JOURNAL OF VIROLOGY, Aug. 1991, p. 4091–4100
0022-538X/91/084091-10$02.00/0

Vol. 65, No. 8

# Cytomegalovirus Assembly Protein Nested Gene Family: Four 3′-Coterminal Transcripts Encode Four In-Frame, Overlapping Proteins†

ANTHONY R. WELCH, LISA M. McNALLY, AND WADE GIBSON*

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

Received 11 March 1991/Accepted 26 April 1991

**The genomic region encoding the assembly protein of simian cytomegalovirus (CMV) strain Colburn has been cloned, sequenced, and found to be organized as a nested set of four in-frame, 3′-coterminal genes, each with its own TATA promoter element and translational start codon, and all using a single 3′ polyadenylation signal. The 3′ end of the longest open reading frame (1,770 bp) was identical to the 930-bp sequence coding for the assembly protein precursor, as determined from a cDNA clone. The assembly protein coding region of human CMV strain AD169 was similarly organized, suggesting that both viral genomes could give rise to four independently transcribed 3′-coterminal RNAs coding for four overlapping, in-frame, carboxy-coterminal proteins. These predictions were tested and confirmed. Four mRNAs corresponding in size and sequence to those predicted were identified in both human and simian CMV-infected cells by using transcript-specific antisense oligonucleotide probes in Northern (RNA blot) assays. The 5′ ends of the three largest of these Colburn transcripts were determined by S1 nuclease protection assays and found to map between the anticipated TATA sequences and corresponding translational start codons. The four predicted overlapping proteins were identified by immunoassays in lysates of simian and human CMV-infected cells by using an antiserum specific for the carboxyl end of the assembly protein precursor. The structural relationship of both sets of proteins was verified by comparing their peptide patterns following protein cleavage at tryptophan residues by *N*-chlorosuccinimide. The similar organization of the homologous coding regions in other herpesviruses into at least two nested, in-frame, 3′-coterminal genes is discussed.**

Assembly and maturation of the herpesvirus capsid appears to involve the phosphorylation (6, 14, 21, 23, 31), proteolytic cleavage (6, 19, 28, 39), and ultimate elimination from the particle of an abundant "procapsid" constituent referred to as the assembly protein (17, 20, 23, 31). Although the chronology and specific consequences of these modifications are not yet established, studies of the cytomegalovirus (CMV) assembly protein have demonstrated that its 40-kDa precursor is synthesized at late times after infection, moves slowly into the nuclear fraction, and undergoes cleavage which eliminates its carboxyl end (19). On the basis of evidence obtained by using temperature-sensitive mutants of herpes simplex virus (HSV) (39, 49) and pseudorabies virus (28), it seems likely that there is a relationship between assembly protein processing and DNA packaging or nucleocapsid maturation and that the assembly protein is both central and essential to virus replication.

In order to pursue structural and functional studies of the CMV assembly protein at the molecular level, we cloned and sequenced a cDNA encoding it (41). The 5′ end of this clone contained a potential TATA promoter element and remained in frame with the assembly protein-coding sequence upstream of its translational start codon. This cDNA was also found to hybridize in Northern (RNA blot) assays to an RNA twice as large as that needed to encode the assembly protein, suggesting that it may have been copied from a larger transcript which contained the promoter region of the assembly protein gene. It has also been shown that the B-capsid 45-kDa protein, which is larger but immunologically and structurally related to the assembly protein and its 40-kDa precursor (15, 31), is an amino-terminal extension of the assembly protein (47), consistent with there being a larger assembly protein-related genomic coding region.

These observations led us to clone and sequence the genomic region encoding the assembly protein of CMV strain Colburn. The nucleotide sequence revealed that this region contains four 3′-coterminal putative genes having the potential to encode four 3′-coterminal, overlapping transcripts and four corresponding in-frame, overlapping proteins. Evidence is presented here that all four predicted mRNAs and proteins are expressed in cells infected with simian (Colburn) or human (AD169) strains of CMV.

(Presentations of this work were made at the 9th meeting of the American Society for Virology, 8 to 12 July 1990, the Federation of American Societies for Experimental Biology Meeting on Viral Assembly, 15 to 20 July 1990, and the 15th International Herpesvirus Workshop, 2 to 8 August 1990.)

## MATERIALS AND METHODS

**Cells and viruses.** CMV strains Colburn (simian) and AD169 (human) were propagated in human foreskin fibroblasts as described before (15).

**DNA cloning and sequencing.** The *Xba*I R fragment of strain Colburn CMV DNA was cloned into the plasmid pUC18, and the nucleotide sequence of both strands was determined by the dideoxy nucleotide chain termination method (45) with appropriate DNA oligonucleotide primers and the Sequenase kit (USB, Cleveland, Ohio).

**RNA isolation and Northern analysis.** Polyadenylated RNA was recovered from approximately $3 \times 10^8$ cells (six 32-oz [ca. 900-ml] bottles), infected with CMV strain Colburn or

* Corresponding author.

† Dedicated to the memory of Bob Honess.

AD169 at a multiplicity of infection of 5 to 20 or mock infected, by using the Fast Track mRNA isolation kit (In Vitrogen, San Diego, Calif.). RNA was prepared from lytically infected cells at the times stated in the text and from mock-infected cells just prior to confluence. Resulting mRNAs were denatured, subjected to electrophoresis in denaturing agarose gels, and blotted to nitrocellulose (12). The blots were blocked by incubation for 1 h at 37°C in hybridization buffer containing 6× SSPE (1× = 0.15 M NaCl, 10 mM $NaH_2PO_4 \cdot H_2O$, 1 mM EDTA [pH 7.4]), 5× Denhardt's solution (1× = 0.02% polyvinylpyrrolidone 360, 0.02% bovine serum albumin, 0.02% Ficoll 400), and 1% sodium dodecyl sulfate (SDS). Single-stranded DNA oligonucleotide probes were end-labeled with [$\gamma$-$^{32}$P]ATP (Du Pont-NEN, Boston, Mass.) by using T4 polynucleotide kinase (Bethesda Research Laboratories [BRL], Gaithersburg, Md.) at 37°C for 1 h and then separated from nonincorporated nucleotides by G-25 Sephadex gel exclusion chromatography (36). The oligonucleotide probes used were antisense to the sequences 5′-End, 5′-Mid, 3′-Mid, and 3′-End indicated in Fig. 1. Radiolabeled probes were incubated with the blots for the times given in the text; washed once in 6× SSPE at 37°C, once in 6× SSPE at 45°C, and once in 3× SSPE at 45°C; sealed in Saran Wrap to prevent drying; and visualized by fluorography (30). Where indicated, labeled probes were stripped from blots by heating at 80°C for 5 min in 1% glycerol; the blots were then washed five times for 5 min each in $H_2O$ at room temperature (12) and fluorographed overnight to verify removal of the probe. Transcript sizes were determined relative to ethidium bromide-stained RNA markers (BRL).

A β-actin riboprobe was synthesized as a *Hin*dIII runoff antisense RNA by using the p18-2 plasmid (a 276-bp *Ava*I fragment of human β-actin in the pGEM3 vector [11]) in an in vitro RNA transcription reaction mix containing 1 μg of *Hin*dIII-digested p18-2 plasmid, 40 mM Tris-HCl (pH 7.5), 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 1 mM dithiothreitol, 500 μM each ATP, GTP, and UTP, 12 μM CTP, 50 μCi of [α-$^{32}$P]CTP, and 20 U of T7 polymerase (36). The reaction mix was incubated at 37°C for 1 h, and the probe was separated from unincorporated nucleotides by G-25 Sephadex gel exclusion chromatography. The blots were prehybridized for 4 h at 62°C in a solution containing 55% formamide, 2.5× Denhardt's solution, 0.1% SDS, 1 mM EDTA, 5× SSPE, 10 μg of tRNA per ml, and 250 μg of denatured salmon sperm DNA per ml and then hybridized overnight at 62°C with approximately 5 × $10^6$ cpm of labeled probe.

**Analysis of mRNA transcriptional start sites.** S1 nuclease protection assays (5) were done with total RNA recovered from noninfected or Colburn CMV-infected cells (10), essentially as described by Kaufer et al. (25). Protection assay reaction mixes contained approximately 20 μg ($A_{260}$) of RNA and 100 U of S1 nuclease (no. 8001SA; BRL).

