Journal of General Virology (2003), **84**, 2511–2516 DOI 10.1099/vir.0.19298-0

Short Communication

The most abundantly transcribed human cytomegalovirus gene (β2.7) is non-essential for growth *in vitro*

Brian P. McSharry,[1] Peter Tomasec,[1] M. Lynne Neale[2] and Gavin W. G. Wilkinson[1]

Section of Infection and Immunity[1] and Department of Medical Microbiology[2], University of Wales College of Medicine, Heath Park, Cardiff CF14 4XY, UK

Correspondence
Gavin Wilkinson
WilkinsonGW1@cardiff.ac.uk

Received 16 April 2003
Accepted 26 June 2003

The most abundantly transcribed HCMV gene (β2.7) encodes a 2·7 kb polyadenylated RNA. Although the laboratory-adapted HCMV strains AD169 and Towne possess two copies of the β2.7 gene within an expanded *b* sequence element, the low passage strain Toledo and all clinical isolates analysed contain only a single copy located within the U_L region. A β2.7 deletion mutant constructed based on a strain Toledo background was shown to replicate with kinetics comparable to those of the parental virus; the β2.7 gene is therefore not essential for virus replication *in vitro*. Sequencing the β2.7 gene from HCMV clinical isolates and the Toledo strain reveals that although the overall gene sequence is highly conserved (>99 %), the RL4 frame originally assigned in strain AD169 was disrupted in each of these viruses. Consequently, the β2.7 transcript does not encode any obvious translation product and thus may not function as an mRNA.

The most abundantly transcribed early gene from the HCMV genome in permissive cells is a 2·7 kb unspliced polyadenylated RNA (β2.7) (McDonough & Spector, 1983; McDonough *et al.*, 1985). The β2.7 gene exhibits early gene kinetics showing maximal amplification between 8 and 14 h post-infection (p.i.) and remains transcriptionally active through the virus replication cycle, accumulating throughout infection (McDonough *et al.*, 1985). The β2.7 gene resides in the inverted repeats flanking the unique long (U_L) segment of the genome in both strains AD169 and Towne and therefore these strains have two copies of the gene. A Towne strain deletion mutant lacking one copy of the β2.7 gene has been constructed and exhibits normal growth kinetics (Spaete & Mocarski, 1987), while the Davis strain of HCMV naturally contains only a single copy of this gene present in the U_L segment of the genome (Demarchi, 1981). The longest open reading frame in the gene, conserved in both strains AD169 and Towne, is 513 bp and was designated IRL4/TRL4 (Chee *et al.*, 1990). This ORF was predicted to encode a 170 amino acid protein (Greenaway & Wilkinson, 1987). RL4 begins 283 bp downstream of the gene's transcriptional start site in strain AD169 and is preceded by two short open reading frames of 24 and 108 bp respectively.

During productive HCMV infection β2.7 RNA can be detected in polysome preparations (Wathen & Stinski, 1982; Wu *et al.*, 1992). Forced expression of an RL4 ORF epitope-tagged fusion protein derived from either strain Towne or AD169 generated a 24 kDa product that trafficked to the nucleoli (Bergamini *et al.*, 1998). However, despite its abundance, no β2.7-associated translation product has yet been detected during HCMV infection (Greenaway & Wilkinson, 1987; Lord *et al.*, 1989).

Concern exists regarding the genomic integrity of the commonly used laboratory strains AD169 and Towne (Akter *et al.*, 2003; Dargan *et al.*, 1997; Davison *et al.*, 2003; Goldmacher *et al.*, 1999; Mocarski *et al.*, 1997), in particular at the ends of the U_L/*b* sequence boundary where the β2.7 gene resides (Cha *et al.*, 1996). We were particularly concerned that a number of genes within these regions exhibit a high degree of sequence variation between clinical isolates (Lurain *et al.*, 1999; A. Davison, personal communication). We wished to evaluate the integrity of the β2.7 gene within the laboratory strains and, more generally, sequence variation in the β2.7 gene. To this end, the complete gene from the Toledo strain and three HCMV clinical isolates was sequenced. Furthermore, in order to investigate the function of the β2.7 gene, an HCMV deletion mutant was generated in the Toledo strain.

