Four Phosphoproteins with Common Amino Termini Are Encoded by Human Cytomegalovirus AD169

DOWAIN A. WRIGHT, SILVIJA I. STAPRANS, AND DEBORAH H. SPECTOR*

Department of Biology C-016, University of California, San Diego, La Jolla, California 92093

Received 5 June 1987/Accepted 14 September 1987

In this report, we identify the proteins encoded by the 2.2-kilobase class of early transcripts arising from a region of the strain AD169 human cytomegalovirus genome (map units 0.682 to 0.713) which contains cell-related sequences. These transcripts, encoded by adjacent *EcoRI* fragments R and d, have a complex spliced structure with 5' and 3' coterminal ends. Antiserum directed against a synthetic 11-amino-acid peptide corresponding to the predicted amino terminus of the proteins was generated and found to immunoprecipitate four infected-cell proteins of 84, 50, 43, and 34 kilodaltons. These proteins were phosphorylated and were associated predominantly with the nuclei of infected cells. The 43-kilodalton protein was the most abundant of the four proteins, and its level of expression remained relatively constant throughout the infection. Expression of the other proteins increased as the infection progressed. Pulse-chase analysis failed to show a precursor-product relationship between any of the proteins. A comparison of the [³⁵S]methionine-labeled tryptic peptide maps of the four proteins from infected cells and an in vitro-generated polypeptide derived from the putative first exon showed that all four infected-cell proteins were of viral origin and contained a common amino-terminal region.

Human cytomegalovirus (HCMV) is a herpesvirus capable of causing a wide variety of pathogenesis in humans. It is recognized as the leading viral cause of birth defects and as a common, opportunistic infectious agent in immunosuppressed individuals (for a review, see reference 18). HCMV is known to establish both persistent and latent infections and may have oncogenic potential (reviewed in references 22 and 26). The replication cycle of HCMV in permissive cells is complex, but it appears to display the sequential order of gene expression common to other herpesviruses (8, 17, 31). Identification of specific transcription units, elucidation of the protein products encoded by these units, and determination of the biochemical characteristics of these proteins will provide the foundation for a greater understanding of the multiple virus-host interactions involved in the replication cycle of HCMV.

In our previous studies (27), we investigated the pattern of transcription of a 2.2-kilobase (kb) class of transcripts arising from the adjacent *EcoRI* fragments R and d (map units 0.682 to 0.713), located at the right end of the long unique segment of the HCMV AD169 genome. This site is near the region of HCMV major immediate-early transcription (14, 17, 28, 29, 32), overlaps the transforming region described by Nelson et al. (19, 20), Clanton et al. (7), and El-Beik et al. (9), and contains the cell-related sequences described by Shaw et al. (24) and further characterized by Rasmussen et al. (23). Staprans and Spector (27) mapped the 5' and 3' limits and complex spliced structure of the 2.2-kb class of early transcripts encoded within the *EcoRI* R-d region. They showed that the 2.2-kb RNAs consisted of invariable 5' and internal exons and a heterogeneous 3' exon (see Fig. 1). The predicted amino acid sequence encoded by the amino-terminal exon of the RNAs was determined and found to be very glycine rich (16.7%). With these data as background, it was important to determine what, if any, proteins were encoded by this complex set of transcripts.

To identify the protein products encoded by the 2.2-kb

transcripts, we generated a polyclonal rabbit antiserum directed against a synthetic peptide corresponding to the putative amino terminus of the protein(s). In this study, we investigated the nature of the polypeptides immunoprecipitated by this antiserum. We showed that the 2.2-kb class of transcripts encodes four distinct proteins and that these proteins are all of viral origin. The kinetics of accumulation and subcellular localization are presented, along with an analysis of potential phosphorylation, glycosylation, and precursor-product relationships between these proteins.

MATERIALS AND METHODS

Virus and cells. HCMV AD169 and human fibroblasts were obtained and propagated as previously described (30).

Generation of antiserum. The peptide used in this study is located at the putative amino terminus of the proteins (Fig. 1). The 11-amino-acid peptide, acetylated at the amino terminus, was synthesized at the Peptide Oligonucleotide Synthesis Facility at the University of California, San Diego. The peptide was attached to bovine serum albumin (BSA) (Sigma Chemical Co., St. Louis, Mo.) with bisdiazobenzidine (Sigma) for tyrosine-to-tyrosine side chain coupling (2). Briefly, 20 mg of BSA was dissolved in 3 ml of 0.13 M NaCl-0.16 M borate (pH 9.0), and a 50-fold molar excess of the peptide was added. To the BSA-peptide solution was added 0.8 ml of bisdiazobenzidine in 0.2 M HCl with stirring in an ice bath; the pH was adjusted to 9.0 by the addition of 0.5 M NaOH, and stirring was continued at 4°C for 2 h. The conjugate was dialyzed against phosphate-buffered saline (PBS) for 2 days at 4°C. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was used to verify that coupling had occurred. Antipeptide serum was obtained after immunization of female rabbits by intradermal injections at 20 different sites with 1 mg of peptide-BSA conjugate in 0.5 ml of PBS emulsified with 0.5 ml of complete Freund adjuvant (GIBCO Laboratories, Grand Island, N.Y.). The animals were boosted every 4 weeks by combined intramuscular and subcutaneous injections, and antisera were ob-

* Corresponding author.

FIG. 1. Summary of the 2.2-kb RNA exon structure and the nucleotide sequence of the putative promoter-regulatory region and exon 1. (Top) A diagram of the HCMV AD169 genome structure and *EcoRI* restriction endonuclease sites (25) is presented along with a more detailed map of the R-d region. The portion of the sequence represented is indicated by the black bar above the restriction map of *EcoRI* fragments R and d. The positions of the component exons of the 2.2-kb RNA are summarized below, with sizes given in nucleotides. (Bottom) The DNA sequence was derived from the data of Kouzarides et al. (15) (nucleotides 1 to 573) and Rasmussen et al. (23) (nucleotides 574 to 1084). The putative TATA box for this transcription unit is underlined. The nucleotide to which the 5' end was mapped (27) is marked with an asterisk. The putative methionine initiation codon and surrounding sequence context are underlined, as are the putative splice donor sequence and the stop codon in the reading frame shown. The 11-amino-acid peptide which was used to generate the antiserum utilized in our studies is shown in the box. Nucleotide and amino acid numbers are indicated on the right- and left-hand margins, respectively.

tained at 2 and 4 weeks after each injection. This protocol was followed for 9 months with a good responder.

