Structural Analysis of the Major Immediate Early Gene of Human Cytomegalovirus

RICHARD M. STENBERG, DARRELL R. THOMSEN, AND MARK F. STINSKI*

Department of Microbiology, School of Medicine, University of Iowa, Iowa City, Iowa 52242

Received 11 July 1983/Accepted 28 September 1983

The most abundant species of human cytomegalovirus (Towne) immediate early polysome-associated RNA originates from a region of ca. 2.8 kilobases (0.739 to 0.755 map units) within the *Xba*I-E DNA fragment. These sequences code for a 1.95-kilobase mRNA and are referred to as immediate early coding region one (M. F. Stinski, D. R. Thomsen, R. M. Stenberg, and L. C. Goldstein, *J. Virol.* **46**:1-14, 1983). We have utilized the nuclease mapping technique of Berk and Sharp (A. J. Berk and P. A. Sharp, *Cell* **12**:721-732, 1977) to examine this gene in detail. Cloned fragments of human cytomegalovirus DNA, either labeled with ³²P in vivo or end labeled in vitro at the 5' or 3' termini, were hybridized to immediate early polysome-associated RNA. The hybrids were treated with single-strand-specific nuclease and subjected to electrophoresis in either neutral or denaturing gels. The major transcript was shown to be a spliced molecule containing a 3' terminal exon of 1,341 nucleotides. Upstream of the major body of the mRNA are three small exon sequences of 185, 88, and 121 nucleotides. The sequence of the exons as well as the locations of the intron-exon splice junctions were determined. Based on the DNA sequence, the viral mRNA molecule has one open reading frame which begins within the second exon and extends for 491 amino acid residues. The predicted molecular weight of the polypeptide originating from this region was estimated to be 64,000. It is hypothesized that this viral gene codes for the major regulatory protein controlling transcription of the viral genome at early times. The properties of the viral gene and its protein product are discussed.

Human cytomegalovirus (CMV), like other members of the herpesviruses, has a propensity for the establishment of latent infections (37). The first genes expressed after reactivation or after primary infection are the immediate early (IE) viral genes. An understanding of how the IE genes regulate subsequent viral gene expression is necessary to understand the underlying mechanisms of latency and reactivation.

Infection of permissive cells with CMV in the presence of inhibitors of protein synthesis (i.e., cycloheximide or anisomycin) leads to restricted transcription of the viral genome (5-8, 18, 30, 31, 34, 35). RNA made under these conditions is referred to as IE. The IE viral RNA in highest abundance originates from a restricted region of the long unique section which maps between ca. 0.66 and 0.77 map units on the prototype molecule for both the Towne and Davis strains of CMV (6-8, 30, 31, 34, 35). In contrast, viral RNAs that are transcribed at early times (i.e., in the absence of inhibitors of protein synthesis and in the presence of inhibitors of viral DNA synthesis) and at late times (i.e., after the onset of viral DNA synthesis) originate from all regions of the genome (5, 6, 8, 18, 34, 35).

We have previously analyzed the CMV IE polysome-associated RNA by R-loop and Northern blot analysis (30). The IE polysome-associated RNAs originate from three regions clustered in one area of the genome. These regions are referred to as IE coding regions one, two, and three and are located between map coordinates 0.739 to 0.751, 0.732 to 0.739, and 0.709 to 0.728, respectively, on the prototype molecule of the Towne strain of CMV. IE region one codes for the most abundant viral RNA molecule, a 1.95-kilobase (kb) species which is transcribed from right to left on the prototype arrangement. In vitro translation of mRNA selected by hybridization to IE region one demonstrated that this region codes for the predominant 72,000-dalton IE protein found in the infected cell within 1 h after infection (2, 28, 30).

The 72,000-dalton protein is a phosphorylated protein (11) that accumulates primarily in the nucleus (20, 32) and is chromatin associated (S. Michelson, personal communication). IE proteins of CMV increase chromatin transcriptional activity and cause alterations in chromatin conformation within 1 h after infection (14). These observations suggest that the 72,000-dalton protein may be a regulatory protein with the potential to influence viral as well as cellular transcription.

In this communication the structure of the gene coding for the CMV major IE 72,000-dalton protein is described. We employed the nuclease mapping technique of Berk and Sharp (1) in conjunction with DNA sequencing (17) to analyze this gene and its RNA product. The major IE mRNA was found to be a spliced RNA molecule containing four exon sequences. The sequence of the gene as well as the sequences at the splice junctions permitted us to deduce the structure of the mRNA molecule and its corresponding amino acid sequence.

MATERIALS AND METHODS

Cell culture and virus propagation. Primary human foreskin fibroblasts were prepared and cultured as previously described (28). Human CMV (Towne) was propagated by serial passage at low multiplicity in primary human foreskin cells as previously described (28). The amount of infectious virus was determined by either plaque assay (38) or tissue culture infectious doses (10).

IE RNA. IE polysome-associated RNA was prepared as described by Stinski et al. (30).

Recombinant plasmid DNA. The CMV recombinant plasmid DNAs that were used in this study and the methods for their purification have been previously published (30, 33).

Restriction enzymology and isolation of DNA fragments. All enzymes were purchased from Bethesda Research Laboratories, Bethesda, Md. Plasmid DNAs were digested with the

* Corresponding author.

appropriate restriction enzyme under the conditions recommended by the supplier. When the 5' termini of the DNA were to be labeled in vitro (see below), the termini were dephosphorylated with 5 units of bacterial alkaline phosphatase per pmol of fragment ends during the restriction enzyme digestion period. Digested DNA was subjected to electrophoresis in either 0.8 to 1.5% agarose or 5% acrylamide gels, the gels were stained in ethidium bromide, and the desired band was visualized and excised from the gel. The DNA was electroeluted from the gel by using 1/20 $\times$ electrophoresis buffer, and it was then ethanol precipitated.

Radioactive DNA probes. Plasmid DNAs were uniformly labeled in vivo by growing *Escherichia coli* in dephosphorylated LB broth in the presence of $^{32}\text{P}_i$. LB broth was dephosphorylated by the addition of 1 ml of 1 M MgCl_2 and 1 ml of concentrated NH_4OH to 100 ml of media. After stirring at room temperature for 30 min, the precipitate was filtered out, Tris buffer was added to a final concentration of 5 mM, and the pH was adjusted to 7.2. Dephosphorylated LB broth was sterilized by filtration and stored at 4°C.

