A Comparative Analysis of the Sequence of the Thymidine Kinase Gene of a Gammaherpesvirus, Herpesvirus Saimiri

By R. W. HONESS,* M. A. CRAXTON,† L. WILLIAMS and U. A. GOMPELS‡

Division of Virology, National Institute for Medical Research, Mill Hill, London NW7 1AA, U.K.

(Accepted 27 June 1989)

SUMMARY

We present the nucleotide sequence of a region from the genome of the A + T-rich gammaherpesvirus, herpesvirus saimiri (HVS), which includes the coding sequences for the viral thymidine kinase (TK) gene. The organization of genes in this region resembles the homologous region of the Epstein–Barr virus (EBV) genome and is very compact, using overlapping coding sequences and with nucleotides shared by initiation and termination codons of neighbouring reading frames. The HVS TK is predicted to contain a 527 residue polypeptide with the first part of the presumptive nucleotide-binding site [(L, I, V)(F, Y)(I, L)(D, E)(G)(X)(X)(G)(L, I, V, M)(G)(K)(T, S)(T, S)] located at residues 212 to 224. This motif is close to the amino terminus of the TK polypeptides of alphaherpesviruses and the polypeptides of the cellular and poxvirus-encoded enzymes. The corresponding reading frame of the human gammaherpesvirus (EBV) also has a long amino-terminal extension but significant amino acid sequence similarities between the HVS and EBV sequences are not observed until the region of the nucleotide-binding site. Comparisons of these homologous carboxy-terminal sequences of the HVS- and EBV-encoded proteins with those from six alphaherpesviruses and proteins encoded by Marek’s disease virus (MDV) and the herpesvirus of turkeys (HVT) confirm that the HVS and EBV sequences are products of a distinct lineage. The sequences of the MDV and HVT encoded enzymes are significantly more similar to sequences of alphaherpesvirus enzymes than to those of HVS and EBV. Comparison of these 10 highly divergent TK sequences extends and refines the identification of essential features of this family of herpesvirus enzymes and defines 19 positions at which all sequences have identical residues.

INTRODUCTION

The replication of progeny genomes during productive cycles of herpesvirus growth involves the biosynthetic activities of a core of indispensable replicative proteins (e.g. DNA polymerase, DNA-binding proteins: Wu et al., 1988; McGeoch et al., 1988) which act on pools of precursors supplied or supplemented by the activities of a range of ancillary virus-specified enzymes. A number of these latter enzymes have substrate specificities which differ significantly from the analogous enzymes of uninfected cells and consequently provide potential targets for antiviral chemotherapy. The thymidine kinase (TK) activity of herpes simplex virus type 1 (HSV-1) currently provides the best example of the exploitation of such a target. The unusual substrate specificity of this enzyme includes nucleoside analogues such as acyclovir, which are not

* Present address: Howard Hughes Medical Institute, Duke University Medical Center, Box 3054, Durham, North Carolina 27710, U.S.A.
† Present address: Division of Virology, Department of Pathology, Addenbrooke’s Hospital, Hills Road, Cambridge CB2 2QQ, U.K.
substrates for the cellular enzyme (Elion et al., 1977; Kit, 1985). Moreover, the HSV-1 enzyme also has thymidylate kinase activity (Chen & Prusoff, 1978; Chen et al., 1979, 1984) and can also phosphorylate a number of analogue monophosphates to the corresponding diphosphates (Fyfe, 1981; Chen et al., 1984). The distinctive substrate specificity of the herpesvirus enzyme originates in an amino acid sequence which differs markedly from that of other known thymidine kinases. Comparisons of the sequences of several herpesvirus enzymes with those of higher eukaryotes and of the poxviruses have shown that the herpesvirus TK enzymes are not homologous to those of the higher eukaryotes and poxviruses (e.g. Kit, 1985; Boyle et al., 1987), with which they share only those convergent features common to a range of phylogenetically and functionally diverse nucleotide-binding proteins (Walker et al., 1982; Wu et al., 1988; Gorbalenya et al., 1989). However, although the available sequences of herpesvirus enzymes (see Kit, 1985; Robertson & Whalley, 1988, and this paper) show that portions of these proteins are recognizably homologous, the enzymes are highly divergent, resulting in significant differences in sensitivity of a range of herpesviruses to nucleoside analogues (Fyfe, 1981; Kit, 1985; Kit et al., 1987; Machida, 1986), therefore these differences are of practical as well as of theoretical interest.