The complete β2.7 gene was amplified by PCR from genomic DNA of the Toledo strain and three recent HCMV clinical isolates, 3157, 6397 and Merlin. Clinical isolates were cultured from urine samples of new-born infants and passaged on human foetal foreskin fibroblasts (HFFFs) fewer than three times prior to analysis. The gene was amplified as two separate overlapping fragments of 1215

The GenBank accession numbers of the HCMV sequences reported in this paper are AY325309–AY325312.

and 1697 bp (covering 2421 nucleotides; 4575–2574 of the strain AD169 sequence, GenBank accession no. X17403) using an enzyme with proof reading activity (Dynazyme EXT; Finnzymes). PCR-generated products were cloned into pCR2.1-TOPO (Invitrogen) and sequenced using the BigDye Terminator Cycle Sequencing Kit (PE Applied Biosystems). Overall, the gene exhibited greater than 99 % conservation throughout its entire length compared to the strain AD169 sequence. However, the Toledo strain and each of the three clinical isolates exhibited variations in the sequence corresponding to the RL4 ORF leading to disruption and truncation of the predicted protein sequence (Fig. 1A). Each of the sequence variations in RL4 was confirmed by sequencing directly three independently amplified PCR products from the same DNA samples.

The first major change in the RL4 ORF sequence is found in the 3157 sequence with a 4 bp insertion just 30 bp into the putative coding sequence. A single base-pair deletion in the Merlin sequence compared to the published strain AD169 sequence at position 173 leads to a frameshift and premature truncation of the postulated ORF. Both the Toledo strain and clinical isolate 6397 have a single base-pair insertion at position 184 in the RL4 sequence that leads to a frameshift and truncation of the ORF. This additional T residue is also found in the 3157 sequence that exhibited a previous frameshift. Therefore, each of the clinical isolates exhibits a distinct change in the sequence that leads to frameshifts and truncation of the RL4 ORF.

The putative protein products of the ORFs with the same start codon as RL4 in strain AD169 are aligned in Fig. 1(B). The protein products of the Toledo strain, 6397 and Merlin have homology with the amino terminus of the strain AD169 putative protein product; however, each of these viruses exhibits premature truncation of the postulated

