Transcription of Herpes Simplex Virus Genes In Vivo: Overlap of a Late Promoter with the 3' End of the Early Thymidine Kinase Gene

JANICE A. SHARP,¹ MICHAEL J. WAGNER,^{1,2} AND WILLIAM C. SUMMERS^{1*}

Department of Therapeutic Radiology, Department of Molecular Biophysics and Biochemistry, and Department of Human Genetics, Yale School of Medicine, New Haven, Connecticut 06510,¹ and Department of Biology, University of Houston, Houston, Texas 77004²

Received 1 June 1982/Accepted 30 August 1982

We identified in herpes simplex virus type 1-infected cells six cytoplasmic transcripts which were complementary to *Bam*HI restriction endonuclease fragment Q. Two transcripts appeared in major amounts compared with the other four. One major transcript of about 1.4 kilobases was the mRNA for the viral thymidine kinase, was synthesized at intermediate times, and was classified as a beta transcript. The other major transcript was synthesized at late times and was classified as a gamma transcript. This late transcript was about 3 kilonucleotides long and was transcribed in the same direction as the gene for thymidine kinase. The 5' end of this late RNA was located by RNA sequence analysis and was 23 nucleotides downstream from the polyadenylation site for the thymidine kinase mRNA. This finding led to the conclusion that the control region for the 3-kilobase gamma transcript is contained within the 3' untranslated region of the thymidine kinase transcript.

The transcriptional regulation of gene expression in herpes simplex virus (HSV) involves a complex network of cascade controls (12, 21). One major feature of this regulation involves positive transcriptional activation of the "early" (or beta) and "late" (or gamma) classes of viral genes by at least one gene product of other viral genes belonging to the "immediate-early" (or alpha) class. To investigate the molecular mechanisms of this type of regulation, we studied the transcription patterns of a small part of the HSV type 1 (HSV-1) genome in the region of the viral thymidine kinase (TK) locus. In this descriptive study we determined the major transcription units in this region, as well as the locations of some possible control signals in the DNA sequence. Other regions of the HSV-1 genome have been subjected to similar studies by other workers (2, 6, 10, 11, 23).

On *Bam*HI fragment Q of HSV-1, which contains the gene for HSV TK, we found the promoters for two major transcripts, the previously identified TK transcript (21) and an adjacent transcript of about 3 kilobases (kb). These two adjacent transcripts are transcribed in the same direction but are synthesized at different times after infection. The novel finding reported in this paper is that the control region of the 3-kb RNA is embedded in the DNA encoding the 3' region of the TK transcript.

The region between the polyadenylation site for the upstream (TK) transcript and the 5' "cap" site for the adjacent mRNA is 23 nucleotides long, whereas the distance from the end of the TK protein-encoding sequence to the 5' cap site is 91 nucleotides. This implies that most of this intercistronic region has a dual function, first, as a template for transcription of TK mRNA and second, as the promoter for the 3-kb late transcription unit. Our work confirmed for one region at the nucleotide level the dense packing arrangement of HSV genes noted by Frink et al. (10) from their hybridization studies of HSV transcripts. In addition to the two major transcripts, we observed that four other RNAs from this region occur at about one-tenth the abundance of the two major transcripts.

MATERIALS AND METHODS

Virus. HSV-1 strain Cl 101 and mutant TK43 were propagated on Vero cells at a multiplicity of infection of 0.05 PFU/cell. Infectious virus was released from cell pellets by freeze-thawing and gentle sonication. Virus stocks were stored at -70°C.

Cytoplasmic RNA preparation. RNA was prepared by a modification of the procedure described by Cremer et al. (7) and Rosen et al. (20).

Mouse LMTK cells in roller bottles were infected at multiplicity of infection of 15 to 20 PFU/cell. At 8 h after infection or at the times indicated below, cells were harvested onto crushed, frozen phosphate-buff-

