JOURNAL OF VIROLOGY, May 1987, p. 1358–1367
0022-538X/87/051358-10$02.00/0

Vol. 61, No. 5

Map Position and Nucleotide Sequence of the Gene for the Large Structural Phosphoprotein of Human Cytomegalovirus

G. JAHN,[1]* T. KOUZARIDES,[2] M. MACH,[1] B.-C. SCHOLL,[1] B. PLACHTER,[1] B. TRAUPE,[1] E. PREDDIE,[2]† S. C. SATCHWELL,[2] B. FLECKENSTEIN,[1] AND B. G. BARRELL[2]

Institut für Klinische Virologie der Universität Erlangen-Nürnberg, D-8520 Erlangen, Federal Republic of Germany,[1] and Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, United Kingdom[2]

Received 3 November 1986/Accepted 14 January 1987

Human cytomegalovirus particles contain a phosphoprotein of 150,000 (pp150) apparent molecular weight in their matrix; the protein appears particularly reactive in Western blot analyses with human antisera. The gene for pp150 was mapped by screening a bacteriophage lambda gt11 cDNA expression library with monospecific rabbit antisera. Subsequent hybridization of cDNA with cosmid and plasmid clones containing the human cytomegalovirus strain AD169 genome mapped the gene to *Hin*dIII fragments J and N. The gene is transcribed into a late 6.2-kilobase RNA. The nucleotide sequence of this region was determined, and a transcription initiation site and two polyadenylation sites of an abundant transcript were located by primer extension and nuclease protection experiments. The reading frame for pp150, deduced from computer analyses, gives rise to a polypeptide of 1,048 amino acids in length; protein secondary structure analysis revealed multiple β-pleated sheets in hydrophilic clusters, providing a possible explanation for the immunogenic properties of the polypeptide.

Human cytomegalovirus (HCMV), a ubiquitous pathogenic herpesvirus of considerable clinical importance, has at least 25 structural proteins in the molecular weight range of 200,000 (200K protein) to 18,000. Several of these proteins are modified by glycosylation or phosphorylation (4, 11, 15, 30, 32, 40). Though identification of genes for structural proteins will be required to develop novel diagnostics and vaccines, very few of these genes have been mapped to date. The HCMV gene coding for an abundant 65K virion polypeptide has been identified by hybrid-selected in vitro translation (30, 34), and a gene for a polypeptide similar in size has been found by hybridization of synthetic oligonucleotides derived from the known polypeptide sequence (31). A second matrix protein (71K phosphoprotein) was mapped by hybrid selection combined with immunoprecipitation of in vitro translation products, using monoclonal antibodies (30, 34). Eucaryotic expression cloning of viral DNA sequences under the control of the simian virus 40 early promoter in COS-I cells and detection of gene products with monoclonal antibodies allowed mapping of the gene of a major viral tegument protein (Towne strain) of about 67K (8, 9). The coding sequence for a virion envelope glycoprotein (gp58) of HCMV was determined with a bacteriophage lambda expression system (24). This procedure has recently been successfully used to map the gene for an HCMV-DNA-binding protein (28).

Like other herpesviruses, purified HCMV contains two prominent large proteins with apparent molecular weights of about 150,000. One of these is assumed to be the major nucleocapsid protein (15). The other polypeptide is phosphorylated and was designated as basic phosphoprotein or pp150, and it probably represents one of the matrix components (15, 33). Western blot analyses with human sera indicated that the 150K phosphoprotein is highly immunogenic, apparently more so than any other of the HCMV structural proteins (22; G. Jahn, B. C. Scholl, B. Traupe, and B. Fleckenstein, unpublished data). This suggests that pp150 is a primary candidate for developing diagnostic reagents by expression cloning. This study describes physical mapping, nucleotide sequence, and preliminary transcription analyses of the coding region for pp150.

MATERIALS AND METHODS

Virus, cell culture, and virion purification procedures. HCMV AD169 was provided by U. Krech, St. Gallen, Switzerland. Propagation of virus in human foreskin fibroblasts followed standard methods. The virus was purified by centrifugation through a glycerol-tartrate gradient (41).

Protein gel electrophoresis and Western blot analysis. The proteins of purified HCMV particles were fractionated in sodium dodecyl sulfate-containing 8.5 and 10% (wt/vol) polyacrylamide gels essentially as described by Laemmli (21). For some experiments, the gels were cross-linked with a higher amount of methylene bisacrylamide/acrylamide (1:28) to resolve the large proteins of around 150K (19). Gels were stained with Coomassie brilliant blue or stained with silver nitrate (26). Standard proteins of known molecular weight (Sigma Chemical Co., St. Louis, Mo.) and *Escherichia coli* RNA polymerase (Boehringer GmbH, Mannheim, Federal Republic of Germany) were run on the same gel. For Western blot analyses, proteins were electrophoretically transferred onto nitrocellulose (42) and blocked with NET buffer (0.15 M NaCl, 5 mM EDTA, 50 mM Tris hydrochloride, pH 7.4, 0.25% gelatin, 0.05% Nonidet P-40, 2% bovine serum albumin). After reaction with the antibody, the next incubation was performed with horseradish peroxidase-

* Corresponding author.

† Present address: Lady Davis Institute for Medical Research, Montreal, Quebec, Canada.

conjugated protein A (Sigma), and the staining was done with 4-chloro-1-naphthol and H_2O_2.

Preparation of antisera. Antiserum to the 150K protein was raised in rabbits. Extracts of virion proteins (about 1.5 mg) were fractionated on preparative 8.5% (wt/vol) polyacrylamide gels; proteins were visualized by Coomassie blue staining. The 150K gel slice was excised and homogenized. Proteins were eluted with 0.1 M $(NH_4)HCO_3$, pH 9.5, containing 0.1% sodium dodecyl sulfate and inoculated intracutaneously and subcutaneously together with Freund complete adjuvant. Booster injections were administered intramuscularly with Freund incomplete adjuvant in 4-week intervals over a period of 7 months. Antibody titers of sera, obtained before immunization of the rabbits and of immune sera, were tested by enzyme-linked immunosorbent assay and immunoblotting.

Construction and screening of the cDNA library. Double-stranded cDNA was synthesized from RNA isolated at 96 to 120 h after infection, as described by Gubler and Hoffman (16). The cDNA was methylated with *Eco*RI methylase and *Eco*RI linkers were added. The cDNA was inserted between *Eco*RI-cleaved dephosphorylated lambda gt11 arms (46) without size fractionation. The DNA was packaged in vitro, and *E. coli* strain Y1090 cells were infected (46). A library of 5×10^5 independent recombinants, containing about 20% wild-type phages, was obtained from approximately 10 ng of cDNA. Screening of the library with antibody probes was carried out as described previously (46), except that horseradish peroxidase coupled to protein A and 4-chloro-1-naphthol were used as the detection system (24). Plaques containing immunoreactive phages were picked and purified by two or three subsequent screening steps. DNA from recombinant phages was prepared from a plate lysate according to published procedures (36). Purified recombinant DNA was digested with *Eco*RI and subcloned directly into bacteriophage M13 vectors (27).

Induction of fusion proteins. Fusion proteins were synthesized in *E. coli* strain Y1089. An overnight culture of Y1089 was infected with recombinant phages at a multiplicity of infection of 2 to 3 per cell for 20 min at room temperature. After dilution in 3 ml of LB medium, cells were grown at 32°C to a density of 0.2 A_{600} unit. The cultures were induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) and incubated at 42°C for 15 min. After an additional 2 h at 37°C, the cells were collected in 100 μl of polyacrylamide gel electrophoresis (PAGE) buffer.

RNA extraction. RNA was isolated from infected cells as described before (6). Cells were washed free of media with phosphate-buffered saline, collected by low-speed contrifugation for 5 min, and washed once more with ice-cold phosphate-buffered saline. About 10^9 cells were lysed in 6 ml of 5 M guanidium isothiocyanate containing 50 mM lithium citrate, 0.1 M β-mercaptoethanol, and 0.1% lithium lauroyl sulfate. The lysate was centrifuged through a 4.5-ml cushion of 5.7 M CsCl containing 0.1 M EDTA (pH 7.0) in a Spinco SW41 rotor for approximately 20 h at 28,000 rpm and 20°C. The sediment was air dried, suspended in double-distilled water, and precipitated with ethanol.

Hybridization analyses. Purifications of recombinant plasmid DNA and Southern blot hybridizations were done by standard procedures (25). Cosmid cloned HCMV DNA, strain AD169 (14), was subcloned in vectors pACYC184 (5) and pUC8 (43); and cDNA clone BB-8 was subcloned in vector M13mp19 (45).

