Characterization of a Structurally Tricistronic Gene of Human Cytomegalovirus Composed of U_S18, U_S19, and U_S20

YAW-WEN GUO¹ AND ENG-SHANG HUANG^{1,2,3,4*}

*Curriculum of Genetics,¹ Department of Medicine,² Department of Microbiology and Immunology,³
and Lineberger Comprehensive Cancer Center,⁴ University of North Carolina
at Chapel Hill, Chapel Hill, North Carolina 27599-7295*

Received 22 October 1992/Accepted 16 December 1992

A tricistronic gene mapped between 0.91 and 0.93 map units within the *Eco*RI D fragment of the human cytomegalovirus unique short region (U_S) has been cloned, sequenced, and expressed in vitro. Cloned cDNAs of 2.3, 1.8, and 1.1 kb derived from this region were isolated from a λ gt11 cDNA library made from virus-infected fibroblasts and used for this study. Two major classes of 3'-coterminal mRNAs, 2.8 and 1.1 kb, were transcribed from this region. Sequence analysis of the cDNAs and the upstream genomic DNA revealed three open reading frames (ORFs), U_S18, U_S19, and U_S20, and a common polyadenylation signal located 15 bases upstream of the poly(A) tail of both the 2.85- and 1.1-kb mRNAs. Protein structure analyses predicted the existence of multiple hydrophobic moieties, suggesting that the U_S18, U_S19, and U_S20 polypeptides were transmembrane proteins. The major transcription initiation site, determined by primer extension and S1 nuclease mapping, for the 2.85-kb transcript was located right at the first initiation codon of the U_S20 ORF. There was no typical TATA box or CAAT box upstream of the 2.85-kb mRNA cap site except for a TATAAGA sequence that was found about 210 bp downstream from the major cap site. The 1.1-kb transcript was initiated 33 bp upstream of the U_S18 translation initiation site, and an atypical TATA box sequence (GATAAGA) was found 22 bp upstream of the transcription start site. Differences in transcription kinetics and sensitivities to metabolic inhibitors suggest that they were regulated by different mechanisms; the 2.85-kb mRNA belongs to the early (β) class of transcripts, while the 1.1-kb mRNA is a late (γ) message. Subgenomic DNA segments derived from the U_S18, U_S19, and U_S20 ORFs were subcloned and overexpressed in *Escherichia coli* as fusion proteins with glutathione-S-transferase. Western immunoblot analysis with antibodies against the U_S18, U_S19, and U_S20 fusion proteins detected virus-specific polypeptides with molecular sizes of 36, 32, and 43 kDa, respectively. All three antibodies also exhibited a positive immunofluorescence reaction with human cytomegalovirus-infected cells harvested at late stages of infection.

Human cytomegalovirus (HCMV) is the fourth herpesvirus to be sequenced and has the greatest genomic complexity among the human viruses discovered to date (3). The viral genome is a double-stranded DNA 230 kb in length. More than 200 open reading frames (ORFs) can be derived from this viral genome (3). Like that of herpes simplex viruses, the HCMV genome is composed of a unique long (U_L) and a unique short (U_S) segment, each bounded by terminal repeats that permit U_L and U_S inversions to form four isomeric molecules (9, 13, 14). The HCMV U_S region is very different from that of other human herpesviruses. The 38-kb U_S region of HCMV contains at least six families of homologous ORFs, and except for the G protein-coupled receptor (GCR) family, none of these show homology to other known human herpesviruses ORFs (25). Each family contains 2 to 13 members or ORFs which occur either as tandem arrays (the U_S2, U_S6, and U_S12 families) or dispersed in the U_S region (the U_S1, U_S22, and GCR families).

The U_S2 family, composed of the U_S2 and U_S3 ORFs, produces a group of abundant immediate-early (IE) transcripts encoding a group of putative membrane-associated glycoproteins (24). The U_S6 family is composed of six members, U_S6 to U_S11, which encode a group of early transcripts (including two bicistronic mRNAs) for a group of virion-associated envelope glycoprotein complexes, designated gCII (gp47-52 [6]). The U_S22 family has 12 members,

only 4 of which, U_S22, U_S23, U_S24, and U_S26, are located within the U_S region. This family encodes a group of early viral proteins which contain N-linked glycosylation sites and C-terminal charged residues but no obvious hydrophobic transmembrane region. HCMV protein ICP22 was identified as a member of this U_S22 family. This protein is an early viral protein predominantly localized in the nucleus but also secreted into the culture medium from infected cells (15). The GCR family produces membrane-spanning glycoproteins which have sequence homology with the opsin family of cellular receptors (3). All three members of the GCR family, U_S27, U_S28, and U_S33, have seven potential membrane-spanning regions. Members of this diverse family transduce different signals in a variety of systems that have roles in vision, olfaction, memory, learning, and regulation of the circulatory system (16). Except for data obtained from computer analysis of viral DNA sequences, there is little information about the remaining two large gene families, U_S1 and U_S12. The U_S12 family has 10 members (U_S12 to U_S21), which are clustered and tandemly arranged between the U_S6 and U_S22 families in the U_S region. Members of this family are grouped together by the characteristic of having multiple hydrophobic regions within their products, which are tentatively classified as membrane-associated proteins with potential glycosylation sites. Each has up to seven highly hydrophobic putative transmembrane regions. The HCMV ORFs studied in this communication belong to this family.

Most virus genes are monocistronic, encoding only one

* Corresponding author.

polypeptide in one gene. HCMV is an exception to this rule, as polystronic messages are not uncommon in this human herpesvirus. For example, pp65 and pp71 (U_L82 and U_L83) are encoded by bicistronic genes in HCMV (3, 18). In this communication, we show that the unique short region of HCMV between 0.91 and 0.93 map units encodes a structurally tricistronic gene for the U_S18, U_S19, and U_S20 polypeptides. Transcription of polycistronic genes in prokaryotic organisms is regulated in a sequential manner. In general, the coding region closest to the promoter regulatory region produces more products. Therefore, at the beginning of this study, we assumed that regulation of the transcription of this HCMV tricistronic gene might be similar to that of a prokaryotic polycistronic gene. Northern (RNA blot) hybridization, however, revealed two major mRNAs with different transcription kinetics. Therefore, the transcriptional and translational regulations of these messages were major subjects for this study.

Sequence analysis of these three nonoverlapping ORFs (transcription is leftward, from U_S20 to U_S18) predicted three polypeptides of 342, 240, and 274 amino acid residues, respectively. To verify whether these three ORFs are functional and to confirm that this tricistronic gene is translated in HCMV-infected cells, subgenomic DNA fragments of these three ORFs were engineered and subcloned into a bacterial expression vector containing the carboxy terminus of the glutathione-S-transferase (GST) gene from *Schistosoma japonicum* under the control of a *tac* promoter, the pGEX-GST system (21). These bacterially expressed fusion proteins were used as antigens to prepare polyclonal antibodies in rabbits for use in Western immunoblot and immunofluorescence analyses.

MATERIALS AND METHODS

Virus, cells, and bacterial strains. Human embryonic lung (HEL) fibroblasts (ATCC HEL 229; passages 15 to 24) were used for propagation of the virus and all biochemical studies. HEL cells were cultured in Eagle's minimal essential medium supplemented with 10% fetal calf serum. The Towne strain of HCMV (passage 34 to 39) was propagated in HEL cells at a low multiplicity of infection (0.001 PFU/cell) for seed stocks, and the virus titer was determined by a plaque assay in HEL cells. For most experiments, HEL cells were grown to near confluence in a culture flask and infected with Towne strain HCMV at a multiplicity of infection of approximately 2 PFU/cell. Virus absorption was carried out in a CO₂ incubator at 37°C for 2 h. Virus-infected cells were harvested at the times indicated. Two bacterial strains were used in these experiments. *Escherichia coli* JM107 was used in the propagation of plasmid DNA and M13 single-stranded phage DNA template. *E. coli* Y1090 was used in the propagation of λ gt11 cDNA clones (8).

