Identification of Sequences in the Herpes Simplex Virus Thymidine Kinase Gene Required for Efficient Processing and Polyadenylation.

Charles N. Cole  
*Dartmouth College*

Terryl P. Stacy  
*Dartmouth College*

Follow this and additional works at: https://digitalcommons.dartmouth.edu/facoa

Part of the Medical Biochemistry Commons, Medical Cell Biology Commons, Medical Genetics Commons, and the Virus Diseases Commons

Dartmouth Digital Commons Citation  
https://digitalcommons.dartmouth.edu/facoa/1279

This Article is brought to you for free and open access by the Faculty Work at Dartmouth Digital Commons. It has been accepted for inclusion in Open Dartmouth: Peer-reviewed articles by Dartmouth faculty by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.
Identification of Sequences in the Herpes Simplex Virus Thymidine Kinase Gene Required for Efficient Processing and Polyadenylation

CHARLES N. COLE* AND TERRYL P. STACY

Department of Biochemistry, Dartmouth Medical School, Hanover, New Hampshire 03756

Received 8 March 1985/Accepted 21 May 1985

The herpes simplex virus (HSV) type I thymidine kinase gene (tk) was resected from its 3' end with BAL 31 exonuclease. Two sets of plasmids were isolated that lacked information distal to the two copies of the hexanucleotide 5'-AATAAA-3' located at the 3' end of the HSV tk gene. The presence of a simian virus 40 origin of DNA replication in each plasmid facilitated analysis of patterns of transcription in transfected Cos-1 monkey cells. Transcription analyses were performed with an S1 nuclease protection assay. Efficient processing and polyadenylation at the normal site still occurred when all sequences more than 44 or 46 base pairs (bp) downstream from the first AATAAA were removed (pTK311R/SV010 and pTK209R/SV010). Removal of an additional 7 bp (pTK312R/SV010) decreased the amount of tk mRNA processed at that normal site, and tk mRNA polyadenylated at a cryptic site within pBR322 sequences began to appear. The normal processing and polyadenylation was not used at all when an additional 12 bp was removed (pTK314R/SV010); the small amount of tk mRNA produced was processed and polyadenylated at the cryptic pBR322 site. The region of the tk gene critical for efficient processing and polyadenylation of tk mRNA is located 20 to 38 bp downstream from the first AATAAA, distal to the polyadenylation site, and as RNA can form a stem-loop structure containing AAUAAA. Similar G+T-rich elements were located in DNA fragments which substitute efficiently for the HSV tk processing and polyadenylation signal and were not found in AATAAA-containing DNA fragments which substitute inefficiently for the HSV tk signal.

The biosynthesis of eucaryotic mRNA requires a series of processing steps which convert the precursor heterogeneous nuclear RNA into functional mRNA (see reference 38 for a review). These processing steps include the addition of caps at the 5' end of the mature mRNA, the removal of intervening sequences by splicing, and the addition of polyadenyllic acid at the 3' end of the mature mRNA. Although transcription by RNA polymerase II initiates at the cap site, evidence from several systems (17, 18, 22, 39) indicates that transcription extends beyond the site of polyadenylation and that endonucleolytic cleavage creates the site at which a poly(A) polymerase adds approximately 200 A residues.

The hexanucleotide 5'-AAUAAA-3' forms at least a part of the signal for processing and polyadenylation. AATAAA or a closely related variant sequence is located 8 to 32 bases upstream from the actual site of processing and polyadenylation in the vast majority of gene sequences examined (5, 8, 13, 23, 44, 49, 51). When this hexanucleotide sequence is deleted from the late transcription unit of simian virus 40 (SV40), the ability to produce late mRNA is lost (16). We showed previously that processing and polyadenylation of the herpes simplex virus (HSV) type I thymidine kinase gene (tk) ceased when the gene was resected from its 3' end to a point where the AATAAA was deleted (12). Furthermore, in several systems, single-base mutations of the AATAAA hexanucleotide reduce the level of processed polyadenylated mRNA to 1 to 10% of the level observed with the wild-type signal (36, 56). Under these conditions, all of the mRNA produced is both processed and polyadenylated, reflecting the tight coupling in vivo between processing and polyadenylation.

Many genes contain internal copies of AATAAA that do not appear to function as processing and polyadenylation signals (15, 43, 46, 57). This suggests that AATAAA alone may not be a sufficient signal for processing and polyadenylation. Examination of DNA sequences in the vicinity of both functional polyadenylation signals and unused AATAAAAs reveals no obvious homologies other than the AATAAAs. SV40 contains such an unused AATAAA in the middle of the gene for large T antigen. Although this hexanucleotide does not appear to signal processing and polyadenylation in cells infected with or transformed by SV40 (45; G. Khoury, personal communication; S. M. Weissman, personal communication), insertion of an 88-base-pair (bp) DNA fragment containing this hexanucleotide into a DNA construction containing the promoter and coding region of the tk gene restored the production of polyadenylated [poly(A)]n tk mRNA (12). Although the level of tk mRNA produced was only 5% of the level observed with the wild-type tk gene, tk mRNA polyadenylated at a site approximately 20 bases downstream from the AAUAAA could be detected easily. This AAUAAA could have become functional either by its coming into proximity with sequences in the tk construction, which provided a functional, although weak, processing and polyadenylation signal or by separation from sequences within SV40 which prevented it from serving as a processing and polyadenylation signal. This latter explanation includes the possibility that this AAUAAA does not function in SV40 because it is located within a transcription unit containing another, much stronger, polyadenylation signal. In such cases, competition between two or more possible polyadenylation signals, each containing an AAUAAA, might often result in exclusive use of only one of the signals.

