Identification and Nucleotide Sequence of a Gene in Equine Herpesvirus 1 Analogous to the Herpes Simplex Virus Gene Encoding the Major Envelope Glycoprotein gB

By J. MILLAR WHALLEY,* GRAHAM R. ROBERTSON, NORMA A. SCOTT, GRANT C. HUDSON,† CHRISTOPHER W. BELL AND LYNN M. WOODWORTH‡
School of Biological Sciences, Macquarie University, Sydney, New South Wales 2109, Australia

(Accepted 21 October 1988)

SUMMARY

A gene in equine herpesvirus 1 (EHV-1; equine abortion virus) equivalent to the gB glycoprotein gene of herpes simplex virus (HSV) has been identified by DNA hybridization and nucleotide sequencing. A 4.3 kbp EHV-1 PstI–ClaI sequence (0.40 to 0.43 map units) contained an open reading frame flanked by appropriate control elements and was capable of encoding a polypeptide of 980 amino acids. This had 50 to 60% identity over a 617 amino acid conserved region with the gB gene products of HSV and three other alphaherpesviruses, and 20 to 30% identity with those of human cytomegalovirus and Epstein–Barr virus. Analysis of the amino acid sequence predicts a long signal peptide, hydrophobic and hydrophilic domains and N-glycosylation sites, and has identified a probable internal proteolytic cleavage site. The EHV-1 gB open reading frame appears to be overlapped at its 5' end by 135 nucleotides of the 3' end of an upstream open reading frame the potential translation product of which has approximately 50% identity with HSV gene ICP 18.5 and VZV gene 30 products.

INTRODUCTION

The alphaherpesvirus equine herpesvirus 1 (EHV-1) is a major cause of abortion in mares and severe neonatal disease in foals, and has also been associated with neurological disorders (O'Callaghan et al., 1983; Campbell & Studdert, 1983; Sabine et al., 1983; Allen & Bryans, 1986). The genome of EHV-1 (Whalley et al., 1981; Henry et al., 1981) is generally collinear with the genomes of herpes simplex virus (HSV), pseudorabies virus (PRV) and varicella-zoster virus (VZV), as determined by molecular hybridization experiments (Davison & Wilkie, 1983). Analysis of the organization and function of the EHV-1 genome is therefore not only relevant for elucidating the mechanisms underlying EHV-1 infection, but also can help to identify key features of herpesvirus genomes through comparative molecular biology.

Among many herpesvirus genes being studied in detail, those encoding virion structural glycoproteins (reviewed for HSV by Spear, 1985) have received attention due to their roles in the infectious process and their ability to invoke an immune response (Courtney, 1984). One of these, the major envelope glycoprotein gB, which occurs as prominent 14 nm spikes on virion surfaces (Stannard et al., 1987), is required for virus entry and for cell fusion (Delaux et al., 1982). It has been suggested that in the case of Epstein–Barr virus (EBV), the gB counterpart (gp110) may also have a role in reorganizing cell membranes to permit virus egress (Gong et al., 1987).

HSV gB has been shown, either by temperature-sensitive mutant studies (Glorioso et al., 1984) or as an immunopurified protein (Chan et al., 1985), to invoke circulating antibody and cell-mediated responses which protected mice against lethal challenge with the virus. A role for gB in

