JOURNAL OF VIROLOGY, June 1989, p. 2699–2708
0022-538X/89/062699-10$02.00/0

Vol. 63, No. 6

Regulated Expression of Early and Late RNAs and Proteins from the Human Cytomegalovirus Immediate-Early Gene Region

RICHARD M. STENBERG,[1]†* ALISON S. DEPTO,[1]† JAMES FORTNEY,[1] AND JAY A. NELSON[2]

Department of Microbiology and Immunology, West Virginia University Health Sciences Center, Morgantown, West Virginia 26506,[1] and Department of Immunology, Research Institute of Scripps Clinic, La Jolla, California 92037[2]

Received 24 August 1988/Accepted 2 March 1989

Expression of RNA and protein from the human cytomegalovirus immediate-early (IE) gene region (map units 0.732 to 0.751) was analyzed at early and late times after infection. The level of RNA present at late times (48 to 72 h after infection) was significantly higher than that present at IE times (5 h after infection). The profile of IE RNA in the cytoplasm of infected cells was different from that previously reported on polysomes (R. M. Stenberg, P. R. Witte, and M. F. Stinski, J. Virol. 56:665–675, 1985). The data indicate that the 1.95-kilobase (kb) major IE region 1 mRNA, which codes for the 72-kilodalton (kDa) protein, and the 1.7-kb IE region 2 (IE2) spliced mRNA, which codes for the IE2 55-kDa protein, may be preferentially associated with polysomes. However, the IE2 2.2-kb unspliced mRNA, which codes for an 86-kDa protein, may be preferentially excluded. This RNA was abundant in the cytoplasm under IE conditions but was not present on polysomes in significant quantities. This indicates that IE gene products may be involved in translational control of cytomegalovirus RNA. At late times, new transcription takes place within region 2. A 1.5-kb RNA is transcribed from a late promoter in region 2 that apparently does not function in cells infected with DNA-negative mutant *ts*66. These results demonstrate that the IE gene region is transcribed throughout infection and that multiple levels of regulation exist.

Regulation of gene expression in human cytomegalovirus (CMV)-infected cells has been the subject of extensive investigation in recent years. Human CMV infection of permissive cells results in coordinated expression of viral genes that have been characterized as immediate early (IE), early, and late (7, 25, 46, 47). IE genes are expressed in the absence of de novo protein synthesis (7, 17, 25, 41, 46, 47, 49) and are stimulated by a component of the virion (32, 40). The IE gene region has been subdivided into several subregions, of which regions 1 and 2 (IE1 and IE2, respectively) are the most extensively characterized (38, 39, 41). IE1 codes for a 1.95-kilobase (kb) mRNA that is abundantly expressed on the polysomes under IE conditions and has been shown to code for the 72-kilodalton (kDa) major IE protein (38, 41). IE2 codes for a series of mRNAs which share upstream exons with IE1 and code for a series of proteins, some of which share amino-terminal homology with the 72-kDa protein (39). In addition to IE1 and IE2, two additional IE genes have been recently identified (20, 48).

Previous studies demonstrated that mRNA levels for the 72-kDa protein reached a maximum at 5 h after infection and subsequently declined (37). However, exons corresponding to region 2 RNAs increased during CMV infection, reaching a maximum at 72 h after infection (39). These events were explained by autoregulation of the major IE gene and the subsequent activation of a putative promoter existing in region 2 that functions at either early or late times (37, 39).

Regulation of early and late genes is less well understood. A number of early and late genes have been described, including herpesvirus homologs for glycoproteins B (4, 22, 24) and H (5), DNA polymerase (13, 21, 22), the major DNA-binding protein ICP8 (1, 19), several phosphorylated virion proteins (6, 18, 26, 30), and others of unknown function (12, 15, 16, 34). However, many of these studies have involved gene mapping rather than specific questions concerning regulation. Studies by Goins and Stinski (11) and Geballe et al. (10) implicate posttranscriptional regulation as having a major role in CMV gene expression. Our recent studies using a DNA-negative mutant of CMV, *ts*66, demonstrate that, in addition to posttranscriptional regulation, true late genes exist and require CMV DNA replication before their expression (unpublished data).

In this study, we attempted to address the regulation of gene expression from the IE gene region throughout the CMV replicative cycle. Our data demonstrate that this region is expressed maximally at late times after infection, that new species of RNA and protein are produced, and that multiple levels of regulation may be involved.

MATERIALS AND METHODS

Cell culture and virus propagation. Growth of human foreskin fibroblast (HF) cells and propagation of human CMV strain Towne have been described previously (41, 46, 47). CMV mutant *ts*66 was isolated from HF cells infected with CMV Towne in the presence of 20 μg of nitrosoguanidine per ml. This concentration of nitrosoguanidine effectively reduced the yield of infectious virus to 1% of that of control cultures (data not shown). Isolation of mutant virus was by plaque morphology. Plaques that failed to enlarge at 39.5°C (nonpermissive temperature) after temperature shift up were shifted back down to 33.5°C (permissive temperature) and picked 5 days later. Virus was plaque purified again and tested for its ability to replicate at 33.5 and 39.5°C.

Analysis of viral DNA replication. Viral DNA replication at 39.5°C was analyzed by radiolabeling wild-type (WT) CMV- or *ts*66-infected cells with 10 μCi of [^{3}H]thymidine (Dupont, NEN Research Products) per ml from 48 to 72 h after infection. Whole-cell nucleic acid was isolated as previously

* Corresponding author.
† Present address: Eastern Virginia Medical School, Norfolk, VA 23501.

described (36) and digested with *Xba*I restriction endonuclease. The samples were subjected to electrophoresis in 0.8% agarose gels, the gels were fluorographed with Amplify (Amersham Corp.), dried, and autoradiographed.

Recombinant plasmids. The plasmids used in this study and the methods for purification have been previously described (37–39, 41, 44).

Preparation of RNA. Cytoplasmic RNA was prepared as previously described (35, 37). Whole-cell RNA was prepared essentially as described by Strohman et al. (42), as previously described (37, 39). RNA prepared under IE conditions in the presence of anisomycin (38, 41) was as previously described. By our definition, *ts*66 RNA prepared at 72 h at the nonpermissive temperature is early RNA, whereas WT CMV RNA prepared at 72 h at the nonpermissive temperature is late RNA. RNA prepared in the presence of phosphonoacetic acid (PAA) was as previously described (46, 47).

RNA analysis. Single-strand nuclease analysis (2) of CMV RNA using mung bean (MB) nuclease and formaldehyde gel and Northern analysis were as previously described (37–39).

DNA sequence analysis. DNA sequencing was performed by the dideoxy-chain termination method as previously described (31), except that Klenow polymerase (Boehringer Mannheim Biochemicals) reactions were incubated at 50°C.

Radiolabeling of infected cells. Infected or uninfected cells were labeled for 2 h with 50 μCi of [^{35}S]methionine (Dupont, NEN) per ml in methionine-free medium containing 5% fetal bovine serum. IE proteins were labeled following overnight treatment of infected cells with cycloheximide as previously described (41). Cycloheximide-treated cells were washed three times with methionine-free minimal essential medium before labeling with [^{35}S]methionine.

