Overlapping transcription structure of human cytomegalovirus UL140 and UL141 genes

YANPING MA, MALI LI, BO ZHENG, NING WANG, SHUANG GAO, LIN WANG, YING QI, ZHENGRONG SUN and QIANG RUAN*

Virus Laboratory, the Affiliated Shengjing Hospital, China Medical University, Shenyang, Liaoning, P.R. China, 110004

**Corresponding author (Fax, +86-024-23892617; Email, ruanq@sj-hospital.org)*

Transcription of human cytomegalovirus UL/b′ region has been studied extensively for some genes. In this study, transcripts of the UL140 and UL141, two of the UL/b′ genes, were identified in late RNAs of three HCMV isolates using Northern blot hybridization, cDNA library screening and RACE-PCR. At least three transcripts with length of 2800, 2400 and 1700 nt, as well as a group of transcripts of about 1000–1300 nt, were found in this gene region with an accordant 3′ ends. Among the transcripts, two initiated upstream of the start code of the UL140 gene and contained the UL140 and UL141 open reading frame (ORF), one initiated in the middle of the UL140 gene, and could encode short ORFs upstream of the UL141 ORF. A group of transcripts initiated upstream or downstream of the start code of the UL141 gene, and could encode 'nested' ORFs, including the UL141 ORF. These 'nested' ORFs possess different initiation sites but the same termination site as that of the UL141 ORF.

[Ma Y, Li M, Zheng B, Wang N, Gao S, Wang L, Qi Y, Sun Z and Ruan Q 2013 Overlapping transcription structure of human cytomegalovirus UL140 and UL141 genes. *J. Biosci.* **38** 35–44] **DOI** 10.1007/s12038-012-9293-4

1. Introduction

The human cytomegalovirus (HCMV; species: human herpesvirus five; genus: *Cytomegalovirus*; family: Betaherpesvirinae of Herpesviridae) genome consists of 230 to 235 kb of double-stranded DNA and had been predicted to encode 165 open reading frames (ORFs) in the laboratory strain of AD169 (Chee *et al.* 1990; Murphy *et al.* 2003). The UL/b′ region of HCMV DNA, containing at least 19 ORFs (UL133–UL151), was found in the Toledo strain (EMBL, U33331) and several other low-passage clinical strains, but was deleted in the laboratory AD169 strain (EMBL, X17403) (Cha *et al.* 1996). Some of these new ORFs and their predicted products have been proven crucial for viral infection *in vivo*. The products of UL133 and UL138 genes contribute to HCMV latency (Goodrum *et al.* 2007; Petrucelli *et al.* 2009; Petrucelli *et al.* 2012). The UL144 gene encodes type 1 membrane protein similar to the tumour necrosis factor receptor that is involved in immune regulation in HCMV infection (Poole *et al.* 2008). The UL146 gene encodes a viral CXC (α) chemokine, which binds with high affinity to the CXCR2 receptor and contributes to the induction of neutrophil chemotaxis (Penfold *et al.* 1999; Saederup and Mocarski 2002). It has been reported that all these important genes have complicated transcript structures (Lurain *et al.* 2006; Grainger *et al.* 2010; Wang *et al.* 2011; He *et al.* 2012). Studies on the transcription characteristics of these ORFs have provided further information to understand their functions.

Another member of the UL/b′ region, the UL141 gene, is capable of suppressing natural killer (NK) cell recognition and induces protection against NK cell-mediated cytolysis by down-regulating cell surface expression of CD155 and CD112 (Prod'homme *et al.* 2010; Tomasec *et al.* 2005). It has been reported that the UL141 ORF and its neighbour, the UL140 ORF, are located within one transcription unit (Qi *et al.* 2011). However, the transcription characteristic of this region has not been investigated extensively. In the current study, the transcripts of the UL140 and UL141 genes were characterized in three clinical strains.

Keywords. Cytomegalovirus; transcription; UL140; UL141

Supplementary materials pertaining to this article are available on the *Journal of Biosciences* Website at *http://www.ias.ac.in/jbiosci/mar2013/supp/Ma.pdf*

Published online: 21 January 2013

2. Materials and methods

2.1 *Virus and specimens*

Three HCMV clinical strains of H, C and X were isolated from urine samples of three infants less than 5 months old, who had been hospitalized in Shengjing Hospital of China Medical University. The strains were passaged 10 times in MRC5 cells, which were maintained in MEM medium supplemented with 2% fetal calf serum and 100 units penicillin/streptomycin at 37°C and 5% CO_2 in a humidified incubator. MRC5 cells were inoculated with the strains at a multiplicity of infection (MOI) of 3-5.

2.2 *DNA amplification and sequencing of the UL139-UL141 region of HCMV C and X strains*

Virus DNAs were extracted from the infected cells by boiling the cells with lysis buffer for 15 min. Based on H strain sequence (Accession number: Q981646), four pairs of primers were designed to amplify the DNA sequences of C and X strains from nt 7350 to nt 9931, which containing UL139–UL141 gene region. The sequences of the primers are listed in table 1 and the positions are shown in figure 1.

