JOURNAL OF VIROLOGY, Aug. 1988, p. 2850–2858
0022-538X/88/082850-09$02.00/0
Copyright © 1988, American Society for Microbiology
Vol. 62, No. 8

Characterization of an Equine Herpesvirus Type 1 Gene Encoding a Glycoprotein (gp13) with Homology to Herpes Simplex Virus Glycoprotein C†

GEORGE P. ALLEN* AND LINDA D. COOGLE

Department of Veterinary Science, University of Kentucky, Lexington, Kentucky 40546-0099

Received 18 February 1988/Accepted 2 May 1988

The molecular structure of the equine herpesvirus type 1 (EHV-1) gene encoding glycoprotein 13 (gp13) was analyzed. The gene is contained within a 1.8-kilobase *Acc*I-*Eco*RI restriction fragment mapping at map coordinates 0.136 to 0.148 in the U_L region of the EHV-1 genome and is transcribed from right to left. Determination of the nucleotide sequence of the DNA fragment revealed a complete transcriptional unit composed of typical regulatory promoter elements upstream to a long open reading frame (1,404 base pairs) that encoded a 468-amino-acid primary translation product of 51 kilodaltons. The predicted protein has the characteristic features of a membrane-spanning protein: an N-terminal signal sequence, a hydrophobic membrane anchor region, a charged C-terminal cytoplasmic tail, and an exterior domain with nine potential N-glycosylation sites. The EHV-1 DNA sequences expressed in λgt11 as gp13 epitopes were present in the open reading frame. Amino acid sequences composing a major antigenic site, recognized by 35% of a panel of 42 anti-gp13 monoclonal antibodies, were identified in the N-terminal surface domain of the deduced gp13 molecule. Comparison of the EHV-1 gp13 DNA sequence with that encoding glycoproteins of other alphaherpesviruses revealed no detectable homology. However, a search for homology at the amino acid level showed regions of significant sequence similarity between the amino acids of the carboxy half of EHV-1 gp13 and those of the same region of gC-like glycoproteins of herpes simplex virus (gC-1 and gC-2), pseudorabies herpesvirus (gIII), and varicella-zoster virus (gp66). The sequences of the N-terminal portion of gp13, by contrast, were much less conserved. The results of these studies indicate that EHV-1 gp13 is the structural homolog of herpes simplex virus glycoprotein C and further suggest that the epitope-containing N-terminal amino acid sequences of the herpesvirus gC-like glycoproteins have undergone more extensive evolutionary divergence than the C-terminal sequences.

Equine herpesvirus type 1 (EHV-1) is a ubiquitous herpesvirus of horses associated with epidemics of abortion, respiratory tract disease, and central nervous system disorders (2). Two cross-neutralizing, molecular subtypes of EHV-1 exist which differ in antigenicity, restriction endonuclease cleavage profile, and pathogenicity for horses (for a review, see reference 2). The genome of each of the two EHV-1 subtypes consists of a linear, double-stranded DNA of approximately 96×10^6 daltons, with an estimated coding capacity of 80 to 100 genes (62; A. Cullinane and A. Davison, Equine Infect. Dis., in press). Of interest for this report are the genes encoding the surface glycoproteins of the virion envelope. Herpesvirus glycoproteins are important in interactions of the virus with the host immune system as well as in mediation of essential viral functions such as cellular attachment, entry, and pathogenicity (50, 51).

Both subtypes of EHV-1 specify six high-abundance glycoproteins designated gp2, gp10, gp13, gp14, gp17/18, and gp21/22a (2, 4, 55). While considerable progress has been made in the characterization of glycoproteins of a variety of other alphaherpesviruses (1, 7, 8, 15, 16, 18, 19, 25, 26, 30, 37, 39, 40, 41, 43, 54, 57, 58, 61), little is known about the molecular or antigenic structure of EHV-1 glycoproteins. We recently determined the genetic loci of the DNA sequences encoding the six major glycoproteins of EHV-1 (4). Three of these EHV-1 glycoproteins (gp14, gp13, and gp17/18) mapped to positions colinear with those for herpes simplex virus (HSV) glycoproteins B, C, and E, respectively. EHV-1 gp13, whose coding sequences were localized to the U_L region of the viral genome between 0.11 and 0.15 map units, was selected as the glycoprotein upon which to focus our current research. By using a panel of 42 monoclonal antibodies, many of the antigenic characteristics of gp13 have been determined and reported (G. Allen, M. Yeargan, and L. Coogle, Equine Infect. Dis., in press). The antigenic epitopes of EHV-1 gp13 are predominantly subtype specific, exhibit much interisolate variability, and may elicit neutralizing antibody.

In this article, we report the further characterization of EHV-1 gp13, including the determination of the complete nucleotide sequence of the coding gene for this major EHV-1 glycoprotein, the localization on the protein of an immunodominant antigenic site, and the demonstration of significant amino acid homology with gC-like glycoproteins of several other alphaherpesviruses.

MATERIALS AND METHODS

Virus, cells, and viral DNA isolation. The Kentucky T431, subtype 1 strain of EHV-1 (5) was propagated in equine dermal cells as previously described (56). Extracellular virus was concentrated from EHV-1-infected cell cultures and purified by isopycnic banding in potassium tartrate gradients (56). EHV-1 DNA was isolated from purified virions as described previously (3).

Construction of λgt11 library expressing EHV-1 gp13 determinants. The *Eco*RI site in the β-galactosidase gene of

* Corresponding author.

† Paper no. 88-4-41 of the Kentucky Agricultural Experiment Station.

λgt11 was used to construct an expression library of the sequences contained within the *Bam*HI H fragment of EHV-1 DNA. The *Bam*HI H fragment encompasses the EHV-1 DNA sequences that encode gp13 (4). Random DNA fragments of 200 to 400 base pairs (bp) in length were generated by sonication of plasmid pIBI24 (International Biotechnologies, Inc. [IBI], New Haven, Conn.) DNA containing the EHV-1 *Bam*HI H fragment as a cloned insert. The resulting fragments were fitted with *Eco*RI linkers and then inserted into the lambda phage vector λgt11 for expression in *Escherichia coli*, as detailed by Huynh et al. (23). The expression library was amplified, and recombinant phage expressing gp13 isotopes were identified with monoclonal antibodies as reported previously (4, 63, 64). The preparation and characterization of the panel of 42 anti-gp13 monoclonal antibodies has been previously detailed (G. Allen et al., in press).

Construction of recombinant plasmids and M13 phage. The *Bam*HI H fragment of the genomic DNA (0.114 to 0.158 map units) of EHV-1 strain Kentucky T431 was excised from a 0.6% low-melting-temperature agarose gel, purified by Elutip-D chromatography (46; Schleicher & Schuell, Keene, N.H.), and then cloned into plasmid pIBI24. After the right terminal 1.0 kilobase of DNA was removed from the H fragment with *Eco*RI, the remaining 6.0-kilobase *Bam*HI-*Eco*RI fragment of EHV-1 DNA was amplified in pIBI24 and subcloned into M13mp18 for sequencing.