† Present address: School of Biological Sciences, Flinders University, Bedford Park, South Australia 5042, Australia.
‡ Present address: Mauri Research Centre, 69 Epping Road, North Ryde, New South Wales 2113, Australia.
human cell-mediated immunity has been indicated by activation of HSV-specific T cells by
cloned HSV-1 gB expressed in mammalian cell culture (Zarling et al., 1986). Vaccination of mice
with L cell lines expressing gB resulted in protection from viral challenge and inhibited the
establishment of latency (Blacklaws et al., 1987). Vaccination of rabbits with a recombinant
vaccinia virus expressing the human cytomegalovirus (HCMV) ‘gB’ gene produced antisera
which could neutralize HCMV infectivity in vitro (Cranage et al., 1986). The conservation of this
gene has been demonstrated for the human herpesviruses by DNA hybridization and nucleotide
sequencing (HSV-1 and -2, Bzik et al., 1984, 1986; Pellet et al., 1985b; VZV, Keller et al., 1986;
EBV, Pellet et al., 1985a; HCMV, Cranage et al., 1986) and also for PRV by Robbins et al.
(1987). EHV-1 glycoproteins have been characterized by Turtinen & Allen (1982) and mapped
on the genome using expression in Jgt11 vectors and a panel of monoclonal antibodies (Allen &
Yeargen, 1987). Evidence for a gB-like protein in EHV-1 has come from antigenic cross-
reactivity and peptide analysis (Snowden & Halliburton, 1985; Snowden et al., 1985). Here we
describe the identification and nucleotide sequence of an EHV-1 gB gene and compare it with
the analogous genes in several other herpesviruses.

METHODS

Viral and plasmid DNA preparation. EHV-1 (isolate HVS 25A) was grown in BHK-21 cells and DNA was
prepared by a standard method (Whalley et al., 1981). Recombinant plasmids pMAC209 and pMAC221 containing
EHV-1 BamHI I and BamHI IBG fragments, respectively, inserted in pBR322 (Robertson & Whalley, 1985), were
grown in Escherichia coli HB101 cells. Plasmid pGX37 was kindly donated by Dr V. Preston and Professor J.
Subak-Sharpe (MRC Virology Unit, Institute of Virology, University of Glasgow, Glasgow, U.K.) and contained the
BamHI G fragment of HSV-1 inserted in pAT153. Plasmids were amplified with chloramphenicol and purified by
SDS lysis and caesium chloride gradient ultracentrifugation, essentially as described by Maniatis et al.
(1982). To construct a hybridization probe containing only sequences coding for HSV-1 gB, pGX37 was digested
with the restriction enzymes BamHI and SalI, ligated into M13 mp18 followed by selection of a subclone containing a 1.74 kbp region of the HSV-1 gB gene (deduced from Bzik et al., 1984). For initial identification of gB-
like regions in EHV-1 by DNA sequencing, a 0.6 kbp HindIII/ClaI fragment and a 2.9 kbp BamHI/ClaI fragment
of pMAC209 were subcloned in defined orientations into M13 mp10 or mp11 phages.

DNA mapping and hybridization. Restriction endonuclease BamHI, SalI, CiaI, HindIII and PstI (Boehringer
Mannheim) were used under conditions recommended by the manufacturer, and the resulting DNA fragments
were characterized by agarose gel electrophoresis. For hybridization experiments EHV-1 DNA bands were
Southern-transferred to nitrocellulose (Schleicher & Schuell, BA85), and HSV-1 plasmid DNA was labelled by
nick translation using a kit from Bresapect. Recombinant HSV gB M13 template DNA was also used in
hybridization experiments and was labelled with [\(^{32}\)P]dATP using a modified sequencing reaction mix which
lacked any dideoxynucleotide. Hybridization was in 6 x SSC, 0.5% SDS, 5 x Denhardt’s solution and 100 µg/ml
denatured salmon sperm DNA at 55°C for 16 to 20 h. Washes were in 2 x SSC, 0.1% SDS at 55°C.

DNA sequencing. DNA from recombinant M13 phages was sequenced by the dideoxynucleotide chain
termination method. For the initial identification of the EHV-1 gB gene, sequence was obtained from BamHI,
ClaI or HindIII restriction sites using forced clones. The subsequent sequencing strategy was based on the method
of Dale et al. (1985) with the generation of several deletion series of M13 mp9 and mp18 overlapping clones
covering both DNA strands. Reactions containing dITP were also run to clarify or reveal zones of GC
compression on gels, particularly for analysis of the sequence around possible translation start sites. Nucleotides,
dideoxynucleotides, [\(^{32}\)P]dATP and Klenow enzyme were from Bresapect or Pharmacia. DNA sequence reading
and analyses was done using a range of computer programs, including DNA RODENT (Breen et al., 1988), the
Cornell package (Fristensky et al., 1982), Staden programs (Staden, 1982), MTX analysis package (Reisner &
Bucholtz, 1986) and the PHYLIP phylogeny inference package (J. Felsenstein, University of Washington, Seattle,
Wash., U.S.A.), modified by G. McKay, Macquarie University.

