large number of non-polar amino acids and, after removal of the hydrophilic carbohydrate moieties, may bind irreversibly to the HPLC column. As described above, peptide T22 elutes at two positions in the RCM CL44 tryptic map presumably as a result of conversion of the N-terminal glutamine to pyroglutamic acid. The retention times of both of the T22 peaks were altered in the deglycosylated material produced by treatment with both PNGase F and endo H, confirming that the difference between these forms of peptide T22 in the RCM CL44 tryptic map was not due to carbohydrate heterogeneity.

The tryptic map of endo H-treated RCM CL44 (Fig. 5c) indicated that six of the 13 tryptic glycopeptides were endo H-susceptible (peptides T14, T16, T22, T24, T28, and T31). In addition, a small amount of peptide T15 showed endo H susceptibility. For each of these glycopeptides, the elution time of the endo H-treated glycopeptide was earlier than that of the corresponding PNGase F-treated glycopeptide. This is due to the hydrophilic N-acetylglucosamine residue that remains attached to the asparagine residue following endo H treatment. Peptide T16 was not identified in the tryptic map of endo H-treated RCM CL44. This peptide contains three potential glycosylation sites and was poorly recovered under any circumstances.

Conclusions as to the type of glycosylation present on each of the tryptic glycopeptides based on susceptibility to PNGase F and endo H are summarized in Table IV. Seven of the 13 glycopeptides identified in the tryptic map of RCM CL44 contain only a single glycosylation site and thus could be characterized unambiguously with regard to enzyme susceptibility. Peptides T14 (Asn-58), T26 (Asn-326), and T32 (Asn-433) were deglycosylated only by PNGase F and, therefore, contain attached complex-type oligosaccharide structures. Peptides T22 (Asn-302), T24 (Asn-309), and T31 (Asn-418) were susceptible to both PNGase F and endo H and, therefore,

**TABLE IV**

Assignment of glycosylation type to RCM CL44 tryptic peptides by susceptibility to PNGase F and endo H

Susceptibility to PNGase F or endo H was determined by an increase in the retention time of a peptide in the tryptic map of RCM CL44. PNGase F releases all types of N-linked oligosaccharide structures, whereas endo H releases only high mannose and hybrid oligosaccharide structures.
carry high mannose-type and/or hybrid-type oligosaccharide structures. Peptide T15 is only partially susceptible to endo H; therefore, Asn-246 carries primarily complex-type oligosaccharides but must also have some attached high mannose-type and/or hybrid-type oligosaccharide structures.

Peptides T6, T9, and T11 each contain two potential glycosylation sites. Each peptide was deglycosylated by PNGase F but not by endo H indicating the presence of mostly complex-type oligosaccharide structures. In order to determine whether one or both of the potential glycosylation sites in each peptide were actually glycosylated, the PNGase F-treated glycopeptides were subjected to either FAB-MS or Edman degradation. Treatment with PNGase F converts the attachment asparagine residue to aspartic acid during deglycosylation (Tarentino et al., 1985). This conversion can be detected by FAB-MS as an increase of 1 atomic mass unit in the mass of the peptide for each site deglycosylated (Carr and Roberts, 1986) or by Edman degradation by the appearance of the PTH derivative of aspartic acid at the appropriate cycles. FAB-MS of deglycosylated peptide T5,6 revealed an ion corresponding to the peptide mass plus 2 atomic mass units ([MH]+ observed: m/z 1772.6; calculated: m/z 1772.7). FAB-MS of deglycosylated peptide T9 gave similar results ([MH]+ observed: m/z 1301.8; calculated: m/z 1301.5). Edman degradation was performed instead of FAB-MS on deglycosylated peptide T11 because of its high molecular weight (>2000 a.m.u.). Aspartic acid was observed in cycles 8 (derived from Asn-156) and 19 (derived from Asn-167). These combined results indicate the presence of complex-type oligosaccharide structures attached to Asn residues 106, 111, 126, 130, 156, and 167.