McDevitt et al. (31) examined the adenovirus type 2 E2A gene and found that sequences between 20 and 35 bases downstream from the processing site were essential for efficient processing and polyadenylation at the normal site. While 90% of the mRNA produced by the mutant containing 35 bp downstream from the polyadenylation site was

* Corresponding author.
polyadenylated at the normal site and only 10% was readthrough transcript, 90% of the mRNA produced was readthrough transcript when the mutant contained only 20 bp downstream from the polyadenylation site. The other 10% of the mRNA produced was processed and polyadenylated at the normal site. A similar result was reported by Gil and Proudfoot (20) with the rabbit beta-globin gene.

In this report, we describe experiments performed to investigate further the requirements for processing and polyadenylation of tk mRNA. The HSV tk gene (see Fig. 1) contains two copies of AATAAA located 46 to 51 and 59 to 64 bp downstream from the end of the tk coding region (32, 52). The tk gene was resected from its 3' end, and a series of constructs was isolated lacking different amounts of HSV sequence downstream from the AATAAA hexanucleotides. tk mRNA production was examined in Cos-1 cells transfected by these constructions. Efficient production of poly(A)^+ tk mRNA required not only AATAAA but also a G+T-rich sequence located just downstream from the processing and polyadenylation site. The entire coding region of the tk gene could be deleted without any effect on either the amount or 5' or 3' ends of the tk mRNA produced, indicating that internal sequences are not important for tk mRNA biosynthesis.

**MATERIALS AND METHODS**

Usage of abbreviations for TK. We follow the convention of Ostrander et al. (41), which uses the abbreviations TK^+ and TK^- for the phenotype of cell lines and tk for the genotype or when referring to nucleic acid (DNA, mRNA) coding for thymidine kinase or derived from the thymidine kinase gene.

**Plasmids, bacteria, and plasmid DNA.** Established methods were used for transfection of bacterial cells (29), analysis of minilysate preparations of DNA (7), and preparation of purified plasmid DNA (10, 51). Some of the plasmids used in these studies were described previously (12). Plasmid pSV010 (37) was provided by R. Myers and R. Tjian, University of California at Berkeley. All plasmids were propagated in cells of Escherichia coli HB101.

**Construction of plasmid gene constructs.** The construction of plasmid pTK2 (see Fig. 1A), the resection of pTK2 with BAL 31 exonuclease to generate pTK201 through pTK209, and the insertion of the tk information from some of these plasmids into pSV010 (37), yielding pTK2/SV010 (see Fig. 1B), pTK206/SV010 (see Fig. 1C), and so forth have been described previously (12).

(i) Construction of pTK2/SPV010, pTK202/SPV010, etc. Vector pSPV010 DNA (15 μg) was digested to completion with BamHI in a volume of 100 μl. After digestion, the cohesive termini were filled in by adding 10 U of the large fragment of DNA polymerase I and the four deoxyribonucleoside triphosphates (final concentration of deoxyribonucleoside triphosphates, 10 μM) and incubating at room temperature for 10 min. DNA polymerase I was inactivated by heating the reaction mixture to 70°C for 15 min. After cooling to 37°C, DNA was digested with BglII. The DNA was subjected to electrophoresis in a 1.0% agarose gel in Tris-borate-EDTA buffer (TBE buffer; 89 mM Tris hydroxide, 89 mM boric acid, 2.5 mM EDTA [pH 8.3]), and the large fragment was isolated by electroelution in a dialysis bag in E buffer (40 mM Tris acetate [pH 8.0], 2.5 mM EDTA). pTK2, pTK202, pTK203, pTK204, pTK205, pTK206, and pTK209 DNAs were digested simultaneously with PvuII and BamHI and subjected to electrophoresis in a 1.0% agarose gel in TBE buffer. The smaller fragment, spanning most of the tk gene, was isolated by electroelution as described above. Each ligation mixture contained 15 μg of the DNA fragment spanning the tk gene per ml, 2 μg of the pSV010 vector fragment per ml, 600 μM ATP, and 5 U of T4 DNA ligase (Collaborative Research, Inc., Waltham, Mass.) in a total volume of 50 μl. After ligation overnight at 6°C, competent E. coli HB101 was transfected with ligated DNA, and colonies were selected on plates containing ampicillin (20 μg/ml). Minilysate preparations of DNA were screened by digestion with different restriction endonucleases. Approximately one-third of the samples examined did not contain a BamHI site which should have been formed by ligation of a PvuII blunt end with a filled-in BamHI site. Samples with the predicted restriction endonuclease cleavage patterns were identified, and purified DNA was prepared for subsequent use.