2.3 *RNA preparations*

For preparation of HCMV immediate-early (IE) RNA, 100 μL/mL of protein synthesis inhibitor cycloheximide (CHX, Sigma, USA) was added to the culture medium 1 h before infection and the cells were harvested at 24 h post infection (hpi). For early (E) RNA preparation, 100 μL/mL of DNA synthesis inhibitor phosphonoacetic acid (PAA, Sigma, USA) was added to the medium immediately after infection and the cells were harvested at 48 hpi. HCMV late (L) RNA and mock-infected cellular RNA were derived from infected and uninfected cells, respectively, cultured in parallel and harvested at 96 hpi. Total RNAs were isolated from approximately 10^7 infected or uninfected MRC5 cells using TRIzol agent (Invitrogen, Carlsbad, CA). The isolated RNAs were treated with DNA-Free reagent (Ambion, Austin, USA) to remove possible contaminating DNA. The integrity and size of the isolated RNAs were analysed by formaldehyde agarose gel electrophoresis. The quantity and purity of the RNAs were estimated by optical density value detection. To evaluate the inhibitory effects of CHX and PAA, the expressions of UL123, UL55 and UL99, which are IE, E and L genes respectively, were detected by RT-PCR. Only those with UL123-positive but UL55-negative are believed to be real IE RNA preparations; and those with

Table 1. The primers used in the present study

Experiments	Primer names	Primer sites (5′–3′)	Primer sequences (5′-3′)
DNA amplification	UL139 F	7350–7349	TAACGCTGATTGTTACGACA
	UL139 R	7968–7951	CCGAAGCCTTAGCCTCTA
	UL140 F	7866–7883	CCACGTTGAATCTGACGC
	UL140 R	8644–8625	ATTGCTACCTCGGATGTTTA
	UL141 F1	8614–8633	ACGTCCCTTCTTAAACATCC
	UL141 R1	9388–9369	CTGCGTGAGAATTACAAAGC
	UL141 F2	9364–9383	GCAACGCTTTGTAATTCTCA
	UL141 R2	9929–9908	ATGCGTGTTCTCACCTCTTCAT
5′RACE	UL140 5′outer	8595–8576	CTGATGAAGCTGCCAAGAGT
	UL140 5′inner	8543–8526	CGTTTCACCGCTTTCTGC
	UL141 5′outer	9789–9768	CATAGGAAACATACGGTGA AAA
	UL141 5′inner	9494–9477	GATTCCGTAGACGCCAGG
3′RACE	UL140 3′ outer	8280–8298	TTCCCAGAGCGACGACGAG
	UL140 3′ inner	8302–8322	CGTACTAACAGCGTGTCATCG
	UL141 3′ outer	8939–8958	CTGTTCTGGGTGCTGTTGAG
	UL141 3′ inner	9205–9222	TCGGCTGATGAACGGACT
Northern blot probes	UL140 Northern F	8280–8298	TTCCCAGAGCGACGACGAG
	UL140 Northern R*	8595–8576	CTGATGAAGCTGCCAAGAGT
	UL141 Northern F	9477–9494	CCTGGCGTCTACGGAATC
	UL141 Northern R*	9789–9768	CATAGGAAACATACGGTGA AAA

*The additional oligonucleotide of 5′-AATACGACTCACTATAGG-3′ was added to the 5′ ends of the reverse primers for the templates of the Northern blot probes and acted as the promoter of T7 RNA polymeras.

