JOURNAL OF VIROLOGY, Aug. 1996, p. 5455–5465 Vol. 70, No. 8
0022-538X/96/$04.00+0

Disulfide Bond Structure Determination and Biochemical Analysis of Glycoprotein C from Herpes Simplex Virus

ANN H. RUX,[1,2,3]* WILLIAM T. MOORE,[4] JOHN D. LAMBRIS,[4] WILLIAM R. ABRAMS,[2,5] CHARLINE PENG,[1,2] HARVEY M. FRIEDMAN,[6] GARY H. COHEN,[1,2] AND ROSELYN J. EISENBERG[2,3]

Department of Microbiology,[1] Center for Oral Health Research,[2] and Department of Anatomy and Histology,[5] School of Dental Medicine, Department of Pathobiology, School of Veterinary Medicine,[3] and Department of Pathology and Laboratory Medicine[4] and Department of Medicine,[6] School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Received 29 February 1996/Accepted 6 May 1996

A biochemical analysis of glycoprotein C (gC) of herpes simplex virus was undertaken to further characterize the structure of the glycoprotein and to determine its disulfide bond arrangement. We used three recombinant forms of gC, gC1(457t), gC1(Δ33-123t), and gC2(426t), each truncated prior to the transmembrane region. The proteins were expressed and secreted by using a baculovirus expression system and have been shown to bind to monoclonal antibodies which recognize discontinuous epitopes and to complement component C3b in a dose-dependent manner. We confirmed the N-terminal residues of each mature protein by Edman degradation and confirmed the internal deletion in gC1(Δ33-123t). The molecular weight and extent of glycosylation of gC1 (457t), gC1(Δ33-123t), and gC2(426t) were determined by treating each protein with endoglycosidases and then subjecting it to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and mass spectrometric analysis. The data indicate that eight to nine of the predicted N-linked oligosaccharide sites on gC1(457t) are occupied by glycans of approximately 1,000 Da. In addition, O-linked oligosaccharides are present on gC1(457t), primarily localized to the N-terminal region (amino acids [aa] 33 to 123) of the protein. gC2(426t) contains N-linked oligosaccharides, but no O-linked oligosaccharides were detected. To determine the disulfide bond arrangement of the eight cysteines of gC1(457t), the protein was cleaved with cyanogen bromide. SDS-PAGE analysis followed by Edman degradation identified three cysteine-containing fragments which are not connected by disulfide linkages. Chemical modification of cysteines combined with matrix-assisted laser desorption ionization mass spectrometry identified disulfide bonds between cysteine 1 (aa 127) and cysteine 2 (aa 144) and between cysteine 3 (aa 286) and cysteine 4 (aa 347). Further proteolysis of the cyanogen bromide-generated fragment containing cysteine 5 through cysteine 8, combined with mass spectrometry and Edman degradation, showed that disulfide bonds link cysteine 5 (aa 386) to cysteine 8 (aa 442) and cysteine 6 (aa 390) to cysteine 7 (aa 419). A similar disulfide bond arrangement is postulated to exist in gC homologs from other herpesviruses.

Herpes simplex virus (HSV) is an enveloped double-stranded DNA virus which encodes information for at least 11 glycoproteins, 10 of which are found on the virion envelope as well as on the surface of infected mammalian cells (52, 68, 69, 76). Because of their surface location, HSV glycoproteins act as major antigenic determinants for the cellular and humoral immune responses of the host (25, 48, 54, 61). In addition, three of the glycoproteins appear to modulate the immune response by interacting with components of the humoral immune system (15). Glycoproteins E and I (gE and gI) function as a complex to bind the Fc portion of immunoglobulin G (4, 5, 16, 19, 28, 37, 38, 58). gC binds the C3b and iC3b fragments of the third component of human complement (17, 20, 24, 30, 35, 36, 42, 49, 64, 65, 73). This activity protects the virus from complement-mediated attack (22). Although gE, gI, and gC are not absolutely required for infection in cell culture, gE and gI have been shown to play important roles in pathogenicity and virus spread (12, 15). HSV gC, as well as gC homologs in other herpesviruses, interacts with heparan sulfate proteoglycan, thereby mediating the first step of virus-cell interaction (31, 32, 45, 50, 55, 63, 69, 72). In addition, gC is rarely absent from clinical isolates (3, 21, 33), suggesting that gC's role in modulating the complement pathway and/or its role in virus attachment is important for viral pathogenesis in vivo.

The goal of this study was to extend our understanding of gC structure in order to relate that to its various functions. HSV gC1 contains 511 amino acids (aa), of which the first 25 constitute a cleavable signal sequence (23, 40). The protein has nine consensus sites for N-linked oligosaccharides (N-CHO) (23) as well as numerous O-linked oligosaccharides (O-CHO) (11, 39, 57). It is unknown if all nine N-CHO sites are utilized. It has been postulated that the O-CHO on gC1 are clustered near the N terminus (11, 57). gC2 contains 480 aa with seven potential N-CHO sites (14, 71). Little is known about the location of O-CHO present on gC2, although they have been demonstrated to exist (81). gC1 and gC2 are most homologous in the carboxy-terminal four-fifths of the protein. The spacing of five of the N-CHO sites and all eight cysteines are completely conserved in gC1 and gC2 from HSV (71).

Binding of purified gC1 and gC2 to C3b is dependent on protein conformation and maintenance of disulfide bonds (17), yet little else is known about the folded structure of the pro-

* Corresponding author. Mailing address: Department of Microbiology, School of Dental Medicine, University of Pennsylvania, 4010 Locust St., Philadelphia, PA 19104-6002. Phone: (215) 898-6553. Fax: (215) 898-8385. Electronic mail address: ahrux@biochem.dental.upenn.edu.

tein. In one study, native (nonreduced) gC1 was cleaved with cyanogen bromide (CNBr) and the fragments were separated and identified by sequencing (40). Cysteines 1 and 2 were found in one CNBr fragment, cysteines 3 and 4 were found in a second, and cysteines 5 through 8 were in a third. No data were presented to show whether any of the cysteines were disulfide bonded. Using the results of this study, we proposed a model consistent with disulfide bonds between Cys-1 (aa 127) and Cys-2 (aa 144) and between Cys-3 (aa 286) and Cys-4 (aa 347) (36). In addition, we postulated that these disulfide bonds help to stabilize the conformation of the four C3b binding regions (36) and the two major antigenic sites which surround these cysteines (47, 78). The disulfide bond status of Cys-5 through Cys-8 was not depicted. The region containing these four cysteines does not directly participate in C3b binding, nor is it recognized by a panel of monoclonal antibodies (MAbs) which bind to discontinuous epitopes (36, 78). However, deletion of this region appears to alter the conformational stability of gC1 (36). As a primary step in elucidating the conformation of gC, the present study emphasizes the determination of the disulfide bond arrangement.

To carry out these studies, we used truncated forms of gC1 and gC2 which were produced in the baculovirus-insect cell system (gCBac) (67, 72). This system allows for the production of the required quantities of gC for biochemical analysis. In a previous study (72), we described the expression and initial characterization of these recombinant proteins. They lack the transmembrane region, thus eliminating the need for solubilizing detergents and simplifying their purification from the culture system. Each truncated form of gC binds both to MAbs which recognize discontinuous epitopes and to C3b. These findings suggest that the respective conformations of the baculovirus-expressed forms of gC are similar to those of their full-length counterparts produced in HSV-infected cells. Thus, as gCBac should have the same disulfide bond arrangement as the wild-type virion protein, we used the recombinant proteins to solve the disulfide bond structure of gC. Our results show that the eight cysteines of gC form four disulfide bonds.

In addition, we obtained information concerning both the N-CHO and O-CHO on gCBac. In the case of gC1Bac, we found that most or all of the consensus sites for addition of N-CHO are modified by an approximately 1,000-Da oligosaccharide and that most of the O-CHO sites are located near the N terminus.

MATERIALS AND METHODS

Production and purification of gC1(457t), gC1(Δ33-123t), and gC2(426t). As described previously (72), plasmids pCP247 (gC1t), pCP248 (gC1-Δ33-123t), and pCP249 (gC2t), each encoding a truncated form of the gC gene, were recombined into baculovirus (*Autographa californica* nuclear polyhedrosis virus) by using Baculogold (Pharmingen) as the source of baculovirus DNA (67). The recombinant proteins were designated gC1(457t), gC1(Δ33-123t), and gC2(426t). Truncated proteins gC1(457t) and gC1(Δ33-123t) were purified from culture supernatants of baculovirus-infected cells by immunoaffinity chromatography (72). gC2(426t) was purified by heparin chromatography as previously described (72). These purification methods typically yielded 10 to 15 mg of purified protein per liter of infected cell medium. All experiments were carried out with purified proteins.

Antibodies. Rabbit polyclonal anti-gC sera R46 (17) and R81 (36) were used for detection of gC1 and gC2, respectively, on Western blots (immunoblots). R46 was prepared by immunizing with purified full-length gC1 from HSV, and R81 was prepared by immunizing with purified full-length gC2 from HSV.

Glycosidase treatment and Western blot analysis. Purified glycoproteins were digested with endoglycosidase F/N-glycosidase F (endo F), neuraminidase, and/or *O*-glycosidase (Boehringer Mannheim). One microgram of purified recombinant gC was incubated with 50 mU of endo F, 50 mU of *O*-glycosidase, and/or 0.5 mU of neuraminidase for 2 h at 37°C in 0.1 M sodium phosphate buffer (pH 7.3). Digests were examined by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) under reducing conditions on 8% acrylamide gels as previously described (10, 44). Proteins were transferred to nitrocellulose and probed first with either an anti-gC1 serum (R46) or an anti-gC2 serum (R81) and then with ^{125}I-protein A (ICN) (9).

CNBr digestion of gC. Cleavage at methionine residues was achieved by dissolving purified gC in 70% formic acid containing a 340-fold molar excess of CNBr (Pierce) over methionine residues (27). For example, 20 μg of gC was dissolved in 20 μl of a 5-mg/mL solution of CNBr in 70% formic acid. After 18 h in the dark at room temperature, maximal digestion was achieved, and samples were diluted 10-fold with water to stop the reaction and dried by evaporation. The final dilution-evaporation step was repeated twice to ensure removal of formic acid for electrophoresis.

Tricine SDS-PAGE analysis. Analysis of the CNBr-generated peptides was carried out with Tricine-SDS-gels (16% acrylamide; EMSCO) according to the method of Schägger and von Jagow (60), a gel system which separates low-molecular-weight polypeptides.

