Equine herpesvirus type 1 unique short fragment encodes glycoproteins with homology to herpes simplex virus type 1 gD, gI and gE

Jean-Christophe Audonnet,¹ Joseph Winslow,¹ George Allen² and Enzo Paoletti*¹

¹Virogenetics Corporation, Rensselaer Technology Park, 465 Jordan Road, Troy, New York 12180-8349 and
²Department of Veterinary Science, University of Kentucky, Lexington, Kentucky 40546-0076, U.S.A.

The nucleotide sequence of a 6.4 kbp portion of the 10-6 kbp BamHI fragment D contained in the unique short region of the equine herpesvirus type 1 (EHV-1) genome has been determined. Analysis of this sequence revealed five open reading frames (ORFs), four complete and one incomplete, which were encoded by the same sense strand. Comparison of the EHV-1 DNA sequence with that encoding glycoproteins of other alphaherpesviruses has revealed no significant homologies. Comparison at the amino acid level, however, has demonstrated regions of significant sequence similarity between the three complete EHV-1 ORFs 2, 3 and 4, and the herpes simplex virus type 1 (HSV-1) glycoprotein gD encoded by the US6 gene, the HSV-1 glycoprotein gI encoded by the US7 gene and the HSV-1 glycoprotein gE encoded by the US8 gene, respectively. The interrupted ORF 5 was found to display partial homology with the HSV-1 US9-encoded protein, but no homology was found between the protein encoded by ORF 1 and other proteins. The three collinear EHV-1 ORFs encoding putative glycoproteins with homology to the HSV-1 glycoproteins were therefore designated EHV-1 gD, gI and gE, respectively. Moreover, further similarities were found between EHV-1 gD and pseudorabies virus (PRV) gp50, between EHV-1 gI and PRV gp63 and varicella-zoster virus (VZV) gpIV, and between EHV-1 gE and PRV gI and VZV gpI. It is concluded that EHV-1, PRV, HSV-1 and VZV encode homologous glycoprotein genes in the small unique components of their genomes and that the genetic organization of these regions is conserved.

Introduction

Alphaherpesviruses equine herpesvirus types 1 and 4 (EHV-1 and EHV-4) are significant viral pathogens of horses (O'Callaghan et al., 1983). Efforts to control economic losses resulting from EHV-1 (respiratory diseases, abortion and paralysis) and EHV-4 (respiratory diseases) infections in horses are based on antigen-induced stimulation of antiviral immune responses of the horse with either inactivated or live vaccines (Bryans, 1978; Bryans & Allen, 1986; Purdy et al., 1978). EHV-1 and EHV-4, although cross-neutralizing, can be distinguished by their antigenicity profiles, restriction endonuclease fingerprints and pathogenic properties in horses (Allen & Bryans, 1986). Both express six major glycoproteins designated gp2, gp10, gp13, gp14, gp17/18 and gp21/22 with Mᵦ values of 200K, 125K, 95K, 90K, 68K and 45K, respectively, and at least five minor glycoproteins (Turtinen & Allen, 1982). The genomic portions of the DNA sequence encoding the six major EHV-1 glycoproteins have been identified by using λgt 11 expression libraries and monoclonal antibodies (MAbs) specific to the individual glycoproteins (Allen & Yeargan, 1987). EHV-1 appears unique among the alphaherpesviruses whose glycoprotein genes have been mapped in that five of its six major glycoproteins are encoded from sequences within the UL region, whereas only one (gp17/18) has been mapped to the US region. Analysing these data, Allen & Yeargan (1987) predicted that some of the low abundance glycoproteins identified in EHV-1 virions as well as EHV-1 glycoproteins not yet identified may map to the S component of the genome. Thus far, the only known nucleotide sequences for EHV-1 glycoproteins are those of gp13 and gp14 (Allen & Coogle, 1988; Whalley et al., 1989; Guo et al., 1989, 1990). Analysis of the predicted amino acid sequences of gp13 and gp14 glycoproteins revealed significant homology to the herpes simplex virus type 1 (HSV-1) gC and gB glycoproteins, respectively. Complete and partial nucleotide sequences have recently been described respectively for the EHV-4 gB glycoprotein (Riggio et al., 1989) and gE glycoprotein (Cullinane et al., 1988) genes.