Single-stranded, 5′-end-labeled antisense DNA probes overlapping the regions between each of the TATA sequences and their corresponding ATGs were prepared as follows. Single-stranded DNA oligonucleotide primers (100 to 200 ng, see below) were end-labeled with [$\gamma$-$^{32}$P]ATP as outlined above and combined with 15 μg of plasmid XbaR/pUC18 that had been denatured by boiling in 0.2 M NaOH for 2 min, ethanol precipitated, washed, and dried. The denatured plasmid and labeled primer were then annealed at 37°C for 15 min in 8 μl of $H_2O$ and 2 μl of 5× Sequenase buffer (1× = 40 mM Tris-HCl [pH 7.5], 20 mM $MgCl_2$, 50 mM NaCl). Primer extensions from the end-labeled oligonucleotides were done in 20-μl reaction volumes containing 5 mM dithiothreitol, 500 μM each of the deoxynucleoside triphosphates, and 3.25 U of the modified T7 polymerase Sequenase. The reaction mixtures were incubated at room temperature for 5 min, at 37°C for 5 min, and finally at 65°C for 10 min to inactivate the polymerase. The probes were given defined 3′ ends by treating primer extension reaction mixtures with appropriate restriction endonucleases (see below) in a 100-μl volume at 37°C for 1 h. Following ethanol precipitation, washing, and drying, the pellets were suspended in alkaline loading buffer (30 mM NaOH, 1 mM EDTA [pH 8.0], 0.02% bromophenol blue, 50% glycerol) and boiled, and the 5′-end-labeled, single-stranded S1 probes were resolved and recovered from an alkaline denaturing agarose gel (12).

The following oligonucleotides and endonucleases were used to generate the probes: (i) APNG1 S1 probe, defined by 5′-End oligo (Fig. 1) and *Bgl*II, is antisense to nucleotides 96 to 367 in Fig. 1; (ii) APNG.7 S1 probe, defined by LK-3 oligo (5′ATTCCGCACTTTCCTTCGCG3′) and *Ava*II, is antisense to nucleotides 515 to 822; (iii) APNG.5 S1 probe, defined by 3′-Mid oligo (Fig. 1) and *Ava*I, is antisense to nucleotides 960 to 1195; and (iv) APNG.4 S1 probe, defined by PC-3 oligo (5′ACTGTAGTCAAAGTGACGTGAGGAG GGGTCCTCATAATTCATTAC3′) and *Taq*I, is antisense to nucleotides 1113 to 1318.

**Protein analyses: gel electrophoresis, immunoassay, and peptide comparisons.** Proteins were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) essentially as described by Laemmli (29). Towbin immunoassays (50) were done as described before (16), but proteins were electrotransferred to 0.22-μm-pore-size nitrocellulose (Schleicher & Schuell, Keene, N.H.), Pronase was omitted, blocking was done in 0.5% nonfat dried milk (Carnation Co., Los Angeles, Calif.), and the antiserum was used at dilutions of 1:50 to 1:150 in 5% bovine serum albumin. Antigen-antibody complexes were detected by using $^{125}$I-protein A and visualized by fluorography (30). Peptide comparisons were made by cleaving SDS-PAGE-separated proteins in situ with *N*-chlorosuccinimide (NCS; Aldrich Chemical Co., Milwaukee, Wis.) (32) and subjecting the resulting fragments to second-dimension SDS-PAGE (3, 4), followed by electrotransfer to nitrocellulose and sequential reaction with a rabbit antiserum, Anti-C1 (47), and $^{125}$I-protein A.

**Nucleotide sequence accession number.** The nucleotide sequence of the Colburn CMV APNG1 gene shown in Fig. 1 has been entered into GenBank as cytomegalovirus APNG1 and has been given the accession number M64627.

## RESULTS

**Assembly protein-coding sequence is 3′ end of larger ORF.** Nucleotide sequence analysis of the CMV (Colburn) genomic *Xba*I R fragment confirmed the cDNA sequence previously determined for the assembly protein-coding region (43) and revealed that the 930-bp coding sequence for the assembly protein precursor (nucleotides 1072 to 2001) is the 3′ end of a 1,770-bp open reading frame (ORF) (nucleotides 232 to 2001) that begins with a methionine and, together with its upstream regulatory region, was designated assembly protein nested gene 1 (APNG1) (Fig. 1). APNG1 includes an upstream potential TATA promoter element, contains three internal potential TATA promoters and three corresponding ATG translational start codons in addition to its own, and is followed by a single downstream polyadenylation signal. This organization indicated that the APNG1 region could give rise to four 3′-coterminal mRNAs able to

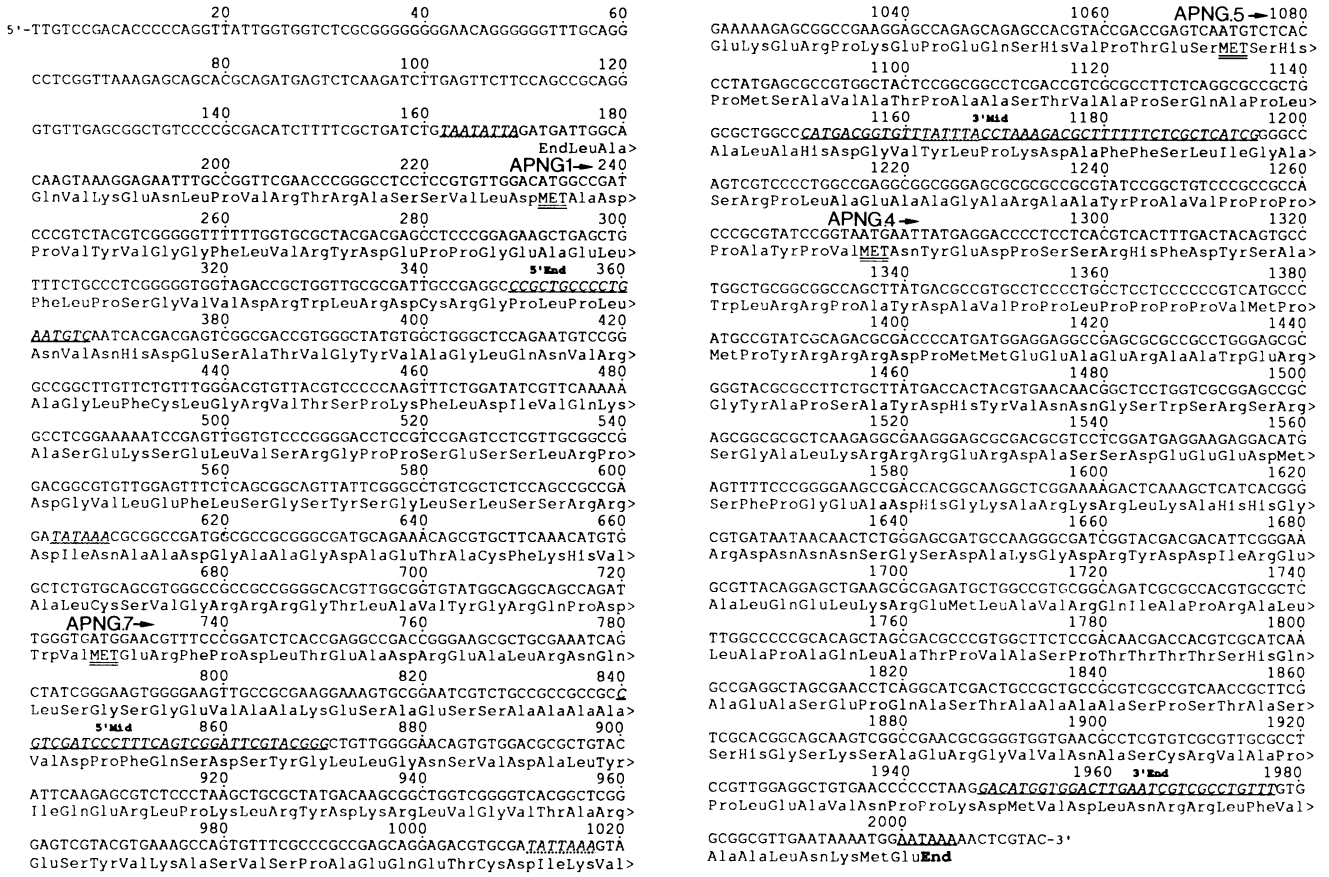


FIG. 1. Nucleotide and predicted amino acid sequences of CMV Colburn genomic region containing the assembly protein gene at the 3′ end of a larger, 1,770-bp ORF. This sequence is a portion of the strain Colburn *Xba*I R fragment (41). The following features are indicated for reference: (i) proposed TATA promoter elements are italicized and dot underlined, (ii) proposed translational start methionines for the coding sequence in each of the nested genes are capitalized and doubly underlined, and the designation of the corresponding assembly protein nested gene (APNG) is indicated above each (i.e., fractional length of APNG1 coding sequence), (iii) the single polyadenylation signal at the 3′ end is underlined, and (iv) sequences antisense to the probes used in the Northern analyses, i.e., 5′-End, 5′-Mid, 3′-Mid, and 3′-End (see Materials and Methods), are italicized and underlined.

