Complex alternative processing of human cytomegalovirus UL37 pre-mRNA

Richard Adair,1 Gregory W. Liebisch1 and Anamaris M. Colberg-Poley1,2

Center for Cancer and Immunology Research, Children’s Research Institute, Children’s National Medical Center1 and Department of Pediatrics2, George Washington University School of Medicine and Health Sciences, 111 Michigan Avenue, NW, Washington, DC 20010, USA

Differentially processed human cytomegalovirus (HCMV) UL37 RNAs encode biologically significant proteins. Due to the recent discovery of alternative UL37 exon 3 (UL37x3) splice donors, permissively infected cells were thoroughly examined for additional alternatively spliced UL37 RNAs. Newly described donors within UL37 exon 1 (nt 52520) and intron 1 (nt 52209) as well as UL37x3 di (nt 50770) and dii (nt 50782) were differentially spliced to known downstream UL37 acceptors. The alternatively spliced UL37x1, UL37L, UL37 di and UL37dii RNAs predictably encode proteins of 83, 163, 217 and 213 residues, respectively, which share UL37x1 N-terminal sequences but differ downstream in their C termini. Moreover, temporal expression of the alternatively spliced UL37 RNAs differs during HCMV infection. The complexity of UL37 pre-mRNA processing is evidenced by the detection of 11 UL37 spliced and unspliced UL37x1 RNAs in HCMV-infected cells. Based upon these data, a revised HCMV UL37 gene map is presented, which incorporates all RNA species detected during permissive infection.

The human cytomegalovirus (HCMV) UL37 immediate-early (IE) locus encodes proteins that play important roles for anti-apoptosis, viral DNA replication and growth (Colberg-Poley, 1996; Colberg-Poley et al., 1992, 1998; Goldmacher et al., 1999; Hayajneh et al., 2001a, b; Smith & Pari, 1995). The UL37 exon 1 (UL37x1) protein (pUL37x1 or vMIA) and UL37 glycoproteins (gpUL37 and gpUL37 M) contain two domains, at aa 5–34 and 118–147, which are required and sufficient for anti-apoptotic activities (Goldmacher, 2002; Goldmacher et al., 1999; Hayajneh et al., 2001a). HCMV UL37 proteins unconventionally traffic into the endoplasmic reticulum and mitochondria, where they inhibit apoptosis by blocking cytochrome C release (Al-Barazi & Colberg-Poley, 1996; Colberg-Poley et al., 2000; Goldmacher, 2002; Goldmacher et al., 1999; Hayajneh et al., 2001a; McCormick et al., 2003).

The UL37 gene is known to encode three alternatively processed transcripts (Chee et al., 1990; Goldmacher et al., 1999; Kouzarides et al., 1988; Tenney & Colberg-Poley, 1991a, b). Because of our recent discovery of unreported UL37 exon 3 (UL37x3) splice donors (Su et al., 2003), we thoroughly examined HCMV-infected cells for more alternatively processed UL37 RNA species. To that end, we generated a UL37 cDNA library, screened 100 clones by PCR (as described below) and sequenced any previously unreported UL37 spliced cDNAs. The UL37 spliced junctions and cDNAs found in HCMV-infected cells are shown in Table 1 and Fig. 1(A). pUL37x1, gpUL37 and gpUL37 M, as well as the newly described UL37 ORFs, are represented in Fig. 1(B).

To generate the UL37 cDNA library, 2 × 10⁸ human foreskin fibroblasts (HFFs) were infected with HCMV strain AD169 in the absence of any inhibitor treatment. To insure that all cells were synchronously infected, an m.o.i. of 3 p.f.u. per cell was used. Cells were harvested at 8 h post-infection (p.i.) and poly(A)⁺ RNA was isolated using the Poly(A) Pure Isolation kit (Ambion). Reverse transcription of 800 ng RNA was performed using oligo(dT) and SuperScript II reverse transcriptase (Invitrogen), as described previously (Tenney et al., 1993). The UL37 cDNA library was generated by PCR amplification of HCMV-infected HFF poly(A)⁺ cDNAs, as described by Su et al. (2003), using primers against the outermost UL37 gene sequences (Fig. 1A). Forward primer 244 (nt 52738–52712) is positioned upstream of the UL37x1 ATG codon, while reverse primer 245 (nt 48127–48101) is positioned downstream and retains the UL37 (and UL36) RNA cleavage site. The cDNA library was amplified by PCR using Advantage cDNA polymerase (Clontech), as recommended by the manufacturer.

