Sequence of a fusogenic herpes simplex virus, RH2, for oncolytic virotherapy

Gen Takahashi, Noritoshi Meshii, Masakazu Hamada, Soichi Iwai and Yoshiaki Yura

Department of Oral and Maxillofacial Surgery, Osaka University Graduate School of Dentistry, Osaka University, 1-8 Yamadaoka, Suita, Osaka 565-0871, Japan

RH2 is a novel oncolytic herpes simplex virus type 1 (HSV-1) produced by simultaneous infection with neurovirulent γ134.5 gene-deficient HSV-1 R849 derived from strain F and the spontaneously occurring, fusogenic HSV-1 HF in cell culture. The genome of RH2 was studied using Genome Sequencer FLX. RH2 comprised 149,643 bp and it was shown that the lacZ gene was inserted into the γ134.5 gene of R849. Comparison of ORFs revealed that RH2 had 100 % identity with strain F in 21/58 unique long (UL) genes (36.2 %) and 1/13 unique short (US) genes (7.7 %). RH2 had 100 % amino acid identity with HF10 in 24/58 UL genes (41.4 %) and 9/13 US genes (69.2 %). Twelve genes, including UL27 (gB), US4 (gG) and UL6 (gD), had amino acid changes unique to RH2. Amino acid changes in gB occurred at positions 459 (T→A) and 817 (L→P). Other unique features were the amino acids missing in UL36 (VP1/2) and UL46 (VP11/12). Thus, RH2 is an HF10-based vector preserving the fusogenic amino acid changes of gB but lacking the γ134.5 gene. RH2 is expected to be a version of HF10 useful for the treatment of brain tumours as well as oral squamous cell carcinoma. Spontaneously occurring HSV-1 mutants may also be useful clinically, as their genome sequences can easily be determined by this genome sequencing system.

INTRODUCTION

A number of oncolytic strains of herpes simplex virus type 1 (HSV-1) have been developed and studied for their anti-tumour ability (Mineta et al., 1995; MacKie et al., 2001; Kemeny et al., 2006; Rowan, 2010). Although HSV-1 vectors have been constructed using specific strains such as strain F (Andreansky et al., 1997), the virulence of the original virus is important. For example, OncoVEXGSM-CSF, which is under clinical trials, was produced using a new isolate exhibiting high growth potential in tumour cells. It has had the γ134.5 and ICP47 genes deleted, which otherwise block antigen presentation. ICP47 also increases expression of the unique short (US) gene sequence US11, which enhances viral growth in tumour cells. The coding sequence for human granulocyte macrophage-colony stimulating factor has been inserted (Liu et al., 2003; Harrington et al., 2010). It has been shown that addition of a cell fusion capability to an oncolytic HSV can significantly increase the anti-tumour potency of the virus (Fu & Zhang, 2002; Fu et al., 2003). The fusogenic activity of the virus was generated by random mutagenesis of a well-characterized oncolytic HSV (Fu & Zhang, 2002) or by inserting a hyperfusogenic membrane glycoprotein from gibbon ape leukemia virus into the viral genome (Fu et al., 2003; Simpson et al., 2006). HF10, a clone of strain HF, is a spontaneously occurring, highly attenuated laboratory virus with the γ134.5 gene (Wheeler, 1958; Takakuwa et al., 2003). It has multiple mutations compared with the wild-type strain 17. A segment spanning from unique long (UL) gene UL52 to UL55 is translocated to the terminal repeat and UL36 is absent (Ushijima et al., 2007). However, HF10 forms large syncytia in a variety of cell types and has a potent anti-tumour effect. In clinical trials, HF10 was found to be effective for the treatment of breast cancer, pancreatic cancer, and head and neck cancer (Nakao et al., 2004, 2011; Fujimoto et al., 2006). R849 is a γ134.5 gene-deficient HSV-1 derived from strain F and has a lacZ gene at the deleted site (Andreansky et al., 1997).