(ii) Reinsertion of pTK209/SPV010 with BAL 31 exonuclease. pTK209/SPV010 DNA was digested to completion with XbaI. This DNA was treated with BAL 31 exonuclease (New England Biolabs, Inc., Beverly, Mass.) under the following conditions: 200 μg of DNA per ml, 300 μg of bovine serum albumin per ml, 600 mM NaCl, 20 mM Tris chloride (pH 8.0), 12 mM CaCl2, 12 mM MgCl2, 1.0 mM EDTA, and 28 μl of BAL 31 exonuclease per ml. Digestion was permitted to proceed at 30°C, and at various times, portions were removed and transferred to tubes containing a twofold excess of the amount of EGTA [ethylene glycol-bis(beta-aminoethy1 ether)-N,N'-tetraacetic acid] needed to chelate completely the Ca^2+ in the reaction mixture. A small portion was withdrawn from each sample and digested with PstI. These PstI-digested samples were analyzed by electrophoresis on a 1.2% agarose gel in TBE buffer. Samples which had been digested by BAL 31 to the desired extent (8.75, 10, 11, or 12 min of digestion) were pooled, extracted twice with phenol-chloroform-isooamyl alcohol (25:24:1), extracted four times with ether, and ethanol precipitated with 2.2 volumes of ethanol. The ends of these treated linear molecules were made flush by incubation with the large fragment of DNA polymerase I and four deoxyribonucleoside triphosphates in the appropriate reaction buffer. After 15 min at room temperature, DNA polymerase was inactivated by incubation at 70°C for 15 min. Phosphorylated Xbal linkers were added, the DNA was digested with XbaI and ligated at a low concentration, which permitted polymerization and a codon pTK209 was transfected. Minilysate preparations of plasmid DNA were analyzed by restriction endonuclease digestion to identify colonies containing derivatives of pTK209/SPV010 which had lost approximately the desired number of nucleotides. Several were chosen for further characterization. These were designated pTK3110R/SPV010 through pTK3130R/SPV010.

(iii) Reinsertion of deleted plasmid and polylinker sequences in pTK3120R/SPV010, pTK3130R/SPV010, and pTK3140R/SPV010. In the construction of pTK3120R/SPV010 and related plasmids, the action of BAL 31 nuclease led to the removal of nucleotides in both directions beginning at the XbaI site. Since this created different plasmid and polylinker sequences at the boundary between HSV and non-HSV DNA, an additional set of plasmids was constructed that has the same polylinker-plasmid sequences at the HSV DNA–non-HSV boundary. Portions (20 μg) of pTK209/SPV010 DNA, pTK3120R/SPV010 DNA, pTK3130R/SPV010, and pTK3140R/SPV010 were each digested simultaneously with XbaI and BglIII, producing two fragments. Digested samples were subjected to electrophoresis in a preparative 1.0% agarose gel in TBE buffer. The large fragment of the digest of
pTK209R/SVO10 DNA, spanning all non-HSV sequences and the 5' portion of the HSV tk gene, and the small fragments of the other plasmids, spanning most of the HSV sequences, were isolated by electroelution in E buffer. Ligation mixtures contained 5 μg of the fragment per ml from pTK209R/SVO10, 15 μg of the small fragment from one of the other plasmids per ml, 600 μM ATP, ligation buffer (as recommended by the supplier), and 0.25 Weiss units of T4 DNA ligase in a total volume of 50 μl. After ligation overnight at 4°C, E. coli HB101 was transfected, colonies were spotted on plates containing ampicillin (25 μg/ml), and minilysates were analyzed by restriction endonuclease digestion. These plasmids were designated pTK412R/SVO10, pTK413R/SVO10, and pTK414R/SVO10.

Tissue culture cells and mammalian DNA. Previously described methods were used for the culture and maintenance of simian Cos-1 cells (12) and mouse LTK" APRT" (adenine phosphoribosyltransferase-negative) cells (48, 55). Carrier high-molecular-weight cellular DNA was prepared as described previously (48, 55, 56).

Transfection of Cos-1 cells, preparation of RNA, and S1 nuclease analysis. Cos-1 cells were subcultured 1:4 to 30 h before transfection. Cells were transfected as described previously (48) with DEAE-dextran (500 μg/ml). The frequency of transfection was increased to 20 to 30% by exposing cells to 100 μM chloroquine diphosphate (Sigma Chemical Co., St. Louis, Mo.) for 4 h after the removal of the DNA-DEAE-dextran solution (27). Cytoplasmic RNA was isolated 48 to 72 h later by the method of White et al. (53). Poly(A)+ mRNA was isolated by chromatography on columns of oligodeoxythymidylylcellose by the method of Aviv and Leder (4). S1 nuclease analysis of the transcripts was carried out by the method of Berk and Sharp (6) as modified by Favaloro et al. (14), as described in detail previously (12). Probes were 3' or 5' end-labeled at the BgIII, BssIII, or Smal sites by standard procedures as described previously (12). Approximately 50 to 100 ng of end-labeled probe was denatured with the RNA sample [20 to 40 μg of total cytoplasmic RNA or 1 to 2 μg of poly(A)+ RNA containing 25 μg of carrier Saccharomyces cerevisiae tRNA] at 90°C for 15 min, annealed for 12 to 18 h at temperatures listed in the figure legends, and treated with S1 nuclease at room temperature for 45 min, followed by ethanol precipitation, and electrophoresis in either alkaline agarose or acrylamide-7 M urea gels. Gels were fixed in 7% trichloroacetic acid, dried, and exposed to Kodak XAR-5 film with an intensifying screen (Cronex Lightning-Plus; Du Pont Co., Wilmington, Del.) for 12 h to 2 days. To ensure that hybridizations were conducted in excess DNA, a constant amount of RNA from cells transfected by pTK2R/SVO10 was annealed with different amounts of probe (2 to 200 ng) and analyzed as described above. In all experiments, the amount of probe used was at least threefold greater than the amount needed to give a maximum signal, indicating that all hybridizations were performed in excess DNA. Additional control experiments (data not shown) demonstrated that probes end labeled within the tk gene hybridized only with tk mRNA; no transcripts originating from the other DNA strand were detected. Therefore, it was not necessary to separate the probe for tk mRNA from the probe which could hybridize to mRNAs originating from the opposite DNA strand.