The remaining three glycopeptides identified in the tryptic map of RCM CL44 contained multiple potential glycosylation sites and were endo H susceptible. Peptides T14, T16, and T28 account for a total of 10 potential glycosylation sites. Characterization of each glycosylation site was achieved by Edman degradation of HPLC-purified peptides that had been subjected to treatment with endo H followed by PNGase F (Scheme 2). When endo H releases the high mannose-type and hybrid-type oligosaccharide structures, it leaves an N-acetylglucosamine residue attached to the asparagine residue of the peptide (Tarentino et al., 1974). PNGase F will not remove this N-acetylglucosamine residue but will release the remaining N-linked oligosaccharide structures by cleavage of the β-asparagylcosylamine bond, resulting in conversion of the attachment asparagine residue to aspartic acid (Chu, 1986). Therefore, treatment with endo H followed by PNGase F will yield asparagine at an unglycosylated site, GlcNAc-Asn at a glycosylation site to which had been attached high mannose or hybrid oligosaccharide structures, or an Asp residue at a glycosylation site which had carried complex type oligosaccharide structures.

The results of these experiments are summarized in Table V and indicate that CL44 contains complex-type oligosaccharide

## Table V

| Tryptic Peptide | Asn Residue# | Residue Observed | Glycosylation Type |
|----------------|-------------|-----------------|-------------------|
| T14            | 515         | GlcNAc-Asn      | High Mannose and/or Hybrid |
| T28            | 356         | GlcNAc-Asn      | High Mannose and/or Hybrid |
| T16            | 259         | GlcNAc-Asn      | High Mannose and/or Hybrid |
| T28            | 362         | GlcNAc-Asn      | High Mannose and/or Hybrid |

**Fig. 6.** Schematic representation of gp120 showing disulfides and glycosylation sites. Glycosylation sites containing high mannose-type and/or hybrid-type oligosaccharide structures are indicated by a (Ⅵ) and glycosylation sites containing complex-type oligosaccharide structures are indicated by a (Ⅴ). Roman numerals label the five disulfide-bonded domains. The five hypervariable regions of Modrow et al. (1987) are enclosed in boxes and labeled V1-V5.
structures at Asn residues 271, 367, and 376 and high mannosetype and/or hybrid-type oligosaccharide structures at Asn residues 204, 211, 232, 259, 265, 356, and 362.

Peptide T13, which contains the remaining glycosylation site was not identified in any of the tryptic maps presented in this paper. However, FAB-MS data obtained from the void peak of a tryptic map of RCM CL44 treated with endo H followed by PNGase F revealed an ion corresponding to MH* for that peptide containing an attached N-acetylgalcosamine residue (observed: m/z 740.1; calculated: m/z 740.4). The presence of peptide T13 in the void peak was further confirmed by AAA. Therefore, we conclude that Asn-200 is glycosylated and carries primarily high mannosetype and/or hybrid-type oligosaccharide structures.

The data presented here demonstrate that all 24 potential glycosylation sites of gp120 are utilized, that 13 sites contain primarily complex-type oligosaccharide structures while 11 sites contain primarily high mannosetype and/or hybrid-type oligosaccharide structures. The type of glycosylation at each site is summarized in Fig. 6.

DISCUSSION

We have determined the disulfide bonding pattern and the attachment positions of oligosaccharide moieties of rgp120 from the IIIb isolate of HIV-1. A schematic representation of this information is presented in Fig. 6. The gp120 molecules from which the structural data were obtained possess the functional properties attributed to gp120 produced by HIV-1 virions including high affinity CD4 binding (Lasky et al., 1987), and HIV-1 neutralizing antigenicity (Lasky et al., 1986). We therefore conclude that the CHO-expressed gp120 is properly folded and that the disulfide-bonded domains reported here for the recombinant molecules are representative of those occurring in gp120 produced by HIV-1 virions.

Functional Aspects of gp120 Structure—The gp120 molecule comprises five disulfide-bonded loop structures. The first and fourth are simple loops formed by single disulfide bonds while the second, third, and fifth are more complex arrays of loops formed by nested disulfide bonds. The fourth disulfide-bonded domain (residues 266-301) has been shown to contain significant type-specific neutralizing epitopes (Matsushita et al., 1988; Rusche et al., 1988; Goudsmit et al., 1988; Javaherian et al., 1989) and the fifth disulfide-bonded domain (residues 348-415) has been shown to be important for CD4 binding (Lasky et al., 1987; Kowalski et al., 1987). No direct functional correlates have been described for the other three disulfide-bonded domains. The amino acid sequence of gp120 varies to a large extent between different viral isolates but the majority of the variability is localized in hypervariable regions which punctuate the otherwise relatively conserved sequences (Wiley et al., 1986; Modrow et al., 1987). Modrow et al. (1987) have identified five hypervariable regions which are characterized by sequence variation, insertions, and deletions. Four of these hypervariable regions correspond to well-delineated loops as indicated in Fig. 6. With the exception of the third hypervariable loop (disulfide-bonded domain IV) the functional significance of these regions is unknown.

