Characterization of human telomeric repeat sequences from human herpesvirus 6 and relationship to replication

U. A. Gompels* and H. A. Macaulay

Viral Pathogenesis Unit, Department of Clinical Sciences, London School of Hygiene and Tropical Medicine, University of London, Keppel Street, London WC1E 7HT, UK

Here we examine by polymerase chain reaction amplification followed by cloning and sequence analyses selected regions of the human herpesvirus 6 (HHV-6) genome which contain human telomeric repeats (TTAGGG). We determine the relative number, arrangement and orientation of the repeats in the unit length genome, in concatemeric replicative intermediates and in heterogeneous (het) regions. We also examine distribution of the repeats in the entire genome (159 kb) and their orientation relative to DNA packaging motifs and the origin of lytic replication. In the prototype orientation the HHV-6 repeat is the related complement, TAACCC. We show that tandem arrays of this repeat are present in the right and left long direct repeats (DRL and DRR, 8 kb each) which bound the long unique sequence (UL, 143 kb). Within each DR there is a left terminal imperfect tandem array and a right terminal perfect tandem array (58 copies). In DR they are each adjacent to DNA packaging motifs, pac1 and pac2, described for herpes simplex virus and human cytomegalovirus, in the arrangement pac1-imperfect repeat-7.2 kb-perfect repeat-pac2. Five independent clones were isolated and sequence determined from junctions of concatemeric replicative intermediates which showed adjacent pac2 and pac1 motifs surrounded by telomeric repeats. Favoured cleavage sites for unit length genomes were indicated which avoided cleavage within the repeats. Analyses of the complete genome showed no tandem repeats within UL but did show a polar distribution of monomeric copies and related sequences around the origin of replication, with an effect on the overall base composition. The implications for virus replication are discussed.

Human herpesvirus 6 (HHV-6) is one of the most recently characterized of the family of seven human herpesviruses. Like other herpesviruses it can establish a latent or persistent infection which remains for the lifetime of the host and can reactivate during immunosuppression. HHV-6 infects up to 90% of the population as infants where infection is asymptomatic or causes exanthem subitum, a mild skin rash (Yamanishi et al., 1988; Okuno et al., 1989). The virus has a cellular tropism for CD4+ T lymphocytes and there is evidence for the monocyte as a possible site for latency. HHV-6 strains were first isolated from blood samples from AIDS patients where the virus had reactivated (Salahuddin et al., 1986; Downing et al., 1987). Given the similar cellular tropisms of human immunodeficiency virus (HIV) and HHV-6 (Tedder et al., 1987; Takahashi et al., 1989) there has been speculation that interactions occur between these viruses in AIDS or that HHV-6 has a role in immunodeficiencies, and some in vitro interactions have been observed (Lusso et al., 1989; Carrigan et al., 1990; Enosi et al., 1989). Evidence for secondary reactivated infections has been observed in immunosuppressed bone marrow transplant patients and in a ‘normal’ adult, which show an association with pneumonitis and bone marrow suppression (Carrigan et al., 1991; Cone et al., 1993; Drobyski et al., 1993; Gompels et al., 1994a).

HHV-6 has characteristics of betaherpesviruses, sharing both encoded amino acid similarities and overall gene organization with human cytomegalovirus (HCMV) (Lawrence et al., 1990; Neipel et al., 1991; Gompels et al., 1992; Nicholas & Martin, 1994; Gompels et al., 1994b). HHV-6 has a distinct genetic structure composed of two G+C-rich terminal direct repeats (DR), DL (left) and DR (right), which bound a long A+T-rich unique sequence (UL) (Gompels et al., 1994b; Martin et al., 1991). For HCMV two motifs for DNA packaging,
pacl and pac2, which are similar to those described for herpes simplex virus (HSV) have been identified at the ends of the genome within C+G-rich repetitive regions and at the junction of these sequences between concatemeric subunits (Kemble & Mocarski, 1989; Deiss et al., 1986). In studies of HHV-6 a repeat unit, GGGTTA, has been identified at the right end of DR₄; this repeat is related to the human telomeric repeat, TTAGGG, and it has been inferred and later shown by hybridization studies that it is present at the genomic termini (Martin et al., 1991; Kishi et al., 1988). It has been suggested that one role for these repeats is to participate in the DNA packaging motif in a manner analogous to the DR2 repeat in HSV (Martin et al., 1991). Another possible role is in maintenance of the latent state preserving the DNA as a linear mini-chromosome with the telomeric repeat units functioning as on the host chromosome. In this paper we use PCR amplification, cloning and nucleotide sequencing, coupled with analyses of the complete genome from our collaborative sequencing studies on strain U1102 in order to characterize these repeats further. The results suggest possible roles of the repeats during replication and latency.