The envelope glycoproteins are the principal immunogens of herpesviruses involved in eliciting both humoral and cellular host immune responses (Ben Porat et al., 1986; Cantin et al., 1987; Glorioso et al., 1984;
Wachsmann et al., 1989) and therefore provide the immediate targets for the design of efficacious vaccines. MAbs directed against certain EHV-1 glycoproteins have been shown to be neutralizing (Sinclair et al., 1989) and passive immunization experiments have demonstrated that MAbs directed against gp13, gp14 and gp17/18 (Shimizu et al., 1989; Stokes et al., 1989) protect hamsters against a lethal EHV-1 challenge. The genes encoding gp13 and gp14 have been expressed separately or in association in recombinant vaccinia viruses and these recombinant viruses have elicited protective immunity in hamsters (Guo et al., 1989, 1990). Such results have clearly demonstrated the role of the gp13- and gp14-specific immune responses in protection against EHV-1 infection.

Since other EHV-1-encoded glycoproteins may be relevant to the induction of protective immunity in horses, it is important to identify the genes which express such products. In this communication, we report the nucleotide sequence of 6.4 kbp of the Us region of the EHV-1 genome and the identification and analysis of EHV-1 open reading frames (ORFs), four complete and one incomplete. Three of these ORFs display significant homology with HSV-1 Us-encoded glycoproteins.

**Methods**

**Preparation of EHV-1 DNA.** EHV-1 genomic DNA (Kentucky D strain) was obtained from Rhône-Merieux. The DNA was extracted from EHV-1 virions as described previously (Henry et al., 1981).

**Cloning of the EHV-1 BamHI D fragment.** EHV-1 genomic DNA (Kentucky D strain) was digested with BamHI according to the manufacturer's specifications and the 10-6 kbp BamHI D fragment, representing most of the Us region of the EHV-1 genome and the identification and analysis of EHV-1 open reading frames (ORFs), four complete and one incomplete. Three of these ORFs display significant homology with HSV-1 Us-encoded glycoproteins.

**Identification of DNA sequences encoding EHV-1 gD, gI and gE.** The complete nucleotide sequence for both strands of a 6400 bp portion of BamHI D was obtained from several subclones of the BamHI D fragment inserted into pHBl24, using the modified T7 enzyme Sequenase (U.S. Biochemical Corporation) (Tabor & Richardson, 1987). Standard dideoxynucleotide chain termination reactions (Sanger et al., 1977) were performed on double-stranded templates that had been denatured in 0.4 M NaOH (Hattori & Sakaki, 1986). The M13 forward and reverse primers were used to obtain the initial sequence of each clone. Custom primers (18-mer), synthesized by using standard chemistries (Biosearch 8700 and Applied Biosystems 380B), were used for subsequent sequence reactions. Sequences of the junctions between consecutive fragments were confirmed on the initial clone, pHV BamHID. The PC/GENE (Intelligenetics) and IBI Pustell software packages were used in all sequence data analyses. Homology searches were done with the FASTP program (Lipman & Pearson, 1985) against Swissprot release 11.0 (Intelligenetics) and with the FASTN program (Lipman & Pearson, 1985) against GenBank release 61.0 (Intelligenetics).

**Results**

**Restriction map of the EHV-1 BamHI D fragment**

A restriction map of the BamHI D fragment was constructed using the restriction enzymes AatII, ClaI, EcoRI, EcoRV, HindIII, PstI, PvuII, SacII, SalI and Smal. It is shown in Fig. 1.

**Nucleotide sequence analysis of the 6400 bp fragment**

Sequence data were obtained for a 6400 bp region of the BamHI D fragment. The sequence is presented in Fig. 2 as the rightward 5' to 3' strand. The base composition is 50.44% G + C. Analysis of the nucleotide sequence in all

**Table 1. Summary of major ORFs in the 6.4 kbp sequenced fragment of the BamHI D fragment of EHV-1 Kentucky D strain virus**

