JOURNAL OF VIROLOGY, Mar. 1995, p. 1521–1531 Vol. 69, No. 3
0022-538X/95/$04.00+0

Analysis and Mapping of a Family of 3′-Coterminal Transcripts Containing Coding Sequences for Human Cytomegalovirus Open Reading Frames UL93 through UL99

BRET A. WING[1,2]* AND ENG-SHANG HUANG[1,2,3,4]

Lineberger Comprehensive Cancer Center,[1] Department of Microbiology and Immunology,[2] Department of Medicine,[3] and Curriculum of Genetics and Molecular Biology,[4] University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Received 24 August 1994/Accepted 21 November 1994

Human cytomegalovirus (HCMV) open reading frames (ORFs) UL93 through UL99 are contained within a region of viral genome that is well conserved in all herpesviruses. Previous reports detailing the expression of ORF UL99 (also referred to as the 28-kDa virion phosphoprotein or pp28) indicated that the pattern of transcription proximal to pp28 is extremely complex and involves a number of large overlapping transcripts, none of which have been characterized. We have used an RNA-mapping approach consisting of Northern (RNA) hybridization, RNase protection, and primer extensions to determine the coding capacity of several large-molecular-weight transcripts which overlap the 1.3- and 1.6-kb UL99-specific transcripts. Our results suggest that six differentially regulated transcripts with sizes of 2.6, 4.7, 5.6, 7.3, 9.1, and 10.5 kb, and derived from the same strand of the viral genome overlap, are 3′-coterminal with the smaller UL99-specific transcripts. On the basis of 5′-end mapping via primer extension and RNase protection, we have determined that the 2.6- to 10.5-kb messages initiate upstream of each of the potential ORFs in this region, UL98, UL97, UL96, UL95, UL94, and UL93. By using cycloheximide and ganciclovir [9-(1,3-dihydroxy-2-propoxymethyl)guanine] to block de novo viral protein synthesis and viral DNA replication, respectively, we have determined that the 2.6-, 4.7-, 5.6-, and 7.3-kb messages have characteristics of early or early-late transcripts, whereas the 9.1- and 10.5-kb messages appear to be true late transcripts. The evolutionary conservation of ORFs UL93 through UL99 and their transcriptional regulation in other herpesviruses are discussed.

Human cytomegalovirus (HCMV) is a large, double-stranded DNA virus with the largest genome (240 kb) of the known human herpesviruses (14). It is widely disseminated throughout the population and is associated with a variety of clinical manifestations, including encephalitis, retinitis, enteritis, hepatitis, pneumonitis, and immune system dysregulation, which result from both primary and persistent viral infection (reviewed in references 2 and 83). Additionally, HCMV is thought to play a major role in the acquisition of symptomatic disease by immunocompromised hosts, including neonates, patients with AIDS, and organ transplant recipients (reviewed in reference 2). The wide range of disease syndromes associated with HCMV infection is thought to be due at least in part to the complex interactions between virus and the host cell which are consequential to the establishment and maintenance of persistent viral infection. As an example, HCMV has been shown to encode proteins homologous to several regulators of cellular cytokine function, including the T-cell receptor γ subunit (14) and the major histocompatibility class I antigen (HLA) (8), which are thought to play a direct role in the activation of an abnormal and protracted immune response seen in some HCMV patients.

Infection of permissive cells in culture by HCMV leads to a productive viral infection resulting from a coordinated cascade of viral gene expression that is similar, albeit protracted, to that exhibited by herpes simplex virus, the archetypal human herpesvirus (reviewed in reference 59). Three broad classes of gene expression, referred to as immediate-early, early, and late, can be observed. Immediate-early genes are functionally defined as genes which are transcriptionally activated in the absence of de novo protein synthesis. The protein products of immediate-early genes are often regulators of subsequent viral gene expression; examples include the IE72 and IE86 proteins (35, 67–70, 77, 78), as well as TRS1 (66), US3 (16, 80), and UL36-38 (17, 73, 81). Early genes, on the other hand, are defined as those genes that are transcriptionally activated only following prior viral gene expression. The protein products of viral early genes are often directly involved in replication of the viral genome and as such are expressed prior to or concomitant with viral DNA replication. The principal example of an HCMV early gene is the viral DNA polymerase gene, whose expression is transcriptionally regulated by the viral IE72 and IE86 proteins (67); other early genes have also been documented (3, 8, 12, 23, 24, 34, 41, 49, 54, 60, 65, 75, 77, 78, 80–82). Finally, the late genes are defined as those genes that are transcriptionally activated only after the onset of viral DNA replication; the proteins encoded by late genes are generally structural constituents of mature virus progeny or are enzymatically involved in the maturation process itself (19, 20, 23, 33, 40, 45–49, 52, 53, 55, 75, 77). One example of a known HCMV late gene is pp28 (UL99) (25, 52, 53, 55, 56).

HCMV gene expression also appears to be regulated at the posttranscriptional level, as demonstrated in studies of the *Eco*RI R and d fragments (82) and the ICP36 family of proteins (30, 31). In these studies, two families of proteins encoded on overlapping and differentially processed transcripts appeared with different kinetics during infection, despite being transcribed with similar kinetics. This has led to the specula-

* Corresponding author. Mailing address: 117 Lineberger Comprehensive Cancer Center, Campus Box 7295, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7295. Phone: (919) 966-4323. Fax: (919) 966-4303. Electronic mail address: wing@med.unc.edu.

tion that posttranscriptional controls are common regulatory features of HCMV gene expression. The ultimate result of this complex set of viral gene expression is the production of infectious virus progeny, which are secreted from the cell fully competent to infect other cells and reiterate the entire process.

As summarized above, the regulation of HCMV gene expression appears to be controlled through a complex set of transcriptional and posttranscriptional mechanisms. However, many of the controls, including both important *cis*-regulatory sequences and *trans*-acting proteins (viral and cellular), as well as the organization of HCMV transcription units, are at present poorly defined. In this report, we describe the existence of a large family of structurally polycistronic transcripts having the capacity to encode HCMV open reading frames (ORFs) UL99, UL98, UL97, UL96, UL95, UL94, and UL93. All of these transcripts appear to utilize a common polyadenylation site downstream of ORF UL99, and yet they are not coordinately expressed during the HCMV replication cycle. The results of this study demonstrate an extreme example of the organization of HCMV ORFs into structurally polycistronic, overlapping transcription units and provide evidence for a family of nested 3′-coterminal transcripts that are not coordinately regulated. The implications of these findings in terms of the organization of HCMV transcription units, the potential functions of the UL93 through UL99 gene products, and the evolutionary conservation of this phenomena in other human herpesviruses are discussed.

MATERIALS AND METHODS

Cells and virus. Human embryonic lung cells (HELs) were maintained in Dulbecco modified Eagle medium supplemented with 10% fetal calf serum, tetracycline, penicillin, and streptomycin. Infection of HELs with HCMV (Towne strain) was carried out in Dulbecco modified Eagle medium supplemented with 6% heat-inactivated fetal calf serum and the antibiotics listed above as described previously (37). For drug block experiments, the medium described above was supplemented with either 10 μM [9-(1,3-dihydroxy-2-propoxymethyl) guanine; ganciclovir] or 50 μg of cycloheximide per ml.

DNA cloning and plasmids. HCMV genomic clones containing sequences for the UL93 through UL99 region were subcloned from plasmids pHD1 and pHD2 (22). Plasmid p7zXS was generated by digesting pHD2 with *Xho*I and *Sal*I; the 2.3-kb *Sal*I-*Xho*I fragment was gel purified, eluted from agarose by using Geneclean (Midwest Scientific), and cloned into *Sal*I-cut pGEM-7zf(+) (Promega). Plasmid p7z97P was generated by cloning a 1.0-kb *Eco*RI-*Bam*HI fragment from pHD2 into *Eco*RI-*Bam*HI-cut pGEM-7zf(+). Plasmid p97Sal-2 was subsequently generated from p7z97P by digestion of this plasmid with *Sal*I and subcloning the 324-bp *Sal*I fragment into *Sal*I-digested pBS(+). Plasmid p97ES-1 was also generated from p7z97P by digestion with *Eco*RI plus SalI and subsequent cloning of the 529-bp fragment into *Eco*RI-SalI-cut pBS(+).

For subcloning of plasmid pBSUL95P, the 4.0-kb *Eco*RI S fragment was cloned from HCMV (Towne) genomic DNA by digestion with *Eco*RI and subsequent ligation into *Eco*RI-cut pBS(+). White colonies were screened by colony hybridization with [α-^{32}P]dATP (10 mCi/ml; Amersham)-labeled, nick-translated UL94-specific probe generated by PCR amplification from HMCV genomic DNA (81b). The resultant plasmid (pBSES-1) was digested with *Eco*RI plus *Bam*HI, and the 1.0-kb fragment was subcloned into *Bam*HI-*Eco*RI-cut pBS(+) to generate plasmid pBSUL95P.

Plasmid p93S1-10 was generated by PCR amplification from plasmid pHD1 (22), using the primers 5′ACGTACGTGAATTCCCAGAACGCGTTTACCGA GGAGATCCAG3′ and 5′ACGTACGTCTGGAGCGCCAGATCCGAATAC AGGTGCGTTTCC3′. The amplified product was digested with *Eco*RI and *Pst*I and cloned into pBS(+) cut with *Eco*RI and *Pst*I.