To prepare in vivo-labeled plasmid DNA, *E. coli* cells were grown until the culture reached 100 Klett units. At that time, synthesis of the plasmid DNA was amplified with the appropriate antibiotic (300 μg of spectinomycin per ml for pACYC184 vectors and 200 μg of chloramphenicol per ml for pBR322 vectors), and 30 min later, 200 μCi of $^{32}\text{P}_i$ per ml (Amersham Corp., Arlington Heights, Ill.) was added to the culture. Amplification was for 12 to 16 h. To isolate plasmid DNA, a modification of the rapid boiling method of Holmes and Quigley (13) was employed. The 10-ml culture was transferred to a 30-ml centrifuge tube, and the bacteria were pelleted at 5,000 rpm for 15 min. The bacteria were suspended in 250 μl of STET buffer (8% sucrose, 5% Triton X-100, 50 mM EDTA, 50 mM Tris [pH 8.0]), followed by the addition of 20 μl of a solution (10 mg/ml) of lysozyme (Sigma Chemical Co., St. Louis, Mo.). The sample was boiled for 60 s, chilled on ice, and centrifuged for 15 min at 10,000 $\times g$. The supernatant was removed, and the nucleic acids were precipitated at -20°C with an equal volume of isopropanol. The nucleic acids were recovered by centrifugation, digested with the appropriate restriction enzyme, and subjected to fractionation by gel electrophoresis. The specific activity of the plasmid DNA was ca. 1×10^6 to 2×10^6 cpm/ μg .

To label the 5' termini of a DNA fragment, the forward reaction (17) was employed, using dephosphorylated DNA (see above) in the presence of [γ - ^{32}P]ATP (3,000 Ci/mM; Amersham Corp.) and T4 polynucleotide kinase (Bethesda Research Laboratories). Fragments to be 3' labeled were "filled in" with the appropriate [α - ^{32}P]deoxynucleoside triphosphate (3,000 Ci/mM; Amersham Corp.) in the presence of the large fragment of *E. coli* DNA polymerase I (Bethesda Research Laboratories) as previously described (19). Uniquely end-labeled DNA fragments were generated by performing a second restriction enzyme digestion or by separating the single strands (17), followed by isolation of the desired fragment from a 5% polyacrylamide gel as described above.

Structural analysis of RNA. The structure of the IE RNA was examined by the nuclease mapping technique as described by Berk and Sharp (1), using exonuclease VII (Bethesda Research Laboratories) and mung bean nuclease (P-L Biochemicals Inc., Milwaukee, Wis.) rather than S1 nuclease. Hybridizations containing the appropriate DNA probe (10^4 to 10^5 cpm/assay) and 5 to 25 μg of IE polysome-associated RNA were performed in 80% deionized formamide for 4 h at 50 to 60°C. The hybridizations were diluted into 20 volumes of either mung bean buffer (30 mM sodium

acetate [pH 4.6], 50 mM NaCl, 1 mM ZnCl_2 , 5% glycerol) or exonuclease VII buffer (10 mM Tris [pH 8], 100 mM NaCl, 10 mM EDTA, 5 mM 2-mercaptoethanol) and treated with an appropriate amount of the specified enzyme for 30 min at 37°C. After digestion, 10 μg of tRNA was added as a carrier and the nuclease-resistant hybrids were ethanol precipitated.

Gel electrophoresis. Nuclease-resistant hybrids were electrophoresed in neutral or alkaline 1.5% agarose gels as previously described (1). The gels were dried and exposed at -70°C by using Kodak XAR-5 film with Dupont Cronex intensifying screens. Denaturing urea acrylamide gels were prepared and subjected to electrophoresis as previously described (17, 25). A recombinant plasmid DNA consisting of the entire SV40 genome inserted into the *EcoRI* site of pBR325 was a gift of L. Villarreal. This plasmid was used to generate DNA fragments of known size for use as electrophoretic markers.

DNA sequencing. DNA sequence analysis was by the chemical degradation method of Maxam and Gilbert (17) as modified by Smith and Calvo (25), except that the AG modification reaction was performed for 5 min at 80°C in 34 mM sodium citrate (pH 4.5).

RESULTS

Physical map of IE region one. A detailed physical map showing the restriction enzyme cleavage sites, their relationship with respect to immediate early regions one, two, and three and the recombinant plasmids used in this study has been previously published (30). Figure 1 shows a detailed restriction enzyme cleavage map of IE region one generated by the method of Smith and Birnstiel (26) in conjunction with double restriction enzyme digestion of recombinant plasmids. Also shown is the sequencing strategy which permitted confirmation of the physical map in addition to allowing for the determination of subsequent restriction enzyme sites.

Nuclease mapping of the abundant IE mRNA. We previously demonstrated that the 3' end of the major IE mRNA was slightly left of the *Bam*HI site in IE region one (30) (Fig. 1). To analyze the structure of this mRNA in greater detail, we uniformly labeled plasmid DNA containing the *Bam*HI-A fragment (*Bam*HI-*Xba*I) of *Xba*I-E (30) in vivo with $^{32}\text{P}_i$ and isolated the 5.8-kb *Bam*HI-*Sal*I fragment (Fig. 2). This fragment was hybridized to IE polysome-associated RNA at 60°C and treated with either mung bean nuclease or exonuclease VII. The data in Fig. 2 show that treatment with mung bean nuclease (Fig. 2A) yielded a band of approximately 1,200 nucleotides (n) that shifts to a major band of 2,600 n plus two other minor bands of 1,800 n and 1,600 n after digestion with exonuclease VII (Fig. 2B). These data suggest that the abundant IE mRNA is a spliced molecule.