Construction and screening of cDNA libraries. Total and cytoplasmic RNAs were isolated from HCMV-infected HEL cells at 72 h postinfection (p.i.) by the guanidinium isothiocyanate method (10). A cDNA library was constructed from this RNA as described previously (8). Essentially, polyadenylated [poly(A)⁺] RNAs were purified by using oligo(dT)-cellulose column chromatography. Double-stranded cDNA was synthesized from poly(A)⁺ RNA by using oligo(dT) as the primer. The cDNAs were methylated with *Eco*RI methylase before the *Eco*RI linker was added. After fractionation through a Sepharose 4B column, cDNAs larger than 0.5 kb were inserted into the *Eco*RI site of λ gt11 and packaged in vitro with packaging extracts obtained from Promega (Mad-

ison, Wis.) as described before (8). Two cDNA libraries with 1.5×10^6 and 2×10^6 recombinants were constructed from total and cytoplasmic mRNA, respectively.

DNA labeling. The radioactive viral DNA probe was labeled by the random-primed DNA labeling method (Random Primer Labeling System U1110; Promega). To label the 5' termini of an oligonucleotide probe, 20 pmol of dephosphorylated oligonucleotides in an oligonucleotide kinase buffer with 0.1 mCi of [γ -³²P]ATP (3,000 Ci/mmol; Amersham Corp.) and 4 U of T4 polynucleotide kinase was incubated at 37°C for 30 min. After heat inactivation at 65°C for 10 min, the labeled oligonucleotides were purified by three rounds of ethanol precipitation.

Southern hybridization. Southern hybridizations were performed by standard methods (20). Prehybridization was carried out in a mixture containing 25 mM KPO₄ (pH 7.4), 6× SSC (0.9 M NaCl, 0.09 M sodium citrate), 5× Denhardt's solution (0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin), 50 μ g of denatured salmon sperm DNA per ml, and 0.1% sodium dodecyl sulfate (SDS) at 42°C for 6 h. After prehybridization, denatured ³²P-labeled cDNA probe (5 × 10⁵ cpm/ml) was added and incubated at 42°C overnight. Following hybridization, filters were washed twice in 1× SSC–0.1% SDS and twice in 0.25× SSC–0.1% SDS at room temperature for 15 min each.

Northern hybridization. Total RNAs (8 μ g per well) prepared from mock- and HCMV-infected HEL cells by guanidinium isothiocyanate cell lysis and cesium chloride equilibrium centrifugation were subjected to electrophoresis in a 1.2% agarose gel containing 6% formaldehyde and then transferred to nitrocellulose as described before (10, 20). The resulting filter was prehybridized at 42°C for 18 h and then hybridized with a random-primed or nick-translated ³²P-labeled cDNA probe (5 × 10⁵ cpm/ml) at 42°C for 24 h. After being washed, the hybridized filter was autoradiographed overnight at –70°C. The same blot was stripped of the labeled probe by incubation at 90°C in TE buffer (10 mM Tris-HCl [pH 7.5], 1 mM EDTA) for 15 min. The blot was then hybridized with the 5'-end γ -³²P-labeled antisense oligonucleotides or α -³²P-labeled cDNA probe for the yeast 28S rRNA gene. For the study of mRNA synthesis in the presence of drugs, IE RNAs were prepared from cycloheximide (50 μ g/ml)-treated HEL cells harvested at 10 h p.i. Early RNAs were prepared from phosphonoacetic acid (100 μ g/ml)-treated HEL cells harvested at 24 h p.i. Late RNAs were harvested at 60 h p.i. without drug treatment.

DNA sequencing. cDNA and genomic DNA sequencings were performed by the dideoxy chain termination method (19), with the Sequenase kit from U.S. Biochemical Corp. (Cleveland, Ohio) and Sequentide (NEG-034; New England Nuclear, Boston, Mass.). Single- and double-stranded templates were obtained from appropriate subclonings in M13mp18 and mp19. The enzymes used for the subcloning include *Pvu*II, *Rsa*I, *Pst*I, and *Xho*I. All sequences were obtained and confirmed by sequencing data taken from both strands. Eight synthetic oligonucleotides were used as sequencing primers to resolve sequence gaps and to confirm sequence ambiguities. ORF analysis was carried out with the DNA Inspector II (Textco, West Lebanon, N.H.). Two-dimensional protein structures (4) and hydrophobicity and hydrophilicity profiles of the deduced peptide were determined by using a software package provided by the Genetics Computer Group (University of Wisconsin, Madison).

Primer extension. Primer extension experiments were performed by the method described by Kingston (11), with minor modification. Four 30-bp oligonucleotides, including

primer 517 (5'-ACACAAGCGAGCGAGTGGGGCACGGT GACG-3') and primer 520 (5'-CCACTCGAGAGCCTCCAT GCGGGAGAGCAG-3'), both complementary to the 2.85-kb message, and primer 130 (5'-TGTTTCGGAACCGAGGCG GTGTCGCCATGC-3') and primer 136 (5'-TCGGCCACC AGCGCGTGGCTGCGATGGAGC-3'), both complementary to the 1.1-kb message, were synthesized on an Applied Biosystems synthesizer and end labeled with [γ - 32 P]ATP by using T4 polynucleotide kinase. Seven micrograms of total RNA and 10^4 cpm of labeled primers were mixed for 5 min at 80°C and then precipitated with alcohol. Pellets were dissolved in 20 μ l of 1 $\times$ hybridization buffer [80% formamide, 40 mM piperazine-*N,N'*-bis(2-ethanesulfonic acid) (PIPES, pH 6.4), 400 mM NaCl, 1 mM EDTA] and allowed to anneal at 30°C overnight. After ethanol precipitation, reverse transcriptase extension reactions were carried out at 42°C for 90 min. These were done in a final volume of 25 μ l containing 18 U of avian myeloblastosis virus reverse transcriptase, 50 mM Tris-HCl (pH 8.0), 5 mM MgCl₂, 5 mM dithiothreitol, 50 mM KCl, 50 U of RNasin, and 50 μ M each of the four deoxynucleoside triphosphates. After RNase digestion, phenol extraction, and ethanol precipitation, the pellets were dissolved in 4 μ l of TE and 4 μ l of gel loading buffer, heated at 80°C for 3 min, and loaded on a 6% polyacrylamide sequencing gel containing 7 M urea. The extension products were analyzed by comparison with genomic DNA sequencing reactions, which were run in parallel. The control genomic DNA was derived either from a 1,008-bp (bp 3 to 1010) *NcoI-PstI* fragment upstream of the 2.85-kb mRNA or from a 602-bp (bp 1722 to 2333) *PvuII-PvuII* fragment upstream of the 1.1-kb mRNA.