Oligonucleotide probes. Single-stranded oligonucleotides were synthesized at the Lineberger Comprehensive Cancer Center Oligonucleotide Facility. Partially purified oligonucleotides were purified on polyacrylamide-urea gels as described previously (5) and visualized by UV shadowing. Bands were excised from the gel, crushed with a spatula, and eluted overnight at 37°C in elution buffer (0.5 M ammonium acetate, 1 mM EDTA). The molar concentration of oligonucleotides was calculated by determining the optical density at 260 nm and molar extinction coefficient for each oligonucleotide. Oligonucleotides were end labeled for hybridization experiments, using 8 U of polynucleotide kinase (Boehringer Mannheim), 5 μl of [γ-^{32}P]ATP (50 μCi/ml; NEN/Dupont), and 10 pmol of oligonucleotide in a 20-μl reaction mixture for 1 h at 37°C. Labeled oligonucleotides were subsequently purified from reaction mixtures by G-25 spin column chromatography (Boehringer Mannheim). Oligonucleotides were designed to be complementary to the coding strand of the UL93 through UL99 region of the HCMV genome on the basis of the complete genomic sequence for HCMV (AD169) determined by Chee et al. (14). All computer sequence analysis was carried out by using the Genetics Computer Group program (32).

TABLE 1. Sequences of oligonucleotides used in this study

Oligonucleotide	Sequence
UL99-1	GGA CAG GTA AAT TGA GGT GTC CTC TTC GTC
PolyA-1	CTG TTC TTC CAC CAG CTA CTT TAC GTG
PolyA-2	GCG AAG AAG AGG ACG ACG ACG ATG ATG AAG
PolyA-3	CTT CAT CAT CGT CGT CGT CCT CTT CTT CGC
UL98-1	AAG CCC TCG TGT TTG ATA CCG AAC TGG ATG
UL98-2	TCG TAG TCC AAA CTC GAG ACG CCC CAC ATG
UL97-1	GAT GAT CTG GTC GTT GGA ACA GGT GCA ATG
UL97-2C	ATA GTG GCG ACA GTC TTC CCT GTT CCT AGA G
UL96-1	GTG TCC TTC AGC TCG GTT AAA GAG TCC AGC
UL96-2	GAT GCT GCT GCT GCT CAA GGT CGA CGC GC
UL95-1	ATT GCG TAA CGC ATA ATC ACG CAC ACA AAG
UL95-2	GAT CCG CAG TTT CAG CAG ATC AAC AAC TTC
UL94-1	AAC CTA CGG CGC CAT AGG TAC TTG GCG TCA T
UL94-3	CAC CAC GTC AGC GTA CCA AGT CTG TTC
UL93-1	CGC CAG ATC CGA ATA CAG GTG CGT TTC C
UL92-1	AAC TGG ATC TCC TCG GTA AAC GCG TTC TGG

RNA isolation and Northern (RNA) blot analysis. At each time point, polyadenylated RNA was isolated from approximately 10^7 HELs (one T150 flask) infected with HCMV (Towne) at a multiplicity of approximately 10 or mock infected, using the polyAttract mRNA isolation system (Promega) or MicroFastTract system (Invitrogen). RNA samples were standardized by using several techniques, including (i) spectrophotometric quantitation (optical density at 260 nm), (ii) ethidium bromide staining of RNA gels following size fractionation, and (iii) hybridization of Northern blots with internal standard probes.

Northern blots were prepared by size fractionating mRNA on denaturing agarose gels as described previously (5), followed by an overnight transfer to a nitrocellulose membrane (MSI). Blots were prehybridized for at least 2 h at 37°C in oligonucleotide prehybridization buffer (6× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate], 0.05% sodium pyrophosphate, 5× Denhardt's reagent, 0.1% sodium dodecyl sulfate, 100 μg of salmon sperm DNA per ml, 100 μg of yeast RNA per ml) and hybridized overnight at 37°C in oligonucleotide hybridization buffer (6× SSC, 0.05% sodium pyrophosphate, 1× Denhardt's reagent, 100 μg of yeast RNA per ml, 5×10^6 to 1×10^7 cpm of each end-labeled oligonucleotide listed in Table 1 per ml). Blots were washed two to four times for 5 min at room temperature in 6× SSC–0.05% sodium pyrophosphate and once at 65°C for 15 min in 6× SSC–0.05% sodium pyrophosphate, as needed. Blots were subsequently autoradiographed for 12 h to 3 days at −70°C, using Cronex intensifying screens. For most experiments, blots were successively reprobed with different oligonucleotides by stripping blots with 70% formamide for 30 min at 37°C and subjecting them to a 20-min wash in 0.1× SSC at room temperature. Blots were confirmed to be stripped by demonstrating a complete lack of signal upon overnight autoradiography at −70°C. Northern hybridizations shown in this report are representative of results obtained in at least three repetitions using independently isolated RNA samples.

Primer extension. Primer extension mapping of the 5′ ends of some RNAs was carried out essentially as described previously (5), using RNA purified from HCMV- or mock-infected HELs (one T150 flask per time point). The sequences for oligonucleotides used in primer extensions are shown in Table 1. Extended products were size fractionated on polyacrylamide-urea sequencing gels (5); gels were fixed for 20 min in 10% methanol–7% acetic acid, dried, and autoradiographed overnight at −70°C, using Cronex (Dupont) intensifying screens. The sizes of extended products were determined by comparing band mobilities against an M13 sequencing ladder generated by using the reagents, single-stranded DNA, and primer supplied with the Sequenase (United States Biochemical) DNA sequencing kit.

RNase protection assays. RNase protection was carried out essentially as described previously (5). Single-stranded RNA probes were generated from linearized plasmids with T7, SP6, and T3 RNA polymerase sites containing HCMV (Towne) genomic subclones (described above). Riboprobes were generated by using reagents supplied with the Riboprobe kit (Promega) and [α-^{32}P]CTP (10 mCi/ml; Amersham) as suggested by the manufacturer. Labeled probes were separated from unincorporated nucleotides by G-50 spin column chromatography (Boehringer Mannheim). Labeled probe (10^5 cpm) was hybridized overnight at 45°C with 10 μg of whole cell RNA purified from mock-infected or 72-h-postinfection (hpi) infected cells as described above. In some experiments, 10 μg of yeast tRNA (Boehringer Mannheim) was also used as a negative control. RNA hybrids were digested with 40 μg of RNase A per ml plus 2 μg of RNase T_1 per ml for 60 min at 30°C and then subjected to a 15-min digestion

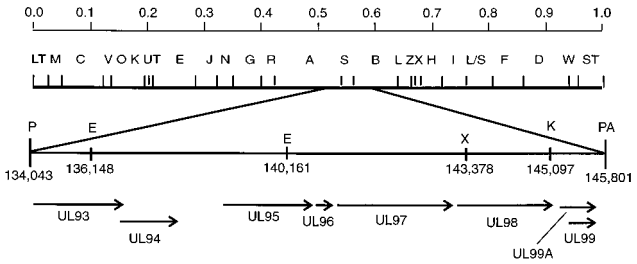

FIG. 1. Schematic representation of the HCMV (Towne) genome in standard configuration showing the locations of ORFs UL93 through UL99. (A) Unit-length map; (B) *Eco*RI map as described elsewhere (25); (C) subfragment of the HCMV genome containing ORFs UL93 through UL99 encompassing sequences from nt 134043 to 145801 (14), showing locations of restriction sites (P, *Pst*I; E, *Eco*RI; X, XhoI; K, *Kpn*I) and a polyadenylation signal (PA; AATAAA).

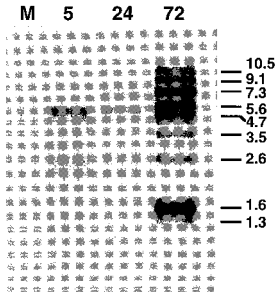

FIG. 2. Northern blot analysis of polyadenylated RNA from mock-infected (lane M) and infected HELs at the times indicated (hours postinfection), probed with end-labeled oligonucleotide UL99-1, specific for ORF UL99. Approximate sizes are shown in kilobases. Photoimages of autoradiographs shown in this report were captured by using NIH Gel Capture 2.0 for Macintosh (National Institutes of Health). Figures were constructed by using MacDraw 2.0 (Claris) for Macintosh and CorelDraw! 3.0 (Corel Corp.) for IBM-compatible personal computers.

with proteinase K, phenol-chloroform extraction, and ethanol precipitation. Products were separated on denaturing polyacrylamide-urea gels; the gels were subsequently dried and subjected to autoradiography. Sizes of protected fragments were estimated in comparison with a radiolabeled single-stranded DNA ladder (Bethesda Research Laboratories). In our hands, RNA products had mobilities approximately 5 to 15% higher than those of the single-stranded DNA markers.

RESULTS

Analysis of RNA transcribed from the UL93 through UL99 region. HCMV ORFs UL93 through UL99 are located within the long unique region of the viral genome spanning bp 134043 to 145801 (Fig. 1). Previous studies indicated that HCMV ORF UL99 is transcribed on two 3′-coterminal transcripts of 1.3 and 1.6 kb which terminate at a consensus polyadenylation signal located immediately downstream of UL99 (45, 46, 52, 53, 55, 56). The smaller message encodes the 28-kDa virion phosphoprotein (UL99), while the 1.6-kb message encodes a 58-kDa frameshift variant of UL99 (UL99A). These studies and others (25) also demonstrated that a series of larger-molecular-weight transcripts hybridized with UL99-specific probes; however, the nature of these transcripts was not investigated. To determine the exact number and sizes of the larger-molecular-weight transcripts, Northern blots of polyadenylated RNA isolated from HCMV-infected HELs at immediate-early (5 hpi), early (24 hpi), and late (72 hpi) times of infection, as well as mock-infected-cell RNA, were probed with an end-labeled oligonucleotide internal and specific to UL99 coding sequences. The results of this experiment (Fig. 2) indicated that, in addition to the 1.6- and 1.3-kb transcripts, a series of seven larger transcripts having molecular sizes of approximately 10.5, 9.1, 7.3, 5.6, 4.7, 3.5, and 2.6 kb hybridized with the UL99-specific oligonucleotide. All of the transcripts were maximal at 72 hpi; in addition, the 7.3-, 5.6-, 4.7-, and 2.6-kb transcripts were detected as early as 5 hpi (also see Fig. 4 for a longer exposure of 5-hpi transcripts, which more clearly demonstrates the presence of the four transcripts). Since an oligonucleotide probe was used in this analysis, our results indicate that all of the transcripts arise from the same strand of the viral DNA as the 1.6- and 1.3-kb UL99-specific transcripts.

