Identification and Nucleotide Sequence of the Glycoprotein gB Gene of Equine Herpesvirus 4

MARCELLO P. RIGGIO,[1] ANN A. CULLINANE,[2] AND DAVID E. ONIONS[1]*

Department of Veterinary Pathology, University of Glasgow Veterinary School, Bearsden Road, Bearsden, Glasgow G61 1QH, Scotland,[1] and Irish Equine Research Centre, Johnstown, Naas, County Kildare, Ireland[2]

The nucleotide sequence of the glycoprotein gB gene of equine herpesvirus 4 (EHV-4) was determined. The gene was located within a *Bam*HI genomic library by a combination of Southern and dot-blot hybridization with probes derived from the herpes simplex virus type 1 (HSV-1) gB DNA sequence. The predominant portion of the coding sequences was mapped to a 2.95-kilobase *Bam*HI-*Eco*RI subfragment at the left-hand end of *Bam*HI-C. Potential TATA box, CAT box, and mRNA start site sequences and the translational initiation codon were located in the *Bam*HI M fragment of the virus, which is located immediately to the left of *Bam*HI-C. A polyadenylation signal, AATAAA, occurs nine nucleotides past the chain termination codon. Translation of these sequences would give a 110-kilodalton protein possessing a 5' hydrophobic signal sequence, a hydrophilic surface domain containing 11 potential N-linked glycosylation sites, a hydrophobic transmembrane domain, and a 3' highly charged cytoplasmic domain. A potential internal proteolytic cleavage site, Arg-Arg/Ser, was identified at residues 459 to 461. Analysis of this protein revealed amino acid sequence homologies of 47% with HSV-1 gB, 54% with pseudorabies virus gpII, 51% with varicella-zoster virus gpII, 29% with human cytomegalovirus gB, and 30% with Epstein-Barr virus gB. Alignment of EHV-4 gB with HSV-1 (KOS) gB further revealed that four potential N-linked glycosylation sites and all 10 cysteine residues on the external surface of the molecules are perfectly conserved, suggesting that the proteins possess similar secondary and tertiary structures. Thus, we showed that EHV-4 gB is highly conserved with the gB and gpII glycoproteins of other herpesviruses, suggesting that this glycoprotein has a similar overall function in each virus.

Equine herpesvirus 1 (EHV-1) and EHV-4 are alphaherpesviruses which cause serious disease in the horse, EHV-1 being predominantly associated with abortion and neurological disease and EHV-4 being a major cause of respiratory disease (28). The envelope glycoproteins are important immunogens of herpesviruses involved in producing a protective immune response (40, 41). EHV-1 and EHV-4 are known to contain at least eight highly abundant envelope glycoproteins (1). However, little is known about the structure of these glycoproteins, although the six major glycoproteins of EHV-1 have recently been mapped (2). The major glycoproteins gB and gC of herpes simplex virus type 1 (HSV-1) have been sequenced and characterized (8, 16, 31), and they have been shown to possess determinants generating cytolytic and virus-neutralizing antibody (17, 41). Glycoprotein gB is essential for the production of infectious virus, since temperature-sensitive mutants have been mapped and isolated in gB (8, 25, 27), and it is thought to have a role in cell fusion and viral penetration of infected cells (25, 38). Syncytial (27) and fast-entry (13) phenotypes have been mapped within HSV-1 gB and have been shown to be the result of single amino acid substitutions—a Val-to-Ala substitution at residue 552 and an Arg-to-His substitution at residue 857 account for the fast-entry and syncytial phenotypes, respectively (9).

Wild-type gB exists as a dimer (38) and appears on the surface of virions and virus-infected cells.

In localizing EHV-4 gB, we took advantage of the known colinearity of the EHV-4 genome with the I_{SL}-I_L arrangement of the HSV-1 genome (12). The available map position for HSV-1 gB enabled us to locate the region of the genome containing the EHV-4 gB gene, and the 5' and 3' boundaries

were determined with DNA probes derived from the HSV-1 gB gene. In this report, we present the nucleotide sequence of the EHV-4 gB gene and report that this protein shares amino acid homology with the gB glycoproteins of HSV-1 (8, 31), Epstein-Barr virus (29), and human cytomegalovirus (11) and the gpII glycoproteins of pseudorabies virus (PRV) (34) and varicella-zoster virus (VZV) (21).

MATERIALS AND METHODS

Recombinant DNA methods. The *Eco*RI F restriction fragment of HSV-1 (19) cloned in plasmid pACYC184 (26) was kindly provided by J. B. Clements (MRC Institute of Virology, Glasgow), and this recombinant plasmid was designated pACYC-EcoRI(F). The entire HSV-1 gB gene was excised from plasmid pACYC-EcoRI(F) as a 3.3-kilobase (kb) *Xho*I-*Kpn*I fragment, which was subsequently directionally cloned into the *Xho*I and *Kpn*I sites of pIC20R to generate the plasmid pICgB.

Digestion of pICgB with *Nar*I yielded a 650-base-pair fragment (corresponding to amino acids 121 to 336 of HSV-1 gB), and digestion of pICgB with *Pst*I yielded an 810-base-pair fragment (corresponding to amino acids 530 to 799 of HSV-1 gB), which were used as 5' and 3' HSV-1 gB DNA hybridization probes, respectively, in dot-blot and Southern blot analyses (see Fig. 2).

The genomic library of EHV-4 strain 1942 (12) contained the 13.6-kb *Bam*HI C fragment in the *Bam*HI site of pUC9. The 2.95-kb *Bam*HI-*Eco*RI fragment at the left end of *Bam*HI-C was excised from pUC9 as a 2.95-kb *Eco*RI fragment and cloned into the *Eco*RI site of Bluescript M13+ vector. This recombinant plasmid was designated pBSgB.

Bacterial strains. *Escherichia coli* JM83 and JM101 were used for recombinant DNA experiments and were grown in L broth medium (L broth contains 10 g of tryptone, 5 g of
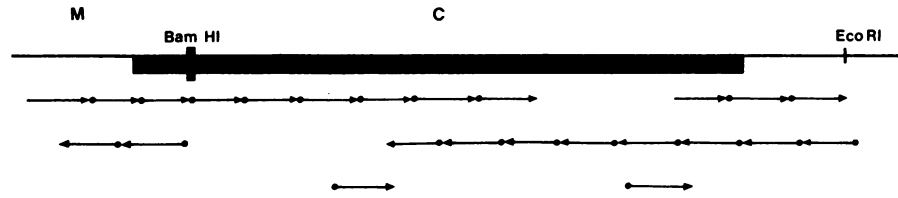
---

* Corresponding author.

FIG. 1. Diagram showing strategy for sequence determination of the EHV-4 gB gene. Sequence data were generated with synthetic deoxyoligonucleotide primers. Dots represent the 5' end and arrowheads show the 3' end of each portion of the sequence. The EHV-4 gB ORF is denoted by a thick bar.

yeast extract, and 10 g of NaCl per liter). Bluescript M13+ recombinant DNA was propagated in JM101 cells, and all other recombinant plasmids were propagated in JM83 cells. Cells containing recombinant plasmids were selected on L broth agar plates (1.5% agar in L broth) containing ampicillin at 100 μg/ml or tetracycline at 12.5 μg/ml.

Plasmid DNA isolation. Plasmid DNA was isolated from bacteria by the boiling method (5) for small-scale preparations. For large-scale isolations, plasmid DNA was extracted from bacteria by the alkaline lysis method (26) and further purified by banding on cesium chloride gradients containing 50% (wt/vol) CsCl and 200 μg of ethidium bromide per ml. Ethidium bromide was removed by multiple extractions with isopropanol, and the DNA was dialyzed against 0.1× TE (1 mM Tris hydrochloride, 0.1 mM EDTA [pH 8.0]), followed by ethanol precipitation.

Dot-blot analysis. The C, F, and M viral restriction fragments cloned in the BamHI site of pUC9 were excised with BamHI, purified through low-melting-point agarose gels, and spotted onto Gene-Screen hybridization membranes in 100-, 200-, and 400-ng amounts. In a later experiment, the four subfragments of BamHI-C obtained by digestion with SmaI were spotted onto membranes as above. Samples (200 ng) of vector DNAs pUC8 and pBR322 were used as negative hybridization controls and 200 ng of the appropriate probe DNA was used as a positive hybridization control in all experiments.

Membranes were prehybridized in 10 ml of hybridization buffer (40% formamide, 5× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate], 5× Denhardt solution [0.1% (wt/vol) Ficoll, 0.1% (wt/vol) polyvinylpyrrolidone, 0.1% (wt/vol) bovine serum albumin], 10% dextran sulfate, 50 mM sodium PP$_i$ [pH 6.5], 100 μg of denatured salmon sperm DNA per ml) at 42°C for 16 to 20 h.

HSV-1 gB 5' and 3' probe DNA was prepared by nick translation (33) to a specific activity of 10^8 cpm/μg of DNA and was added at a concentration of 10^6 cpm/ml of hybridization buffer (10 ng/ml) to the prehybridized membranes. Hybridization was carried out at 42°C for 16 to 20 h. Membranes were prewashed in 2× SSC for 15 min at room temperature and then washed in 2× SSC–0.1% sodium dodecyl sulfate (SDS), 1× SSC–0.1% SDS, or 0.1× SSC–0.1% SDS for three 20-min washes at 65°C. Finally, membranes were rinsed in 0.1× SSC, air dried, and exposed to Fuji-RX X-ray film at −70°C for 24 h.

Southern blot analysis. The BamHI-C viral restriction fragment contained in pUC9 was digested with BamHI and then double digested with BamHI and each of the restriction enzymes BglI, EcoRI, PstI, PvuII, and SmaI to release viral DNA sequences from vector sequences. Digested DNA was run on 0.8% agarose gels in 1× TEA buffer (50× TEA is 2 M Tris base, 0.1 M EDTA, 1 M sodium chloride, 1 M sodium acetate, pH 8.1), denatured, neutralized, and transferred to Gene-Screen hybridization membranes (39). Membranes

were hybridized to HSV-1 gB 5' and 3' DNA probes as outlined above.

DNA sequencing. The 2.95-kb BamHI-EcoRI EHV-4 DNA fragment contained in plasmid pBSgB was sequenced by the dideoxy chain termination method (37) with [α-^{35}S]dATP as a label (4). Extensive use was made of synthetic deoxyoligonucleotide primers to rapidly generate sequencing data, in addition to Bluescript M13+ specific sequencing primers, and the sequencing strategy is shown in Fig. 1.

Since the BamHI-M viral restriction fragment contains the start of the gB gene, the right-hand end of this fragment was sequenced in pUC9 with the M13 reverse primer and synthetic deoxyoligonucleotide primers.