encode four corresponding in-frame, overlapping proteins. These nested coding sequences are numbered according to their fractional length relative to that of the longest, APNG1. Homologous coding sequences in the other herpesviruses were similarly numbered for comparative purposes (see Table 2). Figure 1 presents the nucleotide and amino acid sequences of the APNG1 region and shows the positions of (i) the four putative TATA promoter elements and corresponding translational start ATGs, (ii) the single 3′ polyadenylation signal, and (iii) the four nucleotide sequences complementary to the antisense oligonucleotide probes described below.

**Identification of four 3′-coterminal transcripts from the Colburn nested genes.** The organization of the four RNAs and corresponding proteins potentially encoded by the 3′-coterminal nested genes is shown in Fig. 2A. To test for the predicted transcripts, antisense oligonucleotides complementary to sequences within each of the four nested genes were used as probes (Fig. 1, underlined and italicized). In the first experiment, three RNAs were detected by sequentially probing, stripping, and reprobing a single blot with oligonucleotides specific for the 5′ end of APNG1 (5′-End), an internal sequence near the 5′ end of APNG.7 (5′-Mid), and an internal sequence near the 5′ end of APNG.5 (3′-Mid) (Fig. 3A). The 5′-End probe hybridized to a 2.2-kb transcript, the 5′-Mid probe hybridized to the 2.2-kb and a scarce 1.6-kb transcript, and the 3′-Mid probe hybridized to both the 2.2-kb and the 1.6-kb transcripts as well as to a very abundant 1.2-kb species. Because the smallest RNA was predicted to be close to the size of the 1.2-kb RNA, a second experiment was done with a higher-percentage agarose gel (1.6% instead of 1.2%) and a sequence near the 3′ end of all four assembly protein nested genes (3′-End) as a probe. Under these conditions, a scarce 1-kb transcript was reproducibly detected as a weak band slightly ahead of the 1.2-kb RNA (arrow, Fig. 3B). Consistent with the overlapping organization suggested in Fig. 2A, when the blot was stripped and rehybridized with the 3′-Mid probe, the 1-kb transcript was no longer detected (Fig. 3B). The sizes of these four transcripts are consistent with those predicted from their sequences (Fig. 2A), allowing for a 200- to 300-nucleotide polyadenosine [poly(A)] tail.

**Identification of probable start sites for the Colburn overlapping transcripts.** The 5′ ends of the overlapping transcripts were mapped by S1 nuclease protection assays (Fig.

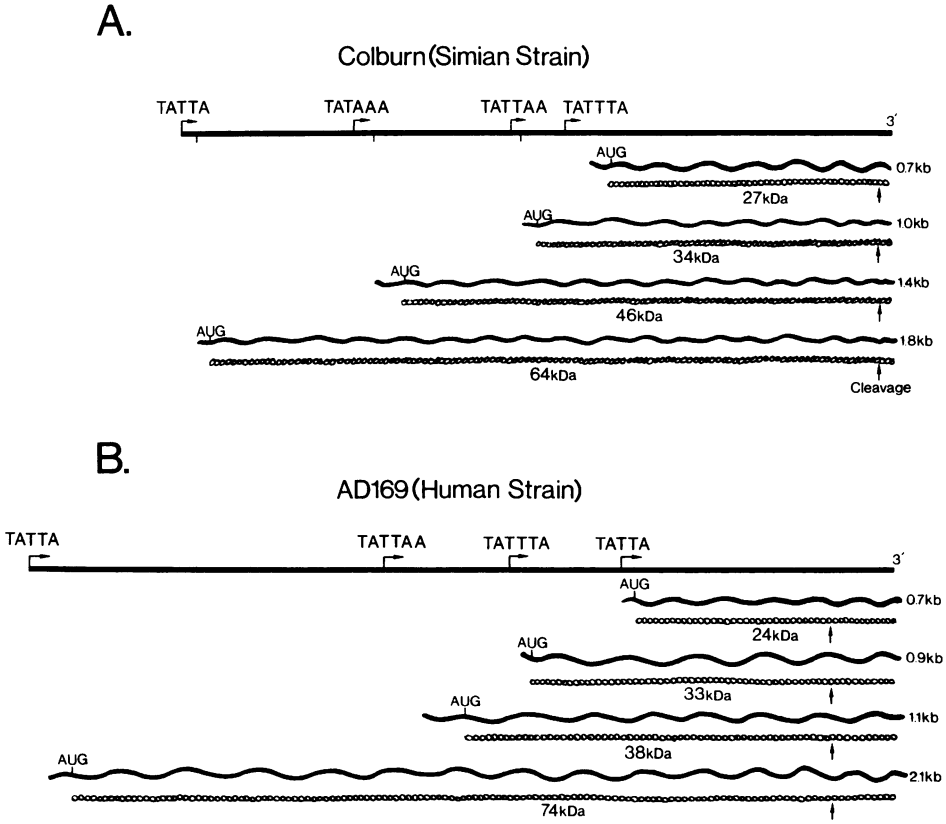


FIG. 2. Model for the transcriptional and translational expression of the 3'-coterminal, nested assembly protein-related genes of CMV strains Colburn (A) and AD169 (B). This model is based on the nucleotide sequence of strain Colburn shown in Fig. 1 and that of AD169 (CMV UL80; EMBL accession no. X17403). Potential TATA promoter elements are indicated above the top line representing DNA in both panels, and putative mRNA start sites are indicated as tic marks beneath the top line in panel A. The mRNAs (wavy lines) and corresponding proteins (beaded lines), along with their predicted sizes, are also indicated. Arrow at carboxyl end of proteins indicates putative maturational cleavage site (19).

4) by using the probes described in Materials and Methods. The 2.2-kb transcript protected a 168-nucleotide fragment, corresponding to a transcriptional start site 36 nucleotides upstream of the predicted translational start site of APNG1. No additional fragments were detected with this probe, as would be expected if the 2.2-kb RNA were the largest one transcribed from this region and there were no splicing in the region of the probe. The 1.6-kb transcript protected 168- and 169-nucleotide fragments, corresponding to transcriptional start sites 72 and 73 nucleotides upstream of the predicted translational start site of APNG.7, respectively. In addition, a fragment representing the entire 308-nucleotide probe was detected due to its protection by the overlapping 2.2-kb transcript. The 1.2-kb transcript protected a 148-nucleotide fragment corresponding to a transcriptional start site 27 nucleotides upstream of the translational start site of APNG.5. Again, the full-length 234-nucleotide probe was detected due to its protection by the larger overlapping 2.2-kb and 1.6-kb transcripts. Mapping of the 5' end of the scarce 1.0-kb transcript was inconclusive because of the comparatively high level of "background" fragments resulting from nonspecific S1 nuclease nicking of the probe hybridized to the very abundant overlapping 1.2-kb and 2.2-kb transcripts. Each of the three 5' ends determined mapped between the predicted promoter elements and corresponding translational start ATGs of the respective genes, as indicated by tic marks in Fig. 2A.

**Colburn nested genes are transcribed late.** To determine whether the nested genes were transcribed early (i.e., in the absence of viral DNA synthesis) or late (i.e., viral DNA synthesis required), their presence was tested for in infected cells treated with 500 μM phosphonoformic acid (PFA), a concentration that prevents viral DNA and late viral protein synthesis (1). Noninfected (mock) and nontreated infected cells were processed in parallel for comparison. RNA was isolated 4 days after infection, when the nontreated infected cells showed strong cytopathic effects but remained attached to the glass. The results of a Northern assay of the three mRNA preparations probed with $^{32}$P-labeled 3'-End oligonucleotide are shown in Fig. 5A. All four transcripts were present in the nontreated infected cells, but none was detected in the PFA-treated cells, indicating that all four RNAs are transcribed late. Similar amounts of β-actin mRNA were detected in all three lanes, demonstrating that the samples contained approximately equivalent amounts of mRNA (Fig. 5B).