A nested set of overlapping deletion subclones of the *Bam*HI-*Eco*RI fragment inserted in mp18 was prepared for sequencing by the exonuclease III method of Henikoff (20), using the protocols and reagents (Erase-A-Base kit) prepared by Promega, Inc., Madison, Wis.

DNA sequencing methods. DNA sequence analysis was performed by the dideoxy chain termination method (45) with single-stranded M13mp18 phage DNA as the template. M13 DNA was prepared from 6-h LB broth cultures as described by Dale et al. (11), with *E. coli* NM522 as the host (IBI). The DNA synthesis reactions were primed with either a 20-residue universal primer (IBI) that hybridized to M13 sequences adjacent to the inserted DNA or with custom 18-residue oligonucleotides complementary to a specific region of the EHV-1 insert DNA.

The latter primers were synthesized by the University of Kentucky Macromolecular Structure Analysis Facility by using the phosphoramidite procedure and a model 380B DNA synthesizer (Applied Biosystems, Foster City, Calif.). Primers were extended with the Klenow fragment of DNA polymerase in the presence of ^{35}S-labeled dATP. Wedge-shaped sequencing gels containing 8% polyacrylamide were used to resolve the products. Compression anomalies occurring in DNA regions of high G+C content were resolved by substituting 7-deaza-dGTP (Boehringer Mannheim Biochemicals, Indianapolis, Ind.) for dGTP in the sequencing reactions (34).

Determination of the nucleotide sequence of EHV-1 DNA cloned into λgt11 was performed either directly by using a 15-nucleotide primer complementary to the β-galactosidase gene 10 nucleotides upstream from the *Eco*RI site (New England BioLabs, Inc., Beverly, Mass.) or after subcloning into mp18.

Analysis of DNA sequence data. A sonic digitizer (IBI) was used to read and transfer nucleotide sequence data to a computer file. Regions of sequence overlap were located, and the final sequence was assembled with the IBI Pustell sequence analysis software (38). Analysis of the DNA sequence data was performed with the PC/GENE software package (Intelligenetics, Inc., Mountain View, Calif.) and with the programs available through the BIONET national computer resource. Searches of data bases for DNA and protein sequences homologous with those of EHV-1 gp13 were performed with the FASTN/P programs of Lipman and Pearson (31) by scanning the GenBank and NBRL sequence data bases, respectively. Alignments of homologous amino acid sequences were performed with the BIONET GENALIGN program, which uses the method of Sobel and Martinez (49).

RESULTS

Strategy of EHV-1 DNA sequence determination. We previously showed that the DNA in a *Bam*HI-*Eco*RI restriction fragment between 0.114 and 0.148 map units on the EHV-1 genome encodes a major glycoprotein designated gp13 (4). To identify and characterize the gp13 coding sequences, the determination of the nucleotide sequence of this 6-kilobase fragment was undertaken. Preliminary analysis of the data identified, within the sequences generated from the right (*Eco*RI) end of the fragment, the subset of EHV-1 DNA sequences that were inserted and expressed in λgt11 as gp13 epitopes, thus further localizing the gene for gp13 to this region on the fragment. A 1,765-bp segment of DNA from the *Acc*I site at 0.136 map units to the *Eco*RI site at 0.148 map units was therefore completely sequenced.

Generation of subclones of the *Acc*I-*Eco*RI fragment in M13mp18 for sequencing was accomplished by the unidirectional deletion technique of Henikoff (20), in which an ordered set of overlapping deletion subclones of the insert DNA to be sequenced is generated with exonuclease III. The locations on the parent sequence of the individual subregions of DNA that were sequenced are shown in Fig. 1. Each nucleotide of the DNA was sequenced at least twice; generally, three to four determinations were made for each region. Gaps in the sequence were filled in by using custom oligonucleotide primers to extend the DNA sequence from existing clones through the sequence gaps (53).

Analysis of nucleotide sequence of gp13 gene. The entire nucleotide sequence of the 1,765-bp *Acc*I-*Eco*RI fragment (map coordinates 0.136 to 0.148) was determined (Fig. 2). An open reading frame (ORF) of 1,404 bp was present that contained the 200- to 400-bp EHV-1 DNA subsequences cloned and expressed in λgt11 as epitopes reactive with gp13-specific monoclonal antibodies. At least 16 stop codons were present in each of the other five reading frames within the same region of DNA (Fig. 3). The coding potential of the 1,404 ORF was confirmed by a codon usage analysis (52) using EHV-1 gp14 as a standard for codon usage bias (M. Whalley, personal communication). Final verification of the authenticity of this reading frame for gp13 was obtained by sequencing of the EHV-1 DNA inserted into λgt11 and expressed as an in-frame fusion protein with β-galactosidase (24).

The overall base composition of the single, long ORF was 24.7% A, 21.6% T, 25.2% G, and 28.4% C. The first of five ATGs in the 1,404-bp ORF conforms sufficiently to Kozak's consensus sequence of a strong translation initiation signal (27, 28). Although the second in-frame ATG is also in a favorable context for translation initiation, it is not followed by nucleotides encoding a membrane insertion signal sequence. The proposed chain termination codon, TAG, is found at nucleotide position 1528.

The DNA sequence for 123 bases upstream and 237 bases downstream of the coding sequence was also determined to

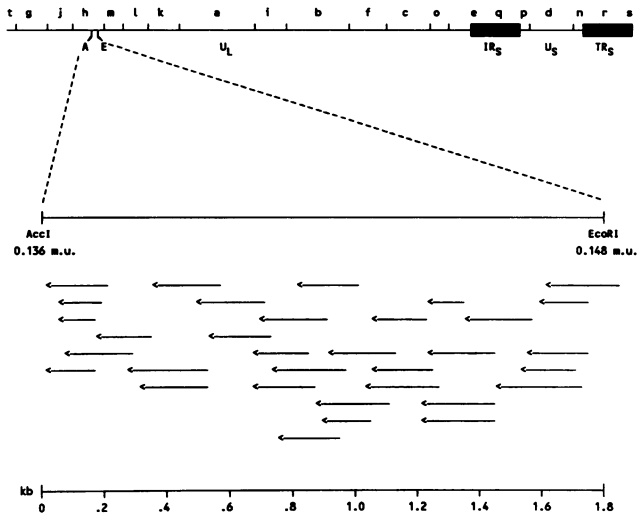

FIG. 1. Strategy for determining the nucleotide sequence of the EHV-1 gp13 gene. The upper part of the figure shows a conventional representation of the genome of EHV-1 (62). The long unique (U_L) and short unique (U_S) segments (—) and the short repeat (IR_S and TR_S) sequences (■) are shown. The *Bam*HI cleavage sites are shown, and the *Bam*HI fragments are designated with lowercase letters. The relative location of the *Acc*I-*Eco*RI fragment in the EHV-1 genome is indicated (A to E). The lower part of the figure expands the 1,765-bp *Acc*I-*Eco*RI region and shows the relative positions of the 33 clones used to determine the sequence of the fragment. The arrows indicate the locations, lengths, and directions of the sequences read from each sequencing template. kb, Kilobase; m.u., map units.

identify putative noncoding, regulatory sequences of the gp13 transcriptional unit. The upstream sequence contains possible CAT (nucleotides 8 to 12) and TATA boxes (nucleotides 55 to 59) characteristic of many eucaryotic promoters (10). By using the observed upstream distances of such promoter elements from transcriptional start sites (i.e., −70 to −85 and −20 to −35, respectively), the transcriptional start site of the gp13 gene can be predicted to be in the region of nucleotides 75 to 90 (10). With these putative transcriptional and translational start signals, a nontranslated leader region for gp13 mRNA of approximately 50 nucleotides in length can be predicted.

