Identification of homologues to the human cytomegalovirus US22 gene family in human herpesvirus 6

Stacey Efstathiou,1* Glenda L. Lawrence,2† Carol M. Brown2 and Bart G. Barrell2

1Division of Virology, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP and 2MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

The sequence of 10079 bp corresponding to the overlapping Sall H and Smal G restriction fragments of the genome of human herpesvirus 6 (HHV-6) strain U1102 was determined. The sequence contains six complete open reading frames (ORFs) and two incomplete ORFs located at the 5' and 3' ends of the Sall H and Smal G fragments respectively. Seven of these ORFs have recognizable homologues only in the betaherpesvirus human cytomegalovirus (HCMV), no obvious counterparts being detectable in the genomes of the human alphaherpesviruses, varicella-zoster virus and herpes simplex virus type 1 or the gammaherpesvirus Epstein-Barr virus. The DNA sequenced is located proximal to the left repeat of the HHV-6 genome outside the well recognized region encompassing conserved herpesvirus gene blocks. A close collinear relationship is evident between the HHV-6 ORFs identified in this study and their counterparts in HCMV, ORFs UL23, UL24 and UL27 to UL31. Four of the HHV-6 ORFs, SHL1, SHL2, SFL1 and SSL2, are related to members of the HCMV US22 family of proteins, which are themselves tandemly arranged and located predominantly within the unique short and the left end of the unique long region of the prototype HCMV strain AD169 genome. Two adjacent HHV-6 ORFs, SSL1 and SHL3, are related to HCMV UL27. The identification of this gene set in addition to the HHV-6 ORFs with amino acid sequence similarity to the HCMV US22 family indicates a particularly close relationship between these two human herpesviruses, and suggests that the clustering of these related tandemly arranged genes may be a general feature of betaherpesvirus-type genomes.

Introduction

Human herpesvirus 6 (HHV-6) was first isolated following the in vitro cultivation of peripheral blood lymphocytes from six patients with lymphoproliferative disorders, two of whom were infected with human immunodeficiency virus (HIV) (Salahuddin et al., 1986). Subsequent isolates, which have been shown to be closely related to the original GS isolate by Southern blot hybridization, have been recovered from the peripheral blood of HIV-infected patients from Uganda (strain U1102) (Downing et al., 1987), the Gambia (strain AJ) (Tedder et al., 1987) and Zaire (strain Z29) (Lopez et al., 1988), and the saliva of healthy adults (Harnett et al., 1990; Levy et al., 1990; Fox et al., 1990). Serological studies have shown that more than 80% of healthy adults have antibodies to the virus and that seroconversion commonly occurs within the first 2 years of life (Okuno et al., 1989; Levy et al., 1990; Briggs et al., 1988). Primary HHV-6 infection has been shown to cause exanthem subitum (roseola infantum; sixth disease) (Takahashi et al., 1988; Yamanishi et al., 1988), a common mild febrile illness normally acquired within the first 2 years of life. As yet, no specific disease has been attributed to HHV-6 in adults, although there have been reports of infectious mononucleosis-like illness, lymphadenopathy and mild hepatitis (Niederman et al., 1988; Irving & Cunningham, 1990; Steeper et al., 1990), and the virus may cause complications in organ transplant recipients (Ward et al., 1989; Okuno et al., 1990; Morris et al., 1989; Steeper et al., 1990).

