JOURNAL OF VIROLOGY, Feb. 1989, p. 669–676
0022-538X/89/020669-08$02.00/0

Vol. 63, No. 2

Primate Cytomegalovirus Assembly Protein: Genome Location and Nucleotide Sequence

LORI ROBSON AND WADE GIBSON*

Virology Laboratories, Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205

Received 6 September 1988/Accepted 2 November 1988

A cDNA encoding the 37-kilodalton (kDa) capsid assembly protein of cytomegalovirus (CMV) strain Colburn was isolated from a λgt11 library constructed from CMV Colburn-infected human fibroblast RNA. RNA transcribed in vitro from this cDNA was translated in vitro to give a 40-kDa protein whose electrophoretic mobility during sodium dodecyl sulfate-polyacrylamide gel electrophoresis and fragmentation pattern following partial proteolysis were indistinguishable from those of authentic assembly protein precursor. The position of the assembly protein gene was mapped to the *Eco*RI F, *Xba*I R, and *Sal*I U restriction fragments, near the middle of the CMV Colburn genome, by Southern hybridizations using the cloned assembly protein cDNA as a probe. Similar sequences were identified by cross-hybridizations in colinear regions of the genomes of human CMV strains Towne and AD169: specifically, in the *Hin*dIII H, *Bam*HI V, and *Eco*RI A fragments of Towne and in the *Hin*dIII L and *Bgl*II S fragments of AD169. The predominant transcript of the assembly protein gene was determined to be approximately 1 kilobase in size; however, a larger transcript (1.8 kilobases) was also identified. The nucleotide sequence of the assembly protein cDNA was determined and found to contain a single long open reading frame predicted to encode a polypeptide of 36.6 to 37 kDa, close to the 40-kDa size determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the assembly protein precursor. The presence of a single cysteine residue at the carboxy-terminal end of this open reading frame is consistent with data from biochemical studies and indicates that processing of the assembly protein precursor includes a proteolytic cleavage that removes its carboxyl end.

The principal intranuclear capsid form recovered from fibroblast cells infected with human or simian strains of cytomegalovirus (CMV) is referred to as a B-capsid because of its similarity to a previously described particle in herpes simplex virus (HSV)-infected cells (8, 13). The most distinguishing characteristic of B-capsids is the presence of an abundant protein constituent, the assembly protein, that has apparent counterparts in other herpesviruses and is thought to be a group-common species (8, 13). Unlike the other abundant capsid proteins, the assembly protein is derived by cleavage from a precursor, is phosphorylated, is in the size range of 35 to 40 kilodaltons (kDa), and is not found in the mature virion. Although the function of this protein is not known, several lines of evidence are consistent with it having a role in capsid assembly or DNA packaging or both (16, 21a, 26, 30).

To learn more about the structure of the assembly protein and as a first step toward studying its function through genetic manipulation, it was necessary to identify its gene. In the present study, we describe the isolation of a cDNA encoding the assembly protein of CMV strain Colburn from a cDNA library constructed in λgt11. This cDNA was used to map the position of the assembly protein gene in the CMV Colburn genome and to localize homologous regions in the genomes of two human CMV (HCMV) strains. The number and size of assembly protein transcripts in infected fibroblasts and the nucleotide sequence of the cDNA were also determined.

MATERIALS AND METHODS

Cells, viruses, and plasmids. Human foreskin fibroblasts were prepared, cultured, and infected with CMV Colburn (simian strain) and HCMV strains Towne and AD169 as described previously (8, 9). The strains of *Escherichia coli* used were Y1090 (r^- Δ*lacU169* $ProA^+$ Δ*lon araD139 strA supF* [*trpC22*::Tn*10*] *hsdR* $HsdM^+$ [pMC9]) (12), XL1-Blue (*recA1 endA1 gyrA96 thi hsdR17* [rk^- mk^+] *supE44 relA1* λ^- lac [F′ *proAB* $lacI^qA$ ΔM15 Tn*10* {Tet^r}]) (4), and K-12 JM101 [*supE thi* Δ(*lac proAB*) (F′ *traD36 proAB* $lacI^qZ$ ΔM15)] (34). The CMV cDNA library was constructed in λgt11 (35). The *E. coli* plasmids pBluescript KS+ (referred to here as pBSKS+) (Stratagene, San Diego, Calif.) and pUC18 (34) were also used in this study.

Culture media and growth conditions. For the purpose of plating out λgt11, *E. coli* Y1090 was grown in TB medium containing 10 g of tryptone (Difco Laboratories, Detroit, Mich.) and 5 g of NaCl per liter; the medium was adjusted to pH 7.4 and supplemented with 10 mM $MgSO_4$, 0.2% maltose, and 100 μg of ampicillin per ml after autoclaving. *E. coli* Y1090 was otherwise grown in liquid or agar-solidified NZYM medium (24) containing 100 μg of ampicillin per ml. *E. coli* XL1-Blue and K-12 JM101 were routinely grown in L broth containing 10 g of tryptone, 5 g of yeast extract (Difco), and 10 g of NaCl per liter; the medium was adjusted to a final pH of 7.2. *E. coli* K-12 JM101 was maintained on glucose-minimal medium agar containing the following (per liter): 6 g of Na_2HPO_4, 3 g of KH_2PO_4, 1 g of NH_4Cl, 0.5 g of NaCl, 0.493 g of $MgSO_4 \cdot 7H_2O$, 14.7 mg of $CaCl_2 \cdot 2H_2O$, 1 mg of thiamine hydrochloride, 0.2 g of glucose, and 15 g of minimal agar; the medium was adjusted to pH 7.0. Bacterial cultures were grown in liquid with vigorous aeration at 37°C. Growth was measured spectrophotometrically at 550 nm.

