JOURNAL OF VIROLOGY, Sept. 1995, p. 5304–5310 Vol. 69, No. 9
0022-538X/95/$04.00+0

The Human Cytomegalovirus UL98 Gene Transcription Unit Overlaps with the pp28 True Late Gene (UL99) and Encodes a 58-Kilodalton Early Protein

BAO-LING ADAM,[1] TABMITHA Y. JERVEY,[1] CATHERINE P. KOHLER,[1]
GEORGE L. WRIGHT, JR.,[1] JAY A. NELSON,[2]
AND RICHARD M. STENBERG[1]*

Department of Microbiology and Immunology, Eastern Virginia Medical School, Norfolk, Virginia 23501,[1] and Department of Microbiology and Immunology, Oregon Health Sciences University, Portland, Oregon 97201[2]

Received 8 March 1995/Accepted 15 May 1995

A murine monoclonal antibody (I2) reacts strongly with the nucleus of human cytomegalovirus (HCMV)-infected human fibroblasts. Western blot (immunoblot) analysis using I2 demonstrated that a protein with an apparent molecular mass of 58-kDa (E58) was expressed at 5 h after infection, and levels increased through 72 h. Immunoblot screening of an early cDNA expression library resulted in a positive clone which hybridized to the right end of the *Xba*I C fragment of the HCMV Towne strain. Further analysis demonstrated that the E58-specific clone was homologous to the putative UL98 open reading frame, which has been proposed to encode the viral alkaline exonuclease homolog. RNA analysis demonstrated a 3.0-kb RNA which is expressed at early times after infection, as well as in the absence of viral DNA replication, and which is 3′ coterminal with the pp28 (UL99) gene region. Insertion of the UL98 genomic sequence into a eucaryotic expression vector and subsequent Western blot analysis using I2 demonstrated that the expressed protein comigrated with E58 from infected cells. E58 also reacts specifically with a previously described antibody, anti-$P_{2\text{-}1}$, which was proposed to recognize a putative late 58-kDa protein. E58 comigrates with the putative late 58-kDa protein, indicating that these two proteins are likely the same. Analysis of the UL98 promoter revealed a TATATAA sequence located at nucleotide 142525. Insertion of the putative promoter 5′ to a reporter gene demonstrated that the UL98 promoter was activated in cotransfection experiments with IE1 and IE2 proteins. These studies demonstrate that UL98 is a bona fide early gene, which is consistent with its probable role as the viral alkaline exonuclease gene.

Human cytomegalovirus (HCMV) is an opportunistic pathogen affecting immunocompromised or immunosuppressed individuals (1). HCMV is generally acquired early in life and in healthy individuals causes benign or asymptomatic infections. The virus remains latent in infected individuals and reactivates to cause numerous clinical syndromes, some of which are life threatening. In addition, HCMV crosses the placenta and can infect the fetus in utero, which often results in congenital birth defects of varying severity.

HCMV is a large, double-stranded, linear DNA virus whose genome measures approximately 230 kbp. The sequence of the AD169 strain has been determined, thereby offering significant information regarding the genetic complexity of this virus (3). However, relatively few genes have been analyzed, and a considerable effort will be necessary to characterize this complex virus. As in other herpesviruses, replication proceeds in a coordinated way, with characteristic classes of viral genes being expressed in a cascade-like manner. Three phases of virus gene expression occur, and they have been described as immediate early (IE), early, and late (5, 22, 40). From a broad perspective, IE genes are thought to be responsible for regulating the course of virus replication (reviewed in reference 32), early genes are responsible for DNA replication (reviewed in references 29 and 32), and late proteins are the structural components of the virus (reviewed in reference 8). This simplistic version of virus replication probably represents a gross underestimate of the complexity of these three classes of virus genes.

While early genes have been defined as genes expressed in the presence of viral protein synthesis and the absence of viral DNA replication, they represent a complex class of genes that have varied patterns of expression (reviewed in references 29 and 32). Some genes are expressed rapidly and at high levels soon after infection, whereas others are expressed maximally during the late phase of virus replication. In addition, early genes are expressed at either increasing, constant, or decreasing levels as infection progresses (2, 6, 9–11, 19, 23, 28, 31, 41). Clearly, numerous influences must impact on the regulation of early gene expression.

To date, only a few early genes have been extensively characterized with respect to their regulation during infection. Collectively, these studies imply that viral IE proteins are important for activation of early promoters and that specific promoter sequences and cell-protein interactions are important for the activation process (2, 6, 13–15, 20, 30, 34, 38). Importantly, the early phase of the virus life cycle is necessary for preparing the cell for the onset of viral DNA replication. As virus replication and spread are necessary for causing disease and because poor or inefficient replication could result in latency or persistence, it is vital that we understand the regulation of this important phase of the virus life cycle.

In an effort to expand our knowledge of HCMV early genes and promoters, we have identified a murine monoclonal antibody (I2) that reacts with a viral protein in infected cells within

* Corresponding author. Mailing address: Department of Microbiology and Immunology, Eastern Virginia Medical School, P.O. Box 1980, Norfolk, VA 23501. Phone: (804) 446-5684. Fax: (804) 624-2255.

the first 5 h after infection. We have used expression vector cloning to identify the gene that encodes the I2-reactive protein. Our studies indicate that I2 reacts with an early 58-kDa protein (E58) originating from UL98, the open reading frame (ORF) that corresponds to the putative alkaline exonuclease homolog (3, 18). The E58 transcription unit is 3′ coterminal with the true late pp28 (UL99) gene region, demonstrating nested 3′ ends for genes of different kinetic classes. Lastly, our studies raise questions about a previous report on the mapping of the E58 protein (16).

MATERIALS AND METHODS

Cells and virus. The growth of human fibroblasts (HF) and the propagation of HCMV (Towne) have been described previously (36, 37, 40). The DNA-negative temperature-sensitive mutant *ts*66 has also been described elsewhere (33). In all experiments, virus infections were carried out at 10 to 20 PFU per cell and cells were harvested as described below.

Western blot analysis of viral proteins. HF were infected with HCMV and harvested at the indicated times for total cell protein and Western blot (immunoblot) analysis as previously described (33). The proteins were reacted with a 1/500 dilution of monoclonal antibody I2 and detected by using the ECL chemiluminescent system (Amersham). Murine monoclonal antibody I2 was generated as described in a previous study (27) and characterized in J. A. Nelson's laboratory. Antibody anti-$P_{2\text{-}1}$ was the gift of Stephen St. Jeor and has been described previously (16). Molecular masses of proteins were determined by using rainbow marker standards (Amersham).