We have previously examined the cytolytic effect of HF on human oral squamous cell carcinoma (SCC) cells and produced a recombinant virus, RH2, by infecting Vero cells simultaneously with R849 and HF (Ogawa et al., 2008; Takaoka et al., 2011). A PCR-based analysis revealed that RH2 was deficient in the γ134.5 gene and had a lacZ gene. In oral SCC cells and nude mouse tumours, RH2 produced larger syncytia than HF. Thus, it was suspected that the genome was composed of genes derived from either R849 or HF, and that other changes associated with recombination might have occurred. HSV-1 strain 17 has long been the standard virus since its entire sequence was determined (McGeoch et al., 1988). However, the genome structures of
HF10 and strain F were published in 2007 and 2010, respectively (Ushijima et al., 2007; Szpara et al., 2010). This makes it possible to compare RH2 with HSV-1 strains, because the parental R849 and HF are related to strain F and HF10, respectively. The recently developed 454 pyrosequencing technology, both the Genome Sequencer 20 (GS 20) and FLX (GS FLX), has been applied successfully to complete microbial genome sequencing, epigenetic studies, genome surveys and gene expression profiling (Albert et al., 2007; Hiller et al., 2007; Swaminathan et al., 2007; Torres et al., 2008). The sequence of a mildly virulent strain of gallid herpesvirus type 2 was determined by this technology, indicating that sequencing of a complete genome of HSV is easier than before (Spatz & Rue, 2008). Sequencing is

![Genomic structure of RH2.](http://vir.sgmjournals.org)
essential to characterize the genome structure of an oncolytic virus prior to clinical use. In the present study, we attempted to determine the ORFs of RH2 in the UL and US regions using the GS FLX system and compared the results with published data for strain F, HF10 and strain 17. We discuss the possible role of fusogenic virus proteins with amino acid changes in RH2 as well as in HF10.