We are interested in the relationship between the evolution of the molecular and biological properties of the herpesviruses and have previously shown that the A + T-rich gammaherpesvirus, herpesvirus saimiri (HVS), specifies a TK activity which also accepts iododeoxyuridine and bromoethyldeoxyuridine as substrates but not acyclovir or iododeoxycytidine (Hones et al., 1982). In this paper, we present the nucleotide sequence of the region of the HVS genome which encodes the viral TK gene and also a comparison of the predicted sequence of the encoded protein with proteins encoded by the Epstein–Barr virus (EBV), by a range of alphaherpesviruses and proteins predicted to be encoded by Marek's disease virus (MDV) and the herpesvirus of turkeys (HVT). The results of these comparisons have shown that, even with respect to TK proteins, the products of HVS and those of EBV share features which are not found in enzymes of the alphaherpesviruses or in products of MDV and HVT.

METHODS

**DNA sequencing.** The DNA for sequencing was prepared from the cloned EcoRI B fragment of HVS strain 11[Onc] DNA (KK102; Knust et al., 1983). Subclones of the MspI I (4.8 kbp) and PstI M (3.4 kbp) subfragments of EcoRI B were used to prepare random libraries by sonication, end-repair and cloning into the SmaI site of M13mp18. In addition, the EcoRI-PstI fragment of EcoRI B was cloned directly into M13mp18 and M13mp19 cut with EcoRI and PstI (see Gompels et al., 1988a, for maps of PstI sites in EcoRI B). The DNA from recombinant single-stranded phage was isolated and sequenced using the dideoxynucleotide chain-termination method, using [35S]ATP (Amersham) as the radioactive label (Bankier & Barrell, 1983), with the labelled products fractionated on Tris–borate–EDTA buffer gradient gels of 6% polyacrylamide (Biggin et al., 1983).

**Assembly and analysis of the DNA sequence and analysis of protein sequences.** Sequence data were compiled and analysed on a DEC 20-60 (Digital Equipment Corporation) computer using the programs of Staden (1982, 1984, 1986) as previously described (Gompels et al., 1988a, b). Database searches and dot matrix comparisons were performed with a modification of the procedure of Lipman & Pearson (1985) and with a dot matrix program similar to DIAGON of Staden (1986), implemented in the Molecular Genetics and Sequencing (MGS) package at the National Institute for Medical Research, London (B. Greer, P. Gillett & R. Mott). Multiple alignments of protein sequences were compiled using the program MULTAL, described by Taylor (1988) and operated on a μVAX computer and pairwise comparisons of the natural protein sequences and randomized versions of these sequences were made by using the MGS programs. The measures of protein divergence obtained (see Table 1 and text) were used to construct a ‘tree’ of inter-relatedness using the programs in the PHYLIP package distributed by Dr Joseph Felsenstein.

RESULTS AND DISCUSSION

**The DNA sequence and its interpretation**

Fig. 1 indicates the location and orientation of the major open reading frames (ORFs) within a 5 kbp region of the EcoRI B (MspI I) fragment of the 'light' (L) DNA component of HVS strain 11[Onc] DNA as deduced from the sequence given in Fig. 2, together with the sequence of the HVS homologue of the glycoprotein H gene (Gompels et al., 1988b). The sequence shown in
Fig. 1. Location and organization of the HVS TK gene. (a) Maps of restriction endonuclease cleavage sites and the resulting fragments within the 111 kb L DNA component of HVS strain 1 l[Onc] (Stamminget et al., 1987). Note that the orientation given here is inverted relative to the conventional orientation of the HVS genome such that genes common to HVS and EBV are shown in the same relative order and orientation as the orientation of the EBV genome given by Baer et al. (1984) (see Cameron et al., 1987; Gompels et al., 1988a). (b) Organization of major ORFs in the MspI I fragment of the HVS-1 l[Onc] genome as deduced from the sequence given in Fig. 2, together with the sequence of the glycoprotein H homologue of HVS (gH; Gompels et al., 1988b). Each ORF is indicated by an arrowed box directed from the proposed initiation codon (ATG) to the proposed termination codon and use of overlapping sequences for the initiation and termination codons of the TK ORF (hatched) and the initiation codons of neighbouring reading frames are indicated. Occurrences of the AATAAA polyadenylation signal on each strand of the sequence are indicated by the half-barbed arrows above (strand going right) or below (strand going left) for the ORFs. (c) Organization of major ORFs encoded by the region from nt 146000 to 147200 of the genome of the B95-8 strain of EBV (Baer et al., 1984). The two ORFs directed towards the left (BXLF2, BXLF1) and the two ORFs directed towards the right (BXRFR1, BVRF1) are each homologous to the HVS reading frames with which they are aligned in this figure.