RNA and DNA isolation. Total-cell RNA was prepared from CMV Colburn-infected fibroblasts 70 h postinfection by disrupting the cells in guanidine thiocyanate (5) and, subsequently, recovering the RNA by pelleting through a CsCl

* Corresponding author.

cushion (11). Poly(A)$^+$ RNA was selected by column chromatography on poly(dT)-cellulose (1). CMV DNA was prepared from extracellular virions as described previously (10, 19). DNA was isolated from bacteriophage λ recombinants as described by Leder et al. (21). Plasmid DNA was isolated by an alkaline lysis procedure (3) as modified by Maniatis et al. (24). Covalently closed supercoiled DNA was purified by CsCl-ethidium bromide equilibrium density gradient centrifugation, *n*-butanol extractions, and dialysis against 10 mM Tris hydrochloride–1 mM EDTA (pH 8.0) (24, 27).

Restriction endonuclease digestions, ligation reactions, and transformation protocols. Restriction endonuclease digestions, dephosphorylation reactions using calf intestinal alkaline phosphatase, and ligations using T4 DNA ligase were carried out as specified by the manufacturers. Agarose gel electrophoresis was used to analyze plasmids, phage DNA, endonuclease restriction fragments, and ligation reaction products. Agarose gels were prepared and run in 40 mM Tris-acetate–2 mM disodium EDTA–0.5 μg of ethidium bromide per ml. DNA was detected by illumination with a short-wavelength UV transilluminator. *Hin*dIII-digested λ DNA and *Hae*III-digested φX174 replicative-form DNA were used as molecular size markers. Low-melting-point agarose (FMC Corp., Marine Colloids Div., Rockland, Maine), Elutip-d columns (Schleicher & Schuell, Inc., Keene, N.H.), and Spin-X centrifuge filter units (Costar, Cambridge, Mass.), were used to recover and purify DNA fragments from preparative gels as specified by the manufacturers. Transformation of *E. coli* strains was carried out by a $CaCl_2$ shock procedure (23).

Construction of the cDNA library. cDNA was synthesized from poly(A)$^+$ RNA obtained from fibroblasts infected with CMV Colburn using a commercial cDNA synthesis system (Amersham Corp., Arlington Heights, Ill.) according to manufacturer recommendations. Following methylation of internal *Eco*RI restriction endonuclease sites in the cDNAs by *Eco*RI methylase (New England BioLabs, Inc., Beverly, Mass.), *Eco*RI linkers were attached to the cDNAs. The cDNA molecules were ligated into the *Eco*RI site of predigested λgt11 (Protoclone GT; Promega Biotec, Madison, Wis.). Packaging of recombinant phage was carried out using an in vitro packaging extract (Stratagene). *E. coli* Y1090 (r^-) was used to plate out and amplify recombinant phage (12).

Screening protocol. Recombinant phage containing cDNA encoding the assembly protein were identified by a modification of the immunoassay protocol of Huynh et al. (12). Monospecific antisera, previously raised against the purified 37-kDa assembly protein (W. Gibson, A. Irmiere, and J. Lee, submitted for publication), biotinylated goat anti-rabbit immunoglobulin G, streptavidin-conjugated peroxidase, and 4-chloro-1-naphthol were used in an in situ enzymatic immunoassay according to manufacturer specifications (Bethesda Research Laboratories, Inc., Gaithersburg, Md.; BRL Products for Immunodetection, Applications Guide, May 1984).

Hybridizations. Radiolabeled, denatured double-stranded assembly protein cDNA probes were synthesized by oligolabeling with [α-^{32}P]dCTP and the Klenow fragment of DNA polymerase I according to manufacturer specifications (Pharmacia, Inc., Piscataway, N.J.). Conditions for Southern transfer (31) were as described by Maniatis et al. (24). Southern hybridizations and washes were performed in aqueous conditions at 68°C when both probe DNA and target DNA were complementary. In cross-hybridization experiments, annealing reactions were performed at 37°C in 40% formamide and washes were carried out at 50°C. Denaturing agarose gels containing 0.66 M formaldehyde were used for Northern (RNA) transfers (22, 33) as described by Davis et al. (6). Northern blots were routinely hybridized at 50°C in 50% formamide solutions and washed in aqueous solutions at 50°C. Hybrid selections (28) and hybrid-arrested translations (25) of assembly protein mRNA were carried out using agarose gel-purified assembly protein cDNA and CMV Colburn-infected fibroblast total cell RNA.

cDNA transcriptions. RNA transcripts were prepared from recombinant pBSKS+ using T7 RNA polymerase as described in the Bluescript Exo/Mung DNA Sequence System Instruction Manual (Stratagene). Both sense and antisense RNA transcripts were synthesized.