**RESULTS**

**Pyrosequencing of the genome of RH2 and gap closing**

For sequencing of RH2 DNA, pyrosequencing technology was used. A sequencing run was carried out using the GS FLX system, providing a total of 11 356 116 bases. In de novo
assembly, short sequence reads were assembled into larger blocks by using overlapping stretches of homology between the reads. To improve the de novo assembly process, we identified and removed host sequences that always contaminated the viral DNA preparations. The number of reads analysed was 31,885, and 10,408 reads were used for assembly, resulting in 2077 contigs. The mean length of a contig was 1985 nt and the number of contigs with >500 nt was 93. There were six contigs of >4 kb. The GC content was 45.8 mol%. After alignments using concatenated reference sequences, eight contigs were produced. To determine the gaps between contigs, primers were designed and gap closing was performed by the Sanger method. This procedure joined the large gaps, and sequences encoding a total of 71 ORFs of UL and US were determined (Fig. 1). However, 12 regions located in RL and RS sequences with lengths ranging from 21 to 854 nt were undetermined, because of a failure of the PCR. The total length of RH2 was 149,643 bp. In the RH2 genome, two copies of the lacZ gene were found in RL1 at nt 248–3394 and nt 119,799–122,945.
<table>
<thead>
<tr>
<th>ORF</th>
<th>Identity (%)</th>
<th>Function of gene product</th>
<th>ORF</th>
<th>Identity (%)</th>
<th>Function of gene product</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strain F</td>
<td>HF10</td>
<td></td>
<td>Strain F</td>
<td>HF10</td>
</tr>
<tr>
<td>UL1</td>
<td>96</td>
<td>100</td>
<td>Glycoprotein L, forms complex with gH</td>
<td>UL21</td>
<td>100</td>
</tr>
<tr>
<td>UL2</td>
<td>97.9</td>
<td>100</td>
<td>Uracil DNA glycosylase</td>
<td>UL22</td>
<td>100</td>
</tr>
<tr>
<td>UL3</td>
<td>99.6</td>
<td>100</td>
<td>Nuclear phosphoprotein</td>
<td>UL23</td>
<td>99.2</td>
</tr>
<tr>
<td>UL4</td>
<td>99.5</td>
<td>100</td>
<td>Nuclear protein</td>
<td>UL24</td>
<td>99.6</td>
</tr>
<tr>
<td>UL5</td>
<td>99.5</td>
<td>100</td>
<td>DNA helicase, forms complex with UL8 and UL52</td>
<td>UL25</td>
<td>99.5</td>
</tr>
<tr>
<td>UL6</td>
<td>100</td>
<td>99.7</td>
<td>Capsid portal protein, involved in DNA cleavage and packing</td>
<td>UL26</td>
<td>99.4</td>
</tr>
<tr>
<td>UL7</td>
<td>100</td>
<td>99.7</td>
<td>Associated with intracellular capsid</td>
<td>UL26.5</td>
<td>99.1</td>
</tr>
<tr>
<td>UL8</td>
<td>99.7</td>
<td>99.1</td>
<td>Helicase-primase, forms complex with UL5 and UL52</td>
<td>UL27</td>
<td>98.8</td>
</tr>
<tr>
<td>UL9</td>
<td>99.9</td>
<td>99.3</td>
<td>DNA replication origin-binding helicase</td>
<td>UL28</td>
<td>99.7</td>
</tr>
<tr>
<td>UL10</td>
<td>100</td>
<td>99.6</td>
<td>Glycoprotein M</td>
<td>UL29</td>
<td>100</td>
</tr>
<tr>
<td>UL11</td>
<td>100</td>
<td>97.9</td>
<td>Myristylated tegument protein</td>
<td>UL30</td>
<td>99.4</td>
</tr>
<tr>
<td>UL12</td>
<td>100</td>
<td>99.5</td>
<td>Alkaline exonuclease</td>
<td>UL31</td>
<td>100</td>
</tr>
<tr>
<td>UL13</td>
<td>99.8</td>
<td>99.6</td>
<td>Serine/threonine protein kinase</td>
<td>UL32</td>
<td>99.8</td>
</tr>
<tr>
<td>UL14</td>
<td>100</td>
<td>99.5</td>
<td>Tegument protein</td>
<td>UL33</td>
<td>100</td>
</tr>
<tr>
<td>UL15</td>
<td>100</td>
<td>99.9</td>
<td>DNA packaging terminase subunit 1</td>
<td>UL34</td>
<td>99.6</td>
</tr>
<tr>
<td>UL16</td>
<td>100</td>
<td>99.7</td>
<td>Tegument protein</td>
<td>UL35</td>
<td>100</td>
</tr>
<tr>
<td>UL17</td>
<td>100</td>
<td>99.7</td>
<td>Tegument protein, DNA cleavage and packing</td>
<td>UL36</td>
<td>98.8</td>
</tr>
<tr>
<td>UL18</td>
<td>100</td>
<td>99.7</td>
<td>VP23, forms triplex with VP19C</td>
<td>UL37</td>
<td>99.8</td>
</tr>
<tr>
<td>UL19</td>
<td>100</td>
<td>99.4</td>
<td>VP5, major capsid protein</td>
<td>UL38</td>
<td>99.6</td>
</tr>
<tr>
<td>UL20</td>
<td>99.5</td>
<td>98.6</td>
<td>Membrane protein, involved in egress</td>
<td>UL39</td>
<td>99.6</td>
</tr>
<tr>
<td>UL21</td>
<td>99.4</td>
<td>100</td>
<td>Small subunit of ribonucleotide reductase</td>
<td>US1</td>
<td>98.1</td>
</tr>
<tr>
<td>UL22</td>
<td>99</td>
<td>99.8</td>
<td>Tegument protein, virion host shut-off protein</td>
<td>US2</td>
<td>99</td>
</tr>
<tr>
<td>UL23</td>
<td>98.4</td>
<td>99.8</td>
<td>Accessory subunit of DNA polymerase</td>
<td>US3</td>
<td>99.2</td>
</tr>
<tr>
<td>UL24</td>
<td>98.3</td>
<td>99.2</td>
<td>None</td>
<td>US4</td>
<td>99.2</td>
</tr>
<tr>
<td>UL25</td>
<td>97.5</td>
<td>100</td>
<td>Glycoprotein C</td>
<td>US5</td>
<td>100</td>
</tr>
<tr>
<td>UL26</td>
<td>100</td>
<td>100</td>
<td>Membrane protein</td>
<td>US6</td>
<td>99</td>
</tr>
<tr>
<td>UL27</td>
<td>96.9</td>
<td>99.9</td>
<td>VP11/12, tegument phosphoprotein</td>
<td>US7</td>
<td>95.9</td>
</tr>
<tr>
<td>UL28</td>
<td>99.9</td>
<td>100</td>
<td>VP13/14, tegument phosphoprotein</td>
<td>US8</td>
<td>98.7</td>
</tr>
<tr>
<td>UL29</td>
<td>99.4</td>
<td>99.6</td>
<td>VP16, tegument protein z-transinducing factor</td>
<td>US8.5</td>
<td>97.5</td>
</tr>
<tr>
<td>UL30</td>
<td>98</td>
<td>100</td>
<td>VP22, tegument protein</td>
<td>US9</td>
<td>98.9</td>
</tr>
<tr>
<td>UL31</td>
<td>96.7</td>
<td>None</td>
<td>Membrane-associated protein</td>
<td>US10</td>
<td>97.1</td>
</tr>
<tr>
<td>UL32</td>
<td>98.4</td>
<td>100</td>
<td>Deoxyuridine triphosphatase</td>
<td>US11</td>
<td>96.8</td>
</tr>
</tbody>
</table>
DNA-level sequence variation of RH2