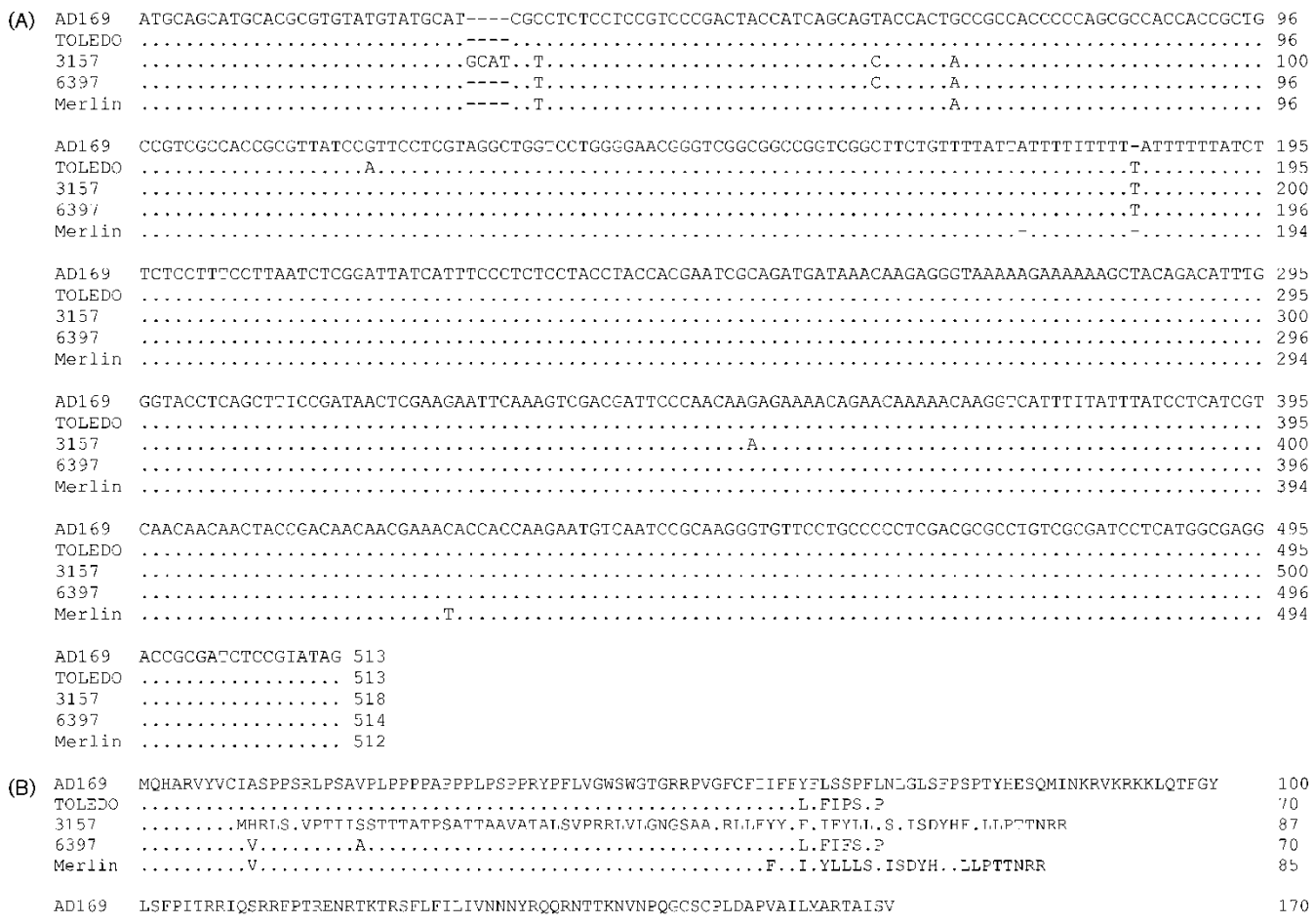

Fig. 1. Alignment of the DNA and protein sequence corresponding to the RL4 ORF. (A) Alignment of the RL4 DNA sequence with numbering of the nucleotide positions illustrated from the strain AD169, strain Toledo and three HCMV clinical isolates, 3157, 6397 and Merlin. Changes from the strain AD169 sequence are detailed. (B) Alignment of the putative protein products in the RL4 region. The full strain AD169 sequence is illustrated and differences from the strain AD169 sequence in strain Toledo and the clinical isolates 3157, 6397 and Merlin are detailed. The amino acid numbering of the putative protein products is illustrated.

protein. The RL4 sequence encoded by isolate 3157 diverges from strain AD169 after only nine amino acids resulting in a major disruption and truncation of the postulated open reading frame. No major conserved ORFs were generated by changes detected in the gene sequence. The sequences corresponding to the two short upstream ORFs found in strain AD169 were conserved in strain Toledo and the clinical isolates (not shown).

To determine the copy number of the β2.7 gene in the Toledo strain and three clinical isolates, DNA was prepared from purified virus stocks. Virus harvested in supernatants from infected-cell monolayers was centrifuged at 12 000 ***g***, resuspended in TE buffer (10 mM Tris, 1 mM EDTA) pH 8·0, 1 % NP40 and digested overnight at 37 °C with proteinase K (200 μg ml^{-1}). DNA was then extracted gently with phenol/chloroform/isoamyl alcohol (25 : 24 : 1) before being precipitated and resuspended in TE buffer. Viral DNA was digested with *Hin*dIII and *Kpn*I to identify the copy number of the gene. Digested DNA was run on 0·5–0·8 % agarose/1 × TBE (Tris/borate/EDTA) gels. The DNA was then transferred to a nitrocellulose membrane by capillary action in 20 × SSC overnight and fixed by baking at 80 °C for 2 h. The membrane was then hybridized with a DNA probe labelled with [α-^{32}P]CTP using HighPrime (Roche) and washed as described previously (Phillips *et al.*, 1998) before autoradiography.