HHV-6 has CD4+ T cell tropism both in vitro (Lusso et al., 1988; Takahashi et al., 1989) and in vivo (Takahashi et al., 1989), resulting in the lytic infection of such cells. As yet there is no evidence for the ability of HHV-6 to establish a latent non-productive infection of lympho-
cyttes, which is a major feature of the well studied lymphotropic gammaherpesviruses Epstein–Barr virus (EBV) and herpesvirus saimiri (HVS), and recent evidence suggests that monocytes may be an important reservoir of latent HHV-6 (Kondo et al., 1991) and its closest genetic relative, the prototypic betaherpesvirus human cytomegalovirus (HCMV) (Taylor-Wiedeman et al., 1991). Restriction endonuclease mapping of HHV-6 strain U1102 has shown the genome to consist of a dsDNA molecule of approximately 161 kbp composed of 141 kbp of largely unique sequences flanked by 10 kbp direct repeats (Martin et al., 1991a), a structure which differs from that determined for the five previously characterized human herpesviruses and which is similar to that proposed for the Z29 isolate of HHV-6 (Lindquester & Pellett, 1991). The genome has a mean G+C content of 43 to 44% as estimated by buoyant density cesium chloride centrifugation (Lopez & Honess, 1990). Nucleotide sequence analysis of approximately 22 kbp of the HHV-6 (strain U1102) genome (Lawrence et al., 1990) has identified a number of open reading frames (ORFs) which are conserved among the human representatives of the alpha-, beta- and gammaherpesvirus subgroups, and has demonstrated that HHV-6 is more closely related to HCMV, a betaherpesvirus, than to the neurotropic alphaherpesviruses herpes simplex virus (HSV) and varicella-zoster virus (VZV), or the lymphotropic gammaherpesvirus EBV. In this paper we present the nucleotide sequence of 10079 bp of HHV-6 strain U1102, and define a region of the HHV-6 genome which contains ORFs which have recognizable homologues only in the HCMV genome and which are absent from the genomes of the human alpha- and gammaherpesviruses.

Methods

Two overlapping cloned restriction fragments, SalI H and SmaI G, of HHV-6 strain U1102 (Martin et al., 1991a) were chosen for nucleotide sequence analysis. The 6.0 kb SalI H fragment was cloned into pUC13 (pSE6); the 5.2 kb SmaI G fragment cloned into pUC13 (pSMD5) was kindly provided by R. W. Honess. Purified insert DNA was sonicated following self-ligation to produce randomly sheared fragments (Deininger, 1983). These were then cloned into the SmaI site of bacteriophage M13mp8 (Messing & Vieira, 1982). Single-stranded DNA templates were prepared as described previously (Sanger et al., 1980) and sequenced by the dideoxynucleotide chain termination method (Sanger et al., 1977) as described by Bankier et al. (1987).

Sequence data were assembled by using the computer programs DBAUTO and DBUTIL (Staden, 1980, 1982, respectively), and analysed for the presence of ORFs using the programs NIP (Staden, 1990a), and DIANA (J. Crooke, T. Horsnell & B. Barrell, unpublished). Protein sequences were analysed using the PIP program (Staden, 1990a), and predicted translation products were compared against a library of herpesvirus protein sequences and the Protein Identification Resource protein library (George et al., 1986) using the program FASTA (Pearson & Lipman, 1988). Pairwise sequence comparisons were performed using MULTALIGN (Barton & Sternberg, 1987). All computer programs were run on DEC VAX and micro VAX computers.

Results

The DNA sequence of a 10079 bp region of the HHV-6 strain U1102 genome was determined for both strands, each base being sequenced an average of six times. The overall G+C content of the sequence is 47.5% and the observed CpG dinucleotide frequency does not differ significantly from that expected on the basis of random associations between mononucleotides (data not shown). This observation is consistent with those for other sequenced regions of the U1102 genome (Lawrence et al., 1990; Martin et al., 1991b; Teo et al., 1991), with the exception of a localized region of CpG dinucleotide suppression proximal to the putative HHV-6 immediate early gene locus (Martin et al., 1991b), and is in contrast to eukaryotic DNA and the genomes of the gammaherpesviruses EBV and HVS, which have an overall CpG dinucleotide deficiency (Bird, 1980; Honess et al., 1989; Gompels et al., 1988). A search for restriction enzyme cleavage sites revealed a small (242 bp), previously unmapped SmaI restriction fragment located between the SmaI J and Q fragments of the published restriction maps of strain U1102 (Martin et al., 1991a); we have designated this fragment SmaI S (Fig. 1).