No apparent polyadenylation signal (AATAAA) could be found following the translation stop codon. A second set of possible CAT and TATA motifs, located at the 3′ end of the coding sequence at nucleotides 1717 and 1738, respectively, probably represent promoter elements for a contiguous EHV-1 gene homologous to the HSV gene encoding an 18-kilodalton protein and immediately 3′ to the HSV gC gene (16, 54).

Nucleotide sequences within the 5′ untranslated leader of the predicted gp13 mRNA that are complementary to the 18S rRNA sequence 3′-AGGAAGGCGT-5′ could not be identified. Likewise, no sequences with the properties of consensus donor or acceptor splice junction sequences were present in the ORF (35).

Predicted amino acid sequence of EHV-1 gp13. The deduced amino acid sequence for the polypeptide encoded by the 1,404-bp ORF is indicated in Fig. 2. The codon usage for the 468 amino acids is given in Table 1. Only one codon (CGT [Arg]) was not used at all. Although strong codon bias exists, it cannot be wholly explained by the preferred use of codons with G or C in the third base position that has been noted for herpesvirus genomes with much higher G+C content than EHV-1 DNA. For example, the most frequently used codons for Val and Phe have third base Ts.

The molecular mass of the 468-amino-acid primary translation product is 50,889 daltons. The protein predicted from the DNA sequence has features characteristic of a membrane-spanning glycoprotein (14). The first 30 amino acids possess the consensus sequence of an N-terminal, hydrophobic signal peptide, as discussed by McGeoch (33). Using the −3, −1 rule of von Heijne (59), the cleavage site between the signal peptide and the mature protein was predicted to lie after Ala_{30}.

A hydropathic analysis of the protein encoded by this ORF is shown in Fig. 4. A strongly hydrophobic region of 27 amino acids present near the carboxy terminus at positions 428 to 454 has the properties of a membrane insertion sequence. The carboxy-terminal 14 amino acids have a net positive charge similar to that of the cytoplasmic domains of other herpesvirus glycoproteins (7, 16, 61).

The 397-amino-acid surface domain of gp13 is predicted from secondary structure analysis (9) to consist predominantly of extended β-strands linked by regions of β-turns (data not shown). This surface domain contains regions of both extended hydrophobicity and hydrophilicity and may represent the internal and surface regions, respectively, of the exterior, globular domain of gp13. Nine potential N-linked glycosylation sites (Asn-X-Ser/Thr) are present in this surface domain between the putative signal sequence and transmembrane anchor of the protein (see Fig. 2).

Localization of a region of the gp13 gene encoding an immunodominant antigenic site. Recombinant λgt11 phage expressing EHV-1 gp13 epitopes were constructed as described previously (4) and identified by immunoassay with gp13-specific monoclonal antibodies (G. Allen et al., in press). One subset of 12 λgt11 recombinants expressed an antigenic determinant reactive with each of 15 members of a panel of 42 independently derived monoclonal antibodies to EHV-1 gp13. The insert EHV-1 DNA present in four such lambda clones expressing gp13 epitopes was subcloned into M13mp18 and sequenced. The maximal boundaries of the gp13 epitope-encoding sequences that were expressed in *E. coli* by λgt11 were thus localized within the gp13 gene to a region between nucleotides 529 and 694 (Fig. 2). This region encodes amino acids 136 to 191 in the N-terminal surface domain of the predicted glycoprotein molecule. Contained within this protein subregion is the most hydrophilic sequence of six contiguous amino acids present in the gp13 molecule (positions 145 through 150; Glu-Arg-Lys-Lys-Ser-Arg) that is predicted by the algorithm of Hopp and Woods (22) to comprise an antigenic site. Studies using synthetic oligopeptides to define more precisely the amino acids within this protein region that contribute to the antigenic site are under way.

Homology between gp13 and other herpesvirus glycoproteins. Data presented by Davison and Wilke (12) indicate that the L segment of the EHV-1 genome is colinear with that of the I_L isomer of HSV. The coding sequences of EHV-1 gp13 therefore map to a position on the genome that is colinear with the genes for HSV gC-1 and gC-2 (i.e., 0.12 to 0.14 map units on the HSV I_L isomer; 16, 54). To verify that gp13 is the EHV-1 structural homolog of gC-like glycoproteins of other herpesviruses, its DNA and protein sequences were compared with those of the gC-like glycoproteins of the three other alphaherpesviruses for which nucleotide sequence data is available: HSV, pseudorabies herpesvirus (PRV), and varicella-zoster virus (VZV) (16, 26, 41, 54).