Carboxyamidomethylation and reduction of CNBr-generated peptides. For reaction with iodoacetamide, typically 10 μg of peptide was dissolved in 100 μl of 50 mM sodium phosphate buffer (pH 7.8) containing 1 mM EDTA. A twofold molar excess of iodoacetamide (Sigma) over cysteines was added (5.6 μl of 1 mM iodoacetamide), and the reaction was carried out in the dark at room temperature for 30 min under argon. To prepare reduced peptides, dithiothreitol (DTT; Sigma) was first added in a 50-fold molar excess over cysteines (2.6 μl of 50 mM DTT). Samples were heated for 4 h at 37°C under agron. Then, iodoacetamide was added as a twofold molar excess over thiols (5.7 μl of 50 mM iodoacetamide). After reaction with iodoacetamide, all samples were quickly frozen and stored at −80°C. Samples were desalted for mass spectrometry by reversed-phase high-pressure liquid chromatography (HPLC).

Enzyme digestion of CNBr-generated peptides. For endoprotease Lys-C (endo Lys-C) digestion, peptide samples were dissolved in 50 mM sodium phosphate buffer (pH 7.8). Endo Lys-C (sequencing grade; Sigma) was added at a 1:20 (wt/wt) enzyme/peptide ratio, and the mixture was heated for 18 h at 37°C. Samples were desalted for mass spectrometry by reversed-phase HPLC. For trypsin digestion, the peptide sample was dissolved in 100 mM Tris buffer pH 7.8. L-1-*p*-tosylamino-2-phenylethyl chloromethyl ketone-treated trypsin (Sigma) was added at a 1:25 (wt/wt) ratio, and the mixture was heated for 24 h at 37°C. The digest was purified by reversed-phase HPLC.

Reversed-phase HPLC. For HPLC, nonreduced CNBr peptides were redissolved in 0.1% trifluoroacetic acid and then separated on a 5-μm-particle size Vydac C_{18} column (4.6 by 250 mm) at 1.5 ml/min with a linear gradient of 5 to 45% acetonitrile (Fisher) in 0.1% trifluoroacetic acid. The system used consisted of two Waters 510 pumps, a WISP 700 auto injector, and a Waters 484 tunable absorbance detector set at 214 nm. The same system was used to desalt and purify enzymatic digests (see above).

Mass spectrometry. To determine the mass of recombinant proteins and peptides, samples were analyzed by mass spectrometry. Matrix-assisted laser desorption ionization mass spectrometry (34) was performed on a Fisons Instruments (Beverly, Mass.) VG TofSpec time-of-flight mass spectrometer (0.6-m flight tube) outfitted with an N_2 (337 nm, 4-nm pulse) laser. The accelerating voltage was set to 20 kV, and the detector voltage was set to 1.5 kV. Positive ion data were collected in the linear mode, and each spectrum was derived from the accumulation of 20 to 50 laser shots. External calibrations were performed with equine cytochrome *c*. External calibrant data were acquired into sample data. Data were analyzed with Fisons Instruments Opus software. Samples were dissolved in 50 μl of 50% acetonitrile containing 0.1% trifluoroacetic acid and diluted before analysis, using either α-cyano-4-hydroxycinnamic acid or 2-(4-hydroxyphenylazo)benzoic acid (Aldrich) as the matrix. The use of external calibrants potentially leads to a 0.1 to 0.5% error.

N-terminal sequencing. (i) Samples of gC1(457t), gC1(Δ33-123t), and gC2 (426t) were sequenced by adsorptive or covalent protocols on a Millipore ProSeq 6625 microscale protein/peptide N-terminal sequencer as previously detailed (46). The peptide fragment containing residues 359 to 419 (generated by successive cleavage with CNBr, endo Lys-C, and trypsin) was isolated by reverse-phase HPLC and covalently attached to a DITC-Sequelon (Millipore) membrane via the free amino group of the C-terminal lysine, using a previously described protocol (46). (ii) Other CNBr peptides were separated by SDS-PAGE, electroblotted to a ProBlott membrane (Perkin Elmer/Applied Biosystems), and stained with a 0.2% solution of Ponceau S in 0.1% acetic acid, and N-terminal sequencing was performed with a Perkin-Elmer/Applied Biosystems 473A microsequencer.

RESULTS

N-terminal sequencing of baculovirus-expressed proteins. The truncated forms of gCBac used in this study are shown in Fig. 1A. The predicted amino acid sequence for the first 15 aa of gC1(457t) is DLETASTGPTITAGA, and that for gC1(Δ33-123t), an N-terminal deletion mutant, it is DLETASTGPV WCDRR. The primary sequence obtained by automated Edman degradation for each protein agrees with the predicted sequence (data not shown) and indicates that the mellitin sig-

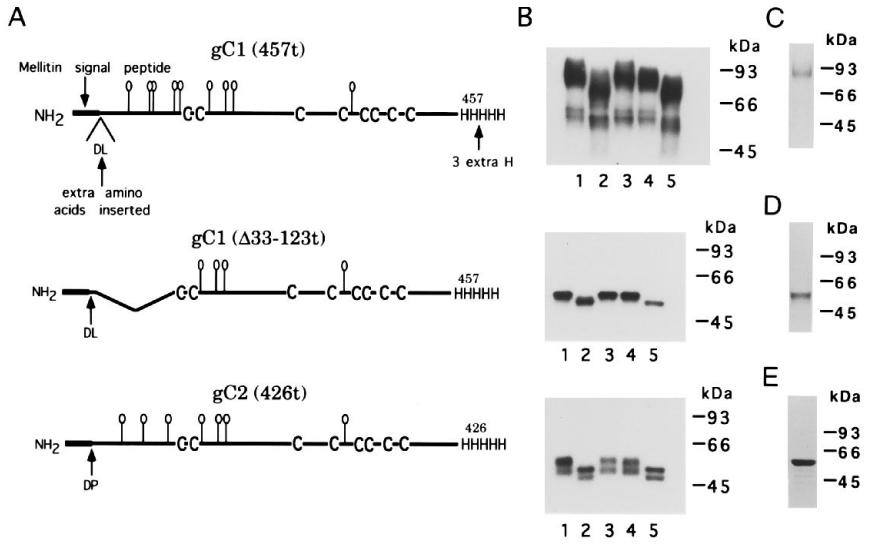


FIG. 1. (A) Schematic diagrams of truncated forms of HSV gC1 and gC2 produced by baculovirus-infected insect cells. In each case, the wild-type signal peptide of gC (23, 71) has been replaced with that of the honeybee mellitin signal. As a result of cloning, cleavage at the mellitin signal leaves two additional amino acids on the N terminus of gC, aspartic acid (D) and leucine (L) in gC1 constructs and aspartic acid and proline (P) in gC2. The consensus sites for N-CHO are shown as balloons. Both forms of gC1 were truncated at residue 457, and gC2 was truncated at residue 426. (B) Western blot analysis showing the effects of carbohydrate-modifying enzymes on each of the purified gC proteins. Each Western blot corresponds to the protein diagrammed to its left. Lanes: 1, untreated proteins; 2, proteins treated with endo F to remove N-linked glycans; 3, proteins treated with neuraminidase to remove sialic acid; 4, proteins treated with neuraminidase followed by *O*-glycosidase to remove O-linked glycans; 5, proteins treated with endo F followed by neuraminidase followed by *O*-glycosidase. Results similar to those in lanes 4 were observed upon *O*-glycosidase treatment alone (not shown). Polyclonal antibody R46 was used to detect gC1(457t) and gC1(Δ33-123t), and R81 was used to detect gC2(426t). (C to E) SDS-PAGE analysis of gC1(457t) (C), gC1(Δ33-123t) (D), and gC2(426t) (E). Purified glycoproteins were electrophoresed on 10% acrylamide gels and stained with Coomassie blue (C and D) or silver (E).

nal peptide was cleaved in the mature proteins. The first two amino acids of both proteins are DL, which result from cloning gC into pVTBac. In the case of gC1(Δ33-123t), the sequence data also confirmed that the internal deletion begins after residue 9 (P) as expected. For gC2(426t), the observed amino acid sequence for the first 12 residues was DPNASPGRTITV, which agreed with the predicted sequence. The amino acids DP at the N terminus resulted from cloning gC2 into pVTBac. In addition, a serine just prior to the glutamic acid at residue 3 of both proteins was removed as a result of the cloning. Attempts to sequence gC2 purified from HSV-infected cells failed because of a blocked N terminus (unpublished data). In the case of gC2 from the baculovirus expression system, it is unknown if the asparagine was not blocked because the protein was made in insect cells or because insertion of DP prior to the asparagine circumvented the blocked residue. This is the first time that the N terminus of gC2 protein has been sequenced.

Molecular mass and carbohydrate analysis of gCBac. To examine the sizes of the glycoproteins and probe the extent of glycosylation, gC1(457t), gC1(Δ33-123t), and gC2(426t) were treated with glycosidases and analyzed by SDS-PAGE and mass spectrometry. Figure 1B shows a Western blot analysis of the three proteins either untreated or treated with endo F, neuraminidase, and/or *O*-glycosidase. In addition, each of the proteins was separated by SDS-PAGE and stained for protein (Fig. 1C to E). In the case of gC1(457t), Western blot analysis showed that the untreated protein contained a major band migrating at approximately 90,000 Da and a minor band at approximately 60,000 Da (Fig. 1B, upper panel, lane 1). On the protein-stained gel (Fig. 1C), no band is seen at 60,000 Da, indicating that while this species reacts with anti-gC antibody, it represents an insignificant proportion of the purified protein sample. In the case of untreated gC1(Δ33-123t), only one band was seen on both the Western blot (Fig. 1B, middle panel, lane 1) and the protein-stained gel (Fig. 1D). For untreated gC2 (426t), the Western blot also contains two bands (Fig. 1B, lower panel, lane 1), but only one major band was seen (at 64,000 Da) on the protein-stained gel (Fig. 1E), again indicating that the faster-migrating species on the Western blot represents an insignificant proportion of the purified protein sample.

Endo F treatment increased the mobility of each protein (Fig. 1B, lanes 2), indicating the presence of N-CHO. Neuraminidase treatment (lanes 3), however, had no effect on mobility, indicating the absence of sialic residues on truncated gC. Neuraminidase followed by *O*-glycosidase treatment (lanes 4) slightly increased the mobility of gC1(457t) but had no effect on gC1(Δ33-123t) or gC2(426t). These results indicate that gC1(457t) contains O-CHO, but none were detected in gC1 (Δ33-123t) or gC2(426t). The data suggest that O-CHO are localized at the N terminus of gC1.

Mass spectrometry was used to determine the accurate molecular mass of each protein (Fig. 2; Table 1). The masses of gC1(457t), gC1(Δ33-123t), and gC2(426t) determined by mass spectrometry were much closer to the formula masses based on the amino acid sequences than those calculated from SDS-PAGE (Table 1). The 90-kDa form of gC1 had an anomalously slow mobility by SDS-PAGE, which has previously been ascribed to the high content of carbohydrate and proline (23). By mass spectrometry, only one peak (58,000 Da) was detected for gC1(457t), and the mass was 11,000 Da higher than the polypeptide formula mass. Most of this difference is accounted for by N-CHO, since treatment with endo F reduced the observed mass by 8,200 Da. The remainder is in part due to O-CHO, as determined from the finding that *O*-glycosidase increased the mobility of gC1(457t) on a denaturing gel. The broad peaks for gC1(457t) (Fig. 2A) are probably caused by heterogeneity of the carbohydrates.

