The carboxyl terminus of the human cytomegalovirus UL37 immediate-early glycoprotein is conserved in primary strains and is important for transactivation

Wail A. Hayajneh,1,2 Despina G. Contopoulos-Ioannidis,1,2† Marci M. Lesperance,1,3‡ Ana M. Venegas1 and Anamaris M. Colberg-Poley1

Center for Virology, Immunology and Infectious Disease Research, Children's Research Institute1, Department of Infectious Diseases2 and Department of Otolaryngology,3 Children's National Medical Center, George Washington University School of Medicine and Health Sciences, 111 Michigan Avenue, NW, Washington, DC 20010, USA

The human cytomegalovirus (HCMV) UL37 exon 3 (UL37x3) open reading frame (ORF) encodes the carboxyl termini of two immediate-early glycoproteins (gpUL37 and gpUL37M). UL37x3 homologous sequences are not required for mouse cytomegalovirus (MCMV) growth in vitro; yet, they are important for MCMV growth and pathogenesis in vivo. Similarly, UL37x3 sequences are dispensable for HCMV growth in culture, but their requirement for HCMV growth in vivo is not known. To determine this requirement, we directly sequenced the complete UL37x3 gene in multiple HCMV primary strains. A total of 63 of the 310 amino acids in the UL37x3 ORF differ non-conservatively in one or more HCMV primary strains. The HCMV UL37x3 genetic diversity is non-random: the N-glycosylation (46/186 aa) and basic (9/15 aa) domains have the highest proportion of non-conservative variant amino acids. Nonetheless, most (15/17 signals) of the N-glycosylation signals are retained in all HCMV primary strains. Moreover, new N-glycosylation signals are encoded by 5/20 primary strains. In sharp contrast, the UL37x3 transmembrane (TM) ORF completely lacks diversity in all 20 HCMV sequenced primary strains, and only 1 of 28 cytosolic tail residues differs non-conservatively. To test the functional significance of the conserved carboxyl terminus, gpUL37 mutants lacking the TM and/or cytosolic tail were tested for transactivating activity. The gpUL37 carboxyl-terminal mutants are partially defective in hsp70 promoter transactivation even though they trafficked similarly to the wild-type protein into the endoplasmic reticulum and to mitochondria. From these results, we conclude that N-glycosylated gpUL37, particularly its TM and cytosolic domains, is important for HCMV growth in humans.

Introduction

Human cytomegalovirus (HCMV) is a medically significant pathogen that causes congenital infections as well as seriously debilitating or life-threatening infections in immunocompromised patients (Britt & Alford, 1996). The UL36–38 locus appears to be essential for HCMV origin-dependent DNA replication and HCMV growth in vitro (Pari et al., 1995; Smith & Pari, 1995). The UL36–38 locus encodes three UL37 immediate-early (IE) proteins: the UL37 exon 1 (UL37x1) protein (pUL37x1) and two UL37 N-glycoproteins (gpUL37 and gpUL37M), which differ in the portion of the N-glycosylation domain which they retain (Wilkinson et al., 1984; Kouzarides et al., 1988; Chee et al., 1990; Al-Barazi & Colberg-Poley, 1996; Goldmacher et al., 1999). pUL37x1, gpUL37 and gpUL37M play significant roles in anti-apoptosis and in the regulation of HCMV DNA replication genes and of cellular genes (Colberg-Poley et al., 1992, 1998; Tenney et al., 1993; Colberg-Poley, 1996; Zhang et al., 1996; Biegalke, 1999; Goldmacher et al., 1999; Hayajneh et al., 2001).
All UL37 IE proteins share their amino-terminal signal sequence (aa 1–22), a strongly charged acidic domain (aa 81–108) and two domains (aa 54–34 and aa 118–147) that are required and sufficient for anti-apoptotic activity (Kouzarides et al., 1988; Goldmacher et al., 1999; Hayajneh et al., 2001). Although UL37x1 homologous sequences are not found in other herpesviruses (Baer et al., 1984; Davison & Scott, 1986; McGeoch, 1989; Gompels et al., 1995; Nicholas, 1996; Rawlinson et al., 1996; Russo et al., 1996; Vink et al., 2000), they are very well conserved in HCMV primary strains, indicating their biological importance for growth in vivo (Hayajneh et al., 2001).

