Amino-Terminal Sequence of Glycoprotein D of Herpes Simplex Virus Types 1 and 2

ROSELYN J. EISENBERG,^{1,2*} DEBORAH LONG,² RUTH HOGUE-ANGELETTI,³ AND GARY H. COHEN^{2,4}

Department of Pathobiology, School of Veterinary Medicine,¹ Department of Microbiology⁴ and Center for Oral Health Research,² School of Dental Medicine, and Department of Pathology, School of Medicine,³ University of Pennsylvania, Philadelphia, Pennsylvania 19104

Received 18 April 1983/Accepted 27 September 1983

Glycoprotein D (gD) of herpes simplex virus is a structural component of the virion envelope which stimulates production of high titers of herpes simplex virus type-common neutralizing antibody. We carried out automated N-terminal amino acid sequencing studies on radiolabeled preparations of gD-1 (gD of herpes simplex virus type 1) and gD-2 (gD of herpes simplex virus type 2). Although some differences were noted, particularly in the methionine and alanine profiles for gD-1 and gD-2, the amino acid sequence of a number of the first 30 residues of the amino terminus of gD-1 and gD-2 appears to be quite similar. For both proteins, the first residue is a lysine. When we compared our sequence data for gD-1 with those predicted by nucleic acid sequencing, the two sequences could be aligned (with one exception) starting at residue 26 (lysine) of the predicted sequence. Thus, the first 25 amino acids of the predicted sequence are absent from the polypeptides isolated from infected cells.

Herpes simplex virus (HSV), a large enveloped DNA virus, contains four distinct glycoproteins designated gB, gC, gD, and gE (21). These molecules, which are found in the virion envelope, differ in electrophoretic mobility in sodium dodecyl sulfate-polyacrylamide gels, biological properties, and genetic map positions (11, 13, 17-19, 21, 22). At least two of these proteins, gB and gD, appear to be very similar antigenically in the oral (HSV type 1 [HSV-1]), and genital (HSV type 2 [HSV-2]) types of the virus (2, 5, 6, 8, 9, 21).

We previously showed that purified gD isolated from either HSV-1 (gD-1) or HSV-2 (gD-2) stimulates the production of high titers of type-common virus-neutralizing antibody (2, 5, 10). It is therefore likely that gD plays an important role in virus infection. Furthermore, gD appears to be one of the targets of the immune response to HSV infections, at both the humoral and the cellular level (18, 21). Previous studies in our laboratory have shown that gD-1 and gD-2, though clearly not identical in structure, do appear to share a number of tryptic peptides (8) and antigenic determinants (9). Our overall conclusion from structural studies of the two glycoproteins was that the antigenic relatedness was probably matched by a high degree of amino acid sequence similarity.

Recently, Watson and co-workers (22) carried out nucleic acid sequencing studies of the protein-coding region corresponding to gD-1 of HSV-1 strain Patton. To evaluate the degree of amino acid similarity in gD-1 and gD-2, it will be necessary to carry out a similar analysis for gD-2 as well as other strains of both HSV-1 and HSV-2. In the present report, we carried out N-terminal amino acid sequencing studies of radiolabeled gD-1 and gD-2. These studies show that the sequences of the first 30 amino acids of gD-1 and gD-2 are very similar but not identical.

The general procedure used for labeling gD was to infect cells with either HSV-1 or HSV-2 and then to metabolically label the cells with the particular radioactive amino acid. For methionine, arginine, and lysine, a 15-min pulse carried out

at 6 h postinfection was sufficient to obtain enough radioactive label incorporated into gD for sequencing (Fig. 1A). Under these conditions of labeling, most of the radioactivity was found in the precursor forms (2, 3, 8) of gD-1 (53,000 daltons) and gD-2 (52,000 daltons) (Fig. 1A). For alanine incorporation into gD-1 and leucine incorporation into both gD-1 and gD-2, it was necessary to label for an additional 2 h to get a sufficient amount of labeled gD. Under these conditions of labeling, both the precursor and product forms (3, 8) of the glycoproteins were labeled (Fig. 1B, lanes 1 to 3). At the end of the labeling period, cytoplasmic extracts were prepared and immunoprecipitated with a polyclonal antibody prepared against purified gD-1 (2, 10). To carry out sequencing studies of labeled tyrosine, gD-1 and gD-2 were purified from infected cell extracts by immunoabsorbant chromatography (10), and the purified proteins were iodinated with ¹²⁵I, using the chloramine T procedure (12).