To examine how this mRNA molecule is oriented with regard to the physical map of the viral genome, we performed a similar experiment with the *Bam*HI-*Sal*I probe uniquely labeled with [γ - ^{32}P]ATP at the 5' end of the *Bam*HI site. Hybridization of this probe to polysome-associated IE RNA gave results similar to those obtained with a uniformly labeled DNA probe. Digestion with mung bean nuclease (Fig. 2C) yielded a major band at 1,200 n and a minor band at 1,600 n (Fig. 2B). This minor band may represent a minor species of RNA. Digestion with exonuclease VII (Fig. 2D) resulted in a major band at 2,600 n plus three other minor bands of lower molecular weight (Fig. 2D). Therefore, the major body of the abundant IE mRNA extends ca. 1,200 n to the right of the *Bam*HI site in region one with the 5' end of the RNA at a position ca. 2,600 n to the right of the *Bam*HI site. The minor bands generated after exonuclease VII digestion (Fig. 2B and D) were found to be due to the

FIG. 1. Restriction enzyme map and DNA sequencing strategy for IE region one. Restriction enzyme sites were generated as described in the text. The following abbreviations are used: A (*Ava*I), Av (*Ava*II), B (*Bam*HI), Bg (*Bgl*II), D (*Dde*I), H (*Hinf*I), P (*Pst*I), S (*Sst*I), S9 (*Sau*96I) and T (*Taq*I). The horizontal arrows at the bottom of the figure represent the sequencing strategy for the coding sequences of IE region one. Samples of DNA were labeled at either the 5' or 3' termini (tail of the arrow) and sequenced in the designated direction. Heavy arrows represent fragments which were analyzed with both 5' and 3'-labeled termini. Thin arrows represent fragments labeled at the 5' termini that were either cut with a secondary restriction enzyme or strand separated and then sequenced in the designated direction.

FIG. 2. Nuclease mapping of IE polysome-associated RNA. Polysome-associated RNA was hybridized to the *Bam*HI-*Sal*I fragment as described in the text. The location of the *Bam*HI-*Sal*I probe and its relationship to the IE region one physical map are shown at the bottom of the figure. The probes used in this study were prepared by either *in vivo* labeling the plasmid DNA with 32 P, or end labeling *in vitro* the 5' termini of the *Bam*HI site followed by secondary digestion with *Sal*I. The hybrids were treated with either mung bean nuclease or exonuclease VII and subjected to electrophoresis in either 1.5% alkaline (lanes A through D) or neutral (lane E) agarose gels. The gels were dried and subjected to autoradiography. Lanes: A, uniformly labeled probe digested with mung bean nuclease; B, same as lane A except exonuclease VII was used; C, end-labeled probe digested with mung bean nuclease; D, same as lane C except exonuclease VII was used; E, same as lane C. n, Nucleotides.

instability of the small exon sequences at 60°C (see below). Although the temperature at which the DNA becomes single stranded of the overall region is ca. 57°C under the hybridization conditions, our studies indicated that the 5' small exon sequences were unstable at this temperature. This type of variability in the integrity of the DNA:RNA hybrid yielded additional bands after exonuclease VII digestion. This explanation is consistent with the location of the 5' ends of the small exons with respect to their relative position on the physical map as described below.

Mung bean nuclease-digested hybrids were also subjected to electrophoresis on neutral agarose gels (Fig. 2E). A major band of 1,600 n was observed with two minor bands also being present. Again, these minor bands are presumably due to "breathing" of the DNA:RNA hybrid at the small exon sequences since the sizes of the bands are consistent with the size and location of the small exons as described below.

To analyze the size and location of the small exon sequences, plasmid DNA was uniformly labeled in vivo with $^{32}\text{P}_i$, the *Bam*HI-*Sal*I fragment was isolated and hybridized to IE polysome-associated RNA at 50°C. The hybrids were digested with mung bean nuclease and subjected to electrophoresis on an 8% acrylamide-7 M urea gel. Figure 3 (lane A) shows the mung bean digestion pattern with the *Bam*HI-*Sal*I probe. Exon sequences of 185, 120, and 85 n were detected. These data indicate that the major IE mRNA of CMV is a spliced molecule containing four exon sequences separated by three introns.

To map these sequences with respect to the CMV genome, IE region one was subdivided by using the restriction enzymes *Pst*I, *Sst*I, and *Ava*II. Equimolar amounts of the individual subfragments were hybridized to IE polysome-associated RNA at 50°C and analyzed as described above. The map location and sizes of the ^{32}P -labeled fragments are illustrated in Fig. 3. Hybridization of IE RNA to the *Pst*I 2.0-kb fragment (probe B) resulted in the detection of a single band of 120 n (Fig. 3B). A subfragment of this region (*Sst*I 0.7-kb fragment, probe E) also detected the 120-n exon (Fig. 3, lane E). These data indicate that the 120-n exon is within the *Sst*I 0.7-kb fragment. This was considered to represent the 5' leader of the RNA because exonuclease VII analysis (Fig. 2) also places the 5' end within this fragment. Hybridization of IE polysome-associated RNA to the *Pst*I 0.65-kb fragment (probe C) and subsequent nuclease treatment resulted in two bands of 350 and 150 n (Fig. 3C). The 350 n band represents the right-hand end of the major body that was truncated by the left-hand *Pst*I site. The 150 n band is the left hand 150 n of the 185 n exon that was truncated by the right-hand *Pst*I site (Fig. 3). If this interpretation is correct, then the 185 n exon should extend approximately 35 n into the adjacent *Pst*I fragment. Hybridization of the *Pst*I 0.25-kb fragment (probe D) to IE polysome-associated RNA resulted in a band of ca. 30 n in addition to a band of 85 n (Fig. 3D). These data also indicate that the 85 n exon originated exclusively from the 0.25-kb *Pst*I fragment. Analyses performed in the absence of RNA detected only bands corresponding to the size of the reassociated probe present at the top of each lane (data not shown).

To map the 120-n leader in more detail, uniformly labeled *Pst*I 2.0-kb fragment was isolated and redigested with *Ava*II (Fig. 3). Hybridization of IE polysome-associated RNA to the 500-n *Pst*I-*Ava*II (probe FG) fragment was not detectable. The intense band at the top is due to the reassociation of the probe in the absence of RNA (Fig. 3, lanes G and I). However, hybridization of IE RNA to the adjacent 0.34-kb *Ava*II-*Ava*II fragment (probe HI) resulted in a band of 60 n

(Fig. 3, lane H). This indicates that the right-hand *Ava*II site within the *Sst*I 0.7-kb fragment splits the 120-n leader in half, allowing for a determination of its precise location. The size and location of the exon sequences determined from the above data is summarized at the bottom of Fig. 3.

