A novel herpes simplex virus gene (UL49A) encodes a putative membrane protein with counterparts in other herpesviruses

Barbara C. Barnett,*† Aidan Dolan,† Elizabeth A. R. Telford, Andrew J. Davison† and Duncan J. McGeoch†

Institute of Virology, University of Glasgow, Church Street, Glasgow G11 5JR, U.K.

Comparative analysis of DNA sequences located between the coding regions of genes UL49 and UL50 of herpes simplex virus types 1 and 2 (HSV-1 and -2) has revealed a small open reading frame (ORF) of 91 and 87 codons respectively with the characteristics of a genuine protein-coding region. The predicted protein products are clearly related and exhibit features of membrane-inserted proteins, with potential N-proximal signal peptides and C-proximal membrane anchor regions. Counterparts are present in the other sequenced alphaherpesviruses, namely varicella-zoster virus (a previously undescribed gene, 9A) and equine herpesvirus type 1 (gene 10), in the betaherpesvirus human cytomegalovirus (gene UL73) and in the gammaherpesvirus Epstein–Barr virus (gene BLRF1). Therefore, we consider that this ORF represents an additional HSV gene (UL49A) with counterparts in all sequenced alpha-, beta- and gammaherpesviruses.

The complete sequence of herpes simplex virus type 1 (HSV-1) strain 17 was predicted to contain 70 distinct genes (McGeoch et al., 1988). Determination of the genomic coding potential was based on critical sequence interpretation with gene designations requiring evidence beyond the presence of an open reading frame (ORF), of which many existed other than those considered genuinely functional. Specifically, the protein-coding potential of ORFs was evaluated on the basis of codon usage and triplet periodicity in base frequencies compatible with the high G+C content of the genome. Supporting evidence for functional ORFs came from experimental studies on gene structure, transcript layout, viral mutants and protein mapping, and from comparative sequence analyses with other herpesviruses. Two additional genes have since been described, namely UL26.5 (Liu & Roizman, 1991) and RL1 (Chou & Roizman, 1990; Dolan et al., 1992).

On the basis of the above criteria, the relatively large region of sequence (618 bp) between genes UL49 and UL50 was not assigned a protein-coding function. To characterize further the coding potential of these sequences we have analysed the equivalent region of the HSV-2 genome. We have identified a small conserved ORF with the characteristics of a functional gene in each HSV serotype. This gene potentially encodes a virus membrane protein, and has counterparts in other herpesvirus genomes.

We investigated the coding potential of the DNA sequence between HSV genes UL49 and UL50 by comparative analysis of published HSV-1 (McGeoch et al., 1988) and previously unreported HSV-2 sequences. The HSV-2 (strain HG52) BamHI w restriction fragment was isolated and purified from the recombinant plasmid pG226, containing the larger HindIII a fragment cloned into the vector pAT153. The DNA sequence was obtained using M13 dideoxynucleotide chain termination technology (Bankier & Barrell, 1989). Clones were sequenced using [α-35S]dATP (Amersham) and Sequenase version 2.0 T7 DNA polymerase (United States Biochemicals). Sequences which presented problems under standard conditions were resolved as described by McGeoch et al. (1988). A DEC Microvax II computer was used to assemble a database (Staden, 1987) and the complete sequence was interpreted with the aid of programs in the University of Wisconsin Genetics Computer Group (GCG) package (Devereux et al., 1984). The complete BamHI w fragment sequence contained 2851 bp and was derived from approximately $3 \times 10^4$ nucleotides of data with 84% of the sequence determined on both strands. Comparison with the HSV-1 sequence identified coding regions equivalent to HSV-1 gene UL49 and to the 939 bp at the 5' end of gene UL50, encoding the viral dUTPase (Fig. 1).

Alignment of the DNA sequences between genes UL49 and UL50 (Fig. 2) revealed a conserved ORF in the same orientation as UL49 and located close to the 5' end of UL50. The HSV-1 and -2 ORFs contain 91 and 87...
joined segments, termed the long (L) and short (S) regions, each consisting of a unique sequence (UL; US) flanked by inverted repeat sequences (TR; and IR; TS and IR). Unique sequences are shown as continuous lines and major repeats as open boxes. The lower part is an expansion of the BamHI w fragment showing the predicted coding regions UL49, UL49A and UL50.

