JOURNAL OF CLINICAL MICROBIOLOGY, Nov. 1991, p. 2494–2502
0095-1137/91/112494-09$02.00/0

Vol. 29, No. 11

Comparative Sequence Analysis of Human Cytomegalovirus Strains

ROBERT LEHNER, THOMAS STAMMINGER, AND MICHAEL MACH*

Institut fuer Klinische und Molekulare Virologie, Friedrich-Alexander-Universitaet Erlangen/Nuernberg, Loschgestrasse 7, 8520 Erlangen, Germany

Received 25 April 1991/Accepted 14 August 1991

Three regions of DNA from five low-passage clinical isolates of human cytomegalovirus were amplified by polymerase chain reaction. The DNA sequences as well as the predicted amino acid sequences were compared with those of the laboratory strains AD169 and Towne. The genomic regions consisted of (i) three regions from the major glycoprotein (gp58/116, unique long [U_L] 55), (ii) three regions from the integral membrane protein (IMP, U_L100), and (iii) a region from the major immediate-early 1 and 2 (IE-1/2) enhancer/promoter. Homologies ranged from 75.8 to 100.0% on the nucleotide level and from 47 to 100% on the amino acid level. The following two patterns were observed. (i) There are regions with a high degree of conservation with few scattered point mutations (mainly in the IE-1/2 enhancer/promoter and in the IMP gene). (ii) There are clusters of highly variable regions (parts of the gp58/116 gene and of the IMP gene). Within the areas of high variability, the strains could be classified into a limited number of subtypes.

Human cytomegalovirus (HCMV) is capable of establishing lifelong persistent infections which normally remain asymptomatic. Infections with this virus, however, can also result in a wide variety of clinical manifestations, especially in immunologically impaired individuals or in newborns. In immunocompromised individuals, symptoms can range from mild disease to life-threatening interstitial pneumonia (38). Infections in newborns can be asymptomatic or lead to severe damage of the child (52). The mechanisms that determine the type of clinical manifestation are not understood. Host factors such as development of a potent cellular and/or humoral immune response might play an important role. It has been shown that the synthesis of cytotoxic T lymphocytes in bone marrow transplant patients correlates with the clinical outcome of HCMV infection (41). Antibodies may also have a role in protective immunity against HCMV disease since prophylaxis of transplant patients has been successfully carried out by the administration of HCMV hyperimmune globulin (13, 48). By using immunoprecipitation and immunoblotting analyses, it has also been established that the humoral immune response varies greatly among individuals (1, 21, 27, 28, 61). On the other hand, genetic variability among different virus strains could influence clinical manifestations of HCMV infections. DNA structure variants were first demonstrated when patterns of restriction endonuclease digestions of viral DNA from different sources were compared (24). Although the differences were found to be dispersed over the entire genome (10), particular regions, such as the *a* sequence, could be identified which contain a higher rate of mutations (62). Chou (12) recently reported the differentiation of 40 HCMV strains by restriction analysis of DNA fragments from clinical specimens amplified by polymerase chain reaction (PCR). Restriction analyses, however, do not allow conclusions about individual differences in viral proteins or regulatory regions.

Different isolates have also been shown to be antigenically variable (22, 59). In a number of cases, monoclonal antibodies which are capable of neutralizing established laboratory strains failed to react with various clinical isolates (2, 33, 42). In other virus systems, antigenic variations have been reported which determine cell or organ tropism of different virus strains and types, respectively. In different human immunodeficiency virus type 1 strains, for example, regions on gp120 can be defined which determine the selective tropism of the respective variant for T cells and macrophages (47). In herpes simplex virus type 1 (HSV-1), a single amino acid substitution within glycoprotein C is responsible for a serologic type conversion to an HSV-2 epitope (25).

As a first step to address the questions of strain-specific differences on the sequence level, we compared selected parts of DNA sequences of five low-passage HCMV clinical isolates with those of the laboratory strains AD169 and Towne. The entire DNA sequence of HCMV AD169 has been recently established and can be used as a prototype sequence (11). We used the PCR method to amplify DNA fragments derived from the following: (i) the gp58/116 (unique long [U_L] 55) gene which codes for the major glycoprotein complex on the viral surface (6) and represents an important target for the humoral (26, 58) and the cellular (4) immune responses, (ii) the integral membrane protein (IMP) (U_L100) gene which codes for an envelope protein with multiple membrane-spanning domains (29), and (iii) the immediate-early 1 and 2 (IE-1/2) enhancer/promoter region which is an important regulatory element of viral immediate early gene expression (5, 57).

MATERIALS AND METHODS

Cells and viruses. Human foreskin fibroblasts were used for the propagation of HCMV AD169 and Towne as well as for wild-type isolates. All clinical strains have been isolated from urine specimens examined in our diagnostic laboratory. Strains 1 and 2 were derived from asymptomatically infected newborns. Strains 3, 4, and 5 were isolated from patients undergoing active HCMV infection after renal transplantation. Eagle's minimal essential medium (GIBCO, Glasgow, Scotland), supplemented with 5% fetal calf serum, 350 mg of glutamine per ml, and 100 mg of gentamicin per ml, was used as the cultivation medium. Cells were infected with 1 to 5 PFU of strain AD169 or Towne per cell. Infection with the wild-type isolates was performed by mixing equal numbers of infected cells and noninfected cells. Infected cells were harvested after 5 to 12 days of cultivation depending on the

* Corresponding author.

TABLE 1. PCR primer pairs

Primer[a]	Sequence 5′ to 3′[b]	Position[c]
58–20	acggttGAATTCCGACACGCAAGAGACCACGA	−57 to −38
58–21	acggttAAGCTTGTAGCTGGCATTGCGATTGG	867 to 848
58–23	acggttGAATTCGTGTTCTGGCAAGGCATCAA	1285 to 1304
58–24	acggttAAGCTTCAGTACCCTGAAGTCGGTAT	1989 to 1970
IMP-N	aaccgtGAATTCACCCGCGAGCCTGTCGTCAT	−67 to −48
IMP-C	tcagtcAAGCTTATGTCCATGTCTCCAAGTCT	1157 to 1138
IE-A	catcGGATCCGAATCCGCGTTCCAATGCAC	104 to 85
IE-B	gaggAAGCTTAGAGGCGACATCAAGCTGG	−789 to −770

[a] IMP-N, amino-terminal primer of the IMP gene; IMP-C, carboxy-terminal primer of the IMP gene; IE-A, 3′ primer of the IE-1/2 enhancer/promoter; IE-B, 5′ primer of the IE-1/2 enhancer/promoter.

[b] Recognition sites for the restriction endonucleases are underlined. Lowercase letters indicate the primer clamps.

[c] The position numbers refer to the first base of the translational initiation codon in the case of gp58/116 and IMP, respectively. For the immediate-early primer pair, the numbers give the location relative to the cap site of the IE-1 transcript.

capability of the respective virus strain to induce full cytopathic effect in cell culture.

DNA preparation from HCMV-infected cells. Infected cells were scraped from the bottom of 75-cm^2 tissue culture flasks and collected by centrifugation. The cells were washed twice with phosphate-buffered saline and subsequently incubated in 300 μl of sarcosyl lysis buffer (2% sarcosyl, 50 mM Tris-HCl [pH 7.5], 10 mM EDTA) for 30 min at 56°C. RNA and proteins were enzymatically degraded by sequential incubation with RNase A (10 min at 37°C; final concentration, 0.3 mg/ml) and proteinase K (1 h at 37°C; final concentration, 0.7 mg/ml), respectively. The lysate was passed through a 12-gauge needle twice and extracted with phenol-chloroform twice. The DNA was precipitated by the addition of 2 volumes of ethanol in the presence of 0.1 M NaCl and dissolved in 0.5 ml of distilled water.