The mass of gC1(Δ33-123t) was determined by mass spectrometry to be 42,500 Da, and its polypeptide formula mass is 38,300 Da. When gC1(Δ33-123t), which has four potential

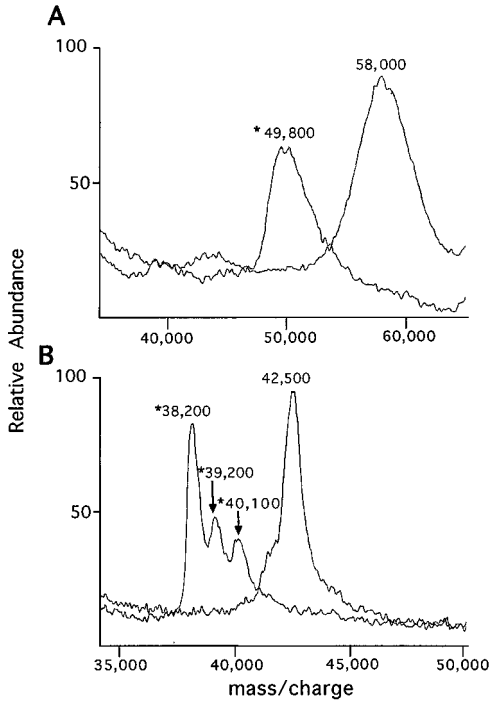


FIG. 2. Mass spectrometric analysis of gC1(457t) (A) and gC1(Δ33-123t) (B), with and without endo F treatment to remove N-linked glycans. The asterisks indicate endo F-treated species.

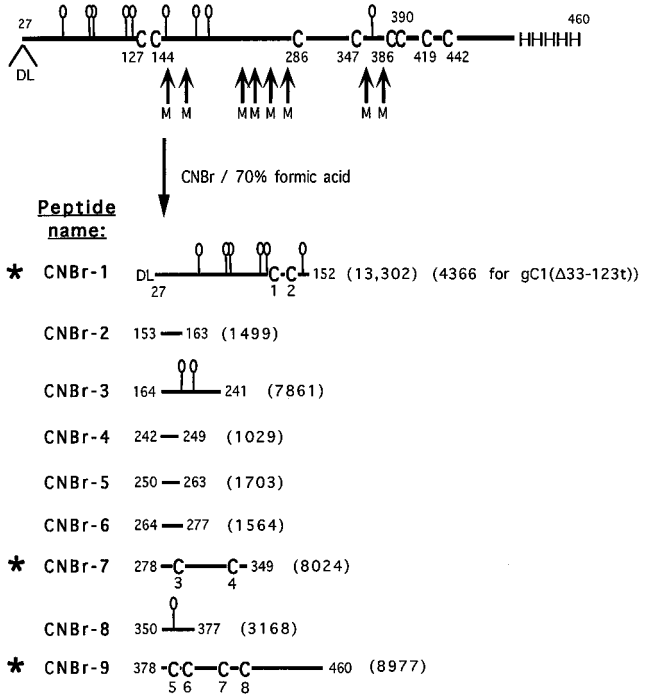


FIG. 3. Predicted CNBr cleavage fragments of gC1(457t) generated under reducing conditions. Residue 27 corresponds to the N-terminal amino acid of the mature gC1(457t) protein. Residues D and L which follow it result from cloning (see Materials and Methods). CNBr cleavage sites are indicated by the small arrows at the methionine (M) residues. The balloons indicate predicted sites for addition of N-CHO. The predicted CNBr fragments generated under reducing conditions are labeled CNBr-1 through CNBr-9, originating from the N terminus to the C terminus of the protein, respectively. The eight cysteines (C) are numbered 1 through 8. The polypeptide formula weight (from amino acids only) of each fragment is indicated in parentheses. The asterisks indicate the cysteine-containing fragments.

N-CHO sites, was treated with endo F, mass spectrometric analysis of the digest (Fig. 2B) revealed three peaks. The mass of peak 1 was the same as the polypeptide formula mass (Table 1), indicating that all N-CHO were removed. Peaks 1, 2, and 3 differed from each other in mass by approximately 1,000 Da. Assuming that each N-CHO on gC1(Δ33-123t) has an approximate mass of 1,000 Da, then the difference in each case is due to a single N-CHO. The untreated protein was approximately 4,000 Da larger than peak 1 of the endo F digest. This experiment suggests that each of the four potential sites is used and that virtually all of the secondary modifications on gC1(Δ33-123t) are N-CHO. Since treatment of gC1(457t) with endo F reduced the mass by 8,200 Da, and assuming that each N-CHO on gC1(457t) has a mass of approximately 1,000 Da, the endo F data suggest that eight or all nine N-CHO sites on gC1(457t) are used.

In the case of gC2(426t), the observed mass from the mass spectrometric analysis was approximately 7,000 Da larger than the polypeptide formula mass (Table 1). Since no mass spectrometric signal was obtained from the endo F-treated protein, we were unable to estimate how much of this difference could be attributed to the N-CHO present on gC2(426t).

TABLE 1. Molecular mass determinations and carbohydrate content

Sample	Mass (Da)			
	Formula[a]	SDS-PAGE[b]	Mass spectrometry[c]	N-CHO[d]
gC1(457t)	47,222	90,000	58,000	
Endo F-treated gC1(457t)	47,222	77,000	49,800	8,200
gC1(Δ33-123t)	38,216	58,000	42,500	
Endo F-treated gC1(Δ33-123t)				
Peak 1	38,216	54,000	38,200	4,300
Peak 2	38,216		39,200	3,300
Peak 3	38,216		40,100	2,400
gC2(426t)	43,974	64,000	51,300	ND

[a] Based on amino acid sequence, no carbohydrate included.
[b] Calculated from Fig. 1B.
[c] Data for gC1(457t) and gC1(Δ33-123t) are those indicated in Fig. 2. A similar experiment using endo F was performed with gC2(426t). However, no signal was detected by mass spectrometry following endo F treatment of the protein.
[d] Determined as the difference between the mass of the untreated protein (from mass spectrometry) and the mass of endo F-treated protein (from mass spectrometry). ND, not determined.

Strategy for cleavage of gC for further analysis. Our next goal was to cleave the protein between cysteines, separate the fragments, and compare them under nonreducing and reducing conditions to identify which contain disulfide bonds. CNBr cleaves between several of the cysteines of gC1 (Fig. 3) and generates a limited number of large fragments (40). Only three of these fragments, CNBr-1, CNBr-7, and CNBr-9, contain cysteine. For gC1(Δ33-123t), CNBr-1 is missing residues 33 to 123, including five of the potential N-CHO. However, this fragment does contain the first two cysteines.

CNBr cleavage of gC under nonreducing and reducing conditions. CNBr digests of purified gC1(457t) and gC1(Δ33-123t) were electrophoresed on an SDS–tricine–16% acrylamide gel (Fig. 4A). Lane 1 shows the control intact gC1(457t). Lanes 2 and 3 show gC1(457t) after treatment with CNBr in formic acid. When the sample was electrophoresed under nonreducing conditions (lane 2), four major bands, labeled A to D (approximate sizes are 40,000, 22,000, 12,000, and 6,000 Da), were

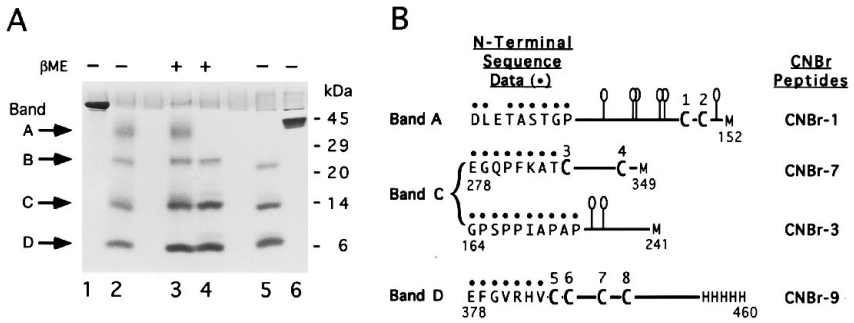


FIG. 4. CNBr cleavage of gC1(457t) and gC1(Δ33-123t). (A) Analysis of CNBr digests of gC1(457t) and gC1(Δ33-123t) run on an SDS–tricine–16% acrylamide gel under nonreducing (− βME [β-mercaptoethanol]) or reducing (+ βME) conditions. The gel was stained with Coomassie blue. Lanes: 1, untreated gC1(457t); 2 and 3, gC1(457t) digested with CNBr; 4 and 5, gC1(Δ33-123t) digested with CNBr; 6, untreated gC1(Δ33-123t). (B) Residues identified by N-terminal sequencing of bands in lane 2 of panel A are indicated by filled circles.

detected. When the sample was separated under reducing conditions (lane 3), there was no change in mobility of any of the four bands; thus, none of the peptides in the four bands are connected by disulfide bonds. Similar results were obtained for gC2(426t) (data not shown).

Treatment of gC1(Δ33-123t) with CNBr (Fig. 4A) resulted in three gel bands under both reducing (lane 4) and nonreducing (lane 5) conditions. The three bands correspond in size to bands B, C, and D of gC1(457t) in lanes 2 and 3. Band A was missing from the CNBr digest of gC1(Δ33-123t), suggesting that it corresponds to the N-terminal region of gC1(457t), i.e., CNBr-1. The data also suggest that Cys-1 and Cys-2, which are present in this fragment of gC1(457t), are not disulfide bonded to any other cysteines.

To identify the cysteine-containing peptides, the CNBr fragments of gC1(457t) obtained under nonreducing conditions were transferred to a polyvinylidene difluoride membrane and the N-terminal amino acids were sequenced. Sequences were obtained for three of the four bands. In this experiment, there was insufficient material to sequence band B. The residues identified are indicated by filled circles in Fig. 4B. Band A contained CNBr-1, which includes Cys-1 and Cys-2. Thus, the sequence data confirm the SDS-PAGE data which showed that digested gC1(Δ33-123t) was missing this band. Band C contained two N termini corresponding to CNBr-7, which contains Cys-3 and Cys-4, and CNBr-3 (no cysteines [Fig. 4B]). Band D contained CNBr-9, the C-terminal portion of gC which contains Cys-5 through Cys-8. Since bands, A, C, and D accounted for all eight cysteines, we did not pursue the identity of band B any further. We speculate that it represents a partial CNBr digest of gC1(457t).