gpUL37, encoded mostly by the downstream UL37 exon 3 (UL37x3) open reading frame (ORF), is an integral membrane protein with a large N-glycosylated domain, a basic domain, a typical hydrophobic transmembrane (TM) domain and a compact cytosolic tail (Kouzarides et al., 1988; Chee et al., 1990; Al-Barazi & Colberg-Poley, 1996). gpUL37 traffics through the endoplasmic reticulum (ER) and Golgi apparatus, to the plasma membrane, and to mitochondria (Al-Barazi & Colberg-Poley, 1996; Colberg-Poley et al., 2000a). gpUL37 is modified by N-glycosylation in the ER and Golgi apparatus during HCMV infection of permissive cells (Al-Barazi & Colberg-Poley, 1996). gpUL37M differs from gpUL37 by the absence of aa 178–262, which contain the first six N-glycosylation signals within the N-glycosylation domain (Goldmacher et al., 1999).

In contrast to the UL37x1 gene, UL37x3 homologous sequences are conserved in other betaherpesviruses (Nicholas & Martin, 1994; Gompels et al., 1995; Nicholas, 1996; Rawlinson et al., 1996). UL37x3 homologous sequences are not required for mouse cytomegalovirus (MCMV) growth in culture (Lee et al., 2000). Nonetheless, an MCMV UL37x3 mutant is severely hindered both in growth and in pathogenesis in mice (Lee et al., 2000).

It has been found that HCMV UL37 exons 2 and 3 are not required for HCMV growth in human diploid fibroblasts (HFF) in culture (Borst et al., 1999; Goldmacher et al., 1999). As the UL37x3 ORF is important for MCMV growth and pathogenesis in vivo but not for its growth in culture, we tested whether UL37x3 sequences are important for HCMV growth in vivo. The limited host range of HCMV and the lack of a suitable animal model which fully recapitulates HCMV replication and pathogenesis in vivo makes it difficult to determine whether or not a gene product is essential for HCMV growth. We therefore determined the nucleotide sequences of the complete UL37x3 gene (930 nt) in 20 HCMV primary strains to assess conservation of its ORF in vivo. Sequence diversity of subdomains within ORFs considered to be highly conserved has been found to occur in HCMV primary strains (Chou & Dennison, 1991; Chou, 1992; Meyer-König et al., 1998; Zwygberg-Wirgert et al., 1998). Conservation of ORFs or specific subdomains in primary strains at the amino acid level suggests a crucial role of the product for HCMV growth in humans. Conversely, if HCMV is able to grow in vivo despite non-conservative substitutions or deletions in ORFs or subdomains, those ORFs or subdomains are predictably non-essential for HCMV growth in humans.

We tested the functional significance of the conserved gpUL37 carboxy-terminal domains of primary strains by examining mutants in the TM and cytosolic domains for transactivating activity and their ability to traffic like the parental gpUL37 protein.

Methods

HCMV primary strains. Clinical specimens used directly for PCR sequencing of HCMV primary strains included a lung biopsy (specimen 1), urine (specimens 6, 11, 13, 15, 32, 52, 53 and 54) and a purified buffy coat preparation from peripheral blood (specimen 72) (Hayajneh et al., 2001). Other HCMV primary strains were PCR-sequenced either from urine cultured only once in HFF cells (specimens 31, 33, 34, 36 and 38) or from supernatant of shell vial assays that were performed on urine specimens (specimens 63, 64, 65, 67 and 68). The specimens had scored HCMV-positive by shell vial assay or culture and were not subcultured further in HFF cells prior to PCR sequencing. HCMV DNA purified from HFF cells infected with HCMV (strain AD169, ATCC) as previously described (Lesperance et al., 1998) served as a positive control for all PCR reactions.