*Kpn*I cuts 584 bp downstream of the transcriptional start site of the β2.7 gene and in the U_L segment of the genome in strain AD169 (Chee *et al.*, 1990). When *Kpn*I-digested strain AD169 DNA is hybridized against a PCR-generated probe specific for the first 527 bp of the gene, two DNA fragments are recognized owing to the presence of the gene in both copies of the repeats flanking U_L (Fig. 2A). However, when *Kpn*I-digested DNA of each of the other isolates is similarly probed only a single band is recognized, indicating that the β2.7 gene is only single copy in

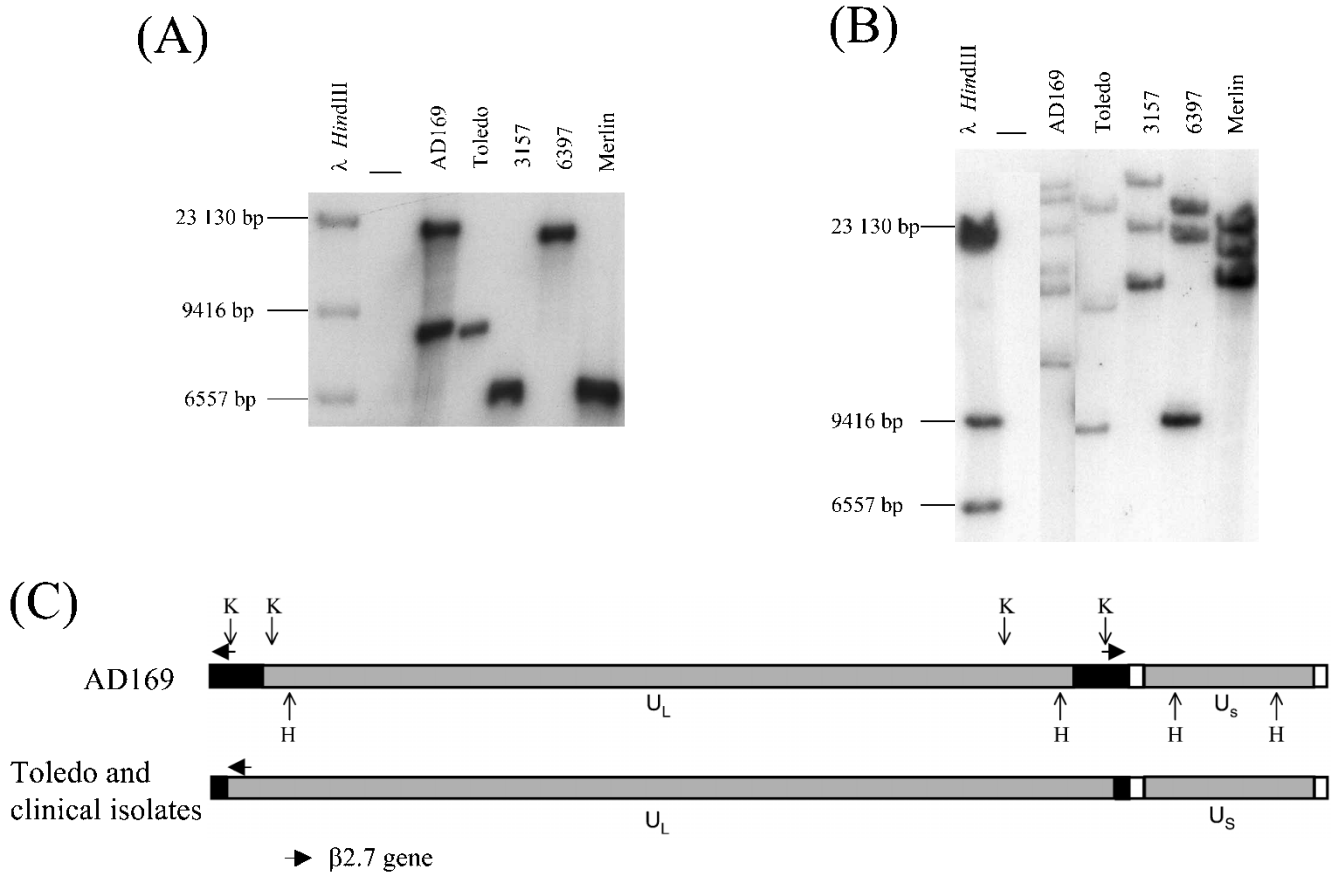

Fig. 2. Copy number of the β2.7 gene. *Kpn*I- (A) or *Hin*dIII- (B) digested HCMV DNA was run on a 0·8 or 0·5 % agarose gel respectively, blotted onto nitrocellulose before being hybridized against a 527 bp β2.7-specific radioactive probe. The molecular mass marker, *Hin*dIII-digested λ DNA, is shown. (C) Schematic representation of the location of the β2.7 gene in strain AD169 genome in comparison to the Toledo strain and clinical isolates. In strain AD169 the two copies of the β2.7 gene (■) reside in the inverted repeats while in the other viruses the gene is single copy in the U_L segment of the genome. The *Hin*dIII (H) and *Kpn*I (K) sites within the strain AD169 sequence are illustrated.

