JOURNAL OF VIROLOGY, Feb. 1992, p. 1098–1108
0022-538X/92/021098-11$02.00/0

Vol. 66, No. 2

Characterization of Human Cytomegalovirus UL84 Early Gene and Identification of Its Putative Protein Product

YUN SHENG HE,[1] LIAN XU,[1] AND ENG-SHANG HUANG[1,2,3,4]*

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

Received 7 August 1991/Accepted 7 November 1991

The DNA sequence and transcription pattern of human cytomegalovirus early gene UL84 were analyzed. This gene was mapped within a 2.6-kb *Pst*I fragment located between 0.534 and 0.545 map unit of the large unique segment of the human cytomegalovirus genome, which is adjacent to the pp65 and pp71 genes. A 2.0-kb mRNA was transcribed from this region in the same leftward direction as the mRNAs of the pp65 and pp71 genes. The message was first detected at 2.5 h postinfection and reached a maximal level between 72 and 96 h postinfection. The nucleotide sequences of the 2.6-kb *Pst*I genomic DNA fragment and the cDNA derived from this region were determined. The resulting data revealed a polyadenylation signal (AATAAA) located 14 nucleotides upstream from the poly(A) tail of the cDNA and a 1,761-bp open reading frame capable of encoding a 65-kDa polypeptide. A potential leucine zipper was found in the N-terminal half of the peptide molecule between amino acids 114 and 135. In addition, a different periodic leucine repeat with leucine at every eighth position was found between amino acids 325 and 373. The transcriptional initiation site of this early gene was determined by primer extension analysis. A putative TATA box (TATTTAA) located 24 bp upstream of the cap site and several inverted repeats were found in the region further upstream of the TATA box. To test whether the open reading frame of this cDNA encodes a virus-specific protein, the cDNA was overexpressed in *Escherichia coli* as a fusion protein used to generate antibodies in rabbits. A protein with a molecular size of 65 kDa was detected in the infected-cell extracts harvested at 6 to 72 h postinfection, but not in purified virions, using immunoblot analysis. Both nuclear and cytoplasmic fluorescences were found at late stages of virus infection. From the results obtained, we postulate that UL84 may be a stable, virus-specific, nonstructural protein capable of forming a homo- or heterodimeric molecule.

Human cytomegalovirus (CMV) is a linear double-strand DNA virus with a genome of approximately 150×10^6 Da or 240 kb. The genomic complexity of CMV makes it one of the largest animal DNA viruses discovered to date. This virus is associated with many clinical manifestations, ranging from developmental abnormality to fatal CMV pneumonitis in organ transplant and AIDS patients. The latency, reactivation, and oncogenic potentials of human CMV are comparable to those of other herpes group viruses. Human CMV gene expression is regulated in a cascade manner in permissive cells. Viral genes have been classified into three major groups based on their temporal order and the functional requirements for their expression (24, 49, 50). The first group of genes, designated immediate-early (IE) or alpha genes, is expressed within 2 to 4 h after infection without prior viral protein synthesis. These genes encode predominantly regulatory proteins with transactivating activities (1, 10, 12, 13, 37, 41, 42, 44–48). The next group of genes, designated early or beta genes, requires prior viral IE protein synthesis. This group of genes encodes predominantly nonstructural proteins, many of which are enzymes involved in viral DNA replication (2, 3, 6, 8, 11, 21, 41). With the onset of viral DNA replication, the expression of a third group of genes is initiated. This group of genes consists of late or gamma genes encoding virus-specific structural proteins (15, 16, 22, 23, 26–31, 36).

A number of CMV early genes have recently been studied, and viral IE proteins were commonly found to be needed for the activation of early promoters (2, 9, 16, 18, 19, 38). Analyses of sequences responsible for transactivation by IE proteins suggest that both direct and inverted repeats within promoter regions of these early genes might be involved in the regulation of early gene expression. However, no significant homology or consensus sequence could be detected among the repeat regions of various CMV early promoter sequences. The promoter region of a 2.2-kb RNA transcribed from the *Eco*RI fragments R and d contains two 9-base direct repeats, GCGGAAAGG and ACGTTGTTT (43). A palindromic octamer sequence, TCACGTGA, that is similar to the consensus binding site of MLTF/USF (CCACGTGA) of the promoter of adenovirus major late genes was found located between −112 and −105 of the promoter of a 2.7-kb RNA expressed from the *Eco*RI fragments O and W (18, 19). Likewise, the regulatory region of pp65 has an octameric sequence, ATTCGGG (9). In the case of the UL84 early gene, four pairs of inverted repeat sequences and a cyclic AMP (cAMP) response element were found within the promoter-enhancer sequences of the UL84 gene.

The UL84 gene is located within *Eco*RI fragment A, mapped within 0.534 to 0.545 map unit of the large unique segment of the human CMV genome. It is adjacent to the region encoding tegument proteins pp65 and pp71, as described by Ruger et al. (37). This gene was named UL84 in conformity with the nomenclature of Chee et al. (4). Given the unique promoter-enhancer sequence of UL84 and the extremely early appearance of its transcriptional and translational products, as compared with other CMV early genes,

* Corresponding author.

we performed detailed DNA sequence, transcription, and translation analyses of this unique gene.

MATERIALS AND METHODS

Virus and cells. Human embryonic lung (HEL) fibroblasts (ATCC HEL 229, passages 16 to 24) were used for the propagation of virus and biochemical studies. HEL cells were cultured in Eagle's minimal essential medium supplemented with 10% fetal calf serum. The Towne strain of human CMV was propagated in HEL cells by standard methods. All experiments were conducted at multiplicities of 1 to 2 PFU per cell. Extracellular virions were purified by sucrose and CsCl gradient centrifugation as described before (17).

Construction and screening of cDNA libraries. Total and cytoplasmic RNAs were isolated from human CMV-infected HEL cells at 72 h postinfection by the standard method (33). Polyadenylated RNAs were selected by oligo(dT)-cellulose column chromatography. Double-stranded cDNA was synthesized from poly(A) RNA by using oligo(dT) as a primer. The cDNA synthesized was methylated with *Eco*RI methylase prior to ligation with *Eco*RI linkers. After size fractionation through a Sepharose 4B column, cDNAs larger than 0.5 kb were inserted into the *Eco*RI sites of lambda gt11 and packaged in vitro, using packaging extracts obtained from Promega (Madison, Wis.). *Escherichia coli* Y1088 cells were used as hosts to amplify the recombinant phage. Two libraries which contain independent recombinants, 1.5×10^6 and 2×10^6, were constructed from total and cytoplasmic mRNAs, respectively.

Approximately 4×10^4 recombinant phages from the total mRNA library were screened with a ^{32}P-labeled CMV genomic DNA subclone, pHD1013-211, containing the pp65 and pp71 genes and the upstream region of the pp65 gene as described by Huynh et al. (14a). Positive plaques obtained from the first high-density screening were further purified through three successive rounds of DNA hybridization screening with low plaque densities. Recombinant phage DNA was prepared from plate lysates, digested with *Eco*RI, and then subcloned directly into bacteriophage M13mp19 at the *Eco*RI site. Another 4×10^4 recombinant phages from a cytoplasmic mRNA library were screened with the cDNA probe.

DNA cloning and Southern hybridization. Plasmid pHD1, containing the *Eco*RI A fragment of the CMV genome, and pHD1013-211, containing a 6.5-kb *Bam*HI fragment from pHD1, were constructed by Davis and Huang (7). pHD1 was digested with *Eco*RI, and the 27-kb *Eco*RI fragment A was separated by agarose gel electrophoresis. This fragment was digested with *Bam*HI and *Pst*I and transferred to nitrocellulose, and Southern hybridization was performed by standard methods (32). Prehybridization was carried out in a mixture of 25 mM KPO_4 (pH 7.4), 6× SSC (0.9 M NaCl, 0.09 M sodium citrate), 5× Denhardt's solution (0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin), 50 μg of salmon sperm DNA per ml, and 0.1% sodium dodecyl sulfate (SDS) at 42°C for 4 h. After prehybridization, ^{32}P-labeled cDNA (2×10^5 cpm/ml) in prehybridization buffer was added. Following hybridization, filters were washed twice in 1× SSC–0.1% SDS and twice in 0.25× SSC–0.1% SDS at room temperature for 15 min and then once in 0.25× SSC–0.1% SDS at 60°C for 1 h. A 2.6-kb *Pst*I fragment which hybridized with the cDNA probe, designated pHD1A, was recovered from an agarose gel and subcloned into bacteriophage M13mp19.