Telomeric repeat sequences had been identified at the end of the terminal DR₄ at the junction between unique and repeat sequences (Martin et al., 1991; Kishi et al., 1988). This region was analysed further by direct sequencing of a 2.3 kb PCR-amplified product plus a plasmid clone of the amplified product. Primer sets, with BamHI sites to facilitate cloning, were used to amplify DNA sequences from JHan cells infected with the earliest passages of strain U1102 available (P2) (Downing et al., 1987) to avoid effects of continuous culture. The HHV-6 Smal J sequence at the DR₄-U₄ junction was amplified (conditions described in Gompels et al., 1993) using two primers, 5' CATGAGAGGATCTGGAAGCTTAAGTTAGGAGTTACCTGGCAGCAGTCGAGTCGAGGGAGATCGGAGATCCTGGACGGACGGAGTACTC 3' and 5' CGAGGATCTGAGATCCTGGCAGCAGGGAGATCGGAGATCCTGGACGGACGGAGTACTC 3', derived from sequence in the adjacent Smal L and SalI H fragments (Lawrence, 1991; Efstatious et al., 1992); the amplified sequence was then cloned into pUC18 and the resulting construct was designated pSMJ6. Sonicated fragments of insert DNA and the original amplification product were end-repaired with Klenow polymerase and ligated with Smal-digested/ phosphatased M13mp18 (Messing, 1983). Using the dideoxynucleotide chain termination method with [³²P]dATP (Sanger et al., 1977; Bankier et al., 1988), the sequence was determined from both sets of clones and then imperfect telomeric repeats; t₂, perfect telomeric repeats then pac2 signal. "-, Regions shown in Fig. 2.

sequence, in an internal 348 bp region (Figs 1 and 2). Directly downstream of this region sequences similar to the HSV-1 and HCMV DNA packaging signals were identified as pac2 (Fig. 2) (Varmuza & Smiley, 1985; Deiss et al., 1986; Kemble & Mocarski, 1989). Simple repeats including A/CAG motifs are found further downstream of the telomeric repeats and the pac2 signal.

Analyses for coding regions by the positional base preference method (Gleeson & Staden, 1991) suggested that the region may be coding. Screening for open reading frames (ORFs) encoding > 100 amino acids with initiating methionine allowed us to identify two rightwards ORFs, SJRF1 and SJRF2. SJRF2 overlapped by 36 bp with the leftward ORF SHL1, previously described as a member of the HCMV US22 family of transcriptional activator genes (Efstatious et al., 1992). The telomeric repeat region was open in two frames from either strand, the ORF encoding the RV repeated amino acids was depicted as the initiating methionine conforms to consensus (Kozak, 1984). Using the FASTA program (Pearson & Lipman, 1988) the amino acid sequences were compared against the SWISSPROT and PIR (releases 28.3.94 and 40.5.94, respectively) protein sequence databases. SJRF1 showed highest similarity with the splicing factor SC35 (score 89, 21.4% identity in a 98 amino acid overlap) and SJRF2 with a U1snRNP 70K protein (score 100, 39% identity in an 81 amino acid overlap). Both proteins are members of the SR family of splicing factors and the alignments favour these residues; SJRF2 additionally has two consensus sequences for phosphorylation of these splicing factor phosphoproteins (S/T-P-X-R/K at TPRR and TPTR) (Gu et al., 1994). Using the TFASTA program to translate nucleotide sequences in the databases and then make amino acid sequence comparisons, SJLF1 shows most similarity with potential sequences also encoded by human telomeric repeats and associated sequences which are similar in organization to this region (Brown et al., 1990; Weber et al., 1990).