Nested PCR and PCR sequencing. Nested PCR (NPCR) was performed as previously described (Hayajneh et al., 2001). The primers used for PCR and NPCR amplifications are represented in Fig. 1(a). The sequencing strategy involved NPCR, purification of the NPCR products using QIAquick gel extraction kit (Qiagen) and PCR sequencing using UL37 sequencing primers (Fig. 1(a)) as described previously (Hayajneh et al., 2001). Primers 113 (nt 50764–50786), 114 (nt 50156–50176), 115 (nt 50156–50176), 134 (nt 49833–49852), 154 (nt 50765–50786), 173 (nt 50873–50894), 174 (nt 50919–50939) and 175 (nt 49803–49823) allowed for direct PCR sequencing of the complete UL37x3 ORF (nt 49910–50842). UL37x3 sequences from HCMV primary strains were compared to those obtained for HCMV strain AD169, sequenced in parallel. Variant sequences in primary strains were sequenced at least three times to verify their identities independent of PCR amplification.

Phylogenetic tree analysis. This was performed on the nucleotide sequences of the HCMV UL37x3 ORF (nt 49910–50842) of the sequenced primary strains and HCMV strain AD169. The analysis was conducted by complete alignment of the nucleotide sequences using the Clustalx program (NCBI) and construction of an unrooted tree using the PHYLIP 3.5 DrawTree program.

gpUL37 carboxy-terminal mutants. gpUL37 truncation mutants were used to test the functional importance of the conserved carboxy-terminal TM and cytosolic tail (Fig. 1b). Mutant gpUL37 aa 1–438 (p605) was created by insertion of a stop linker (5’ GTA GCC TCT AGC GGC CCC TAG 3’) into an Ndel site (nt 50062) within the TM ORF. This mutant gene was then subcloned under the control of the HCMV major IE enhancer/promoter, generating p627. Mutant gpUL37 aa 1–461 (p612), under the control of the major IE promoter, was generated by insertion of the termination linker into the unique Ndel site (nt 49992) in the UL37x3 cytosolic ORF (Colberg-Poley et al., 1998). The mutant constructions were verified by nucleotide sequencing (US Biochemical) and the ORFs were independently verified by in vitro transcription and translation (unpublished results).

Enzymatic assays. Transactivation of the human hsp70 promoter–chloramphenicol acetyltransferase (pHB-CAT) construction by gpUL37
Fig. 1. (a) The UL37 gene and PCR and sequencing primers. The UL37 exons 1, 2 and 3 are represented at the top of the figure. The approximate nucleotide position of the exons, start (nt 52706) and stop (nt 49910) of translation as well as the splice junctions (nt 52219/50989 and nt 50947/50842) are indicated on the gene. The nucleotide numbering is from the HCMV AD169 sequence (EMBL accession no. X17403, Chee et al., 1990). The arrows on UL37x3 represent PCR and sequencing primers and their polarity. (b) Wild-type and mutant UL37 proteins. gpUL37, gpUL37M and pUL37x1 are represented at the top. The leader (first cylinder) and the acidic domain (open rectangle) are common to all UL37 proteins and are encoded by UL37x1 sequences (Kouzarides et al., 1988; Goldmacher et al., 1999). The N-glycosylation signals (lollipops), a basic domain (cross-hatched hexagon), TM (second cylinder) and cytosolic tail (line) are encoded by UL37x3 sequences. The dashed line represents the sequences absent from gpUL37M. The motifs retained by truncation mutants gpUL37 aa 1–461 and gpUL37 aa 1–438 are represented below.