DNA sequencing and computer analysis. cDNA and genomic DNA sequencing was performed by the chain termination method (39), using the Sequenase kit (U.S. Biochemical Corp.) and Sequentide (NEG-034N; New England Nuclear). Single- and double-stranded templates were obtained from subclones constructed with appropriate restriction enzymes, including *Bam*HI, *Hin*dIII, *Sal*I, *Pst*I, and *Sph*I. All sequences were obtained and confirmed by using data from both directions. Eight synthetic oligonucleotides were applied as sequencing primers to resolve the sequence gaps and to confirm the strand inconsistencies. Open reading frame (ORF) analyses were carried out with the DNA Inspector II (Textco, West Lebanon, N.H.). Two-dimensional protein structure as well as hydrophilicity and hydrophobicity profiles of the deduced peptide were determined by using a software package provided by the Genetics Computer Group (University of Wisconsin, Madison).

Northern (RNA) hybridization. Total RNAs were prepared from mock- and CMV-infected HEL cells by the method described above. To prepare authentic IE RNA, cells were treated with 50 μg of cycloheximide (CH) per ml for 1 h before infection and then infected with virus in the presence of CH. Cells were harvested at 2.5 h postinfection. Total RNA (8 μg per well) was subjected to electrophoresis on a 1.2% agarose gel containing 6% formaldehyde and then transferred to nitrocellulose. The resulting filter was prehybridized at 42°C for 24 h and then hybridized with a ^{32}P-labeled cDNA probe at 42°C for 14 h as described above for Southern hybridization. After washing, the hybridized filter was autoradiographed overnight at −70°C.

The same blot was stripped of the ^{32}P-labeled cDNA probe by boiling in TE buffer (10 mM Tris-HCl [pH 7.5], 1 mM EDTA) for 15 min and reprobed with a ^{32}P-labeled cDNA of the gene for 28S rRNA.

Primer extension. Primer extension experiments were performed by the method of Kingston (34), with minor modification. A 20-bp oligonucleotide, 5′-TGCACCCGAGGCTGA CGGAT-3′, was synthesized on an Applied Biosystems synthesizer and end labeled with [γ-^{32}P]ATP by T4 polynucleotide kinase. Six micrograms of total RNA and 2×10^4 cpm of primer were mixed for 5 min at 80°C and then precipitated with alcohol. Pellets were dissolved in 20 μl of 1× hybridization buffer (80% formamide, 40 mM piperazine-*N*,N′-bis)2-ethanesulfonic acid) (PIPES; pH 6.4), 400 mM NaCl, 1 mM EDTA) and allowed to anneal at 30°C overnight. After ethanol precipitation, reverse transcriptase extension reaction was carried out at 42°C for 1.5 h in a final volume of 25 μl containing 18 U of avian myeloblastosis virus reverse transcriptase, 50 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$, 5 mM dithiothreitol, 50 mM KCl, 50 U of RNasin, and 50 μM deoxynucleoside triphosphates. After RNase digestion, phenol extraction, and ethanol precipitation, the pellets were dissolved in 4 μl of TE and 4 μl of gel loading buffer, heated at 80°C for 3 min, and loaded on a 6% polyacrylamide sequencing gel containing 7 M urea. The extension products were analyzed by comparison with genomic DNA sequencing reactions run in parallel. The control genomic DNA sequencing template was from a pHD1 subclone containing a 810-bp *Bam*HI-*Pst*I fragment in M13mp19.

Expression of the cDNA in *E. coli*. The glutathione *S*-transferase (GST) gene fusion system (pGEX) was used to express the cDNA as a fusion protein (40). The vector contains the carboxy terminus of the GST gene from *Schistosoma japonicum* under the control of a tac promoter. Downstream from this gene is a multiple cloning site con-

taining *Bam*HI, *Sma*I, and *Eco*RI restriction sites. To construct an in-frame clone, cDNA was inserted into the *Eco*RI site of pGEX-3X in the sense orientation to yield the expression plasmid pGEX-UL84A. Overnight cultures of pGEX-UL84A-transformed *E. coli* JM107 were diluted 1:10 with fresh medium and grown for 1.5 h at 37°C; 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) was added to the culture. After a further 1 to 6 h of growth, cells were pelleted and lysed in a 0.25 volume of SDS sample buffer (0.006 M Tris-HCl [pH 6.8], 4% SDS, 40% glycerol, 3% dithiothreitol, 0.005% bromphenol blue). The samples were boiled, applied to a 10% polyacrylamide gel, and subjected to polyacrylamide gel electrophoresis (PAGE).

Inclusion bodies containing the insoluble UL84A-GST fusion protein were prepared as described by Harlow and Lane (12a), with modification. The induced *E. coli* JM107 (pGEX-UL84A) cultures were pelleted by centrifugation at 7,000 × *g* for 5 min, and the cell pellets were resuspended in 100 mM NaCl–1 mM EDTA–50 mM Tris-HCl (pH 8.0) to a final concentration of 10% (vol/vol). The crude inclusion bodies were obtained by adding lysozyme to the suspension (1 mg/ml) and incubating the mixture at room temperature for 20 min. Subsequently, the suspension was sonicated for three 20-s cycles at full power with Branson model 200 cell disruptor with a microtip and centrifuged at 5,000 × *g* for 10 min. The inclusion bodies collected were washed subsequently with ice-cold washing buffer (100 mM NaCl, 1 mM EDTA, 50 mM Tris-HCl [pH 8.0]) with 0.1% SDS, washing buffer with 1% Nonidet P-40, and washing buffer without detergent. The final pellets were suspended in phosphate-buffered saline, analyzed for purity, and used as the antigen for antibody production.

Antibody preparation. Antiserum against the UL84A-GST fusion protein was made by immunizing rabbits with 100 μg of the fusion protein in the form of inclusion bodies in an equal volume of Freund's complete adjuvant (Calbiochem Corp.) and boosted three times with fusion proteins in an equal volume of Freund's incomplete adjuvant (Calbiochem) at 2-week intervals. At 12 days after the final injection, immune rabbit serum was tested for the presence of the specific anti-CMV antibodies by Western immunoblot analysis and immunofluorescent staining.

Western blot and indirect immunofluorescence. HEL cells were infected with human CMV as described above. At 6, 24, and 72 h postinfection, cells were washed twice with ice-cold phosphate-buffered saline and then lysed with SDS sample buffer. The lysates were sheared by passage four times through a 20-gauge needle and then boiled for 5 min. The purified virions were also lysed with the SDS sample buffer. Twenty micrograms of protein was loaded onto an SDS–10% polyacrylamide gel. Control mock-infected cultures were harvested with the same procedure. After electrophoresis, proteins were transferred to nitrocellulose in transfer buffer (20 mM Tris-HCl [pH 8.0], 150 mM glycine, 20% methanol) overnight at 14 V. The blot was washed four times with washing buffer (4% dry milk in phosphate-buffered saline) at room temperature for 25 min each time and then probed with the UL84A-GST antiserum (diluted 1:100 in washing buffer) at room temperature for 1 h. The antiserum was removed, and the blot was rinsed three times with washing buffer. Detection of bound antibody-antigen complexes was achieved by incubating the filter with ^{125}I-protein A (0.3 μCi/ml; Amersham) in 10 ml of washing buffer at room temperature for 1 h. After drying, the filter was exposed to X-ray film at −70°C for overnight.

