JOURNAL OF VIROLOGY, Sept. 2004, p. 9579–9583 Vol. 78, No. 17
0022-538X/04/$08.00+0 DOI: 10.1128/JVI.78.17.9579–9583.2004

Characterization of the Human Cytomegalovirus UL34 Gene

B. J. Biegalke,* E. Lester, A. Branda, and R. Rana

Department of Biomedical Sciences, College of Osteopathic Medicine, Ohio University, Athens, Ohio

Received 9 March 2004/Accepted 16 April 2004

UL34 encodes the transcriptional repressor of the human cytomegalovirus immune evasion gene, US3, and is essential for viral replication in tissue culture. Two different monocistronic transcripts originate from UL34 at early and late times postinfection and encode two predominant proteins and a third, minor protein. The UL34 proteins are differentially expressed throughout the viral replication cycle, with both proteins localizing to the nucleus and repressing expression of the US3 gene.

Human cytomegalovirus (HCMV) has a large genome containing ~200 predicted open reading frames (4, 6, 13), with replication of the virus requiring the coordinated expression of the genome during the 5-day replication cycle (see reference 12 for a review). A relatively small number of open reading frames have been identified as containing viral genes essential for replication in cell culture (7, 13). Interestingly, although the only known function for UL34 is transcriptional repression of the nonessential US3 immune evasion gene (9), the global analyses of the HCMV genome identified UL34 as an essential gene (7, 15).

Expression of the US3 gene results in down-regulation of major histocompatibility class I complexes by retaining the protein complexes in the endoplasmic reticulum (1, 8, 10). The US3 gene is transcribed maximally at 3 h postinfection (hpi), with the level of expression declining to minimally detectable levels by 5 hpi (10). The down-regulation of US3 expression is mediated by the UL34 gene, with the UL34 protein binding to the transcriptional repressive element (*tre*) of the US3 gene (3, 9).

Transcriptional analysis of the UL34 gene. The transcription pattern of the UL34 gene was determined by performing Northern blot analyses of RNA harvested from mock-infected cells and HCMV-infected cells at different times postinfection, hybridizing blots to a UL34-specific probe. Experiments examining UL34 RNA levels throughout a single replication cycle demonstrated that UL34 is abundantly transcribed at 4 hpi with a predominant transcript of 1.6 kb and a larger transcript of 3.2 kb (Fig. 1A). The 1.6-kb transcript corresponded in size to the predicted UL34 transcript, while the 3.2-kb transcript corresponded in size to the predicted UL33-UL34 read-through transcript (14). A low level of UL34 transcripts was detected in cells treated with the protein synthesis inhibitor cycloheximide, suggesting that de novo synthesis of viral proteins is not absolutely required for UL34 transcription. At later times in infection, the 1.6-kb RNA decreased in abundance, and a 1.4-kb transcript appeared by 24 hpi and became predominant by 48 hpi. The UL33-UL34 transcript also became more abundant late in infection (Fig. 1A). Treatment with phosphonoformic acid inhibited transcription of UL34, suggesting that newly replicated DNA genomes are required for UL34 transcription at late times of infection. A comparison of infected cell RNAs harvested 4 and 96 hpi is depicted in Fig. 1C, clearly illustrating the change in transcript sizes from a 1.6-kb to a 1.4-kb UL34 transcript and the presence of the 3.2-kb UL33-UL34 read-through transcript. Further analyses of UL34 transcription were performed at very early times postinfection (0.5 to 4 hpi), demonstrating that the 1.6-kb UL34 transcript is detectable by 2 hpi, becoming maximal by 3 hpi, with a slight diminution by 4 hpi (Fig. 1B).

The pattern of UL34 expression correlated directly with the decrease in US3 transcription (10). Somewhat surprisingly, two smaller transcripts of 1.6 and 1.4 kb originated from the UL34 open reading frame at different times postinfection, suggesting that UL34 may encode more than one protein.