Antibodies. Peptide antibodies were prepared as previously described (9). Antibody 8528 is to peptide IE1-1 (PDNPDGLPSSKVPRPETP) in the amino terminus of IE1 from amino acids 11 to 28 (38), 8575 is to peptide IE1-2 (AIVAYTLATAGASSSDSLVSPPESPVPATIPLSSVIVA) in the carboxy terminus of IE1 from amino acids 383 to 420 (38), 1219 is to peptide IE2-1 (KQEDIKPEPDFTIQYRN KII) in IE2 from nucleotides 3169 to 3228 (see Fig. 9), and 1218 is to peptide IE2-4 (QFEQPTETPPEDLDTLSC) in the carboxy-terminal region of IE2 and has been previously described (14).

Immune precipitation of cell extracts. Extracts for immune precipitation were prepared by washing cells with Tris-buffered saline (30 mM Tris hydrochloride [pH 7.4], 150 mM NaCl), followed by lysis in radioimmune precipitation assay buffer (30 mM Tris hydrochloride [pH 7.4], 150 mM NaCl, 1% Nonidet P-40, 0.5% deoxycholic acid, 0.1% sodium dodecyl sulfate [SDS]). Extracts were clarified by centrifugation and reacted with a 1:1,000 dilution of anti-peptide antibody on ice for 1 h. Immune complexes were precipitated on ice for 1 h with Pansorbin (Staph A; Calbiochem-Behring) as recommended by the supplier. Staph A immune complexes were pelleted, washed five times with radioimmune precipitation assay buffer, and subjected to SDS-polyacrylamide gel electrophoresis (SDS-PAGE) as previously described (27, 36).

Western blot (immunoblot) analysis. To determine steady-state levels of virus proteins, infected cells were washed with Tris-buffered saline and harvested in lysis buffer (50 mM Tris hydrochloride [pH 7.5], 1% SDS). Samples were subjected to SDS-PAGE as described above and transferred to nitrocellulose as described by Towbin et al. (45). The blots were treated with a protein-blocking solution (3) to block unbound sites and then reacted with a 1:1,000 dilution of specific antibody for 1 h at room temperature. After extensive washing with Tris-buffered saline, the blots were reacted with a 1:1,000 dilution of horseradish peroxidase-conjugated anti-rabbit immunoglobulin (Southern Biotechnologies). After being washed five times with Tris-buffered saline, the blots were developed with horseradish peroxidase developing reagent (Bio-Rad Laboratories) as recommended by the supplier.

Primer extension of RNA. An oligonucleotide primer (5′-GAGGTCACTCGCGGCGTGCC-3′) complementary to region 2 sequences beginning at nucleotide (nt) 3348 was synthesized on an Applied Biosystems Oligonucleotide Synthesizer and labeled at the 5′ end with [γ-^{32}P]ATP as previously described (38). Approximately 50,000 cpm of primer was mixed with 25 μg of the appropriate RNA in 1× hybridization buffer (2, 38) minus formamide and incubated overnight at 37°C. The product was precipitated with 2.5 volumes of ethanol, washed twice with 70% ethanol, and suspended in 50 mM Tris hydrochloride (pH 8.3)–75 mM KCl–3 mM $MgCl_2$–10 mM dithiotreitol–500 μM each dATP, dCTP, dGTP, and dTTP–10,000 U of Moloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories, Inc., Gaithersburg, Md.) per ml. The reaction was incubated at 37°C for 60 min, terminated with 10 mM EDTA, and precipitated with ethanol. The primed cDNA was suspended in sequencing gel loading buffer and applied to an 8% sequencing gel prepared as previously described (38). A known sequence ladder was also applied as a molecular size marker.

RESULTS

Replication of CMV mutant *ts*66. CMV mutant *ts*66 was tested for its ability to replicate at both the permissive and nonpermissive temperatures. The data in Table 1 demonstrate that at the permissive temperature, normal levels of *ts*66 relative to WT virus were produced. In contrast, at the nonpermissive temperature, no detectable virus was produced in *ts*66-infected cells. While we cannot conclusively say that no virus was produced, we were unable to detect measurable virus production at the nonpermissive temperature.

We also compared viral DNA replication in WT-infected versus *ts*66-infected cells at the nonpermissive temperature. *ts*66 failed to produce detectable levels of viral DNA (Fig. 1), indicating that this virus falls into the DNA-negative group of *ts* virus.

Expression of the IE gene region throughout CMV replication. To analyze expression of the IE gene region throughout the course of CMV infection, cells were infected and whole-cell RNA was prepared at the indicated times. To compare relative levels of abundancy, RNA from an equivalent number of infected cells was analyzed at each time. Initially, we used *Eco*RI fragment I as a DNA probe because it contains regions 1, 2, 2B, and part of 3. Early in infection (5 to 10 h after infection), relatively little RNA originated from this region, but as infection proceeded, this region became highly transcribed (Fig. 2A). Major RNA species of 9.5, 4.4, and 1.5 kb, as well as less abundant species of 2.2 and 1.8 kb, were detected. It is interesting that the 1.95-kb major IE mRNA was virtually undetected relative to RNAs expressed at later times and was seen as part of a faint doublet migrating from 1.9 to 2.2 kb. The IE RNAs were easily detected after longer exposures (data not shown). As infection proceeded, the lower half of the doublet corresponding to the 1.95-kb major

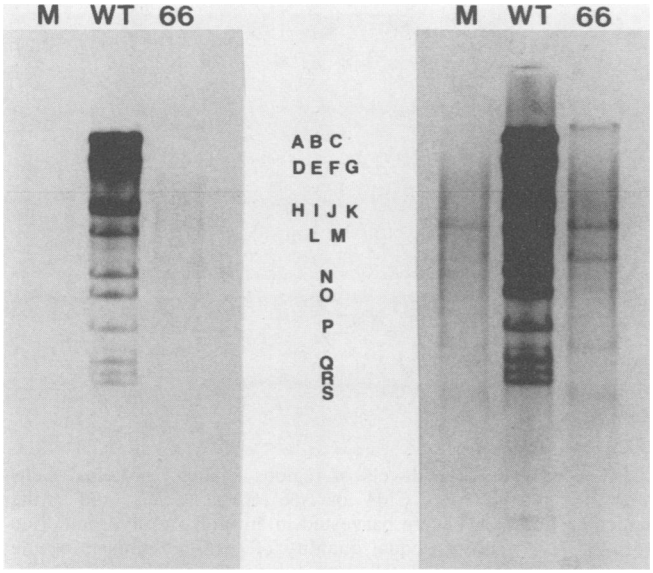

FIG. 1. Viral DNA synthesis in WT CMV-infected and *ts*66-infected cells. HF cells were infected with WT CMV or *ts*66 (66), labeled with [^{3}H]thymidine from 48 to 72 h, and analyzed as described in Materials and Methods. The letters between the panels represent the various *Xba*I restriction fragments. The panel on the right is a longer exposure of the panel on the left. M, Mock-infected cells.

IE mRNA disappeared. This is consistent with earlier studies on the expression of this mRNA (37).

Our studies with CMV DNA-negative mutant *ts*66 demonstrated that some CMV genes require viral DNA replication prior to expression (unpublished data). To investigate potential late genes present in this region, we infected cells with WT CMV or *ts*66 at 39.5°C and prepared cytoplasmic RNA at 72 h after infection. The RNA was subjected to Northern (RNA) blot analysis and hybridized with *Eco*RI fragment I. Figure 2B demonstrates that the 1.5-kb RNA was expressed only in WT-infected cells, suggesting that this RNA originates from a true late gene. We have also compared whole-cell RNA from WT-infected versus *ts*66-infected cells and demonstrated that the 1.5-kb RNA was not transcribed in *ts*66-infected cells (unpublished data; see Fig. 8).