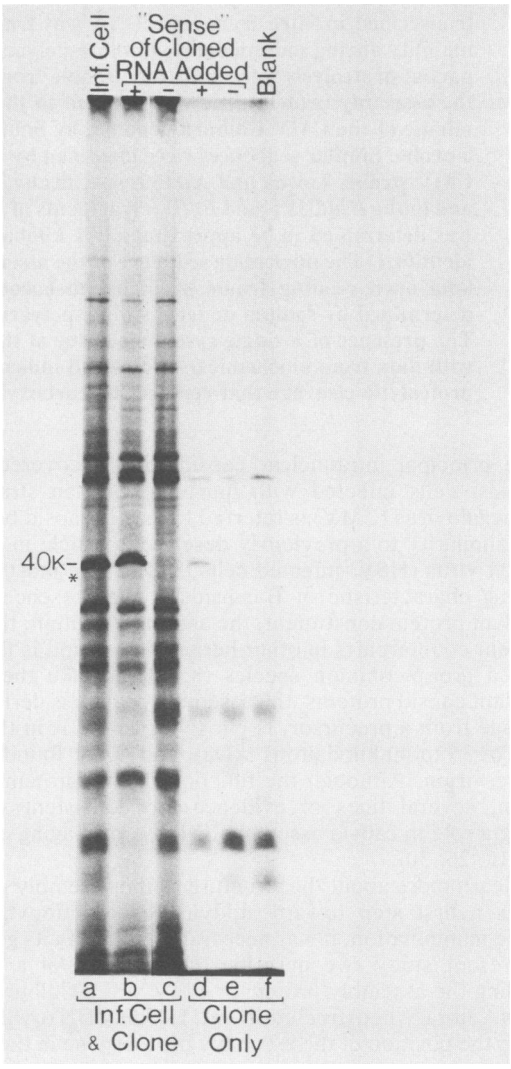

FIG. 1. APcDNA1 antisense transcript inhibits in vitro translation of the 40-kDa assembly protein precursor, and sense transcript is translated to give a 40-kDa protein. Sense and antisense transcripts were prepared in vitro from cloned APcDNA1. Each was added to RNA from CMV Colburn-infected cells, heated, allowed to hybridize, translated in a rabbit reticulocyte lysate, and subjected to SDS-PAGE. Lanes: a, Colburn-infected-cell RNA; b, infected-cell RNA plus sense transcript of APcDNA1; c, infected-cell RNA plus antisense transcript; d, sense transcript alone; e, antisense transcript alone; f, blank translation with no RNA added. The asterisk indicates a protein whose position is shown by an arrow in Fig. 2. Inf. Infected. K, Kilodaltons.

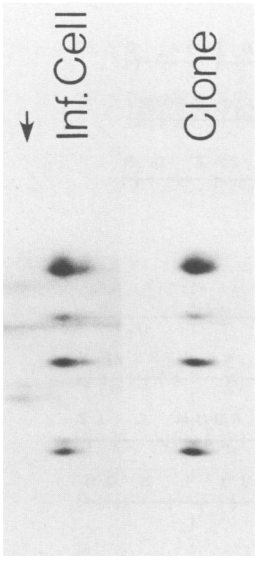

FIG. 2. Product made by APcDNA1 sense transcript is the 40-kDa assembly protein precursor. The 40-kDa assembly protein precursor translated from infected-cell RNA (Inf. Cell) and the 40-kDa band translated from APcDNA1 sense RNA (Clone) were labeled with [^{35}S]methionine, separated by SDS-PAGE, and individually subjected to partial proteolysis using *S. aureus* V8 protease. The resulting peptide patterns were indistinguishable, demonstrating that APcDNA1 sense RNA encodes the assembly protein. The arrow indicates the position of the protein translated from infected-cell RNA but not from the APcDNA1 transcript (see asterisk in Fig. 1).

In vitro translations, SDS-PAGE, and peptide comparisons. In vitro translations were carried out using a rabbit reticulocyte lysate preparation (Amersham) and [^{35}S]methionine according to manufacturer recommendations. In vitro translation products were analyzed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) as described by Laemmli (17) in gels with an acrylamide monomer concentration of 10%. Samples were prepared for electrophoresis by adding an equal volume of a solution containing 4% SDS, 20% (vol/vol) β-mercaptoethanol, 10% (vol/vol) glycerol, 0.02% bromphenol blue, and 0.1 M Tris hydrochloride (pH 7.0) and by heating at 100°C for 3 min. [^{35}S]methionine-labeled proteins were detected in gels by fluorography (20) with prefogged Kodak X-Omat AR film (Eastman Kodak Co., Rochester, N.Y.). Peptide comparisons were done as described previously (2); *Staphylococcus aureus* V8 protease was used as the cleavage enzyme.

DNA sequence determination. cDNA insert APDNA1, from which the assembly protein precursor could be produced in vitro, was excised with *Eco*RI from a λgt11 recombinant DNA, purified by preparative agarose gel electrophoresis, and subcloned into pBSKS+. Restriction endonuclease fragments of this cDNA were also purified and subcloned into pBSKS+. Competent *E. coli* XL1-Blue cells were transformed with these recombinant plasmids. The transformants were infected with the M13 helper phage R408, which preferentially packages single-stranded recombinant pBSKS+ DNA over its own genome. Single-stranded recombinant plasmid DNA was prepared from the culture supernatants of the R408-infected cells by a polyethylene glycol-ammonium acetate precipitation procedure as outlined in the Bluescript Exo/Mung DNA Sequence System Instruction Manual. The dideoxynucleotide chain termination method of Sanger et al. (29) was used for DNA sequence determinations; however, the protocol was modified for use with Sequenase, and altered T7 DNA polymerase (United States Biochemical Corp., Cleveland, Ohio). DNA was labeled with [α-^{35}S]dATP (Amersham) during the sequencing reaction. Both strands were sequenced.