RNase protection assays were used to characterize the different UL34 transcripts. The results of the RNase protection experiments (data not shown) demonstrated that the 1.6- and 1.4-kb transcripts differed at their 5′ ends. To precisely identify the 5′ ends of the UL34 transcripts, rapid amplification of cDNA ends (RACE) reactions were performed using the FirstChoice RLM-RACE kit (Ambion, Austin, Tex.), oligonucleotides 308 (outer primer, 5′ GTCTCGGGTGGTGATGATG AAGT 3′) and 309 (inner primer, 5′ GGTAATAGTATAGG TCCGTGCGA 3′), and RNA extracted from infected cells either 4 or 96 hpi. The resulting amplimers were cloned and grouped according to size. Multiple examples of each group of amplimers were sequenced. At early times postinfection, two predominant classes of amplimers were seen: the majority (75%) of the amplified transcripts originated at nucleotide 44616, with a smaller but significant number (25%) originating at nucleotide 44682. These data demonstrated that transcription from the UL34 gene initiates at two sites early in infection and support the proposed promoter role of the TATA box at positions 44584 to 44588 (Fig. 2A and C).

Two predominant classes of amplimers were also detected at late times postinfection. One class of amplimers contained UL33 as well as UL34 sequences and corresponded to the UL33-UL34 read-through transcripts. The second class of late amplimers contained UL34 sequences, beginning at nucleotide 44794 (Fig. 2A). Thus, the late UL34 transcript initiates within the predicted open reading frame of the UL34 gene, beginning immediately after the translation initiation codon for the early

* Corresponding author. Mailing address: Department of Biomedical Sciences, Ohio University College of Osteopathic Medicine, 228 Irvine Hall, Athens, OH 45701. Phone: (740) 593-2377. Fax: (740) 597-2778. E-mail: biegalke@ohiou.edu.

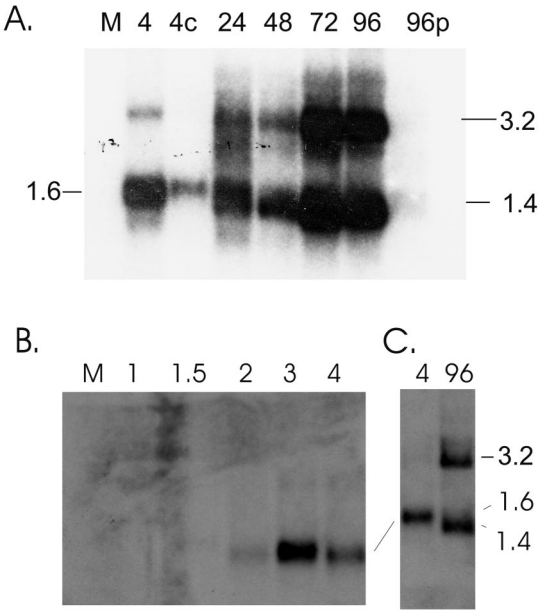

FIG. 1. Northern blot analysis of UL34 transcription. Total cellular RNA was harvested from mock-infected cells (M) or infected cells at the indicated times postinfection. Following electrophoresis of equal amounts of RNA and transfer, blots were hybridized to a ^{32}P-radiolabeled UL34 probe. (A) Expression of UL34 during a single-step replication cycle. 4c, RNA harvested at 4 hpi with infection and transcription occurring in the presence of cycloheximide; 96p, RNA harvested at 96 hpi in the presence of phosphonoformic acid. (B) Analysis of UL34 transcripts at early times of infection. (C) Comparison of RNA harvested 4 and 96 hpi. Equivalent amounts of RNA were electrophoresed in each lane, as demonstrated by ethidium bromide staining of rRNA (data not shown).

UL34 open reading frame. The difference in sizes between the early and late UL34 transcripts seen by Northern blot analyses correlated with the transcription initiation sites identified by the 5′ RACE analysis.

UL34 protein expression. The early 1.6-kb UL34 transcript contains an open reading frame beginning with the ATG at positions 44791 to 44793; this open reading frame corresponds to the open reading frame encoding the US3 repressor (9). The late monocistronic UL34 transcript contains an open reading frame that begins with the ATG at positions 44854 to 44856. The open reading frame found in the late UL34 transcript is in the same reading frame and is contained within the open reading frame of the early UL34 protein. The protein predicted to be encoded by the late UL34 transcripts would lack 21 amino-terminal amino acids compared to the open reading frame of the early transcript (Fig. 2).

Antisera to UL34 proteins were obtained by immunizing rabbits with purified, six-His-tagged UL34 proteins. UL34 protein expression was characterized using Western blot analyses and immunoprecipitation of ^{35}S or ^{32}P-labeled cells as described previously (11).