ered saline and collected by centrifugation at $700 \times g$ for 5 min. Cell pellets were washed with cold phosphate-buffered saline and then suspended in 4 volumes of 10 mM Tris-hydrochloride (pH 7.4)–150 mM NaCl–2 mM $MgCl_2$ –0.5% Nonidet P-40. After 5 min on ice, the cells were lysed by Dounce homogenization. Nuclei were removed by centrifugation at $1,500 \times g$ for 5 min. The supernatant was diluted with an equal volume of 10 mM Tris-hydrochloride (pH 8.0)–25 mM EDTA–75 mM NaCl (TNE) containing 0.5% sodium dodecyl sulfate and extracted with an equal volume of redistilled phenol equilibrated with TNE. The phases were separated by centrifugation at 12,000 rpm for 10 min. The aqueous phase and the interphase were collected and extracted twice more with an equal volume of phenol-chloroform (1:1, vol/vol). The final aqueous phase was collected and digested with 100 μg of proteinase K per ml for 30 min at 37°C. The proteinase K (2.5 mg/ml) had been preincubated at 37°C in 10 mM Tris-hydrochloride (pH 7.5)–5 mM EDTA–1% sodium dodecyl sulfate for 60 to 90 min. After incubation the digest was extracted with an equal volume of phenol-chloroform, and the aqueous phase was precipitated at $-20^\circ C$ overnight by adding 0.1 volume of 0.5 M Tris-hydrochloride (pH 7.5)–1.0 M NaCl–50 mM EDTA and 2.5 volumes of ethanol.

The precipitate was collected by centrifugation at $10,000 \times g$ for 10 min, dried in vacuo, and suspended in water to a final concentration of 2 mg/ml. High-molecular-weight RNA was precipitated by adding cold 4 M sodium acetate (pH 6.0) to give a final concentration of 3 M sodium acetate. After 30 min on ice, the precipitate was collected by centrifugation at $10,000 \times g$ for 10 min. The pellet was dissolved in water and precipitated with 2.5 volumes of ethanol at $-70^\circ C$ for 15 min. The precipitate was collected by centrifugation at $10,000 \times g$ for 10 min, and the pellet was washed twice with 70% ethanol containing 40 mM NaCl, dried in vacuo, and suspended in 10 mM Tris-hydrochloride (pH 7.5)–1 mM EDTA–0.2% sodium dodecyl sulfate at a concentration of 1 mg/ml. The polyadenylated RNA was purified from the total cytoplasmic RNA by chromatography on oligodeoxythymidylic acid cellulose. The polyadenylated RNA fraction was precipitated by adding 40 mM NaCl and 2.5 volumes of ethanol at $-20^\circ C$; this fraction was collected by centrifugation at $10,000 \times g$ for 15 min, washed with 70% ethanol containing 40 mM NaCl, and dried in vacuo. The polyadenylated RNA was suspended in water and stored at $-70^\circ C$.

Plasmid DNAs. Plasmid pX1 contains the entire *Bam*HI-Q fragment of the HSV-1 genome cloned in plasmid pBR322 (8). Numbering of nucleotides in *Bam*HI fragment Q starts at the *Bam*HI site nearer the 5' end of the TK gene. The complete nucleotide sequence from this *Bam*HI site to nucleotide 2,720 is available on request. Our previous numbering system (21) started at nucleotide 318. Subregions of *Bam*HI fragment Q were cloned in pBR322 as follows. *Pvu*II fragment B (plasmid pYS-J2) comprises region V (see Fig. 2); *Bgl*II fragment A (plasmid pJS-1) comprises the region labeled IV (that is, all sequences to the left of the unique *Bgl*II site); *Pst*I fragment B (plasmid pC-1) comprises region III; *Pst*I fragment D (plasmid pT2-1) comprises region II; and *Pvu*II fragment C (plasmid pYS-J4) comprises region I. The plasmids constructed in our laboratory are designated pYS as registered

with the Plasmid Reference Center, Stanford University (E. Lederberg). Plasmids pJS-1, pC-1, and pT2-1 were generous gifts from James Smiley, McMaster University.

Preparation of labeled DNAs. Plasmid DNA prepared by equilibrium centrifugation in CsCl in the presence of ethidium bromide was labeled by nick translation, using the method of Rigby et al. (19). Labeled M13 phage DNA was prepared by *in vivo* labeling of infected cells, using $^{32}P_i$ in low-phosphate medium as previously described (13). The specific activities of these hybridization probes varied between 0.05×10^8 and 1×10^8 cpm/ μg .

