# Identification of the Major Capsid Protein Gene of Human Cytomegalovirus

MARK CHEE,[1] STEFANIE-ANJA RUDOLPH,[2] BODO PLACHTER,[2] BART BARRELL,[1] AND GERHARD JAHN[2]*

*Institut für Klinische und Molekulare Virologie der Universität Erlangen-Nürnberg, Loschgestraße 7, D-8520 Erlangen, Federal Republic of Germany,[2] and Laboratory of Molecular Biology, Medical Research Council, Cambridge CB2 2QH, England[1]*

The coding region for the major capsid protein (MCP) of human cytomegalovirus (HCMV) was identified by comparing the protein sequence with the respective sequences of herpes simplex virus (HSV), Epstein-Barr virus, and varicella-zoster virus. The predicted length of the HCMV MCP was 1,370 amino acids. Comparison of the MCP sequences of the different human herpesviruses showed a homology of 25% to the MCP of HSV type 1, a homology of 29% to the MCP of Epstein-Barr virus, and a homology of 23% to the MCP of varicella-zoster virus. A subfragment of the HSV type 1 KpnI i fragment encoding the MCP VP5 cross-hybridized with the HCMV HindIII U fragment containing part of the MCP gene. Northern (RNA) blot analyses with subclones out of the coding region for the HCMV MCP detected one large transcript of about 8 kilobases. A portion of the open reading frame was expressed in Escherichia coli plasmid pBD2 IC2OH as a β-galactosidase fusion protein and was used to generate polyclonal antibodies in New Zealand White rabbits. The obtained antisera reacted in Western immunoblots with the MCP of purified HCMV virions. A monoclonal antibody against the human MCP and a monospecific rabbit antiserum against strain Colburn of simian cytomegalovirus detected the fusion protein as well as the MCP of purified virions in immunoblots.

Human cytomegalovirus (HCMV), a member of the herpesvirus family, is a major pathogen in immunosuppressed individuals and a common infectious cause of congenital abnormalities. Conventional methods of HCMV diagnostics are thus far based on insufficiently defined viral antigens. Additionally, the study of the regulation of late gene expression of HCMV has been hampered by the limited number of mapped late genes (8, 9, 21, 24, 33, 41, 46). For studying the regulation of late gene expression and its influence on the pathogenicity of this virus, a detailed analysis of the virion components and the identification of the genes coding for them are prerequisites. This appears to be especially important for HCMV, since posttranscriptional means of late gene expression have been proposed recently (16, 17, 19).

Purified HCMV particles contain a minimum of 25 structural proteins, some of which are modified by glycosylation or phosphorylation (18, 26, 43, 52). At least two prominent large proteins with an apparent molecular weight of about 150,000 can be detected in virion preparations. One of these polypeptides was identified as the basic phosphoprotein pp150 and was considered a matrix component (18, 24, 45). The other protein is thought to represent the major capsid protein (MCP) of HCMV. This was deduced from an electrophoretic mobility similar to that of the MCP of simian cytomegalovirus (SCMV) strain Colburn in one-dimensional and two-dimensional gel electrophoreses (18). In addition, a polyvalent monospecific antiserum reactive with the putative MCP of SCMV cross-reacted with a polypeptide of about 150,000 molecular weight (150K polypeptide) in virion preparations of HCMV (55). The MCPs of other human herpesviruses, such as herpes simplex virus (HSV), varicella-zoster virus (VZV), and Epstein-Barr virus (EBV), also migrate at between 149K and 160K in polyacrylamide gel electrophoresis (10).

For a number of viruses, the capsid proteins serve as

dominant antigens in the course of infection. In HSV, the MCP has been shown to be highly reactive with human antisera (14). For HCMV, the immunodominant proteins appear to be two phosphorylated matrix proteins of 150K and 65K (25) and the glycosylated envelope protein of 58K (8, 20, 33). In contrast to the other human herpesviruses, the MCP of HCMV appears to be of limited antigenicity in its natural host; only a few sera from patients infected with HCMV detect the respective band in Western immunoblots (25). The differences in immune response to structural proteins of various herpesviruses might be important in terms of clinical and biological aspects of a particular virus.

This article describes the identification of the MCP gene of HCMV. The approach taken was to compare sequences of HCMV with the known genes for the MCPs of HSV, EBV, and VZV. A subfragment from the 5' end of the putative HCMV open reading frame was cloned in a bacterial expression vector. The resulting fusion protein was used to elicit antibodies in rabbits. These sera, together with other monoclonal and monospecific antibodies against the MCPs of HCMV and SCMV, were taken for a definite identification of the MCP gene of HCMV.

## MATERIALS AND METHODS

**Virus and cells.** HCMV AD169 was provided by U. Krech. Propagation of virus in human foreskin fibroblasts was done by standard methods. Purification of HCMV particles from the supernatant of infected cells was done by centrifugal separation in glycerol-tartrate gradients as described before (23, 53).

**Plasmids and Southern blot hybridization.** The HindIII U and a fragments of HCMV used for genomic sequencing were derived from the pAT153 recombinant plasmids H1G1 and V, respectively (44). These plasmids were the kind gift of Jon Oram. Other work used the HindIII U and a fragments subcloned from the cosmid pCM1007 (15) into dephosphorylated pACYC184 and pUC19 by standard meth-

---

* Corresponding author.

ods (34). The HSV type 1 (HSV-1) *Kpn*I i fragment encoding the MCP VP5 was kindly provided by B. Matz. Southern blot hybridizations were probed with a nick-translated *Dde*I subfragment of *Kpn*I-i. Hybridizations and washings were done under conditions of low stringency only.

**DNA sequencing and data analysis.** Sequencing was by the chain termination method of Sanger et al. (47). Random subfragments of *Hin*dIII-U and -a were generated, cloned in M13mp8 (40), and sequenced by the method of Bankier et al. (3). Sequence data were compiled by using the programs DBAUTO (50) and DBUTIL (49) and analyzed for the presence of open reading frames with ANALYSEQ (51). Predicted protein sequences were compared with protein library sequences by using FASTP (32). Multiple alignment of the MCP sequences of HCMV, EBV, VZV, and HSV-1 was performed by using the AMPS suite of programs (4). Diagonal homology plots were generated with the program COMPARE (13). All programs were run on DEC VAX computers operating VMS.

**RNA extraction and Northern (RNA) blot analysis.** Total late RNA of HCMV was isolated by previously published procedures (6, 24) 6 days postinfection with a cytopathic effect of 70%. Uninfected-cell RNA (mock RNA) was purified 7 days after passage. RNA was fractionated on denaturing agarose gels containing formaldehyde (31) and transferred to nitrocellulose filters. The RNA was hybridized with nick-repair-labeled plasmid DNA of the corresponding clones. Filter hybridization and washings were done as described previously (24).

**Expression cloning, induction, and purification of fusion proteins.** The construction of plasmid pSH1 was carried out by ligating a 1,142-base-pair *Stu*I-*Hin*dIII subfragment from the *Hin*dIII a fragment of HCMV into the *Sma*I-*Hin*dIII-cleaved expression plasmid pBD2 IC2OH (5, 36). Plasmid pNB1 was made by cloning a 928-base-pair *Nae*I-*Bam*HI subfragment from the *Hin*dIII U fragment into the *Eco*RV-*Bgl*II-restricted plasmid pBD2 IC2OH. pSH1 and pNB1 were transformed into *Escherichia coli* BMH7118 (39). For induction of fusion proteins, 1 ml of overnight culture of BMH7118 was diluted in 5 ml of Luria-Bertani medium. Bacteria were grown at 37°C to an optical density of 1.0 at 600 nm. The cells were induced by adding isopropyl-β-D-thiogalactosidase (IPTG) to a final concentration of 1 mM. After 3 to 5 h, cells were collected by centrifugation and suspended in 100 μl of polyacrylamide gel loading buffer as described by Laemmli (30). By this procedure, pNB1 produced no distinct fusion protein and the DNA construct proved unstable. For purification of the stable fusion protein SH1, induced 50-ml cultures were pelleted and suspended in 2 ml of phosphate-buffered saline containing 2 μl of DNase (1 mg/ml) and 5 μl of lysozyme (0.1 mg/ml). After incubation for 30 min on ice, the lysate was sonicated for 30 s and centrifuged at high speed. The pellet was washed for 30 min in phosphate-buffered saline, centrifuged, and suspended in 1 ml of 1 M urea. The whole procedure was repeated three times with 1 ml of 2 M, 4 M, and 7 M urea, respectively. The recombinant protein proved soluble only in the 7 M urea fraction. The supernatant of the last fraction was then dialyzed against phosphate-buffered saline.

**Monoclonal and monospecific antibodies.** A monoclonal antibody, directed against the MCP of HCMV, was kindly provided by W. Britt. A monospecific antibody against the MCP of SCMV strain Colburn was kindly provided by W. Gibson (9). The SH1 monospecific antiserum was raised in New Zealand White rabbits. The fusion protein SH1 was treated with increasing concentrations of urea as described

above and was used for immunization by previously published procedures (25).

**Protein gel electrophoresis and Western blot analysis.** Proteins were fractionated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis essentially as described previously (30). Gels were stained with Coomassie brilliant blue or with silver nitrate (38). Purified virions were separated in gels containing 8.5% (wt/vol) polyacrylamide with an acrylamide/methylenebisacrylamide ratio of 28:1 (high bis) (23). Separation of fusion proteins was done in 10% polyacrylamide gels. For Western blot analysis, proteins were electrophoretically transferred to nitrocellulose filters (54) and blocked with phosphate-buffered saline containing 1% bovine serum albumin and 0.05% Nonidet P-40. After reaction with the antibody, sheets were incubated with horseradish peroxidase-conjugated protein A or antiimmunoglobulin G. Fusion proteins were detected by using 4-chloro-1-naphthol as a substrate. For detection of virion antigens, a more sensitive procedure using diaminobenzidine in combination with $CoCl_2$ and $Ni(NH_3)_6SO_4$ was used (12).

