NEWLY IDENTIFIED TRANSCRIPTS OF UL4 AND UL5 GENES OF HUMAN CYTOMEGALOVIRUS

SHUANG GAO, SHAU RUAH, YAPING MA, MALI LI, LIN WANG, BO ZHENG, YING QI, ZHENGCHONG SUN, YUJING HUANG AND QIANG RUAN

Virus Laboratory, Affiliated Shengjing Hospital, China Medical University, Shenyang, Liaoning, 110004, China

Human cytomegalovirus (HCMV) UL4 and UL5 genes are two members of the RL11 gene family. In an earlier study, three UL4 transcripts of about 1.7, 1.5 and 1.4 kb were found in early and late classes after infection by the Towne strain by nuclease protection and primer extension analyses. In the present study, two UL4 transcripts (1.5 and 1.7 kb) were found by cDNA library screening, Northern blot, 3’ and 5’ RACE analyses to appear initially in the immediate early phase and one UL4 transcript (1.4 kb) in the late phase in a low-passage clinical isolate. Furthermore, two novel low-abundance UL5 transcripts with the same 3′ terminus as the identified UL4 transcripts in the UL4-UL5 gene region were found in late classes.

Key words: HCMV, UL4, UL5, transcript

Received: 30 June, 2014; revised: 12 November, 2014; accepted: 16 December, 2014; available on-line: 02 March, 2015

INTRODUCTION

Human cytomegalovirus (HCMV) contains a linear double-stranded genome of approximately 230 kilo base pair (kb), which has a potential to encode more than 165 proteins (Murphy et al., 2003; Dolan et al., 2004; Ma et al., 2012). As other herpesviruses, HCMV genes are expressed in a temporal cascade comprising immediate early (IE), early (E) and late (L) phases. The IE proteins are required for subsequent early gene expression, and the early products are required for viral DNA replication. After viral DNA replication, late gene expression occurs (Spector, 1996).

Nine multigene families have been identified in the AD169 strain of HCMV (Chec et al., 1990) and in three low-passage isolates (Davison et al., 2003a). The RL11 family consists of 12 genes, including RL11, RL12, RL13, UL1, UL4, UL5, UL6, UL7, UL8, UL9, UL10 and UL11, oriented from left to right near the left-hand terminus of the genome (Davison et al., 2003b).

As a member of the RL11 family, UL4 has been defined as an early gene encoding a 48-kDa subgenus-specific virion glycoprotein (Chang et al., 1989a). Two early (1.7 and 1.5 kb) and one late (1.4 kb) transcript regulated by three inducible promoters at various times have been identified in the UL4 region of the Towne strain by nuclease protection and primer extension analyses (Chang et al., 1989a; 1989b). In 2007, Zhang et al. found three transcripts overlapping the UL5 sequence with the 3′ end at nucleotide (nt) 14748 and the 5′ ends at nt 13758, nt 13906 and nt 13925, respectively, in a late cDNA library (Zhang et al., 2007). These results indicate a transcription unit in the UL4-UL5 gene region. However, there has been no report about the transcript structures of UL4 and UL5 in HCMV strains isolated from patients so far. In the current study, the transcription characteristics of the UL4-UL5 gene region were investigated in a low-passage isolate by cDNA screening, northern blotting and rapid amplification of cDNA ends (RACE). Two novel low-abundance UL5 transcripts with the same 3′ terminus as the earlier UL4 transcripts identified were found in late class RNAs.

MATERIALS AND METHODS

Cell, virus and RNA preparation. HCMV low-passage strain H was isolated from a urine sample of a congenitally HCMV-infected infant in the Pediatrics Department at the Affiliated Shengjing Hospital of China Medical University. MRC-5 cells were routinely cultured in minimal essential medium (MEM) containing 15% fetal calf serum and penicillin-streptomycin at 37°C, 5% CO2. After inoculation with virus, the MRC-5 cells were maintained in MEM supplemented with 2% fetal calf serum and penicillin-streptomycin at 37°C, 5% CO2 in an incubator.

For HCMV IE infection, 100 μl/ml of cycloheximide (Sigma, USA) was added to culture medium prior to infection and the cells were harvested 24 hours post infection (hpi). For E infection, 100 μl/ml of DNA synthesis inhibitor phosphonoacetic acid (Sigma, USA) was added to the medium immediately after virus inoculation, and the infected cells were harvested at 48 hpi. For L infection, the cells were harvested at 72 hpi without any drug treatment. Total RNA was isolated from infected and uninfected MRC-5 cells using TriZol reagent (Invitrogen, CA) according to the manufacturer’s instruction. Possible DNA contamination was removed from RNA preparations using DNA-free reagents (Ambion, USA). The quantity and purity of the RNA preparations were estimated by optical density measurements.