To determine the potential coding capacity of the 10.5- to 2.6-kB transcripts, Northern blots of mock-infected-cell and infected-cell RNA were subsequently probed with a series of oligonucleotide probes specific for predicted ORFs proximal to UL99 and encoded from the same strand of the viral genome. Within 11 kb upstream of UL99, there are six putative ORFs colinear with UL99 (UL98, UL97, UL96, UL95, UL94, and UL93), and probes for all of these ORFs were used (see Table 1 for probe sequences). The hybridization pattern seen in these experiments (Fig. 3) is indicative of a series of overlapping, 3′-coterminal transcripts containing coding sequences for ORFs UL98 through UL93. These messages appeared with kinetics identical to those of the identically sized transcripts shown in Fig. 2; therefore, only the 72-hpi time point is shown in Fig. 3. The 10.5-kb message hybridized to all probes specific for UL98 through UL93 (lanes 1 to 6) but did not hybridize to a probe specific for UL92 (lane 7), indicating that UL93 defines the upstream boundary for this series of transcripts. In addition, it would appear that the upstream boundaries for the 9.1-, 7.3-, 5.6-, 4.7-, and 2.6-kb messages are UL94, UL95, UL96, UL97, and UL98, respectively, since in each case, probes to the next most proximal upstream ORF failed to hybridize to the smallest message detected with the prior probe.

The UL96-specific probe appeared to hybridize slightly with an approximately 4.2- to 4.7-kb band (Fig. 3, lane 3); however, detailed RNA mapping by RNase protection (see below) did not detect any initiation or splice sites within ORF UL96 consistent with a message of this size. More likely, the UL96 probe cross-hybridized with contaminating rRNA in the polyadenylated RNA preparations; in support of this idea, we could also detect an approximately 4.7-kb transcript in mock-infected-cell RNA with the UL96 probe but not with probes to UL97 to UL99, which also detected a 4.7-kb band (Fig. 2 and data not shown). Several smaller transcripts were also weakly detected with probes to UL97 in some experiments (lane 2); however, we could not consistently detect these transcripts with either oligonucleotide or double-stranded DNA probes (data not shown). Thus, it is likely that the UL97 oligonucleotide cross-reacted with several small, unrelated viral or cellular transcripts. More detailed RNA mapping (presented below) supported this conclusion. The 3.5-kb message was detected only with the UL99-specific probe; in addition, this transcript was not detected with an oligonucleotide probe (polyA-1) specific for 3′ untranslated sequences found in the UL99 transcripts (lane 8). This result suggested that the 3.5-kb message is an unrelated viral or cellular transcript containing some sequence homology to UL99. Probe polyA-2 located downstream of the UL99 polyadenylation site did not detect the UL99 family of messages (lane 9); however, a probe to the opposite strand of the viral DNA (polyA-2) detected several unrelated transcripts originating from the opposite strand of the viral DNA (not shown), indicating that reaction conditions

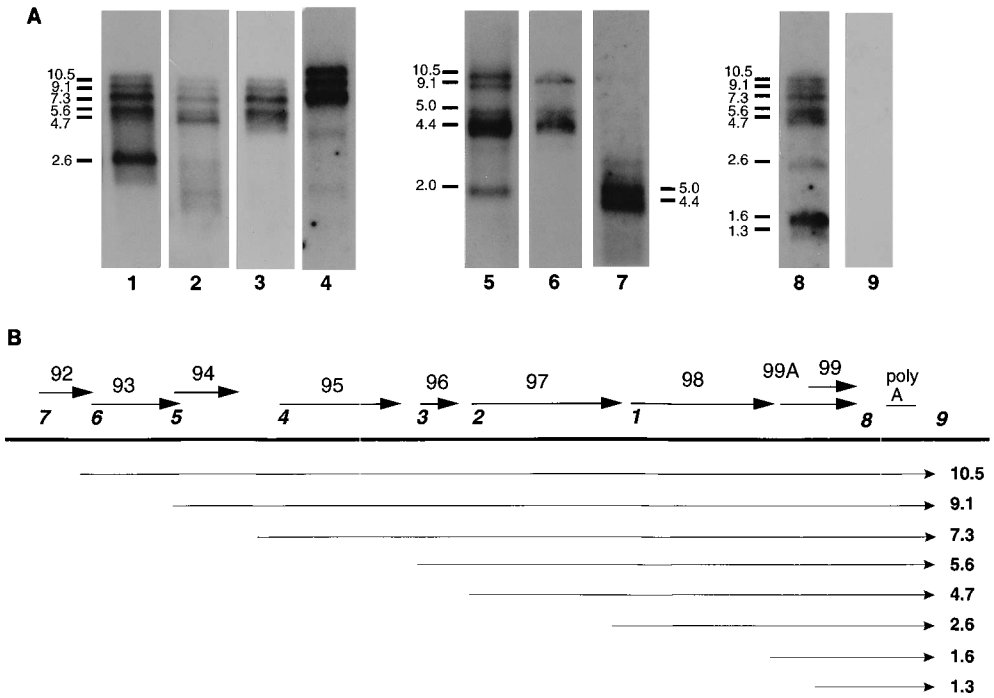

FIG. 3. (A) Composite Northern blot analysis of polyadenylated RNA isolated from 72-hpi infected fibroblasts and probed with oligonucleotide probes UL98-1 (lane 1), UL97-1 (lane 2), UL96-1 (lane 3), UL95-1 (lane 4), UL94-1 (lane 5), UL93-1 (lane 6), UL92-1 (lane 7), polyA-1 (lane 8), and polyA-2 (lane 9). (B) Schematic representation of ORFs UL93 through UL99 and the downstream polyadenylation signal; numbers indicate relative positions of probes used in panel A and proposed structures for UL99 coterminal transcripts; sizes of transcripts given at the right in kilobases.

were sufficient to detect any hybridization to polyA-2. This result suggests that the UL99-proximal polyadenylation signal defines the 3′ boundary for all of the transcripts overlapping UL99 and is consistent with the preceding Northern blots, which indicate that the larger messages initiate upstream of UL99. These results are also consistent with previous studies which mapped the 3′ termini of the 1.3- and 1.6-kb UL99 transcripts (52, 53, 55, 56).

In addition to the UL99 family of transcripts, probes to UL92, UL93, and UL94 (Fig. 3, lanes 5 and 6) detected transcripts of approximately 5.0 and 4.4 kb (probes UL92, UL93, and UL94) as well as a 2.0-kb transcript which did not conform to the colinear, 3′-coterminal arrangement described above. Preliminary characterization of these transcripts suggests that they are members of a second family of overlapping transcripts which utilize a second polyadenylation signal located immediately downstream of UL94. This polyadenylation signal, on the basis of its primary sequence, is predicted to be less efficient than the UL99 polyadenylation signal (data not shown), which probably accounts for the appearance of the 10.5- and 9.1-kb messages. Whatever the case may be, it is clear that the smaller UL93- and UL94-specific transcripts represent novel species which do not terminate 3′ coterminal with the UL99 family of transcripts. These transcripts are likely to be specific for proteins encoded from open reading frames UL91 through UL94, and experiments are in progress to confirm this hypothesis.

Assignment of the UL99 3′-coterminal transcripts to kinetic classes. To define the kinetic classes of the larger UL99-coterminal transcripts, polyadenylated RNA was isolated from HCMV-infected cells at 5 hpi in the presence of cycloheximide and at 72 hpi in the presence of ganciclovir (DHPG), as well as at both time points in the absence of drugs. The results shown in Fig. 4 were obtained by using UL98-specific probe UL98-1 (Table 1), which detects all of the novel UL99 3′-coterminal messages described above. Hybridizations with probes to reading frames UL93 through UL97 yielded similar results (data not shown). Figure 4 indicates that none of the messages are transcribed in the presence of cycloheximide at 5 hpi (lane 3), suggesting that there are no immediate-early transcripts in this family. In the presence of DHPG at 72 hpi, however, the 10.5- and 9.1-kb messages are not transcribed (lane 4), indicating that these messages are of a late classification. Levels of the 4.7-kb transcript appear to be unaltered by DHPG treatment (lane 4), implying that this transcript is of an early classification. The 7.3-, 5.6-, and 2.6-kb messages are transcribed, albeit at reduced levels, in the presence of DHPG (lane 4), indicating that these are early or early-late transcripts. Interestingly, the levels of the 5.6- and 2.6-kb messages are significantly higher at

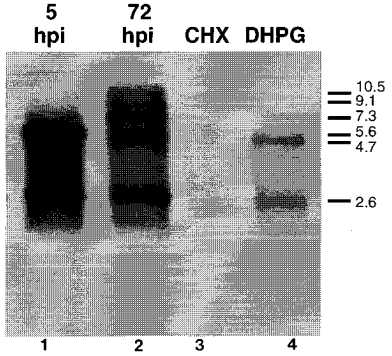

FIG. 4. Northern blot analysis of polyadenylated RNA from HCMV-infected HELs at 5 hpi (lane 1), 72 hpi (lane 2), 5 hpi with cyclohixamide (CHX; 50 μg/ml; lane 3), and 72 hpi with DHPG (10 μM; lane 4) probed with end-labeled oligonucleotide UL98-1. Sizes are indicated in kilobases.

5 hpi than at 72 hpi in the presence of DHPG (lanes 1 and 4); this result suggests that these messages may be under the control of multiple promoter elements allowing expression in more than one kinetic class. This phenomenon has been observed in the transcriptional control of other HCMV gene products (12, 34, 47, 48).