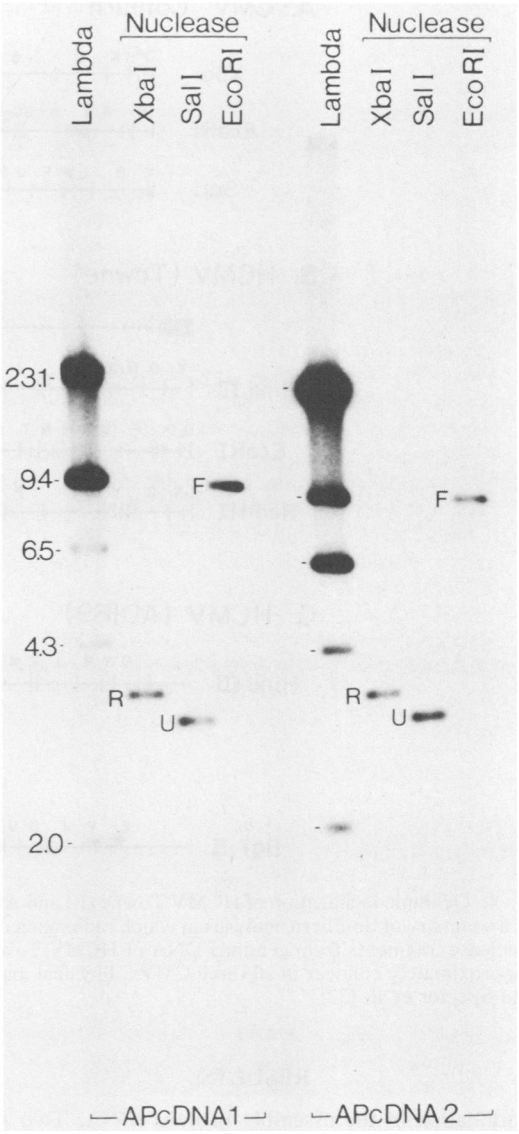

FIG. 3. Southern analysis localizes CMV Colburn assembly protein gene. Two independent cDNA clones of the Colburn assembly protein (APcDNA1, left panel; APcDNA2, right panel) were radiolabeled and used as probes against genomic CMV Colburn DNA that had been cleaved with one of the three restriction endonucleases indicated at the top. Both probes recognized the same restriction fragments: *Xba*I R, *Sal*I U, and *Eco*RI F. The sizes (in kilobases) of λ marker fragments (Lambda) are indicated to the left.

Chemicals and reagents. Restriction endonucleases were purchased from New England BioLabs, Bethesda Research Laboratories, and Boehringer-Mannheim Biochemicals (Indianapolis, Ind.). T4 DNA ligase was purchased from Promega-Biotec. Calf intestinal alkaline phosphatase was obtained from Boehringer-Mannheim. All other chemicals were of standard or ultrapure reagent grade.

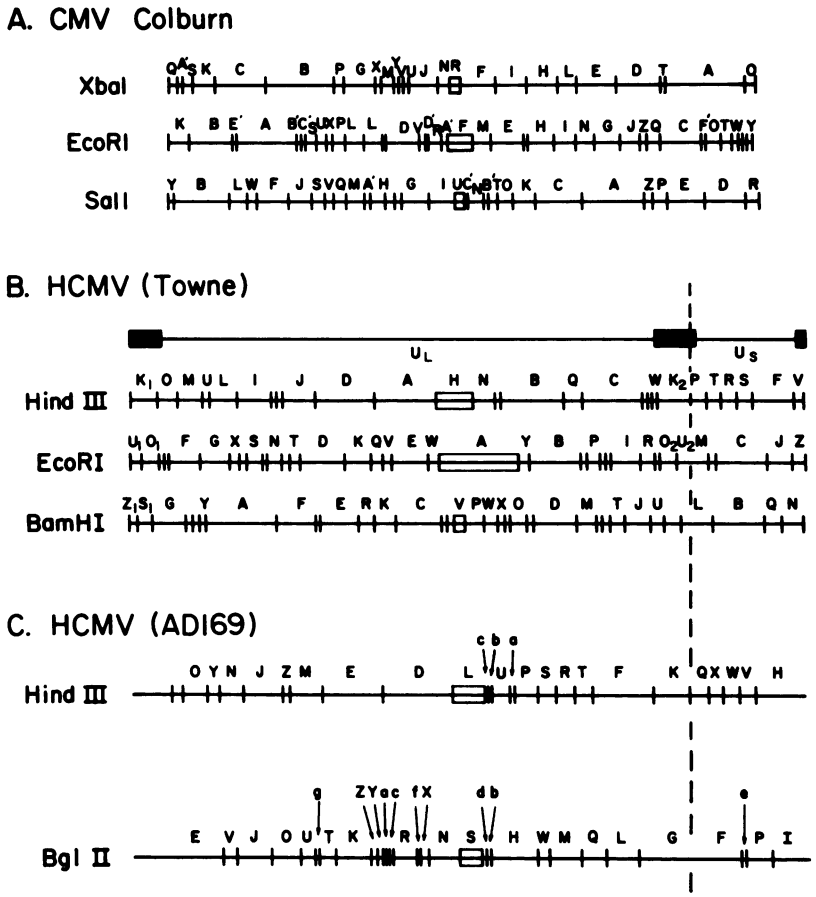

FIG. 4. Genomic localization of HCMV Towne (B) and AD169 (C) fragments that cross-hybridize to CMV Colburn (A) APcDNA1. Shown here is a summary of Southern analyses in which radiolabeled Colburn APcDNA1 was used as a probe to identify cross-hybridizing restriction endonuclease fragments from genomic DNA of HCMV Towne and AD169. The cross-hybridizing fragments (indicated by open rectangles) were approximately colinear in all three CMVs. Physical maps of the Colburn, Towne, and AD169 DNAs are from LaFemina and Hayward (18) and Spector et al. (32).

RESULTS

Identification of the assembly protein cDNA. Two recombinant λgt11 phage, each of which produced a fusion protein that reacted positively with anti-assembly protein serum, were selected for further study and plaque purified. The cDNA inserts, designated here as APcDNA1 and APcDNA2 (AP = assembly protein), were excised from λgt11 with *Eco*RI and subcloned into the *E. coli* plasmids pUC18 and pBSKS+ to facilitate DNA preparation. Electrophoresis of the cDNA inserts in agarose gels revealed that APcDNA1 was slightly smaller than APcDNA2, approximately 1,050 base pairs (bp) versus 1,150 bp, respectively.

To verify that APcDNA1 encodes the strain Colburn assembly protein, sense and antisense transcripts were prepared from the APcDNA1 insert by using the T3 and T7 promoters in the pBluescript KS+ vector. Each transcript was allowed to anneal with RNA prepared from CMV Colburn-infected cells. In vitro translation of the resulting RNA mixtures (Fig. 1) showed that the 40-kilodalton (kDa) assembly protein precursor was synthesized in the preparations containing infected-cell RNA either alone (lane a) or with the APcDNA1 sense transcript (lane b) but not in the preparation containing infected-cell RNA and the APcDNA1 antisense transcript (lane c). This inhibitory effect of the antisense transcript was both complete and highly selective for assembly protein synthesis, consistent with APcDNA1 being from the coding region of the assembly protein gene.