Structural analysis of the 3' end of the major IE mRNA. Our earlier studies indicated that the 3' end of the RNA extended to just left of the *Bam*HI site in region one (30). To determine the precise location of the 3' terminus of the major IE mRNA, we used [α - ^{32}P]dGTP and the large fragment of *E. coli* DNA polymerase I to label the 3' end of the *Bam*HI site of the 0.3-kb *Sst*I-*Bam*HI fragment (Fig. 2). This probe was used for RNA:DNA hybridization and mung bean nuclease analysis in addition to Maxam and Gilbert sequencing, and the reaction products were subjected to electrophoresis on a 12.5% acrylamide-7 M urea sequencing gel. Figure 4 demonstrates that the single mung bean digestion product aligns with the A in the sequence (5')TACGG(3'). Taking into account the 1- to 1.5-n correction factor (12), the 3' terminus was assigned to the C within the above sequence at a position 85 n left of the *Bam*HI site. On the complementary strand, this would read (3')ATGCC(5') with the G residue being the approximate 3' end. Supportive evidence for this being the 3' end is the presence of the sequence (5')AA-TAAA(3') at a position 22 to 27 n upstream (Fig. 4). This sequence is usually present 10 to 30 n upstream of the polyadenylation sites of eucaryotic mRNAs (23).

Structural analysis of the 5' end of the major IE RNA. Similar studies were performed to accurately map the 5' terminus of the major IE mRNA. Preliminary sequencing data had demonstrated the presence of a Hogness-Goldberg (4, 16) type TATATA structure upstream of the right-hand *Sst*I site within the *Pst*I 2.0-kb fragment. The *Ava*II-*Hinc*II fragment, uniquely labeled at the 5' end of the *Ava*II site, was hybridized to IE polysome-associated RNA in addition to being analyzed for DNA sequence. Figure 5A demonstrates that a single major band is present in the mung bean lane adjacent to the sequence. By using the 1- to 1.5-n correction factor (12), the band aligns with the A in the sequence (3')CAGTC(5'). The complementary strand, which would correspond to the sequence of the mRNA, would read (5')GTCAG(3') with the approximate 5' terminal nucleotide being at the T position.

Analysis of splice junction sequences. Figures 5B and 6 show the schematic representations of the DNA probes used for analyzing the sequences surrounding the donor and acceptor sites of the exons which make up the major IE mRNA. The RNA:DNA hybridizations, mung bean analyses and sequencing protocols were as described above. A representative pattern of a splice junction determination is shown in part B of Fig. 5. We used the *Ava*II site located within the 120-n leader and labeled at the 3' end to analyze the donor splice site. A major band was detected adjacent to the A in the sequence (5')GCATT(3') which would indicate a splice junction between the G and T within the sequence (5')GACG/TAAGT(3') of the opposite strand. This sequence is very close to the consensus sequence

(21, 24), but the splice junction appears to be 1 off in the 3' direction. A minor band occurs adjacent to the major band at a position consistent with the donor splice site. It is likely that this is the true splice junction and that the digestion conditions were insufficient in this situation.

FIG. 3. Analysis of the small exon sequences of the major IE mRNA. Plasmid DNA containing IE region one was grown in the presence of ^{32}P , and digested with the appropriate restriction enzyme. The locations of the ^{32}P -labeled fragments and their relationship to the physical map are shown at the bottom of the figure. The designated letter of each probe corresponds to its appropriate lane on the autoradiogram. Lanes: A, *Bam*HI-*Sall* (see Fig. 2); B, *Pst*I 2.0 kb; C, *Pst*I 0.65 kb; D, *Pst*I 0.25 kb; E, *Sst*I 0.7 kb; F and G, *Pst*I-*Ava*II 0.5 kb; H and I, *Ava*II-*Ava*II 0.34 kb. Lanes G and I represent reassociation of the probes in the absence of RNA. Each probe was hybridized to IE polysome-associated RNA at 50°C, digested with mung bean nuclease and subjected to electrophoresis in 8% polyacrylamide-7M urea gels. The gels were dried and subjected to autoradiography. Bands designated as (P) represent reassociated probe. n, Nucleotides.

The remaining intron-exon splice junctions were analyzed in a similar manner by using the end-labeled DNA probes depicted in Fig. 6. The data obtained from these studies are summarized in Fig. 7. Figure 7A shows the structure of the major IE mRNA with its corresponding intron sequences. Figure 7B shows the sequences around each intron-exon splice junction. Each intron-exon splice junction occurs at or around a consensus splice signal (21, 24). However, the bands were not always located at the theoretical splice junctions. This is presumably due to artifactual digestion

owing to various sequence anomalies or possible steric hindrance. Therefore, we have indicated the experimentally determined splice sites (Fig. 7) as well as the theoretical splice site based on consensus sequence. In the case of multiple bands, it is possible that the experimentally determined splice sites are due to alternative splice patterns. However, this situation is unlikely since the splice junctions would be inconsistent with the AG-GT rule (21, 24).

Predicted amino acid sequence of the major IE protein. To analyze in greater detail the structure of the major IE

FIG. 4. Structural analysis of the 3' terminus of the major IE mRNA. The 300-n *Sst*I to *Bam*HI fragment representing the left end of region one was uniquely labeled at the 3' end of the *Bam*HI site with [α - 32 P]dGTP as described in the text. This probe was hybridized to IE polysome-associated RNA at 50°C and digested with mung bean nuclease (MB). The probe was also subjected to the chemical degradation reactions as described by Maxam and Gilbert (17) for DNA sequence analysis. The products of these reactions were subjected to electrophoresis in a 12.5% acrylamide-7 M urea sequencing gel and autoradiographed. The sequence of the radioactive coding strand is shown adjacent to the autoradiogram. Also shown is the sequence of the RNA deduced from the sequence of the complementary strand. The presence of the consensus signal for polyadenylation (23) is shown at the bottom of the figure. The star designates the approximate 3' terminus.

mRNA, we sequenced the entire coding region of the major IE gene. The strategy for the sequencing is shown in Fig. 1. Using the consensus splice junctions (Fig. 7), we constructed the mRNA molecule which is shown in Fig. 8. The actual exon sizes of the mRNA by sequence are 121, 88, 185, and 1,341 n. A single open reading frame begins at nucleotide 139 at the sequence ATG and extends 491 amino acids to nucleotide 1,611 (excluding intron sequences) at which it meets the stop signal sequence TAA. By adding up the molecular weight of the individual amino acids, we obtained a molecular weight of ca. 64,000, which is close to the estimated molecular weight of 68,000 to 75,000 as previously reported (2, 3, 11, 27-30, 35).