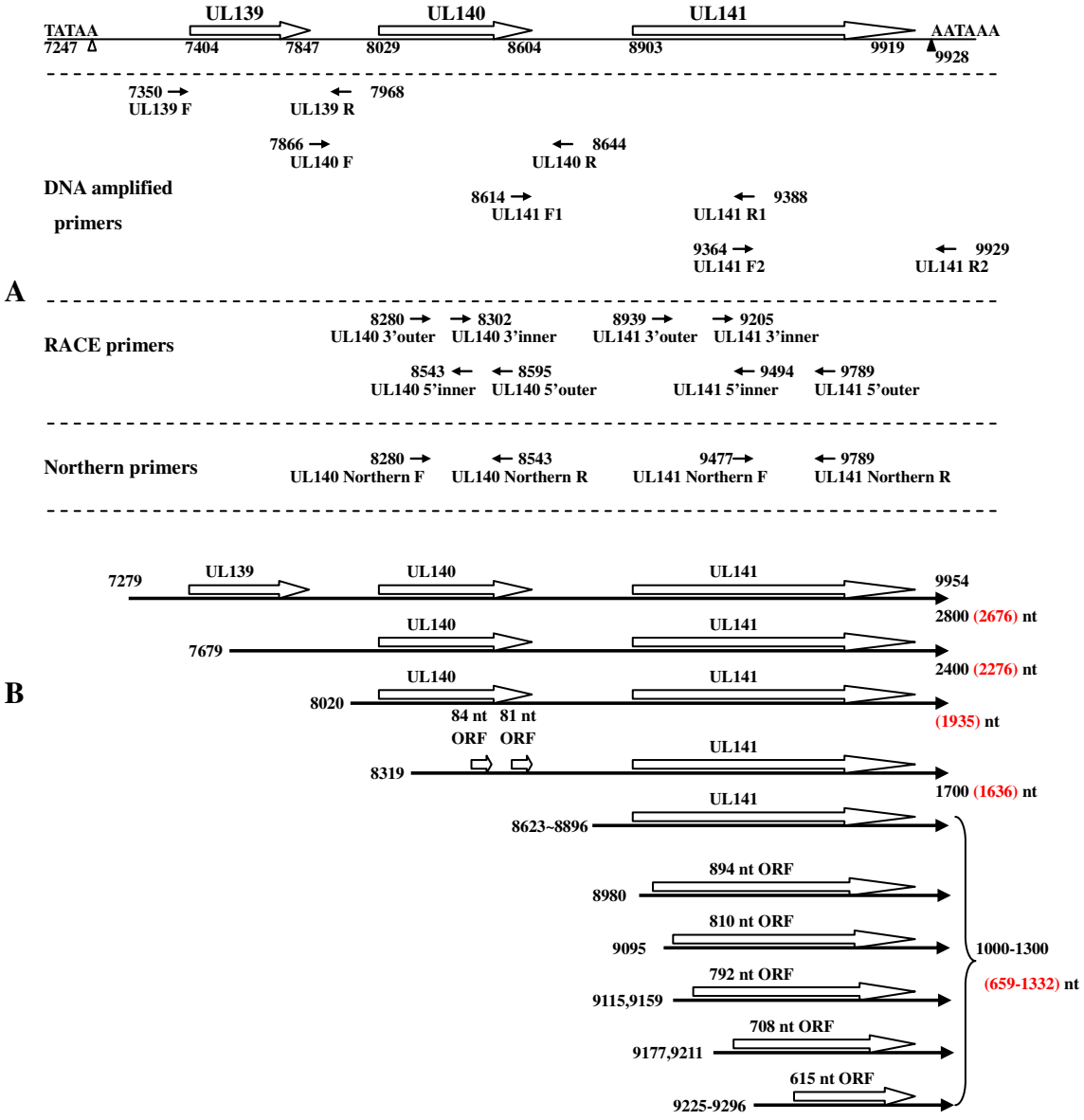

Figure 1. Genome structure of the UL139–UL141 gene region of H strain (GenBank GQ981646). The blank arrows show ORFs. The TATA element and polyA signal are marked. (**A**) Relative positions of the primers to the sequence of GenBank GQ981646 used in this study. The relative positions of the DNA amplified primers, RACE primers and Northern primers were marked respectively in three lines. (**B**) Transcripts obtained in this study. The predicted ORFs are marked on the transcripts. The 5′ ends of the transcripts are labelled on left side of the transcripts, and the lengths of transcript achieved in Northern blot are marked on right side; those in brackets are the actual sizes of the transcripts.

UL55-positive and UL99-negative are used as E RNA preparations in the study.

2.4 *Northern blot*

For Northern blot analysis, 10 μg per lane of IE, E and L RNA preparations of the three clinical strains and total RNA from mock-infected MRC5 cells were subjected to denaturing agarose gel (1% [wt/vol]) electrophoresis in the presence of formaldehyde, alongside the digoxigenin-labelled RNA molecular weight marker I (Roche, Indianapolis, IN, USA). Probes were labelled using a DIG Northern starter kit (Roche, Indianapolis, IN, USA) according to the manufacturer's instructions. Primers for producing the probes are listed in table 1 and shown in figure 1. The separated RNA

fragments were transferred onto positively charged nylon membranes using capillary transfer. Then, the nylon membranes were baked at 80°C for 2 h followed by prehybridization for 30 min at 65°C using the Dig EasyHyb-buffer (Roche, Indianapolis, IN, USA). After overnight hybridization to the probes at 65°C, the membranes were washed according to the manufacturer's instructions. The hybridized probes were incubated with anti-digoxigenin conjugated to alkaline phosphatase and were then visualized with the chemiluminescence substrate CDP-Star (Roche, Indianapolis, IN, USA). The membranes were exposed using ChemiDocTM XRS+ (Bio RAD, USA). To ensure that equal amounts of RNA were loaded, the RNA preparations were adjusted by comparing to the quantities of the 28S and 18S rRNAs in the same RNA preparations estimated by electrophoresis and ethidium bromide staining.

2.5 *RACE*

Experiments of rapid amplification of cDNA 3′ ends (3′ RACE) and 5′ ends (5′ RACE) were performed with 3′-Full RACE Core Set Ver.2.0 and 5′-Full RACE Kit (TaKaRa, Dalian, China), respectively. L RNA preparations of the three strains were used as templates. First-strand cDNAs were synthesized with MMLV reverse transcriptase using oligo-dT-adaptor primers (3′RACE) and random 9-mer primers (5′RACE), respectively. Nested PCR amplifications were carried out using LA Taq (TaKaRa, Dalian, China) after reverse transcription. All of the primers are listed in table 1 and shown in figure 1. The reactions were carried out at 94°C for 4 min, 30 cycles of 94°C for 30 s, 55–58°C (depending on primers) for 30 s, and 72°C for 30 s–3 min, with a final extension at 72°C for 10 min. In 5′RACE experiments, two control reactions were performed in strict accordance with kit instructions: (i) TAP(-), omitting tobacco acid pyrophosphorylase, to rule out non-specific ligation of incomplete mRNA, tRNA and rRNA with the Adaptor. (ii) MMLV(-), omitting MMLV reverse transcriptase, to rule out non-specific reaction caused by contaminating genomic DNA.

2.6 *Cloning and sequencing*

Products of PCR and RACE were separated by agarose gel electrophoresis and were purified using the DNA Purification Kit (Promega, Madison, WI, USA). Recovered PCR products were ligated into a pCR 2.1 TA vector (Invitrogen, China) with T4 ligase at 14°C overnight. The ligation products were transformed into E. coli DH/5α competent cells. Three to six clones of each purified PCR product were selected randomly for sequencing using the M13 primers and the ABI PRISM 3730 DNA analyser (Applied Biosystems, Carlsbad, CA).

2.7 *Screening a HCMV cDNA library*

An HCMV cDNA library had been constructed previously using the SMART technique (Clontech, USA) and the L RNA of HCMV H strain (Ma *et al.* 2011). To select specific cDNA clones from the cDNA library by polymerase chain reaction (PCR), a graded PCR was set up as previously described (Qi *et al.* 2011; Sun *et al.* 2010). Five-thousand cDNA clones were screened by graded PCR using the primers of UL141 Northern F and UL141 Northern R (table 1; figure 1). The PCR condition was initial denature of templates at 94°C for 4 min, 30 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 1 min, followed by a final elongation of 72°C for 10 min. Inserts of the selected clones were sequenced using vector primers (M13F and M13R) and the ABI PRISM 3730 DNA analyzer (Applied Biosystems, Carlsbad, CA). The screening results allowed us to obtain cDNA clones containing gene-specific transcript sequences.