## RESULTS

**Sequence analysis and relationship of the HCMV MCP to the MCPs of other herpesviruses.** The DNA sequences of the *Hin*dIII U, a, P, and S fragments of HCMV AD169 have been determined. Computer analysis was used to identify the presence of open reading frames within the sequence. One long open reading frame was found to span the junction of *Hin*dIII fragments a and U (Fig. 1 and 2) with a predicted translation product of 1,370 amino acids (Fig. 3). By comparing this amino acid sequence with a library containing EBV, HSV, and VZV sequences, matches to the MCPs of all three viruses were found. The percent amino acid identity within the overlapping regions of each pair of sequences was determined from the alignment shown in Fig. 3. Thus, identities of 29% between the MCPs of HCMV and EBV, 25% between the MCPs of HCMV and HSV-1, and 23% between the MCPs of HCMV and VZV were found. Dot matrix homology plots are also shown in Fig. 4.

**Genomic localization of the MCP of HCMV.** Computer analysis placed the coding region for the MCP of HCMV to the junction region of *Hin*dIII fragments a and U (Fig. 1). To see whether the degree of homology between the different herpesviruses was sufficient for cross-hybridization, cosmid clones of the whole HCMV genome were hybridized to a 3.1-kilobase (kb) *Dde*I subfragment of the HSV-1 *Kpn*I i fragment encoding the MCP (VP5) of this virus (Fig. 1). Hybridization above background was detected only with the 5.8-kb *Hin*dIII U fragment of HCMV (data not shown). This part of the open reading frame corresponds to the portion of the *Dde*I subfragment from the MCP gene of HSV-1.

**Northern blot analysis.** To identify the RNA for the MCP of HCMV, Northern blot analyses were carried out with subclones from its coding region. Late RNA from HCMV-infected fibroblasts was purified at about 6 days postinfection and subjected to electrophoresis on denaturing 1% agarose-formaldehyde gels. After transfer, the filters were hybridized to the α-$^{32}$P-labeled 2.1-kb *Hin*dIII a fragment and the 5.8-kb *Hin*dIII U fragment. The *Hin*dIII a fragment hybridized with an abundant RNA of about 8 kb and a minor one of about 5 kb (Fig. 5, lane 1). The *Hin*dIII U fragment revealed a predominant RNA of about 8 kb and two minor transcripts of about 3.3 and 2.2 kb (Fig. 5, lane 3). RNA from uninfected cells did not hybridize with the HCMV DNA probes (Fig. 5, lanes 2 and 4). Since only the 8-kb RNA was
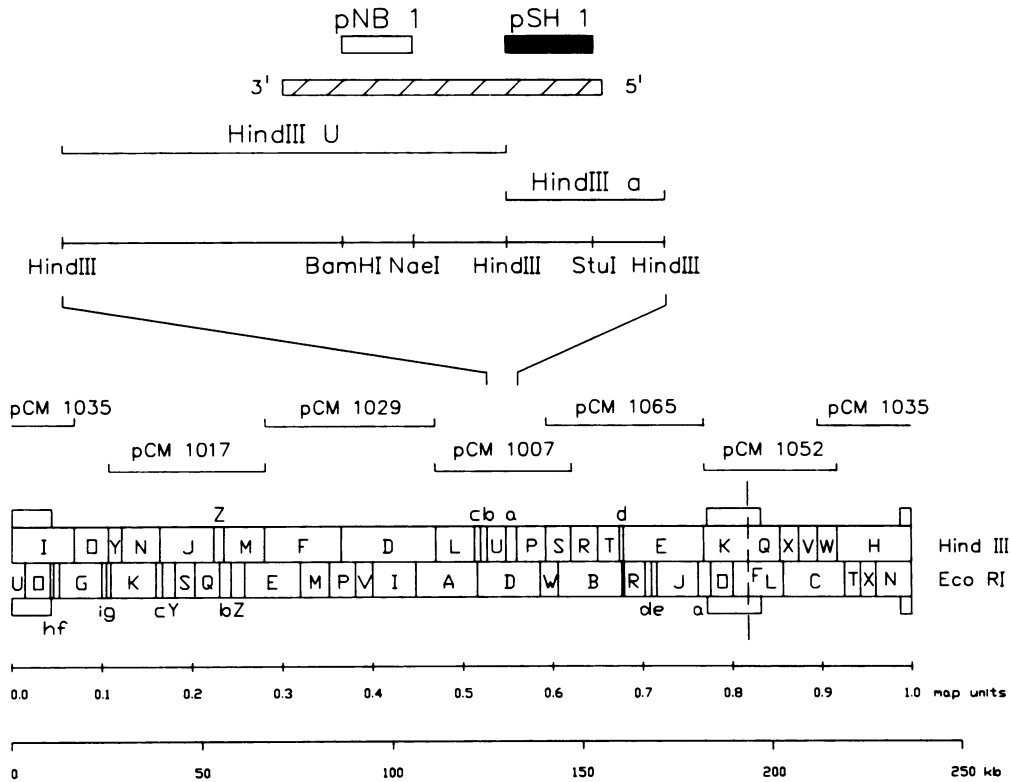
FIG. 1. Structure of the HCMV AD169 genome, with prototype arrangement of the HCMV genome with restriction sites for *Hin*dIII and *Eco*RI (bottom) and relative positions of the cosmids (15) and plasmids used in this study. Symbols: ▨, location of the open reading frame coding for the HCMV MCP; ▬, location of the expression clone pSH1; ☐, location of the unstable expression clone pNB1.

seen with these two adjacent fragments representing the coding region of the MCP open reading frame, this transcript appears to be the RNA coding for the MCP of HCMV.

**Verification of the MCP gene by procaryotic expression cloning.** To prove that the open reading frame does code for the 155K MCP of HCMV, parts of the open reading frame were cloned into the *E. coli* expression vector pBD2 IC2OH. A *Stu*I-*Hin*dIII DNA fragment of the putative MCP gene was ligated into the plasmid pBD2 IC2OH cut with *Sma*I-*Hin*dIII, giving rise to the clone pSH1 (Fig. 1). The vector harbors the DNA sequence for a truncated N-terminal portion of β-galactosidase of 375 amino acids, adjacent to a polylinker region which facilitates the in-frame ligation of gene fragments to the amino terminus of β-galactosidase. The recombinant clone pSH1 was transformed into *E. coli* BMH7118. The plasmid could be stably propagated in the bacteria, and a fusion protein of the expected size of about 92K (Fig. 6A, lane 2) was induced by IPTG. However, another construct with a *Nae*I-*Bam*HI fragment of 928 base pairs from a downstream region in the open reading frame proved to be unstable in our system (Fig. 1).

The SH1 fusion protein was used to raise antisera in New Zealand White rabbits. To reduce *E. coli*-specific proteins, the fusion protein was purified by urea fractionation before injection as described in Materials and Methods. A serum obtained after primary injection and two booster injections detected the MCP of purified virions as well as the fusion protein SH1 in Western blot assays (Fig. 6B). For the resolution of purified virions, we used 8.5% denaturing polyacrylamide gels with a high bisacrylamide ratio to separate the upper band with a molecular weight of about 155,000, corresponding to the MCP, and the lower band

corresponding to the basic phosphoprotein pp150. Additionally, a monospecific antiserum directed against the MCP of SCMV strain Colburn (54) reacted in Western blots with the HCMV MCP of purified virions and with the fusion protein SH1 (Fig. 6C). The fusion protein SH1 was also detected by a monoclonal antibody against the HCMV MCP (Fig. 6D). Besides the strong specific reaction with the MCP band of purified virions, the monoclonal antibody reacted weakly with an additional band of about 65K. In some cases this 65K band was also seen in Western blots when the rabbit antisera directed against the fusion protein SH1 and the MCP of SCMV were used. This kind of cross-reactivity was observed reproducibly and remains unclear.

Taken together, these data demonstrate that the long open reading frame spanning the *Hin*dIII a and U junction codes for the MCP of HCMV.

## DISCUSSION

For a number of genes, sequence conservation between the different members of the herpesvirus family has been shown either at the nucleic acid or the amino acid level. By computer search, for instance, Davison and Taylor (11) found 29 of 67 putative proteins deduced from the known genomic sequences of EBV and VZV to be homologous. Several genes of HCMV encoding structural and nonstructural proteins have been identified by searching for homologies in other herpesviruses. Examples of proteins identified in this way are the HCMV glycoproteins gB and gH, the major DNA-binding protein, and the DNA polymerase (9, 22, 27, 28).