Translation of the APcDNA1 sense transcript alone in vitro (Fig. 1) resulted in the synthesis of a protein that comigrated with the 40-kDa assembly protein precursor (lane d); no protein was translated from antisense APcDNA1 alone (lane e). Comparison by partial proteolysis of the APcDNA1 protein product with authentic assembly protein precursor translated from infected-cell RNA (e.g., Fig. 1, lanes d and a, respectively) showed that the two were indistinguishable (Fig. 2). Taken together, these results indicate that APcDNA1 encodes the complete assembly protein precursor.

Consistent results were obtained in initial experiments that used denatured APcDNA1 to selectively (i) interfere with translation of assembly protein mRNA and (ii) hybridize translatable assembly protein mRNA from a preparation of total RNA from CMV Colburn-infected fibroblasts (data not shown).

Southern analysis. Purified CMV Colburn DNA was digested to completion with the restriction endonucleases *Xba*I, *Sal*I, and *Eco*RI and subsequently hybridized in a Southern blot to radiolabeled APcDNA1, APcDNA2, or pUC18. Both APcDNA1 and APcDNA2 hybridized to the

same fragments of CMV Colburn DNA (Fig. 3): *Xba*I R (3.6 kilobases [kb]), *Sal*I U (3.3 kb), and *Eco*RI F (9.3 kb). Each of these restriction endonuclease fragments fell within the same area of the CMV Colburn genome (Fig. 4). Radiolabeled pUC18 served as a negative control and did not hybridize to any of the CMV Colburn DNA fragments.

Cross-hybridization experiments done by probing restriction endonuclease digestion fragments of HCMV Towne and AD169 with radiolabeled APcDNA1 identified an approximately colinear region of these HCMV genomes that had sequence similarity to CMV Colburn APcDNA1. The specific HCMV fragments that hybridized to APcDNA1 were *Hin*dIII H (12.5 kb), *Bam*HI V (3.9 kb), and *Eco*RI A (27.7 kb) of HCMV Towne and *Hin*dIII L (12.0 kb) and *Bgl*II S (8.0 kb) of HCMV AD169. Results consistent with the above were obtained when the cloned HCMV Towne *Hin*dIII H fragment, excised from plasmid pRL108A (kindly provided by R. LaFemina and G. Hayward), and an *Eco*RI-*Xba*I digest of genomic HCMV AD169 were hybridized to radiolabeled APcDNA1 (autoradiogram not shown). Assignment of restriction fragments of CMV Colburn and HCMV Towne and AD169 was based on the data of LaFemina and Hayward (18) and Spector et al. (32). Under similar hybridization conditions APcDNA1 did not cross-hybridize to any of the *Bam*HI or Asp-718 restriction fragments from HSV type 1 (HSV-1) strain 17 DNA (kindly provided by S. Silverstein) (data not shown).

Transcript analysis. Unfractionated CMV Colburn-infected fibroblast RNA was subjected to electrophoresis in a denaturing formaldehyde-agarose gel, blotted to nitrocellulose (Northern transfer), probed with radiolabeled, denatured double-stranded APcDNA1, washed, and visualized by fluorography (Fig. 5). Two RNA transcripts were identified, and their sizes were calculated on the basis of their electrophoretic migration relative to those of RNA marker molecules. The smaller transcript (ca. 1 kb) was predominant. The intensity of the larger transcript (ca. 1.8 kb) was less than 15% that of the 1-kb band. Although the position of the 1.8-kb band was close to that of 18S rRNA, it is unlikely that it was the result of spurious hybridization of APcDNA1 to rRNA because (i) its relative intensity was essentially the same at different hybridization stringencies; (ii) APcDNA1 did not hybridize to species in the mouse RNA preparation, even though it had comparable amounts of 18S rRNA; and (iii) the relative amounts of the 1- and 1.8-kb bands were the same in nonselected (i.e., whole-cell RNA) and poly(dT)-cellulose-selected [i.e., poly(A)$^+$ mRNA] RNA preparations (data not shown).

DNA sequence analysis. DNA sequence analysis established that APcDNA1, which encodes the CMV Colburn assembly protein, consists of 1,054 bp, including a 16-bp poly(dA · dT) sequence at the 3′ end. The complete nucleotide sequence and predicted amino acid sequence are shown in Fig. 6. A computer analysis of this sequence revealed only one open reading frame long enough to encode the approximately 40-kDa assembly protein precursor detected by SDS-PAGE. This sequence begins at nucleotide 3 and ends at nucleotide 1025. There are two ATG codons, beginning at positions 96 and 108 near the 5′ end of the open reading frame, which could serve as start sites of translation. Assuming that translation initiates at either the first or the second of these two codons, a polypeptide of ca. 37 or 36 kDa, respectively, would be synthesized.

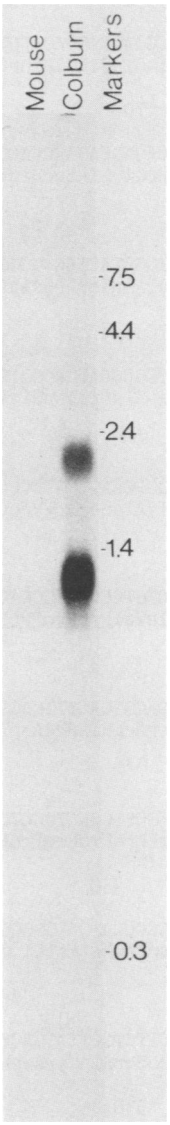

FIG. 5. Two RNA transcripts identified by APcDNA1. Infected-cell RNA was subjected to electrophoresis in a formaldehyde-agarose gel, transferred to nitrocellulose, and probed with radiolabeled APcDNA1 (Colburn). Two transcripts with estimated sizes of 1.8 and 1.0 kb were identified. No cross-hybridizing transcripts were detected in a similarly prepared RNA sample from mouse 3T3 cells (Mouse). The positions of marker RNAs of known size (in kilobases) are indicated to the right (Markers).