Radiolabeling of infected-cell proteins. Nearly confluent monolayers of human foreskin fibroblasts (5×10^5 cells) in 60-mm tissue culture dishes were infected with AD169 at a multiplicity of infection of 1 to 3. For [35 S]methionine labeling, at 6.5, 22.5, or 70.5 h postinfection (p.i.), cells were washed twice and incubated for 30 min with 4 ml of methionine-free Dulbecco modified Eagle medium containing 10% fetal bovine serum (FBS) (Irvine Scientific, Irvine, Calif.).

The medium was replaced with 0.67 ml of methionine-free Dulbecco modified Eagle medium containing 10% FBS and 400 μ Ci of [35 S]methionine (Translabel; 1,011 Ci/mmol; ICN Pharmaceuticals, Inc., Irvine, Calif.; except for tryptic peptide mapping for which we used pure [35 S]methionine; 1,270 Ci/mmol; Amersham Corp., Arlington Heights, Ill.) per ml and incubated for 1 h at 37°C. For [32 P]_i labeling; at 5.5 and 23.5 h p.i., 2×10^6 cells in 100-mm dishes were washed two times and incubated for 30 min with 10 ml of phosphate-free minimum essential medium (Flow Laboratories, Inc.,

McLean, Va.) containing 10% dialyzed FBS (GIBCO). The medium was replaced with 2 ml of phosphate-free minimal essential medium containing 10% dialyzed FBS and 1 mCi of $^{32}\text{P}_i$ (285 Ci/mg; ICN) per ml and incubated for 1 h at 37°C. For [^3H]glucosamine labeling; at 4 h p.i., cells in a 60-mm dish were washed once and incubated with Dulbecco modified Eagle medium containing 10% FBS and 50 μCi of [^3H]glucosamine (40 Ci/mmol; Amersham) per ml at 37°C for 24 h. All subsequent manipulations were done at 4°C. Cells were harvested by first washing them three times with 5 ml of PBS (pH 7.4), followed by the addition of 0.33 ml of cell lysis buffer (50 mM Tris hydrochloride [pH 8.0], 10 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.5% deoxycholate, 0.1% SDS, 100 mM NaCl, 0.5 trypsin-inhibiting units of aprotinin [Sigma] per ml, 250 μg of BSA per ml, 25 mM Na_2HPO_4 , 2 mM phenylmethylsulfonyl fluoride, 100 μM sodium orthovanadate). Cells were allowed to lyse for 30 min before being scraped from the dish with a rubber policeman. Lysates were clarified by centrifugation in an Eppendorf centrifuge for 30 min. Cleared lysates were then used directly for immunoprecipitation or were stored at -70°C .

Pulse-chase analysis of infected-cell proteins. Cell monolayers were prepared as described above. At 24 h p.i., cells were pulsed for 10 min with 0.67 ml of Dulbecco modified Eagle medium containing 10% FBS and 400 μCi of [^{35}S]methionine (Translabel; 1,011 Ci/mmol) per ml. One dish was washed twice with PBS and stored at -70°C . Other dishes were washed twice with 5 ml of Dulbecco modified Eagle medium containing 10% FBS and incubated with 5 ml of the same medium at 37°C for 30, 60, or 120 min. Monolayers were then washed with PBS and stored at -70°C . The cells were then thawed and lysed, and cleared lysates were collected as described above.

Nuclear-cytoplasmic fractionation. Cell monolayers were labeled with [^{35}S]methionine as described above. All subsequent steps were at 4°C. The cells were washed twice with PBS and scraped from the dish with a rubber policeman in 1 ml of PBS. The cells were then pelleted by centrifugation at 2,000 rpm for 5 min in a clinical centrifuge. The pellet was suspended in 0.25 ml of cytoplasmic lysis buffer (30 mM Tris hydrochloride [pH 7.6], 150 mM NaCl, 1.5 mM MgCl_2 , 0.2% Nonidet P-40, 0.5 trypsin-inhibiting units of aprotinin per ml, 1 mM phenylmethylsulfonyl fluoride, 100 μM sodium orthovanadate) and incubated at 4°C for 10 min. The cell nuclei were pelleted by centrifugation as described above. The supernatant was considered the cytoplasmic fraction. The nuclear pellet was suspended in cytoplasmic lysis buffer and adjusted to a final concentration of 0.8% Nonidet P-40 and 0.4% deoxycholate. The nuclei were pelleted as described above, and the supernatant was considered the nuclear wash or membrane fraction. Both supernatants were adjusted to $1\times$ cell lysis buffer by the addition of 0.25 ml of $2\times$ cell lysis buffer. The nuclei were suspended in 0.5 ml of cell lysis buffer and incubated at 4°C for 30 min. Lysates were cleared by centrifugation in an Eppendorf centrifuge for 30 min and used directly for immunoprecipitation or stored at -70°C .

Immunoprecipitation and SDS-PAGE. Radiolabeled lysates (80 to 125 μl) were precleared by the addition of 8 to 12 μl of 50% Pansorbin (Calbiochem-Behring, La Jolla, Calif.) and incubated at 4°C for 30 min. The lysates were centrifuged at 4°C for 5 min in an Eppendorf centrifuge, and the pellets were discarded. Preimmune rabbit serum was added to the supernatant at a dilution of 1:12 and incubated at 4°C for 1 h. Antibody was cleared with Pansorbin as above, and the pellet was considered the preimmune fraction. Immune

rabbit serum was added to the supernatant at a dilution of 1:30 and incubated at 4°C for 1 to 12 h. Immune complexes were collected with Pansorbin as above. The Pansorbin pellets were washed three times with TNDTS (10 mM Tris hydrochloride [pH 7.4], 1% deoxycholate, 1% Triton X-100, 150 mM NaCl, 0.1% SDS), suspended in sample buffer (2% SDS, 5 mM Na_2HPO_4 [pH 7.0], 10% glycerol, 5% 2-mercaptoethanol, 0.1 M dithiothreitol, 0.01% bromophenol blue), and heated to 100°C for 5 min.