Figure 1 shows a sodium dodecyl sulfate-polyacrylamide gel electrophoresis analysis of the radiolabeled preparations used for automated N-terminal sequencing. It can be seen that when metabolic labeling was employed (Fig. 1A and B), over 95% of the radioactive label was present in either the precursor or product form (or both) of gD-1 and gD-2. In the case of iodinated gD-1 (Fig. 1C), some label was present in lower-molecular-weight polypeptides. We determined by tryptic peptide analysis that these lower-molecular-weight polypeptides represented various portions of the N-terminal region of gD-1 (G. H. Cohen, D. Long, and R. J. Eisenberg, unpublished data). It is not clear whether these fragments of gD were generated as a result of iodination or were due to proteolytic digestion of purified gD-1 which occurred before iodination. A similar result was obtained for iodinated gD-2 (data not shown).

Figures 2 and 3 show profiles of automated Edman degradations of radiolabeled gD-1 and gD-2, respectively. The sequences derived from these profiles are shown in Table 1. The profiles of [³H]lysine (Fig. 2B and 3B) and ¹²⁵I-labeled gD (presumably tyrosine; Fig. 2F and 3F) were identical for gD-1 and gD-2. The data indicate that the N-terminal amino acid for both glycoproteins is lysine. Differences were noted

* Corresponding author.

TABLE 1. Sequences derived from profiles of automated Edman degradations of radiolabeled gD-1 and gD-2

Protein (technique)	Amino acid sequence at indicated residue no. ^a										
	(26) 1	(27) 2	(28) 3	(29) 4	(30) 5	(31) 6	(32) 7	(33) 8	(34) 9	(35) 10	(36) 11
gD-1 (nucleic acid sequencing)	Lys	Tyr	Ala	Leu	Ala	Asp	Ala	Ser	Leu	Lys	Met
gD-1 (Edman degradation)	Lys	Tyr	Ala	Leu	Ala	X	Ala	Met	Leu	Lys	Met
gD-2 (Edman degradation)	Lys	Tyr	Ala	Leu	Ala	X	X	X	X	Lys	Met

FIG. 1. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis analysis of gD-1 and gD-2 preparations used for N-terminal sequencing. Fluorograms of 10% diallyltartardiamide cross-linked polyacrylamide gels (3, 8–10) are shown. Conditions for the growth of KB and BHK cells and preparation and titration of virus have been previously described (2, 3, 5, 8–10). For infection, an input multiplicity of 20 PFU of HSV-1 (HF) or 10 PFU of HSV-2 (SAVAGE) per cell was employed. For each radioactive label shown (A, lanes 1 to 6; B, lane 4), a 75-cm² bottle of confluent KB or BHK cells was infected with HSV-1 or HSV-2. At 2 h postinfection, the cells were overlaid with Eagle minimal medium containing 1/10 the normal concentration of methionine, arginine, or lysine. Pulse-labeling (2) was carried out at 6 h postinfection by incubating infected cells for 15 min in 4.5 ml of Hanks salts containing one of the following isotopes: [³⁵S]methionine (specific activity, 600 Ci/mmol; 1 mCi); [2,3-³H]arginine (specific activity, 15 Ci/mmol; 1 mCi); or [4,5-³H]lysine (specific activity, 60 to 80 Ci/mmol; 1 mCi). The monolayers were washed with iced saline and lysed (2), and cytoplasmic extracts were prepared as described previously (2, 3, 8–10). For each radioactive label shown in (B), lanes 1 to 3, infected cells were pulse-labeled at 6 h postinfection for 15 min in Hanks salts and then overlaid with Eagle minimal medium and incubated at 37°C for an additional 2 h. The following radioisotopes were used: [4,5-³H]leucine (specific activity, 50 Ci/mmol; 1 mCi) and [3-³H]alanine (specific activity, 75 Ci/mmol; 500 μ Ci). Each sample was immunoprecipitated with anti-gD-1 serum (10; see below). To determine the positions of tyrosine residues, gD-1 and gD-2 were each purified by immunoabsorbant chromatography (8), and 15 μ g of each protein was iodinated by the chloramine T procedure (12). Sodium dodecyl sulfate-polyacrylamide gel electrophoresis was carried out in slabs of 10% acrylamide cross-linked with 0.4% *N,N'*-diallyltartardiamide by essentially the same method described by Spear (20). After electrophoresis, the gels were stained with Coomassie brilliant blue (2), dried on filter paper, and exposed to Kodak XAR-5 film. For fluorography, the gels were impregnated with 2,5-diphenyloxazole, placed in contact with X-ray film, and stored at -70°C (1). (A) Lane 1, gD-1 labeled with [³⁵S]methionine; lane 2, gD-2 labeled with [³⁵S]methionine; lane 3, gD-1 labeled with [³H]arginine; lane 4, gD-2 labeled with [³H]arginine; lane 5, gD-1 labeled with [³H]lysine; lane 6, gD-2 labeled with [³H]lysine. (B) For lanes 1 to 3, the same procedure was followed as described above, except that after the 15-min pulse in Hanks salts, the cells were overlaid with Eagle minimal medium and incubated at 37°C for 2 h. Lane 4, the same procedure was followed as in (A). Lane 1, gD-1 labeled with [³H]leucine; lane 2, gD-2 labeled with [³H]leucine; lane 3, gD-1 labeled with [³H]alanine; lane 4, gD-2 labeled with [³H]alanine. (C) gD-1 was purified (8) and iodinated with ¹²⁵I (12). The iodinated product was immunoprecipitated with anti-gD-1 serum (10).