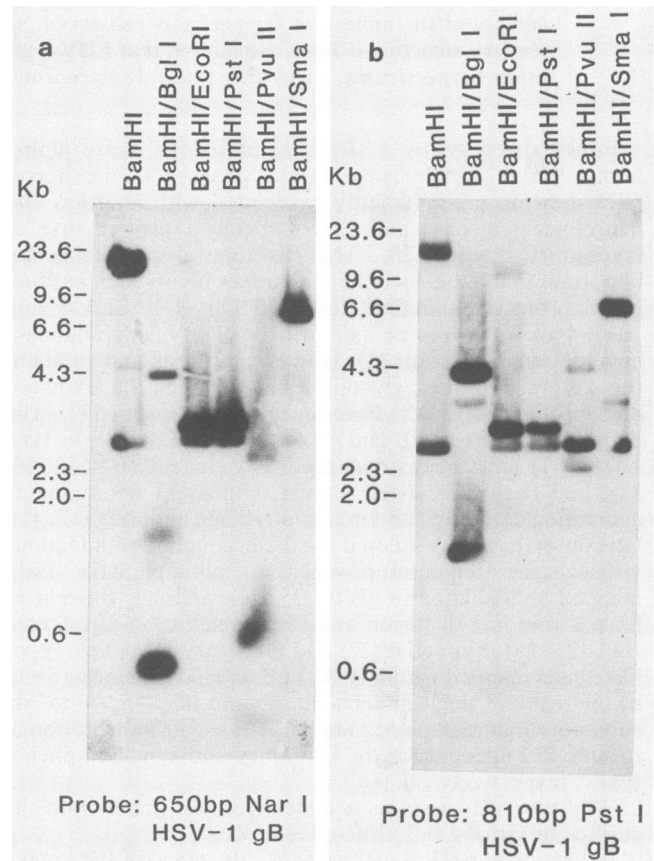FIG. 2. Southern blot analysis of EHV-4 DNA. (a) Southern blot analysis of BamHI-C DNA with the 650-base-pair (bp) NarI HSV-1 gB (5') DNA as a hybridization probe. (b) As in panel a, but with the 810-base-pair PstI HSV-1 gB (3') DNA as a hybridization probe. All procedures were done as described in the text. Hybridization was only observed when membranes were washed at low stringency (2× SSC–0.1% SDS at 65°C).

Single-stranded template for sequencing was produced by alkaline denaturation of 1 pmol of plasmid DNA in 0.2 M NaOH–0.2 mM EDTA in a volume of 20 µl for 5 min at room temperature, followed by neutralization with 2 µl of 2 M ammonium acetate (pH 5.3) and ethanol precipitation.

The resulting single-stranded DNA template was suspended in 8 µl of 0.1× TE (pH 8.0) and mixed with 2 pmol of sequencing primer, 3 µl (30 µCi) of [α-^{35}S]dATP, and 1.5 µl of 10× Klenow annealing buffer (100 mM Tris hydrochloride [pH 8.0], 50 mM MgCl₂) in a total volume of 15 µl. The mixture was annealed for 15 min at 37°C and then used in standard sequencing reactions at 37°C. Samples were run on 6% polyacrylamide wedge-shaped gels containing 7 M urea at 60°C. Gels were bonded to glass plates by treatment of the plates with 2% dimethylchlorosilane in 1,1,1-trichloro-ethane, fixed in 10% acetic acid–10% methanol, and dried in an 80°C oven before being exposed to X-ray film at −70°C for 18 to 24 h.

RESULTS

Identification of DNA fragment containing the EHV-4 gB gene. Since the EHV-4 genome is colinear with the I$_L$ arrangement of the HSV-1 genome (12) and the exact location and sequence of the HSV-1 gB gene was known (8, 31), we were able to predict the region of the EHV-4 genome which would probably contain the gB gene.

Hybridization of DNA fragments from this region to both 5' and 3' HSV-1 gB DNA probes in dot-blot analyses localized the gB gene to a 6.5-kb subfragment of BamHI-C (data not shown).

Southern blot analysis of BamHI-C more precisely located the gene to the left terminal 2.95-kb BamHI-EcoRI subfragment of BamHI-C (Fig. 2a and b). The 5' probe hybridized to 2.95-kb BamHI-EcoRI, 0.56-kb BamHI-PvuII, 6.5-kb BamHI-SmaI, 3-kb BamHI-PstI, and 0.42-kb BamHI-BglI fragments; weak hybridization was seen to 4.2-kb BglI and 0.14-kb PvuII fragments. The 3' probe hybridized to 2.95-kb BamHI-EcoRI, 2.7-kb PvuII, 6.5-kb BamHI-SmaI, 3-kb BamHI-PstI, and 4.2-kb BglI fragments. Some hybridization of probes to vector DNA (pUC9) was observed which was probably due to cross hybridization of G+C-rich regions of pUC9 to G+C-rich regions of HSV-1 DNA.

These observations strongly suggested that the gene is located predominantly within a 2.95-kb BamHI-EcoRI fragment at the left terminus of BamHI-C with a left-to-right transcription orientation (Fig. 3d and e).

Since the 5' HSV-1 gB probe does not contain the first 357 nucleotides of the coding region of the HSV-1 gB gene, it was predicted that a corresponding portion of the EHV-4 gB gene and its transcriptional control domains reside within the BamHI M fragment, which is immediately to the left of BamHI-C.
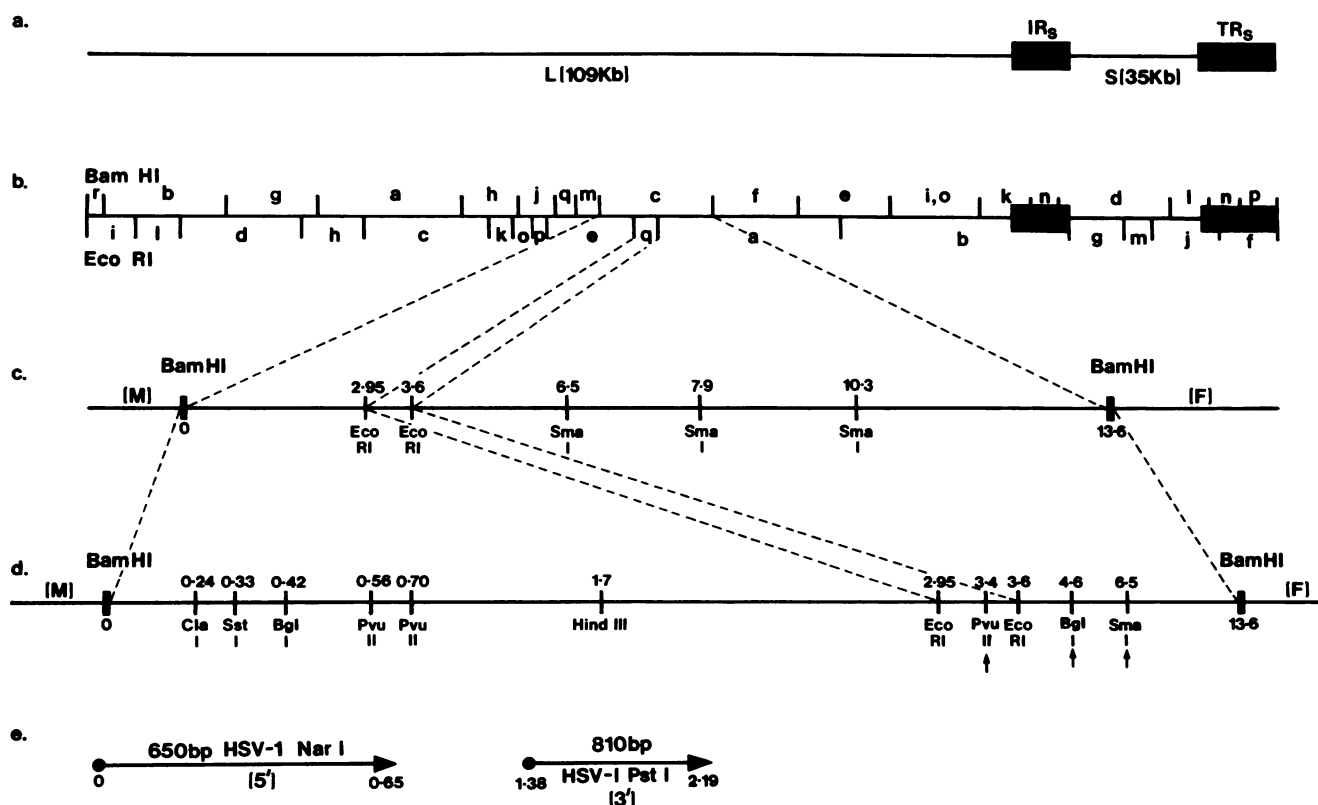
FIG. 3. Map of the EHV-4 genome showing restriction enzyme sites and DNA fragments used in studying the glycoprotein gB gene. (a) Structure of EHV-4 genome. This consists of a unique long region (L) and a unique short region (S) bounded by inverted repeats (IR$_s$, TR$_s$). (b) Arrangement of BamHI and EcoRI restriction enzyme sites along the genome. (c) Location of EcoRI and SmaI restriction enzyme sites in the 13.6-kb BamHI C fragment of the genome. (d) Detailed restriction enzyme mapping of the 2.95-kb BamHI-EcoRI subfragment of BamHI-C, which contains the predominant portion of the gB gene. The first PvuII, BglI, and SmaI restriction enzyme sites to the right of the 2.95-kb BamHI-EcoRI subfragment are indicated by arrows. (e) Arrows show the regions of the 2.95-kb BamHI-EcoRI subfragment of BamHI-C to which the 5' and 3' HSV-1 gB DNA probes hybridized. bp, Base pairs.