This article describes the identification of the MCP gene of HCMV. The gene was identified by comparing the deduced

TAGCGGAGCCGCCAACGAC[TATTAGA]GCGTCACAGCCGAGGCGGCGCGGCGCAGCGCCGGCCATTCTCCGGGTCGCGCTGCTTTCCCCGAGCTCCGACGCCTCGCGCTCACGCCGCCGCC 120

```
          M  E  N  W  S  A  L  E  L  L  P  K  V  G  I  P  T  D  F  L  T  H  V  K  T  S  A  G  E  E  M  F  E  A  L  R  I  Y  Y
GCGATGGAGAACTGGTCGGCGCTCGAGCTCCTGCCTAAAGTAGGCATTCCCACTGACTTTCTGACGCACGTCAAGACCAGCGCCGGCGAGGAGATGTTTGAGGCCTTACGCATCTACTAC 240

     G  D  D  P  E  R  Y  N  I  H  F  E  A  I  F  G  T  F  C  N  R  L  E  M  V  Y  F  L  T  S  G  L  A  A  A  A  H  A  I  K
GGCGATGACCCCGAGCGTTACAATATTCACTTCGAAGCCATCTTCGGCACCTTCTGCAATCGCCTGGAATGGGTTTATTTTTTGACCAGCGGCTTGGCCGCGGCGGCGCACGCCATCAAG 360

     F  H  D  L  [N  K]  L  T  T  G  K  M  L  F  H  V  Q  V  P  R  V  A  S  G  A  G  L  P  T  S  R  Q  T  T  I  M  V  T  K  Y
TTCCACGACCTC[AATAAA]CTGACTACGGGCAAAATGCTGTTCCACGTGCAGGTGCCGCGTGTGGCCAGCGGCGCCGGTTTGCCCACTAGTCGTCAGACCACCATTATGGTCACCAAATAC 480

     S  E  K  S  P  I  T  I  P  F  E  L  S  A  A  C  L  T  Y  L  R  E  T  F  E  G  T  I  L  D  K  I  L  N  V  E  A  M  H  T
AGTGAAAAATCCCCCATCACCATCCCCTTCGAGCTCAGCGCAGCCTGCCTCACGTATCTGCGCGAGACATTCGAGGGCACCATCTTGGACAAGATCCTCAACGTGGAAGCCATGCACACG 600

     V  L  R  A  L  K  N  T  A  D  A  M  E  R  G  L  I  H  S  F  L  Q  T  L  L  R  K  A  P  P  Y  F  V  V  Q  T  L  V  E  N
GTGCTGCGCGCGCTTAAAAACACGGCCCGACGCCATGGAGCGCGGCTTGATTCACAGCTTTTTACAGCTTTACTACGCAAGCGCCGCCGGTACTTTGTGGTGCAGACGCTGGTAGAAAAC 720

     A  T  L  A  R  Q  A  L  N  R  I  Q  R  S  N  I  L  Q  S  F  K  A  K  M  L  A  T  L  F  L  L  N  R  T  R  D  R  D  Y  V
GCCACACTGGCGCGACAGGCACTGAACCGCATCCAGCGAAGCAACATCTTGCAGAGCTTCAAGGCCAAAATGCTGGCCACGCTCTTTTTGCTGAATCGCACGCGCGATCGCGACTATGTG 840

     L  K  F  L  T  R  L  A  E  A  A  T  D  S  I  L  D  N  P  T  T  Y  T  T  S  S  G  A  K  I  S  G  V  M  V  S  T  A  N  V
CTCAAGTTCCTCACGCGTCTGGCCGAAGCGGCCACCGATAGCATCCTAGACAACCCCACCACGTACACCACGAGCTCGGGGGCCAAGATCAGCGGCGTGATGGTAAGCACGGCCAACGTT 960

     M  Q  I  I  M  S  L  L  S  S  H  I  T  K  E  T  V  S  A  P  A  T  Y  G  N  F  V  L  S  P  E  N  A  V  T  A  I  S  Y  H
ATGCAGATCATCATGTCGCTGCTCTCAAGCCACATCACCAAAGAAACGGTCTCGGCGCCGGCCACCTACGGCAATTTTGTACTCAGTCCCGAAAATGCCGTCACTGCCATATCCTATCAC 1080

     S  I  L  A  D  F  N  S  Y  K  A  H  L  T  S  G  Q  P  H  L  P  N  D  S  L  S  Q  A  G  A  H  S  L  T  P  L  S  M  D  V
TCGATCCTGGCCGATTTCAACTCGTACAAGGCTCATCTGACTTCGGGGCCAACCGCATTTGCCCAACGACTCGCTTTCGCAGGCCGGGGCGCACAGCTTAACGCCGCTGAGCATGGACGTG 1200

     I  R  L  G  E  K  T  V  I  M  E  N  L  R  R  V  Y  K  N  T  D  T  K  D  P  L  E  R  N  V  D  L  T  F  F  F  P  V  G  L
ATCCGACTGGGCGAAAAAACGGTGATTATGGAAAACTTGCGGCGCGTCTACAAAAACACCGATACCAAAGATCCGCTAGAACGTAACGTGGACCTGACGTTTTTCTTCCCCGTAGGGCTG 1320

     Y  L  P  E  D  R  G  Y  T  T  V  E  S  K  V  K  L  N  D  T  V  R  N  A  L  P  T  T  A  Y  L  L  N  R  D  R  A  V  Q  K
TATCTGCCCGAGGATCGCGGTTACACCACGGTGGAAAGCAAAGTTA↑AGCTTAACGACACCGTTCGCAACGCGTTGCCCACCACGGCCTACCTGCTAAACCGGGACCGTGCCGTACAAAAA 1440

     I  D  F  V  D  A  L  K  T  L  C  H  P  V  L  H  E  P  A  P  C  L  Q  T  F  T  E  R  G  P  P  S  E  P  A  M  Q  R  L  L
ATAGACTTTGTGGACGCGCTCAAGACGCTGTGCCACCCGGTGCTGCACGAGCCCGCGCCCTGTCTACAGACCTTCACGGAACGCGGACCGCCCTCGGAACCCGCCATGCAACGACTGCTC 1560

     E  C  R  F  Q  Q  E  P  M  G  G  A  A  R  R  I  P  H  F  Y  R  V  R  R  E  V  P  R  T  V  N  E  M  K  Q  D  F  V  V  T
GAGTGTCGTTTTCAGCAGGAACCCATGGGCGGTGCGGCGCGCCGCCATACCGCATTTTTACCGCGTACGACGCGAGGTCCCGCGCACGGTGAACGAAATGAAGCAGGACTTTGTGGTCACC 1680

     D  F  Y  K  V  G  N  I  T  L  Y  T  E  L  H  P  F  F  D  F  T  H  C  Q  E  N  S  E  T  V  A  L  C  T  P  R  I  V  I  G
GACTTCTACAAGGTGGGCAACATCACGCTCTACACCGAGCTGCACCCCTTCTTCGACTTTACGCACTGCCAGGAAAACAGCGAAACGGTGGCGCTGTGTACACCGCGTATCGTCATCGGT 1800

     N  L  P  D  G  L  A  P  G  P  F  H  E  L  R  T  W  E  I  M  E  H  M  R  L  R  P  P  P  D  Y  E  E  T  L  R  L  F  K  T
AATTTGCCAGATGGACTGGCACCGGGCCCCTTCCACGAACTGCGCACCTGGGAGATCATGGAACACATGCGACTGCGGCCACCACCCGACTACGAAGAGACGTTGCGACTTTTCAAAACC 1920

     T  V  T  S  P  N  Y  P  E  L  C  Y  L  V  D  V  L  V  H  G  N  V  D  A  F  L  L  I  R  T  F  V  A  R  C  I  V  N  M  F
ACCGTGACAAGCCCCAATTACCCCGAGCTTTGCTACCTGGTCGACGTGCTGGTACACGGCAACGTCGACGCTTTTTTGCTAATCCGCACTTTCGTAGCACGTTGTATCGTCAACATGTTT 2040

     H  T  R  Q  L  L  V  F  A  H  S  Y  A  L  V  T  L  I  A  E  H  L  A  D  G  A  L  P  P  Q  L  L  F  H  Y  R  N  L  V  A
CACACACGTCAGCTGCTGGTTTTTGCGCACAGCTACGCGCTGGTGACGTTGATCGCCGAACACCTGGCTGATGGCGCCCTACCACCGCAGTTGCTCTTTCACTACCGCAATCTGGTCGCC 2160