For immunoprecipitation experiments, mock-infected cells and HCMV-infected cells were labeled with [^{35}S]methionine or [^{32}P]phosphorous at various times postinfection, cell lysates were harvested, and UL34 proteins were immunoprecipitated. At 4 hpi, a 43-kDa protein was immunoprecipitated from HCMV-infected cells, along with a minor amount of a 45-kDa protein (Fig. 3A and B; antisera from two different rabbits were used for these experiments). As the infection progressed, another smaller protein of 40 kDa in mass was also precipitated. The smaller 40-kDa protein corresponded to the predicted size of the protein encoded by the late UL34 transcript.

The Western blot analysis of infected cell lysates revealed a pattern of UL34 proteins similar to that seen with immunoprecipitation: a predominant early UL34 protein of 43 kDa in mass and a predominant late protein of 40 kDa in mass (Fig. 3D). The early UL34 protein corresponded in size to the protein produced from in vitro transcription and translation of the early UL34 open reading frame (Fig. 3D, lane 7). In addition to the predominant early and late proteins detected of 43 and 40 kDa in mass, respectively, a minor amount of a slightly larger protein (45 kDa) was also reactive with the anti-UL34 antisera (Fig. 3A, B, and D).

In addition to analyzing infected cells for UL34 expression, we also examined virion preparations for reactivity with the anti-UL34 antisera. We were unable to detect UL34 in preparations of virions, suggesting that UL34 is not a component of mature virions (data not shown).

The proteins encoded by the UL34 gene contain numerous negatively charged amino acids and predicted nuclear localization signals. In addition, the UL34-encoded proteins are predicted to be phosphorylated on serine, threonine, and tyrosine residues. The potential phosphorylation of the UL34 proteins was examined using immunoprecipitation of ^{32}P-labeled proteins. No phosphorylated proteins corresponding to the 40-, 43-, or 45-kDa UL34 proteins were detected (Fig. 3C). Thus, UL34 proteins are not phosphorylated, in contrast to the predicted results.

Interestingly, in all of the immunoprecipitation experiments, using either antiserum to UL34 (generated in two separate rabbits to two different preparations of antigen), at late times postinfection (72 and 96 hpi), a coprecipitating protein of approximately 60 kDa was detected following labeling with either ^{35}S or ^{32}P (Fig. 3A to C). The anti-UL34 antisera did not recognize the 60-kDa protein in Western blots (Fig. 3D), suggesting that detection of the 60-kDa protein occurred through associations with pUL34 with resulting coprecipitation.

Localization of UL34 proteins. The early UL34 open reading frame, when expressed as a fusion protein with green fluorescent protein (GFP), localizes to the nucleus (9). The late pUL34 (pUL34^{40kDa}) contains the consensus nuclear localization signals found in pUL34^{43kDa}, suggesting that both early and late forms of the proteins would localize to the nucleus. Experiments were performed to determine the intracellular location of the late UL34 protein. The early and late UL34 proteins were expressed as GFP fusion proteins in transfected human diploid fibroblasts and visualized using a fluorescence microscope. The results of these experiments are depicted in Fig. 4. GFP alone had a widespread distribution in the expressing cells (Fig. 4B), while expression of either the early or the late UL34-GFP fusion construct resulted in nuclear fluorescence (Fig. 4D and F), confirming the predicted localization patterns of the UL34 proteins.

Functional analysis of the UL34 proteins. Experiments were conducted to determine if pUL34^{40kDa} functions as a sequence-specific DNA-binding protein, able to repress expres-

FIG. 2. UL34 transcription initiation sites. (A) The nucleotide sequence of the 3′ end of UL33 and the 5′ end of UL34 is depicted (5). The UL33 stop codon is indicated by an open rectangle; the predicted early TATA box of UL34 is indicated with a gray rectangle. The early transcription initiation sites are indicated by open arrowheads; the late transcription initiation start site is indicated by a filled arrowhead. The first ATG codons in the early and late transcripts are indicated in bold; the ACG postulated to initiate translation of the 45-kDa UL34 protein is indicated by underlined letters. (B) Diagram of the UL34 gene with the TATA box indicated by a T, the early transcription initiation start sites indicated by open arrowheads, and the late transcription start site indicated by a filled arrowhead. The proposed open reading frames are depicted, with the dotted rectangle corresponding to translation initiation at ACG, the open rectangle corresponding to initiation of translation at the first methionine on the early UL34 transcripts, and the gray rectangle corresponding to the late open reading frame. All of the predicted proteins are in frame with each other and share the amino acid sequences indicated by the gray rectangle. T, TATA box; A, polyadenylation signal. (C) Schematic diagram of the early UL34 transcripts. (D) Schematic diagram of the late UL34 transcripts.