Immunoprecipitated samples were subjected to SDS-PAGE essentially as described by Laemmli (16). Except when otherwise noted, SDS-PAGE gels consisted of a 5% acrylamide–0.13% bisacrylamide stacking gel and a 10% acrylamide–0.13% bisacrylamide resolving gel. Gels were stained with Coomassie brilliant blue. Molecular weights were estimated by comparison with known molecular weight standards (Bio-Rad Laboratories, Richmond, Calif.). ^{35}S and ^3H gels were impregnated with PPO (2,5-diphenyloxazole; Sigma) by soaking in 10% (wt/vol) PPO in dimethyl sulfoxide followed by extensive rinsing with water before drying. All gels were fluorographed at -70°C except for the ^{32}P -labeled gel, which was autoradiographed at -70°C with an intensifying screen.

Immunoblotting. Human foreskin fibroblasts (1.5×10^6 cells per 100-mm dish) were either infected with AD169 at a multiplicity of infection of 3.0 or were left uninfected. Uninfected- and infected-cell monolayers at 2, 4, 8, 24, 48, and 72 h p.i. were released from the dish with trypsin and pelleted in a clinical centrifuge (2,000 rpm, 5 min). The cell pellets were suspended in 30 μl of lysis buffer (50 mM Tris hydrochloride [pH 8.0], 10 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.5% deoxycholate, 0.1% SDS, 100 mM NaCl, 0.5 trypsin-inhibiting units of aprotinin [Sigma] per ml, 25 mM Na_2HPO_4 , 2 mM phenylmethylsulfonyl fluoride) and incubated at 4°C for 30 min. The lysates were clarified by centrifugation in an Eppendorf centrifuge at 4°C for 30 min, an equal volume of $2\times$ sample buffer was added to the supernatants, and the samples were heated to 100°C for 5 min.

Lysates (5×10^5 cell equivalents per sample) were separated by SDS-PAGE as described above, transferred to nitrocellulose, incubated in blocking buffer (Tris-buffered saline with 50 g of nonfat dry milk [Carnation] per liter, 0.1 ml of antifoam A emulsion [Sigma] per liter, 0.02% NaN_3) at 4°C for 12 h, and then incubated with a 1:50 dilution of immune rabbit serum in the above buffer for 1 h at 22°C. Blots were washed four times with 100 ml of blocking buffer for 15 min at 22°C, incubated with 25 ml of 10^6 -cpm/ml ^{125}I -labeled protein A (40 $\mu\text{Ci}/\mu\text{g}$; ICN) for 1 h at 22°C, washed four more times, and autoradiographed at -70°C with an intensifying screen.

Tryptic peptide mapping. Tryptic peptide analysis was performed essentially as described by Beemon and Hunter (3). Briefly, protein bands were excised from the gel and homogenized in a Ten Broeck homogenizer in 0.05 M NH_4HCO_3 containing 0.1% SDS. Proteins were precipitated in 25% trichloroacetic acid and oxidized with performic acid. After lyophilization, samples were digested with 30 μg of tosylsulfonyl phenylalanyl chloromethyl ketone-trypsin (TPCK-trypsin; Sigma) in 0.05 M NH_4HCO_3 for 18 h at 20°C followed by an additional 20 μg for 4 h at 20°C. The samples were lyophilized, and the tryptic peptides were dissolved in water and spotted on thin-layer cellulose chromatography plates (EM Science, Cherry Hill, N.J.). Electrophoresis was performed at 1.0 kV for 27 min with 1% ammonium carbonate buffer (pH 8.9). The plates were then chromatographed in the second dimension in buffer containing *n*-butanol,

pyridine, acetic acid, and water (97:75:15:60, vol/vol). Plates were sprayed with En³Hance (New England Nuclear Corp., Boston, Mass.) and fluorographed at -70°C .

Generation of pGemRd-1. All enzymes were obtained from Bethesda Research Laboratories, Inc. (Gaithersburg, Md.) and used according to the recommendations of the supplier. pRd, containing the R-d fusion fragment (27), was linearized with *Bgl*II (about 0.7 kilobase pairs upstream of the start site of transcription of the 2.2-kb early RNAs). *Bal* 31 exonuclease was used to produce deletions through the upstream and leader sequences of the 2.2-kb early RNA transcription unit. The resulting fragments were treated with T4 DNA polymerase and subsequently cleaved with *Hind*III (3' of the R-d sequences, within the pUC9 polylinker). The deletion fragments were isolated by gel purification and ligated to *Hinc*II-*Hind*III-cleaved pUC13. Recombinant clones were analyzed by acrylamide gel electrophoresis. A clone terminating at nucleotide 212, located between the transcription start site and the putative initiation codon of the 2.2-kb RNAs (Fig. 1), was used to generate pGemRd-1. This was achieved by cleaving the 5' deletion clone with *Xba*I (5' of the 2.2-kb RNA coding sequences, within the pUC13 polylinker) and *Hind*III and gel purifying the 5'-deleted R-d fusion fragment and ligating it into *Xba*I-*Hind*III-cleaved pGem-1 (Promega Biotec, Madison, Wis.).