The sequenced region, which lies adjacent to the left terminal repeat of the genome (Martin et al., 1991a), contains two partial and six complete ORFs (Fig. 1) predicted to encode protein on the basis of both positional base preference and codon usage methods (Staden, 1990b). The protein sequences of the predicted ORFs are shown relative to the nucleotide sequence in Fig. 2. A summary of the location of the ORFs together with the positions of putative TATA boxes and transcription start sites are shown in Table 1. The location of TATA boxes is clearly speculative and is complicated by the possibility of RNA splicing occurring across this region of DNA. This situation will be resolved by future mRNA mapping studies.

Homology of HHV-6 ORFs to those of HCMV

Each of the eight amino acid sequences was screened against a library of herpesvirus protein sequences and the Protein Identification Resource protein library (George et al., 1986) using the computer program FASTA (Pearson & Lipman, 1988) with a K-tuple value of 1. An optimized FASTA score of greater than 100 was considered to indicate a significant degree of amino acid sequence similarity. Of the eight HHV-6 ORFs, seven
Human cytomegalovirus US22 homologues

Fig. 1. Restriction map of the region sequenced showing Smal and SalI restriction sites. The DNA sequence was determined from the shaded restriction fragments, SalI H and Smal G, based on the restriction endonuclease cleavage maps of strain U1102 (Martin et al., 1991a). The position of termination codons in each of the three possible reading frames for each strand of the region is indicated by bars, the arrows showing the location and direction of the major ORFs. ORF designations are based on the SalI restriction fragment in which ORF start codons are located and the direction of transcription, and are numbered from left to right for each strand.

Table 1. Summary of data: ORFs, putative translation start sites, TATA consensus sequences, lengths and Mr,s of predicted translation products

<table>
<thead>
<tr>
<th>Name</th>
<th>ORF start</th>
<th>ORF end*</th>
<th>ATG position</th>
<th>ATG context</th>
<th>TATA position</th>
<th>TATA sequence</th>
<th>Length (amino acids)</th>
<th>Mr,</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHL1</td>
<td>1102</td>
<td>(2)†</td>
<td>1072</td>
<td>AAAATGT</td>
<td>1133</td>
<td>TATATT</td>
<td>357†</td>
<td>&gt;39-0K</td>
</tr>
<tr>
<td>SHL2</td>
<td>2574</td>
<td>1414</td>
<td>2532</td>
<td>GTAATGG</td>
<td>2610</td>
<td>TATATTA</td>
<td>373</td>
<td>43.7K</td>
</tr>
<tr>
<td>SHL3</td>
<td>4351</td>
<td>2744</td>
<td>4348</td>
<td>GTGATGG</td>
<td>4366/4574</td>
<td>TATATA/TATAAT</td>
<td>535</td>
<td>62.0K</td>
</tr>
<tr>
<td>SSL1</td>
<td>6233</td>
<td>4473</td>
<td>?</td>
<td>?</td>
<td>587</td>
<td>67.8K</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSL2</td>
<td>7258</td>
<td>6167</td>
<td>7192</td>
<td>TTATAGT‡</td>
<td>7214/7473</td>
<td>TATTTA/TATATT</td>
<td>342</td>
<td>39-4K</td>
</tr>
<tr>
<td>SFL1</td>
<td>8476</td>
<td>7280</td>
<td>8347</td>
<td>AAGATGG</td>
<td>8400</td>
<td>TATATT</td>
<td>356</td>
<td>41.1K</td>
</tr>
<tr>
<td>SFL2</td>
<td>8823</td>
<td>8497</td>
<td>8808</td>
<td>GATATGG</td>
<td>8970</td>
<td>GATATTA</td>
<td>104</td>
<td>11.8K</td>
</tr>
<tr>
<td>SFR1</td>
<td>8854</td>
<td>(10077)†</td>
<td>8860</td>
<td>CAGATGG</td>
<td>8622</td>
<td>TATATAA</td>
<td>406†</td>
<td>&gt;46-7K</td>
</tr>
</tbody>
</table>

* Exclusive of stop codon.
† Incomplete ORF.
‡ ATG context does not conform to the consensus sequence RNNATG/NNNATGG (Kozak, 1984).