PCR amplification. Amplification of HCMV-specific DNA fragments was performed by the standard procedures described previously (32, 44). DNA isolated from mock-infected cells was used as a negative control. The oligonucleotide primers were prepared with a Cyclone DNA synthesizer (Biosearch, Inc., San Rafael, Calif.), applying the phosphoamidite method (34). Sequences and locations of the primer pairs on the AD169 genome are listed in Table 1. All of the oligonucleotides contained a recognition motif for a restriction endonuclease and a four- or six-base clamp sequence at their 5′ ends. Thirty cycles of amplification were performed (annealing, at 45°C for 2 min; elongation, at 72°C for 2 min; denaturation, at 95°C for 0.5 min) on a Biomed Thermocycler (B. Braun, Theres, Germany). Aliquots of the amplification assays (1/10 volume) were analyzed on 1% agarose gels.

Fragment purification and DNA cloning. PCR amplification products were separated on 1% agarose gels. Bands containing the amplified fragments were visualized under UV illumination and excised, and nucleic acids were purified by using a Gene Clean kit (Bio 101, La Jolla, Calif.) according to the manufacturer's instructions. Sticky ends suitable for cloning were created by digestion of the fragments with the appropriate restriction endonucleases. The cleaved ends were removed by an additional purification step with the Gene Clean kit, and the PCR-amplified fragments were ligated into the plasmid pUC18. All procedures were performed as described by Maniatis et al. (32).

DNA sequencing. Nucleotide sequences were determined by the method of Sanger et al. (46). A T7 DNA polymerase kit from Pharmacia (Freiburg, Germany) was used by following the recommended protocol.

GenBank accession numbers for the reported sequences. The following GenBank accession numbers have been assigned: (i) for the *AD-2* region of gp58/116, M64956 (strain 1), M64949 (strain 3), M64937 (strain 4), M64938 (strain 2), M64939 (strain 5); (ii) for the cleavage region of gp58/116, M64955 (strain 1), M64946 (strain 3), M64948 (strain 4), M64951 (strain 2), M64953 (strain 5); (iii) for the *AD-1* region of gp58/116, M64945 (strain 1), M64947 (strain 3), M64950 (strain 4), M64952 (strain 2), M64954 (strain 5); (iv) for the amino-terminal region of IMP, M64973 (strain 1), M64959 (strain 3), M64962 (strain 4), M64974 (strain 2), M64964 (strain 5), M64968 (Towne); (v) for the internal region of IMP, M64958 (strain 1), M64961 (strain 3), M64957 (strain 4), M64965 (strain 2), M64969 (strain 5), M64972 (Towne); (vi) for the carboxy-terminal region of IMP, M64971 (strain 1), M64960 (strain 3), M64963 (strain 4), M64966 (strain 2), M64967 (strain 5), M64970 (Towne); (vii) for the IE-1/2 enhancer/promoter region, M64940 (strain 1), M64941 (strain 3), M64942 (strain 4), M64943 (strain 2), M64944 (strain 5).

Computer analyses. DNA sequence editing, translation, comparison, and alignment of multiple sequences was performed by use of the UWGCG (University of Wisconsin Genetics Computer Group) software package (15) on a VAX/VMS V5.0-2 computer.

RESULTS

Comparison of sequences from the gp58/116 gene. The gp58/116 gene is located at the right end of the *Hin*dIII fragment R of HCMV AD169 (Fig. 1) and has the capacity to encode a polypeptide of 102 kDa (30). The DNA sequences of the open reading frames in strains AD169 and Towne have been established (14, 50). The two genes have overall similarities of 94% on the nucleotide level and 95% on the amino acid level. The majority of the differences in the amino acid sequence occur in two clusters between amino acid residues 28 and 67 and between residues 453 and 472. Of the overall differences, 62% are located in these two areas. In strain Towne, the precursor protein is proteolytically cleaved by a cellular endopeptidase at the amino acid sequence motif RTKR/S_{460} (50). A similar motif RTRR/S is located at an identical position within the gene of strain AD169.

The primer pairs for the amplification of gp58/116 sequences were designed such that the amplification products contained the following regions which are known to be functionally and/or immunologically important (Fig. 1): two neutralizing antibody-binding sites, namely, *AD-1* which is located in the region around amino acid 616 on gp58 (26, 58) and *AD-2* which spans amino acids 68 to 76 on gp116 (37), and the proteolytic cleavage site of the precursor molecule at amino acid 460. The amplification products were cloned into plasmid pUC18 and subjected to sequence analysis. The nucleotide and deduced amino acid sequences of five clinical isolates were aligned and compared with the published sequences of the laboratory strains AD169 and Towne. To assess the error rate of the *Taq* polymerase in the amplification process, a control assay with human foreskin fibroblast cells infected with AD169 was performed.

The nucleotide sequences of the clinical isolates from the amino-terminal 250 bp containing the coding sequence for *AD-2* showed between 97.2 and 87.4% homology to the sequences of strain AD169. Alignment of the sequences of the two isolates with the lowest homology (isolates 3 and 4)

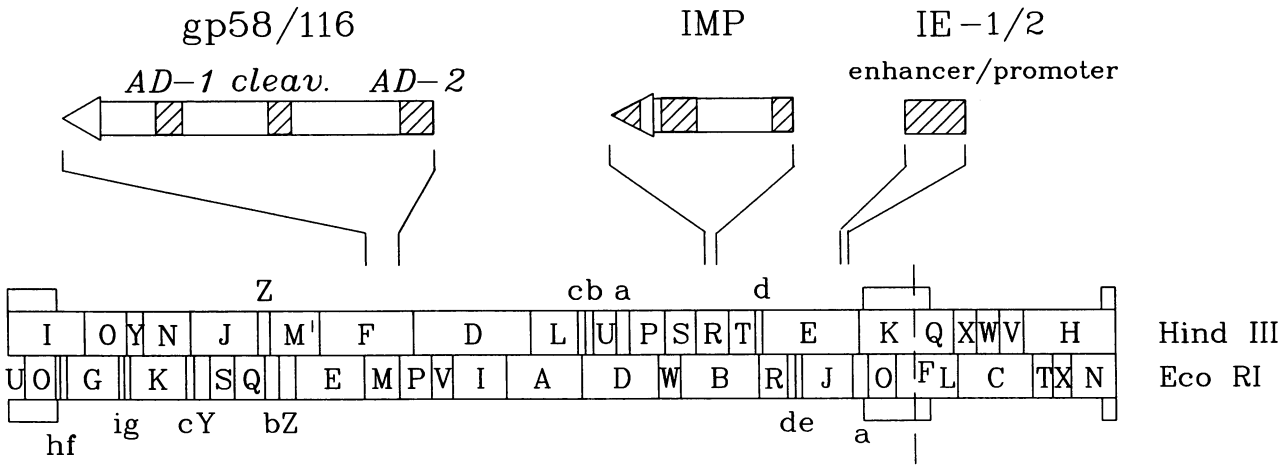

FIG. 1. Schematic representation of the HCMV genome with restriction maps given for *Hin*dIII and *Eco*RI (lower part) and the examined genomic regions shown in an expanded scale (upper part). Arrows indicate the direction of transcription of gp58/116 and IMP, respectively. Hatched areas show the regions which have been subjected to sequence comparison.