S1 nuclease mapping. The four oligonucleotides used in primer extension were also used in S1 nuclease mapping (1). Primer 517 and primer 520 were used for the 5'-end mapping of the 2.85-kb message. Primer 130 and primer 136 were used for the 5'-end mapping of the 1.1-kb messages. The procedure used was as follows: 5'-end-labeled oligonucleotides were annealed to single-stranded templates derived from the corresponding *NcoI-PstI* region (for the 2.85-kb message) or *PvuII-PvuII* region (for the 1.1-kb mRNA), extended with Klenow enzyme, and restricted to generate uniform restriction fragments. A polyacrylamide gel electrophoresis-purified probe (2×10^4 cpm) was annealed to 10 μ g of total RNA in 30 μ l of hybridization buffer (80% deionized formamide, 0.4 M NaCl, 10 mM PIPES [pH 6.5], 1 mM EDTA). This annealing mixture was placed in a 55°C water bath for 5 min and then incubated at 30°C for 16 h. To the annealed mixture, 200 μ l of S1 buffer (30 mM sodium citrate [pH 4.5], 0.25 mM NaCl, 1 mM ZnSO₄, 5% glycerol) and 100 U of S1 nuclease were added. The mixture was incubated at 16°C for 1 h. The S1 nuclease-resistant fragments were precipitated with ethanol, denatured in 4 μ l of formamide sample buffer (90% formamide, 0.1% bromophenol blue, 0.1% xylene cyanol), and electrophoresed in a 6% polyacrylamide sequencing gel with 7 M urea. The nuclease-resistant products were analyzed by comparison with genomic DNA sequencing reactions run in parallel.

Overexpression of U_S18, U_S19, and U_S20 in the pGEX-GST system. The GST gene fusion system was used to express viral DNA subclones as fusion proteins (21). To construct in-frame clones, DNA fragments were inserted into the *Bam*HI and *Eco*RI sites of one of three expression plasmids, pGEX-1N, -2T, or -3X (21), with the reading frame matched in the sense orientation to yield recombinant plasmids pGEX-1N-US18, pGEX-1N-US19, and pGEX-3X-US20. Overnight cultures of plasmid-transformed *E. coli* JM107

were diluted 1:10 with fresh medium and grown at 37°C for 1.5 h. They were subsequently induced with isopropyl- β -D-thiogalactopyranoside (IPTG) at a final concentration of 0.1 mM. After 1 to 6 h of growth, cells were pelleted and lysed in 0.25 volume of SDS sample buffer (0.006 M Tris-HCl [pH 6.8], 4% SDS, 40% glycerol, 3% dithiothreitol, 0.005% bromophenol blue). The samples were boiled and subjected to 12% polyacrylamide gel electrophoresis.

The majority of the U_S18-GST, U_S19-GST, and U_S20-GST fusion proteins were expressed as insoluble inclusion bodies. The inclusion bodies were prepared as described by Harlow and Lane (7) with modification. In brief, the induced plasmid-transformed *E. coli* JM107 cultures (pGEX-1N-US18, pGEX-1N-US19, and pGEX-3X-US20, respectively) were pelleted by centrifugation at $8,000 \times g$ for 5 min. The cell pellets were washed twice and resuspended at 1/10 their original volume in a washing buffer containing 100 mM NaCl, 1 mM EDTA, and 50 mM Tris-HCl (pH 8.0). Lysozyme was added to the suspension at a concentration of 1 mg/ml, and the suspension was incubated at room temperature for 20 min. The lysate was then sonicated for three 30-s cycles at full power with a Branson Sonifier (model 200 cell disrupter) equipped with a microtip and centrifuged at $5,000 \times g$ for 10 min. The supernatant containing soluble fusion protein was further purified on a glutathione-Sepharose column. The pellets containing inclusion bodies were washed with an ice-cold washing buffer (100 mM NaCl, 1 mM EDTA, 50 mM Tris-HCl [pH 8.0]) with 0.1% SDS, then with washing buffer containing 1% Nonidet P-40, and finally with washing buffer alone. The inclusion bodies were then further purified by 10 to 50% sucrose gradient centrifugation. Fractions containing inclusion bodies were pelleted and resuspended in phosphate-buffered saline (PBS), analyzed for purity, and used as antigens for antibody production in rabbits.

Antibody preparation. Antisera against the U_S18-GST, U_S19-GST, and U_S20-GST fusion proteins were made by individually immunizing New Zealand White rabbits. Inclusions with fusion protein aggregates were denatured in the presence of 7 M guanidine HCl solution and dialyzed against PBS with several changes of buffer solution to remove the guanidine HCl. A mixture of 100 μ g of the fusion proteins with an equal volume of Freund's complete adjuvant (Calbiochem Corp.) was injected into the rabbits subdermally at multiple sites near the hind legs. Booster injections with fusion proteins in an equal volume of Freund's incomplete adjuvant (Calbiochem Corp.) were given intradermally to the rabbits three times at three-week intervals. Ten days after the final injection, rabbits were bled, and the sera were tested for the presence of HCMV-specific antibodies by Western blot analysis and immunofluorescence staining.

Western blot and indirect immunofluorescence analysis. HEL cells were infected with HCMV as described above. At the indicated times, cells were washed twice with ice-cold PBS and then lysed with 2 $\times$ SDS sample buffer (0.125 M Tris [pH 6.8], 20% glycerol, 0.2% SDS, 0.28 M β -mercaptoethanol, 10 μ g of bromophenol blue). The lysates were sheared by passage through a 20-gauge needle four times and then boiled for 5 min. After electrophoresis, proteins were transferred to nitrocellulose paper in transfer buffer (20 mM Tris-HCl [pH 8.0], 150 mM glycine, 20% methanol). The blot was washed three times with blocking buffer (2% dry milk in PBS) at room temperature for 10 min each and then probed with antiserum against the fusion protein (diluted 1:100 in blocking buffer) at room temperature for 1 h. The antiserum was removed, and the blot was rinsed three times with a

FIG. 1. Physical map of HCMV Towne strain DNA and relative locations of the *EcoRI* D fragment, cDNAs, and various transcriptional products derived from this tricistronic gene. (A) Fractional map of the HCMV DNA genome. (B) Schematic representation of the HCMV genome, with the restriction map for *EcoRI*. (C) *EcoRI* D fragment of the HCMV genome, designated pHD4 (5). The transcription orientation of the 2.3-kb cDNA is indicated by an arrow, and the magnified map is shown underneath. (D) Transcripts and relative locations of DNA primers used for primer extension and S1 mapping. (E) Schematic diagram showing locations of the 2.85-kb and 1.1-kb mRNAs and cDNAs and their locations relative to transcription initiation sites. A 628-bp sequence was missing in the 1.1-kb cDNA compared with their mRNAs. (F) The location and arrangement of putative gene products. Data were obtained from the experiments shown in Fig. 2 through 8.

blocking buffer. Bound antigen-antibody complexes were detected by incubating the filter in buffer solution with protein A-peroxidase-labeled secondary antibody (Boehringer Mannheim Co.) at a 1:500 dilution for 2 h. After incubation, the filter was washed as above and reacted with substrate (10 mg of diaminobenzidine in 100 ml of PBS with 10 μ l of 30% H_2O_2) for 30 min.

For the indirect immunofluorescence test, antibodies were preabsorbed with HEL cellular lysates at 37°C for 1 h. HCMV-infected HEL cells grown on eight-well chamber slides (Nunc, Naperville, Ill.) were harvested at the indicated times p.i. The cells were washed with PBS three times and fixed with ice-chilled methanol-acetone (1:1) at -20°C for 15 min. The fixed cells were then blocked with 20% normal goat serum (Jackson Immuno Research Laboratory Inc.) in PBS for 30 min at room temperature. Cells were then reacted with diluted rabbit antiserum (1:50 in PBS) for 1 h. After PBS washes, biotin-labeled goat anti-rabbit immunoglobulin G secondary antibodies (Vector Laboratories Inc., Burlingame, Calif.) at a 1:200 dilution were added to the

cells and incubated at room temperature for 1 h. After a PBS wash, fluorescein-conjugated avidin D (Vector Laboratories Inc.) was added at a concentration of 12.5 μ g/100 ml. The slide was mounted with 50% glycerol in PBS and examined under a fluorescent microscope.