Northern blotting. Northern blot analysis was carried out according to a standard procedure established in...
our laboratory. Aliquots of 5 μg of RNA per lane from IE, E and L infected and mock-infected MRC-5 cells were subjected to denaturing agarose gel (1.5% [wt/vol]) electrophoresis in the presence of 5.4% formaldehyde, alongside digoxigenin-labeled RNA molecular weight marker I (Roche, USA). Probes were labeled with digoxigenin using a DIG Northern Starter kit (Roche, USA). Primers used to produce the RNA probes are listed in Table 1. In order to evaluate the effects of cycloheximide and phosphonoacetic acid on HCMV replication, transcripts of UL123, UL34 and UL99, which are IE, E and L genes, respectively, were detected in the same RNA preparations by Northern Blot. The separated RNA was transferred onto a positively charged nylon membrane by capillary transfer. Then, the membrane was baked at 80°C for 2 hours followed by pre-hybridization for 30 min at 60°C using the Dig EasyHybl buffer (Roche, USA) and hybridization with RNA probes for 16 hours at 60°C. Following washing twice with 2% SSC, 0.1% SDS buffer and 0.1% SSC, 0.1% SDS buffer at 50°C under constant agitation, the membrane was incubated with anti-digoxigenin antibodies conjugated to alkaline phosphatase and probes were visualized with the chemiluminescence substrate CDP-Star (Roche, USA) using ChemiDocTM XRS+ (Bio RAD, USA). To ensure equal RNA loading, the RNA preparations were adjusted basing on 28S and 18S rRNA levels estimated by ethidium bromide staining.

Screening of cDNA library. A full-length cDNA library of HCMV H strain has been constructed in pBlue-script SK vector before (Ma et al., 2011). Recombinants of the cDNA library were transferred into JM109 (Promega, USA). A total of 8600 clones were randomly picked up and inoculated into LB medium. A pair of UL5 gene specific primers (Table 1) were used to screen for UL5 clones from the cDNA library by graded polymerase chain reaction (PCR) as described before (Sun et al., 2010; Qi et al., 2011). The PCR reaction conditions were as follows: 95°C for 5 min, 30 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 30 s, followed by a final elongation at 72°C for 10 min. Inserts of the selected clones were sequenced using vector primers M13F and M13R on an ABI PRISM 3730 DNA analyzer (Applied Biosystems, CA).

5’ RACE (rapid amplification of cDNA 5’ends). For mapping 5’ ends of transcripts, 5’ RACE was performed using the 5’-full RACE Kit (TaKaRa, China) according to the manufacturer’s instructions. First-strand cDNA was synthesized with M-MLV reverse transcriptase using random 9-mer primers provided in the kit. Reactions without TAP (to account for the interference caused by 5’ phosphate of tRNA, rRNA and incomplete mRNA) and M-MLV (to determine the interference by contaminating DNA) were performed as two controls. UL4-UL5 gene specific cDNA sequences were amplified by nested PCR using specific primers R1 and R2, together with the 5’ RACE adaptor primer provided in the kit. The PCR conditions were as follows: initial denaturation at 94°C for 5 min, 30 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 1 min, followed by final elongation at 72°C for 10 min.

3’ RACE. For mapping 3’ end of transcripts, 3’ RACE was performed using 3’-full RACE Core Set Ver.2.0 (TaKaRa, China) according to the manufacturer’s instructions. First-strand cDNA was prepared using oligo-dT-adaptor primer and M-MLV. Nested PCR was then performed using specific primers F1 and F2, together with the oligo-dT-adaptor primer provided in the kit.

Cloning and sequencing. RACE products were gel-purified and inserted into PCR2.1 vector PCR2.1 (Invitrogen, China) with T4 ligase at 14°C overnight. The ligation products were transformed into E. coli DH5α competent cells. Ten to twenty clones of each purified PCR product were selected randomly and identified by PCR. Then, the insert sequences of the clones were confirmed by DNA sequencing on an ABI PRISM 3730 DNA analyzer (Applied Biosystems, USA). The nucleotide positions referred to in this study are in reference to the sequence of the HCMV Towne strain (GenBank: FJ616285.1). Open reading frame (ORF) was predicted using Editseq program of the DNA star package.

