JOURNAL OF VIROLOGY, Jan. 1987, p. 125–133
0022-538X/87/010125–09$02.00/0

Vol. 61, No. 1

Sequence and Transcription Analysis of the Human Cytomegalovirus DNA Polymerase Gene

TONY KOUZARIDES,†* ALAN T. BANKIER, SANDRA C. SATCHWELL, KATHLEEN WESTON, PETER TOMLINSON, AND BARCLAY G. BARRELL

Medical Research Council, Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

Received 19 May 1986/Accepted 8 September 1986

DNA sequence analysis has revealed that the gene coding for the human cytomegalovirus (HCMV) DNA polymerase is present within the long unique region of the virus genome. Identification is based on extensive amino acid homology between the predicted HCMV open reading frame HFLF2 and the DNA polymerase of herpes simplex virus type 1. We present here a 5280-base-pair DNA sequence containing the HCMV *pol* gene, along with the analysis of transcripts encoded within this region. Since HCMV *pol* also shows homology to the predicted Epstein-Barr virus *pol*, we were able to analyze the extent of homology between the DNA polymerases of three distantly related herpesviruses, HCMV, Epstein-Barr virus, and herpes simplex virus. The comparison shows that these DNA polymerases exhibit considerable amino acid homology and highlights a number of highly conserved regions; two such regions show homology to sequences within the adenovirus type 2 DNA polymerase. The HCMV *pol* gene is flanked by open reading frames with homology to those of other herpesviruses; upstream, there is a reading frame homologous to the glycoprotein B gene of herpes simplex virus type 1 and Epstein-Barr virus, and downstream there is a reading frame homologous to BFLF2 of Epstein-Barr virus.

Human cytomegalovirus (HCMV) is a member of the herpesvirus family and ubiquitous in human populations. Although most HCMV infections do not lead to any apparent disease, under certain circumstances HCMV infection can have serious consequences. The virus is known to be a major cause of congenital abnormalities (44), and HCMV-induced disease can cause severe problems in immunosuppressed individuals, such as transplant recipients (15).

The search for therapeutic agents to be used in the treatment of herpesvirus infections has yielded a number of nucleoside analogs which are able to preferentially inhibit the virus-encoded DNA polymerase. In herpes simplex virus (HSV), this selectivity is apparently due to the activation of these agents by the HSV-encoded thymidine kinase (10). HCMV is thought not to possess a thymidine kinase gene (11), which may explain why some nucleoside analogs are less effective against it. However, certain nucleoside analogs do inhibit HCMV replication, although their mode of activation is still obscure (13, 24–26, 42).

Since the herpesvirus DNA polymerase plays an important role in the mechanism of action of antiherpetic agents, the functional characterization of this enzyme is of interest. Although the HCMV DNA polymerase (19) has not been studied as extensively as that of HSV, the two virus polymerases are known to possess a number of characteristics in common: their polymerase activity is stimulated severalfold by high concentrations of salt, they both possess a 3′-to-5′ exonuclease activity, and they are both sensitive to certain drugs such as phosphonoacetic acid and aphidicolin (21, 29). The HSV DNA polymerase is a 150-kilodalton (kDa) protein which copurifies with the 54-kDa major DNA-binding protein (32, 43). Purified HCMV DNA polymerase consists of two polypeptides of 140 and 58 kDa (25), suggesting that the larger of the two polypeptides is probably the DNA polymerase.

We report here the sequence and transcription analysis of an HCMV open reading frame (HFLF2) which shows extensive amino acid homology to the polymerase of HSV type 1 (HSV-1) and Epstein-Barr virus (EBV). The HCMV *pol* gene is present within the long unique region of the virus genome and is found adjacent to genes conserved in other herpesviruses.

MATERIALS AND METHODS

DNA sequencing and data analysis. The *Hin*dIII F fragment of HCMV was isolated from a pAT153 recombinant plasmid (30) and sonicated to produce randomly sheared fragments (8). These were then made blunt-ended with T4 DNA polymerase and cloned into the *Sma*I site of bacteriophage M13mp8 vector (28). Single-stranded DNA templates were prepared as described previously (36) and sequenced by the dideoxynucleotide chain termination method (35, 37) with a complementary synthetic oligonucleotide primer (9). Fractionation of the single-stranded DNA products of the primer elongation reactions was carried out on TBE buffer gradient polyacrylamide gels (3). The sequences were aligned and overlapped by computer (39), and in this way the complete sequence of both DNA strands was determined. Analysis of open reading frames and signal sequences was carried out with the program ANALYSEQ (40). The program DIAGON (38) was used to identify homologous sequences.

Virus and cell culture. HCMV strain AD169 was obtained from J. Oram, Centre for Applied Microbiology and Research, Porton Down, United Kingdom, and propagated in diploid human embryo lung (MRC-5) cells. To prepare immediate-early infected-cell RNA, 80% confluent monolayers of MRC-5 cells were incubated for 1 h before infection with fresh medium containing cycloheximide (100 μg/ml) and then infected with HCMV at 10 PFU/cell. After infection the cells were maintained in medium containing cycloheximide (100 μg/ml), and RNA was prepared at 12 h postinfection.

* Corresponding author.
† Present address: Department of Biochemistry, New York University Medical Center, New York, NY 10016.

For early RNA, 80% confluent monolayers of MRC-5 cells were infected with HCMV at 10 PFU/cell; the cells were maintained in medium containing phosphonoacetic acid (100 μg/ml), and RNA was prepared at 23 h postinfection. For late RNA, MRC-5 cells were infected with HCMV at 10 PFU/cell, and RNA was prepared at 90 h postinfection.

RNA preparation. HCMV-infected MRC-5 cells were harvested by trypsinization and centrifugation, washed in phosphate-buffered saline, and then lysed in 10 volumes of 0.1 M NaCl–0.01 M Tris (pH 7.5)–1 mM EDTA–0.5% Nonidet P-40. The nuclei were removed by centrifugation at 10,000 rpm for 10 min. The supernatant was then made 1% with sodium dodecyl sulfate (SDS) and extracted three times with phenol-chloroform (1:1, vol/vol) and once with chloroform. Total cytoplasmic RNA was then recovered by ethanol precipitation at −20°C.

Northern blot analysis. RNA samples (10 μg) dissolved in MOPS buffer (20 mM MOPS [morpholinepropanesulfonic acid], 1 mM EDTA, 5 mM sodium acetate) containing 50% formamide and 2 M formaldehyde were incubated at 60°C for 5 min and then applied to a 1.5% agarose gel prepared in MOPS buffer with 2 M formaldehyde. Electrophoresis was done at 10 V/cm for 2.5 h in MOPS buffer. Human 18S and 28S rRNA (1.74 and 4.85 kilobases [kb], respectively) were used as markers. Following electrophoresis, the RNA was transferred to nitrocellulose membranes as described by Thomas (41). RNA blots were prehybridized in 5× Denhardt solution–5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sarcosyl–100 μg of denatured salmon sperm DNA per ml at 65°C for at least 1 h. Hybridization of single-stranded "prime-cut" probes to the RNA blots was carried out in 1× Denhardt solution–5× SSC–0.1 sarcosyl–100 μg of denatured salmon sperm DNA per ml at 65°C for 16 h. Blots were washed at 65°C first in 1× SSC–0.1% sarcosyl and then in several changes of 0.1× SSC–0.1% sarcosyl.

Prime-cut probes. The single-stranded probes used in RNA mapping experiments (Table 1) were prepared by the prime-cut method (12). About 0.5 μg of template DNA (3 μl of single-stranded M13 clone DNA) was added to 2 μl of sequencing primer (0.2 pmol) and 0.5 μl of TM buffer (100 mM Tris, pH 8.0, 50 mM $MgCl_2$) and incubated in a 60°C oven for 30 min. The primer annealed to the template DNA was then extended by a 10-min room temperature incubation with 5 U (1 μl) of Klenow DNA polymerase, in the presence of 2.5 μl of 0.5 mM each dGTP, dCTP, and dTTP and 2 μl of [α-^{32}P]dATP (10 mCi/ml, 800 Ci/mmol). After the addition of 1 μl of 0.5 mM each dATP, dCTP, dGTP, and dTTP and further incubation for 10 min, the enzyme was heat killed by a 10-min incubation at 70°C. The primer-extended M13 clone was then digested with the appropriate restriction enzyme in a total volume of 20 μl. After the addition of 25 μl of formamide containing 0.1% bromophenol blue and 0.1% xylene cyanol FF, the mixture was denatured by a 3-min incubation at 100°C and then electrophoresed through a 6% polyacrylamide–8 M urea sequencing gel. The single-stranded radioactively labeled probe was then located by autoradiography, excised, and eluted from the gel by incubating the gel slice in elution buffer (0.5 M ammonium acetate, 10 mM $MgCl_2$, 1 mm EDTA, 0.1% SDS) for 2 h in a 37°C orbital shaker.