DNA sequence analysis. DNA sequence analysis was performed by the chemical modification method of Maxam and Gilbert (30).

Alternate method for mapping resection endpoints. In some cases, the endpoint of BAL 31 resection was determined indirectly rather than by DNA sequence analysis. Plasmid DNA (e.g., pTK315R/SVO10) was digested with XbaI, 3' end-labeled with the large fragment of DNA polymerase I, unlabeled dATP, dTTP, and dCTP, and alphar-[32P]-dGTP in standard buffer conditions, and digested with MspI, and the digested DNA was subjected to electrophoresis in an 8% acrylamide-7 M urea gel in TBE buffer with a sequencing ladder. Alignment of the MspI-XbaI fragments with the sequencing ladder allowed the exact length of the MspI-XbaI fragment to be determined, thus marking the resection endpoint.

Gene transfer. The transformation of LTK" cells to TK+ cells was performed by modifications of the method of Wigler et al. (55, 56) as described previously (48).

Enzymes and chemicals. Enzymes were purchased from New England BioLabs, Beverly, Mass.; Pharmacia/P-L Biochemicals, Milwaukee, Wis.; Boehringer Mannheim Biochemicals, Indianapolis, Ind.; or Collaborative Research, Inc., Waltham, Mass., and were used according to the directions of the supplier. All radiochemicals were purchased from Amersham Corp., Chicago, Ill. Oligodeoxythymidylylcellose was purchased from Collaborative Research.

RESULTS

Construction of plasmids containing the wild-type and 3' deletions of the HSV thymidine kinase gene. The HSV type 1 thymidine kinase gene has been cloned and sequenced (32, 52), and its upstream regulatory regions have been subjected to detailed molecular genetic analysis (33, 34). The gene encodes a 1.4-kilobase (kb) mRNA which is synthesized at intermediate times after infection of cells by HSV. The tk gene contains several unique restriction endonuclease cleavage sites (Fig. 1) which we have used in the preparation of probes for analysis of tk mRNA by S1 nuclease analysis. The BgIII site near the 5' end of the tk gene was used for preparation of both 5' and 3' end-labeled probes. The unique Smal and BssIII sites near the end of the tk coding region were used for the preparation of 3' end-labeled probes.

Figure 1 also shows the structure of several families of plasmids used in these studies. All regulatory signals needed for production of tk mRNA are located within a 2.2-kb PvuII subfragment of HSV DNA (11, 25) which we previously cloned into pBR328 and have called pTK2 (Fig. 1A). Although synthesis of tk mRNA in HSV-infected cells requires the action of early gene products, tk mRNA is synthesized efficiently in Cos-1 cells transfected by plasmids containing an SV40 origin of DNA replication and the PvuII fragment of HSV DNA containing the tk gene (pTK2/SVO10, Fig. 1B) (12). (Cos-1 cells contain an integrated copy of a replication-defective SV40 genome and express SV40 large T antigen constitutively [21].)

The nucleotide sequence of the 3' end of the HSV tk gene is shown in Fig. 1E. The gene contains two copies of the hexanucleotide 5'-AATAAA-3'. We previously described the resection of the HSV tk gene from its 3' end, beginning at the downstream PvuII site, and the isolation of a series of deleted clones, pTK201 through pTK209, lacking different amounts of sequence downstream from the tk coding sequences. One of these deletions, pTK206, lacks both copies of the AATAAA. When it was transferred to a vector containing an SV40 origin of replication, yielding pTK206/SVO10 (Fig. 1C), the deleted gene was unable to direct the synthesis of stable polyadenylated tk mRNA in Cos-1 cells, demonstrating a role for the AATAAA
hexanucleotides or downstream sequences or both in processing and polyadenylation of tk mRNA.

The tk information from several deleted clones, pTK201 through pTK206 and pTK209, was transferred to a vector capable of replication in Cos-1 cells. The resulting plasmids were designated pTK2R/SV010, pTK202R/SV010 (Fig. 1D), etc. The R indicates that the tk information is in the reverse orientation from that of pTK2/SV010 (Fig. 1B). The amount of HSV DNA present in these deleted clones was determined by DNA sequence analysis or restriction endonuclease analysis and is listed in Table 1.