<table>
<thead>
<tr>
<th>ORF</th>
<th>Start</th>
<th>Stop</th>
<th>Amino acids</th>
<th>Mr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>193</td>
<td>786</td>
<td>198</td>
<td>2163K</td>
</tr>
<tr>
<td>2</td>
<td>971</td>
<td>2176</td>
<td>402</td>
<td>4524K</td>
</tr>
<tr>
<td>3</td>
<td>2287</td>
<td>3558</td>
<td>424</td>
<td>4639K</td>
</tr>
<tr>
<td>4</td>
<td>3794</td>
<td>5449</td>
<td>552</td>
<td>6149K</td>
</tr>
<tr>
<td>5</td>
<td>6172</td>
<td>*</td>
<td>76</td>
<td>ND†</td>
</tr>
</tbody>
</table>

* Designates incomplete ORF at right end of sequence. Precise location of translation termination is unknown.
† ND, Not determined.
six reading frames revealed five potential major ORFs encoded by the positive strand and four minor ORFs encoded by the negative strand (Fig. 1). Two other minor ORFs of greater than 70 amino acids do occur within the primary and secondary structure did not reveal any significant homology with other viral DNA sequences. However, comparison of EHV-1 DNA segment 4685 to 6400 showed high levels of homology with the EHV-4 DNA segment 1 to 1669 (Cullinane et al., 1988). There is 82-2% identity between the DNA segments encoding the carboxy termini of EHV-1 ORF 4 and EHV-4 ORF 1, 64-3% identity between the EHV-1 ORF 4-ORF 5 and EHV-4 ORF 1-ORF 2 intergenic segments and a lower homology (55.5±%) between the DNA segments encoding EHV-1 ORF 5 and EHV-4 ORF 2.

Analysis of the major ORFs

The first complete ORF (ORF 1) extends from nucleotide positions 193 to 786. The sequence context surrounding the proposed initiation codon (TATAATGG, nucleotides 190 to 196) qualifies as a functional sequence context for translation initiation of eukaryotic mRNA (Kozak, 1983, 1986). There is a second possible initiation codon located at positions 301 to 303; however the sequence context of this ATG (CTGATGC) is not favourable with respect to Kozak’s rules. A TATA box-like element having the sequence AATAGAGA is located at positions 117 to 124. No evident CAAT box was found upstream of this TATA box. A typical polyadenylation signal (AATAAA) (Proudfoot & Brownlee, 1976) was found at positions 959 to 964, 75 nucleotides downstream of the TAG termination codon. The EHV-1 ORF 1 encodes a 198 amino acid long polypeptide with a calculated Mr of 21.6K. Analysis of the amino acid sequence of the putative EHV-1 ORF 2-encoded glycoprotein with proteins in the Swissprot database revealed a number of features common to membrane-associated glycoproteins. A transmembrane segment from residues 356 to 372 was predicted as a membrane anchor element. Comparison of the amino acid sequence revealed significant homology with HSV-1 gD (McGeoch et al., 1985; Watson et al., 1982) and pseudorabies virus (PRV) gp50 (Petrovskis et al., 1986a) glycoproteins (Table 2).

The second ORF (ORF 2) extends from nucleotide positions 821 to 2176. Five possible initiation codons are located in the first 200 bp of this ORF at positions 821, 971, 989, 992 and 1016. However only initiation codons at positions 971, 992 and 1016 are favourable with respect to Kozak’s rules (Kozak, 1983, 1986). Putative transcriptional regulatory signals are situated in the region 5' to the probable initiation codon at position 971 (see Discussion). A TATA box-like element having the sequence TATATTTA (nucleotides 871 to 878) was located 60 nucleotides downstream from a putative CAAT box element (TGACACAT) at position 811 to 817. No polyadenylation signal (AATAAA) was found downstream of the TAA termination codon (nucleotides 2177 to 2179). The EHV-1 ORF 2 (starting at position 971) encodes a 402 amino acid long polypeptide with a calculated Mr of 45.2K. Analysis of the amino acid sequence revealed a number of features common to membrane-associated glycoproteins. A transmembrane segment from residues 356 to 372 was predicted as a membrane anchor element. Comparison of the amino acid sequence of the putative EHV-1 ORF 2-encoded glycoprotein with proteins in the Swissprot database revealed significant homology with HSV-1 gD (McGeoch et al., 1985; Watson et al., 1982) and pseudorabies virus (PRV) gp50 (Petrovskis et al., 1986a) glycoproteins (Table 2).