These results show that the three cysteine-containing peptides are in separate bands and that none of the cysteine-containing fragments are disulfide bonded to each other. These findings confirm those of Kikuchi et al. (40), providing additional evidence that the structure of gC produced in insect cells (gCBac) is the same as that of gC expressed in mammalian cells.

N-CHO content of gC. gC1 has nine consensus N-CHO sites. Generation and identification of the CNBr fragments allowed us to obtain some information about which N-CHO sites are utilized. CNBr-1 has six N-CHO consensus sites, CNBr-3 has two sites, and CNBr-8 has one site (Fig. 3). Following CNBr cleavage, gC1(457t) and gC1(Δ33-123t) were treated with endo F and analyzed by SDS-PAGE (Fig. 5). In the case of gC1 (457t), the mobility of bands A, B, and C increased (compare lanes 2 and 3), indicating the presence of N-CHO on CNBr-1 and CNBr-3. Thus, CNBr-1 may contain as many as six N-linked glycans and CNBr-3 may contain up to two. As predicted from its sequence, the mobility of band D was unaffected by endo F treatment (lane 3). Similar results were seen for bands B, C, and D of gC1(Δ33-123t) (compare lanes 5 and 6). Interestingly, endo F treatment of this protein revealed an additional band migrating below band D (lane 6). The size of this band is consistent with the predicted size of the shortened form of CNBr-1 present in gC1(Δ33-123t).

Separation of CNBr fragments of gC1(457t) by HPLC. Reversed-phase HPLC was used to obtain the Cys-containing fragments of gC1(457t) in the absence of SDS (data not shown). Analysis of the individual HPLC peaks by SDS-PAGE (data not shown) confirmed that CNBr-1, CNBr-7, and CNBr-9 were separated by HPLC, allowing us to collect enough of these cysteine-containing fragments for further disulfide bond analysis.

Disulfide bond analysis of CNBr-7. Fragment CNBr-7 contains Cys-3 and Cys-4. SDS-PAGE showed (Fig. 4) that these cysteines are not disulfide bonded to any other cysteines in gC. To determine if Cys-3 is bonded to Cys-4, the isolated peptide was reacted with iodoacetamide, which adds only to cysteines that are not in a disulfide bond. The strategy was to use mass spectrometry to accurately detect the shift in peptide mass due to the addition of iodoacetamide to cysteine (i.e., a mass shift assay). Thus, if Cys-3 is not disulfide bonded by Cys-4, each free thiol should react with iodoacetamide and cause the mass

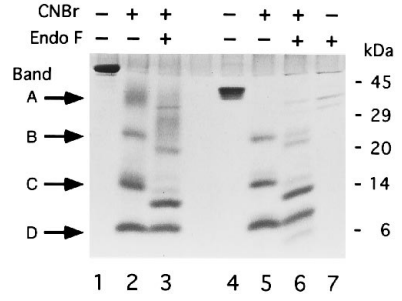


FIG. 5. Endo F treatment of CNBr fragments of gC1(457t) and gC1(Δ33-123t). gC1(457t) (lanes 1 to 3) and gC1(Δ33-123t) (lanes 4 to 6) were first cleaved with CNBr and then treated with endo F and analyzed on an SDS–tricine–16% acrylamide gel under nonreducing conditions. The gel was stained with Coomassie blue. Lanes: 1 and 4, untreated protein; 2 and 5, proteins were cleaved with CNBr; 3 and 6, proteins were cleaved with CNBr, and the cleaved fragments were then treated with endo F; 7, endo F alone (no gC).

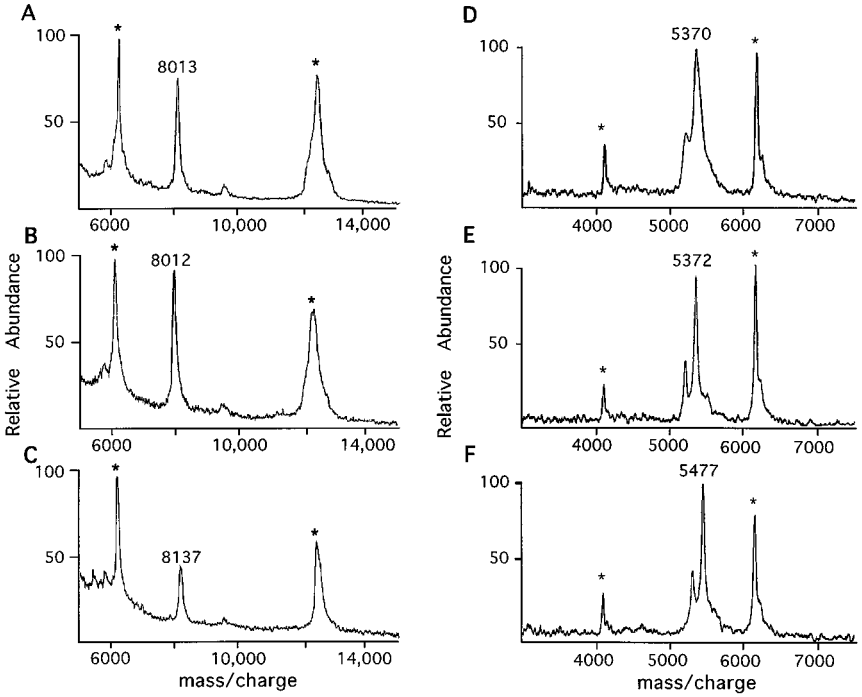


FIG. 6. Mass spectrometric analysis of fragment CNBr-7 of gC1(457t) and fragment CNBr-1 of gC1(Δ33-123t) treated with iodoacetamide. (A) Untreated CNBr-7; (B) CNBr-7 treated with iodoacetamide; (C) CNBr-7 reduced with DTT prior to treatment with iodoacetamide; (D) untreated CNBr-1 of gC1(Δ33-123t); (E) CNBr-1 treated with iodoacetamide; (F) CNBr-1 reduced with DTT prior to treatment with iodoacetamide. In panels A to C, the mass for CNBr-7 is indicated; in panels D to F, the mass of the most abundant peak of CNBr-1 is indicated. In each case, a small shoulder which was 145 Da less than the mass of the main peak was also seen, suggesting that the former was a different processed form of CNBr-1. Theoretical masses of polypeptide components (in daltons) are as follows: (A) 8,024; (B) 8,024 (assuming that a disulfide bond is present) or 8,140 (assuming that no disulfide bond present); (C) 8,140; (D) 4,341; (E) 4,341 (assuming that a disulfide bond is present) or 4,457 (assuming that no disulfide bond present); (F) 4,457. In panels D to F, the difference between the observed mass and the theoretical mass in approximately 1,000 Da in each case, probably representing the mass of one N-CHO. Asterisks indicate peaks due to the singly charged (at 12,360), doubly charged (at 6,180), or triply charged (at 4,120) forms of the cytochrome *c* standard.

of the peptide to increase by 116 Da, from 8,024 to 8,140 Da. If Cys-3 is disulfide bonded to Cys-4, then the mass of CNBr-7 should remain the same upon treatment with iodoacetamide. As a control, CNBr-7 was first reacted with DTT to force reduction of any disulfide bond, and then iodoacetamide was added to produce the expected increase in mass.

Untreated CNBr-7 exhibited a sharp peak in the mass spectrum, with an observed mass of 8,013 Da (Fig. 6A). After treatment with iodoacetamide, the mass of CNBr-7 remained the same (Fig. 6B), indicating that no free thiols were present. We conclude that Cys-3 is disulfide bonded to Cys-4. In the control experiment, in which CNBr-7 was treated with DTT and then with iodoacetamide (Fig. 6C), the observed mass was 8,137 Da, confirming that the method was sensitive enough to detect the carboxyamidomethylation of free thiols. Theoretical masses for this experiment are given in the legend to Fig. 6.

Disulfide bond analysis of CNBr-1. A similar approach was attempted with CNBr-1 of gC1(457t), which contains Cys-1 and Cys-2. The fragment was isolated by HPLC and analyzed by mass spectrometry. The mass spectrum contained a very weak, very broad peak with an approximate mass of 19,000 Da (data not shown). The broadness and poor quality of the ion signal were probably due to the heterogeneity and extent of glycosylation in the glycopeptide. This fragment contains N-CHO (Fig. 5, lane 3) and probably contains a number of O-CHO (Fig. 1). In an attempt to sharpen the signal, the peptide was treated with endo F to remove the N-CHO. The endoglycosidase-treated peptide still displayed a very broad peak. The low quality of the ion signal for this fragment precluded a mass spectrometric approach in this case.

To enable a mass spectrometric approach for this region, we used CNBr-1 obtained from the N-terminal deletion mutant, gC1(Δ33-123t). This mutant form of gC is predicted to have the same disulfide bond structure as gC1(457t) because it recognizes MAbs to discontinuous epitopes and binds to complement component C3b as does gC1(457t) (72). CNBr-1 of gC1(Δ33-123t) contains both Cys-1 and Cys-2 but is missing five of the six N-CHO consensus sites which are present in CNBr-1 of gC1(457t). Reversed-phase HPLC was used to isolate this fragment from other CNBr fragments of gC1(Δ33-123t) (data not shown). This fragment comigrated with CNBr-9 on a denaturing gel as seen previously (Fig. 5, lane 5), and its mobility increased with endo F treatment as in Fig. 5, lane 6. N-terminal sequencing of the fragment confirmed its identity (data not shown). CNBr-1 of gC1(Δ33-123t) had a stronger and sharper mass spectral signal (Fig. 6D) than did CNBr-1 from gC1(457t) because of the lower oligosaccharide content. As expected, the mass of this peptide was larger (by 1,000 Da) than the mass predicted by the sum of the amino acid masses alone (Fig. 6, legend) because of the presence of the N-CHO. Reaction of CNBr-1 with iodoacetamide did not increase the mass of the fragment (Fig. 6E); therefore, cysteine 1 is disulfide bonded to cysteine 2. In the control experiment, CNBr-1 was first treated with DTT prior to reaction with iodoacetamide (Fig. 6F). The fragment showed the expected increase in mass.