DISCUSSION

In this study a λgt11 cDNA library was prepared from CMV Colburn-infected-cell RNA and screened for recombinants expressing antigens recognized by a monospecific rabbit antiserum to the B-capsid assembly protein. Initial characterizations of the resulting antigen-positive clones indicated that APcDNA1 contained an insert just large enough to encode the 40-kDa assembly protein precursor, and it was therefore selected for further study.

Direct evidence that this insert does encode the assembly protein was provided by showing that its sense transcript could be translated in vitro to give a product having the same electrophoretic mobility during SDS-PAGE and the same

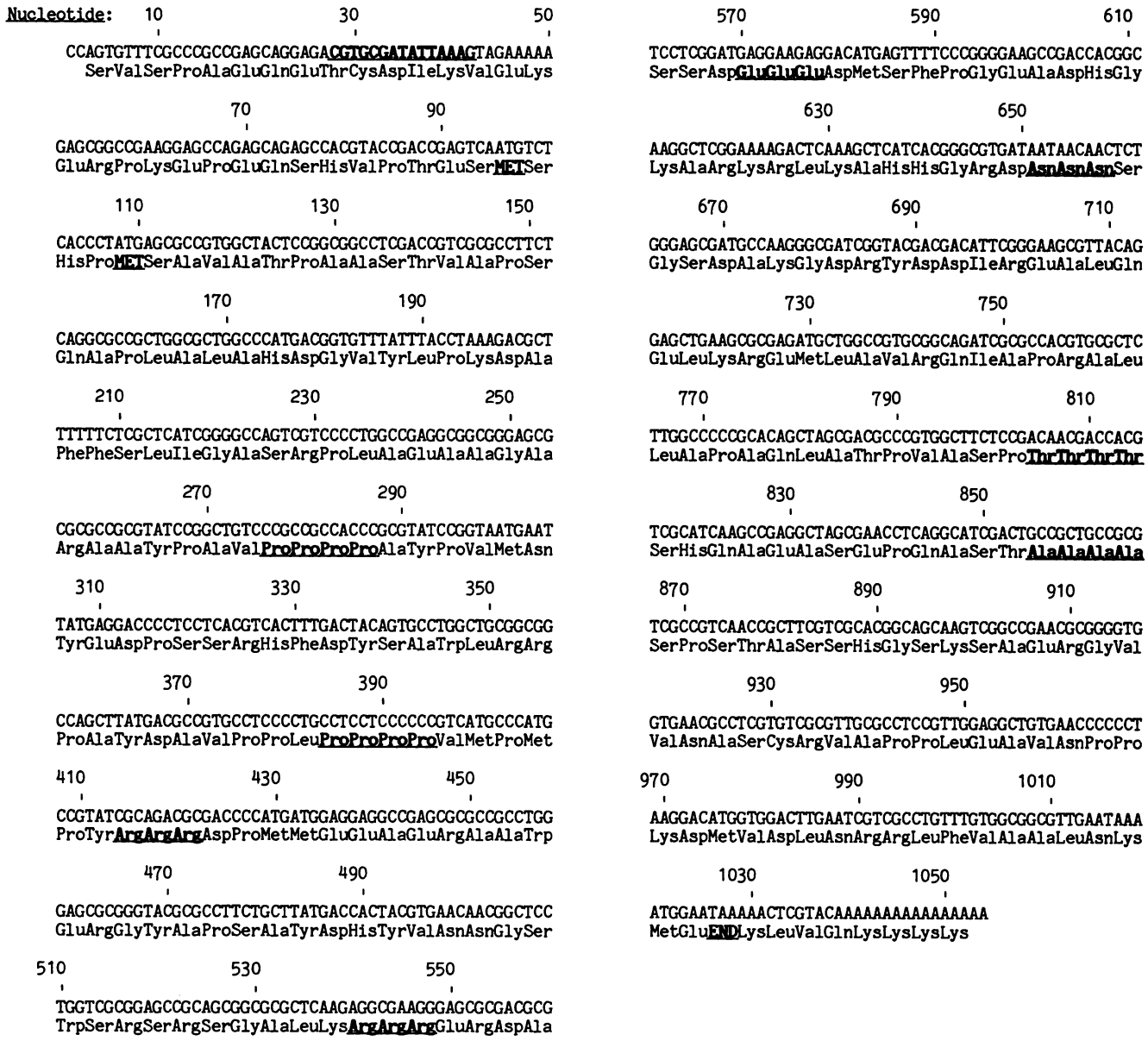

FIG. 6. Nucleotide sequence of APcDNA1 and deduced amino acid sequence of the longest open reading frame. Boldface methionines indicate the presumed start sites of translation. Other highlighting indicates the locations of a TATA box (TATTAAA)-containing sequence and several three- and four-amino-acid runs mentioned in the text.

fragmentation pattern following partial proteolysis as the authentic assembly protein precursor. In addition, the APcDNA1 antisense transcript hybridized to assembly protein mRNA from infected cells and blocked its translation in vitro. Given that APcDNA1 is about the same size as the predominant RNA species detected by Northern assays (i.e., about 1 kb; Fig. 5) and that this size is close to that required to encode the 40-kDa assembly protein precursor, it is likely that APcDNA1 represents an essentially full-length copy of the 1-kb RNA and that the 40-kDa assembly protein precursor is the primary translation product of that predominant transcript.

The assembly protein gene (i.e., the APcDNA1 sequence) was determined to be located near the middle of the CMV Colburn genome (i.e., *Xba*I R fragments), and cross-hybridization experiments identified a similar sequence at an approximately colinear position in the genomes of HCMV AD169 and Towne (Fig. 3). As discussed below, this region of the HCMV genome contains a sequence with similarities to Colburn APcDNA1 and is presumed to encode the HCMV assembly protein (9, 13). Although the gene for the apparent counterpart HSV-1 assembly protein (i.e., p40) did not cross-hybridize under the conditions used, nucleotide sequence comparison of APcDNA1 and the p40 gene (D. McGeoch, personal communication) revealed similarities, as discussed below.