each of these viruses. This was confirmed by digesting with *Hin*dIII and probing with a β2.7-specific sequence (Fig. 2B). Digested DNA of strain AD169 produces six bands because *Hin*dIII cuts outside of the repeats and the genome inverts to yield four individual isomers (Fig. 2C). When DNA from the other viruses is similarly hybridized the β2.7-specific probe anneals to only three fragments; two of these fragments are present in 0·25 M internal repeats and the other fragment detected denotes a 0·5 M terminal fragment. This indicates that the β2.7 gene of HCMV clinical isolates lies in the U_L segment of the genome close to flanking repeats.

The DNA sequence indicated that although the RL4 ORF was disrupted the β2.7 gene was highly conserved and thus likely to be functional in all the characterized strains and isolates. Since the Toledo strain contains only a single copy of the gene, a deletion mutant based on this strain was generated. Regions both upstream (nucleotides 5455–4576) and downstream (nucleotides 2155–1200) of the β2.7 gene from the AD169 strain were amplified by PCR and used to flank the reporter gene EGFP to target the complete deletion of the β2.7 gene, generating the targeting plasmid pAL325. This strategy placed the reporter gene under the transcriptional control of the β2.7 promoter (Spaete & Mocarski, 1987). pAL325 was then linearized and transfected into HFFFs with purified strain Toledo genomic DNA. Recombinant viruses expressing EGFP were then isolated as previously described (Wang *et al.*, 2002).

A virus mutant containing the predicted recombination was confirmed by Southern analysis (Fig. 3). When *Eco*RV-digested DNA of Toledo and TOL$\Delta\beta$2.7 was probed with a β2.7 promoter fragment (nucleotides 5455–4576) a size difference of ~1700 bp is detected, which is consistent with targeted replacement of 2421 bp of the β2.7 gene with the 720 bp of EGFP (Fig. 3A). When similarly digested DNA of both viruses is probed with a 2466 bp PCR-generated fragment covering the complete β2.7 gene a band is only detected from Toledo DNA, confirming the