Generation of an early cDNA expression library. To generate an early cDNA library, poly(A)$^+$ RNA was isolated with the Fast Track system (Invitrogen) from cells infected with HCMV for 24 h. Double-stranded cDNAs were synthesized with the ZAP cDNA synthesis kit (Stratagene) and cloned into a lambda ZAP XR expression vector exactly as recommended by the supplier. Analysis of the primary library demonstrated that 9.1×10^5 independent clones were represented.

Immunoscreening with antibody I2. The HCMV early cDNA library was plated at approximately 3,000 to 5,000 plaques per plate and transferred to nitrocellulose membranes. The membranes were pretreated as recommended by the supplier and reacted with a 1/500 dilution of I2. Individual plaques were picked and subjected to three more rounds of screening to isolate pure plaques. Plasmid pI2-3.1, containing the cDNA of interest, was rescued with helper phage and subsequently characterized.

Characterization of I2-specific cDNA. To identify the location of the gene encoding the I2-specific protein, the insert of pI2-3.1 was radiolabeled and hybridized to Southern blots containing plasmid DNAs representing *Xba*I fragments of the HCMV genome (provided by M. Stinski). DNA fragments that proved positive for hybridization were digested with additional restriction enzymes and subjected to Southern blot analysis to further localize the I2-specific region to the HCMV genome. The I2-positive cDNA clone was found to be ~1.8 kb long and was subsequently characterized by DNA sequence analysis using the dideoxy chain termination method as described previously (33). Nucleotide sequence designations were as described elsewhere (3).

RNA isolation and Northern blot analysis. Cells were infected with wild-type Towne strain or *ts*66, and total cell RNA was isolated by previously published procedures (33). Northern (RNA) blot analysis was performed as described previously (33), using a probe corresponding to the I2-specific insert which was excised from the vector by using *Xho*I and *Eco*RI sites in the polylinker.

Expression of UL98 in eucaryotic cells. The construction of vectors expressing E58 is described below (also see Fig. 5A). HCMV IE1 expression vector pSVCC3 (6, 34) was digested with *Bam*HI and *Sst*II to remove the coding sequences of IE72. The DNA was recircularized by using a synthetic double-stranded oligonucleotide (BLA-1) containing *Bam*HI and *Sst*II cohesive ends and internal *Xho*I, *Eco*RI, and *Mlu*I restriction enzyme sites. To generate pE58, the plasmid pMIEP-BLA1 was digested with *Xho*I and *Mlu*I, and HCMV genomic sequences between the *Xho*I (nucleotide [nt] 142712) and *Mlu*I (nt 147055) sites were inserted. BLA-1, which contains no UL98-specific sequence, was generated by annealing two complementary oligonucleotides (5′-GGCTC GAGCTGCAGGAATTCGACGCGTG-3′ and 5′-GATCCACGCGTCGAAT TCCTGCAGCTCGAGCCGC-3′). To generate pE58.2, pE58 was digested with *Sst*II and *Xho*I to remove 5′ linker sequences, and a double-stranded oligonucleotide (BLA-2) was inserted into the *Sst*II and *Xho*I sites. Oligonucleotide BLA-2 contains E58-specific sequences 5′ to the *Xho*I site (nt 142712) to reconstruct the ATG at nt 142701 of the HCMV sequence. BLA-2 was generated by annealing oligonucleotides 5′-GGAGATCTATCATCATCTGGGGCGTC-3′ and 5′-TCGAGACGCCCCACATGATGATAGATCTCCGC-3′. These constructs place the E58 coding sequence immediately 3′ to the major IE promoter (MIEP). To generate a construct expressing the E58 cDNA, pI2-3.1 was digested with *Xho*I and *Eco*RI and cloned into the analogous sites of the pMIEP-BLA1 construct. This places the cDNA coding sequences adjacent to the MIEP. Plasmids were transfected into COS cells and analyzed for protein as previously described (35). Protein was detected by Western blotting using I2 as described above.

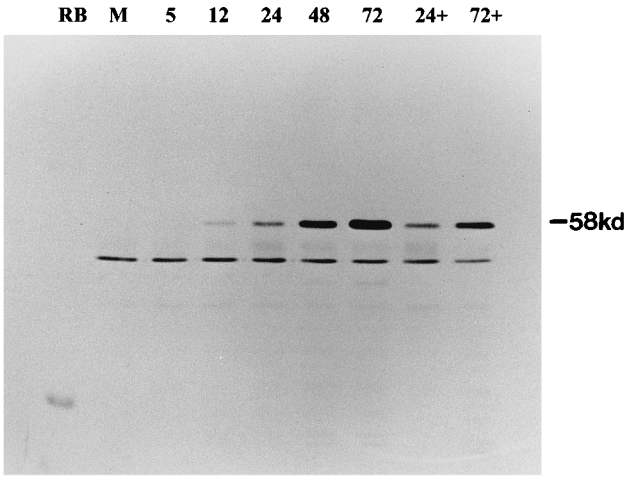

FIG. 1. Time course analysis of I2-reactive proteins. HFs were infected with HCMV in the presence (+) or absence of phosphonoacetic acid or were mock infected (lane M). Cells were harvested at the indicated times (in hours) after infection and analyzed by Western blotting as described in Materials and Methods. The molecular mass of the I2-reactive protein is indicated (in kilodaltons). Lane RB, rainbow markers.

Cloning of the UL98 promoter and expression in eucaryotic cells. The promoter of the UL98 gene was isolated as a fragment by digestion with *Bsr*FI (New England Biolabs) at nt 142245 and 142666. The DNA was blunted, *Hin*dIII linkers were added to the DNA, which was subsequently digested with *Hin*dIII, and the fragment was cloned into the *Hin*dIII site of pSVOCATd as previously described (6, 34). The promoter was tested in HF for activation by IE proteins exactly as described previously (6, 34).

RESULTS

Characterization of I2-specific proteins in HCMV-infected cells. We have long been interested in the early events that influence HCMV replication (6, 12, 34). In an effort to identify viral genes expressed early in infection, we evaluated a number of uncharacterized murine monoclonal antibodies to HCMV-infected cells. In immunofluorescence analyses, one antibody, I2, reacted strongly with the nuclei of AD169-infected cells at 12 h after infection. Control uninfected cells did not react with this antibody (data not shown).