Nucleotide sequence accession number. The GenBank accession number for the sequence reported in this communication is L04998.

RESULTS

cDNA cloning. In our first attempt to isolate *EcoRI* D fragment-related sequences, a 2.3-kb cDNA was isolated from our λ gt11 cDNA library made from HCMV-infected fibroblasts. This 2.3-kb cDNA was mapped in the *EcoRI* D fragment. The mapping was done by Southern hybridization and further confirmed by DNA sequence analysis at both ends of the 2.3-kb cDNA, as shown below (Fig. 1). In subsequent screenings with the 2.3-kb cDNA clone as the probe, 80 positive clones were isolated from 4×10^4 recom-

FIG. 2. Transcription analysis with the 2.3-kb cDNA and strand-specific oligonucleotide probes. Total RNA (8 μ g per well) was prepared from mock- and HCMV-infected HEL cells. Cytoplasmic, nuclear, and total RNAs were purified from HEL cells at 60 h p.i. Cytoplasmic and nuclear RNAs were further fractionated by oligo(dT)-cellulose column chromatography to obtain poly(A)⁺ mRNA. Sizes are indicated on the left, and arrows on the right indicate the messages detected and their sizes (in kilobases). (A) Probed with a strand-specific probe (primer 136) close to the 3' end. Two messages, 2.85 and 1.1 kb, were observed. (B) Probed with a strand-specific probe (primer 520) close to the 5' end of the 2.85-kb mRNA. This probe detected only one message of 2.85 kb (see Fig. 1 for relative locations of oligonucleotide probes used in hybridization studies). (C) Probed with labeled 2.3-kb cDNA. This probe detected 2.8-kb, 1.1-kb, and a minor 1.8-kb message.

binant phages. Twenty positive clones were analyzed by restriction enzyme mapping and DNA sequencing. The clones were categorized into three groups, with a cDNA insert of 2.3, 1.8, and 1.1 kb, respectively. Three cDNA clones were further cloned into M13mp18 for 3'-end and 5'-end analysis. The results of 3'-end analysis reveal that the 2.3-kb, 1.8-kb, and 1.1-kb cDNAs have an identical polyadenylation signal (AATAAA) and a 14-bp downstream sequence (GCTTTTTCACGCTT) and are 3' coterminal.

Transcription pattern. To analyze the transcripts corresponding to the 2.3-kb cDNA, RNA samples isolated at 60 h after infection were probed with the nick-translated and random-primed labeled 2.3-kb cDNA fragment. Two major groups of mRNAs of 2.85 and 1.1 kb were detected (Fig. 2C). The detection of a minor band of 1.8 kb was not consistent and varied from batch to batch of RNA preparations. Hybridization with an antisense oligonucleotide probe (primer 520) derived from the 5' end of the 2.3-kb cDNA detected the 2.85-kb message but not the 1.1-kb transcript (Fig. 2B). When an antisense oligonucleotide probe (primer 136) derived from the 3' end of the 2.3-kb cDNA was used as the probe, both the 2.85-kb and the 1.1-kb messages were detected (Fig. 2A). This result suggests that the 1.1-kb transcript was transcribed within the 3'-end portion of the 2.85-kb message. The DNA sequence between the poly(A) signal sequence and the 5' end of the 1.1-kb cDNA was equivalent to the length detected in the 1.1-kb RNA transcript. These data suggest that the 2.85-kb and 1.1-kb mRNAs are 3' coterminal and that the 1.1-kb transcript is unspliced. These 3'-coterminal transcripts were further confirmed by 3'-end sequencing of the cDNA as well as by the method of the rapid amplification of the cDNA ends (RACE) (data not shown).

5'-end mapping by primer extension and S1 nuclease mapping. Two antisense oligonucleotides (primer 517 and primer 520), complementary to two areas close to the 5' end of the

2.85-kb mRNA, were used in the primer extension studies (data not shown). Both of these primers showed the same major transcription initiation site right at the translation initiation codon of the U_S20 ORF, even though these two primers are 80 bp apart. These two oligonucleotides were also used in S1 nuclease mapping. The major transcription initiation site for the 2.85-kb mRNA was similar to the primer extension site (Fig. 3A). However, when primer 517 was used for S1 nuclease mapping, a minor band also appeared about 243 bp downstream from the major initiation site indicated above, which is approximately 26 bp downstream from a putative TATA box and approximately 21 bp upstream of a potential second initiation codon for the U_S20 ORF (data not shown).

Another set of antisense oligonucleotides (primer 130 and primer 136), complementary to the 5'-end region of the 1.1-kb mRNA, were used in the primer extension studies to detect the transcription initiation site of the 1.1-kb mRNA. This set of primers picked up four transcription initiation sites with equal intensity within a 7-base range (data not shown). S1 nuclease mapping with the same oligonucleotides also detected a major transcription initiation site in the same region (Fig. 3B). This site was 33 bp upstream of the U_S18 translation start site. The 5' end of this transcript has a TATA-like box, GATAAGA (bp 2115 to 2121), 22 to 28 bp upstream of the transcription initiation site. A CAAT-like box, CGTCAATCA (bp 2085 to 2092), was 50 to 58 bp upstream of the transcription initiation site. The transcription initiation site was 33 bp upstream of the U_S18 translation initiation codon. The schematic diagrams in Fig. 1E and F outline the transcription initiation sites found and their relative locations in the cDNAs. The details are indicated in Fig. 6.

Attempts to map the 5' end of the minor 1.8-kb mRNA were not successful when primer 156 (5'-GTAGCAGCAG GAACA-3') was used for primer extension and S1 nuclease

FIG. 3. S1 nuclease mapping of the 2.85-kb and 1.1-kb messages. The 5'-end-labeled oligonucleotides (primer 520 for the 2.85-kb mRNA, and primer 136 for the 1.1-kb mRNA) were annealed to corresponding sense single-stranded DNA templates for the synthesis of ^{32}P -labeled antisense probes. The preparation of radioactive single-stranded DNA fragments and the procedure for S1 mapping are described in Materials and Methods. The dideoxynucleotides (T, G, C, and A) used in the reactions are shown above the lanes. The DNA sequencing data from parallel experiments are shown to the left of each panel; possible transcription initiation sites are marked with asterisks (*). (A) S1 mapping of the 2.85-kb mRNA. Late RNA (L) isolated at 60 h p.i. was used in this analysis. (B) S1 mapping of the 1.1-kb mRNA. RNAs isolated at 24 (E) and 60 (L) h p.i. were used in this study. The transcription initiation site is indicated by an arrow and marked by an asterisk (*) in the sequence.

mapping, perhaps because of the low transcription level of this message.

3'-end mapping by cDNA sequencing and RACE. In order to further verify that these mRNAs were 3'-end coterminal, two different methods were used: sequencing different batches of cDNA, and RACE. Twenty positive cDNA clones were sequenced at both ends, and the sequence results indicated that these three groups of cDNA (2.85, 1.8, and 1.1 kb) use the same polyadenylation signal (AATAAA), which is 15 bp upstream of the poly(A) tail. On the other hand, these three cDNAs have different lengths of poly(A) tail. The results of the RACE experiment revealed a single band of 290 bp, the predicted size. Subsequent sequence analyses of this RACE-amplified DNA fragment further verified the results (data not shown). Both the cDNA sequence and RACE amplification provide supporting data for 3'-end coterminal of the major 2.85-kb and 1.1-kb mRNAs and the minor 1.8-kb mRNA.