For indirect immunofluorescence, CMV-infected HEL cells on eight-well chamber slides (Nunc, Naperville, Ill.) were harvested at the indicated times after infection. Cells were washed with 0.01 M phosphate-buffered saline (pH 7.0) three times and fixed with ice-chilled acetone for 15 min. Infected and mock-infected cultures were stained first with rabbit antiserum and subsequently with fluorescein-conjugated goat anti-rabbit antibody for 1 h each at 37°C.

Nucleotide sequence accession number. The GenBank accession number for the sequence reported in this communication is M81432.

RESULTS

Construction of cDNA libraries and isolation of cDNA clones. As described in Materials and Methods, two cDNA libraries were constructed from the total RNA and cytoplasmic RNA of CMV-infected HEL cells. The first six clones with positive signals were obtained from the screening of 4 × 10^4 recombinant phages from the total RNA library by using a ^{32}P-labeled 6.5-kb *Bam*HI fragment of CMV DNA which contains pp65 and pp71 genes and the upstream regions of these two genes. After further plaque purification and *Eco*RI digestion, all six isolates were shown to have DNA inserts with the same molecular size of 1.6 kb. Digestion of these isolates with *Bam*HI, *Hin*dIII, *Sal*I, and *Pst*I revealed that all six clones contained the same restriction fragment. The insert from one lambda gt11 clone was subsequently cloned into the *Eco*RI site of plasmid M13mp19 (designated M13-UL84A) for DNA sequence analysis.

To better understand the transcription of this cDNA, we screened the cytoplasmic RNA library by using the 1.6-kb cDNA probe. Fourteen clones with positive signals were obtained from the screening of 4 × 10^4 recombinants. After *Eco*RI digestion of the DNA isolated from positive recombinants, inserts of different molecular sizes (from 0.5 to 2.0 kb) were found. Four DNA inserts (2.0, 1.85, 1.7, and 1.65 kb) larger than UL84A were subsequently cloned into *Eco*RI site of M13mp19 (designated M13-UL84 and M13-UL84-1, -2, and -3, respectively) for DNA sequence analysis.

To map the locations of these cDNAs within the CMV genome, the radiolabeled 1.6-kb cDNA (UL84A) fragment was hybridized to the *Eco*RI A fragment digested with *Pst*I and *Bam*HI. The result revealed that the cDNA was located within a 2.6-kb *Pst*I-*Pst*I subfragment or within a 6.5-kb *Bam*HI-*Bam*HI subfragment of *Eco*RI-A (data not shown). This 2.6-kb *Pst*I subfragment was then subcloned into M13mp19 to determine the nucleotide sequence. This construct was designated pHD1A.

Nucleotide and predicted amino acid sequences. Plasmids M13-UL84A and pHD1A, which contain the first 1.6-kb cDNA fragment and the 2.6-kb *Pst*I genomic fragment, respectively, were sequenced by the dideoxy-chain termination method. The first group of DNA sequencing data indicated that 1.6-kb cDNA was located upstream of the pp65 gene (Fig. 1). Comparison of M13-UL84A and pHD1A nucleotide sequences revealed a 31-bp intron (GCGGAAC TTACCATCTAATCCAGTTGCACCT) located within the very 5′ end of this 1.6-kb cDNA. The first exon of the 1.6-kb cDNA contains only 8 bp. Subsequently, the sequences of M13-UL84 and M13-UL84-1, -2, and -3 were completed. All of the clones show no intron and coterminal 3′ ends. These results indicate that the 31-bp intron appearing in the UL84A 1.6-kb cDNA may have be an artifact generated during cDNA synthesis. The complete nucleotide sequences of the 2.0-kb UL84 cDNA and the corresponding 2.6-kb *Pst*I fragment of genomic DNA are shown in Fig. 2. The genomic

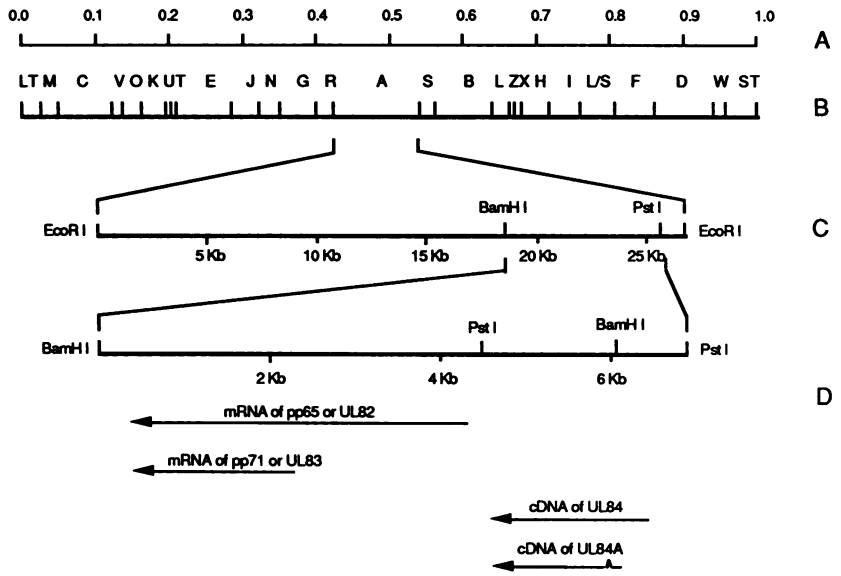

FIG. 1. Physical map of the Towne strain human CMV genome and localization of the UL84 gene. (A) Fractional length of the CMV DNA molecule; (B) schematic representation of the CMV genome with the restriction map for *Eco*RI; (C) 27-kb *Eco*RI fragment of the HCMV genome, designated pHD1; (D) RNA transcripts of pp65, pp71, and UL84. The orientation of transcription is indicated by arrows.

*Pst*I DNA fragment contains 2,600 bp, and the cDNA sequence (shown under the genomic DNA sequence) has 1,990 bp. A single ORF capable of encoding a 587-amino-acid peptide with a predicted molecular size of 65 kDa can be found within this region. The first in-frame initiation codon is found at nucleotide 439, and nucleotides surrounding the first ATG of the ORF follow the Kozak rules for the initiation of effective translation (20). The coding region ends with a TAA termination codon immediately downstream of the leucine residue at nucleotide 2200. It is followed by 143 bp of 3'-end untranslated region containing a putative polyadenylation signal, AATAAA, which is 124 bp from the termination codon.

Figure 3 shows the predicted secondary peptide structure, including hydrophilicity and hydrophobicity, antigenic domains, and peptide flexibility of the UL84 peptide obtained by computer analysis using the method of Chou and Fasman (5). Study of the hydropathy plot reveals several distinct hydrophilic regions located within the N-terminal half of the peptide molecule, particularly the end and the region from amino acid residues 150 to 185. The hydrophobic regions were clustered at the C terminus. The antigenic domains were centered at each terminus and also at regions with higher flexibility. Five putative N-linked glycosylation sites (NXT and NXS) at positions 28, 279, 354, 376, and 570 were found. In addition, this peptide contains a high number of leucine residues (13%). A putative leucine zipper with periodic leucine repeats at every seventh position was found between amino acids 114 and 135. Another form of periodic leucine repeats on every eighth amino acid residues was found in the region between amino acid residues 325 and 373. A cluster of six threonine residues was observed at residues 53 to 58. A highly charged region with the amino acid sequence EKKKEKEEKKEEEDEDD was found beginning at amino acid 161.

Nucleotide sequence variations between strains Towne and AD169 in the UL84 gene are shown in Fig. 2. Sequence variations of approximately 2.8% (nucleotide) and 2.7% (amino acid) were found.