Strand-specific probes were constructed by primed synthesis on single-stranded DNA templates as described previously for DNA sequencing (21). M13 phage DNA into which *Eco*RI fragment N had been cloned was used as template. A restriction fragment just 5' to the region to be labeled was annealed to the single-stranded template, and DNA chain extension with labeled nucleotides was carried out as described previously (21). The specific fragment with only one strand labeled was then cleaved with the appropriate restriction endonuclease and purified by gel electrophoresis. The probe Pr100 is the DNA between nucleotide 716 and nucleotide 814, and this probe has the same polarity as the TK coding strand and hence is complementary to TK mRNA. The probe Hha7 is a mixture of two *Hha*I fragments between nucleotides 2,377 and 2,437 and nucleotides 2,450 and 2,512, both of which have the same polarity as Pr100 (that is, complementary to the TK mRNA-like DNA strand).

S1 nuclease mapping. S1 nuclease digestion of RNA-DNA hybrids were performed by a modification of the method of Berk and Sharp (3). A uniformly labeled M13 clone containing *Eco*RI fragment N (21) from the HSV-1 genome was hybridized to polyadenylated RNA (10 μg) by first denaturing at $80^\circ C$ for 5 min and then annealing at $42^\circ C$ for 18 h in 20 μl of 60% formamide–400 mM NaCl–20 mM PIPES [piperazine-*N,N'*-bis(2-ethanesulfonic acid)] (pH 6.5)–0.5% sodium dodecyl sulfate. Upon completion of hybridization, samples were diluted 10-fold with 60 mM sodium acetate (pH 4.6)–100 mM NaCl–2 mM $ZnCl_2$ –10% glycerol and digested with 500 to 1,000 U of S1 nuclease per ml of reaction mixture for 30 min at room temperature and then for 15 min at $4^\circ C$. DNA-RNA hybrids protected from S1 digestion were precipitated by adding 2.5 volumes of ethanol and 10 μg of yeast tRNA per ml. Samples were suspended in 10 mM Tris-hydrochloride (pH 7.5)–0.1 mM EDTA.

Gel electrophoresis. Electrophoretic separations of DNA were carried out on slab gels (18 by 18 by 0.2 cm). The ratio of acrylamide monomer to methylene bisacrylamide monomer was 30:1, and the buffer in the gel and in the electrode tanks was 40 mM Tris-acetate (pH 8.3)–1 mM EDTA. DNA or DNA-RNA hybrids were denatured by heating to $90^\circ C$ for 5 min in 80% formamide–1 mM EDTA–0.1% bromophenol blue–xylene cyanol FF. RNA samples were denatured for 1 h at $50^\circ C$ in 1 M glyoxal–50% dimethyl sulfoxide–10 mM sodium phosphate (pH 7.0) (16) and fractionated in 1.4% agarose gels by using 10 mM sodium phosphate (pH 7.0)–1 mM EDTA as the buffer. After electrophoresis the RNA was transferred to diazobenzoyloxymethyl paper as described by Alwine et al. (1).

RNA sequence analysis. 5' sequences of RNA were

determined by using reverse transcriptase, as described by Wagner et al. (21) and Bina-Stein et al. (4). Briefly, cytoplasmic polyadenylated RNA was hybridized to a labeled restriction nuclease-generated DNA fragment which served as a primer for reverse transcriptase. Synthesis in the presence of deoxynucleotides and dideoxynucleotides resulted in a DNA sequence that was complementary to the RNA template and extended to the 5' end of the RNA. The primer fragment used to analyze the 3-kb late transcript was the 57-nucleotide *Hae*III fragment from nucleotide 2,117 to nucleotide 2,174.

RESULTS

Time course of transcription. To survey the transcripts from HSV-1-infected cells for RNAs complementary to the *Bam*HI fragment Q region, we isolated polyadenylated RNA from LMTK cells infected with HSV-1 at a multiplicity of 15 to 20 PFU/cell. Samples were prepared during infection at intervals up to 16 h. Figure 1 shows a gel electrophoretic analysis of one such experiment from 0 to 12 h in which the interval was 2 h. The RNAs were denatured in glyoxal and separated by electrophoresis in a 1.4% agarose gel. The RNA was transferred to diazobenzyloxymethyl paper and hybridized to labeled DNA from the entire *Bam*HI-Q restriction endonuclease fragment. Two major RNA species and three minor species were consistently observed. The sizes of these RNAs were estimated from their electrophoretic mobilities compared with the mobilities of 18S and 28S rRNAs (1,950 and 5,000 nucleotides long, respectively [25]). The two major mRNA species, which were 1.4 and 3.1 kb long, accumulated at different times during the lytic cycle. The 1.4-kb transcript was not observed until 4 h after infection and disappeared by 10 h. However, the larger 3.1-kb RNA appeared first in very low amounts at 6 h and was still present at 12 h. The other transcripts observed were present in at least a 10-fold-lower amounts or represented hybridization to very small regions of *Bam*HI fragment Q. The species with very low mobility appeared at late times. The apparent size of this RNA was about 10 to 12 kb, and this RNA may have been a nuclear precursor which contaminated the cytoplasmic fraction late during infection, when the nuclei were most fragile. Two other minor species (4.8 and 5.5 kb) appeared with the kinetics of the 1.4-kb major beta transcript.