These newly assembled genomes were used to assess DNA-level sequence variation across the genome. When nucleotide sequences were compared with those of strain F (15,2151 bp), RH2 was found to have 88.2% identity with the strain F genome. The lacZ gene was present in the RL region of RH2 but not of strain F, and many single-nucleotide differences were found in the remaining genome except for nucleotides encoding UL10–UL16 (Fig. 2a). In the case of HF10 (13,1774 bp), DNA sequence identity with RH2 was 86.6%. However, the differences were relatively minor, except for the regions encoding UL6–UL22 and UL29 (Fig. 2b). As the published HF10 sequence has a large deletion at both ends of UL and US (GenBank accession no. DQ889502), compatible sequences showed a wide range of nucleotide differences.

Amino acid coding-level identity and changes

The amino acids of all ORFs in UL and US were compared among HSV-1 strains. RH2 had 100% identity with strain F in 21/58 genes (36.2%) in the UL region and 1/13 genes (7.7%) in the US region. In comparison with HF10, RH2 had 100% identity in 24/58 genes (41.4%) in the UL region and 9/13 genes (69.2%) in the US region (Table 1).

Based on amino acid differences among HSV-1 strains, RH2 genes could be categorized into three groups: those different from strain F, those different from HF10 and those different from both strain F and HF10 (Fig. 3). If a gene of RH2 had many amino acid sequences that were different from strain F but not from HF10, the gene was considered to be derived from HF10. The genes of UL1–UL5, UL23–UL28, UL37–UL40 and US1–US12 contained amino acid sequences different from strain F, and those of UL6–UL20 had sequences different from HF10. Moreover, 12 genes had amino acids different from both strain F and HF10, i.e. unique to RH2. These were UL8, UL9, UL13, UL20, UL27 (gB), UL36 (VP1/2), UL41, UL42, UL46 (VP11/12), US4 (gD), US6 (gD) and US12 (ICP47) (Table 2). Five putative proteins encoded by UL31, UL33, UL35 (VP26), UL45 and US5 (gJ) of RH2 were common to these three strains. UL56, which is deleted in HF10, was present in RH2. When the 12 genes unique to RH2 were compared with strain 17, there were amino acid changes as observed in strain F (Table 2).

Specific gene variability

Alterations in certain HSV-1 glycoproteins are responsible for fusogenic ability (Cai et al., 1988; Ligas & Johnson 1988; Baines et al., 1991; Forrester et al., 1992; Roop et al., 1993; Dolter et al., 1994). The 12 gene products with amino acid changes unique to RH2 included the envelope glycoproteins UL27 (gB), US4 (gI) and US6 (gD). RH2 had 100% sequence identity with HF10 in UL10 (gM), UL44 (gC), UL53 (gK), US7 (gI) and US8 (gE), whereas 100% identity with strain F was observed in UL10 (gM) and UL22 (gH). US5 (gJ) was a common gene among all three
strains. It has been reported that a change at position 817 (L→P) at gB is responsible for the syncytial phenotype (Dikadiki-Kosta et al., 2003). This change was observed in RH2 as well as in HF10 (Fig. 4). RH2 had an additional change at position 459 (T→A) of gB. UL36 (VP1/2), UL41 and UL46 (VP11/12) encode tegument proteins. UL36 and UL46 of RH2 showed a deletion at positions 582–583 and 2907–1910, respectively, as a unique change.