Nucleotide sequence analysis of APcDNA1 showed that it is 1,054 bp in length and, as expected for a cDNA, has a poly(dA · dT) sequence at its 3′ end. A TATA box (i.e., TATTAAA) is present near the 5′ end of the sequence. Two ATG codons, either of which may be a start site of translation, are also present near the 5′ terminus. Neither of these

codons, however, falls within a sequence resembling the apparent consensus sequence for the initiation of translation in vertebrates [i.e., (GCC)GCC$^{A}/_{G}$CCATGGG] (14). The long open reading frame which follows these codons has a coding capacity for a 36.6- to 37-kDa protein, depending on which of the two is utilized. The slight discrepency between this calculated size and the 40-kDa size determined by SDS-PAGE for the assembly protein precursor translated in vitro from either APcDNA1 or infected-cell mRNA (Fig. 1) may be due to anomolous migration of the protein (e.g., due to proline runs) or to posttranslational modification (e.g., phosphorylation), which may increase the apparent size of the protein.

An analysis of the predicted amino acid sequence of the CMV Colburn assembly protein revealed several unusual features, including three- and four-residue runs of arginine, glutamic acid, alanine, asparagine, threonine, and proline as well as an extensive region around the center of the protein which scores as very hydrophilic when a hydrophobicity algorithm is used (15). Additionally, the proline and arginine runs are located in regions that are rich, overall, in those residues. These features are reflected in the amino acid composition of the protein. For example, the assembly protein contains more than twice the average amount of alanine, proline, arginine, and methionine residues (7). In contrast, it is much lower than average in its lysine and isoleucine content. It is interesting to note that the low lysine level in the protein may be compensated for by the high level of arginine, thus positioning the assembly protein near the average with regard to the amount of basic amino acid residues it contains. In keeping with its generally hydrophilic character, the assembly protein contains a lower than average amount of hydrophobic amino acid residues.

The predicted amino acid sequence of the assembly protein supports the results of earlier biochemical studies which indicated that the mature assembly protein is processed from a slightly larger precursor (Gibson et al., submitted). These studies showed that the precursor contains cysteine but that the mature form of the assembly protein, which is about 3 kDa smaller, does not. The presence of a sole cysteine residue 32 amino acid residues from the carboxy terminus of the protein indicates that this processing event occurs at the carboxy-terminal end of the assembly protein primary translation product.

Lindenmaier and colleagues (A. Necker, S. Krause, B. Khattab, and W. Lindenmaier, 13th International Herpesvirus Workshop, abstr. no. 190, 1988; personal communication) have sequenced a region of the AD169 *Hin*dIII L fragment, the HCMV fragment that cross-hybridized with Colburn APcDNA1 (Fig. 4). Comparison of this sequence with that of Colburn APcDNA1 showed clear similarities, the most striking of which were as follows: (i) there was a perfectly conserved nucleotide sequence that includes a TATA box (i.e., CGTGCGATATTAAAG) 5′ to the putative initiation codon; (ii) 9 of the first 13 amino-terminal residues were the same; and (iii) 19 of the last 21 carboxy-terminal residues were the same.

Nucleotide sequence comparisons revealed that the gene for HSV-1 p40, the counterpart assembly protein of HSV (8, 9, 13), shares many of these unusual features, including runs of arginine, proline, alanine, and threonine residues and the seemingly compensatory ratio of arginine to lysine. However, in spite of these shared general characteristics, p40 and the CMV Colburn assembly protein exhibit only limited amino acid sequence similarity, the best matched sequences being predominantly in the carboxy-terminal portions of the proteins. Furthermore, neither the TATA box nucleotide sequence nor the amino- and carboxy-terminal amino acid sequence similarities noted above for the HCMV and Colburn genes are shared with HSV-1 p40. The lack of extensive similarity between these CMV and HSV proteins is consistent with the lack of hybridization observed in Southern assays when APcDNA1 was used to probe the HSV-1 genome (data not shown).

It is of considerable interest that p40, the 40-kDa HSV-1 counterpart of the 37-kDa mature CMV Colburn assembly protein, is produced by a gene with the coding potential for a 76-kDa protein (D. McGeoch, personal communication). This theoretical 76-kDa product may correspond to the 80-kDa protein p80, identified by Zweig et al. (36), which is immunologically related to p40. The peptide maps of p40 and p80 were shown to exhibit similarities as well as differences, and it was concluded that p80 is not a precursor of p40, although the two do share antigenic determinants. The open reading frame which encodes p40 does contain an internal methionine (amino acid 307) which, if used as a translational initiation site, would give rise to a 39.4-kDa polypeptide. If the HSV and CMV assembly protein genes are analogously transcribed and translated, it is possible that the 1.8-kb RNA detected in CMV Colburn-infected cells (Fig. 5) represents the coding transcript for a CMV counterpart to HSV-1 p80.

ACKNOWLEDGMENTS

We thank S. Silverstein for kindly providing us with HSV DNA for cross-hybridization experiments; D. McGeoch for sharing the nucleotide sequence of the HSV-1 p40 gene prior to its publication; and W. Lindenmaier for sharing the nucleotide sequence of the HCMV AD169 p38 gene prior to its publication. We also thank M. Amzel, D. Anders, and P. Collins for helpful discussions.

This work was aided by Public Health Service research grants AI13718 and AI22711 from the National Institute of Allergy and Infectious Diseases. L.R. was a postdoctoral fellow in the Anticancer Drug Development Training Program and was supported by Public Health Service training grant PHS T32 CA09243 awarded by the National Cancer Institute.