2.8 *BLAST search and sequence analysis*

Standard nucleotide-nucleotide BLAST was performed on the NCBI Website. The nucleotide positions in this study are in reference to the sequence of the HCMV H strain UL/b′ region (GenBank: GQ981646.1). DNA alignment was done by MegAlign using Clustal W algorithms. ORFs prediction was performed using Editseq program of the DNAstar package.

3. Results

3.1 *UL139-UL141 DNA sequences of the HCMV C and X strains*

The DNA sequences comprising the UL139-UL141 gene region of C and X strains were obtained, which consisted of nt 7350 to nt 9931 using the HCMV H strain (GenBank: GQ981646.1) as the reference sequence. The UL139–UL141 sequence of C strain and the sequence comprising the UL140–UL141 gene region of X strain showed a high level of identity to that of H strain. However, the UL139 sequence of X strain had 22 nt deletions and 49 nt substitutions, and had an 84.0% identity score with respect to the H strain. The UL139–UL141 sequences of C and X strains were deposited in the GenBank database (GenBank: JX399887, JX399888).

3.2 *Northern blot*

To investigate the sizes and kinetics of UL140 and UL141 transcripts, Northern blots were performed using the total RNAs from MRC5 cells infected with the three HCMV strains of H, C and X at IE, E and L phases. The total RNA of mock-infected cells was used as a control. To amplify the DNA templates for labelling the Northern blot probes, the primers of UL140 Northern F and UL140 Northern R, as well as UL141 Northern F and UL141 Northern R were designed within the predicted UL140 and UL141 genes, respectively (table 1; figure 1).

The RNA blots were first hybridized with the digoxigenin-labelled UL141 probe, which was complemented to a 313 bp sequence within the predicted UL141 gene at nt 9477–9789. Four bands were detected in the L RNAs of all the three strains, but not in the IE, E and mock-infected RNAs (figure 2). The longest transcript was about 2800 nt, and the shorter two were about 2400 nt and 1700 nt, respectively. The shortest one was about 1000–1300 nt, which looked as being composed of a group of transcripts.

Then, the L RNAs of the three strains were hybridized with the digoxigenin-labelled UL140 probe, which was complemented to a 316 bp sequence within the predicted UL140 gene at nt 8280–8595. The UL140 probe detected all the bands hybridized with the UL141 probe, except for the shortest one.

3.3 *Identification of the 3′ termini of the UL140 and UL141 transcripts by 3′RACE*

To identify the 3′ ends of the UL140 and UL141 transcripts by 3′RACE, gene-specific primers of UL140 3′ outer and inner primers, as well as UL141 3′ outer and inner primers, were designed in the region of the predicted UL140 and UL141 genes, respectively (table 1; figure 1). cDNAs were reverse-transcribed from L RNAs of the three strains with 3′RACE adaptor primer containing oligo-dT17 (table 1). The 3′RACE products using the UL140 primers were about 1800 bp in all the three strains, and those using the UL141 primers were about 750 bp (figure 3A). Sequencing results demonstrated that both the 3′ ends of the UL140 and UL141 transcripts were located at nt 9954, which is downstream of a consensus poly (A) signal at nt 9928–9933.

3.4 *Identification of the 5′ termini of the UL140 and UL141 transcripts by 5′RACE*

To identify the 5′ ends of the UL140 and UL141 transcripts by 5′RACE, gene-specific primers of UL140 5′ outer and inner primers, as well as UL141 5′ outer and inner primers, were designed in the region of the predicted UL140 and UL141 genes (table 1; figure 1). cDNAs were reverse-transcribed from L RNAs of the three strains with Radom 9 mers. After nested PCR amplification, four main products of about 260, 500, 900 and 1200 bp were obtained with UL140 primers in all the three strains, respectively. All of the four products were recovered. Multiple products were obtained in all the three strains when amplified with UL141 primers, and the products less than 750 bp were recovered respectively. The recovered DNA segments were purified, cloned and sequenced. The 5′ ends obtained by UL140-specific primers were accordant among all of the three strains (table 2). Multiple 5′ ends were obtained using UL141-specific primers in the three strains (table 3). They were distributed from nt 8623 to nt 9296. This result is coincident with the shortest band of about 1000–1300 nt found in the Northern blot. Among them, the 5′ end at nt 8767 was identified both in H and C strains, and another 5′ end at nt 9225 was identified both in H and X strains.

3.5 *cDNA library screening*

Since the transcripts from the UL140 and UL141 gene region are 3′ co-terminal located downstream of the UL141 gene, the primers of UL141 Northern F and UL141 Northern R (table 1; figure 1) were used to screen the transcripts derived from the UL140 and UL141 gene region in a HCMV cDNA library by graded PCR. Four cDNA clones were identified to contain sequences congruent with the UL141 gene region from 5000 clones of the library. All of the four sequences possessed a polyA tail that was not coded by the HCMV genome. Compared with the DNA sequences of H strain, the 3′ ends of the four sequences were all located at nt 9954 downstream of a polyA signal (AATAAA), at nt 9928-9933, which is consistent with the results of 3′RACE. The 5′ ends of the four sequences were located at nt 8293, 8665, 8718 and 8877, respectively. They were not consistent with any 5′ end obtained by 5′RACE. These results may due to the truncated mRNAs. Furthermore, the cDNA library could not contain all of the transcripts accumulated during infection, especially those expressed in small quantities.