To extend these studies, cells were infected with HCMV Towne strain and at various times after infection were harvested for sodium dodecyl sulfate-polyacrylamide gel electrophoresis and Western blot analysis. The data in Fig. 1 demonstrate that a 58-kDa protein is readily detected by 12 h after infection and that the level of this protein gradually increases throughout the course of infection. Longer exposures of the blot demonstrated that the protein was detectable in infected cells at 5 h after infection (data not shown). In the presence of phosphonoacetic acid, an inhibitor of virus DNA replication, levels also increase at 72 h but not to the extent as in untreated infected cells.

Localization of E58 coding sequences. To identify the gene encoding the E58 protein, a cDNA library was prepared with HCMV-infected cell RNA isolated at 24 h after infection. A number of positive clones were identified and isolated after multiple rounds of plaque purification (data not shown). One clone, I2-3.1, was selected for further analysis. To identify the location of the E58 coding sequences, the cDNA insert was hybridized to Southern blots of HCMV recombinant clones. Our studies demonstrated that the cDNA insert hybridized to

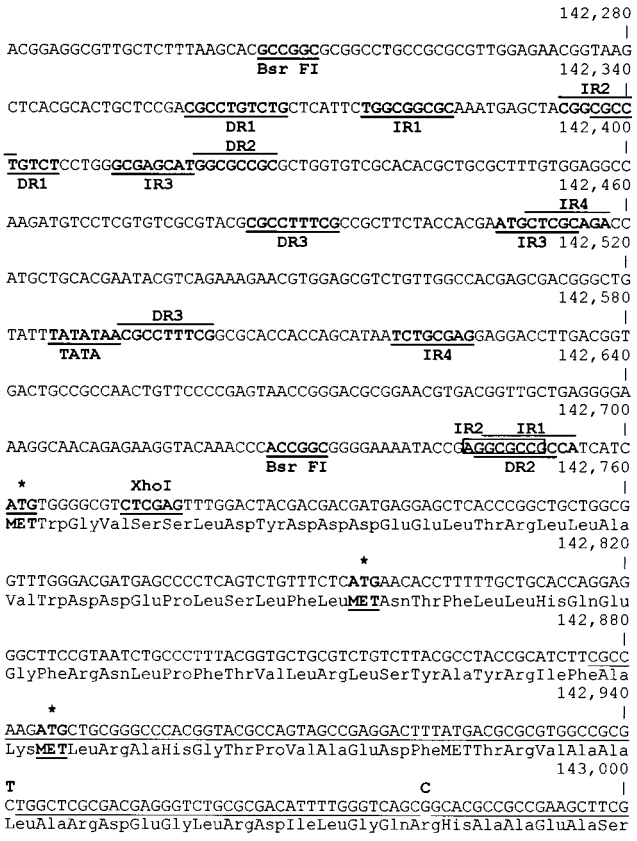

FIG. 2. Nucleotide sequence analysis of pI2-3.1 and comparison with the AD169 sequence from nt 142221 to 143000. Homology of pI2-3.1 with the published sequence (underlined), divergent bases characteristic of the Towne strain (boldface bases above the sequence), the putative TATA moiety (underlined), and potential ATG translation initiation codons (asterisks) are indicated. The *Xho*I site used for cloning manipulations and the locations of the *Bsr*FI sites used to clone the promoter are indicated. Direct repeats (DR) and inverted repeats (IR) of interest are also shown. The ATG methionine at position 142701 is underlined.

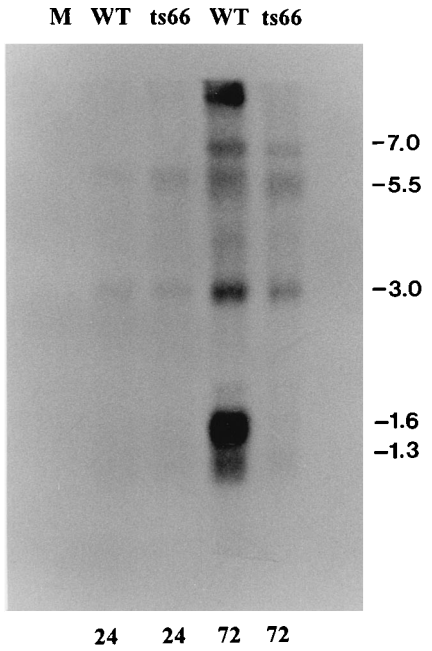

FIG. 3. Analysis of pI2-3.1-specific RNA at early and late times after infection. Cells were mock infected (lane M) or infected with HCMV Towne strain (lanes WT) or a DNA-negative temperature-sensitive mutant (lanes *ts*66) at 39.5°C and harvested at the indicated times (in hours). Molecular sizes of the transcripts are indicated on the right in kilobases.

a region of the HCMV genome overlapping ORFs UL98 and UL99, within *Xba*I-C (data not shown). To confirm the location of the E58 cDNA, sequence analysis of the 5′ end of the cDNA was performed. The data in Fig. 2 demonstrate that the 5′ end of the cDNA is homologous to the 5′ end of the predicted UL98 coding sequence beginning at nt 142877. While the 3′ end of the cDNA was not confirmed, the cDNA extended ~1.8 kb from the 5′ end at nt 142877. With the exception of two mismatched bases at nt 142941 and 142981, the Towne sequence is identical to that of the AD169 strain within the amino terminus of the protein. These data indicate that E58 likely corresponds to the UL98 gene product.

Analysis of RNA from the E58 gene region. To address the RNA originating from the E58 (UL98) gene region, Northern blot analysis of RNA from HCMV-infected cells was performed. Cells were harvested at 24 and 72 h after infection and hybridized to the E58 cDNA. As a comparison, cells infected with *ts*66, a DNA-negative temperature-sensitive mutant (33), were analyzed in parallel. The data in Fig. 3 show an ~3.0-kb RNA present in infected cells at 24 and 72 h after infection. This is approximately the expected size of an RNA that would initiate downstream of a putative TATA sequence at nt 142523 and terminate downstream of the pp28 (UL99) poly(A) sequence at nt 145,135. Consistent with the latter point is the presence of the true late 1.6- and 1.3-kb pp28 (UL99) RNAs that have been previously characterized (7, 16, 21, 24). These data imply that the UL98 transcripts initiate immediately upstream of the UL98 ORF and are 3′ coterminal with the downstream UL99 transcripts. Additional, larger transcripts of 5.5 and 7.0 kb are also present. These RNAs hybridize to UL98- and UL99-specific probes (Fig. 3) (7), and recent studies demonstrate that they originate from further-upstream promoters (41).