and the gpUL37 mutant was measured using a modification of the previously published procedure (Colberg-Poley et al., 1992). Briefly, HeLa cells were transfected with pHB-CAT and the expression vectors for gpUL37, mutant gpUL37 or negative control glycoprotein B (gB) using calcium phosphate coprecipitation. Transfection efficiency was normalized by cotransfection of pCH110 (Pharmacia), carrying E. coli lacZ under the control of the SV40 early promoter (Colberg-Poley et al., 1992). Cells were harvested at 48 h after transfection and assayed for both CAT and β-galactosidase (β-gal) activities as previously described (Colberg-Poley et al., 1992). Briefly, CAT reactions contained 80 µg of protein extract, 0.25 M Tris–HCl pH 8.0, 1 mM acetyl-coenzyme A (Pharmacia) and [14C]chloramphenicol. The acetylated forms of [14C]chloramphenicol were resolved from the unacetylated [14C]chloramphenicol by thin layer chromatography and each was quantified by scintillation counting. Transfected cells were assayed for β-gal activity by conversion of chlorophenol red β-galactopyranoside (Boehringer Mannheim) (Eustice et al., 1991). Protein concentrations of the extracts were determined using Bio-Rad Protein Determination Reagent as previously described (Colberg-Poley et al., 2000b). Fold inductions were calculated by dividing the normalized CAT/β-gal activity of each group by the normalized CAT/β-gal activity of the negative control group (gB) (Colberg-Poley et al., 1992).
Confocal laser scanning microscopy. HFF cells were transiently transfected with expression plasmids for mutant gpUL37 aa 1–461 (p612), mutant gpUL37 aa 1–438 (p627), wild-type gpUL37 (p414) or control gB (p370) using LipofectAmine as previously described (Colberg-Poley et al., 2000a). Cells were harvested 48 h after transfection, fixed in methanol and stained simultaneously with rabbit polyvalent Ab1064 (1:200, against gpUL37 aa 27–40), mouse anti-protein disulphide isomerase (PDI, 1:25, StressGen) and human autoimmune serum against mitochondria (1:25, ImmunoVision) at 37 °C for 1 h. Secondary antibodies were fluorescein isothiocyanate (FITC)-conjugated goat anti-rabbit IgG (1:50, Southern Biotechnology), Texas red (TR)-conjugated goat anti-mouse IgG (1:50, Kirkegaard and Perry Laboratories) and cyanine 5 (Cy5)-conjugated goat anti-human IgG (1:100, Jackson ImmunoResearch). Analyses were performed with a Bio-Rad MRC1024 confocal laser scanning microscope (Center for Microscopy and Image Analysis, George Washington University, USA) as previously described (Colberg-Poley et al., 2000a). Emission signals were measured at 520 (FITC), 615 (TR) and 670 (Cy5) nm. Individual signals were captured sequentially to avoid spurious overlap of the emission signals. Individual optical sections were examined to determine co-localization of gpUL37 mutants with cellular markers. Images were generated using Adobe Photoshop (version 4.0), Bio-Rad confocal microscopy plug-ins and Microsoft Publisher 98.

Results and Discussion

Fig. 1 represents the HCMV UL37 genomic sequences, the UL37x3 primers and the wild-type UL37 proteins and truncation mutants. The HCMV strain AD169 UL37x3 sequence was as previously published (Kouzarides et al., 1988).

The UL37x3 ORF is maintained in all sequenced HCMV primary strains. Although substitutions, in-frame insertions and in-frame deletions were found within the HCMV UL37x3 ORF of primary strains, none resulted in termination codons or frame shifts of the UL37x3 ORF. The nucleotide and amino acid variations are not randomly scattered throughout the UL37x3 ORF; rather, they are located predominantly in the extracellular domains of gpUL37/gpUL37M (Fig. 2). The UL37x3 genetic diversity spanned the first three-quarters of the ORF. A total of 88 of the 310 aa in the UL37x3 ORF differ from AD169, 63 of them non-conservatively in one or more strains (Table 1).