Analysis of tk mRNA production in Cos-1 cells transfected with tk gene constructions. To determine whether any of the 3' deletion mutants of the tk gene which retained the AATAAAAs would be impaired in their ability to direct the synthesis of poly(A)+ tk mRNA, Cos-1 cells were transfected with the wild-type tk gene (pTK2R/SV010) and various deletion derivatives of the tk gene. Cos-1 cells were exposed to DNA (1 µg/100-mm plate) in the presence of DEAE-dextran, followed by treatment of transfected monolayers with 100 µM chloroquine diphosphate for 4 h, to increase the efficiency of transfection. RNA was isolated 48 h after transfection and analyzed (Fig. 2). The levels of tk mRNA were similar and the polyadenylation site was identical in all cases except that no tk mRNA was detected in cells transfected by pTK206/SV010, which lacks AATAAA.

FIG. 1. Structure of plasmids containing the HSV tk gene and derivatives of the tk gene. (A) pTK2 contains the entire tk gene on a 2.2-kb PvuII fragment inserted into the PvuII site of pBR328. (B) pTK2/SV010 contains the same 2.2-kb PvuII fragment of HSV DNA. It was constructed by taking the PvuI-to-BamHI fragment of pTK2 and ligating it with a BamHI-to-PvuI fragment of pSV010. Thus, pTK2/SV010 contains, starting at the PvuII site, 2.2 kb of HSV DNA, 1,692 bp of pBR328 (PvuI-to-BamHI site), the SV40 origin of replication region, pBR328 sequences (between nucleotide 2440 of pBR328 and the PvuI site at nucleotide 3738), and pBR328 sequences (941 bp) between the PvuI site and the PvuII site of pBR328. (C) pTK206/SV010 is similar to pTK2/SV010 but lacks the most distal 590 bp of HSV DNA, including both copies of 5'-AATAAA-3' and also lacks the 1,692-bp segment of pBR328 DNA between the PvuII and BamHI sites. (D) pTK202R/SV010 and related plasmids contain tk gene derivatives in an opposite orientation from those shown in panels B and C. They contain, proceeding clockwise from the BamHI site, the same SV40 origin of replication region, pBR328 from nucleotide 2440 to 4362, and a polylinker region into which the tk gene and derivatives of the gene have been inserted. (E) Nucleotide sequence of the 3' end of the HSV tk gene. Two copies of the hexanucleotide 5'-AATAAA-3' are shown. The arrow in this area indicates the region of processing and polyadenylation. The sequence endpoints of various resected tk genes are indicated. The AATAAA sequence functioning in processing and polyadenylation is enclosed in a box. The G+T-rich block is enclosed in a stippled box.
TABLE 1. Transformation of LTK\(^-\) cells to TK\(^+\) by using derivatives of the HSV type 1 thymidine kinase gene

<table>
<thead>
<tr>
<th>Construct</th>
<th>No. of expts(^a)</th>
<th>Relative gene transformation frequency(^b)</th>
<th>Distance from AATAAA to BamHI site(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pTK2</td>
<td>4</td>
<td>100</td>
<td>552</td>
</tr>
<tr>
<td>pTK201</td>
<td>3</td>
<td>111</td>
<td>175</td>
</tr>
<tr>
<td>pTK202</td>
<td>2</td>
<td>65</td>
<td>160</td>
</tr>
<tr>
<td>pTK203</td>
<td>3</td>
<td>80</td>
<td>140</td>
</tr>
<tr>
<td>pTK204</td>
<td>4</td>
<td>53</td>
<td>111</td>
</tr>
<tr>
<td>pTK205</td>
<td>3</td>
<td>16</td>
<td>70</td>
</tr>
<tr>
<td>pTK209</td>
<td>6</td>
<td>6</td>
<td>46</td>
</tr>
<tr>
<td>pTK206</td>
<td>5</td>
<td>1.4</td>
<td>-22</td>
</tr>
</tbody>
</table>

\(^a\) Several different tk gene constructs were examined in each experiment. Each experiment contained the wild-type tk gene, pTK2, as well as plates which received only carrier DNA. In each experiment, either 5 or 10 plates (100 mm) were exposed to DNA for each construct being tested.

\(^b\) Frequency of transformation is expressed relative to the frequency observed with the wild-type gene, pTK2, which was set at 100% in each experiment. The absolute transformation frequency for pTK2 in different experiments ranged from 1,000 to 9,500 colonies per \(\mu\)g of plasmid DNA.

\(^c\) The distance from the AATAAAA hexanucleotides is known precisely, from DNA sequence analysis, for pTK2, pTK204, pTK205, pTK206, and pTK2. For pTK201, pTK202, and pTK203, the distance was estimated from the mobilities in polyacrylamide gels of DNA fragments generated in a series of digestions with different restriction endonucleases. These estimates are accurate to \(\pm\)10 bp.

Removal of an additional 5 bp abolished the ability of the tk processing and polyadenylation signal to function, and all mRNAs detected were extended transcripts (pTK313R/SV010 through pTK319R/SV010; Fig. 3A, lanes 5 through 12). The extended tk transcripts amount to approximately 3 to 7% as much tk mRNA as was seen in cells transfected by pTK2R/SV010 (Fig. 3A, lane 1).