to those of strain AD169 revealed that the differing nucleotides are not distributed randomly over the entire sequence but rather are clustered in a defined region of 117 bp (nucleotide [nt] 84 to 201; data not shown). This is also reflected in the deduced amino acid sequences (Fig. 2A). The first 28 amino acid residues which most probably contain the signal sequence are almost completely conserved among all strains. However, the differences that were found include a Cys-to-Ser exchange at position 18 in strain 4. Between amino acids 29 and 68, strains 3 and 4 in particular showed considerable divergence from AD169, including exchanges of amino acid residues as well as insertions and deletions. However, these strains were identical to strain Towne. Strains 1, 2, and 5 had two (strains 1 and 5) or five (strain 2) amino acid exchanges, compared with AD169. The amino acid sequence following the variable cluster has been shown to be recognized by the neutralizing human monoclonal antibody C23 (36). This region was conserved among all tested strains. The sequence amplified from AD169-infected cells was identical to the published sequence. This is true not only for this region but also for all other analyzed regions, indicating that the error rate of the *Taq* polymerase can be neglected in our analyses (Fig. 3 to 5).

The second region that was examined (nt 1321 to 1440) contains the coding sequence for the proteolytic cleavage site (Fig. 1). On the DNA level, the degree of conservation ranged between 75.8 and 90.1%. The overall sequence variability in this area was higher than that for the amino-terminal part of the gene. It is dispersed over the entire sequence (data not shown). Isolates 1 and 5 had the lowest homology to AD169 in the cleavage region (75.8 and 76.1%, respectively), whereas they were similar to AD169 in the *AD-2* region (97.2 and 96.2%, respectively). The alignments of the encoded peptide sequences reflected this situation (Fig. 2B). In general, on the amino acid level, the strain variations were more complex in this area (ranging between 47 and 82% for amino acid residues 448 to 480) than in the variable cluster within the *AD-2* region (ranging between 62 and 97% for amino acid residues 18 to 78). The patterns of strains 2 to 4 were more similar to the Towne sequence than to that of AD169. The patterns of strains 1 and 5 were different from those of both AD169 and Towne. The putative recognition motif for the endopeptidase which is RTRR/S for AD169 was not conserved. It was mutated to RTKR/S in strains 1, 3, 4, and 5, which is the sequence found in strain Towne.

The third region examined within the gp58/116 gene spans nt 1792 to 1968. This area has been shown to contain an immunodominant antibody-binding site (*AD-1*) which can be destroyed by insertional mutagenesis of four amino acids at position 616 (58). The nucleotide sequences of the five wild-type strains showed homologies between 94.2 and 100.0% to the sequence of AD169 in this area. This is a higher degree of conservation than was found for *AD-2* and the cleavage region. The degree of conservation was again reflected on the amino acid level, where only few dispersed amino acid mutations (zero to three amino acid residue exchanges) were detected (Fig. 2C). Because of the strong conservation in this region, no classification into AD169 and Towne types, respectively, can be made.

Comparison of sequences from the IMP gene. Less information is available concerning the biological and immunological importance of the IMP (U_L100) than of the gp58/116 for the viral replication cycle. However, there is the following indirect evidence that the IMP has functional relevance. (i) There are homologous open reading frames in the other human herpesviruses, and the genes are located in a conserved arrangement on their respective genomes (29). The homologous protein of HSV-1 (U_L10) has been shown to influence the replication rate of the virus in vitro (3). (ii) The IMP is expressed at late times in the infection and is a structural component of the virion (29). One particular feature of the IMP is a hydrophobic overall structure containing eight hydrophobic regions which presumably are transmembrane domains.

Three areas of the IMP genes have been subjected to sequence comparison (Fig. 1): (i) the amino-terminal part (nt 1 to 156) consisting of the putative signal peptide and 66 5′-nontranslated nucleotides, (ii) a region from the center of the coding sequence (nt 694 to 903) containing three hydrophobic domains and four adjacent hydrophilic motifs, and (iii) the hydrophilic carboxyl terminus (nt 985 to 1119) and additional 26 3′-nontranslated nucleotides. Since the DNA sequence of the IMP gene of strain Towne was not available, the corresponding fragment was amplified from infected cells and sequenced in parallel with the five clinical isolates. To