Fig. 2 consists of a 2880 nucleotide segment of EcoRI B, beginning at the EcoRI site that separates EcoRI F and B (Fig. 1). This sequence has a mean mononucleotide composition of 36.5% (G + C) and an observed frequency of 1.42% of CpG dinucleotides, relative to the random expectation of 3.33% of these values are close to those observed for a large sample of sequences from the L DNA component of HVS DNA (Honess et al., 1989). The entire sequence of Fig. 2 is predicted to code for protein; it contains two complete ORFs and parts of two others (Fig. 1, Fig. 2), all of which conform to the general pattern of biased codon usage observed in studies of other genes from HVS (Cameron et al., 1987; Nicholas et al., 1988). All these HSV ORFs have collinear homologues encoded by the region from nt 140600 to 147200 of the genome of the B95-8 strain of EBV (Baer et al., 1984). One complete ORF [nucleotides (nt) 1122 to 214] and part of another (nt 581 to 1) overlap on the reverse strand of the sequence shown and the predicted products of these ORFs are homologous to products of the EBV reading frames BXRF1 and BVRF1, respectively (homologues of reading frames 35 and 34 of varicella-zoster virus (VZV) (Davison & Scott, 1986; Davison & Taylor, 1987) and of UL24 and UL25 of HSV-1 (McGeoch et al., 1988; Jacobson et al., 1989)). The major ORF on the forward strand (nt 1121 to 2701) is the homologue of EBV reading frame BXLFF1 (encoding TK (Littler et al., 1986); reading frame 56 of VZV and UL23 of HSV-1) and the proposed amino terminus of the HVS...
Fig. 2. Sequence of 2880 nt from the L DNA component of HVS-1 (Onc) which includes the sequences coding for the HVS TK gene. The sequence is shown as the message-sense strand for the HVS TK gene for the region beginning at the EcoRI restriction site that separates EcoRI B and F and extends into MspI I (Fig. 1).

Amino acid sequences predicted for products of major ORFs are indicated in the single-letter code above the sequence for the strand shown (i.e. HVS TK gene, beginning at nt 1121 and the gH gene beginning at nt 2701) and below the sequence for the complementary strand (HVS-encoded homologues of EBV genes BXRF1, beginning at nt 1122 and BVRF1, beginning at nt 581). Positions of some restriction endonuclease recognition sites and occurrences of the polyadenylation signal (AATAAA) are indicated above the sequence. Note that the initiation codons proposed for the HVS TK gene and for the HVS homologue of BXRF1 on the opposite strand utilize complementary nucleotides (1121 and 1122) and that the C-terminal and termination codons for the HVS TK ORF overlap the amino-terminal codons proposed for the HVS gH gene (nt 2701 to 2704; see also Fig. 1).
Fig. 3. Alignments of the amino acid sequences predicted for the TK polypeptides of HVS and EBV with the homologous regions of TK genes of alphaherpesviruses and of MDV and HVT. Sequences were aligned using a multiple sequence alignment program (Taylor, 1988). Positions at which identical residues are observed in eight or more of the 10 aligned sequences are shaded and positions at which
Herpesvirus thymidine kinase

homologue of EBV BXL2F2 [encoding glycoprotein H (Gompels et al., 1988b)] begins at nt 2701. This region of the HVS genome is very densely packed, using common nucleotides for initiation of divergent reading frames (nt 1121 and 1122) and for the carboxy-terminal amino acid and termination codons of one reading frame with the proposed amino-terminal codons of the next (i.e. nt 2701 to 2704) as well as one extended region of overlap between reading frames (nt 581 to 211). These particular features of the arrangements of initiation and termination codons are very similar to the arrangement of the homologous reading frames in the genome of EBV (Fig. 2) and a compact arrangement is also observed in HSV-1 and VZV (McGeoch et al., 1988; Jacobson et al., 1989). Conservation of these features seems noteworthy in contrast to the significant changes within parts of the coding sequences. For example the amino-terminal portions of EBV BXL1F1 and of the HVS TK ORF are each extended relative to the other TK ORFs, but these extensions are not significantly similar in sequence. In addition the HVS homologue of EBV BXRF1 contains a 70 residue insertion relative to the EBV ORF (data not shown).