LITERATURE CITED

1. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**:1408–1412.
2. **Benko, D. M., R. S. Haltiwanger, G. W. Hart, and W. Gibson.** 1988. Virion basic phosphoprotein from human cytomegalovirus contains O-linked N-acetylglucosamine. Proc. Natl. Acad. Sci. USA **85**:2573–2577.
3. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
4. **Bullock, W. O., J. M. Fernandez, and J. M. Short.** 1987. XL1-Blue: a high efficiency plasmid transforming recA *Escherichia coli* strain with beta-galactosidase section. Biotechniques **5**:376–379.
5. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
6. **Davis, L. G., M. D. Dibner, and J. F. Battery.** 1986. Basic methods in molecular biology. Elsevier Science Publishing, Inc., New York.
7. **Dayhoff, M. O.** 1972. Atlas of protein sequence and structure, vol. 5. The National Biomedical Research Foundation, Silver Spring, Md.
8. **Gibson, W.** 1981. Structural and nonstructural proteins of strain Colburn cytomegalovirus. Virology **111**:516–537.
9. **Gibson, W.** 1983. Protein counterparts of human and simian

cytomegaloviruses. Virology **128**:391–406.
10. **Gibson, W., R. van Breemen, A. Fields, R. LaFemina, and A. Irmiere.** 1984. DL-α-Difluromethylornithine inhibits human cytomegalovirus replication. J. Virol. **50**:145–154.
11. **Glisin, V., R. Crkvenjakov, and C. Byus.** 1974. Ribonucleic acid isolated by cesium chloride centrifugation. Biochemistry **13**: 2633–2637.
12. **Huynh, T. V., R. A. Young, and R. W. Davis.** 1985. Constructing and screening cDNA libraries in λ gt11, p. 56–110. *In* D. M. Glover (ed.), DNA cloning, vol. 1. IRL Press Ltd., Oxford.
13. **Irmiere, A., and W. Gibson.** 1985. Isolation of human cytomegalovirus intranuclear capsids, characterization of their protein constituents, and demonstration that the B-capsid assembly protein is also abundant in noninfectious enveloped particles. J. Virol. **56**:277–283.
14. **Kozak, M.** 1987. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. **15**:8125–8148.
15. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
16. **Ladin, B. F., S. Ihara, H. Hampl, and T. Ben-Porat.** 1982. Pathway of assembly of herpesvirus capsids: an analysis using DNA^+ temperature-sensitive mutants of pseudorabies virus. Virology **116**:544–561.
17. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
18. **LaFemina, R. L., and G. S. Hayward.** 1980. Structural organization of the DNA molecules from human cytomegalovirus, p. 39–55. *In* B. N. Fields, and R. Jaenisch (ed.), Animal virus genetics. Academic Press, Inc., New York.
19. **LaFemina, R. L., and G. S. Hayward.** 1983. Replicative forms of human cytomegalovirus DNA with joined termini are found in permissively infected human cells but not in non-permissive BALB/c-3T3 mouse cells. J. Gen. Virol. **64**:373–389.
20. **Laskey, R. A., and A. D. Mills.** 1975. Quantitative film detection of ^3H and ^{14}C in polyacrylamide gels by fluorography. Eur. J. Biochem. **56**:335–341.
21. **Leder, P., D. Tiemeier, and L. Enquist.** 1977. EK2 derivatives of bacteriophage lambda useful in the cloning of DNA from higher organisms: the λgtWES system. Science **196**:175–177.

21a. **Lee, J. Y., A. Irmiere, and W. Gibson.** 1988. Primate cytomegalovirus assembly: Evidence that DNA packaging occurs subsequent to B-capsid assembly. Virology **167**:87–96.

22. **Lehrach, H., D. Diamond, J. Wozney, and H. Doedtker.** 1977. RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry **16**:4743–4751.
23. **Mandel, M., and A. Higa.** 1970. Calcium dependent bacteriophage DNA infection. J. Mol. Biol. **53**:159–162.
24. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
25. **Paterson, B. M., B. E. Roberts, and E. L. Kuff.** 1977. Structural gene identification and mapping by DNA · mRNA hybrid-arrested cell-free translation. Biochemistry **74**:4370–4374.
26. **Preston, V. G., J. A. V. Coates, and F. J. Rixon.** 1983. Identification and characterization of a herpes simplex virus gene product required for encapsidation of virus DNA. J. Virol. **456**:1056–1064.
27. **Radloff, R., W. Bauer, and J. Vinograd.** 1967. A dye-buoyant-density method for the detection and isolation of closed circular duplex DNA in HeLa cells. Proc. Natl. Acad. Sci. USA **57**:1514–1521.
28. **Ricciardi, R. P., J. S. Miller, and B. E. Roberts.** 1979. Purification and mapping of specific mRNAs by hybridization-selection and cell-free translation. Proc. Natl. Acad. Sci. USA **76**:4927–4931.
29. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
30. **Sherman, G., and S. L. Bachenheimer.** 1988. Characterization of intranuclear capsids made by ts morphogenic mutants of HSV-1. Virology **163**:471–480.
31. **Southern, E.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**: 503–517.
32. **Spector, D. H., L. Hock, and J. C. Tamashiro.** 1982. Cleavage maps for human cytomegalovirus DNA strain AD169 for restriction endonucleases *Eco*RI, *Bgl*II, and *Hin*dIII. J. Virol. **42**: 558–582.
33. **Thomas, P.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77**:5201–5205.
34. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **23**:103–119.
35. **Young, R. A., and R. W. Davis.** 1983. Efficient isolation of genes by using antibody probes. Proc. Natl. Acad. Sci. USA **80**: 1194–1198.
36. **Zweig, M., C. J. Heilman, H. Rabin, and B. Hamper.** 1980. Shared antigenic determinants between two distinct classes of proteins in cells infected with herpes simplex virus. J. Virol. **35**:644–652.