Comparative analysis of antibodies I2 and anti-P_{2-1}. Earlier studies of this region using an antipeptide antibody (anti-P_{2-1}) had identified a 58-kDa late protein that was concluded to be the product of the 1.6-kb late mRNA from the pp28 gene region (16, 17). This protein was proposed to be generated by way of a translational read-through because the coding region of the late 1.6-kb RNA was unable to encode a protein of 58 kDa (17). These studies were performed prior to the availability of the HCMV sequence, which subsequently revealed an ORF (UL98) that overlapped this region (3). In fact, P_{2-1} is a peptide within UL98. To determine whether E58 (UL98) was the same protein as previously described, we tested antibody I2 versus anti-P_{2-1} in a time course analysis of viral proteins. Our data in Fig. 4 demonstrate that these two antibodies give identical profiles when tested directly. Both antibodies react with an early 58-kDa protein which accumulates with similar kinetics. In addition, other, smaller proteins are evident and react with both antibodies. The origins of these proteins are presently unclear. Also, use of anti-P_{2-1} demonstrates E58 as early as 5 h after infection. These data suggest that the two proteins are the same.

Identification of the translation start site for E58. Several potential ATG start sites exist at nt 142701, 142794, and 142884. To determine the location of the translation start site for the E58 protein, we tested several expression constructs for their ability to produce the E58 protein in eucaryotic cells. A

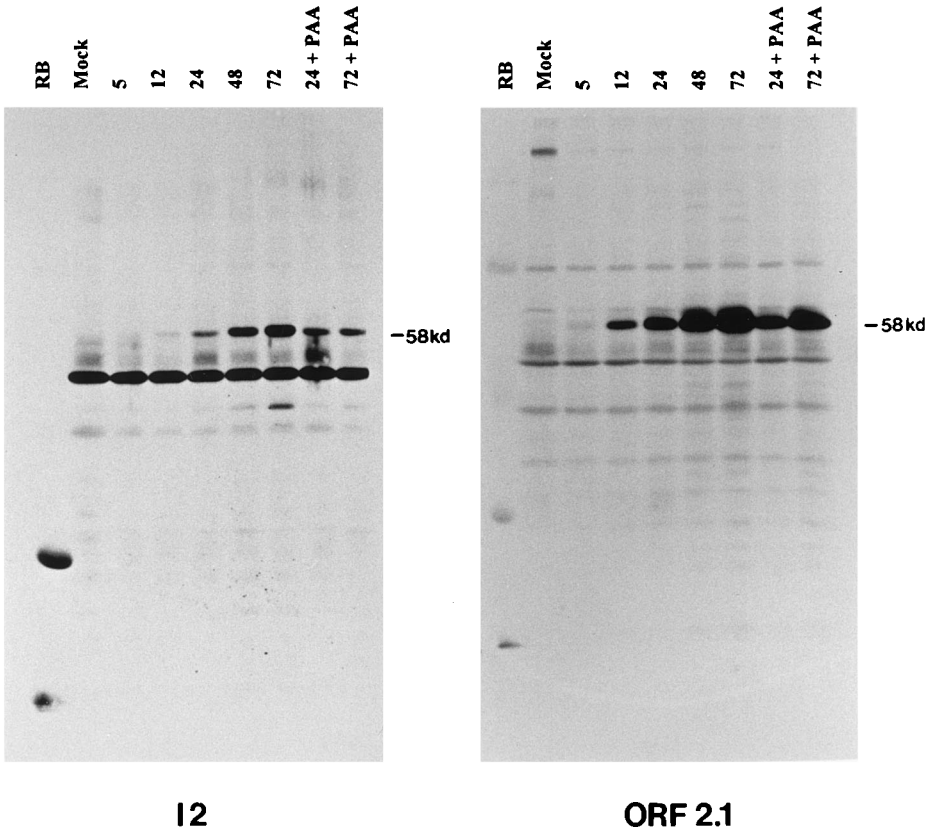

FIG. 4. Comparative time course analysis of I2- and anti-P_{2-1}-reactive proteins. Analyses were carried out exactly as described in the legend to Fig. 1 and in Materials and Methods. Cells were harvested at the indicated times (in hours) after infection, and blots were reacted with I2 or anti-P_{2-1} (ORF 2-1). Lanes RB, rainbow markers. PAA, phosphonoacetic acid.

construct from *Xho*I (nt 142712) to *Mlu*I (nt 147055) was cloned into the IE1 expression vector pSVCC3 (6, 34), replacing the IE1 coding sequences with the E58 coding sequences as described in Materials and Methods and the legend to Fig. 5A. This places the E58 coding sequences under control of the MIEP. This construct is capable of initiating at the two most downstream ATG codons (nt 142794 and 142884) but not the most 5′ codon (at nt 142701). Transfection of this construct into COS cells and comparison with native protein produced in HCMV-infected cells demonstrated that the COS cell-expressed protein migrated slightly faster than the native protein (Fig. 5B). In addition, I2 and anti-P_{2-1} gave identical profiles, demonstrating that P_{2-1} is a component of E58.

To directly test the ATG codon at nt 142701, we modified our construct to insert sequences missing from the ATG to the *Xho*I (nt 142712) site (pE58.2). Transfection of this construct in COS cells yielded a protein that comigrated with native E58 (Fig. 5C). These data demonstrate that the ATG at nt 142701 is likely to be the translation initiation site for the E58 protein and that the ORF extends to a TGA sequence at nt 144453. This generates a theoretical protein of 65,273 Da according to computer analysis.

In addition to testing the genomic constructs (pE58 and pE58.2), the E58 cDNA beginning at nt 142877 was inserted downstream of the MIEP by using *Xho*I and *Eco*RI sites located in the polylinker of the plasmid vector. Surprisingly, this construct failed to express any detectable protein that reacted with I2 or anti-P_{2-1} (Fig. 5B). This implies that (i) the ATG at position 142884 may not be a functional start codon and the truncated protein described above is likely to originate from the ATG at nt 142794 or (ii) the E58 ORF in the cDNA is incomplete. Additional studies will be necessary to resolve these possibilities.