in the methionine and alanine profiles for gD-1 and gD-2. In each of these cases, however, several residues were present in both glycoproteins, and one or more residues were present in one and possibly missing in the other. Thus, for example, in the case of alanine (Fig. 2E and 3E), both proteins were found to have alanine at residues 3, 5, and 12. However, only gD-1 contained an alanine at position 7. Both proteins had methionine residues at position 11 (Fig. 2A and 3A), but only gD-1 had a methionine residue at position 8. In certain cases, it was difficult to interpret the data unambiguously owing to high backgrounds of radioactivity. For example, both proteins appeared to contain arginine residues at positions 16 and 18. For gD-1, position 20 was unambiguously determined to be a lysine residue. For gD-2, position 20 could be either a lysine or an arginine (compare Fig. 3B and C). For both proteins, the leucine profiles (Fig. 2D and 3D) also exhibited a high background of radioactivity. This may have been due to the very long labeling time required to obtain sufficient incorporation of this particular amino acid label. However, in spite of the high background, the [³H]leucine peaks for gD-1 appear to correlate precisely with the positions of leucine in the deduced amino acid sequence (22) (Table 1). For gD-2, the only leucine residue that appears to be unambiguous is at position 4.

Table 1 shows that our data for gD-1 can be aligned fairly well with the deduced amino acid sequence (22) of gD-1, beginning at residue 26 of the deduced sequence. One difference between our sequence and the deduced sequence is at residue 8 (residue 33 of the deduced sequence). Our data indicate that gD-1 (strain HF) contains a methionine residue at that position; however, gD-2 (strain SAVAGE) did not. The residue predicted by nucleic acid sequencing (with strain Patton of HSV-1) is a serine. The differences noted at this position might be due to strain and type variation. However, an alteration from a methionine to the serine predicted would require three base changes.

Our data indicate that the first 25 amino acids of the predicted sequence are not present in the protein as isolated from infected cells. This stretch of amino acids is largely hydrophobic, the only exceptions being an arginine at residues 7 and 24 and a histidine at residue 21. It has been proposed (22) that the first 20 amino acids in this stretch might constitute a signal peptide (16) which is commonly present at the amino terminus of many transmembrane glycoproteins.

In addition to the profiles shown in Fig. 2 and 3, we also used [2-³H]mannose and [³⁵S]cysteine as radioactive probes for sequence analysis of gD-1. In both cases, no radioactivity was detected in the first 30 residues (data not shown). According to the deduced amino acid sequence for gD-1, the first cysteine would be expected to occur at residue 66, and the first asparagine that has the appropriate sequence (Asn-X-Thr or Ser) to be a glycosylation site (15) would be expected to occur at residue 94. Thus, our negative data correlate with what would be predicted from the sequence of gD-1. In other studies, we showed that both gD-1 and gD-2 contained three *N*-asparagine-linked tryptic glycopeptides

TABLE 1—Continued

Amino acid sequence at indicated residue no. ^a																			
(37) 12	(38) 13	(39) 14	(40) 15	(41) 16	(42) 17	(43) 18	(44) 19	(45) 20	(46) 21	(47) 22	(48) 23	(49) 24	(50) 25	(51) 26	(52) 27	(53) 28	(54) 29	(55) 30	
Ala	Asp	Pro	Asn	Arg	Phe	Arg	Gly	Lys	Asp	Leu	Pro	Val	Leu	Asp	Gln	Leu	Thr	Asp	
Ala	X	X	X	Arg	X	Arg	X	Lys	X	Leu	X	X	Leu	X	X	Leu	X	X	
Ala	X	X	X	Arg	X	Arg	X	Lys (Arg)	X	X	X	X	X	X	X	X	X	X	

^a The numbers in parentheses refer to residue numbers according to nucleic acid sequencing (22). The numbers outside parentheses refer to residue numbers obtained from Edman degradation of radiolabeled gD-1 and gD-2. See Fig. 2 and 3 for data.