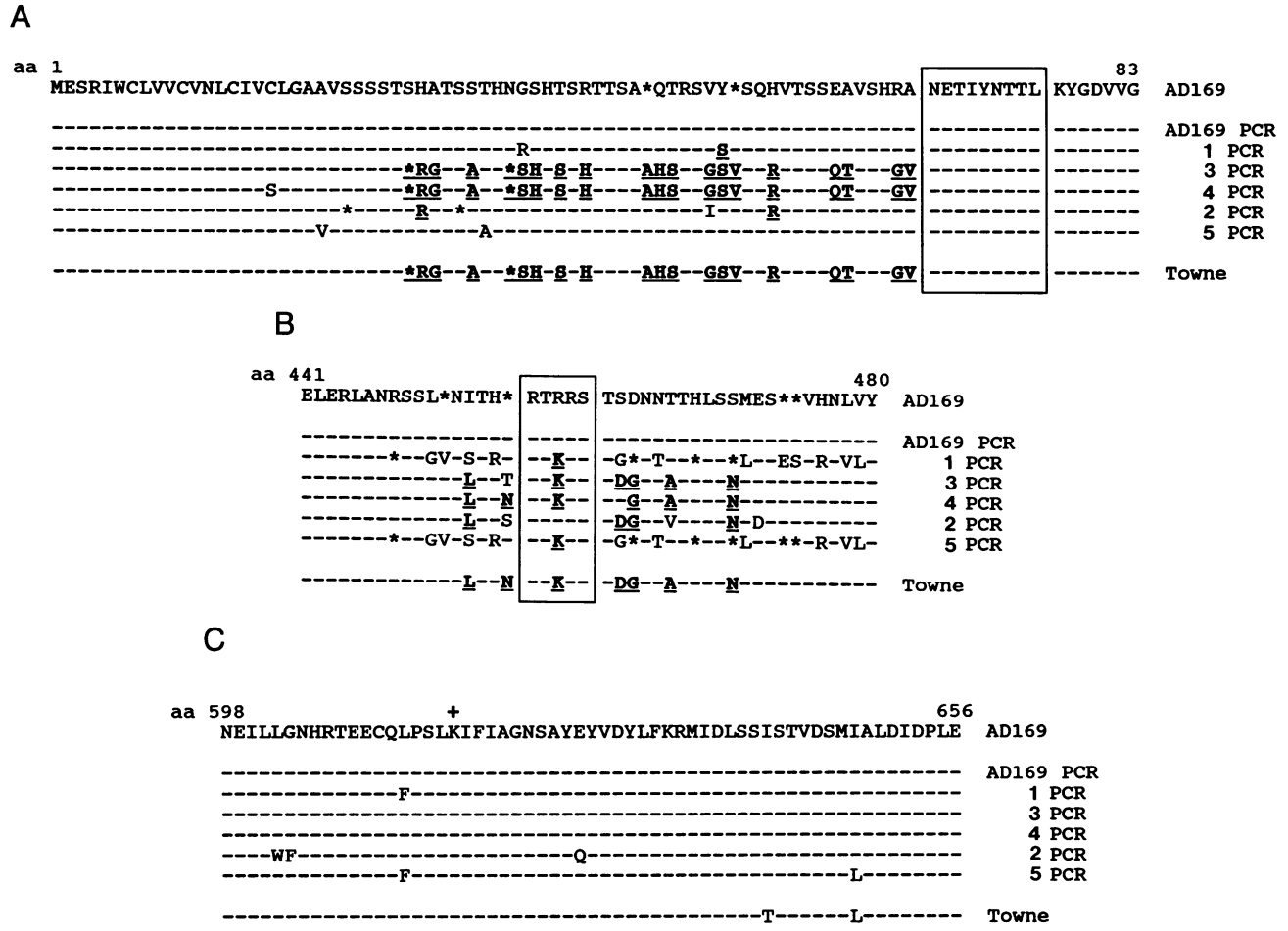

FIG. 2. Amino acid sequence comparison of selected areas within the gp58/116 gene of the laboratory strains AD169 and Towne with those of clinical isolate 1 PCR, 3 PCR, 4 PCR, 2 PCR, and 5 PCR (control sequence, AD169 PCR). (A) *AD-2* region; (B) cleavage site region; (C) *AD-1* region. Dashes indicate sequence identity to strain AD169; only different amino acids are listed. Underlined letters in boldface demonstrate identity to the sequence of strain Towne. Stars symbolize amino acid deletions. The box in panel A indicates the domain which is recognized by a neutralizing monoclonal antibody; the box in panel B shows the recognition motif of a cellular endopeptidase. The cross in panel C denotes the position which is essential for the integrity of the domain *AD-1*.

establish the identity of the Towne strain, the DNA fragment from the *AD-2* region of the gp58/116 was amplified and the DNA sequence was determined. The sequence was identical to the published Towne sequence in this region (data not shown).

The aligned DNA sequences revealed homologies to the AD169 prototype sequence of between 93.4 and 100.0%. This is a smaller range of variability than that of the gp58/116 genes. The isolate 3 sequence showed no differences when compared with the AD169 sequence. Only seven nucleotide exchanges were found in strain 1. Strains 2, 4, and 5 had nearly the same mutation patterns (Fig. 3). A cluster of nucleotide sequence variability was detected in these three strains between nt 863 and 887 (Fig. 3B). The cluster of variation does not correspond to that of the sequence in strain Towne. No significant difference was found in the frequencies of mutations between translated and nontranslated regions (Fig. 3A and C).

A comparison of the deduced amino acid sequences showed high conservation (Fig. 4) because most of the nucleotide exchanges were located in the third positions of the respective codons and therefore represent translationally silent mutations (Fig. 3). However, the cluster of variation detected in strains 2, 4, and 5 (between nt 863 and 887) resulted in six amino acid exchanges and one amino acid insertion (Fig. 3B and 4). This region is located in a hydrophilic motif of the polypeptide chain, which might represent a domain exposed on the surface of the protein. The mutations did not result in a hydrophilicity profile different from that of strain AD169 (data not shown).

Comparison of sequences from the IE-1/2 enhancer/promoter region. It has been shown that differences in the pathogenicity of certain murine leukemia viruses are the result of tissue-specific transcription preference due to variations within the long-terminal-repeat sequences (9). Therefore, the IE-1/2 enhancer/promoter region of HCMV (Fig. 1) was chosen as an important element of viral immediate early gene transcription. It drives expression of IE-1 and IE-2 polypeptides that have been shown to transactivate both homologous and heterologous promoters (31, 39, 54). The IE-1/2 enhancer/promoter sequence has been established for both HCMV AD169 and HCMV Towne (5, 57). It contains

A

```
                                                                M  A  P
CCCGCGAGCCTGTCGTCATCGGCGCGCCCCCATCGCCTCCCGAGCGAGCGGGCCGCCGCTATCGCCATGGCCCCC  IMP N-terminus
                                                                              AD169 PCR
                                                                                 1 PCR
                                                                                 3 PCR
                                                                                 4 PCR
                                                                                 2 PCR
                                                                                 5 PCR
                                                                              Towne PCR

S  H  V  D  K  V  N  T  R  T  W  S  A  S  I  V  F  M  V  L  T  F  V  N  V
TCGCACGTGGATAAGGTGAATACACGGACATGGAGCGCTTCTATCGTTTTCATGGTGCTGACTTTTGTCAACGTC  IMP N-terminus
                                                                              AD169 PCR
                                                                                 1 PCR
                                                                                 3 PCR
                                                                                 4 PCR
                                                                                 2 PCR
                                                                                 5 PCR
                                                                              Towne PCR

S  V  H  L  V  L  S  N  F  P  H  L  G  Y  P  C  V  Y  Y  H  V  V  D  F
AGCGTGCATCTAGTGCTGAGCAATTTTCCGCACCTGGGCTACCCCTGCGTCTACTATCACGTCGTGGACTTT    IMP N-terminus
                                                                              AD169 PCR
          G                                                                      1 PCR
                                                                                 3 PCR
          G                        T                                             4 PCR
          G                                                                      2 PCR
          G                                                                      5 PCR
          G                                                                   Towne PCR
```

B

```
  T  G  H  M  V  L  A  V  F  V  V  Y  A  I  I  S  I  I  Y  F  L  L  I
GTACAGGCCACATGGTGTTAGCCGTCTTCGTGGTCTACGCTATCATCTCCATCATCTACTTTTTACTGAT    IMP-Seq 694 - 903
                                                                              AD169 PCR
                                                                  A              1 PCR
                                                                                 3 PCR
                  G                       A                                    4 PCR
                  G                       A                                    2 PCR
                  G                       A                                    5 PCR
                                                                              Towne PCR

 E  A  V  F  F  Q  Y  V  K  V  Q  F  G  Y  H  L  G  A  F  F  G  L  C
CGAGGCCGTCTTTTTTCAATACGTCAAGGTGCAATTCGGCTACCACCTGGGCGCCTTCTTTGGACTCTGC    IMP-Seq 694 - 903
                                                                              AD169 PCR
                                                                                 1 PCR
                                                                                 3 PCR
  A                       A                             C                      4 PCR
  A                       A                             C                      2 PCR
  A                       A                             C                      5 PCR
                                                                              Towne PCR

G  L  I  Y  P  I  V  Q  Y  D  T  F     L  S  N  E  Y  R  T  G  I  S  W
GGCCTCATCTACCCCATCGTGCAGTACGATACCTT...CCTCAGCAACGAATACCGCACCGGCATCAGCTGGT    IMP-Seq 694 - 903
                                                                              AD169 PCR
                                                                                 1 PCR
                                                                                 3 PCR
                  GCG    TAC A  G    G    T    T                              4 PCR
                  GCG    TAC A  G    G    T    T                              2 PCR
                  GCG    TAC A  G    G    T    T                              5 PCR
                                                                              Towne PCR
```