Of the eight HHV-6 ORFs identified in the current analyses, four (SHL1, SHL2, SSL2 and SFL1) are related to the HCMV US22 family of proteins and two (SHL3 and SSL1) are distantly related to HCMV UL27. Of the remaining HHV-6 ORFs, SFR1 is homologous to HCMV UL31 and represents the most highly conserved gene in the region sequenced. In contrast, the adjacent ORF SFL2, although analogous to HCMV UL30 in terms of its size, position and orientation, does not share significant amino acid sequence similarity with its putative HCMV counterpart (Table 2). The partially sequenced HHV-6 ORF SFR1 and HCMV ORF UL31...
encode six and three potential N-linked glycosylation sites respectively, but no obvious hydrophobic transmembrane anchor sequences are evident in either protein sequence, suggesting that these related genes are unlikely to encode membrane glycoproteins.

**HHV-6 ORFs with homology to the HCMV US22 gene family**

The HHV-6 ORFs SHL1, SHL2, SSL2 and SFL1 are related to one or more members of the HCMV US22 gene family, which consists of ORFs US22, US23, US24, US26, IRS1/TRS1, UL23, UL24, UL28, UL29, UL30, UL31 and UL36 (Weston & Barrell, 1986; Kouzarides et al., 1988; Chee et al., 1990). A summary of the overall amino acid sequence similarity between these proteins is shown by the FASTA score matrix (Table 3). In the case of HHV-6 ORF SHL1, the highest FASTA score (105) was Table 2. *Summary of optimized FASTA scores observed in comparison between HHV-6 ORFs and the homologous genes from HCMV*

<table>
<thead>
<tr>
<th>HHV-6 ORF</th>
<th>Identity (%)</th>
<th>FASTA score</th>
<th>HCMV ORF</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHL1†</td>
<td>20.9†</td>
<td>97</td>
<td>UL23</td>
<td>US22 family member</td>
</tr>
<tr>
<td>SHL2</td>
<td>18.6</td>
<td>137</td>
<td>UL24</td>
<td>US22 family member</td>
</tr>
<tr>
<td>SHL3</td>
<td>16.6</td>
<td>81</td>
<td>UL27</td>
<td>Most similar to HHV-6 SSL1</td>
</tr>
<tr>
<td>SSL1</td>
<td>17</td>
<td>165</td>
<td>UL27</td>
<td>Most similar to HHV-6 SHL3</td>
</tr>
<tr>
<td>SSL2</td>
<td>17.4</td>
<td>189</td>
<td>UL28</td>
<td>US22 family member</td>
</tr>
<tr>
<td>SFL1</td>
<td>20.1</td>
<td>154</td>
<td>UL29</td>
<td>US22 family member</td>
</tr>
<tr>
<td>SFL2</td>
<td>12.8§</td>
<td>346</td>
<td>UL30</td>
<td>Positional homologue</td>
</tr>
<tr>
<td>SFR1†</td>
<td>26.8‡</td>
<td>346</td>
<td>UL31</td>
<td></td>
</tr>
</tbody>
</table>

* Percentage identities were calculated using the SIP alignment program (Staden, 1990b).
† Incomplete ORF.
‡ For the incomplete ORFs SHL1 and SFR1, percentage identities represent values obtained from the FASTA alignment output for overlaps of 258 and 369 amino acids respectively.
§ NS, No significant homology.
Human cytomegalovirus US22 homologues

Table 3. FASTA† score matrix showing overall amino acid sequence similarity between US22 family members