TABLE 1. Probes used in transcript mapping experiments

Probe[a]	HCMV sequence positions[b]	Total length (bases)[c]
F653/*Aat*II	194–513	367
F161/*Aat*II	194–557	411
F653/*Nsi*I	295–513	266
F209/*Eco*RI	~3512–4212	~751
F74/*Eco*RI	~4627–5127	~551
F74/*Sac*I	4653–5127	522
F228/*Tth*III-I	4784–5166	430
F599/*Eco*RI	5042–5272	751

[a] Prime-cut probes are made by primed second-strand synthesis on M13, followed by restriction of the resulting double-stranded molecule by an appropriate enzyme. Each probe is therefore named after the M13 clone used for the primed second-strand synthesis and the enzyme used to cut it.

[b] HCMV sequence included in each probe. All probes were in the rightward orientation. See Fig. 2.

[c] Total length of each probe includes 48 bases of M13 sequence (51 bases in the *Eco*RI-cut probes) in addition ot the HCMV sequences.

S1 nuclease mapping. Approximately 50,000 cpm of the prime-cut probe was annealed to 10-μg samples of total cytoplasmic RNA in 30 μl of hybridization buffer (80% deionized formamide, 0.4 M NaCl, 10 mM PIPES [piperazine-*N*,*N*'-bis(2-ethanesulfonic acid), pH 6.5], 1 mM EDTA) and placed in a 55°C water bath; the temperature was then reduced to 42°C, and the mixture was incubated at this temperature for 16 h. To the annealed mixture was added 200 μl of S1 buffer (30 mM sodium acetate, pH 4.5, 0.25 mM NaCl, 1 mM $ZnSO_4$, 5% glycerol) followed by ca. 100 U of S1 nuclease, and the mixture was incubated at 15°C for 1 h. The S1 nuclease-resistant fragments were precipitated with ethanol, denatured in 4 μl of formamide (containing 0.1% bromophenol blue and 0.1% xylene cyanol), and electrophoresed through a 6% acrylamide–8 M urea sequencing gel.

Primer extension analysis. The procedure used for primer extension analysis was essentially that described by Bina-Stein et al. (4). The primer (50,000 cpm) was annealed to 10-μg samples of total cytoplasmic RNA as described for S1 nuclease mapping. The annealed mixture was then ethanol precipitated and redissolved in 10 μl of a buffer containing 0.1 M Tris, pH 8.3, 0.15 M KCl, 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM each dATP, dGTP, dCTP, and dTTP, and 60 μg of actinomycin D per ml. The primer was extended with 5 U of reverse transcriptase for 30 min at 42°C. The resultant extensions were electrophoresed through 6% acrylamide–8 M urea sequencing gels as described for S1 nuclease mapping.

RESULTS

Sequence analysis. The seqence of the *Hin*dIII F fragment of HCMV stain AD169 (Fig. 1) revealed the presence of an open reading frame showing extensive amino acid homology to the HSV-1 DNA polymerase (14, 34) and to reading frame BALF5 of EBV (2). A 5,280-base-pair (bp) sequence encompassing the HCMV DNA polymerase (*pol*) gene is presented in Fig. 2. The ATG codon at position 660 marks the beginning of the 3,726-bp open reading frame identified as the HCMV *pol*. This frame terminates at position 4385. Two polyadenylation signal sequences, AATAAA (33), are present downstream of the *pol* reading frame, at positions 4962 to 4967 and 5031 to 5036.

The *pol* gene is bound by two reading frames which also have homologous counterparts in other herpesviruses; part of the sequence giving rise to these frames is presented in Fig. 2. Upstream of *pol* (and in the same direction) there is an uninterrupted reading frame extending from the beginning of the sequence in Fig. 2 to a termination codon at position 517. This reading frame shows homology to glycoprotein B of HSV-1 (6) and reading frame BALF4 of EBV (2, 31).

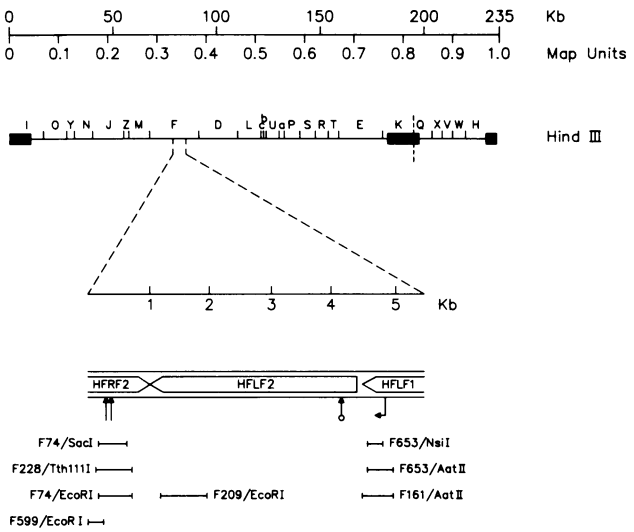

FIG. 1. Position of the HFLF2 reading frame within the HCMV genome. Arrows indicate polyadenylation sites of the type AATAAA, and the arrow with a circle indicates a polyadenylation site of the type ATTAAA. The bent arrow represents a promoter. The positions of the prime-cut probes used in the transcription analysis are indicated below the HFLF2 reading frame; see Table 1 for their precise positions.

Downstream from the *pol* gene there is another uninterrupted reading frame extending from the end of the sequence shown in Fig. 2 to a termination codon at position 4368; this reading frame is therefore arranged in a tail-to-tail configuration with respect to *pol*, the 3′ termini of the two polypeptides overlapping by 5 amino acids. This reading frame shows homology to BFLF2 of EBV (20). The full sequence of these two additional genes will be presented elsewhere (manuscript in preparation).

Transcript analysis. All the single-stranded probes used in the transcription analysis described below were made by the prime-cut method (12). The accuracy of the bands generated by S1 nuclease and primer extension analysis is estimated to be within 5 bases of the size shown; the sizes of the smaller S1-protected fragments are taken as being most accurate.

The 5′ terminus of the *pol* mRNA was mapped with probes F161/*Aat*II and F653/*Aat*II. The major S1-resistant fragment of 281 bases generated by probe F161/*Aat*II (Fig. 3b) correlates with the 237-base S1-resistant fragment produced by F653/*Aat*II (Fig. 3a), indicating the presence of a 5′ RNA-DNA discontinuity at position 276. The potential TATA box sequence (7), TATAATT (positions 250 to 256), present about 20 bp upstream of this point suggests that this S1-mapped end may represent the 5′ terminus of an early RNA. It is worth noting that in the above S1 experiment (Fig. 3b) a less abundant S1-resistant fragment can be detected within several nucleotides of the major protected fragment, suggesting that multiple initiation sites may exist in this region. To confirm the existence of a major 5′ RNA terminus at position 276, a primer extension experiment was carried out. Probe F653/*Nsi*I (total length 266 bases, including M13 sequence) was annealed to total cytoplasmic early HCMV RNA and extended with reverse transcriptase. The 282-base extension product (Fig. 3e) is consistent with the presence of a 5′ RNA terminus at position 276.

There are two closely situated polyadenylation signal sequences following the *pol* reading frame. The first one is present 575 bp downstream of the *pol* reading frame (position 4962 to 4967), and the second one is 64 bp further on, at position 5031 to 5036. Two prime-cut probes, whose sequence extends over both polyadenylation signal sequences, were used in S1 nuclease analysis to establish whether any transcripts terminate after either of these sites. Probe F228/*Tth*III-I generated four S1-resistant fragments of 266, 256, 242, and 176 bases (Fig. 3c), whereas probe F74/*Sac*I gave S1-resistant fragments of 397, 387, 373, and 307 bases (Fig. 3d). These results are interpreted as follows. The S1-resistant fragments of 266 and 397 bases identify a 3′ RNA-DNA discontinuity at position 5050. This falls 14 bases downstream of the polyadenylation signal sequence at position 5031 to 5036, confirming that this is a 3′ RNA terminus. The S1-resistant fragments of 256 and 387 bases identifies a heterogeneous 3′ RNA-DNA discontinuity around position 5040. This falls 4 bases downstream of the polyadenylation signal sequence at 5031 to 5036 and therefore represents the 3′ terminus of an abundant late RNA; the AT-rich sequence in this area is probably responsible for local melting and subsequent S1 nuclease digestion, resulting in the heterogeneity seen in this region (Fig. 3c). The S1-resistant fragments of 242 and 373 bases identify a 3′ RNA-DNA discontinuity at position 5026. The presence of a splice donor-like sequence around this point (position 5026 to 5034) suggests that this may represent a splice donor site of an abundant late RNA species. Finally, the S1-resistant fragments of 176 and 307 bases identifies a 3′ RNA-DNA discontinuity at position 4959. A sequence showing significant identity to a splice donor consensus sequence is present at this point (position 4957 to 4965).