UL37 cDNAs (3–4 kbp) amplified by PCR were excised from a 1·2 % agarose gel, recovered using the QIAquick Gel Extraction kit (Qiagen) and cloned into the pCR4-TOPO cloning vector (Invitrogen). Mini-prep plasmid DNAs were screened by PCR amplification using combinations of the forward UL37x1 primers 264 (nt 52541–52520) or 256 (nt 52330–52312) with one of several specific reverse primers complementary to the UL37x1/x3A [primer 257 (nt 50587–50574/50782–50770)], UL37dii/x3A [primer 258 (nt 50587–50574/50791–50782)] or UL37x2/x3A [primer 259 (nt...
Table 1. Sequence of newly described HCMV UL37 alternatively spliced RNA junctions

<table>
<thead>
<tr>
<th>UL37 spliced junctions (splice site prediction score*)</th>
<th>Splice donor (exon/intron)</th>
<th>Splice acceptor (intron/exon)</th>
<th>Spliced junction (nt)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon 1s (0-99)/exon 2 (0-99)</td>
<td>UUGACC/GUGAG</td>
<td>AUUUCUUUUCAGUGGCCGGGU</td>
<td>UUGACC/GUGCGC (52520/50989)</td>
</tr>
<tr>
<td>Exon 1 (0-91)/exon 2 (0-99)</td>
<td>UCACCA/GUAG</td>
<td>AUUUCUUUUCAGUGGCCGGGU</td>
<td>UACCAC/GUGCGC (52219/50989)</td>
</tr>
<tr>
<td>Exon 1s (0-54)/exon 2 (0-99)</td>
<td>GCAGGG/GUAA</td>
<td>AUUUCUUUUCAGUGGCCGGGU</td>
<td>GCAGGG/GUGCGC (52209/50989)</td>
</tr>
<tr>
<td>Exon 2 (0-97)/exon 3 (0-75)</td>
<td>UUUCAG/GUAA</td>
<td>GCUCUCGCAGAG/GUUCUGUGU</td>
<td>UUUCAG/GUUUUC (50947/50842)</td>
</tr>
<tr>
<td>Exon 2 (0-97)/exon 3A (0-32)</td>
<td>UUUCAG/GUAA</td>
<td>AUUGUGUUCAGUGAACCUGG</td>
<td>UUUCAG/GUAAC (50947/50587)</td>
</tr>
<tr>
<td>Exon 3a (0-94)/exon 3A (0-32)</td>
<td>UCAGGA/GUACG</td>
<td>AUUGUGUUCAGUGAACCUGG</td>
<td>UCAGGA/GUAAC (50782/50587)</td>
</tr>
<tr>
<td>Exon 3a (0-99)/exon 3A (0-32)</td>
<td>ACCAGG/GUACG</td>
<td>AUUGUGUUCAGUGAACCUGG</td>
<td>ACCAGG/GUAAC (50770/50587)</td>
</tr>
</tbody>
</table>

*The prediction score for use of the splice site is based on the Berkeley Drosophila Genome Project splice and acceptor site prediction program [http://www.fruitfly.org/seq_tools/splice.html] for human genes.
†Numbering refers to HCMV genomic sequence (EMBL accession no. 17403).

50587–50574/[50956–50947]) spliced junctions (Su et al., 2003). Primer specificity for the newly described UL37 spliced junctions was verified by PCR amplification of cloned cDNAs containing the UL37 di or UL37 dii spliced junctions by primers 264/257 and 264/258, respectively (Fig. 2A).

UL37 cDNA clones, whose PCR products were not predicted by currently known UL37 transcripts, were initially sequenced using primers 99 (nt 50840–50819) and 154 (nt 50786–50765). Selected clones were fully sequenced using automated DNA sequencing, as described previously (Hayajneh et al., 2001a). DNA sequencing of UL37 cDNAs revealed the use of two additional splice donors located within UL37x1 and UL37 intron 1 (Table 1). The first, UL37x1s, is located at nt 52250 and is positioned upstream of the known UL37x1 donor (Kouzarides et al., 1988; Tenney & Colberg-Poley, 1991a). Three alternatively spliced UL37 variants using the UL37x1s donor were detected: UL37 S , UL37 Sdi and UL37 Sdii (Fig. 1A). Curiously, these UL37 splice variants are predicted to encode an identical 83-residue protein spanning the UL37x1 hydrophobic leader and first anti-apoptotic domain (aa 5–34) and a previously unreported C terminus (Fig. 1B). In addition to these UL37S RNAs, we detected at 8 h p.i. a PCR product that corresponded in size to UL37 SM RNA (unpublished results). The predicted UL37 SM ORF spans aa 1–62 of UL37x1 but differs from the other UL37 S products in the last 12 residues of its C terminus.