Transcript analyses by primer extension and nuclease protection. The probes used for RNA analyses were made by the "prime cut" method as previously described (12). The primer extension and S1 nuclease digestion were carried out according to Biggin et al. (2), and the products were analyzed on 6% polyacrylamide–8 M urea sequencing gels.

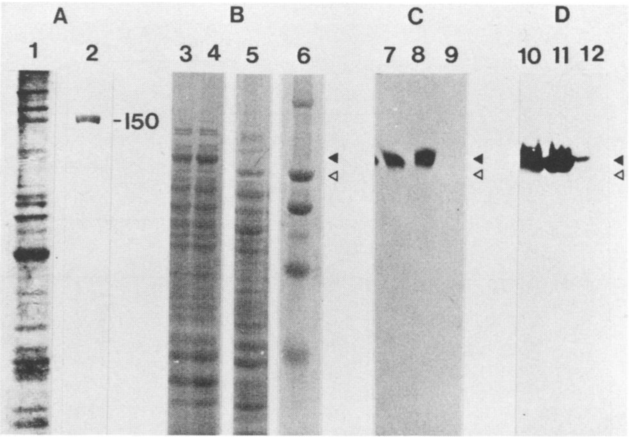
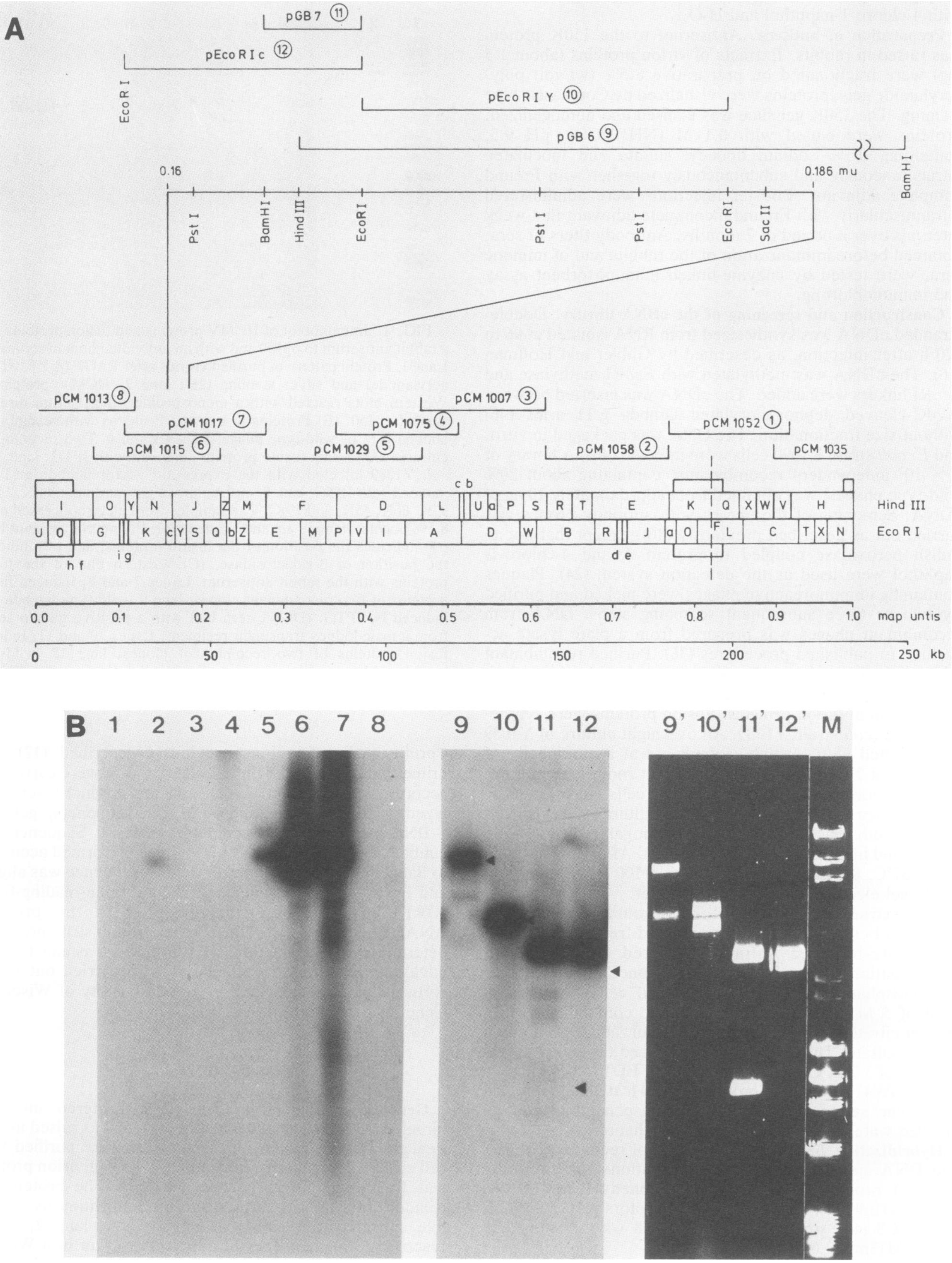


FIG. 1. Immunoblot of HCMV proteins and fusion proteins with a rabbit antiserum to pp150 and with an individual human serum. (A) Lane 1, Protein pattern of purified virions after PAGE (8.5%, wt/vol, acrylamide) and silver staining (26); lane 2, HCMV proteins in Western blots reacted with a monospecific rabbit serum directed against pp150. (B) Proteins in induced lysogens with recombinant lambda gt11 or wild-type phage. Lanes 3 and 4, Two recombinant cultures expressing fusion proteins, induced with IPTG; lane 5, *E. coli* Y1089 infected with the expression vector lambda gt11 and induced with IPTG; lane 6, molecular weight markers (200K, 116K, 97K, 66K, 45K, and 29K). The samples were electrophoresed on an 8.5% polyacrylamide gel and stained with Coomassie brilliant blue. (◀) indicates the position of the fusion proteins, and (◁) indicates the position of β-galactosidase. (C) Western blot of the fusion proteins with the rabbit antiserum. Lanes 7 and 8, Induced fusion proteins of two recombinant clones; lane 9, wild-type lambda gt11 induced by IPTG. (D) Western blot with a positive human serum from a male kidney transplant recipient. Lanes 10 and 11, Induced fusion proteins of two recombinant clones; lane 12, wild-type lambda gt11.

DNA sequencing and computer analyses. Sequencing of lambda gt11 cDNA and viral DNA was performed according to Sanger et al. (35). The HCMV DNA sequence was aligned and overlapped by computer (37), and open-reading-frame (ORF) analysis was carried out with the program ANALYSER (38). Two-dimensional protein structures were determined by the method of Chou and Fasman (7), and nucleotide sequence comparisons were carried out with a software package provided by the University of Wisconsin Genetics Computer Group (10).

RESULTS

Genomic location of the gene. A high-titered antiserum directed against the 150K phosphoprotein was raised in New Zealand White rabbits. Virus particles were purified from cell culture supernatants, and about 1.5 mg of virion proteins was fractionated by preparative PAGE. The protein was isolated from the gel and used for immunization by at least five subsequent inoculations. Figure 1A, lane 2, shows reactivity of a monospecific rabbit antiserum in a Western blot analysis with proteins of purified virions. The same serum was used to screen a cDNA library in the bacterio-


A
pGB 7 ⑪
pEco RI c ⑫
Eco R I
pEco RI Y ⑩
pGB 6 ⑨
Bam H I
0.16
0.186 mu
Pst I
BamHI
Hind III
EcoRI
Pst I
Pst I
Eco RI
Sac II
pCM 1013 ⑧
pCM 1017 ⑦
pCM 1015 ⑥
pCM 1029 ⑤
pCM 1075 ④
pCM 1007 ③
pCM 1058 ②
pCM 1052 ①
pCM 1035
c b
I O Y N J Z M F D L U a P S R T E K Q X W V H
Hind III
U O G K c Y S Q b Z E M P V I A D W B R J a O F L C T X N
Eco R I
h f
i g
d e
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 map untis
0 50 100 150 200 250 kb
B
1 2 3 4 5 6 7 8
9 10 11 12
9' 10' 11' 12' M

phage lambda gt11 expression vector; the cDNA had been synthesized from oligo(dT)-selected RNA that was extracted from HCMV-infected fibroblasts in the late phase of virus replication. Eight bacteriophage clones producing β-galactosidase–HCMV fusion proteins were identified by the initial plaque-staining immunoblot procedure. Protein extracts from individual clones were analyzed by PAGE. Fusion proteins larger than the 118K β-galactosidase could be detected in each extract (Fig. 1B, lanes 3 and 4) and were recognized by rabbit antiserum against pp150 and by human anti-HCMV sera (Fig. 1C and D). Antibodies were subsequently raised against gel-purified fusion protein; they reacted with pp150 exclusively when tested in Western blots with proteins of purified virions (data not shown).