Transcript analysis. Figure 4 shows the results of Northern blot hybridization. An RNA species of 2.0 kb was first observed as early as 2.5 h postinfection, but this message was not seen in the RNA sample prepared in the presence of the protein synthesis inhibitor CH, which implies that de novo protein synthesis is required for the transcription of UL84. When the sample was hybridized to the same amount of RNA from 6 to 24 h postinfection, the level of this 2.0-kb message was lower than that at 2.5 h postinfection. However, the maximal level of transcription appeared between 72 and 96 h postinfection. This bimodal expression might be due to the second run of gene expression induced by the infection from the newly synthesized progeny. By late times in the infection, additional transcripts of 3.2 and 8.5 kb were seen. To confirm these observations and monitor the amount of RNA applied, the same blot was stripped of the UL84 cDNA probe and reprobed with a cDNA probe of the 28S rRNA gene as an internal control. As shown in Fig. 4b, the amounts of RNA samples loaded from mock-infected cells, CH-treated infected cells harvested at 2.5 h postinfection, and CMV-infected culture harvested at various times after infection are in a very narrow range. Results from these studies suggest that the UL84 gene is a very early gene which can be detected as early as 2.5 h after virus infection. The IE gene products are needed for the transactivation of UL84 gene expression.

To determine the location and sequence of the transcription initiation site, primer extension experiments were performed. Total RNAs were hybridized with a ^{32}P-end-labeled oligonucleotide primer (20-mer) complementary to the genomic DNA sequences from bases 728 to 747 (Fig. 2). The result revealed that the RNA start site was located around 302 bp upstream from the oligonucleotide primer, at base 426 (Fig. 5). A putative TATA box (TATTTAA) was found 24 bases upstream from the cap site. Several inverted repeat sequences, designated IR1, IR2, IR3, and IR4 (Fig. 2), also were found at the upstream region from nucleotides 233 to 334. No typical CAAT box was found around −75. However, a classical CAAT box (GGCCAATCT) was located

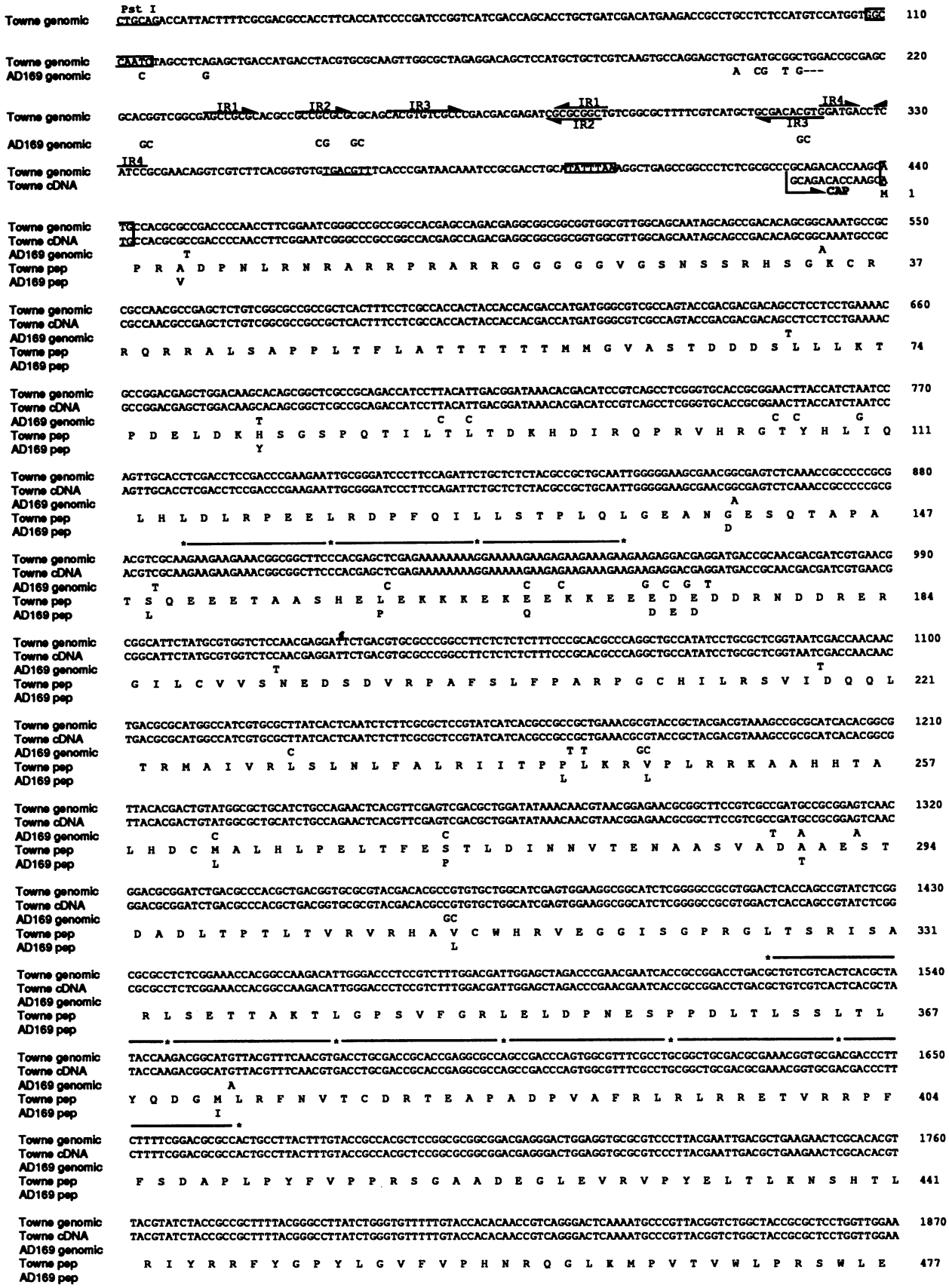

FIG. 2. Comparison of the nucleotide and deduced amino acid sequences of the UL84 genes of strains Towne and AD169. A potential TATA consensus sequence and a CAAT box are boxed; the transcription initiation site identified by primer extension is indicated by an arrow plus CAP; several inverted repeat sequences (IR1, -2, -3, and -4) are shown; a cAMP response element consensus sequence (TGACGTT) is underlined; the translation initiation site and the polyadenylation signal sequence are also underlined; and the leucine repeats are indicated by asterisks and lines.

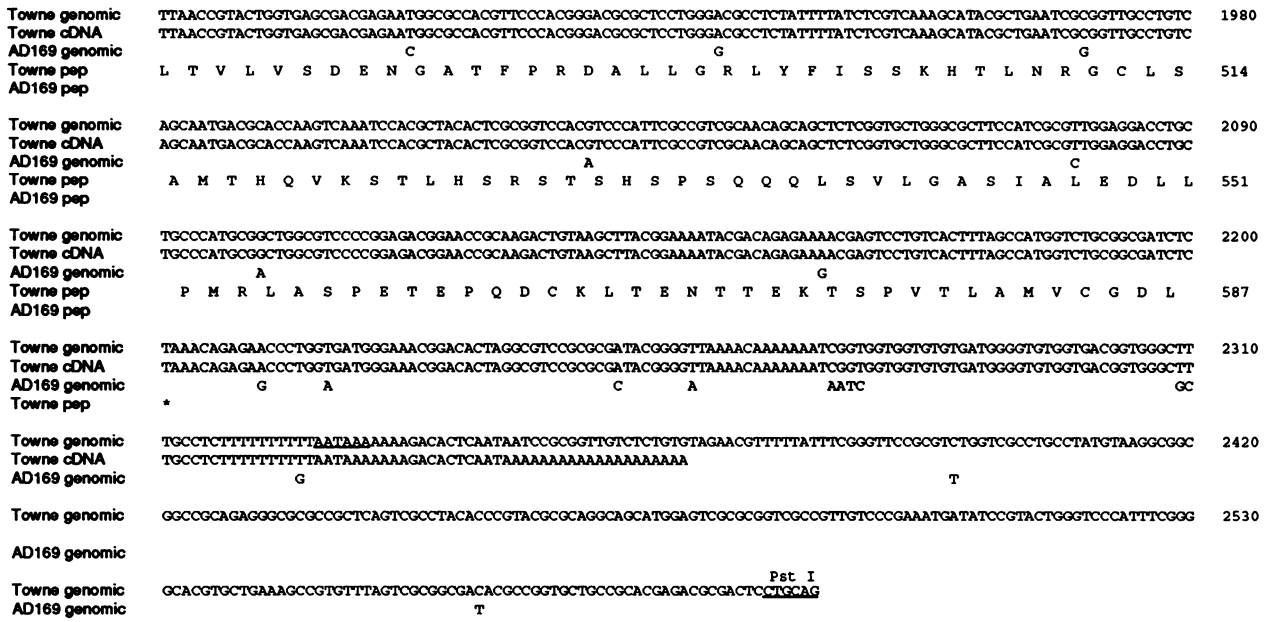

FIG. 2—*Continued.*

further upstream from the cap site of the inverted repeats at −311 to −319.