**Colburn nested genes encode multiple in-frame, overlapping proteins.** Four in-frame overlapping proteins could be encoded by the four transcripts detected (Fig. 2A). To determine whether these proteins are made, a rabbit antiserum, Anti-C1, was prepared to a synthetic peptide representing the carboxy-terminal 21-amino-acid sequence of the Colburn assembly protein precursor (47). Anti-C1 recognizes only the predominantly cytoplasmic precursor form of the assembly

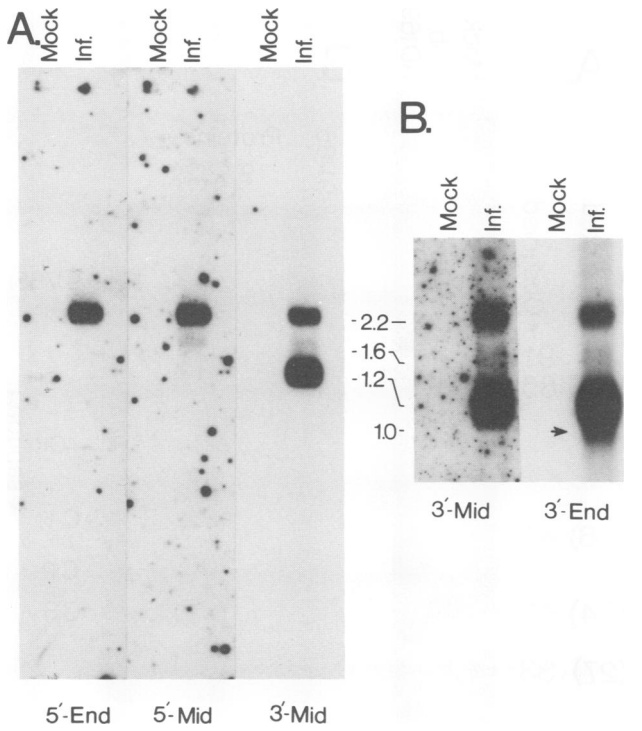


FIG. 3. Four overlapping, 3'-coterminal transcripts detected in CMV Colburn-infected cells. Northern blot analyses were done as described in Materials and Methods with the indicated $^{32}$P-end-labeled oligonucleotide probes (antisense to the sequences underlined and italicized in Fig. 1). Each lane contained approximately 1 μg of RNA, as determined by $A_{260}$, from mock-infected (Mock) or infected (Inf.) cells. Overnight hybridization at 37°C and subsequent washings were done as described in Materials and Methods. (A) Collage of three resulting fluorograms produced by probing the same blot sequentially with the probes indicated. (B) Collage of two fluorograms prepared from a different blot that was probed sequentially with the probes indicated. RNA sizes are indicated (in kilobases) between the panels; arrow indicates position of the 1.0-kb band. Fluorograms were aligned by matching background spots that were not removed by the stripping procedure.

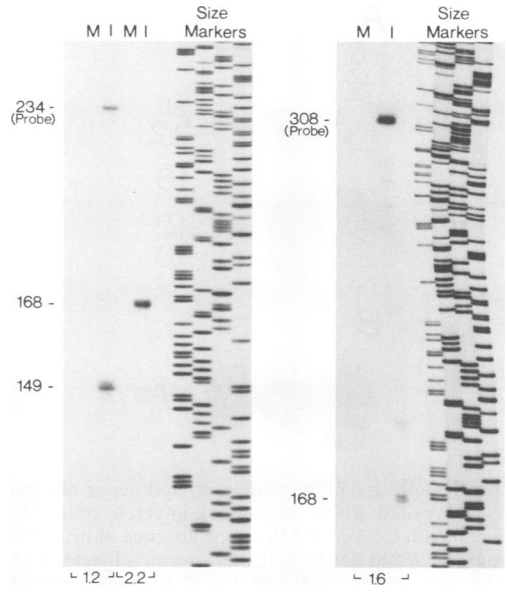


FIG. 4. Transcriptional start sites of three CMV Colburn nested genes. S1 nuclease protection assays were done as described in Materials and Methods on RNA from mock-infected (M) and infected (I) cells to map the 5' ends of the Colburn 2.2-, 1.6-, and 1.2-kb transcripts. The sizes of the labeled oligonucleotide probes that were protected from S1 nuclease digestion were estimated by using an M13mp8 sequence (Size Markers lane) produced according to the USB protocol with the control DNA supplied with the Sequenase kit (USB). The resulting autoradiograms shown here reveal the sizes (in nucleotides) of the probe sequences that were S1 resistant. The calculated locations of the three transcriptional start sites are indicated in Fig. 2A by tic marks beneath the top line.

protein because maturational cleavage eliminates its carboxyl end, including the C1 antigenic sites (19, 47). Anti-C1 was used in Towbin immunoassays to detect antigenically related species in the NP-40 cytoplasmic fractions of Colburn-, AD169- (discussed below), and mock-infected cells (Fig. 6A). Five bands, having estimated sizes of 91, 85, 49, 40, 33, and 23 kDa, were detected in Colburn-infected cells. The sizes of the 49-, 40-, and 33-kDa bands approximated those predicted from the sequences of APNG.7, APNG.5, and APNG.4, respectively, as indicated in Fig. 6A. The two largest bands, 91 and 85 kDa, were significantly (42 and 33%, respectively) larger than predicted from the nucleotide sequence of APNG1, suggesting that they may have posttranslational modifications or comparatively unusual structural characteristics or both.

To verify that these immunologically cross-reactive proteins were structurally related and corresponded to the predicted nested gene products, peptide comparisons were made following protein cleavage at tryptophan residues by NCS, as described in Materials and Methods. An immunoimage showing the resulting peptides detected by Anti-C1 (Fig. 6B) demonstrated the relatedness of the Colburn proteins at their carboxyl ends. All proteins shared at least some carboxyl peptides, and the number and sizes of the observed peptides from each protein were consistent with the positions of tryptophans in the predicted proteins (Fig. 1; also see Fig. 1B in reference 47).

**Structure and expression of human CMV AD169 assembly protein coding region is similar to that of simian strain Colburn.** The assembly protein gene of the human CMV strain AD169 lies within the *Hin*dIII L fragment (43) and corresponds to the 3' portion of the human CMV UL80 ORF (9). Examination of UL80 (EMBL accession no. X17403) revealed a genomic organization similar to that of Colburn: a large 2,121-bp ORF beginning with methionine and referred to as UL80a, with four potential TATA elements, four corresponding translational start codons, and one 3' polyadenylation signal. This organization again suggested the potential to encode four overlapping 3'-coterminal transcripts and four corresponding in-frame overlapping proteins (Fig. 2B). It is worth noting that the UL80 ORF (2,235 bp) also begins with a methionine and would code for a protein 38 amino acids longer at its amino end than that of UL80a. However, it is questionable that the UL80 ORF is transcribed, because the single TATA element within 600 bp upstream of its translational start codon is only one nucleotide away—considered too close to be functional. In addition, the ATG beginning the UL80 ORF has a pyrimidine at both its −3 and +1 positions, indicating that it is in a poor translational start context (26) and is not likely to be utilized even if it is transcribed.

To determine whether the predicted transcripts were

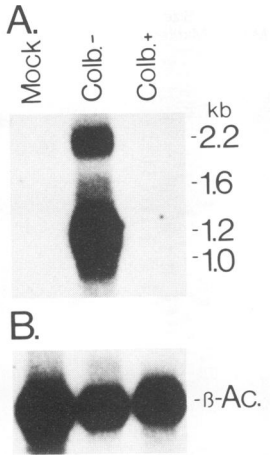


FIG. 5. Nested, assembly protein-related genes are transcribed late. Polyadenylated RNA from mock-infected cells (Mock) and cells infected with Colburn CMV in the absence of drug (Colb.−) or in the presence of 500 μM PFA (Colb.+) was subjected to Northern assay. The blot was probed with the 3′-End oligonucleotide (A), stripped, and reprobed with a riboprobe to β-actin mRNA (B). Hybridization with the 3′-End probe was done as described in Materials and Methods and in the legend to Fig. 3. Hybridization with the β-actin probe was done overnight as described in Materials and Methods; the blot was washed twice in 1× SSPE–0.1% SDS and twice in 0.1× SSPE–0.1% SDS and then visualized by a 1-h fluorographic exposure. RNA sizes and the position of β-actin (β-Ac.) are indicated in the right-hand margin.