Mapping of transcripts by hybridization. Labeled subcloned regions of *Bam*HI fragment Q were hybridized to mRNA blots to obtain a better idea of the map locations of these transcripts. Figure 2 shows the subregions used and the transcripts which hybridized. After hybridization to a specific subregion, the same blot was then hybridized to the entire *Bam*HI fragment Q

FIG. 1. Hybridization analysis of *Bam*HI fragment Q-related cytoplasmic RNAs isolated at different times after infection of LMTK cells with HSV-1. At the times indicated, cytoplasmic polyadenylated RNA was prepared, denatured in glyoxal, and subjected to electrophoresis. The RNA was transferred to diazobenzyloxymethyl paper and hybridized to ^{32} P-labeled DNA from plasmid pX1, which contains *Bam*HI fragment Q of HSV-1. Detection of labeled material was by contact autoradiography. The approximate sizes of the RNAs (indicated on the right, in kilobases) were estimated from a plot of log size versus migration, with stained bands of rRNA (5,000 and 1,950 kb) serving as internal standards at the positions indicated on the left.

region, and then the two autoradiographs were superimposed for comparison.

By using probes with higher specific activities and relatively long exposure times, we detected an additional transcript, which was about 1.0 kb long. This transcript appeared with kinetics similar to those of the major 1.4-kb transcript and hence was designated a beta transcript.

The 1.4-kb transcript was localized to the central region of *Bam*HI fragment Q between a *Pst*I site and a *Pvu*II site (Fig. 2). This region corresponds to the structural gene for HSV TK, which we sequenced previously (21). To test the polarity of transcription of the 1.4-kb RNA, we hybridized the blotted electropherogram to a small strand-specific probe (Pr100) constructed to be complementary to the TK mRNA strand by primed synthesis on single-stranded M13 phage carrying inserted *Eco*RI fragment N (21).

Figure 3 shows that Pr100 hybridized to the 1.4-kb transcript and confirms that this transcript has the same polarity as TK mRNA.

The 3.1-kb late transcript (designated 3KBL) mapped to the left of the TK region and therefore must also have contained sequences that were not represented on *Bam*HI fragment Q, which was only 3.5-kb long in toto. Hybridization to a short, strand-specific probe (Hha7) made by primed synthesis of a single-stranded M13 clone of *Eco*RI fragment N showed that 3KBL RNA had the same polarity as the 1.4-kb TK mRNA (Fig. 3). Hybridization of the total cytoplasmic polyadenylated RNA to labeled single-stranded *Eco*RI fragment N cloned in M13 phage, followed by digestion of the unhybridized DNA with S1 nuclease and gel electrophoretic analysis of the protected DNA, showed that two major transcripts approximately 1,200 and 850 base pairs long protected *Eco*RI fragment N from S1 digestion (Fig. 4). Together with the location and polarity of the 1.4-kb TK transcript, this result suggests that the other major leftward transcript starts about 850 nucleotides from the left end of *Eco*RI fragment N, a location close

FIG. 2. Hybridization analysis of RNAs from the *Bam*HI fragment Q region. RNAs isolated at different times were electrophoresed and transferred to diazobenzoyloxymethyl paper as described in the legend to Fig. 1. Hybridization probes from the indicated regions of *Bam*HI fragment Q were then hybridized to the fractionated RNAs. The track for the 8-h RNA sample is shown in each case. A schematic map of *Bam*HI fragment Q in the standard P orientation is shown with the regions (regions I through V) covered by the probes as described in the text. The central fragment between the two indicated *Eco*RI (RI) cleavage sites is *Eco*RI fragment N from the whole viral genome. The sizes (in kilobases) of the major RNAs are indicated. These RNAs were identified as the RNAs in Fig. 1 by rehybridization of the blot to the entire *Bam*HI fragment Q probe and subsequent superimposition of the two autoradiographs. Sma, *Sma*I; Pvu, *Pvu*II; Pst, *Pst*I.