To identify the transcripts originating from the *pol* gene, prime-cut probe F209/*Eco*RI, containing about 700 bases of the *pol* coding sequence, was hybridized to a Northern blot of total cytoplasmic RNA prepared from HCMV-infected cells. Three transcripts of 1.6, 4.7, and about 7.5 kb can be detected (Fig. 4a). The sizing of the 7.5-kb transcript is only an approximation, since it could not be accurately sized by the markers used in the experiment. The expression of these RNAs is relatively low at early times, but increases late in infection. All three of the above RNAs plus an RNA larger than the ~7.5-kb transcript also hybridize to probe F74/*Eco*RI; this probe begins approximately 220 bp downstream of the *pol* reading frame and extends to position 5112, exactly 76 bases downstream of the polyadenylation site at 5031 to 5036 (Fig. 4b). However, none of the above RNAs can be detected with probe F599/*Eco*RI, which begins 5 bp downstream of this polyadenylation signal sequence (Fig. 4c), suggesting that all these RNAs terminate at or before this point. The abundance of the four transcripts detected in Northern blots correlate with the abundance of the four S1-resistant fragments generated by probes which span the two polyadenylation signal sequences (Fig. 3c and d).

The low-abundance transcript of 4.7 kb most likely represents the RNA coding for the DNA polymerase for the following reasons. First, the distance between the *pol* promoter and the polyadenylation site at 5031 to 5036 is 4,755 bp; second, the low abundance of the 4.7-kb transcript is consistent with the abundance of the early RNA expressed from the *pol* promoter; third, the two transcripts of over 7.5 kb hybridize to sequences upstream of the *pol* promoter (data not shown); fourth, the 1.6-kb transcript is too small to encode the *pol* polypeptide. It is unclear where the 3′ end of the *pol* mRNA lies. It may be at position 5050 (14 bp downstream of the polyadenylation site at 5031), since a low-abundance transcript has been mapped at this point.

```
GCGGTGGCCTCCGTGGTCGAAGGCGTTGCCACCTTCCTCAAAAACCCCTTCGGAGCCTTCACCATCATCCTCGTGGCCATAGCCGTAGTCATTATCACTTATTTGATCTATACTCGACAG  120

CGGCGTCTGTGCACGCAGCCGCTGCAGAACCTCTTTCCCTATCTGGTGTCCGCCGACGGGACCACCGTGACGTCGGGCAGCACCAAAGACACGTCGTTACAGGCTCCGCCTTCCTACGAG  240

GAAAGTGTTTATAATTCTGGTCGCAAAGGACCGGGACCACCGTCGTCTGATGCATCCACGGCGGCTCCGCCTTACACCAACGAGCAGGCTTACCAGATGCTTCTGGCCCTGGCCCGTCTG  360

GACGCAGAGCAGCGAGCGCAGCAGAACGGTACAGATTCTTTGGACGGACAGACTGGCACGCAGGACAAGGGACAGAAGCCTAACCTGCTAGACCGGCTGCGACATCGCAAAAACGGCTAC  480

AGACACTTGAAAGACTCCGACGAAGAAGAGAACGTCTGAACCAGGAGGAAAAAAAAACTAGACAAAAAATATTGACACAGAGACTTGTGATATACGGGGTTAAACTGGATATCTAGGTGC  600

                                                                M  F  F  N  P  Y  L  S  G  G  V  T  G  G  A  V  A  G  G  R   20
TGCATGTGTATTTTCTTTGTGATTTTGCTTCGTAAGCTGTCAGCCTCTCACGGTCCGCTATGTTTTTCAACCCGTATCTGAGCGGCGGCGTGACCGGCGGTGCGGTGCGGTGGCCGGC  720

R  Q  R  S  Q  P  G  S  A  Q  G  S  G  K  R  P  P  Q  K  Q  F  L  Q  I  V  P  R  G  V  M  F  D  G  Q  T  G  L  I  K  H   60
GTCAGCGTTCGCAGCCCGGCTCCGCGCAGGGCTCGGGCAAGCGGCCGCCACAGAAACAGTTTTTGCAGATCGTGCCGCGAGGTGTCATGTTCGACGGTCAGACGGGGTTGATCAAGCATA  840

K  T  G  R  L  P  L  M  F  Y  R  E  I  K  H  L  L  S  H  D  W  V  W  P  C  P  W  R  E  T  L  V  G  R  V  V  G  P  I  R  100
AGACGGGACGGCTGCCTCTCATGTTCTATCGAGATTAAACATTTGTTGAGTCATGACATGGTTTGGCCGTGTCCTTGGCGCGAGACCCTGGTGGGTCGCGTGGTGGGACCTATTCGTT  960

F  H  T  Y  D  Q  T  D  A  V  L  F  F  D  S  P  E  N  V  S  P  R  Y  R  Q  H  L  V  P  S  G  N  V  L  R  F  F  G  A  T  140
TTCACACCTACGATCAGACGGACGCCGTGCTCTTCTTCGACTCGCCCGAAAACGTGTCGCCGCGCTATCGTCAGCATCTGGTGCCTTCGGGGAACGTGTTGCGTTTCTTCGGGGCCACAG 1080

E  H  G  Y  S  I  C  V  N  V  F  G  Q  R  S  Y  F  Y  C  E  Y  S  D  T  D  R  L  R  E  V  I  A  S  V  G  E  L  V  P  E  180
AACACGGCTACAGTATCTGCGTCAACGTTTTCGGGCAGCGCAGCTACTTTTACTGTGAGTACAGCGACACCGATAGGCTGCGTGAGGTCATTGCCAGCGTGGGCGAACTAGTGCCCGAAC 1200

P  R  T  P  Y  A  V  S  V  T  P  A  T  K  T  S  I  Y  G  Y  G  T  R  P  V  P  D  L  Q  C  V  S  I  S  N  W  T  M  A  R  220
CGCGGACGCCATACGCCGTGTCTGTCACGCCGGCCACCAGACCTCCATCTATGGGTACGGGACGCGACCCGTGCCCGATTTGCAGTGTGTGTCTATCAGCAACTGGACCATGGCCAGAA 1320

K  I  G  E  Y  L  L  E  Q  G  F  P  V  Y  E  V  R  V  D  P  L  T  R  L  V  I  D  R  R  I  T  T  F  G  W  C  S  V  N  R  260
AAATCGGCGAGTATCTGCTGGAGCAGGGTTTCCCGTGTACGAGGTCCGTGTGGATCCGCTGACGCGTTTGGTCATCGATCGGCGGATCACCACGTTCGGCTGGTGCTCCGTGAATCGTT 1440

Y  D  W  R  Q  Q  G  R  A  S  T  C  D  I  E  V  D  C  D  V  S  D  L  V  A  V  P  D  D  S  S  W  P  R  Y  R  C  L  S  F  300
ACGACTGGCGGCAGCAGGGTCGCGCGTCGACTTGTGATATCGAGGTAGACTGCGATGTCTCTGACCTGGTGGCTGTGCCCGACGACAGCTCGTGGCCGCGCTATCGATGCCTGTCCTTCG 1560

D  I  E  C  M  S  G  E  G  G  F  P  C  A  E  K  S  D  D  I  V  I  Q  I  S  C  V  C  Y  E  T  G  G  N  T  A  V  D  Q  G  340
ATATCGAGTGCATGAGCGGCGAGGGTGGTTTTCCCTGCGCCGAGAAGTCCGATGACATTGTCATTCAGATCTCGTGCGTGTGCTACGAGACGGGGGGAAACACCGCCGTGGATCAGGGGA 1680

I  P  N  G  N  D  G  R  G  C  T  S  E  G  V  I  F  G  H  S  G  L  H  L  F  T  I  G  T  C  G  Q  V  G  P  D  V  D  V  Y  380
TCCCAAACGGGAACGATGGTCGGGGCTGCACTTCGGAGGGTGTGATCTTTGGGCACTCGGGTCTTCATCTCTTTACGATCGGCACCTGCGGGCAGGTGGGCCCAGACGTGGACGTCTACG 1800

E  F  P  S  E  Y  E  L  L  L  G  F  M  L  F  F  Q  R  Y  A  P  A  F  V  T  G  Y  N  I  N  S  F  D  L  K  Y  I  L  T  R  420
AGTTCCCTTCCGAATACGAGCTGCTGCTGGGCTTTATGCTTTCTTTCAACGGTACGCGCCGGCCTTTGTGACCGGTTACAACATCAACTCTTTTGACTTGAAGTACATCCTCACGCGTC 1920

L  E  Y  L  Y  K  V  D  S  Q  R  F  C  K  L  P  T  A  Q  G  G  R  F  F  L  H  S  P  A  V  G  F  K  R  Q  Y  A  A  A  F  460
TCGAGTACCTGTATAAGGTGGACTCGCAGCGCTTCTGCAAGTTGCCTACGGCGCAGGGCGGCCGTTTCTTTTTACACAGCCCCGCCGTGGGTTTTAAGCGGCAGTACGCCGCCGCTTTTC 2040

P  S  A  S  H  N  N  P  A  S  T  A  A  T  K  V  Y  I  A  G  S  V  V  I  D  M  Y  P  V  C  M  A  K  T  N  S  P  N  Y  K  500