To map the 3' endpoint(s) of the extended transcripts, the same RNA preparations analyzed in Fig. 3A were annealed with homologous probes 3' end labeled at their BssHII sites. These RNA molecules had endpoints 300 to 350 bp from the BssHII site, within pBR322 sequences (upper arrow in Fig. 3B). The 3' ends of the extended transcripts map to the vicinity of nucleotide 4165 of pBR322. tk mRNAs with 3' and all downstream sequences. The appearance of multiple bands was due to an artifact with labeling of the probe; in other experiments (for example, see Fig. 3), only a single band was observed. Our S1 nuclease mapping data is consistent with processing and polyadenylation occurring at the site indicated in Fig. 1E, in agreement with the mapping data of McKnight (32). Nucleotide sequence analysis indicated that the clone with the least amount of sequence downstream from the polyadenylation site, pTK209R/SV010, contained 46 bp of tk DNA distal to the downstream AATAAA. This experiment has been repeated several times; the amount of tk mRNA produced in pTK209R/SV010-transfected cells ranges from 30 to 100% of the amount synthesized in pTK2R/SV010-transfected cells. This suggests that tk sequences missing from pTK209R/SV010 are not necessary for efficient processing and polyadenylation of tk mRNA. Analysis of nonpolyadenylated RNA revealed no distinct tk-related species (data not shown).

To determine whether the 46 bp of DNA between the AATAAAAs and the end of HSV information in pTK209R/SV010 contained sequences important for efficient processing and polyadenylation, pTK209R/SV010 (Fig. 1D) DNA was digested with XbaI and resected with BAL 31 exonuclease. Cultures with plasmids which had lost approximately 10 to 100 bp of DNA were identified by restriction endonuclease digestion. These plasmids were designated pTK311R/SV010, pTK312R/SV010, etc. The amount of HSV DNA removed was determined by DNA sequence analysis, and this is shown in Fig. 1E.

S1 nuclease analysis was performed on RNA prepared from Cos-1 cells transfected with these plasmids (Fig. 3). Probes were 3' end-labeled at their BssHII sites. Removal of 2 bp from pTK209R/SV010 did not affect the amount or precision of processing and polyadenylation (pTK311R/SV010; Fig. 3A, lane 3). When an additional 7 bp was removed, some of the tk mRNA continued to be processed and polyadenylated at the normal size, but some mRNA was an extended transcript (pTK312R/SV010; Fig. 3A, lane 4).

**FIG. 2.** Analysis of the level and processing and polyadenylation sites of tk mRNA present in Cos-1 cells transfected by different tk gene constructions. Cytoplasmic RNA isolated from transfected Cos-1 cells 48 h after transfection was mixed with a probe 3' end labeled at the Smal site. Control experiments indicate that all hybridizations were conducted in excess DNA. Hybridization was carried out at 57°C for 18 h, followed by treatment with S1 nuclease, as described in Materials and Methods. Hybrids were analyzed by electrophoresis in a 5% acrylamide–7 M urea gel in TBE buffer (89 mM Tris hydroxide, 89 mM boric acid, 2.5 mM EDTA [pH 8.3]). The gel was fixed in 7% trichloroacetic acid, dried, and exposed to Kodak XAR-5 film. Lanes: 1, pTK2R/SV010; 2, pTK206R/SV010; 3, pTK202R/SV010; 4, mock-transfected; 5, pTK205R/SV010; 6, pTK209R/SV010. Lane m contained \(^{32}P\)-labeled size markers.
ends at the correct site within HSV DNA are indicated by the lower arrow in Fig. 3B. This experiment was also performed with cytoplasmic poly(A)⁺ mRNA isolated by chromatography on oligodeoxythymidylate-cellulose. All of the mRNA species detected in Fig. 3B appeared in the poly(A)⁺ fraction, while no discrete tk mRNA species were detected in the nonpolyadenylated fraction (data not shown).

It is possible that polyadenylation near nucleotide 4165 of pBR322 is occurring in response to the combined action of AATAAA within HSV sequences and other sequences found within pBR322. To examine this possibility, cells were transfected with pTK206R/SVO10, which lacks both copies of the AATAAAs and all downstream HSV sequences. Cytoplasmic RNA was isolated 48 h later, and the poly(A)⁺ fraction was isolated on oligodeoxythymidylate-cellulose. An S1 nuclease protection experiment was performed with a probe pTK206R/SVO10 DNA 3' end labeled at its SmaI site (Fig. 1). It can be seen (Fig. 3C, lane 1) that pTK206R/SVO10-transfected cells contained poly(A)⁺ tk mRNAs whose 3' ends mapped near residue 4165 of pBR322 (large arrow in Fig. 3C; smaller arrows indicate location of minor mRNA species). Although there is no AATAAA near residue 4165 of pBR322, we conclude that a functional but weak polyadenylation signal is located within pBR322 sequences in this vicinity. As expected, tk mRNA from pTK206R/SVO10 protected only 56 bases when the probe was derived from pTK206R/SVO10 3' end labeled at the same site (Fig. 3C, lane 2). pTK206R/SVO10 DNA is homologous to pTK2R/SVO10 from the SmaI sites to the boundary between HSV and pBR322 DNA 56 bp downstream from the SmaI site.