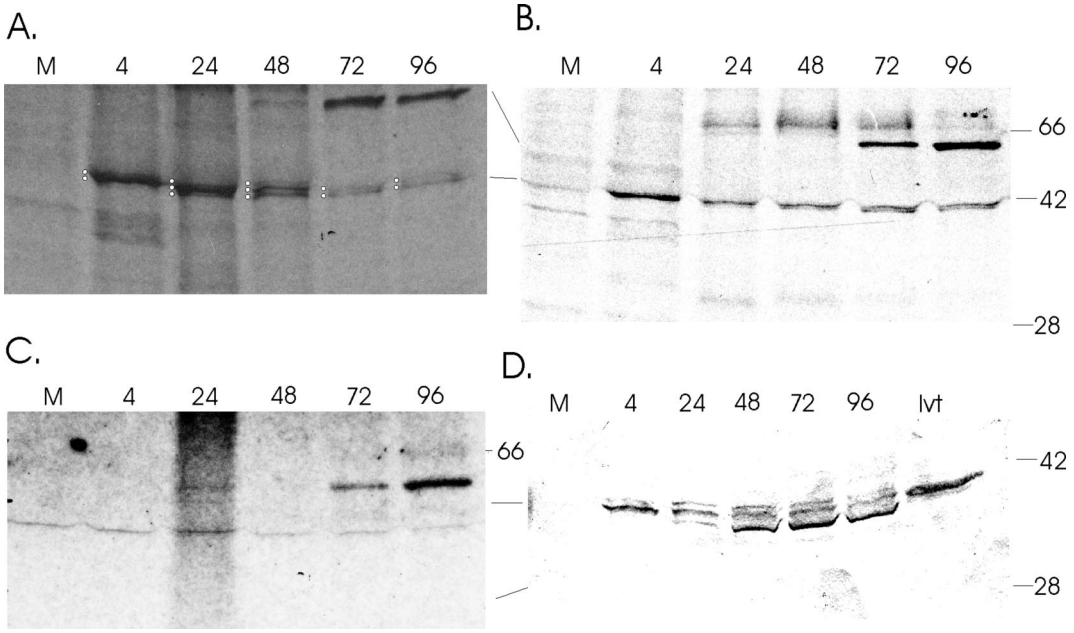

FIG. 3. Analysis of UL34 protein expression. (A and B) Immunoprecipitation of [^{35}S]methionine-labeled UL34 proteins. Mock-infected or HCMV-infected cells were radiolabeled with [^{35}S]methionine at the indicated hours postinfection, immunoprecipitated using antisera to UL34, and analyzed by electrophoresis and autoradiography. Antisera from two different rabbits were used for the experiments illustrated. The three UL34 proteins are indicated by white circles in panel A; the 45-kDa UL34 protein is not easily visualized in panel B because of the exposure time. (C) Immunoprecipitation of [^{32}P]-labeled proteins. (D) Western blot analysis of UL34 proteins. Mock or HCMV-infected cells were harvested at the indicated times postinfection and analyzed by Western blotting. The positions of the molecular weight markers are indicated; M mock-infected; 4, 24, 48, 72, and 96, hours postinfection; ivt, in vitro-translated pUL34.