In vitro transcription and translation of recombinant clone of *Eco*RI R-d fusion fragment. pGemRd-1 was linearized by digestion with *Pvu*II (Bethesda Research Laboratories) according to the recommendations of the supplier. Cleavage with *Pvu*II yields a fragment containing the T7 promoter and approximately 1.5 kilobase pairs of the HCMV genome encoding the 5' half of the primary transcript for the 2.2-kb RNAs. The DNA was extracted with phenol-chloroform-isoamyl alcohol (50:48:2, vol/vol/vol) and precipitated with ethanol. Standard in vitro transcription reaction mixtures (100 μl) contained 5 μg of the linearized DNA template, 40 mM Tris hydrochloride (pH 7.5), 6 mM MgCl_2 , 2 mM spermidine, 10 mM NaCl, 10 mM dithiothreitol, 100 units of RNasin (Promega), 500 μM each ATP, CTP, and UTP, 50 μM GTP, 500 μM $\text{m}^7\text{G}(5')\text{ppp}(5')\text{G}$, plus 2 μCi of [α - ^{32}P]UTP (400 Ci/mmol; Amersham). After the addition of 75 U of T7 RNA polymerase (Promega), the reaction mixtures were incubated at 40°C for 45 min, followed by a 45-min incubation in the presence of 75 additional units of T7 RNA polymerase. This was followed by the addition of 20 U of RNase-free DNase and incubation for 15 min at 37°C . After extraction with phenol-chloroform-isoamyl alcohol (50:48:2, vol/vol/vol) and ethanol precipitation, the labeled RNA was purified from unincorporated nucleotides and salt with the use of NENSORB 20 cartridges (Du Pont Co., Wilmington, Del.) according to the recommendations of the manufacturer. The RNA was eluted from the column with 20% ethanol, and the eluate was adjusted to a final concentration of 0.5 M LiCl and 70% ethanol for precipitation of the RNA. The RNA pellet was washed twice at 4°C with 95% ethanol to remove residual salt and suspended in diethyl pyrocarbonate-treated H_2O for in vitro translation.

Standard in vitro translation reaction mixtures (30 μl) contained 50% nuclease-treated wheat germ extract (Amersham), 100 mM potassium acetate, 67 μM amino acid mixture minus L-methionine (Amersham), 25 units of RNasin (Promega), 30 μCi of [^{35}S]methionine (1,270 Ci/mmol; Amersham), and RNA synthesized from the in vitro transcription reaction described above. The reaction mixtures were incubated for 60 min at 25°C , followed by a 15-min incubation at 37°C in the presence of 60 μg of RNase A

(Pharmacia Fine Chemicals, Piscataway, N.J.). The reaction mixtures were adjusted to a final concentration of 2% SDS, 5 mM Na_2HPO_4 (pH 7.0), 10% glycerol, 5% 2-mercaptoethanol, 0.1 M dithiothreitol, and 0.01% bromophenol blue and boiled for 5 min. The polypeptide products of the reactions were isolated from 15% acrylamide-0.086% bisacrylamide preparative SDS-PAGE gels and used for tryptic peptide mapping as described above.

RESULTS

Generation of anti-peptide antiserum and identification of potential virus-encoded proteins. To determine what, if any, viral proteins were derived from the 2.2-kb class of early RNAs, we prepared antiserum directed against their putative amino terminus. A synthetic 11-amino-acid peptide corresponding to the predicted amino terminus of the protein (Fig. 1) was covalently attached to BSA and inoculated into female rabbits. A resulting antiserum (BSA2-8) was tested for its ability to immunoprecipitate specific viral proteins from AD169-infected cells. Lysates from [^{35}S]methionine-labeled uninfected and infected cells at 8 h p.i. were immunoprecipitated with BSA2-8. Four proteins of approximately 84, 50, 43, and 34 kilodaltons (kDa) were specifically immunoprecipitated only in AD169-infected cells by immune (BSA2-8) and not preimmune (BSA2) antiserum (Fig. 2). The

FIG. 2. Identification of proteins encoded by the 2.2-kb early RNAs. Uninfected (U) and infected (I) cells at 7 h p.i. were labeled with [^{35}S]methionine for 1 h. Cell lysates were immunoprecipitated with preimmune (Pre) antiserum or cleared with preimmune antiserum followed by immunoprecipitation with immune (Imm) antiserum and analyzed by SDS-PAGE. +, Immunoprecipitation with immune serum was done in the presence of 100 μg of free peptide per ml.

immunoprecipitation was specific in the sense that in the presence of excess peptide, all four proteins could be completely competed away. These data show that these proteins are specific to AD169-infected cells and suggest that they contain at least some homologous regions.

Kinetics of appearance of the proteins. In the previous analysis of the 2.2-kb class of RNAs, they were classified as early in the HCMV cascade of immediate-early, early, and late gene expression (27). The following experiments were designed to determine whether the temporal appearance of the proteins encoded by these transcripts in AD169-infected cells paralleled the synthesis of the transcripts. To determine the steady-state concentration of the proteins throughout the infection, uninfected and AD169-infected cells at 2, 4, 8, 24, 48, and 72 h p.i. were separated by SDS-PAGE and transferred to nitrocellulose. Proteins were detected by immunoblotting as described in Materials and Methods (Fig. 3). The 43-kDa protein was present in the highest concentration at all time points. It could be detected by 8 h p.i. and increased in abundance through 72 h p.i. The 50- and 84-kDa proteins could be detected at 24 h p.i., and the 34-kDa protein, which was least abundant, could only be seen at 48 and 72 h p.i.

To determine the relative rates of synthesis of the proteins at different times during the infection, we pulse-labeled cells with [35 S]methionine from 7 to 8, 23 to 24, and 71 to 72 h p.i., and cell lysates were immunoprecipitated with BSA2-8 at dilution of 1:30 and analyzed by SDS-PAGE (Fig. 4). All four proteins were detected by 8 h p.i. The 43-kDa protein was the most abundantly translated at all times throughout the infection. Its rate of synthesis was highest at 8 h p.i. and decreased slightly at late times in the infection. The 84- and 50-kDa proteins were synthesized at detectable levels at 8 h p.i., but their rate of synthesis increased as the infection progressed. The 34-kDa protein was synthesized at low

FIG. 3. Immunoblot analysis of cell lysates from uninfected (U) and infected (I) cells. Lysates (5×10^5 cell equivalents) obtained from uninfected cells or from cells infected with HCMV for 2, 4, 8, 24, 48, and 72 h were separated by SDS-PAGE, transferred to nitrocellulose, probed with immune antiserum, and labeled with 125 I-protein A.