CCTCGGCTTCTCACAACAATCCGGCCAGCACGGCCGCCACCAAGGTGTATATTGCGGGTTCGGTGGTTATCGACATGTACCCTGTATGCATGGCCAAGACTAACTCGCCCAACTATAAGC 2160

L  N  T  M  A  E  L  Y  L  R  Q  R  K  D  D  L  S  Y  K  D  I  P  R  C  F  V  A  N  A  E  G  R  A  Q  V  G  R  Y  C  L  540
TCAACACTATGGCCGAGCTTTACCTGCGGCAACGCAAGGATGACCTGTCTTACAAGGACATCCCGCGTTGTTTCGTGGCTAATGCCGAGGGCCGCCGCCCAGGTAGGCCGTTACTGTCTGC 2280

Q  D  A  V  L  V  R  D  L  F  N  T  I  N  F  H  Y  E  A  G  A  I  A  R  L  A  K  I  P  L  R  R  V  I  F  D  G  Q  Q  I  580
AGGACGCCGTATTGGTGCGCGATCTGTTCAACACCATTAATTTTCACTACGAGGCCGGGCCATCGCGCCGGCTGGCTAAAATTCCGTTGCGGCGTGTCATCTTTGACGGACAGCAGATCC 2400

R  I  Y  T  S  L  L  D  E  C  A  C  R  D  F  I  L  P  N  H  Y  S  K  G  T  T  V  P  E  T  N  S  V  A  V  S  P  N  A  A  620
GTATCTACACCTCGCTGCTGGACGAGTGCGCCTGCCGCGATTTTATCCTGCCCAACCACTACAGCAAAGGTACGACGGTGCCCGAAACGAATAGCGTTGCTGTGTCACCTAACGCTGCTA 2520

I  I  S  T  A  A  V  P  G  D  A  G  S  V  A  A  M  F  Q  M  S  P  P  L  Q  S  A  P  S  S  Q  D  G  V  S  P  G  S  G  S  660
TCATCTCTACCGCCGCTGTGCCCGGCGACGCGGGTTCTGTGGCGCGCTATGTTTCAGATGTCGCCGCCCTTGCAATCTGCGCCGTCCAGTCAGGACGGCGTTTCACCCGGCTCCGGCAGTA 2640

N  S  S  S  S  V  G  V  F  S  V  G  S  G  S  S  G  G  V  G  V  S  N  D  N  H  G  A  G  G  T  A  A  V  S  Y  Q  G  A  T  700
ACAGTAGTAGCAGCGTCGGCGTTTTCAGCGTCGGCTCCGGCAGTAGTGGCGGCGTCGGCGTTTCCAACGACAATCACGGCGCCGGCGGTACTGCGGCGGTTTCGTACCAGGGCGCCACGG 2760

V  F  E  P  E  V  G  Y  Y  N  D  P  V  A  V  F  D  F  A  S  L  Y  P  S  I  I  M  A  H  N  L  C  Y  S  T  L  L  V  P  G  740
TGTTTGAGCCCGAGGTGGGTTACTACAACGACCCCGTGGCCGTGTTCGACTTTGCCAGCCTCTACCCTTCCATCATCATGGCCCACAACCTCTGCTACTCCACCCTGCTGGTGCCGGGTG 2880

G  E  Y  P  V  D  P  A  D  V  Y  S  V  T  L  E  N  G  V  T  H  R  F  V  R  A  S  V  R  V  S  V  L  S  E  L  L  N  K  W  780
GCGAGTACCCTGTGGACCCGCCGACGTATACAGCGTCACGCTAGAGAACGGCGTGACCCACCGCTTTGTGCGTGCTTCGGTGCCGCTCTCGGTGCTCTCGGAACTGCTCAACAAGTGGG 3000

V  S  Q  R  R  A  V  R  E  C  M  R  E  C  Q  D  P  V  R  R  M  L  L  D  K  E  Q  M  A  L  K  V  T  C  N  A  F  Y  G  F  820
TTTCGCAGCGGCGTGCCGTGCGCGAATGCATGCGCGAGTGTCAAGACCCTGTGCGCCGTATGCTGCTCGACAAGGAACAGATGGCGCTCAAAGTAACGTGCAACGCTTTCTACGGTTTTA 3120

T  G  V  V  N  G  M  M  P  C  L  P  I  A  A  S  I  T  R  I  G  R  D  M  L  E  R  T  A  R  F  I  K  D  N  F  S  E  P  C  860
CCGGCGTGGTCAACGGTATGATGCCGTGTCTGCCCATCGCCGCCAGCATCACGCGCATCGGTCGCGACATGCTAGAGCGCACGGCGCGGTTCATCAAAGACAACTTTTCAGAGCCGTGTT 3240

F  L  H  N  F  F  N  Q  E  D  Y  V  V  G  T  R  E  G  D  S  E  E  S  S  A  L  P  E  G  L  E  T  S  S  G  G  S  N  E  R  900
TTTTGCACAATTTTTTTAATCAGGAAGACTATGTAGTGGGAACGCGGGAGGGGGATTCGGAGGAGAGCAGCGCGTTACCGGAGGGGCTCGAAACATCGTCAGGGGGCTCGAACGAACGGC 3360

R  V  E  A  R  V  I  Y  G  D  T  D  S  V  F  V  R  F  R  G  L  T  P  Q  A  L  V  A  R  G  P  S  L  A  H  Y  V  T  A  C  940
GGGTGGAGGCGCGGGTCATCTACGGGGACACGGACAGCGTGTTTGTCCGCTTTCGTGGCCTGACGCCGCAGGCTCTGGTGGCGCGTGGGCCCAGCCTGGCGCACTACGTGACGGCCTGTC 3480

L  F  V  E  P  V  K  L  E  F  E  K  V  F  V  S  L  M  M  I  C  K  K  R  Y  I  G  K  V  E  G  A  S  G  L  S  M  K  G  V  980
TTTTTGTGGAGCCCGTCAAGCTGGAGTTTGAAAAGGTCTTCGTCTCTCTTATGATGATCTGCAAGAAACGTTACATCGGCAAAGTGGAGGGCGCCTCGGGTCTGAGCATGAAGGGCGTGG 3600

D  L  V  R  K  T  A  C  E  F  V  K  G  V  T  R  D  V  L  S  L  L  F  E  D  R  E  V  S  E  A  A  V  R  L  S  R  L  S  L  1020
ATCTGGTGCGCAAGACGGCCTGCGAGTTCGTCAAGGGCGTCACGCGTGACGTCCTCTCGCTGCTCTTTGAGGATCGCGAGGTCTCGGAAGCAGCCGTGCGCCTGTCGCGCCTCTCACTCG 3720

D  E  V  K  K  Y  G  V  P  R  G  F  W  R  I  L  R  R  L  V  Q  A  R  D  D  L  Y  L  H  R  V  R  V  E  D  L  V  L  S  S  1060
ATGAAGTCAAGAAGTACGGCGTGCCACGCGGTTTCTGGCGTATCCTGCGCCGTCTGGTGCAGGCCCGCGACGATCTGTACCTGCACCGTGTGCGTGTCGAGGACCTGGTGCTTTCGTCGG 3840

V  L  S  K  D  I  S  L  Y  R  Q  S  N  L  P  H  I  A  V  I  K  R  L  A  A  R  S  E  E  L  P  S  V  G  D  R  V  F  Y  V  1100
TGCTCTCTAAGGACATCTCGCTGTACCGTCAATCTAACCTGCCGCACATTGCCGTCATTAAGCGATTGGCGGCCCGTTCTGAGGAGCTACCCTCGGTCGGGGATCGGGTCTTTTACGTTC 3960

L  T  A  P  G  V  R  T  A  P  Q  G  S  S  D  N  G  D  S  V  T  A  G  V  V  S  R  S  D  A  I  D  G  T  D  D  D  A  D  G  1140
TGACGGCGCCCGGTGTCCGACGGCGCCGCAGGGTTCCTCCGACAACGGTGATTCTGTAACCGCCGGCGTGGTTTCCCGGTCGGACGCGATTGATGGCACGGACGACGACGCTGACGGCG 4080

G  G  V  E  E  S  N  R  R  G  G  E  P  A  K  K  R  A  R  E  P  P  S  A  V  C  N  Y  E  V  A  E  D  P  S  Y  V  R  E  H  1180
GCGGGGTAGAGGAGAGCAACAGGAGAGGAGGAGAGCCGGCAAAGAAGAGGGCGCGGAAACCACCGTCGGCCGTGTGCAACTACGAGGTAGCCGAAGATCCGAGCTACGTGCGCGAGCACG 4200

G  V  P  I  H  A  D  K  Y  F  E  Q  V  L  K  A  V  T  N  V  L  S  P  V  F  P  G  G  E  T  A  R  K  D  K  F  L  H  M  V  1220
GCGTGCCCATTCACGCCGACAAGTACTTTGAGCAGGTTCTCAAGGCTGTAACTAACGTGCTGTCGCCCGTCTTTCCCGGCGGCGAAACCGCGCGCAAGGACAAGTTTTTGCACATGGTGC 4320

L  P  R  R  L  H  L  E  P  A  F  L  P  Y  S  V  K  A  H  E  C  C                                                          1242
TGCCGCGGCGCTTGCACTTGGAGCCGGCTTTCTGCCGTACAGTGTCAAGGCGCACGAATGCTGTTGAGAAACAGCGGCGGCGGTGGTGACTGGGGACGGTGATGATGCTGCTGAGACTG 4440

AGACTGGTGGTGAGAGTAGTGGTGGGGCTGCGTCGCCTGCGACGGCGGGTGGAGATGAGGCGGCGTGGACTGGGACGAGGAGGAGGAGGGCCGCAGCCGTTGGTGGAAACTACGTGCAACGG 4560

CGACGCGGTTAAGGGAGACCGTATCGCGTAGGACGACGTGGCCTCCTCGTATAGGTTGTTGCCGCTGGACTGACACAGCTCCTGAATGAGCTCTTTGTAGCGCTCAAAGGACTCGCTCAC 4680

GTCGTTGGGAATGTCCATCTCGTCAATCTTGCGTTGCAAAATAGTCACGTCGATCTTGACGCTGCTGGCCGAGACGGCGTGACACAGCACGCTGATAACGACGTGGTCGCGCACGATGTT 4800

GAGCGTGACGCTGTAGTCTTCGCGCGCCGCCGTGAGCATCTGCGTGATGCAGTCGCAGGGGATGTGCACGTCGGGGTTTTCGAAGATGACGTGCATGTGCAACCGGCCGTCCTCGGCGTG 4920

CATAAAGGGCACGATGTCGTGGTTGGCGGCGTGCAGCAGGTAATAAATTTGGAAGATGGGCGGCTGGTTGGTGGAGAGCTTGAGTAGCTCGCTGCTCTTACAGAACCCGTAATAAAAGTT 5040

TCTATTTTTCATGATGTTGGACGTCTGATTGATGAAGGCCACCATGGCGTCGGAGGCGGTGGGCAGGATGCGGTTGTCCTCGCACGAGGTGCAGGCGATGCGCGAGACCTGCTCGTTGCC 5160

GTACGGCGAGATGTCGAGGCAGGTGTGCTGCCGGTGCGAGTGGAAATTGAAGGGTAAGCAGATGGACTCTTTGCCCGTGATGGCCATCTTCATCATGTTAAGGTACTTGAGCTCCAGTTC 5280
```