3.6 *Analysis of the full-length transcripts*

To obtain the full-length transcripts, sequences obtained from 5′RACE and 3′RACE were linked together on the basis of their overlapping sequences. The sequences from RACE were consistent with the corresponding DNA sequences of the three strains, and no introns exist. Detailed information on the linked transcripts is listed in tables 2 and 3. The 2676 nt transcript initiates at −125 nt upstream of the start code of UL139 gene and comprise the full sequence from UL139 to UL141genes. This transcript is consistent

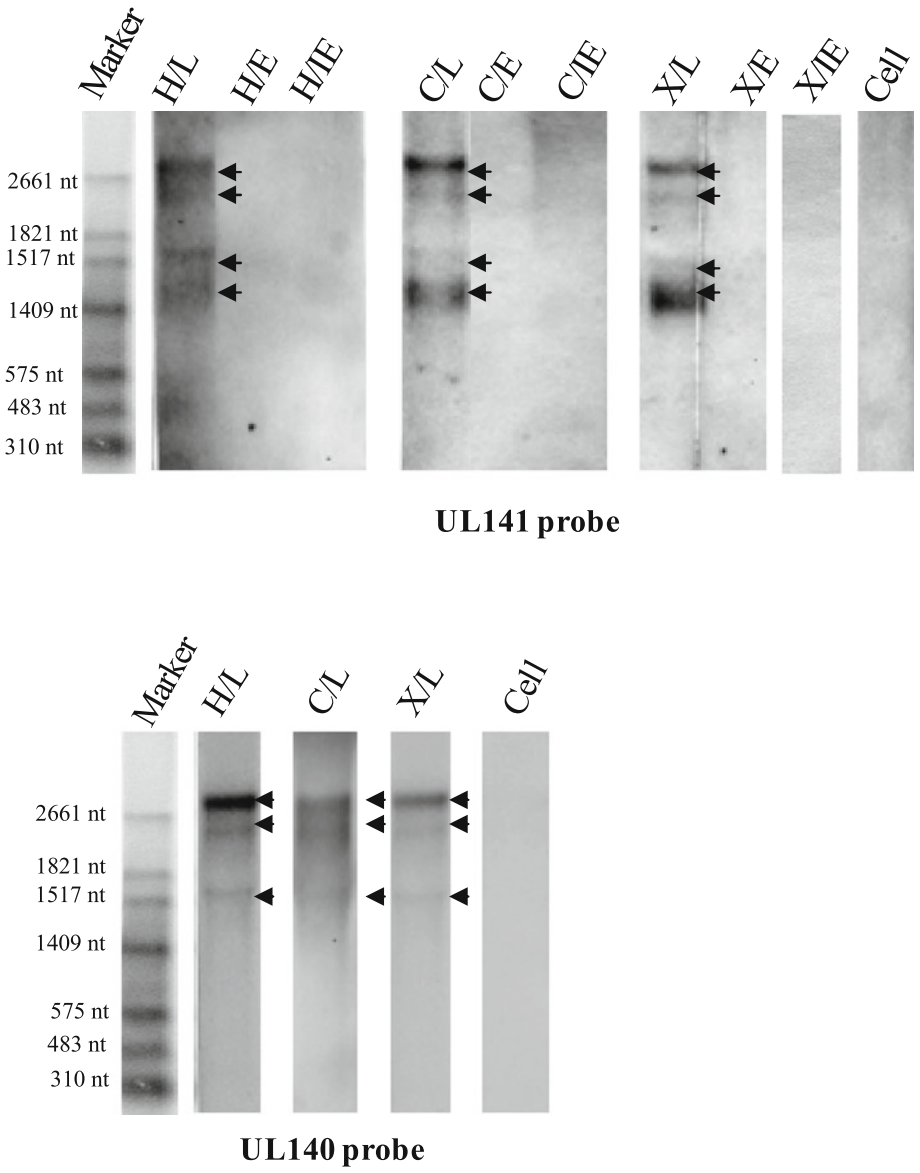

Figure 2. The results of Northern blots. RNA preparations from HCMV clinical strains H, C and X in immediately early (IE), early (E) and late (L) kinetics classes were used in this analysis. RNA from non-infected MRC5 cells was used as a negative control. The Northern blot was performed with two different probes, UL140 probe and UL141 probe. The transcripts are indicated by arrows. The positions of RNA size markers are indicated.

Table 2. Information of the transcripts initiated upstream of the UL140 ORF

Products of 5′ RACE (bp)	Sites of 5′ ends (nt)	Lengths of linked transcripts (nt)	Lengths of transcripts detected by Northern blot (nt)
1200	7279	2676	2800
900	7679	2276	2400
500	8020	1935	–
260	8319	1636	1700