FIG. 4. Time course of expression of viral proteins. Cell lysates from uninfected (U) and infected (I) cells pulse-labeled with [35 S]methionine for 1 h were obtained at 8, 24, and 72 h p.i., immunoprecipitated with preimmune antiserum (Pre) or preimmune followed by immune (Imm) antiserum, and analyzed by SDS-PAGE.

levels at 8 h p.i. and its rate of synthesis increased significantly from early to late times in the infection.

Subcellular localization. To identify the subcellular localization of the viral proteins, we labeled cells with [35 S]methionine at 24 h p.i. and split them into two equal fractions. One fraction was used to generate a total cell lysate, while the other was separated into nuclear, nuclear wash, and cytoplasmic fractions (see Materials and Methods). Immunoprecipitates were prepared from the total cell lysate and subcellular fractions and analyzed by SDS-PAGE (Fig. 5). The 84-, 50-, 43-, and 34-kDa proteins were all associated predominantly with the nuclear fraction.

Pulse-chase analysis. To determine whether there was a precursor-product relationship between any of the proteins, we performed a pulse-chase analysis. Infected cell monolayers at 8 h p.i. were pulsed for 10 min with [35 S]methionine and chased for 0, 30, 60, and 120 min. Cell lysates were prepared as described in Materials and Methods, and immunoprecipitates were analyzed by SDS-PAGE. Figure 6 shows that after the 10-min pulse, all four proteins were present in approximately the same ratio as in cells labeled for 1 h before lysis (Fig. 2). During the chase period, the relative proportions of the proteins did not appear to change. This indicates that there is no apparent precursor-product relationship between any of the proteins under the conditions in which we performed the experiments.

FIG. 5. Subcellular localization of viral proteins. Lysates from cells pulse-labeled with [35 S]methionine from 23 to 24 h p.i. were prepared and separated into total cell (T), cytoplasmic (C), nuclear wash (NW), and nuclear (N) fractions as described in Materials and Methods, immunoprecipitated with preimmune antiserum (Pre) or preimmune followed by immune (Imm) antiserum, and analyzed by SDS-PAGE.

Tryptic peptide mapping. From the above analysis, we were able to show that four infected-cell proteins could be specifically immunoprecipitated with an antiserum directed against a predicted N-terminal peptide from the coding region of the 2.2-kb class of transcripts. Kinetically, none of these proteins appeared to have a precursor-product relationship. We could not determine from the data, however, whether all these proteins were authentic translation products of the 2.2-kb RNAs or whether one or more were virus-induced cellular proteins with homology to the true viral proteins. To differentiate between these possibilities, we utilized tryptic peptide mapping. [35 S]methionine-labeled cell lysates were prepared from AD169-infected cells at 24 h p.i. for the 84-, 50-, and 43-kDa proteins and at 72 h p.i. for the 34-kDa protein. The 72-h time point was chosen because we observed that synthesis of the 34-kDa protein increases greatly at late times in the infection (Fig. 4). The proteins were immunoprecipitated with BSA2-8 and isolated by preparative SDS-PAGE.

To compare the *in vivo* proteins with the known putative amino-terminal-coding region, we prepared an *in vitro* polypeptide derived from the 5' portion of the 2.2-kb early RNA coding sequence (Fig. 1). pGemRd-1 contains a portion of the *EcoRI* R-d fusion fragment extending from 26 base pairs upstream of the putative initiation codon for the 2.2-kb transcripts to the *Bam*HI site in *EcoRI*-d. pGemRd-1 was linearized with *Pvu*II, and an *in vitro* RNA transcript was generated with T7 RNA polymerase. The transcript was then

translated in a wheat germ extract translation system. The expected polypeptide product should correspond to the entire 5' coding region and terminate in intron 1 at position 1042 (Fig. 1). The major product of the *in vitro* translation was a 34-kDa polypeptide which nearly comigrated with the 34-kDa infected-cell protein (data not shown). The *in vitro* product was isolated by preparative SDS-PAGE. Tryptic peptide analysis was then performed on the isolated proteins (see Materials and Methods).

The tryptic peptide maps of the putative viral proteins were compared along with that of the *in vitro* polypeptide (Fig. 7). The maps of the 43- and 50-kDa proteins were virtually identical, showing that these proteins were highly related. Both proteins reproducibly generated 11 major peptides and a minor peptide which comigrated in mixing experiments (Fig. 7A, B, and C). The 84-kDa protein generated a more complex map containing 16 tryptic peptides (Fig. 7D). A mixing experiment with the 50- and 84-kDa tryptic peptides (Fig. 7E) showed that the 84-kDa protein shares all 11 of the major peptides of both the 43- and 50-kDa proteins along with five unique peptides of its own. The 84-kDa protein is missing the minor tryptic spot which is present in the tryptic maps of the 43- and 50-kDa proteins (designated I). These data show that the 84-kDa protein is closely related to and probably overlaps with sequences contained in both the 43- and 50-kDa proteins. Tryptic analysis of the 34-kDa protein immunoprecipitated from cell lysates at 72 h p.i. generated eight major peptides (Fig. 7F). By mixing the *in vivo* 34- and 43-kDa tryptic digests, we showed all eight major spots to be a subset of those generated by the 84-, 50-, and 43-kDa proteins (Fig. 7I).

To confirm that these proteins were indeed of viral origin and encoded by the 2.2-kb class of transcripts, we then compared their tryptic peptide maps with that of the *in vitro*

FIG. 6. Pulse-chase analysis of infected-cell proteins. Uninfected (U) and 8 h p.i. infected (I) cell monolayers were pulsed with [35 S]methionine for 10 min and chased for 0, 30, 60, or 120 min. Cell lysates were immunoprecipitated with preimmune antiserum (Pre) or preimmune followed by immune (Imm) antiserum and analyzed by SDS-PAGE.