To more accurately localize these RNAs to the physical map of the viral genome, cytoplasmic RNAs from WT-infected and *ts*66-infected cells were analyzed by Northern blot and hybridized to region 1 (Fig. 2D), region 2 (Fig. 2C), or region 2B plus 3 (Fig. 2E). The data demonstrate that the 9.5- and 4.4-kb RNAs originate from downstream region 2B or 3 and are probably not associated with region 1 or 2. Region 1 was expressed at relatively low levels, whereas region 2 was expressed at higher levels, and the 1.5-kb late RNA mapped entirely within region 2. These data are consistent with previous studies on IE1 and IE2 expression (37, 39).

To demonstrate that the 1.5-kb RNA can be produced in *ts*66-infected cells, we analyzed expression of region 2 RNA at the permissive temperature. The 1.5-kb RNA was produced in *ts*66-infected cells and at levels greater than those of WT virus (Fig. 3A). The reason for increased production of the 1.5-kb RNA in *ts*66-infected cells at the permissive temperature is unclear, but it may be due to selected growth of the *ts* virus at the lower temperature.

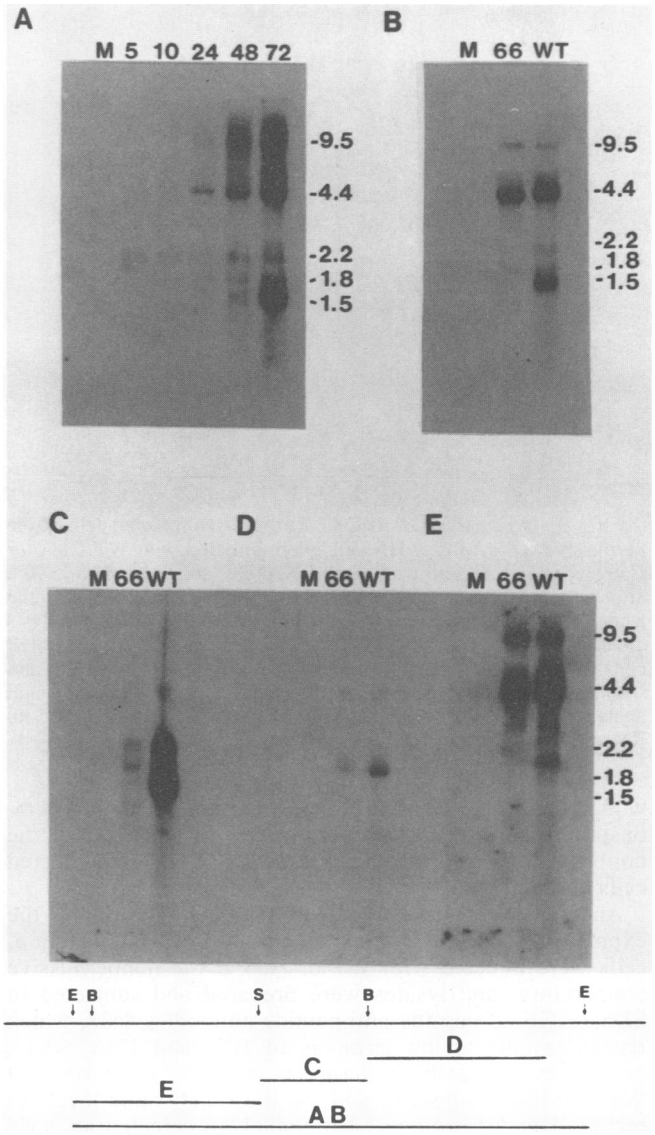

FIG. 2. Early and late gene expression from the IE gene region. (A) HF cells were infected with CMV, and at the indicated times after infection, whole-cell RNA was prepared. Mock-infected (M) cell RNA (1 μg) and RNA from an equivalent number of infected cells were subjected to Northern blot analysis and hybridized to *Eco*RI fragment I (probe AB in the diagram). (B) Cytoplasmic RNAs from mock-infected (M), WT-infected, and *ts*66-infected (66) cells were isolated at 72 h and hybridized to the *Eco*RI-I probe. Infections were performed at 39.5°C. Northern blots identical to those in panel B were hybridized to a region 2 probe (C), a region 1 probe (D), and a region 2B plus 3 probe (E). The autoradiographic exposure times for panels C to E were longer than that of panel B to demonstrate the *ts*66-infected cell RNAs. The locations of the probes are shown in the diagram. RNA sizes (kilobases) were estimated relative to an RNA ladder (Bethesda Research Laboratories). Abbreviations: B, *Bam*HI; E, *Eco*RI; S, *Sal*I.

Also, to demonstrate that our studies with *ts*66 were consistent with more conventional approaches that used DNA synthesis inhibitors, we analyzed RNA from infected cells treated with PAA. The 1.5-kb RNA was produced in infected cells at 72 h only in the absence of PAA (Fig. 3B). This demonstrated that the 1.5-kb RNA is a true late RNA. In support of the data, we have performed similar studies

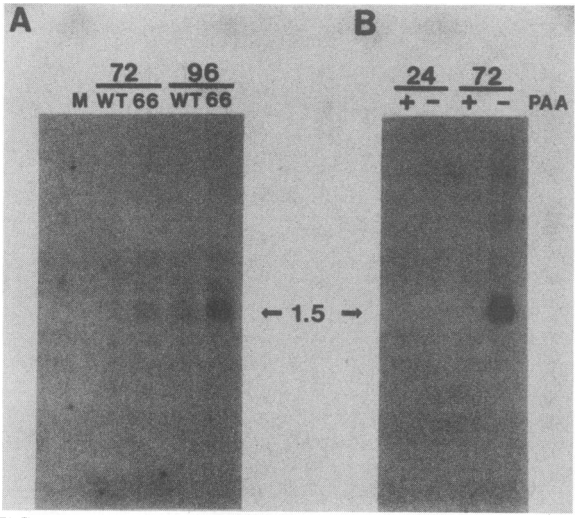

FIG. 3. (A) Analysis of region 2 RNA in *ts*66-infected cells at the permissive temperature. HF cells were infected with WT CMV or *ts*66 (66) at 33.5°C, and whole-cell RNA was isolated at 72 and 96 h after infection. Northern blot analysis was done as described in the legend to Fig. 2, with probe C. M, Mock-infected cells. (B) Analysis of region 2 RNA in WT CMV-infected cells treated with PAA. HF cells were infected in the presence (+) or absence (−) of PAA, and whole-cell RNA was isolated at 24 and 72 h after infection and analyzed as described in the legend to Fig. 2. The size of the 1.5-kb RNA is shown.

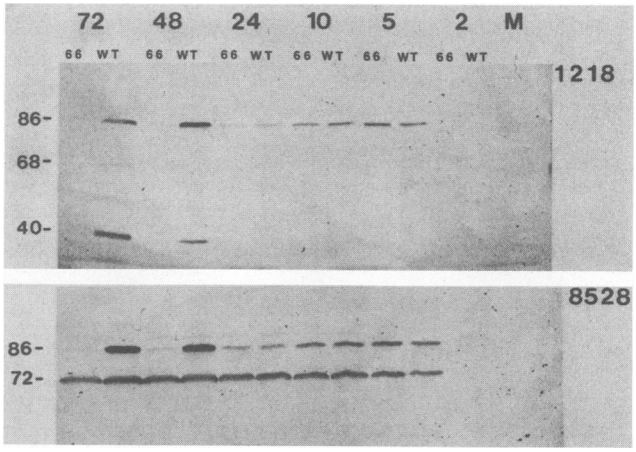

FIG. 4. Steady-state levels of regions 1 and 2 proteins. Cells were infected with WT CMV or *ts*66 (66) at 39.5°C, and at the indicated times they were harvested in 50 mM Tris hydrochloride (pH 7.5)–1% SDS. An equal quantity of extract representing the same number of infected cells was subjected to 15% SDS-PAGE. After transfer to nitrocellulose, the blots were reacted with antibody 1218 or 8528 as described in Materials and Methods. The numbers on the left indicate molecular sizes in kilodaltons.

with the pp65 lower-matrix phosphoprotein gene and demonstrated that the RNA profiles of *ts*66-infected cells at the nonpermissive temperature and PAA-treated WT-infected cells are identical (8).