Disulfide bond analysis of Cys-5 through Cys-8. Fragment CNBr-9 contains four cysteines, and so reaction with iodoacetamide was carried out to determine how many disulfide bonds are present in this peptide (Table 2, experiment 1). If there were not disulfide bonds in CNBr-9, four molecules of

TABLE 2. Summary of mass spectrometric data for CNBr-9

Sample	Mass (Da)		
	Observed	Theoretical[a]	% Difference
Expt 1			
CNBr-9	8,962	8,977	0.17
CNBr-9 + IA[b]	8,948	8,977[c]	0.32
CNBr-9 + DTT + IA	9,197	9,209	0.13
Expt 2			
CNBr-9	9,012	8,977	0.39
CNBr-9 + endo Lys-C	9,037	8,995[d]	0.47
CNBr-9 + endo Lys-C + DTT + IA	3,572	3,580	0.22
	5,641	5,647	0.11

[a] Mass of the polypeptide component.
[b] IA, iodoacetamide.
[c] Mass of the polypeptide if two disulfide bonds are present. If one disulfide bond is present, then the theoretical mass is 9,093 Da. If no disulfide bonds are present, the theoretical mass is 9,209 Da.
[d] Mass of the polypeptide if two disulfide bonds which do not connect Cys-5 to Cys-6 and Cys-7 to Cys-8 are present. This mass is 18 Da greater than that of untreated CNBr-9, since a molecule of water adds to the peptide upon proteolytic cleavage. If disulfide bonds link Cys-5 to Cys-6 and Cys-7 to Cys-8, then cleavage with endo F results in two peptides with theoretical masses of 3,580 and 5,647 Da.

iodoacetamide would be added to the fragment, increasing the mass from 8,977 to 9,209 Da. If it contained one disulfide bond, two molecules of iodoacetamide would be added; if two disulfide bonds were present, there would be no change in mass. Since there was no change in the mass of the fragment upon reaction with iodoacetamide (Table 2, experiment 1), we conclude that the fragment contains two disulfide bonds. The control sample which was reduced prior to carboxyamidomethylation showed the expected increase in mass.

There are three ways in which the two disulfide bonds in CNBr-9 could be arranged. One possibility is that Cys-5 is linked to Cys-6 and Cys-7 is linked to Cys-8. If this were the case, cleavage of CNBr-9 between Cys-6 and Cys-7 with endo Lys-C would yield two smaller peptides (Fig. 7B). We cleaved CNBr-9 with endo Lys-C and measured the mass of the untreated and treated fragments (Table 2, experiment 2). As a control to ensure that the enzyme was cleaving at the predicted site, CNBr-9 was reacted with endo Lys-C, and then DTT was added to force reduction of the disulfide bonds. Finally, iodoacetamide was added to make the reduction irreversible. As expected, cleavage occurred, since two peptides having the predicted masses were detected (Table 2, experiment 2). However, treatment of CNBr-9 with endo Lys-C under nonreducing conditions yielded only one fragment which had the same mass as the untreated fragment. The fact that the two peptides remained disulfide linked indicates that Cys-5 is not disulfide bonded to Cys-6. Therefore, the disulfide pattern must be either Cys-5 to Cys-7 and Cys-6 to Cys-8 or vice versa. To distinguish between these two possibilities, the ideal method would be to use a protease that cleaved between each cysteine. However, there are only three amino acids between Cys-5 and Cys-6, and none of these are substrates for specific proteases that are readily available. Therefore, we used a less direct approach employing Edman degradation.

The endo Lys-C digest of CNBr-9 was isolated by HPLC and then further digested with trypsin (Fig. 7C). This fragment was repurified by HPLC (data not shown). The resulting fragment was predicted to consist of three sequences, peptides A, B, and C, connected by two disulfide bonds arranged according to either model I or model II in Fig. 7C. The fragment was covalently linked to a DITC-Sequelon membrane through the C-terminal lysine's (aa 410) free amino group in peptide B and subjected to automated Edman degradation. As the sequencing proceeded, we expected to see three N termini. As sequencing progressed past cycle 4 (Cys-5), we expected one of the sequences to drop out, as a result of release of the peptide from the membrane. For example, if Cys-5 were disulfide

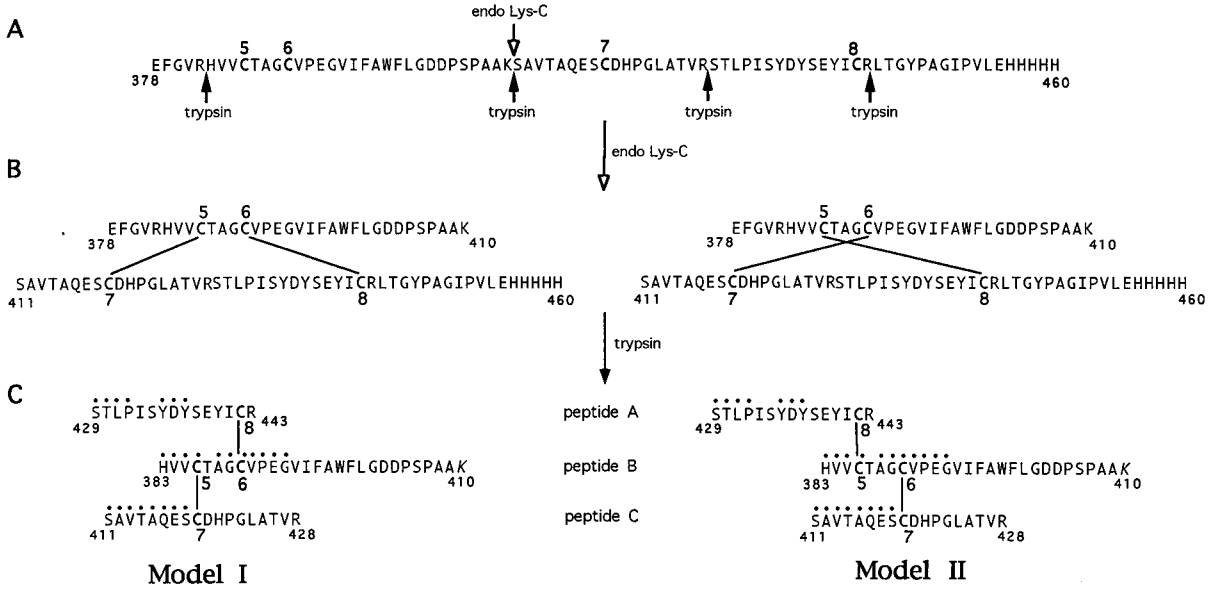


FIG. 7. Proteolysis of CNBr-9. (A) Sequence of fragment CNBr-9 containing cysteines 5 through 8. Endo Lys-C and trypsin cleavage sites are indicated by open and filled arrows, respectively. In panels B and C, models I and II represent two possible disulfide bond arrangements of Cys-5 through Cys-8. CNBr-9 was cleaved first with endo Lys-C at residue 410 (B) and then with trypsin at residues 378, 428, and 443 (C). The disulfide-bonded fragment (aa 383 to 443) generated by trypsin cleavage (C) contains peptides A, B, and C as shown. This fragment was covalently attached to a membrane via the lysine residue (*K*, aa 410), and N-terminal sequencing identified the residues indicated by filled circles.

FIG. 8. Hypothetical model of gC1 protein structure. Residue 26 is the N-terminal amino acid of the mature protein (minus the signal peptide). The four disulfide bonds are indicated in red as follows: C1 (aa 127) to C2 (aa 144), C3 (aa 286) to C4 (aa 347), C5 (aa 386) to C8 (aa 442), and C6 (aa 390) to C7 (aa 419). Hexagonal orange balloons represent the predicted N-CHO sites at residues 42, 70, 74, 108, 109, 148, 181, 197, and 362. The small green circles represent the predicted O-CHO sites based on a neural network algorithm by Hansen et al. (29). The four regions important for C3b binding are shown in blue: I (aa 124 to 137), II (aa 276 to 292), III (aa 339 to 366), and IV (aa 223 to 246) (36). Antigenic sites I (aa 307 to 373) and II (aa 129 to 247) (shown in purple) were determined by Glorioso et al. (47, 78).

bonded to Cys-7 and Cys-6 were disulfide bonded to Cys-8 (model I), then sequence C would drop out after cycle 4. If the disulfide pattern were Cys-5–Cys-8 and Cys-6–Cys-7 (model II), then sequence A would drop after cycle 4. The sequencing data showed that peptides A, B, and C were all present in the first four cycles of degradation, indicated by filled circles in Fig. 7C. The presence of phenylthiohydantoin-derived dehydroserine in cycle 4 is consistent with a cysteine in peptide B. The expected residues from peptide C were also present in cycles 5 through 8, consistent with model II. In contrast, residues from peptide A were absent in cycles 5 and 6, indicating that this peptide had dropped out, consistent with model II. Residues from peptide B were present in cycles 6 to 12, indicating that this peptide remained attached to the membrane for at least 12 cycles. In cycles 7 to 11, several extraneous residues which could not be accounted for were present. However, on the basis of the data for cycles 1 to 6, we believe that model II (Cys-5 to Cys-8 and Cys-6 to Cys-7) is the most likely disulfide bond arrangement for these four cysteines. The overall disulfide bond arrangement of gC1 is shown in Fig. 8.

DISCUSSION

Several functions have been attributed to HSV gC. First, gC mediates the initial interaction between virus and cells (69). Second, gC's interaction with complement component C3b plays an important role in modulation of the immune response against HSV (22, 30, 49). Finally, the protein is a major inducer of both humoral and cellular immunity, and immunization of animals with gC affords some protection against a lethal virus challenge (62). The functions of gC are conformation dependent. Disulfide bonds are essential for the stability of many proteins (75) and occur frequently in membrane-bound proteins with significant ectodomains (13), e.g., HSV gD (46), and also in secreted proteins. However, until this study, little was known about the disulfide bond arrangement of HSV gC or other gC homologs (40).

Disulfide bond arrangement of gC. Chemical modification of cysteine residues followed by mass spectrometric analysis has recently been used to identify disulfide bonds (2, 51, 80). Herein we provide evidence that the eight cysteine residues in gC1 form four disulfide bonds. The disulfide bond arrangement is cysteine 1 (aa 127) to cysteine 2 (aa 144), cysteine 3 (aa 286) to cysteine 4 (aa 347), cysteine 5 (aa 386) to cysteine 8 (aa 442), and cysteine 6 (aa 390) to cysteine 7 (aa 419). These disulfide pairs have been incorporated into a structural model of gC (Fig. 8). The first two disulfide bonds produce loops between adjacent cysteines within the C3b binding domain (36). The last two disulfide bonds result in an extended loop which may stabilize the carboxy terminus. Thus, the assignment of disulfide bonds in gC1 allows us to construct a more accurate model of the folded structure. We postulate that the same arrangement of disulfide bonds is present in gC2, given (i) the fact that the number and spacing of cysteine residues are completely conserved between the two proteins and (ii) our observation that the CNBr cleavage pattern of gC2 is similar to that of gC1 (data not shown).

Domains of gC. Several laboratories, including our own, have set out to define the structural domains of HSV gC involved in its function(s). Towards this end, mutagenesis techniques have been used to construct and study the properties of

mutant forms of gC. Glorioso and coworkers (47, 78) utilized MAbs to generate MAb-resistant (*mar*) mutants. They showed that the amino acids changes in those mutants cluster into two distinct antigenic sites on gC1 which are in the central portion of the protein (purple regions in Fig. 8).