On the basis of hybridization experiments, telomeric repeats have also been identified at the junction of unique sequences and the long DR₄, a region noted for
heterogeneity (het) during culture and including a variable *PstI* site (Martin et al., 1991). The sequence determined across this junction was from a cosmid clone, *BamHI* G, but was shorter than the mapped fragment and was concluded to be a het clone (Thomson et al., 1994; Gompels et al., 1994b). The region was examined by using a 5′ primer outside the *PstI* site (5′ CGCTGAGCCGGATCCATCCCAGGAC 3′) and a 3′ primer containing telomeric repeats in PCR amplification from the early passage (P2) U102-infected cell DNA. Two M13 clones were derived from both strands by independent amplifications, and the inserted DNA was sequenced. The 304 bp sequence derived was identical to the sequence from the cosmid clone, both having perfect and imperfect sets of telomeric repeats present at both containing copies of the imperfect repeat pattern.

We consider this sequence to be correct, the deletion perhaps occurring during cosmid cloning (Fig. 2). The region contained telomeric repeat sequences and a sequence similar to the pac1 DNA packaging motif found in HSV-1 and HCMV. In the *BamHI* G sequence, the telomeric repeats at this junction are imperfect, with related hexamers interrupting and adjacent to the tandem array. They follow the pattern (TAACCC)3(TAGGT-)C)(TAGCCC)(TAACCC)a, followed by imperfect repetitions of the telomeric repeats, the entire region spanning 355 bp. The PCR clones supported this arrangement, both containing copies of the imperfect repeat pattern.

As both the terminal sequences are direct repeats the perfect and imperfect sets of telomeric repeats present at the junctions of DR_L and DR_R are expected to be

---

**Fig. 2.** Sequence of the two junctions of telomeric sequences within DR with unique sequences in HSV. Sequence was analysed using the 'nip' suite of Staden's programs (nucleotide interpretation program, UNIX version 7.1, July 1993) (Gleeson & Staden, 1991). EMBL accession nos X79798 and X79799.

**DR_L junction of telomeric repeats with unique sequences**

**start SJRF1**

M W R S S N Q R G V S R R R D K S M K R Y T R H G N A D R R

**end SJRF1**

G H K E G V R A G S R S A H F F S G M start SJLF1

**DR_R junction of telomeric repeats with unique sequences**

**start SJRF2**

M A P A R E G E R E R D G R

**end SJRF2**

M A P A R E G E R E R D G R

---

present in the primer. Clones mconcat93 and 104 had six and nine copies of tandem telomeric repeats, respectively. Telomeric repeat units. Clones mconcat91, 92 and 105 products cloned contained variable amounts of the annealing in the repeats of the DR the amplification negative orientation except mconcat91 which had the designated mconcat91, 92, 93, 104 and 105. All were in the same orientation as the 5' primer. Due to random amplified DNA was cloned into M13mpl8 and the are at the termini of DR~. From separate PCR reactions would also show that the telomeric repeated sequences are at the termini of DR~. The authors also noted a number of aberrant concatemeric junctional clones, but did not report their sequence. It is possible that these are PCR artifacts that arose due to secondary structures forming across the telomeric repeats, but using the same procedure across the DR~U~ or DR~U~ junctions we did not observe similar deletions (Fig. 2).
Table 1. Base compositional bias is related to the orientation of single telomeric repeats around the origin of replication

<table>
<thead>
<tr>
<th>Repeat unit or percentage base composition</th>
<th>All</th>
<th>DR&lt;sub&gt;r&lt;/sub&gt;</th>
<th>U&lt;sub&gt;L&lt;/sub&gt;</th>
<th>&lt;sub&gt;U&lt;sub&gt;L&lt;/sub&gt;&lt;/sub&gt;</th>
<th>&lt;sub&gt;U&lt;sub&gt;L&lt;/sub&gt;&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAACCCTAACCCT</td>
<td>130</td>
<td>65</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GGGTTAGGGTTA</td>
<td>174</td>
<td>75</td>
<td>24</td>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td>TAAACCCT</td>
<td>22</td>
<td>1</td>
<td>20</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>GGGTTA</td>
<td>21-5</td>
<td>33</td>
<td>20</td>
<td>24-5</td>
<td>17-4</td>
</tr>
<tr>
<td>%C</td>
<td>28-9</td>
<td>22-9</td>
<td>29-5</td>
<td>32-2</td>
<td>27-5</td>
</tr>
<tr>
<td>%A</td>
<td>20-7</td>
<td>25-8</td>
<td>20-2</td>
<td>16-8</td>
<td>22-7</td>
</tr>
<tr>
<td>%G</td>
<td>28-9</td>
<td>18-3</td>
<td>29-9</td>
<td>26-5</td>
<td>32-4</td>
</tr>
</tbody>
</table>