     V  L  R  L  V  T  R  I  S  A  L  P  G  L  N  N  G  Q  L  A  E  E  P  L  S  A  Y  V  N  A  L  H  D  H  R  L  W  P  P  F
GTGCTGCGTCTGGTAACGCGCATCTCAGCGCTGCCGGGACTTAACAACGGTCAACTGGCTGAGGAACCGCTATCGGCTTACGTGAACGCGCTGCACGACCACCGGCTGTGGCCCCCCTTC 2280

     V  T  H  L  P  R  N  M  E  G  V  Q  V  V  A  D  R  Q  P  L  N  P  A  N  I  E  A  R  H  H  G  V  S  D  V  P  R  L  G  A
GTTACGCACCTGCCGCGTAACATGGAAGGCGTACAGGTAGTGGCCGACCGGCAACCGTTGAACCCGGCCAACATCGAGGCGCGACACCACGGCGTTTCGGACGTACCGCGTTTGGGCGCC 2400

     M  D  A  D  E  P  L  F  V  D  D  Y  R  A  T  D  D  E  W  T  L  Q  K  V  F  Y  L  C  L  M  P  A  M  T  N  N  R  A  C  G
ATGGATGCCGACGAGCCGCTGTTTGTGGACGATTACCGTGCCCACCGACGACGAATGGACGTTACAAAAGGTCTTTTATCTCTGCCTCATGCCCGCCATGACCAACAACCGGGCCTGCGGC 2520

     L  G  L  N  L  K  T  L  L  V  D  L  F  Y  R  P  A  F  L  L  M  P  A  A  T  A  V  S  T  S  G  T  T  S  K  E  S  T  S  G
CTGGGACTCAACCTCAAGACGCTGCTAGTAGACCTCTTTTACCGACCGGCGTTTCTCCTGATGCCGGCCGCCACCGCGGTCTCCACGTCAGGTACGACGTCGAAAGAGAGCACAAGCGGT 2640

     V  T  P  E  D  S  I  A  A  Q  R  Q  A  V  G  E  M  L  T  E  L  V  E  D  V  A  T  D  A  H  T  P  L  L  Q  A  C  R  E  L
GTCACCCCCGAAGACTCGATAGCGGCGCAACGTCAAGCCGTAGGCGAGATGCTGACCGAGCTGGTGGAGGACGTCGCCACCGACGCACATACGCCGCTGCTGCAGGCCTGTCGTGAACTG 2760

     F  L  A  V  Q  F  V  G  E  H  V  K  V  L  E  V  R  A  P  L  D  H  A  Q  R  Q  G  L  P  D  F  I  S  R  Q  H  V  L  Y  N
TTTTTGGCCGTGCAATTCGTGGGCGAGCACGTCAAGGTGCTCGAAGTGCGCGCGCCTCTGGACCACGCGCAGCGTCAGGGCCTGCCGGACTTTATCTCACGTCAGCACGTGCTTTACAAC 2880

     G  C  C  V  V  T  A  P  K  T  L  I  E  Y  S  L  P  V  P  F  H  R  F  Y  S  N  P  T  I  C  A  A  L  S  D  D  I  K  R  Y
GGCTGCTGCGTCGTGACGGCGCCTAAGACGTTGATCGAGTACAGCCTACCCGTACCTTTCCACCGTTTCTACTCAAACCCCACCATCTGCGCGGCGTTGAGCGACGACATTAAACGCTAC 3000

     V  T  E  F  P  H  Y  H  R  H  D  G  G  F  P  L  P  T  A  F  A  H  E  Y  H  N  W  L  R  S  P  F  S  R  Y  S  A  T  C  P
GTGACCGAGTTCCCGCACTATCACCGTCACGACGGTGGTTTCCCGCTACCCACAGCCTTCGCGCACGAGTACCACAACTGGCTGCGCTCGCCCTTTTCGCGTTACTCGGCCACCTGTCCC 3120

     N  V  L  H  S  V  M  T  L  A  A  M  L  Y  K  I  S  P  V  S  L  V  L  Q  T  K  A  H  I  H  P  G  F  A  L  T  A  V  R  T
AACGTGTTGCACTCAGTCATGACGCTGGCCGCCATGTTATACAAAATCTCACCAGTTTCACTGGTGCTGCAGACCAAGGCGCACATCCACCCGGGATTCGCACTCACGGCCGTACGCACC 3240

     D  T  F  E  V  D  M  L  L  Y  S  G  K  S  C  T  S  V  I  I  N  N  P  I  V  T  K  E  E  R  D  I  S  T  T  Y  H  V  T  Q
GACACCTTCGAGGTGGACATGCTGCTCTACAGCGGCAAATCATGCACGTCAGTTATCATCAACAACCCTATCGTGACCAAGGAGGAACGCGACATCAGCACTACGTACCACGTCACGCAG 3360

     N  I  N  T  V  D  M  G  L  G  Y  T  S  N  T  C  V  A  Y  V  N  R  T  D  M  G  V  R  V  Q  D  L  F  R  V  F  P  M
AACATTAACACCGTAGACATGGGTCTGGGTTACACGTCAAATACCTGCGTAGCCTACGTAAACCGCGTGCGAACCGACATGGGCGTCCGCGTGCAGGACCTCTTTCGCGTCTTCCCTATG 3480

     N  V  Y  R  H  D  E  V  D  R  W  I  R  H  A  A  G  V  E  R  P  Q  L  L  D  T  E  T  I  S  M  L  T  F  G  S  M  S  E  R
AACGTGTACCGGCACGACGAGGTGGACCGCT↑GGATCCGGCACGCGGCCGGCGTAGAACGGCCGCAACTGTTAGACACGGAGACCATCTCTATGCTGACGTTCGGCAGCATGTCCGAGCGC 3600

     N  A  A  A  T  V  H  G  Q  K  A  A  C  E  L  I  L  T  P  V  T  M  D  V  N  Y  F  K  I  P  N  N  P  R  G  R  A  S  C  M
AATGCCGCCGCCACGGTGCACGGACAAAAGGCCGCGTGCGAACTAATCCTTACGCCGGTGACGATGGACGTTAATTACTTTAAGATCCCCAACAACCCGCGCGGGCGTGCCTCCTGCATG 3720

     L  A  V  D  P  Y  D  T  E  A  A  T  K  A  I  Y  D  H  R  E  A  D  A  Q  T  F  A  A  T  H  N  P  W  A  S  Q  A  G  C  L
CTGGCCGTCGATCCGTACGACACGGAGGCCGCCACCAAGGCCATCTACGACCATCGCGAAGCGGACGCACAGACTTTCGCGGCCACACACAACCCCTGGGCCTCGCAAGCCGGCTGTCTG 3840

     S  D  V  L  Y  N  T  R  H  R  E  R  L  G  Y  N  S  K  F  Y  S  P  C  A  Q  Y  F  N  T  E  E  I  I  A  A  N  K  T  L  F
AGCGACGTACTCTACAACACCCGACATCGCGAACGTCTAGGCTACAACTCCAAATTCTACAGCCCTTGCGCGCAGTACTTTAACACAGAAGAGATTATCGCGGCCAACAAAACGCTCTTC 3960

     K  T  I  D  E  Y  L  L  R  A  K  D  C  I  R  G  D  T  D  T  Q  Y  V  C  V  E  G  T  E  Q  L  I  E  N  P  C  R  L  T  Q
AAAACCATCGACGAGTACCTGCTACGCGCCAAGGACTGCATCCGCGGCGACACGGACACGCAGTACGTCTGCGTCGAAGGCACGGAACAACTCATCGAAAACCCGTGCCGTCTGACGCAG 4080

     E  A  L  P  I  L  S  T  T  T  L  A  L  M  E  T  K  L  K  G  G  A  G  A  F  A  T  S  E  T  H  F  G  N  Y  V  V  G  E  I
GAAGCGCTACCGATCCTCAGCACCACCACGCTGGCGCTGATGGAGACCAAACTCAAGGGTGGCGCCGGGGCTTTTGCTACGTCCGAAACGCACTTTGGTAACTACGTCGTAGGCGAAATC 4200

     I  P  L  Q  Q  S  M  L  F  N  S  *
ATCCCGCTGCAGCAATCCATGTTATTTAACTCGTGAGGCCGTCCGGTGAGGAGGACGGCGACGACCGCAGGTTAGCGGCGAGTCACCTAGACGCAAACGCGGGCCCGGACGCGCCACGCT 4320
```