The compact arrangement of these genes also means that the regulatory sequences for one gene are embedded in the coding sequences of another.

Comparisons of the sequence predicted for the HVS TK polypeptide with TK polypeptides of other herpesviruses

Comparisons of the entire amino acid sequence of the HVS TK with that of EBV or with TK sequences of HSV-1 and VZV show that the region from residue 200 to the carboxy terminus of the HVS sequence is homologous to the region from residue 280 to the carboxy terminus of EBV BXL1F1 and from residues 45 or 10 to the carboxy termini of the HSV-1 and VZV enzymes (data not shown). This region has previously been suggested as the 'minimal' herpesvirus TK gene (Halpern & Smiley, 1984; Haarr et al., 1985; Haarr & Flatmark, 1987).

The sequences of TK genes from more than 10 different herpesviruses have now been determined, representing the largest collection of homologous genes from these viruses now available for comparison. Such comparisons are of value in identifying conserved and divergent features of sequence and structure to guide studies of function and substrate specificity (Darby et al., 1981, 1986; Liu & Summers, 1988). They may serve also as an example of a set of relatively divergent products which may be informative in attempts to deduce evolutionary relationships between herpesviruses. Fig. 3 displays an optimized mutual alignment of the homologous regions of the predicted TK sequences of HVS with those of nine other herpesviruses. The highly conserved or invariant motifs which have been identified previously, i.e. motif I which is residues 5 to 17 of the aligned sequences [(L/I/V)(F/Y)(L/D/E)(G)(X)(X)(G)(L/I/V/M)(G)(K)(T/S)(T/S)], motif II, residues 39 to 44 [(E)(P)(L/I/M)(X)(Y)(W)] and motif III, residues 123 to 138 [(L/I/V/M)(L/I/V/M)(X)(D)(R)(H)(X)(L/I/V)(A/S)(A)(X)(L/I/V)(L/I/V)(C/V)(F/Y)(P)] are clearly defined, as is a conserved glycine (164) followed by patterns of conserved hydrophobic and charged residues before three conserved arginines [181 to 190; (R)(L/I/V)(X)(R)(X)(R)(X)(X)(D/E)] and a further set of conserved hydrophobic residues (Y

identical residues are observed in all 10 sequences are indicated by filled circles below the alignment. The EBV sequence is from the BXL1F1 reading frame of the B95-8 strain of EBV (Baer et al., 1984), the HSV-1 sequence is the UL23 reading frame of strain MP17 (McGeoch et al., 1988), the HSV-2 sequence is of strain 333 (Swain & Galloway, 1983; Kit et al., 1983), the VZV sequence is RF46 of the Dumas strain (Davison & Scott, 1986), the sequence labelled MarHV in this paper is the TK sequence of the squirrel monkey alpha-herpesvirus, first isolated as herpesvirus tamarinus and referred to as the marmoset herpesvirus (marHV) by Kit and his colleagues (Otsuka & Kit, 1984; the sequence given here has been corrected to take account of a number of known errors in the sequence previously published by Dr Saul Kit, personal communication). The sequence of EHV-1 is from Robertson & Whaley (1988), BHV-1 from Mittal & Field (1989) and the sequences of the MDV and HVT enzymes are from Scott et al. (1989). The alignments suggested here identify conserved features similar to those of other recent studies, but we have minimized the use of ‘gaps’, some of which have been introduced in alternative alignments to match residues in pairwise comparisons; however comparisons with randomized sequences suggest that these are likely to be due to chance.
Table 1. *Measures of identity and similarity between homologous regions of herpesvirus TK genes*