Analysis of regions 1 and 2 polypeptides. To address the expression of regions 1 and 2 proteins during CMV infection, cells were infected with WT or *ts*66 at the nonpermissive temperature, and lysates were prepared and subjected to SDS-PAGE. Using the anti-peptide antibodies 8528, which recognizes the amino terminus of IE1, and 1218, which recognizes the carboxy terminus of IE2, we performed Western blot analysis to address the steady-state levels of regions 1 and 2 proteins. Antibody 1218 detected an early 86-kDa protein and a 40-kDa protein that appeared late (48 to 72 h), and 8528 detected the 72-kDa major IE protein and the 86-kDa protein (Fig. 4). The appearance of a minor 68-kDa protein was variable and may represent a breakdown product. The relative levels of the 72-kDa protein remained constant or increased slightly in both WT-infected and *ts*66-infected cells. However, the 86-kDa protein decreased in both WT-infected and *ts*66-infected cells through 24 h but subsequently increased in abundance at 48 h in WT-infected cells. In *ts*66-infected cells, the 86-kDa protein steadily decreased through 72 h. Antibody 8528 also detected a minor protein migrating at approximately 92 kDa. The identity and location of this protein are unclear and require further investigation. In addition to the 86-kDa protein, 1218 detected an abundant late protein of 40 kDa in WT-infected cells that was not present in *ts*66-infected cells. This may represent the protein product of the 1.5-kb late mRNA.

Analysis of protein synthesis in cycloheximide-reversed CMV-infected cells. Hermiston et al. (14) have proposed that an 82-kDa IE protein which reacts with antibody 1218 originates from IE2 and IE2B in cycloheximide-reversed CMV-infected cells. To determine whether this protein was the same as our 86-kDa protein, we infected cells in the presence of cycloheximide and reversed them as described in Materials and Methods. Extracts were prepared and incubated with the indicated anti-peptide antibodies or normal rabbit serum. The 72-kDa protein reacted with both IE1-

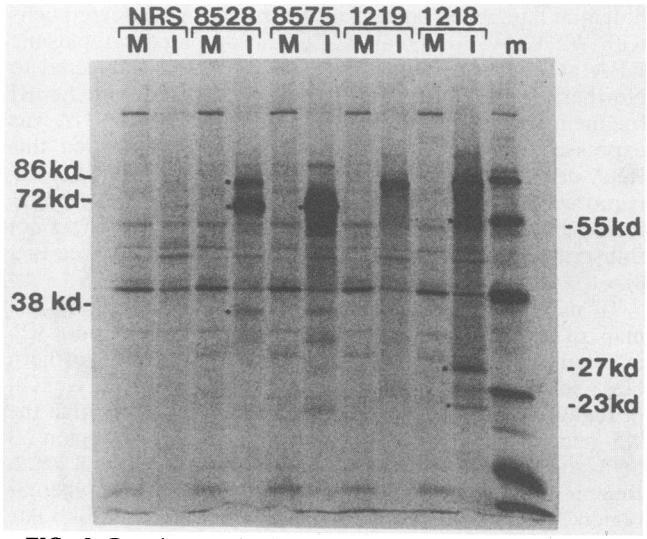

FIG. 5. Proteins synthesized in infected cells after cycloheximide reversal. Infected (I) and mock-infected (M) cells were treated with 200 μg of cycloheximide per ml overnight, washed three times with methionine-free minimal essential medium containing 5% fetal bovine serum, and labeled with [^{35}S]methionine for 2 h. Extracts were prepared as described in Materials and Methods and diluted 1:1,000 with the indicated antibody or normal rabbit serum (NRS). After incubation on ice for 1 h, the immune complexes were collected with Staph A (Pansorbin) and subjected to SDS-PAGE and autoradiography. ^{14}C-labeled protein molecular weight standards were from Bethesda Research Laboratories. The sizes of the proteins are expressed in kilodaltons (kd).

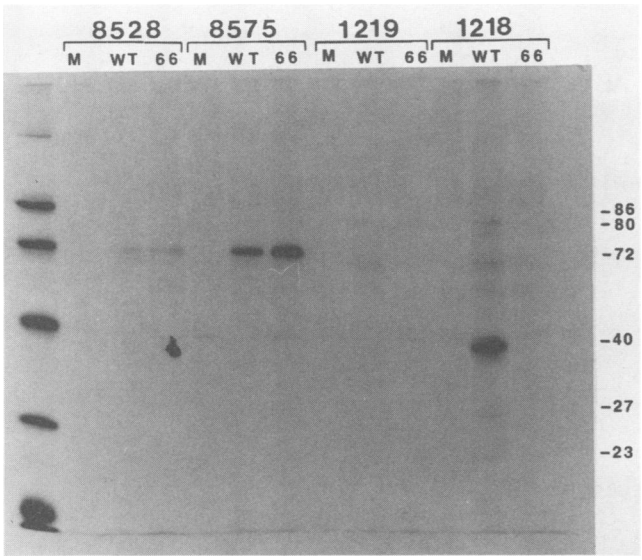

FIG. 6. Protein synthesis in WT-infected and *ts*66 (66)-infected cells at 72 h. Cells were infected with WT CMV or *ts*66 at 39.5°C for 72 h, washed three times with methionine-free-minimal essential medium, and labeled as described in the legend to Fig. 5. The cells were harvested, precleared with normal rabbit serum and Staph A for 1 h on ice, followed by precipitation with the indicated specific antibody and Staph A, and analyzed as described in the legend to Fig. 5. The numbers on the right indicate molecular sizes in kilodaltons.

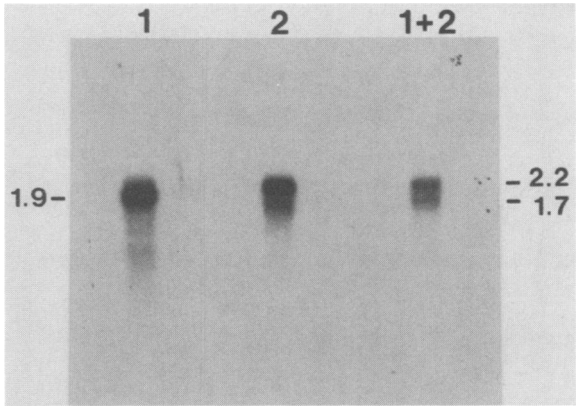

FIG. 7. Northern blot analysis of IE1 and IE2 cytoplasmic RNAs. IE cytoplasmic RNA was subjected to Northern blot analysis as previously described (39) and hybridized to IE1 (lane 1), IE2 (lane 2), and IE1 plus IE2 (lane 1+2) probes. The IE1 and IE2 probes were as described in the legend to Fig. 2. IE1 plus IE2 is a fragment extending from *Cla*I upstream of the IE1 5′ end to the *Sal*I site 3′ to IE2. The numbers on the sides indicate molecular sizes in kilobases.

specific antisera, 8528 and 8575 (Fig. 5). Antibodies 8528, 1219, and 1218 all reacted with the 86-kDa protein, suggesting that this protein originates from both IE1 and IE2. However, the 55-kDa IE protein, which Hermiston et al. (14) demonstrated reacts well with 1219, did not react well in our system. This was probably due to our lysis conditions for immune precipitation, since 1219 readily recognized the 55-kDa protein by Western blot analysis of cycloheximide-reversed infected cells subjected to whole-cell lysis (data not shown). In addition to the 86- and 72-kDa proteins, other minor proteins were evident at 23, 27, 31, and 38 kDa. The former two may represent IE2 proteins previously proposed (39), whereas the 31- and 38-kDa proteins are IE1 proteins recognized by both 8575 and 8528. Since these latter proteins were recognized by both IE1 amino- and carboxy-terminal anti-peptide antibodies, they may originate from minor, uncharacterized spliced IE1 RNAs.