DNA was purified from one of the eight identified bacteriophage clones that produced fusion proteins and was designated lambda BB-8. The inserted DNA was subcloned in M13mp19 (resulting in clone M13-BB8), nick repair labeled with [^{32}P]ATP, and hybridized in Southern blots with *Hin*dIII-digested cosmid clones containing the genome of HCMV strain AD169 in an overlapping fashion (14). Hybridization signals were seen with the *Hin*dIII-J fragment of the overlapping cosmid pCM1015 and pCM1017 (Fig. 2A and B, lanes 6 and 7). Hybridizations with a series of subclones located the sequence homologous to cDNA clone lambda BB-8 within the *Eco*RI-Y fragment (Fig. 2).

Nucleotide sequence. The nucleotide sequence of a 6.36-kilobase (kb) segment (map units 0.160 to 0.186) encompassing the *Eco*RI-Y fragment and the adjacent *Eco*RI-c fragment was determined by the dideoxy chain termination method (35). Figures 3A and B show the entire DNA sequence; the overlap between *Hin*dIII fragments N and J has not been sequenced separately. Analyses of this genomic sequence revealed two large ORFs running in opposite directions (Fig. 4). The longer frame spans 3,144 nucleotides (between nucleotides 524 and 3668 in Fig. 3 and 4), coding for a protein of 112,700 molecular weight. The smaller frame in the opposite orientation was found between nucleotides 5765 and 3751. Nucleotide sequences of the viral 6.36-kb genomic DNA segment and the cDNA clone lambda BB-8 were aligned by computer methods. The region of full homology within the longer ORF is underlined in Fig. 3 and marked as a filled box in Fig. 4. This indicated that the virion phosphoprotein with an apparent molecular weight of 150,000 is encoded by the longer of the two ORFs between nucleotides 524 and 3668 (Fig. 3 and 4).

Protein structure. The DNA sequence encompassing the two opposite ORFs was investigated by the computer program TESTCODE (13). It confirmed that both ORFs correspond to codon usage of eucaryotic genes (data not shown). A remarkable accumulation of hydrophilic regions appeared to be a salient feature of pp150 (Fig. 5a and b). Most hydrophilic clusters were located in the polypeptide sequence between amino acids 370 and 760. Computer analyses by Chou and Fasman (7) indicated that several of those hydrophilic clusters coincide with beta-pleated sheets in the protein (Fig. 5a and b). Antigenic sites are often located within hydrophilic β turns (18). Based on this assumption, strong antigenic sites could be expected around amino acids 430, 530, and 710. First experiments with expressed fusion proteins confirmed this assumption (B.-C. Scholl and G. Jahn, unpublished data).

Transcription signals. Northern blot analyses were performed with total late RNA from infected fibroblasts, using ^{32}P-labeled DNA of the cDNA clone M13-BB8 and various plasmid subclones of cosmid pCM1015 (Fig. 2A). The cDNA hybridized clearly with an RNA size class of about 6.2 kb (Fig. 6, lanes 1 and 2). Overlapping and flanking plasmid clones detected an abundant RNA of identical size (Fig. 6, lane 4); in addition, weaker smaller bands of RNAs of various sizes were seen in the Northern blots (Fig. 6). To confirm that the transcription and termination signals found in the nucleotide sequence were functional late in virus replication, the precise 5′ and 3′ ends of a major transcript were mapped by nuclease protection and primer extension analyses, using late RNA.

Probe J811/*Sac*II (positions 474 to 511; total length, 85 bases including M13 sequences) produced a product of 132 bases when extended with reverse transcriptase (Fig. 7a), indicating that the 5′ terminus of the RNA from the gene is at position 424 (Fig. 3A). Thirty nucleotides upstream of the RNA start site a potential TATA box sequence, TATTAAA, is located (Fig. 4A, positions 391 to 397). S1 nuclease analyses using probe N314/*Eco*RI (positions 6001 to 6347) (Fig. 3B) produced a clearly protected fragment of 260 bases (Fig. 7b, lane L). This indicated that the 3′ end of an abundant late RNA corresponds to nucleotide 6261 (Fig. 3B), which falls 20 bases downstream of the poly(A) signal sequence AATAAA (Fig. 3B). By using the same probe, a much weaker signal was also found, corresponding to a putative minor poly(A) addition site at map position 6310 (Fig. 3B). This site is 24 bases downstream of the sequence ATTAAA, which also has been shown to be functional as a poly(A) signal (1).