FIG. 3. Hybridization of two small 5'-specific probes to HSV-1 RNAs. The methods used are described in the legends to Fig. 1 and 2. The *Eco*RI fragment N (RI N) probe was labeled DNA of M13 phage into which *Eco*RI fragment N had been cloned. The Hha7 probe (nucleotides 2,377 to 2,512) was from the same strand as the *Eco*RI fragment N probe and was made by primed synthesis on a cDNA template (the other strand was cloned in M13). The probe Pr100 (nucleotides 716 to 814) was made in the same fashion and had the same polarity as the other probes.

(150 nucleotides) to the 3' end of the TK gene (15, 21).

Analysis of 5' end of 3KBL RNA. Two methods were used to locate more precisely the 5' end of the 3KBL transcript and to determine its sequence. First, we determined how much of two different DNA fragments were protected from S1 nuclease digestion by hybridization to cytoplasmic polyadenylated RNA. This S1 protection experiment (Fig. 5) showed that an abundant RNA was present, which was colinear with the DNA for at least 338 nucleotides starting at nucleotide 2,054. In addition to this major RNA species, it was clear that at least one other transcript completely covered this region (from the *Msp*I site at nucleotide 2,392 to the *Sma*I site at nucleotide 1,940), since the entire fragment could be protected from S1 digestion by RNA from infected cells.

In the second method of analysis, the nucleotide sequence of the RNA was determined by the dideoxynucleotide termination method, using a labeled DNA primer as previously described (21). Three results were obtained from this experiment, and these results confirmed the S1 protection data (Fig. 6). First, a major termination site for this "run-off" sequencing method indicated the existence of a precise 5' terminus on the major RNA which hybridized to the same primer which was used to label the 3KBL RNA in the experiment shown in Fig. 3. Second, the nucleotide sequence determined for this 3KBL RNA corresponded to the DNA sequence for 30 nucleotides from the 5' end, which confirmed that no RNA splicing events had occurred in this segment of RNA. Third, long exposure of the gel showed that there were some RNAs (about 5%) which had 5' termini further to the right than the major 3KBL transcript. This was shown by the presence of faint sequences at the top of the gel beyond the major stop site. These RNAs may represent the minor, very large 10- to 12-kb RNAs or the minor, 4.8 to 5.5-kb species which also map in this region (Fig. 2).

DISCUSSION

In addition to the 1.4-kb transcript for the 42,000-dalton TK protein, the DNA of *Bam*HI fragment Q of HSV-1 encodes at least four other transcripts. One of these, 3KBL RNA, is present in infected cells at a level comparable to that of the TK transcript. Three other transcripts were consistently detected and were present at about one-tenth the level of the TK RNA. The time course of accumulation of these transcripts suggests that the 3KBL transcript is a gamma or late transcript, whereas the other four transcripts are beta or early transcripts. Since the detection of these RNAs was based on measurements of the total accumulation of RNA of a specific size, we can only estimate the steady-state levels of these transcripts, not their rates of synthesis. This point is worth emphasizing, since it has been suggested recently that control of mRNA levels in adenovirus-infected cells may reflect RNA stabilization rather than differential rates of synthesis (17).

The mapping of these transcripts by hybridization to subregions of *Bam*HI fragment Q cannot rule out the possibility that some of these RNAs may be from overlapping transcription units. The minor transcript about 4.8-kb long hybridized to *Pvu*II fragment B and *Bgl*II fragment A, as did the major TK RNA (*Bgl*II fragment A) and the major 3KBL transcript (*Pvu*II fragment B and *Bgl*II fragment A).