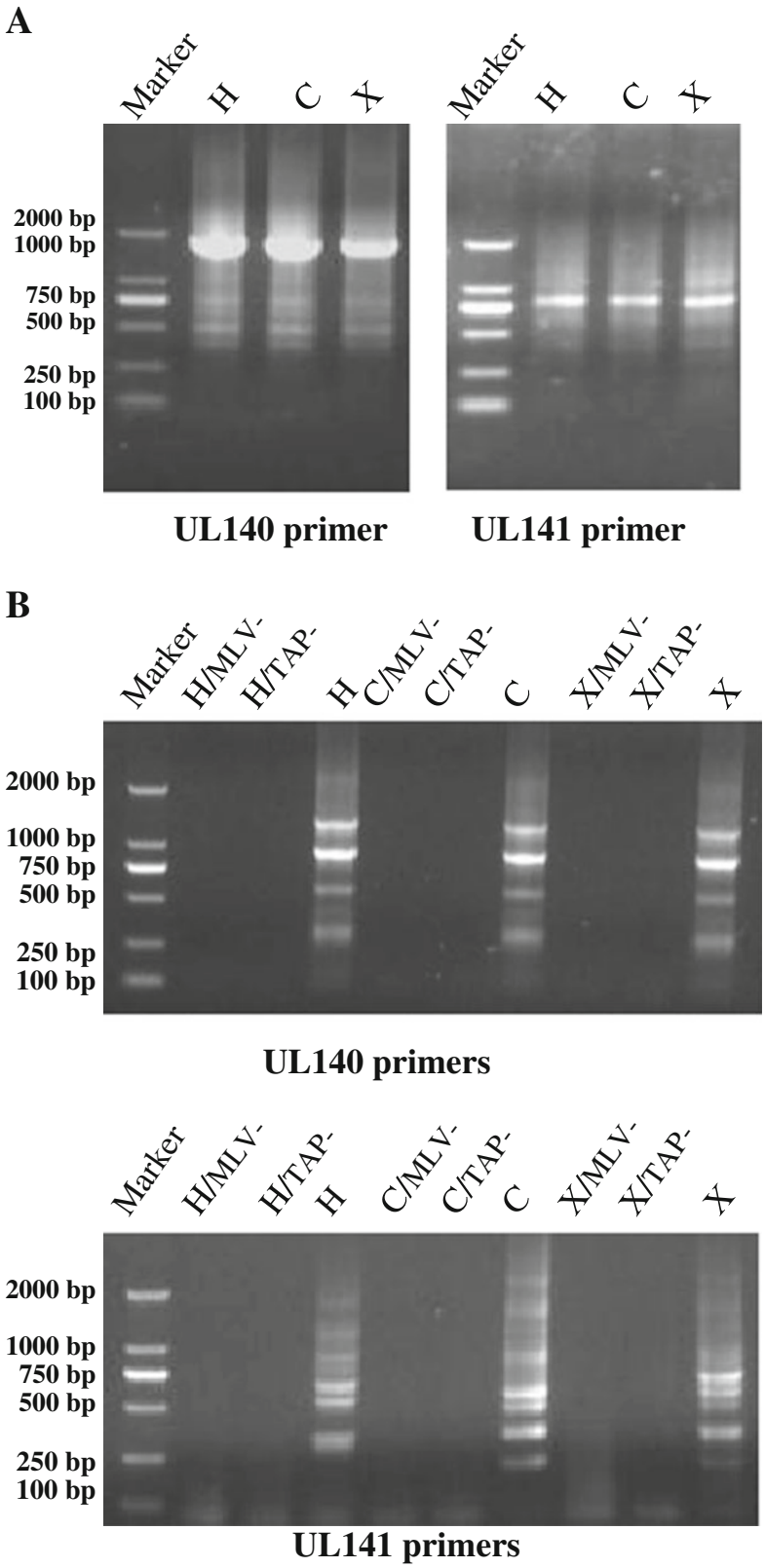

Figure 3. The 3′RACE (**A**) and 5′RACE (**B**) results obtained using UL140 and UL141 primers as well as the L RNA preparations. In the results of 5′RACE, all the controls of TAP(-) and MMLV(-) were negative in three strains.

Table 3. Information of the transcripts initiated from nt 8623–9296 near the start code of UL141 ORF

Sites of 5′ ends (nt)	Lengths of linked transcripts (nt)	Sites of the start code of the encoded ORFs (nt)	Lengths of the encoded ORFs (bp)
8623,8650,8767, 8831,8842,8880, 8896	1058–1332	8903	1017
8980	975	9026	894
9095	860	9110	810
9115	840	9128	792
9159	796		
9177	778	9212	708
9211	744		
9225	730	9305	615
9261	694		
9290	665		
9291	664		
9296	659		

with the 2800 nt band detected by Northern blot when added with a polyA tail. On the DNA sequences of the three strains, a consensus TATA element (TATAA) could be found at −32 nt upstream of the 5′ end of this transcript. The 2276 nt transcript initiated in the middle of the UL139 gene and was detected by Northern blots with the length of 2400 nt. A non-consensus TATA element (TAGTAA), an E2F site (TTTGCCG) and a NIT2 site (TATCT) were located within 90 nt upstream of the 5′ end of this transcript. The 1935 nt transcript initiated at −9 nt upstream of the start code of the UL140 gene. However, this transcript was not detected by Northern blots in all of the three strains. A DNA segment (TCGGTGACGCAGGC) upstream of the 5′ end of this transcript should be a composite element comprising three regulation sites of CREB (TCGGTGACGCAGG), AP-1 (GGTGACGCAG) and ATF (CAGGC). The 1636 nt transcript, which was detected by Northern blot with the length of 1700 nt, initiated in the middle of the UL140 gene. Another composite element (CGAGGTCAT) consisted of a CF1 site (GAGGTCAT) and a RORalp site (CGAGGTCA) could be found at −69 nt upstream of this transcript.

As shown in figure 1, the 2676 nt transcript comprised all the three ORFs from UL139 to UL141, and both of the 2276 nt and 1935 nt transcripts comprised the UL140 and UL141 ORFs. The 1636 nt transcript of all the three strains could encode a short ORF of 81 bp in its 5′ end, spanning nt 8524 to nt 8604, which was upstream of the previously predicted UL141 ORF. Another short ORF of 84 bp, spanning nt 8435 to nt 8518, was predicted in the 1636 nt transcript of H strain, owing to the nucleotide mutation from ‘ACG’ to ‘ATG’ at the DNA sequence of nt 8435–8437 (figure 1; supplementary figure 1). By analysing the corresponding DNA sequences of HCMV strains in GeneBank, the 81 bp ORF was identified in all of the strains listed in the BLAST result. However, the 84 bp ORF was found only in five HCMV strains (GenBank: GQ981646.1; GU585486.1; AY255776.1; AY218857.1; AY218851.1).