Mapping the 5′ ends of novel UL99 family transcripts by primer extension. Mapping of the 10.5- to 2.6-kb messages by Northern hybridization suggested that they were transcribed in a colinear arrangement and that each message initiated upstream of a different ORF. To confirm this observation, we mapped transcription cap sites upstream of each of ORFs UL93 through UL98 by using a combination of primer extension and RNase protection assays. Figure 5 shows the results of primer extensions using probes proximal to UL98, UL94, and UL95. In Fig. 5A, an end-labeled oligonucleotide probe overlapping the predicted translational start site for UL98 was hybridized with RNA from mock-infected and 72-hpi infected-cell RNA and extended with Moloney murine leukemia virus reverse transcriptase as outlined in Material and Methods. The results indicate that an infected-cell-specific transcript initiates 177 nucleotides (nt) upstream of the UL98 initiation codon; this site is also 25 nt downstream of a consensus TATA element, indicating that it is a favorable start site for transcription. In addition, the size of an unspliced transcript originating at this site and extending to the UL99 polyadenylation site is approximately 2,639 nt; this correlates well with the estimated size of the transcript shown in by Northern hybridization to initiate between the UL97 and UL98 probes (2.6 kb). Analysis of the genomic sequence upstream of the TATA box indicated the presence of a consensus CREB/ATF site (5′TACGTCA G3′) at −82, relative to the RNA initiation site, as well as an E2F site (5′TTTCGCCG3′) at −127 and a serum response element (5′GATGTCC3′) at position −152. All of these elements have been identified in other HCMV transcription units (24, 29, 36, 38, 43, 51, 63). CREB/ATF binding sites appear to play a role in activation of the major immediate-early enhancer mediated through the viral pp71 (UL82) gene product (51). CREB/ATF sites are also found upstream of the HCMV DNA polymerase (a viral early gene product) as well as the UL112-UL113 early gene products (41, 64). In addition, E2F sites are found upstream of several viral early transcription units (36, 43, 44, 64, 82). Since the protein product of ORF UL98 is thought to encode another enzyme involved in viral DNA replication or nucleotide metabolism (an alkaline DNase), it too is likely to be expressed at early times of infection; this possibility is supported by the appearance of the cycloheximide-sensitive 2.6-kb transcript by 5 hpi.

To map cap sites upstream of UL94, we used an oligonucleotide probe (UL94-3; Table 1) located approximately 200 nt upstream of the predicted translational start site for this ORF, since preliminary Northern hybridizations indicated that a larger 5′ untranslated region was likely to be seen in messages initiating upstream of UL94. The oligonucleotide probes used for primer extension hybridized in Northern blots with identically sized transcripts as probes internal to the UL94 reading frame, and no extended products were detected with more downstream probes (data not shown). Figure 5B indicates that a transcripts initiating 127 nt upstream of the UL94 probe could be detected after hybridization of end-labeled UL94-3 to 72-hpi infected-cell RNA and extension with reverse transcriptase; no extended products were detected with mock-infected RNA. This cap site is approximately 300 nt upstream of the predicted initiation codon for UL94. The mapped cap site was also found 30 nt downstream of a consensus TATA element (Fig. 5D). This cap site is in agreement with the Northern hybridization data in Fig. 3, which suggest that the 5′ end of the 9.1-kb message is located between UL93 and UL94. Furthermore, the size of an unspliced transcript extending from this start site to the UL99 polyadenylation site is 9,120 nt; again, these results are in agreement with the Northern hybridization data. Analysis of the genomic sequence proximal to the UL94 TATA box indicated the presence of consensus binding sites for Sp1 (5′GGGCGG3′), CREB/ATF (5′CGTCA3′), NF-κB (5′GGGAACTTT3′), and p53 (5′GAACATGTCC3′) at −58, −86, −97, and −120 relative to the RNA initiation site, indicating that the control of expression from this promoter could be quite complex. Since the 9.1-kb message appears only at late times of infection and is sensitive to treatment with DHPG, the UL94 protein is likely to be a late viral gene product. The mechanisms regulating expression of late gene products are unclear, but it is possible that cellular factors such as those mentioned above contribute in part to this regulation.

To map RNA cap sites upstream of UL95, we used an oligonucleotide which hybridized with sequences located 352 nt upstream of the predicted initiation codon for UL95 (UL95-2; Table 1). This oligonucleotide hybridized with identically sized products as internal UL95 probes in Northern blots (data not shown). Following hybridization of end-labeled UL95-2 with RNA from mock-infected and 72-hpi infected cells and extension with reverse transcriptase, we were able to detect an extended product of 121 nt specifically with infected-cell RNA (Fig. 5C). Transcripts initiating at this site and extending to the UL99 polyadenylation site would be 7,232 nt in length, which is in agreement with the 7.3-kb size of transcripts shown by Northern hybridization to initiate between UL95 and UL94 probes. Analysis of the genomic sequence proximal to this initiation site indicates the presence of a consensus TATA box at −28 with respect to the initiation site, as well as consensus CREB/ATF (5′TTGGTCA3′ and GTGACGTCG3′) and serum response element (5′GATGTCG3′) sites at −39, −166, and −124, respectively. The 7.3-kb message is detected at weak levels by 5 hpi and accumulates to higher levels by 24 hpi; thus, regulation of its expression as an early gene product could be mediated in part through the CREB/ATF sites, as has been postulated for the viral DNA polymerase and UL112-UL113 gene products (41, 64).

Analysis of UL99 family transcripts by RNase protection. RNase protection assays were used to map transcripts initiating proximal to ORFs UL96, UL97, and UL93, since high GC content in these regions yielded high background with primer extensions (data not shown); the UL98-proximal cap site was also mapped by RNase protection. Figures 6 and 7 show the results of this set of experiments. To map transcripts proximal to UL96 and UL97, we subcloned a series of overlapping HCMV genomic fragments spanning a 2-kb region inclusive of partial UL95, UL96, and UL97 coding sequences (Figure 6F). These genomic fragments were used to generate a series of riboprobes which were hybridized to RNA from mock-infected and 72-hpi infected cells, digested with RNase, and analyzed on sequencing gels as described in Materials and Methods. Figures 6A and B show the results of experiments obtained with riboprobes generated from plasmid p7z97P, which contains a 1,051-bp insert spanning sequences 513 bp upstream of the predicted UL96 initiation codon to 67 bp within ORF UL97. When digested with *Sal*I, this plasmid generated a 210-nt antisense probe initiating 67 nt within ORF UL97 and extending to 78 nt upstream of the UL97 initiation codon. Figure 6A indicates that a protected product of approximately 134 nt was obtained with *Sal*I-digested p7z97P riboprobe. The size of this product is consistent with a transcript initiation or splice site located approximately 57 to 77 nt upstream of the

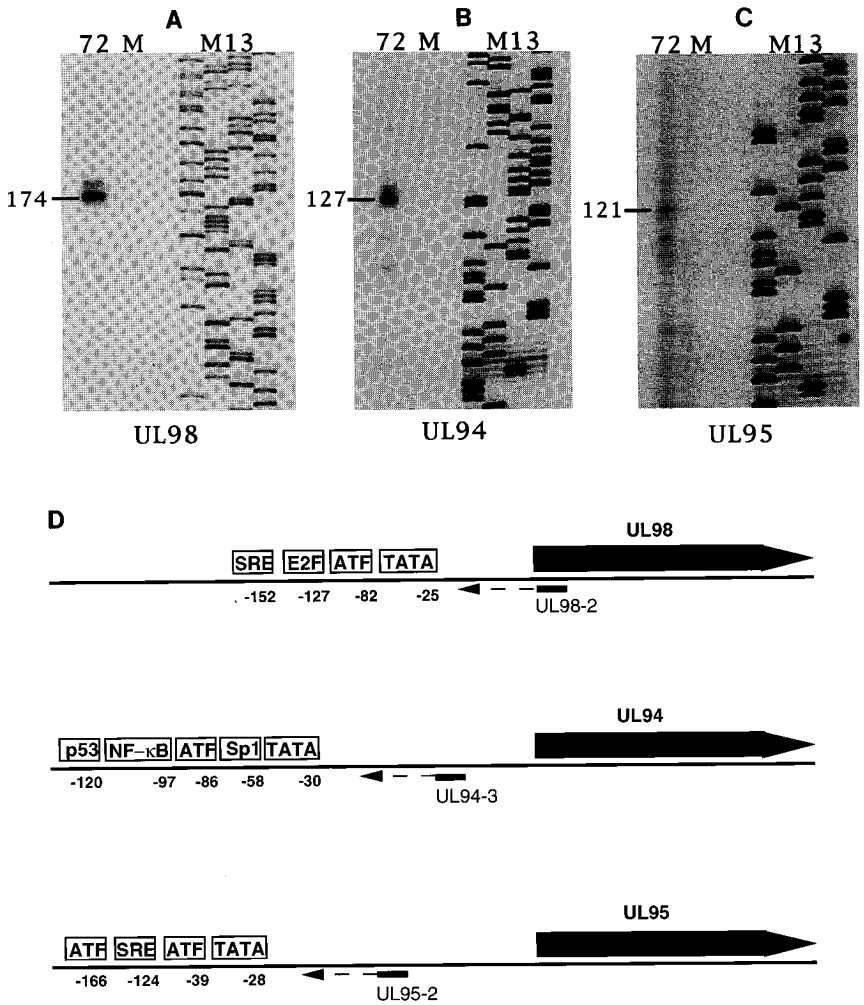

FIG. 5. Primer extension analysis of RNA initiating upstream of ORFs UL98, UL94, and UL95. (A) Primer extension results with oligonucleotide UL98-2 hybridized to whole cell RNA isolated at 72 hpi (lane 72) or mock infected (lane M) and M13 sequence ladder (A, C, G, T; lane M13); (B) primer extension with oligonucleotide UL94-3 hybridized to RNA isolated at 72 hpi (lane 72) or mock infected (lane M) and M13 sequence ladder (lane M13); (C) primer extension with end-labeled oligonucleotide UL95-2 hybridized to RNA isolated at 72 hpi (lane 72) or mock-infected (lane M) and M13 sequence ladder (lane M13). (D) Sequences of probes used. Sizes are indicated in nucleotides.

