JOURNAL OF VIROLOGY, Jan. 1991, p. 373–381
0022-538X/91/010373-09$02.00/0

Vol. 65, No. 1

Characterization of a Human Cytomegalovirus 1.6-Kilobase Late mRNA and Identification of Its Putative Protein Product

ROYA S. LAHIJANI, ELMER W. OTTESON, JOHN D. ADLISH, AND STEPHEN C. ST. JEOR*

Department of Microbiology and the Cell and Molecular Biology Program, University of Nevada, Reno, Nevada 89557-0046

Received 25 June 1990/Accepted 26 September 1990

In a previous study (J. Martinez, R. S. Lahijani, and S. C. St. Jeor, J. Virol. 63:233–241, 1989), we identified a late, unspliced 1.6-kb mRNA that maps to the *Hin*dIII R fragment of human cytomegalovirus (HCMV) AD169. In the present study, the direction of transcription of this mRNA was determined by Northern (RNA) analysis with strand-specific probes. Primer extension was used to precisely map the 5′ end of the mRNA. An open reading frame (ORF) designated ORF 2-1, located 176 nucleotides downstream from the cap site of the 1.6-kb mRNA, was identified. A synthetic peptide was made representing a hydrophilic region in the amino terminus of ORF 2-1. Immunoprecipitation and Western immunoblot analysis of infected HEL cell lysates, using affinity-purified antibody to the peptide (anti-P_{2-1}), detected a viral protein with an apparent molecular mass of 58 kDa late in infection. Further support for the presence of this protein in infected-cell lysates was obtained by an enzyme-linked immunosorbent assay. Expression of viral antigens in intact infected HEL cells was assessed by immunofluorescence. General cytoplasmic staining was observed at 62 h postinfection, in contrast to a localized staining observed in the nuclear and perinuclear region at 96 h postinfection.

Human cytomegalovirus (HCMV), a ubiquitous member of the herpesvirus family, is associated with a wide variety of diseases, particularly when infection occurs in developing fetuses or in immunocompromised individuals (1, 14, 17). HCMV contains a linear double-stranded DNA genome of approximately 240 kb, which is the largest and most complex of the human herpesvirus genomes. HCMV gene expression in permissive human fibroblast cells exhibits a sequential and regulated pattern, which can be broadly categorized into three phases designated as immediate-early, early, and late (6, 7, 27, 36, 41, 42; for recent reviews, see references 24 and 37).

The immediate-early genes, the first genes expressed after infection, are transcribed from limited regions in the genome in the absence of preceding viral protein synthesis (35, 39, 44). The primary regulatory mechanism of the immediate-early genes is at the transcriptional level (2, 5, 34, 38, 43). After IE proteins have been synthesized, there is a switch from restrictive to extensive transcription of the genome. Transcripts made at this time represent both early and late genes. Early mRNAs are transcribed, transported into the cytoplasm, and translated prior to viral DNA replication (3, 4, 11, 12, 19, 28, 32). In contrast, late mRNAs are sequestered in the nucleus following transcription. After the onset of viral DNA replication, these mRNAs are processed, transported into the cytoplasm, and subsequently translated into viral proteins. This suggests posttranscriptional regulation for late genes (10, 13, 22). However, true late genes, which stringently require DNA replication for expression (22, 25, 29, 33), as well as early-late (leaky-late), genes which require DNA replication for maximum expression (8, 25), have been reported.

Late genes, which code for viral structural proteins, play an important role in the assembly of infectious virus. Structural proteins are important targets for the humoral and cellular immune response. Thus, an understanding of late gene regulation could yield valuable information for designing strategies to control HCMV replication. To this end, we attempted to identify and characterize HCMV late genes. In our previous studies, three cDNA clones were isolated from a lambda gt11 library by using antibody against HCMV virion proteins (26). One clone (C3) mapped to the *Hin*dIII R fragment. Human convalescent sera reacted with the β-galactosidase–C3 fusion protein (fpC3), and antisera made to fpC3 identified a 25-kDa virion protein (25). This protein has also been identified by Meyer et al. (29) as the highly immunogenic 28-kDa structural phosphoprotein (pp28). Two late transcripts (1.3 and 1.6 kb) homologous to the C3 cDNA were subsequently identified. S1 nuclease analysis located the 5′ ends of these mRNAs, and DNA sequencing revealed CAAT and TATA boxes located at a proper distance from each of the 5′ ends (25). To verify these findings, primer extension and Northern (RNA) blot analyses with strand-specific probes were conducted to precisely locate the start of transcription and to identify the strand from which transcription was initiated. Here we report that the 1.3- and 1.6-kb mRNAs are transcribed from the same strand and that their transcription initiation sites are within 240 nucleotides of each other.

Reading frame analysis of the first 3,000 nucleotides of the *Hin*dIII R fragment identified the initiation codon of an open reading frame (ORF) capable of coding for a 21-kDa protein located 233 nucleotides downstream from the 1.3-kb 5′ end. Meyer et al. (29) showed that a protein synthesized from this ORF is recognized by a monoclonal antibody to pp28. As a result, this ORF has been named ORF 28 K. In addition to ORF 28 K, two short ORFs (ORF 2-1 and ORF 2-2) are located in this region. The initiation codon of ORF 2-1 is located 176 nucleotides downstream of the 5′ end of the 1.6-kb mRNA. To verify whether this ORF codes for a protein, we synthesized a synthetic peptide representing a hydrophilic region in the amino terminus of ORF 2-1. In this report we show that antibody against this peptide recognizes

* Corresponding author.

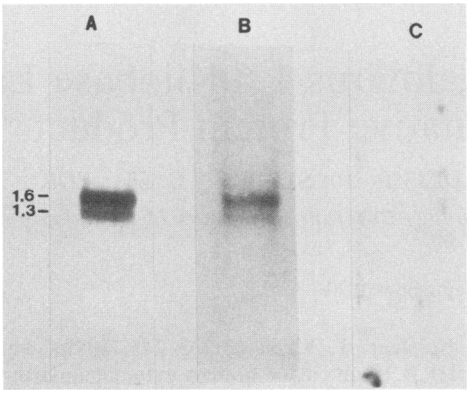

FIG. 1. Northern analysis with strand-specific probes. Cytoplasmic poly(A)$^+$ RNA was isolated from HCMV-infected HEL cells 62 h p.i., fractionated on a 1% formaldehyde–agarose gel, and transferred to GeneScreen Plus membrane. Each lane contained 10 μg of RNA. C3 cDNA was cloned into a M13mp18 vector in both orientations to generate strand-specific probes. Individual lanes were probed with ^{32}P-labeled C3 cDNA (lane A), single-stranded DNA from M13 clone C3-4 (lane B), and single-stranded DNA from M13 clone C3-6 (lane C). C3-4 DNA is complementary to C3-6 DNA. The size of the detected RNAs is given in kilobases.

a protein with an apparent molecular mass of 58 kDa in HCMV-infected cells late in infection.

MATERIALS AND METHODS

Cells and virus. Human embryonic lung (HEL) cells were grown in monolayers and maintained on Dulbecco modified Eagle medium supplemented with 10% fetal calf serum. HEL cells were infected with HCMV (AD169) at a multiplicity of infection of 10 PFU per cell.