FIG. 2. Structure of the HCMV AD169 genome and map position of the pp150 gene. (A) Schematic representation of the HCMV genome with restriction maps for *Hin*dIII and *Eco*RI. The orientation of the long unique region of the HCMV genome is reversed in this graph relative to earlier publications (14, 20, 30). The nomenclature of cosmid clones pCM1007, pCM1013, pCM1015, pCM1017, pCM1029, pCM1035, pCM1052, pCM1058, and pCM1075 covering the entire HCMV AD-169 genome is taken from the original report (14). Brackets at the top indicate cloned DNA fragments that were used in the Southern blot hybridization experiments shown in (B). The DNA segment of 6.36 kb between map units 0.160 and 0.186 of HCMV AD169 is shown in expanded scale, including relevant restriction endonuclease cleavage sites that were used in following experiments. The 6.36-kb DNA segment corresponds to the nucleotide sequence shown in Fig. 3. (B) Physical mapping of the pp150-encoding DNA sequence by Southern blot hybridizations. The cDNA clone M13-BB8, derived from the bacteriophage lambda expression clone lambda BB-8 (Fig. 1, lanes 3, 7, and 10), was labeled with ^{32}P and hybridized with cosmids and plasmid clones. Lane 1, pCM1052; lane 2, pCM1058; lane 3, pCM1007; lane 4, pCM1075; lane 5, pCM1029; lane 6, pCM1015; lane 7, pCM1017; lane 8, pCM1013: all cleaved with *Hin*dIII; lane 9, *Hin*dIII- and *Bam*HI-digested DNA of clone pGB6 in vector pACYC184; lane 10, cloned viral *Eco*RI-Y fragment and vector pACYC184; lane 11, *Eco*RI- and *Bam*HI-digested DNA of clone pGB7 in vector pUC8; lane 12, cloned *Eco*RI-c fragment and vector pUC8. Lanes 9′ to 12′ show the ethidium bromide-stained gel with the restricted clones pGB6, pEcoRI-Y, pGB7, and pEcoRI-c, respectively. Lanes 9 and 10 indicate homology between the cDNA clone M13-BB8 and the inserted virion DNA in clones pGB6 and pEcoRI-Y. No hybridization is found with the viral DNA fragment of clones pEcoRI-c and pGB7. The hybridization signals in lanes 11 and 12 are due to sequence homologies of labeled M13mp19 vector (in M13-BB8) and pUC8 vector of clones pEcoRI-c and pGB7. The positions of all viral fragments (9 to 12) are indicated by triangles. M, Marker (Sigma).

```
A TAGATCACCGATAGAAATTTACACGAGGCCACGCCGGCCGGCAACAGCCACTGGTTGCTGAGTACGATAAAGGGTAGCACAGTAAGCGTGAGAAAATTAGTAGAGTAGAGGTTGGTCATG   120

  TAAATGGTGGGCGTCGAATAGCCAAGCACGCGATTCGTGAGCAGCTGCGTGATCAACACTATGGCGTTAAGTGGACCGCCCACGAAGATGATGAATGTGTTGAGTACGGCTTCGGTGGTT   240

  CGAATGGCGAATAGCGGCCCTGTCATGTTGCAAGTGTCATTGATGTGCGGAGGAGTGTTGTTGCGGGTCTGGGCGGAACAGCACACGGGGCGAAAAAACAGAAGAAACAAGTCAGCGGCG   360

                                >>>>>>>                      --------->
  CTTAAAAGAAAACCGCGTATCCGCCTCCGCTATTAAACTACCCCCCCTCCCTCTAGGTGGGGCGCTCACCGAGTTGTGGATGATGGTGTCCATCGTGGGCGAATAGCAGACCGCGGGCGC   480
                                                                                                                 Sac II
                                              M  S  L  Q  F  I  G  L  Q  R  R  D  V  V  A  L  V  N  F  L  R  H  L  T  Q  K
  AGTCCGGGGCGACGACGCTTCCGGGTTCTGGAGAAAAGCCAGCATGAGTTTGCAGTTTATCGGTCTACAGCGGCGCGATGTGGTAGCCCTGGTCAACTTTCTGCGCCATCTCACGCAAAA   600

   P  D  V  D  L  E  A  H  P  K  I  L  K  K  C  G  E  K  R  L  H  R  R  T  V  L  F  N  E  L  M  L  W  L  G  Y  Y  R  E  L
  GCCCGACGTGGATCTCGAGGCACACCCCAAGATCCTGAAAAAATGTGGCGAAAAACGCCTGCACCGGCGTACGGTGCTGTTCAACGAGCTCATGCTTTGGTTGGGATACTACCGCGAGCT   720

   R  F  H  N  P  D  L  S  S  V  L  E  E  F  E  V  R  C  V  A  V  A  R  R  G  Y  T  Y  P  F  G  D  R  G  K  A  R  D  H  L
  GCGTTTTCACAACCCCGACCTCTCCTCAGTGCTCGAGGAGTTCGAGGTGCGTTGCGTGGCCGTGGCGCGTCGCGGCTACACTTACCCGTTCGGTGATCGTGGTAAGGCGCGTGACCACCT   840

   A  V  L  D  R  T  E  F  D  T  D  V  R  H  D  A  E  I  V  E  R  A  L  V  S  A  V  I  L  A  K  M  S  V  R  E  T  L  V  T
  GGCTGTGCTAGACCGTACCGAATTCGATACGGACGTGCGCCACGATGCCGAGATCGTGGAACGCGCGCTCGTAAGCGCGGTCATTCTGGCCAAGATGTCGGTGCGCGAGACGCTGGTCAC   960
                      Eco RI
   A  I  G  Q  T  E  P  I  A  F  V  H  L  K  D  T  E  V  Q  R  I  E  E  N  L  E  G  V  R  R  N  M  F  C  V  K  P  L  D  L
  AGCCATCGGCCAGACGGAACCCATCGCCTTTGTGCACCTCAAGGATACGGAGGTGCAGCGCATTGAAGAAAACCTGGAGGGTGTGCGCCGTAACATGTTCTGCGTGAAACCGCTCGACCT   1080

   N  L  D  R  H  A  N  T  A  L  V  N  A  V  N  K  L  V  Y  T  G  R  L  I  M  N  V  R  R  S  W  E  E  L  E  R  K  C  L  A
  TAACCTGGACCGGCACGCCAACACGGCGCTGGTCAACGCCGTCAACAAGCTCGTGTACACGGGCCGTCTCATCATGAACGTGCGCAGGTCTTGGGAGGAGCTGGAGCGCAAATGTCTGGC   1200

   R  I  Q  E  R  C  K  L  L  V  K  E  L  R  M  C  L  S  F  D  S  N  Y  C  R  N  I  L  K  H  A  V  E  N  G  D  S  A  D  T
  GCGCATTCAGGAGCGCTGCAAGCTGCTGGTCAAGGAGCTGCGCATGTGCCTTTCCTTTGATTCCAACTACTGTCGCAATATCCTCAAGCACGCCGTGGAAAACGGCGACTCGGCCGACAC   1320

   L  L  E  L  L  I  E  D  F  D  I  Y  V  D  S  F  P  Q  S  A  H  T  F  L  G  A  R  S  P  S  L  E  F  D  D  D  A  N  L  L
  GCTGTTGGAGCTGCTCATCGAGGACTTTGATATCTACGTGGACAGCTTCCCACAGTCGGCGCACACGTTTTTGGGCGCGCGCTCGCCGTCGTTGGAGTTTGACGATGACGCCAATCTCCT   1440

   S  L  G  G  G  S  A  F  S  S  V  P  K  K  H  V  P  T  Q  P  L  D  G  W  S  W  I  A  S  P  W  K  G  H  K  P  F  R  F  E
  CTCGCTCGGCGGCGGTTCGGCCTTCTCGTCGGTACCCAAGAAACATGTCCCCACGCAGCCGCTGGACGGCTGGAGCTGGATCGCCAGTCCCTGGAAGGGACACAAACCGTTCCGCTTCGA   1560

   A  H  G  S  L  A  P  A  A  E  A  H  A  A  R  S  A  A  V  G  Y  Y  D  E  E  E  K  R  R  E  R  Q  K  R  V  D  D  E  V  V
  GGCCCATGGTTCTCTGGCACCGGCCGCCGAAGCCCACGCTGCCCGTTCGGCGGCCGTCGGCTATTACGACGAAGAGGAAAAGCGTCGCGAGCGGCAGAAACGGGTGGACGACGAGGTGGT   1680

   Q  R  E  K  Q  Q  L  K  A  W  E  E  R  Q  Q  N  L  Q  Q  R  Q  Q  Q  P  P  P  P  A  R  K  P  S  A  S  R  R  L  F  G  S
  GCAGCGTGAGAAACAGCAGCTGAAGGCTTGGGAGGAGAGGCAGCAGAACCTGCAGCAACGTCAGCAGCAACCACCGCCCCCGGCACGTAAACCGAGCGCCTCCCGGAGGCTCTTTGGCTC   1800
                                                    Pst I
   S  A  D  E  D  D  D  D  D  D  D  E  K  N  I  F  T  P  I  K  K  P  G  T  S  G  K  G  A  A  S  G  G  G  V  S  S  I  F  S
  CAGTGCCGATGAGGACGACGACGATGATGATGACGAGAAAAACATCTTTACGCCCATCAAGAAACCGGGAACTAGCGGCAAGGGCGCCGCTAGTGGTGGCGGTGTTTCCAGCATTTTCAG   1920

   G  L  L  S  S  G  S  Q  K  P  T  S  G  P  L  N  I  P  Q  Q  Q  Q  R  H  A  A  F  S  L  V  S  P  Q  V  T  K  A  S  P  G
  CGGCCTGTTATCCTCGGGCAGTCAGAAACCGACCAGCGGTCCCTTGAACATCCCGCAACAACAACAGCGTCACGCGGCTTTCAGTCTCGTCTCCCCGCAGGTGACCAAGGCCAGCCCGGG   2040

   R  V  R  R  D  S  A  W  D  V  R  P  L  T  E  T  R  G  D  L  F  S  G  D  E  D  S  D  S  S  D  G  Y  P  P  N  R  Q  D  P
  AAGGGTCCGTCGGGACAGCGCGTGGGACGTGAGGCCGCTCACGGAGACCAGAGGGGATCTTTTCTCGGGCGACGAGGATTCCGACAGCTCGGATGGCTATCCCCCCAACCGTCAAGATCC   2160

   R  F  T  D  T  L  V  D  I  T  D  T  E  T  S  A  K  P  P  V  T  T  A  Y  K  F  E  Q  P  T  L  T  F  G  A  G  V  N  V  P
  GCGTTTCACCGACACGCTGGTGGACATCACGGATACCGAGACGAGCGCCAAACCGCCCGTCACCACCGCGTACAAGTTCGAGCAACCGACGTTGACGTTCGGCGCCGGAGTTAACGTTCC   2280

   A  G  A  G  A  A  I  L  T  P  T  P  V  N  P  S  T  A  P  A  P  A  P  T  P  T  F  A  G  T  Q  T  P  V  N  G  N  S  P  W
  TGCTGGCGCCGGCGCTGCCATCCTCACGCCGACGCCTGTCAATCCTTCCACGGCCCCCGCTCCGGCCCCGACACCTACCTTCGCGGGTACCCAAACCCCGGTCAACGGTAACTCGCCCTG   2400

   A  P  T  A  P  L  P  G  D  M  N  P  A  N  W  P  R  E  R  A  W  A  L  K  N  P  H  L  A  Y  N  P  F  R  M  P  T  T  S  T
  GGCTCCGACGGCGCCGTTGCCCGGGGATATGAACCCCGCCAACTGGCCGCGCGAACGCGCGTGGGCCCTCAAGAATCCTCACCTGGCTTACAATCCCTTCAGGATGCCTACGACTTCCAC   2520

   A  S  Q  N  T  V  S  T  T  P  R  R  P  S  T  P  R  A  A  V  T  Q  T  A  S  R  D  A  A  D  E  V  W  A  L  R  D  Q  T  A
  GGCTTCTCAAAACACCGTGTCCACCACCCCTCGGAGGCCGTCGACTCCACGCGCCGCGGTGACACAAACAGCGTCTCGGGACGCCGCTGATGAGGTTTGGGCTTTAAGGGACCAAACTGC   2640
                                                                                                                       Pst I
   E  S  P  V  E  D  S  E  E  E  D  D  D  S  S  D  T  G  S  V  V  S  L  G  H  T  T  P  S  S  D  Y  N  N  D  V  I  S  P  P
  AGAGTCACCGGTCGAAGACAGCGAGGAGGAAGACGACGACTCCTCGGACACCGGCTCCGTCGTCAGCCTGGGACACACAACACCGTCGTCCGATTACAACAACGACGTCATTTCGCCTCC   2760

   S  Q  T  P  E  Q  S  T  P  S  R  I  R  K  A  K  L  S  S  P  M  T  T  T  S  T  S  Q  K  P  V  L  G  K  R  V  A  T  P  H
  CAGTCAGACGCCCGAGCAGTCGACGCCGTCCAGAATACGTAAAGCTAAGTTATCGTCTCCAATGACGACGACATCCACGAGCCAGAAACCGGTGCTGGGCAAGCGAGTCGCGACGCCGCA   2880

   A  S  A  R  A  Q  T  V  T  S  T  P  V  Q  G  R  L  E  K  Q  V  S  G  T  P  S  T  V  P  A  T  L  L  Q  P  Q  P  A  S  S
  CGCGTCCGCCCGAGCGCAGACGGTGACGTCGACGCCGGTTCAGGGAAGGCTAGAGAAACAGGTGTCGGGCACGCCGTCGACGGTACCCGCCACGCTGTTGCAACCTCAACCGGCTTCGTC   3000

   K  T  T  S  S  R  N  V  T  S  G  A  G  T  S  S  A  S  S  A  R  Q  P  S  A  S  A  S  V  L  S  P  T  E  D  D  V  V  S  P
  TAAAACGACGTCATCAAGGAACGTGACTTCTGGCGCGGGAACCTCTTCCGCTTCTTCGGCTCGACAGCCGTCAGCCTCGGCGTCCGTTTTGTCGCCCACGGAGGATGATGTCGTGTCCCC   3120

   A  T  S  P  L  S  M  L  S  S  A  S  P  S  P  A  K  S  A  P  P  S  P  V  K  G  R  G  S  R  V  G  V  P  S  L  K  P  T  L
  CGCCACATCGCCGCTGTCCATGCTTTCGTCAGCCTCTCCGTCCCCGGCCAAGAGTGCCCCCCCGTCTCCGGTGAAAGGCCGGGGCAGCCGCGTCGGTGTTCCTTCCTTGAAACCTACTTT   3240

   G  G  K  A  V  V  G  R  P  P  S  V  P  V  S  G  S  A  P  G  R  L  S  G  S  S  R  A  A  S  T  T  P  T  Y  P  A  V  T  T
  GGGCGGCAAGGCGGTGGTAGGTCGACCGCCCTCGGTCCCCGTGAGCGGTAGCGCGCCCGGGTCGCCTGTCCGGCAGCAGCCGGGCCGCCTCGACCACGCCGACGTATCCCGCGGTAACCAC   3360
```