Identification of the promoter for the UL98 gene. As stated above, a putative TATA motif exists at nt 142523. This coupled with the location of the ATG at nt 142701 indicated that the promoter might lie within this general vicinity. To test this, sequences between nt 142245 and 142666 were cloned 5′ to a chloramphenicol acetyltransferase (CAT) reporter gene and transfected into HF in the presence or absence of pSVH, a construct that expresses IE1 and IE2 proteins (6, 34). The data in Fig. 6 demonstrate that IE proteins efficiently activate the E58 promoter but that, like other early promoters studied, the E58 promoter is poorly activated in the absence of IE proteins. Inspection of the putative E58 promoter sequence reveals a number of direct and inverted repeats which could be important for regulating E58 promoter activity (Fig. 2). This is consistent with other early and late promoters which have been shown to regulate expression through these types of repeats. Collectively, these data indicate that the sequences between nt 142245 and 142666 contain a functional promoter that behaves like other HCMV early promoters.

DISCUSSION

In this study, we identified the protein product for the UL98 ORF, which encodes the putative alkaline exonuclease homolog. In addition, we have examined its expression through-

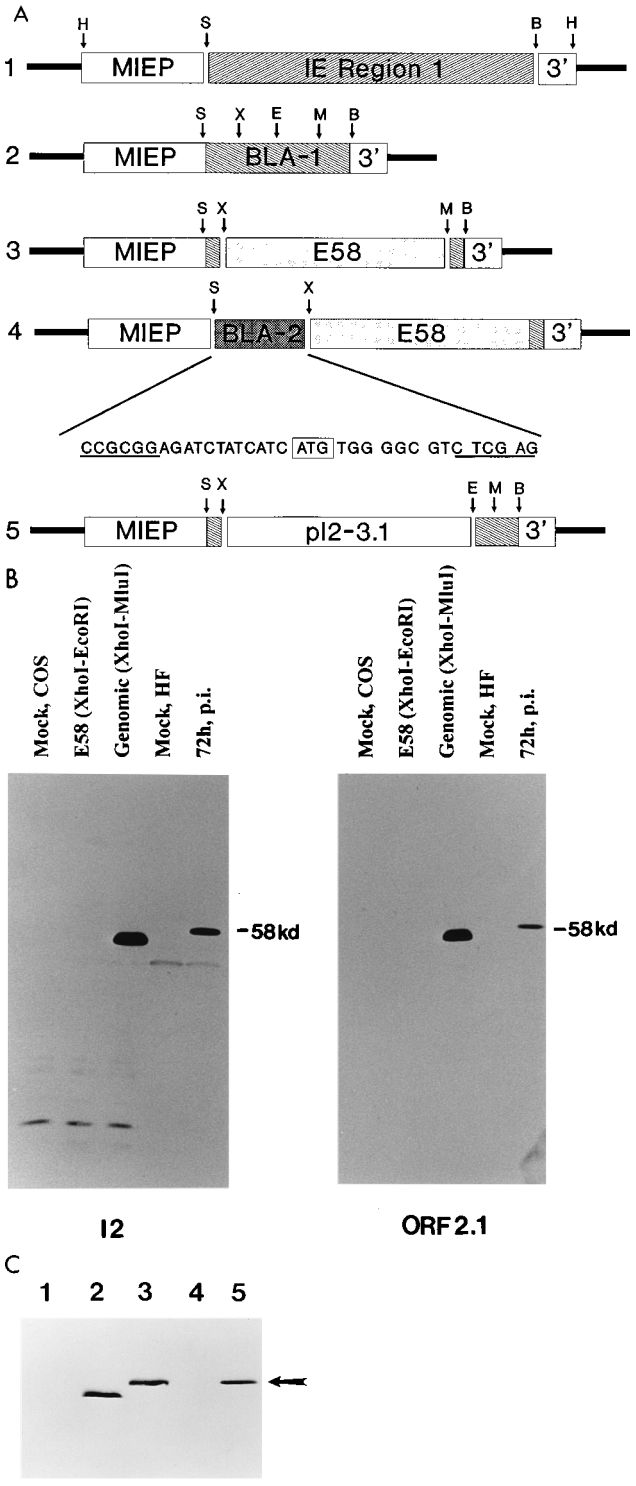

FIG. 5. Expression of E58 sequences in COS cells. (A) Generation of expression constructs. Line 1, parental IE1 expression vector pSVCC3, which has been described previously (6, 34). Line 2, insertion of oligonucleotide BLA-1 into the *Sac*II (S) and *Bam*HI (B) sites of pSVCC3 to generate pMIEP-BLA1. Line 3, insertion of E58 genomic sequences from *Xho*I (X) (Fig. 2) to *Mlu*I (M) at compatible internal sites within BLA-1 to construct the vector pE58. Line 4, modification of pE58 to include upstream sequences (BLA-2) missing from pE58 (shown below the line). BLA-2 was inserted from *Sac*II to *Xho*I to create the vector pE58.2. The positions of the MIEP and IE1 3′ end and the ATG at nt 142701 (boxed) are indicated. Line 5, pI2-3.1 cDNA cloned into pMIEP-BLA1 at the *Xho*I and *Mlu*I sites as described in Materials and Methods. H, *Hin*dIII; E, *Eco*RI. (B) Western blot analysis of E58 expression in COS cells. COS cells were transfected with the indicated plasmids or mock transfected and analyzed as previously described (33). Immunoblot analysis was done as for Fig. 1. Lanes Mock, COS, mock-transfected COS cells; lanes E58 (XhoI-EcoRI), E58 cDNA (pI2-3.1) cloned from the *Xho*I site in the polylinker to the *Eco*RI site in the polylinker; lanes Genomic (*Xho*I to *Mlu*I), genomic sequences from *Xho*I to *Mlu*I as described for panel A (pE58); lanes Mock, HF, mock-infected HFs; lanes 72h,p.i., HFs infected with HCMV Towne for 72 h. (C) Comparative analysis of pE58 and pE58.2 in COS cells. Analyses were done as for panel B. Lane 1, mock-transfected COS cells; lane 2, pE58-transfected COS cells; lane 3, pE58.2-transfected COS cells; lane 4, mock-infected HFs; lane 5, 72-h Towne-infected HFs. The position of E58 is indicated (arrow).

out the course of infection and identified its promoter. Consistent with a recent report, RNAs expressed from this region are 3′ coterminal with RNAs encoded by the pp28 gene region (UL99) demonstrating nested 3′ ends which are regulated by different promoters. As is typical for other early promoters, the E58 promoter is poorly expressed when transfected into cells but is readily activated by a plasmid expressing viral IE proteins. Lastly, we have raised a question concerning the mapping of this protein by a previous study.