<table>
<thead>
<tr>
<th></th>
<th>HCMV</th>
<th>HHV-6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>US22</td>
<td>US23</td>
</tr>
<tr>
<td>US22</td>
<td>—</td>
<td>461</td>
</tr>
<tr>
<td>US23</td>
<td>—</td>
<td>302</td>
</tr>
<tr>
<td>US24</td>
<td>—</td>
<td>206</td>
</tr>
<tr>
<td>US26</td>
<td>—</td>
<td>85</td>
</tr>
<tr>
<td>TRS1‡</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>UL23</td>
<td>—</td>
<td>116</td>
</tr>
<tr>
<td>UL24</td>
<td>—</td>
<td>142</td>
</tr>
<tr>
<td>UL28</td>
<td>—</td>
<td>92</td>
</tr>
<tr>
<td>UL43</td>
<td>—</td>
<td>*</td>
</tr>
<tr>
<td>SHL1</td>
<td>—</td>
<td>141</td>
</tr>
<tr>
<td>SHL2</td>
<td>—</td>
<td>97</td>
</tr>
<tr>
<td>SFL1</td>
<td>—</td>
<td>97</td>
</tr>
<tr>
<td>SSL2</td>
<td>—</td>
<td>97</td>
</tr>
</tbody>
</table>

† Optimized FASTA scores obtained using a K-tuple of 1.
‡ Asterisks indicate optimized FASTA scores < 80.
§ IRS1, which is identical to TRS1 over all conserved amino acid domains common to US22 family members, is excluded from this matrix.
¶ Exons 1 and 2 of UL36 were joined.

The overall similarity of FASTA scores between this partially sequenced HHV-6 ORF and other members of the HCMV US22 family has led us to assign this ORF tentatively as a positional homologue of HCMV UL23.

Comparing the HCMV US22 family and related HHV-6 proteins has proved difficult, largely owing to the low overall amino acid sequence similarity observed between family members and to their differing sizes, between 143 and 788 amino acids. However, certain features of this family of distantly related proteins are apparent. The HCMV family members encoded in the unique short (Us) region are most closely related to each other both in terms of overall amino acid sequence similarity and in length, between 500 and 603 amino acids, whereas, with the exception of HCMV UL43 [which is likely to represent part of a spliced gene product (Chee et al., 1990)], the HCMV unique long (UL) region-encoded and HHV-6 US22 family members are more closely related in terms of their size (342 to 379 amino acids), relative genomic location and overall amino acid sequence similarity. A multiple alignment of all HCMV and HHV-6 US22 family members assembled using MULTALIGN (Barton & Sternberg, 1987) (data not shown) identified three sequence motifs previously recognized in this family (Kouzarides et al., 1988; Chee et al., 1990). The previously unrecognized domain I consists of a GXXOXOXWP core motif, where O is any hydrophobic residue and X is any residue (Fig. 4a), and is found in all of the HCMV Us family members, UL23, -24, -28, -29, -36 and -43. This core motif is less well conserved in the HHV-6 US22 family members, with only ORF SHL2 encoding a complete motif, the remaining HHV-6 ORFs, SHL1, SSL2 and SFL1, encoding domains in which similarity is restricted to the final four amino acids (OXWP). Domain I is poorly conserved in the HCMV Us and TRS1/IRS members, the only recognizable similarity in the case of US22, -23 and -24 being restricted to a sequence lacking the conserved WP found in all HCMV Ul and HHV-6 members, and comprising the sequence motif OXOXPPXXW.

Domain II consists of the highly conserved motif OOCCXXXLXXOG (Kouzarides et al., 1988; Chee et al., 1990), which is found in all of the HCMV US22 family members except UL28, UL29 and TRS1/IRS. Of the three completely sequenced HHV-6 ORFs (SHL2, SSL2 and SFL1) related to the HCMV US22 family, only SHL2 encodes this motif (Fig. 4b), whereas partially sequenced ORF SHL1 encodes the closely related amino acid sequence YGCCPGMDLTVIG within the region encompassing domain II.