predicted UL97 initiation codon. Figure 6B indicates that a similarly sized product (154 nt) was obtained with full-length 1,051-nt p7z97P probe digested with *Eco*RI; the approximately 20-nt difference seen between the two probes is due to slight differences in gel conditions which result in sequence-dependent 5 to 15% differences in the mobilities of RNA products with respect to the single-stranded DNA size markers. In addition, the product at approximately 330 nt is likely to be due to either incomplete digestion or a nonspecific interaction and was not detected in further RNase protection analysis (described below). We also detected complete protection of p7z97P subfragment p97Sal-2 in RNase protection assays of infected-cell RNA (Fig. 6C). The small shift in mobility seen with this probe is due to the presence of approximately 50 nt of vector sequence at the ends of the probe. Analysis of the genomic sequence in the vicinity of this site indicates the presence of a consensus TATA element located at approximately −26 with respect to the initiation site, as well as a consensus CREB/ATF site (5′CGTCA3′) at −48 and an E2F site (5′TTTCGGCG3′) at −79. The size of an unspliced transcript initiating at this site and extending to the UL99 polyadenylation signal would be approximately 4,714 nt, which is in agreement with the 4.7-kb band seen in Northern hybridizations. This transcript was shown to exhibit early kinetics with respect to its expression during viral infection. Since ORF UL97 is thought to encode a ganciclovir kinase (50, 71), it is likely to be involved in nucleotide metabolism and thus would be expected to be expressed at early times of infection.

To examine transcriptional initiation proximal to the UL96 start site, probes p97ES-1 and pBSUL95P were used (Fig. 6D and E). *Bgl*II-digested p97ES-1 allowed generation of a 245-nt riboprobe which overlaps the first 12 nt of ORF UL96 and extends upstream. As seen in Fig. 6D, this probe was fully protected from RNase digestion when hybridized to infected-cell RNA, but not mock-infected-cell RNA, except for vector sequences located at the 5′ end of the probe. This result is consistent with those obtained with full-length p7z97P probe (see above), which entirely overlaps the p97ES-1 *Bgl*II probe. With *Sal*I-digested pBSUL95P, a 445-nt probe extending into UL95 coding sequences was generated. This probe produced three major bands in RNase protection of infected-cell RNA of approximately 140, 145, and 280 nt. Analysis of the genomic

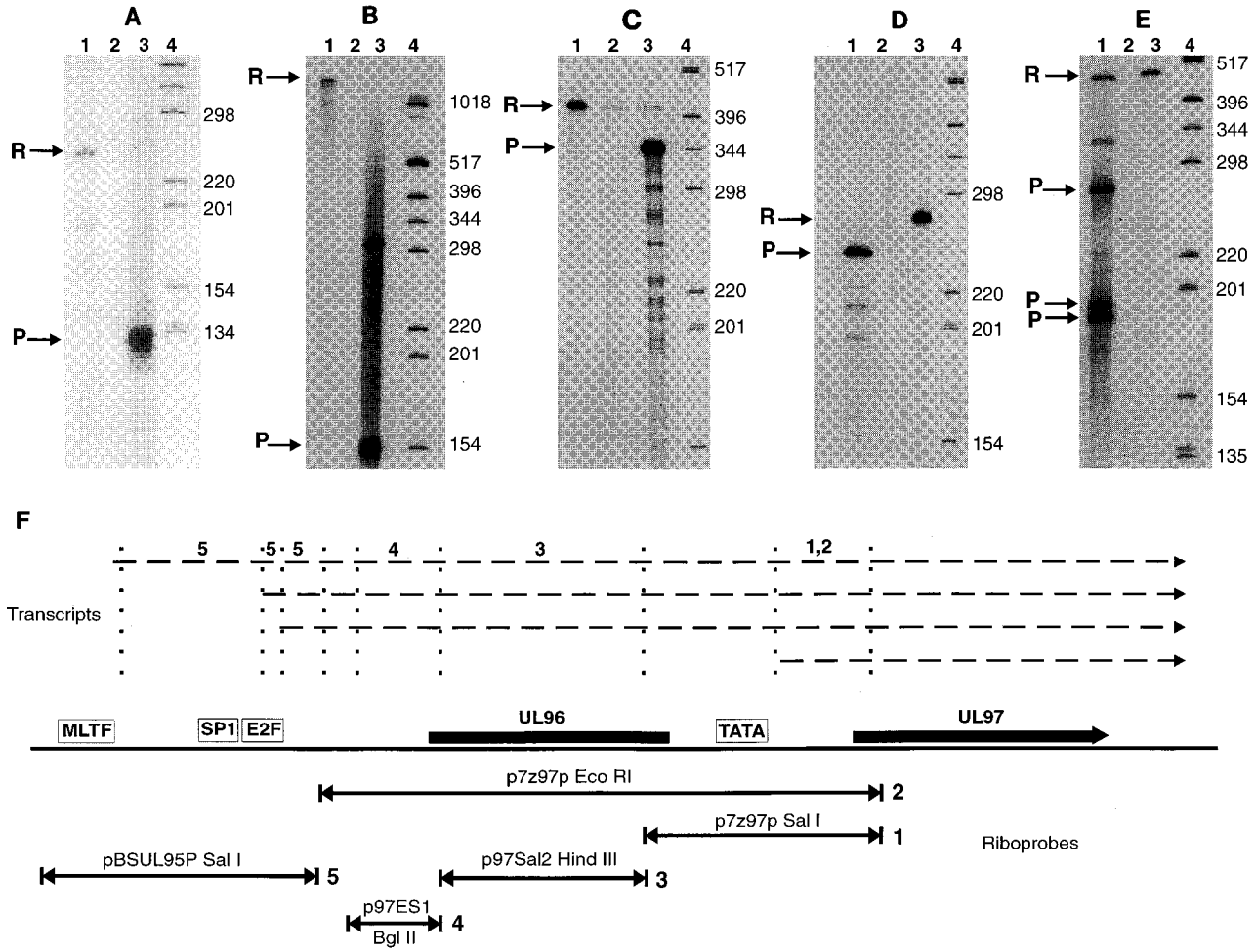

FIG. 6. RNase protection analysis of RNA initiating proximal to ORFs UL96 and UL97. Riboprobes used were *Sal*I-digested p7z97p (A), *Eco*RI-digested p7z97p (B), *Hin*dIII-digested p97Sal2 (C), *Bgl*II-digested p97ES1 (D), and *Sal*I-digested pBSUL95P (E). (A to D) Lanes: 1, unreacted riboprobe (R); 2, RNase protection using 72-hpi infected-cell RNA (protected products designated P); 3, RNase protection using mock-infected-cell RNA; 4, end-labeled single-stranded DNA ladder markers (sizes shown in nucleotides). (E) Lanes: 1, RNase protection using 72-hpi infected-cell RNA; 2, RNase protection with mock-infected-cell RNA; 3, unreacted free probe; lane 4, DNA ladder (sizes shown in nucleotides). (F) Schematic diagram of the UL96 to UL97 genomic region showing the relative position of each probe used as well as relative locations of putative promoter elements (MLTF, Sp1, E2F, and TATA); also shown are the transcripts mapped by this analysis showing regions of each RNA mapped by the probes in panels A to E.

sequence in this region showed no obvious splice sites proximal to these sites and no consensus TATA boxes. However, consensus sites for the cellular transcription factors E2F (5′TTT TGCGCG3′) and Sp1 (5′CCGCGC3′) spaced 10 bases apart were detected within 18 bases of the sites mapped by the 140- and 145-nt protected fragments. Additionally, a consensus adenovirus major late transcription factor (MLTF) binding site (5′TCACGTGG3′) was located approximately 10 bases upstream of the site mapped by the 280-nt fragment. Both E2F and Sp1 have been demonstrated to direct initiation of transcription from cellular TATA-less promoters (10, 11, 61, 72), and an MLTF binding site is thought to play a crucial role in activation of the HCMV $\beta_{2.7}$ transcript (44). Thus, it is likely that these factors play some role in activation of transcription from the UL96 promoter and that the 5.6-kb transcript observed in Northern hybridizations represents a class of transcripts initiating at three distinct sites. It is also worth noting that the 5.6-kb transcripts accumulate to high levels by 5 hpi, consistent with their classification as early messages.

To map potential cap sites upstream of UL93, a 562-nt riboprobe generated from *Eco*RI-cut p93S1-10 was used (Fig. 7C). This probe specifically protected a fragment of approximately 160 nt when hybridized to RNA from 72-hpi infected cells (Fig. 7A). The size of the protected fragment is consistent with a cap site 80 nt upstream of the predicted UL93 translational start site. Analysis of the genomic sequence in the vicinity of this site indicates that a TATA-like (5′TATTTA3′) element is located 31 nt upstream of the mapped cap site. In addition, two potential CREB/ATF sites are located at −127 (5′TACCTCAT3′) and −254 (5′CGTCA3′). In all, five of the six putative promoters mapped in this study contain putative ATF elements, the exception being the apparently TATA-less UL96 promoter. Thus, it is possible that CREB binding represents an important control mechanism regulating HCMV gene expression.