Analysis of late proteins in WT-infected and *ts*66-infected cells. We also used these antibodies to address protein synthesis in WT-infected and *ts*66-infected cells at 72 h after infection. Only two proteins were predominantly synthesized in late infected cells, the 72-kDa major IE protein and the 40-kDa region 2 late protein (Fig. 6). Other proteins were also expressed, including the 86-kDa protein, although it was expressed at low levels at 72 h. Although the 86-kDa protein was not actively synthesized at 72 h after WT infection, it was abundant under steady-state conditions (Fig. 4). These data, taken together, indicate that the late burst of 86-kDa protein expression in WT-infected cells occurred after DNA replication, but by 72 h this burst of expression virtually ceased.

Nuclease analysis of region 2 early and late RNAs. Previous studies demonstrated that the IE1 major IE mRNA was depressed but detectable at late times (37) and that IE polysomal RNA from IE2 was overwhelmingly represented by the spliced species that codes for the 55-kDa IE protein (39). However, analysis of region 2 whole-cell RNA at 5 h after CMV infection demonstrated equal levels of spliced and unspliced RNAs, and as the CMV replicative cycle progressed, region 2 RNA showed a predominance of the unspliced exon (39). These earlier studies, in combination with the studies reported here, show that new species of region 2-specific RNA appear at early and late times after infection. Because earlier studies of IE cytoplasmic RNA dealt strictly with polysome-associated RNA, we examined total cytoplasmic IE RNA by Northern blot analysis. The profile of IE cytoplasmic RNA was different from that previously published for polysomal RNA. IE1 hybridized to the abundant 1.95-kb major IE mRNA, but hybridization with IE1 and IE2 demonstrated that the 1.7- and 2.2-kb RNAs were present in higher relative quantities (Fig. 7). Therefore, analysis of total cytoplasmic IE RNA indicated that IE2 RNAs were present at levels similar to or greater than those of IE1 RNA. In addition, IE2 spliced RNA species tended to accumulate on the polysomes more efficiently than did unspliced IE2 RNAs, resulting in different RNA profiles for IE polysomal and cytoplasmic RNAs.

To extend these studies, nuclease analyses of RNAs isolated at IE, early, and late times were performed to determine the levels of specific region 2 species. We used a 3′-labeled *Nco*I site in region 2 to analyze IE and late cytoplasmic RNAs. We have used this same probe in the past to analyze IE polysomal RNA (39). This probe detected both the spliced and unspliced species of IE RNA (Fig. 8A). However, the relative ratio of spliced to unspliced region 2 RNAs was greatly different from that which occurs with polysomal RNA (39) and is consistent with that reported in earlier studies of whole-cell RNA present at 5 h after infection (39). The ratio of spliced to unspliced cytoplasmic RNAs was about 1:1 whereas the ratio of spliced to unspliced RNAs on polysomes is usually 20:1 to 50:1 (39). At late times, the RNA was virtually 100% unspliced (Fig. 8A).

To complement this study, we used a 5′-labeled *Nco*I probe to determine 5′ exon structure in region 2. A new species of RNA was present in IE cytoplasmic RNA that was not demonstrated in IE polysomal RNA (Fig. 8B). The