FIG. 3. Nucleotide sequence of the 6,360-base pair DNA segment between map units 0.160 and 0.186. The cDNA clone M13-BB8 corresponds to the virion DNA sequence between nucleotides 3209 and 3389 (underlined); the clone was derived from the expression clone lambda BB-8 (see Fig. 1). A typical TATA consensus sequence is marked >>>>; the transcription initiation site identified by primer extension (see Fig. 7) is indicated by – – →. The single poly(A) signal sequence AATAAA and the possibly alternative poly(A) signal ATTAAA are marked by ****. The 3′ ends of the transcript determined by S1 nuclease protection (see Fig. 7) are indicated by – – ⅃ . The amino acid sequence of the single large ORF (see Fig. 4) is given above the nucleotides in a one-letter code. (A) The sequence between nucleotides 1 and 3360 and (B) the sequence between nucleotides 3361 and 6360.

```
    V  Y  P  P  S  S  T  A  K  S  S  V  S  N  A  P  P  V  A  S  P  S  I  L  K  P  G  A  S  A  A  L  Q  S  R  R  S  T  G  T
B CGTTTACCCACCGTCGTCTACGGCCAAAAGCAGCGTATCGAATGCGCCGCCTGTGGCCTCCCCCTCCATCCTGAAACCGGGGGCGAGCGCGGCTTTGCAATCACGCCGCTCGACGGGGAC   3480
    A  A  V  G  S  P  V  K  S  T  T  G  M  K  T  V  A  F  D  L  S  S  P  Q  K  S  G  T  G  P  Q  P  G  S  A  G  M  G  G  A
  CGCCGCCGTAGGTTCCCCCGTCAAGAGCACGACGGGCATGAAAACGGTGGCTTTCGACCTATCGTCGCCCCAGAAGAGCGGTACGGGGCCGCAACCGGGTTCTGCCGGCATGGGGGGCGC   3600
    K  T  P  S  D  A  V  Q  N  I  L  Q  K  I  E  K  I  K  N  T  E  E
  CAAAACGCCGTCGGACGCCGTGCAGAACATCCTCCAAAAGATCGAGAAGATTAAGAACACGGAGGAATAGTTAAGAAACACACACGCAGACGTACTTTTTAATGAAACCATCGGATAGTG   3720

  ACGTGTCGGGAAAGGAGGACGGACGGAGGGTCAGGGATGGGGAGACGTGAGAAAGTTGTCCGCGGGCAATTGCATGTCGCCCAGAAAGAACGTGGTTGTTCCGGCGGCGTGCATCTGCCG   3840

  AAACACCGTGTGGTGGTTGTACGAGTACACGTTACCGTCGCCCTCGGTAATTTGATACAACGTGGCGATGGGGGTGCCCTGCGGGATCACGATGGAACGCGTGCGCGTCCACAGCGTGAC   3960

  TTTGAGCGGCTCGCCGCCGCGCCACACGCTGAGCCCCGTGTAAAAGGCGTCCTCGTGTGGCAAGTTGGCCACCAAGAAACACCGGTCTGTGATCTGCACGTAGCGCAAGTCCAACTCCAC   4080

  CGTCTGCCGCGGTTGCACCCCGAAGTGGATATCGTAAGGCGCGTGCACCGTGAGCGAAAACACGTTGGGCTCATTGAGAAGCGGACAGTTGAGCGCGTCGCCGCTAAAAAAGAGTGACGG   4200

  GTTGCGGCTGAATCGCAGGTCGTACCCGCGCTGCGCGCTCGTCAGCAGGTAGAAGGAAAAAGCGCGCGGCATGTTGCGCGCCGTGATCTTGTCCGAGACGCGGTGACAGAAGGAGGTGGC   4320

  CACGGTGCCCAGCAGTTGGCGCTGTTCCGCGTCCACGCATAGTGAATCCACGTTGACGGTGAAAATGAGACCCATGAATTCGTACTGCACGTTTTTGGACGCGATCCACGCTTCGTCCTC   4440
                                                                             Eco RI
  GCCGGGTAGCGCTGCCTCGTCGTCGTCCATCGTGCCGCGGAACTGCGCGAGGTAGCGCGTAATTTTTTTGTGTCCGTACGTGGTTACGCGCTTACTGATCCAGGTCAGATGGTCCACGCG   4560

  ACATAGCAGCGTCGCGCCATGCCGCGTGACGCTGACCCGTCCAAAGGGCGCCGCCTCCTCCAACCCCGCAACGCCGCTCGGAGCACCGCCGCAGCCCGGCTTTCCCGGCGTCGTGAAAGG   4680

  CACGGCGTAATGCGGGCAGGCGTGCGGCACGAAGGGCACCATGACCAGTTGTGTGTGCAGAAAACCGATCTGCACCGCCTGCGACTGCCGCATGGTTTCCTCGTCGTAAACCGCCATGGA   4800

  CGAGCAGAGCCCGCCCTTGGTGATGAGCGGTTGCAGCACCACGGAGCTCTCGCTGGTGGAGCAGAGCAGAAAGAAGAGCTCGGCGTACGCCGCCTTGGGCGTCACCACGTTGGACCAGTC   4920

  GTACTTGTAGCCGCAGCCCTGCGTGTTGTTGTAAATGACGGGAAACGAGAGAAAGATGCAGCCCTGCACGTACGAAGCTTTCTCCGTCACGTTCGAGGCCGTGTTGTACTGCTCGGTGAT   5040
                                                                            Hind III
  GGACACCAAGTACGACTCGTAGGCCGTCAGGTGCGAGGCCGAACGGTGAATCTTGGCGTGGCGCACGCAGCGACCGTAGTTGTCGCGGTCCGCGTCGCGTAGCGCTTCGATCCACGAGGT   5160

  CACCACGTCCTGCGCCGGCAGACGATAGTCCTGCTCGGGGTCCATGTGGCGGCACAGCCGCAGGCGCTCTGCCAGTTGGCGAGGGATACCGTCGTGCGACCTTTTGACCGCGGTGGTGCC   5280

  TGTCGTCCTCGTCTCCCCTCCTTCGTTCTCCCTGTTTTCTCTTCTCTCATTCCCGGTCTCCGGATCCGCAGCCGCTACCTCTTGCTCCGCGGTTTTCTCGCCCACCTCGCTCGTCGCTGT   5400
                                                               Bam HI
  CGCCGCCACCGCAGCGGCGGCGACGGACGGCGGCGGTAACAACAGCTCCGTGAAGCTGACGAGCGGCAGCGGCGACGACGGTGGCGGCGACGACACGGCGACGGTCAACAGGGTCACAAG   5520

  CGTGGGTTTGTCCCCCATAATCTGGTCGCCGCCACCGCCGTCGTTGCCGGTCCCCGTTTCCTCCGGCGTCGCGGTTTCCGCCGTCTCCGGATGAGCGGCCGCGGCGCGGGCTCGGCGTCC   5640

  CGCCGTCCGAGACGGTGTATATAAACCGCGTCGGCCTCGCCGGCCCGAGCGCGCCGGGGAGAAGAACCTCTTCCCGGGCCCCGCGTTCAAGACGGCGTGCCGTGACGCTCGATGGGTCCG   5760

  CTTCATCAGACTGCGTACGCTTTGGAGCGTCAGACCCAGGGCGCATGTAGCCGACTTGGAGGACTTTGCCGCCTTTTATCGCACCCTCTCGGACAGTGAGCAGCAGGAGTTCGAGCAAGA   5880

  AGCCGAACTCGCCTCCCGCTCACAACGCGTGCAACACCTGCGCGAGGCCCGGCGCCAGCTCAAGATGGACCTGATGTGTCACGGCGGTTGAAAACGCGCATGATCTCGCGAAGCCATCTA   6000

  CGCGCCTGTCAGGGCGATGACGACATCAGCGATGACGGCTCCTGATACGCGCCGGCAGCTGCAGCACGTGGAGACGCTGCGTCGGTTTCTGCGCGGCGACAGCTGCTTTGTGCACGATCT   6120
                                                            Pst I
                                                                                                                     *****
  CCGGGGCATGATGGACTATCACGACGGGCTCTCGCGCCGTCAACAGCGTGCCTTTTGCCGCGCGAGTCGCGTGTTGACGGACCCGGAGCCCATCCAGAGCGAAGCGGAGGGGGAGAATAA   6240
  *              -----┘                   ******                  -----┘
  ACAGTTTACGGAGCACACACACAAAGTAGTCTCGTTTTTTATTAAAAGTGTCTTTGTATTTCCCTATCTTGTGTTGCCCAACTGCTGTCAGGTCTCCGTAGATCGCTCCCGGGTGCCCGA   6360
```