To confirm the cap site mapped by primer extension upstream of UL98, RNase protection was performed with a 502-nt riboprobe generated from *Fsp*I-cut p7zXS (Fig. 7C). This probe overlaps the putative UL98 start site by 17 bases and extends upstream relative to this position. When this probe was hybridized to 72-hpi infected-cell RNA, a protected fragment of approximately 200 nt was detected (Fig. 7B). No

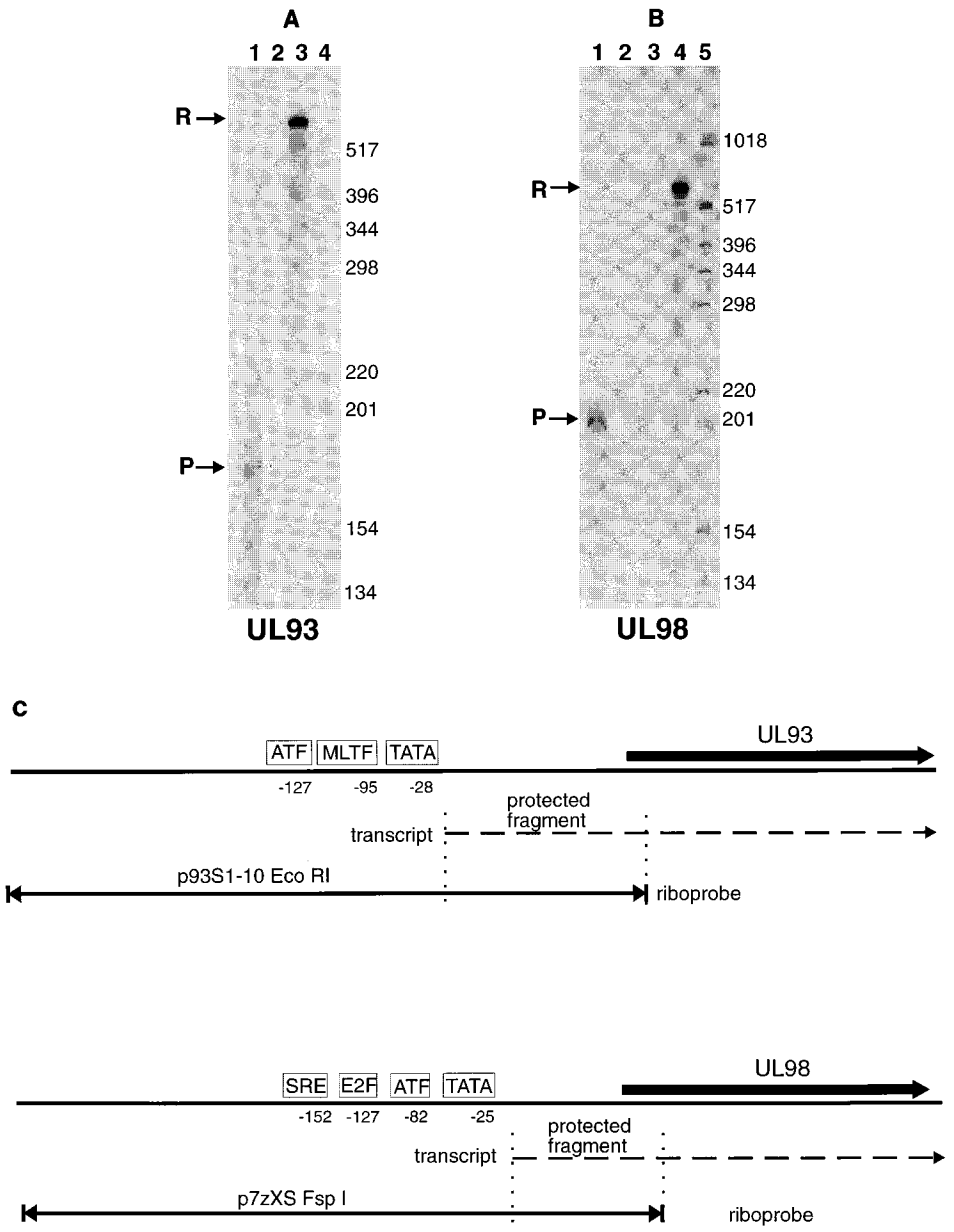

FIG. 7. RNase protection analysis of RNA initiating proximal to ORFs UL93 and UL98 mapped with riboprobes generated from *Eco*RI-digested p93S1-10 (A) and *Fsp*I-digested p7zXS (B). Lanes: 1, 72-hpi infected-cell RNA; 2, mock-infected-cell RNA; 3, unreacted probe; lane 4, end-labeled single-stranded DNA ladder (sizes shown in nucleotides). (C) Schematic representation of the RNase protection showing genomic positions of ORFs and putative promoter elements, protected regions of transcripts mapped in this study, and riboprobes used to map transcripts.

protection was observed with either mock-infected-cell RNA or tRNA controls. The 200-nt fragment places the RNA initiation site at approximately the same position which was mapped by primer extension, confirming this result as well as results obtained by differential hybridization and genomic sequence analysis. As mentioned above, we were able to detect consensus ATF, E2F, and serum response elements proximal to the TATA box in the UL98 promoter.

DISCUSSION

In this report, we provide evidence for the existence of a large family of overlapping mRNAs transcribed from the unique long region of the HCMV genome, encompassing ORFs UL93 through UL99. A summary of the proposed structures for these transcripts is shown in Fig. 3. The smallest two messages, having sizes of 1.3 and 1.6 kb, have previously been shown to have the capacity to encode ORFs UL99 and UL99A (45, 46, 52, 53, 55, 56). The other messages, having sizes of 2.6, 5.6, 7.3, 9.1, and 10.5 kb, are shown in this report to be 3′ coterminal with the UL99 messages and to have the capacity to encode ORFs UL93 through UL98. Using primer extension and RNase protection analysis, we have mapped the transcription initiation sites for the 2.6- to 10.5-kb messages; each of the messages except for the 5.6-kb transcript appears to initiate from a TATA-containing promoter, on the basis of the presence of consensus TATA elements in the genomic sequence 20

to 30 bases upstream of the mapped initiation sites. The 5.6-kb transcript appears to initiate from TATA-less promoters; however, there are consensus E2F, Sp1, and MLTF sites in favorable positions upstream from the mapped 5′ start sites which could direct the initiation of these transcripts. E2F and Sp1 have both been demonstrated to direct activation of transcription from cellular TATA-less promoters (10, 11, 61, 72). Interestingly, many cellular promoters containing E2F sites control the expression of gene products required for DNA replication, such as dihydrofolate reductase (10, 11, 61, 72), as well as genes involved in cell cycle control, such as c-*myc* (74). An S-phase-specific complex containing E2F, cyclin A, $p33^{cdk2}$, and p107 has been shown to be induced at early times of HCMV infection of fibroblasts (76), and cellular E2F-responsive genes such as the dihydrofolate reductase, DNA polymerase α, and topoisomerase II genes are also known to be induced by HCMV infection (9, 28, 37, 39). Additionally, at least three families of HCMV early gene products have E2F-like sites in their promoters, including the 2.2- and 2.7-kb early transcription units (43, 44, 65) and the UL84 gene product (34, 36). The protein product of ORF UL84 has recently been shown to be required for viral DNA replication (57), thus establishing a possible link between E2F, HCMV early gene expression, and viral DNA replication. This effect is likely to be mediated through the HCMV IE72 and IE86 gene products, potentially in a manner analogous to the adenovirus E1A and E1B proteins, since cotransfection experiments demonstrate IE72 and IE86 responsiveness of viral early promoters. The 5.6-kb transcripts are also transcribed at early times of infection, and it is possible that proteins translated from this family of messages are also involved in viral DNA replication.

The 2.6- and 4.7-kb messages initiating upstream of UL98 and UL97, respectively, also contain putative E2F sites, and it is possible that these gene products are also involved in viral DNA replication. Since the promoters for both of these transcripts appear to contain TATA boxes, their mode of regulation is probably distinct from that of the 5.6-kb transcripts. Indeed, the 4.7-kb transcript appears to exhibit true early kinetics, since it is present at early times of infection, is sensitive to cycloheximide treatment, and is unaffected by the absence of viral DNA replication. The 2.6-kb message exhibits early-late kinetics similar to that of the 5.6-kb family, in that it is present at early times of infection, is sensitive to cycloheximide treatment, and requires viral DNA replication for maximal expression at late times of infection. Other HCMV transcription units exhibiting similar dual kinetics have been well documented (12, 47, 48). The protein product of ORF UL97 has recently been shown to encode a putative ganciclovir kinase activity (50, 71), and the UL98 gene has homology to the herpes simplex virus alkaline DNase (14, 18, 27). Both of these gene products are also highly conserved in all of the human herpesviruses as well as in equine herpesvirus and herpesvirus saimiri (13–15, 18, 21, 26, 27, 58, 62). The UL97 and UL98 homologs in other herpesviruses are also transcribed as a series of overlapping messages reminiscent of what is seen with HCMV (18, 26, 27); thus, the acquisition of these ORFs appears to have occurred early in the evolution of the herpesviruses, and their maintenance throughout the divergence of the herpesviruses suggests that they are involved in shared aspects of herpesvirus biology. It should be pointed out, however, that none of the ORFs in this family appear to be essential for viral DNA replication in permissive fibroblasts (57); thus, what role they might play in DNA replication is unclear. Also, it is unlikely that E2F responsiveness is the sole determinant controlling HCMV early gene expression, since not all early promoters contain E2F sites. Rather, E2F responsiveness probably provides one convenient mechanism by which high levels of viral gene products needed for viral DNA replication can be maintained at an appropriate point during the replicative cycle. Other mechanisms besides E2F responsiveness are certain to exist.

The 7.3-kb transcript also appears to exhibit early-late expression, since it is transcribed at early times of infection, is sensitive to cycloheximide treatment, and is partially dependent upon viral DNA replication. The putative promoter for the 7.3-kb transcript contains ATF/CRE elements as well as a serum response element. All of the TATA-containing promoters mapped in this study contain consensus CRE elements, suggesting that CREB binding could play an important role in regulating gene expression from these promoters. CREB binding sites have been shown to be located in the promoters of several other viral gene products (38, 51, 63). Interestingly, the recently cloned CREB-associated protein CBP has sequence homology with the E1A-associated p300 protein (4), which is itself a TFIID adaptor protein (27a). Since the activity of these factors is cell cycle regulated, it is possible that S-phase induction by HCMV (1, 76) could lead to expression of the inducible CREB-binding activity and the subsequent activation of HCMV promoters with ATF elements.