Analysis of transcription patterns. Total RNAs collected at different time points (0, 4, 8, 12, 24, 48, 60, and 72 h p.i.) were electrophoresed through a 1.2% agarose gel containing 6% formaldehyde and analyzed by Northern hybridization. The transcription patterns were derived from the use of two antisense oligonucleotide probes, one from the 5' end of the 2.85-kb mRNA and another from the 3' end of both transcripts. The 5'-end probe detected only the 2.85-kb mRNA (Fig. 4), while the 3'-end probe detected both the 2.85-kb and 1.1-kb messages (Fig. 4B). RNA levels were quantitated by densitometer scanning and normalized to that of the internal control, rRNA. Transcription of the 2.85-kb mRNA first appeared at 12 h p.i., reached a maximum level at 48 h p.i.,

and then gradually decreased. The 1.1-kb mRNA appeared between 24 and 48 h p.i. and remained at a constant level until 72 h p.i. (Fig. 4C). The increase in transcription at 96 h p.i. was probably due to virus reinfection.

Transcription pattern in the presence of drugs. To further characterize these two mRNAs, the transcription patterns were studied in the presence of protein and viral DNA synthesis inhibitors. In the presence of the protein synthesis inhibitor cycloheximide (50 $\mu\text{g}/\text{ml}$), both the 2.85-kb and 1.1-kb transcripts became undetectable (Fig. 5). These data suggested that the 2.85-kb and 1.1-kb messages might require HCMV IE gene products for the initiation of their transcription. Therefore, we conclude that these two messages were not IE transcripts.

In the presence of phosphonoacetic acid (100 $\mu\text{g}/\text{ml}$), a viral DNA synthesis inhibitor, the 2.85-kb mRNA accumulated to a level comparable to that in the control (without drug) at 60 h p.i. The transcriptional level of the 1.1-kb message in the presence of phosphonoacetic acid, however, decreased to less than one-fifth of control levels (Fig. 5). These data suggest that the 2.85-kb mRNA is actually an early, or β -type, message of HCMV, which did not require viral DNA replication for expression. The 1.1-kb mRNA was expressed at late times during infection (24 to 48 h p.i.) and required viral DNA replication for maximum expression; it is clearly a late, or γ -type, message. The results show that these two transcripts were expressed at different times after infection and were regulated by different mechanisms.

Computer analysis of DNA sequence. A 3,158-bp DNA sequence (Fig. 6) was assembled from the sequencing data determined from the analysis of both strands of various

FIG. 4. Transcription pattern analysis of the 2.85-kb and 1.1-kb mRNAs by Northern blot hybridization. Total RNA was prepared from mock- and HCMV-infected HEL cells at various times (0, 4, 8, 12, 24, 48, 60, 72, and 96 h p.i.) as indicated above each lane. Eight micrograms of total RNA was subjected to electrophoresis through a 1.2% agarose gel containing 6% formaldehyde and then transferred to nitrocellulose paper. (A) The resulting filter was probed with an antisense oligonucleotide primer located at the 5' end of the 2.85-kb mRNA (see Fig. 1D, primer 520). (B) The filter was reprobed with an antisense oligonucleotide primer (primer 136) located at the 3' end of both mRNAs. (C) The blot was probed with yeast rRNA as an internal control. The arrows on the right indicate the messages detected, and the arrows on the left indicate the positions of size markers (in kilobases).

restricted genomic or cDNA fragments with an M13mp18 subcloning system and primer extension approach. This DNA sequence includes a 2,290-bp cDNA fragment and a 1,008-bp genomic region upstream of the 5' end of the 2.3-kb cDNA. These two DNA fragments overlap tail to head, with 140 bp at the 3' end of the 1,008-bp upstream genomic region and at the 5' end of the 2,290-bp cDNA sequence. ORF analyses of the sequences shown in Fig. 1F and Fig. 6 were done with the DNA Inspector II program. Three independent nonoverlapping ORFs of Towne strain HCMV, equivalent to U_{S18} , U_{S19} , and U_{S20} of AD169 strain HCMV, were derived. A typical polyadenylation signal (AATAAA, bp 3138 to 3143) was located 15 bp upstream of the common poly(A) tail. All of these ORFs have a typical consensus Kozak sequence (12) around the translation initiation codon.

Sequence comparison with AD169. The sequences of the Towne and AD169 strains of HCMV revealed 99.3% homology. The Towne strain is an HCMV strain isolated from the urine of a congenitally infected baby. The AD169 strain is a prototype of HCMV isolated from adrenal glands almost 25 years ago that has been subcultured in vitro for more than 280 passages (17). The sequence comparison between these two strains revealed 20 base mismatches (Fig. 6). No insertions or deletions were found. Nine of the 20 mismatched bases are located within noncoding regions. Ten of the 11 other single-base mismatches do not result in any amino acid sequence alteration, and the other mismatch changes amino acid residue 238 of U_{S20} from alanine in AD169 to valine in

FIG. 5. Transcription in the presence of protein and viral DNA synthesis inhibitors. Eight micrograms of total RNA from mock- and HCMV-infected cells treated with cycloheximide (CH; 50 μ g/ml) for 10 h or phosphonoacetic acid (PAA; 100 μ g/ml) for 60 h or untreated cells at 60 h p.i. (LATE) was applied to each lane for Northern blot RNA analysis. The resulting filter was probed with the 2.3-kb cDNA probe. The arrows on the right indicate messages detected, and the arrows on the left indicate the migration of size markers.

the Towne strain. The results of this experiment show that the products of this tricistronic gene are very well conserved.

Two-dimensional protein structure analysis. Three ORFs (U_{S18} , U_{S19} , and U_{S20}) were determined from DNA sequence analysis. Two-dimensional structure analysis of the putative U_{S18} polypeptide reveals seven hydrophobic segments. Each is more than 20 amino acids long and may span the cell membrane. One strong antigenic site was found within one of the two potential glycosylation sites, where the hydrophilic moiety is also obvious. The N-terminal end of this peptide is hydrophilic, whereas the C-terminal end is more hydrophobic. This ORF has 40 leucine residues among 274 amino acids. The putative U_{S19} polypeptide has only six hydrophobic segments, but one of these is a long stretch of 50 hydrophobic amino acids. No potential glycosylation sites were found in the U_{S19} polypeptide, but two highly antigenic clusters with hydrophilic moieties were found between the first and the second and between the second and third hydrophobic domains from the N-terminal side of this peptide. Both the N- and C-terminal ends of this peptide are hydrophobic. The putative peptide contains 52 leucines among 240 amino acid residues. Two-dimensional structure analysis of U_{S20} revealed seven hydrophobic segments, two potential glycosylation sites, and two highly hydrophilic as well as highly antigenic moieties located around amino acid residues 50 and 250. The putative U_{S20} peptide contains 47 leucines among 342 amino acid residues. The hydrophobicity profiles of these peptides show characteristics of membrane-associated proteins: long stretches of hydrophobic amino acids connected by a length of random coil or hydrophilic amino acids.