However, it is equally possible that the *pol* mRNA is spliced at position 4959, since a low-abundance RNA has also been mapped at this point. The proximity of these two sites makes it difficult to discriminate between these two possibilities by Northern blot analysis.

HCMV *pol* reading frame. The 5′ terminus of the *pol* mRNA has been mapped to position 276 (Fig. 2). At 383 bp downstream of the mRNA start site is the ATG (position 660) which leads into the 3,725-bp open reading frame identified as that potentially coding for the HCMV DNA polymerase. Within this 383-base leader sequence there are three ATG codons (positions 290, 337, and 604) which lead into open reading frames of 158, 179, and 74 bp, respectively. The sequence around these three ATG codons does not conform to the consensus preferred for initation codons as described by Kozak (23), whereas the ATG (position 660) leading into the DNA *pol* reading frame lies within such a consensus sequence. A similar situation is apparent in the HSV-1 *pol* mRNA leader sequence, in which an ATG codon (which does not conform to Kozak's rules) leads into a 30-bp open reading frame.

The 3,726-bp HCMV *pol* open reading frame is capable of coding for a 1,242-amino acid polypeptide with a molecular weight of 137,103. This compares favorably with the 140-kDa polypeptide seen when purified HCMV DNA polymerase is fractionated on SDS-polyacrylamide gels (6). The HSV-1 DNA polymerase has a polypeptide of very similar length (1,235 amino acids), whereas that of EBV is 18.3% smaller, with a polypeptide 1,015 amino acids long.

Homology to HSV-1 and EBV DNA polymerases. The HCMV DNA polymerase was identified by the extensive homology between the HCMV reading frame HFLF2 (within the *Hin*dIII F fragment) and the HSV-1 DNA polymerase (14, 34). The same HCMV reading frame also showed homology to EBV reading frame BALF5, which has previously been shown to be homologous to the HSV-1 DNA polymerase (34). The homologies were detected by using the program DIAGON (38). From these diagonal plots, it is evident that most of the regions within HCMV polymerase which show homology to HSV-1 polymerase correspond precisely to the regions which are homologous to EBV polymerase. The homology between these three polymerases can be seen more clearly in Fig. 5, where the polymerase polypeptides of HCMV, HSV-1, and EBV have been aligned to show maximum homology. The regions within HCMV polymerase which show homology to both HSV and EBV polymerase in the DIAGON plots are boxed. These areas clearly represent the most conserved regions of the polymerase polypeptide; outside these blocks of homology the sequence shows little or no conservation and, in addition, a number of major insertions and deletions are evident.

In the alignment shown in Fig. 5, 347 of the 1,242 residues of HCMV polymerase have an identically matched pair in the HSV-1 polymerase and 388 residues have identities in the EBV polymerase. Of the 1,235 residues of HSV-1 polymerase, 354 have an identically matched pair in EBV polymerase. Since EBV polymerase is the smallest of the three polypeptides, with 1,015 residues, it can be said that 38.2% of its sequence is identical to that of HCMV polymerase and 35% to that of HSV-1 polymerase. The number of residues with identities in all three of these herpesvirus polymerases is 247, which represents 24.3% of the EBV polymerase sequence.