FIG. 7. Tryptic peptide mapping of [35 S]methionine-labeled infected-cell proteins and an in vitro polypeptide generated from exon 1 of the 2.2-kb class of RNAs. (A) The 43-kDa protein. The origin of the tryptic peptides is indicated (ori) along with the orientation of anode (+) and cathode (-). The direction of chromatography is indicated by an arrow. Conditions for electrophoresis and chromatography were identical for all tryptic peptide maps. (B) The 50-kDa protein. (C) Mix of the tryptic peptides from the 43- and 50-kDa proteins. (D) The 84-kDa protein. (E) Mix of the tryptic peptides from the 50- and 84-kDa proteins. (F) Infected-cell (ic) 34-kDa protein. (G) In vitro (iv) 34-kDa protein. (H) Mix of the tryptic peptides from the 43- and 34-kDa (iv) proteins. (I) Mix of the tryptic peptides from the 43- and 34-kDa (ic) proteins. Spots are labeled for purposes of discussion.

34-kDa polypeptide. The 34-kDa polypeptide should contain the entire coding region of the putative first exon and terminate in the first intron (Fig. 1). The tryptic peptide map of the 34-kDa in vitro polypeptide and its mix with the

43-kDa protein showed that it contained a set of eight spots which were also present in the maps of the 84-, 50-, 43-, and 34-kDa infected-cell proteins (Fig. 7G and H).

These data showed that the 84-, 50-, 43-, and 34-kDa

proteins which are immunoprecipitated with BSA2-8 are of viral origin and share a common amino terminus. This result is similar to other data obtained by partial proteolysis of the [35 S]methionine-labeled proteins with *Staphylococcus aureus* V-8 protease (data not shown).

32 P_i and [3 H]glucosamine labeling. To further characterize the proteins encoded from the 2.2-kb class of RNAs, we labeled infected cells with 32 P_i from 23 to 24 h p.i. Cell lysates were immunoprecipitated with BSA2-8 at a dilution of 1:150 for 1 h at 4°C, and immune complexes were isolated. The supernatant was subjected to a second round of immunoprecipitation for 12 h, and all immune complexes were analyzed by SDS-PAGE (see Materials and Methods) (Fig. 8). All four proteins were phosphorylated under these conditions. The second round of immunoprecipitation was performed to determine whether there was antibody excess in the initial reaction. The addition of an equal amount of antiserum to the supernatant and subsequent incubation for 12 h brought down equivalent amounts of the 84-, 50-, and 43-kDa proteins and an increased amount of the 34-kDa protein. The appearance of the 34-kDa protein only after a 12-h incubation with antiserum could be the result of a relative inaccessibility of the amino terminus in the native

protein. The 34-kDa protein was also detected by immunoblot analysis of infected-cell lysates (Fig. 3) and increased in abundance at late times in the infection. Thus, appearance of the 34-kDa protein is not an artifact of the immunoprecipitation protocol. All previously mentioned immunoprecipitations were performed at an antibody dilution of 1:30 for 12 h.

Cells were also labeled from 4 to 24 h p.i. with [3 H]glucosamine (see Materials and Methods), and cell lysates were immunoprecipitated and analyzed by SDS-PAGE (data not shown). Under these conditions, although [3 H]glucosamine was incorporated into infected-cell proteins, there was no apparent incorporation of label into any of the proteins encoded by the 2.2-kb RNAs.

DISCUSSION

In this report, we present the results of studies to identify and characterize the protein products encoded by the 2.2-kb class of transcripts arising from the adjacent *Eco*RI fragments R and d (map units 0.682 to 0.713) within the long unique segment of the HCMV AD169 genome. The 2.2-kb RNAs have been classified as early in the HCMV cascade of immediate-early, early, and late gene expression (27). Utilizing a polyclonal rabbit antiserum directed against the putative amino terminus of the proteins encoded by the 2.2-kb RNAs, we identified four proteins with approximate molecular sizes of 84, 50, 43, and 34 kDa whose kinetics of appearance followed a pattern similar to that of the RNAs. Pulse-labeling experiments showed that synthesis of the proteins could be detected by 4 h p.i. (data not shown), and all four proteins were clearly present by 8 h p.i. (Fig. 4). This distinguishes the 2.2-kb transcription unit from those described by Geballe et al. (10) in which the RNAs may be present at early times but the proteins are not detected until late in the infection.

The 43-kDa protein was the most abundantly translated of the four proteins at all times during the infection and most closely mirrored the kinetics of the 2.2-kb RNAs. Its maximum rate of synthesis was achieved by 8 h p.i. and remained constant or decreased slightly as the infection progressed to at least 72 h p.i. In contrast, the rate of synthesis of both the 50- and 84-kDa proteins increased from 8 to 72 h p.i., and although the 34-kDa protein was synthesized at detectable levels at 8 h p.i., high levels of this protein were only detected at late times in the infection. As discussed below, these results suggest that this transcription unit has an additional level of regulation which occurs posttranscriptionally and may involve alternative RNA splicing or perhaps differential posttranslational processing.

From a comparison of the tryptic peptide maps of the [35 S]methionine-labeled 84-, 50-, 43-, and 34-kDa infected-cell proteins with the map generated by an in vitro-synthesized 34-kDa polypeptide corresponding to the putative amino terminus of the proteins (Fig. 7), we were able to establish that all four infected-cell proteins were of viral origin and contained homologous regions. We found that the in vitro-synthesized polypeptide generated eight tryptic peptides (labeled A to H in Fig. 7) and that this core set of peptides was present in all four in vivo proteins. It should be noted, however, that a subset of these eight peptides is probably incomplete digestion products since the sequence (Fig. 1) predicts that there should be only five methionine-containing peptides. One set of partial digestion products is likely due to the Arg-Lys doublet at amino acid positions 9 and 10, but the origin of the other incomplete cleavage products is not obvious. Further analysis of the tryptic maps

FIG. 8. 32 P-labeled infected-cell proteins. Uninfected (U) and infected (I) cells at 6 and 24 h p.i. were labeled for 1 h with 32 P_i. Lysates were immunoprecipitated with preimmune antiserum (Pre) or preimmune followed by immune (Imm) antiserum. Immune complexes were collected, and the supernatants were immunoprecipitated a second time with immune antiserum (U/2x and I/2x). All immunoprecipitates were then analyzed by SDS-PAGE.

from proteins labeled with alternative radioactive amino acids should facilitate precise identification of each peptide, and these studies are in progress. Nevertheless, since the maps of both in vivo and in vitro tryptic peptides are identical with respect to the presence of all eight peptides, we can conclude that all four in vivo proteins share a common amino terminus which corresponds to the putative first exon as mapped by Staprans and Spector (27).