We used linker insertion mutagenesis to identify four noncontiguous regions, I to IV (Fig. 8), within the central domain of gC1 which are important for C3b binding (36). Three of these regions correspond to similar regions in gC2 (64). We also showed that the interaction between gC and C3b is highly dependent on the proper conformational structure of gC (17, 36). Assignment of the disulfide bond structure allows us to explain some of our data on C3b binding. The disulfide bond between Cys-1 and Cys-2 produces a loop which forms C3b binding region I (Fig. 8). While C3b binding regions II and III are separated by 46 aa, they are held in close proximity by the disulfide bond between cysteine 3 and cysteine 4 and so may act together to bind C3b. The first two disulfide bonds are critical for maintaining the structure of the C3b binding domain, since mutagenesis of either Cys-2 or Cys-4 to serine produces proteins which no longer bind to C3b (36). These Cys mutants also failed to recognize some MAbs which bind to discontinuous epitopes on gC, a further indication that these disulfide bonds are important in maintaining the native, functional conformation of gC. Also, a previous study (21) showed that a *mar* mutant of gC with an additional cysteine inserted between Cys-1 and Cys-2 did not bind C3b. Insertion of a cysteine in the C3b binding site formed by a loop between Cys-1 and Cys-2 probably disrupted the binding site, perhaps by scrambling the disulfide bonds.

Previous reports have shown that the N-terminal region of gC (prior to Cys-1) is important for efficient attachment to cells (72). In addition, the ability of gC1 to interfere with properdin binding to C3b was mapped to the N terminus of gC1 (35). This interference is postulated to be the reason for the down regulation of the complement cascade by gC (35).

While no function has been associated with the region of gC containing Cys-5 through Cys-8, this region is important in maintaining the native, functional conformation of the protein. gC which lacks this region binds less well to C3b than the wild-type protein, and a virus mutant containing this deletion cannot be neutralized by MAbs to gC (35). In addition, the spacing of cysteines 5 through 8 is completely conserved among the alphaherpesviruses (18).

Disulfide bond structure of gC homologs. Fitzpatrick et al. (18) aligned gC sequences of seven alphaherpesviruses (1, 6, 41, 59, 71) reported up to that time and noted that 51 of the 65 residues (78%) conserved in all seven sequences were in the carboxy-terminal half of the proteins. The alignment showed that this region of each gC homolog contained six invariant cysteines around which were clustered other conserved residues. Fitzpatrick et al. (18) noted that this pattern of conservation suggests that the gC homologs from all alphaherpesviruses have the same disulfide bond arrangement for the six cysteines corresponding to Cys-3 through Cys-8 of HSV gC. According to their alignment, Cys-1 and Cys-2 of the gC homologs from equine herpesvirus 1, varicella-zoster virus, and Marek's disease virus do not align with Cys-1 and Cys-2 of gC from HSV. However, we note that the number of amino acids between Cys-1 and Cys-2 is the same in these homologs as in HSV gC. Thus, we propose that three of the disulfide bonds are arranged identically in the ectodomain of these homologs and that the disulfide bond between Cys-1 and Cys-2 is also present but displaced. However, the gC homolog from bovine herpesvirus 1 has an additional cysteine near the N terminus. The effect of this Cys on the disulfide bond pattern of this protein is unknown. Clearly, biochemical analysis is needed to verify our predictions.

N-CHO of gC. gC1 from HSV-infected cells contains N-linked glycans (8) which are proposed to be predominately di- and triantennary complex N-CHO (66). However, little was known about which of the nine potential N-CHO sites are occupied. We constructed gC1 and gC2 mutants which lack the final N-CHO consensus site and showed that this site is normally occupied in gC1 (aa 362) and gC2 (aa 331) (36). Endo F treatment of CNBr-1 and CNBr-3 of gC1(457t) indicates that one or more N-CHO sites on each of these fragments are occupied (Fig. 5). In addition, we showed that the potential N-CHO site at residue 148 is occupied, since endo F treatment of the first CNBr fragment of gC1(Δ33-123t), which contains only one N-CHO consensus site, increased the mobility of the fragment on SDS-PAGE (Fig. 5, lane 6). Moreover, the mass spectrometric data indicate that at least eight of the nine N-CHO consensus sites on gC1(457t) are utilized. If an N-CHO site on gC1(457t) is unoccupied, it must be within the N-terminal region, since all of the N-CHO consensus sites on gC1(Δ33-123t) are used. Our data also indicate that the N-CHO on gC1Bac are approximately 1,000 Da, a size consistent with N-glycosidically linked oligosaccharide structures reported for insect cells (26, 43, 53, 77). The N-linked glycans on proteins produced in insect cells are commonly shorter and less complex than those produced in mammalian cells (7, 26, 43, 70, 79).

O-CHO of gC. Our data allow speculation about the contribution of O-CHO to the structure of gC produced in insect cells. The mass of endo F-treated gC1(457t) (49,800 Da [Table 1]) minus the formula mass of the protein from the amino acid sequence (47,222 Da) yields a mass of 2,600 Da contributed by other posttranslational modifications. *O*-Glycosidase treatment indicated that O-CHO are present on gC1(457t) but not on gC1(Δ33-123t). Thus, most or all of the O-CHO are located within residues 33 to 123 of gC1(457t). This finding agrees well with studies which proposed that most of the O-CHO on gC1 produced in mammalian cells are clustered in the N-terminal region (11, 56).

Previous studies identified the structures GalNAc (203 Da) and Gal-GalNAc (365 Da) present in insect cell produced O-CHO (74, 77), where GalNAc is *N*-acetylgalactosamine and Gal is galactose. Considering that the O-CHO on gC1(457t) contribute a maximum of 2,600 Da, we speculate that 7 to 13 O-CHO are present on gC1(457t).

Recently, Hansen et al. (29) developed an artificial neural network algorithm to recognize the pattern of O-CHO sites on mucin-type mammalian glycoproteins based on primary sequence alone. This method predicts 29 O-CHO sites for gC1 from HSV (Fig. 8). Our data suggest that no more that 13 of these sites are actually occupied in gC1(457t) produced by insect cells. In addition, 83% percent of the predicted O-CHO sites for gC1 are in the first 25% of the sequence. This observation agrees with the fact that we detected no O-CHO on gC1 (Δ33-123t). It is also consistent with studies of gC1 from HSV-infected mammalian cells (11, 56). The number of O-CHO predicted, coupled with their clustered locations in the amino-terminal portion of gC, supports the idea that this region of gC1 has a mucin-like structure (11, 39, 56, 72).

gC2 from HSV-infected cells contains O-linked oligosaccharides (81), although none were detected on gC2Bac. The neural network predicts only 11 occupied O-CHO sites for gC2, and only 36% of these are located prior to the first cysteine. Thus, gC2 may not possess the mucin-like domain of gC1. This difference may account for some of the differences in the biological properties of the two proteins. For, example, gC1 but not gC2 inhibits the binding of properdin to C3b (35). This

function has been mapped to the N-terminal region of gC1 (35), and the proteins have only 30% amino acid homology in this region (71).

The assignment of disulfide bonds in gC allows us to construct more accurate models of its tertiary structure. Further structural analysis of gC awaits solution of the three-dimensional structure by X-ray crystallography. Efforts are under way in our laboratory to crystallize gCBac.

ACKNOWLEDGMENTS

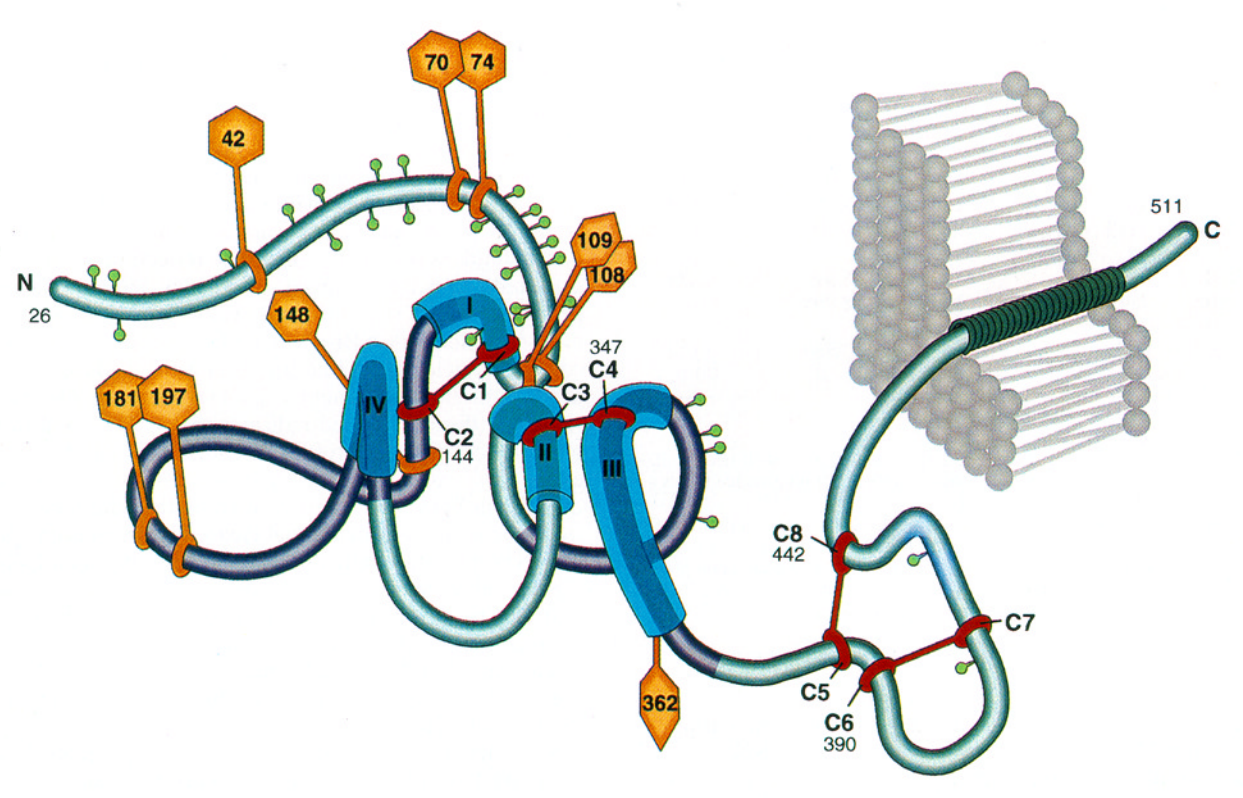
This investigation was supported by Public Health Service grants HL-28220 from the National Heart, Lung, and Blood Institute and NS-30606 from the National Institute of Neurological Diseases and Stroke. Automated Edman degradation was provided by the Biopolymer Analysis Laboratory (W.R.A.) of the School of Dental Medicine, supported by Public Health Service grant DE-08239 from the National Institute of Dental Research. Mass spectrometry and additional automated Edman degradation were provided by the Protein Chemistry Laboratory (W.T.M. and J.D.L.) of the School of Medicine of the University of Pennsylvania, supported by core grants of the Diabetes and Cancer Centers (Public Health Service grants DK-19525 and CA-16520). A.H.R. is a postdoctoral fellow supported by Public Health Service grants AI-07324 and AI-07278 from the National Institute of Allergy and Infectious Diseases.