The third ORF (ORF 3) extends from nucleotide positions 2287 to 3558. The sequence context surrounding the proposed initiation codon (GCTATGG) is in a favourable context with respect to Kozak’s rules. There are at least two other possible ATG initiation codons at positions 2305 and 2332. However the sequence context of these codons (GGGATGT and TCTATGG) is not as favourable. No putative transcriptional regulatory signals were found in the region 5' to the ATG initiation codon. Two polyadenylation signals (ATTAAA) were

<table>
<thead>
<tr>
<th>EHV-1 ORF</th>
<th>Homologue protein</th>
<th>Alignment score (initial/optimized)* (%)</th>
<th>Homology (%)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HSV-1 gD</td>
<td>114/244 (21-6/268)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>HSV-1 gI</td>
<td>210/247 (25-1/227)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>HSV-1 gE</td>
<td>116/273 (23-9/226)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>HSV-1 gI</td>
<td>169/412 (26-3/437)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>HSV-1 US9</td>
<td>139/463 (26-5/400)</td>
<td></td>
</tr>
</tbody>
</table>

* Sequences were aligned using the FASTP algorithm of Lipman & Pearson (1985) using the PAM250 weighting matrix (Schwartz & Dayhoff, 1978). Scores are presented before (initial) and following (optimized) gap insertion for alignment optimization.
† Percentage homology is expressed as % identity/no. of amino acid overlap.
§ The homology search was made against the EHV-4 gE carboxy-terminal 255 amino acids since only this part of the protein is known (Cullinane et al., 1988).
ND, Not determined (interrupted ORFs).
Fig. 2. Nucleotide sequence of the 600 bp region contained in the 10.6 kbp EHV-1 Us region. Presumptive TATA and CAT boxes, putative signal elements and putative transmembrane elements are underlined. The potential cleavage sites of the signal peptides are noted with an arrow (T). Linked glycosylation sites are defined as Asn-X-Ser/Thr ( ). The sequence data were obtained as described in Methods. Amino acids are numbered at the right of the sequence. Glycoproteins of EHV-1 Us region
found at positions 3585 to 3590 and 3617 to 3622, respectively at 30 and 60 nucleotides downstream of the TAG termination codon. The EHV-1 ORF 3 encodes a polypeptide of 424 amino acids with a calculated Mr of 46-4K. Analysis of the amino acid sequence revealed a number of features common to membrane-associated glycoproteins. A transmembrane segment from residues 320 to 336 was predicted as a membrane anchor element. Comparison of the amino acid composition of the EHV-1 ORF 3 revealed significant homology with the HSV-1 gI (McGeoch et al., 1985; Longnecker et al., 1987), varicella-zoster virus (VZV) gpIV (Davison, 1983) and PRV gp63 (Petrovskis et al., 1986b) glycoproteins (Table 2).

The fourth ORF (ORF 4) extends from nucleotide positions 3794 to 5449. The sequence context surrounding the proposed initiation codon (ACAATGG) is favourable with respect to Kozak’s rules (Kozak, 1983, 1986). Putative transcriptional regulatory signals were found in the region 5' to the ATG initiation codon at position 3794. A TATA box-like element having the sequence GTTTAAA (nucleotides 3703 to 3709) was located 50 nucleotides downstream of a putative CAT box-like element (GCAATG) at positions 3647 to 3652. No evident polyadenylation signal was found immediately downstream of the TGA termination codon (nucleotides 5450 to 5452). A typical polyadenylation signal, AATAAA, was found at positions 5832 to 5837, 390 nucleotides downstream of the TGA codon. The EHV-1 ORF 4 encodes a polypeptide of 552 amino acids with a calculated Mr of 61-5K. Analysis of the amino acid sequence revealed a number of features common to membrane-associated glycoproteins. A transmembrane segment from residue 411 to 427 was predicted as a membrane anchor element. Comparison of the amino acid composition of the EHV-1 ORF 4 protein revealed significant homology with the HSV-1 gE (McGeoch et al., 1985), VZV gpI (Davison, 1983) and PRV gI (Petrovskis et al., 1986b) glycoproteins (Table 2). Comparison with the partial amino acid sequence published for EHV-4 ORF 1 (gE) (Cullinane et al., 1988) revealed 84-7% identity between both sequences (Table 2).