Our finding that the in vitro and in vivo 34-kDa proteins were similar with respect to size and tryptic peptide maps suggests a common origin. The in vitro product was generated from a synthetic RNA starting 26 base pairs upstream of the initiating ATG codon through the first *PvuII* site in *EcoRI*-d (Fig. 1). The first in-frame stop signal occurs at nucleotide 1042, 48 base pairs downstream of the putative splice donor. We propose that the 34-kDa infected-cell protein could be generated in a similar manner; mRNAs which have failed to splice out intron 1 are translated to yield this protein. In this regard, it has recently been shown that the region of the genome encoding these proteins can generate an unspliced transcript of about 2.8 kb and that the concentration of this longer transcript increases at late times (R. D. Rasmussen, unpublished data). This is consistent with the time course of expression of the 34-kDa protein (Fig. 3 and 4), i.e., it is present in very low amounts at early times and is most abundant at late times in the infection.

The derivation of the other proteins is somewhat more obscure. We showed that the 84-, 50-, and 43-kDa proteins share a common amino-terminal region. They also share a set of three tryptic peptides (labeled J, K, L in Fig. 7), which are likely generated from the sequences within exon 2 of this transcription unit. However, we are still left with two known transcripts and three proteins. It is possible that two of the proteins are derived from the fully processed messages and that the third is derived from a minor transcript which has removed the first but not the second intron. A second possibility is that there is an alternative splice acceptor very close to one of the two previously mapped splice acceptors for exon 3 and that the two bands were not resolved in previous mapping studies (27). Although this type of post-transcriptional regulation generating alternatively spliced RNAs, which in turn generate multiple proteins, is unique, so far, for HCMV, there are several notable examples of genomic economy for other viruses. Two well-known examples are the polyomavirus T antigens and the adenovirus E1A proteins. For the polyomavirus T antigens, three proteins with a common amino terminus but different carboxy termini are generated from three differentially spliced mRNAs (12). For the adenovirus E1A-coding unit, an alternative splice donor site is used to generate two overlapping mRNAs (5, 6) encoding two nuclear phosphoproteins with identical amino and carboxy termini (21). In addition, Beisel et al. (4) have recently shown that the two major outer envelope glycoproteins of Epstein-Barr virus are encoded by transcripts with common 5' and 3' ends, one which is unspliced and the other which has an internal splice but maintains the same open reading frame. Finally, Goins and Stinski (11) have shown that an HCMV late gene is regulated posttranscriptionally by utilization of an upstream cleavage and polyadenylation signal only at late times in the infection.

An alternative explanation for the derivation of the 50- and 43-kDa proteins is that they are different forms of the same protein. Although pulse-chase analysis (Fig. 6) failed to show a precursor-product relationship between any of the four proteins and thus suggested that they are encoded by individual transcripts, it is possible that a rapid modification

could occur on a fraction of one or more of the proteins either cotranslationally or immediately after synthesis. None of the proteins were labeled when infected cells were incubated with [³H]glucosamine (data not shown); thus, it appears that they are not glycosylated. However, all four proteins were phosphorylated in vivo (Fig. 8), with the 43-kDa protein proportionally less phosphorylated than the 50- and 84-kDa proteins. Studies are currently in progress to determine the role of alternative processing of the RNA transcripts and differential phosphorylation in the generation of the four proteins.

From our analysis of the subcellular distribution of these proteins, we were able to show that all four proteins were predominantly associated with the nuclear fraction of infected cells (Fig. 5). This common subcellular compartmentalization suggests that these proteins contain homologous regions which determine localization, possibly within exon 1. It should be noted, however, that with the fractionation procedures used, we cannot exclude the possibility that the proteins are associated with cytoskeletal elements which collapse on the nucleus during the extraction. Immunofluorescent staining of infected cells with immune sera should verify the nuclear localization of these proteins, and these experiments are in progress.

An important question yet to be answered concerns the function of these proteins. Given their nuclear localization, a number of potential roles are suggested, including the regulation of expression of other viral or host cellular genes; a component of virions or some intermediate in the assembly process; or perhaps a role in viral DNA replication. In this regard, we noted that the sizes of the four proteins do not correspond to those of any previously described HCMV virion proteins (13), and preliminary analysis in our laboratory has not shown any association of these proteins with intact virions. However, it is possible that these proteins are the counterparts to the strain Colburn CMV-specified proteins described by Anders et al. (1). In particular, they identified several early, nucleus-associated, nonstructural proteins with approximate molecular sizes of 86, 42, and 35 kDa. We are currently determining whether any of the above-described HCMV proteins have the ability to bind to viral or host cellular DNA and expect that the further characterization of this complex class of HCMV proteins will give a greater insight into their potential functions and importance in the replication cycle of HCMV.

ACKNOWLEDGMENTS

We thank E. Wade for determining the nucleotide sequence at the 5' boundary of the pGemRd-1 deletion clone; D. Jaquish for excellent technical assistance; K. Klucher, D. Rabert, and R. Rasmussen for helpful discussions; and D. David for preparation of the manuscript. We also thank B. Sefton and J. Buss for help with the peptide mapping.

This investigation was supported by Public Health Service grant 34729 from the National Cancer Institute to D. H. Spector, and Public Health Service training grants GM07198 and GM07240 from the National Institutes of Health.