Our antibody, I2, reacted with a 58-kDa protein expressed early in infection. This protein was of particular interest as it was expressed by 5 h after infection and was present in the nuclei of infected cells. In addition, this protein is readily detectable by immunohistochemistry in the brains of AIDS patients infected with HCMV (25a). Consequently, the I2 antibody may be a valuable reagent for diagnostics. Our results reported here demonstrate that E58 is the product of the UL98 ORF. Previous sequence comparisons demonstrated that UL98 was the likely homolog for the alkaline exonuclease gene (3, 18), an essential gene which is thought to be important for viral DNA maturation (4, 25). A homology search demonstrates that HCMV UL98 amino acids 276 to 286 (VYEIKCRYKYL) are almost completely conserved among the various human herpesviruses and may represent an active site for the viral protein.

Our studies demonstrate that the E58 protein is initiated at an ATG sequence at position 142701. Inclusion of this ATG in an expression vector resulted in a protein which migrated similarly to the protein produced in infected cells. In contrast, initiation at an ATG immediately 3′ resulted in a protein that was truncated relative to E58 from infected cells. The protein-coding content of the UL98 ORF is sufficient to encode a protein of 584 amino acids, which would yield a theoretical molecular mass of 65,273 Da.

A previous study had mapped a putative late 58-kDa protein to the 1.6-kb late RNA from the pp28 gene region (16, 17). The rationale was that RNA and protein were both expressed from this region at late times and that an antibody to a 5′ ORF (ORF 2-1) was reactive to a 58-kDa protein. As the coding capacity of the 1.6-kb RNA was insufficient to encode a 58-kDa protein, the investigators proposed a translational read-through of an in-frame stop codon to a downstream ORF. Inspection of the sequence reveals two stop codons (Fig. 7) which would have to be suppressed in order for a 58-kDa protein to be expressed from the UL99 1.6-kb late mRNA. However, when the HCMV sequence became available, we predicted that the UL98 ORF was contiguous with ORF 2-1. Because the peptide to which anti-P_{2-1} reacts lies within the UL98 ORF, questions arose as to the origins of the 58-kDa late protein. Our studies demonstrate that an early 58-kDa protein is encoded by UL98 and that this same protein reacts with anti-P_{2-1}. Both antibodies used in this study (I2 and anti-P_{2-1}) detect 58-kDa proteins which have identical kinetic profiles during infection. In addition, they detect the same protein in UL98-transfected cells.

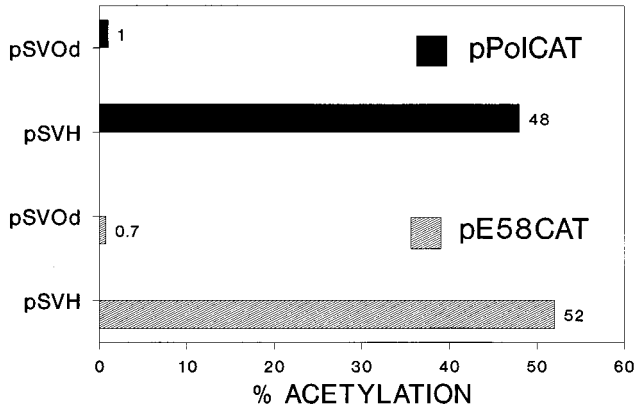

FIG. 6. Analysis of the E58 promoter region and activation of the E58 promoter by viral IE proteins. HFs were transfected with 5 μg of the indicated target promoter-CAT construct in the presence (pSVH) or absence (pSVOd) of viral IE proteins as described previously (6, 34). Cells were harvested and assayed for CAT activity as described in the text, and the data are expressed as percent acetylation. The data are from a single representative experiment that was performed five times.

While we cannot exclude the possibility of a second true late 58-kDa protein which migrates with E58, our attempts to confirm this by using an antibody to the 1.6-kb mRNA downstream ORF (ORF 2-2) were unsuccessful (data not shown).

Like other HCMV genes, UL98 is expressed as a nested RNA which appears to be 3′ coterminal with RNAs from the pp28 (UL99) gene region. Figure 7 represents a proposed summary of the structure of the UL98-UL99 gene region. In addition, there are other, larger RNAs of 5.5 and 7.0 kb that also overlap this gene region. A recent study demonstrates that several transcription units overlap within this region and are also 3′ coterminal with the pp28 gene region (UL99) (41). Our data are consistent with that study and extend those results by characterizing the UL98 gene region and its protein product. This multiple use of 3′ ends by RNAs of different kinetic classes has been demonstrated previously and appears to be a common means for HCMV to utilize genetic information (2, 6, 7, 9, 10, 19, 21, 28, 41). These data indirectly imply that HCMV genes and RNAs may be regulated mostly at the level of their promoters and 5′ sequences, since RNAs of different kinetic classes share 3′ ends.

Our results suggest the presence of a functional early promoter immediately 5′ of the UL98 gene region. At a position 87 bp upstream of the ATG lies a consensus TATA moiety. Our location for the E58 promoter is consistent with studies that mapped the 5′ end of the UL98 mRNA (41). Insertion of sequences from nt 142245 to 142666 into a reporter plasmid resulted in an E58 promoter-CAT construct responsive to IE proteins in a manner similar to that of other early promoters (2, 6, 12, 14, 20, 34). Within the promoter sequences are a number of conserved direct and inverted repeats of 8 bp or more (Fig. 2). Sequences of this type have been shown previously to be important for activation of early and late promoters by viral *trans*-acting proteins (2, 6, 7, 12, 31). In addition, consensus sites for ATF/CREB (−82) and E2F (−127) are located within the promoter (41). These transcription factors have been implicated in regulating HCMV early genes (11a, 39, 41). Studies to identify the sequences important for E58 promoter activation are in progress.