Expression of UL84A cDNA as a fusion protein in *E. coli*. The expression plasmid pGEX-UL84A was constructed by inserting a UL84 cDNA *Eco*RI fragment into the *Eco*RI cloning site of pGEX-3X. The vector pGEX-3X contains a 26-kDa GST gene derived from the parasitic helminth *S. japonicum* and is under the control of a tac promoter. After inspection of the cDNA sequences, we found that the 1.6-kb UL84A cDNA can be directly inserted into the *Eco*RI site of pGEX-3X to express it as an in-frame fusion protein with the 31-bp intron at the 5′ end. The reading frame and coding sequence were identical to those of the full-length 2.0-kb cDNA UL84 after amino acid 115. Therefore, a C-terminal fusion protein with a molecular size of 80 kDa (26 kDa of GST and 54 kDa of UL84) was successfully overexpressed. Figure 6 showed the time course of the expression of this fusion protein in *E. coli* after IPTG induction (an 80-kDa Coomassie band is indicated by an arrow). To determine whether the expressed fusion protein was aggregated in inclusion bodies, the cells were lysed with lysozyme and then subjected to several differential low-speed centrifugations to purify inclusion bodies. The inclusion body pellet and the cytoplasmic supernatant were analyzed by SDS-PAGE. Data from this experiment revealed that more than 90% of the fusion protein synthesized was present as inclusion bodies.

Figure 7a shows that a 65-kDa protein could be detected in infected-cell extracts harvested between 6 to 72 h postinfection by Western blot analysis using the rabbit antiserum against the UL84A-GST fusion protein. This protein was not present in mock-infected cell extracts. An additional protein with molecular size of 90 kDa was seen in the infected-cell extracts at 72 h postinfection. Figure 7b shows the Western blot analysis of purified virus. UL84 protein was not detected from at least two virion preparations, which indicated that the UL84 gene product is not a structural protein.

Figure 8 shows the synthesis and location of UL84 protein at various times after infection, as determined by indirect immunofluorescence. Perinuclear localization of this protein was observed at 6 to 7 h after virus infection. By 24 h, nuclear fluorescence could be demonstrated. Heavy but specific fluorescent staining was exhibited at both nuclear and cytoplasmic regions of CMV-infected cells at later stages of infection (72 h postinfection).

DISCUSSION

The goal of this study was to identify early or late transcripts of human CMV genes in an attempt to better understand the control of viral gene expression in CMV-infected cells. Recently, a number of early and late genes have been mapped (3, 6, 7, 12, 14–16, 21–31, 36, 37, 41), allowing for specific analysis of their regulated expression. The results suggested that various CMV early or late genes are regulated differently (2, 9, 18, 19, 22, 38, 42). In this report, we describe the complete nucleotide sequences, RNA mapping, and expression of a human CMV early transcript, UL84.

Sequence analysis of UL84 cDNA revealed an ORF capable of encoding a 65-kDa peptide. According to the analysis of the protein-coding content of the sequence of human CMV strain AD169 described by Chee et al. (4), a UL84 ORF encoding a 65-kDa peptide and located at the immediate upstream region of the UL83 or pp65 gene was predicted. Because of the identities between our full-length cDNA and the UL84 ORF predicted by Chee et al. with respect to peptide size and the position of the translation initiation codon, we use the same name, UL84, to designate the gene for this 2.0-kb cDNA. The cDNAs UL84-1, -2, and -3 were non-full-length cDNAs, as judged by the features of 3′ ends coterminal with that of the 2.0-kb cDNA. However, the first cDNA, UL84A, isolated from a total RNA library, contains an 31-bp intron. An ORF encoding a 364-amino-acid, 41-kDa peptide could be deduced from the exon 2 region of the UL84A 1.6-kb cDNA sequence. This 41-kDa peptide has a carboxy end coterminal with that of the peptide deduced from the full-length UL84 cDNA, but the typical

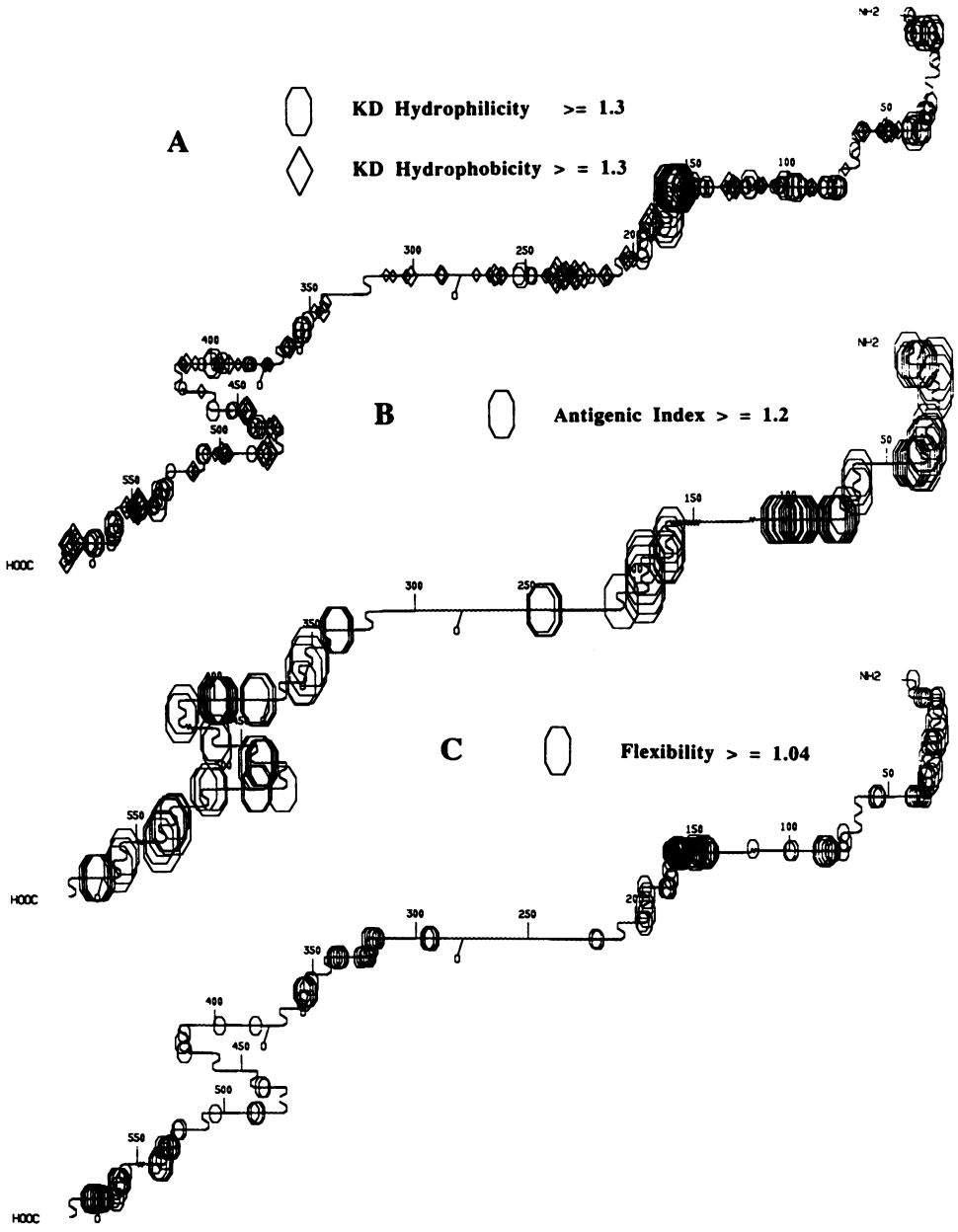

FIG. 3. Two-dimensional structure of the UL84 peptide obtained by Chou-Fasman computer analyses. The amino and carboxy termini are indicated. KD, Kyte-Doolittle.

sequences of splicing donor and acceptor were not found at the 3′ end of exon 1 and 5′ end of exon 2 of the UL84A sequence. This implies that 31-bp splicing within the 1.6-kb cDNA might be an artifact. Western blot analysis using a rabbit anti-UL84A fusion protein further confirmed this observation; a protein of 41-kDa could not be detected in the infected-cell extract.