RNA isolation and Northern analysis. Confluent monolayers of cells infected with HCMV were harvested at 62 h postinfection (p.i.), and cytoplasmic poly(A)$^+$ RNA was isolated as previously described (25). RNA was size fractionated on 1% agarose gels containing 2.2 M formaldehyde and transferred to GeneScreen Plus (Dupont, NEN Research Products, Boston, Mass.). Northern blots were hybridized and washed as specified by the manufacturer.

Oligonucleotide preparation. Oligonucleotides were made on a Du Pont Coder 300 as recommended by the manufacturer. Crude preparations from the synthesizer were purified on 7 M urea–17% polyacrylamide gels. The purified oligomers were 5′ end labeled with T4 kinase and [γ-^{32}P]ATP for 1 h at 37°C. Labeled oligomers were separated from unincorporated nucleotides as previously described (25). Two primers were synthesized. Primer 8784 is a 17-base primer complimentary to sequences 182 nucleotides downstream from the putative 5′ end of the 1.6-kb mRNA. Primer 6723 is a 17-base primer complementary to sequences 41 nucleotides downstream from the putative 5′ end of the 1.3-kb mRNA.

DNA sequencing. Sequencing was done by the dideoxy-chain termination method (31). The *Hin*dIII-*Kpn*I subfragment of *Hin*dIII-R was cloned into the M13mp18 vector, and single-stranded phage DNA was obtained. Primers 8784 and 6723 (mentioned above) were used in sequencing reactions.

Primer extension. Radiolabeled primers (3×10^4 cpm) were annealed to 10 μg of poly(A)$^+$ RNA and extended for 1 h at 42°C by using 10 U of reverse transcriptase (Life Sciences, St. Petersburg, Fla.). Reaction mixtures contained 50 mM Tris hydrochloride (pH 8.3), 10 mM $MgCl_2$, 10 mM dithiothreitol, and the four deoxynucleoside triphosphates (0.1 mM each). The extended products were phenol extracted and ethanol precipitated, and the pellets were dissolved in 5 μl of a solution containing 80% deionized formamide, 10 mM NaOH, 0.1% xylene cyanol, and 0.1% bromophenol blue. Samples were subjected to electrophoresis through a 6% polyacrylamide–7 M urea gel and analyzed by autoradiography.

Strand-specific M13 probes. C3 cDNA was cloned into the M13mp18 vector in both orientations. Single-stranded phage DNA generated from these clones contained the complementary strands of the C3 cDNA clone. Radiolabeled hybridization probes were made from single-stranded phage DNA, as described by Hu and Messing (18), by using a 17-base probe primer (Bethesda Research Laboratories, Gaithersburg, Md.) which was hybridized to M13 DNA at a point immediately downstream from the cloned insert. These probes were not denatured prior to hybridization.

Antibody production and purification. ORF 2-1 was translated by using the software SEQNCE (Delaney Software Ltd., Vancouver, Canada). The hydrophobicity plot of the deduced amino acid sequence was obtained by using the Kyte and Doolittle Hydrophobicity program (21). A synthetic peptide corresponding to residues 12 to 26 was synthesized by MilliGen/Biosearch, Navato, Calif. The synthetic peptide ($P_{2\text{-}1}$) was coupled to keyhole limpet hemocyanin (Sigma, St. Louis, Mo.) through tyrosine residues by using bis-diazobenzidine (BDB) (16). Rabbits were injected with coupled $P_{2\text{-}1}$ mixed with Freund complete adjuvant (Difco Laboratories, Detroit, Mich.) and boosted every 14 days with coupled $P_{2\text{-}1}$ mixed with Freund incomplete adjuvant. Rabbits were given two booster injections, and serum was collected 2 weeks after the second booster injection. Anti-peptide affinity-purified antibody (anti-$P_{2\text{-}1}$) was obtained by passing the rabbit serum over a column which was prepared by immobilizing 50 mg of peptide on the surface of cyanogen bromide-activated Sepharose CL-4B (Sigma). Purified antibody was assessed for reactivity to $P_{2\text{-}1}$ by enzyme-linked immunosorbent assay (ELISA) and double immunodiffusion (16).

ELISA of infected-cell lysates and purified virion proteins. HEL cells were either infected with HCMV or mock infected. At 62 and 96 h p.i., cells were harvested, resuspended in phosphate-buffered saline (PBS) containing 2 mM phenylmethylsulfonyl fluoride, and lysed by using a Dounce homogenizer. Fourfold serial dilutions of cell lysates were applied to 96-well microtiter plates and assessed for reactivity to anti-$P_{2\text{-}1}$ by ELISA. The anti-$P_{2\text{-}1}$ was used at a 1:1,000 dilution as the primary antibody, and horseradish peroxidase-labeled anti-rabbit immunoglobulin G (Southern Biotechnology, Birmingham, Ala.) was used as the secondary antibody. The concentration of anti-$P_{2\text{-}1}$ was 2 mg/ml prior to dilution. The substrate for the ELISA was 3,3′,5,5′-tetramethylbenzidine (Kirkegaard & Perry Laboratories, Inc., Gaithersburg, Md.), which was used as specified by the manufacturer. The optical density was read at 450 nm.

Immunofluorescence. HEL cells, grown on cover slips, were either infected with HCMV at a multiplicity of infection of 0.5 or mock infected. At 62 and 96 h p.i., the cells were washed in PBS, rinsed in water, fixed for 5 min with acetone, and allowed to air dry. A 1:10 dilution of anti-$P_{2\text{-}1}$ was applied to the cells, which were then incubated for 30 min at 37°C. The cells were washed four times in PBS, and a fluorescein-conjugated affinity-purified goat anti-rabbit anti-