The positions of the cysteine residues and, presumably, the disulfide bonding pattern in gp120 are highly conserved between isolates. Among HIV-1 isolates, the only exception to this conservation is the Z3 isolate (Wiley et al., 1986) which has an additional pair of cysteine residues in the fourth hypervariable domain (residues 383-384). These residues most likely form a tenth disulfide bond in the gp120 from this isolate. The presence of this extra bond in such a hypervariable region probably has no more effect on the structure and function of the molecule than the other sequence variations that occur in that region. In HIV-2 and SIV, the positions of the cysteine residues in disulfide-bonded domains I, II, IV, and V are conserved (Mycers et al., 1989). In domain III there are two additional pairs of cysteine residues (three in SIV isolate MM142) which are presumed to be disulfide bonded within a finger-like domain III structure analogous to that illustrated in Fig. 6. Another major difference between HIV-1, HIV-2, and SIV is that hypervariable region V2 is reduced to five amino acids in HIV-2 and SIV. The functional significance of the differences between HIV-1, HIV-2, and SIV is unknown at this time.

One of the most important functions of gp120 is its ability to bind to CD4 and thereby mediate the attachment of virions to susceptible cells (Klatzman et al., 1984; Dalgleish et al., 1984). The CD4-binding function has been localized by mutagenesis and structural studies (Lasky et al., 1987; Kowalski et al., 1987) to the region between residues 320 and 450, which includes the fifth disulfide-bonded domain. Lasky et al. (1987) showed that deletion of residues 396 to 407 and mutagenesis of Ala-402 to Asp abolished CD4 binding. They also mapped the epitope of a monoclonal antibody that blocks gp120-CD4 binding to residues 392-402. Kowalski et al. (1987) identified three regions as being involved with CD4 binding. Insertions between residues 333-334, 388-390, and 442-443 abolished CD4 binding. In addition, a deletion of residues 441-479 abolished CD4 binding while deletion of residues 362-369 within the fourth hypervariable region had no effect on binding. Cordier et al. (1989) have shown that mutagenesis of Trp-397 to Tyr or Phe decreases CD4 binding and changes to Ser, Gly, Val, or Arg abolish binding. Nguyen et al. (1988) have reported that a proteolytic fragment of gp120 from residue 322 to near the C terminus retains the ability to bind to CD4. The results of these studies indicate that the CD4 binding capacity of gp120 is localized to the region between residues 320 and 450 and more specifically to the residues around 333-334, 442-443, and the sequence between 388 and 407.

In the course of efforts to map the epitope of monoclonal antibody 5C2-E5 which blocks gp120-CD4 binding, Lasky et al. (1987) treated rgp120 (CL44) with acetic acid to cleave the protein at aspartic acid residues (Ingram, 1963) and isolated the peptide fragment 383-426 from a column of immobilized anti-gp120 monoclonal antibody 5C2-E5. Digestion of reduced rgp120 yielded the same fragment. Consequently, it was concluded that a disulfide bond existed between Cys residues 388 and 415. In the analysis reported here we have failed to find this disulfide bond and, instead, have consistently found the disulfide bonds between Cys-355 and Cys-388, and between Cys-348 and Cys-415 as summarized in Fig. 6. We believe that the true disulfide-bond assignment is as indicated in Fig. 6 and that the acetic acid digestion produced some disulfide bond rearrangement (Ryle and Sanger, 1955) in the earlier work.