UL84, a member of the first subclass of human CMV early genes, is transcribed into a 2.0-kb mRNA which appears as early as 2.5 h postinfection. Primer extension analysis of this mRNA revealed a putative transcription initiation site within a region about 14 nucleotides upstream of the first ATG of the UL84 ORF, which was only one base further than that of the UL84 cDNA start site. This analysis suggested that the UL84 cDNA was short one nucleotide at the 5′ end. The mRNA size estimated from the primer extension and the sequence of the UL84 cDNA matches perfectly the size of the message detected from Northern hybridization analysis. To investigate whether a putative functional promoter is located within the upstream region of the cap site of the UL84 gene, we have analyzed the sequences from the cap site to the 5′-end *Pst*I site. Three structural features were observed: (i) a TATTTAA sequence located 23 nucleotides upstream of the cap site that might represent a TATA motif, (ii) several inverted repeat sequences located within the upstream region of the putative TATA box (IR1 [AGC CGCGC], IR2 [GCCGCGCG], IR3 [CACGTGTC], and IR4 [CGTGGATGACCTTCATCCGCG]), and (iii) the absence

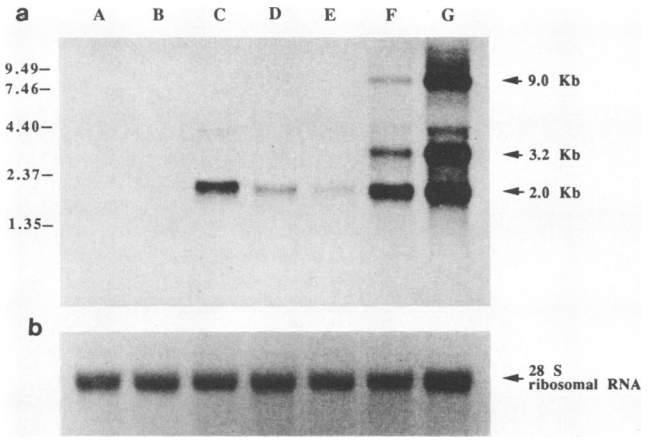

FIG. 4. Northern blot analysis of the UL84 transcript. Total RNA was prepared from mock-infected (lane A) and virus-infected (2.5 h plus CH, 2.5 h, 6 h, 24 h, 72 h, and 96 h postinfection; lanes B to H, respectively) HEL cells by the guanidinium isothiocyanate-cesium chloride method. Eight micrograms of total RNA was subjected to electrophoresis through a 1.2% agarose gel containing 6% formaldehyde and then transferred to nitrocellulose. The resulting filter was hybridized with ^{32}P-labeled UL84A cDNA (a) and with ^{32}P-labeled cDNA of 28S rRNA (b). Locations of the 2.0-, 3.2-, and 9.0-kb UL84 transcripts and of 28S rRNA are indicated by arrows. The RNA ladder on the left is shown in kilobases.

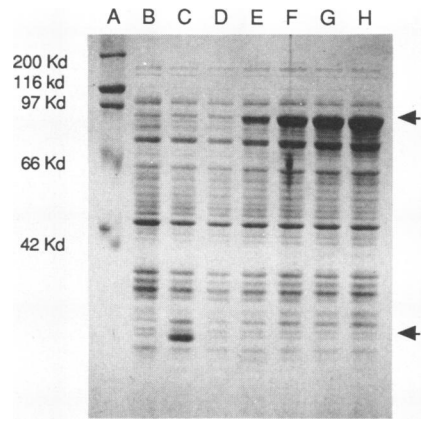

FIG. 6. Expression of UL84A cDNA in *E. coli* JM107. Overnight cultures of JM107 (B) and of JM107 transformed with expression vector pGEX-3X (lane C) and recombinant vector pGEX-UL84A (lanes D to H) were diluted 1:10 in YT medium, grown for 1 h, and then induced with IPTG. Samples were harvested at 0, 0.5, 1, 2, and 4 h (lanes D to H, respectively), subjected to SDS-PAGE analysis, and stained with Coomassie blue. Lane A indicates protein molecular sizes in kilodaltons. Arrows show the positions of the expressed fusion protein (80 kDa) and GST (26 kDa).

of a typical CAAT box around −75 from the cap site; instead, there is a classic CAAT box sequence (GGCCAA TCT) located upstream of the inverted repeats at −311 to −319 from the cap site. These inverted repeat sequences may be involved in the regulation of expression of the UL84 gene, but computer analysis of a number of other CMV early promoters and cellular promoters did not reveal any significant homology to these inverted repeats. It is possible that only a few nucleotides are needed for specific interactions between DNA and proteins for promoter activation. To verify the existence of promoter activity within this region, we constructed a chloramphenicol acetyltransferase (CAT) expression plasmid by fusing a *Pst*I-*Sph*I fragment, containing the upstream region of the UL84 gene, to the CAT indicator plasmid pKK232-8. Significant promoter activity was detected from this sequence after CMV induction (12b).

Prepared UL84 antiserum and immunofluorescence stain-

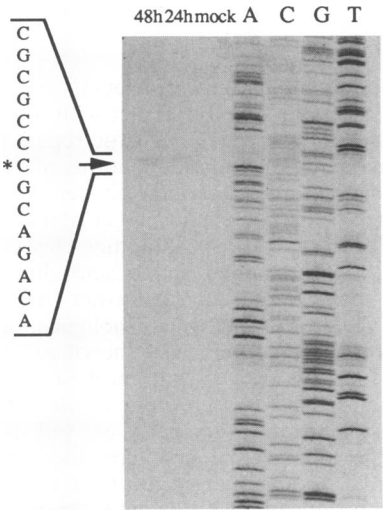

FIG. 5. Primer extension analysis. A synthetic 20-base oligonucleotide, 5′-TGCACCCGAGGCTGACGGAT-3′, was end labeled by T4 polynucleotide kinase and [^{32}P]ATP. The end-labeled primer was annealed to 6 μg of total RNA from mock- and virus-infected HEL cells at the indicated times. Avian myeloblastosis virus reverse transcriptase was used for primer extension. The corresponding complementary nucleotides were sequenced in parallel to allow direct reading of the coding sequence. The transcription initiation site (asterisk) predicted from the primer extension is shown at the left.

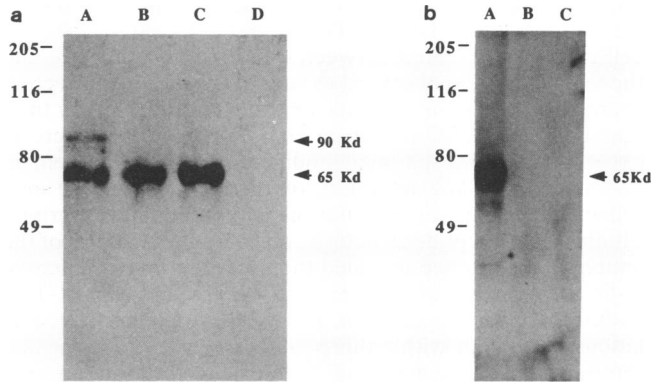

FIG. 7. Immunoblot of the purified extracellular human CMV virions and the proteins extracted from human CMV-infected cells. After SDS-PAGE separation, the proteins were transferred to nitrocellulose and probed with rabbit antiserum against the UL84A-GST fusion protein. ^{125}I-protein A was used to monitor the reacted antibody. (a) Proteins extracted from human CMV-infected (72, 24, and 6 h postinfection [lanes A to C, respectively]) and mock-infected (lane D) HEL cells; (b) proteins extracted from human CMV-infected cells harvested at 24 h postinfection (lane A), from purified virions (lane B), and from mock-infected cells (lane C). Positions of molecular weight markers are shown at the left in kilodaltons; the arrows indicate the immunoreactive proteins (65 and 90 kDa).