<table>
<thead>
<tr>
<th></th>
<th>EBV</th>
<th>HSV-1</th>
<th>HSV-2</th>
<th>VZV</th>
<th>MarHV</th>
<th>EHV-1</th>
<th>BHV-1</th>
<th>MDV</th>
<th>HVT</th>
</tr>
</thead>
<tbody>
<tr>
<td>HVS</td>
<td>37.4</td>
<td>15.0</td>
<td>16.8</td>
<td>18.0</td>
<td>16.5</td>
<td>21.4</td>
<td>15.4</td>
<td>19.5</td>
<td>16.0</td>
</tr>
<tr>
<td></td>
<td>(54.9)</td>
<td>(33.0)</td>
<td>(33.2)</td>
<td>(33.3)</td>
<td>(36.0)</td>
<td>(30.2)</td>
<td>(35.4)</td>
<td>(30.2)</td>
<td>(30.2)</td>
</tr>
<tr>
<td></td>
<td>18.3</td>
<td>19.6</td>
<td>18.2</td>
<td>15.6</td>
<td>18.8</td>
<td>17.0</td>
<td>18.6</td>
<td>16.4</td>
<td></td>
</tr>
<tr>
<td>EBV</td>
<td>(31.6)</td>
<td>(33.6)</td>
<td>(35.3)</td>
<td>(31.1)</td>
<td>(34.7)</td>
<td>(28.2)</td>
<td>(35.6)</td>
<td>(30.0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29.0</td>
<td>28.6</td>
<td>40.1</td>
<td>36.9</td>
<td>30.7</td>
<td>28.4</td>
<td>26.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSV-1</td>
<td>(85.8)</td>
<td>(43.5)</td>
<td>(53.7)</td>
<td>(53.8)</td>
<td>(44.4)</td>
<td>(47.3)</td>
<td>(45.4)</td>
<td>(45.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29.3</td>
<td>42.1</td>
<td>36.6</td>
<td>32.7</td>
<td>26.5</td>
<td>26.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSV-2</td>
<td>(46.4)</td>
<td>(57.1)</td>
<td>(55.6)</td>
<td>(45.1)</td>
<td>(46.3)</td>
<td>(46.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>28.1</td>
<td>37.6</td>
<td>27.2</td>
<td>25.7</td>
<td>24.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VZV</td>
<td>(44.1)</td>
<td>(54.0)</td>
<td>(42.3)</td>
<td>(43.3)</td>
<td>(43.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>38.4</td>
<td>33.7</td>
<td>27.0</td>
<td>24.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MarHV</td>
<td>(51.9)</td>
<td>(45.8)</td>
<td>(43.4)</td>
<td>(43.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>34.0</td>
<td>30.3</td>
<td>29.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EHV-1</td>
<td>(45.7)</td>
<td>(47.1)</td>
<td>(45.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>23.9</td>
<td>24.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BHV-1</td>
<td>(38.6)</td>
<td>(37.0)</td>
<td>(59.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* In each case the number of identical and of identical and similar (in parentheses) amino acids obtained in pairwise comparisons between the homologous regions of the TK sequences as aligned in Fig. 3 are indicated as a percentage of the number of residues in the pairwise comparison. For example, the total number of positions in the aligned comparison of the EBV and HVS sequences is 326, 122 of these positions are occupied by identical residues and 57 are occupied by similar residues. Similar residues were assigned as members of the following sets: L = I = V = A = M; R = K = H; T = S; F = Y = W; D = N; E = Q; P; G; C. Estimates of the significance of these measures of identity and similarity were made by scoring the number of identical and similar residues in alignments of each of these TK sequences (Fig. 3) with 10 or 20 randomly permuted versions of the other sequences. These comparisons gave mean identities of 9 to 10% and mean similarities of 14 to 16%. For example, comparisons of the EBV sequence (283 to 607 nt) with 10 randomly permuted sequences from residues 208 to 527 nt of the HVS sequence gave a mean identity of 9.0% with a s.d. of $\pm 1.1\%$.

Quantitative expressions of similarity and a TK 'tree'