Thomson et al. (1994) show linkage of the termini with the perfect telomeric repeats by cloning the DR<sub>r</sub> terminus; they did not report the sequence of the telomeric repeats of a cloned DR<sub>L</sub> terminus or use PCR primers to link the imperfect repeats at the DR<sub>L</sub> terminus with concatemeric junctions. We have extended their work by showing linkage of the imperfect telomeric repeats to the left end of the genome. Our studies also show that the perfect telomeric repeats identified at the DR<sub>r</sub> terminus by Thomson et al. are also at the DR<sub>L</sub>—U<sub>L</sub> junction in the Sma I clone described in the previous section. This sequence includes the pac2 motif and the terminal nucleotides are in agreement with the concatemeric clones mconcat92 and 104 (Fig. 3). However, their U1102 sequence from the right terminus (DR<sub>r</sub>) and in the concatemer is different from that presented here and has eight substitutions and one insertion. This could be from use of later passage virus or from the sequencing method which used direct cycle sequencing with dye terminators. An alignment with consensus sequences for pac1 and pac2 shows 93% and 97% identity between the Z29 and the U1102 sequences shown here, respectively, whereas the pac2 sequence for the U1102 sequence in Thomson et al. has only 84% identity. Overall, taken together the data for these different strains show that the imperfect and perfect tandem arrays of telomeric related repeat sequence are present bounding both ends of the long direct repeats and are present at concatemeric junctions between unit length genomes, the distribution throughout the HHV-6 genome was also examined. Our collaborative studies have led to a description of the complete nucleotide sequence of HHV-6 (Gompels et al., 1994b). Like other herpesviruses, the G + C content across the genome is not uniform. The U<sub>L</sub> segment has a relatively low G + C composition (40%) while DR<sub>L</sub> and DR<sub>r</sub> have a higher G + C composition (59%). The telomeric repeat TTAGGG or its complement CCCTAA (or TAACCC in the case of HHV-6) is 50% G + C bias and could be expected at uneven frequencies in the genome. What is observed, however, are marked differences in the distribution of telomeric repeats between the DR and U<sub>L</sub> regions.

Tandem telomeric repeats are only observed in the two regions bounding each DR and are completely absent from the rest of the genome (Table 1). In contrast, monomers of the repeat unit are represented in the U<sub>L</sub> region, but are limited in the DR regions. Furthermore, the monomeric repeats show a marked polarity around the origin for lytic replication (Table 1, Fig. 4). This is further shown by examining the occurrences of GGG or CCC triplets, which show markedly this polarity around the origin (Fig. 4). If the base composition in U<sub>L</sub> is examined on either side of the origin it can be seen that preceding the origin of replication the bias is in agreement with the bias in the repeat unit TAACCC, in that the C composition and to a lesser extent the A composition (less easy to distinguish from an overall AT-rich region) is increased relative to the overall base composition. Conversely, in the U<sub>L</sub> region downstream of the origin the bias follows the complement of the repeat unit, GGGTTA, such that the G composition and to a lesser extent the T composition are now increased relative to the overall composition (Table 1).