The second new UL37 donor, UL37x1L, is located at nt 52209, is positioned downstream of the previously described UL37x1 donor (nt 52219) and is located within UL37 intron 1 (Kouzarides et al., 1988; Chee et al., 1990). Three alternatively spliced UL37 variants using the UL37x1L donor were detected: UL37 L , UL37 LM or UL37 Ldi. Although the low-abundance RNA splice variants UL37 L , UL37 LM and UL37 Ldi were detected throughout HCMV infection and verified by sequencing of the cDNA products, only the full-length UL37 L cDNA has been cloned. Each of these splice variants is predicted to encode a 163-residue protein identical to pUL37x1.

Fig. 1. (A) HCMV alternatively processed UL37 transcripts. The RNA map indicates the direction of transcription from the HCMV genome, exons (grey boxes), introns (white boxes), 3′ untranslated region (thin lines) and poly(A) tails (arrowheads). The three upper UL37 RNAs, UL37x1, UL37 and UL37x3, have been reported previously (Goldmacher et al., 1999; Kouzarides et al., 1988; Tenney & Colberg-Poley, 1991a, b). Full-length cDNA clones of UL37x1, UL37x1L, UL37x1S and UL37x1L have been isolated. Partial cDNAs of UL37x1, UL37x3 and UL37x3 have verified splicing of UL37x1 S/x2/di/x3A, UL37x1 L/x2/x3A and UL37x1/x2/di/x3A, respectively, but the full-length cDNAs have not been isolated yet. The predominant temporal expression of each UL37 RNA, as determined by RT-PCR of HCMV-infected cell RNAs at 8 (IE), 16 (IE), 24 (E), 48 (EL), 72 (L) h p.i. [see Fig. 2(B) and, as indicated by an asterisk, unpublished results], is indicated on the right. The approximate physical locations and orientations of primers 244 and 245 used to generate the UL37 cDNA libraries are indicated on the UL37 RNA. (B) HCMV UL37 proteins and ORFs. The UL37 ORFs derived from the newly described UL37 RNAs are shown. The N-terminal hydrophobic leader (cyt, aa 1–22), immediately downstream basic residues (aa 23–29) and the first anti-apoptotic domain (aa 5–34) (Hayajneh et al., 2001a) are present in all UL37 ORFs. pUL37x1, gpUL37x1, pUL37x1, gpUL37x1 and gpUL37x1 and UL37x1 di ORFs also include the UL37x1 acidic domain (aa 81–108), the second anti-apoptotic domain (aa 118–147) and the UL37x1 C terminus. gpUL37x1, pUL37x1 and gpUL37x1 further share the UL37x2 (aa 163–177) and the upstream UL37x3 (aa 178–197) sequences. The unique C-terminal sequences of pUL37x1, pUL37x1 and UL37x1 di ORFs are encircled. pUL37x1, pUL37x1 and UL37x1 di encode 83-residue proteins, which include aa 1–62 of the UL37x1 ORF and the encircled unique C terminus. The UL37x3 C-terminal N-glycosylation domain, downstream basic residues, and transmembrane and cytosolic tails are present in gpUL37x1 and gpUL37x1.
Since the UL37M cDNA was not isolated by our initial screening of the HCMV at 8 h p.i. UL37 cDNA library and because Goldmacher et al. (1999) cloned the UL37M cDNA at 27 h p.i., we examined temporal expression of the newly described UL37 alternatively spliced RNAs during HCMV infection. For these experiments, poly(A)+ RNA from HCMV-infected HFFs at 8, 16, 24, 48, 72 and 96 h p.i. was used to generate cDNAs. cDNAs (5 µg) were then amplified by PCR using primers 264 (nt 52541–52520) and 196 (nt 50542–50521), which can amplify the spliced junctions of all UL37 RNAs identified. The PCR product (703 bp) corresponding to UL37 RNA was predominantly detected at IE (16 h p.i.) times of infection, prior to HCMV oriLyt DNA replication, but was detectable as a low-abundance RT-PCR product throughout HCMV infection. This predominant IE expression of UL37 RNA appears to be slightly delayed after the 8 h p.i. period observed previously (Tenney & Colberg-Poley, 1991a, b), and might result from small
differences in cell physiology or multiplicity of HCMV infection. Consistent with UL37M cDNA isolation at 27 h p.i. by Goldmacher et al. (1999), the UL37M RNA was predominantly detected as an 448 bp PCR product at 24 h p.i. We also detected UL37M RNA in HCMV-infected HFFs at 96 h p.i. PCR products corresponding to the newly described UL37di (521 bp) and its related UL37 Sdi (220 bp) RNAs were detected at IE times (8 and 16 h p.i.) and at late times (96 h p.i.) of HCMV infection; the UL37S PCR product (402 bp) was predominantly detected at late times. To independently verify their identities, all of the indicated UL37 cDNAs detected at 96 h p.i. were cloned and sequenced. The predominant temporal expression of the newly described UL37 splice variants, based upon Fig. 2(B) and data not shown, is summarized in Fig. 1(A). The presence of the UL37x1 RNA from IE to late times of HCMV infection was verified by PCR amplification of these cDNAs using primers 264 and 140 (nt 52259–52236) (Fig. 2B, bottom).