The Oligosaccharides of gp120—Approximately 50% of the apparent molecular mass of gp120 is carbohydrate. The structures of the oligosaccharide moieties released by hydrazinolysis of CL44 rgp120 have been exhaustively analyzed (Mizochi et al., 1988a; Mizochi et al., 1988b). These authors found that 53% of the N-linked oligosaccharides were of the high mannosetype, 4% were of the hybrid type, and 63% were of the complex type. Of the complex oligosaccharides 90% were fucosylated and 94% were sialylated. The complex structures were approximately 4% monoantennary, 61% biantennary, 19% triantennary, and 6% tetraantennary. No O-linked oligosaccharides were found. Geyer et al. (1988) have...
analyzed the oligosaccharides of gp120 from the IIIb isolate of HIV-1 infected human cells. They found that high mannosetype oligosaccharides accounted for approximately 60% of the carbohydrate structures. The remaining structures were fucosylated, partially sialylated bi-, tri-, and tetraantennary complex-type oligosaccharides. No novel carbohydrate structures, or moieties that would be expected to act as heterophilic antigens in man, have been isolated from gp120 from either source.

We have shown here that all 24 glycosylation sites are utilized, and that 13 of the 24 sites contain complex-type oligosaccharides as the predominant structures while 11 contain primarily hybrid and/or high mannosetype structures. The demonstration of endo H-susceptible structures at 11 of the 24 sites is consistent with the earlier results of Mizuochi et al. (1988a, 1988b) who determined that nearly 40% of the total oligosaccharide structures released from gp120 were hybrid and/or high mannosetype oligosaccharides.

The 24 potential N-linked glycosylation sites in the gp120 sequence are conserved to a large extent between different viral isolates (Wille et al., 1986; Modrow et al., 1987). Based on the gp120 sequence comparisons in these references, 13 of the sites on gp120 from the IIIb isolate of HIV-1 are absolutely conserved: these include eight of the 11 sites that carry predominantly hybrid-type and/or high mannosetype oligosaccharides. Thus, the less fully processed (i.e. endo H-susceptible) oligosaccharides of gp120 are found preferentially at the most conserved glycosylation sites. The remaining sites (eight complex and three hybrid/high mannose) are relatively conserved, even though many of them occur in the hypervariable regions. The positions of these sites may shift or be deleted, but there is always one or more new site(s) within 5-10 residues of the reference IIIb site. Studies by Wille et al. (1988) demonstrated that mutagenesis of Asn-232 to Gln decreased the infectivity of virions containing the mutant gp120 molecules without affecting CD4 binding or syncytium formation. At this time, no particular functional significance can be attributed to the type of oligosaccharide structure at any of the sites.

The role of the carbohydrate moieties on gp120 in CD4 binding has been investigated by several authors (Lifson et al., 1986; Rosen, C. A., Haseltine, W. A., Lee, T. D., and Shively, J. E. (1986) Science 228, 1091-1094).

Arthur, L. O., Pyle, S. W., Nara, P. L., Bess, J. W., Jr., Gonda, M. A., Kelliler, J. C., Gilden, R. V., Robey, W. G., Bolognesi, D. P., Gallo, R. C., and Fischinger, P. J. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 8583-8587

Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Champsy, S., Gruest, J., Dauquet, C., Axler-Blin, C., Vezenet-Brun, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983) Science 220, 862-871.

Berman, P. W., Groopman, J. E., Gregory, T. J., Weiss, R. A., Clapham, P. R., Ferriani, R., Riddle, L., Shimasaki, C., Lucas, C., Lasky, L. A., and Eichberg, J. W. (1989) Proc. Natl. Acad. Sci. U. S. A. 85, 5200-5204.

Bersudsky, Y., and Shively, J. E. (1990) J. Biol. Chem. 265, 839-840.

Bersudsky, Y., Boily, Y., and Hounard, J. (1972) J. Biol. Chem. 247, 6720-6726.

Elliott, L. H. (1959) Arch. Biochem. Biophys. 82, 70-77.

Fenouillet, E., Clerget-Raslain, B., Gluckman, J. C., Guetard, D., Montagnier, L., and Bahrouk, S. J. (1986) J. Immunol. 141, 3719-3727.

Drapeau, G. (1987) J. Biol. Chem. 262, 383-386.

Paxton, R. J., Mooser, G., Pande, H., Lee, T. D., and Shively, J. E. (1986) Anal. Biochem. 157, 396-406.

Lasky, L. A., Nakamura, G., Smith, D. H., Fennie, C., Shimasaki, C., Stemp, J. L., and Burakoff, S. J. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 993-998.

Kornfeld, R., and Kornfeld, S. (1985) Annu. Rev. Biochem. 54, 631-665.