Comparisons between all possible pairs of sequences in the alignment provides an internally consistent measure of their relative degrees of similarity; a summary of the results of such comparisons is given in Table 1. It is clear that although divergent the HVS- and EBV-encoded proteins are much more similar to each other than they are to proteins of any of the other viruses and are no more similar to the proteins of MDV or HVT than they are to those of the alphaherpesviruses. Moreover, the proteins of MDV and HVT which are relatively closely related are more similar to some alphaherpesvirus sequences [e.g. MDV compared with equine herpesvirus type 1 (EHV-1), 30.3% identity] than some alphaherpesviruses are to other alphaherpesviruses [e.g. VZV compared with bovine herpesvirus type 1, (BHV-1), 27.1% identity]. Measurements of the number of identical and similar residues in alignments of each of the observed sequences with randomized versions of the others have shown that a mean of about 9% of residues are identical in such comparisons (i.e. $I_{ran} = 9\%$ with a standard deviation of about 1%). Thus the most distantly related sequences (HVS and HSV-1, 15% observed identity, $I_{obs}$, HVS and HVT, 16% identity) have only 6 to 7% of identical residues in excess of those attributable to chance (i.e. the 'effective' identity; $I_{eff} = I_{obs} - I_{ran} = 6$ to 7%). Almost all of these residues are accounted for as parts of the conserved motifs noted above.
Fig. 4. A tree displaying the relative divergence of the sequences of homologous portions of herpesvirus TK polypeptides (as aligned in Fig. 3). Sequences of proteins from each virus are located at the tips of the branches of a tree, with the distances separating each tip calculated to give the best fit to the simple percentage of non-identical residues observed in the corresponding pairwise comparisons of the aligned sequences (i.e. the distances shown are \(100 - \frac{x}{100}\) identical residues; values for the \(\%\) identical residues are given in Table 1). The lengths of the branches and the positions of the nodes were obtained by minimizing the square of the deviations of these branch lengths from the values obtained in pairwise comparisons, using a program in the PHYLIP suite distributed by Dr Joseph Felsenstein. The virus sequences at the ends of each branch are shown separated on a plane, one axis of which represents the mean mononucleotide composition of the corresponding virus genome \(x\) axis; mean \(\%\)(G + C) and the other axis the ratio of the observed to the expected frequencies of CpG dinucleotides in these genomes \(z\) axis; CpG). The genomes of the alphaherpesviruses and of MDV and HVT all have observed frequencies of CpG dinucleotides which are close to those expected from their mononucleotide compositions and the corresponding tips are located at points on the \(z\) axis for which observed CpG/expected CpG > +1.0. In contrast, genomes of HVS and of EBV are each relatively deficient in CpG dinucleotides \(i.e.\) expected CpG/observed CpG is \(-3.0\) for HVS and \(-1.8\) for EBV: see Honess et al. (1989) for a fuller discussion of deviations of observed from expected frequencies of CpG dinucleotides in herpesvirus genomes.

We have used the data of Table 1 as the basis for computing a ‘tree’ to provide a pictorial representation of the mutual degrees of relationship between these TK sequences. For this purpose we have used the simple percentage of non-identical residues \((100 - I_{ob})\%\) as the measure of distance, rather than the more usual transformation to an evolutionary distance matrix \(D = -\ln I_{ev}\). We have discussed elsewhere some of the evidence that the mutational influences on the different members of the family are likely to be biased by independent mechanisms (Honess, 1984; Honess et al., 1989) and therefore, notwithstanding some interesting attempts (Gentry et al., 1988), we do not believe there is currently any valid method to calibrate the time scales for independent branches of such a ‘tree’ of relationships. However this form of display does allow the ready assessment of the major subgroupings of related sequences. The separation of the HVS and EBV sequences from the main cluster is evident, as is the linkage of the HVT/MDV branch as an outlying member of this main sequence group. It should perhaps be emphasized that many of these sequences are at comparably large distances one from another. Therefore at present the best ‘tree’ may not differ significantly from other
trees with differing placements of some of the closely spaced nodes from the main branch of the tree. It will be of interest to compare the topology of the present 'TK tree' with those based on comparisons between less highly divergent proteins when a sufficiently representative data set is available.

We are most grateful to Dr Saul Kit for providing us with a corrected version of his sequence of the TK gene from the squirrel monkey alphaherpesvirus (‘marmoset herpesvirus’, herpesvirus tamarinus, saimiriine herpesvirus-1) and Simon Scott for a prepublication copy of his sequences for the TK genes from MDV and HVT. Willie Taylor and Richard Mott provided assistance with aspects of the construction and phylogenetic analysis of the multiple sequence alignments.

REFERENCES


CAMERON, K. R., STAMMINGER, T., CRAXTON, M., BODEMER, W., HONSEN, R. W. & FLECKENSTEIN, B. (1987). The 160,000-Mr virion protein encoded at the right end of the herpesvirus saimiri genome is homologous to the 140,000-Mr membrane antigen encoded at the left end of the Epstein-Barr virus genome. Journal of Virology 61, 2063–2070.


Herpesvirus thymidine kinase

internal AUG codons for initiation of protein synthesis directed by mRNAs from normal and mutant genes encoding herpes simplex virus thymidine kinase. *Journal of Virology* 56, 512–519.


(Received 21 April 1989)