HHV-6 ORFs related to HCMV UL27

HHV-6 ORF SSL1 is related to HCMV UL27 both in terms of its overall amino acid sequence similarity (Table 2), and on the basis of its position, size and orientation (Fig. 3). This ORF was also found to be related to the adjacent HHV-6 ORF, SHL3, the optimized FASTA score observed in comparison between these protein sequences being 181. Thus the neighbouring HHV-6 ORFs SSL1 and SHL3 form a gene set and are likely to have arisen by the well recognized mechanism involving duplication and divergence (re-
S. Efstathiou and others

(a)  

UL24 79 YLRRFE SC S G SQCIYVV  
UL43 136 L VRR VS TV

(b)  

UL23 79 S P R P K H P L C W R C L P  
UL24 79 Y L R R F E SC S G SQCIYVV  
UL28 88 Y L R R F E SC S G SQCIYVV  
UL29 65 Y L R R F E SC S G SQCIYVV  
UL36 27 FLGECGQ Y R P K I F Q  
UL43 136 L V R R S V

Fig. 4. (a) An alignment of domain I found in the US22 family members grouping the HCMV UL, HCMV Us and HHV-6 members separately, and showing the high conservation of the core motif GXXOXOXWP (where O is a hydrophobic amino acid and X any amino acid) among the UL members. This boxed motif is less well conserved in the HHV-6 proteins and poorly conserved in the HCMV Us family members, as described in the text. (b) An alignment of the previously described domain II (Kouzarides et al., 1988; Chee et al., 1990) found in the US22 family members and the related HHV-6 ORF SHL2.

viewed in McGeoch, 1989). ORF SHL3 is more distantly related to HCMV UL27 than is ORF SSL1, the optimized FASTA scores between these two ORFs being 81. Nonetheless, an alignment of protein sequences encoded by these three ORFs (Fig. 5) demonstrates a convincing relatedness over the entire length of these amino acid sequences.

ORF SSL1 is likely to represent a spliced reading frame because the well conserved N-terminal region of the triple alignment, occurring between amino acids 87 and 121 of the SSL1 protein sequence, occurs before the first methionine encoded by this ORF which is located at amino acid number 144. Furthermore, two pyrimidine-rich potential splice acceptor sites (Senapathy et al., 1990), T0CTTACAACAGG and TGTGTTCTGTGGCCAGA, are located proximal to the start of this ORF at nucleotide positions 6134 and 6194 respectively. A search for possible upstream splice donor sequences has identified candidate donor signals close to the 3' ends of three ORFs: SSL2 (GTAG at nucleotide number 6237), SFL1 (GTAGTAG, two potential overlapping donor signals at nucleotide number 7284) and SFL2 (GTGAG at nucleotide number 8497). Interestingly, the potential donor sequences located at the ends of ORFs

HHV-6 SSL1 1 ASAYVRAKQGVG C V I C R P E R I I I R C W R L P A N F P S P N N R L V P E K I R A V R L D E E D S T V F K T I L P V F G V L R P A T  
HCMV UL27 1 MNPVQDFQPHFLFTQFQEQQAEKHDGGDElrFGRFDLTFDTLDGHOlLCDCHGQFClORFLYRlTyHlNHRlEQA A SlLAClCRFEDlAAAGQKKGFPKlPHK  
HHV-6 SHL3 1 MELGHD  

HCMV UL27 1 ANPVIDQQFFQPDQFTQFQFEQQAEKHDGGDElrFGRFDLTFDTLDGHOlLCDCHGQFClORFLYRlTyHlNHRlEQA A SlLAClCRFEDlAAAGQKKGFPKlPHK  
HHV-6 SHL3 44 A L D H D P S N D S G P L S N N A E K V R L E G D E D T L G H D Y E Y  