**DISCUSSION**

The GS 20 system has been used successfully to sequence galid herpesvirus type 2 and was found to be rapid and cost-effective (Spatz & Rue, 2008). The major concerns have been relatively short read lengths (i.e. a mean of 100–200 nt compared with 800–1000 nt for the Sanger method), the lack of a paired end protocol and the accuracy of individual reads for repetitive DNA, particularly in the case of monoplymer repeats (Hutchinson, 2007). In the present study, we used the GS FLX system, which provides longer read lengths and lower per-base error rates than the GS 20 system (Margulies et al., 2005). Indeed, approximately 76-fold coverage of the genome size was analysed in a single run and 96.4% of contigs had >500 nt. The assembling of contigs and alignment with the reference sequences revealed that eight gap spaces lay in the RL and RS regions. Thus, we performed PCR for the gaps and sequenced the products using the Sanger method. It should be stated that most unique regions (UL and US) can be determined in a single analysis with the GS FLX system. The sequencing of terminal variable regions is a point to address in this technology.

Variation among viral genomes reflects the processes of mutation and recombination. RH2 was obtained by simultaneous infection with R849 and HF and selecting fusogenic viruses with the *lacZ* gene. The genome of RH2 was expected to be a mixture of R849 and HF. Indeed, we found a full-length *lacZ* gene, indicating that R849 provided *lacZ* located at RL1. In comparison with strain F and HF10, the identity was 88.2 and 86.2%, respectively. HF10 has a deletion at the internal terminal of UL, and parts of UL and US are not included in the published data. These sequences were judged by alignment to be different from RH2, which resulted in the reduced sequence identity between HF10 and RH2. Nevertheless, UL sequences, except for UL6–UL22, and US sequences were highly homologous to those of HF10. After conversion of the sequences to amino acids, we found 100% identity in 22/71 genes (30.9%) in strain F and 33/71 genes (46.4%) in HF10. Overall, we concluded that RH2 can be considered an HF10-based HSV-1 vector with additional modifications. Another finding was that RH2 had many genes whose amino acids were different from the two reference strains and strain 17. These were considered to be unique to the recombinant virus.

Although wild-type viruses cause a limited amount of cell fusion, certain mutations cause extensive cell fusion. Several oncolytic HSV-1 vectors kill their target cells through syncytial formation (Nakamori et al., 2003; Israyelyan et al., 2008). The presence of syncytial mutations within different viral genes suggests that virus-induced cell fusion is mediated by the concerted actions and interactions of gD, gB, gH/gL, gK and the membrane protein UL20 (Cai et al., 1988; Ligas & Johnson, 1988; Baines et al., 1991; Forrester et al., 1992; Roop et al., 1993; Dolter et al., 1994). As RH2 produced large syncytia in cell culture and nude mouse tumours, it was suspected that genes encoding viral glycoproteins were altered. Specific gD binding to its receptors herpesvirus entry mediator (HVEM), nectin-1 and heparan sulfate triggers membrane fusion, in which gB and the gH/gL complex are key participants (Montgomery et al., 1996; Geraghty et al., 1998; Shukla et al., 1999, 2000; Tiwari et al., 2005; Atanasiu et al., 2010). Single amino acid changes within two regions of the intracellular cytoplasmic domain of gB cause syncytium formation and are designated region I (aa 816 and 817) and region II (aa 853, 854 and 857) (Bzik et al., 1984; Cai et al., 1988; Gage et al., 1993; Walev et al., 1994). The deletion of the terminal 28 aa of gB also causes extensive virus-induced cell fusion, predominantly because it alters the extracellular conformation (Foster et al., 2001). As gB was altered at position 817 (L→P) in HF10 and RH2, their fusogenic capability can be ascribed to this change. Chouljenko et al. (2010) reported that the N-terminal 82 aa of gK physically interacted with the extracellular portions of gB (aa 30–748) and gH (aa 1–792) but not gD (aa 1–340). We found an additional unique amino acid change in gB at position 459 (T→A). This may affect the interaction of gB with gK to provide a highly fusogenic capability to RH2.