N-glycosylation subdomain (aa 206–391)

As viral glycoproteins are potentially subject to the host immune response, we anticipated that the gpUL37/gpUL37M N-glycosylated subdomain (aa 206–391) might contain the largest number of variant amino acids. A total of 67 residues of the 186 aa differed: of these, 46 residues differed non-conservatively in one or more primary strains (Table 1). Thus, we found that the N-glycosylation ORF differs from AD169 more frequently than does the upstream UL37x1 ORF (Hayajneh et al., 2001) but comparably with the HCMV UL144 ORF (Lurain et al., 1999).

gpUL37 (including signals 1–17) and gpUL37M (including signals 7–17) have large N-glycosylation domains (Fig. 1b). As
Table 1. Proportion of UL37x3 variant amino acids in primary strains

<table>
<thead>
<tr>
<th>gpUL37 aa residue no.</th>
<th>Total variant residues</th>
<th>Conservative variants</th>
<th>Non-conservative variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>UL37x3 ORF</td>
<td>178–487</td>
<td>88/310 (28.4%)</td>
<td>25/310 (8.1%)</td>
</tr>
<tr>
<td>N-glycosylation</td>
<td>206–391</td>
<td>67/186 (36.0%)</td>
<td>21/186 (11.3%)</td>
</tr>
<tr>
<td>Basic domain</td>
<td>368–382</td>
<td>13/15 (86.7%)</td>
<td>4/15 (26.7%)</td>
</tr>
<tr>
<td>Transmembrane</td>
<td>433–459</td>
<td>0/27 (0%)</td>
<td>–</td>
</tr>
<tr>
<td>Cytosolic tail</td>
<td>460–487</td>
<td>2/28 (7.1%)</td>
<td>1/28 (3.6%)</td>
</tr>
</tbody>
</table>

Table 2. Conservation of gpUL37 N-glycosylation signals in HCMV primary strains

<table>
<thead>
<tr>
<th>N-glycosylation Signals</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11*</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD169</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain 13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Sequence of a new consensus N-glycosylation signal is shown.
† Deletion (Δ) of an N-glycosylation signal is indicated.

the UL37x1 hydrophobic leader lacks diversity in all HCMV primary strains sequenced (Hayajneh et al., 2001), targeting of gpUL37/gpUL37M encoded by HCMV primary strains to the ER is predicted. To test this prediction and the importance of post-translational modification of gpUL37/gpUL37M in vivo, we examined the conservation of N-glycosylation signals. Two N-glycosylation signals, signals 6 and 15, are deleted in-frame in 10/20 and 9/20 primary strains, respectively (Table 2). Nonetheless, the majority of the N-glycosylation signals (NXT/S) either lack diversity (signals 1–5, 7–12 and 17) or are conserved (signals 13, 14 and 16) in all sequenced primary strains because of second (X) amino acid variants, which do not alter the recognition signal. Moreover, five strains contain an in-frame insertion of consensus N-glycosylation signals either between signals 11 and 12 (signal 11*, NAT, 1/20) or between signals 13 and 14 (signal 13*, NVT, 4/20). Thus, while overall genetic diversity within the UL37x3 N-glycosylation domain is notable, the N-glycosylated character of gpUL37/gpUL37M
is retained by virtue of the conserved UL37x1 hydrophobic leader and multiple N-glycosylation signals.

Three cysteines (aa 291, 366, 453) encoded by the HCMV UL37x3 gene are conserved in the human herpesvirus 6 UL37x3 homologous ORF (Nicholas & Martin, 1994). These residues are conserved in all the HCMV primary strains sequenced (data not shown).

The divergent basic subdomain (aa 368–382)

The UL37x3 basic subdomain partially overlaps with the end of the N-glycosylation subdomain. The basic residues are the most diverse within the UL37x3 ORF, 13 of the 15 residues differ, nine non-conservatively (Table 1). The conservative variants include $^{371}$K → R, $^{373}$T → S or A, $^{374}$V → L or I, $^{376}$L → I, $^{377}$T → A, $^{378}$R → K, $^{380}$K → R and $^{382}$K → R (Table 3). The non-conservative variants include $^{369}$F → L, $^{372}$R → G, $^{375}$K → S, $^{376}$L → F, $^{377}$T → M or deletion, $^{379}$N → R or K, $^{380}$K → N, Q or deletion, $^{381}$T → Q or deletion and $^{382}$K → I. Thus, four basic subdomain residues ($^{371}$K, $^{373}$T, $^{374}$V and $^{378}$R) varied only conservatively while nine residues ($^{369}$F, $^{372}$R, $^{375}$K, $^{376}$L, $^{377}$T, $^{379}$N, $^{380}$K, $^{381}$T and $^{382}$K) varied both conservatively and non-conservatively.