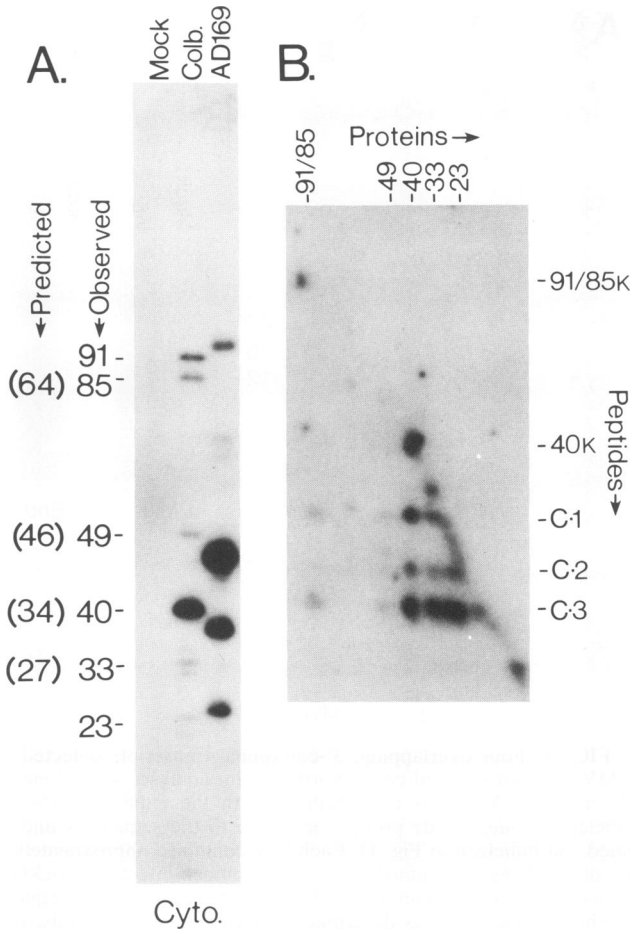


FIG. 6. CMV Colburn assembly protein-related overlapping, in-frame proteins identified. Cytoplasmic proteins from mock-infected cells (Mock) or cells infected with CMV Colburn (Colb.) or AD169 (AD169) were analyzed by Towbin immunoassay with Anti-C1 as described in Materials and Methods (A). A fluorogram of the resulting nitrocellulose replica is shown here, and the observed and predicted sizes of the proteins are indicated in the left-hand margin. A peptide comparison (B) was done by using Anti-C1 to identify carboxyl fragments of the proteins identified in panel A following their cleavage in situ with NCS, as described in Materials and Methods. The positions of proteins separated in the first gel are indicated by their sizes (in kilodaltons) at the top of the figure; the positions of noncleaved proteins (i.e., 91/85K and 40K) and carboxyl peptides C-1 to C-3 (47) are indicated in the right-hand margin.

synthesized, a Northern assay was done with a 43-mer oligonucleotide probe, AD-End: (5′ATTCAGATCTACCATG TCTTTGGGCGGGCTCTGGGAAGCTGCG3′), complementary to the 3′ end of AD169 UL80. Four transcripts were detected (Fig. 7), and their sizes were estimated to be 2.4, 1.5, 1.2, and 0.9 kb, consistent with those predicted from the sequence (Fig. 2B), allowing for the presence of 5′ and 3′ untranslated regions. The apparent absence of a second high-molecular-weight RNA is taken as additional evidence that the UL80 ORF is not transcribed in its entirity. When the upstream putative promoter sequences and the relative lengths of the genes for these AD169 RNAs were compared with those of the Colburn nested genes, it was found that the 2.4-, 1.5-, and 1.2-kb RNAs of AD169 appear to correspond, respectively, to the 2.2-, 1.2-, and 1.0-kb RNAs of Colburn (Table 1). The 0.9-kb RNA of AD169 and the 1.6-kb RNA of Colburn do not have such counterparts.

To determine whether the four predicted protein products of these transcripts were also made, the NP-40 cytoplasmic fractions of mock- and AD169-infected cells were subjected to immunoassay with Anti-C1 as described above for strain Colburn. Nineteen of the 21 amino acids of the strain Colburn C1 peptide are conserved at the carboxyl end of the strain AD169 assembly protein precursor (19, 41), making the Anti-C1 serum cross-reactive between the two. Four proteins were detected by Anti-C1, and their sizes were estimated to be 94, 46, 37, and 29 kDa (Fig. 8A). With the exception of the largest protein, as noted above for strain Colburn, these sizes approximate those predicted from the four nested genes (see predicted versus observed values in Fig. 8A). Peptide comparisons of these proteins were made following NCS cleavage as described above, and the resulting immunoimmage (Fig. 8B) showed related peptide patterns that differed in accordance with the relative positions of the tryptophans in each overlapping gene. These results confirm the predicted relatedness of the AD169 overlapping proteins and are consistent with the model presented in Fig. 2B.

## DISCUSSION

A cDNA clone known to encode the strain Colburn CMV assembly protein precursor contained a potential TATA promoter sequence upstream of the assembly protein translational start codon (41), suggesting that it had been copied from an RNA larger than required to code for the assembly protein. To determine the significance of this observation, we cloned and sequenced the genomic region containing the assembly protein gene and identified a larger, 1,770-bp ORF (APNG1) that included the assembly protein-coding se-

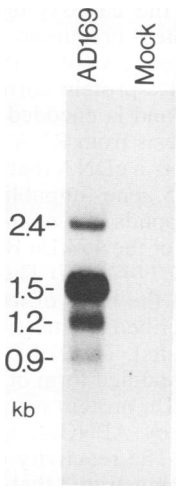


FIG. 7. Four similarly nested 3′-coterminal transcripts detected in CMV AD169-infected cells. RNA prepared from mock-infected cells (Mock) and cells infected with CMV strain AD169 (AD169) was subjected to Northern assay with $^{32}$P-end-labeled AD3′-End probe. Each lane contained approximately 1 μg of RNA, as determined by $A_{260}$. The blot was hybridized overnight at 37°C and further processed as described in Materials and Methods. Shown here is a fluorogram of the resulting blot. RNA sizes are indicated in the left-hand margin.

quence at its 3′ end. The complete sequence (Fig. 1) revealed a potential TATA promoter element upstream of APNG1, three more within APNG1, each followed by a corresponding translational start codon, and a single polyadenylation signal at the 3′ end of APNG1, suggesting that this region could give rise to four 3′-coterminal transcripts that encode four in-frame, overlapping, carboxy-coterminal proteins (Fig. 2A). The experiments described here support this model for both simian (Colburn) and human (AD169) strains of CMV.

The transcriptional portion of the model was substantiated by the following evidence. First, four RNAs having sizes close to those expected for the transcripts of the four nested genes were detected by using Northern assays (Fig. 3 and 7). Second, the nested, overlapping relationship of the transcripts was confirmed by demonstrating that specific oligonucleotide probes reacted in Northern assays with only the

TABLE 1. Correspondence of transcripts with specific assembly protein nested genes

| RNA species (kb)[a] | | Nested gene[b] | | Putative promoter region[c] |
|---|---|---|---|---|
| Colburn | AD169 | Colburn | AD169 | |
| **2.2** | **2.2** | APNG1 | UL80a | ATCTGTAATATTAGATGATTGGC |
| 1.6 | | APNG.7 | | GCCGAGATATAAACGCGG |
| **1.2** | **1.5** | APNG.5 | UL80.5 | CGTGCGATATTAAAG |
| **1.2** | **1.2** | APNG.4 | UL80.4 | GTTATTTACC |
| | 0.9 | | UL80.3 | ACCGGCCTATTACCGTCG |

[a] Boldfaced numbers indicate RNA species that have counterparts in Colburn and AD169.

[b] Assignment of Colburn RNAs to specific nested genes (i.e., APNG1, APNG.7, APNG.5, and APNG.4) is based on data from transcript mapping experiments summarized in Fig. 2, 3, and 4. Assignment of AD169 RNAs to specific nested genes (i.e., UL80a, UL80.5, UL80.4, and UL80.3) is extrapolated from data shown in Fig. 7, 8, and 9. The numerical designations of the genes indicate their relative lengths within the nested set.