EHV-4 gB
HSV-1 gB

```
1       GA--ACT-------AGCT--------------------CGGTTTAT------------------GGTTACTGCGG---CTAAA---------------
        II III         IIII                  II I III               II I IIII I   II II
1       GTCAACGGGCCCCTCTTTGATCACTCCACCCACAGCTTCGCCCAGCCCCCCAACACCGCGCTGTATTACAGCGTCGAGAACGTGGGGCTCCTGCCGCACCTGAAGGAGGAGCTCGCCCGG

                                                    CAT                                                                           TATA
34      --------------------GGTG---ATTGGTCAATTAGCGAGTTTCAAAGGTTTTATTGCTTTGAGGGTGTGACAGGTGTGACGGCCACGCAACGGCTGGCGTGGAAATATATC
                            IIII   IIIIII I I IIIII IIIII IIIIIIIII I I III I I III IIII II II I III IIIIII
121     TTCATCATGGGGGCGGGGGGCTCGGGTGCTCATTGGGCCGTCAGCGAATTTCAGAGGTTTTACTGTTTTGACGGCATTTCCGGAATAACGCCCACTCAGCGCGCCGCCTGGCGATATATT

        mRNA init
127     GGGGAGCTCATTCTAGCTACCGCAGTATTCTCTTCGGTTTTCCACTGCGGAGAGGTGCGCCTTCTGCGCGCAGATCGTACATATCCAAACACCAACGGCGCACAGCGCTGCGCTAGCGGC
        I IIIII IIII III I II II II I II I II I IIIII I II IIIII III I   I I IIIII III I      II IIII II II   I III I I   IIII
241     CGCGAGCTGATTATCGCCACCACACTCTTTGCCTCGGTCTACCGGTGCGGGGAGCTCGAGTTGCGCCGCCCGGACTGCAGCCGCCCGACCTCCGAAGGTCGTTACCGTTACCCGCCCGGC

                  start
247     ATTTACATAACATACGAGACGTCATGTCCACTTGTTGCCGTGCTATTTGTGGCCCCCAACGGTGTTATTGGCGAAGAGACTGTGGTAATTTACGACAGCGACGTGTTCTCGCTTCTATAC
        I II I II IIIII I   IIIII II II III I I     IIIIII IIII  III III   I I II II I IIIIII  IIIIII IIIIII I II III
361     GTATATCTCACGTACGACTCCGACTGTCCGCTGGTGGCCATCGTCGAGAGCGCCCCCGACGGCGTGTATCGGCCCCCGGTCGGTCGTGGTCTACGACGCCGACGTTTTCTCGATCCTCTAC

367     ACCGTACTCCAGCAGCT-GGCTCCTGGCT--CT----GGAGCCA--ATTAGGAAATGTAAACTTGCCA-GCTAC----CTCCCCCATGT-------CTAAAGACTCGAC--AT-----CT
        I III IIIIIIIII II I I II IIII II   II II I  I       III I IIII II I  I IIIII II       I I III I II         II
481     TCGGTCCTCCAGCACCTCGCCCCCAGGCTACCTGACGGGGGGGCACGACGGGCCCCCGTAGTCCCGCCATGCACCAGGGCGCCCCCTCGTGGGGGCGCCGGTGGTTCGTCGTATGGGCGCT
                                                                                         start

459     C-TGGGGGTGA------GAACAATAGT--CAT-TGCGTGTTTG-GTTCTCTTGGGA----TGTTGTATTGTG------------GAA-GCTGTACCAACCACGCCAAGTTCTCAGCCCA
        I IIIII III     I I II I I IIII I I I I III   II II          II II         I IIIIII III II I IIIII III III II IIIII
601     CTTGGGGTTGACGCTGGGGGTCCTGGTGGCGTCGGCGGCTCCGAGTTCCCCCGGCACGCCTGGGGTCGCGCGCGACCCAGGCGGCGAACGGGGGCCCTGCCACTCCGGCGCCGCCGCCCT

550     GTACTCC-------CGCGTCAACCCAGTCCG--------C-TAAAACCG-TTGACCAAACGCTTCTA----CCAACTGAAAC-ACCAGACC--CG-CTC-AGACTGG---CTGTACGCGA
        I II       II I I II I III          I IIIIIII   III IIIII I I    II I I II II IIII IIII I I I     I IIIII
721     TGGCGCCGCCCCAACGGGGGACCCGAAACCGAAGAAGAACAAAAAACCGAAAAACCCAACGCCACCACGCCCCGCCGGCGACAACGCGACCGTCGCCGCGGGCCACGCCACCCTGCGCGA

641     G----TCCGG--TAT--ACTCGCAGA----GGATGGAGACTTTTACACCTGCCCGCCGCCTACTGGATCCACAGTTGTACGCATTGAACCCCCACGGTCATGTCCCAAGTTTGATCTGGG
        I  I III  II I II IIII      IIII I IIIIIII   IIIII III II II II   IIII II II  I III I III I        I III
841     GCACCTGCGGGACATCAAGGCGGAGAACACCGATGCAAACTTTTACGTGTGCCCACCCCCCACGGGCGCCACGGTGGTGCAGTTCGAGCAGCCGCGCCGCTGCCCGACCCGGCCCGAGGG

749     GAGGAACTTCACGGAGGGCATTGCTGTTATTTTCAAGGAAAACATAGCCCCGTACAAATTTAGAGCAAACGTCTACTACAAAGACATTGTAGTGACAAAGGTTTGGAAAGGATACAGCCA
        IIIII IIIIIIIIIII II II I IIIIIII IIIII IIIIIIIIIIII II I II I I IIIIIIIIIIII   II I IIII III II II II I
961     TCAGAACTACACGGAGGGCATCGCGGTGGTCTTCAAGGAGAACATCGCCCCGTACAAGTTCAAGGCCACCATGTACTACAAAGACGTCACCGTTTCGCAGGTGTGGTTCGGCCACCGCTA

869     CACCTCTTTATCCGATAGATACAATGACAGAGTGCCAGTTTCAGTGGAGGAGATATTCACTCTCATCGATAGCAAAGGAAAATGTTCTTCTAAGGCAGAGTACCTCCGAGATAACATTAT
        I II  II   I I I  I IIII I I II I I  I IIIIIII I II  III I   III II   III  II I III IIIII I II  I III I
1081    CTCCCAGTTTATGGGGATCTTTGAGGACCGCGCCCCCGTCCCCTTCGAGGAGGTGATCGACAAGATCAACGCCAAGGGGGTCTGTCGGTCCACGGCCAAGTACGTGCGCAACAACCTGGA

989     GCATCACGCTTACCACGACGACGAAGACGAGGTGGAGCTCGACCTGGTTCCGTCTAAGTTTGCTACTC-CTGGGGCCAGAGCATGGCAAACCACTAACGAC-ACCACGTCTTATGTCGGA
        I    III I III IIIII IIIII I II III    IIII II IIIII II I I IIII I II IIIIIII II II I III II I I I I IIII
1201    GACCACCGCGTTTCACCGGGACGACCACGAGACCGACATGGAGCTGAAACCGGCCAACGCCGCGACCCGCACGAGCCGGGGC-TGGCACACCACCGACCTCAAGTACAACCCCTCGCGGG

1107    TGGATGCCATGGAGGCACTACACATCAACCTCTGTCAACTGCATTGTCGAAGAGGTAGAAGCGCGGTCTGTTTACCCATACGACTCCTTTGCCCTATCGACCGGTGATATTGTGTACACC
        IIII II  I I I I III   II I III IIIIIII IIIII IIIII II IIIII II II IIIII IIIIII IIII II IIII II II IIIIIIIII
1320    TGGAGGCGTTCCA-CCGGTAC---GGGACGACGGTAAACTGCATCGTCGAGGAGGTGGACGCGCGCTCGGTGTACCCGTACGACGAGTTTGTGCTGGCGACTGGCGACTTTGTGTACATG

1227    TCACCGTTTTACGGCCTTCGGTCAGCTGCTCAGTTAGAACACAATAGCTACGCACAGGAGCGCTTTAGACAAGTTGAAGGATACCAACCAAGAGACTTGGACAGTAAATTACAGGCCGGA
        II IIIIIIIIIIII III I I II IIIIIIIII I IIIII II IIIIII I II II II II I I I IIII I II II I IIII
1436    TCCCCGTTTTACGGCTACCGGGAGGGGTCGCACACCGAACACACCACGTACGCCGCCGACCGCTTCAAGCAGGTCGACGGCTTCTACGCGCGCGACCTCACCACCAAGGCCCGGGCCACG

1347    GAGCCAGTTACCAAAAACTTTATTACTACACCTCATGTTACAGTCAGCTGGAACTGGACTGAAAAAAAGATAG-AGGCGTGTACACTAACTAAATGGAAGGAGGTTGACGAACTTGTCAG
        I III   III   IIII I II I IIII II I II I I IIII I II II IIIII II I I IIIII    III   IIIII I II III I II III I I
1556    GCGCCGACCACCCGGAACCTGCTCACGACCCCCAAGTTCACCGTGGCCTGGGACTGG-GTGCCAAAGCGCCCGTCGGTCTGCACCATGACCAAGTGGCAGGAAGTGGACGAGATGCTGCG

1466    AGATGAGTTTCGGGGGTCCTACAGGTTTACTATTCGA-TCCATTTCGTCCACGTTTATTAGCAACACTACTCAATTTAAGCTAGAAGATGCCCCACTCACCGACTGTGTGTCAAAAGAAG
        IIII  I II IIIII I I II  I I III  IIII II IIIII II III I I II    II I I     III   I I I II IIII I I I   I I
1675    CTCCGAGTACGGCGGCTCCTTCCGATTCTC-CTCCGACGCCATATCCACCACCTTCACCACCAACCTGACCGAGTACCCGCT-----CTCGCGCGTGGAC---CTGGGGGACTGCA-TCG

1585    CCAAAGATGCCATAGACTCTATATACCGAAAACAGTATGAGTCTACACACGTTTTTAGTGGGGATGTGGAATTTTACTTGGCACGTGGAGGGTTCTTAATCGCATTTAGACCGATGATTT
        III II III   III I II IIII  II    IIII I I I       I III II  II IIII I III     I IIII I
1785    GCAAGGACGCCCGCGACGCCATGGACCG----CA-------TCTTCGCCC-----------GCAGGTACAACGCGAC---GCACATCAAGGTGGGCCAGCCGCAGT-------------

1705    CTAACGAACTTGCCAGGCTGTACCTAAACGAGCTTGTGAGATCTAACCGCACCTATGACCTAAAAAATCTGTTAAACCCCAACGCAAACCATAATACCAATCGAACACGCAGGTCGCTAC
        II I II IIII II I I   II IIII I I IIIII II I III II       III       IIIIII I II I
1866    ---ACTACCTGGCCA--ATG--------GGGGCTT-TCTGATC---GCGTACC-AGCCCCT------TCT---------CA--GC----------------AACACGCTCG-CGGAGC
```
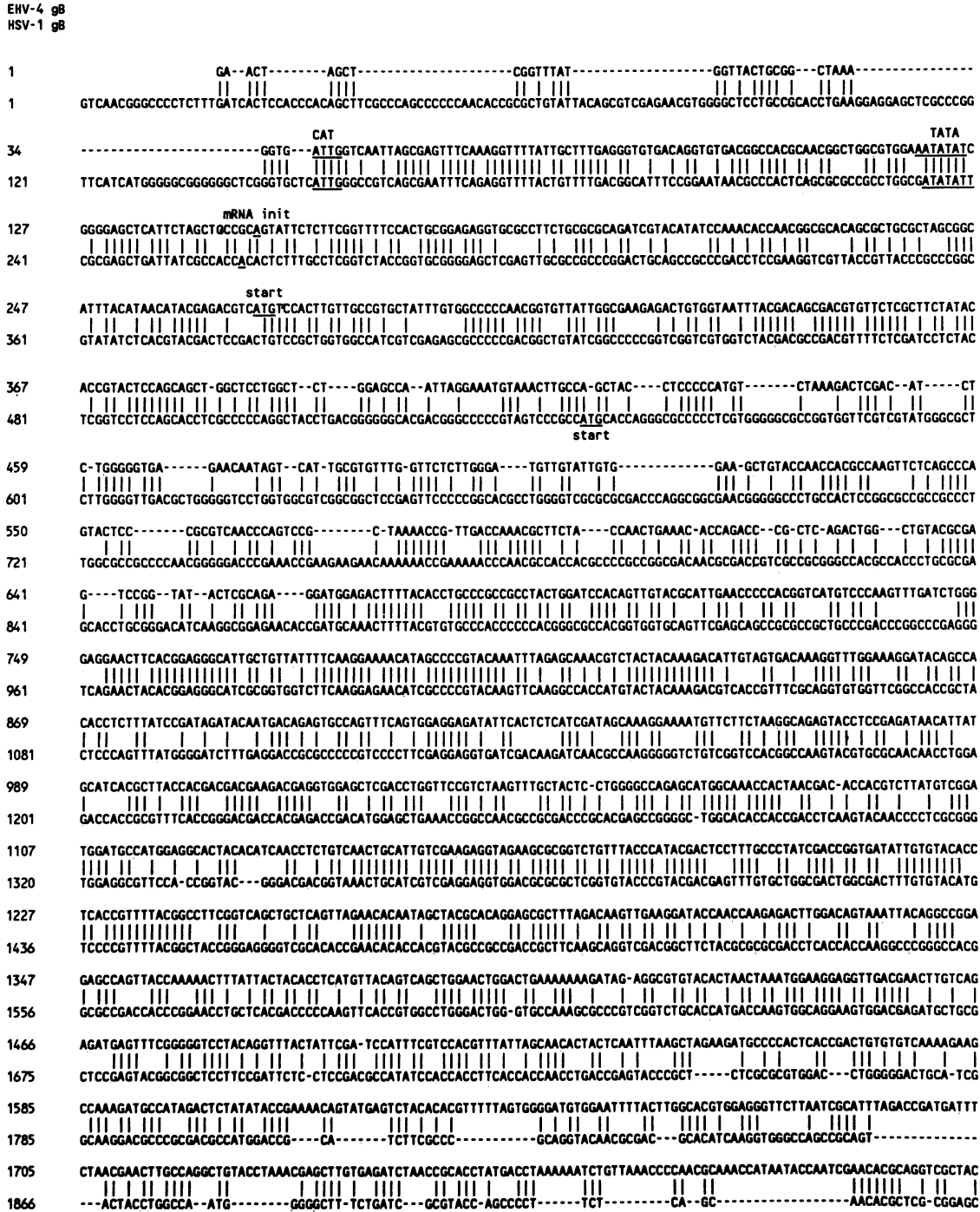
FIG. 4. Alignment of EHV-4 gB and HSV-1 gB (KOS) DNA sequences. The top line is the DNA sequence of EHV-4 gB, and the bottom line is the HSV-1 gB DNA sequence. Identical residues are indicated with a vertical bar, and dashes mark spaces introduced to maximize homology between the two sequences. The salient features of the sequences (CAT box, TATA box, mRNA cap site, termination codon [ter], and polyadenylation signal [polyA]) are indicated for each sequence.