Consistent with our previous detection of the UL37di and UL37dii donors in HCMV-infected cells at 16 h p.i. (Su et al., 2003), thorough screening of full-length UL37 cDNAs with a common forward primer and reverse primers against the newly described UL37di/x3A, UL37 dii/x3A or UL37x2/x3A junctions confirmed splicing of these donors in UL37 RNAs at 8 h p.i. Nonetheless, only the abundances of UL37x1, UL37Sdi and UL37 RNAs were sufficiently high to be readily detected by RT-PCR amplification of HCMV-infected cell poly(A)+ RNA (Fig. 2B). Consistent with these results, we note that some UL37 spliced variant cDNAs were only rarely obtained in the UL37 cDNA library (unpublished results). In addition, a prediction program for human splice sites (Berkeley Drosophila Genome Project, http://www.fruitfly.org/seq_tools/splice.html) predicts UL37x1S and UL37x1 L to be used as splice donors with scores of 0.0099 and 0.054, respectively (Table 1). The fact that more clones using the UL37x1S than with UL37x1L donor were isolated is in accord with their score for predicted

Fig. 2. (A) Specificity of primers spanning the UL37 di/x3A, dii/x3A and x2/x3A spliced junctions. To determine the suitability of primers specifically against the UL37 newly described spliced (di/x3A, dii/x3A and x2/x3A) junctions for screening of the UL37 cDNA library, cloned and sequenced UL37 Sdi (p1129) or UL37 dii (p1133) cDNAs (2 pg) were amplified by PCR with primers 264 and 257 (di/x3A), 258 (dii/x3A) or 259 (x2/x3A) at an annealing temperature of 68 °C. PCR amplification products were resolved by agarose gel electrophoresis. Markers (123 bp) served as molecular size standards. The approximate physical location and orientation of the primers on the corresponding UL37 spliced junctions are shown below the gel. (B) Differential temporal expression of alternatively spliced UL37 RNAs during permissive HCMV infection. cDNAs (~5 μg), reverse-transcribed from HCMV-infected HFF poly(A)+ RNA at 8, 16, 24, 48, 72 and 96 h p.i., were amplified by PCR using primers 264 and 196 (top) or 264 and 140 (bottom) at an annealing temperature of 60 °C. Poly(A)+ cDNA from uninfected HFFs and water served as negative controls for the PCR. The UL37 spliced cDNAs indicated on the right (top) were sequenced to verify their identities. The approximate physical location and orientation of primers 264, 140 and 196 on UL37 pre-mRNA are shown below the gels.
usage. The recently described UL37x3 donors, di and dii (Su et al., 2003), were also predicted with scores of 0.99 and 0.94, respectively. Both the UL37x3 and UL37x3A acceptors are also predicted with scores of 0.75 and 0.32, respectively, which correlates with their respective use. In addition to the newly described UL37 RNAs detected in HCMV permissively infected cells, the program also predicts the use of another donor splice site in UL37 intron 1 (0.97) at nt 51998–51984 (5′-CUACGGGUGAGGAA-3′), whose use we have not detected yet.