RESULTS

Identification and location of EHV-1 DNA sequences with homology to HSV-1 gB

A restriction map for BamHI sites in the EHV-1 genome and a more detailed map for the
region spanning 0-40 to 0-46 map units are shown in Fig. 1. In initial hybridization experiments
the plasmid pGX37 containing HSV-1 gB and DNA-binding protein sequences was found to
anneal strongly to the EHV-1 BamHI A and I fragments (not shown). The pGX37 subclone
containing a 1.74 kb SalI/BamHI fragment in M13 hybridized to the EHV-1 BamHI I fragment,
**Equine herpesvirus 1 gB**

385

Fig. 1. (a) Restriction map of the EHV-1 genome showing BamHI sites [from Whalley et al. (1981) except that the locations of fragments F and G have been corrected and interchanged]. (b) Expanded map of the BamHI I fragment and part of the adjacent BamHI A fragment (B, BamHI; C, ClaI; H, HindIII; P, PstI; S, SalI), along with positions of the gB and ICP 18.5 ORFs within the region sequenced. The common fragment hybridizing to an HSV gB-specific probe (deduced from Fig. 2) is shown by the cross-hatched region.

Fig. 2. (a) Agarose gel electrophoresis of cloned BamHI I DNA restricted with combinations of enzymes (B, BamHI; C, ClaI; H, HindIII; P, PstI S, SalI). (b) Southern blot hybridization of the gel transfer of (a) with 32P-labelled SalI/BamHI 1-74 kb fragment containing the sequence of HSV-1 gB.

within a common region of 1.5 kbp (SalI-HindIII) defined by hybridization against the double digestion combinations shown in Fig. 2. If the orientation of the genes coding for HSV-1 gB and the DNA-binding protein is the same in EHV-1, the hybridization data indicated that the gB gene homologue in EHV-1 coded in the 5' to 3' direction from the BamHI A fragment into the BamHI I fragment. The nucleotide sequence of 573 bp between HindIII and ClaI sites revealed homology between the first 200 nucleotides from the HindIII site of the EHV-1 sequence and
Fig. 3. Nucleotide sequence of the 4.3 kbp \textit{Phl}–\textit{Clal} DNA region (Fig. 1) and predicted amino acid sequences of the EHV-1 gb polypeptide and of part of the EHV-1 gene product analogous to HSV ICP 18S. Putative CAT, TATA and poly(A) signal elements are shown by underlining, a sequence with 7/9 match to putative HSV-1 mRNA start site is shown by a dotted line, and strongly hydrophobic stretches of amino acids (Kyte & Doolittle, 1982) are shown by boxes.
Equine herpesvirus 1 gB

regions of the last 200 nucleotides of the protein-coding (C-terminal) regions of the genes for HSV-1 gB (Bzik et al., 1984) and VZV gpII (gB) (Keller et al., 1986).

Analysis of the 4.3 kbp nucleotide sequence of the PstI/BamHI and BamHI/ClaI fragments revealed two overlapping open reading frames (ORFs) specifying the EHV-1 gB gene product and part of a gene upstream which appears to be the EHV-1 homologue of the HSV-1 ICP 18-5 gene (Pellett et al., 1986) and the VZV gene 30 (Davison & Scott, 1986). The complete 4.3 kbp nucleotide sequence between the PstI and ClaI sites and the corresponding amino acid sequence is shown in Fig. 3. The G + C content of the sequence is 53.4%, which is close to the estimates of 55 to 57% for total EHV-1 DNA as determined from ultracentrifugation data (Goodheart & Plummet, 1975). Putative regulatory signals include a TATA box which aligns on the sequence (relative to conserved sequences in the gB ORF) exactly with that predicted for HSV (Bzik et al., 1984; Pellett et al., 1986), a CAT sequence at approximately 80 nucleotides upstream of this, and, at 20 nucleotides downstream from the TATA element, a sequence with a 7/9 bp match to the predicted HSV mRNA start site. The ATG codon furthest upstream in the main ORF of the EHV-1 gB gene is at nucleotide 951 in the sequence which, with a termination codon at 3891, gives a total length for the primary translation product of 980 amino acids. The flanking sequence around this ATG codon has some of the features proposed by Kozak (1987) to enhance translation initiation, namely a purine (G) at −3, and C at −1 and −4, although it lacks a G at +4.