The TM ORF (aa 433–459) lacks diversity in HCMV primary strains

In strong contrast to the N-glycosylation and basic subdomains, no amino acid differences were observed in the UL37x3 TM domain of any of the 20 sequenced HCMV primary strains (Fig. 2 and Table 1). Because the 27 TM residues of gpUL37/gpUL37_M serve as an anchor for this type I integral membrane protein (Al-Barazi & Colberg-Poley, 1996), we expected that all residues encoded by primary strains would be hydrophobic. However, the complete lack of diversity in the primary strain UL37x3 TM ORF was surprising and suggests an important role beyond that of a hydrophobic anchor, likely involving direct interactions with other membrane proteins.
The conserved cytosolic tail (aa 460–487)

The compact gpUL37/gpUL37M cytosolic tail has consensus PKC (\(^{176}\)STK), CKII (\(^{177}\)TKND) and Tyr (\(^{461}\)RDLLD-FRY) phosphorylation sites. The residues contained within these consensus phosphorylation sites lack diversity in all the sequenced primary strains (Fig. 2 and Table 1). Two cytosolic tail residues (\(^{176}\)S → G and \(^{461}\)R → W), which are not contained within consensus phosphorylation sites, were found to differ infrequently (2/20 and 3/20, respectively). The \(^{176}\)S → G variation is conservative while the other is not.

Phylogenetic analysis of HCMV UL37x3 DNA sequences

Visual inspection of the HCMV UL37x3 ORF indicated that some strains were closely related in nucleotide sequence. To test this suggestion, we performed a phylogenetic analysis of the HCMV UL37x3 nucleotide sequences to generate an unrooted tree (Fig. 3). Based upon this analysis, five primary branches of HCMV primary strains were observed. These are named I–V starting with the cluster containing strain AD169. Group I viruses (strains 34, 52, 53, 64, 65 and AD169) differ by no more than seven of the 930 nucleotides in the UL37x3 ORF from strain AD169. Group II (strains 1, 6, 11, 13, 32, 33, 63 and 68) differ by 93–120/930 nt from AD169 UL37x3 ORF. Group III (strains 15 and 36), group IV (strains 31, 54, 67 and 72) and group V (strain 35) differ by 100–143/930 nt from the strain AD169 UL37x3 ORF. We also generated an unrooted tree using the amino acid sequences of the UL37x3 ORF from the HCMV primary strains. All the strains remained in the same groups, as seen in the nucleotide phylogenetic analysis, indicating that translationally silent mutations had not resulted in any significant convergence of the UL37x3 amino acid sequence (data not shown).

gpUL37 carboxy-terminal mutants are partially defective for hsp70 promoter transactivation in transfected cells

The low variability of gpUL37 carboxy-terminal TM and cytosolic sequences in most primary strains suggested their functional importance for HCMV growth in vivo. Therefore, we set out to determine the contribution of the TM and cytoplasmic subdomains to the known regulatory activity of gpUL37 for hsp70 promoter transactivation (Colberg-Poley et al., 1992, 1998; Tenney et al., 1993; Zhang et al., 1996). We examined two gpUL37 truncation mutants, gpUL37 aa 1–461 and gpUL37 aa 1–438 (Fig. 1b), for their ability to transactivate hsp70 promoter activity (Fig. 4). The CAT activities of triplicate wild-type gpUL37 extracts converted about 90–94% (29238, 31131 and 33952 c.p.m.) of total \(^{14}\)C]chloramphenicol (31133, 34578 and 37000 c.p.m.) to its acetylated forms. Mutant gpUL37 aa 1–461, which retains the TM domain and 2 aa of the cytosolic tail, transactivated hsp70 expression (10.7 ± 1.1-fold) to levels about 30% of wild-type gpUL37 (38.5 ± 0.48-fold). Mutant gpUL37 aa 1–438, which lacks an intact TM and the cytosolic domain, transactivated hsp70 promoter-driven expression (3.75 ± 0.6-fold) to levels about 10% of wild-type gpUL37. gB, an HCMV structural glycoprotein, served as a negative control for these experiments (10.0 ± 0.0-fold). These results suggest that the gpUL37 carboxy-terminal cytosolic tail and TM domain, encoded by
the UL37x3 sequences which lack diversity, play key functional roles in transactivating activity.