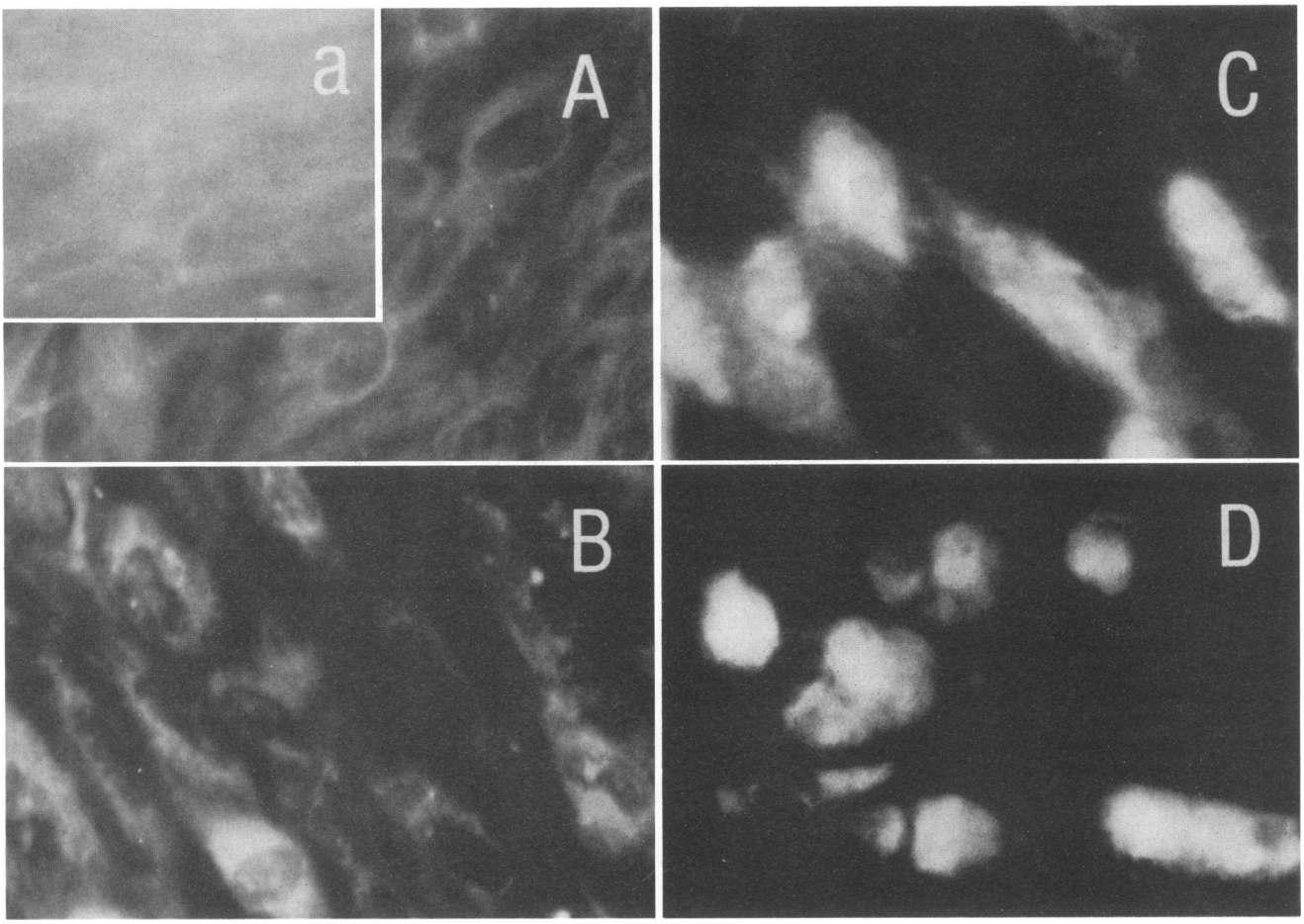

FIG. 8. Indirect immunofluorescence test. The mock-infected (a) and human CMV-infected (6, 24, 48, and 72 h postinfection [A to D, respectively]) HEL cells were reacted with the first rabbit antibody and stained with fluorescein-conjugated goat anti-rabbit antibody.

ing detected functional virus-specific protein in CMV-infected HEL cells. The 65-kDa protein detected in infected-cell extracts harvested between 6 to 72 h postinfection and the 90 kDa protein observed at later stages of virus infection share some common antigenic epitopes, as reflected by their reactivity to UL84 antiserum. It is possible that there is extensive posttranslational modification of this protein in late stages of CMV infection, or this 90-kDa peptide may reflect the translation product of the additional transcripts. Studies of the peptide structure and computer analysis of the amino acid sequence revealed the presence of five potential N-linked glycosylation sites (Asn-X-Ser and Asn-X-Thr) within the ORF. In addition, many other O-linked glycosylations may occur within this peptide because the serine and threonine residues make up 13% of the total amino acid composition, and they are the acceptor sites for O-linked oligosaccharides. These two amino acid residues are heavily distributed at both N- and C-terminal ends of this peptide.

The leucine zipper found in the region between amino acids 114 and 135 at the N-terminal half of molecule implies that this peptide may be involved in downstream gene regulation via dimerization, with itself or other molecules, and DNA binding. In addition, another periodic array of leucine on every eighth amino acid was found in the region between amino acids 325 and 349. The arrangement of this periodic leucine repeat is $L(X_7)L(X_7)L(X_7)L(X_7)P(X_7)L(X_7)L(X_7)$. This unique class of leucine repeat might be able to form a leucine coil to interact with other macromolecules. The significance of this leucine repeat has not been analyzed. Two additional unique features of the UL84 peptide were observed: a six-threonine cluster at positions 53 to 58 and the highly charged amino acid cluster at positions 161 to 177. The function of the six-threonine cluster is not clear. Whether the highly charged amino acid cluster associates with nuclear localization requires further study.

UL84 is a very early gene whose biological function is not clear. The translated protein was shown to be very stable and was not found associated with purified virions. It could, therefore, be a regulatory protein or an enzyme which performs its function at very early as well as late stages. Studies of the biochemical characteristics of the protein encoded by UL84 and identification of the essential *cis*-acting elements controlling UL84 gene expression are under way.

ACKNOWLEDGMENTS

We thank Timothy Kowalik, Bret Wing, David Jensen, and Belinda Wilson for useful discussions and help in preparation of the manuscript and Shu-Mei Huong, Yaw-Wen Guo, and Belinda Wilson for help in antibody production and technical assistance.

This investigation was supported by Public Health Service grants

AI 12717, AI 25868, and CA 21773 from National Institutes of Health.