C

```
S  V  K  Y  Q  A  L  A  T  A  S  G  E  E  V  A  V  L  S  H  H  D  S  L  E  S  R  R
AGTGTCAAGTACCAGGCGCTGGCCACAGCCTCCGGCGAAGAAGTCGCTGTGCTCAGTCACCACGACAGCTTGGAAAGCCGTC  IMP C-terminus
                                                                              AD169 PCR
                                       C                                         1 PCR
                                                                                 3 PCR
                                                      AG                         4 PCR
                                                                                 2 PCR
                                                                                 5 PCR
  T                                    CA                                    Towne PCR

 L  R  E  E  E  D  D  D  D  D  E  D  F  E  D  A  *
GCCTCCGCGAAGAAGAGGACGACGACGATGATGAAGACTTCGAGGACGCTTAACCCCGCCGCCACCCGCACCAGACTTG  IMP C-terminus
                                                                              AD169 PCR
  T                                    G                    AG                1 PCR
                                                                                 3 PCR
                                       G                                         4 PCR
                                       G                                         2 PCR
                  CC                   G                                         5 PCR
                                       G                                      Towne PCR
```

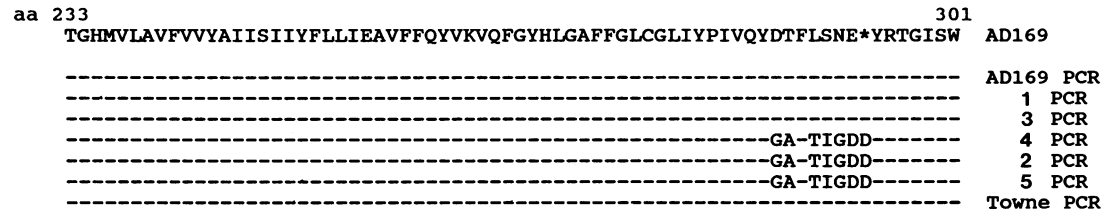

FIG. 4. Amino acid sequence alignment of IMP nucleotide sequences deduced from Fig. 3B. Dashes indicate identity to the AD169 prototype sequence; only the different amino acids are shown. Asterisks indicate deletions of amino acids.

four types of repetitive sequence elements; some are targets for *trans*-acting proteins of the cell. In particular, the palindromic repeat motif of 19 bp and the 18-bp motif bind proteins of the adenovirus transcription factor/cAMP-responsive element binding protein (ATF/CREB) and NF-kappa-B families of transcription factors and are targets for cellular signal transmission pathways which modulate enhancer activities in a cell-type-specific manner (for reviews, see references 53 and 55).

Amplification of the enhancer/promoter region of five wild-type strains together with strain AD169 as a control was achieved by using primers complementary to nt −789 to −770 and nt 85 to 104 relative to the IE-1 transcription start site (Table 1). Comparison with the prototype sequence of strain AD169 revealed a high degree of homology ranging between 97.8 and 98.4%. This was the same degree of homology observed between the strain AD169 and the strain Towne enhancer/promoter regions. The sequences of the wild-type strains differed from each other. In addition to randomly distributed nucleotide exchanges (for example, strain 2, nt 428 [Fig. 5A]), positions could be identified which were altered in several strains (for example, nt 446 and 445 [Fig. 5A] and nt 41 to 39 [Fig. 5B]). However, no classification into subgroups of viruses that have a higher degree of homology to each other within the enhancer/promoter region was possible. Mutations within repetitive sequence elements of the enhancer (for example, nt 110 [Fig. 5B]) did not alter the central core element that is known to be important for protein binding within the 18- and 19-bp elements (17, 55). Variations were always localized in parts of the motif flanking the central core.

DISCUSSION

Considering the complexity of the HCMV genome, comparative sequence analysis of different strains will probably not be practicable in the near future. However, with accumulating data on genomic regions with respect to regulatory or immunological functions, these regions can be selected for sequence analysis. As a first step to address the question whether or not such a limited analysis could produce meaningful results, we have analyzed five clinical isolates in selected genomic areas which represent either regulatory regions or code for immunologically important proteins.

In general, the following two patterns were observed on the DNA level: (i) a high degree of conservation with few scattered point mutations (immediate-early region, most parts of IMP, *AD-1* of gp58/116) and (ii) clusters of highly variable regions (*AD-2* of gp58/116, gp58/116 cleavage region). The mutations have different effects on the encoded amino acid sequences. In regions with high DNA conservation, most mutations are silent. The clustered mutations result in changes in the encoded proteins. The variations are of limited complexity and can be arranged into types, two of which are represented by the laboratory strains AD169 and Towne, respectively. From our analysis, it cannot be decided how many additional types of isolates occur. However, the uniform pattern of variation among isolates suggests a limited number of additional classes of isolates. The classification into types, however, is not uniform within a given isolate. Strain 1, for example, represents an AD169 type in *AD-2*, whereas in the cleavage region it cannot be assigned to an AD169 or Towne type. Another example of an additional, independent type of isolate was found in the variable cluster of IMP.

From previous analyses it is known that *AD-1* and *AD-2* contain sequences capable of inducing neutralizing antibodies and, therefore, represent targets of immunological pressure (26, 37). In this respect it was interesting to observe that the neutralizing epitope in *AD-2* and the entire *AD-1* region are highly conserved among all strains. This is in concordance with immunological assays which have shown that all strains tested so far, including the isolates presented in this study, were reactive with the monoclonal antibodies directed against these epitopes (data not shown). The high degree of conservation could mean that the respective protein domains are of functional importance and that any change is deleterious for viral growth. Another possibility, however less likely, is that these areas are not of significant importance for the humoral immune response during natural infection and therefore do not represent targets for immunological pressure. The neutralizing epitope in *AD-2* is flanked by highly variable sequences. In fact, in strains 3 and 4, the areas of variability with respect to strain AD169 include residues immediately upstream of the epitope. The significance of this divergence is not clear. It does not influence the binding of neutralizing monoclonal antibodies to the isolates. We have, however, evidence that this area is part of an additional antigenic region (36). Major differences between glycoprotein B molecules of HSV-1 and HSV-2, the homologous counterparts to gp58/116, are also found in the amino-terminal portion of the molecule. However, in contrast to the

FIG. 3. Nucleotide sequence comparison of selected areas of the IMP gene. The sequence of the laboratory strain AD169 was aligned with PCR-amplified sequences of strain Towne and of the clinical isolates 1, 3, 4, 2, and 5 (control sequence, AD169 PCR). (A) Amino-terminal end; (B) internal region; (C) carboxy-terminal end. The nucleotides which are different from those of the AD169 prototype sequence are listed. The lines above the AD169 sequences show the corresponding amino acid sequences. Points indicate deletions of nucleotides. Underlined letters show mutations which result in an amino acid alteration.