HHV-6 SSL1 268 R F S Y M N K R K F P G K S P C T P C G F M G Q Q R  
HHV-6 SHL3 229 ASFRIYLGKSDR  

HCMV UL27 393 NAVLAVLHRAKRLTACIRMVGKHLFKEGAEALNLGR LA O N P V Q R G L R K R H V K K E N T E D E R V E T C F D  

HHV-6 SHL3 408 R A L R E H L C K I K L G D  

HHV-6 SSL1 550 R A F Y A Y L K L G S L V D F D D Y K Y K T E T P D C P G E L M G G O O D  

Fig. 5. An alignment of the related HHV-6 ORFs SSL1 and SHL3 with HCMV UL27 using the MULTALIGN program (Barton & Sternberg, 1987). Amino acids are boxed according to the grouping ILVM, DE, KRH and YFW. The first methionine encoded in the potentially spliced HHV-6 ORF SSL1 is indicated by an arrowhead.
SFL1 and SFL2 encompass their respective stop codons, TAG and TGA. Scrutiny of the corresponding sequence of HCMV reveals a similar pattern of potential splice donor and acceptor sequences. A potential acceptor sequence, CCTTCTTCTTCCTTCAGT, is located at nucleotide number 34677, proximal to the proposed Kozak methionine codon (Kozak, 1984) of the HCMV UL27 ORF (Chee et al., 1990). Thus, the related ORFs HHV-6 SSL1 and HCMV UL27 both encode potential splice acceptor signals in similar locations. In addition, potential splice donor sequences are adjacent to the 3′ ends of the HCMV ORFs UL28 (GTGAG at nucleotide number 34770) and UL29 (GTAGGTGAG, corresponding to potentially overlapping splice donor sequences at nucleotide number 35927). In a manner similar to its homologous HHV-6 ORF SFL1, the potential overlapping splice donor signal located at the end of HCMV UL29 encompasses its TGA stop codon. However, in contrast to the presence of a potential splice donor signal at the end of the HHV-6 SFL2 ORF, no such signals could be identified in the region of the HCMV sequence encoding the positional homologue HCMV UL30. It is clear from these analyses and has been suggested elsewhere (Chee et al., 1990) that a complicated pattern of splicing involving members of the US22 family is likely to occur in the genomes of both HHV-6 and HCMV.

**Discussion**

Sequences have been determined for the genomes of the lymphotropic gammaherpesvirus EBV (Baer et al., 1984), the neurotropic alphaherpesviruses VZV (Davison & Scott, 1986) and HSV-1 (McGeoch et al., 1988), and the betaherpesvirus HCMV (Chee et al., 1990). Comparisons of these sequences have allowed the identification of genes common to all these viruses as well as the recognition of genes specific to either a particular virus or virus subgroup (Davison & McGeoch, 1986; Davison & Taylor, 1987; Kouzarides et al., 1987; Chee et al., 1990). Previous large scale sequence analysis of a 22 kbp region of HHV-6 strain U1102 (Lawrence et al., 1990) identified 17 ORFs, including those encoding highly conserved herpesvirus proteins such as the major capsid protein and alkaline exonuclease, and clearly demonstrated that HHV-6 and HCMV encode closely related proteins and have a similar organization of their coding sequences. More recent sequencing studies of HHV-6 strain U1102 have confirmed these initial findings of overall collinearity and amino acid sequence similarity between HHV-6 and HCMV ORFs, as well as identifying a number of potential genes which have no detectable homologues in the betaherpesvirus HCMV, the alphaherpesviruses HSV-1 and VZV, or the gammaherpesvirus EBV (Teo et al., 1991; Martin et al., 1991b; Neipel et al., 1991; Thomson et al., 1991). Interestingly, large scale sequence analyses of two non-contiguous regions encompassing a total of 16698 bp of HHV-6 strain GS (Josephs et al., 1991) have identified 16 ORFs, of which one (RF2) is clearly homologous to the highly conserved herpesvirus glycoprotein H gene and another, ORF RF5, is homologous to the HCMV tegument protein UL48, but lacks detectable homology to the related gene products of the other human herpesviruses. Surprisingly the amino acid sequences encoded by the remaining 12 ORFs identified in this study showed no significant similarity to ORFs encoded by HCMV or any of the other sequenced human herpesviruses. Since both large scale (Lawrence et al., 1990; Teo et al., 1991; Martin et al., 1991b; B. J. Thomson, M. E. D. Martin & J. Nicholas, personal communication) and short sequence analysis (Neipel et al., 1991) of HHV-6 strain U1102 have shown that a large proportion of the unique region of the virus genome is essentially collinear with HCMV in terms of gene organization, it will be of interest to determine whether the breakdown in collinearity between HHV-6 strain GS and HCMV reported by Josephs et al. (1991) is strain- or virus type-dependent.