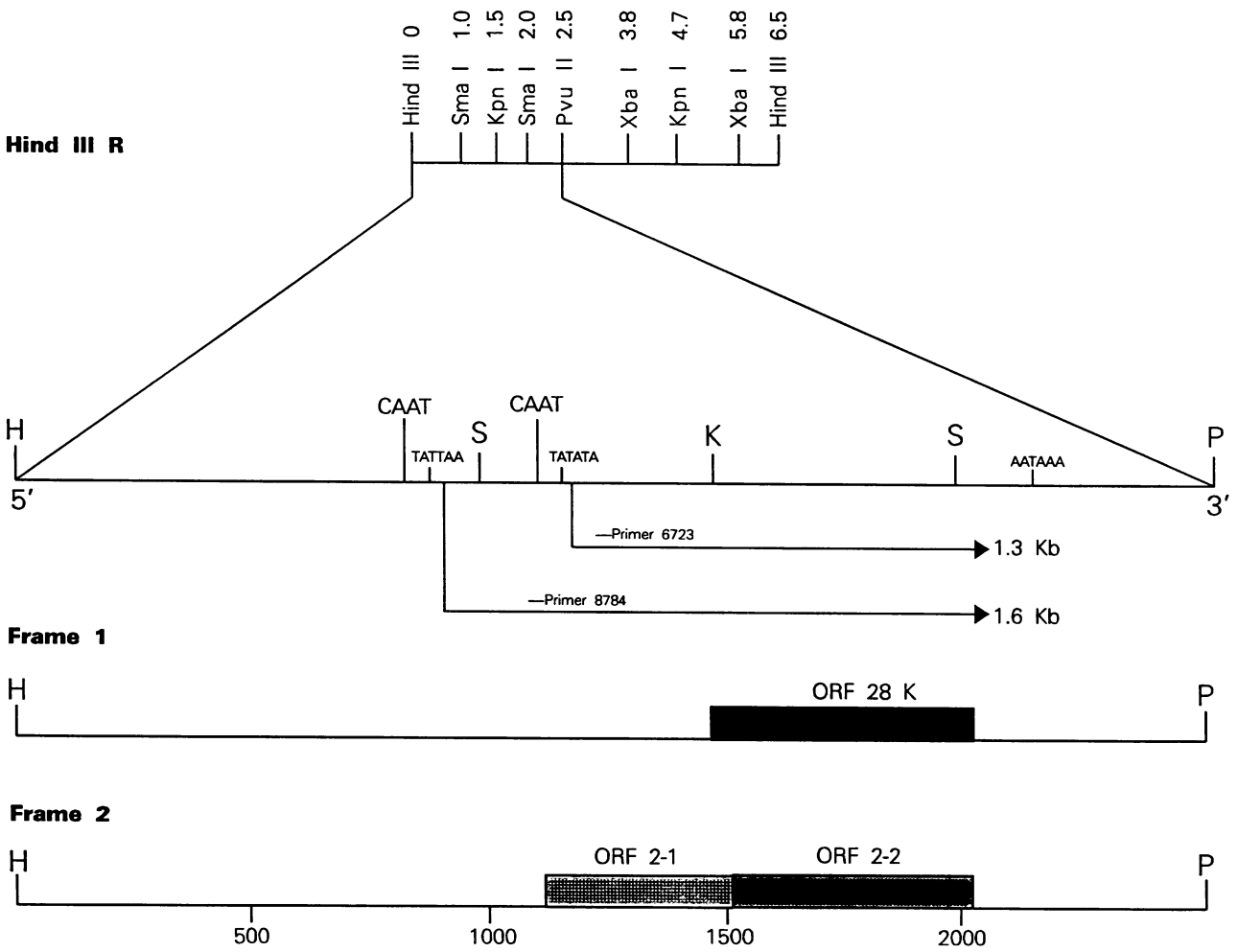

FIG. 2. Organizational summary of gene expression from the region encoding the 58-kDa protein. The restriction map of the 6.5-kb *Hin*dIII R fragment is shown. The region encoding the 1.6- and 1.3-kb transcripts is expanded, and selected restriction recognition sites are given. The 1.6- and 1.3-kb transcripts are represented by arrows from their respective 5′ ends. The promoter consensus sequences and the polyadenylation signal are indicated. Positions of oligomers used in primer extension analysis are shown. The location of ORF 28 K as well as the predicted ORFs in frame 2 are indicated.

body (Southern Biotechnology) was applied for 30 min at 37°C. After incubation, the cells were washed four times with PBS and rinsed once with water. They were air dried, mounted on slides, and observed under a fluorescence microscope.

Immunoprecipitation. HEL cells infected with HCMV or mock infected were starved for 1 h with cysteine-free Dulbecco modified Eagle medium containing 2% dialyzed fetal calf serum. The medium was replaced with Dulbecco modified Eagle medium containing 2% fetal calf serum and [^{35}S]cysteine (50 μCi/ml) and left on the cells for 4 h. At 62 h p.i., the cells were harvested and resuspended in a buffer containing 50 mM Tris hydrochloride (pH 7.4), 150 mM NaCl, and 1% Triton X-100. The lysates were precleared by incubation with normal rabbit serum and protein A-Sepharose (Sigma) for 60 min at 4°C. To 50 μl of each lysate (5 × 10^6 cpm) were added 30 μl of protein A-Sepharose and 10 μl of 2-mg/ml anti-P_{2-1}. The total volume was adjusted to 500 μl, and the mixture was incubated overnight at 4°C. The immune complexes formed were pelleted and washed three times in a buffer containing 50 mM Tris hydrochloride (pH 7.4), 5 mM EDTA, 5% sucrose, 1% Nonidet P-40, and 50 mM NaCl. The final pellet was resuspended in 30 μl of sample buffer (100 mM Tris hydrochloride [pH 6.8], 1% sodium dodecyl sulfate [SDS], 0.1 M β-mercaptoethanol) and assessed by SDS-polyacrylamide gel electrophoresis (PAGE) and autoradiography.

Isolation of infected-cell proteins and Western immunoblot analysis. HEL cells were infected with HCMV or mock infected. Infected and mock-infected cells were harvested at 96 h p.i., lysed in SDS-PAGE sample buffer, boiled for 3 min, and sonicated. Gradient SDS-PAGE gels (7.5 to 20%) were prepared as described by Hames (15). Infected-cell protein (100 μg) and mock-infected cell protein (100 μg) were electrophoresed on a 7.5 to 20% gradient polyacrylamide gel. Rainbow Protein markers (Amersham Corp., Arlington Heights, Ill.) were used as molecular weight standards. Proteins were blotted onto nitrocellulose membrane filters, as described by Towbin et al. (40), for 3 h at 300 mA. Blots were probed with anti-P_{2-1} at a 1:100 dilution. The concentration of anti-P_{2-1} was 2 mg/ml prior to dilution. A biotinylated goat anti-rabbit immunoglobulin G (Southern Biotech-

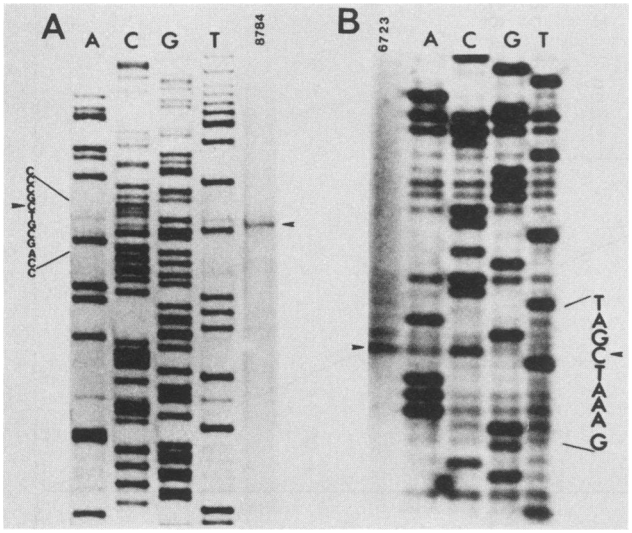

FIG. 3. Primer extension analysis. Primers 8784 and 6723 5′ end labeled with ^{32}P were annealed to late cytoplasmic poly(A)$^+$ RNA extracted from infected HEL cells and extended. The primer-extended products were separated on a sequencing gel adjacent to dideoxy sequencing ladders (lanes A, C, G, and T) generated with the primers and single-stranded *Hin*dIII-*Kpn*I DNA. A portion of the sequence ladder in the vicinity of the 5′ end is shown. The sequence is that of the template strand and hence the complement of that shown in Fig. 4. The 5′ ends are indicated by arrowheads.

nology) served as the secondary antibody. Bound antibodies were detected by subsequent addition of alkaline phosphatase-labeled streptavidin and stained with the Bio-Rad AP color-developing system as specified by the manufacturer.