A second possible initiation site at Met 103 was considered less likely to be used because the succeeding residues, although somewhat hydrophobic in character, do not sum to the high overall level of hydrophobicity expected for a membrane-spanning segment. S1 nuclease mapping (C. W. Bell, unpublished data) locates a probable 5' end of the EHV-1 gB mRNA to approximately 30 nucleotides downstream of the predicted TATA sequence, which is also consistent with the proposed translation start codon.

A polyadenylation signal occurs within a few nucleotides from the termination codon of the EHV-1 gB ORF. In contrast, no obvious poly(A) signal was evident 3' to the EHV-1 'ICP 18-5' ORF suggesting that as in HSV-1 (Holland et al., 1984) the 3' terminus of the 'ICP 18-5' transcript is coterminial with that of the EHV-1 gB mRNA.

Features of the EHV-1 gB polypeptide sequence

Analysis of hydropathy (Fig. 4) shows most of the polypeptide to be relatively hydrophilic, although strongly hydrophobic domains occur near the N and C termini. A potential signal peptide cleavage site is after the Val-Arg-Ala sequence (residues 83 to 85, at +7 from the end of the hydrophobic core, Fig. 3), as it has several of the consensus features described by Von Heijne (1984) and Perlman & Halvorson (1983), i.e. cleavage after an alanine (−1) with a charged residue at −2 and valine at −3, in a ‘window’ of four to 10 residues downstream from the hydrophobic core. The features of the EHV-1 gB N terminus are also consistent with the criteria ascertained by McGeoch (1985) for membrane-translated proteins. The length of the cleaved (mature) EHV-1 polypeptide would then be 895 amino acids, giving a non-glycosylated Mr of 101K prior to further internal cleavage (see below). Other strongly hydrophobic sequences of amino acids near the C terminus (particularly from residues 853 to 868) are consistent with transmembrane anchor sequences characteristic of these glycoproteins; hydropathy plots (Fig. 4) highlight these regions and show the overall structural similarity between the gB proteins of EHV-1 and HSV-1. Chou & Fasman (1978) analysis of secondary structure potential for EHV-1 gB also shows that some α-helical domains occur in regions similar to those in the HSV-1 gB, and analyses of amino acid polarity show substantial stretches of non-polar or polar (neutral) residues associated with the predicted transmembrane domains and signal cleavage sequence (data not shown).

Comparison of EHV-1 gB with ‘gB’ polypeptides of five other herpesviruses

Alignment of the EHV-1 gB amino acid sequence with analogous sequences published for HSV-1, PRV, VZV, EBV and HCMV (Fig. 5) clearly identifies the EHV-1 gene as a member of the herpesvirus gB gene family, with regions of substantial sequence identity over all but the N-
and C-terminal regions. There are major blocks of highly conserved amino acids, particularly among the four alphaherpesviruses, with an alignment of 10 cysteine residues across the six viruses. Certain other amino acids such as phenylalanine are also at highly conserved positions, as are certain short 'motifs' such as Ser-Pro-Phe-Tyr at residues 324 to 327, Cys-Tyr-Ser-Arg-Pro at 668 to 672, and Glu-(Tyr)-Gly-Gln-Leu-Gly at 689 to 693. Six out of 12 potential N-linked glycosylation sites on the EHV-1 gB protein are conserved in at least three of the other viruses, indicating that these sites are likely to be utilized. A number of other potential glycosylation sites, e.g. at residue 621, may not be active due to the presence of a proline residue within or adjacent to the Asn-Xaa-Ser/Thr signal (Bause, 1983).