DISCUSSION

The most abundant region of CMV IE transcription is in the long unique region of the viral genome between map coordinates 0.739 and 0.751. We had previously defined this area of the genome as IE region one and demonstrated that it coded for the predominant IE 72,000-dalton protein (30). It is

hypothesized that this IE protein plays a dominant role in the subsequent regulation of CMV gene expression. If the synthesis of this protein is inhibited, then early and late viral gene expression fails to occur (5-8, 18, 30, 31, 34, 35). A study with various concentrations of interferon demonstrated that the amount of 72,000-dalton protein synthesized is directly related to the level of early gene expression (29). Therefore, we believe this to be an important regulatory protein controlling transcription of other viral genes.

The 5' end of the viral RNA molecule was found to initiate at or near a U residue located 15 n to the left of the right-hand *Sst*I site (0.751 map units) in the *Pst*I 2.0-kb fragment (Fig. 1 and 5). If one assigns the 5' termini at position 1 on the DNA sequence, a (5')TATATA(3') sequence is present at nucleotides -28 through -23 and a (5')CAAAT(3') sequence at -61 through -57 (D. R. Thomsen, R. M. Stenberg, W. F. Goins, and M. F. Stinski, Proc. Natl. Acad. Sci. U.S.A., in press). Therefore, the typical Hogness-Goldberg boxes and CAAT boxes were found at the predicted distance and in the expected orientations for a eucaryotic gene (4, 16). Although the presence of a U at the position for the 5' terminal nucleotide is unusual, it does have a precedent. Green and Roeder (12) have reported that the 5' terminal nucleotide for the major viral RNA of adeno-associated virus also occurs at a U. In addition, the sequence around the start site of the major adeno-associated virus mRNA, (5')GTCAG(3'), is the same as that of the major IE gene of CMV. Whether or not these similarities have some molecular significance is presently unclear. The 3' terminus of the mRNA molecule occurred at a position 85 n left of the *Bam*HI site in region one. As is the case with other eucaryotic and viral mRNAs, the consensus sequence for polyadenylation, (5')AATAAA(3') (23), occurs at a position 22 to 27 n upstream. The location of the 3' end is in strong agreement with our previously published data (30).

The major mRNA that originates from IE region one is a spliced molecule containing four exon sequences. The approximate locations of the splice junctions were established by hybridizing defined 32 P-labeled DNA fragments to viral RNA and treating the hybrids with mung bean nuclease. The nuclease-resistant hybrids were fractionated in acrylamide-urea sequencing gels adjacent to chemically cleaved DNA fragments. Because the nuclease-resistant hybrids terminated at or near the consensus sequences for donor and acceptor sites, we were able to propose the theoretical splice points for the various intron-exon junctions. The 5' terminal leader sequence measured 121 n in length. Two small exons of 88 and 185 n exist downstream in addition to the major body of the RNA which measures 1,341 n.

The structures of the donor and acceptor sites correspond closely to the consensus sequence, and they always obey the AG-GT rules previously put forth (21, 24). The GT sequence of the G/GT donor splice junction was present at all three donor splice sites. This requirement appears to be very rigid as only a few eucaryotic mRNAs have been shown to diverge from this rule. In addition, at a position 5' to each acceptor site, we found the pyrimidine-rich sequence $(\frac{1}{2})_n$ which is consistent with published findings (21, 24). The purine dinucleotide AG occurred exclusively at the theoretical splice site of the acceptor sequence.

From these studies on the structure of the mRNA and its corresponding DNA sequence, the mRNA sequence and the amino acid composition of the viral protein can be predicted (Fig. 8). The first AUG initiation codon is located within the second exon at nucleotide 139 on the RNA molecule. A single open reading frame extends for 491 amino acids to

FIG. 5. Sequence analysis of the 5' end and downstream donor splice site of the 120-n leader. The labeling and sequencing strategy is described at the bottom of the figure. The *Ava*II site within the 120-n leader was utilized in this study. To examine the 5' terminus of the DNA, the *Ava*II site was labeled at the 5' end with [γ - 32 P]ATP, and the uniquely end-labeled fragment was generated by a secondary digestion with *Hinc*II. To examine the donor splice site, the 3' terminus of the *Ava*II site was labeled with [α - 32 P]dGTP, and the uniquely end-labeled fragment was generated by secondary digestion with *Sau*96I. These probes were used in analyses with polysome-associated RNA as described in the legend to Fig. 4. A, Analysis of the 5' terminus; B, analysis of the donor splice junction. The approximate locations of the 5' (closed star) and 3' (open star) ends of the 120-n leader are indicated on the RNA sequence or noncoding strand. The location of the consensus splice junction (SJ) is indicated. Also shown is the location of the TATATA sequence upstream from the 120-n leader.

nucleotide 1,612 at which a UAA stop signal sequence is located. A second AUG initiation codon exists at nucleotide 163 which is 24 n downstream from the proposed initiator codon at position 139 n. Therefore, it is possible that either one or both of these signals could represent the NH₂ terminus of the protein in that both obey the consensus start signal $\hat{A}$ NNATG(3') (15). Since most eucaryotic messages initiate at the first AUG, it is likely that the NH₂ terminus occurs at the AUG at nucleotide 139.

The predicted amino acid sequence indicates several interesting characteristics of this protein. (i) The protein has a high number of charged amino acids, both positive and negative, which results in a protein that is slightly acidic. This is consistent with our findings that the isoelectric point of the protein is ca. 5.9 (M. F. Stinski, unpublished data).