The color drawing of gC was prepared by Denise Hansen of Graphicae, Inc., Philadelphia, Pa. We thank Lynn Spruce and Yang Wang for excellent technical assistance.

REFERENCES

1. **Allen, G. P., and L. D. Coogle.** 1988. Characterization of an equine herpesvirus type 1 gene encoding a glycoprotein (gp13) with homology to herpes simplex virus glycoprotein C. J. Virol. **62:**2850–2858.
2. **Angeletti, R. H., L. Bibbs, L. F. Bonewald, G. B. Fields, J. S. McMurray, W. T. Moore, and J. T. Stults.** 1995. Formation of a disulfide bond in an octreotide-like peptide: a multicenter study, p. 261–274. *In* D. R. Marshak (ed.), Techniques in protein chemistry, vol. VII. Academic Press, San Diego, Calif.
3. **Arvin, A. M., C. M. Koropchak, A. S. Yeager, and L. Pereira.** 1983. Detection of type-specific antibody to herpes simplex virus type 1 by radioimmunoassay with herpes simplex virus type 1 glycoprotein C purified with monoclonal antibody. Infect. Immun. **40:**184–189.
4. **Baucke, R., and P. Spear.** 1979. Membrane proteins specified by herpes simplex viruses. V. Identification of an Fc-binding glycoprotein. J. Virol. **32:** 779–789.
5. **Bell, S., M. Cranage, L. Borysiewicz, and T. Minson.** 1990. Induction of immunoglobulin G Fc receptors by recombinant vaccinia viruses expressing glycoproteins E and I of herpes simplex virus type 1. J. Virol. **64:**2181–2186.
6. **Binns, M. M., and N. L. J. Ross.** 1989. Nucleotide sequence of the Marek's disease virus (MDV) RB-1B A antigen gene and the identification of the MDV A antigen as the herpes simplex virus-1 glycoprotein C homologue. Virus Res. **12:**371–382.
7. **Butters, T. D., and R. C. Hughes.** 1981. Isolation and characterization of mosquito cell membrane glycoproteins. Biochim. Biophys. Acta **640:**655–671.
8. **Campadelli-Fiume, G., and F. Serafini-Cessi.** 1985. Processing of the oligosaccharide chains of herpes simplex virus type 1 glycoproteins, p. 357–382. *In* B. Roizman (ed.), The herpesviruses, vol. 3. Plenum Press, New York.
9. **Cohen, G. H., B. Dietzschold, M. Ponce de Leon, D. Long, E. Golub, A. Varrichio, L. Pereira, and R. J. Eisenberg.** 1984. Localization and synthesis of an antigenic determinant of herpes simplex virus glycoprotein D that stimulates production of neutralizing antibody. J. Virol. **49:**102–108.
10. **Cohen, G. H., V. J. Isola, J. Kuhns, P. W. Berman, and R. J. Eisenberg.** 1986. Localization of discontinuous epitopes of herpes simplex virus glycoprotein D: use of a nondenaturing ("native" gel) system of polyacrylamide gel electrophoresis coupled with Western blotting. J. Virol. **60:**157–166.
11. **Dall'Olio, F., N. Malagolini, V. Speziali, G. Campadelli-Fiume, and F. Serafini-Cessi.** 1985. Sialylated oligosaccharides O-glycosidically linked to glycoprotein C from herpes simplex virus type 1. J. Virol. **56:**127–134.
12. **Dingwell, K. S., L. C. Doering, and D. C. Johnson.** 1995. Glycoproteins E and I facilitate neuron-to-neuron spread of herpes simplex virus. J. Virol. **69:** 7087–7089.
13. **Doms, R. W., R. A. Lamb, J. K. Rose, and A. Helenius.** 1993. Folding and assembly of viral membrane proteins. Virology **193:**545–562.
14. **Dowbenko, D. J., and L. A. Lasky.** 1984. Extensive homology between the herpes simplex virus type 2 glycoprotein F gene and the herpes simplex virus type 1 glycoprotein C gene. J. Virol. **52:**154–163.
15. **Dubin, G., N. O. Fishman, R. J. Eisenberg, G. H. Cohen, and H. M. Friedman.** 1992. The role of the herpes simplex viruses in immune evasion, p. 111–120. *In* B. T. Rouse (ed.), Herpes simplex viruses. Springer-Verlag, New York.
16. **Dubin, G., I. Frank, and H. M. Friedman.** 1990. Herpes simplex virus type 1 encodes two Fc receptors which have different binding characteristics for monomeric immunoglobulin G (IgG) and IgG complexes. J. Virol. **64:**2725–2731.
17. **Eisenberg, R. J., M. Ponce de Leon, H. M. Friedman, L. F. Fries, M. M. Frank, J. C. Hastings, and G. H. Cohen.** 1987. Complement component C3b binds directly to purified glycoprotein C of herpes simplex virus types 1 and 2. Microb. Pathog. **3:**423–435.
18. **Fitzpatrick, D. R., L. A. Babiuk, and T. J. Zamb.** 1989. Nucleotide sequence of bovine herpesvirus type 1 glycoprotein gIII, a structural model for gIII as a new member of the immunoglobulin superfamily, and implications for the homologous glycoproteins of other herpesviruses. Virology **173:**46–57.
19. **Frank, I., and H. M. Friedman.** 1989. A novel function of the herpes simplex virus type 1 Fc receptor: participation in bipolar bridging of antiviral immunoglobulin G. J. Virol. **63:**4479–4488.
20. **Friedman, H. M., G. H. Cohen, R. J. Eisenberg, C. A. Seidel, and D. B. Cines.** 1984. Glycoprotein C of herpes simplex virus 1 acts as a receptor for the C3b complement component on infected cells. Nature (London) **309:**633–635.
21. **Friedman, H. M., J. C. Glorioso, G. H. Cohen, J. C. Hastings, S. L. Harris, and R. J. Eisenberg.** 1986. Binding of complement component C3b to glycoprotein C of herpes simplex virus type 1: mapping of gC-binding sites and demonstration of conserved C3b binding in low-passaged clinical isolates. J. Virol. **60:**470–475.
22. **Fries, L. F., H. M. Friedman, G. H. Cohen, R. J. Eisenberg, C. H. Hammer, and M. M. Frank.** 1986. Glycoprotein C of herpes simplex virus 1 is an inhibitor of the complement cascade. J. Immunol. **137:**1636–1641.
23. **Frink, R. J., R. J. Eisenberg, G. H. Cohen, and E. K. Wagner.** 1983. Detailed analysis of the portion of the herpes simplex virus type 1 genome encoding glycoprotein C. J. Virol. **45:**634–647.
24. **Ghosh-Choudhury, N., M. Butcher, and H. P. Ghosh.** 1990. Expression from cloned DNA of biologically active glycoprotein C of herpes simplex virus type 1 in mammalian cells. J. Gen. Virol. **71:**689–699.
25. **Glorioso, J., U. Kees, G. Kümel, H. Kirchner, and P. H. Krammer.** 1985. Identification of herpes simplex virus type 1 (HSV-1) glycoprotein gC as the immunodominant antigen for HSV-1-specific memory cytotoxic T lymphocytes. J. Immunol. **135:**575–582.
26. **Grabenhorst, E., B. Hofer, M. Nimtz, V. Jäger, and H. S. Conradt.** 1993. Biosynthesis and secretion of human interleukin-2 glycoprotein variants from baculovirus-infected Sf21 cells. Eur. J. Biochem. **215:**189–197.
27. **Gross, E.** 1967. The cyanogen bromide reaction. Methods Enzymol. **11:**238–255.
28. **Hanke, T., F. L. Graham, V. Lulitanand, and D. C. Johnson.** 1990. Herpes simplex virus IgG Fc receptors induced using recombinant adenovirus vectors expressing glycoproteins E and I. Virology **177:**437–444.
29. **Hansen, J. E., O. Lund, J. Engelbrecht, H. Bohr, J. O. Nielsen, J.-E. S. Hansen, and S. Brunak.** 1995. Prediction of O-glycosylation of mammalian proteins: specificity patterns of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase. Biochem. J. **308:**801–813.
30. **Harris, S. L., I. Frank, A. Yee, G. H. Cohen, R. J. Eisenberg, and H. M. Friedman.** 1990. Glycoprotein C of herpes simplex virus type 1 prevents complement-mediated cell lysis and virus neutralization. J. Infect. Dis. **162:** 331–337.
31. **Herold, B. C., S. I. Gerber, T. Polonsky, B. J. Belval, P. N. Shaklee, and K. Holme.** 1995. Identification of structural features of heparin required for inhibition of herpes simplex virus type 1 binding. Virology **206:**1108–1116.
32. **Herold, B. C., D. WuDunn, N. Soltys, and P. G. Spear.** 1991. Glycoprotein C of herpes simplex virus type 1 plays a principal role in the adsorption of virus to cells and in infectivity. J. Virol. **65:**1090–1098.
33. **Hidaka, Y., S. Sakuma, Y. Kumano, H. Minagawa, and R. Mori.** 1990. Characterization of glycoprotein C-negative mutants of herpes simplex virus type 1 isolated from a patient with keratitis. Arch. Virol. **113:**195–207.
34. **Hillenkamp, F., M. Karas, R. C. Beavis, and B. T. Chait.** 1991. Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. Anal. Chem. **63:**1193A–1203A.
35. **Hung, S.-L., C. Peng, J. Lambris, H. M. Friedman, R. J. Eisenberg, and G. H. Cohen.** 1994. The interaction of glycoprotein C of herpes simplex virus types 1 and 2 with the alternative complement pathway C3 convertase. Virology **203:**299–312.
36. **Hung, S.-L., S. Srinivasan, H. M. Friedman, R. J. Eisenberg, and G. H. Cohen.** 1992. The structural basis of C3b binding by glycoprotein C of herpes simplex virus. J. Virol. **66:**4013–4027.
37. **Johnson, D. C., and V. Feenstra.** 1987. Identification of a novel herpes simplex virus type 1-induced glycoprotein which complexes with gE and binds immunoglobulin. J. Virol. **61:**2208–2216.
38. **Johnson, D. C., M. C. Frame, M. W. Ligas, A. M. Cross, and N. D. Stow.** 1988. Herpes simplex virus immunoglobulin G Fc receptor activity depends on a complex of two viral glycoproteins, gE and gI. J. Virol. **62:**1347–1354.
39. **Johnson, D. C., and P. G. Spear.** 1983. O-linked oligosaccharides are ac-