Fig. 4. Relative hydropathy of EHV-1 (a) and HSV-1 (b) gB primary polypeptides using the program of Novotny & Auffray (1984) modified to include the parameters of Kyte & Doolittle (1982) with a seven-point moving window and 20 smoothing reiterations. Points above the horizontal line represent regions of above average hydrophobicity and are candidate signal or membrane anchor sequences.

Fig. 5. Alignment of amino acids of EHV-1 gB and gB-like polypeptides of five other herpesviruses, obtained initially using the MTX program (Reisner & Bucholtz, 1986) as a basis for pairwise alignments, followed by adjustments to allow maximum alignment across the six herpesviruses. Due to their low levels of homology the N- and C-terminal regions were written by extension from the nearest region of clear alignment. Amino acids with identity to the corresponding EHV-1 gB amino acid are shown by shading in blocks, potential N-linked glycosylation sites (Asn-Xaa-Thr/Ser) by cross-hatching, and conserved cysteine residues by asterisks. The conserved 'core regions' used in Table 1 and Fig. 6 are bounded by arrows and numbers refer to EHV-1 gB residues at the ends of each line. Conserved N-glycosylation sites free of internal or adjacent proline residues are marked by O. The predicted internal proteolytic cleavage site is indicated by V, and the corresponding residues are underlined. The sources of amino acid sequences for the other gB polypeptides were: PRV, Robbins et al. (1987); VZV, Keller et al. (1986), Genbank, 20/5/87; HSV-1 (KöS), Bzik et al. (1984), Genbank 20/5/87; HCMV, Cranage et al. (1986); EBV, Pellett et al. (1985a). The number of amino acids in each sequence is recorded at the end of the sequence.
<table>
<thead>
<tr>
<th>Virus</th>
<th>Protein</th>
<th>Access Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>EHV-1</td>
<td>VZV</td>
<td>MG515617</td>
</tr>
<tr>
<td>EHV-1</td>
<td>PRV</td>
<td>MG515618</td>
</tr>
<tr>
<td>EHV-1</td>
<td>HSV-1</td>
<td>MG515619</td>
</tr>
<tr>
<td>EHV-1</td>
<td>HCMV</td>
<td>MG515620</td>
</tr>
<tr>
<td>EHV-1</td>
<td>EBV</td>
<td>MG515621</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Access Number</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MG515617</td>
<td>VZV gB</td>
</tr>
<tr>
<td>MG515618</td>
<td>PRV gB</td>
</tr>
<tr>
<td>MG515619</td>
<td>HSV-1 gB</td>
</tr>
<tr>
<td>MG515620</td>
<td>HCMV gB</td>
</tr>
<tr>
<td>MG515621</td>
<td>EBV gB</td>
</tr>
</tbody>
</table>

Equine herpesvirus 1 gB
Table 1. Percentage homologies of amino acid sequences of gB 'core' polypeptides* of seven herpesviruses

<table>
<thead>
<tr>
<th></th>
<th>PRV</th>
<th>VZV</th>
<th>HSV-1</th>
<th>HSV-2†</th>
<th>HCMV</th>
<th>EBV</th>
</tr>
</thead>
<tbody>
<tr>
<td>EHV-1</td>
<td>62-4</td>
<td>57-5</td>
<td>52-2</td>
<td>52-0</td>
<td>29-0</td>
<td>26-6</td>
</tr>
<tr>
<td>PRV</td>
<td>59-8</td>
<td>56-9</td>
<td>57-1</td>
<td>31-0</td>
<td>28-0</td>
<td></td>
</tr>
<tr>
<td>VZV</td>
<td>50-6</td>
<td>50-2</td>
<td>28-2</td>
<td>30-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSV-1</td>
<td>95-3</td>
<td>28-5</td>
<td>30-0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSV-2</td>
<td>28-5</td>
<td>30-0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCMV</td>
<td></td>
<td></td>
<td>30-6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Sequences aligned with EHV-1 residues 137 to 492 and 574 to 825.
† Sequence derived from Bzik et al. (1986). Sources of other sequences as described in legend to Fig. 5.