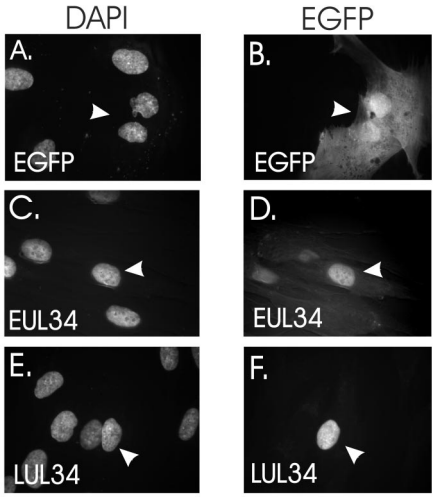

FIG. 4. Analysis of the intracellular location of the early and late UL34 proteins. The early and late UL34 open reading frames were expressed as enhanced green fluorescent protein (EGFP) fusion proteins following transfection of human diploid fibroblasts with pBJ507 (9) or pBJ581, respectively. The late open reading frame was cloned using oligonucleotides 316 (5′ CCGGAATTCGAGATGCGTGACAACGTG 3′) and 317 (5′ CGCGTCGACTTATTGTTCTCCAGTGACG 3′), inserting the amplimer into pEGFP-C2 (Clontech). Constructs expressing fluorescent proteins were transfected into human diploid fibroblasts using Effectene (QIAGEN, Valencia, Calif.) as directed by the manufacturer. Following transfection, cells were stained with 4′,6′-diamidino-2-phenylindole. Fluorescence was visualized using a Nikon epifluorescence microscope. (A, C, and E) DAPI staining of cell nuclei; (B, D, and F) green fluorescence of EGFP; (A and B) EGFP; (C and D) the early pUL34-EGFP fusion protein; (E and F) the late pUL34-EGFP fusion protein. The transfected cells are indicated by arrows.

sion from the US3 promoter, similar to pUL34^{43kDa} (9). The DNA binding activities of the early and late forms of pUL34 were investigated using electrophoretic mobility shift assays. The early and late forms of pUL34 and the control protein, luciferase, were synthesized by in vitro transcription-and-translation experiments. Successful synthesis of the proteins was confirmed by protein gel electrophoresis and autoradiography (data not shown). The synthesized proteins were incubated with radiolabeled double-stranded DNA fragments containing the US3 *tre* (3, 9); DNA-protein mixtures were analyzed by native gel electrophoresis. As depicted in Fig. 5A, both early and late pUL34 formed complexes with US3 *tre*, while the control protein, luciferase, did not. The DNA-protein complexes formed were sequence specific, since utilization of a mutant form of *tre* (9) did not result in the formation of DNA-protein complexes (data not shown).

pUL34^{40kDa} was also analyzed for effects on US3 expression. Transient-expression experiments were performed, utilizing reporter constructs that express the *lacZ* gene under the control of the US3 promoter and containing either a wild-type or a mutant form of *tre* (2). A promoterless plasmid containing the *lacZ* gene was also used to control for background levels of enzyme activity. The reporter plasmids were transfected into human diploid fibroblasts along with plasmids expressing either pUL34^{43kDa} or pUL34^{40kDa}. A plasmid expressing HCMV immediate-early 1 and immediate-early 2 proteins was included in the transfections to maximize levels of gene expression. Levels of beta-galactosidase activity were determined by measuring the fluorescence of the cleavage product of methylumbelliferyl-β-D-galactoside that was released into the medium of transfected cells ~48 h after transfection. The early and late forms of pUL34 were equally able to repress expression from the US3 promoter in the presence of functional *tre* (Fig. 5B).