[c] Also see Table 2.

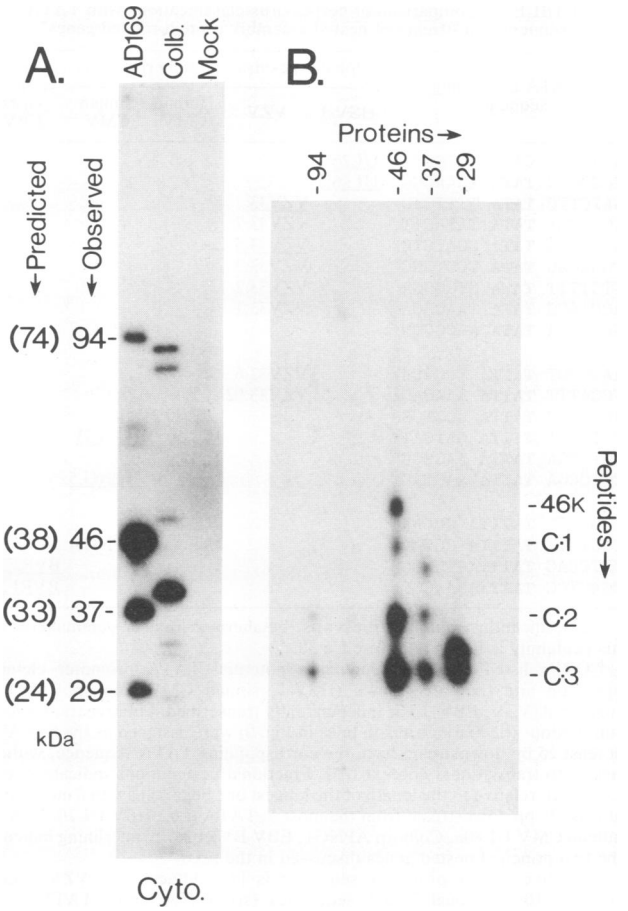


FIG. 8. Four CMV AD169 assembly protein-related, overlapping, in-frame proteins identified. Cytoplasmic fractions of mock-infected cells (Mock) and cells infected with CMV strains AD169 (AD169) and Colburn (Colb.) were subjected to Towbin immunoassay with Anti-C1 as described in Materials and Methods (A). A fluorogram of the resulting nitrocellulose replica is shown here, and the observed and predicted sizes of the proteins are indicated in the left-hand margin. This is the same fluorogram shown in Fig. 6A but printed backwards for convenience of labeling and discussion. A peptide comparison (B) was done by using Anti-C1 to identify carboxyl fragments of the proteins identified in panel A following their cleavage in situ with NCS, as described in Materials and Methods. The positions of proteins separated in the first gel are indicated by their sizes (in kilodaltons) at the top of the figure; the positions of noncleaved proteins (i.e., 46K) and carboxyl peptides C-1 to C-3 are indicated in the right-hand margin.

predicted RNAs. Third, the 5′ ends of the three largest Colburn transcripts were located between the predicted TATA promoter sequences and translational start sites by using S1 nuclease protection assays (Fig. 2A). The good correspondence between predicted and observed transcript sizes argues against the possibility that these 5′ ends represent splice acceptor sites rather than true 5′ termini. Finally, the presence of a single polyadenylation signal at the 3′ end of the coding region, considered together with the sizes and overlapping nature of the RNAs, is consistent with the transcripts being 3′ coterminal.

None of these RNAs were detected in PFA-treated infected cells (Fig. 5, data from parallel experiments with AD169 not shown), identifying them as late transcripts.

TABLE 2. Comparison of herpesvirus classification[a] with TATA sequences upstream of nested assembly protein-related genes[b]

| TATA-containing sequence[c] | Alphaherpesvirus | | Betaherpesvirus | | Gamma-herpes-virus |
|---|---|---|---|---|---|
| | HSV-1 | VZV | Human CMV | Simian CMV | EBV |
| AAAAGGT **CATA** CCCCGTA | UL26 | | | | |
| GCGCGGG **TATA** AGAACGG | UL26.5 | | | | |
| GATCTTC **TATA** TTTTGTC | | VZV33 | | | |
| CGTGCGT **TATA** TATACTT | | VZV33.7 | | | |
| GGGCCCA **TATA** CGATCTC | | VZV33.5 | | | |
| CGGACAG **TATA** TCGTCGG | | VZV33.3 | | | |
| TTCTTTT **TATA** GTGGGGA | | VZV33.2 | | | |
| ACCAACG **TATA** TAACGCA | | VZV33.1 | | | |
| GCCGAGA **TATA** AACGCGG | | | | APNG.7 | |
| AAGGTGT **TATTA** TCCACA | | VZV33.6 | | | |
| TCGATTT **TATTA** AAACAG | | VZV33.02 | | | |
| ACCGGCC **TATTA** CCGTCG | | | UL80.3 | | |
| TCTGTAA **TATTA** GATGAT | | | UL80a[d] | APNG1 | |
| CGTGCGA **TATTA** AAGCGC | | | UL80.5 | | |
| CGTGCGA **TATTA** AAGTAG | | | | APNG.5 | |
| CGGAGTT **TATTTA** CCCAA | | | UL80.4 | | |
| CGGTGTT **TATTTA** CCTAA | | | | APNG.4 | |
| TTTCCAG **TATTTA** TCACG | | | | | BVRF2a[d] |
| CGAGTCG **TATTTA** AAGGC | | | | | BVRF2.6 |

[a] Classification as alphaherpesvirus, betaherpesvirus, or gammaherpesvirus subfamily is based on reference 38.

[b] ORFs listed are those having a potential TATA promoter element upstream and that are known (HSV-1, simian CMV, human CMV) or suggested (VZV, EBV) to be independently transcribed. Putative translational start codons (tic marks beneath lines in Fig. 9) were assigned as the first ATG at least 25 bp downstream from the corresponding TATA sequence, without regard to translational context (25). Fractional designations indicate size of each ORF relative to the length of the longest one beginning with a methionine at least 25 bp downstream from the most 5′ TATA (i.e., HSV UL26, VZV33, human CMV UL80a, Colburn APNG1, EBV BVRF2a). Underlining indicates the two principal nested genes discussed in the text.

[c] Putative TATA promoter sequence is in boldface. The VZV33 gene contains 10 additional TATA sequences (six TATA, three TATTA, one TATTTA) that are not listed because they adjoin the translational start codon (one) or because they were upstream of a more ATG-proximal TATA (nine) and therefore were considered secondary.

[d] The human CMV UL80 and EBV BVRF2 ORFs were originally designated without regard to translational start codons and consequently begin upstream of the probable amino termini of these proteins. The designations UL80a and BVRF2a are used to indicate the longest ORFs in UL80 and BVRF2, respectively, that begin with methionine.

Their different relative amounts during late lytic infection, however, indicate that their synthesis is differentially regulated, most probably through the respective TATA-containing sequences. Evidence that these sequences are important regulatory domains came from the observation that three of the four are identically conserved between the Colburn and AD169 nested genes: (i) ATCTGTAATATTAGATGATTG GC upstream of both Colburn APNG1 and AD169 UL80a, (ii) CGTGCGATATTAAAG upstream of Colburn APNG.5 and AD169 UL80.5, and (iii) GTTTATTTACC upstream of Colburn APNG.4 and AD169 UL80.4 (Tables 1 and 2). With the apparent exception of TATTTA (51), these presumptive regulatory sequences differ from consensus binding domains for known host transcription factors, as might be expected if they interact with viral rather than cellular factors. Furthermore, they are not conserved upstream of the assembly protein gene homologs of the other sequenced herpesviruses (Table 2), consistent with their being responsive to evolutionarily diverged viral regulatory factors.