When pTK209R/SVO10 was resected from its XbaI site with BAL 31 exonuclease, information was removed in both directions from the XbaI site. Therefore, it is possible that the effects on processing and polyadenylation seen in Fig. 3 were due in part to loss of non-HSV sequences to the left of the XbaI site (Fig. 1D). Furthermore, each of the resected clones contained different sequences at the boundary between HSV DNA and vector DNA. Therefore, derivative constructions were prepared from representative resected clones (pTK312R/SVO10, pTK313R/SVO10, and pTK314R/SVO10) such that all new resected constructions would contain the same sequences at the boundary between HSV DNA and the vector. These constructions were designated pTK412R/SVO10, pTK413R/SVO10, and pTK414R/SVO10. Analysis of mRNA synthesized in Cos-1 cells transfected by these constructions gave results identical to those seen in Fig. 3 (data not shown). Therefore, non-HSV sequences adjacent to the XbaI site in pTK209R/SVO10 and deleted to various extents in the resected derivatives of pTK209R/SVO10 did not play any role in processing and polyadenylation of tk mRNA.

Do sequences within the tk gene or within pBR322 affect efficiency of processing and polyadenylation? To determine whether sequences internal to the tk gene were essential for processing and polyadenylation, additional constructions were prepared in which much of the coding information of pTK2R/SVO10 and pTK209R/SVO10 was deleted. These plasmids, pTK2del/SVO10 and pTK209del/SVO10, lack tk sequences between the BglII and SmaI sites. Cos-1 cells were transfected with these plasmids and the parental plasmids, and tk mRNA was analyzed by S1 nuclease protection analysis (Fig. 4). Although the deletion plasmids contained neither the BglII nor the SmaI site found in the tk gene (Fig. 1), probes prepared from pTK2R/SVO10 and 3' end labeled at the SmaI site or 5' end labeled at the BglII site were protected by tk mRNA synthesized from these deletion templates. The internal deletion constructions directed the synthesis of the same approximate amount of tk mRNA (compare Fig. 4A and 4B; lanes 1 with lanes 2, lanes 3 with lanes 4) with the same 3' end points (Fig. 4A) and the same 5' end points (Fig. 4B) as the wild-type gene (pTK2R/SVO10) or pTK209R/SVO10. Furthermore, the 5' end analysis indicates that three
gene to transform mouse LTK− cells to TK+ is another measure of gene activity. LTK− cells were transfected with various tk gene constructions in the presence of carrier LTK− cellular DNA and calcium phosphate. Colonies were counted after 14 days. The results of several experiments are tabulated in Table 1. pTK201, which contains only 175 bp downstream from the two AATAAAs, transformed TK− cells to TK+ as well as the wild-type construction, pTK2. Removal of additional information resulted in a progressive decrease in frequency of transformation. pTK209 transformed only 6% as well as pTK2. The very low but positive frequency observed with pTK206, which lacks a processing and polyadenylation signal, probably reflects the frequency with which a signal from carrier DNA was joined to the pTK206 DNA during the transformation process. Similar results have been obtained with tk gene derivatives inserted into a Cos-1 cell vector (pTK2/SV010, pTK202/SV010, pTK206/SV010, etc.).

The finding that pTK209 was only 6% as effective as pTK2 at transforming LTK− cells, contrasts with the ability of pTK209/SV010 to produce approximately 30 to 100% as much tk mRNA as pTK2/SV010. This probably reflects the fact that multiple copies of the tk gene function to produce tk mRNA immediately after transfection of LTK− cells. Although cells transfected by pTK209 may have, on the average, half as much tk mRNA as cells transfected by pTK2, some cells will have much more and others much less tk mRNA. To survive hypoxanthine-aminopterin-thymidine selection probably requires a level of tk mRNA (protein) found only in a small minority of cells receiving pTK209 DNA and in a much greater fraction of cells taking up pTK2 DNA. Thus, there is a direct but nonlinear relationship between the level of tk mRNA produced by a series of constructs in Cos-1 cells and the ability of these constructs to transform LTK− cells to TK+.

**DISCUSSION**

Formation of mature eucaryotic mRNA molecules requires a detailed series of processing steps through which primary transcripts are converted into functional mRNAs. Available data from several systems indicates that transcription generally proceeds well beyond the polyadenylation site (17, 18, 22, 39), and for a few systems, regions have been identified in which transcription termination is thought to occur (9, 19, 22). Thus, processing is required to generate the site where poly(A) is added.

We showed previously that tk mRNA was not processed or polyadenylated when both copies of the hexanucleotide 5′-AATAAA-3′ and all downstream sequences were deleted from the HSV tk gene (12). When DNA fragments containing AATAAA were substituted for HSV sequences, processing and polyadenylation of tk mRNA was restored. These DNA fragments included several known to function as polyadenylation signals in their normal backgrounds and one, from the middle of the early region of SV40, which does not signal processing and polyadenylation in either the lytic or transforming infection by SV40. The levels of poly(A)+ tk mRNA produced in cells transfected by the construct with this alternate SV40 signal were low, suggesting that alone, the hexanucleotide AATAAA was a sufficient signal for processing and polyadenylation but that additional sequences might play a role in increasing the efficiency of the processing and polyadenylation reaction.