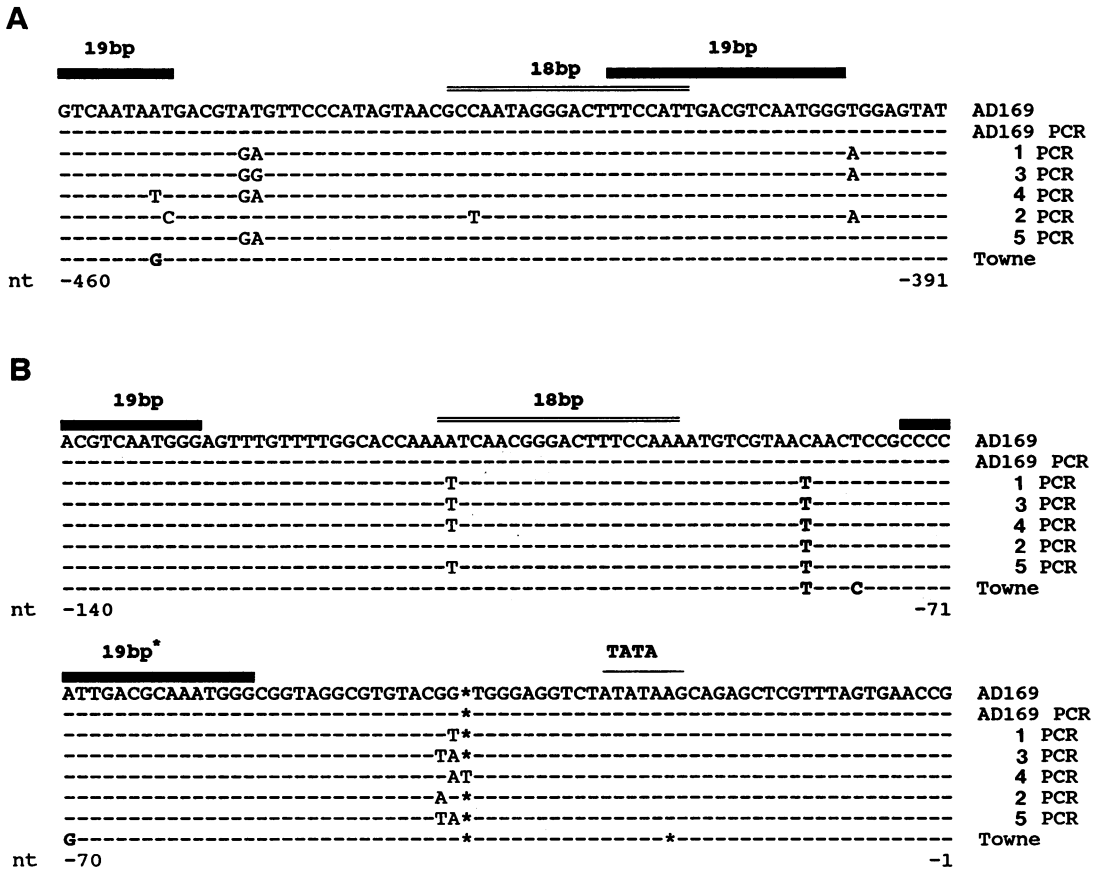

FIG. 5. Nucleotide sequence comparison of selected areas of the IE-1/2 enhancer/promoter of the laboratory strains AD169 and Towne with those of clinical isolates 1 PCR, 3 PCR, 4 PCR, 2 PCR, and 5 PCR (control sequence, AD169 PCR). (A) Region between nt −460 and nt −391; (B) region between nt −141 and −1 (numbering from the cap site of the IE-1 mRNA). Dashes indicate identity to the AD169 prototype sequence; only different nucleotides are shown. Double lines and filled bars above the AD169 sequence symbolize the location of 18- and 19-bp repeat motifs. The TATA box of the IE-1 promoter is indicated.

case with HCMV, the divergence extends into the leader peptide sequence (8, 56).

The significance of the proteolytic cleavage of the gp58/116 precursor for viral replication has not been established. Experiments with synthetic peptide analogs as inhibitors point to an impaired release of virus when processing is inhibited (7). In other systems, such as influenza A virus (60) or human immunodeficiency virus (35), proteolytic processing of envelope glycoproteins is an essential step in the production of infectious virus. On the other hand, cleavage does not occur in gp58/116 analogs of other human herpesviruses such as HSV gB (51) and Epstein-Barr virus gp110 (20). For HCMV Towne, the sequence motif RTKR is essential for proteolytic cleavage. Mutation of any of the basic residues to an uncharged amino acid leads to loss of processing (49). From our analysis we can conclude that both motifs RTRR and RTKR represent functional cleavage sites, regardless of the surrounding residues, since all isolates were shown to produce processed gp58 (data not shown). In this respect it should be noted that sites for N-linked glycosylation (residues, NXT/S) are conserved in this area although mutations occur in some of the motifs. The mutations, however, do recreate potentially functional glycosylation sites. Glycosylation has been shown to be important for the processing of hemagglutinin of influenza A virus (23). The reason for the divergence in the area around the cleavage site is not clear. The region does not contain linear binding sites for human antibodies (35a). It could, however, be part of a conformational epitope and consequently subject to immunological pressure. In this case, antibodies would react with a limited number of isolates, and, in fact, gp58/116-specific monoclonal antibodies with restricted strain recognition have been described (33, 42).

The remaining analyzed areas from the immediate-early region and from IMP were genetically remarkably stable. The conservation of sequences from the IMP was somewhat unexpected since this protein is a target of the host immune response (28a). It is possible, however, that structural constraints of this highly membrane-associated protein do not allow changes in protein composition. It remains to be established whether the identified cluster of variation within an internal hydrophilic domain of the IMP is immunogenic in humans. The high degree of conservation within the regulatory region of the major immediate-early genes fits with the assumption that this element is essential for initiation of the viral replication cycle as it drives the expression of *trans*-acting polypeptides (31, 39, 54). No major deletions or insertions were found within the enhancer/promoter, and the few scattered point mutations did not affect sites that are known from previous analyses to be of functional impor-

tance or to bind cellular transcription factors (17–19, 45). Therefore, it is assumed that the observed sequence variations do not alter the function of the IE-1/2 enhancer/promoter.

The question arises as to whether the method that has been used to generate the data is sufficiently accurate to allow the conclusions that have been drawn. Of course, we cannot definitely exclude the possibility that our sequence data contain errors caused by the *Taq* polymerase. Estimates on error frequencies of the enzyme have ranged from 10^{-4} to 10^{-5} per replication cycle (16). However, we think that for our analysis the error frequency can be neglected for the following reasons: (i) we did not observe nucleotide exchanges when we compared our sequences with published sequences in the controls from AD169- or Towne-infected cells; (ii) we have determined selected sequences from different regions from four independent plasmid clones and did not find differences in the sequences (data not shown).

Therefore, the reported differences between clinical isolates are real and represent characteristic features of the different isolates. The finding of a limited number of related types of isolates is surprising, given the different geographic areas and time of isolation (40, 43). However, since the strains that were used to determine the nucleotide sequence have been passaged in tissue culture, the possibility cannot be excluded that this procedure selects for a limited number of HCMV types. Further analysis of isolates, amplified directly from clinical material, will answer this question. In summary, our data have shown that, on the HCMV genome, regions of both high conservation and of remarkable variability exist. The knowledge of location and structure of variable regions may allow in future studies epidemiological analyses of functionally important domains as well as a correlation to pathogenesis.

ACKNOWLEDGMENTS

We thank H. Moch for excellent technical help and B. Fleckenstein and M. F. Stinski for critically reading the manuscript.

This work was supported by grants from the Bundesministerium für Forschung und Technologie, the Deutsche Forschungsgemeinschaft, and the Wilhelm Sander Stiftung.