The sequence of the N terminus of gD determines whether nectin-2 or 3-O-sulfated heparan sulfate, as well as the HVEM, can serve as entry/fusion receptors. The amino acids of gD directly in contact with HVEM include residues 7–15 and 24–32 in the N-terminal hairpin (Carfì, 2001). Single amino acid changes and deletions within this region eliminate functional interactions of gD with HVEM. Changes at positions 215, 222 and 223 also eliminate physical and functional interactions with nectin-1 and nectin-2, but have little or no effect on such interactions with HVEM (Manoj et al., 2004). In the present study, unique amino acid changes in gD were observed at positions 50, 52 and 312, but they were not included in the functional domains. gB and gH/gL interact in response to receptor binding by gD (Chowdary et al., 2010). As gH was identical in RH2 and strain F, the involvement of gH alteration in cell fusion is unlikely. Mutational analysis of gL using C-terminal deletions showed that the first 147 aa were sufficient for associating with gH (Peng et al., 1998; Klyachkin et al., 2006). Amino acid changes in gL at positions 90, 100, 109 and 115 relative to strain F occurred in RH2 as well as in HF10, suggesting a role for gL in the fusogenicity of these viruses.

gK and UL20 proteins are strictly required for virus-induced cell fusion. The gK–UL20 protein complex modulates the fusogenic properties of gB and gH via direct physical
interactions, and mutation within either the gK or the UL20 gene causes extensive cell fusion (Hutchinson et al., 1992; Melancon et al., 2004; Chouljenko et al., 2010). Syncytial mutations in gK have been found predominantly within extracellular domains (Foster et al., 2003). Israyelyan et al. (2008) introduced a syncytial mutation in the gK gene at position 40 (A→V) to endow fusogenicity on the oncolytic HSV-1 OncSyn. In HF10 and RH2, one amino acid change in gK was observed at position 226 (F→V), but this change was located in the non-functional transmembrane domains. With regard to UL20, both N- and C-terminal portions, which are predicted to lie within the cytoplasmic side of cellular membranes, function in virus-induced cell fusion (Melancon et al., 2004). RH2 had a unique change in UL20, but it was located at position 186 (A→V) of transmembrane region 4.

All herpesviruses have a tegument, a layer of protein located between the virus capsid and membrane. The tegument is thought to function promptly after infection, before the expression of virus genes can take place (McKnight et al., 1987; Zahariadis et al., 2008; Newcomb & Brown, 2010). UL36 is able to bind the major components by way of documented direct (UL37 and UL48) and indirect (UL46, UL47 and UL49) contacts. UL48 in turn interacts with gH, gB and gD (Ko et al., 2010; Wagner & Smiley, 2011). We found that UL36 and UL46 of RH2 had a deletion as a unique change. This may affect the replication of RH2 at an early step of infection.

Major features of HF10 are the deletion of the right end of UL and internal repeat, including UL56 and LAT, and the insertion of the segment corresponding to UL52–UL55 in the terminal repeat (Ushijima et al., 2007). It has been reported that the UL43 and UL49.5 homologues inhibit the replication of RH2 at an early step of infection.

### Table: Coding sequence variation across the HSV-1 genome

<table>
<thead>
<tr>
<th>Protein</th>
<th>HF10</th>
<th>RH2</th>
</tr>
</thead>
<tbody>
<tr>
<td>UL1</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL2</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL3</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL4</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL5</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL6</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL7</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL8</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL9</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL10</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL11</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL12</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL13</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL14</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL15</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL16</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL17</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL18</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL19</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL20</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL21</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL22</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL23</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL24</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL25</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL26</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL27</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL28</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL29</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL30</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL31</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL32</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL33</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL34</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL35</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL36</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL37</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL38</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL39</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL40</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL41</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL42</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL43</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL44</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL45</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL46</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL47</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL48</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL49</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL50</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL51</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL52</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL53</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>UL54</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>US1</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>US2</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>US3</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>US4</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>US5</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>US6</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>US7</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>US8</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>US8.5</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>US9</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>US10</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>US11</td>
<td>Filled</td>
<td>Filled</td>
</tr>
<tr>
<td>US12</td>
<td>Filled</td>
<td>Filled</td>
</tr>
</tbody>
</table>