Western blot analysis and indirect immunofluorescence tests. Antibodies to the bacterially expressed fusion proteins were able to react with virus-specific products in Western blot analyses. The results obtained with HCMV-infected cell lysates showed that anti- U_{S18} -GST antibodies detected a 36-kDa virus-specific band (Fig. 7A). This protein appeared at 34 h p.i. and coincided with the late transcription of the 1.1-kb mRNA, a γ -class transcript. The U_{S19} -GST antibody detected a 32-kDa virus-specific peptide which appeared at a

Towne Genomic	NcoI CATGGTGGCCACAGTAAAGCTATTTTGTTCCTTAGGCTCGCTACGTATCTGTTTGCCCGGACGTGGAAATACAGCGTTCCTATTGACGGGCGCATCCACGCTGATTCTATTACTGTT	120
AD169 Genomic	G C C	
Towne Genomic	CGCCGATTTTCTCTTTTCCGAGGCGGTAAAGTAGATCTGGTGTATGATAGCCGGTTTGGCCGTCATCGTGAACCTCGGTGGTGGTGGATACCGAGGACATCTGCACGACATCGAGTA	240
AD169 Genomic		
Towne Peptide	→ CAP 2.85 Kb mRNA	
Towne Genomic	SmaI TGAGAGTTATATCCCGGCGCGCTCTGCTGTACATGGACCTCATGTACCTCTTTGTCTCCGTGCTCTACTTCATGCGCTCCGAGCCCGGACAGCGCCACACTGCTCAGACGAGGTGCG	360
AD169 Genomic		
Towne Peptide	R V I S R A R S A C T W T S C T S L S P C S T S C P P S P A A P T L L R R R S L	
Towne Genomic	IR5 TGCCACAGCAGCGCGGTGCCCCAGTTCTGCTCTCTAATCGAGGTGTCGCGCGGTGACCACTGTCACCGTCCCTCGCTCGCTGTGTTATAAAGTAAGGGTAGGTGCGCCGACGCGC	480
AD169 Genomic		
Towne Peptide	P Q Q R R R P S S S P N R R V R G V T T S P C P T R S L V Y K R R V G A P Q R L	
Towne Genomic	minor TTTGTGCGGAGACCGTGCACCATGTCAGGCGCAGAGGGCTAAAGCGGTGCTGCTCTCCGATGAGGGCTCTCGAGTGGTTCAAAAAGTTACCGTATGGCTGCGGTGTACGCCATCT	600
AD169 Genomic	2 nd transl. init.	
Towne Peptide	L C R D R R H M Q A Q E A N A L L L S R M E A L E W F K K F T V W L R V Y A I F	
AD169 Peptide		
Towne Genomic	TCATCTTTCAGCTGGCTTTCAGCTTCGGCTTGGGAAGCGTTTGTGGTGGGTTCCCAAAACGCAACTTTTGGCTGAGAACTACAGCTTCTTCTCACCGTGTCTGTCGCCATCG	720
AD169 Genomic		
Towne Peptide	I F Q L A F S F G L G S V F W L G F P Q N R N F C V E N Y S F F L T V L V P I V	
AD169 Peptide		
Towne Genomic	TCTGCATGTTTCATCAGCTACGTTGGGCAACGAACCCCTAGTAACGCCACGGTGTCTTTCATCTATCTGTTGGCCAAAGCGCTGACGGCGGCCATCTTCCAGATGTGCTCTGAAAGCC	840
AD169 Genomic		
Towne Peptide	C M F I T Y T L G N E H P S N A T V L F I Y L L A N S L T A A I F Q M C S E S R	
AD169 Peptide		
Towne Genomic	cDNA 2.3 kb GCGTACTAGTAGTTCCTACGTGATGACCTTGGCGTGTGTTATCTCTTTACGGGCTGGCGTTTCTAGGTGGCGGTGACCGAGCTCGCTGGAAAATGCATCAGCTGCGTCTACGTGGTGA	960
AD169 Genomic		
Towne Peptide	V L V G S Y V M T L A L F I S F T G L A F L G G R D R R R W K C I S C V Y V V M	
AD169 Peptide		
Towne Genomic	PstI TGCTGCTTTTCGTTCTCTACGCTCGCTCTGCTAAGCGACGGCGATTGGCTGCAG	1080
AD169 Genomic		
Towne Peptide	L L S F L T L A L L S D A D W L Q K I V V T L C A F S I S F F L G I L A Y D S L	
AD169 Peptide		
Towne cDNA	TCATGGTATCTTTTCTGCCCCAATGCAATGCAATCCGTACGCGCTCTGCTCTACCTGGACAGCATGGCCATCTTCTCAGCTTGTGCTCATGCTCTCGGGTCCCGTTGGATTAT	1200
AD169 Genomic		
Towne Peptide	M V I F F C P P N Q C I R H A V C L Y L D S M A I F L T L L M L S G P R W I S	
AD169 Peptide		
Towne cDNA	TATATAAATCCCGGCCCGTTTTCACAGCAGCGCGCTTGTGCGCATGCTTT	1320
AD169 Genomic		
Towne Peptide	L S D G V P L D N G T L T A A S T T G K S *	
AD169 Peptide		
Towne cDNA	CATGTCTGTCCTAGAAATGACCGTCCGAGGAGTAGTCCCTACCTAGAGAGGCTGGCGTATGCTGCGGCGCTCGCTCTCTGCTGCTTCCAGCTTACCGCCACGCTCGCCCTCAGC	1440
AD169 Genomic		
Towne Peptide	H V V P L E W T V E E V V P Y L E R L A V W L R A S V L V A F Q L T A T V A L S	
AD169 Peptide		
Towne cDNA	GTGCTCAGCTGGTGGCTGATGCCGCCACCGTGGCGAGCTATGCGAGCGGGACGGGACGACGCCACCCCGCTTTCTCATCTCAGCTCTGCTGTCGCCGTGGCGCTGTCCTG	1560
AD169 Genomic		
Towne Peptide	V L S W W L M P P P V A E L C E R G R D D D P P P L S H L S L V V P V G C L F L	
AD169 Peptide		
Towne cDNA	CTGCTACGCGGTCCGTGATCGACCGTTGTCCCGCAAACTCCCACTCTGTGTGGCTTACTGCTGCGCAGCGCTAGCCCTTCTGACCTTGTGATGTGCCAGCGCTCGCGCAGGCC	1680
AD169 Genomic		
Towne Peptide	L L R G P S I D R C P R K L P L L L A Y C L P H A L A F L T L L M C Q P S P Q A	
AD169 Peptide		
Towne cDNA	TTTGTGCGCGCGCTCTTCTGCGCTGGCGGTAGATCTCAGCTGCGCTGGGGCCCTCGCTTTTGGGTGCGATCTGGCGGCTCTCCGCGGCTGTGGCTACCGTGGTCTCTGCTCTG	1800
AD169 Genomic		
Towne Peptide	F V G A A L L A L A V D L S C L G A S L L G C D P G A S L R R L W L P S V L S L	
AD169 Peptide		
Towne cDNA	Sp1/IR7 CTCTGCGCTACGCGCTGGACTTTGGCTGCTGCGGCGCGCGCGCGCTTTCTTCGGTTTACAGCGACGCGCTGTGACGCTGACGCTGATGCTAATTCACGATCTGTCTGATC	1920
AD169 Genomic		
Towne Peptide	L C A T A L G L W L L R A A A P F F L G L H A T T L L T V T L M L I H D L S L I	
AD169 Peptide

FIG. 6. Nucleotide sequence of the structurally tricistronic gene encoding the 2.85-kb message and the putative polypeptides U_S18, U_S19, and U_S20. The 2,290-bp cDNA (2.3-kb cDNA) and 1,008 bp of genomic sequence located upstream of this cDNA were combined and are shown in this figure with an overlap of 140 bp. Sequence mismatches between the Towne and AD169 strains are noted in the AD169 genomic sequence. The polypeptide sequences are indicated below the DNA sequence. The transcription initiation sites identified by primer extension and S1 nuclease mapping are indicated by solid bent arrows. The minor cap site for the 2.85-kb mRNA is also indicated. The beginnings of the two cDNAs used in this study are also marked with open bent arrows. Upstream sequence analysis of the 2.85-kb and 1.1-kb mRNAs revealed several indirect repeats (IR) and direct repeats (DR), which are numbered sequentially. Solid horizontal arrows indicate the IR and open horizontal arrows indicate the DR upstream of U_S18 and U_S20. The polyadenylation signal sequence and potential glycosylation sites are underlined. The possible second translation initiation codon for U_S20 is indicated by a bent arrow.

very low level at 18 h p.i. and was clearly demonstrated at 34 h p.i. (Fig. 7B). The anti-U_s20-GST antibody detected a major 43-kDa virus-specific product at 18 h p.i. (Fig. 7C). This translation result also matched the 2.85-kb mRNA transcription data. A minor band of 36 kDa was found at a late time (34 h) in virus-infected lysates. This 36-kDa protein might result from the use of the second translation initiation codon at later stages of virus infection.