FIG. 2. Nucleotide sequence of the HCMV MCP gene. The complete sequence of the open reading frame is shown together with a translation from the first ATG, which fits the Kozak consensus (29). The probable TATA signal for this gene is boxed, as well as a poly(A) signal which is presumed to be cryptic for this reading frame. The orientation of the sequence is opposite to that shown in Fig. 1. The HindIII and BamHI restriction sites are marked by vertical arrows (positions 1366 and 3512, respectively).

```
         1
VZV40    MTTVSCPANVITTTESDRIAGLFNIPAGIIPTGNVLSTIEVCAHRCIFDFFKQIRSDDNSLYSAQFDILLGTYCNTLNFVRFLELGLSVACICTKFPELA
UL19     MAAPNRDPPGYRYAAAMVPTGSLLSTIEVASHRRLFDFFSRVRSDANSLYDVEFDALLGSYCNTLSLVRFLELGLSVACVCTKFPELA
BcLF1    MASNEGVENRPFPYLTVDADLLSNLRQSAAEGLFHSFDLLVGKDAREAGIKFEVLLGVYTNAIQYVRFLETALAVSCVNTEFKDLS
HaLF1    MENWSALELLPKVGIPTDFLTHVKTSAGEEMFEALRIYYGDDPERYNIHFEAIFGTFCNRLEWVYFLTSGLAAAAHAIKFHDLN
CONSENSUS          *  *   L*  *    *   *F       *       F* * G * N *  V FL *L* *     F *L

         101
VZV40    YVRDGVIQFEVQQPMIARDGPHPVDQPVHNYMVKRIHKRSLSAAFAIASEALSLLSNTYVDGTEIDSSLRIRAIQQMARNLRTVLDSFERGTADQLLGVL
UL19     YMNEGRVQFEVHQPLIARDGPHPIEQPTHNYMTKIIDRRALNAAFSLATEAIALLTGEALDGTGIGAHRQLRAIQQLARNVQAVLGAFERGTADQMLHVL
BcLF1    RMTDGKIQFRISVPTIAHGDGRRPSKQRTFIVVKNCHKHHISTEMELSMLDLEILHS  IPETPVEYAEYVGAVKTVASALQFGVDALERGLINTVLSVK
HaLF1    KLTTGKMLFHVQVPRVASGAGLPTSRQTTIMVTKYSEKSPITIPPFELSAACLTYLRET FEGTILDKILNVEAMHTVLRALKNTADAMERGLIHSFLQTL
CONSENSUS       * G * F * P *A  *      * K  *  *  ** * L      T *  * A* * * *  * ERG   L

         201
VZV40    LEKAPPLSLLSPINKFQPEGH  LNRVARAALLSDLKRRVCADMFFMTRHARE     PRLISAYLSDMVSC TQPSVMVSRITHTNTRG RQVDGVLVTT
UL19     LEKAPPLALLLPMQRYLDNGR  LATRVARATLVAELKRSFCETSFFLGKAGHR    REAVEAWLVDLTTA TQPSVAVPRLTHADTRG RPVDGVLVTT
BcLF1    LRHAPPMFILQTLADPTFTER  GFSKTVKSDLIAMFKRHLLEHSFFLDRAENMGSGFSQYVRSRLSEMVAAVSGESVLKGVSTYTTAKGGEPVGGVFIVT
HaLF1    LRKAPPYFVVQTLVENATLARQALNRIQRSNILQSFKAKMLATLFLLNRTRD       RDYVLKFLTRLAEA ATDSILDNPTTYTTSSG AKISGVMVST
CONSENSUS    L *APP ** **    *   *  ** **  K   * F * *       * L  *  * S*   T  * G  * GV * T

         301
VZV40    ATLKRQLLQGIL QIDDTAADVPVTYGEMVLQGTNLVTALVMGKAVRGMDDVARHLLDITDPNTLNIPSIPPQSNSDSTTAGLPVNARVPADLVIVGDKL
UL19     APIKQRLLQSFL KVEDTEADVPVTYGEMVLNGANLVTALVMGKAVRSLDDVGRHLLEMQEEQ    LDLNRQTLDELESA PQTTRVRADLVSIGEKL
BcLF1    DNVLRQLLT FLGEEADNQIMGPSSYASFVVRGENLVTAVSYGRVMRTFEHFMARIVDSPEKAGSTKSDLPAVAAGVEDQP     RVPISAAVIKLGNHA
HaLF1    ANVMQIIMSLLSSHITKETVSAPATYGNFVLSPENAVTAISYHSILADFNSYKAHLTSGQ       PHLPNDSLSQAGAH    SLTPLSMDVIRLGEKT
CONSENSUS     *  **           P *Y* V* * N VTA*   *  *  **          * *          * ** *G**

         401                                                                       ^
VZV40    VFLEALERRVYQATRVAYPLIGNIDITFIMPMGVFQANSMDRYTRHAGDFSTVSEQ DPRQFPPQGIFFYNKDGILTQLTLRDAMGTICHSSLLDVEATL
UL19     VFLEALEKRIYAATNVPYPLVGAMDLTFVLPLGLFNP VMERFAAHAGDLVPAPGHPDDPRAFPPRQLFFWGKDRQVLRLSLEHAIGTVCHPSLMNVDAAV
BcLF1    VAVESL QKMYNDTQSPYPLNRRMQYSYYFPVGLFMPNPK YTTSAA IKMLD  NPTQQLPVEAWIVNKNNLLLAFNLQNALKVLCHPRLHTPAHTL
HaLF1    VIMENL RRVYKNTDTKDPLERNVDLTFFFPVGLYLPEDRG YTTVESKVKLND   TVRNALPTTAYLLNRDRAVQKIDFVDALKTLCHPVLHEPAPCL
CONSENSUS V *E L **Y T  PL  *  ** P*G** *    **  *          P  *  **  *      A*  *CH* L     *

         501
VZV40    VALRQQHLDRQCYFGVYVAEG TEDTLDVQMGRFMETWADMMPH HPHWVNE HLTILQFIAPSNPRL RFELNPAFDFFVAPGDVDLPGPQRPPEAMPT
UL19     GGLNRDPVEAANPYGAYVAA  PAGPAADMQQLFLNAWGQRLAHGRVRWVAEGQMTPEQFMQPDNANL ALELHPAFDFFVGVADVELPGGDVPPAGPGE
BcLF1    NSLNAAPAPRDRR ETYSLQHRRPNHMNVLV IVDEFYDNKYAAPVTDIALKCGLPTEDFLHPSNYDLLRLELHPLYDIYIG        RDAGERAR
HaLF1    QTFTERGPPSEPAMQRLLECRFQQEPMGGAARRIPHFYRVRREVPRTVNEMKQDFVVTDFYKVGNITLYT ELHPFFDF          THCQENSETV
CONSENSUS *                          *                        *F   N  L  EL P *D                       *

         601                                                                     ^
VZV40    VNATLRIINGNIPVPLCPISFRDCRGTQLGLGR   HTMT PATIKAVKDTFEDRAYPTIFYMLEAVIHGNERNFCALLRLLTQCIRGYWEQSHRVAFVN
UL19     IQATWRVVNGNLPLALCPAAFRDARGLELGVGR   HAMA PATIAAVRGAFDDRNYPAVFYLLQAAIHGSHVFCALARLVVQCITSYWNNTRCAAFVN
BcLF1    HRAVHRLMVGNLPTPLAPAAFQEARGGQFETATSLAHVVD QAVIETVQDTAYDTAYPAFFYVVEAMIHGFEEKFVMNVPLVSLCINTYWERSGRLAFVN
HaLF1    ALCTPRIVIGNLPDGLAPGPFHELRTWEIMEHMRLRPPPDYEETLRLFKTTVTSPNYPELCYLVDVLVHGNVDAFLLIRTFVARCIVNMFHTRQLLVFAH
CONSENSUS     R** GN*P *L P *F * R* *        *      *    YP  Y*** *HG   F   * CI  *      F

         701
VZV40    NFHMLMYITTYLGNGELPEVCINIYRDLLQHVRALRQTITDFTIQGEGHNGETSEALNNILTDDTFIAPILWDCDALIYRDEAARDRLPAIRVSGRNG Y
UL19     DYSLVSYVVTYLG GDLPEECMAVYRDLVAHVEALAQLVDDFTLTGPELGGQAQAELNHLMRDPALLPPLVWDCDALMRRAALDRHRDCRVSAGGHDPVY
BcLF1    SFSMIKFICRHLGNNAISKEAYSMYRKIYGELIALEQALMRLA GSDVVGDESVGQYVCALLDPNLLPPVAYTDIFTHLLTVSDRAPQIIIGNEVYADTL
HaLF1    SYALVTLIAEHLADGALPPQLLFHYRNLVAVLRLVTRISALPGLNNGQLAEEPLSAYVNALHDHRLWPPFV       THLPRNMEGVQVVADRQPLN
CONSENSUS * **  *  L* **  YR * * *    *     **    * D  *P          R

         801                                     ^
VZV40    QALHFVDMAGHNFQRRDNVL IHGRPVRGDTGQGIP  ITPHHDREWG ILSKIYYYIVIPAFSRGSCCTMGVRYDRLYPAL   QAVIVPEIPADEEAP
UL19     AAA  CNVATADFNRNDGQL LHNTQARAADAADDR   PHRGADWT VHHKIYYYVMVPAFSRGRCCTAGVRFDRVYATL   QNMVVPEIAPGEECP
BcLF1    AAPQFIERVGNMDEMAAQFVALYGYRVNGDHDHDFRLHLGPYVDEGHADVLEKIFYYVFLPTCTNAHMCGLSGVDFQHVAQTL            AYNGP
HaLF1    PA       NIEARHHG   VSDVPRLGAMDADEPLFVDDYRATDDEWTLQKVFYLCLMPAMTNNRACGLGLNLKTLLVDLFYRPAFLLMPAATAVSTS
CONSENSUS A          *                        K**Y * *P* *   C* G*    * L                             *

         901
VZV40    TTP EDPRHPLHAHQLVPNSLNVYFHNAHLTVDGDALLTLQELMGDMAERTTAI LVSSAPDAGAA TATTRNMRIYDGALY  HGLIMMAYQAYDETIA
UL19     SDPVTDPAHPLHPANLVANTVNAMFHNGRVVVDGPAMLTLQVLAHNMAERTTAL LCSAAPDAGAN TASTTNMRIFDGALH  AGILLMAPQHLDHTIQ
BcLF1    AFSHHFTRDEDILDNLENGTLRDLLEISDLRPTVGMIRDLSASFMTCPTFTRAV RVSVDNDVTQQLAPNPADKRTEQTVLV  NGLVAFAFSERTRAVT
HaLF1    GTTSKESTSGVTPEDSIAAQRQAVGEMLTELVEDVATDAHTPLLQACRELFLAVQFVGEHVKVLEVRAPLDHAQRQGLPDFISRQHVLYNGCCVVTAPKT
CONSENSUS * *  *     *                              *   A* **      **   R  *      ** *

         1001
VZV40    TGTFFYPVPVNPLFACPEHLASLRGMTNARRV LAKMVPPIPPF     LGANHHATIRQPVAYHVTHSKSDFNTLTYSLLGGYFKFTPISLTHQLRTGF
UL19     NGDYFYPLPVHALFAGADHVANAPNFPPALRD LSRQVPLVPP      ALGANYFSSIRQPVVQHVRESAAGENALTYALMAGYFKISPVALHHQLKTGL
BcLF1    QC LFHAIPFHMFYGDPRVAATMHQDVATFVM RNPQQRAVEAFNRPEQLFAEYREWHRSPMGKYAAECLPSLVSIS GHTAMHIKMSPMAYIAQAKLKI
HaLF1    LIEYSLPVPFHRFYSNPTICAALSDDIKRYVTEFPHYHRHDGGFPLPTAFAHEYHNWLRSPFSRYSATCPNVLHSVM TLAAMLYKISPVSLVLQTKAHI
CONSENSUS          **P    ** * A            *       *    R P          ** ** *  K *P** Q *

         1101
VZV40    HPGIAFTVVRQDRFATEQLLYAERASESYFVGQIQVHHHDA IGGVNFTLTQPRAHVDLGVGYTAVCATAALRCPLTDMGNTAQNLFFSRGGVPMLHDNV
UL19     HPGFCGFTVVRQDRFVTENVLFSERASEAYFLGQLQVARHET GGGVNFTLTQPRANVDLGVGYTAVVATATVRNPVTDMGNLPQNFYLGRGAPPLLDNAA
BcLF1    HPGVAMTVVRTDEILSENILFSSRASTSMFIGTPNVSRREARVDAVTFEVHHEMASIDTGLSYSSTMTPARVAAITTDMGIHTQDFFSVFPAEAFGNQQV
HaLF1    HPGFALTAVRTDTFEVDMLLYSGKSCTSVIINNPIVTKEERDIST TYHVTQNINTVDMGLGYTSNTCVAYVNRVRTDMGVRVQDLFRVFPMNVYRHDEV
CONSENSUS HPG * T VR D   * *L** ** *  *   V *    * * *     *D G**Y** A *   TDMG Q* *  *   *

         1201                                    ^
VZV40    TESLRRITAS GGRLNPTE PLPIFGGLRPATSAGIARGQASVCEFVAMPVSTDLQYFRTACNPRGRASGMLYMGDRDADI EAIMFDHTQSDVAYTDR
UL19     AVYLRNAVVA GNRLGPAQ PVPVFGCAQVPRRAGMDHGQDAVCEFIATPVSTDVNYFRRPCNPRGRAAGGVYAGDKEGDV TALMYDHGQSDPSRAFA
BcLF1    NDYIKAKVGAQRNGTLLRDPRTYLA GMTNVNGAPGLCHGQQATCEIIVTPVTADVAYFQKSNSPRGRAACVVSCENYNQEVAEGLIYDHSRPDAAYEYR
HaLF1    DRWIRHAAGVERPQLLDTETISMLTFGSMSERNAAATVHGQKAACELILTPVTMDVNYFKIPNNPRGRASCMLAVDPYDTEAATKAIYDHREADAQ TFA
CONSENSUS   **       L        G      ** *GQ * CE *  PV* D* YF  *  PRGRA* *    * *    **DH *D

         1301                       ^                                        ^
VZV40    ATLNPWASQKHSYGDRLYNGTYN LTGASPIYSPCFKFFTPAEVNTNCNTLDRLLMEAKA VASQSSTDTEYQFKRPPGSTEMTQDPCGLFQEAYPPL
UL19     ATANPWASQRFSYGDLLYNGAYH LNGASPVLSPCFKFFTSADIAAKHRCLERLIVETGS AVSTATAASDVQFKRPPGCRELVEDPCGLFQEAYPLT
BcLF1    STVNPWASQLGSLGDIMYNSSYR QTAVPGLYSPCRAFFNKEELLRNNRGLYNMVNEYSQRLGGHPATSNTEVQFVVIAGTDVFLEQPCSFLQEAFPAL
HaLF1    ATHNPWASQAGCLSDVLYNTRHRERLGYNSKFYSPCAQYFNTEEIIAANKTLFKTIDEYLL RAKDCIRGDTDTQYVCVEGTEQLIENPCRLTQEALPIL
CONSENSUS *T NPWASQ     D *YN*       *   SPC *F **    L  * E          * ** Q*    G     **PC  QEA P

         1401
VZV40    CSSDAAMLRTAHAGETGA  DEVHLAQYLIRDASPLRGCLPLPR
UL19     CASDPALLRSARNGEAHA  RETHFAQYLVYDASPLKGLAL
BcLF1    SASSRALIDEFMSVKQTH  APIHYGHYIIEEVAPVRRILKFGNKVVF
HaLF1    STTTLALMETKLKGGAGAFATSETHFGNYVVGEIIPLQQSMLFNS
CONSENSUS   **   A**          H * Y** * P*
```