**Fig. 3.** Coding sequence variation across the HSV-1 genome. A histogram depicting the total number of amino acid differences observed in RH2 relative to the reference strain F and HF10, for each protein in the HSV-1 genome. Proteins encoded in the UL and US regions are listed in spatial order of occurrence along the HSV-1 genome. Shaded boxes depict the proportion of amino acids that were different from strain F and open boxes indicate those that were different from HF10. Filled boxes depict amino acids that were different from both strain F and HF10, i.e., those unique to RH2. Five genes, UL31, UL33, UL35, UL45 and US5, had no amino acid changes.

http://vir.sgmjournals.org
UL56 encodes a C-terminal anchored type II membrane protein and may be involved in viral neuroinvasiveness. Takakuwa et al. (2003) and Ushijima et al. (2007) proposed that the absence of UL43 and UL49.5 expression was involved in the fusogenicity of HF10, and ascribed the low virulence of HF10 to the lack of the UL56 gene. However, this is not the case in RH2, because these genes were demonstrated in the RH2 genome. The attenuation of RH2 should be ascribed to a lack of the c134.5 gene.

In conclusion, the genome of RH2 is an admixture of the strain F and HF10 genomes. RH2 retains most US regions of HF10 and the fusogenic amino acid change of gB, but is deficient in the c134.5 gene and contains an UL56 gene. Although RH2 has a genome structure with the backbone of HF10, it can be used for brain tumours as well as oral SCC, because of its defect in the neurovirulent c134.5 gene. This study also indicated that spontaneously occurring HSV-1 mutants suitable for oncolytic virotherapy may be transferred to clinical studies, because the genome structure of HSV-1 can easily be determined by genome sequencing.

**METHODS**

**Viral DNA.** For the isolation of viral DNA, confluent monolayers of Vero cells were infected with RH2 at an m.o.i. of 2 and harvested at 48 h after infection. Infected cells were resuspended and subjected to two cycles of freezing/thawing. The supernatant was centrifuged and pelleted using an Ultra Centrifuge High Mac (Hitachi). Viral DNA was isolated using a Pure Link DNA kit (Invitrogen).

**DNA sequencing and de novo assembly.** Sequencing of 7 μg RH2 DNA was carried out commercially using a pyrosequencing platform, GS FLX (454 Life Sciences). This involved the construction of a random library of RH2 DNA using the method described by Margulies et al. (2005). RH2 DNA was mechanically sheared into fragments and adaptor oligonucleotides were added to the ends. The
Sequence of a herpes simplex virus vector

peak size of purified ssDNA containing adaptors was 686 nt. These were hybridized to DNA capture beads to ensure only one DNA fragment per bead and clonally amplified using emulsion PCR. The DNA beads were then distributed into picolitre-sized wells on a fibre-optic slide (PicoTitrePlate; 454 Life Sciences), along with a mixture of smaller beads coated with the enzyme required for the pyrosequencing reaction, including the firefly enzyme luciferase. The four DNA nucleotides were then flowed sequentially over the plate. Sequence reads, quality scores for sequence and contigs were obtained from Takara Bio. Individual reads were assembled initially into large contigs representing the UL, US, RL and RS genomic segments. To generate a new HSV-1 genome sequence, all blocks of assembled sequence were from the new genome. All blocks of assembled sequence were used to generate amino acid translations of assembled sequence locations were used to generate amino acid translations of amino acids of the four HSV-1 strains were aligned. Substitutions observed in RH2 are circled.

**Fig. 4.** Amino acid alignment of the UL27 (gB) protein in strain F, strain 17, HF10 and RH2. The amino acid sequences of the four HSV-1 strains were aligned. Substitutions observed in RH2 are circled.

**ACKNOWLEDGEMENTS**

This investigation was supported in part by a Grant-in-aid for Scientific Research from the Ministry of Education, Science and Culture of Japan. The authors declare that they have no competing interests.

**REFERENCES**


Identification and characterization of a novel herpes simplex virus glycoprotein, gK, involved in cell fusion. 


Foster, T. P., Melancon, J. M. & Kousoulas, K. G. (2001). An α-helical domain within the carboxyl terminus of herpes simplex virus type 1 (HSV-1) glycoprotein B (gB) is associated with cell fusion and resistance to heparin inhibition of cell fusion. Virology 287, 18–29.


G. Takahashi and others