REFERENCES

1. **Boshart, M., F. Weber, G. Jahn, K. Dorsch-Hasler, B. Fleckenstein, and W. Schaffner.** 1985. A very strong enhancer element located upstream of an immediate early gene of human cytomegalovirus. Cell **41**:521–530.
2. **Chang, C.-P., C. L. Malon, 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.
3. **Chang, C.-P., D. H. Vesole, J. Nelson, M. B. A. Oldstone, and M. F. Stinski.** 1989. Identification and expression of a human cytomegalovirus early glycoprotein. J. Virol. **63**:3330–3337.
4. **Chee, M. S., A. T. Bankier, S. Beck, R. Bohni, C. M. Brown, R. Cerny, T. Horsnell, C. A. Hutchison III, T. Kouzarides, J. A. Martignetti, E. Preddie, S. C. Satchwell, P. Tomlinson, K. M. Weston, and B. G. Barrell.** 1990. Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. Curr. Top. Microbiol. Immunol. **154**:125–169.
5. **Chou, P. Y., and G. D. Fasman.** 1974. Prediction of proteins conformations. Biochemistry **13**:222–245.
6. **Cranage, M. P., G. L. Smith, S. E. Bell, H. Hart, C. Brown, A. T. Bankier, P. Tomlinson, B. G. Barrell, and T. C. Minson.** 1988. Identification and expression of a human cytomegalovirus glycoprotein with homology to the Epstein-Barr virus BXLF2 product, varicella-zoster virus gpIII, and herpes simplex virus type 1 glycoprotein H. J. Virol. **62**:1416–1422.
7. **Davis, M. G., and E. S. Huang.** 1985. Nucleotide sequence of a human cytomegalovirus DNA fragment encoding a 67-kilodalton phosphorylated viral protein. J. Virol. **56**:7–11.
8. **Davis, M. G., E. C. Mar, Y. M. Wu, and E. S. Huang.** 1984. Mapping and expression of a human cytomegalovirus major viral protein. J. Virol. **52**:129–135.
9. **Depto, A. S., and R. M. Stenberg.** 1989. Regulatory 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.
10. **Ghazal, P., H. Lubon, and L. Hennighausen.** 1988. Specific interaction between transcription factors and the promoter-regulatory region of the human cytomegalovirus major immediate-early gene. J. Virol. **62**:1076–1079.
11. **Ghazal, P., H. Lubon, and L. Hennighausen.** 1988. Multiple sequence-specific transcription factors modulate cytomegalovirus enhancer activity in vitro. Mol. Cell. Biol. **8**:1809–1811.
12. **Greenaway, P. J., and G. W. G. Wilkinson.** 1987. Nucleotide sequence of the most abundantly transcribed early gene of human cytomegalovirus strain AD 169. Virus Res. **7**:17–31.

12a. **Harlow, E., and D. Lane.** 1988. Antibodies—a laboratory manual, p. 90–91. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

12b. **He, Y. S., et al.** Unpublished data.

13. **Hennighausen, L., and B. Fleckenstein.** 1986. Nuclear factor 1 interacts with five DNA elements in the promoter region of the human cytomegalovirus major immediate early gene. EMBO J. **5**:1367–1371.
14. **Hunnighake, G. W., M. M. Monick, B. Liu, and M. F. Stinski.** 1989. The promoter-regulatory region of the major immediate-early gene of human cytomegalovirus responds to T-lymphocyte stimulation and contains functional cyclic AMP response elements. J. Virol. **63**:3026–3033.

14a. **Huynh, T. V., R. A. Young, and R. N. Davis.** 1985. Constructing and screening cDNA libraries in lambda gt10 and lambda gt11. DNA Cloning **1**:49–78.

15. **Jahn, G., T. Kouzarides, M. Mach, B.-C. Scholl, B. Plachter, B. Traupe, E. Preddie, S. C. Satchwell, B. Fleckenstein, and B. G. Barrell.** 1987. Map position and nucleotide sequence of the gene for the large structural phosphoprotein of human cytomegalovirus. J. Virol. **61**:1358–1367.
16. **Jahn, G., B.-C. Scholl, B. Traupe, and B. Fleckenstein.** 1987. The two major structural phosphoproteins (pp65 and pp150) of human cytomegalovirus and their antigenic properties. J. Gen. Virol. **68**:1327–1337.
17. **Kilpatrick, B. A., and E. G.-S. Huang.** 1977. Human cytomegalovirus genome: partial denaturation map and organization of genome sequence. J. Virol. **24**:261–276.
18. **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.
19. **Klucher, K. M., and D. H. Spector.** 1990. The human cytomegalovirus 2.7-kilobase RNA promoter contains a functional binding site for the adenovirus major late transcription factor. J. Virol. **64**:4189–4198.
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. **Kouzarides, T., A. T. Bankier, S. C. Satchwell, K. Weston, P. Tomlinson, and B. G. Barrell.** 1987. Sequence and transcription analysis of the human cytomegalovirus DNA polymerase gene. J. Virol. **61**:125–133.
22. **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. **Mach, M., U. Utz, and B. Fleckenstein.** 1986. Mapping of the major glycoprotein gene of human cytomegalovirus. J. Gen. Virol. **67**:1461–1467.
26. **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.
27. **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. **62**:531–538.
28. **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.
29. **Mocarski, E. S., L. Pereira, and N. Michael.** 1985. Precise localization of genes on large animal virus genomes: use of gt11 and monoclonal antibodies to map the gene for a cytomegalovirus protein family. Proc. Natl. Acad. Sci. USA **82**:1266–1270.
30. **Nowak, B., A. Gmeiner, P. Sarnow, A. Levine, and B. Fleckenstein.** 1984. Physical mapping of human cytomegalovirus genes: identification of DNA sequences coding for a virion phosphoprotein of 71 KDa and for a viral 65 KDa polypeptide. Virology **134**:91–102.
31. **Pande, H., S. W. Baak, A. D. Riggs, B. R. Clark, J. Shavely, and J. A. Zaia.** 1984. Cloning and physical mapping of a gene fragment coding for a 64-kilodalton major late antigen of human cytomegalovirus. Proc. Natl. Acad. Sci. USA **81**:4965–4969.
32. **Selden, R. F.** 1987. Analysis of DNA sequences by blotting and hybridization, p. 2.9.1.–2.9.10. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology. John Wiley & Sons, New York.
33. **Kingston, R. E.** 1987. Guanidinium method for total RNA preparation, p. 4.2.1.–4.5.3. Current Protols In Molecular Biology. John Wiley & Sons, New York.
34. **Kingston, R. E.** 1987. Primer extension, p. 4.8.1.–4.8.3. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology. John Wiley & Sons, New York.
35. **Kingston, R. E.** 1987. Harvest and assay for chloramphenicol acetyltransferase, p. 9.6.3.–9.6.4. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and

K. Struhl (ed.), Current protocols in molecular biology. John Wiley & Sons, New York.
36. **Lahijani, R. S., Elmer, W. Otteson, J. D. Adlish, and S. C. St. Jeor.** 1991. Characterization of a human cytomegalovirus 1.6-kilobase late mRNA and identification of its putative protein product. J. Virol. **65**:378–381.
37. **Ruger, B., S. Klages, B. Walla, J. Albrecht, B. Fleckenstein, P. Tomlinson, and B. Barrell.** 1987. Primary structure and transcription of the genes coding for the two virion phosphoproteins pp65 and pp71 of human cytomegalovirus. J. Virol. **61**:446–453.
38. **Sambucetti, L., J. M. Cherrington, G. W. G. Wilkinson, and E. S. Mocarski.** 1989. NF-κB activation of the cytomegalovirus enhancer is mediated by a viral transactivator and by T cell stimulation. EMBO J. **8**:4251–4258.
39. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
40. **Smith, D. B., and K. S. Johnson.** 1988. Single-step purification of polypeptides expressed in Escherichia coli as fusion with glutathione S-transferase. Gene **67**:31–40.
41. **Spaete, R. R., and E. S. Mocarski.** 1985. Regulation of cytomegalovirus gene expression: α and β promoters are *trans*-activated by viral function in permissive human fibroblasts. J. Virol. **56**:135–143.
42. **Staprans, S. I., D. K. Rabert, and D. H. Spactor.** 1988. Identification of sequence requirements and *trans*-acting functions necessary for regulated expression of a human cytomegalovirus early gene. J. Virol. **62**:3463–3473.
43. **Staprans, S. I., and D. H. Spector.** 1986. 2.2-Kilobase class of early transcripts encoded by human cytomegalovirus strain AD169. J. Virol. **57**:591–602.
44. **Stenberg, R. M., and M. F. Stinski.** 1985. Autoregulation of the human cytomegalovirus major immediate-early gene. J. Virol. **56**:676–682.
45. **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.
46. **Stenberg, R. M., P. R. Witte, and M. F. Stinski.** 1984. Multiple spliced and unspliced transcripts from human cytomegalovirus immediate-early region 2 and evidence for a common initiation site within immediate-early region 1. J. Virol. **56**:665–675.
47. **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.
48. **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.
49. **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.
50. **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.