The fifth ORF (ORF 5) starts at position 6172 and extends beyond the BamHI site at position 6396. The sequence context of the proposed initiation codon (TTCATGG) is favourable with respect to Kozak’s rules. A TATA box-like element having the sequence GATAAACT was found at positions 6026 to 6033. No evident CAAT box was found upstream of this TATA box. The 76 amino acid sequence encoded by this interrupted ORF was used for a homology search against the Swissprot database. No significant homology was found between ORF 5 and the database. However, this truncated polypeptide was found to be partially homologous to proteins encoded by HSV-1 US9 (19-7% identity) (McGeoch et al., 1985) and by EHV-4 ORF 2 (25% identity) (Cullinane et al., 1988) (Table 2).

The amino acid sequences of the four small ORFs (ORFs 6, 7, 8 and 9) encoded by the negative strand were analysed for homology to proteins contained in the Swissprot database, but no significant homology was found. These ORFs will not be discussed further.

Discussion

In this report, we describe the nucleotide sequence of a 6-4 kbp segment of the 10-6 kbp Us component of the EHV-1 genome. Analysis of the DNA sequence has revealed five major ORFs, all orientated in the same direction. Three of these appear to encode membrane-associated glycoproteins based upon biochemical characteristics deduced from the amino acid sequence and their apparent homology with glycoproteins of other alphaherpesviruses. We propose to designate the three putative EHV-1 glycoproteins gD, gI and gE, on the basis of their respective homologies with the HSV-1 gD, gI and gE glycoproteins.

EHV-1 ORF 1 encodes a 198 amino acid polypeptide which displays no particular features. The genomic map position of this ORF relative to the genomic organization seen in other alphaherpesviruses suggested that ORF 1 could possibly encode an HSV-1 US5 protein homologue. However no such homology was found when the ORF 1 amino acid sequence was compared to the HSV-1 US5 protein. Furthermore, no significant homology was revealed when ORF 1 was compared to proteins contained in the Swissprot database.

The EHV-1 ORF 2 (gD homologue) encodes a 402 amino acid polypeptide. This ORF has several features characteristic of a membrane-associated glycoprotein and contains four potential N-linked glycosylation sites. The amino acid sequence encoded by this ORF demonstrated a significant homology with PRV gp50 and with HSV-1 gD. The PRV gp50 and HSV-1 gD sequences both align with the ORF 2 sequence starting approximately at ATG 971, 989 or 992 (Fig. 3). No homology was revealed between the amino acid sequence stretch starting with ATG 821 through to ATG 971. Although initiation of ORF 2 at ATG 821 cannot be ruled out, Kozak’s rules and protein alignment data suggest that initiation of translation occurs at the ATG at position 971. The amino acid homology found between EHV-1 ORF 2 and gD homologues in other herpesviruses is not displayed throughout the entire ORFs but confined to the central regions (Fig. 4). These data may hint at the significance of these amino acid sequences...
with respect to the structure/function of these glycoproteins. Furthermore, as previously observed between PRV gp50 and HSV-1 gD (Petrovskis et al., 1986a), all but one cysteine in HSV-1 gD and EHV-1 ORF 2 (gD homologue) align. The sequence following the C terminus coding of the EHV-1 gD gene does not include the AATAAA or ATTAAAA polyadenylation signal sequences (Proudfoot & Brownlee, 1976). There are only 107 bp between the termination codon of EHV-1 ORF 2 (gD homologue) and the initiation codon of the next downstream gene, EHV-1 ORF 3 (gI homologue). It is notable that neither the PRV gp50 nor the HSV-1 gD gene has a polyadenylation signal immediately downstream (Petrovskis et al., 1986a; McGeoch et al., 1985). The HSV-1 gD transcript includes the next downstream gene (US7) (McGeoch et al., 1985). Surprising data from the transcriptional mapping of the Us region of the PRV genome suggested that gp50 and gp63 were translated from a single coterminial transcript (Kost et al., 1989). Based on the lack of evident transcriptional regulatory signals upstream to EHV-1 ORF 3 and the short ORF 2–ORF 3 intergenic region, it is possible that the transcriptional organization for EHV-1 gD and gI genes is similar to the proposed PRV gp50 and gp63 genes' transcriptional organization, but transcriptional data from the EHV-1 Us region are needed to confirm such a model, which is quite different from the transcriptional organization of the HSV-1 gD and gI genes described by McGeoch et al. (1985).