In this report we have presented the analysis of 10079 bp of HHV-6 strain U1102 which lies at the left end of the HHV-6 genome outside the region of conserved herpesvirus gene-blocks. Analysis of this region has identified eight ORFs, of which seven have recognizable homologues in the betaherpesvirus HCMV and no obvious counterparts in the alphaherpesviruses VZV and HSV-1, or the gammaherpesvirus EBV. The remaining ORF, SFL2, is of a similar size, position and orientation to HCMV UL30 and is therefore considered to be a positional homologue.

The close collinear relationship between the HHV-6 ORFs identified in this study and their HCMV counterparts both confirms and extends the observations made by Neipel et al. (1991), who by short sequence analysis found that the collinearity between the HHV-6 and HCMV genomes extended as far to the left of the HHV-6 genome as to include a homologue of HCMV UL31. Interestingly, HHV-6 ORF SHL3, which is positionally analogous to HCMV UL26, shows no significant amino acid sequence similarity to this ORF but is clearly related to the adjacent HHV-6 SSL1 ORF, which is itself related to HCMV UL27 in terms of size, orientation and overall amino acid similarity. Therefore HHV-6 ORFs SHL3 and SSL1 form a gene set which is likely to have arisen by a mechanism involving gene duplication and divergence of a single progenitor gene.
The identification of this gene set in addition to the four HHV-6 ORFs with low level amino acid sequence similarity to members of the HCMV US22 gene family suggests that the clustering of such tandemly arranged homologous genes, which mainly occur adjacent to the terminal repeats or within the U3 region of HCMV (Weston & Barrell, 1986; Chee et al., 1990), may be a general feature of herpesvirus genomes. This argument is supported by the identification of an HHV-6 ORF as an HCMV US22 family member which is located proximal to the right terminal repeat (Thomson & Hones, 1992) and the observed clustering of distantly related glycoprotein genes in the U5 regions of alphaherpesvirus genomes (Davison & McGeoch, 1986; McGeoch, 1990).

At present, no biological function has been ascribed to any member of the HCMV US22 family, and the gene product of only one member, HCMV US22, has been identified in infected cells and shown to be an early nuclear protein that is secreted from cells (Mocarski et al., 1988). The identification of four ORFs related to this gene family in the HHV-6 genome, and the absence of family members in sequenced alpha- or gammaherpesvirus genomes indicates that the close genetic relationship between HHV-6 and HCMV is not limited to the highly conserved blocks of herpesvirus genes identified previously (Lawrence et al., 1990; Teo et al., 1991) and it will be of interest to determine whether genes related to this family are common to all betaherpesviruses or indicate a particularly close relationship between HCMV and HHV-6.

We would like to thank Dr B. J. Thomson for unpublished sequence data and helpful discussion, and Mrs Charli Bayley for the typing of this manuscript. Dr Stacey Efstatiou wishes to thank Professor A. C. Minson for helpful discussions during the course of this work.

References


(Received 3 February 1992; Accepted 11 March 1992)