Consequently, the 2.95-kb *Bam*HI-*Eco*RI subfragment of *Bam*HI-C and the right-hand end of *Bam*HI-M were sequenced and analyzed for an open reading frame (ORF) which contains the gB gene.

Analysis of DNA and amino acid sequences. (i) Nucleotide sequence of the EHV-4 gB gene showing homology to HSV-1 gB. The DNA insert of pBSgB and the right-hand end of

*Bam*HI-M were sequenced as described above. An ORF of 2,925 nucleotides was found in the sequence with an ATG initiation codon at position 270 and a chain termination codon at position 3195. The predicted translation product of this ORF would be a protein of 975 amino acids, which is larger than the 903 amino acids predicted for HSV-1 gB and the 913 amino acids predicted for PRV gpII.

```
1825  TATCAATACCAGAACCTACTCCAACCCAAGAGAGCCTCCACAGAGAACAAATACTACATCGCCTACACAAACGAGCAGTGGAGGCTGCGAATAGTACAAACTCTTCCAACGTCACCGCCA
      | |  | | ||| | |||| | |    |  |   | |   | | |   || | ||   ||  | | || ||  | | ||   | ||| || |  |  || | | ||| | |      |||||| |||
1932  TGTACGTGCGGGAA-CACCTCCGA--GAGCAGAG---CCGCA-AGCCCCCAAACCCCA-CGCCCCCGC---CGCCC---GGGGCCAGCG--------------CCAACGCGTCCG---

1945  AACAACTAGAGCTAATCAAAACAACGTCCTCTATTGAGTTTGCTATGCTACAGTTTGCATACGATCACATCCAATCCCACGTTAATGAGATGCTAAGTAGGATAGCAACTGCGTGGTGTA
        | ||||  ||||| || | ||| ||   ||||| ||  ||    | || | ||  ||  ||| |||||  |  || ||||| | |    | | | |  | |  | |   |    ||||||||
2019  ------TGGAGCGCATCAAGACCACCTCCTCCATCGAGTTCGCCCGGCTGCAGTTTACGTACAACCACATACAGCGCCATGTCAACGATATGTTGGGCCGCGTTGCCATCGCGTGGTGCG

2065  CACTACAAAACAAAGAGCGGACCCTCTGGAATGAGATGGTAAAGGTTAACCCAAGCGCTATTGTTTCCGCCACTCTTGACGAGCGAGTTGCGGCAAGGGTTTTGGGAGACGTTATAGCCA
      |||||| ||   | |||| ||||||| ||||| ||| ||||| |||   ||| | ||  |  | ||||| | || | || ||| || | || |  |   || || | | || ||||| || |||
2133  AGCTACAGAATCACGAGCTGACCCTGTGGAACGAGGCCCGCAAGCTGAACCCCAACGCCATCGCCTCGGTCACCGTGGGCCGGCGGGTGAGCGCGCGGATGCTCGGCGACGTGATGGCCG

2185  TAACACATTGTGTAAAAAT---AGAGGGCAATGTGTACTTACAAAACTCTATGCGC-TC--CTCGGACAGCAACACGTGCTACTCCCGCCCACCTGTAACGTTTACCATTACTAAAAATG
      | |  ||| ||    |  |       | ||| |||||| || |||||| | |   | ||    | ||| |  |   | | |  | |  ||||| |  ||||||   ||||| |  || | |  |||
2253  TCTCCACGTGCGTGCCGGTGGCCGCGGACAACGTGATCGTCCAAAACTCGATGCGCATCAGCTCGCGGCCCGGGGCCTGCTACAGCCGCCCCCTGGTCAGCTTT--CGGTACGAA----G

2299  CAAACAGCAGAGGGACGATAGAGGGCCAGTTGGGAGAAGAAAACGAGGTTTATACGGAGCGCAAGCTTATCGAGCCGTGCGCTATCAATCAAAAACGATACTTTAAGTTTGGCAAAGAGT
      | ||| |   |  | | |  |  || ||||| |  ||| || | | |||| |  |  | | |      | |||| |  ||||||||||| |  |   ||  ||||||| | || | | |
2367  --ACCAG-GGCCCGTTGGTCGAGGGGCAGCTGGGGGAGAACAACGAGCTGCGGCTGACGCGCGATGCGATCGAGCCGTGCACCGTGGGACACCGGCGCTACTTCACCTTCGGTGGGGGCT

2419  ATGTTTACTATGAGAACTACACGTACGTTCGCAAAGTGCCCCCGACTGAAATCGAAGTGATCAGCACCTACGTTGAACTAAACTTAACTCTTTTGGAAGACCGCGAGTTTCTACCCCTGG
      | || |||  ||| || ||| || | ||| | ||  | || | | ||| || | || |||          |||||||||| | | ||  || ||| | | |  |||| || | ||||||||| | |||||||
2484  ACGTGTACTTCGAGGAGTACGCGTACTCCCACCAGCTGAGCCGCGCCGACATCACCACCGTCAGCACCTTCATCGACCTCAACATCACCATGCTGGAGGATCACGAGTTTGTCCCCCTGG

2539  AGGTTTACACGCGAGCTGAGCTTGAAGACACGGGGCTATTGGATTACAGCGAGATACAGCGCCGTAACCAGCTTCACGCCCTCCGATTCTACGATATAGACAGCGTTGTCAACGTGGACA
      |||| ||||| |||  | | |  || | || ||| |||| |||| | |   |  |  || | |   |||| ||| || | ||  || | | | |  || ||| || | || || |||||
2604  AGGTGTACACCCGCACGAGATCAAGGACAGCGGCCTGCTGGACTACACGGAGGTCCAGCGCGCAACCAGCTGCACGACCTGCGCTTCGCCGACATCGACACGGTCATCCACGCCGACG

2659  ACACTGCTGTCATTATGCAGGGAATTGCCACCTTTTTTAAAGGCCTTGGTAAGGTGGGAGAGGCAGTTGGGACGCTTGTACTTGGAGCGGCTGGCGCGGTTGTTTCTACAGTATCGGGTA
       ||  ||| ||||  | |||   |  ||| || || |  |||| || | ||  ||    ||  | | ||||||  |  ||| |||| |    ||||| || || || || | | || |||||
2724  CCAACGCCGCCATGTTCGCGGGCCTGGGCGCGTTCTTCGAGGGGATGGGCGACCTGGGCGCGCGCGGTCGGCAAGGTGGTGATGGGACTCGTGGGCGGCGTGGTATCGGCCGTGTCGGGCG

2779  TAGCCTCATTTATAAACAACCCATTTGGGGGGCTCGCAATAGGCCTGTTGGTAATTGCGGGCTTAGTGGCTGCGTTTTTTGCCTACCGGTATGTAATGCAACTGCGCAGCAACCCCATGA
      | ||||| || |  ||||||| || ||||||||| | |||| | |  ||||||| || | || ||||||| |||     | || || || ||| ||| ||||||| |||| ||||||||||||||||||
2844  TGTCCTCCTTCATGTCCAACCCCTTTGGGGCGCTGGCCGTGGGTCGTGGTCCTGGCCGGGCCTGGCGGCGGCGCCCTTCTTCGCCTTTCGTTACGTCATGCGGCTGCAGAGCAACCCCATGA

2899  AAGCTCTATACCCAATAACAACCAGGAGCCTTAAAAACAAAGCCAAAGCCTCATACGGCCAAAACGACGATGATGACACTAGCGACTTCGATGAAGCCAAGCTGGAGGAGGCACGCGAAA
      | || || |||||  |||| ||| | |  | || ||||  | || || | || ||||||||| | ||||||| | || | |  | | |  ||||  | | ||||| | ||||| ||||||
2964  AGGCCCTGTACCCTCTAACCACCAAGGAGCTCAAGAACCCCACC-AACCCGGACGCGTCC--GGGGAGGGCGAGGAGGGCGGCGACTTTGACGAGGCCAAGCTAGCCGAGGCCAGGGAGA

3019  TGATCAAATATATGTCTATGGTTTCTGCCCTGGAAAAACAGGAAAAAAAGGCAATGAAGAAAACAAAGGGGGTTGGACTTATTGCCAGCAACGTTTCAAAACTCGCACTGCGCAGGCGCG
      ||||    || ||| |  |||| || || | |   | |   ||    ||| |  | | |  ||  | ||  |  || | || |||  | |  || | |||||  ||||||
3081  TGATACGGTACATGGCCCTGGTGTCGGCCATGGAGCGCACGGAACACAAGGCCAAGAAGAAGGGGCA--CGAGCCGG-CTGCTCAGCGCCAAGGTCACCGACATGGTCATGCGCAAGCGCC

                                                              ter      poly A
3139  G--TC-CGAAATATACCC--GTCTTCGA-GAAGACGATCCCATGGAAAGCGAAAAAATG-GTTTAAAAATGTTAAATAAAT-ATTTTGACACGTACTTGTGGGTTGACTCATATTTGCAT
      |  | ||| || ||||  || |  | |||||| | |  || || | | || |     |  | |||||| || | | |  | |  | || ||| |
3198  GCAACACCAACTACACCCAAGTTCCCAACAAAGACGGTGACGCCGACGAGGACGACCTGTGACGGGGGGTTTGTTGTAAATAAAACCACGGGTGTTAAACCGCATGCGCATCTTTTGGT
                                                            ter      poly A

3251  AACATCTTTCTAGTTCCGGCTATAAGCCTATTTAAGCCTAGTATTTTTGCCAAAAGTTTATCATCCTCTACAAGCGCACATCCTCTCAAAAGAGTTGAATTTTGCTGTTTATTACGCTAT
        | || ||  | || | |  ||| ||| |||||  ||   ||    ||| |  | ||| || |  | ||| | |  ||| | |
3318  --TTTTTTGTTTGGTCAGCCT-TTTGTGTGTGTGTGGGAAGAAAGAAAAAGGAACACATAAACTCCCCCGGGTGTCCGCGGCCTGT--TTCCTCTTTCCTTTCCCGTGACAAAACG-GAC

3371  CCTAAAGCTAAACGCCTGTAATGGAATCTCAATGCAAAACTTCTACATCAGCCGCTGATG-AAACT-CTGTTGGCTGCATCGCTACCGCGGCGGAAATCCAAATAAAAACAGAAGCACCC
      ||   ||| | ||| | |  |    |||| |  | | |   | || || ||| | |  |||| |     | | | || ||||| || | | ||||| | || |
3432  CCCCTTGGTCAGTGCCGATTTCCTCCCCCCCCACGC-CTTCCTCCACGTCAAAGGCTTTTGCATTGTAAAGCTACCCGCCTACCCGCCCCTCCCAATAAAAAAAAAAGAACATACACCAAT

3489  GATTCAGACACGCCCGCTGCCACGGGGTGT--CAAGA-CCACACCTACGC-TCGCCGGCTCACCGAGAATGGTGC-AATCGAAGAGATAAACACGGCTGATCTACTGGAAATGGTGCTGG
      |  || |  |       |     || || |         | |  |||| |    | |   | |  | || || | || || |       |  | ||||  || |    | |    | | ||
3551  GGGTCTTATTTGGTATTACCTGGTTTATTTAAAAAGATATACAGTAAGACATCCCATGGTACCAAAGACCGGGGCGAATCAGCGGGCCCCCATCATCTGAGAGAC-GAACAAATCGGCGG

3604  CTTCTGAAAACGCTCAAAGCGAACCCGGAATTC
      | | |   ||||| |  ||   | | |
3670  C-GCGGGCCGTG-TCAACGTCCACGTGTGCTGCGCTGCTGGCGTTGACAAGGCCCCGGCCTCCGCGTTGGATGCTCCGGTTGGGATCC
```