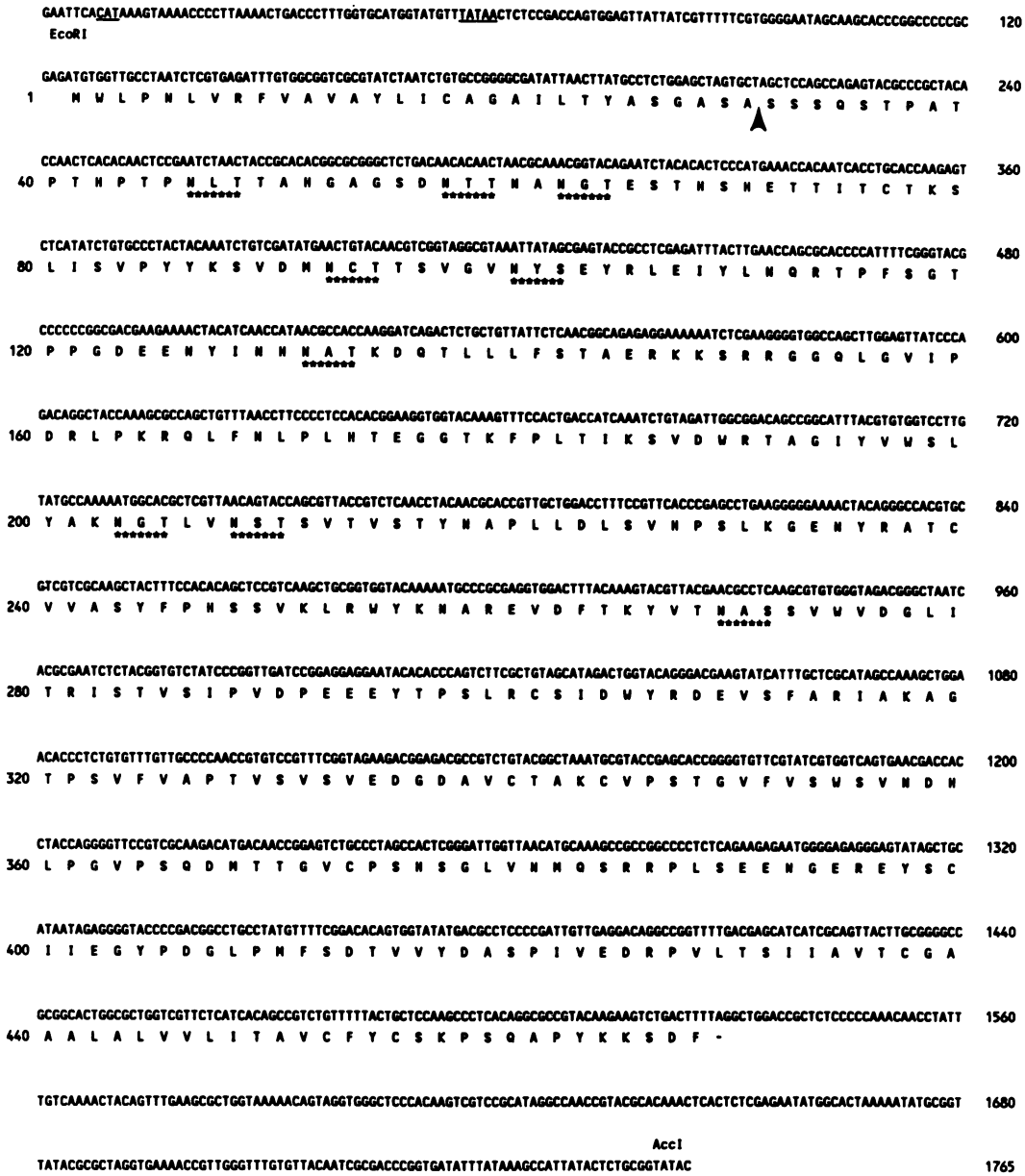

FIG. 2. Nucleotide sequence of the 1,765-bp, *Eco*RI-*Acc*I segment of EHV-1 DNA encoding glycoprotein 13. The sequence represented is that of the noncoding (messenger sense) strand and includes the entire gp13 protein-coding sequence, 123 nucleotides of the 5′-flanking sequence, and 237 nucleotides of the 3′-flanking sequence. Nucleotides are numbered at the right of the sequence. The derived amino acid sequence of the 1,404-bp ORF is indicated in the conventional one-letter code, and the amino acids are numbered at the left of the sequence. The presumptive TATA and CAT boxes are underlined. The arrow indicates the position at which the 30-residue signal sequence is cleaved from the mature polypeptide. The starred segments are the Asn-X-Thr/Ser motif, proposed to be N-glycosylation sites.

By using parameters in the BIONET FASTN and GENALIGN DNA homology search programs set at values to detect minimal homology, no statistically significant homology could be detected between the nucleotide sequences encoding gp13 and those encoding the gC glycoproteins of HSV type 1 (HSV-1) and HSV-2, PRV, or VZV. A comparison of the predicted amino acid sequences for EHV-1 gp13, HSV-1 gC (16), PRV gIII (41), and VZV gp66 (26) is shown in Fig. 5. The sequences have been aligned by the BIONET GENALIGN multialignment program (49), with gaps introduced to maximize the amino acid homology. Statistical analysis performed by the computer program on the alignments indicated that the homology shown in Fig. 5 is 20-fold greater than the number of amino acid matches present in randomly permuted subsegments of the protein molecules, suggesting that the homology is highly significant.

Of the 281 amino acids at the carboxy terminus of gp13, 107 (38%) were shared with PRV gIII, 69 (25%) with HSV-1 gC, and 82 (29%) with VZV gp66. Of these 281 C-terminal amino acid sequences, 28 (10%) are common among all four glycoproteins. The most extensive glycoprotein homology within the C-terminal portion of the molecules was present in a stretch of amino acids between residues 297 and 346 in gp13. Of the amino acids in this region, 34% are matched in two or more of the herpesvirus glycoproteins, while 22% are perfectly matched among all four glycoproteins. Strong

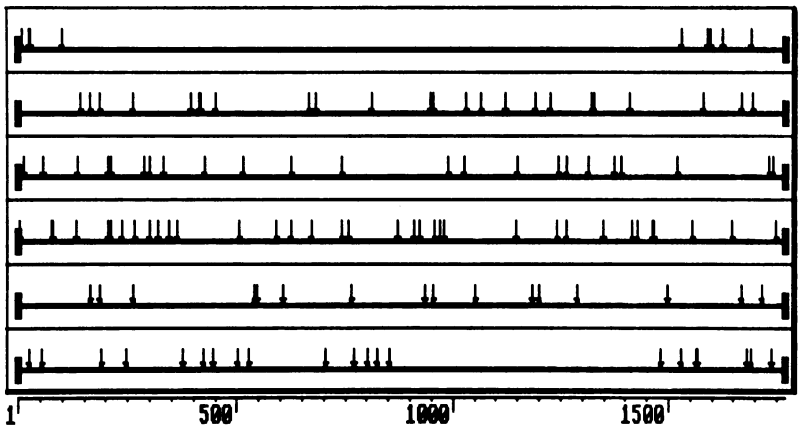

FIG. 3. Identification of the ORFs within the 1,765-bp *Acc*I-*Eco*RI fragment (0.136 to 0.148 map units) of EHV-1 DNA. Each panel represents one of six reading frames. The vertical arrows indicate the termination codons present in each reading frame. Reading frame 1 (top panel) contains a 1,404-bp region without stop codons. Numbers at the bottom indicate base pairs.