The EHV-1 ORF 3 (gI homologue) encodes a 424 amino acid polypeptide. This ORF has several features characteristic of a membrane-spanning glycoprotein with seven potential N-linked glycosylation sites. The amino acid sequence encoded by this ORF displays significant homology with HSV-1 gI, VZV gp4V and PRV gp63. It is notable that the homology exists in the amino-terminal portion of the ORFs. Homology restricted to the amino-terminal portion is also found between PRV gp63 and VZV gp4V and between HSV-2 gI and VZV gp4V (Davison & Scott, 1986; McGeoch et al., 1985; Petrovskis et al., 1986b). The conserved regions of PRV gp63, HSV-1 gI, VZV gp4V and EHV-1 gI shown in Fig. 5 may indicate, as for the gD-like glycoproteins, amino acid sequences important for the biological function of these glycoproteins. Since no classical transcriptional regulatory signals were observed upstream of the initiation codon of EHV-1 ORF 3, this ORF could be translated from an mRNA initiated upstream of ORF 2 and terminated downstream of ORF 3. This is similar to the transcriptional organization which has been described for the PRV gp50 and gp63
Fig. 5. Comparison of PRV gp63, HSV-1 gI, VZV gpI and EHV-1 ORF 3 (gI homologue) displaying the region of highest homology between their amino acid sequences. The HSV-1 gI amino acid sequence is from McGeoch et al. (1985), the PRV gp63 amino acid sequence from Petrovskis et al. (1986b) and the VZV gpI amino acid sequence from Davison & Scott (1986). The numbers at the left of the sequences indicate the positions in the amino acid sequences, with the initiation codon being 1 in all cases. The identity of an amino acid in two sequences is indicated by a vertical line. The cysteine residues are boxed. The sites for N-linked glycosylation are underlined. The symbol * indicates conserved amino acid in three out of four sequences.

**Fig. 6.** Comparison of PRV gI, HSV-1 gE, VZV gpI and EHV-10RF 4 (gE homologue) displaying the region of highest homology between their amino acid sequences. The HSV-1 gE amino acid sequence is from McGeoch et al. (1985), the PRV gE amino acid sequence is from Petrovskis et al. (1986b) and the VZV amino acid sequence from Davison & Scott (1986). The numbers at the left of the sequences indicate the positions in the amino acid sequences, with the initiation codon being 1 in all cases. The identity of an amino acid in two sequences is indicated by a vertical line. The cysteine residues are boxed. The sites for N-linked glycosylation are underlined. The symbol * indicates conserved amino acid in three out of four sequences.
possible that the same transcriptional organization exists for EHV-1 gE mRNA.

The EHV-1 ORF 5 was only analysed for homology with the 76 amino-terminal amino acids, since it is interrupted by the BamHI cloning site. This product, although incomplete, displays some homology to the HSV-1 US9-encoded protein. Interestingly, relative to EHV-1 ORFs 2, 3 and 4, the EHV-1 ORF 5 would have the same genomic location as the HSV-1 US9 relative to HSV-1 gD (US6), gI (US7) and gE (US8).

In summary, the organization of the five ORFs described in the U5 region of the EHV-1 genome follows the general blueprint previously described for HSV-1 (McGeoch et al., 1985), VZV (Davison, 1983) and PRV (Petrovskis et al., 1986a, b). It is possible that the other alpha herpesviruses have a similar genomic organization and therefore it could be predicted that genes encoding glycoprotein analogues of HSV-1 gD, gI and gE exist in the U5 region of feline herpesvirus type 1 and Marek’s disease virus.

We thank H. Martinez and G. Johnson for their contributions to various aspects of this study and J. Tartaglia and M. Riviere for critical reading of the manuscript.

References


(Received 18 May 1990; Accepted 15 August 1990)