REFERENCES

1. **Alford, C. A., K. Hayes, and W. J. Britt.** 1988. Primary cytomegalovirus infection in pregnancy: comparison of antibody responses to virus encoded proteins between women with and without intrauterine infection. J. Infect. Dis. **158**:917–924.
2. **Baboonian, C., K. Blake, J. Booth, and C. N. Wiblin.** 1989. Complement-independent neutralising monoclonal antibody with differential reactivity for strains of human cytomegalovirus. J. Med. Virol. **29**:139–145.
3. **Baines, J. D., and B. Roizmann.** 1991. The open reading frames U_L3, U_L4, U_L10, and U_L16 are dispensable for the replication of herpes simplex virus 1 in cell culture. J. Virol. **65**:938–944.
4. **Borysiewicz, L. K., J. K. Hickling, S. Graham, J. Sinclair, M. P. Cranage, G. L. Smith, and J. G. P. Sissons.** 1988. Human cytomegalovirus-specific cytotoxic T cells. Relative frequency of stage-specific CTL recognizing the 72-kD immediate early protein and glycoprotein B expressed by recombinant vaccinia viruses. J. Exp. Med. **168**:919–931.
5. **Boshart, M., F. Weber, G. Jahn, K. Dorsch-Häsler, B. Fleckenstein, and W. Schaffner.** 1985. A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. Cell **41**:521–530.
6. **Britt, W. J.** 1984. Neutralizing antibodies detect a disulfide linked glycoprotein complex within the envelope of human cytomegalovirus. Virology **135**:369–378.
7. **Brücher, K. H., W. Garten, H. D. Klenk, and K. Radsack.** 1990. Inhibition of endoproteolytic cleavage of cytomegalovirus (HCMV) glycoprotein B by palmitoyl-peptidyl-chloromethyl ketone. Virology **178**:617–620.
8. **Bzik, D. J., C. Debroy, B. A. Fox, N. E. Pederson, and S. Person.** 1986. The nucleotide sequence of the gB glycoprotein of HSV-2 and comparison with the corresponding gene of HSV-1. Virology **155**:322–333.
9. **Celander, D., and W. A. Haseltine.** 1984. Tissue-specific transcription preference as a determinant of cell tropism and leukaemogenic potential of murine retroviruses. Nature (London) **312**:159–162.
10. **Chandler, S. H., and J. K. McDougall.** 1986. Comparison of restriction site polymorphisms among clinical isolates and laboratory strains of human cytomegalovirus. J. Gen. Virol. **67**: 2179–2192.
11. **Chee, M. S., A. T. Bankier, S. Beck, R. Bohni, C. M. Brown, R. Cerny, T. Horsnell, C. A. Hutchinson III, T. Kouzarides, J. A. Martignetti, E. Preddie, S. C. Satchwell, P. Tomlinson, K. M. Weston, and B. G. Barrell.** 1990. Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. Curr. Top. Microbiol. Immunol. **154**:125–169.
12. **Chou, S.** 1990. Differentiation of cytomegalovirus strains by restriction analysis of DNA sequences amplified from clinical specimens. J. Infect. Dis. **162**:738–742.
13. **Condie, R. M., and R. J. O'Reilly.** 1984. Prevention of cytomegalovirus infection in bone marrow transplant recipients by prophylaxis with an intravenous, hyperimmune cytomegalovirus globulin, p. 327–344. *In* S. A. Plotkin et al. (ed.), CMV: pathogenesis and prevention of human infection. Alan R. Liss, Inc., New York.
14. **Cranage, M. P., T. Kouzarides, A. T. Bankier, S. Satchwell, K. Weston, P. Tomlinson, B. Barrel, 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.
15. **Devereux, J.** 1989. Program manual for the sequence analysis software package of the Genetics Computer Group, version 6.0. University of Wisconsin, Madison, Wis.
16. **Eckert, K. A., and T. A. Kunkel.** 1990. High fidelity DNA synthesis by the Thermus aquaticus DNA polymerase. Nucleic Acids Res. **18**:3739–3744.
17. **Fickenscher, H., T. Stamminger, R. Rüger, and B. Fleckenstein.** 1989. The role of a repetitive palindromic sequence element in the human cytomegalovirus major immediate early enhancer. J. Gen. Virol. **70**:107–123.
18. **Ghazal, P., H. Lubon, B. Fleckenstein, and L. Hennighausen.** 1987. Binding of transcription factors and creation of a large nucleoprotein complex on the human cytomegalovirus enhancer. Proc. Natl. Acad. Sci. USA **84**:3658–3662.
19. **Ghazal, P., H. Lubon, and L. Hennighausen.** 1988. Specific interactions between transcription factors and the promoter-regulatory region of the human cytomegalovirus major immediate-early gene. J. Virol. **62**:1076–1079.
20. **Gong, M., T. Ooka, T. Matsuho, and E. Kieff.** 1987. Epstein-Barr virus glycoprotein homologous to herpes simplex virus gB. J. Virol. **61**:499–508.
21. **Hayes, K., C. A. Alford, and W. J. Britt.** 1987. Antibody response to virus encoded proteins after cytomegalovirus mononucleosis. J. Infect. Dis. **156**:615–621.
22. **James, G. S., and M. J. Cloonan.** 1990. Demonstration of antigenic variation among isolates of human cytomegalovirus using monoclonal antibodies and indirect ELISA. J. Virol. Methods **30**:301–310.
23. **Kawaoka, Y., and R. G. Webster.** 1989. Interplay between carbohydrate in the stalk and the length of the connecting peptide determines the cleavability of influenza virus hemagglutinin. J. Virol. **63**:3296–3300.
24. **Kilpatrick, B. A., E.-S. Huang, and J. S. Pagano.** 1976. Analysis of cytomegalovirus genomes with restriction nucleases *Hin*dIII and *Eco*RI. J. Virol. **18**:1095–1105.
25. **Kimmel, K. A., K. E. Dolter, G. M. Toth, M. Levine, and J. C. Glorioso.** 1990. Serologic type conversion of a herpes simplex