The 9.1- and 10.5-kb transcripts appear only at late times of infection, and their expression is completely abolished in the absence of viral DNA replication, suggesting that these are true late transcripts. These transcripts are likely to encode ORFs UL94 and UL93, respectively, on the basis of the positions of their cap sites. Indeed, the protein product of ORF UL94 is found in infected cells only at late times of infection (81b). UL94 and UL93 coding sequences are also found on a series of smaller, overlapping transcripts which appear to define a novel transcription unit. No precursor-product relationship between the larger and smaller messages is evident, however. Rather, it is more likely that the smaller messages utilize a second polyadenylation site located in the genomic sequence immediately downstream of UL94. Transcripts initiating upstream of UL94 at the site mapped in this study and terminating at this upstream polyadenylation signal would be approximately 2.0 kb in size, which is consistent with the 2.0-kb transcript seen with UL94 probes. Interestingly, the putative promoters upstream of UL93 and UL94 appear to be quite complex in that both sequences contain consensus binding sites for a variety of cellular factors, including NF-κB, Sp1, ATF, MLTF, and p53. Deletion analysis of the UL99/UL99A (another HCMV late gene) promoter did not indicate a dependence on sequences containing any of these binding sites (25). Thus, it is not yet clear whether these factors are required for UL93 and UL94 transcription. We are particularly intrigued by the consensus p53 binding site in the putative UL94 promoter, though, since only a very small number of these sites are found within the HCMV genome (81a). The UL94 gene product is highly conserved in the other human herpesviruses; the herpes simplex virus homolog of UL94 (UL16) was recently shown to be nonessential for productive viral infection in tissue culture (6, 7).

The pattern of translation which occurs from the UL99 family of transcripts is subject to speculation at present; however, since cap sites for each of the transcripts are located immediately upstream of different ORFs, it is likely that the first ATG encountered determines which protein product is translated from a given message (42). We cannot rule out the possibility that small splicing events which delete sequences for downstream reading frames are present in the mature transcripts, since the transcripts were not mapped contiguously by using overlapping probes spanning the entire UL93 to UL99 region; however, this is highly unlikely since the sizes of the transcripts

described in this report are consistent with sizes of unspliced transcripts initiating at the mapped cap sites and extending to the UL99 polyadenylation signal, and detailed mapping of the UL96 to UL97 region did not detect any splicing events. Structurally, the appearance of unspliced, overlapping transcripts utilizing a common polyadenylation signal and encoding distinct viral proteins has been documented for other HCMV transcription units (34, 60, 79). Additionally, it is possible that multiple reading frames are translated from the apparently polycistronic transcripts; however, coding sequences for the UL99 gene product (pp28) are found on all of the 2.6- to 10.5-kb messages, as shown by Northern hybridization, yet the pp28 protein is seen in infected cells only at late times of infection (25, 52, 53, 55, 56). Thus, pp28 is unlikely to be translated from the larger messages, which appear at earlier times of infection. The putative upstream polyadenylation signal located proximal to UL94 could represent the evolutionary emergence of a new signal intended to allow processing of messages upstream of UL95; indeed, in other herpesviruses, the UL94 homolog does not appear to be present on large transcripts containing the more downstream open reading frames (16). Alternatively, the use of multiple polyadenylation signals could provide a means of regulating the levels of proteins translated from these messages.

ACKNOWLEDGMENTS

We thank Eric Poma and Andrew Yurochko for critical reading of the manuscript, Tim Kowalik for helpful discussion, and Shu-Mei Huong for excellent technical assistance. B.A.W. thanks A. Romero for moral support.

This work supported by grants AI-12717 and CA 21773 from the National Institutes of Health to E.-S.H.

REFERENCES

1. **Albrecht, T., I. Boldough, M. Fons, C. H. Lee, S. AbuBakar, J. M. Russell, and W. W. Au.** 1989. Cell-activation responses to cytomegalovirus infection: relationship to the phasing of CMV replication and to the induction of cellular damage. Subcell. Biochem. **15:**157–202.
2. **Alford, C. A., and W. J. Britt.** 1990. Cytomegalovirus, p. 1981–2010. *In* B. N. Fields, D. M. Knipe, R. M. Chanock, M. S. Hirsch, J. L. Melnick, T. P. Monath, and B. Roizman (ed.), Fields virology. Raven Press, New York.
3. **Anders, D. G., and W. Gibson.** 1988. Location, transcript analysis, and partial nucleotide sequence of the cytomegalovirus gene encoding an early DNA-binding protein with similarities to ICP8 of herpes simplex virus type 1. J. Virol. **62:**1364–1372.
4. **Arany, Z., W. R. Sellers, D. M. Livingston, and R. Eckner.** 1994. E1A-associated p300 and CREB-associated CBP belong to a conserved family of coactivators. Cell **77:**799–800.
5. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1987. Current protocols in molecular biology. John Wiley & Sons, New York.
6. **Baines, J. D., and B. Roizman.** 1991. The open reading frames U_L3, U_L4, U_L10, and U_L16 are dispensable for the replication of herpes simplex virus 1 in cell culture. J. Virol. **65:**938–944.
7. **Baines, J. D., and B. Roizman.** 1992. The cDNA of U_L15, a highly conserved herpes simplex virus 1 gene, effectively replaces the two exons of the wild-type virus. J. Virol. **66:**5621–5626.
8. **Beck, S., and B. G. Barrell.** 1988. Human cytomegalovirus encodes a glycoprotein homologous to MHC class I antigens. Nature **331:**269–272.
9. **Benson, J. D., and E.-S. Huang.** 1990. Human cytomegalovirus induces expression of cellular topoisomerase II. J. Virol. **64:**9–15.
10. **Blake, M. C., and J. C. Azizkahn.** 1989. Transcription factor E2F is required for efficient expression of the hamster dihydrofolate reductase gene in vitro and in vivo. Mol. Cell. Biol. **9:**4494–5002.
11. **Blake, M. C., R. C. Jambou, A. G. Swick, J. W. Kahn, and J. C. Azizkahn.** 1990. Transcriptional initiation is controlled by upstream GC-box interactions in a TATAA-less promoter. Mol. Cell. Biol. **10:**6632–6641.
12. **Chang, C.-P., C. L. Malone, and M. F. Stinski.** 1989. A human cytomegalovirus early gene has three promoters that are regulated differently at various times after infection. J. Virol. **63:**281–290.
13. **Chee, M.** 1991. The HCMV genome project: what has been learned and what can be expected in the future. Transplant. Proc. **23**(Suppl. 3)**:**174–180.
14. **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. Tomlison, 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:**125–169.
15. **Chee, M. S., G. L. Lawrence, and B. G. Barrell.** 1989. Alpha-, beta- and gammaherpesviruses encode a putative phosphotransferase. J. Gen. Virol. **70:**1151–1160.
16. **Chen, M.-R., T.-Y. Hsu, S.-W. Lin, J.-Y. Chen, and C.-S. Yang.** 1991. Cloning and characterization of cDNA clones corresponding to transcripts from the BamHI G region of the Epstein-Barr virus genome and expression of BGLF2. J. Gen. Virol. **72:**3047–3055.
17. **Colberg-Poley, A. M., L. D. Santomenna, P. P. Harlow, P. A. Benfield, and D. J. Tenny.** 1992. Human cytomegalovirus US3 and UL36-38 immediate-early proteins regulate gene expression. J. Virol. **66:**95–105.
18. **Costa, R. H., K. G. Draper, L. Banks, K. L. Powell, G. Cohen, R. Eisenberg, and E. K. Wagner.** 1983. High-resolution characterization of herpes simplex virus type 1 transcripts encoding alkaline exonuclease and a 50,000-dalton protein tentatively identified as a capsid protein. J. Virol. **48:**591–603.
19. **Cranage, M. P., T. Kouzarides, A. T. Bankier, S. Satchwell, K. Weston, P. Tomlinson, B. Barrell, H. Hart, S. E. Bell, and T. Minson.** 1986. Identification of the human cytomegalovirus glycoprotein B gene and induction of neutralizing antibodies via its expression in recombinant vaccinia virus. EMBO J. **5:**3057–3063.
20. **Cranage, M. P., G. L. Smith, S. E. Bell, H. Hart, C. Brown, A. T. Bankier, P. Tomlinson, B. G. Barrell, and T. C. Minson.** 1988. Identification and expression of a human cytomegalovirus glycoprotein with homology to the Epstein-Barr virus BXLF2 product, varicella-zoster virus gbIII, and herpes simplex type 1 glycoprotein H. J. Virol. **62:**1416–1422.
21. **Cunningham, C., A. J. Davison, A. Dolan, M. C. Frame, D. J. McGeoch, D. M. Meredith, H. W. M. Moss, and A. C. Orr.** 1992. The UL13 virion protein of herpes simplex virus type 1 is phosphorylated by a novel virus-induced protein kinase. J. Gen. Virol. **73:**303–311.
22. **Davis, M. G., E. C. Mar, Y. M. Wu, and E.-S. Huang.** 1984. Mapping and expression of a human cytomegalovirus major viral protein. J. Virol. **52:**129–135.
23. **DeMarchi, J. M., C. A. Schmidt, and A. S. Kaplan.** 1980. Pattern of transcription of human cytomegalovirus in permissively infected cells. J. Virol. **35:**277–286.
24. **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.
25. **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.
26. **deWind, N., J. Domen, and A. Berns.** 1992. Herpesviruses encode an unusual protein-serine/threonine kinase which is nonessential for growth in cultured cells. J. Virol. **66:**5200–5209.
27. **Draper, K. G., G. Devi-Rao, R. H. Costa, E. D. Blari, R. L. Thompson, and E. K. Wagner.** 1986. Characterization of the genes encoding herpes simplex virus type 1 and type 2 alkaline exonucleases and overlapping proteins. J. Virol. **57:**1023–1036.