The possible overlap between the TK transcript and the 3KBL RNA was resolved by S1 nuclease protection experiments and by se-

FIG. 4. S1 nuclease protection experiment. RNA from *Escherichia coli* (0 RNA), HSV-1-infected LMTK cells (IF RNA), or uninfected LMTK cells (UF RNA) was hybridized to labeled single-stranded M13 phage DNA containing *Eco*RI fragment N of HSV-1 (Fig. 2). The hybrids were digested with S1 nuclease and then denatured before electrophoresis. The size standards used were stained *Sma*I fragments of plasmid pX1 DNA which had been denatured before electrophoresis.

quence analysis of the 3KBL RNA. Although the 5' end of the 3KBL RNA lies 23 nucleotides beyond the polyadenylation site for the TK mRNA, the control region for the 3KBL transcript must be embedded in the 3' region of the TK transcription region. The faint sequence data beyond the strong stop observed in run-off sequencing of 3KBL RNA suggest that transcripts which have their 5' ends "upstream" from the 3KBL 5' end, and which span the intercistronic region and hence overlap the TK and 3KBL sequences, exist at very low levels. The sequence obtained for the 5' end of 3KBL RNA coincides with the DNA sequence from the 5' end at nucleotide 2,054 to nucleotide 2,090, which was as far as we determined the RNA sequence. From the S1 protection data, it appears that this transcript lacks introns within 338 nucleotides of the 5' end. The first initiator codon (AUG) occurs at nucleotide 2,266, which is 212 bases from the 5' end of the mRNA. Starting with this AUG, there are no in-phase

FIG. 5. S1 protection experiment to locate the 5' end of the 3KBL in vivo transcript. Cytoplasmic polyadenylic acid-containing RNA from HSV-1-infected cells was annealed to two different ^{32}P -end-labeled DNA fragments. Part of the sample was digested with S1 nuclease (tracks C), and part was left undigested (tracks A). Annealing in the absence of cytoplasmic RNA but with an equivalent amount of *E. coli* rRNA, followed by digestion with S1 nuclease, gave the samples which were analyzed in tracks B. The end-labeled DNAs were subjected to the chemical method of DNA sequence analysis (14) (tracks S; order of reactions from left to right: G, G+A, C+T, T). (A) DNA fragment from the *Sma*I site at nucleotide 1,940 to the *Taq*I site at nucleotide 2,092. (B) DNA fragment from the *Sma*I site at nucleotide 1,940 to the *Msp*I site at nucleotide 2,394.

terminator codons for as far as we have determined the DNA sequence (464 nucleotides beyond the AUG) (Fig. 7). In the other two translational phases, stop codons are found at several loci in this sequence. Thus, it appears that this late gene, like the TK gene, may not contain introns. This absence of splicing is not a constant feature of HSV-1 transcripts, however, since other workers (10, 24) have shown that some HSV-1 transcripts are spliced.

The intercistronic region between the TK gene and the 3KBL locus is 24 base pairs long, and so the regulatory sequences (5) for the 3KBL transcription unit are within the 3' un-

translated region of the TK gene (Fig. 8). However, the important sequences at about -80 and -30 base pairs upstream from the 5' end of 3KBL RNA do not overlap with the TK coding sequences. Furthermore, it appears that the repeated sequence AATAAA after the TK gene is used for two different functions; the first, at nucleotide 2,009, is the polyadenylation signal for the TK transcript, and the second, at nucleotide 2,022, is part of the 3KBL promoter.

If polyadenylation occurs by a processing event at a previously transcribed sequence (9), it is likely that the promoter region of 3KBL RNA is transcribed during the beta phase of the lytic

cycle. It may be that such transcription termination activity is not compatible with initiation of RNA synthesis from the same region and thus activation of this particular gamma gene may require cessation of beta gene transcription.

It should be noted that transcription of the 3KBL region might result from suppression of termination of the TK transcript, followed by processing of this polycistronic RNA to give the observed 5' terminus of 3KBL. This is possible since the second polyadenylation promoter signal is within the 11 to 30 nucleotides before the 5' terminus of 3KBL mRNA, which is also at an appropriate CA cleavage site for processing and polyadenylation (9). Frink et al. (10) have observed at least one HSV transcript which ap-