RESULTS

Direction of transcription of the 1.6- and 1.3-kb mRNAs. To determine the direction of transcription of the 1.6- and 1.3-kb mRNAs, Northern blots of cytoplasmic poly(A)$^+$ RNA extracted from HEL cells at 62 h p.i. were hybridized with strand-specific probes. A control lane was probed with C3 cDNA. The results of the Northern analysis are presented in Fig. 1. Nick-translated C3 cDNA hybridized to both the 1.3- and 1.6-kb mRNAs that were present at 62 h p.i. (Fig. 1, lane A). Single-stranded DNA from the M13 clone C3-4 also hybridized to these transcripts (lane B). On the other hand, single-stranded DNA from the M13 clone C3-6, the complementary strand to C3-4 DNA, did not hybridize to either the 1.3- or 1.6-kb RNA (lane C). These results demonstrate that both the 1.6- and 1.3-kb mRNAs are transcribed from left to right in the genome arrangement shown in Fig. 2.

Defining the 5′ end of the 1.6- and 1.3-kb mRNAs by primer extension. Previously, we identified the 5′ ends of the 1.3- and 1.6-kb mRNAs by S1 nuclease protection (25). A primer extension experiment was conducted to precisely map the nucleotide initiating transcription. End-labeled primers 8784 and 6723 (Fig. 2) were annealed to poly(A)$^+$ RNA extracted from HCMV-infected cells at 62 h p.i. and extended with reverse transcriptase. The extension products were compared with dideoxy sequencing ladders generated with the same primers and single-stranded *Hin*dIII-*Kpn*I DNA. The nucleotide initiating transcription of the 1.6-kb mRNA was identified to be the G residue (Fig. 3A) located 913 bases downstream from the *Hin*dIII restriction site (Fig. 2). The nucleotide initiating the transcription of the 1.3-kb mRNA was identified as the G residue (Fig. 3B) located 1,153 bases downstream from the *Hin*dIII site (Fig. 2). The location of each initiation site relative to the *Hin*dIII restriction site shows that the 1.6-kb mRNA initiates at a site 240 nucleotides upstream of the initiation site of the 1.3-kb mRNA (Fig. 2).

Nucleotide sequence and reading frame analysis. Analysis of the first 3,000 nucleotides of the *Hin*dIII R fragment identified two short ORFs in reading frame 2 (ORF 2-1 and ORF 2-2). These ORFs are distinct from the previously described ORF 28 K (29) in reading frame 1 (Fig. 2). ORF 2-1 is able to encode a 120-residue polypeptide with a molecular mass of approximately 14 kDa. The initiation codon of ORF 2-2 is separated from the stop codon of ORF 2-1 by 3 nucleotides. ORF 2-2 is able to encode a 170-residue polypeptide with a molecular mass of approximately 18 kDa. The initiation codon of ORF 2-1 is the first AUG codon downstream from the cap site of the 1.6-kb RNA 5′ end. The 1.6-kb mRNA would contain a 176-base 5′ untranslated leader sequence. The polyadenylation signal (AATAAA) is located at nucleotides 2144 through 2149 with respect to the *Hin*dIII restriction site. Downstream of this signal, a consensus GT cluster is observed (Fig. 4). The same polyadenylation signal has been reported to be used by the 1.3-kb mRNA (29). This indicates that these two mRNAs have heterogenous 5′ and coterminal 3′ ends.

Predicted polypeptide sequence of ORF 2-1 and the production of antibody. To determine whether ORF 2-1 encodes a viral protein, the nucleotide sequence of ORF 2-1 was translated into its corresponding amino acid sequence and the hydrophobicity profile of the predicted protein was determined (Fig. 5). Molecular modeling techniques were used to search for areas of high charge density with a high probability of forming random or loop structures. This information was used to choose a very hydrophilic region located on the surface of the protein. A synthetic peptide ($P_{2\text{-}1}$) was synthesized, corresponding to amino acid residues 12 to 26, which met the above-mentioned criteria (Fig. 4). Seventy five percent of ths region is located upstream of the 5′ end of the 1.3-kb mRNA. Antibody was raised against $P_{2\text{-}1}$ in rabbits and was affinity purified as described in Materials and Methods.

Immunoreactivity of infected-cell lysate proteins. The affinity-purified anti-peptide antibody (anti-$P_{2\text{-}1}$) was tested for its ability to recognize virus-specific antigens in infected cells. Infected HEL cell lysates or mock-infected cell lysates obtained at 62 and 96 h p.i. were analyzed by ELISA. Infected-cell lysates obtained 62 h p.i. reacted with anti-$P_{2\text{-}1}$ (Fig. 6). The mock-infected cell lysates did not show appreciable reactivity with anti-$P_{2\text{-}1}$ (Fig. 6). In addition, ELISA demonstrated that cell lysates obtained at 96 h p.i. reacted with anti-$P_{2\text{-}1}$ (data not shown).