Among the transcripts with the lengths of 1000–1300 nt, some initiated upstream of the start code (nt 8903) of the UL141 gene, and encode the intact UL141 ORF. The others initiated downstream of the start code, and could encode some ‘nested’ ORFs, which possess different initiation sites but the same termination site as that of the UL141 ORF. These ‘nested‘ ORFs were identified in all of the HCMV strains published in GeneBank, except for the 615 bp ORF, which was absent in C strain and some other strains owing to the start code of ‘ATG’ mutated to ‘ACA’. Detailed information on the transcripts initiated from nt 8623 to 9296 is depicted in table 3 and supplementary figure.

4. Discussion

At least four large families of 3′-coterminal transcripts are transcribed from the HCMV UL/b′ region, containing UL133–UL138, UL139–UL141, UL142–UL145 and UL146–UL132. It has been reported that the UL139-UL141 transcription unit comprises UL139, UL140 and UL141 genes. One long transcript composed of all the three genes has been identified to initiate upstream of the UL139 ORF (Qi *et al.* 2011). In this study, other transcripts from this unit comprising the UL140 and UL141 genes were investigated extensively on their kinetics and structures in three HCMV clinical strains.

In Northern blots, four bands were detected by the UL141 probe and three bands by the UL140 probe in the L RNAs of all the three strains This result demonstrated that the transcripts from this unit are transcribed in late infected phase. Combined with the results of 3′RACE, these transcripts were identified to terminate with a polyA signal (AATAAA)

downstream of the UL141 ORF. Consistent with the result of Qi et al (Qi *et al.* 2011), the longest transcript of 2676 nt was initiated at nt 7279 upstream of the UL139 ORF. The 5′ end of nt 8020 was obtained in the UL140 5′RACE results of all the three strains. However, the 1935 nt transcript initiated with this 5′ end could not be detected by Northern blots. This result could be due to a low expression quantity of this transcript.

The 2276 nt and 1935 nt transcripts encode the same gene products but with distinct 5′-untranslated region (UTR). Different promoters should regulate the two transcripts. Overlapping mRNAs encoding the same products have been observed in some HCMV gene regions, such as the UL4 and UL44 gene regions (Chang *et al.* 1989a; Chang *et al.* 1989b). It has been presumed that heterogeneous 5′ ends of mRNA may play a role in posttranscriptional regulation of gene expression, e.g. at different stages of the infectious cycle (Chang *et al.* 1989a). However, the two transcripts detected in this study are both transcribed in late infected phase. Possibly, this reduplicate transcription pattern could be propitious to increase the quantity of protein product.

The 1636 nt transcript could encode short ORFs of 84 bp and 81 bp upstream of the UL141 ORF. It has been reported that three analogous upstream ORFs (uORFs) are present on the 5′ end of the UL4 transcript. Among them, the uORF2 could inhibit the translation initiation by the downstream UL4 AUGs (Degnin *et al.* 1993). The function of uORF may be to down-regulate expression of the downstream ORF (Geballe and Morris 1994; Cao and Geballe 1995; Sachs and Geballe 2006). The expressions and functions of the short uORFs upstream of the UL141 gene should be investigated in different HCMV strains in future.

A group of transcripts of 1000–1300 nt were detected by Northern blot. According to the results of 5′RACE and cDNA library screening, it could be presumed that these transcripts should not initiate at some specific sites but within a range. Similar to that of the UL145 transcripts (Wang *et al.* 2011), this initiation pattern could be inherent in HCMV. A cluster of nested organized ORFs could be predicted with 3′ co-terminal and in the same frame as that of the UL141 ORF. The UL141 protein had been detected with its monoclonal antibodies. Although not depicted by the author, major and minor products with proximity molecular masses were visual in the immunoblot results of a previous study (Tomasec *et al.* 2005). In addition, the mature UL141 proteins could be modified. So, it is necessary to study the expression of the assembly protein by detecting its precursor form. The similar organization has been found in the HCMV UL80 gene region (Welch *et al.* 1991). These ‘nested’ ORFs possess different initiation sites but the same termination site, which are usually in frame with a common sequence. The co-linear mRNAs are translated to overlapping polypeptides with identical carboxyl domains. Synthesis of related and overlapping proteins is one strategy used by the virus to generate additional proteins without an increase in genome size. The genetic organization of the virus may provide a transcriptional means for encoding and regulating a family of proteins able to interact with each other through their identical carboxyl domains and with other molecules through unique domains at their amino ends. The expressions of the ‘nested’ ORFs predicted in the present study remain to be confirmed.

Complicated transcript structures were identified in the UL140 and UL141 gene region in three HCMV clinical strains. The transcription kinetics and transcript structures were accordant among the three strains. Although multiple nucleotide mutations were observed in the UL139 sequence of X strain, these mutations should not affect the transcription regulation of the UL140 and UL141 transcripts. In this transcription family, multiple transcripts could encode the same or similar products. Revealing the transcriptional and posttranscriptional regulation mechanism of these transcripts is one key point to understand the signification of this genetic organization.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (No.30672248, 30801254, 30770109, 30700916, 30901625, 81171580; 81171581).