FIG. 4—*Continued.*

```
EHV-4 ICP 18.5
HSV-1 ICP 18.5
```

```
  1      E  L  A  R  F  -  M  V  T  A  A  K  G  -  D  W  S  I  S  E  F  Q  R  F  Y  C  F  E  G  V  T  G  V  T  A  Q  R  L  A
         |  |  |  |  |     |              |     |     |     |  |  |  |  |  |  |  |     |        |     |        |     |  |  |  |
638      E  L  A  R  F  I  M  G  A  G  G  S  G  A  H  W  A  V  S  E  F  Q  R  F  Y  C  F  D  G  I  S  G  I  T  P  T  Q  R  A  A

 39      W  K  Y  I  G  E  L  I  L  A  A  A  V  F  S  S  V  F  H  C  G  E  V  R  L  L  R  A  D  R  T  Y  P  N  T  N  G  A  Q  R
         |     |  |     |     |  |  |        |        |     |  |        |  |  |        |     |     |                 |        |
678      W  R  Y  I  R  E  L  I  I  A  T  T  L  F  A  S  V  Y  R  C  G  E  L  E  L  R  R  P  D  C  S  R  P  T  S  E  G  R  Y  R

 79      C  A  S  G  I  Y  I  T  Y  E  T  S  C  P  L  V  A  V  L  F  V  A  P  N  G  V  I  G  E  E  T  V  V  I  Y  D  S  D  V  F
               |     |  |              |  |  |  |  |           |        |  |        |  |              |  |     |  |  |  |
718      Y  P  P  G  V  Y  L  T  Y  D  S  D  C  P  L  V  A  I  V  E  S  A  P  D  G  C  I  G  P  R  S  V  V  V  Y  D  A  D  V  F

119      S  L  L  Y  T  V  L  Q  Q  L  A  P  G  S  G  A  N  (TERM)
         |  |     |  |     |  |        |     |  |  |
758      S  I  L  Y  S  V  L  Q  H  L  A  P  R  L  P  D  G  G  H  D  G  P  P  (TERM)
```

FIG. 5. Alignment of the predicted carboxy-terminal amino acids of the ICP18.5 protein of EHV-4 and HSV-1 (KOS). The top line is the EHV-4 sequence, and the bottom line is the HSV-1 sequence, shown in the single-letter amino acid code. Residue 1 in the EHV-4 sequence aligns with residue 638 in the HSV-1 sequence. Identical residues are indicated with a vertical bar, and dashes have been introduced to maximize homology between the two sequences.
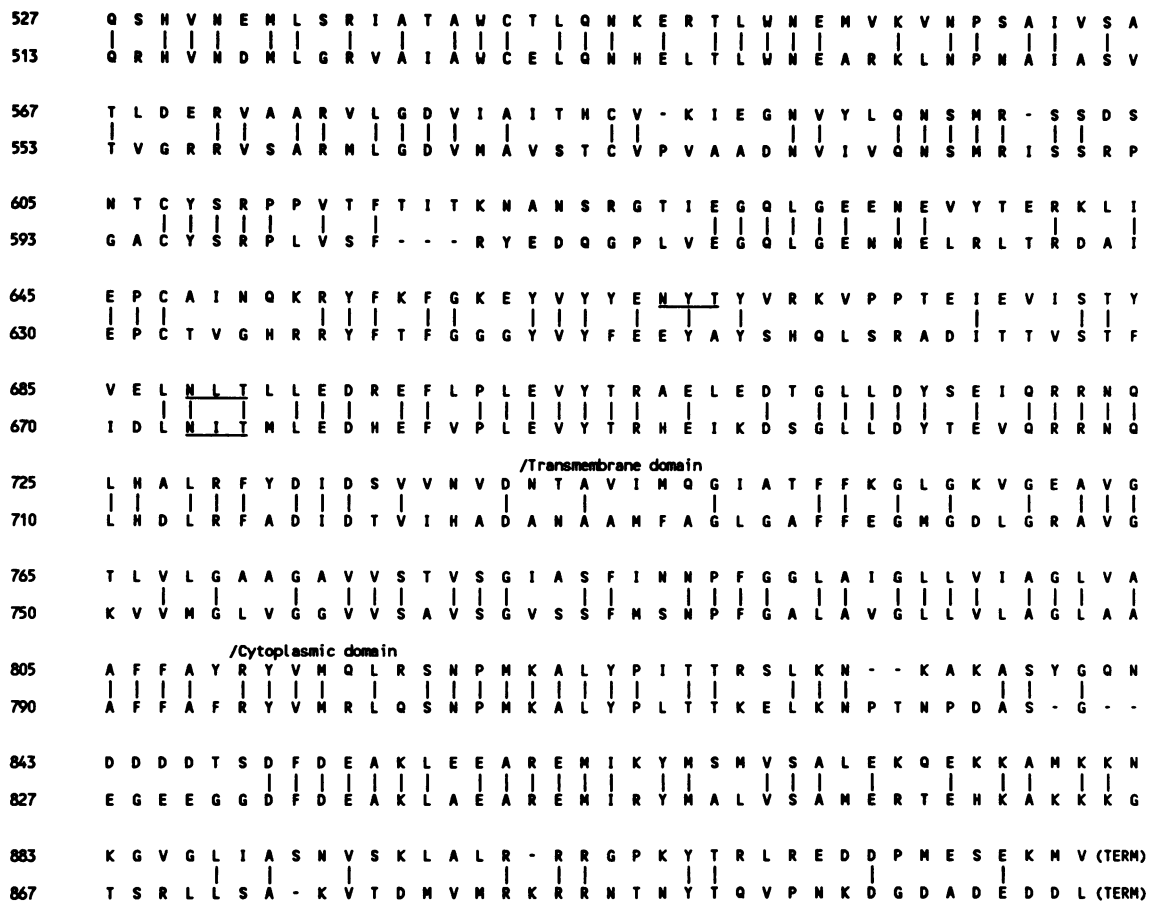
(-56 MSTCCRAICGPQRCYWRRDCGNLRQRRVLASIHRTPAAGSWLWSQLGNVNLPATSP -1)