If the whole of U<sub>L</sub> is examined there is no obvious bias to the mononucleotide composition (Table 1), except increased A + T content. However, if the mononucleotide composition for individual nucleotides is plotted (not shown), the C composition decreases after the origin of replication and the G composition increases. This is also

![Fig. 4. Sites of repeats in the HHV-6 genome show a polarity of the telomeric repeat and related repeats. ori, origin of replication; Rtaacc, tandem telomeric repeats. A vertical line marks an exact match; where matches are frequent they appear as a block.](image-url)
observed, but less clearly due to the high A + T content, with the A composition decreasing after the origin of replication (ori) and the T composition increasing. Using all four nucleotides to examine the compositional bias (nip program), equal deviations in mono-, di- and trinucleotide compositions are observed across $U_L$, but are actually arising from different biases pre- or post-ori (not shown). Therefore, the scale of the deviation pre- or post-ori is the same, but due to complementary nucleotides: C, A increased one side; G, T increased the other side. Analyses of all other completely sequenced herpesvirus genomes for deviations in mononucleotide composition did not reveal any evidence for a similar genome polarity, although distribution of internal repetitive sequences may mask this effect. The data are consistent with the repeat unit providing an uneven mutational bias within $U_L$, although this distribution may also be interpreted as a consequence of the polarity in mononucleotide composition. However, if this was so, related hexamers of the same composition should have the same marked polarity; this is observed, for example with the human repeat TTAGGG, but not as clearly around the ori as the HHV-6 repeat GGGTTA (not shown).

Overall our studies complement and extend those of Thomson et al. (1994) which focused on the repeats in $DR_t$ (perfect terminal repeats) while we have analysed $DR_r$ (imperfect terminal repeats) and analysed telomeric repeats through the entire genome. Both $DR_t$ and $DR_r$ are bounded by tandem arrays of the telomeric repeat monomer. On the left end these are imperfect, interrupted by two related hexamers, and on the right end they are perfect, 58 repeats. The tandem arrays are terminated with motifs for DNA packaging such that neither the exact genomic termini nor the junction with $U_L$ contain telomeric repeats. Therefore the exact boundaries of the DR and $U_L$ are sequences which include the pacl and pac2 DNA packaging motifs. Genomic termini were identified by examination of clones derived from concatemeric replicative intermediates (Fig. 2) and comparisons with sequence from a terminal (DR) clone (Thomson et al., 1993, 1994). Therefore the termini are 80 bp, left end, and 56 bp, right end, from tandem telomeric repeats. Within $U_L$, 143 kbp, (Gompels et al., 1994b) there are no tandem telomeric repeats; however, there are monomeric copies of the repeat unit as described above.

The arrangement of the telomeric repeat unit and its relationship with respect to cis-acting sites for replication and DNA packaging leads to certain speculations concerning its function in the virus relative to its function in the host genome. Telomeric repeats function in at least two roles: protecting the ends of chromosomes from exonucleases or ligases, and preventing shortening of the chromosome which can occur during replication from lack of primer sequences at the 5′ end. The shortening is overcome by the action of telomerase, a telomere-specific DNA polymerase which can add telomeric repeats to chromosomes ends using an RNA template with complementarity to the telomeric repeat unit (Blackburn, 1991; Greider, 1993). In the host genome, the chromosomal ends have several kb of the tandem repeats (TTAGGG) with the G-rich strand oriented at the 3′ end with a terminal single-stranded region (Blackburn, 1991; Greider, 1993).

The telomeric repeats in HHV-6 have some features that are distinct from those observed in the host chromosome. The repeats are smaller, 0-3 kb; they are near, 21 to 80 bp, but not precisely at the genomic termini; and the G-rich strand at the 3′ end is the imperfect array of repeats (the het region which we have shown at the end of the genome) not the perfect 58 copies array. However, variants of the simple telomeric repeat have been observed in the terminal array of host chromosomes (Weber et al., 1990) as in the HHV-6 imperfect telomeric array. Some of the concatemeric clones indicate that DNA cleavage may occur closer to the telomeric repeats, the closest being clone mconcat105, which has a concatemeric junction only 21 bp from tandem telomeric repeats (Fig. 3). In studies using oligonucleotides containing at least two tandem telomeric repeats, these ends can be ‘healed’ by being templates for telomerase, allowing subsequent repeat addition, even if separated from the ends by as much as 36 bp (Harrington, 1991). Therefore under certain conditions, the G-rich 3′ end of the HHV-6 genome (left end) may serve as a template for telomerase resulting in establishment of a persisting, linear ‘mini-chromosome’. Whether this takes place in a natural infection or during latency is not known, but these repeats may be utilized in a defective replicating recombinant virus vector combining the origin of replication and DNA packaging motifs. These studies are in progress to ascertain whether the telomeric repeats allow the DNA to persist as an artificial chromosome.