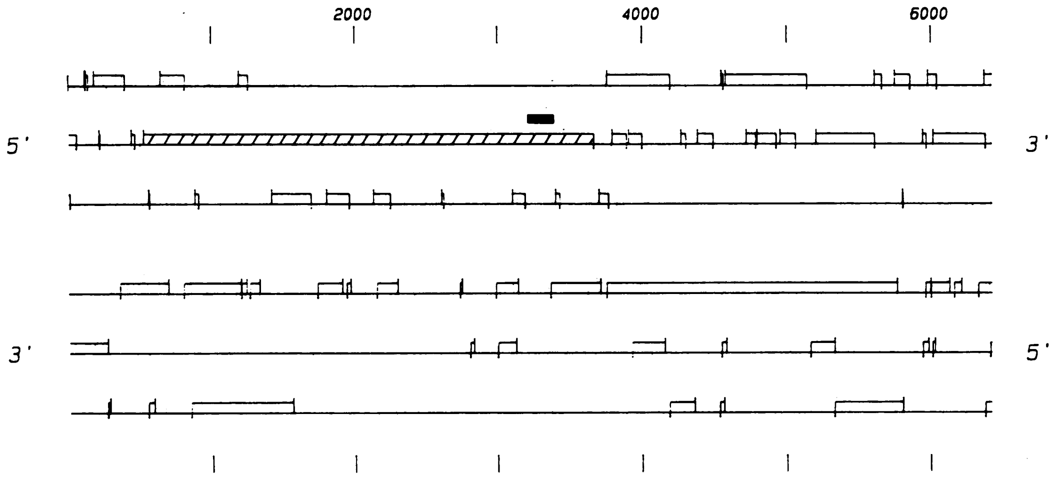


FIG. 4. Reading frame analysis of the DNA region between map positions 0.160 and 0.186. Bars above the horizontal lines indicate the start codons; bars below the lines indicate stop codons. Southern blot hybridizations and nucleotide sequence analyses located the cDNA clone M13-BB8 (filled box) within the long reading frame (hatched box) which extends between nucleotides 524 and 3668.

DISCUSSION

HCMV possesses at least 25 structural proteins. Western blot analyses using various individual patient sera and hyperimmune serum pools have shown that the most reactive polypeptide is a phosphoprotein (pp150) of 150,000 apparent molecular weight. The protein had been identified as a component of the virion matrix (15). This paper describes physical map position, nucleotide sequence, and preliminary transcription analyses of the gene encoding pp150. The gene-coding sequence was found by screening a lambda gt11 cDNA library with a monospecific rabbit antiserum.

The nucleotide sequence comparison between the genomic DNA and the DNA of the positive expression clone of the library (Fig. 1) showed that the cDNA clone lambda BB-8 represents an internal part of the long reading frame. This is most easily explained since a lambda gt11 clone from the long 3'-untranslated sequence could not express virus-specific fusion proteins detectable by the plaque-staining immunoblot procedure with the antibody (rabbit antiserum). That the cDNA did not contain the poly(A) site is not surprising since all generated douple-stranded sequences were cloned without attempt to enrich for full-length cDNA.

The coding sequence for pp150 is located in the long unique segment of the viral genome within *Hin*dIII fragments J and N between map positions 0.160 and 0.186 of HCMV strain Ad169. The gene is transcribed into an abundant late RNA which appears to consist of 6.2 kb. Primer extension and nuclease protection experiments identified a transcription initiation site which is preceded by a TATA consensus sequence. It might be possible that the minor RNA bands seen in Fig. 6 are 5' coterminal with the major RNA of 6.2 kb. A poly(A) site from the pp150 coding sequence was localized by S1 analyses. It is 20 nucleotides downstream of the single canonical poly(A) signal AATAAA that was seen within the entire sequenced region. The sequence has the typical structure of a poly(A) addition site (3). An alternative poly(A) signal, ATTAAA, was found 40 nucleotides downstream of the sequence AATAAA. S1 analysis revealed a second protected fragment of low intensity (Fig. 7b, lane L), suggesting a poly(A) site 24 nucleotides downstream of the ATTAAA motif. The possible second poly(A) signal, however, did not coincide with a known downstream addition sequence for faithful polyadenylation (3). The same signal appears to be occasionally used in another herpesvirus, Epstein-Barr virus (EBV) (1).