Fig. 1. Predicted ORFs in the BamHI w fragment of HSV-2 (strain HG52). The upper part depicts HSV DNA comprising two covalently joined segments, termed the long (L) and short (S) regions, each consisting of a unique sequence (UL; US) flanked by inverted repeat sequences (TR; and IR; TS and IR). Unique sequences are shown as continuous lines and major repeats as open boxes. The lower part is an expansion of the BamHI w fragment showing the predicted coding regions UL49, UL49A and UL50.

codons, respectively, and show 70% identity of aligned bases and 56% identity of predicted amino acids. Analysis of the nucleotide sequences revealed features consistent with each HSV ORF representing a functional protein-coding region. First, HSV protein-coding sequences generally show characteristic triplet periodicities of base frequencies. Owing to the high G + C content of HSV DNA and the nature of the genetic code, the set of codon third positions have the highest G + C content and the set of second positions the lowest (Perry & McGeoch, 1988); in the HSV-1 and -2 ORFs, the G + C frequencies at positions 1, 2 and 3 are 71, 57 and 79%, and 71, 54 and 79%, respectively. Second, HSV-1 and -2 ORFs encoding corresponding proteins generally show the highest level of substitutions in the third position of aligned codons (McGeoch et al., 1987); at the first, second and third codon positions respectively the aligned HSV-1 and -2 sequences are mismatched by 30, 34 and 41%. Finally, sequences putatively involved in transcriptional regulation of the HSV-1 ORF are conserved in HSV-2 (Fig. 2); Hall et al. (1982) precisely mapped the common 5' terminus of two candidate HSV-1 transcripts of 1-8 and 3-8 kb, and located sequences recognizable as CAAT and TATA boxes. The 3' termini of these transcripts lie downstream of the UL49 and UL48 ORFs.

Thus, this ORF, which we designate UL49A, has the characteristics of a genuine protein-coding region. We consider that UL49A represents the only functional ORF in the sequence between HSV genes UL49 and UL50; no other ORF in this region possesses the features associated with protein-coding sequences. Counterparts of HSV ORF UL49A are present in the other sequenced alphaherpesviruses, varicella-zoster virus (VZV; Davison & Scott, 1986) and equine herpesvirus type 1 (EHV-1; Telford et al., 1992). A previously undescribed VZV coding region, designated 9A, and ORF 10 in EHV-1 overlap the oppositely oriented dUTPase ORF (VZV gene 8; EHV-1 gene 9) by 26 and 28 bases, and potentially encode proteins of 87 and 100 amino acids respectively. The four predicted alphaherpesvirus proteins show low conservation of aligned amino acid sequences (Fig. 3a). However, the hydrophobicity profiles of the proteins are similar and are characteristic of membrane proteins (Fig. 4). Specifically, the amino carboxytermini of each protein are highly hydrophobic. The former satisfies the criteria for a signal sequence as defined by McGeoch (1985), and the latter meets the requirements of a transmembrane anchor domain according to the parameters of Kyte & Doolittle (1982) and Eisenberg et al. (1984). The HSV-1, HSV-2 and EHV-1 sequences each contain at least one candidate cleavage site for removal of the signal sequence (von Heijne, 1986). The most probable cleavage site in the VZV protein, after residue 24, does not score above the accepted minimum value.

Significant sequence similarity is observed in the carboxy-terminal halves of the products encoded by pairs of viruses, most notably between those of HSV-1 and -2, and those of VZV and EHV-1; the HSV sequences are 76% identical, VZV and EHV-1 55%. Alignment of the sequences of all four alphaherpesvirus proteins reveals few conserved residues: one cysteine residue in the central hydrophilic domain, and phenylalanine, tyrosine and valine residues within the potential anchor domain.