FIG. 2. Amino acid sequence analysis of gD-1. For preparation of samples for sequencing, each cytoplasmic extract was immuno-precipitated with anti-gD-1 serum (10). *Staphylococcus aureus* Cowan I (IgSorb; New England Enzyme Center) was employed to collect antigen-antibody complexes (3). The precipitates were washed, and the antigen-antibody complexes were disrupted as described previously (2). A portion was analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Bovine serum albumin (100 μ g) was added to the remainder, and the protein was precipitated with 25% trichloroacetic acid at 4°C for 18 h. The precipitates were collected by centrifugation at $13,000 \times g$ for 30 min, dissolved in 1 ml of 0.1 N NaOH, dialyzed exhaustively against distilled water, and lyophilized. A similar procedure was employed to immunoprecipitate iodinated gD-1 and gD-2. Stepwise Edman degradation of radiolabeled gD-1 and gD-2 was accomplished in a Beckman 890 B protein sequencer (7, 14). The lyophilized radiolabeled samples were dissolved in water and mixed with 50 mmol of sperm whale myoglobin. The samples taken at each step were dried, resuspended in 100 μ l of acetone, and transferred to scintillation vials. The tubes were washed with an additional 50 μ l of acetone and 100 μ l of ethyl acetate, and the vials were dried over N_2 and analyzed by scintillation counting. The samples containing ^{125}I were analyzed directly in a gamma counter. In each run, the positions of all labeled and several unlabeled steps were confirmed by high-pressure liquid chromatography of the myoglobin carrier protein-derived amino acids. (A) [^{35}S]methionine; (B) [3H]lysine; (C) [3H]arginine; (D) [3H]leucine; (E) [3H]alanine; (F) ^{125}I -labeled gD-1 (tyrosine).

FIG. 3. Amino acid sequence analysis of gD-2. (A) [^{35}S]methionine; (B) [3H]lysine; (C) [3H]arginine; (D) [3H]leucine; (E) [3H]alanine; (F) ^{125}I -labeled gD-2 (tyrosine). For details, see the legend to Fig. 2.

(4). Each of these peptides had the same molecular weight in the two glycoproteins. Thus, we would expect that gD-1 and gD-2 are similarly organized with respect to the positions of N-linked glycosylation sites.

One interesting feature of the deduced amino acid sequence (22) is that there is an asparagine residue close to the N terminus (residue 40 of the deduced sequence, or residue 15 of the protein). According to the sequence of the adjacent amino acids, this asparagine is not a potential glycosylation site (15). Since we detected no [$2-^3H$]mannose label at this position, it appears that this asparagine is not glycosylated in the protein.

Our overall conclusion from these experiments is that gD-1 and gD-2 appear to be quite similar although not identical in sequence in the N-terminal region of the protein. Thus, the sequencing data have led us to the same conclusions drawn previously about the structures of gD-1 and gD-2 (8–10). In addition, we have defined the N terminus for gD-1 as being 25 residues downstream from the initiating methionine

site. Our sequence data correlate extremely well with the sequence predicted for gD-1 based on nucleic acid sequencing (22). Only one difference (methionine at residue 8) was noted between the predicted sequence and the actual sequence. Since different strains of HSV-1 were used for the two studies, the data emphasize the overall conservation in sequence of gD between different strains of HSV-1. Further experiments are under way in our laboratory to align the tryptic peptides of gD-1 with those predicted to be present from the overall sequence.

ACKNOWLEDGMENTS

This investigation was supported by Health and Human Services grant AI-18289 from the National Institute of Allergy and Infectious Diseases, grant DE-02623 from the National Institute of Dental Research, and grant NS-13201 from the National Institute of Neurological Communicative Disorders and Stroke.

We thank Randy Birenbaum and Nora Guerrero for help in performing the amino acid sequencing studies, Manuel Ponce de Leon for preparing purified gD-1 and gD-2, and Madeline Cohen for excellent technical assistance.