EHV-4 gB
HSV-1 gB

```
                                                      Signal sequence/
1    M S K D S T S L G V R - T I V I A C L V L - L G C C I V E A - - - V P T T P - -
     |         |   |   |       |   |   |   |   |   | |       |   |         |     | |
1    M H Q G A P S W G R R W F V V W A L L G L T L G V L V A S A A P S S P G T P G V

34   S S Q P - - - - - - - - S T P A S T Q S A K T V D - - - Q T L L P T E T P D P L
     |             |         |       |   |       |   |               |         | |
41   A R D P G G E R G P C H S G A A A L G A A P T G D P K P K K N K K P K N P T P P

63   R L A - - - - - V - - - - - - - R E - - S G I L A E - - D G D F Y T C P P P T G
     | |           |               | |       | |       | |       | |   | | | | | |
81   R P A G D N A T V A A G H A T L R E H L R D I K A E N T D A N F Y V C P P P T G

87   S T V V R I E P P R S C P K F D L G R N F T E G I A V I F K E N I A P Y K F R A
     | | |       |   | |   | |       |     |   | | | | | | | | | | | | | | | |   |
121  A T V V Q F E Q P R R C P T R P E G Q N Y T E G I A V V F K E N I A P Y K F K A

127  N V Y Y K D I V V T K V W K G Y S H T S L S D R Y N D R V P V S V E E I F T L I
     |   | | | |   | |   |   | |           | |   |       | |   | |   | |   |   |   |
161  T M Y Y K D V T V S Q V W F G H R Y S Q F M G I F E D R A P V P F E E V I D K I

167  D S K G K C S S K A E Y L R D N I M H H A Y H D D E D V E L D L V P S K F A T
       | |   | | | | | | | | | | | |       | | |   | |       | |       |   | |
201  N A K G V C R S T A K Y V R N M L E T T A F H R D D H E T D M E L K P A N A A T

207  P G A R A W Q T T N D T T S Y V G W M P W R H Y T S T S V N C I V E E V E A R S
     |   | |   | |   | |   |           |       | |   | | | | | | | | | |   | | |
241  R T S R G W H T T - D L K Y N P S R V E A F H R Y G T T V N C I V E E V D A R S

247  V Y P Y D S F A L S T G D I V Y T S P F Y G L R S A A Q L E H N S Y A Q E R F R
     | | | |   | | |   | | |   | |   | | | | | |   |       | |       | |       | |
280  V Y P Y D E F V L A T G D F V Y M S P F Y G Y R E G S H T E H T T Y A A D R F K

287  Q V E G Y Q P R D L D S K L Q A G E P V T K N F I T T P H V T V S W N W T E K K
     | |   |       | | | | |   | |       | |     | |       | | |   | | | | | |   | |
320  Q V D G F Y A R D L T T K A R A T A P T T R N L L T T P K F T V A W D W V P K R

327  I E A C T L T K W K E V D E L V R D E F R G S Y R F T I R S I S S T F I S N I T
         | |   | | |   | | | |   | |       | |   | |   | |       | | |   | |   | | |
360  P S V C T M T K W Q E V D E M L R S E Y G G S F R F S S D A I S T T F T T N L T

367  Q F K L E D A P L T D C V S K E A K D A I D S I Y R K Q Y E S T H V F S G D V E
         | |     | | | | |   | |   | | | | | |   |       |   | | |   | |   |
400  E Y P L S R V D L G D C I G K D A R D A M D R I F A R R Y N A T H I K V G Q P Q

407  F Y L A R G G F L I A F R P M I S N E L A R L Y L N E L V R S N R T Y D L K N L
     | | |   | | | | | | | | | |   | |       | | | |   | | | |       | |       |
440  Y Y L A N G G F L I A Y Q P L L S N T L A E L Y - - - - V R - - - - E H L - - -

447  L N P N A N H N T N R T R R S L L S I P E P T P T Q E S L H R E Q I L H R L H K
                         |             |           |   | | |
469  - - - - - - - - - - R - E Q S - R K P P N P T P - - - - - - - - - - - - - P P P

487  R A V E A A N S T N S S N V T A K Q L E L I K T T S S I E F A M L Q F A Y D H I
       |                               |   | | | | | | | | |   | | |   |   | |   | |
484  G A - - - - - S A N A S - V - - - - - E R I K T T S S I E F A R L Q F T Y N H I
```

FIG. 6. Alignment of the predicted amino acid sequences of the gB proteins of EHV-4 and HSV-1 (KOS). The top line is the EHV-4 sequence, and the bottom line is the HSV-1 sequence, shown in the single-letter amino acid code. Identical residues are indicated with a vertical bar, and dashes show spaces introduced to maximize homology between the two sequences. The predicted signal peptidase cleavage site is indicated for both proteins, as are the hydrophilic surface, hydrophobic transmembrane, and cytoplasmic domains of the proteins. Potential N-linked glycosylation sites are underlined. The additional 56 amino acids that appear to form the start of an uncharacteristically long 84-amino-acid signal sequence for EHV-4 gB, as discussed in the text, are indicated in parentheses above the alignment (residues −56 to −1).

The aligned DNA sequences of EHV-4 gB and HSV-1 gB are shown in Fig. 4. The two genes are well conserved and show DNA homology of 52%.

The main features of the transcriptional and translational control signals of the EHV-4 gB gene are as follows. (a) A TATA box, AATATAT, is found at nucleotides 119 to 125. This aligns well with the HSV-1 gB TATA box (ATATATT) at nucleotides 234 to 240 in the HSV-1 gB sequence. The first

```
527  Q S H V N E N L S R I A T A W C T L Q N K E R T L W N E N V K V N P S A I V S A
513  Q R H V N D N L G R V A I A W C E L Q N H E L T L W N E A R K L N P N A I A S V

567  T L D E R V A A R V L G D V I A I T H C V - K I E G N V Y L Q N S N R - S S D S
553  T V G R R V S A R N L G D V N A V S T C V P V A A D N V I V Q N S N R I S S R P

605  N T C Y S R P P V T F T I T K N A N S R G T I E G Q L G E E N E V Y T E R K L I
593  G A C Y S R P L V S F - - - R Y E D Q G P L V E G Q L G E N N E L R L T R D A I

645  E P C A I N Q K R Y F K F G K E Y V Y Y E N Y T Y V R K V P P T E I E V I S T Y
630  E P C T V G H R R Y F T F G G G Y V Y F E E Y A Y S H Q L S R A D I T T V S T F

685  V E L N L T L L E D R E F L P L E V Y T R A E L E D T G L L D Y S E I Q R R N Q
670  I D L N I T N L E D H E F V P L E V Y T R H E I K D S G L L D Y T E V Q R R N Q

                      /Transmembrane domain
725  L H A L R F Y D I D S V V N V D N T A V I N Q G I A T F F K G L G K V G E A V G
710  L H D L R F A D I D T V I H A D A N A A N F A G L G A F F E G N G D L G R A V G

765  T L V L G A A G A V V S T V S G I A S F I N N P F G G L A I G L L V I A G L V A
750  K V V N G L V G G V V S A V S G V S S F N S N P F G A L A V G L L V L A G L A A

         /Cytoplasmic domain
805  A F F A Y R Y V N Q L R S N P N K A L Y P I T T R S L K N - - K A K A S Y G Q N
790  A F F A F R Y V N R L Q S N P N K A L Y P L T T K E L K N P T N P D A S - G - -

843  D D D D T S D F D E A K L E E A R E N I K Y N S N V S A L E K Q E K K A N K K N
827  E G E E G G D F D E A K L A E A R E N I R Y N A L V S A N E R T E H K A K K K G

883  K G V G L I A S N V S K L A L R - R R G P K Y T R L R E D D P N E S E K N V (TERM)
867  T S R L L S A - K V T D N V N R K R R N T N Y T Q V P N K D G D A D E D D L (TERM)
```

FIG. 6—Continued.

transcribed base of eucaryotic mRNA is usually an A residue surrounded by pyrimidine residues (7).

The sequence CAGT was observed at nucleotides 149 to 152. Since the A residue at position 150 is 25 nucleotides downstream from the TATA box, it is likely that this is the first transcribed base of EHV-4 gB mRNA. This would be in keeping with the typical 24- to 32-nucleotide spacing between the TATA box and mRNA cap site (7). Furthermore, this presumptive mRNA cap site aligns well with the HSV-1 gB mRNA cap site.

(b) The sequence ATTG at nucleotides 38 to 41 reads CAAT on the other strand and is likely to function as a CAT box, since it is about 100 nucleotides upstream from the mRNA cap site (6). This CAT box is identical to, and aligns perfectly with, the HSV-1 gB CAT box.

(c) The modified scanning hypothesis of translation predicts that the first AUG in a mRNA serves as the initiation codon, assuming that the local environment will allow this to occur (23). The first AUG in the mRNA occurs at nucleotides 270 to 272 and is in frame with two nearby AUGs in the sequence. Use of this first ATG as the initiation codon would give rise to an unusually long signal sequence of 84 amino acids, and similarly long signal sequences of 53 and 67 amino acids have also been reported for the gB homologs of PRV and bovine herpesvirus 1, respectively (34, 44).

The second available ATG is at nucleotides 412 to 414 but is immediately followed by an in-frame chain termination codon at nucleotides 415 to 417. The next available ATG is at nucleotides 438 to 440, and the DNA sequences around both this ATG and the ATG at nucleotides 270 to 272 suggest that either ATG could function as an efficient initiation site (22). Thus, two initiation codons can be predicted for EHV-4 gB, although we consider the first ATG as the most likely initiation codon for this gene.

(d) The stop codon, TAA, occurs at nucleotides 3195 to 3197. The first polyadenylation signal, AATAAA (3), occurs 9 nucleotides past the termination codon and occurs about 20 nucleotides before the poly(A) addition site (15). A second polyadenylation signal occurs at nucleotides 3470 to 3475, suggesting that other mRNA(s) which may be present in this region would terminate at this point.

(e) An ORF encoding the protein ICP18.5 in HSV-1 has been shown to terminate 10 nucleotides before the HSV-1 gB ATG initiation codon. ICP18.5 has been reported to have a role in the transport of viral glycoproteins (30).

On examining the EHV-4 sequence, we located the termination codon (TAG) of an upstream ORF at nucleotides 406 to 408, occurring 136 nucleotides after the predicted EHV-4 gB ATG initiation codon. Analysis of this ORF back to nucleotide 1 revealed that this region encodes a protein showing 48% homology to HSV-1 ICP18.5 (Fig. 5). A similar protein has also been reported for PRV (34) and bovine herpesvirus 1 (44), in which the ICP18.5 protein-coding sequence actually overlaps the gB homolog-coding sequence by 44 and 47 codons, respectively, which compares with the 45-codon overlap predicted here for EHV-4.

Thus, in all four viruses, there is a great constraint on this region of the genome since it probably serves as the tran-
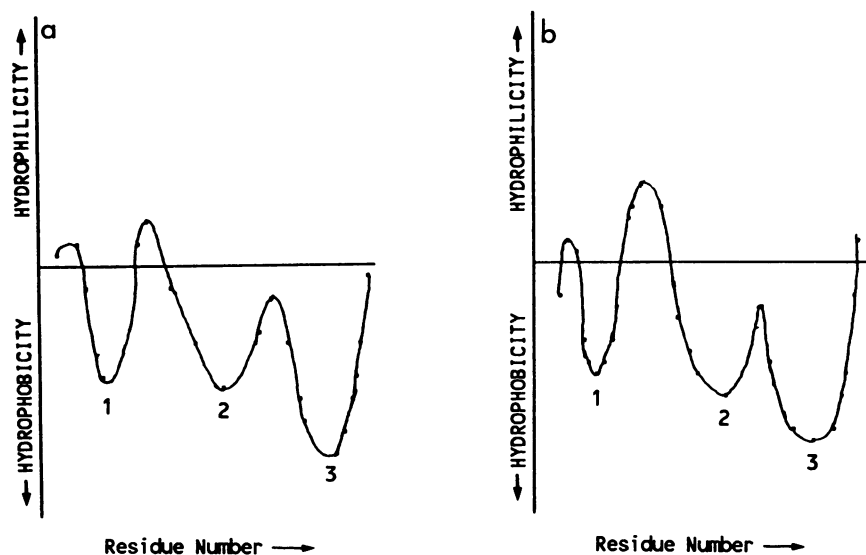
FIG. 7. Hydropathic analysis of the transmembrane region of the gB proteins of EHV-4 (residues 734 to 812) (a) and HSV-1 (KOS) (residues 720 to 799) (b), as determined by the method of Kyte and Doolittle (24). The area above the *x* axis denotes hydrophilicity and that below the *x* axis denotes hydrophobicity. The predicted three membrane-spanning segments of this domain, which are indicated by hydrophobic peaks, are numbered for both proteins. These segments are perfectly aligned between the two proteins and are highly conserved regions of the molecules (see Fig. 6 and text).

scriptional control domain of the gB gene and also encodes the carboxy-terminal amino acids of the protein ICP18.5. Further constraints are imposed on this region in EHV-4, PRV, and bovine herpesvirus 1 as it must also encode the amino-terminal amino acids of the unusually long signal sequence of the gB homolog of these viruses.