The comparison of amino acid sequences has also revealed a likely internal proteolytic cleavage site for the EHV-1 gB polypeptide. VZV gp11 is known to be processed to a pair of glycoproteins of approximately 60K, one of which terminates at an Arg-Arg with a serine residue at +1 (Keller et al., 1986). An identical cleavage site (Arg-Arg-Ser) is also present in HCMV (R. R. Spaete, personal communication) and has been predicted for PRV (Robbins et al., 1987). Inspection of the aligned sequences in Fig. 5 shows the same three amino acids in EHV-1 gB at residues 520 to 522 in a location virtually identical to those in PRV, VZV and HCMV, but not in HSV-1 (whose gB is not cleaved) or EBV (therefore not expected to be cleaved). The sequence Arg-Arg-Ser does not occur anywhere else in any of these polypeptides. The predicted (cleaved) EHV-1 gB polypeptides would have unglycosylated Mr values of 50K and 51K. Another feature highlighted in Fig. 5 is an additional stretch of amino acids (48 relative to HSV-1 and 25 relative to EBV) between residues 512 and 560 in the EHV-1 gB polypeptide.

A matrix of similarities of gB 'core' polypeptides (Table 1), based on amino acid sequence in conserved regions of the proteins (EHV-1 residues 137 to 492, 574 to 825), shows the EHV-1 gene product to be most closely related to that of PRV, followed by VZV, HSV, HCMV and EBV in that order. The dendrogram of Fig. 6 illustrates the relatedness of the gB 'core' polypeptides,
Equine herpesvirus 1 gB

showing as expected the alphaherpesvirus genes to be more closely related to each other than to HCMV or EBV, which in turn are not closely related to one another. This analysis does show a divergent evolution for the HSV gB and those of EHV-1, PRV and VZV.

The potential translation product from the EHV-1 overlapping upstream ORF shares similar levels of amino acid identity with the ICP 18.5 equivalents of HSV-1 (101 out of 201 amino acids available for alignment), VZV (106/201) and PRV (65/142).

DISCUSSION

We have described the identification, location and sequence determination of a gene in EHV-1 analogous to gB-like genes in HSV and several other herpesviruses. The map location of the EHV-1 gB gene is consistent with the general collinearity of herpesvirus genomes (Davison & Wilkie, 1983). In addition, identification of several EHV-1 homologues of HSV and VZV genes, including thymidine kinase, major capsid protein and glycoprotein H, has confirmed their collinearity within a 25 kbp region at the level of individual genes (G. R. Robertson & J. M. Whalley, unpublished results). Regions of collinearity containing some of these genes have also been described for the lymphotropic herpesviruses EBV and herpesvirus saimiri (Gompels et al., 1988; Baer et al., 1984). The position of the EHV-1 gB gene, spanning the BamHI A and I fragments, indicates that this gene is the same as that coding for the EHV-1 gpl4 glycoprotein mapped by Allen & Yeargan (1987).

Despite a wide variation in base composition among the genomes of the different herpesviruses, which also occurs in their gB genes, alignment of amino acid sequences shows a pattern of conserved regions or blocks which are therefore likely to have particular functional significance. Even within divergent sequences near N and C termini, all six viruses compared have strongly hydrophobic domains in approximately the same locations relative to the conserved regions. This is consistent with their essential roles in transporting the nascent polypeptide across the endoplasmic reticulum and/or anchoring the mature protein to cell or virus membranes. Claesson-Welsh & Spear (1987) have identified the signal cleavage sequence of HSV-1 gB by comparing N-terminal amino acids determined from protein sequencing with those predicted from nucleotide sequence data. Some evidence supporting a model of three membrane-spanning segments near the C terminus (Pellett et al., 1985b) was also provided by the nature of fragments produced by trypsin cleavage. From the similarity of hydropathy profiles (Fig. 4) the EHV-1 gB would be expected to function in a similar manner to that of HSV-1, although the levels of peak hydrophobicities are such in both cases as to suggest another possibility in which a single non-polar sequence with the potential to form a helix (approximately residues 853 to 868) anchors the protein in the membrane (see Engelman et al., 1986). In the case of HSV-1 gB, truncation of 194 C-terminal amino acids and resulting secretion of the product from transfected cells showed directly that this region contained sequences essential for this anchoring function (Pachl et al., 1987).