**gpUL37 carboxyl-terminal mutants traffic to the ER and to mitochondria in HFF cells**

As mutants gpUL37 aa 1–461 and gpUL37 aa 1–438 are partially defective in hsp70 promoter transactivation, we examined their trafficking in HFF cells to determine if it was similar to that of the wild-type gpUL37 (Fig. 5). Wild-type gpUL37 is known to traffic to the ER and to mitochondria (Colberg-Poley et al., 2000a). For these studies, HFF cells were transfected with expression vectors encoding gpUL37 aa 1–461 (Fig. 5 a–f) or gpUL37 aa 1–438 (Fig. 5 g–l) and stained simultaneously with anti-gpUL37 (Ab1064), anti-PDI and anti-mitochondrial antibodies. gpUL37 aa 1–461 (Fig. 5 a) and gpUL37 aa 1–438 (Fig. 5 g) were detected using Ab1064. The ER (Fig. 5 b, h) and mitochondria (Fig. 5 d, j) were clearly stained by their respective antibodies. In the optical sections shown, gpUL37 aa 1–461 (Fig. 5 c, e) and gpUL37 aa 1–438 (Fig. 5 i, k) co-localized with the ER and mitochondrial markers as indicated by the yellow and aquamarine overlaps, respectively. Overlap between the two compartments containing the mutant glycoproteins (white overlap) was observed (Fig. 5 f, l). Taken together, these results suggest that gpUL37 aa 1–461 and gpUL37 aa 1–438 traffic through the secretory apparatus and to mitochondria, as wild-type gpUL37 does.

The UL37x3 gene has been deleted in an HCMV mutant with no apparent effect on the viability of HCMV in cultured fibroblasts (Borst et al., 1999). Although the dispensability of the UL37x3 gene has been tested in cell culture, it has not been determined for HCMV growth in humans. This is particularly pertinent to HCMV, as the MCMV UL37x3 homologous sequences were also found not to be essential for MCMV growth in mouse fibroblasts in culture; yet, the MCMV UL37x3 mutant was severely attenuated for its growth and pathogenicity in mice.

Identification of HCMV essential genes has been hindered by the limited host range of the virus and the lack of a suitable animal model for studying HCMV infection in vivo. To determine the requirement for UL37x3 sequences for HCMV growth in vivo, we therefore directly sequenced the complete UL37x3 gene in HCMV primary strains and compared it to that of HCMV strain AD169, sequenced in parallel. HCMV (AD169) is a laboratory strain, which has been extensively passaged for more than 40 years in laboratories. This prolonged passage of strain AD169 in culture has resulted in its loss of multiple genes encoding other HCMV glycoproteins (Cha et al., 1996) but not of UL37x3 sequences (Kouzarides et al., 1988). Sequence diversity of subdomains within ORFs considered to be highly conserved has been found to occur in HCMV primary strains (Chou & Dennison, 1991; Chou, 1992; Meyer-König et al., 1998; Zweyergberg Wirgart et al., 1998). Thus, when ORFs or specific subdomains are conserved in primary strains at the amino acid level, a crucial role of the product for HCMV growth is implied. Conversely, non-conservative substitutions or deletions of ORFs or subdomains in primary strains suggest a non-essential role of the ORF or its subdomain for HCMV growth in humans.