TABLE 1. Codon usage in the gp13 gene of EHV-1

Codon (amino acid or TER)[a]	Frequency of occurrence	% of total[b]	Codon (amino acid or TER)	Frequency of occurrence	% of total
TTT (Phe)	11	2.3	TAT (Tyr)	6	1.2
TTC (Phe)	2	0.4	TAC (Tyr)	16	3.4
TTA (Leu)	2	0.4	TAA (TER)	0	0
TTG (Leu)	6	1.2	TAG (TER)	0	0
CTT (Leu)	4	0.8	CAT (His)	2	0.4
CTC (Leu)	7	1.4	CAC (His)	8	1.7
CTA (Leu)	5	1.0	CAA (Gln)	2	0.4
CTG (Leu)	10	2.1	CAG (Gln)	6	1.2
ATT (Ile)	3	0.6	AAT (Asn)	6	1.2
ATC (Ile)	11	2.3	AAC (Asn)	15	3.2
ATA (Ile)	6	1.2	AAA (Lys)	8	1.7
ATG (Met)	5	1.0	AAG (Lys)	10	2.1
GTT (Val)	14	2.9	GAT (Asp)	4	0.8
GTC (Val)	10	2.1	GAC (Asp)	17	3.6
GTA (Val)	9	1.9	GAA (Glu)	10	2.1
GTG (Val)	12	2.5	GAG (Glu)	11	2.3
TCT (Ser)	11	2.3	TGT (Cys)	5	1.0
TCC (Ser)	8	1.7	TGC (Cys)	7	1.4
TCA (Ser)	7	1.4	TGA (TER)	0	0
TCG (Ser)	7	1.4	TGG (Trp)	7	1.4
CCT (Pro)	3	0.6	CGT (Arg)	0	0
CCC (Pro)	10	2.1	CGC (Arg)	7	1.4
CCA (Pro)	8	1.7	CGA (Arg)	2	0.4
CCG (Pro)	10	2.1	CGG (Arg)	3	0.6
ACT (Thr)	7	1.4	AGT (Ser)	5	1.0
ACC (Thr)	13	2.7	AGC (Ser)	14	2.9
ACA (Thr)	15	3.2	AGA (Arg)	1	0.2
ACG (Thr)	12	2.5	AGG (Arg)	7	1.4
GCT (Ala)	6	1.2	GGT (Gly)	5	1.0
GCC (Ala)	14	2.9	GGC (Gly)	8	1.7
GCA (Ala)	7	1.4	GGA (Gly)	6	1.2
GCG (Ala)	7	1.4	GGG (Gly)	8	1.7

[a] TER, Chain termination signal.
[b] Total, 468 amino acids.

amino acid homology was also observed within the hydrophobic membrane anchor sequence of the four herpesvirus glycoproteins.

Substantially fewer sequence similarities were present among the amino acids composing the N-terminal half of the

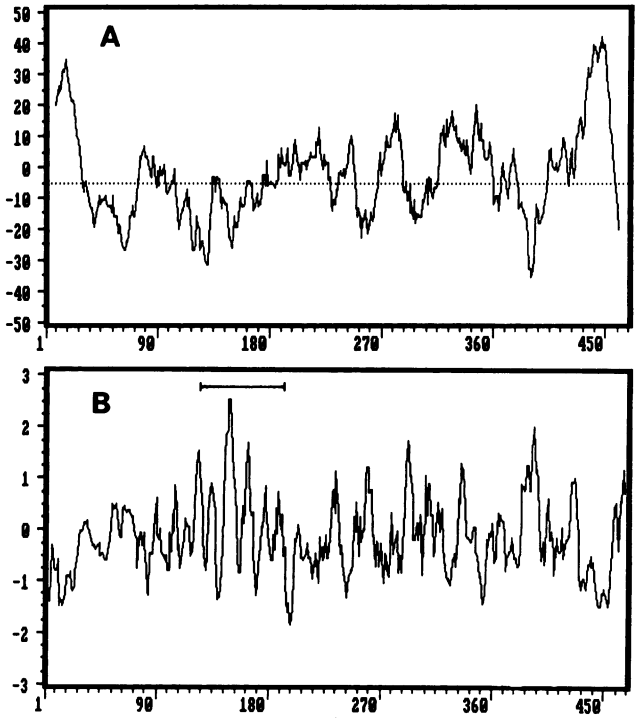

FIG. 4. Hydropathy plots of the sequence of 468 amino acids composing EHV-1 gp13. The hydrophobicity profile in panel A was determined with the values of Kyte and Doolittle (29) by using a 15-amino-acid window. The two most hydrophobic regions at the N- and C-termini are predicted to represent the signal sequence and transmembrane spanning region, respectively, of the glycoprotein molecule. The hydrophilicity profile in panel B was generated with the values of Hopp and Woods (22) by using a 6-amino-acid window. The horizontal bar delineates the limits for the gp13 sequences expressed in λgt11 as immunodominant epitopes. The upward-pointing peaks represent the most hydrophobic regions in panel A and the most hydrophilic regions in panel B.