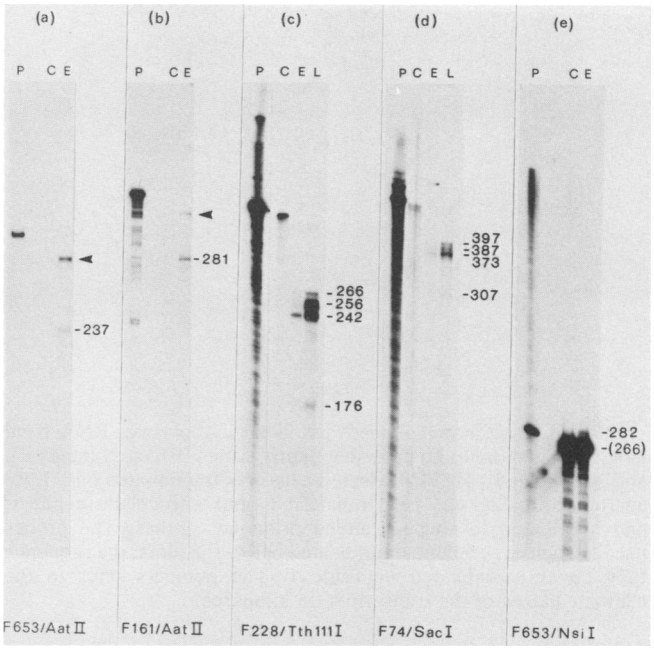

FIG. 3. Nuclease S1 digestion and primer extension analysis. The prime-cut probe used for each experiment is indicated underneath; the precise position of each probe is shown in Table 1. The probe (lanes P) was annealed with RNA from mock-infected cells (lanes C), RNA from early infected cells (lanes E), or RNA from late infected cells (lanes L) and then digested with S1 nuclease (a to d) or extended with reverse transcriptase (e). Arrows indicate bands generated by full-length protection of the probe (after S1 digestion of M13 sequences). The sizes of bands generated by protected fragments are indicated (in base). The number in parentheses for panel e represents the size of the primer used in the extension analysis.

The area showing the most extensive conservation among the three polymerases lies between residues 712 and 736 in the HCMV enzyme. In this 25-residue stretch, there are 20 positions in which an identical residue is present in all three polymerases, 4 positions in which there is an identity in only two and 1 position in which each polymerase has a different but conserved amino acid. A second area of noteworthy homology is at position 905 to 920 in the HCMV sequence. This region of the HSV and EBV polymerase has previously been shown to have a homologous counterpart within the adenovirus type 2 (Ad2) DNA polymerase (14, 34). The addition of the HCMV sequence to this list of homologous polymerases defines more precisely the boundary of this homology and the extent of conservation (Fig. 6a). A second stretch of amino acid conservation between the herpesvirus and Ad2 polymerases has also been identified (positions 810 to 820 in the HCMV sequence; Fig. 6b); this region has recently been shown to possess a homolog in the polymerase gene of bacteriophage ϕ29 (1). The two small stretches of amino acid conservation shown in Fig. 6a and b are present in approximately the same relative positions in the herpesvirus and Ad2 polymerase polypeptides.

FIG. 2. A 5,280-bp sequence containing the HCMV *pol* gene. The amino acid sequence of reading frame HFLF2 (*pol*) is represented by the one-letter amino acid code above the nucleotide sequence. The TATA box sequence present upstream of the mapped 5′ RNA terminus is boxed with solid lines, and polyadenylation signal sequences are boxed with dashed lines. The termination codon of the two open reading frames on either side of HFLF2 is underlined (see Fig. 1).

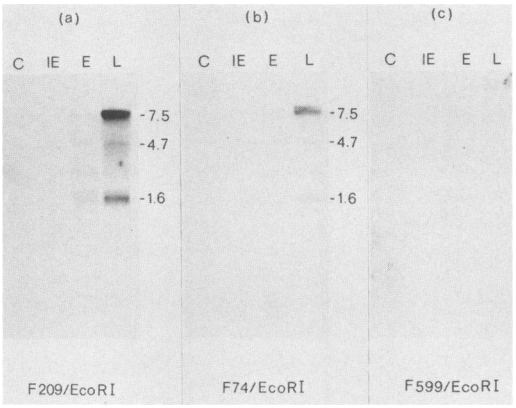

FIG. 4. Northern blot analysis. Total cytoplasmic RNA from mock-infected (lanes C) immediate-early (lanes IE), early (lanes E), and late (lanes L) HCMV-infected cells was fractionated on a 1.5% agarose–formaldehyde gel, transferred onto nitrocellulose filters, and hybridized to single-stranded prime-cut probes. The probes used in each experiment are indicated below the blots; the sequence they cover is indicated in Table 1. The numbers refer to the estimated sizes of the transcripts (in kilobases).

The polymerases of both HCMV and HSV are about 18% larger than that of EBV. The extra sequence can be accounted for by a number of major internal insertions and longer N and C termini. Notably, there are five positions within the core of the HCMV polymerase at which major insertions are apparent with respect to the EBV sequence. At two of these positions (starting at 453 and 881), HCMV has insertions relative to both EBV and HSV, whereas at the other three positions (starting at 341, 608, and 1109) both HCMV and HSV have insertions relative to EBV. The largest insertion within the HCMV enzyme started at position 608, where there are 89 residues without a counterpart in EBV. Starting at the same relative position, HSV had 49 residues which are not present in EBV. The 89-residue insertion with HCMV shows no homology to the 49-residue insertion in HSV; furthermore, the strikingly hydrophilic nature of the HSV insertion is not conserved in HCMV. The second largest insertion within HCMV starts at position 1109, where there are 47 residues with no counterpart in EBV. An insertion of similar size is present in HSV at the same relative position, but again there was no evidence of conservation between the two sequences. Other relatively nonconserved areas of the polymerase polypeptides are the N and C termini. In all three viruses the relative length of the termini is also different, with HSV possessing the longest and HCMV the second longest N and C termini.

DISCUSSION

We report here the sequence and transcription analysis of the HCMV DNA polymerase gene. The predicted *pol* polypeptide corresponds to reading frame HFLF2, present within the *Hin*dIII F fragment of HCMV strain AD169. This reading frame shows extensive amino acid homology to the DNA polymerase gene of two distantly related herpesviruses, HSV-1 and EBV; this homology provides convincing evidence that reading frame HFLF2 encodes the HCMV DNA polymerase. Supportive of this conclusion is the fact that the putative HCMV *pol* gene is present in a conserved location with respect to EBV *pol*: the *pol* genes of both viruses are directly preceded by the glycoprotein B gene, a gene of unknown function, and the gene coding for the major DNA-binding protein (32; unpublished data).

The homology between the three herpesvirus polymerases consists of several long stretches of conserved or highly similar amino acid sequence, with bordering regions showing little or no homology. At present there is no evidence that any of these conserved regions correspond to functionally important parts of the protein. However, such striking conservation among the DNA polymerases of three distantly related herpesviruses is suggestive and provokes the speculation that at least some of these blocks of conservation will turn out to be part of functional domains of the enzyme. Analysis of temperature-sensitive and altered-sensitivity mutations of the HSV-1 DNA polymerase gene provide the only guide to which parts of the polypeptide are functionally important (14, 18, 22). On the basis of such mapping data, Gibbs et al. (14) have suggested that the substrate-binding site of the enzyme falls within its C-terminal portion. More precisely, they suggest that much of the C terminus, including the 228-amino-acid segment between 696 and 924, folds to form a domain containing the deoxynucleoside triphosphate- and PP_i-binding sites. As seen in Fig. 5, this part of the polypeptide contains the most highly conserved region within the three herpesvirus polymerases (HCMV position 712 to 736) and the two short stretches of sequence conserved in the Ad2 polymerase gene (position 905 to 920 and 810 to 820 in HCMV). Verification of a functional role for these sequences awaits the sequence analysis of *pol* mutants.

The detailed comparison of the HCMV, HSV-1, and EBV DNA polymerases strongly suggests that they have a common ancestor: they are homologous along much of their length, more than 24% of their sequence is precisely conserved, and a number of large deletions and insertions that appear to have occurred are all in the nonconserved areas of the protein. An interesting observation, in evolutionary terms, is the fact that there were no insertions of sequence in the EBV *pol* relative to the *pol* of HSV and HCMV (Fig. 5).

The *pol* genes of HCMV and HSV-1 also possess homology at the DNA level (data not shown). Evidence for this comes from Heilbronn et al. (17). Using the HSV-1 *pol* gene as a probe, they identified an 850-bp HCMV fragment showing significant cross-hybridization on Southern blotting. The sequence of this fragment is identical to part of the *pol* gene reported here.

Nuclease S1 protection and primer extension analysis identified the 5′ end of a low-abundance early RNA, 383 bases upstream of the *pol* coding region. A 4.7-kb RNA shown to hybridize to the *pol* sequence is potentially capable of encoding the entire *pol* polypeptide (assuming that no splicing occurs). We have been unable to identify unequivocally the 3′ end of this RNA due to the presence of other transcripts of equivalent abundance in the vicinity. If the transcript is unspliced, its 3′ terminus probably lies beyond the polyadenylation site at 5031 to 5036. Heilbronn et al. (17), who also analyzed the transcripts originating from this region, identified the *pol* mRNA as a 5.4-kb transcript. Their mapping of the 5′ and 3′ ends of the 5.4-kb mRNA is consistent with our S1 analysis of the 5′ and probable 3′ termini of the 4.7-kb transcript, so the discrepancy may reflect an error in sizing.