Detection of HCMV-specific antigen in infected cells by immunofluorescence. Immunofluorescence experiments were conducted to identify the cellular location of the viral protein reactive with anti-$P_{2\text{-}1}$. HEL cells grown on cover slips were either mock infected or infected with HCMV at a multiplicity of infection of 0.5. At 62 and 96 h p.i., cells were fixed in acetone and examined by immunofluorescence with anti-$P_{2\text{-}1}$ as the primary antibody. No significant fluorescent staining was observed with mock-infected HEL cells (Fig. 7A and B). Infected cells at 62 h p.i. exhibited diffuse cytoplasmic staining (Fig. 7C). At 96 h p.i., the staining appeared to be

```
GCA GTA CCT GGC CGA TCT GCT GTA TCT CAA TAA GGC CGA GTG TTC

GGA AGT GAT CGT GTT TGA CGC CAA GCA CCT GAG TGA CGA CAA CAG

CGA CGG GGA CGC CAC GAT CAC TAT TAA CGC GAG TCT CGG CCT AGC
        +1
CGC GGG CGA CGC TGG CGC CGG CGC TGA TCA CCA CCT GCG GGG CAG  +38
  SmaI
CCC GGG CGA TTC GCG CCG CCG ATA CCT TTC GAG GAC GAA AAC ACG  +83

CCC GAG CTG CTG GGC CGG CTC ACG TGT ACG AGG TAG CGC GCT TTT  +128

CAC TGC CGG CTT TTG TCA ATC CGG CGT CAC CAG TAT TAC TTT CAG  +173

ATG CTC ATT CAG CAG TAC GTG CTC AGC CAA TAC TAT ATA AAG AAG  +218
Met Leu Ile Gln Gln Tyr Val Leu Ser Gln Tyr Tyr Ile Lys Lys  (15)

CAT CCG GAC CCG GAG CGG ATC GAT TTC CGC GAC CTG CCT ACC GTC  +263
His Pro Asp Pro Glu Arg Ile Asp Phe Arg Asp Leu Pro Thr Val  (30)

TAC CTG GTC TCG GCC ATC TTC CGC GAG CGC GAG GAA AGC GAA CTG  +308
Tyr Leu Val Ser Ala Ile Phe Arg Glu Arg Glu Glu Ser Glu Leu  (45)

GGC TGC GAG TTG CTG GCC GGC GGT CGC GTT TTC CAC TGC GAC CAC  +353
Gly Cys Glu Leu Leu Ala Gly Gly Arg Val Phe His Cys Asp His  (60)

ATC CCG CTC CTG CTC ATC GTC ACG CCC GTG GTC TTT GAC CCT CAG  +398
Ile Pro Leu Leu Leu Ile Val Thr Pro Val Val Phe Asp Pro Gln  (75)

TTT ACG CGC CAT GCC GTC TCT ACC GTG CTA GAC CGT TGG AGT CGC  +443
Phe Thr Arg His Ala Val Ser Thr Val Leu Asp Arg Trp Ser Arg  (90)

GAC CTG TCC CGC AAG ACG AAC CTA CCG ATA TGG GTG CCG AAC TCT  +488
Asp Leu Ser Arg Lys Thr Asn Leu Pro Ile Trp Val Pro Asn Ser  (105)
                                KpnI
GCA AAC GAA TAT GTT GTG AGT TCG GTA CCA CGC CCG GTG AGC CCC  +533
Ala Asn Glu Tyr Val Val Ser Ser Val Pro Arg Pro Val Ser Pro  (120)

TGA AAG ATG CTC TGG GTC GCC AGG TGT CTC TAC GCT CCT ACG ACA  +578
Stop
ACA TCC CTC CGA CTT CCT CCT CGG ACG AAG GGG AGG ACG ATG ACG  +623

ACG GGG AGG ATG ACG ATA ACG AGG AGC GGC AAC AGA AGC TGC GGC  +668

TCT GCG GTA GTG GCT GCG GGG GAA ACG ACA GTA GTA GCG GCA GCC  +713

ACC GCG AGG CCA CCC ACG ACG GCT CCA AGA AAA ACG CGG TGC GCT  +758

CGA CGT TTC GCG AGG ACA AGG CTC CGA AAC CGA GCA AGC AGT CAA  +803

AAA AGA AAA AGA AAC CCT CAA AAC ATC ACC ACC ATC AGC AAA GCT  +848

CCA TTA TGC AGG AGA CGG ACG ACC TAG ACG AAG AGG ACA CCT CAA  +893

TTT ACC TGT CCC CGC CCC CGG TCC CCC CCG TCC AGG TGG TGG CTA  +938

AGC GAC TGC CGC GGC CCG ACA CAC CCA GGA CTC CGC GCC AAA AGA  +983
                                SmaI
AGA TTT CAC AAC GTC CAC CCA CCC CCG GGA CAA AAA AGC CCG CCG  +1028

CCT CCT TGC CCT TTT AAC TCA TAA ACT TTC AGG TCT CGC GTA CGA  +1073

TTC GCG AGT CGG GAA TGG GAC ACC CGT GGG TGT TTC TCC GTG TGT  +1118

ATA TTA TTT TTT TTT TTT GTG TGT GTT TGC GCC CCC GTG TGT CTA  +1163

ATG TGC TGT TTG AAA CAC GTA AAG TAG CTG GTG GAA GAA CAG ATA  +1208

AAC CTT TAA TAA AAA AAA AAG TAT GTG CTC CCG ACC CAC GGT CTG  +1253
```

FIG. 4. DNA sequence and predicted amino acid sequence of ORF-2-1. The nucloetide sequence (top line) and the amino acid sequence of ORF 2-1 (bottom line) are numbered. The transcription start site of the 1.6-kb RNA is numbered +1. The sequence of the synthesized peptide is boxed. Relevant restriction sites, CAAT motif, TATA motif, and polyadenylation signal are underlined.

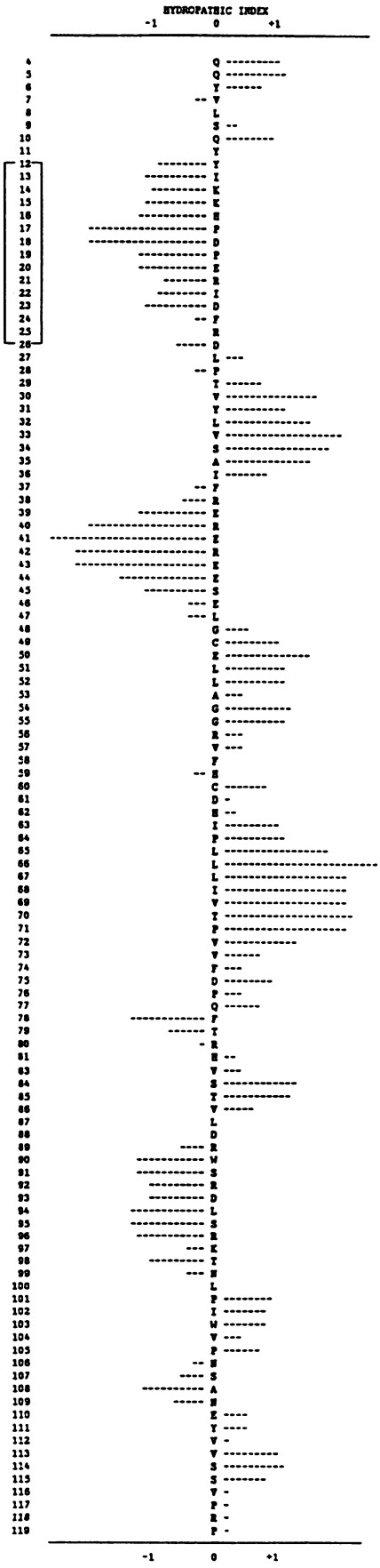

FIG. 5. Hydrophobicity profile of ORF 2-1. The predicted amino acid sequence of ORF 2-1 is given by the single-letter code. The hydropathy plot was obtained by using the Kyte-Doolittle program (21). The hydrophilicity (−) and hydrophobicity (+) are indicated by the horizontal bars.

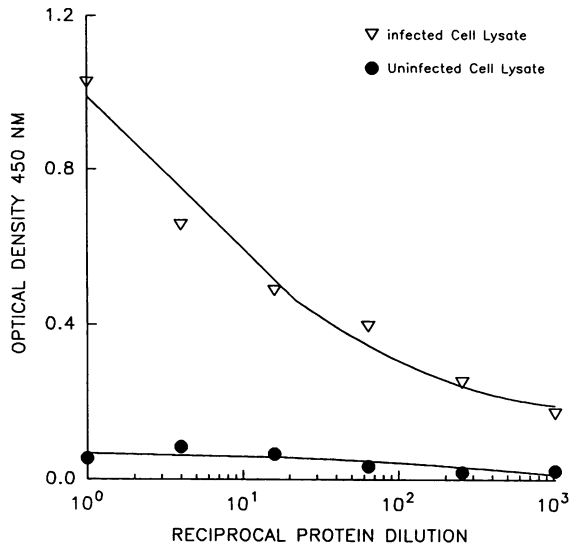

FIG. 6. Assessment of reactivity of anti-peptide antibody (anti-P_{2-1}). The Reactivity of anti-P_{2-1} to HCMV-infected HEL cell lysates, obtained at 62 h p.i., was determined by ELISA. Mock-infected HEL cell lysates were used as the control. Serial fourfold dilutions of each protein were applied to the wells of microtiter plates and assayed. Anti-P_{2-1} at an initial concentration of 2 mg/ml was diluted 1:1,000 and used as the primary antibody.

granular and localized to the nuclear and perinuclear region (Fig. 7D).