The transcribed DNA sequence reveals two large ORFs. The longer frame codes for a polypeptide of 1,048 amino acids (113K) and runs in the direction of the major late RNA from this region. This raises the question of whether this single long ORF can code for the entire mature pp150 in virions. The size discrepancy may be explained by the high content of proline residues (39), which are 10% of the total amino acids in the putative polypeptide. Also, the basic nature seen from computer analyses and the very high degree of phosphorylation (33) may contribute to an overestimation of its size in PAGE. In addition, electrophoresis with more suitable size markers, particularly *E. coli*-RNA polymerase (160K and 150K), led us to a reassessment of the apparent molecular weight in sodium dodecyl sulfate-PAGE to about 140,000. All of this could explain the error in molecular weight estimation for pp150. However, it is still possible that the mature mRNA combines the long open frame with smaller ORFs near the 3' end of the transcript (Fig. 4). Yet, a computer search for splice donor consensus sequences (29) did not recognize such a signal within the entire large ORF. The single poly(A) signal (AATAAA) was found 2.6 kb downstream of the large ORF. On the other side, there is a possible poly(A) signal, AATGAA, detectable close to the 3' end of the large ORF at nucleotides 3701 to 3706 (Fig. 3B). This motif had been shown by in vitro mutagenesis to be a very weak poly(A) cleavage signal (44). Future studies will discriminate among primary transcript, various possible RNA precursors, and the functional mRNA as a basis for comprehensive transcript mapping.

Phosphoproteins of about 150,000 apparent molecular weight have been shown in the matrix or internal envelope of other herpesviruses (23, 33). A computer search of available nucleotide and amino acid sequences did not show any significant homology within the entire sequence of EBV. EBV also contains a membrane protein similar in size which has been termed the 140K nonglycosylated membrane antigen and is thought to be encoded by BNRF1 of the EBV strain B95-8 genome (1). However, there is no recognizable

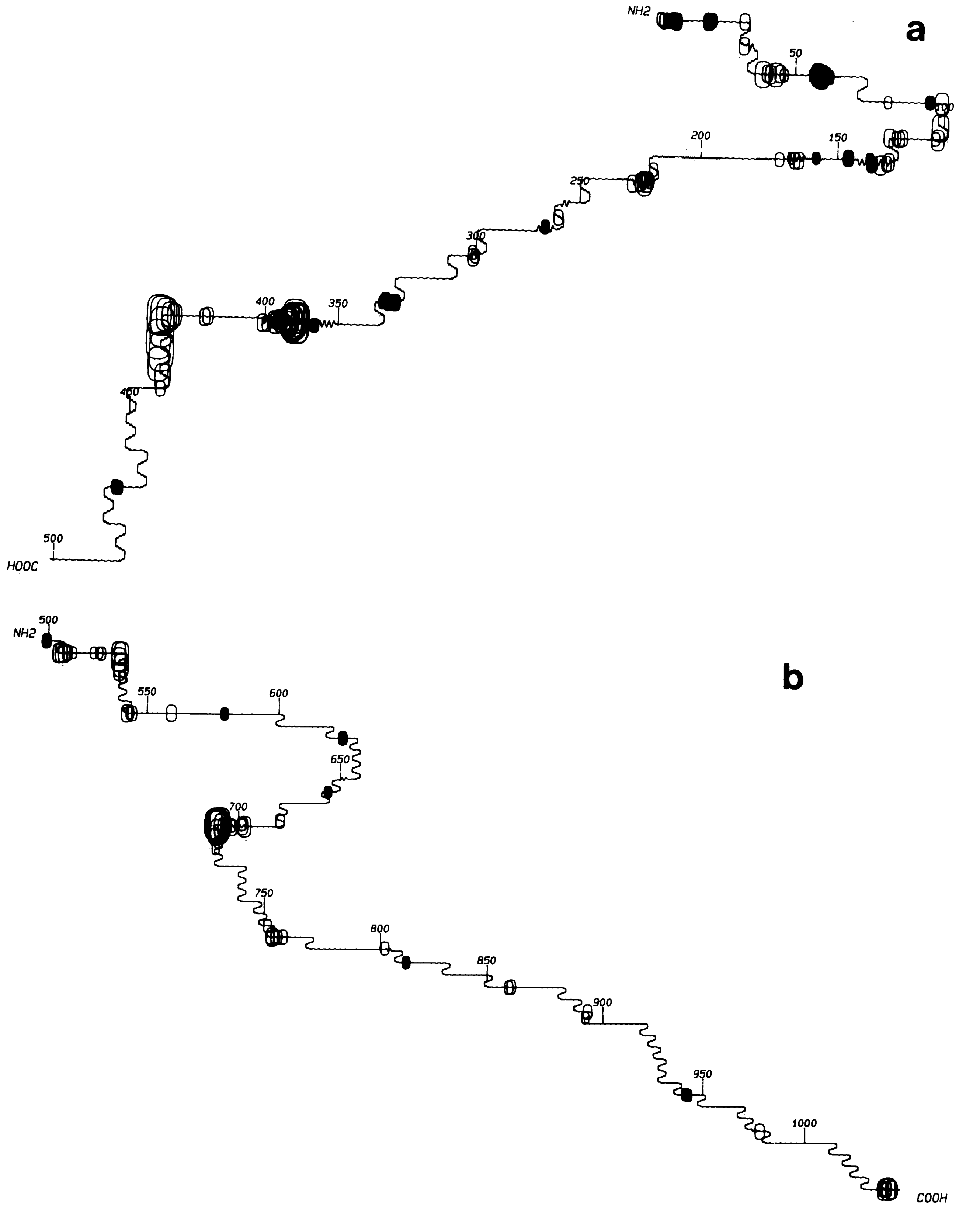


FIG. 5. Secondary structure of phosphoprotein pp150 obtained by computer analyses by Chou and Fasman (7). The amino-(NH_2) and carboxy-(COOH) termini are indicated. The hydrophilic areas are marked by open rectangles; the filled rectangles indicate hydrophobic regions. β-Turns are indicated by line turns of 180°. Hydrophilic β-turns are located between amino acids 420 to 440, 500 to 550, and 700 to 750. Amino acids were numbered every 50th residue. (a) The structure from the amino terminus to amino acid 500 and (b) the structure from amino acid 500 to the putative carboxy terminus.

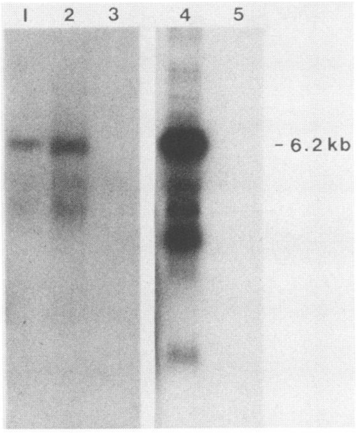


FIG. 6. Northern blot hybridizations of ^{32}P-labeled cDNA clone M13-BB8 and of a cloned virion DNA fragment with whole-cell RNA extracted from fibroblasts late after HCMV infection. RNA was purified 5 days postinfection, fractionated by electrophoresis on a 1% agarose gel, and transferred onto nitrocellulose filters. Lanes 1 and 2, Hybridization of cDNA clone M13-BB8 with late RNA; lane 4, hybridization of the *Hin*dIII/*Bam*HI clone pGB6 (see Fig. 2) with late RNA. Lanes 3 and 5, RNA from mock-infected cells. Molecular weight standards: *E. coli* rRNA (Boehringer) of 23S (3.3 kb) and 16S (1.7 kb) and human cell rRNA of 28S (5.2 kb) and 18S (2.0 kb).

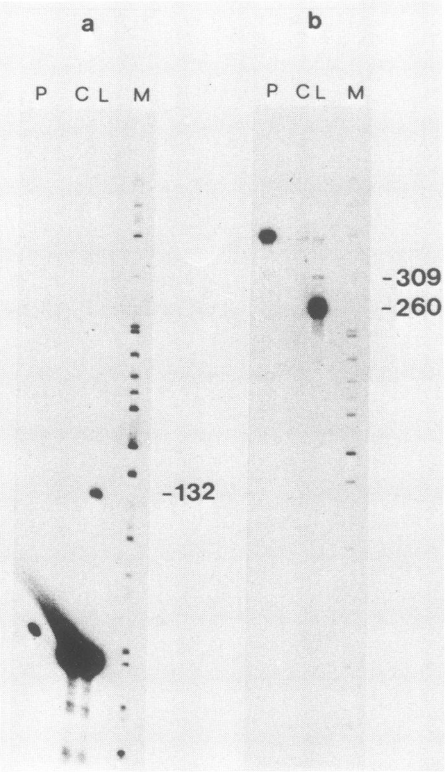


FIG. 7. Identification of the transcription initiation site by primer extension and determination of poly(A) sites with S1 nuclease analysis. The prime-cut probes used were as follows: (a) J811/*Sac*II, nucleotides 474 to 511; (b) N314/*Eco*RI, nucleotides 6001 to 6347. The nucleotide positions correspond to Fig. 3. The *Eco*RI cleavage site of N314 falls 50 nucleotides into the M13 sequence of that clone. In each experiment, the probe (P) was annealed to mock RNA (C) or late RNA from HCMV-infected cells (L) and was either extended with reverse transcriptase (a) or digested with S1 nuclease (b). *Hpa*II-digested pBR322 was used as the molecular weight standard (M). The numbers indicate the estimated size (in bases) of the S1 nuclease-protected fragments and the primer extension product. The primer extension product shown in (a) (132) locates the 5′ end of a major late HCMV RNA to nucleotide 427. The protected fragment of lane L in (b) (260) shows that a 3′ end of HCMV-coded late RNA corresponds primarily to nucleotide 6261 (see Fig. 3), 20 nucleotides downstream of the poly(A) signal AATAAA. A second protected fragment of low intensity is seen in (b) (309 base pairs), suggesting the motif ATTAAA to be used as a second, minor poly(A) signal.

sequence homology, and it is not located in any recognizable colinear fashion. Other HCMV genes, e.g., for the DNA polymerase (17) and for the major glycoprotein gene (24), do have, in contrast, significant sequence homologies with EBV DNA in colinear arrangement.