Antibodies generated in a rabbit against the bacterially expressed fusion proteins (U_S18-, U_S19-, and U_S20-GST) gave positive immunofluorescence staining in HCMV-infected cells. This reaction was not found in mock-infected cells or at 12 h after HCMV infection. The anti-U_S18-GST antibody detected positive immunofluorescence only at 48 and 72 h p.i. (Fig. 8). This result suggests that this virus-

specific product was translated at 48 h, which coincided with the transcription data for the 1.1-kb mRNA. The U_S19-GST antibodies detected positive immunofluorescence at 24, 48, and 72 h p.i., with patterns similar to that with the anti-U_S20-GST antibody (data not shown). This result indicates that the U_S19 virus-specific products appear earlier than the U_S18 virus-specific products. The U_S20-GST antibodies detected positive immunofluorescent reactions at 24, 48, and 72 h p.i., like the U_S19-GST antibodies, but the reaction was stronger than it was with the U_S19-GST antibodies at 24 h p.i. (data not shown). This result indicates that the U_S20 virus-specific products appeared earlier than the U_S18 and U_S19 virus-specific products. The early translation of the U_S20 virus-specific products also matches the 2.85-kb mRNA transcription data.

DISCUSSION

Genomic repetition is a common characteristic of the human herpesvirus group, but gene clusters in the U_S region are a unique characteristic in HCMV and herpes simplex virus (3, 24). Before the HCMV genome had been completely sequenced, we originally speculated that the Towne strain *EcoRI* D fragment located within the U_S region encoded the HCMV counterparts of herpes simplex virus gD and gC. In our first attempt to screen for gD-related cDNA, a 2.3-kb cDNA was isolated from a λ gt11 cDNA library. This cDNA has been used as a tool to characterize the corresponding gene. To our surprise, instead of isolating gC or gD homologs, we identified the U_{S18} , U_{S19} , and U_{S20} tricistronic gene.

In the study of the transcription products related to the 2.3-kb cDNA, two major messages (a 2.85-kb mRNA and a 1.1-kb mRNA) and a minor 1.8-kb mRNA were identified. Detection of the minor message by the 2.3-kb cDNA probe was not consistent, varying from one preparation to another. From DNA sequencing and ORF analysis, the minor 1.8-kb product might be derived from the second ORF, which does not have a strong classical promoter-enhancer sequence and might require special cellular or viral products for maximal expression. The expression of this group of messages might be under very tight control. The presence of the 1.8-kb cDNA reflects the existence of this weak message. Meanwhile, the detection of 2.85-kb mRNA suggested that the 2.3-kb cDNA is not a full-length cDNA, since approximately 628 bp of the 5' upstream sequence was missing. Also, the 1.1-kb mRNA could represent the 1.1-kb cDNA.

As determined by primer extension, the 2.85-kb mRNA has two transcription initiation sites; one is right at the first translation initiation codon of the U_{S20} ORF, and the other is 243 bp downstream from the first site. There is a typical TATA sequence (TATAAGA) 21 bp upstream of the first transcription initiation site, but no typical CAAT sequence was observed. There were two GC boxes and a direct repeat located about 40 to 80 bp upstream of this typical TATA box. On the other hand, no typical TATA [TATA(A/T)A(A/T)] box or CAAT [GG(C/T)CAATCT] box upstream of the second cap site was found. It is unclear why the major transcription initiation site for the 2.85-kb mRNA is located right at the first translation initiation codon of U_{S20} . In general, leader sequences are 20 to 100 nucleotides long on most vertebrate mRNAs except for mRNAs derived from proto-oncogenes, which have very long leader sequences (12). In addition, the ATG codon is not well recognized when it occurs close to the cap (12). Therefore, the translation regulation of the U_{S20} ORF needs further investigation before any conclusions can be made.

The transcription initiation site of the 1.1-kb mRNA was 33 bp upstream of the translation initiation site. In the region upstream of this transcript, there is an atypical TATA box sequence (GATAAGA) and an atypical CAAT box sequence (CGTCAATCA) within 22 to 28 bp and 50 to 58 bp upstream of the cap site, respectively. Promoter activity upstream of the 2.85-kb mRNA was detected during HCMV infection (data not shown). The promoter activity of the region upstream of the 1.1-kb mRNA is still under investigation. With regard to the U_{S19} ORF, there are typical TATA and GC boxes upstream of the translation initiation site and a minor 1.8-kb mRNA which could correspond to a transcript initiating near these potential transcription start sites. However, this message was not clearly demonstrated by S1 mapping analysis.

FIG. 7. Western blot analysis of HCMV-infected HEL cells with anti- U_{S18} -GST, anti- U_{S19} -GST, and anti- U_{S20} -GST antibodies. DNA fragments corresponding to ORFs U_{S18} , U_{S19} , and U_{S20} were isolated by restriction digestion and then subcloned into the pGEX-GST gene fusion system for expression in *E. coli* (21). The fusion proteins obtained were used to generate antibodies in rabbits for immunological analysis of these peptides. HCMV-infected HEL cells (2×10^5 cells) harvested at the times indicated above the lanes (hours p.i.) were used for Western blot analysis. Arrows on the right indicate virus-specific bands, and arrows on the left show the positions of size markers (in kilodaltons). Antibodies derived from the (A) U_{S18} , (B) U_{S19} , and (C) U_{S20} fusion proteins detected virus-specific bands of 36, 32, and 43 kDa, respectively.

Sense-strand oligonucleotides targeting either the 5' end or the 3' end of the 2.85-kb mRNA probes did not detect any antisense messages transcribed from this genomic map region (data not shown). An antisense oligonucleotide (primer 136) targeting the 3' end of the 2.3-kb cDNA could detect both the 2.85-kb and 1.1-kb messages. This result suggested that these two messages were transcribed in the same direction. The transcription pattern of this structurally tricistronic gene is very similar to that of HCMV tegument proteins pp65 and pp71, in which two messages are transcribed from the corresponding genomic region. A large 4.0-kb mRNA and a small unspliced 1.9-kb mRNA are transcribed with a 3'-coterminal end. The pp65 matrix tegument protein is translated from a 4.0-kb mRNA, and the pp71 high-matrix tegument is made from a 1.9-kb mRNA (18).

Transcription kinetics were studied with RNAs harvested at designated time points and after appropriate drug treatment. These studies revealed that the 2.85-kb and 1.1-kb mRNAs belong to two different classes of transcript. The 2.85-kb message is an early or β -type mRNA which does not require virus DNA replication for maximal expression. The 1.1-kb mRNA, expressed late (48 h p.i.) after virus infection and requiring viral DNA replication for maximum expression, is a late or γ -type message. This differential expression occurs even though these transcripts are derived from the same genomic map region and overlap at the 3' end. This is common in the transcriptional regulation of early and late HCMV mRNAs, in which early and late mRNAs are translated after immediate-early mRNAs (2, 5, 22, 23).