Identification of the viral protein encoded by ORF 2-1. To identify the specific protein which reacts with anti-P_{2-1}, HEL cells were infected with HCMV and labeled with [^{35}S] cysteine. Labeled infected or mock-infected cell lysates obtained at 62 h p.i. were immunoprecipitated with anti-P_{2-1} and analyzed by SDS-PAGE (12% acrylamide). Five protein bands, with apparent molecular masses of 58, 46, 44, 36, and 33 kDa, were immunoprecipitated in infected-cell lysates (Fig. 8, lane 3). The 44-kDa protein was also detected in the mock-infected cell lysate (Fig. 8, lane 1). In addition, the 46-kDa protein was detected in the mock-infected cell lysate after a longer exposure time (data not shown). Therefore, these proteins were considered nonspecific. Of the three remaining protein bands, the 58-kDa protein was the most prominent protein immunoprecipitated with anti-P_{2-1}.

To identify the protein that specifically reacts with anti-P_{2-1}, we performed a competition assay with increasing concentrations of peptide. The 58-kDa protein is the only protein that was effectively competed out in the presence of the peptide (Fig. 8, lanes 3 to 6). Of the two minor protein bands, the immunoprecipitation of the 36-kDa protein was not affected by the peptide, whereas the 33-kDa protein was competed out to some degree in the presence of high concentrations of the peptide (Fig. 8, lane 6). When peptide conjugated to keyhole limpet hemocyanin was used as the competitor, the 58-kDa protein was the only protein that was not precipitated with anti-P_{2-1} (data not shown). This con-

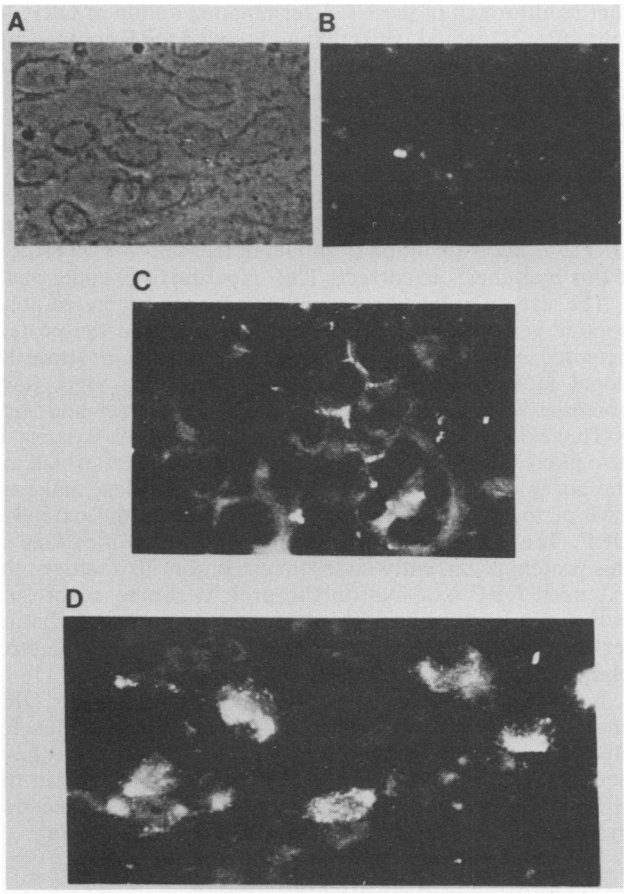

FIG. 7. Indirect immunofluorescence of intact HEL fibroblasts. HCMV-infected or mock-infected HEL cells were treated with anti-P_{2-1} and stained with fluorescein-labeled antibody. (A and B) Phase-contrast and fluorescence microscopy, respectively, of mock-infected HEL cells. (C and D) Immunofluorescence of infected HEL cells at 62 and 96 h p.i., respectively. Magnification, ×400.

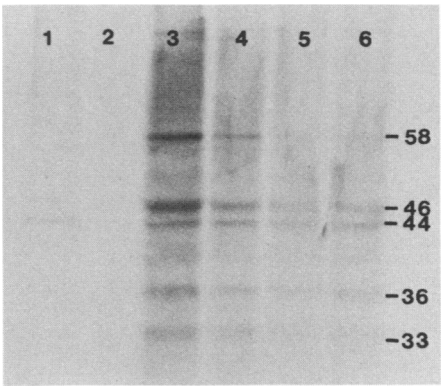

FIG. 8. Immunoprecipitation of HCMV-infected cell lysates. HEL cells were infected with HCMV or were mock infected. Cell lysates were obtained at 62 h p.i., precipitated with anti-P_{2-1}, and analyzed on an SDS–12% polyacrylamide gel. Lanes: 1, mock-infected cell lysates; 3 to 6, infected-cell lysates in the presence of 0, 0.2, 0.4, and 0.6 μg of peptide, respectively. The apparent molecular sizes of the precipitated proteins are indicated in kilodaltons.

firmed that the 58-kDa protein was specifically recognized by anti-P_{2-1}. The minor proteins that react with anti-P_{2-1} may represent breakdown products of the major 58-kDa protein.

Western analysis was also conducted to identify the anti-P_{2-1}-specific protein and to investigate its presence in the virion. Infected-cell proteins from 96 h p.i. and mock-infected cell proteins were fractionated on a 7.5 to 20% gradient polyacrylamide gel and transferred to nitrocellulose. The nitrocellulose filter was probed with anti-P_{2-1}, and a prominent protein band of 58 kDa was observed in infected-cell lysates (Fig. 9, lane 2). This protein was not present in mock-infected cell lysates (Fig. 9, lane 1).

DISCUSSION

In earlier studies we had isolated three cDNA clones from a lambda gt11 cDNA library (26). One clone (C3) mapped to the *Hin*dIII R fragment. Antisera made to the C3 fusion protein detected a 25-kDa virus protein (vp25) found in infected cells and purified virions (25). We believe that vp25 is identical to the highly immunogenic 28-kDa structural phosphoprotein (pp28) identified by Meyer et al. (29). When Northern blots prepared from HCMV late mRNAs were probed with the C3 cDNA, two mRNAs, 1.3 and 1.6 kb in size, were found to initiate within the *Hin*dIII R fragment (25). In contrast, Meyer et al. identified a 1.3-kb late mRNA which they report as the mRNA coding for pp28. They also detected a late 1.5-kb mRNA transcribed in the opposite direction with respect to the 1.3-kb mRNA. Northern analyses with strand-specific probes indicated that our 1.6-kb mRNA is not identical to the 1.5-kb mRNA reported by Meyer et al. In fact, Lehner et al. (23) recently reported that the 1.5-kb mRNA maps to the *Hin*dIII-R, *Sma*I-*Xba*I subfragment (Fig. 2), which is completely outside the region where the 1.3- and 1.6-kb mRNAs originate. In addition, we performed a primer extension experiment to locate the nucleotides initiating the transcription of these two mRNAs. We found that the transcription initiation sites of the 1.6- and 1.3-kb mRNAs are within 240 nucleotides of each other.