RESULTS

UL4-UL5 transcripts detected by Northern blot

Northern blots were repeated three times with different RNA preparations of IE, E and L viral expression phases. The RNA blots were first hybridized with

Table 1. Gene specific primers used in this study

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’-3’)</th>
<th>Genomic coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern Blot</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UL4-NF</td>
<td>AGATATGCTTTATGGTCTGAAG</td>
<td>13446-13469</td>
</tr>
<tr>
<td>UL4-NR</td>
<td>ATTTTGCGACGCTGTTGTTG</td>
<td>13898-13908</td>
</tr>
<tr>
<td>UL5-NF</td>
<td>TGCCCCGCTCTCGGAAATACC</td>
<td>14445-14465</td>
</tr>
<tr>
<td>UL5-NR</td>
<td>CAAGCACACCCCCAGCCAC</td>
<td>14631-14651</td>
</tr>
<tr>
<td>UL123-NF</td>
<td>GCCTTCCATAGCCCAAT</td>
<td>172316-172326</td>
</tr>
<tr>
<td>UL123-NR</td>
<td>ATTTTGGGCTAGACCTAAC</td>
<td>172216-172239</td>
</tr>
<tr>
<td>UL34-NF</td>
<td>TGCGTATACGCGACTGCGT</td>
<td>45494-45514</td>
</tr>
<tr>
<td>UL34-NR</td>
<td>AATGCTCGCGCTGCTGCC</td>
<td>44758-45740</td>
</tr>
<tr>
<td>UL99-NF</td>
<td>CGCGTCCTGAGGTGTTCC</td>
<td>144702-144721</td>
</tr>
<tr>
<td>UL99-NR</td>
<td>GATGGTGGTATTGGTTGTTG</td>
<td>145235-145252</td>
</tr>
<tr>
<td>cDNA library screening</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WS</td>
<td>TGTCCTTTTGCTGAGTTGTCG</td>
<td>14072-14093</td>
</tr>
<tr>
<td>WR</td>
<td>CGTCTAAACCTCGTCTGATGAT</td>
<td>14319-14296</td>
</tr>
<tr>
<td>5’ RACE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1</td>
<td>CACGCCCCACCGTGCTCTTA</td>
<td>14621-14601</td>
</tr>
<tr>
<td>R2</td>
<td>ACTACCGGAAAGCCCTCCCC</td>
<td>14594-14573</td>
</tr>
<tr>
<td>3’ RACE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>GACATCACCAGGCGGAGAG</td>
<td>14138-14158</td>
</tr>
<tr>
<td>F2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Note: Sequence positions are of the Towne strain (GenBank: FJ616285.1). Primer sequences for synthesis of UL123 probe were the same as those used in the reference (Stamminger et al., 1991).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Transcripts of UL4 and UL5 genes of human cytomegalovirus

The results showed that besides the major transcript of 1600~1800 nt detected in all three expression phases, a weak transcript of about 650 nt was detected by the UL5-specific probe in the L RNA preparation. No band was found in RNA preparation from mock-infected cells. Transcripts of UL123, UL34 and UL99 genes, which are IE, E and L genes (Welch et al., 1991; Stamminger et al., 1991; Adam et al., 1995), respectively, were detected in the corresponding RNA preparations (Fig. 2a). To determine whether the 1600~1800 nt transcript contains UL4 sequence or not, UL4-specific RNA probe was used. As shown in Fig. 2b, only the 1600~1800 nt transcript was detected by the UL4 probe.

UL4-UL5 transcripts screened from the HCMV cDNA library

Five cDNA clones selected by graded PCR were identified to contain the sequence congruent with the UL4-UL5 gene region. Sequencing results showed that all the five cDNA sequences were unspliced and had the same 3' end but different 5' ends. The 3' terminus was at nt 14750, which is downstream of a poly(A) signal (AATAAA) at nt 14728-14733, by comparing with the sequence of HCMV Towne strain. The 5' ends of the cDNA sequences were at nt 13028, 13226, 13233, 13328 and 13813, respectively. The corresponding lengths of the cDNA sequences were 1723 bp, 1525 bp, 1518 bp, 1423 bp and 938 bp, respectively (Fig. 1b).

5' and 3' termini of UL4-UL5 transcripts identified by RACE

In order to further identify the 5' and 3' termini of the UL4-UL5 transcripts detected by cDNA library screening and Northern blot, RACE experiments were employed with IE, E and L class RNAs of HCMV H. In 5' RACE experiments, a product of 1600 bp was obtained in all three classes of RNAs and three additional products of about 1300, 500 and 400 bp were amplified in the L class RNA using specific primers R1 and R2 (Fig. 3a). Sequencing results
showed that two 5' ends, at nt 13229 and 13026, were obtained in all three classes of RNAs, and three other 5' ends at nt 13328, 14071 and 14188 were detected in the L RNA only.