virus type 1 (HSV-1) to an HSV-2 epitope caused by a single amino acid substitution in glycoprotein C. J. Virol. **64**:4033–4036.
26. **Kniess, N., M. Mach, J. Fay, and W. J. Britt.** 1991. Distribution of linear antigenic sites on glycoprotein gp55 of human cytomegalovirus. J. Virol. **65**:138–146.
27. **Landini, M. P., G. Mirolo, P. Coppoleccia, M. Re, and M. LaPlaca.** 1986. Serum antibodies to individual cytomegalovirus structural polypeptides in renal transplant recipients during viral infections. Microbiol. Immunol. **30**:683–695.
28. **Landini, M. P., M. C. Re, G. Mirolo, G. Baldassarri, and M. LaPlaca.** 1985. Human immune response to cytomegalovirus structural polypeptides studied by immunoblotting. J. Med. Virol. **17**:303–311.
28a. **Lehner, R.** Unpublished results.
29. **Lehner, R., H. Meyer, and M. Mach.** 1989. Identification and characterization of a human cytomegalovirus gene coding for a membrane protein that is conserved among human herpesviruses. J. Virol. **63**:3792–3800.
30. **Mach, M., U. Utz, and B. Fleckenstein.** 1986. Mapping of the major glycoprotein gene of human cytomegalovirus. J. Gen. Virol. **67**:1461–1467.
31. **Malone, C. L., D. H. Vesole, and M. F. Stinski.** 1990. Transactivation of a human cytomegalovirus early promoter by gene products from the immediate-early gene IE2 and augmentation by IE1: mutational analysis of the viral proteins. J. Virol. **64**:1498–1506.
32. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
33. **Masuho, Y., Y.-I. Matsumoto, T. Sugano, S. Fujinaga, and Y. Minamishima.** 1987. Human monoclonal antibodies neutralizing human cytomegalovirus. J. Gen. Virol. **68**:1457–1461.
34. **Matteucci, M. D., and M. H. Caruthers.** 1981. Synthesis of deoxynucleotides on a polymer support. J. Am. Chem. Soc. **103**:3185–3191.
35. **McCune, J., L. B. Rabin, M. F. Feinberg, M. Lieberman, J. C. Kosek, G. R. Reyes, and I. L. Weissman.** 1988. Endoproteolytic cleavage of gp160 is required for the activation of human deficiency virus. Cell **53**:55–67.
35a. **Meyer, H.** Unpublished results.
36. **Meyer, H., and M. Mach.** 1991. The glycoprotein gp116 of human cytomegalovirus contains two antibody binding sites in close proximity. J. Cell. Biochem. (Suppl.)**15E**:101.
37. **Meyer, H., Y. Masuho, and M. Mach.** 1990. The gp116 of the gp58/116 complex of human cytomegalovirus represents the amino-terminal part of the precursor molecule and contains a neutralizing epitope. J. Gen. Virol. **71**:2443–2450.
38. **Onorato, I., D. Morens, W. Martone, and S. Stansfield.** 1985. Epidemiology of cytomegalovirus infections: recommendations for prevention and control. Rev. Infect. Dis. **7**:479–497.
39. **Pizzorno, M. C., P. O'Hare, L. Sha, R. L. LaFemina, and G. S. Hayward.** 1988. Transactivation and autoregulation of gene expression by the immediate-early region 2 gene products of human cytomegalovirus. J. Virol. **62**:1167–1179.
40. **Plotkin, S. A., T. Furukawa, N. Zygraich, and C. Huygelen.** 1975. Candidate cytomegalovirus strain for human vaccination. Infect. Immun. **12**:521–527.
41. **Quinnan, G. V., Jr., N. Kirmani, A. H. Rook, J. F. Manischewitz, L. Jackson, G. Moreschi, G. W. Santos, R. Saral, and W. B. Burns.** 1982. HLA-restricted T-lymphocyte cytotoxic responses correlate with recovery from cytomegalovirus infection in bone-marrow-transplant recipients. N. Engl. J. Med. **307**:7–13.
42. **Rasmussen, L., J. Mullenax, R. Nelson, and T. Merigan.** 1985. Viral polypeptides detected by complement-dependent neutralizing murine monoclonal antibody to human cytomegalovirus. J. Virol. **55**:274–280.
43. **Rowe, W. P., J. W. Hartley, S. Waterman, H. C. Turner, and R. J. Huebner.** 1956. Cytopathic agent resembling human salivary gland virus recovered from tissue cultures of human adenoids. Proc. Soc. Exp. Biol. Med. **92**:418.
44. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1987. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**:487–491.
45. **Sambucetti, L. C., J. M. Cherrington, G. W. G. Wilkinson, and E. S. Mocarski.** 1989. NF-kB activation of the cytomegalovirus enhancer is mediated by a viral transactivator and by T cell stimulation. EMBO J. **8**:4251–4258.
46. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
47. **Shioda, T., J. A. Levi, and C. Cheng-Mayer.** 1991. Macrophage and T cell-line tropisms of HIV-1 are determined by specific regions of envelope gp120 gene. Nature (London) **349**:167–169.
48. **Snydman, D. R., B. G. Werner, B. Heinze-Lacey, and G. F. Grady.** 1987. Use of cytomegalovirus immune globulin to prevent cytomegalovirus disease in renal transplant recipients. N. Engl. J. Med. **317**:1049–1054.
49. **Spaete, R. R., A. Saxena, P. I. Scott, G. J. Song, W. S. Probert, W. J. Britt, W. Gibson, L. Rasmussen, and C. Pachl.** 1990. Sequence requirements for proteolytic processing of glycoprotein B of human cytomegalovirus strain Towne. J. Virol. **64**:2922–2931.
50. **Spaete, R. R., R. M. Thayer, W. S. Probert, F. R. Masiarz, S. H. Chamberlain, L. Rasmussen, T. C. Merrigan, and C. Pachl.** 1988. Human cytomegalovirus strain Towne glycoprotein B is processed by proteolytic cleavage. Virology **167**:207–225.
51. **Spear, P.** 1985. Glycoproteins specified by herpes simplex virus, p. 315–356. *In* B. Roizman (ed.), The herpesviruses, vol. 3. Plenum Press, London.
52. **Stagno, S., R. F. Pass, E. D. Meyer, W. J. Britt, and C. Alford.** 1984. Congenital and perinatal cytomegalovirus infections: clinical characteristics and pathogenic factors, p. 65–85. *In* S. A. Plotkin et al. (ed.), CMV: pathogenesis and prevention of human infection. Alan R. Liss, Inc., New York.
53. **Stamminger, T., and B. Fleckenstein.** 1990. Immediate-early transcription regulation of human cytomegalovirus. Curr. Top. Microbiol. Immunol. **154**:3–19.
54. **Stenberg, R. M., J. Fortney, S. W. Barlow, B. P. Magrane, J. A. Nelson, and P. Ghazal.** 1990. Promoter-specific *trans* activation and repression by human cytomegalovirus immediate-early proteins involves common and unique protein domains. J. Virol. **64**:1556–1565.
55. **Stinski, M. F., C. L. Malone, T. W. Hermiston, and B. Liu.** 1991. Regulation of human cytomegalovirus transcription, p. 246–260. *In* E. K. Wagner (ed.), Herpesvirus transcription and regulation. CRC Press, Inc., Boca Raton, Fla.
56. **Stuve, L. L., S. Brown-Shimer, C. Pachl, R. Najarian, D. Dina, and R. L. Burke.** 1987. Structure and expression of the herpes simplex virus type 2 glycoprotein B gene. J. Virol. **61**:326–335.
57. **Thomsen, D. R., R. M. Stenberg, W. F. Goins, and M. F. Stinski.** 1984. Promoter-regulatory region of the major immediate early gene of human cytomegalovirus. Proc. Natl. Acad. Sci. USA **81**:659–663.
58. **Utz, U., W. Britt, L. Vugler, and M. Mach.** 1989. Identification of a neutralizing epitope on glycoprotein gp58 of human cytomegalovirus. J. Virol. **63**:1995–2001.
59. **Waner, J. L., and T. H. Weller.** 1978. Analysis of antigenic diversity among human cytomegaloviruses by kinetic neutralization tests with high-titered rabbit antisera. Infect. Immun. **21**:151–157.
60. **Webster, R. G., and R. Rott.** 1987. Influenza virus A pathogenicity: the pivotal role of hemagglutinin. Cell **50**:665–666.
61. **Zaia, J. A., S. J. Foreman, Y. P. Ting, E. Vanderwal-Urbina, and K. G. Blume.** 1986. Polypeptide specific antibody response to human cytomegalovirus after infection in bone marrow transplant recipients. J. Infect. Dis. **153**:780–786.
62. **Zaia, J. A., G. Gallez-Hawkins, M. A. Churchill, A. Morton-Blackshere, H. Pande, S. P. Adler, G. M. Schmidt, and S. J. Foreman.** 1990. Comparative analysis of human cytomegalovirus *a*-sequence in multiple clinical isolates by using polymerase chain reaction and restriction fragment length polymorphism assays. J. Clin. Microbiol. **28**:2602–2607.