Computer analyses of polypeptide secondary structures in pp150 indicated multiple beta-pleated sheets in hydrophilic amino acid clusters, offering an explanation for the highly antigenic properties (18). Therefore, this phosphoprotein of HCMV should be used for scaled-up production by expression vector cloning to obtain larger amounts of the polypeptide and to improve serodiagnostic reactions.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft and, in part, the Bundesministerium für Forschung und Technologie.

We thank Bärbel Rüger and Jens Albrecht for help in analyzing cDNA clones and Thomas Iftner for help in computer studies.

LITERATURE CITED

1. **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. Séguin, P. S. Tuffnell, and B. G. Barrell.** 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature (London) **310**:207–211.
2. **Biggin, M., P. J. Farrell, and B. G. Barrell.** 1984. Transcription and DNA sequence of the BamHI L fragment of B95-8 Epstein-Barr virus. EMBO J. **3**:1083–1090.
3. **Birnstiel, M. L., M. Busslinger, and K. Strub.** 1985. Transcription, termination and 3′ processing: the end is in site. Cell **41**:349–359.
4. **Britt, W. J.** 1984. Nuetralizing antibodies detect a disulfide-linked glycoprotein complex within the envelope of human cytomegalovirus. Virology **135**:369–378.
5. **Chang, A. C. Y., and S. N. Cohen.** 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. **134**:1141–1156.
6. **Chirgwin, J. A., 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.
7. **Chou, P. Y., and G. D. Fasman.** 1974. Prediction of protein conformation. Biochemistry **13**:222–245.
8. **Davis, M. G., and E.-S. Huang.** 1985. Nucleotide sequence of a human cytomegalovirus DNA fragment encoding a 67-kilodalton phosphorylated viral protein. J. Virol. **56**:7–11.
9. **Davis, M. G., E.-C. Mar, Y.-M. Wu, and E.-S. Huang.** 1984. Mapping and expression of a human cytomegalovirus major viral protein. J. Virol. **52**:129–135.
10. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
11. **Farrar, G. H., and J. D. Oram.** 1984. Characterization of the human cytomegalovirus envelope glycoproteins. J. Gen. Virol. **65**:1991–2001.

12. **Farrell, P. J., P. L. Deininger, A. Bankier, and B. Barrell.** 1983. Homologous upstream sequences near Epstein-Barr virus promoters. Proc. Natl. Acad. Sci. USA **80:**1565–1569.
13. **Fickett, J. W.** 1982. Recognition of protein coding regions in DNA sequences. Nucleic Acids Res. **10:**5303–5318.
14. **Fleckenstein, B., I. Müller, and J. Collins.** 1982. Cloning of the complete human cytomegalovirus genome in cosmids. Gene **18:**39–46.
15. **Gibson, W.** 1983. Protein counterparts of human and simian cytomegaloviruses. Virology **128:**391–406.
16. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25:**263–269.
17. **Heilbronn, R., G. Jahn, A. Bürkle, U.-K. Freese, B. Fleckenstein, and H. zur Hausen.** 1987. Genomic localization, sequence analysis, and transcription of the putative human cytomegalovirus DNA polymerase gene. J. Virol. **61:**119–124.
18. **Hopp, T. P., and K. R. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78:**3823–3828.
19. **Irmiere, A., and W. Gibson.** 1983. Isolation and characterization of a noninfectious virion-like particle released from cells infected with human strains of cytomegalovirus. Virology **130:**118–133.
20. **Jahn, G., E. Knust, H. Schmolla, T. Sarre, J. A. Nelson, J. K. McDougall, and B. Fleckenstein.** 1984. Predominant immediate-early transcripts of human cytomegalovirus AD169. J. Virol. **49:**363–370.
21. **Laemmli, U. K.** 1970. Cleavage of structural protein during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
22. **Landini, M. P., M. C. Re, G. Mirolo, B. Baldassarri, and M. La Placa.** 1985. Human immune response to cytomegalovirus structural polypeptides studied by immunoblotting. J. Med. Virol. **17:**303–311.
23. **Lemaster, S., and B. Roizman.** 1980. Herpes simplex virus phosphoproteins. II. Characterization of the virion protein kinase and of the polypeptides phosphorylated in the virion. J. Virol. **35:**798–811.
24. **Mach, M., U. Utz, and B. Fleckenstein.** 1986. Mapping of the major glycoprotein gene of human cytomegalovirus. J. Gen. Virol. **67:**1461–1467.
25. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Merril, C. R., D. Goldman, S. A. Sedman, and M. H. Ebert.** 1981. Ultransensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. Science **211:**1437–1438.
27. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selection either DNA strand of double-digest restriction fragments. Gene **19:**269–276.
28. **Mocarski, E. S., L. Pereira, and N. Michael.** 1985. Preicse localization of genes on large animal virus genomes: use of gt11 and monoclonal antibodies to map the gene for a cytomegalovirus protein family. Proc. Natl. Acad. Sci. USA **82:**1266–1270.
29. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10:**459–472.
30. **Nowak, B., A. Gmeiner, P. Sarnow, A. J. Levine, and B. Fleckenstein.** 1984. Physical mapping of human cytomegalovirus genes: identification of DNA sequences coding for a virion phosphoprotein of 71 kDa and a viral 65-kDa polypeptide. Virology **134:**91–102.
31. **Pande, H., W. S. Baak, A. D. Riggs, B. R. Clark, J. E. Shively, and J. A. Zaia.** 1984. Cloning and physical mapping of a gene fragment coding for a 64-kilodalton major late antigen of human cytomegalovirus. Proc. Natl. Acad. Sci. USA **81:**4965–4969.
32. **Pereira, L., M. Hoffman, M. Tatsuno, and D. Dondero.** 1984. Polymorphism of human cytomegalovirus glycoproteins characterized by monoclonal antibodies. Virology **139:**73–86.
33. **Roby, C., and W. Gibson.** 1986. Characterization of phosphoproteins and protein kinase activity of virions, noninfectious enveloped particles, and dense bodies of human cytomegalovirus. J. Virol. **59:**714–727.
34. **Rüger, B., S. Klages, B. Walla, J. Albrecht, B. Fleckenstein, P. Tomlinson, and B. Barrell.** 1987. Primary structure and transcription of the genes coding for the two virion phosphoproteins pp65 and pp71 of human cytomegalovirus. J. Virol. **61:**446–453.
35. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
36. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
37. **Staden, R.** 1982. Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucleic Acids Res. **10:**4731–4751.
38. **Staden, R.** 1984. A computer program to enter DNA gel reading data into a computer. Nucleic Acids Res. **12:**499–503.
39. **Stenberg, R. M., D. R. Thomsen, and M. F. Stinski.** 1984. Structural analysis of the major immediate early gene of human cytomegalovirus. J. Virol. **49:**190–199.
40. **Stinski, M. F.** 1976. Human cytomegalovirus: glycoproteins associated with virions and dense bodies. J. Virol. **19:**594–609.
41. **Talbot, P., and J. D. Almeida.** 1977. Human cytomegalovirus: purification of enveloped virions and dense bodies. J. Gen. Virol. **36:**345–349.
42. **Towbin, H., T. Stachelin, 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.
43. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19:**259–268.
44. **Wickens, M., and P. Stephenson.** 1984. Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3′ end formation. Science **226:**1045–1051.
45. **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 **33:**103–119.
46. **Young, R. A., and R. W. Davis.** 1983. Efficient isolation of genes by using antibody probes. Proc. Natl. Acad. Sci. USA **80:**1194–1198.