The COOH-terminal half of the peptide is extremely rich in glutamic acid and aspartic acid and results in an acidic region, whereas the center and NH₂ terminus are basic. (ii) The protein is slightly hydrophobic in nature with one distinct region containing high concentrations of hydrophobic or nonpolar amino acids. It occurs from amino acids 383 through 420, at which 36 of the 38 amino acids are either hydrophobic or nonpolar, with 23 being hydrophobic. Since the 72,000-dalton IE protein is highly associated with chromatin (S. Michelson, personal communication), this region may have a role in transport of the protein across the membranes into the nucleus or possibly in its association with chromatin. (iii) The protein is also very rich in proline at both the NH₂ and COOH termini. At the NH₂ terminus there is a stretch of 18 amino acids (11 through 28) containing six

A

B

Exon		Intron		Exon	
	3'	5'	3'	5'	
1	GAC	<u>G</u> TAAGT . . . (792n) . . . TCTTTCTGC	<u>A</u> G	TCAC	2
2	ACG	<u>G</u> TACGT . . . (90n) . . . TTTCTCATGTGTT	<u>A</u> G	GCCC	3
3	TCG	<u>G</u> TAAGT . . . (151n) . . . TCCTCCTCTAC	<u>A</u> G	TCAA	4
	CAG	<u>G</u> TGAGT	(C) T _n N _T CAG	G	

FIG. 7. Physical map and the nucleotide sequences of the splice junctions of the major IE mRNA. A, Summary of the structure of the major IE mRNA showing the introns and exons and their locations. The exon and introns are numbered in order as I1 or E1 for intron 1 and exon 1, respectively. The numbering originates from the 5' end of the mRNA. B, Sequences of the splice junctions are shown with the vertical line present at the theoretical consensus splice junction (21, 22). The approximate splice junctions determined by mung bean nuclease analyses are indicated by the arrows. The 5' and 3' designations indicate the termini of the introns and exons. The numbers in brackets indicate the approximate intron size.

-30	-20	-10	+1		
TCTATATAGC	AGAGCTCGTT	TAGTGAACCG	TCAGATCGCC	TGGAGACGCC	ATCCACGCTG
30					
TTTTGACCTC	CATAGAAGAC	ACCGGACCG	ATCCAGCCTC	CGCGCCGGG	AACGGTGCA
40					
TGGAACGGCG	ATTCCCCCTG	CCAAGAGTGA	CTCACCGTCC	TTGACACG	ATG GAG TCC
147					Met Glu Ser
(3)					
TCT GCC AAG AGA AAG ATG GAC CCT GAT AAT CCT GAC GAG GGC CCT TCC TCC					
Ser Ala Lys Arg Lys Met Asp Pro Asp Asn Pro Asp Glu Gly Pro Ser Ser					
198					
(20)					
AAG GTG CCA CGG CCC GAG ACA CCC GTG ACC AAG GCC ACG ACG TTC CTG CAG					
Lys Val Pro Arg Pro Glu Thr Pro Val Thr Lys Ala Thr Thr Phe Leu Gln					
249					
(37)					
ACT ATG TTG AGG AAG GAG GTT AAC AGT CAG CTG AGT CTG GGA GAC CCG CTG					
Thr Met Leu Arg Lys Glu Val Asn Ser Gln Leu Ser Leu Gly Asp Pro Leu					
300					
(54)					
TTT CCA GAG TTG GCC GAA GAA TCC CTC AAA ACT TTT GAA CGA GTG ACC GAG					
Phe Pro Glu Leu Ala Glu Glu Ser Leu Lys Thr Phe Glu Arg Val Thr Glu					
351					
(71)					
GAT TGC AAC GAG AAC CCC GAG AAA GAT GTC CTG GCA GAA CTC GTC AAA CAG					
Asp Cys Asn Glu Lys Pro Glu Lys Asp Val Leu Leu Val Lys Gln					
402					
(88)					
ATT AAG GTT CGA GTG GAC ATG GTG CGG CAT AGA ATC AAG GAG CAC ATG CTG					
Ile Lys Val Arg Val Asp Met Val Arg His Arg Ile Lys Glu His Met Leu					
453					
(105)					
AAA AAA TAT ACC CAG ACG GAA GAG AAA TTC ACT GGC GCC TTT AAT ATG ATG					
Lys Lys Tyr Thr Gln Thr Glu Val Asn Ser Gln Leu Ser Leu Gly Ala Phe Asn Met Met					
504					
(122)					
GGA CGA TGT TTG CAG AAT GCC TTA GAT ATC TTA GAT AAG GTT CAT GAG CCT					
Gly Gly Cys Leu Gln Asn Ala Leu Asp Ile Leu Asp Lys Val His Glu Pro					
555					
(139)					
TTC GAG GAG ATG AAG TGT ATT GGG CTA ACT ATG CAG AGC ATG TAT GAG AAC					
Phe Glu Glu Met Lys Cys Ile Gly Leu Thr Met Gln Ser Met Tyr Glu Asn					
606					
(156)					
TAC ATT GTA CCT GAG GAT AAG CGG GAG ATG TGG ATG GCT TGT ATT AAG GAG					
Tyr Ile Val Pro Glu Asp Lys Arg Glu Met Try Met Ala Cys Ile Lys Glu					
657					
(173)					
CTG CAT GAT GTG AGC AAG GGC GCC GCT AAC AAG TTG GGG GGT GCA CTG CAG					
Leu His Asp Val Ser Lys Gly Ala Ala Asn Lys Leu Gly Gly Ala Leu Gln					
708					
(190)					
GCT AAG GCC CGT GCT AAA AAG GAT GAA CTT AGG AGA AAG ATG ATG TAT ATG					
Ala Lys Ala Arg Ala Lys Lys Asp Glu Leu Arg Lys Met Tyr Met					
759					
(207)					
TGC TAC AGG AAT ATA GAG TTC TTT ACC AAG AAC TCA GCC TTC CCT AAG ACC					
Cys Tyr Arg Asn Ile Glu Phe Phe Thr Lys Asn Ser Ala Phe Pro Lys Thr					
810					
(224)					
ACC AAT GGC TGC AGT CAG GCC ATG GCG GCA TTG CAG AAC TTG CCT AAG TGC					
Thr Asn Gly Cys Ser Glu Ala Met Ala Leu Gln Asn Leu Pro Gln Cys					
861					
(241)					
TCC CCT GAT GAG ATT ATG GCT TAT GCC CAG AAA ATA TTT AAG ATT TTG GAT					
Ser Pro Asp Gly Ile Met Ala Tyr Ala Gln Lys Ile Phe Lys Ile Leu Asp					
912					
(258)					
GAG GAG AGA GAC AAG GTG CTC ACG CAC ATT GAT CAC ATA TTT ATG GAT ATC					
Glu Glu Arg Asp Lys Val Leu Thr His Ile Asp His Ile Phe Met Tyr Met					
963					
(275)					
CTC ACT ACA TGT GTG GAA ACA ATG TGT AAT GAG TAC AAG GTC ACT AGT GAC					
Leu Thr Thr Cys Val Glu Thr Met Cys Asn Glu Tyr Lys Val Thr Ser Asp					
1014					
(292)					
GCT TGT ATG ATG ACC ATG TAC GGG GGC ATC TCT CTC TTA AGT GAG TTC TGT					
Ala Cys Met Met Thr Met Tyr Gly Gly Ile Ser Leu Leu Ser Glu Phe Cys					
1065					
(309)					
CGG GTG CTG TCC TGC TAT GTC TTA GAG GAG ACT AGT GTG ATG CTG GCC AAG					
Arg Val Leu Cys Cys Tyr Val Leu Glu Glu Thr Ser Val Met Leu Ala Lys					
1116					
(326)					
CGG CCT CTG ATA ACC AAG CCT GAG GTT ATC AGT GTA ATG AAG CGC CGC ATT					
Arg Pro Leu Ile Thr Lys Pro Glu Val Ile Ser Val Met Lys Arg Arg Ile					
1167					
(343)					
GAG GAG ATC TGC ATG AAG GTC TTT GCC CAG TAC ATT CTG GGG GCC GAT CCT					
Glu Glu Ile Cys Met Lys Val Phe Ala Gln Tyr Ile Leu Gly Ala Asp Pro					
1218					
(360)					
CTG AGA GTC TGC TCT CCT AGT GTG GAT GAC CTA CGG GCC ATC GCC GAG GAG					
Leu Arg Val Cys Ser Pro Ser Val Asp Asp Leu Arg Ala Ile Ala Glu Glu					
1269					
(377)					
TCA GAT GAG GAA GAG GCT ATT GTA GCC TAC ACT TTG GCC ACC CGT GGT GCC					
Ser Asp Glu Glu Glu Ala Ile Val Ala Tyr Thr Leu Ala Thr Ala Gly Ala					
1320					
(394)					
AGC TCC TCT GAT TCT CTG GTG TCA CCC CCA GAG TCC CCT GTA CCC GCG ACT					
Ser Ser Ser Asp Ser Leu Val Ser Pro Pro Glu Ser Pro Val Pro Ala Thr					
1371					
(411)					
ATC CCT CTG TCC TCA GTA ATT GTG GCT GAG AAC AGT GAT CAG GAA GAA AGT					
Ile Pro Leu Ser Ser Val Ile Val Ala Glu Asn Ser Asp Gln Glu Glu Ser					
1422					
(428)					
GAG CAG AGT GAT GAG GAA GAG GAG GGT GCT CAG GAG GAG CGG GAG GAC					
Glu Gln Ser Asp Glu Glu Glu Glu Glu Gly Ala Gln Glu Glu Arg Glu Asp					
1473					
(445)					
ACT GTG TCT GTC AAG TCT GAG CCA GTG TCT GAG ATA GAG GAA GTT GCC CCA					
Thr Val Ser Val Lys Ser Glu Glu Pro Val Ser Glu Ile Glu Glu Val Ala Pro					
1524					
(462)					
GAG GAA GAG GAG GAT GGT GCT GAG GAA CCC ACC GCC TCT GGA GGC AAG AGC					
Glu Glu Glu Glu Asp Gly Ala Glu Glu Pro Thr Ala Ser Gly Gly Lys Ser					
1575					
(479)					
ACC CAC CCT ATG GTG ACT AGA AGC AAG GCT GAC CAG TAA ACTATTGTAT					
Thr His Pro Met Val Thr Arg Ser Lys His Ala Asp Gln stop					
1624					
(491)					
ATATATATCA GTTACTGTGA TGGATCCAC GTCACTATTG TATACTCTAT ATTACTCTCT					
1684					
ATGTTACTACT CTGTAATCCT ACTCAATAAA CGTGTACACG CTGTGAACC G					
1735					