The genomes of beta- and gammaherpesviruses contain counterparts to UL49A adjacent to the gene equivalents of HSV UL50. The human cytomegalovirus (HCMV) ORF UL73 (Chee et al., 1990) and the Epstein–Barr virus (EBV) ORF BLRF1 (Baer et al., 1984) are predicted to encode proteins of 138 and 102 amino acids respectively (Fig. 3b, c). Moreover, EBV transcripts for this gene have been mapped (Baer et al., 1984). The six herpesvirus proteins show poor conservation of amino acid sequence; only the cysteine residue within the central hydrophilic domain is invariant. However, the hydrophobicity profiles of these proteins are very similar (Fig. 4), with the HCMV and EBV proteins also possessing putative signal sequences and transmembrane domains. Therefore we consider that the UL49A family of genes in the alpha-, beta- and gammaherpesvirus genomes represents an additional member of the set of genes with counterparts in all sequenced herpesviruses.

The function of the UL49A protein in the virus replicative cycle is unknown and searches of the NBRF (release 31) and Swissprot (release 19) databases revealed no obvious non-herpesvirus counterpart. However, features of the amino acid sequence suggest that the product of the UL49A gene is a previously unrecognized virus membrane-spanning protein, and as such is a potential...
component of the virion envelope. The UL49A protein and its counterparts might be glycosylated at residues within the hydrophilic domain which are potentially exposed at the external surface of the membrane. Only the HCMV protein has potential N-glycosylation sites in this domain (see Fig. 3), but all contain serine and threonine residues which are potential sites for O-glycosylation. The HCMV protein, in particular, could be heavily O-glycosylated; its hydrophilic domain is at least 30 residues longer than the other herpesvirus proteins and 43% of amino acids in this region are serine or threonine. The proteins may also be palmitoylated at the inner surface of the membrane; all six herpesvirus proteins have cysteine residues close to the C terminus of the proposed anchor domain which could be putative sites of palmitoylation (Grand, 1989).

Fig. 3. Predicted amino acid sequences of HSV-1 and HSV-2 UL49A and counterparts in VZV, EHV-1, EBV and HCMV. (a) Alignment of the alphaherpesvirus amino acid sequences. The 'Consensus' indicates residues identical in all four proteins. Residues conserved in the HSV proteins are shown on the 'HSV con.' line, residues identical in the VZV and EHV-1 proteins on the line marked 'V/E con.' The amino acid sequences of the UL49A counterparts in EBV and HCMV are shown in (b) and (c), respectively. The potential sites for cleavage of the signal sequence (von Heijne, 1986) are indicated by arrowheads; filled arrowheads (▼) represent the most probable cleavage sites, empty arrowheads (▲) possible sites of cleavage. The only residue conserved between all six proteins is marked by an asterisk.
protein NC-7 (Cohen et al., 1980). However, direct amino acid sequence data have shown that NC-7 is encoded by the UL35 gene (Davison, et al., 1992).

The herpesviruses encode a number of integral membrane proteins, and detailed functional analyses of several of the major HSV glycoproteins have indicated that the virion envelope displays an array of structurally and functionally distinct viral proteins which, nevertheless, are highly interactive (Stannard et al., 1987; Johnson et al., 1988; Foa-Tomasi et al., 1991). The findings reported here raise the possibility that the UL49A family of herpesvirus membrane proteins may contribute to this complexity. Including UL49A, the HSV-1 genome is now considered to contain 73 distinct genes.

We thank L. J. E. Kattenhorn for help in preparing the text. B.C.B. is funded by SmithKline Beecham, E.A.R.T. by the Equine Virology Research Foundation.

References


Fig. 4. Hydropathy profiles (Kyte & Doolittle, 1982) of the UL49A protein of HSV-1 (a) and -2 (b) and the predicted counterparts from VZV (c), EHV-1 (d), EBV (e) and HCMV (f). The plots were obtained using the GCG program PEPPLOT and a moving window of 12 amino acids.
Short communication


(Received 19 February 1992; Accepted 30 April 1992)