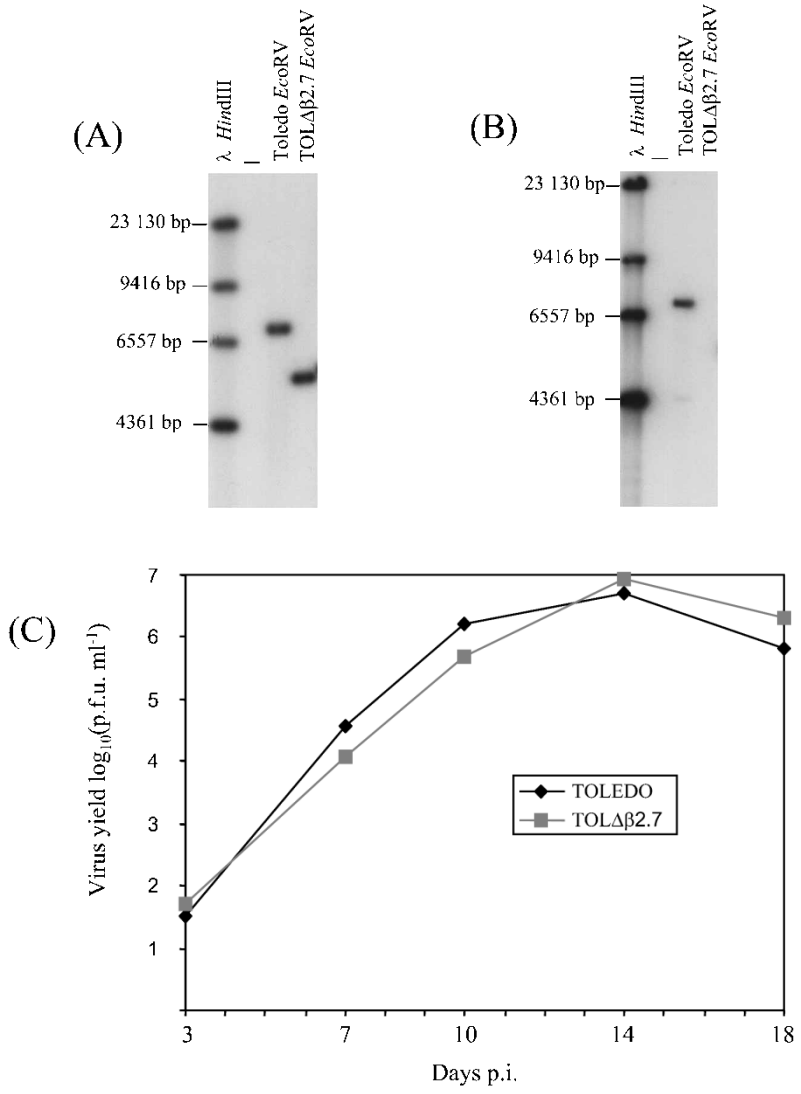

Fig. 3. Genomic structure and growth kinetics of TOL$\Delta\beta$2.7. *Eco*RV-digested DNA of Toledo and TOL$\Delta\beta$2.7 was probed with β2.7 early promoter (A) or the complete β2.7 gene (B). The λ *Hin*dIII size markers are indicated. (C) HFFFs were infected at an m.o.i. of 0·01 with either strain Toledo or TOL$\Delta\beta$2.7. Medium was harvested from the infected monolayers at the indicated times p.i. and titrated by plaque assay.

complete deletion of β2.7 from the mutant virus (Fig. 3B). Additional digests confirmed the predicted genomic structure of TOL$\Delta\beta$2.7 (not shown).

To investigate the growth kinetics of the virus mutant, HFFFs were infected in parallel with either strain Toledo or TOL$\Delta\beta$2.7 at an m.o.i. of 0·01. Virus-containing supernatant was then harvested at 3, 7, 10, 14 and 18 days p.i. and the virus titrated by plaque assay on HFFFs. The growth kinetics of TOL$\Delta\beta$2.7 were unaltered in comparison to its parental strain (Fig. 3C), indicating that the β2.7 gene is non-essential for virus replication *in vitro*.

The complete β2.7 gene sequence is clearly highly conserved not only in HCMV clinical isolates but also in the common laboratory strains, thus demonstrating that the gene sequence was not substantially modified during prolonged *in vitro* passage. In contrast, the major predicted ORF TRL4/IRL4 is not conserved, which would be consistent with the β2.7 gene not encoding a functional translation product. A number of viruses encode RNAs that are apparently not translated but function at the transcript level including the human herpesvirus-8 abundant 1·1 kb nuclear untranslated RNA (*nut-1* RNA) (Zhong & Ganem, 1997), the EBERs of Epstein–Barr virus, the LAT transcript of herpes simplex virus-1 and VA_1 and VA_{II} RNAs of adenovirus (Sharp *et al.*, 1993; Perng *et al.*, 2000). An interaction between the β2.7 transcript and ribonucleoproteins La and Ro has been reported (Lord *et al.*, 1989). However no major alteration could be detected in the subcellular localization or relative abundance of these proteins following infection with Toledo or TOL$\Delta\beta$2.7 (not shown).

The 2.7 kb transcript is found predominantly in the cytoplasm, does associate with polysomes and reporter genes cloned within the gene are expressed at an extremely high level (Gawn & Greaves, 2002; Geballe *et al.*, 1986; Spaete & Mocarski, 1985) (this study, not shown). While it seems probable that the β2.7 transcript does not encode a functional translation product, one cannot yet preclude a potential role for translation products of the short conserved upstream 24 and 108 bp ORFs. However, a comparative analysis of Toledo- or TOL$\Delta\beta$2.7-infected cell extracts by MALDI–TOF mass spectroscopy was unable to detect peptides corresponding to either of the short predicted translation products (data not shown).