DNA sequencing analysis identified an ORF capable of coding for a 21-kDa protein within the *Hin*dIII R fragment, 233 nucleotides downstream from the 5′ end of the 1.3-kb mRNA. Meyer et al. (29) showed that a protein synthesized from this ORF is recognized by a monoclonal antibody to pp28. As a result, this ORF has been named ORF 28 K. Reading frame analysis of the *Hin*dIII R fragment identified two short ORFs (ORF 2-1 and ORF 2-2) in addition to ORF 28 K (Fig. 2). According to Kozak (20), the first AUG is used as an initiator codon in most eucaryotic genes. The initiation codon of ORF 2-1 is the first AUG downstream from the 5′ end of the 1.6-kb mRNA. To determine whether this ORF codes for a protein, a synthetic peptide representing a hydrophilic region in the amino terminus of ORF 2-1 was synthesized and antibody was raised against this peptide in rabbits. The majority of this region is located upstream of the transcription initiation site of the 1.3-kb mRNA. Previously we have shown that Northern analysis with a primer complementary to sequences upstream of the 5′ end of the 1.3-kb mRNA hybridizes only to the 1.6-kb mRNA (25). As a result, any protein recognized by this antibody is encoded by the 1.6-kb mRNA.

Antibody against the synthetic peptide was initially analyzed by an ELISA. It reacted with purified peptide as well as HCMV-infected cell lysates obtained at 62 and 96 h p.i. Subsequently, the antibody was used in immunoprecipitation and Western analyses to identify the viral protein

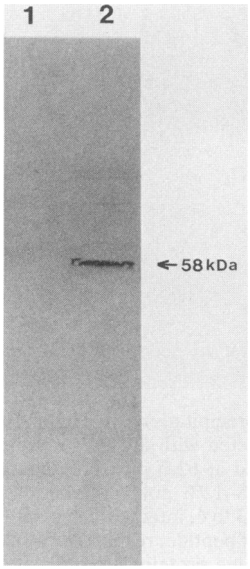

FIG. 9. Western analysis of infected cell lysates. HEL cells were infected with HCMV or were mock infected. Cell lysates obtained at 96 h p.i. were separated on a 7.5 to 20% gradient polyacrylamide gel, transferred to nitrocellulose, and incubated with anti-P_{2-1} (diluted 1:100). Lanes 1, mock-infected cell lysate; 2, infected-cell lysate. The protein detected by anti-P_{2-1} is indicated by an arrow.

encoded by ORF 2-1. This antibody detected a 58-kDa infected-cell protein late in infection.

ORF 2-1 is capable of coding for a protein of 14 kDa. The discrepancy between the sequence-predicted site and the observed size of the detected 58-kDa protein cannot be explained by posttranslational modifications alone. However, ORF 2-1 is followed by ORF 2-2 (Fig. 2), which is capable of coding for a protein of 18 kDa. If the termination codon of ORF 2-1 is avoided by some mechanism, an extended ORF capable of coding for a protein of 32 kDa will be generated. Subsequent posttranslational modifications could yield a protein of 58 kDa. The removal of the termination codon could occur by a splicing event. However, S1 nuclease and exonuclease VII protection analyses have demonstrated that the 1.6-kb RNA is unspliced (25).

Another possible mechanism of bypassing the stop codon is the shift of reading frame at the site of the UGA codon. However, a shift into reading frame 3 would create frequent termination codons in downstream positions. A shift into reading frame 1 would generate a protein that is 3′ coterminal with pp28 and hence should be recognized by pp28-specific antibody. However, a monoclonal antibody to pp28 does not recognize a protein of 58 kDa (30).

Several recent discoveries have indicated that nonsense codons do not necessarily signal the termination of polypeptide synthesis. It has been found that in mitochondria and mycoplasmas, the nonsense codon UGA actually codes for tryptophan (9). In addition, the nonsense codon UGA can be read by tRNAs charged with the modified amino acids selenocysteine and phosphoserine, which are incorporated directly into the polypeptide chain (46). Moreover, it has been shown that in several viral systems, for example Moloney murine leukemia virus, nonsense codons UGA and UAG can be misread by a normal charged tRNA by a process called translational readthrough (45). We speculate that readthrough of the UGA termination codon of ORF 2-1 into ORF 2-2 would generate an extended ORF coding for a protein of 32 kDa. In an effort to substantiate this hypothesis, we carried out a hybrid selection in vitro translation experiment and immunoprecipitated the in vitro translation product with anti-P_{2-1}. The preliminary data from this study indicated that the in vitro translation product is approximately 30 kDa. In addition, this possibility can be verified by raising antibody against a synthetic peptide representing ORF 2-2. Such an antibody should detect the 58-kDa protein if this hypothesis is correct. This experiment is underway.

The size difference between the sequence-predicted polypeptide generated from the extended ORF and the protein detected in infected-cell lysates suggests that posttranslational modifications might have occurred in vivo. Two potential N-linked glycosylation sites (Asn-X-Ser and Asn-X-Thr) exist in the extended ORF. O-linked glycosylation is also likely because serine and threonine residues, which are the acceptor sites for O-linked oligosaccharides, make up 18% of the total amino acid composition of the extended ORF. The nature of the posttranslational modifications of this protein is currently under investigation. In addition, the extended ORF is a hydrophilic protein composed of 20% basic amino acids and 12% proline residues. These characteristics could result in anomalous migration during electrophoresis and yield an overestimated size of 58 kDa.

It is interesting that the initiation codon of the ORF encoding this protein does not follow Kozak's rules for effective translational start sites (20). Since these rules have been used to predict the number of possible ORFs in the HCMV genome, the potential number of proteins encoded by the virus could be greater than currently estimated.

ACKNOWLEDGMENTS

We thank William Welch for assistance in molecular modeling and Mike Wilson for critical reading of the manuscript.

This work was supported by Public Health Service grant AI25755 from the National Institutes of Health. R.S.L. and J.D.A. were supported by training grants from the Reno Cancer Center and the National Institutes of Health. E.W.O. was supported by Public Health Service grant AI24357 from the National Institute of Allergy and Infectious Diseases.