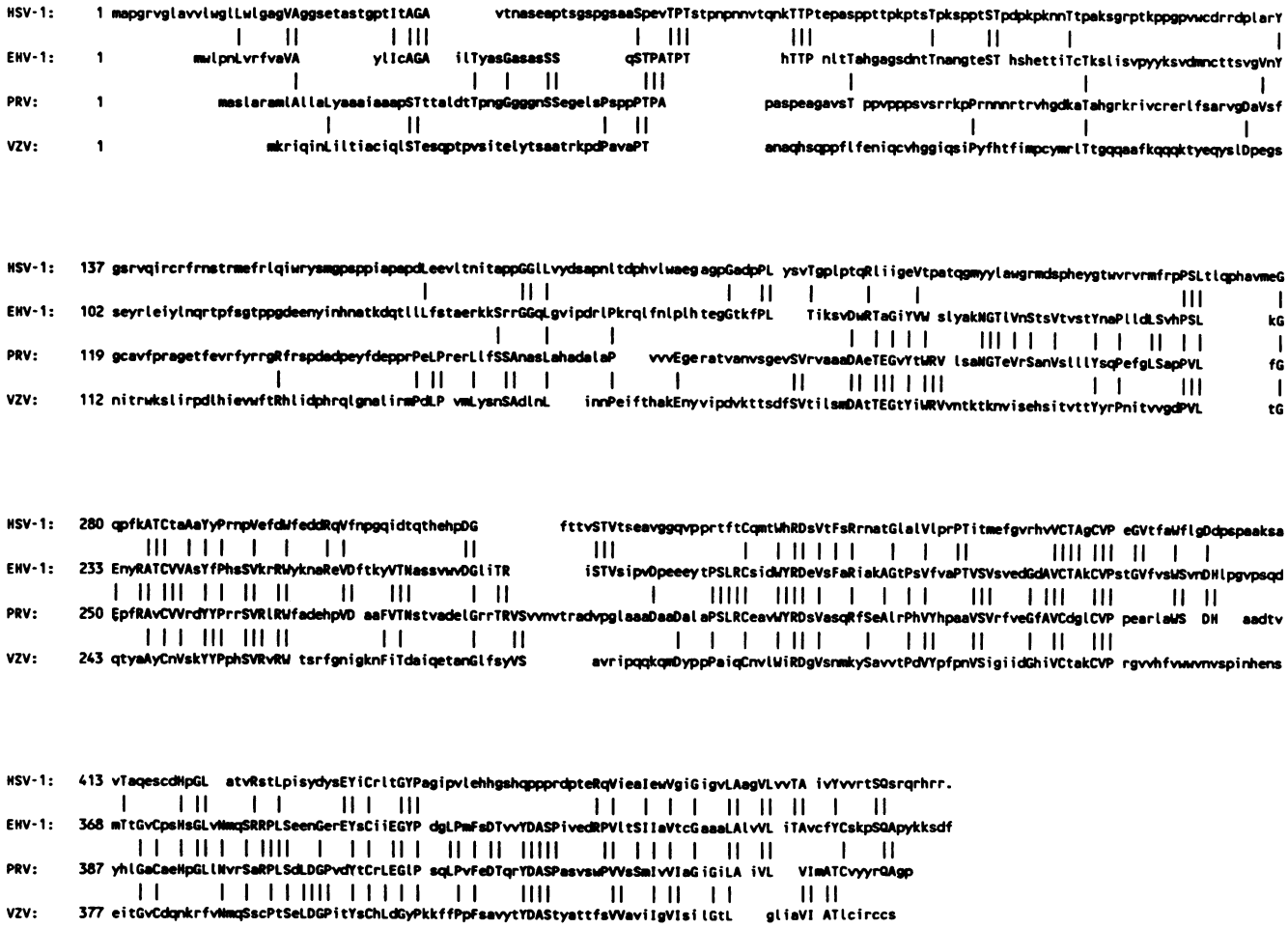

FIG. 5. Homology of EHV-1 gp13 with other herpesvirus glycoproteins. The amino acid sequences of EHV-1 gp13 have been compared with those of HSV-1 gC (16), PRV gIII (41), and VZV gp66 (26). The sequences were aligned with the BIONET GENALIGN program of Sobel and Martinez (49). Locations of the amino acid sequences shown for the proteins are given at the left of each sequence. Gaps have been introduced by the alignment program to maximize the homology. Only one of the eight copies of the 14-amino-acid repeat present in VZV gp66 (26) was included in the alignment. Amino acids that are identical in any two aligned sequences are indicated by the connecting vertical lines.

four glycoproteins. Of the 187 amino acids in this region of the gp13 molecule, only 14 (7%), 27 (14%), and 21 (11%) were identical in gIII, gC, and gp66, respectively. Furthermore, none of the 187 N-terminal amino acids of gp13 were perfectly matched in all four glycoproteins. The difference between the N- and C-terminal conservation of amino acid sequences among the four gC-like glycoproteins was visually illustrated by the matrix homology plot of Pustell and Kafatos (38). Plots of the gp13 amino acid sequence against that of each of the other three gC-like glycoproteins revealed the better matching of amino acids in the carboxy-terminal sequences, as indicated by the more solid diagonal lines in that region (data not shown).

There are nine potential N-glycosylation sites in gp13, eight in both HSV gC and PRV gIII, and five in VZV gp66. Although none of these possible N-linked oligosaccharide sites are at conserved positions in all four proteins, two of the gp13 sites (amino acids 103 and 269) are present in matching locations in PRV gIII.

DISCUSSION

As part of a long-term research project designed to characterize the antigens of EHV-1 responsible for stimulating or suppressing protective immunity in horses, we have begun to examine the molecular properties of the major glycoproteins of EHV-1. In this report, we have extended our analysis of an EHV-1 glycoprotein that shows significant homology to glycoprotein C of human HSV. The importance of gC-like glycoproteins in eliciting immune responses has been well documented for other herpesviruses. Glycoprotein C of HSV-1, for example, has been reported to be the immunodominant antigen for eliciting cytotoxic T-cell responses in mice (17, 44). Likewise, the majority of the virus-neutralizing activity against PRV present in convalescent serum of PRV-infected swine has been demonstrated to be directed against the PRV counterpart of gC (gIII) (6). Although the function of EHV-1 gp13 is unknown, its importance in the immune response to EHV-1 infection is indicated by the presence of anti-gp13 antibody in the serum of convalescing horses and by the finding that several anti-gp13 monoclonal antibodies will neutralize viral infectivity (G. Allen et al., in press).

The present study has allowed the precise identification and mapping of the EHV-1 gene for gp13 and has demonstrated unequivocally that the gene encodes a glycoprotein homologous to HSV-1 gC. Except for gp14 (M. Whalley,

personal communication), no other glycoprotein gene of EHV-1 has been sequenced. The glycoprotein deduced from the nucleotide sequence of the gene has the standard membrane protein arrangement of an N-terminal signal sequence, an external domain containing a number of possible N-glycosylation sites, a C-terminal membrane anchor region, and a short cytoplasmic domain. The assignment of the coding sequence for gp13 rests on the assumption that RNA splicing does not occur. Most HSV mRNAs, including that for HSV gC-1, are unspliced (60).

The gp13 gene is defined by an ORF of 1,404 nucleotides or 468 codons. The primary translation product predicted for the DNA sequence is therefore similar in size to that of PRV gIII (479 amino acids) and HSV-2 gC (480 amino acids), an observation consistent with the identical apparent molecular masses of the three fully glycosylated glycoproteins (i.e., 90 kilodaltons). The gC coding regions of HSV-1 and VZV are 43 and 123 codons longer, respectively, than that of EHV-1 gp13.

The cytoplasmic domain of gp13 is relatively small (14 amino acids). Small cytoplasmic domains have also been observed in gC-like glycoproteins of HSV-1, HSV-2, PRV, and VZV (16, 26, 41, 54) and appear to be a characteristic feature of this family of homologous glycoproteins.

In HSV, the gC gene has no polyadenylation signal immediately after it. Instead, the gC mRNA is 3′ coterminal with that of the transcript of an adjacent gene encoding an 18-kilodalton protein (16, 54). Similarly, there is no canonical polyadenylation sequence (AATAAA) following the gp13 termination codon.

The overall homologies between gp13 and its homologs gC-1, gIII, and gp66 are 21, 26, and 22%, respectively. A similar amount of homology exists between gp13 and HSV-2 gC-2 (data not shown). By comparison, the gC glycoproteins of HSV-1 and HSV-2 are 73% homologous (54), while the gp14 glycoproteins of the two subtypes of EHV-1 share 85% of their amino acid sequences (M. Whalley, personal communication; M. Riggio, personal communication). The homology observed between gp13 and its structural homologs was patchy and scattered over most of the molecule. Only 28 amino acids were perfectly matched in all four gC-like glycoproteins examined.

Of special interest is the observation that amino acid sequences composing the N-terminal segment of EHV-1 gp13 have undergone more extensive evolutionary divergence than those at the C-terminal half of the molecule. This appears to be a universal phenomenon for the gC-like glycoproteins of all alphaherpesviruses whose gC sequences have been determined (16, 26, 41, 54). The environmental pressures responsible for selecting a relatively greater number of amino acid substitutions in this domain of the glycoprotein molecule are unknown. However, as demonstrated by the present studies on gp13, an immunodominant antigenic site is located within this subregion of the glycoprotein and no amino acids within the six-residue subsequence predicted by the alogrithm of Hopp and Woods (22) to comprise this antigenic epitope have been conserved among the four herpesviruses. The greater sequence diversity present in the N-terminal amino acids of gC-like glycoproteins may therefore reflect a region of the molecule that interacts with the host to elicit protective immune responses and the subsequent positive effect of such immune mechanisms in selecting amino acid substitutions that occur in this region.