The highly conserved location of cysteine residues highlights the role of disulphide bridges in the secondary and tertiary structure of the protein, whether the primary polypeptide chain is cleaved or not. Our prediction of the HSV-1 gB internal cleavage site is consistent with the finding that the mature protein is cleaved into two species of 76K and 58K (D. M. Meredith, J.-M. Stocks, G. R. Whittaker, R. A. Killington & I. W. Halliburton, personal communication). The difference between the apparent Mr of these two products compared to the virtually identical sizes for the polypeptide entities (50K and 51K) is likely to be related to the differing numbers of potential N-glycosylation sites in each component (Fig. 5). The features of the EHV-1 gB gene suggest that assembly and processing of the mature glycosylated protein could follow steps similar to those described for gpII of VZV (Montalvo & Grose, 1987).

The EHV-1 gB gene has a segment of DNA coding for an additional 48 amino acids compared with HSV-1, which includes the predicted internal proteolytic cleavage site. Codon usage analysis (Staden, 1984) of the EHV-1 gB gene (data not shown) indicates that this stretch of DNA may have undergone a frameshift or may have a different evolutionary origin from the rest of the gene. Although the functional significance of this region is as yet unknown, an almost identical sequence also occurs in the gB gene of the closely related EHV-4 (EHV-1, respiratory
subtype), which has 85 to 90% overall amino acid identity with the EHV-1 sequence described here (M. Riggio, A. Cullinan & D. Onions, personal communication).

The EHV-1 gB gene region is also distinguished from that of HSV by the apparent existence of two overlapping ORFs, in which a gene equivalent to a possible glycoprotein transport protein of HSV-1 (ICP 18.5 (Pellett et al., 1986)) overlaps the translation start site of EHV-1 gB by 135 nucleotides. The amino acids in the EHV-1 and EHV-4 ICP 18-5 gene analogues are conserved to a level of approximately 85%, while those in the other ORF (gB) show only 50% identity, consistent with the less stringent sequence requirements of a signal peptide. The occurrence of these overlapping ORFs in association with unusually long signal sequences is also a conserved evolutionary feature between EHV-1 and PRV (Robbins et al., 1987). Inspection of the published VZV sequence (Davison & Scott, 1986; Keller et al., 1986) in the same region also reveals a potential overlap, with an ATG at 63 codons upstream from the published start site for gpII, at a virtually identical position to that predicted here for EHV-1. Thus among the alphaherpesviruses for which these sequences are currently available, the organization in HSV appears to be the exception rather than the rule. The HSV ICP 18-5 C-terminal coding region does in fact contain promoter and upstream regulatory regions of the HSV gB gene. The somewhat divergent evolution of the HSV gB is supported by the phylogenetic analysis of 'core' amino acid sequences. It is of interest that the EHV-1 gene is slightly more closely related to that of VZV (an α2 herpesvirus) than to HSV-1 (an α1 herpesvirus).

The identification and sequence analysis of the EHV-1 gB gene forms the basis for assessment and development of the EHV-1 gB protein as a potential vaccine antigen through, for example, the use of vaccinia virus vectors or synthetic peptides. In addition, sequence data on conserved genes such as those of the gB family described here are of particular value in determining evolutionary relationships among the herpesviruses.

The authors thank Dr A. J. Gibbs for assistance with phylogeny analysis, Professor J. H. Subak-Sharpe for helpful discussion, and Michelle Thorpe for typing the manuscript. This research was supported in part by grants from the Australian Research Grants Scheme and the Australian Equine Research Foundation. C.W.B. and G.R.R. are holders of Commonwealth Post-Graduate Research Awards.

REFERENCES


Equine herpesvirus 1 gB


(Received 20 June 1988)