FIG. 6. Sequence analysis of the 5' end of the 3KBL transcript by the primer extension method (21). The sequence is the complement of that shown in Fig. 7, and the easily readable sequence at nucleotide 2,085 is shown for reference. There is a strong stop at nucleotide 2,054.

```

1901                                     Sma I
TCCATACCGACGATCTGCBACCTGGCGCAGCTTTGCCCGGAGATGGG
                                STOP
GGAGGGCTAACTGAACACGGAAGGAGACAATACCGGAAGGAACCGCGCT
2001                                     3' POLY A
ATGACGGCAATAAAAGACAGAATAAAACGCACGGGTGTTGGGTCGTTTG
5' 3KBL mRNA
TTCATAAACGCGGGGTTTCGGTCCCGAGGGCTGGCACTCTGTGATACCCCA
2101
CCGAGACCCCATTTGGGGCAATACGCCGCGCTTCTCTCTTTTCCCCACC
CCAACCCCAAGTTCGGGTGAAGGCCAGGGCTCGACGCCAACGTCGGGG
2201
CGGCAAGCCCGCCATAGCCACGGGCCCGTGGGTTAGGGACGGGTCGCC
MET
CATGGGGAATGGTTTATGGTTCTGGGGGTTATTCTTTTGGGCGTTGCGT
2301
GGGGTCAGGTCCACGACTGGAGTGAACAGACAGCCCATGGTTTTTGGAT
GGCTGGGCATGGACCCGATGTAAGTGGCGACACGAACCCGGCGTCT
2401
GTGGCTGCCAACACCCCGACCCCAAAACACCACGCGGATTTCTGG
CGCGCCGGACGAACATAACCTGACTACGGCATCTCTGCCCTTCTTTCG
2501
TGGTACGAGGAGCGCTTTTGTGTTTGTATGTTGTCACCACGGCGAGTTTC
Pvu II
CGGGGACCCCGGCGAGCTGCTTTACATCTGAAAGACTACCTACTCGGCC
2601
GGCCCCCGAACGCGAGGCTGCCCGCCCCATCACGGTCGAGCCGACCGCC
CAGCCTCCCCCGCGGTGCGCCCCCTTAAGGGTCTCTTGACAAATCCAAC
2701
CGCCTCGGTGTTGCTGCGTT

```

FIG. 7. Nucleotide sequence from nucleotide 1,901 to nucleotide 2,720 of *Bam*HI fragment Q of HSV-1 strain Cl 101. Numbering starts at the *Bam*HI site closer to the start of the TK gene. The previous numbering system (21) started at nucleotide 318. The *Sma*I and *Pvu* II sites near the 3' end of the TK gene are shown, as is the 5' end of the 3KBL transcript. POLY A, Polyadenylic acid.

pears to be heterogeneous at the 3' end, which is consistent with this notion. The 3KBL RNA could be synthesized not from the nearby promoter, but from the TK promoter more than 1,300 nucleotides upstream. The 5' terminus of 3KBL then could be generated by recognition of the AATAAA sequence at nucleotide 2,022 and cleavage at the CA at nucleotide 2,054. Further processing by recognition of the AATAAA at nucleotide 2,009, cleavage at nucleotide 2,031, and polyadenylation would produce the mature 3' terminus of the TK mRNA. However, this possibility does not seem likely since it predicts that syntheses of TK mRNA and 3KBL would follow similar kinetics. Hybridizations to time course RNA blots have indicated that there is a 2-h delay in synthesis of 3KBL after TK mRNA synthesis and that TK mRNA is not detected at late times in infection, when the accumulation of 3KBL is maximal. Throughout the time course the expected 1:1 ratio of TK to 3KBL is not observed. However, this argument is made with a caveat regarding differential RNA stabilities (17). A second line of evidence comes from *in vitro* transcriptions, which suggest that the promoter sequences located within the 3' terminus of the TK mRNA are recognized by RNA polymerase II (18).

FIG. 8. Region between TK RNA and 3KBL RNA. Note that although transcription is shown from left to right here for purposes of clarity, these transcripts run from right to left on the standard P orientation of the genetic and molecular map of HSV-1, as shown in Fig. 2. POLY A, Polyadenylic acid.

ACKNOWLEDGMENTS

We thank Wilma P. Summers for generous help with virus and cell culture and Sheila Cronin and Leena Kalghatgi for expert technical assistance. James Smiley was most generous with his gift of recombinant plasmid DNAs.

This work was supported by Public Health Service grants CA 13515, CA 16038, CA 06519, and CA 09259 from the National Institutes of Health.

ADDENDUM IN PROOF

Wilkie et al. (Proc. R. Soc. London Ser. B 210:411–421, 1980) have recently reported related observations on some of the transcripts reported here.