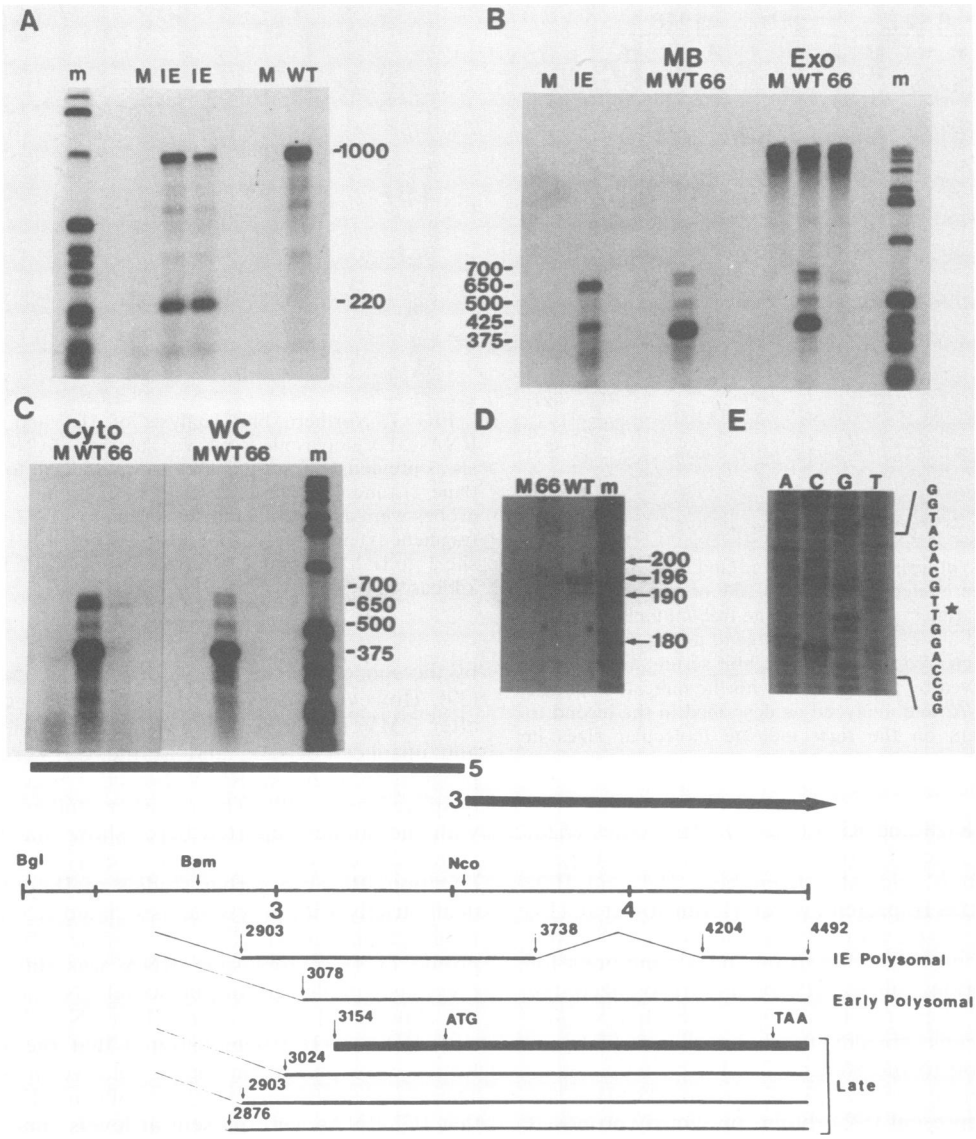

FIG. 8. Structural analysis of region 2 early and late exons. RNAs were prepared as described in Materials and Methods. WT-infected or *ts*66-infected cell RNA was isolated from cells infected at 39.5°C. (A) IE cytoplasmic RNA and WT 72-h late RNA (50 μg) were subjected to MB nuclease analysis as previously described (35, 37, 38), with a probe labeled at the 3′ end of the *Nco*I site (see the diagram). The positions of the spliced (220) and unspliced (1000) exons are indicated. (B) Using a 5′ probe from the *Nco*I site (see the diagram), 50 μg of IE cytoplasmic RNA was analyzed as described for panel A. Whole-cell RNAs (50 μg) from early (66) and late (WT) infected cells were hybridized to the 5′-labeled *Nco*I probe, and the samples were divided and treated with MB nuclease or exonuclease VII (Exo) as previously described (35, 37). (C) Cytoplasmic (Cyto) and whole-cell (WC) RNAs were analyzed with MB nuclease as described for panel B. (D) Whole-cell 72-h RNAs from WT-infected and *ts*66-infected cells were subjected to primer extension as described in Materials and Methods. (E) Sequence analysis within the region 2 intron was as previously described (31). The position of the previously unresolved T doublet is indicated with a star. Abbreviations: m, kilobase ladder from Bethesda Research Laboratories labeled with [γ-^{32}P] ATP and polynucleotide kinase; M, mock-infected cell RNA; Nco, *Nco*I; Bgl, *Bgl*II; Bam, *Bam*HI. The sizes of the protected fragments are given in nucleotides. The structures of the region 2 exons and the locations of the splice donor and acceptor sites are indicated. The potential open reading frame for the late 40-kDa protein extends from the ATG to the TAA indicated. The numbers to the sides of the gels indicate molecular sizes in nucleotides.

5′ end of this new exon exists at a position ~425 nt 5′ to the *Nco*I site. A series of AG dinucleotides exists in this area, with the best dinucleotide (according to consensus) at position 3077. In addition, the 5′ end of the exon at position 2903 is demonstrated by the band at ~650 nt and is consistent with data from previous studies (28).

Using the same 5′-labeled *Nco*I probe, we also examined 72 h whole-cell RNAs from WT-infected and *ts*66-infected cells. MB nuclease analysis generated a major band at ~375 nt and three minor bands at ~500, 650, and 700 nt. Exonuclease VII analysis generated bands identical to those created by MB nuclease, indicating that no upstream splicing of the major late exon occurs as far as the *Bgl*II site in region 1. Since Northern blot analysis demonstrated that this late RNA did not hybridize to upstream sequences, the 1.5-kb RNA is located within region 2. The three minor bands are likely to be 5′ ends of exons at positions 3024, 2903, and 2876. These ends are compatible with donor sites at posi-

FIG. 9. (A) Sequence of the 3′ end of exon 3 of IE1. The two potential donor splice junctions are indicated. The first donor is used in the formation of IE polysomal RNA. The second donor is likely used for the formation of the 2.2-kb IE cytoplasmic RNA. (B) Sequence of region 2, showing the 40-kDa open reading frame extending from ATG at position 3478 to TAA at 4384. The putative +1 nucleotide for the 1.5-kb late RNA and the TTATTAAA sequence located 32 nt upstream are indicated. Methionine residues and donor (GT) and acceptor (AG) sites are underlined.

tions 1358 and 1374, both of which occur at the end of exon 3 of IE1 (Fig. 9). The donor site at position 1374 has a high probability of usage on the basis of computer analysis of potential splice junctions (33). It is not clear why this diversity occurs at late times, and the complete structure of these late RNAs remains speculative, but a detailed study of these sequences is in progress.

To verify that cytoplasmic RNA was similar to whole-cell RNA, we analyzed cytoplasmic RNAs from WT (late)-infected and *ts*66 (early)-infected cells by hybridization to the 5′ *Nco*I probe. MB nuclease analysis demonstrated that cytoplasmic RNA was virtually identical to whole-cell RNA, except that the 5′ exon at position 2903 predominated in the cytoplasm and the 5′ exons at positions 3024 and 2876 were less abundant (Fig. 8C). Again, the major late transcript in this region initiated approximately 375 nt 5′ to the *Nco*I site.

Sequence analysis of region 2. It was not clear to us how region 2 was capable of coding for a 40-kDa late protein that originates solely from unspliced region 2 sequences. The coding capacity of region 2 was apparently insufficient to code for a 40-kDa protein from an unspliced mRNA (39). This was further complicated by the fact that the 40-kDa late protein was immune precipitated by the region 2 carboxy-terminal peptide antibody. This was also true for the 86-kDa protein, which further complicated interpretation of the data. Inspection of the sequence upstream of exon 6 of region 2 demonstrated that the reading frame was open to the TAG termination codon of exon 7 which occurs in the intron of region 2 (39). Therefore, we investigated the sequence around the TAG termination codon by using a modified extended Klenow procedure at 50°C to resequence the region. We have been successful at resolving compressed regions of DNA sequences with this method (unpublished data). This method resolved a T doublet that was previously unresolved which resulted in a +1 frameshift of the published sequence (39) and a continuous reading frame through region 2 (Fig. 8E). Therefore, the 40- and 86-kDa proteins can be produced by the region 2 unspliced exon.

Since peptide antibody 1219 did not react with the 40-kDa protein (Fig. 6), the initiation site for translation must occur downstream of this peptide domain. In addition, another region 2 peptide antibody, 1221 (9), did not react with the 40-kDa protein (data not shown). Initiation of a methionine residue at position 3478 would encode a protein a 302 amino acids and 33,286 daltons. This methionine residue also obeys the Kozak (23) rule for eucaryotic translation initiation, whereas the Met residue at position 3370 does not conform. Another methionine residue exists at position 3154 but would encode a protein significantly larger than 40 kDa. Also, this lies very close to the approximate 5′ end of the 1.5-kb late mRNA. To accurately map the 5′ end, we conducted primer extension by using an oligonucleotide extending from position 3328 to position 3347. A band of 196 nt from WT-infected cells was undetectable in *ts*66-infected cells (Fig. 8D). As a result, the 5′ nucleotide is very close to the A in the ATG codon at position 3154, thereby making it unlikely that this ATG could initiate translation. These data are summarized in Fig. 8 and 10, and the complete sequences of region 2 and the 40-kDa protein are shown in Fig. 9B.