References

Cao J and Geballe AP 1995 Translational inhibition by a human cytomegalovirus upstream open reading frame despite inefficient utilization of its AUG codon. *J. Virol.* **69** 1030–1036

Cha TA, Tom E, Kemble GW, Duke GM, Mocarski ES and Spaete RR 1996 Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. *J. Virol.* **70** 78–83

Chang CP, Malone CL and Stinski MF 1989a A human cytomegalovirus early gene has three inducible promoters that are regulated differentially at various times after infection. *J. Virol.* **63** 281–290

Chang CP, Vesole DH, Nelson J, Oldstone MB and Stinski MF 1989b Identification and expression of a human cytomegalovirus early glycoprotein. *J. Virol.* **63** 3330–3337

Chee MS, Bankier AT, Beck S, Bohni R, Brown CM, Cerny R, Horsnell T, Hutchison CA, Kouzarides T and Martignetti JA 1990 Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. *Curr. Topic. Microbiol. Immunol.* **154** 125–169

Degnin CR, Schleiss MR, Cao J and Geballe AP 1993 Translational inhibition mediated by a short upstream open reading frame in the human cytomegalovirus gpUL4 (gp48) transcript. *J. Virol.* **67** 5514–5521

Geballe AP and Morris DR 1994 Initiation codons within 5'-leaders of mRNAs as regulators of translation. *Trends Biochem. Sci*. **19** 159–164

Goodrum F, Reeves M, Sinclair J, High K and Shenk T 2007 Human cytomegalovirus sequences expressed in latently infected individuals promote a latent infection in vitro. *Blood* **110** 937–945

Grainger L, Cicchini L, Rak M, Petrucelli A, Fitzgerald KD, Semler BL and Goodrum F 2010 Stress-inducible alternative translation initiation of human cytomegalovirus latency protein pUL138. *J. Virol.* **84** 9472–9486

He R, Ma Y, Qi Y, Jiang S, Wang N, Li M, Ji Y, Sun Z and Ruan Q 2012 Characterization of human cytomegalovirus UL146 transcripts. *Virus Res*. **163** 223–228

Lurain NS, Fox AM, Lichy HM, Bhorade SM, Ware CF, Huang DD, Kwan SP, Garrity ER and Chou S 2006 Analysis of the human cytomegalovirus genomic region from UL146 through UL147A reveals sequence hypervariability, genotypic stability, and overlapping transcripts. *Virol. J*. **3** 4

Ma YP, Ruan Q, Ji YH, Wang N, Li ML, Qi Y, He R, Sun ZR and Ren GW 2011 Novel transcripts of human cytomegalovirus clinical strain found by cDNA library screening. *Genet. Mol. Res.* **10** 566–575

Murphy E, Rigoutsos I, Shibuya T and Shenk TE 2003 Reevaluation of human cytomegalovirus coding potential. *Proc. Natl. Acad. Sci. USA* **100** 13585–13590

Penfold ME, Dairaghi DJ, Duke GM, Saederup N, Mocarski ES, Kemble GW and Schall TJ 1999 Cytomegalovirus encodes a potent alpha chemokine. *Proc. Natl. Acad. Sci. USA* **96** 9839–9844

Petrucelli A, Rak M, Grainger L and Goodrum F 2009 Characterization of a novel Golgi apparatus-localized latency determinant encoded by human cytomegalovirus. *J. Virol.* **83** 5615–5629

Petrucelli A, Umashankar M, Zagallo P, Rak M and Goodrum F 2012 Interactions between proteins encoded within the human cytomegalovirus UL133-UL138 Locus. *J. Virol.* **86** 8653–8662

Poole E, Atkins E, Nakayama T, Yoshie O, Groves I, Alcami A and Sinclair J 2008 NF-kappaB-mediated activation of the chemokine CCL22 by the product of the human cytomegalovirus gene UL144 escapes regulation by viral IE86. *J. Virol.* **82** 4250–4256

Prod'homme V, Sugrue DM, Stanton RJ, Nomoto A, Davies J, Rickards CR, Cochrane D, Moore M, Wilkinson GW and Tomasec P 2010 Human cytomegalovirus UL141 promotes efficient downregulation of the natural killer cell activating ligand CD112. *J. General Virol.* **91** 2034–2039

Qi Y, Ma Y, He R, Wang N, Ruan Q, Ji Y, Li M, Sun Z and Ren G 2011 Characterization of 3' termini of human cytomegalovirus UL138-UL145 transcripts in a clinical strain. *Microbiol. Immunol*. **55** 95–99

Sachs MS and Geballe AP 2006 Downstream control of upstream open reading frames. *Genes Dev*. **20** 915–921

Saederup N and Mocarski ES Jr 2002 Fatal attraction: cytomegalovirus-encoded chemokine homologs. *Curr. Topic. Microbiol. Immunol*. **269** 235–256

Sun Z, Ren G, Ma Y, Wang N, Ji Y, Qi Y, Li M, He R and Ruan Q 2010 Transcription pattern of UL131A-128 mRNA in clinical strains of human cytomegalovirus. *J. Biosci*. **35** 365–370

Tomasec P, Wang EC, Davison AJ, Vojtesek B, Armstrong M, Griffin C, McSharry BP, Morris RJ, *et al.* 2005 Downregulation of natural killer cell-activating ligand CD155 by human cytomegalovirus UL141. *Nat. Immunol*. **6** 181–188

Wang N, Ma Y, Sun Z, Qi Y, Ji Y, He R, Li M and Ruan Q 2011 Transcriptional features and transcript structure of UL145 in different strains of human cytomegalovirus. *J. Med. Virol.* **83** 2151–2156

Welch AR, McNally LM and Gibson W 1991 Cytomegalovirus assembly protein nested gene family: four 3'-coterminal transcripts encode four in-frame, overlapping proteins. *J. Virol.* **65** 4091–4100

MS received 12 October 2012; accepted 27 November 2012

Corresponding editor: SHAHID JAMEEL