27a. **Eckner, R., M. E. Ewen, D. Newsome, M. Gerdes, J. A. DeCaprio, J. B. Lawrence, and D. M. Livingston.** 1994. Molecular cloning and functional analysis of the adenovirus E1A-associated 33-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. Genes Dev. **8:**869–884.

28. **Estes, J. E., and E.-S. Huang.** 1977. Stimulation of cellular thymidine kinases by human cytomegalovirus. J. Virol. **24:**13–21.
29. **Fickenscher, H., T. Stamminger, R. Ruger, and B. Fleckenstein.** 1989. The role of a repetitive palindrome sequence element in the human cytomegalovirus major immediate early enhancer. J. Gen. Virol. **70:**114–123.
30. **Geballe, A. P., F. S. Leach, and E. S. Mocarski.** 1986. Regulation of cytomegalovirus late gene expression: γ genes are controlled by posttranscriptional events. J. Virol. **57:**864–874.
31. **Geballe, A. P., and E. S. Mocarski.** 1988. Translational control of cytomegalovirus gene expression is mediated by upstream AUG codons. J. Virol. **62:**3334–3340.
32. **Genetics Computer Group.** 1991. Program manual for the GCG package, version 7, April 1991. Genetics Computer Group, Madison, Wis.
33. **Goins, W. F., and M. F. Stinski.** 1986. Expression of a human cytomegalovirus late gene is posttranscriptionally regulated by a 3′-processing event occurring exclusively late after infection. Mol. Cell. Biol. **6:**4202–4213.
34. **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.
35. **Hermiston, T. W., C. L. Malone, P. R. Witte, and M. F. Stinski.** 1987. Identification and characterization of the human cytomegalovirus immediate-early region 2 gene that stimulates gene expression from an inducible promoter. J. Virol. **61:**3214–3221.
36. **Huang, E.-S.** Unpublished data.
37. **Huang, E.-S.** 1975. Human cytomegalovirus. III. Virus-induced DNA poly-

merase. J. Virol. **16:**298–310.
38. **Hunninghake, G. W., M. M. Monick, B. Liu, and M. F. Stinski.** 1989. The promoter-regulatory region of the major immediate-early gene of human cytomegalovirus responds to T-lymphocyte stimulation and contains functional cyclic AMP-response elements. J. Virol. **63:**3026–3033.
39. **Isom, H. J.** 1979. Stimulation of ornithine decarboxylase by human cytomegalovirus. J. Gen. Virol. **42:**265–278.
40. **Jahn, G., T. Kouzarides, M. Mach, B. C. Sholl, B. Plachter, B. Traupe, E. Preddie, S. C. Satchwell, B. Fleckenstein, and B. G. Barrell.** 1987. Map position and nucleotide sequence of the gene for the large structural phosphoprotein of human cytomegalovirus. J. Virol. **61:**1358–1367.
41. **Kouzarides, T., A. T. Bankier, S. C. Satchwell, K. Weston, P. Tomlinson, and B. G. Barrell.** 1987. Sequence and transcription analysis of the human cytomegalovirus DNA polymerase gene. J. Virol. **61:**125–133.
42. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12:**857–872.
43. **Kulcher, 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.
44. **Kulcher, 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.
45. **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.
46. **Lahijani, R. S., E. W. Otteson, and S. C. St. Jeor.** 1992. A possible role for nonsense suppression in the synthesis of a human cytomegalovirus 58-kilodalton virion protein. J. Virol. **186:**309–312.
47. **Leach, F. S., and E. S. Mocarski.** 1989. Regulation of cytomegalovirus late-gene expression: use of three start sites in the transcriptional activation of ICP36 gene expression. J. Virol. **63:**1783–1791.
48. **Leatham, M. P., P. R. Witte, and M. F. Stinski.** 1991. Alternative 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:**6144–6153.
49. **Lehner, R., H. Meyer, and M. Mach.** 1989. Identification and characterization of a human cytomegalovirus gene coding for a membrane protein that is conserved among human herpesviruses. J. Virol. **63:**3792–3800.
50. **Littler, E., A. D. Stuart, and M. S. Chee.** 1992. Human cytomegalovirus UL97 open reading frame encodes a protein that phosphorylates the antiviral nucleoside analogue ganciclovir. Nature (London) **358:**160–162.
51. **Liu, B., and M. F. Stinski.** 1992. Human cytomegalovirus contains a tegument protein that enhances transcription from promoters with upstream ATF and AP-1 *cis*-acting elements. J. Virol. **66:**4434–4444.
52. **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.
53. **Martinez, J., and S. C. St. Jeor.** 1986. Molecular cloning and analysis of three cDNA clones homologous to human cytomegalovirus RNAs present during late infection. J. Virol. **60:**531–538.
54. **McDonough, S., S. Staprans, and D. Spector.** 1985. Analysis of the major transcripts encoded on the long repeat of human cytomegalovirus strain AD169. J. Virol. **53:**711–718.
55. **Meyer, H., A. T. Bankier, M. P. Landini, C. M. Brown, B. G. Barrell, B. Reuger, and M. Mach.** 1988. Identification and prokaryotic expression of the gene coding for the highly immunogenic 28-kilodalton structural phosphoprotein (pp28) of human cytomegalovirus. J. Virol. **62:**2243–2250.
56. **Pande, H., K. Campo, M. A. Churchill, B. R. Clark, and J. A. Zaia.** 1988. Genomic localization of the gene encoding a 32-kDa capsid protein of human cytomegalovirus. Virology **167:**306–310.
57. **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.
58. **Purves, F. C., and B. Roizman.** 1992. The UL13 gene of herpes simplex virus 1 encodes the functions for posttranslational processing associated with phosphorylation of the regulatory protein a22. Proc. Natl. Acad. Sci. USA **89:**7310–7314.
59. **Roizman, B., and A. E. Sears.** 1990. Herpes simplex viruses and their replication, p. 1795–1841. *In* B. N. Fields, D. M. Knipe, R. M. Chanock, M. S. Hirsch, J. L. Melnick, T. P. Monath, and B. Roizman (ed.), Fields virology. Raven Press, New York.
60. **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.
61. **Slansky, J. E., Y. Li, W. G. Kaelin, and P. J. Farnham.** 1993. A protein synthesis-dependent increase in E2F1 mRNA correlates with growth regulation of the dihydrofolate reductase promoter. Mol. Cell. Biol. **13:**1610–1618.
62. **Smith, R. F., and T. F. Smith.** 1989. Identification of new protein kinase-related genes in three herpesviruses, herpes simplex virus, varicella-zoster virus, and Epstein-Barr virus. J. Virol. **63:**340–455.
63. **Stamminger, T., H. Fickenscher, and B. Fleckenstein.** 1990. Cell type-specific induction of the major immediate early enhancer of human cytomegalovirus by cyclic AMP. J. Gen. Virol. **70:**105–113.
64. **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.
65. **Staprans, S. I., and D. H. Spector.** 1986. 2.2-kilobase class of early transcripts encoded by human cytomegalovirus. J. Virol. **57:**591–602.
66. **Stasiak, P. C., and E. S. Mocarski.** 1992. Transactivation of the cytomegalovirus ICP36 gene promoter requires the alpha gene product TRS1 in addition to IE1 and IE2. J. Virol. **66:**1050–1058.
67. **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.
68. **Stenberg, R. M., D. R. Thomsen, and M. F. Stinski.** 1984. Structural analysis of the major immediate early gene of human cytomegalovirus. J. Virol. **64:**1556–1565.
69. **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:**66–675.
70. **Stinski, M. F., D. R. Thomsen, R. M. Stenberg, and L. C. Goldstein.** 1983. Organization and expression of the immediate early genes of human cytomegalovirus. J. Virol. **46:**1–14.
71. **Sullivan, V., C. L. Talarico, S. C. Stanat, M. Davis, D. M. Coen, and K. K. Biron.** 1992. A protein kinase homologue controls phosphorylation of ganciclovir in human cytomegalovirus-infected cells. Nature (London) **358:**162–164.
72. **Swick, A. G., M. C. Blake, J. W. Kahn, and J. C. Azizkahn.** 1989. Functional analysis of GC element binding and transcription in the hamster dihydrofolate reductase gene promoter. Nucleic Acids Res. **17:**9291–9304.
73. **Tenney, D. J., and A. M. Colberg-Poley.** 1991: Expression of the human cytomegalovirus UL36-38 immediate early region during permissive infection. Virology **182:**199–210.
74. **Thalmeier, K., H. Synovzik, R. Mertz, E. L. Winnacker, and M. Lipp.** 1989. Nuclear factor E2F mediates basic transcription and trans-activation by E1a of the human MYC promoter. Genes Dev. **3:**527–536.
75. **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.
76. **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.
77. **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.
78. **Wathen, M. W., D. R. Thomsen, and M. F. Stinski.** 1981. Temporal regulation of human cytomegalovirus transcription at immediate early and early times after infection. J. Virol. **38:**446–459.
79. **Welch, A. R., L. M. McGregor, and W. Gibson.** 1991. Cytomegalovirus homologs of cellular G protein-coupled receptor genes are transcribed. J. Virol. **65:**3915–3918.
80. **Welch, A. R., L. M. McNally, and W. Gibson.** 1991. Cytomegalovirus assembly protein nested gene family: four 3′-coterminal transcripts encode four in-frame, overlapping proteins. J. Virol. **65:**4091–4100.
81. **Weston, K.** 1988. An enhancer element in the short unique region of human cytomegalovirus regulates the production of a group of abundant immediate early transcripts. Virology **162:**406–416.
81a. **Wing, B. A.** Unpublished data.
81b. **Wing, B. A., and E.-S. Huang.** Unpublished data.
82. **Wright, D. A., and D. H. Spector.** 1989. Posttranscriptional regulation of a class of human cytomegalovirus phosphoproteins encoded on an early transcription unit. J. Virol. **63:**3117–3127.
83. **Zaia, J. A.** 1991. Pathogenesis of CMV-associated diseases in 1990. Transplant. Proc. **23**(Suppl. 3)**:**1–4.