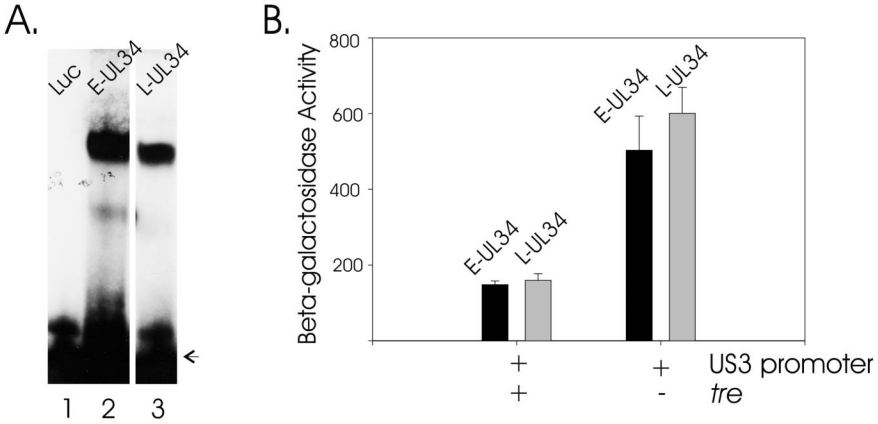

FIG. 5. Comparison of the functions of the early and late UL34 proteins. (A) Electrophoretic mobility shift analysis of the DNA binding activity of in vitro-synthesized early UL34 (E-UL34) and late UL34 (L-UL34) proteins. The unbound DNA is indicated by an arrow; Luc is control luciferase protein. (B) Repression of the US3 promoter by the early and late UL34 proteins. Transient expression assays were used to measure the expression levels of a reporter construct expressing the *lacZ* gene under the control the US3 promoter and *tre* (1) or expression levels of a similar reporter construct that contains a mutant *tre* (pBJ214) (2) in the presence of a construct that expresses the early (E-UL34) (9) or late (L-UL34) form of the UL34 protein (pBJ575). pBJ575, which expresses the late UL34 open reading frame under the control of the HCMV major immediate-early promoter, was constructed by amplifying the late UL34 open reading frame using oligonucleotides 311 (5′ CGTCTAGAGGATCCACTTCTCCAACGACGATTC 3′) and 219 (5′ CTCGTCGACTTAAATACACAACGGGGTTATGG 3′) and inserting the amplimer into pBJ201 (2). This experiment was repeated multiple times with similar results; the results depicted are the averages from two experiments, with error bars representing one standard deviation. Background levels of beta-galactoside activity obtained from cells receiving the promoterless *lac*Z plasmid, pEQ3 (4), were subtracted from the experimental samples.

Discussion. Although sequence analyses have identified approximately 189 genes in the HCMV genome, the synthesis of multiple proteins from one gene allows HCMV to generate far more proteins than predicted. UL34 is yet another HCMV gene that encodes more than one protein. Multiple transcripts are generated from the UL34 open reading frame, with the monocistronic transcripts encoding three highly related proteins. The two major proteins are differentially expressed during infection, with $pUL34^{43kDa}$ predominating early in infection and $pUL34^{40kDa}$ predominating late in infection. A third, low-abundance UL34 ($pUL34^{45kDa}$) protein is also synthesized at early times postinfection. We were unable to detect processing of the UL34 proteins in pulse-chase experiments (data not shown), suggesting that the three proteins arise from individual translation initiation events occurring on the monocistronic UL34 transcripts. $pUL34^{45kDa}$ is predominantly an early protein, leading us to hypothesize that this protein results from translational initiation from the noncanonical ACG codon, present in frame on the early UL34 transcripts. ACG codons have been identified as initiators of translation in multiple systems, particularly in viral systems (see reference 5 for an example).

UL34 functions as a transcription factor, down-regulating expression of the US3 gene. Similar to many other transcription factors, UL34 binds to a sequence-specific DNA element. The activity of UL34 is presumably mediated through additional proteins, either cellular or viral in nature. Intriguingly, at late times postinfection, pUL34 coprecipitates with a phosphoprotein of ~60 kDa. The identity of this protein is unknown; however, the detection of this protein at late times of infection, when host cell protein synthesis is markedly decreased, suggests that it is a virally encoded protein.

UL34 expression is complex, with different transcripts initiating at different sites at different times postinfection, resulting in the synthesis of proteins that to date have no identified differences in function. The identification of UL34 as an essential gene suggests that continued expression throughout infection is critical for viral replication. The utilization of different transcription initiation start sites and presumably different promoters allows the virus to continually transcribe an essential gene throughout infection, despite modifications of the transcriptional machinery that occur during viral gene expression and genome replication.

The identification of UL34 as an essential gene is an intriguing observation, placing UL34 in relatively small group of viral genes (7, 15). We also have shown that UL34 is an essential gene (unpublished results). The majority of essential genes characterized to date contribute to viral replication by providing transcription factors, enzymes for DNA replication, and packaging and structural proteins needed to form new virion particles and initiate a new round of infection. The continual expression of UL34 and the essential nature of this gene suggest that it has roles during viral infection, in addition to US3 regulation. Indeed, an analysis of the HCMV genomic sequence predicts multiple UL34 binding sites within the genome, most notably in sequences flanking the lytic origin of replication. UL34 is not essential for *ori-lyt* replication in the absence of viral infection; however, it may play a role in viral DNA replication in infection. It is clear that additional experiments are needed to ascertain all of the roles this gene and the encoded proteins play in viral replication.