DISCUSSION

The region from map units 0.732 to 0.751 on the CMV genome, commonly referred to as the IE gene region, is more complicated than originally proposed. During CMV infection, relative expression of this region at IE and early times after infection is actually quite low compared with its expression later in infection. Selective enhancement of IE RNAs with inhibitors of protein synthesis has been valuable in mapping their precise locations, their potentially encoded

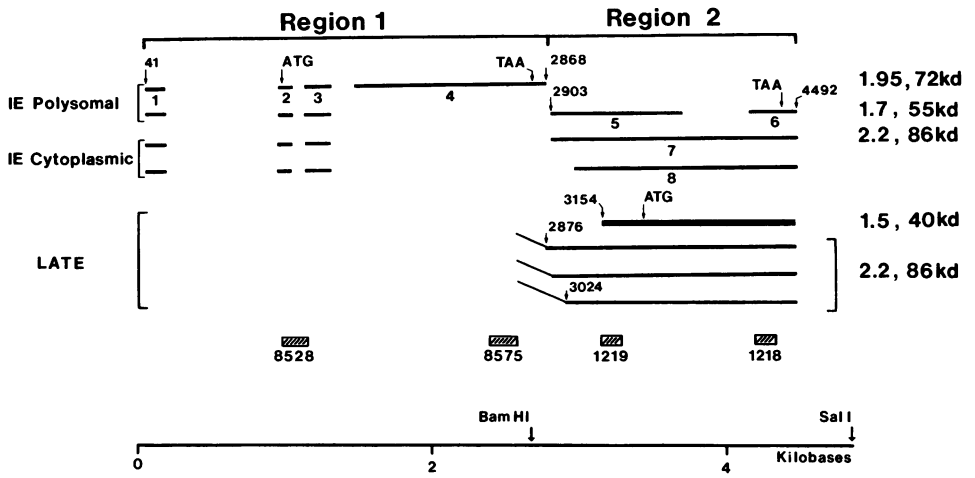

FIG. 10. Summary of RNAs from the IE gene region. IE polysomal, IE cytoplasmic, and late RNAs and the corresponding proteins originating from regions 1 and 2 proteins are indicated. Nucleotides that define the limits of regions 1 and 2 are shown, as well as nucleotides defining the 5′ end of the region 2 late exons. Exons are indicated by block horizontal lines, and the putative splice sites of the minor late RNAs that code for the 86-kDa protein(s) are indicated by incomplete diagonal lines. IE exons are labeled from 1 to 8. Translation initiation (ATG) and termination (TAA) sites are indicated. The locations of the synthetic peptides used to produce the corresponding anti-peptide antibodies are indicated by hatched boxes. The physical map of the IE gene region, indicating the locations of the *Bam*HI and *Sal*I sites, is shown at the bottom. kd, Kilodaltons.

proteins, and their potential roles in infection (17, 38, 39, 41, 49). However, in terms of the overall infection and the impact on the biology of CMV, this region undergoes dramatic changes in gene expression and likely plays a significant role throughout the viral replicative cycle. The structures of the predominant RNAs as determined from this and previous studies are summarized in Fig. 10.

At IE times, a series of mRNAs is produced by differential processing of RNAs that originate from the major IE promoter (39, 43). It has been suggested that the profile of RNA expression at IE times differs between CMV strains Towne (38–39, 41) and AD169 (17, 49). The present study suggests that this is not true and that this difference could be accounted for by the RNA isolation procedures and/or the types of analyses performed. Analysis of polysome-associated RNA demonstrated a series of IE2 RNAs that were at significantly lower levels than the IE1 RNA that codes for the 72-kDa protein (39, 41). Also, spliced IE2 RNA was present on the polysomes at significantly higher levels than unspliced IE2 RNA (39). Here, we examined total cytoplasmic IE RNA and found that the level of IE1 RNA was not higher than that of IE2 RNA and that the ratio of spliced RNA to unspliced RNA in IE2 was approximately 1:1. These data suggest that (i) the RNA for the 72-kDa protein and the spliced RNAs from IE2 are selectively present on polysomes relative to other IE RNAs and (ii) although the 2.2-kb RNA for the 86-kDa protein is transcribed, processed, and transported to the cytoplasm, it is not present on the polysomes. This latter finding is supported by early studies from Stinski et al. (41) in which in vitro translation of IE2 polysomal RNA did not result in synthesis of an 86- or 82-kDa protein. However, Hermiston et al. (14) demonstrated that reversal of cycloheximide-treated cells in the presence of actinomycin D resulted in expression of an 82-kDa protein. These data, taken together with those of our present studies, indicate that differential selection of transcripts and/or translational regulation may take place in CMV-infected cells at IE times and that some IE RNAs may (i) be preferentially excluded from polysomes or (ii) require a viral protein for translation.

At early times, little new gene expression takes place in this region. IE1 mRNA is down regulated, which is consistent with earlier studies (37), and low levels of the 2.2-kb RNA persist throughout early times. The finding that the 72-kDa major IE protein is present at significant levels throughout infection further suggests that its RNA is actively translated and associated with polysomes. This is supported by the finding that a significant level of 72-kDa protein synthesis occurs at 72 h, although RNA levels are reduced.

An early event that does take place is expression of the 86-kDa protein, which we will loosely characterize as an early protein that originates from an IE RNA. This protein is expressed early in infection (by 5 h) but declines continuously until late times. At this time, a new burst of expression occurs that appears to be accompanied by differential splicing at the 5′ end of region 2. Although the normal 5′ end at position 2903 is still preferred, two other 5′ exons exist. Actually, it is likely that several forms of the 86-kDa protein exist at late times because of splicing at the 5′ end of region 2. The reason for this complicated splicing pattern is unclear and requires further investigation. The burst of 86-kDa protein expression is short-lived and occurs at around 48 h after infection. By 72 h, very little 86-kDa protein is synthesized and the steady-state levels remain constant when compared with those at 48 h.

Perhaps the most interesting finding of this study is the discovery of an abundant late protein originating from region 2. It is not clear why a late protein would be encoded by a region that is primarily regulatory in nature. Re-expression of IE1 and its corresponding 89-kDa protein was first reported in murine CMV-infected cells, demonstrating a precedent for this phenomenon (29). However, this does not appear to be true for IE1 in human CMV-infected cells (37). Studies by Hermiston et al. (14) and Pizzorno et al. (28) have shown that IE2 is involved in positive regulation of heterologous promoters. Our recent studies using linker insertion mutagenesis demonstrated that there are two potential functional domains in IE2 that are involved in both positive activation of CMV early promoters and negative repression of the IE promoter (unpublished data). These two domains

may actually be one, since they are located on either side of the IE2 intron and are possibly integral portions of the IE2 55-kDa protein. The 40-kDa protein reads through this intron, thereby physically separating these two important domains (or alternatively disrupting the one domain) and possibly generating some new function. This protein may be involved in maintaining late gene expression down regulating specific early genes or have some other unknown function. Clearly, it requires further investigation.

The characterization of the region 2 late promoter demonstrates a true late CMV promoter that is strictly dependent on viral DNA replication. *ts*66, a DNA-negative mutant, has proven invaluable in identifying a true late class of genes (unpublished data). These late RNAs and proteins are not present in infected cells until ~48 h after infection and are present only after viral DNA replication. These data, together with those of recent studies by Meyer et al. (26), indicate that a true late class of genes exists in the CMV replicative cycle. This is in contrast to the findings of Geballe et al. (10), who postulated that all late genes are transcribed early and regulated posttranscriptionally. We could not detect IE2 late RNA before 48 h or in the absence of CMV DNA replication.

In conclusion, the IE gene region is actually not an IE gene region but expresses IE, early, and late genes in a complicated and coordinated manner. In addition, region 2 is most abundantly expressed at late times. The role of these genes and gene products in regulating subsequent viral gene expression is unclear and remains an important question that requires further study.

ACKNOWLEDGMENTS

This study was supported by Public Health Service grants AI-23313 (R.M.S.) and AI-21640 (J.A.N.) from the National Institutes of Health. Antibody to synthetic peptides was supported by Public Health Service grants AI-07007 and NS-12428 from the National Institutes of Health to M. B. A. Oldstone.

We thank Mark Stinski for the gift of recombinant clones containing various CMV fragments that facilitated our initial studies.