Galerstein, T., Ross, A., Josephs, S., Smith, T., Berzofsky, J., Greenblatt, D., Gherardi, E., Fischer, M., and Feizi, T. (1986) Nature 312, 676-677.

Greaves, M. F., and Wolf, H. (1987) Science 242, 412-415.

Klatzmann, D., Champagne, E., Chambaret, S., Gruest, J., Guetard, D., Holschbach, C., Hunsmann, G., and Schneider, J. (1988) Nature 312, 676-677.

Lasky, L. A., Nakamura, G., Smith, D. H., Fennell, M., Rabson, A., Josephs, S., Smith, T., Bolognesi, D. P., Putney, S. D., and Matthews, T. J. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 6768-6772.

Kornfeld, R., and Kornfeld, S. (1985) Annu. Rev. Biochem. 54, 631-665.

Kuwalewski, M., Potz, J., Basirion, L., Dorfman, T., Goh, W. C., Terwilliger, E., Dayton, A., Rosen, C., Haseltine, W., and Sodroski, J. (1987) Science 237, 1501-1555.

Kawasaki, K., Lasky, L. A., Groopman, J. E., Fennie, C. W., Benz, P. M., Capon, D. J., Dowbenko, D. J., Nakamura, G. R., Nunes, W. M., Renz, M. E., Berman, P. W. (1986) Science 233, 590-592.

Geyer, H., Holschbach, C., Humanski, G., and Schneider, J. (1988) Virology 166, 171-176.

Gonzalez, D. C., Speckman, B. G., Gregory, T. J., Weiss, R. A. (1986) Proc. Natl. Acad. Sci. U. S. A. 84, 8583-8587.

Acknowledgments—We wish to thank Dr. Phil Berman and Dr. Larry Lasky for many helpful discussions, Dr. John Chakel for FAB-MS analysis, and Carol Morita and Anne Stone for preparation of illustrations.

REFERENCES

Allan, J. S., Colligan, J. E., Barin, F., McLane, M. F., Sodroski, J. G., Rosen, C. A., Haseltine, W. A., Lee, T. H., and Essex, M. (1985) Science 228, 1091-1094
Characterization of Recombinant gp120

Gallo, R. C., and Fischinger, P. J. (1985) Science 228, 593-595
Rusche, J. R., Javaherian, K., McDanal, C., Petro, J., Lynn, D. L., Grimaila, R., Langlois, A., Gallo, R. C., Arthur, L. O., Fischinger, P. J., Bolognesi, D. P., Putney, S. D., and Matthews, T. J. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 3198-3202
Ryle, A. P., and Sanger, F. (1955) Biochem. J. 60, 535-540
Sanger, F., and Thompson, E. O. P. (1953) Biochem. J. 53, 366-374
Shalaby, M. R., Krowka, J. F., Gregory, T. J., Hirabayashi, S. E., McCabe, S. M., Kaufman, D. S., Stites, D. P., and Ammann, A. J. (1987) Cell. Immunol. 110, 140-148
Siliciano, R. F., Lawton, T., Knall, C., Karr, R. W., Berman, P., Gregory, T., and Reinherz, E. L. (1988) Cell 54, 561-575
Sodroski, J., Goh, W. C., Rosen, C., Campbell, K., and Haseltine, W. A. (1986) Nature 322, 470-474
Tai, T., Yamashita, K., and Kobata, A. (1977) Biochem. Biophys. Res. Commun. 78, 434-441
Tarentino, A. L., Plummer, T. H., Jr., and Maley, F. (1974) J. Biol. Chem. 249, 818-824
Tarentino, A. L., Gomez, C. M., and Plummer, T. H., Jr. (1988) Biochemistry 24, 4665-4671
Willey, R. L., Rutledge, R. A., Dias, S., Folks, T., Theodore, T., Buckler, C. E., and 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., and Martin, M. A. (1988) J. Virol. 62, 139-147
Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells.

C K Leonard, M W Spellman, L Riddle, R J Harris, J N Thomas and T J Gregory

J. Biol. Chem. 1990, 265:10373-10382.

Access the most updated version of this article at http://www.jbc.org/content/265/18/10373

Alerts:
  • When this article is cited
  • When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/265/18/10373.full.html#ref-list-1