Each of the clinical isolates analysed contains only a single copy of the β2.7 gene with the size of the *b* sequence much reduced compared to the more commonly used laboratory strains and similar to those previously noted in strains Davis (DeMarchi, 1984) and Toledo (Prichard *et al.*, 2001). Our study is consistent with the contention that β2.7 is a haploid gene that has become duplicated due to expansion of the *b* sequence in the heavily passaged strains AD169 and Towne. Although the β2.7 gene is clearly dispensable for efficient HCMV replication in fibroblasts *in vitro*, the observation that a duplication of this abundantly transcribed gene appears to have occurred independently in two heavily passaged strains is suggestive that the gene's function may not be exclusively restricted to a role *in vivo*.

While there is no obvious β2.7 gene homologue in any other characterized β-herpesviruses, it is interesting to note that the equivalent region of chimpanzee CMV (GenBank accession no. AF480884) does contain extended regions with significant sequence homology. However, the homology does not extend into the β2.7 promoter element and it has yet to be established if these sequences are transcribed in the chimpanzee virus. Finally, the high level of sequence conservation of the β2.7 gene combined with its abundant transcription makes it an extremely attractive target for nucleic acid-based detection methods of HCMV in clinical samples (Aono *et al.*, 1998).

ACKNOWLEDGEMENTS

This work was supported by the Wellcome Trust and a Medical Research Council Co-operative Grant. Cardiff Public Health Laboratory provided clinical samples. We are extremely grateful to A. Davison, G. Duke and E. S. Mocarski for helpful discussions and Dr Mike Morton for assistance with mass spectroscopy.

REFERENCES

Akter, P., Cunningham, C., McSharry, B. P. & 8 other authors (2003). Two novel spliced genes in human cytomegalovirus. *J Gen Virol* **84**, 1117–1122.

Aono, T., Kondo, K., Miyoshi, H., Tanaka-Taya, K., Kondo, M., Osugi, Y., Hara, J., Okada, S. & Yamanishi, K. (1998). Monitoring of human cytomegalovirus infections in pediatric bone marrow transplant recipients by nucleic acid sequence-based amplification. *J Infect Dis* **178**, 1244–1249.

Bergamini, G., Reschke, M., Battista, M. C., Boccuni, M. C., Campanini, F., Ripalti, A. & Landini, M. P. (1998). The major open reading frame of the β2.7 transcript of human cytomegalovirus: in vitro expression of a protein posttranscriptionally regulated by the 5′ region. *J Virol* **72**, 8425–8429.

Cha, T. A., Tom, E., Kemble, G. W., Duke, G. M., Mocarski, E. S. & Spaete, R. R. (1996). Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. *J Virol* **70**, 78–83.

Chee, M. S., Bankier, A. T., Beck, S. & other authors (1990). Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. *Curr Top Microbiol Immunol* **154**, 125–169.

Dargan, D. J., Jamieson, F. E., MacLean, J., Dolan, A., Addison, C. & McGeoch, D. J. (1997). The published DNA sequence of human cytomegalovirus strain AD169 lacks 929 base pairs affecting genes UL42 and UL43. *J Virol* **71**, 9833–9836.

Davison, A. J., Dolan, A., Akter, P., Addison, C., Dargan, D. J., Alcendor, D. J., McGeoch, D. J. & Hayward, G. S. (2003). The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. *J Gen Virol* **84**, 17–28.

Demarchi, J. M. (1981). Human cytomegalovirus DNA: restriction enzyme cleavage maps and map locations for immediate-early, early and late RNAs. *Virology* **114**, 23–38.

DeMarchi, J. M. (1984). The physical and transcriptional organization of the human cytomegalovirus genome. *Birth Defects Orig Artic Ser* **20**, 35–47.

Gawn, J. M. & Greaves, R. F. (2002). Absence of IE1 p72 protein function during low-multiplicity infection by human cytomegalovirus results in a broad block to viral delayed-early gene expression. *J Virol* **76**, 4441–4455.

Geballe, A. P., Spaete, R. R. & Mocarski, E. S. (1986). A cis-acting element within the 5′ leader of a cytomegalovirus beta transcript determines kinetic class. *Cell* **46**, 865–872.

Goldmacher, V. S., Bartle, L. M., Skaletskaya, A. & 10 other authors (1999). A cytomegalovirus-encoded mitochondria-localized inhibitor of apoptosis structurally unrelated to Bcl-2. *Proc Natl Acad Sci U S A* **96**, 12536–12541.