In 3' RACE experiments, a band of about 700 bp was obtained from all RNA preparations using specific primers F, and F2 (Fig. 3b). Consistent with the results obtained by cDNA screening, sequencing results of the band demonstrated a 3' end located at nt 14750 downstream of the consensus poly(A) signal (AATAAA) at nt 14728-14733.

Sequence analysis of the UL4 and UL5 transcripts

To obtain the complete sequence of the UL4 and UL5 transcripts, the sequences obtained by 5' RACE and 3' RACE were linked together according to their overlapping sequences. Detailed information of the linked transcripts is showed in Fig. 1c. The full lengths of the UL4 transcripts were 1423, 1522 and 1725 nt with the 5' ends located at nt 13328, 13229 and 13026, respectively. The UL5 transcripts, which initiated at nt 14188 and 14750, comprised 563 and 680 nt, respectively. The lengths of the UL4 transcripts were consistent with those identified in the Northern blots, and their sequences were the same as those obtained by cDNA library screening. However, the lengths of the UL5 transcripts were only in accord with those found in the Northern blots. Based on the DNA sequence, two non-conventional potential TATA promoter elements (TATTA and TATTTA) were predicted at nt 14039 and 14046, upstream of the identified UL5 transcripts (Fig. 4). The UL5 transcripts have a potential to encode an 81-amino-acid protein with a calculated molecular mass of 9 kDa.

DISCUSSION

It has been reported that the UL4 gene is transcribed into three transcripts of 1700, 1500 and 1400 nt, initiated at nt 13026, 13229 and 13313, respectively, and terminated at nt 14750 (GenBank: FJ616285.1). The two longer transcripts are transcribed in E infection phase and the shortest one in L infection phase (Chang et al., 1989a; 1989b). In the present study, the structures of the UL4 transcripts were confirmed in a low-passage HCMV strain. Except for the 1400 nt transcript, the structures of the other two UL4 transcripts were completely consistent with those found in the previous studies (Chang et al., 1989a; 1989b).

HCMV IE genes are expressed as the critical first step in virus infection. Only a few genes, including UL36-UL38, UL115-UL119, UL122-UL123, US3 and RK17/RK17, within the ~230-kbp HCMV genome have been found to be transcribed at the IE phase of infection up to now (Oduro et al., 2012). IE proteins impair many cellular functions to facilitate later phases of infection, including cellular DNA synthesis, STAT signaling, apoptosis, and so on (Murphy et al., 2000; Wiebusch et al., 2001; Skaletskaya et al., 2001; Child et al., 2004; Paulus et al., 2006; Marshall et al., 2009; McCormick et al., 2010). The UL4 gene has been defined as an early gene according to its transcription being detected primarily in E infection phase (Chang et al., 1989a). The UL4 transcripts are regulated differentially by three promoters during the course of infection (Chang et al., 1989b). However, the 1725 and 1522 nt transcripts of UL4 gene, initiated at nt 13026 and 13229, were detected initially in the IE infection phase by 3' RACE, 5' RACE and Northern blot in the present study. This result indicated that transcription of UL4 gene may be active as early as in the IE infection phase. The meaning of the UL4 gene expression during IE phase needs to be investigated further.

In the present study, one monocistron containing the UL5 sequence was confirmed by Northern blot and RACE. These results showed that, in addition to transcription together with the UL4 gene, UL5 can be transcribed independently in late infection phase. The monocistron originates from two different sites and has the same 3' end with that of the UL4 transcripts. However, no UL5 cDNA clone was found in the HCMV cDNA library. The reason for the failure to detect UL5 transcripts in the cDNA library could be the lower abundance in the infected...
MRC5 cells. Northern blot was performed several times. Although the RNA preparations used in these experiments were not from the same batch of cells, the UL5 transcripts were always detected in L cell RNAs. These results indicated that the UL5 gene is a real late class gene. Murphy et al. have identified that UL5 has a potential to encode protein (2003). Existence of two non-canonical TATA elements (TATTAT and TATTTA) upstream of the mapped initiation sites provided evidence for the UL5 transcripts to be conventional mRNAs.

In the present study, two UL4 transcripts were found to be transcribed initially in the IE phase in a low passage HCMV strain, and two UL5 transcripts with the same 3' termini as the UL4 transcripts were identified during the late phase at a relatively low level. Detailed information on HCMV transcripts may benefit understanding HCMV pathogenesis, finding new diagnostic targets and establishing new strategies for prevention of HCMV diseases.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (30672248, 81171580, 81171581 81201274 and 81371788) and the Specialized Research Fund for the Doctoral Program of Higher Education (2011204110012) and the Outstanding Scientific Fund of Shengjing Hospital.

REFERENCES