Despite the large amount of information gained by sequencing the HCMV genome, the majority of viral genes have not been characterized, and it is unclear how they are regulated or how they influence the virus life cycle. In particular, only seven early genes have been previously studied in any detail, and they represent a fraction of the genes expressed at early times after infection. Early genes play a vital role in the preparation of the infected cell for virus DNA replication (26), and it is generally thought that cells which progress successfully through DNA replication go on to produce infectious progeny. Our studies reported here provide important new information on the structure and kinetics of expression of a previously undescribed HCMV early gene. Studies of this type provide the basis for

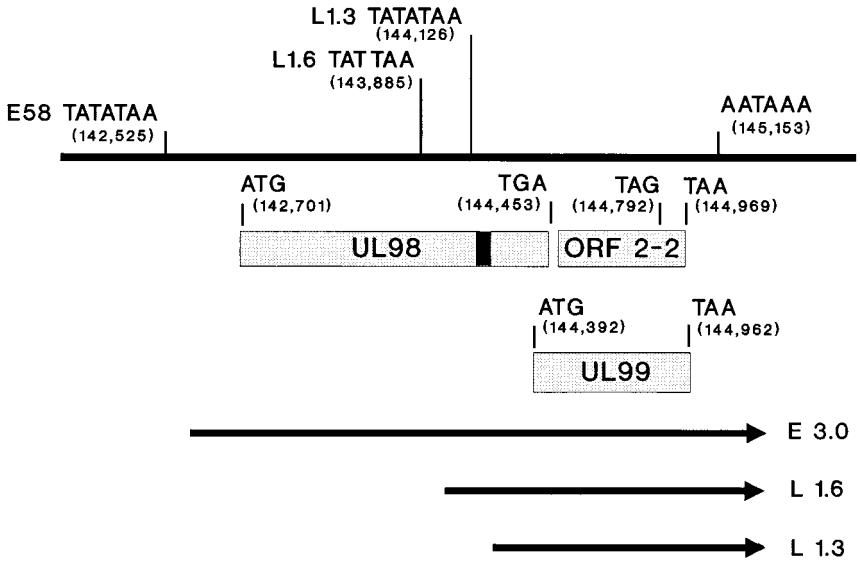

FIG. 7. Proposed structure of the UL98-UL99 gene region. The top line represents the HCMV genome and the relative positions of the E58, L1.6, and L1.3 TATA sequences. Also shown are the polyadenylation signal (AATAAA) and relevant restriction enzyme sites. Sequence designations are as described elsewhere (3). ORFs (stippled boxes) and the respective translation start and stop codons are shown below. The relative positions of the transcripts originating from this region are indicated at the bottom of the figure (arrows). The black box in the UL98 coding region corresponds to the peptide to which anti-$P_{2\text{-}1}$ was prepared.

assessing the regulatory events that influence the expression of viral early genes.

ACKNOWLEDGMENTS

This study was supported by Public Health Service grants AI23313 (R.M.S.), AI21640 (J.A.N.), and CA26659 (G.L.W.) from the National Institutes of Health. Bao-Ling Adam is the recipient of a National Research Service Award (1 F32 CA62908).

We thank Timothy Bos, Ann Campbell, and Julie Kerry for comments and critical reviews of the manuscript. We thank Mark Stinski for the gift of the *Xba*I clones of HCMV and Steve St. Jeor for providing antibodies.

REFERENCES

1. **Alford, C. A., and W. J. Britt.** 1990. Cytomegalovirus, p. 1981–2010. *In* B. N. Fields and D. M. Knipe (ed.), Virology, 2nd ed. Raven Press, New York.
2. **Chang, C.-P., C. L. Malone, and M. F. Stinski.** 1989. A human cytomegalovirus early gene has three inducible promoters that are regulated differentially at various times after infection. J. Virol. **63:**281–290.
3. **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.
4. **Chou, J., and B. Roizman.** 1989. Characterization of DNA sequence-common and sequence-specific proteins binding to *cis*-acting sites for cleavage of the terminal *a* sequence of the herpes simplex virus 1 genome. J. Virol. **63:**1059–1068.
5. **DeMarchi, J. M., C. A. Schmidt, and A. S. Kaplan.** 1980. Patterns of transcription of human cytomegalovirus in permissively infected cells. J. Virol. **35:**277–286.
6. **Depto, A. S., and R. M. Stenberg.** 1989. Regulated expression of the human cytomegalovirus pp65 gene: octamer sequence in the promoter is required for activation by viral gene products. J. Virol. **63:**1232–1238.
7. **Depto, A. S., and R. M. Stenberg.** 1992. Functional analysis of the true late human cytomegalovirus pp28 upstream promoter: *cis*-acting elements and viral *trans*-acting proteins necessary for promoter activation. J. Virol. **66:** 3241–3246.
8. **Gibson, W.** 1993. Molecular biology of human cytomegalovirus, p. 303–329. *In* E.-S. Huang (ed.), Molecular aspects of human cytomegalovirus diseases. Springer-Verlag, Heidelberg, Germany.
9. **Guo, Y.-W., and E.-S. Huang.** 1993. Characterization of a structurally tricistronic gene of human cytomegalovirus composed of US18, US19, and US20. J. Virol. **67:**2043–2054.
10. **He, Y. S., L. Xu, and E.-S. Huang.** 1992. Characterization of human cytomegalovirus UL84 early gene and identification of its putative protein product. J. Virol. **66:**1098–1108.
11. **Hutchinson, N. I., R. T. Sondermeyer, and M. J. Tocci.** 1986. Organization and expression of the major genes from the long inverted repeat of the human cytomegalovirus genome. Virology **155:**160–171.

11a. **Kerry, J. A., et al.** Unpublished data.