Greenaway, P. J. & Wilkinson, G. W. (1987). Nucleotide sequence of the most abundantly transcribed early gene of human cytomegalovirus strain AD169. *Virus Res* **7**, 17–31.

Lord, P. C., Rothschild, C. B., DeRose, R. T. & Kilpatrick, B. A. (1989). Human cytomegalovirus RNAs immunoprecipitated by multiple systemic lupus erythematosus antisera. *J Gen Virol* **70**, 2383–2396.

Lurain, N. S., Kapell, K. S., Huang, D. D., Short, J. A., Paintsil, J., Winkfield, E., Benedict, C. A., Ware, C. F. & Bremer, J. W. (1999). Human cytomegalovirus UL144 open reading frame: sequence hypervariability in low-passage clinical isolates. *J Virol* **73**, 10040–10050.

McDonough, S. H. & Spector, D. H. (1983). Transcription in human fibroblasts permissively infected by human cytomegalovirus strain AD169. *Virology* **125**, 31–46.

McDonough, S. H., Staprans, S. I. & Spector, D. H. (1985). Analysis of the major transcripts encoded by the long repeat of human cytomegalovirus strain AD169. *J Virol* **53**, 711–718.

Mocarski, E. S., Prichard, M. N., Tan, C. S. & Brown, J. M. (1997). Reassessing the organization of the UL42–UL43 region of the human cytomegalovirus strain AD169 genome. *Virology* **239**, 169–75.

Perng, G. C., Jones, C., Ciacci-Zanella, J. & 8 other authors (2000). Virus-induced neuronal apoptosis blocked by the herpes simplex virus latency-associated transcript. *Science* **287**, 1500–1503.

Phillips, A. J., Tomasec, P., Wang, E. C., Wilkinson, G. W. & Borysiewicz, L. K. (1998). Human cytomegalovirus infection downregulates expression of the cellular aminopeptidases CD10 and CD13. *Virology* **250**, 350–358.

Prichard, M. N., Penfold, M. E., Duke, G. M., Spaete, R. R. & Kemble, G. W. (2001). A review of genetic differences between limited and extensively passaged human cytomegalovirus strains. *Rev Med Virol* **11**, 191–200.

Sharp, T. V., Schwemmle, M., Jeffrey, I., Laing, K., Mellor, H., Proud, C. G., Hilse, K. & Clemens, M. J. (1993). Comparative analysis of the regulation of the interferon-inducible protein kinase PKR by Epstein–Barr virus RNAs EBER-1 and EBER-2 and adenovirus VAI RNA. *Nucleic Acids Res* **21**, 4483–4490.

Spaete, R. R. & Mocarski, E. S. (1985). Regulation of cytomegalovirus gene expression: alpha and beta promoters are trans activated by viral functions in permissive human fibroblasts. *J Virol* **56**, 135–143.

Spaete, R. R. & Mocarski, E. S. (1987). Insertion and deletion mutagenesis of the human cytomegalovirus genome. *Proc Natl Acad Sci U S A* **84**, 7213–7217.

Wang, E. C., McSharry, B., Retiere, C., Tomasec, P., Williams, S., Borysiewicz, L. K., Braud, V. M. & Wilkinson, G. W. (2002). UL40-mediated NK evasion during productive infection with human cytomegalovirus. *Proc Natl Acad Sci U S A* **99**, 7570–7575.

Wathen, M. W. & Stinski, M. F. (1982). Temporal patterns of human cytomegalovirus transcription: mapping the viral RNAs synthesized at immediate early, early and late times after infection. *J Virol* **41**, 462–477.

Wu, T. C., Lee, W. A., Pizzorno, M. C., Au, W. C., Chan, Y. J., Hruban, R. H., Hutchins, G. M. & Hayward, G. S. (1992). Localization of the human cytomegalovirus 2.7-kb major early beta-gene transcripts by RNA in situ hybridization in permissive and nonpermissive infections. *Am J Pathol* **141**, 1247–1154.

Zhong, W. & Ganem, D. (1997). Characterization of ribonucleoprotein complexes containing an abundant polyadenylated nuclear RNA encoded by Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8). *J Virol* **71**, 1207–1212.