An alternative explanation for the observed greater conservation of C-terminal amino acid sequences of herpesvirus gC glycoproteins is that essential structural or functional roles carried out by this portion of the glycoprotein molecule impose constraints on the extent of sequence diversity that can be tolerated without loss of those functions. It is not yet known what biological functions gC-like glycoproteins have in the herpesvirus life cycle. Furthermore, it has been shown that HSV gC-1 and gC-2 as well as PRV gIII are nonessential for virus infectivity of cultured cells (21, 42, 65). Examination of this aspect of the question must therefore await further clarification of the function(s) of gC-like glycoproteins in the herpesvirus-host interaction. Pertinent to the latter hypothesis, however, is the observation that some of the most homologous amino acid sequences among the four gC-like glycoproteins examined occur in the structurally essential transmembrane anchor domain.

The overall lack of conservation of homology in the primary structure for gC-like glycoproteins of alphaherpesviruses, in contrast to the highly conserved structure observed for gB-like glycoproteins (8, 13, 36, 47, 48), does not suggest a conserved, essential replicative function for this gC family of glycoproteins. Rather, the highly divergent sequences of gC-like glycoproteins suggest a function that contributes to the adaptation of the particular virus type to its unique ecological niche within its specific host (for a discussion, see reference 32). This interpretation is consistent with the observation that the antigenic epitopes on the gC glycoproteins of the two biotypes of HSV (HSV-1 and HSV-2) and of EHV-1 (EHV-1 subtypes 1 and 2) are predominantly type specific, while those of gB-like glycoproteins are predominantly type common (50; G. Allen et al., in press).

ACKNOWLEDGMENTS

These investigations were supported by grants from the Grayson Foundation, Inc., and from the U.S. Department of Agriculture (85-CRSR-2-2627).

We are grateful to Millar Whalley for providing unpublished data on the nucleotide sequence of EHV-1 gp14.

LITERATURE CITED

1. **Ackermann, M., R. Longnecker, B Roizman, and L. Pereira.** 1986. Identification, properties, and gene location of a novel glycoprotein specified by herpes simplex virus 1. Virology **150:** 207–220.
2. **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.
3. **Allen, G. P., D. J. O'Callaghan, and C. C. Randall.** 1977. Genetic relatedness of equine herpesvirus types 1 and 3. J. Virol. **24:**761–767.
4. **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.
5. **Allen, G., M. Yeargan, L. Turtinen, J. Bryans, and W. McCollum.** 1983. Molecular epizootiologic studies of equine herpesvirus 1 infections by restriction endonuclease fingerprinting of viral DNA. Am. J. Vet. Res. **44:**263–271.
6. **Ben-Porat, T., J. DeMarchi, B. Lomniczi, and A. Kaplan.** 1986. Role of glycoproteins of pseudorabies virus in eliciting neutralizing antibodies. Virology **154:**325–334.
7. **Bzick, 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:**301–307.
8. **Bzik, D. J., C. Debroy, B. Fox, N. Pederson, and S. Person.** 1986. The nucleotide sequence of the gB glycoprotein gene of HSV-2 and comparison with the corresponding gene of HSV-1. Virology **155:**322–333.
9. **Chou, P. Y., and G. D. Fasman.** 1974. Prediction of protein