A number of other transcripts can be detected in the region of the *pol* gene (Fig. 4). The two transcripts of over 7.5 kb have been shown to originate upstream of the *pol* promoter (data not shown) and are probably responsible for the full-length protection of the probes used to map this

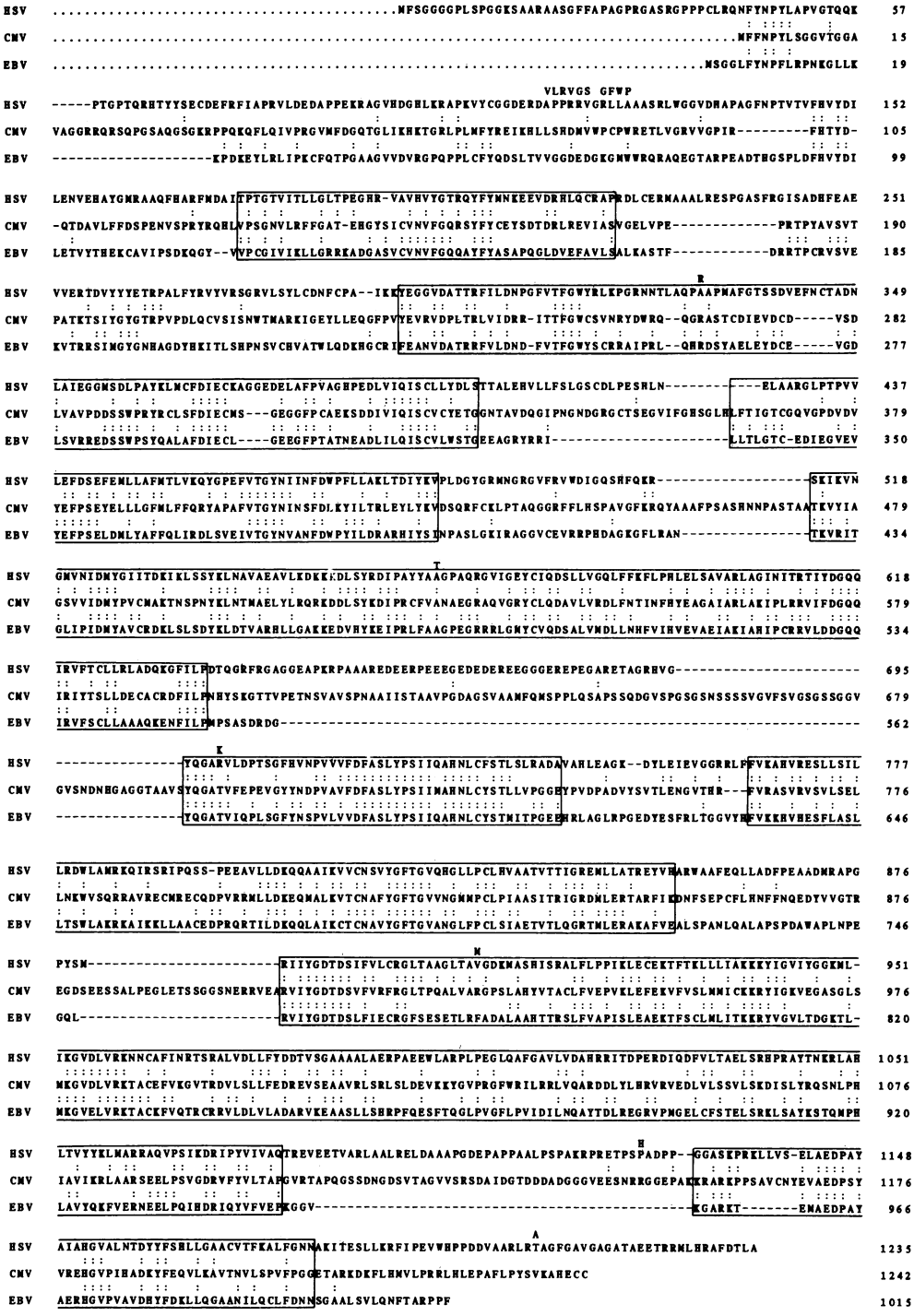

FIG. 5. Alignment of the amino acid sequences of the HSV-1 and EBV polymerases with that of HCMV reading frame HFLF2. Gaps have been introduced into the sequence to account for the different lengths of the three polypeptides and to generate maximum alignment of homologous amino acids. Identical residues are marked (:). The EBV polymerase sequence is from Baer et al. (2), and the HSV-1 sequence is from Quinn and McGeoch (34). The discrepancies between the above HSV-1 sequence and that of Gibbs et al. (14) are presented above the HSV-1 sequence. Boxes represent the regions within the HCMV sequence which, according to the DIAGON plot (38), are homologous to both the HSV and EBV polymerases.

promoter (Fig. 3a and b). Either or both of these two RNAs may be readthrough products of transcription normally terminating at the polyadenylation site present within the *pol* coding region (position 756 to 761 [16]); termination of transcripts at this site has been shown to be inefficient (data not shown). There is no single large open reading frame which could account for the coding capacity for the 1.6-kb transcript (Fig. 4). However, it should be pointed out that a

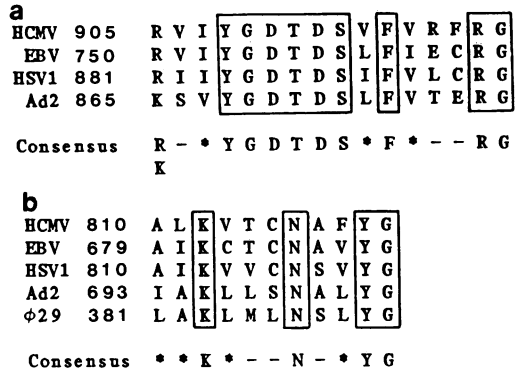

FIG. 6. Short stretch of homology between three herpesvirus (HCMV, HSV-1, and EBV) DNA polymerases and the DNA polymerase of Ad2 (a and b) and φ29 (b). The starting position of each sequence is indicated. Residues shared by all the DNA polymerases are boxed. A consensus sequence is shown below; * represents a hydrophobic amino acid.

389-bp open reading frame (4615 to 5004) ends 45 bp downstream of the mapped splice donor site at 4957 to 4965.

The pattern of transcription within this region of the HCMV genome possesses features also found in the two best-characterized herpesviruses, HSV and EBV. A common transcriptional feature in these two viruses is the presence of nested sets of transcripts having a common 3′ end but a unique 5′ terminus. Two 3′ coterminal RNAs have been mapped at the second polyadenylation signal sequence following the *pol* gene (position 5031 to 5036). Other noteworthy features are the presence of the *pol* promoter within the glycoprotein B coding region and the presence of a functional polyadenylation site in the *pol* coding sequence. Such economical use of sequence has been reported for other herpesviruses (2, 27).

In conclusion, comparison of the amino acid sequence of three distantly related herpesvirus polymerases has identified the most conserved areas of the protein and has highlighted regions which may possess functional importance. We hope this comparison will provide the basis for experiments to analyze the relationship between structure and function in the enzyme. A clear understanding of this relationship may lead to the development of more effective antiherpetic drugs.

ADDENDUM IN PROOF

Recent sequence data from a number of acyclovir-resistant mutants of HSV has shown that their *pol* genes encode a polypeptide with a single amino acid change (Asn → Ser) at position 815 (B. Larder and G. Darby, manuscript in preparation). This Asn residue is conserved in the polymerase of HSV, HCMV, EBV, Ad2, and φ29 (Fig. 6b).