FIG. 3. An alignment of the MCP sequences of VZV (gene 40) (10), HSV-1 (UL19) (37), EBV (BcLF1) (1), and HCMV (HaLF1). The alignment was carried out by using the AMPS programs of Barton and Sternberg (4). Minor changes were made manually. The consensus shows positions where all four sequences contain the same amino acid (single-letter code) or where changes are conservative (*). The groupings used for this purpose were I, L, V, C, and M; F, Y, and W; G, A, S, T, and P; N, Q, D, and E; and H, K, and R. Six conserved cysteine residues are also indicated by a symbol above their alignment positions (^).

## A



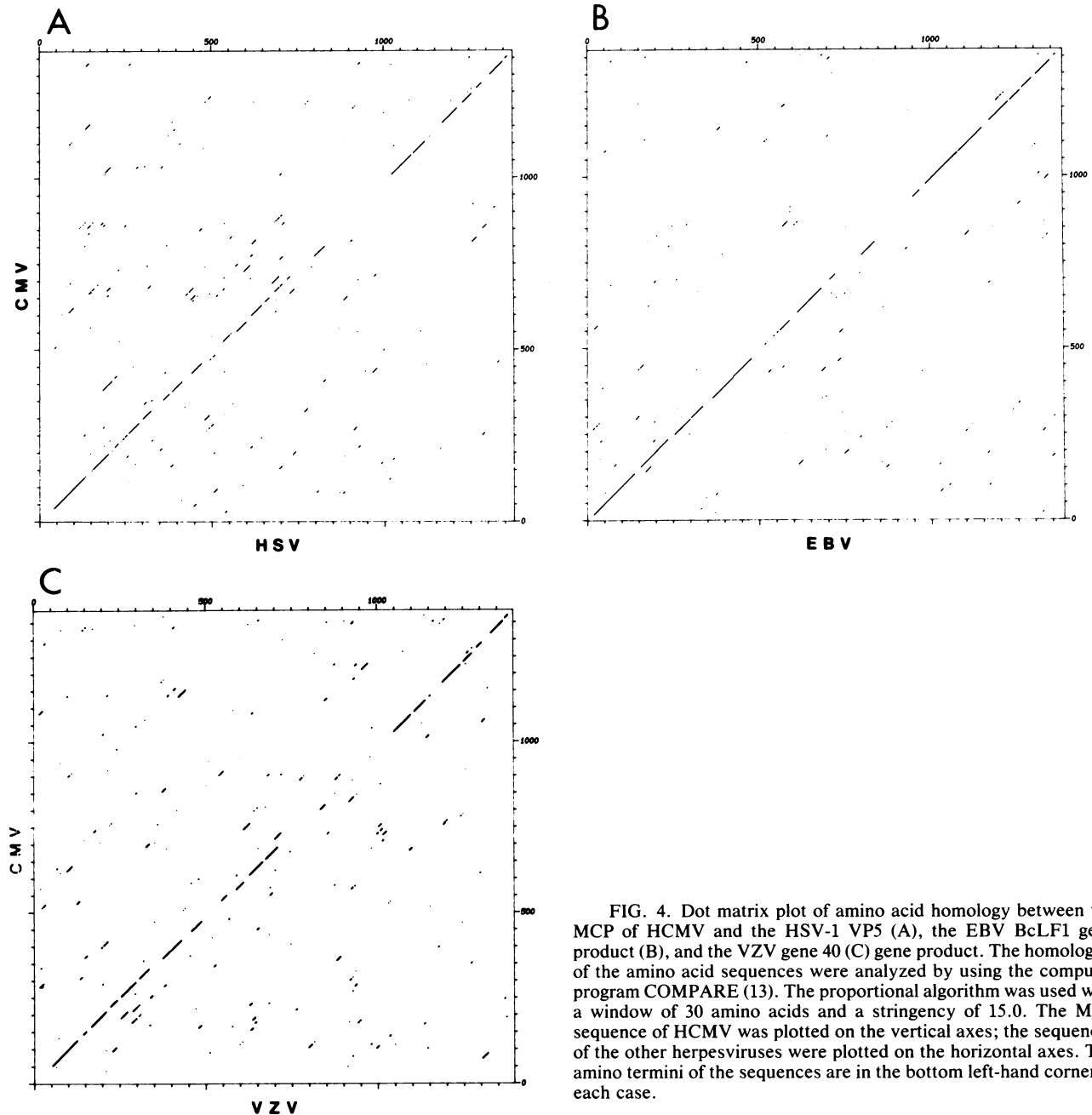**HSV**

## B



**E B V**

## C



**V Z V**

FIG. 4. Dot matrix plot of amino acid homology between the MCP of HCMV and the HSV-1 VP5 (A), the EBV BcLF1 gene product (B), and the VZV gene 40 (C) gene product. The homologies of the amino acid sequences were analyzed by using the computer program COMPARE (13). The proportional algorithm was used with a window of 30 amino acids and a stringency of 15.0. The MCP sequence of HCMV was plotted on the vertical axes; the sequences of the other herpesviruses were plotted on the horizontal axes. The amino termini of the sequences are in the bottom left-hand corner in each case.