LITERATURE CITED

1. **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.
2. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus-mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. Cell **12:**721–732.
3. **Cohen, G. H., V. J. Isola, J. Kuhns, P. W. Berman, and R. J. Eisenberg.** 1986. Localization of discontinuous epitopes of herpes simplex virus glycoprotein D: use of a nondenaturing ("native" gel) system of polyacrylamide gel electrophoresis coupled with Western blotting. J. Virol. **60:**157–166.
4. **Cranage, M. P., T. Kouzarides, A. T. Bankier, S. Satchwell, K. Weston, P. Tomlinson, B. Barrell, H. Hart, S. E. Bell, and A. C. 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.
5. **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 gpIII, and herpes simplex virus type 1 glycoprotein H. J. Virol. **62:**1416–1422.
6. **Davis, M., and E. Huang.** 1985. Nucleotide sequence of a human cytomegalovirus DNA fragment encoding a 67-kilodalton phosphorylated viral protein. J. Virol. **56:**7–11.
7. **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.
8. **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.
9. **Fujinami, R. S., J. A. Nelson, L. Walker, and M. B. A. Oldstone.** 1988. Sequence homology and immunologic cross-reactivity of human cytomegalovirus with HLA-DR β chain: a means for graft rejection and immunosuppression. J. Virol. **62:**100–105.
10. **Geballe, A., F. Leach, and E. Mocarski.** 1986. Regulation of cytomegalovirus late gene expression: γ genes are controlled by posttranscriptional events. J. Virol. **57:**864–874.
11. **Goins, W. F., and M. F. Stinski.** 1986. Expression of a human cytomegalovirus late gene is posttranscriptionally regulated by a 3'-end-processing event occurring exclusively late after infection. Mol. Cell. Biol. **6:**4202–4213.
12. **Greenaway, P. J., and G. W. Wilkinson.** 1987. Nucleotide sequence of the most abundantly transcribed early gene of human cytomegalovirus strain AD169. Virus Res. **7:**17–31.
13. **Heilbronn, R., G. Jahn, A. Burkle, U. Freese, B. Fleckenstein, and H. zur Hausen.** 1987. Genomic localization, sequence analysis, and transcription of the putative human cytomegalovirus DNA polymerase gene. J. Virol. **61:**119–124.
14. **Hermiston, T., C. Malone, P. Witte, and M. 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.
15. **Hutchinson, N., R. Sondermeyer, and M. Tocci.** 1986. Organization and expression of the major genes from the long inverted repeat of the human cytomegalovirus genome. Virology **155:** 160–171.
16. **Hutchinson, N., and M. Tocci.** 1986. Characterization of a major early gene from the human cytomegalovirus long inverted repeat; predicted amino acid sequence of a 30-kDa protein encoded by the 1.2-kb mRNA. Virology **155:**172–182.
17. **Jahn, G., E. Knust, H. Schmolla, T. Sarre, J. A. Nelson, J. K. McDougell, and M. Fleckenstein.** 1984. Predominant immediate early transcripts of human cytomegalovirus AD169. J. Virol. **49:**363–370.
18. **Jahn, G., T. Kouzarides, M. Mach, B.-C. Scholl, 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.
19. **Kemble, G., A. McCormick, L. Pereira, and E. Mocarski.** 1987. A cytomegalovirus protein with properties of herpes simplex virus ICP8: partial purification of the polypeptide and map position of the gene. J. Virol. **61:**3143–3151.
20. **Kouzarides, T., A. T. Bankier, S. C. Satchwell, E. Preddy, and B. G. Barrell.** 1988. An immediate early gene of human cytomegalovirus encodes a potential membrane glycoprotein. Virology **165:**151–164.
21. **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.
22. **Kouzarides, T., A. T. Bankier, S. C. Satchwell, K. Weston, P. Tomlinson, and B. G. Barrell.** 1987. Large scale rearrangement of homologous regions in the genomes of HCMV and EBV. Virology **157:**307–316.
23. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47:** 1–45.
24. **Mach, M., U. Utz, and B. Fleckenstein.** 1986. Mapping of the major glycoprotein gene of human cytomegalovirus. J. Gen. Virol. **67:**1461–1467.
25. **McDonough, S., S. Staprans, and D. Spector.** 1985. Analysis of the major transcripts encoded by the long repeat of human cytomegalovirus strain AD169. J. Virol. **53:**711–718.
26. **Meyer, H., A. T. Bankier, M. P. Landini, C. M. Brown, B. G. Barrell, B. Ruger, and M. Mach.** 1988. Identification and pro-

caryotic expression of the gene coding for the highly immunogenic 28-kilodalton structural phosphoprotein (pp28) of human cytomegalovirus. J. Virol. **62**:2243–2250.
27. **Pizer, L. I., G. H. Cohen, and R. J. Eisenberg.** 1980. Effect of tunicamycin on herpes simplex virus glycoproteins and infectious virus production. J. Virol. **34**:142–153.
28. **Pizzorno, M. C., P. O'Hare, L. Sha, R. L. LaFemina, and G. S. Hayward.** 1988. *trans*-Activation and autoregulation of gene expression by the immediate-early region 2 gene products of human cytomegalovirus. J. Virol. **62**:1167–1179.
29. **Reddehase, M. J., M. R. Fibi, G. M. Keil, and U. H. Koszinowski.** 1986. Late phase expression of a murine cytomegalovirus immediate-early antigen recognized by cytolytic T lymphocytes. J. Virol. **60**:1125–1129.
30. **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.
31. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
32. **Spaete, R. R., and E. S. Mocarski.** 1985. Regulation of cytomegalovirus gene expression: α and β promoters are *trans* activated by viral functions in permissive human fibroblasts. J. Virol. **56**:135–143.
33. **Staden, R.** 1984. Computer methods to locate signals in nucleic acid sequences. Nucleic Acids Res. **12**:505–519.
34. **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.
35. **Stenberg, R. M., and L. I. Pizer.** 1982. Herpes simplex virus induced changes in cellular and adenovirus RNA metabolism in an adenovirus type 5-transformed human cell line. J. Virol. **42**:474–487.
36. **Stenberg, R. M., D. J. Spector, and L. I. Pizer.** 1979. Modulation of herpes simplex virus replication in adenovirus transformed cells. J. Gen. Virol. **44**:297–309.
37. **Stenberg, R. M., and M. F. Stinski.** 1985. Autoregulation of the human cytomegalovirus major immediate-early gene. J. Virol. **56**:676–682.
38. **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.
39. **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.
40. **Stinski, M. F., and T. J. Roehr.** 1985. Activation of the major immediate early gene of human cytomegalovirus by *cis*-acting elements in the promoter-regulatory sequence and by virus-specific *trans*-acting components. J. Virol. **55**:431–441.
41. **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.
42. **Strohman, R., P. Moss, J. Micou-Eastwood, D. Spector, A. Przybyla, and B. Peterson.** 1977. Messenger RNA for myosin polypeptides: isolation from single myogenic cell cultures. Cell **10**:265–273.
43. **Thomsen, D. R., R. M. Stenberg, W. F. Goins, and M. F. Stinski.** 1984. Promoter regulatory region of the major immediate early gene of human cytomegalovirus. Proc. Natl. Acad. Sci. USA **81**:659–663.
44. **Thomsen, D. R., and M. F. Stinski.** 1981. Cloning of the human cytomegalovirus genome as endonuclease Xba I fragments. Gene **16**:207–216.
45. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76**:4350–4354.
46. **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.
47. **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.
48. **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.
49. **Wilkinson, G., A. Akrigg, and P. Greenaway.** 1984. Transcription of the immediate early genes of human cytomegalovirus strain AD169. Virus Res. **1**:101–116.