(ii) Primary structure of the EHV-4 gB protein which shows homology to HSV-1 gB. The alignment of the EHV-4 and HSV-1 gB proteins is shown in Fig. 6. The EHV-4 protein is predicted to have a molecular mass of 110 kilodaltons (kDa), which is larger than the 100.3 kDa reported for HSV-1 gB (8, 31). In this alignment, 47% of the amino acids are perfectly matched between the two proteins.

The codon usage and amino acid composition of the EHV-4 gB protein are shown in Table 1. The G+C content of nucleotides 1, 2, and 3 of the codons is 51, 41, and 50%, respectively. Unlike HSV-1 gB, there is no strong preference for a G or C in the third base of the codons in EHV-4 gB (8), which is a reflection of the lower G+C content of the coding region of EHV-4 gB (47%), as opposed to 66% for the HSV-1 gB-coding region. The EHV-4 gB protein contains features characteristic of all envelope glycoproteins, namely, a 5' hydrophobic signal sequence, an external hydrophilic surface domain, a hydrophobic membrane-spanning domain, and a basic, highly charged cytoplasmic anchor domain, as predicted by the hydropathic analysis of Kyte and Doolittle (24) and the secondary structure analyses of Chou and Fasman (10).

(a) Signal sequence domain. Features characteristic of membrane insertion (signal) sequences are a hydrophobic core preceded by positively charged residues (32), immediately followed by a signal peptidase cleavage site near a predicted beta turn (42). A hydrophobic sequence of amino acids was found from Ile-13 to Val-26 in the EHV-4 gB protein which we believe would be of sufficient length to serve as a signal sequence hydrophobic core which spans the membrane but not to serve as an anchor sequence (32). Furthermore, the protein contains a Leu residue at core position 2, a feature which is conserved in many eucaryotic

and procaryotic signal sequences (32), although the characteristic Val-Val pair at core positions 7 and 8 is replaced by Ile-Val. The hydrophobic core sequence is immediately preceded by 12 amino acids, of which 2 are positively charged, and by a further 56 amino acids which are predominantly hydrophilic (residues −56 to −1, Fig. 6). A similar hydrophilic region has also been reported for the gB homolog of PRV (34) and bovine herpesvirus 1 (44), which has led Robbins et al. (34) to speculate that these unusually long leader sequences are a result of constraints imposed on this region of the viral genome to encode both an upstream protein and a functional signal sequence.

Cleavage of the EHV-4 gB signal sequence is predicted to

TABLE 1. Codon usage and predicted amino acid composition of EHV-4 gB (110 kDa)

1st (%)	2nd				3rd	Residue	No.
	A (32%)	C (25%)	G (17%)	T (26%)			
						Ala	71
A (30)	36 Lys	17 Thr	14 Arg	21 Ile	A	Arg	65
	36 Asn	23 Thr	18 Ser	9 Ile	C	Asn	49
	16 Lys	14 Thr	9 Arg	18 Met	G	Asp	47
	13 Asn	22 Thr	6 Ser	22 Ile	T	Cys	18
						Gln	29
C (20)	16 Gln	19 Pro	13 Arg	19 Leu	A	Glu	69
	12 His	11 Pro	16 Arg	14 Leu	C	Gly	46
	13 Gln	10 Pro	7 Arg	12 Leu	G	His	17
	5 His	5 Pro	6 Arg	16 Leu	T	Ile	52
						Leu	81
G (31)	31 Glu	19 Ala	16 Gly	13 Val	A	Lys	52
	30 Asp	20 Ala	12 Gly	10 Val	C	Met	18
	38 Glu	11 Ala	12 Gly	17 Val	G	Phe	33
	17 Asp	21 Ala	6 Gly	33 Val	T	Pro	45
						Ser	79
T (19)	1 Stop	11 Ser	0 Stop	10 Leu	A	Thr	76
	33 Tyr	17 Ser	5 Cys	6 Phe	C	Trp	12
	0 Stop	6 Ser	12 Trp	10 Leu	G	Tyr	43
	10 Tyr	21 Ser	13 Cys	27 Phe	T	Val	73

occur after Ala-28, and this cleavage site is immediately preceded by a beta turn. That this is a probable cleavage site is also supported by the observations that a helix-breaking residue (Gly or Pro) or a large polar residue (Glu) usually occurs four to eight residues before the cleavage site (43); in the EHV-4 gB protein, a helix-breaking Gly residue occurs at position 22, six residues prior to the cleavage site, and a polar residue (Glu) is found at position 27. Most signal sequences contain either an Ala or Gly at their carboxy terminus (43). The EHV-4 gB signal sequence carries an Ala at its carboxy terminus, and the cleavage site is perfectly aligned with that of HSV-1 gB, suggesting that the cleavage site predicted for EHV-4 gB is likely to be accurate (Fig. 6).

(b) Hydrophobic transmembrane domain. The hydropathic profile of the EHV-4 gB protein suggests the presence of a transmembrane domain (amino acids 741 to 809) containing three antiparallel segments connected to each other by a very short turn region which traverse the membrane (Fig. 7), as has previously been predicted for HSV-1 gB (31). These membrane-traversing segments are thought to assume a helical conformation (14).

The Chou and Fasman (10) analysis of these segments suggests that the first segment (Asn-741 to Lys-754) has the potential to adopt both alpha-helical and beta-sheet conformations, as can the second segment (Leu-766 to Asn-786), while the third segment (Phe-789 to Tyr-809) has an alpha-helical structure. This transmembrane region is highly conserved between EHV-4 gB and HSV-1 gB, with 39 of the 69 amino acids composing this domain being conserved. The first segment of this domain is slightly shorter in the EHV-4 protein (14 amino acids) than in the HSV-1 protein (21 amino acids), while the second and third segments are predicted to be 21 amino acids long in both proteins. A short turn followed by a random coil is predicted to occur between segments 1 and 2 of the EHV-4 protein, although this random coiling does not appear to occur in the corresponding region of the HSV-1 protein.

However, the triple traverse of the membrane by gB in this way has yet to be confirmed experimentally and is predicted solely from analysis of sequence data obtained for both HSV-1 and EHV-4 gB proteins.

(c) Hydrophilic surface domain. The hydrophilic surface domain of EHV-4 gB is predicted to extend from Val-29 to Asp-740. This domain is thought to reside on the outer surface of the viral envelope and to contain the antigenic determinants against which virus-neutralizing antibody is directed (20, 31). This surface domain contains 11 potential N-linked glycosylation sites, Asn-X-Ser, in which X is not proline (35).

Four of the six potential N-linked glycosylation sites in HSV-1 gB are perfectly conserved with EHV-4 gB, and all 10 cysteine residues outside the signal sequence are also conserved, suggesting that the proteins possess similar secondary and tertiary structures. As predicted for HSV-1 gB, most of the glycosylation sites are located at the junction of alpha-helical and beta-sheet domains, suggesting that these sites are exposed to the outer surface of the molecule (31).

Eighteen major alpha-helical domains, nine major beta-sheet domains, and five major turns containing at least seven amino acids are predicted for the amino acid sequence of EHV-4 gB.

(d) Cytoplasmic domain. The remaining 110 amino acids of EHV-4 gB constitute the cytoplasmic domain, which is predicted to be hydrophilic and to adopt an alpha-helical conformation. This carboxy-terminal segment is thought to function as a cytoplasmic anchor domain.

This domain shares a primary amino acid homology of 49% with the cytoplasmic domain of HSV-1 gB, with 54 of the 110 amino acids constituting this domain being conserved.

Of the 110 amino acids in the cytoplasmic domain of EHV-4 gB, 24 are positively charged and 17 are negatively charged, giving an overall positive charge (the HSV-1 gB cytoplasmic domain is also positively charged and contains 109 amino acids).

DISCUSSION

We identified the glycoprotein gB gene within a genomic library of EHV-4 using both 5' and 3' HSV-1 gB DNA probes in dot-blot and Southern blot analyses. The gene was localized to a 2.95-kb *Bam*HI-*Eco*RI subfragment of the *Bam*HI C viral restriction fragment and the right-hand end of the adjacent *Bam*HI M fragment. We sequenced this region and identified an ORF of 2,925 nucleotides, the primary translation product of which would be a 975-amino-acid protein with a molecular mass of 110 kDa, slightly larger than the 903-amino-acid (100.3-kDa) protein predicted for HSV-1 gB. The genes for these proteins show a DNA homology of 52%, and the transcriptional control signals of the two genes are highly conserved. This transcriptional control domain also encodes the carboxy-terminal amino acids of the protein ICP18.5, with EHV-4 ICP18.5 showing 48% homology at the amino acid level to HSV-1 ICP18.5 in this region.

EHV-4 gB and HSV-1 gB proteins are 47% homologous and possess highly conserved hydrophilic surface, hydrophobic transmembrane, and cytoplasmic anchor domains, while the signal sequence domains show the least homology. Two signal sequences may be predicted from the data. A sequence of 28 amino acids would be similar to the 30-amino-acid sequence of HSV-1 (KOS) gB and would be in keeping with the length and structure of other glycoprotein signal sequences. However, an alternative signal sequence of 84 amino acids is likely given the precedent of an 85-amino-acid signal sequence predicted by M. Whalley for EHV-1 gB (personal communication) and signal sequences of 53 and 67 amino acids which have been predicted for the PRV and bovine herpesvirus 1 gB homologs, respectively (34, 44). Whichever sequence is adopted, the mature form of EHV-4 gB is predicted to be 891 amino acids (100.8 kDa) in size.

The hydrophobic domain of both proteins is predicted to be 69 amino acids long and to contain three antiparallel segments passing through the membrane (31). From these predictions, segments 2 and 3 appear to be identical in size (21 amino acids), whereas segment 1 is thought to be somewhat shorter in EHV-4 gB (14 amino acids compared with 21 amino acids for HSV-1 gB). Furthermore, the two proteins possess similar-sized hydrophilic surface domains which contain epitopes for virus-neutralizing antibody (20, 31), and all 10 cysteine residues and four of six potential N-linked glycosylation sites in this domain are perfectly conserved. These observations suggest that the two proteins adopt similar secondary and tertiary structures.