open reading frames from the HindIII fragments U and a of HCMV AD169 with a library of herpesvirus protein sequences which included the MCP sequences of HSV-1, VZV, and EBV. The gene for the MCP VP5 of HSV-1 was originally localized by intertypic recombinant studies (35, 42) and subsequently verified by immunological means (7) and by sequence analysis (10). The open reading frame for VP5 was found to code for a polypeptide with a deduced molecular weight of 149,000. This was in good agreement with the electrophoretic mobility of the MCP of HSV-1 in polyacrylamide gel electrophoresis. Using the predicted amino acid sequence of the VP5 gene product, Davison and Scott found homologous genes in the genomic sequences of VZV (gene 40) and EBV (BcLF1) (1, 10, 11). The predicted molecular weights according to the sequences were 155,000

and 154,000 respectively; this corresponded well with molecular weights as determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. The homology of the amino acid sequences was found to be as high as 50% between HSV-1 and VZV proteins, whereas that between the HSV-1 and the EBV proteins was 27% and that between the VZV and the EBV proteins was 24% (10). This was considered a reflection of the closer relationship between HSV and VZV, which both belong to the same subfamily of herpesviruses, the *Alphaherpesvirinae*. EBV is a member of the gammaherpesvirus subfamily and thus is more distantly related to the latter viruses. HCMV, in contrast, is a member of the *Betaherpesvirinae*. The open reading frame identified in this study potentially codes for a polypeptide of 1,370 amino acids with an estimated molecular weight of 154,000.
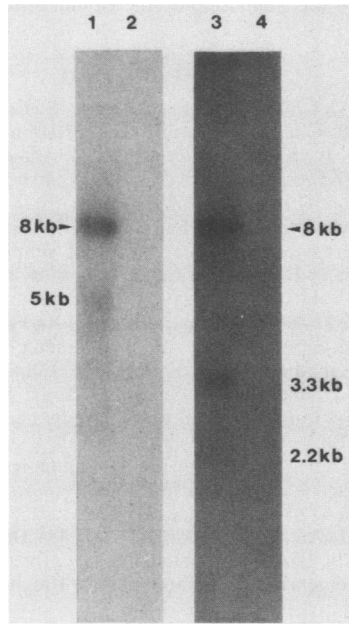
FIG. 5. Northern blot hybridizations with the α-³²P-labeled *Hind*III a and U fragments. Whole-cell RNA was isolated from mock-infected or HCMV-infected fibroblasts. RNA was purified 6 days after infection or mock infection, separated on a 1% agarose-formaldehyde gel, and transferred to nitrocellulose sheets. Lanes: 1, hybridization of the *Hind*III a fragment with late RNA; 2, hybridization of the *Hind*III a fragment with mock RNA; 3, hybridization of the *Hind*III U fragment with late RNA; 4, hybridization of the *Hind*III U fragment with mock RNA. Both DNA clones detected an RNA of about 8 kb.

It shows a homology of 25% and 23%, respectively, to the MCPs of the human alphaherpesviruses HSV-1 and VZV. A slightly higher homology of 29% was found between HCMV and the gammaherpesvirus EBV.

In a preliminary transcriptional analysis using the Northern blot technique, an RNA of about 8 kb hybridized with both the *Hind*III a and the *Hind*III U fragments. Smaller RNAs were also seen in late-RNA preparations. One transcript of about 5 kb hybridized with the *Hind*III a but not with the *Hind*III U fragment. On the other hand, two bands of about 2.2 and 3.3 kb were detectable in Northern blots only when the *Hind*III U fragment was used as a probe. Since the large RNA was the only transcript which hybridized with both DNA clones containing the entire MCP open reading frame, this transcript is the most likely candidate for the mRNA of the MCP. It has been shown that the MCP of HSV-1 is encoded by an abundant 6-kb RNA (7). The fine mapping and characterization of the MCP mRNA together with the question of whether there is overlapping transcription from another gene, as has been reported for HSV-1 (7), has to be addressed in a different study.

Initially we had tried to isolate a recombinant expression clone from a λgt11 cDNA library. This expression library proved to be useful for the identification of three other structural protein genes of HCMV (25, 33, 41). Screening the cDNA library for a MCP expression clone was done with a monoclonal antibody directed against the MCP of HCMV and a cross-reactive rabbit polyclonal serum against the MCP of SCMV. However, no clones could be isolated by this procedure. We do not know the reason for this. It may reflect major differences in structure or kinetics between the MCP RNA and the RNAs of other structural proteins of HCMV.

To express polypeptides representing epitopes of the MCP, two fragments from the open reading frame were
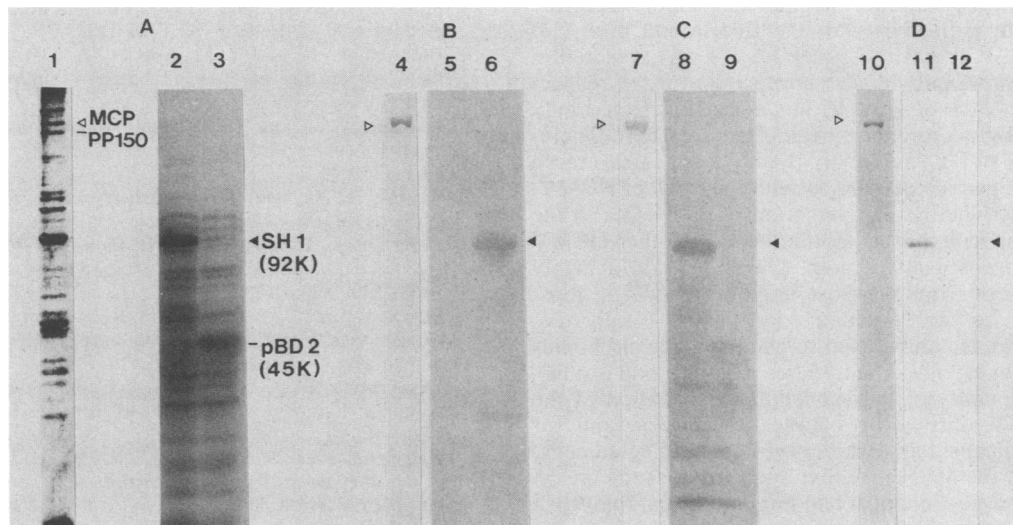


FIG. 6. Western blot of HCMV virions and the fusion protein SH1 with different monospecific antisera and a monoclonal antibody against the MCP. (A) Lanes: 1, protein pattern of purified virions after polyacrylamide gel electrophoresis (8.5% high-bis gel) and silver staining; 2, lysate of *E. coli* BMH7118 expressing the fusion protein SH1; 3, lysate of *E. coli* BMH7118 with induced pBD2 IC2OH. The bacterial lysates were electrophoresed on a 10% acrylamide gel and stained with Coomassie brilliant blue. (B) Western blot with a rabbit antiserum directed against the fusion protein SH1; (C) Western blot with a rabbit antiserum directed against the MCP of SCMV strain Colburn; (D) Western blot with a monoclonal antibody directed against the MCP of HCMV. For lanes 4, 7, and 10, purified virions were separated on an 8.5% high-bis gel and transferred to nitrocellulose filters. For lanes 6, 8, and 11, lysates of *E. coli* BMH7118 expressing the fusion protein SH1 were separated on a 10% polyacrylamide gel and transferred to nitrocellulose. For lanes 5, 9, and 12, *E. coli* lysate (strain BMH7118) expressing pBD2 IC2OH without insert was used. Symbols: ▷, position of MCP of purified virions; ◀, position of the fusion protein SH1.

cloned in a procaryotic expression vector. One *StuI-HindIII* fragment from *HindIII* fragment a of HCMV (Fig. 1) could be stably expressed in *E. coli*, and the fusion protein reacted with a monoclonal antibody against the HCMV MCP as well as with a polyclonal monospecific rabbit serum against the SCMV MCP. However, a *NaeI-BamHI* fragment further downstream (Fig. 1) could not be expressed in bacteria. The phenomenon that particular fragments of HCMV cannot be stably expressed in *E. coli* has been observed on other occasions (48). For epitope mapping, further expression cloning with various fragments is necessary to obtain a representative panel of distinct antigens from the MCP of HCMV. These clones may help to answer a number of questions unresolved thus far. One of these is that despite sequence conservation, no immunologic cross-reactivity has been described between the MCPs of the different human herpesviruses. Balachandran et al. (2) obtained these results using a number of polyvalent rabbit antisera against HSV-1, HSV-2, HCMV, and EBV and a panel of monoclonal antibodies against HSV and EBV proteins. Similar results were also obtained in a study by Weiner and Gibson (55), who detected no cross-reactivity in Western blots using a polyclonal antiserum prepared against the purified 145K MCP of strain Colburn of SCMV to the MCPs of HSV-1, HSV-2, and VZV. However, this antiserum detected the MCPs of several primate cytomegaloviruses as well as the MCPs of HCMVs. Hence, Weiner and Gibson suggested that the MCP of cytomegalovirus might serve as a group-specific antigen. Defining such group-specific epitopes may be relevant in terms of understanding whether the conservation of this specific protein reflects structural constraints in nucleocapsid assembly. In addition, the definition of group-specific epitopes will be helpful for the improvement of specific tools for HCMV diagnostics, which thus far is still dependent upon poorly defined and potentially cross-reactive antigen preparations.

Another subject to be addressed with defined MCP epitopes will be the variation in immune responses of the host to different herpesvirus MCPs. Eberle and Mou (14) found that the MCP of HSV-1 and HSV-2 was detected in Western blots in all individuals seropositive for the respective virus. The MCP of HCMV, however, is detected only inconsistently by acute and convalescent sera (25). Therefore, in the case of HSV infection, the MCP appears to serve as a reliable antigen in comparison with the MCP of HCMV. The question of whether the detection of antibodies to the MCP marks a specific phase in infection or whether HCMV has found a mechanism to more efficiently hide a highly expressed protein from the host immune system is most appropriately assessed with defined antigens. Finally, defined MCP antigens can be used to elicit specific antibodies for certain epitopes. These antibodies might be one promising tool for the study of the subcellular localization of the MCP of HCMV during the course of infection and for determining the capsid assembly of virions in infected cells. The mapping of the MCP gene therefore provides the means for investigating the biological and immunological relevance of the MCP during HCMV infection.