FIG. 8. The nucleotide sequence of DNA coding for the major IE mRNA and the corresponding amino acid sequence. The nucleotides and amino acids (in parentheses) are numbered beginning with the first nucleotide or amino acid. The upstream region is numbered in the opposite direction. Regions of interest are underlined and explained in the text.

proline residues spaced approximately every 4 amino acids. A similar situation exists at the NH₂-terminal end of the major regulatory protein of herpes simplex virus type 1 (Vmw 175 or ICP 4) (22). The NH₂-terminal region of this protein is also proline rich and may impart some type of critical secondary structure to the protein since proline residues are responsible for the formation of beta-pleated sheets in proteins. Since these residues lead to a high degree of intramolecular secondary structure, they may be partly responsible for the anomalous migration of the protein in denaturing polyacrylamide gels and for the overestimation of its molecular weight. A similar overestimation in molecular weight has been reported for glycoproteins C (9) and D (36) of herpes simplex virus, both of which contain high amounts of proline. In addition, the 72,000-dalton protein has been reported to be phosphorylated (11), and therefore, different degrees of phosphorylation may also contribute to a diffused migration in denaturing polyacrylamide gels.

The high level of expression of this region of the genome at IE times and the correlation between synthesis of this protein with extensive transcription of the viral genome suggests that the major IE protein is probably regulatory in nature. The sequence analysis of this major IE gene and the putative protein structure provides valuable information for analyzing this important regulatory protein. Analysis of the protein at the amino acid level will assist in determining the functional domains of this protein. The construction of mutations at this level, either by deletion of amino acids or altering the amino acid composition, will contribute to understanding the role of this protein in CMV replication.

ACKNOWLEDGMENTS

We thank Pamela R. Witte for expert technical assistance and C. Martin Stoltzfus and William F. Goins for intellectual input and critical review of the manuscript. We also thank Mark Urbanowski and George V. Stauffer for assistance in establishing the DNA sequencing technology in our laboratory.

This investigation was supported by Public Health Service grant AI13526 from the National Institute of Allergy and Infectious Diseases and by grant I-697 from the National Foundation of the March of Dimes. M.F.S. is the recipient of Public Health Service research career development award AI100373 from the National Institute of Allergy and Infectious Diseases. R.M.S. is the recipient of a fellowship from the Ladies Auxiliary of the Veterans of Foreign Wars.