Identification of regions of the HCMV genome with low levels of nucleotide variation is also important for designing primers for PCR detection of HCMV DNA in clinical specimens. We discovered significant strain-specific sequence diversity in the 5’-three-quarters of the UL37x3 gene which encode the N-glycosylation and basic subdomains of gpUL37/gpUL37m. These sequences differ comparably with the HCMV UL144 ORF (Lurain et al., 1999). In contrast, the low diversity of the UL37x3 nucleotide sequences encoding the TM and cytosolic tail makes these sequences suitable for diagnostic primer design. Our TM and cytosolic tail primers were found to be unique to HCMV DNA by searching the combined sequence databases (W. A. Hayajneh & A. M. Colberg-Poley, unpublished results).

Although the overall UL37x3 nucleotide diversity is notable, construction of an unrooted tree was possible. The tree indicated that there are at least five primary branches, some of which have multiple members (groups I, II and IV). The length of the branches indicates that UL37x3 sequences have diverged more than UL37x1 sequences (Hayajneh et al., 2001). Notably, the position of strain AD169 in the tree indicates that this laboratory strain has not significantly diverged its UL37x3 sequences during its in vitro culture over the last 40 years from those present in other group I primary strains isolated within the last 5 years.

At the amino acid level, the N-glycosylation subdomain was diverse and comparable to the UL144 ORF (Lurain et al., 1999) in its diversity. In the N-glycosylation subdomain (aa 206–391), we consider 276V/I, 222C, 251Q, 293T, 328H, 338V, 392V and 394S to be consensus because these residues are encoded in $\geq 50\%$ of the HCMV strains. We note, however, that within the N-glycosylation subdomain, the vast majority of the N-glycosylation signals are retained. Although we know that gpUL37 is N-glycosylated (Al-Barazi & Colberg-Poley, 1996), we do not know which N-glycosylation sites are post-translationally modified. The functional significance of N-glycosylation for gpUL37 is unclear, but these N-glycosylation sites are known to be required for modification and processing of gpUL37 in the ER and Golgi apparatus. N-glycosylation of gpUL37 may be needed to ensure proper folding and transport of the protein through the ER and the Golgi apparatus to the cell surface (Colberg-Poley et al., 2000a). A defect in transactivating activity was documented in a gpUL37 mutant lacking the first two-thirds of the N-glycosylation domain (H. Zhang & A. M. Colberg-Poley, unpublished results). Taken together with the previously observed lack of diversity of the gpUL37 hydrophobic leader sequence (Hayajneh et al., 2001), these results suggest that post-translational processing of gpUL37/gpUL37m by N-glycosylation, observed during productive HCMV infection of HFF cells in culture (Al-Barazi &
Colberg-Poley, 1996), is important for its function during HCMV growth in vivo.

Multiple residues within the basic domain ORF differ in most primary strains when compared to AD169. We consider $^{374}R$, $^{374}L$ and $^{374}K$ of the basic subdomain (aa 368–382) to be consensus for HCMV because these residues appear in $\geq 50\%$ of the primary strains. As the basic domain has the highest proportion of non-conservative variant residues and
acquired (AGG) in strains 6, 13, 33, 63 and 68] while group IV [(AGA)] by two different codons [(AGA) in strains 1, 11 and 32 and thus non-essential for gpUL37 function, suggests that specific residues in the TM subdomain which is not required to maintain its role as a hydrophobic anchor, implies that these UL37x3 subdomains have an essential role for HCMV growth in humans not detected during its growth in culture.

The authors wish to thank Drs Holland, Rodríguez and Jantausch for constant encouragement and support, Dr Joe Felsenstein for advice concerning the PHYLIP analysis, and Drs Ted Schutzbank and Joseph Campos for providing clinical specimens. This work was supported by Grant-in-Aid 9750012N from the American Heart Association (to A.M.C-P), Children’s Research Institute (CRI) Discovery Funds, and a CRI Fellowship Award (to W.A.H.).

References