LITERATURE CITED

1. **Argos, P., A. D. Tucker, and L. Philipson.** 1986. Primary structure relationships may reflect similar DNA replication strategies. Virology **149**:208–216.
2. **Baer, R., A. T. Bankier, M. D. Biggin, P. L. Deininger, P. L. 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.
3. **Biggin, M. D., T. J. Gibson, and G.-F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
4. **Bina-Stein, M., M. Thoren, N. Salzman, and J. A. Thompson.** 1979. Rapid sequence determination of late simian virus 40 16S mRNA leader by using inhibitions of reverse transcriptase. Proc. Natl. Acad. Sci. USA **76**:731–735.
5. **Biron, K. K., S. C. Stonat, J. B. Sorrell, J. A. Fyfe, P. M. Keller, C. U. Lambe, and D. J. Nelson.** 1985. Metabolic activation of the nucleoside analog 9-([2-hydroxy-1-(hydroxymethyl)ethoxy]methyl)guanine in human diploid fibroblasts infected with human cytomegalovirus. Proc. Natl. Acad. Sci. USA **82**: 2473–2477.
6. **Bzik, D. J., B. A. Fox, N. A. De Luca, and S. Person.** 1984. Nucleotide sequence specifying the glycoprotein gene, gB, of herpes simplex virus type 1. Virology **133**:301–314.
7. **Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon.** 1980. Promoter sequences of eukaryotic protein coding genes. Science **209**:1406–1414.
8. **Deininger, P. L.** 1983. Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. Anal. Biochem. **129**:216–223.
9. **Duckworth, M. L., M. J. Gait, P. Goelet, G. F. Hong, M. Singh, and R. C. Titmas.** 1981. Rapid synthesis of oligooxyribonucleotides. VI. Efficient, mechanised synthesis of heptadecadeoxyribonucleotides by an improved solid phase phosphotriester route. Nucleic Acids Res. **9**:1691–1706.
10. **Elion, G. B., P. A. Furman, J. A. Fyfe, P. de Miranda, L. Beauchamp, and H. J. Schaeffer.** 1977. Selectivity of action of an antiherpetic agent, 9-(2-hydroxyethoxymethyl)guanine. Proc. Natl. Acad. Sci. USA **74**:5716–5720.
11. **Estes, J., and E. S. Huang.** 1977. Stimulation of cellular thymidine kinase by human cytomegalovirus. J. Virol. **24**:13–21.
12. **Farrell, P. J., P. L. Deininger, A. Bankier, and B. G. Barrell.** 1983. Homologous upstream sequences near Epstein-Barr virus promoters. Proc. Natl. Acad. Sci. USA **80**:1565–1569.
13. **Field, A. K., M. E. Davies, C. De Witt, H. C. Perry, R. Kiou, J. Germershausen, J. D. Karkas, W. T. Ashton, D. B. R. Johnston, and R. L. Tolman.** 1983. 9-([2-Hydroxy-1-(hydroxymethyl)ethoxy]methyl)guanine: a selective inhibitor of herpes group virus replication. Proc. Natl. Acad. Sci. USA **80**:4139–4143.
14. **Gibbs, J. S., H. C. Chiou, J. D. Hall, D. W. Mount, M. J. Retondo, S. K. Weller, and D. M. Coen.** 1985. Sequence and mapping analysis of the herpes simplex virus DNA polymerase gene predict a C-terminal substrate binding domain. Proc. Natl. Acad. Sci. USA **82**:7969–7973.
15. **Glen, J.** 1981. Cytomegalovirus infections following renal transplantation. Rev. infect. Dis. **3**:1151–1177.
16. **Hagenbuchle, O., R. Bovey, and R. A. Young.** 1980. Tissue-specific expression of mouse α-amylase genes: nucleotide sequence of isozyme mRNAs from pancreas and salivary gland. Cell **21**:179–187.
17. **Heilbronn, R., G. Jahn, A. Bürkle, U.-K. Freese, B. Fleckenstein, and H. zur Hausen.** 1987. Genomic localization, sequence analysis, and transcription of the putative human cytomegalovirus DNA polymerase gene. J. Virol. **61**:119–124.
18. **Honess, R. W., D. J. M. Purifoy, D. Young, R. Gopal, N. Cammack, and P. O'Hare.** 1984. Single mutations at many sites within the DNA polymerase locus of herpes simplex virus can confer hypersensitivity to Aphidicolin and resistance to phosphoric acid. J. Gen. Virol. **65**:1–7.
19. **Huang, E.-S.** 1975. Human cytomegalovirus. III. Virus-induced DNA polymerase. J. Virol. **16**:298–310.
20. **Hudson, G. S., T. J. Gibson, and B. G. Barrell.** 1985. The BamHI F region of the B95-8 Epstein-Barr virus genome. Virology **147**:99–109.
21. **Knopf, K. W.** 1979. Properties of the herpes simplex virus DNA polymerase and characterization of its associated exonuclease activity. Eur. J. Biochem. **98**:231–244.
22. **Knopf, K. W., E. R. Kaufman, and C. Crumpacker.** 1981. Physical mapping of drug resistance mutations defines an active center of the herpes simplex virus DNA polymerase enzyme. J. Virol. **39**:746–757.

23. **Kozak, M.** 1983. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
24. **Mar, E.-C., J. F. Chiou, Y.-C. Cheng, and E.-S. Huang.** 1985. Inhibition of cellular DNA polymerase α and human cytomegalovirus-induced DNA polymerase by the triphosphates of 9-(2-hydroxyethoxymethyl)guanine and 9-(1,3-dihydroxy-2-propoxymethyl)guanine. J. Virol. **53**:776–780.
25. **Mar, E.-C., J.-F. Chiou, Y.-C. Cheng, and E.-S. Huang.** 1985. Human cytomegalovirus-induced DNA polymerase and its interaction with the triphosphates of 1-(2′-deoxy-2′-fluoro-β-D-arabinofluranosyl)-5-methyluracil, -5-iodocytosine, and -5-methylcytosine. J. Virol. **56**:846–851.
26. **Mar, E.-C., P. C. Patel, Y.-C. Cheng, J. J. Fox, K. A. Watanabe, and E.-S. Huang.** 1983. Effects of certain nucleoside analogues on human cytomegalovirus replication *in vitro*. J. Gen. Virol. **65**:47–53.
27. **McGeoch, D. J., A. Dolan, S. Donald, and F. J. Rixon.** 1985. Sequence determination and genetic content of the short unique region of the genome of herpes simplex virus type 1. J. Mol. Biol. **181**:1–13.
28. **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.
29. **Nishiyama, Y., K. Maeno, and S. Yoshida.** 1983. Characterization of human cytomegalovirus induced DNA polymerase and the associated 3′ to 5′ exonuclease. Virology **124**:221–231.
30. **Oram, J. D., R. G. Downing, A. Akrigg, A. A. Dollery, C. J. Doggleby, 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.
31. **Pellett, P. E., M. D. Biggin, B. Barrell, and B. Roizman.** 1985. Epstein-Barr virus genome may encode a protein showing significant amino acid and predicted secondary structure homology with glycoprotein B of herpes simplex virus type 1. J. Virol. **56**:807–813.
32. **Powell, K. L., and D. J. Purifoy.** 1977. Nonstructural proteins of herpes simplex virus. I. Purification of the induced DNA polymerase. J. Virol. **24**:618–626.
33. **Proudfoot, N. J., and G. G. Brownlee.** 1976. 3′ non-coding region sequences in eukaryotic messenger RNA. Nature (London) **263**:211–214.
34. **Quinn, J. P., and D. J. McGeoch.** 1985. DNA sequence of the region in the genome of herpes simplex virus type 1 containing the genes for DNA polymerase and the major DNA binding protein. Nucleic Acids Res. **13**:8143–8163.
35. **Sanger, F., and A. R. Coulson.** 1978. The use of thin acrylamide gels for DNA sequencing. FEBS Lett. **87**:107–110.
36. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
37. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5468.
38. **Staden, R.** 1982. An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. Nucleic Acids Res. **10**:2951–2961.
39. **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.
40. **Staden, R.** 1984. Graphic method to determine the function of nucleic acid sequences. Nucleic Acids Res. **12**:521–538.
41. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77**:5201–5205.
42. **Tocci, M. J., T. J. Livelli, H. C. Perry, C. S. Crumpacker, and A. K. Field.** 1984. Effects of the nucleoside analog 2′-nor-2′-deoxyguanosine on human cytomegalovirus replication. Antimicrob. Agents Chemother. **25**:247–252.
43. **Vaughan, P. J., D. J. Purifoy, and K. L. Powell.** 1985. DNA-binding protein associated with the herpes simplex virus DNA polymerase. J. Virol. **53**:501–508.
44. **Weller, T. H.** 1981. Clinical spectrum of cytomegalovirus infection, p. 20–40. *In* A. J. Nahmias, W. R. Dowdle, and R. F. Schinazi (ed.), The human herpesviruses: an interdisciplinary perspective. Elsevier/North-Holland Inc., New York.