REFERENCES

1. **Alford, C. A., and W. J. Britt.** 1990. Cytomegalovirus, p. 1981–2010. *In* B. N. Fields and D. M. Knipe (ed.), Virology, 2nd ed. Raven Press, New York.
2. **Boshart, M., F. Weber, G. Jahn, K. Dorsch-Hasler, B. Fleckenstein, and W. Schaffner.** 1985. A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. Cell **41**:521–530.
3. **Chang, C., C. L. Malone, and M. F. Stinski.** 1989. A human cytomegalovirus early gene has three inducible promoters that are regulated differentially at various times after infection. J. Virol. **63**:281–290.
4. **Chang, C., 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.
5. **Cherrington, J. M., and E. S. Mocarski.** 1989. Human cytomegalovirus iel transactivates the α promoter-enhancer via an 18-base-pair repeat element. J. Virol. **63**:1435–1440.
6. **DeMarchi, J. M.** 1981. Human cytomegalovirus DNA: restriction enzyme cleavage maps and map locations for immediate early, early, and late RNAs. Virology **114**:23–38.
7. **DeMarchi, J. M., C. A. Schmidt, and A. S. Kaplan.** 1980. Pattern of transcription of human cytomegalovirus in permissively infected cells. J. Virol. **35**:277–286.
8. **Depto, A. S., and R. M. Stenberg.** 1989. Regulated expression of the human cytomegalovirus pp65 gene: octamer sequence in the

promoter is required for activation by viral gene products. J. Virol. **63**:1232–1238.
9. **Fox, T. D.** 1987. Natural variation in the genetic code. Annu. Rev. Genet. **21**:67–91.
10. **Geballe, A. P., F. S. Leach, and E. S. Mocarski.** 1986. Regulation of cytomegalovirus late-gene expression: genes are controlled by posttranscriptional events. J. Virol. **57**:864–874.
11. **Geballe, A. P., and E. S. Mocarski.** 1988. Translational control of cytomegalovirus gene expression is mediated by upstream AUG codons. J. Virol. **62**:3334–3340.
12. **Geballe, A. P., R. R. Spaete, and E. S. Mocarski.** 1986. A cis-acting element within the 5′ leader of a cytomegalovirus β transcript determines kinetic class. Cell **46**:865–872.
13. **Goins, W. F., and M. F. Stinski.** 1986. Expression of a human cytomegalovirus late gene is posttranscriptionally regulated by a 3′-end processing event occurring exclusively late after infection. Mol. Cell. Biol. **6**:4202–4213.
14. **Griffiths, P. D., and J. E. Grundy.** 1987. Molecular biology and immunology of cytomegalovirus. Biochem. J. **241**:313–324.
15. **Hames, B. D.** 1981. An introduction to polyacrylamide gel electrophoresis, p. 1–86. *In* B. D. Hames and D. Rickwood (ed.), Gel electrophoresis of proteins: a practical approach. IRL Press, Washington, D.C.
16. **Harlow, E., and D. Lane.** 1988. Antibodies: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
17. **Ho, M.** 1982. Cytomegalovirus, biology and infection. Plenum Publishing Corp., New York.
18. **Hu, N. T., and J. Messing.** 1982. The making of strand-specific M13 probes. Gene **17**:271–277.
19. **Klucher, K. M., D. K. Rabert, and D. H. Spector.** 1989. Sequences in the human cytomegalovirus 2.7-kilobase RNA promoter which mediate its regulation as an early gene. J. Virol. **63**:5334–5343.
20. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
21. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
22. **Leach, F. S., and E. S. Mocarski.** 1989. Regulation of cytomegalovirus late-gene expression: differential use of three start sites in the transcriptional activation of ICP36 gene expression. J. Virol. **63**:1783–1791.
23. **Lehner, R., H. Meyer, and M. Mach.** 1989. Identification and characterization of a human cytomegalovirus gene coding for a membrane protein that is conserved among human herpesviruses. J. Virol. **63**:3792–3800.
24. **Mach, M., T. Stamminger, and G. Jahn.** 1989. Human cytomegalovirus: recent aspects from molecular biology. J. Gen. Virol. **70**:3117–3146.
25. **Martinez, J., R. S. Lahijani, and S. C. St. Jeor.** 1989. Analysis of a region of the human cytomegalovirus (AD169) genome coding for a 25-kilodalton virion protein. J. Virol. **63**:233–241.
26. **Martinez, J., and S. C. St. Jeor.** 1986. Molecular cloning and analysis of three cDNA clones homologous to human cytomegalovirus RNAs present during late infection. J. Virol. **60**:531–538.
27. **McDonough, S. H., and D. H. Spector.** 1983. Transcription in human fibroblasts permissively infected by human cytomegalovirus strain AD169. Virology **125**:31–46.
28. **McDonough, S. H., S. I. Staprans, and D. H. Spector.** 1985. Analysis of the major transcripts encoded by the long repeat of human cytomegalovirus strain AD169. J. Virol. **53**:711–718.
29. **Meyer, H., A. T. Bankier, M. P. Landini, C. M. Brown, B. G. Barrell, B. Ruger, and M. Mach.** 1988. Identification and procaryotic expression of the gene coding for the highly immunogenic 28-kilodalton structural phosphoprotein (pp28) of human cytomegalovirus. J. Virol. **62**:2243–2250.
30. **Re, M. C., M. P. Landini, P. Coppolecchia, G. Furlini, and M. La Placa.** 1985. A 28000 molecular weight human cytomegalovirus structural polypeptide studied by means of a specific monoclonal antibody. J. Gen. Virol. **66**:2507–2511.
31. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
32. **Staprans, S. I., D. K. Rabert, and D. H. Spector.** 1988. Identification of sequence requirements and trans-acting functions necessary for regulated expression of a human cytomegalovirus early gene. J. Virol. **62**:3463–3473.
33. **Stenberg, R. M., A. S. Depto, J. Fortney, and J. A. Nelson.** 1989. Regulated expression of early and late RNAs and proteins from the human cytomegalovirus immediate-early gene region. J. Virol. **63**:2699–2708.
34. **Stenberg, R. M., and M. F. Stinski.** 1985. Autoregulation of the human cytomegalovirus major immediate early gene. J. Virol. **56**:676–682.
35. **Stenberg, R. M., D. R. Thomsen, and M. F. Stinski.** 1984. Structural analysis of the major immediate early gene of human cytomegalovirus. J. Virol. **49**:190–199.
36. **Stinski, M. F.** 1983. Molecular biology of cytomegalovirus, p. 67–113. *In* B. Roizman (ed.), The herpesviruses, vol. 2. Plenum Publishing Corp., New York.
37. **Stinski, M. F.** 1990. Cytomegalovirus and its replication, p. 1959–1980. *In* B. N. Fields and D. M. Knipe (ed.), Virology, 2nd ed. Raven Press, New York.
38. **Stinski, M. F., and T. J. Roehr.** 1985. Activation of the major immediate early gene of human cytomegalovirus by *cis*-acting elements in the promoter-regulatory sequence and by virus-specific *trans*-acting components. J. Virol. **55**:431–441.
39. **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.
40. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76**:4350–4354.
41. **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.
42. **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.
43. **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.
44. **Wilkinson, G. W. G., A. Akrigg, and P. J. Greenaway.** 1984. Transcription of the immediate early genes of human cytomegalovirus strain AD169. Virus Res. **1**:101–116.
45. **Yoshinaka, Y., I. Katoh, T. D. Copeland, and S. Oroszlan.** 1985. Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. Proc. Natl. Acad. Sci. USA **82**:1618–1622.
46. **Zinoni, F., A. Birkmann, T. C. Stadtman, and A. Bock.** 1986. Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from Escherichia coli. Proc. Natl. Acad. Sci. USA **83**:4650–4654.