LITERATURE CITED

- Alwine, J. C., D. J. Kemp, B. A. Parker, J. Reiser, J. Renert, G. R. Stark, and G. M. Wahl. 1979. Detection of specific RNAs or specific fragments of DNA by fractionation in gels and transfer to diazobenzyloxymethyl paper. *Methods Enzymol.* 68:220–242.
- Anderson, K. P., R. J. Frink, G. B. Devi, B. H. Gaylord, R. H. Costa, and E. K. Wagner. 1981. Detailed characterization of the mRNA mapping in the *Hind*III fragment K region of the herpes simplex virus type 1 genome. *J. Virol.* 37:1011–1027.
- Berk, A. J., and P. A. Sharp. 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell* 12:721–732.
- 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 inhibitors of reverse transcriptase. *Proc. Natl. Acad. Sci. U.S.A.* 76:731–735.
- Breathnach, R., and P. Chambon. 1981. Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50:349–383.
- Costa, R. H., G. B. Devi, K. P. Anderson, B. H. Gaylord, and E. K. Wagner. 1981. Characterization of a major late herpes simplex virus type 1 mRNA. *J. Virol.* 38:483–496.
- Cremer, K., M. Bodemer, and W. C. Summers. 1978. Characterization of the mRNA for herpes simplex virus thymidine kinase by cell-free synthesis of active enzyme. *Nucleic Acids Res.* 5:2333–2344.
- Enquist, L. W., G. F. Vande Woude, M. Wagner, J. Smiley, and W. C. Summers. 1979. Construction and characterization of a recombinant plasmid encoding the gene for the thymidine kinase of herpes simplex type 1 virus. *Gene* 7:335–342.
- Fitzgerald, M., and T. Shenk. 1981. The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* 24:251–260.
- Frink, R. J., K. P. Anderson, and E. K. Wagner. 1981. Herpes simplex virus type 1 *Hind*III fragment L encodes spliced and complementary mRNA species. *J. Virol.* 39:559–572.
- Holland, L. E., K. P. Anderson, C. Shipman, Jr., and E. K. Wagner. 1980. Viral DNA synthesis is required for efficient expression of specific herpes simplex virus type 1 mRNA species. *Virology* 101:10–24.
- Honess, R. W., and B. Roizman. 1974. Regulation of herpesvirus macromolecular synthesis. I. Cascade regulation of the synthesis of three groups of viral proteins. *J. Virol.* 14:8–19.
- Ludwig, R. A., and W. C. Summers. 1975. A restriction fragment analysis of the T7 left-early region. *Virology* 68:360–373.
- Maxam, A., and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U.S.A.* 74:560–564.
- McKnight, S. L. 1980. The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. *Nucleic Acids Res.* 8:5949–5964.
- McMaster, G. K., and G. G. Carmichael. 1977. Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proc. Natl. Acad. Sci. U.S.A.* 74:4835–4838.
- Persson, H., H.-J. Monstein, G. Akusjarvi, and L. Philipson. 1981. Adenovirus early gene products may control viral mRNA accumulation and translation *in vivo*. *Cell* 23:485–496.
- Read, G. S., and W. C. Summers. 1982. *In vitro* transcription of the thymidine kinase gene of herpes simplex virus. *Proc. Natl. Acad. Sci. U.S.A.* 79:5215–5219.
- Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg. 1977. Labeling *in vitro* by nick translation with DNA polymerase I. *J. Mol. Biol.* 113:237–251.
- Rosen, J. M., S. L. C. Woo, J. W. Holder, A. R. Means, and B. W. O'Malley. 1975. Preparation and preliminary characterization of purified ovalbumin messenger RNA from hen oviduct. *Biochemistry* 14:69–78.
- Wagner, M. J., J. A. Sharp, and W. C. Summers. 1981. Nucleotide sequence of the thymidine kinase gene of herpes simplex type 1. *Proc. Natl. Acad. Sci. U.S.A.* 78:1441–1445.
- Watson, R. J., and J. B. Clements. 1980. A herpes simplex virus type 1 function continuously required for early and late virus RNA synthesis. *Nature (London)* 285:329–330.
- Watson, R. J., C. M. Preston, and J. B. Clements. 1979. Separation and characterization of herpes simplex virus type 1 immediate-early mRNA's. *J. Virol.* 31:42–52.
- Watson, R. J., M. Sullivan, and G. F. Vande Woude. 1981. Structures of two spliced herpes simplex virus type 1 immediate-early mRNA's which map at the junctions of the unique and reiterated regions of the virus DNA S component. *J. Virol.* 37:431–444.
- Wellauer, P. K., I. B. Dawid, D. E. Kelley, and R. P. Perry. 1974. Secondary structure maps of ribosomal RNA. *J. Mol. Biol.* 89:397–407.