Previously, it had been speculated that telomeric repeats may function as part of a DNA packaging/cleavage signal (Martin et al., 1991). In HSV, which has been the most extensively studied, the ‘a’ repeats, 0.25–0.5 kb, which are part of the direct repeats at the end of the genome and the inverted repeats internally, contain simple repeated sequences (for example, DR2, 10–11 bp) bounded by sequences, pacl and pac2, which direct DNA packaging/cleavage. These motifs are adjacent when there are multiple copies of the ‘a’ sequence or at concatemeric replicative intermediates, where they can serve as packaging/cleavage signals. Endonucleolytic cleavage occurs between the motifs and this process appears linked to copy/repair of the repeats.
and packaging of DNA into preformed virions (Varmuza & Smiley, 1985; Deiss et al., 1986). The HHV-6 DNA packaging motifs are adjacent to the simple telomeric repeats and conform to the minimal regions identified in HCMV as well as a larger consensus identified for HSV, Epstein-Barr virus and related viruses (Fig. 3). In the HHV-6 DR the motifs are separated by 8 kb (pac1–8 kb–pac2), and by 143 kb in the U\textsubscript{r} region but are brought adjacent in the concatemeric replicative intermediates where they appear to mediate DNA cleavage/packaging. Thus the whole DR is analogous to an expanded HSV 'a' sequence in contrast to the model of Thomson et al. (1994) of a–DR\textsubscript{r}–a–U\textsubscript{r}–a–DR\textsubscript{r}–a. The arrangement of pac2/pac1 at concatemeric junctions is also similarly seen in varicella-zoster virus and pseudorabies virus and artificial constructs of HSV carrying this arrangement are functional for DNA packaging (Varmuza & Smiley, 1985; Deiss et al., 1986).

We have shown that no tandem arrays of the telomeric repeat are present within U\textsubscript{r}, whereas monomeric copies are present with a polarity around ori, and this may be a cause or consequence of an overall complementarity in the mononucleotide composition around the ori of the same bias as in the repeat. One model to account for this would be use of the repeat, as RNA copies, in RNA primer formation during synthesis of Okazaki fragments in generation of the lagging replicative strand. Over time, repair replication from the repeat UAACCC may result in complementarity of base composition on either side of the repeat, as RNA copies, in RNA repair replication from the repeat UAACCC may result in generation of the lagging replicative strand. Over time, repair replication from the repeat UAACCC may result in complementarity of base composition on either side of the ori reflecting the composition of the repeat unit as observed in Table 1 and Fig. 4. Pre-formed RNA primers are used in replication of HIV where virion-packaged host tRNAs are used by the reverse transcriptase as primer for synthesis of DNA from an RNA template (Salazar et al., 1993), and genomic complementarity is observed. In Tetrahymena the telomerase has an RNA component with complementarity to 15-telomeric repeats and the complex acts like reverse transcriptase, producing DNA from this RNA template (Greider, 1993). In HHV-6 the hot region of DR\textsubscript{r} at the left end of the genome corresponds to the G-rich 3' end of the host chromosome, which is substrate for the host telomerase. This would be consistent with a model where CCA\textsubscript{3n}A\textsubscript{3n}C\textsubscript{3n} ACC is part of a virus-encoded RNA component which interacts with a telomerase or telomerase-like enzyme acting as template for both repeat additions at the left end of the genome (het), and thus contributing to heterogeneity, as well as participating in RNA priming for lagging strand synthesis during replication of U\textsubscript{r}.

In conclusion, the arrangement of telomeric repeat monomers and tandem direct repeats within the HHV-6 genome suggest that HHV-6 may use these repeats in a unique manner for packaging/cleavage and replication of the genome.

We thank the Medical Research Council UK Genome Mapping Project resource centre and Dr J. Barma (Department of Biochemistry, University of Cambridge) for facilitating computer analyses and the Welcome Trust for support.

References


(Received 4 July 1994; Accepted 8 September 1994)