FIG. 8. Indirect immunofluorescence study with antibodies against HCMV fusion proteins. The rabbit antibodies used in this study were preabsorbed with HEL cellular lysates. (Inset a) Mock-infected HEL cells. (A to D) HCMV-infected HEL cells at 12, 24, 48, and 72 h p.i., respectively, were reacted first with the antibody against the fusion protein and then with biotin-conjugated goat anti-rabbit immunoglobulin G antibody and stained with fluorescein-labeled avidin. Arrows indicate positive staining. Only results for U_S18 are shown. Panels a, A, and B were deliberately overexposed to show cell morphology.

The tricistronic 2.85-kb mRNA was transcribed after 12 h p.i. and reached the maximum transcription level at 24 h p.i. The first 5'-end translational product of the 2.85-kb mRNA, the U_S20 polypeptide, was detected at 18 h p.i., and its level increased slightly over the infection period. The U_S19 polypeptide could be detected at a very low level at 18 h p.i., but higher levels of U_S19 were detected between 24 and 36 h p.i. by immunofluorescence and Western blot analyses. Transcription of the 2.85-kb mRNA decreased after it reached its maximum level. In contrast, translation of this message increased continuously over an extended period. This result probably reflects the stability of the translation products or an increase in the translation efficiency of the 2.85-kb mRNA at late stages of infection.

As pointed out above, a very minor 1.8-kb mRNA species was inconsistently detected when the 2.3-kb cDNA was used as a probe in Northern hybridization. Several attempts to map transcription initiation by S1 mapping and primer extension failed. These results suggest that an extremely low level of transcription, if any, of the 1.8-kb mRNA may occur at the later stages of infection. This speculation is also based upon the isolation of a 1.8-kb cDNA, derived from this region, from the cDNA library of HCMV-infected cells. At this time, we do not have definitive data to conclude whether the 32-kDa U_S19 peptide is derived from the 2.85-kb or the 1.8-kb transcript. The 32-kDa U_S19 peptide could barely be detected at 18 h p.i. but was detected at a level comparable

to that of the 43-kDa U_S20 peptide. If the stability and the levels of translation efficiency are comparable between these transcripts, then the level of mRNA for U_S19 should be comparable to that for U_S20. Immunofluorescent staining of the U_S19 peptide indicated that it appeared as early as the U_S20 peptide, at 24 h p.i., but was less abundant than U_S20. From these data, we believe that the U_S19 ORF may be translated from the 2.85-kb mRNA. Expression of this ORF may require a virus-induced factor(s) for maximal activity. The monocistronic 1.1-kb mRNA was translated at a later time p.i., and its translation subsequently remained constant throughout the infection. This correlated well with the late translation of the U_S18 virus-specific polypeptide.

Antibodies prepared against the three peptides, expressed as fusion proteins, detected virus-specific proteins in HCMV-infected cell lysates. The differences between the sizes determined by Western blot analysis and the sizes predicted from sequence analysis may be due to posttranslational modification. These antibodies gave positive immunofluorescence when tested during the late stages of HCMV infection in HEL cells, but the virus neutralization assay resulted in low neutralization activities. The low neutralization activity of these antibodies could be due to inappropriate folding of the fusion proteins or the lack of posttranslational modification of these virus-specific products expressed in *E. coli*. It is also possible that these proteins are not essential for infectivity. Expression of these ORFs with

a baculovirus vector in insect cells is under way for further functional and antigenic studies.

ACKNOWLEDGMENTS

We thank Timothy Kowalik, Bret Wing, and David Y. Huang for useful discussions and Shu-mei Huang for technical assistance.

This investigation was supported in part by Public Health Service grants AI-12717 and CA-21773 from the National Institutes of Health.

REFERENCES

1. 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.
2. Chang, C. P., D. H. Vesole, J. Nelson, M. B. A. Oldstone, and M. F. Stinski. 1989. Identification and expression of a human cytomegalovirus early glycoprotein. *J. Virol.* 63:3330-3337.
3. Chee, M. S., A. T. Bankier, S. Beck, R. Bohni, C. M. Brown, R. Cerny, T. Horsnell, C. A. Hutchison III, T. Kouzarides, J. A. Martignetti, E. Preddie, S. C. Satchwell, P. Tomlinson, K. M. Weston, and B. G. Barrell. 1990. Analysis of the protein-coding content of the sequence of HCMV strain AD169. *Curr. Top. Microbiol. Immunol.* 154:125-169.
4. Chou, P. Y., and G. D. Fasman. 1974. Prediction of proteins conformation. *Biochemistry* 13:222-245.
5. Davis, M. G., and E.-S. Huang. 1985. Nucleotide sequence of a human cytomegalovirus DNA fragment encoding a 67-kilodalton phosphorylated viral protein. *J. Virol.* 56:7-11.
6. Gretch, D. R., B. Karl, R. Gehr, and M. F. Stinski. 1988. A multigene family encodes the human cytomegalovirus glycoprotein complex gCII (gp47-52 complex). *J. Virol.* 62:1956-1962.
7. Harlow, E., and D. Lane. 1988. Antibodies—a laboratory manual, p. 90-91. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
8. He, Y. S., L. Xu, and E.-S. Huang. 1992. Characterization of human cytomegalovirus UL84 early gene and identification of its putative protein product. *J. Virol.* 66:1098-1108.
9. Kilpatrick, B. A., and E.-S. Huang. 1977. Human cytomegalovirus genome: partial denaturation map and organization of genome sequences. *J. Virol.* 24:261-276.
10. Kingston, R. E. 1987. Guanidinium method for total RNA preparation, p. 4.2.1-4.5.3. In F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), *Current protocols in molecular biology*. John Wiley & Sons, New York.
11. Kingston, R. E. 1987. Primer extension, p. 4.8.1-4.8.3. In F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), *Current protocols in molecular biology*. John Wiley & Sons, New York.
12. Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 15:8125-8132.
13. LaFemina, R., and G. S. Hayward. 1980. Structural organization of the DNA molecules from human cytomegalovirus, p. 39-55. In R. Jaenisch, B. Fields, and C. F. Fox (ed.), *Animal virus genetics*. Academic Press, New York.
14. LaFemina, R., and G. S. Hayward. 1983. Replicative forms of human cytomegalovirus DNA with joined termini are found in permissively infected human cells but not in non-permissive Balb/c-3T3 mouse cells. *J. Gen. Virol.* 64:373-389.
15. Mocarski, E. S., L. Pereira, and L. McCormick. 1988. Human cytomegalovirus ICP22, the product of the HWLF1 reading frame, is an early nuclear protein that is released from cells. *J. Gen. Virol.* 69:2613-2621.
16. Nathans, J., and D. S. Hogness. 1983. Isolation, sequence analysis, and intron-exon arrangement of the gene encoding bovine rhodopsin. *Cell* 34:807-814.
17. Rowe, W. P., J. W. Hartley, S. W. Waterman, H. C. Turner, and R. J. Huebner. 1956. Cytopathogenic agent resembling human salivary gland virus recovered from tissue culture of human adenoids. *Proc. Soc. Exp. Biol. Med.* 92:418-424.
18. 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.
19. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 75:5463-5467.
20. Seldon, R. F. 1987. Analysis of DNA sequences by blotting and hybridization, p. 2.9.1-2.9.9. In F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), *Current protocols in molecular biology*. John Wiley & Sons, New York.
21. Smith, D. B., and K. S. Johnson. 1988. Single-step purification of polypeptides expressed in *Escherichia coli* as fusion with glutathione S-transferase. *Gene* 67:31-40.
22. 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.
23. 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.
24. 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.
25. Weston, K., and B. G. Barrell. 1986. Sequence of the short unique region, short repeats and part of the long repeats of human cytomegalovirus. *J. Mol. Biol.* 192:177-208.