LITERATURE CITED

1. Anders, D. G., A. Irmiere, and W. Gibson. 1986. Identification and characterization of a major early cytomegalovirus DNA-binding protein. *J. Virol.* **58**:253-262.
2. Bassiri, R. M., J. Dvorak, and R. D. Utiger. 1979. Thyrotropin-releasing hormone, p. 46-47. In B. M. Jaffe and H. R. Behrmak (ed.), *Methods of hormone radioimmunoassay*. Academic Press, Inc., New York.
3. Beemon, K., and T. Hunter. 1978. Characterization of Rous

- sarcoma virus *src* gene product synthesized in vitro. *J. Virol.* **28**:551-566.
4. Beisel, C., J. Tanner, T. Matsuo, D. Thorley-Lawson, F. Keady, and E. Kieff. 1985. Two major outer envelope glycoproteins of Epstein-Barr virus are encoded by the same gene. *J. Virol.* **54**:665-674.
 5. Berk, A. J., and P. A. Sharp. 1978. Structure of the adenovirus 2 early mRNAs. *Cell* **14**:695-711.
 6. Chow, L. T., T. R. Broker, and J. B. Lewis. 1979. Complex splicing patterns of mRNAs from the early regions of adenovirus-2. *J. Mol. Biol.* **134**:265-303.
 7. Clanton, D. J., R. J. Jariwalla, C. Kress, and L. J. Rosenthal. 1983. Neoplastic transformation by a cloned human cytomegalovirus DNA fragment uniquely homologous to one of the transforming regions of herpes simplex virus type 2. *Proc. Natl. Acad. Sci. USA* **80**:3826-3830.
 8. DeMarchi, J. M. 1981. Human cytomegalovirus DNA: restriction enzyme cleavage and map locations for immediate-early, early and late RNAs. *Virology* **114**:23-28.
 9. El-Beik, T., A. Razzaque, R. Jariwalla, R. L. Cihlar, and L. J. Rosenthal. 1986. Multiple transforming regions of human cytomegalovirus DNA. *J. Virol.* **60**:645-652.
 10. Geballe, A. P., F. S. Leach, and E. S. Mocarski. 1986. Regulation of cytomegalovirus late gene expression: γ genes are controlled by posttranscriptional events. *J. Virol.* **57**:864-867.
 11. Goins, W. F., and M. F. Stinski. 1986. Expression of a human cytomegalovirus late gene is posttranscriptionally regulated by a 3'-end-processing event occurring exclusively late after infection. *Mol. Cell. Biol.* **6**:4202-4213.
 12. Hunter, T., M. A. Hutchinson, and W. Eckhart. 1978. Translation of polyoma virus T antigens *in vitro*. *Proc. Natl. Acad. Sci. USA* **75**:5917-5921.
 13. Irmieri, A., and W. Gibson. 1983. Isolation and characterization of a noninfectious virion-like particle released from cells infected with human strains of cytomegalovirus. *Virology* **130**:118-133.
 14. Jahn, G., E. Knust, H. Schmolla, T. Sarre, J. A. Nelson, J. K. McDougall, and B. Fleckenstein. 1984. Predominant immediate-early transcripts of human cytomegalovirus AD169. *J. Virol.* **49**:363-370.
 15. Kouzarides, T., A. T. Bankier, and B. G. Barrell. 1983. Nucleotide sequence of the transforming region of human cytomegalovirus. *Mol. Biol. Med.* **1**:47-58.
 16. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**:680-685.
 17. McDonough, S. H., and D. H. Spector. 1983. Transcription in human fibroblasts permissively infected by human cytomegalovirus strain AD169. *Virology* **125**:31-46.
 18. Nankervis, G. A., and M. L. Kumar. 1978. Diseases produced by cytomegaloviruses. *Med. Clin. North Am.* **62**:1021-1035.
 19. Nelson, J. A., B. Fleckenstein, D. A. Galloway, and J. K. McDougall. 1982. Transformation of NIH3T3 cells with cloned fragments of human cytomegalovirus strain AD169. *J. Virol.* **43**:83-91.
 20. Nelson, J. A., B. Fleckenstein, G. Jahn, D. A. Galloway, and J. K. McDougall. 1984. Structure of the transforming region of human cytomegalovirus AD169. *J. Virol.* **49**:109-115.
 21. Perricaudet, M., G. Akusjarvi, A. Virtanen, and U. Pettersson. 1979. Structure of two spliced mRNAs from the transforming region of human subgroup C adenoviruses. *Nature (London)* **281**:694-696.
 22. Rapp, F. 1983. The biology of cytomegaloviruses, p. 1-66. *In* B. Roizman (ed.), *The herpesviruses*. Plenum Publishing Corp., New York.
 23. Rasmussen, R. D., S. I. Staprans, S. B. Shaw, and D. H. Spector. 1985. Sequences in human cytomegalovirus which hybridize with the avian retrovirus oncogene *v-myc* are G+C rich and do not hybridize with the human *c-myc* gene. *Mol. Cell. Biol.* **5**:1525-1530.
 24. Shaw, S. B., R. D. Rasmussen, S. H. McDonough, S. I. Staprans, J. P. Vacquier, and D. H. Spector. 1985. Cell-related sequences in the DNA genome of human cytomegalovirus strain AD169. *J. Virol.* **55**:843-848.
 25. Spector, D. H., L. Hock, and J. C. Tamashiro. 1982. Cleavage maps for human cytomegalovirus DNA strain AD169 for restriction endonucleases *EcoRI*, *BglII*, and *HindIII*. *J. Virol.* **42**:558-582.
 26. Spector, D. H., and S. A. Spector. 1984. The oncogenic potential of human cytomegalovirus. *Prog. Med. Virol.* **29**:45-89.
 27. Staprans, S. I., and D. H. Spector. 1986. 2.2-Kilobase class of early transcripts encoded by cell-related sequences in human cytomegalovirus strain AD169. *J. Virol.* **57**:591-602.
 28. 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.
 29. 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.
 30. Tamashiro, J. C., L. J. Hock, and D. H. Spector. 1982. Construction of a cloned library of *EcoRI* fragments from the human cytomegalovirus genome (strain AD169). *J. Virol.* **42**:547-557.
 31. 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**:446-451.
 32. 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.