conformation. Biochemistry **13**:222–245.
10. **Corden, J., B. Wasylyk, A. Buckwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon.** 1980. Promoter sequences of eukaryotic protein-coding genes. Science **209**:1406–1414.
11. **Dale, R. M., B. A. McClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rDNA. Plasmid **13**:31–40.
12. **Davison, A. J., and N. M. Wilke.** 1983. Location and orientation of homologous sequences in the genomes of five herpesviruses. J. Gen. Virol. **64**:1927–1942.
13. **Edson, C. M., B. A. Hosler, R. A. Respess, D. J. Waters, and D. A. Thorley-Lawson.** 1985. Cross-reactivity between herpes simplex virus glycoprotein B and a 63,000-dalton varicella-zoster virus envelope glycoprotein. J. Virol. **56**:333–336.
14. **Eisenberg, D.** 1984. Three-dimensional structure of membrane and surface proteins. Annu. Rev. Biochem. **53**:595–623.
15. **Ellis, R. W., P. M. Keller, R. S. Lowe, and R. A. Zivin.** 1985. Use of a bacterial expression vector to map the varicella-zoster virus major glycoprotein gene, gC. J. Virol. **53**:81–88.
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., U. Kees, G. Kumel, H. Kirchner, and P. Krammer.** 1985. Identification of herpes simplex virus type 1 (HSV-1) glycoprotein gC as the immunodominant antigen for HSV-1-specific memory cytotoxic T lymphocytes. J. Immunol. **135**: 575–582.
18. **Gompels, U., and A. Minson.** 1986. The properties and sequence of glycoprotein H of herpes simplex virus type 1. Virology **153**: 230–247.
19. **Hampl, H., T. Ben-Porat, L. Erhlicher, K.-O. Habermehl, and A. S. Kaplan.** 1984. Characterization of the envelope proteins of pseudorabies virus. J. Virol. **52**:583–590.
20. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.
21. **Holland, T. C., F. L. Homa, S. D. Marlin, M. Levine, and J. Glorioso.** 1984. Herpes simplex virus type 1 glycoprotein C-negative mutants exhibit multiple phenotypes, including secretion of truncated glycoproteins. J. Virol. **52**:566–574.
22. **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.
23. **Huynh, T. V., R. A. Young, and R. W. Davis.** 1985. Construction and screening cDNA libraries in λgt10 and λgt11, p. 49–78. *In* D. M. Glover (ed.), DNA cloning: a practical approach, vol. 1. IRL Press, Washington, D.C.
24. **Kalnins, A., K. Otto, U. Ruther, and B. Muller-Hill.** 1983. Sequence of the lacZ gene of *Escherichia coli*. EMBO J. **2**:593–597.
25. **Keller, P. M., A. J. Davison, R. S. Lowe, C. D. Bennet, and R. W. Ellis.** 1986. Identification and structure of the gene encoding gpII, a major glycoprotein of varicella-zoster virus. Virology **152**:181–191.
26. **Kinchington, P. R., J. Remenick, J. M. Ostrove, S. E. Straus, W. T. Ruyechan, and J. Hay.** 1986. Putative glycoprotein gene of varicella-zoster virus with variable copy numbers of a 42-base-pair repeat sequence has homology to herpes simplex virus glycoprotein C. J. Virol. **59**:660–668.
27. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
28. **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.
29. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
30. **Lee, G. T.-Y., M. F. Para, and P. G. Spear.** 1982. Location of the structural genes for glycoproteins gD and gE and for other polypeptides in the S component of herpes simplex virus type 1 DNA. J. Virol. **43**:41–49.
31. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
32. **Longnecker, R., and B. Roizman.** 1987. Clustering of genes dispensable for growth in culture in the S component of the HSV-1 genome. Science **236**:573–576.
33. **McGeoch, D. J.** 1985. On the predictive recognition of signal peptide sequences. Virus Res. **3**:271–286.
34. **Mizusawa, S., S. Nishimura, and F. Seela.** 1986. Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. Nucleic Acids Res. **14**:1319–1324.
35. **Mount, S.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
36. **Pellet, 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.
37. **Petrovskis, E. A., J. G. Timmins, and L. E. Post.** 1986. Use of λgt11 to isolate genes for two pseudorabies virus glycoproteins with homology to herpes simplex virus and varicella-zoster virus glycoproteins. J. Virol. **60**:185–193.
38. **Pustell, J., and F. C. Kafatos.** 1984. A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis, and homology determination. Nucleic Acids Res. **12**:643–655.
39. **Rea, T. J., J. G. Timmins, G. W. Long, and L. E. Post.** 1985. Mapping and sequence of the gene for the pseudorabies virus glycoprotein which accumulates in the medium of infected cells. J. Virol. **54**:21–29.
40. **Richman, D. D., A. Buckmaster, S. Bell, C. Hodgman, and A. C. Minson.** 1986. Identification of a new glycoprotein of herpes simplex virus type 1 and genetic mapping of the gene that codes for it. J. Virol. **57**:647–655.
41. **Robbins, A. K., R. J. Watson, M. E. Whealy, W. W. Hays, and L. W. Enquist.** 1986. Characterization of a pseudorabies virus glycoprotein gene with homology to herpes simplex virus type 1 and type 2 glycoprotein C. J. Virol. **58**:339–347.
42. **Robbins, A. K., M. E. Whealy, R. J. Watson, and L. W. Enquist.** 1986. Pseudorabies virus gene encoding glycoprotein gIII is not essential for growth in tissue culture. J. Virol. **59**:635–645.
43. **Roizman, B., B. Norrild, C. Chan, and L. Pereira.** 1984. Identification and preliminary mapping with monoclonal antibodies of a herpes simplex virus 2 glycoprotein lacking a known type 1 counterpart. Virology **133**:242–247.
44. **Rosenthal, K. L., J. R. Smiley, S. South, and D. C. Johnson.** 1987. Cells expressing herpes simplex virus glycoprotein gC but not gB, gD, or gE are recognized by murine virus-specific cytotoxic T lymphocytes. J. Virol. **61**:2438–2447.
45. **Sanger, F., S. Nicklen, and A. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
46. **Schmitt, J. J., and B. N. Cohen.** 1983. Quantitative isolation of DNA restriction fragments from low-melting agarose by Elutip-D affinity chromatography. Anal. Biochem. **133**:462–464.
47. **Snowden, B. W., and I. W. Halliburton.** 1985. Identification of cross-reacting glycoproteins of four herpesviruses by Western blotting. J. Gen. Virol. **66**:2039–2044.
48. **Snowden, B. W., P. R. Kinchington, K. L. Powell, and I. W. Halliburton.** 1985. Antigenic and biochemical analysis of gB of herpes simplex virus type 1 and type 2 and of cross-reacting glycoproteins induced by bovine mammillitis virus and equine herpesvirus type 1. J. Gen. Virol. **66**:231–247.
49. **Sobel, E., and H. M. Martinez.** 1985. A multiple sequence alignment program. Nucleic Acids Res. **14**:363–374.
50. **Spear, P. G.** 1985. Antigenic structure of herpes simplex viruses, p. 425–443. *In* M. H. V. Van Regenmortel and A. R. Neurath (ed.), The basis for serodiagnosis and vaccines, vol. 2. Immunochemistry of viruses. Elsevier Science Publishing, Inc., New York.
51. **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.
52. **Staden, R., and A. D. McLachlan.** 1982. Codon preference and its use in identifying protein coding regions in long DNA sequences. Nucleic Acids Res. **10**:141–156.
53. **Strauss, E. C., J. A. Kobori, G. Siu, and L. Hood.** 1986. Specific-primer-directed DNA sequencing. Anal. Biochem. **154**: 353–360.
54. **Swain, M. A., R. W. Peet, and D. A. Galloway.** 1985. Characterization of the gene encoding herpes simplex virus type 2 glycoprotein C and comparison with the type 1 counterpart. J. Virol. **53**:561–569.
55. **Turtinen, L. W., and G. P. Allen.** 1982. Identification of the envelope surface glycoproteins of equine herpesvirus type 1. J. Gen Virol. **63**:481–485.
56. **Turtinen, L. W., G. P. Allen, R. W. Darlington, and J. T. Bryans.** 1981. Serologic and molecular comparisons of several equine herpesvirus type 1 strains. Am. J. Vet. Res. **42**:2099–2104.
57. **Van Drunen Littel-Van Den Hurk, S., and L. A. Babiuk.** 1986. Synthesis and processing of bovine herpesvirus 1 glycoproteins. J. Virol. **59**:401–410.
58. **Van Drunen Littel-Van Den Hurk, S., J. Van Den Hurk, and L. A. Babiuk.** 1985. Topographical analysis of bovine herpesvirus type-1 glycoproteins: use of monoclonal antibodies to identify and characterize functional epitopes. Virology **144**:216–227.
59. **Von Heijne, G.** 1986. A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. **14**:4683–4697.
60. **Wagner, E. K.** 1985. Individual HSV transcripts: characterization of specific genes, p. 45–104. *In* B. Roizman (ed.), The herpesviruses, vol. 3. Plenum Publishing Corp., New York.
61. **Watson, R. J., J. H. Weis, J. S. Salstrom, and L. W. Enquist.** 1982. Herpes simplex virus type 1 glycoprotein D gene: nucleotide sequence and expression in *Escherichia coli*. Science **218**: 381–385.
62. **Whalley, J. M., G. R. Robertson, and A. J. Davison.** 1981. Analysis of the genome of equine herpesvirus type 1: arrangement of cleavage sites for restriction endonucleases EcoRI, BglII, and BamHI. J. Gen. Virol. **57**:307–323.
63. **Young, R. A., and R. W. Davis.** 1983. Efficient isolation of genes by using antibody probes. Proc. Natl. Acad. Sci. USA **80**:1194–1198.
64. **Young, R. A., and R. W. Davis.** 1983. Yeast RNA polymerase II genes: isolation with antibody probes. Science **222**:778–782.
65. **Zezulak, K. M., and P. G. Spear.** 1984. Mapping of the structural gene for the herpes simplex virus type 2 counterpart of herpes simplex virus type 1 glycoprotein C and identification of a type 2 mutant which does not express this glycoprotein. J. Virol. **49**:741–747.