This work was supported by an American Cancer Society grant and an American Heart Association Grant-In-Aid, 0255229N, to B.J.B.

We thank Dave Bradley for maintaining the epifluorescence microscope, Dan Richardson for performing the anti-pp65 Western blot analysis of UL34 immunoprecipitates, and Radhika Iyer for assistance with generating antisera to pUL34.

REFERENCES

1. **Ahn, K. S., A. Angulo, P. Ghazal, P. A. Peterson, Y. Yang, and K. Früh.** 1996. Human cytomegalovirus inhibits antigen presentation by a sequential multistep process. Proc. Natl. Acad. Sci. USA **93:**10990–10995.
2. **Biegalke, B. J.** 1995. Regulation of human cytomegalovirus US3 gene transcription by a *cis*-repressive sequence. J. Virol. **69:**5362–5367.
3. **Biegalke, B. J.** 1998. Characterization of the transcriptional repressive element of the human cytomegalovirus immediate-early US3 gene. J. Virol. **72:**5457–5463.
4. **Chee, M. S., A. T. Bankier, S. Beck, R. Bohni, C. M. Brown, R. Cerny, T. Horsnell, C. A. Hutchison III, T. Kouzarides, J. A. Martignetti, E. Preddie, S. C. Satchwell, P. Tomlinson, K. M. Weston, and B. G. Barrell.** 1990. Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. Curr. Top. Microbiol. Immunol. **154:**125–170.
5. **Curran, J., and D. Kolakofsky.** 1988. Ribosomal initiation from an ACG codon in the Sendai virus P/C mRNA. EMBO J. **7:**245–251.
6. **Davison, A. J., A. Dolan, P. Akter, C. Addison, D. J. Dargan, D. J. Alcendor, D. J. McGeoch, and G. S. Hayward.** 2003. The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. J. Gen. Virol. **84:**17–28.
7. **Dunn, W., C. Chou, H. Li, R. Hai, D. Patterson, V. Stolc, H. Zhu, and F. Liu.** 2003. Functional profiling of a human cytomegalovirus genome. Proc. Natl. Acad. Sci. USA **100:**14223–14228.
8. **Jones, T. R., E. J. H. J. Wiertz, L. Sun, K. N. Fish, J. A. Nelson, and H. L. Ploegh.** 1996. Human cytomegalovirus US3 impairs transport and maturation of major histocompatibility complex class I heavy chains. Proc. Natl. Acad. Sci. USA **93:**11327–11333.
9. **LaPierre, L., and B. J. Biegalke.** 2001. Identification of a novel transcriptional repressor encoded by human cytomegalovirus. J. Virol. **75:**6062–6069.
10. **Liu, W., Y. Zhao, and B. J. Biegalke.** 2002. Analysis of human cytomegalovirus US3 gene products. Virology **301:**32–42.
11. **Liu, Y., and B. J. Biegalke.** 2002. The human cytomegalovirus UL35 gene encodes two proteins with different functions. J. Virol. **76:**2460–2468.
12. **Mocarski, E. S., Jr.** 1996. Cytomegaloviruses and their replication, p. 2447–2492. *In* B. N. Fields, D. M. Knipe, and P. M. Howley (ed.), Fields virology, vol. 2. Lippincott-Raven, Philadephia, Pa.
13. **Murphy, E., D. Yu, J. Grimwood, J. Schmutz, M. Dickson, M. A. Jarvis, G. Hahn, J. A. Nelson, R. M. Myers, and T. Shenk.** 2003. Coding potential of laboratory and clinical strains of human cytomegalovirus. Proc. Natl. Acad. Sci. USA **100:**14976–14981.
14. **Welch, R. W., L. M. McGregor, and W. Gibson.** 1991. Cytomegalovirus homologs of cellular G protein-coupled receptor genes are transcribed. J. Virol. **65:**3915–3918.
15. **Yu, D., M. C. Silva, and T. Shenk.** 2003. Functional map of human cytomegalovirus AD169 defined by global mutational analysis. Proc. Natl. Acad. Sci. USA **100:**12396–12401.