Tellingly, previous investigations of HCMV UL37 gene expression did not detect these newly described alternatively spliced variants by either S1 or Northern blot analysis (Goldmacher et al., 1999; Kouzarides et al., 1998; Tenney & Colberg-Poley, 1991a, b). However, the UL37x1 and UL37x3 probes were such that they might not efficiently detect some of the UL37 splice variants (Tenney & Colberg-Poley, 1991a, b). This, in combination with the fact that all of the newly identified alternatively UL37 spliced variants are present in exceedingly low abundances relative to the UL37x1 unspliced transcript, some ~100-fold less (unpublished results) and below detection by RT-PCR/gel detection, may account for why these UL37 species were previously undetected in HCMV-infected HFFs.

The reasons underlying the production of several alternatively spliced UL37 RNAs encoding the same polypeptide (e.g. UL37x1, UL37x3, UL37LM and UL37di or UL37o, UL37odi and UL37odii), or indeed why HCMV growth could require several truncated forms of the UL37 protein, are currently unknown. Alternatively, these rare UL37 transcripts may be non-functional transcripts. Nonetheless, alternative processing of known HCMV transcripts results in the production of functionally different gene products. In the best-studied locus, differential processing of the major IE pre-mRNAs leads to the production of multiply spliced and polyadenylated RNAs (reviewed by Mocarski & Tan Courcelle, 2001; Shirakata et al., 2002). Moreover, Alwine and his co-workers (J. A. Isler, S. Awasthi & J. C. Alwine, personal communication) have recently identified novel IE1 RNA splice variants, whose abundances differ during HCMV infection; however, their temporal expression is similar to that of IE1 mRNA. Products of the differentially spliced IE1 and IE2 transcripts differ in function (reviewed by Mocarski & Tan Courcelle, 2001). In addition, pIRS1263, a truncated version of pIRS1, is encoded by a transcript initiated at a promoter within the IRS1 ORF, is in the same reading frame as pIRS1 and is contained within its C-terminal domain (Romanowski et al., 1997). pIRS1263 is part of a regulatory loop, antagonizing transcriptional trans-activation associated with pIRS1 and pTRS1 (Romanowski & Shenk, 1997).

pUL37x1 and gpUL37 traffic through the secretory apparatus and into mitochondria, where they have anti-apoptotic activity (Al-Barazi & Colberg-Poley, 1996; Colberg-Poley et al., 2000; Goldmacher, 2002; Goldmacher et al., 1999; Hayajneh et al., 2001a, McCormick et al., 2003). It may be that truncated versions of the UL37 protein are required for efficient trafficking in different cell types or at different stages of HCMV infection. We are currently investigating this possibility and the production and abundance of UL37 alternatively spliced RNAs in HFFs and in other permissively infected cell types. Importantly, all of the UL37 splice variants are predicted to encode proteins that share N-terminal sequences. Even the short 83-residue products are predicted to retain the first highly conserved anti-apoptotic domain spanning aa 5–34 (Hayajneh et al., 2001a). It is possible that selected UL37 isoforms might be packaged into particles, as some UL37 spliced RNAs are expressed at late times of infection. In support of this possibility is the finding that only a specific isoform of pUL69 is selectively incorporated into HCMV particles (Winkler & Stamminger, 1996). Overall, we have been able to detect 11 UL37 RNA species (UL37x1, UL37, UL37M, UL37o, UL37op, UL37odi, UL37odi, UL37odi, UL37LM and UL37Ldi) by exhaustive cDNA cloning (Fig. 1A). Nonetheless, the UL37LM and UL37Ldi RNAs were only cloned as partial cDNAs during our sequencing analysis of the 96 h p.i. products. Although accurate splicing of the UL37x1/x2/x3A and UL37x1/x2/di/x3A spliced junctions has been verified, the full-length UL37LM and UL37Ldi cDNAs have not been cloned yet. This surprising complexity makes processing of the UL37 pre-mRNA amongst the most complex of any known HCMV transcript, including those from the major IE locus (Rawlinson & Barrett, 1993; Stenberg et al., 1989).

ACKNOWLEDGEMENTS

The authors thank Dr Robert Donaldson for his enthusiastic support during the course of these studies, Dr Jim Alwine for his personal communication and Dr Nancy DiFrancesco for her critical reading of the manuscript. This work was supported by Public Health Service grant AI-46459 from the National Institute of Allergy and Infectious Diseases and the Children’s Research Institutional Funds to A. C.-P.

REFERENCES


Colberg-Poley, A. M., Huang, L., Sotlo, V. E., Iskenderian, A., Schumacher, R.-F. & Anders, D. G. (1998). The acidic domain of pUL37x1 and gpUL37 plays a key role in transactivation of
HCMV DNA replication gene promoter constructions. Virology 246, 400–408.