LITERATURE CITED

- 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.
- Blanton, R. A., and M. J. Tevethia. 1981. Immunoprecipitation of virus-specific immediate-early and early polypeptides from cells lytically infected with human cytomegalovirus strain AD169. *Virology* 112:262-273.
- Cameron, J. M., and C. M. Preston. 1981. Comparison of the immediate early polypeptides of human cytomegalovirus isolates. *J. Gen. Virol.* 54:421-424.
- Chambon, P., and R. Breathnach. 1981. Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50:349-383.
- Chua, C. C., T. H. Carter, and S. St. Jeor. 1981. Transcription of the human cytomegalovirus genome in productively infected cells. *J. Gen. Virol.* 56:1-11.
- De Marchi, J. M. 1981. Human cytomegalovirus DNA: restriction enzyme cleavage maps and map locations for immediate-early, early, and late RNAs. *Virology* 114:23-38.
- De Marchi, J. M. 1983. Post-transcriptional control of human cytomegalovirus gene expression. *Virology* 124:390-402.
- De Marchi, J. M., C. A. Schmidt, and A. S. Kaplan. 1980.

- Patterns of transcription of human cytomegalovirus in permissively infected cells. *J. Virol.* **35**:277-286.
9. Frink, R. J., R. Eisenberg, G. Cohen, and E. K. Wagner. 1983. Detailed analysis of the portion of the herpes simplex virus type 1 genome encoding glycoprotein C. *J. Virol.* **45**:634-647.
 10. Furukawa, T., A. Fioretti, and S. Plotkin. 1973. Growth characteristics of cytomegalovirus in human fibroblasts with demonstration of protein synthesis early in viral replication. *J. Virol.* **11**:991-997.
 11. Gibson, W. 1981. Immediate-early protein of human cytomegalovirus strains AD169, Davis, and Towne differ in electrophoretic mobility. *Virology* **112**:350-354.
 12. Green, M. R., and R. G. Roeder. 1980. Definition of a novel promoter for the major adenovirus-associated virus mRNA. *Cell* **22**:231-242.
 13. Holmes, D. S., and M. Quigley. 1981. A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* **114**:193-197.
 14. Kamata, T., S. Tanaka, and Y. Watanabe. 1978. Human cytomegalovirus induced chromatin factors responsible for changes in template activity and structure of infected cell chromatin. *Virology* **90**:197-208.
 15. Kozak, M. 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiol. Rev.* **47**:1-45.
 16. Liebhaver, S. A., M. J. Goossens, and Y. Wai Kan. 1980. Cloning and complete nucleotide sequence of human 5'- α -globin gene. *Proc. Natl. Acad. Sci. U.S.A.* **77**:7054-7058.
 17. Maxam, A. M., and W. Gilbert. 1980. Sequencing with basic specific chemical cleavages. *Methods Enzymol.* **65**:499-560.
 18. McDonough, S. H., and D. H. Spector. 1983. Transcription in human fibroblasts permissively infected by human cytomegalovirus strain AD169. *Virology* **125**:31-46.
 19. McLauchlan, J., and J. B. Clements. 1982. A 3' co-terminus of two early herpes simplex virus type 1 mRNAs. *Nucleic Acids Res.* **10**:501-512.
 20. Michelson-Fiske, S., F. Horodniceanu, and J. C. Guillon. 1977. Immediate early antigens in human cytomegalovirus-infected cells. *Nature (London)* **271**:615-617.
 21. Mount, S. M. 1982. A catalog of splice junction sequences. *Nucleic Acids Res.* **10**:459-472.
 22. Murchie, M.-J., and D. J. McGeoch. 1982. DNA sequence analysis of an immediate-early gene region of the herpes simplex virus type 1 genome (map coordinates 0.950 to 0.978). *J. Gen. Virol.* **62**:1-15.
 23. Proudfoot, N. J., and G. G. Brownlee. 1976. 3' non-coding region sequences in eucaryotic messenger RNA. *Nature (London)* **263**:211-214.
 24. Sharp, P. A. 1981. Speculation on RNA splicing. *Cell* **23**:643-646.
 25. Smith, D. R., and J. M. Calvo. 1980. Nucleotide sequence of the *E. coli* gene coding for dihydrofolate reductase. *Nucleic Acids Res.* **8**:2255-2274.
 26. Smith, H. O., and M. L. Birnstiel. 1976. A simple method for DNA restriction site mapping. *Nucleic Acids Res.* **3**:2387-2399.
 27. Stinski, M. F. 1977. Synthesis of proteins and glycoproteins in cells infected with human cytomegalovirus. *J. Virol.* **23**:751-767.
 28. Stinski, M. F. 1978. Sequence of protein synthesis in cells infected by human cytomegalovirus: early and late virus-induced polypeptides. *J. Virol.* **26**:686-701.
 29. Stinski, M. F., D. R. Thomsen, and J. E. Rodriguez. 1982. Synthesis of human cytomegalovirus-specified RNA and protein in interferon-treated cells at early times after infection. *J. Gen. Virol.* **60**:261-270.
 30. 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.
 31. Stinski, M. F., D. R. Thomsen, and M. W. Wathen. 1980. Structure and function of the cytomegalovirus genome. p. 72-84. *In* A. Nahmias, W. R. Dowdle, and R. F. Schinazi (ed.), *The human herpesviruses*. Elsevier/North-Holland Publishing Co., New York.
 32. Tanaka, S., M. Otsuka, S. Ihara, F. Maeda, and Y. Watanabe. 1979. Introduction of pre-early nuclear antigen(s) in HEL cells infected with human cytomegalovirus. *Microbiol. Immunol.* **23**:263-271.
 33. Thomsen, D. R., and M. F. Stinski. 1981. Cloning of the human cytomegalovirus genome as endonuclease XbaI fragments. *Gene* **16**:207-216.
 34. 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.
 35. 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.
 36. Watson, R. J., J. H. Weis, J. S. Salstrom, and L. W. Enquist. 1982. Herpes simplex virus type 1 glycoprotein D gene: nucleotide sequence and expression in *Escherichia coli*. *Nature (London)* **218**:381-384.
 37. Weller, T. H. 1971. The cytomegalovirus: ubiquitous agents with protein clinical manifestation. *N. Engl. J. Med.* **285**:267-274.
 38. Wentworth, B. B., and L. French. 1970. Plaque assay of cytomegalovirus strains of human origin. *Proc. Soc. Exp. Biol. Med.* **135**:253-259.