12. **Kerry, J. A., M. A. Priddy, and R. M. Stenberg.** 1994. Identification of sequence elements in the human cytomegalovirus DNA polymerase gene promoter required for activation by viral gene products. J. Virol. **68:**4167–4176.
13. **Klucher, K. M., D. K. Rabert, and D. H. Spector.** 1989. Sequences in the human cytomegalovirus 2.7-kilobase RNA promoter which mediate its regulation as an early gene. J. Virol. **63:**5334–5343.
14. **Klucher, K. M., M. Sommer, J. T. Kadonaga, and D. H. Spector.** 1993. In vivo and in vitro analysis of transcriptional activation mediated by the human cytomegalovirus major immediate-early proteins. Mol. Cell. Biol. **13:**1238–1250.
15. **Klucher, K. M., and D. H. Spector.** 1990. The human cytomegalovirus 2.7-kilobase RNA promoter contains a functional binding site for the adenovirus major late transcription factor. J. Virol. **64:**4189–4198.
16. **Lahijani, R. S., E. W. Otteson, J. D. Adlish, and S. C. St. Jeor.** 1991. Characterization of a human cytomegalovirus 1.6-kilobase late mRNA and identification of its putative protein product. J. Virol. **65:**373–381.
17. **Lahijani, R. S., E. W. Otteson, and S. St. Jeor.** 1992. A possible role for nonsense suppression in the synthesis of a human cytomegalovirus 58-kd virion protein. Virology **186:**309–312.
18. **Lawrence, G. L., M. Chee, M. A. Craxton, U. A. Gompels, R. W. Honess, and B. G. Barrell.** 1990. Human herpesvirus 6 is closely related to human cytomegalovirus. J. Virol. **64:**287–299.
19. **Leatham, M. P., P. R. Witte, and M. F. Stinski.** 1991. Alternate promoter selection within a human cytomegalovirus immediate-early and early transcription unit (UL119-115) defines true late transcripts containing open reading frames for putative viral glycoproteins. J. Virol. **65:**6144–6154.
20. **Malone, C. L., D. H. Vesole, and M. F. Stinski.** 1990. Transactivation of a human cytomegalovirus early promoter by gene products from the immediate-early gene IE2 and augmentation by IE1: mutational analysis of the viral proteins. J. Virol. **64:**1498–1506.
21. **Martinez, J., R. S. Lahijani, and S. C. St. Jeor.** 1989. Analysis of a region of the human cytomegalovirus (AD169) genome coding for a 25-kilodalton virion protein. J. Virol. **63:**233–241.
22. **McDonough, S. H., and D. H. Spector.** 1983. Transcription in human fibroblasts permissively infected by human cytomegalovirus strain AD169. Virology **125:**31–46.
23. **McDonough, S. H., S. I. Staprans, and D. H. Spector.** 1985. Analysis of the major transcripts encoded by the long repeat of human cytomegalovirus strain AD169. J. Virol. **53:**711–718.
24. **Meyer, H., A. T. Bankier, M. P. Landini, C. M. Brown, B. G. Barrell, B. Ruger, and M. Mach.** 1988. Identification and procaryotic expression of the gene coding for the highly immunogenic 28-kilodalton structural phosphoprotein (pp28) of human cytomegalovirus. J. Virol. **62:**2243–2250.
25. **Moss, H.** 1986. The herpes simplex virus type 2 alkaline DNAse activity is essential for replication and growth. J. Gen. Virol. **67:**1173–1178.

25a. **Nelson, J. A.** Unpublished data.

26. **Pari, G. S., and D. G. Anders.** 1993. Eleven loci encoding *trans*-acting factors are required for transient complementation of human cytomegalovirus *ori*-Lyt-dependent DNA replication. J. Virol. **67:**6979–6988.
27. **Rice, G. P. A., R. D. Schrier, and M. B. A. Oldstone.** 1984. Cytomegalovirus infects human lymphocytes and monocytes: virus expression is restricted to immediate-early gene products. Proc. Natl. Acad. Sci. USA **81:**6134–6138.
28. **Ruger, B., S. Klages, B. Walla, J. Albrecht, B. Fleckenstein, P. Tomlinson, and B. Barrell.** 1987. Primary structure and transcription of the genes coding for the two virion phosphoproteins pp65 and pp71 of human cytomegalovirus. J. Virol. **61:**446–453.
29. **Spector, D. H., K. M. Klucher, D. K. Rabert, and D. A. Wright.** 1990. Human cytomegalovirus early gene expression. Curr. Top. Microbiol. Immunol. **154:** 21–45.
30. **Staprans, S. I., D. K. Rabert, and D. H. Spector.** 1988. Identification of sequence requirements and *trans*-acting functions necessary for regulated expression of a human cytomegalovirus early gene. J. Virol. **62:**3463–3473.
31. **Staprans, S. I., and D. H. Spector.** 1986. 2.2-kilobase class of early transcripts encoded by cell-related sequences in human cytomegalovirus strain AD169. J. Virol. **57:**591–602.
32. **Stenberg, R. M.** 1993. Immediate-early genes of human cytomegalovirus: organization and function, p. 330–359. *In* E.-S. Huang (ed.), Molecular aspects of human cytomegalovirus diseases. Springer-Verlag, Heidelberg, Germany.
33. **Stenberg, R. M., A. S. Depto, J. Fortney, and J. A. Nelson.** 1989. Regulated expression of early and late RNAs and proteins from the human cytomegalovirus immediate-early gene region. J. Virol. **63:**2699–2708.
34. **Stenberg, R. M., J. Fortney, S. W. Barlow, B. P. Magrane, J. A. Nelson, and P. Ghazal.** 1990. Promoter-specific *trans* activation and repression by human cytomegalovirus immediate-early proteins involves common and unique protein domains. J. Virol. **64:**1556–1565.
35. **Stenberg, R. M., and M. F. Stinski.** 1985. Autoregulation of the human cytomegalovirus major immediate-early gene. J. Virol. **56:**676–682.
36. **Stenberg, R. M., D. R. Thomsen, and M. F. Stinski.** 1984. Structural analysis of the major immediate-early gene of human cytomegalovirus. J. Virol. **49:**190–199.
37. **Stenberg, R. M., P. R. Witte, and M. F. Stinski.** 1985. Multiple spliced and unspliced transcripts from human cytomegalovirus immediate-early region 2 and evidence for a common initiation site within immediate-early region 1. J. Virol. **56:**665–675.
38. **Wade, E. J., K. M. Klucher, and D. H. Spector.** 1992. An AP-1 binding site is the predominant *cis*-acting regulatory element in the 1.2-kilobase early RNA promoter of human cytomegalovirus. J. Virol. **66:**2407–2417.
39. **Wade, M., T. F. Kowalik, M. Mudryj, E.-S. Huang, and J. C. Azizkhan.** 1992. E2F mediates dihydrofolate reductase promoter activation and multiprotein complex formation in human cytomegalovirus infection. Mol. Cell. Biol. **12:**4364–4374.
40. **Wathen, M. W., and M. F. Stinski.** 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.
41. **Wing, B. A., and E.-S. Huang.** 1995. Analysis and mapping of a family of 3′ coterminal transcripts containing coding sequences for human cytomegalovirus open reading frames UL93 through UL99. J. Virol. **69:**1521–1531.