LITERATURE CITED

- Bonner, W. M., and R. A. Laskey. 1974. A film detection method for tritium labeled proteins and nucleic acids in polyacrylamide gels. *Eur. J. Biochem.* **46**:83-88.
- Cohen, G. H., M. Katze, C. Hydrean-Stern, and R. J. Eisenberg. 1978. Type-common CP-1 antigen of herpes simplex virus is associated with a 59,000-molecular weight envelope glycoprotein. *J. Virol.* **27**:172-181.
- Cohen, G. H., D. Long, and R. J. Eisenberg. 1980. Synthesis and processing of glycoproteins gD and gC of herpes simplex virus type 1. *J. Virol.* **36**:429-439.
- Cohen, G. H., D. Long, J. T. Matthews, M. May, and R. Eisenberg. 1983. Glycopeptides of the type-common glycoprotein gD of herpes simplex virus types 1 and 2. *J. Virol.* **46**:679-689.
- Cohen, G. H., M. Ponce de Leon, and C. Nichols. 1972. Isolation of a herpes simplex virus-specific antigenic fraction which stimulates the production of neutralizing antibody. *J. Virol.* **10**:1021-1030.
- Eberle, R., and R. J. Courtney. 1980. gA and gB glycoproteins of herpes simplex virus type 1: two forms of a single polypeptide. *J. Virol.* **36**:665-675.
- Edman, P., and G. Begg. 1967. A protein sequenator. *Eur. J. Biochem.* **1**:80-91.
- Eisenberg, R. J., M. Ponce de Leon, and G. H. Cohen. 1980. Comparative structural analysis of glycoprotein gD of herpes simplex virus types 1 and 2. *J. Virol.* **35**:428-435.
- Eisenberg, R. J., D. Long, L. Pereira, B. Hampar, M. Zweig, and G. H. Cohen. 1982. Effect of monoclonal antibodies on limited proteolysis of native glycoprotein gD of herpes simplex virus type 1. *J. Virol.* **41**:478-488.
- Eisenberg, R. J., M. Ponce de Leon, L. Pereira, D. Long, and G. H. Cohen. 1982. Purification of glycoprotein gD of herpes simplex virus types 1 and 2 by use of monoclonal antibody. *J. Virol.* **41**:1099-1104.
- Frink, R. J., R. Eisenberg, G. 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.
- Greenwood, F. C., W. M. Hunter, and J. S. Glover. 1963. The preparation of ¹³¹I-labeled human growth hormone of high specific radioactivity. *Biochem. J.* **89**:114-123.
- Halliburton, I. W. 1980. Intertypic recombinants of herpes simplex virus. *J. Gen. Virol.* **48**:1-23.
- Hermanson, M. A., L. H. Ericsson, K. Titani, H. Neurath, and K. A. Walsh. 1972. Application of sequenator and analyses to the study of proteins. *Biochemistry* **11**:4493-4502.
- Hubbard, S. C., and R. J. Ivatt. 1981. Synthesis and processing of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* **50**:555-583.
- Kreil, G. 1981. Transfer of proteins across membranes. *Annu. Rev. Biochem.* **50**:317-348.
- Lee, G. T.-Y., M. F. Para, and P. G. Spear. 1982. Location of the structural genes for glycoproteins gD and gE and for other polypeptides in the S component of herpes simplex virus type 1 DNA. *J. Virol.* **43**:41-49.
- Norrild, B. 1980. Immunochemistry of herpes simplex virus glycoproteins. *Curr. Top. Microbiol. Immunol.* **90**:67-106.
- Ruyechan, W. T., L. S. Morse, D. M. Knipe, and B. Roizman. 1979. Molecular genetics of herpes simplex virus. II. Mapping of the major viral glycoproteins and of the genetic loci specifying the social behavior of infected cells. *J. Virol.* **29**:677-697.
- Spear, P. G. 1976. Membrane proteins specified by herpes simplex viruses. I. Identification of four glycoprotein precursors and their products in type 1-infected cells. *J. Virol.* **17**:991-1008.
- Spear, P. G. 1980. Herpesviruses, p. 709-750. In H. A. Blough and J. M. Tiffany (ed.), *Cell membranes and viral envelopes*, vol. 2. Academic Press, Inc., New York.
- Watson, R. J., J. H. Weis, J. S. Salstrom, and L. W. Enquist. 1982. Herpes simplex virus type 1 glycoprotein D gene: nucleotide sequence and expression in *Escherichia coli*. *Science* **218**:381-384.