quired by herpes simplex virus glycoproteins in the Golgi apparatus. Cell **32:**987–997.
40. **Kikuchi, G. E., J. E. Coligan, T. C. Holland, M. Levine, J. C. Glorioso, and R. Nairn.** 1984. Biochemical characterization of peptides from herpes simplex virus glycoprotein gC: loss of CNBr fragments from the carboxy terminus of truncated, secreted gC molecules. J. Virol. **52:**806–815.
41. **Kinchington, P. R., J. Remenick, J. M. Ostrove, S. E. Straus, W. Ruyechan, and J. Hay.** 1986. Putative glycoprotein gene of varicella-zoster virus with variable copy numbers of a 42-base-pair repeat sequence has homology to herpes simplex virus glycoprotein C. J. Virol. **59:**660–668.
42. **Kubota, Y., T. A. Gaither, J. Cason, J. J. O'Shea, and T. J. Lawley.** 1987. Characterization of the C3 receptor induced by herpes simplex virus type 1 infection of human epidermal, endothelial, and A431 cells. J. Immunol. **134:**1137–1142.
43. **Kuroda, K., H. Geyer, R. Geyer, W. Doerfler, and H.-D. Klenk.** 1990. The oligosaccharides of influenza virus hemagglutinin expressed in insect cells by a baculovirus vector. Virology **174:**418–429.
44. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680.
45. **Liang, X., L. A. Babiuk, and T. J. Zamb.** 1993. Mapping of heparin-binding structures on bovine herpesvirus I and pseudorabies virus gIII glycoproteins. Virology **194:**233–243.
46. **Long, D., W. C. Wilcox, W. R. Abrams, G. H. Cohen, and R. J. Eisenberg.** 1992. Disulfide bond structure of glycoprotein D of herpes simplex virus types 1 and 2. J. Virol. **66:**6668–6685.
47. **Marlin, S. D., T. C. Holland, M. Levine, and J. C. Glorioso.** 1985. Epitopes of herpes simplex virus type 1 glycoprotein gC are clustered in two distinct antigenic sites. J. Virol. **53:**128–136.
48. **Marsden, H. S.** 1987. Herpes simplex virus glycoproteins and pathogenesis, p. 259–288. *In* W. C. Russell and J. W. Almond (ed.), Molecular basis of virus disease. Fortieth symposium of the Society for General Microbiology. Cambridge University Press, Cambridge.
49. **McNearney, T. A., C. Odell, V. M. Holers, P. G. Spear, and J. P. Atkinson.** 1987. Herpes simplex virus glycoproteins gC-1 and gC-2 bind to the third component of complement and provide protection against complement-mediated neutralization of viral infectivity. J. Exp. Med. **166:**1525–1535.
50. **Mettenleiter, T. C., L. Zsak, F. Zuckermann, N. Sugg, H. Kern, and T. Ben-Porat.** 1990. Interaction of glycoprotein gIII with a cellular heparinlike substance mediates adsorption of pseudorabies virus. J. Virol. **64:**278–286.
51. **Ming, D., J. L. Markley, and G. Hellekont.** 1995. Quantification of cysteinyl sulfhydryl residues in peptides and proteins. Peptide Res. **8:**113–114.
52. **Muggeridge, M. I., S. R. Roberts, V. J. Isola, G. H. Cohen, and R. J. Eisenberg.** 1990. Herpes simplex virus, p. 459–481. *In* M. H. V. Van Regenmortel and A. R. Neurath (ed.), Immunochemistry of viruses, vol. II. The basis for serodiagnosis and vaccines. Elsevier Biochemical Press, Amsterdam.
53. **Naim, H. Y., and H. Koblet.** 1992. Asparagine-linked oligosaccharides of Semiliki Forest virus grown in mosquito cells. Arch. Virol. **122:**45–60.
54. **Norrild, B.** 1985. Humoral response to herpes simplex virus infections, p. 69–86. *In* B. Roizman and C. Lopez (ed.), The herpesviruses. Immunobiology and prophylaxis of human herpesvirus infections, vol. 4. Plenum Press, New York.
55. **Okzaki, K., T. Matsuzaki, Y. Sugahara, J. Okada, Y. Iwamura, M. Ohnishi, T. Kanno, M. Shimizu, E. Honda, and Y. Kono.** 1991. BHV-1 adsorption is mediated by the interaction of glycoprotein gIII with heparinlike moiety on the cell surface. Virology **181:**666–670.
56. **Olofsson, S.** 1992. Carbohyhdrates in herpesvirus infections. APMIS **100:** 84–95.
57. **Olofsson, S., I. Sjöblom, M. Lundström, S. Jeansson, and E. Lycke.** 1983. Glycoprotein C of herpes simplex virus type 1: characterization of *O*-linked oligosaccharides. J. Gen. Virol. **64:**2735–2747.
58. **Para, M. F., R. B. Baucke, and P. G. Spear.** 1980. Immunoglobulin G (Fc)-binding receptors on virions of herpes simplex virus type 1 and transfer of these receptors to the cell surface by infection. J. Virol. **34:**512–520.
59. **Robbins, A. K., R. J. Watson, M. E. Whealy, W. W. Hays, and L. W. Enquist.** 1986. Characterization of a pseudorabies virus glycoprotein gene with homology to herpes simplex virus type 1 and type 2 glycoprotein C. J. Virol. **58:**339–347.
60. **Schägger, H., and G. von Jagow.** 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal. Biochem. **166:**369–379.
61. **Schmid, D. S., and B. T. Rouse.** 1992. The role of T cell immunity in control of herpes simplex virus. Curr. Top. Microbiol. Immunol. **179:**57–74.
62. **Schrier, R. K., L. Pizer, and J. Moorhead.** 1983. Type-specific delayed hypersensitivity and protective immunity induced by isolated herpes simplex virus glycoprotein. J. Immunol. **130:**1413–1418.
63. **Sears, A. E., B. S. McGwire, and B. Roizman.** 1991. Infection of polarized MDCK cells with herpes simplex virus 1: two asymmetrically distributed cell receptors interact with different viral proteins. Proc. Natl. Acad. Sci. USA **88:**5087–5091.
64. **Seidel-Dugan, C., M. Ponce de Leon, H. M. Friedman, R. J. Eisenberg, and G. H. Cohen.** 1990. Identification of C3b-binding regions on herpes simplex virus type 2 glycoprotein C. J. Virol. **64:**1897–1906.
65. **Seidel-Dugan, C., M. Ponce de Leon, H. M. Friedman, L. F. Fries, M. M. Frank, G. H. Cohen, and R. J. Eisenberg.** 1988. C3b receptor activity on transfected cells expressing glycoprotein C of herpes simplex virus types 1 and 2. J. Virol. **62:**4027–4036.
66. **Serafini-Cessi, F., F. Dall'Olio, L. Pereira, and G. Campadelli-Fiume.** 1984. Processing of N-linked oligosaccharides from precursor- to mature-form herpes simplex virus type 1 glycoprotein gC. J. Virol. **51:**838–844.
67. **Sisk, W. P., J. D. Bradley, R. J. Leipold, A. M. Stoltzfus, M. Ponce de Leon, M. Hilf, C. Peng, G. H. Cohen, and R. J. Eisenberg.** 1994. High-level expression and purification of secreted forms of herpes simplex virus type 1 glycoprotein gD synthesized by baculovirus-infected insect cells. J. Virol. **68:**766–775.
68. **Spear, P. G.** 1985. Antigenic structure of herpes simplex viruses, p. 425–446. *In* M. V. H. van Regenmortel and A. R. Neurath (ed.), Immunochemistry of viruses. The basis for serodiagnosis and vaccines. Elsevier Science Publishers B. V., Amsterdam.
69. **Spear, P. G.** 1993. Entry of alphaviruses into cells. Semin. Virol. **4:**167–180.
70. **Svoboda, M., M. Przybylski, J. Schreurs, A. Miyajima, K. Hogeland, and M. Deinzer.** 1991. Mass spectrometric determination of glycosylation sites and oligosaccharide composition of insect-expressed mouse interleukin-3. J. Chromatogr. **562:**403–419.
71. **Swain, M. A., R. W. Peet, and D. A. Galloway.** 1985. Characterization of the gene encoding herpes simplex virus type 2 glycoprotein C and comparison with the type 1 counterpart. J. Virol. **53:**561–569.
72. **Tal-Singer, R., C. Peng, M. Ponce de Leon, W. R. Abrams, B. W. Banfield, F. Tufaro, G. H. Cohen, and R. J. Eisenberg.** 1995. The interaction of herpes simplex virus glycoprotein gC with mammalian cell surface molecules. J. Virol. **69:**4471–4483.
73. **Tal-Singer, R., C. Seidel-Dugan, L. Fries, H. P. Huemer, R. J. Eisenberg, G. H. Cohen, and H. M. Friedman.** 1991. Herpes simplex virus glycoprotein C is a receptor for complement component iC3b. J. Infect. Dis. **164:**750–753.
74. **Thomsen, D. R., L. E. Post, and A. P. Elhammer.** 1990. Structure of O-glycosidically linked oligosaccharides synthesized by the insect cell line SF9. J. Cell. Biochem. **43:**67–79.
75. **Thornton, J. M.** 1981. Disulphide bridges in globular proteins. J. Mol. Biol. **151:**261–287.
76. **Ward, P., and B. Roizman.** 1994. Herpes simplex genes: the blueprint of a successful human pathogen. Trends Genet. **10:**267–274.
77. **Wathen, M. W., P. A. Aeed, and A. P. Elhammer.** 1991. Characterization of oligosaccharide structures on a chimeric respiratory syncytial virus protein expressed in insect cell line Sf9. Biochemistry **30:**2863–2868.
78. **Wu, C.-T. B., M. Levine, F. Homa, S. L. Highlander, and J. C. Glorioso.** 1990. Characterization of the antigenic structure of herpes simplex virus type 1 glycoprotein C through DNA sequence analysis of monoclonal antibody-resistant mutants. J. Virol. **64:**856–863.
79. **Yeh, J.-C., J. R. Seals, C. I. Murphy, H. van Halbeek, and R. D. Cummings.** 1993. Site-specific N-glycosylation and oligosaccharide structures of recombinant HIV-1 gp120 derived from a baculovirus expression system. Biochemistry **32:**11087–11099.
80. **Zaluzec, E. J., D. A. Gage, and J. T. Watson.** 1994. Quantitative assessment of cysteine and cystine in peptides and proteins following organomercurial derivatization and analysis by matrix assisted laser desorption ionization mass spectrometry. J. Am. Soc. Mass Spectrom. **5:**359–366.
81. **Zezulak, K. M., and P. G. Spear.** 1984. Mapping of the structural gene for the herpes simplex virus type 2 counterpart of herpes simplex virus type 1 glycoprotein C and identification of a type 2 mutant which does not express this glycoprotein. J. Virol. **49:**741–747.


70
74
42
109
108
148
N
26
I
347
C4
C1
C3
181
197
IV
C2
144
II
III
511
C
C8
442
C7
C5
C6
390
362