EHV-4 gB contains five more potential N-linked glycosylation sites than HSV-1 gB, and it is interesting to note the presence of three such sites immediately adjacent to each other (residues 493 to 501), with the central site being conserved with one such site in HSV-1 gB. It is possible that oligosaccharides are not added at all of these three sites, and no direct evidence exists that fully processed EHV-4 gB contains 11 oligosaccharide chains.

The cytoplasmic anchor domain of EHV-4 gB is composed of 110 amino acids (compared with 109 amino acids for HSV-1 gB) and serves to anchor the protein to the membrane.

It has been proposed that gB is part of a multiprotein complex which determines the social behavior of infected cells and the structure of infected cell membranes (36). The cytoplasmic domain is thought to interact with virion tegument proteins and with other membrane proteins to affect the social behavior of infected cells (29), since *syn* mutations have been mapped to this domain in HSV-1 gB (9, 31) (a single amino acid substitution at position 857 [Arg to His] gives rise to a syncytial phenotype [9]). It has been predicted that a stretch of amino acids before this Arg residue may adopt both a hydrophobic and hydrophilic character and that this region interacts with other proteins (29).

The gB counterpart in VZV, gpII, is known to exist as a disulfide-linked dimer (18). The mature species of VZV gpII has further been suggested to exist as a disulfide-linked heterodimer generated by an in vivo proteolytic cleavage between Arg-431 and Ser-432 of gpII (21). We observed that the residues Asn-426–Thr-427–Arg-428–Ser-429–Arg-430–Arg-431–Ser-432 immediately before and including this cleavage site are aligned with the residues Asn-454–Thr-455–Asn-456–Arg-457–Thr-458–Arg-459–Arg-460–Ser-461 of EHV-4 gB. Six of these residues are perfectly matched, including the cleavage site. It would therefore be interesting to speculate whether EHV-4 gB would also exist as a disulfide-linked dimer as a result of proteolytic cleavage between Arg-460 and Ser-461. This would bisect the molecule, as is the case with VZV gpII proteolytic cleavage.

We compared the predicted amino acid sequence of EHV-4 gB with the VZV gpII, PRV gpII, Epstein-Barr virus gB, and human cytomegalovirus gB amino acid sequences and report homologies of 51, 54, 30, and 29%, respectively. The gB glycoproteins of the alphaherpesviruses show a similar degree of homology to each other (47 to 54%) and significantly less homology to the gB protein of the betaherpesvirus human cytomegalovirus (29%) and the gammaherpesvirus Epstein-Barr virus (30%). Furthermore, Epstein-Barr virus gB and human cytomegalovirus gB are only 30% homologous, suggesting that each family of herpesviruses has separate ancestral origins.

We also report that the EHV-4 gB glycoprotein shows 88% homology to the gB glycoprotein of EHV-1 (M. Whalley, personal communication).

It would now be of interest to produce EHV-4 gB protein in a suitable *E. coli* expression vector system and to assess the efficacy of the purified protein as a subunit vaccine.

The prediction of the structure of potentially immunogenic viral glycoproteins by DNA sequence analysis and the subsequent expression, purification, and testing of such proteins as potential vaccines against the virus will, it is hoped, lead us to a better understanding of viral infection.

ACKNOWLEDGMENTS

LITERATURE CITED

1. **Allen, G. P., and J. T. Bryans.** 1986. Molecular epizootiology, pathogenesis and prophylaxis of equine herpesvirus-1 infections, p. 78–144. *In* R. Pandey (ed.), Progress in veterinary microbiology and immunology, vol. 2. S. Karger, Basel.

2. **Allen, G. P., and M. R. Yeargan.** 1987. Use of λgt11 and monoclonal antibodies to map the genes for the six major glycoproteins of equine herpesvirus 1. J. Virol. **61:**2454–2461.

3. **Berget, S. M.** 1984. Are U4 small nuclear ribonucleoproteins involved in polyadenylation? Nature (London) **309:**179–182.

4. **Biggin, M. D., T. J. Gibson, and C. F. Hong.** 1983. Buffer gradient gels and ³⁵S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80:**3963–3965.

5. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7:**1513–1523.

6. **Breathnach, R., and P. Chambon.** 1981. Organisation and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50:**349–383.

7. **Busslinger, M., R. Portman, J. Irminger, and M. Birnsteil.** 1980. Ubiquitous and gene-specific regulatory 5' sequences in a sea urchin histone DNA clone coding for histone protein variants. Nucleic Acids Res. **8:**957–978.

8. **Bzik, D. J., B. A. Fox, N. A. DeLuca, and S. Person.** 1984. Nucleotide sequence specifying the glycoprotein gene, gB, of herpes simplex virus type 1. Virology **133:**301–314.

9. **Bzik, D. J., B. A. Fox, N. A. DeLuca, and S. Person.** 1984. Nucleotide sequence of a region of the herpes simplex virus type 1 gB glycoprotein gene: mutations affecting rate of virus entry and cell fusion. Virology **137:**185–190.

10. **Chou, P. Y., and G. D. Fasman.** 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Protein Chem. **47:**45–148.

11. **Cranage, M. P., T. Kouzarides, A. T. Bankier, S. Satchwell, K. Weston, P. Tomlinson, B. Barrell, H. Hart, S. E. Bell, A. C. Minson, and G. L. Smith.** 1986. Identification of the human cytomegalovirus glycoprotein B gene and induction of neutralizing antibodies via its expression in recombinant vaccinia virus. EMBO J. **5:**3057–3063.

12. **Cullinane, A. A., F. J. Rixon, and A. J. Davison.** 1988. Characterization of the genome of equine herpesvirus 1 subtype 2. J. Gen. Virol. **69:**1575–1590.

13. **DeLuca, N., D. Bzik, V. C. Bond, S. Person, and W. Snipes.** 1982. Nucleotide sequences of herpes simplex virus type 1 (HSV-1) affecting virus entry, cell fusion, and production of glycoprotein gB (VP7). Virology **122:**411–423.

14. **Engelman, D. M., and T. A. Steitz.** 1981. The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis. Cell **23:**411–422.

15. **Fitzgerald, M., and T. Shenk.** 1981. The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell **24:**251–260.

16. **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.

17. **Glorioso, J., C. H. Schroder, G. Kurnel, M. Szczesiul, and M. Levine.** 1984. Immunogenicity of herpes simplex virus glycoproteins gC and gB and their role in protective immunity. J. Virol. **50:**805–812.

18. **Grose, C., D. P. Edwards, K. A. Weigle, W. E. Friedrichs, and W. L. McGuire.** 1984. Varicella-zoster virus specific gp140: a highly immunogenic and disulfide-linked structural glycoprotein. Virology **132:**138–146.

19. **Holland, L. E., R. M. Sandri-Goldin, A. L. Goldin, J. C. Glorioso, and M. Levine.** 1984. Transcriptional and genetic analyses of the herpes simplex virus type 1 genome: coordinates 0.29 to 0.45. J. Virol. **49:**947–959.

20. **Hopp, T. P., and K. R. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78:**3824–3828.

21. **Keller, P. M., A. J. Davison, R. S. Lowe, C. D. Bennett, and R. W. Ellis.** 1986. Identification and structure of the gene encoding gpII, a major glycoprotein of varicella-zoster virus. Virology **152:**181–191.

22. **Kozak, M.** 1984. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47:**1–45.

23. **Kozak, M.** 1984. Point mutations close to the AUG initiator

codon affect the efficiency of translation of rat preproinsulin *in vivo*. Nature (London) **308**:241–246.

24. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.

25. **Little, S. P., J. T. Jofre, R. J. Courtney, and P. A. Schaffer.** 1981. A virion-associated glycoprotein essential for infectivity of herpes simplex virus type 1. Virology **115**:149–160.

26. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

27. **Manservigi, R., P. G. Spear, and A. Buchan.** 1977. Cell fusion induced by herpes simplex virus is promoted and suppressed by different viral glycoproteins. Proc. Natl. Acad. Sci. USA **74**:3913–3917.

28. **O'Callaghan, D. J., G. A. Gentry, and C. C. Randall.** 1983. The equine herpesviruses, p. 215–318. *In* B. Roizman (ed.), The herpesviruses, vol. 2. Plenum Publishing Corp., New York.

29. **Pellett, P. E., M. D. Biggin, B. Barrell, and B. Roizman.** 1985. Epstein-Barr virus genome may encode a protein showing significant amino acid and predicted secondary structure homology with glycoprotein B of herpes simplex virus 1. J. Virol. **56**:807–813.

30. **Pellett, P. E., F. J. Jenkins, M. Ackermann, M. Sarmiento, and B. Roizman.** 1986. Transcription initiation sites and nucleotide sequence of a herpes simplex virus 1 gene conserved in Epstein-Barr virus genome and reported to affect the transport of viral glycoproteins. J. Virol. **60**:1134–1140.

31. **Pellett, P. E., K. G. Kousoulas, L. Pereira, and B. Roizman.** 1985. Anatomy of the herpes simplex virus 1 strain F glycoprotein B gene: primary sequence and predicted protein structure of the wild type and of monoclonal antibody-resistant mutants. J. Virol. **53**:243–253.

32. **Perlman, D., and H. O. Halvorson.** 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. J. Mol. Biol. **167**:391–409.

33. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick-translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.

34. **Robbins, A. K., D. J. Dorney, M. W. Wathen, M. E. Whealy, C. Gold, R. J. Watson, L. E. Holland, S. D. Weed, M. Levine, J. C. Glorioso, and L. W. Enquist.** 1987. The pseudorabies virus gII gene is closely related to the gB glycoprotein gene of herpes simplex virus. J. Virol. **61**:2691–2701.

35. **Rose, J. K., R. F. Doolittle, A. Anilionis, P. J. Curtis, and W. H. Wunner.** 1982. Homology between the glycoproteins of vesicular stomatitis virus and rabies virus. J. Virol. **43**:361–364.

36. **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.

37. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

38. **Sarmiento, M., M. Haffey, and P. G. Spear.** 1979. Membrane proteins specified by herpes simplex viruses. III. Role of glycoprotein VP7 (B2) in virion infectivity. J. Virol. **29**:1149–1158.

39. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.

40. **Spear, P. G.** 1985. Glycoproteins specified by herpes simplex viruses, p. 315–356. *In* B. Roizman (ed.), The herpesviruses, vol. 3. Plenum Publishing Corp., New York.

41. **Vestergaard, B. F.** 1980. Herpes simplex virus antigens and antibodies. A survey of studies based on quantitative immunoelectrophoresis. Rev. Infect. Dis. **2**:899–913.

42. **von Heijne, G.** 1984. How signal sequences maintain sequence specificity. J. Mol. Biol. **173**:243–351.

43. **Watson, M. E. E.** 1984. Compilation of published signal sequences. Nucleic Acids Res. **12**:5144–5164.

44. **Whitbeck, J. C., L. J. Bello, and W. C. Lawrence.** 1988. Comparison of the bovine herpesvirus 1 gI gene and the herpes simplex virus type 1 gB gene. J. Virol. **62**:3319–3327.