## ACKNOWLEDGMENTS

## 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. Seguin, 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. Balachandran, N., D. E. Oba, and L. M. Hutt-Fletcher. 1987. Antigenic cross-reactions among herpes simplex virus types 1 and 2, Epstein-Barr virus, and cytomegalovirus. J. Virol. 61:1125–1135.
3. Bankier, A. T., K. M. Weston, and B. G. Barrell. 1987. Random cloning and sequencing by the M13/dideoxynucleotide chain termination method. Methods Enzymol. 155:51–93.
4. Barton, G. J., and M. J. E. Sternberg. 1987. A strategy for the rapid multiple alignment of protein sequences. J. Mol. Biol. 198:327–337.
5. Bröker, M. 1986. Vectors for regulated high-level expression of proteins fused to truncated forms of Escherichia coli β-galactosidase. Gen. Anal. Technol. 3:53–57.
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. Costa, R. H., G. Cohen, R. Eisenberg, D. Long, and E. K. Wagner. 1984. Direct demonstration that the abundant 6-kilobase herpes simplex virus type 1 mRNA mapping between 0.23 and 0.27 map units encodes the major capsid protein VP5. J. Virol. 49:287–292.
8. Cranage, M. P., T. Kouzarides, A. T. Bankier, S. Satchwell, K. Weston, P. Tomlinson, B. Barrell, H. Hart, S. E. Bell, A. C. Minson, and G. L. Smith. 1986. Identification of the human cytomegalovirus glycoprotein B gene and induction of neutralizing antibodies via its expression in recombinant vaccinia virus. EMBO J. 5:3057–3063.
9. Cranage, M. P., G. L. Smith, S. E. Bell, H. Hart, C. Brown, A. T. Bankier, P. Tomlinson, B. G. Barrell, and T. C. Minson. 1988. Identification and expression of a human cytomegalovirus glycoprotein with homology to the Epstein-Barr virus BXLF2 product, varicella-zoster virus gpIII, and herpes simplex type 1 glycoprotein H. J. Virol. 62:1416–1422.
10. Davison, A. J., and J. E. Scott. 1986. DNA sequence of the major capsid protein gene of herpes simplex virus type 1. J. Gen. Virol. 67:2279–2286.
11. Davison, A. J., and P. Taylor. 1987. Genetic relation between varicella-zoster virus and Epstein-Barr virus. J. Gen. Virol. 68:1067–1079.
12. DeBlas, A. L., and H. M. Cherwinski. 1983. Detection of antigens on nitrocellulose paper. Anal. Biochem. 133:214–219.
13. Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387–395.
14. Eberle, R., and S.-W. Mou. 1983. Relative titers of antibodies to individual polypeptide antigens of herpes simplex virus type 1 in human sera. J. Infect. Dis. 148:436–443.
15. Fleckenstein, B., I. Müller, and J. Collins. 1982. Cloning of the complete human cytomegalovirus genome in cosmids. Gene 18:39–46.
16. Geballe, A. P., F. S. Leach, and E. S. Mocarski. 1986. Regulation of cytomegalovirus late gene expression: genes are controlled by posttranscriptional events. J. Virol. 57:864–874.
17. Geballe, A. P., R. R. Spaete, and E. S. Mocarski. 1986. A cis-acting element within the 5′ leader of a cytomegalovirus β transcript determines kinetic class. Cell 46:865–872.
18. Gibson, W. 1983. Protein counterparts of human and simian cytomegaloviruses. Virology 128:391–406.
19. Goins, W. F., and M. F. Stinski. 1986. Expression of a human cytomegalovirus late gene is posttranscriptionally regulated by a 3′-end-processing event occurring exclusively late after infection. Mol. Cell. Biol. 6:4202–4213.
20. Gretch, D. R., R. C. Gehrz, and M. F. Stinski. 1988. Character-

ization of a human cytomegalovirus glycoprotein complex (gcI). J. Gen. Virol. **69**:1205–1215.

21. **Gretch, D. R., B. Kari, R. C. Gehrz, and M. F. Stinski.** 1988. A multigene family encodes the human cytomegalovirus glycoprotein complex gcII (gp47-52 complex). J. Virol. **62**:1956–1962.

22. **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.

23. **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.

24. **Jahn, G., T. Kouzarides, M. Mach, B.-C. Scholl, B. Plachter, B. Traupe, E. Preddie, S. C. Satchwell, B. Fleckenstein, and B. G. Barrell.** 1987. Map position and nucleotide sequence of the gene for the large structural phosphoprotein of human cytomegalovirus. J. Virol. **61**:1358–1367.

25. **Jahn, G., B.-C. Scholl, B. Traupe, and B. Fleckenstein.** 1987. The two major structural phosphoproteins (pp65 and pp150) of human cytomegalovirus and their antigenic properties. J. Gen. Virol. **68**:1327–1337.

26. **Kim, K. S., V. J. Sapienza, R. I. Carp, and H. M. Moon.** 1976. Analysis of structural polypeptides of purified human cytomegalovirus. J. Virol. **20**:604–611.

27. **Kouzarides, T., A. T. Bankier, S. C. Satchwell, K. Weston, P. Tomlinson, and B. G. Barrell.** 1987. Sequence and transcription analysis of the human cytomegalovirus DNA polymerase gene. J. Virol. **61**:125–133.

28. **Kouzarides, T., A. T. Bankier, S. C. Satchwell, K. Weston, P. Tomlinson, and B. G. Barrell.** 1987. Large scale rearrangement of homologous regions in the genomes of HCMV and EBV. Virology **157**:397–413.

29. **Kozak, M.** 1983. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.

30. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.

31. **Lehrach, H., D. Diamond, J. M. Wozney, and H. Boedtker.** 1977. RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry **16**:4742–4751.

32. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.

33. **Mach, M., U. Utz, and B. Fleckenstein.** 1986. Mapping of the major glycoprotein gene of human cytomegalovirus. J. Gen. Virol. **67**:1461–1467.

34. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

35. **Marsden, H. S., N. D. Stow, V. G. Preston, M. C. Timbury, and N. M. Wilkie.** 1978. Physical mapping of herpes simplex virus-induced polypeptides. J. Virol. **28**:624–642.

36. **Marsh, J. L., M. Erfle, and E. J. Wykes.** 1984. The pIC plasmids and phage vectors with versatile cloning sites for recombinant selection by insertional inactivation. Gene **32**:481–485.

37. **McGeoch, D. J., M. A. Dalrymple, A. J. Davison, A. Dolan, M. C. Frame, D. McNab, L. J. Perry, J. E. Scott, and P. Taylor.** 1988. The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. J. Gen. Virol. **69**:1531–1574.

38. **Merril, C. R., D. Goldman, S. A. Sedman, and M. H. Ebert.** 1981. Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. Science **211**:1437–1438.

39. **Messing, J., B. Gronenborn, B. Müller-Hill, and P. H. Hofschneider.** 1977. Filamentous coliphage M13 as a cloning vehicle: insertion of a *Hind*II fragment of the lac regulatory region in M13 replicative form *in vitro*. Proc. Natl. Acad. Sci. USA **74**:3642–3646.

40. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269–276.

41. **Meyer, H., A. T. Bankier, M. P. Landini, C. M. Brown, B. G. Barrell, B. Rüger, and M. Mach.** 1988. Identification and prokaryotic expression of the gene coding for the highly immunogenic 28-kilodalton structural phosphoprotein (pp28) of human cytomegalovirus. J. Virol. **62**:2243–2250.

42. **Morse, L. S., L. Pereira, B. Roizman, and P. A. Schaffer.** 1978. Anatomy of herpes simplex virus (HSV) DNA. X. Mapping of viral genes by analysis of polypeptides and functions specified by HSV-1 × HSV-2 recombinants. J. Virol. **26**:389–410.

43. **Nowak, B., C. Sullivan, P. Sarnow, R. Thomas, F. Bricout, J. C. Nicolas, B. Fleckenstein, and A. J. Levine.** 1984. Characterization of monoclonal antibodies and polyclonal immune sera directed against human cytomegalovirus virion proteins. Virology **132**:325–338.

44. **Oram, J. D., R. G. Downing, A. Akrigg, A. A. Dollery, C. J. Duggleby, G. W. G. Wilkinson, and P. J. Greenaway.** 1982. Use of recombinant plasmids to investigate the structure of the human cytomegalovirus genome. J. Gen. Virol. **59**:111–129.

45. **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.

46. **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.

47. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

48. **Scholl, B.-C., J. von Hintzenstern, B. Borisch, B. Traupe, M. Bröker, and G. Jahn.** 1988. Procaryotic expression of immunogenic polypeptides of the large phosphoprotein (pp150) of human cytomegalovirus. J. Gen. Virol. **69**:1195–1204.

49. **Staden, R.** 1980. A new computer method for the storage and manipulation of DNA gel reading data. Nucleic Acids Res. **8**:3673–3694.

50. **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.

51. **Staden, R.** 1986. The current status and portability of our sequence handling software. Nucleic Acids Res. **14**:217–231.

52. **Stinski, M. F.** 1976. Human cytomegalovirus: glycoproteins associated with virions and dense bodies. J. Virol. **19**:594–609.

53. **Talbot, P., and J. D. Almeida.** 1977. Human cytomegalovirus: purification of enveloped virions and dense bodies. J. Gen. Virol. **36**:345–349.

54. **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.

55. **Weiner, D., and W. Gibson.** 1981. Identification of a primate cytomegalovirus group-common protein antigen. Virology **115**: 182–191.