Four proteins could be encoded by the overlapping transcripts, as depicted in Fig. 2, and corresponding proteins were detected in both Colburn- and AD169-infected cells. At least five bands in Colburn-infected cells reacted with Anti-C1, an antiserum to the carboxy-terminal portion of the Colburn assembly protein precursor. Three of these can be identified as the products of APNG1, APNG.7, and APNG.5. (i) The 40-kDa protein corresponds to the assembly protein precursor and is encoded by APNG.5, as demonstrated by its synthesis from RNA transcribed and translated in vitro from either a cDNA that contains APNG.5 (41) or the cloned APNG.5 gene (unpublished results). (ii) The 49-kDa protein corresponds in size and immunological reactivity to the precursor of the 45-kDa B-capsid protein, which is coded for by APNG.7 (47). (iii) The 85-kDa protein is the product of APNG1 on the basis of its comigration with the product of RNA transcribed in vitro from the cloned APNG1 gene (unpublished results). The 91-kDa protein is likely to be a posttranslationally modified form of the 85-kDa product of APNG1, and the 33-kDa protein is approximately the right size to be coded for by APNG.4, but these relationships remain to be proven. The reactivity of these proteins with the Anti-C1 serum demonstrated that they are in frame with one another and have the same carboxy-terminal sequence. This and their overlapping relationship were further supported by the results of peptide comparisons, which demonstrated that the proteins shared the expected carboxyl fragments (Fig. 6B).

Parallel experiments identified the four predicted protein products of the homologous AD169 nested genes and verified their carboxy-coterminal, overlapping relationship (Fig. 8). Like their Colburn counterparts, there was better agreement between predicted and observed sizes for the three smallest proteins than for the largest. The reason that several of the Colburn proteins appeared as doublet bands while their apparent counterparts in AD169 were single bands (Fig. 6 and 8) is not known but may relate to the fact that AD169 was infecting cells of its natural host (human) while the simian strain Colburn was infecting cells of another species. This cross-species infection could result in imperfectly matched signaling systems and cause aberrant processing. In this connection, it is noted that glycoproteins gp119/100 and gp65/61 from Colburn virions produced in human cells also appear as doublets, while their respective human CMV counterparts, gp130 and gp62, are single bands (3).

Homologs of the CMV assembly protein gene have been identified in HSV type 1 (HSV-1) (HSV UL26) (37), Epstein-Barr virus (EBV) (BVRF2) (2), and varicella-zoster virus (VZV33) (13), and they also appear to be organized as overlapping genes (Fig. 9). The homologous region of HSV-1 has only two nested genes. The smaller one, UL26.5, begins about halfway through the larger, UL26, and appears to encode the HSV assembly protein counterpart (i.e., VP22a, p40, NCP-3, ICP35e) (34, 37). A product of HSV UL26, p80, has been identified and, like the overlapping CMV proteins, appears to be in frame with the assembly protein product of UL26.5 (34, 37, 46, 52). The assembly protein homologs of EBV have not been described, but their coding region (2, 9), like that of HSV, contains two potential in-frame, 3′-coterminal genes (Fig. 9). The homologous region of VZV stands out as being potentially the most complicated, with at least eight possible 3′-coterminal, nested genes (Fig. 9). Phosphoprotein products 32 to 38 kDa in size and possibly representing the products of VZV33.5 and VZV33.6 have been described (14). A common feature of these homologous genes in all five viruses is that each contains one large ORF approximately 2 kb in size and an in-frame nested gene approximately one-half as large (see underlined TATAs, Fig. 9). It was noted that the TATA sequences upstream of these two principal genes differ between the alphaherpesvirus (C/

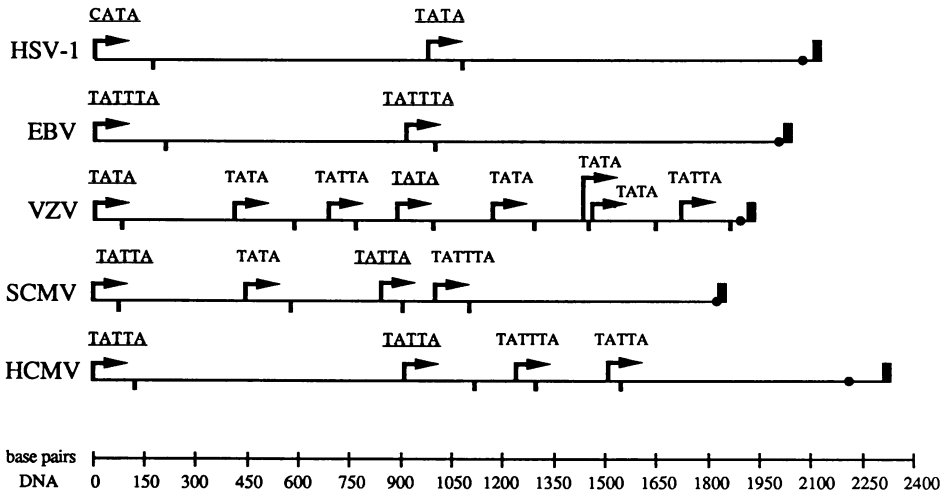


FIG. 9. Genetic organization of assembly protein gene homologs is similar in other herpesviruses. The genetic organization of the assembly protein nested genes of five different herpesviruses is presented schematically. Arrows represent potential TATA promoter elements, with the specific sequence shown above each; tic marks below the line following each TATA represent corresponding in-frame translational start codons at least 25 bp downstream; and the solid circles and solid rectangles at the right-hand end of each line represent translation stop codons and 3′ polyadenylation signals, respectively. Underlined TATAs precede the two known (HSV-1, human CMV [HCMV], and simian CMV [SCMV]) and suggested (VZV and EBV) principal transcription units of each nested set. Each set of the genes is drawn to the scale at the bottom of the figure. Data were from Fig. 1 for CMV Colburn (SCMV) and from published sequences for HSV-1 (37), EBV (2), VZV (13), and CMV AD169 (HCMV) (9).

TATA), betaherpesvirus (TATTA), and gammaherpesvirus (TATTTA) subfamilies (Table 2; principal genes underlined). The TATAs upstream of the corresponding major capsid protein genes (also transcribed late), including that of human herpesvirus 6 (33) which has many characteristics of the betaherpesvirus subfamily (35, 42), fit with this pattern, except that of EBV, which is preceded by TATTA rather than TATTTA. Although this observation suggests the possibility of herpesvirus subfamily-specific TATAs for late genes and may provide some insight into the regulation and evolution of these genes, more information is needed to assess its significance.

Other proteins from evolutionarily diverse viruses, as well as some cellular proteins (7, 44 and references therein), appear to have a similar genetic organization. Specific examples of overlapping viral proteins having the same carboxyl ends but different amino termini include (i) the VP2 and VP3 capsid proteins of polyomavirus (18, 22) and simian virus 40 (40, 43), (ii) the C and Nu3 (scaffold) proteins of bacteriophage lambda (48), (iii) the pre-S1, pre-S2, and S proteins of hepatitis virus (8, 27, 38), and (iv) the p80 and p40 proteins of HSV-1 (34). All of these are involved in virus assembly, and it has been suggested (48) that their genetic organization may provide a means of encoding and regulating a family of proteins able to interact with each other through their identical carboxyl domains and with other molecules through unique domains at their amino ends. Whether and how the four CMV proteins described here interact with each other and with other molecules remains to be determined, but answers to these questions should lead to a better understanding of their function.

## ACKNOWLEDGMENTS

We thank Peter Collins for the oligonucleotides designated PC-3 and 3′-Mid and for helpful advice on using oligonucleotides as probes in Northern assays, and we thank Annie Colberg-Poley and Dan Tenney for the p18-2 plasmid (β-actin). We also appreciatively acknowledge Jody Franklin and Ponniah Shenbagamurthi of the Johns Hopkins DNA/Protein/Peptide Facility, who synthesized the other oligonucleotides, and Lori Robson, who cloned Colburn *Xba*I-R.

A.R.W. was supported by training grant T32 GM07626. This work was aided by Public Health Service research grants AI13718 and AI22711.

## REFERENCES

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. **Baer, R., A. T. Bankier, M. D. Biggin, P. L. Deininger, P. J. Farrell, T. J. Gibson, G. Hatfull, G. S. Hudson, S. C. Satchwell, C. Séquin, P. S. Tufnell, and B. G. Barrell.** 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature (London) **310**:207–211.
3. **Benko, D. M., and W. Gibson.** 1986. Primate cytomegalovirus glycoproteins: lectin-binding properties and sensitivities to glycosidases. J. Virol. **59**:703–713.
4. **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.
5. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell **12**:721–732.
6. **Braun, D., B. Roizman, and L. Pereira.** 1984. Characterization of post-translational products of herpes simplex virus gene 35 proteins binding to the surfaces of full capsids but not empty capsids. J. Virol. **49**:142–153.
7. **Carlson, M., R. Taussig, S. Kustu, and D. Botstein.** 1983. The secreted form of invertase in *Saccharomyces cerevisiae* is synthesized from mRNA encoding a signal sequence. Mol. Cell. Biol. **3**:439–447.
8. **Cattaneo, R., H. Will, N. Hernandez, and H. Schaller.** 1983. Signals regulating hepatitis B surface antigen transcription. Nature (London) **305**:336–338.
9. **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.
10. **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.
11. **Colberg-Poley, A. M., and L. D. Santomenna.** 1988. Selective induction of chromosomal gene expression by human cytomegalovirus. Virology **166**:217–228.
12. **Davis, L. G., M. D. Dibner, and J. F. Battey.** 1986. Basic methods in molecular biology, p. 143–146. Elsevier Science Publishing Co., Inc., New York.
13. **Davison, A. J., and J. E. Scott.** 1986. The complete DNA sequence of varicella-zoster virus. J. Gen. Virol. **67**:1759–1816.
14. **Friedrichs, W. E., and C. Grose.** 1986. Varicella-zoster p32/p36 is present in both the viral capsid and the nuclear matrix of the infected cell. J. Virol. **57**:155–164.
15. **Gibson, W.** 1981. Structural and non structural proteins of strain Colburn cytomegalovirus. Virology **111**:516–537.
16. **Gibson, W.** 1981. Protease-facilitated transfer of high-molecular-weight proteins during electrotransfer to nitrocellulose. Anal. Biochem. **118**:1–3.
17. **Gibson, W.** The molecular biology of human cytomegalovirus. *In* E.-S. Huang and C. A. Alford (ed.), Diseases caused by human cytomegalovirus. Kluwer Academic Publishers, Norwell, Mass., in press.
18. **Gibson, W., T. Hunter, B. Cogen, and W. Eckhart.** 1977. Altered virion proteins of a temperature-sensitive mutant of polyoma virus, ts 59. Virology **80**:21–41.
19. **Gibson, W., A. I. Marcy, J. C. Comolli, and J. Lee.** 1990. Identification of precursor to cytomegalovirus capsid assembly protein and evidence that processing results in loss of its carboxy-terminal end. J. Virol. **64**:1241–1249.
20. **Gibson, W., and B. Roizman.** 1972. Proteins specified by herpes simplex virus. VIII. Characterization and composition of multiple capsid forms of subtypes 1 and 2. J. Virol. **10**:1044–1052.
21. **Gibson, W., and B. Roizman.** 1974. Proteins specified by herpes simplex virus. X. Staining and radiolabeling properties of B-capsid and virion proteins in polyacrylamide gels. J. Virol. **13**:155–165.
22. **Hunter, T., and W. Gibson.** 1978. Characterization of the mRNAs for the polyomavirus capsid proteins VP1, VP2, and VP3. J. Virol. **28**:240–253.
23. **Irmiere, A., and W. Gibson.** 1983. Isolation and characterization of a noninfectious virionlike particle released from cells infected with human strains of cytomegalovirus. Virology **130**:118–133.
24. **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.
25. **Kaufer, N., H. M. Fried, W. F. Schwindinger, M. Jasin, and J. R. Warner.** 1983. Cycloheximide resistance in yeast: the gene and its promoter. Nucleic Acids Res. **11**:3123–3134.
26. **Kozak, M.** 1987. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. **15**:8125–8148.
27. **Kuroki, K., R. Russnak, and D. Ganem.** 1989. Novel N-terminal amino acid sequence required for retention of a hepatitis B virus glycoprotein in the endoplasmic reticulum. J. Virol. **63**:4459–4466.
28. **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.
29. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
30. **Laskey, R. A., and A. D. Mills.** 1977. Enhanced autoradiographic detection of $^{32}P$ and $^{125}I$ using intensifying screens and hypersensitized film. FEBS Lett. **82**:314–316.
31. **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.
32. **Lischwe, M. A., and D. Ochs.** 1982. A new method for partial peptide mapping using *N*-chlorosuccinimide/urea and peptide silver staining in sodium dodecyl sulfate-polyacrylamide gels. Anal. Biochem. **127**:453–457.
33. **Littler, E., G. Lawrence, M.-Y. Liu, B. G. Barrell, and J. R. Arrand.** 1990. Identification, cloning, and expression of the major capsid protein gene of human herpesvirus 6. J. Virol. **64**:714–722.
34. **Liu, F., and B. Roizman.** 1990. The promoter, transcriptional unit, and coding sequence of herpes simplex virus 1 family 35 proteins are contained within and in frame with the $U_L26$ open reading frame. J. Virol. **65**:206–212.
35. **Lopez, C., and R. W. Honess.** 1990. Human herpesvirus-6, p. 2055–2062. *In* B. N. Fields and D. M. Knipe (ed.), Virology, 2nd ed. Raven Press, Ltd., New York.
36. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 466–467. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
37. **McGeoch, D. J., M. A. Dalrymple, A. J. Davison, A. Dolan, M. C. Frame, D. McNab, L. J. Perry, et al.** 1988. The complete sequence of the long unique region in the genome of herpes simplex virus type 1. J. Gen. Virol. **69**:1531–1574.
38. **Neurath, A. R., and S. B. H. Kent.** 1988. The pre-S region of hepadnavirus envelope proteins, p. 65–142. *In* K. Maramorosch, F. A. Murphy, and A. J. Shatkin (ed.), Advances in virus research, vol. 34. Harcourt Brace Jovanovich Publishers, Academic Press, Inc., San Diego, Calif.
39. **Preston, V. G., J. A. V. Coates, and F. J. Rixon.** 1983. Characterization of a herpes simplex virus gene product required for encapsidation of virus DNA. J. Virol. **45**:1056–1064.
40. **Reddy, V. E., R. Dhar, and S. M. Weissman.** 1978. Nucleotide sequence of the genes for the simian virus 40 proteins VP2 and VP3. J. Biol. Chem. **253**:621–630.
41. **Robson, L., and W. Gibson.** 1989. Primate cytomegalovirus assembly protein: genome localization and nucleotide sequence. J. Virol. **63**:669–676.
42. **Roizman, B.** 1990. Herpesviridae: a brief introduction, p. 1787–1793. *In* B. N. Fields and D. M. Knipe (ed.), Virology, 2nd ed. Raven Press, Ltd., New York.
43. **Rozenblatt, S., R. C. Mulligan, M. Gorecki, B. Roberts, and A. Rich.** 1976. Direct biochemical mapping of eukaryotic viral DNA by means of a linked transcription-translation cell-free system. Proc. Natl. Acad. Sci. USA **73**:2747–2751.
44. **Russo, R. N., N. L. Shaper, and J. H. Shaper.** 1990. Bovine β1-4-galactosyltransferase: two sets of mRNA transcripts encode two forms of the protein with different amino-terminal domains. J. Biol. Chem. **256**:3324–3331.
45. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
46. **Schenk, P., S. Pietschmann, H. Gelderblom, G. Pauli, and H. Ludwig.** 1988. Monoclonal antibodies against herpes simplex virus type 1-infected nuclei defining and localizing the ICP8 protein, 65K DNA-binding protein and polypeptides of the ICP35 family. J. Gen. Virol. **69**:99–111.
47. **Schenk, P., A. S. Woods, and W. Gibson.** 1991. The 45-kDa protein of cytomegalovirus (Colburn) B-capsids is an amino-terminal extension form of the assembly protein. J. Virol. **65**:1525–1529.
48. **Shaw, J. E., and H. Murialdo.** 1980. Morphogenetic genes C and Nu3 overlap in bacteriophage λ. Nature (London) **283**:30–35.
49. **Sherman, G., and S. L. Bachenheimer.** 1988. Characterization of intranuclear capsids made by ts morphogenic mutants of HSV-1. Virology **163**:471–480.
50. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76**:4350–4354.
51. **Wobbe, C. R., and K. Struhl.** 1990. Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription in vitro. Mol. Cell. Biol. **10**:3859–3867.
52. **Zweig, M., C. J. Heilman, H. Rabin, and B. Hampar.** 1980. Shared antigenic determinants between two distinct classes of proteins in cells infected with herpes simplex virus. J. Virol. **35**:644–652.