In another study, we inserted BamHI fragments of monkey genomic DNA into a tk construction lacking AATAAA...
and all downstream sequences (48). Of the DNA fragments, 75% (4 to 22.5 kb) were able to restore tk gene expression. In all positive cases, poly(A) + tk mRNA was synthesized in transfected cells; there were no cases in which nonpolyadenylated mRNA was formed. The frequency with which random DNA fragments were able to restore tk gene expression is consistent with a relatively simple sequence for a processing and polyadenylation signal.

Together, these results suggested the existence of other element(s) within the tk gene which acted to permit efficient processing and polyadenylation downstream from AATAAA. In this report, we have described experiments designed to locate these additional polyadenylation signal elements. Our results indicate that there is an element located between 20 and 38 bp downstream from the first AATAAA (Fig. 1E) which is essential for efficient processing and polyadenylation at the normal HSV tk site. This critical region of the tk gene is very G+T rich (18 of the 19 residues are G or T), and the last six of these could, as RNA, form a perfectly base-paired stem with AUAUAAA (Fig. 5).

Our data indicate that the normal processing and polyadenylation site was still used when 6 bp of this G+T-rich block was removed (pTK313R/SV010), but most of the transcripts continued on into pBR322 sequences (Fig. 3B, lane 3). Removal of an additional 7 bp from the G+T-rich block (pTK314R/SV010) resulted in cessation of use of the normal processing and polyadenylation site, and the small amount of transcript produced was polyadenylated at a site within pBR322 (Fig. 3B, lane 4). No further qualitative or quantitative changes occurred as additional sequences were removed. Although pBR322 does not contain an AATAAA in this vicinity, three hexanucleotides in this region differ from AATAAA at only a single base. A low level of polyadenylation does occur in gene constructs containing single-base mutations within AATAAA (36, 54), so one of these hexanucleotides may function as a weak processing and polyadenylation signal for tk mRNA.

Recently, analyses and findings similar to these have been reported for other mammalian genes. McDevitt et al. (31) reported the discovery of a role for the sequence between 34 and 49 bp from the AATAAA in the adenovirus type 2 E2A transcription unit. The critical region in this transcription unit contains a 12-bp sequence able to form an imperfect stem with an upstream sequence overlapping part of the AATAAA. This 12-bp sequence is A+T rich, in contrast to the G+T-rich sequence in the tk gene. A similar result was also demonstrated for the rabbit beta-globin gene (20). When 51 bp of DNA downstream from AATAAA was present, processing and polyadenylation occurred at the normal site and level; when only 15 bp downstream from AATAAA was present, processing and polyadenylation occurred much further downstream at a second copy of the beta-globin polyadenylation region. This portion of the globin gene contains a G+T-rich block (19 of 21 bases are G or T). Similar results were also obtained in studies with the SV40 late transcription unit (47). In contrast, a qualitatively different result was obtained from similar studies of the bovine growth hormone gene (58). In that system, deletion of information just downstream from AATAAA resulted primarily in a shift of the precise site of processing and polyadenylation but did not drastically reduce the total amount of processed and polyadenylated mRNA produced. A highly G+T-rich region is located just downstream from the processing and polyadenylation site of the bovine growth hormone gene (19 of 20 consecutive residues are G or T), and it is the loss of this region that alters the specificity of the processing and polyadenylation reaction in this system. Possibly there are multiple types of processing and polyadenylation signals, with signal elements (other than AATAAA) playing a qualitative role in some and a quantitative role in other systems. Clearly, the other elements needed for efficient processing and polyadenylation downstream from the HSV tk gene, rabbit beta-globin, and adenovirus E2A are not identical, though there may be some sequence similarities.

In our previous studies, we reported that a wide range of tk mRNA levels was produced in cells transfected by tk constructions with different polyadenylation signals (12). The DNA sequences of these constructions in the vicinity of the AATAAA are shown in Fig. 6. A G+T-rich block of sequence is located downstream from the AATAAA in those constructions which functioned efficiently in Cos-1 cells (SV40 early and late, polyomavirus early polyadenylation signal), while no highly G+T-rich block is found downstream from AATAAA in those constructions which produced much lower levels of poly(A) + tk mRNA (polyomavirus late, polyomavirus alternate, SV40 alternate polyadenylation signals). Thus, a role for a G+T-rich sequence in efficient processing and polyadenylation is consistent with our previous studies of polyadenylation signaled by several AATAAA-containing DNA fragments. A G+T-rich block is located in pBR322 DNA immediately downstream from the site at which tk extended transcripts were polyadenylated.

McLauchlan et al. (35) compared the DNA sequences of the 3' untranslated regions of more than 75 genes and found that 67% contained the consensus sequence 5'-YGGTGGY-3' 24 to 30 bp downstream from AATAAA. They note that the HSV tk gene contains a variant of this sequence, 5'-GTTGTTGG-3', downstream from AATAAA at the 5' end of the G+T-rich block shown in Fig. 1E. The results presented here suggest that sequences at the 5' end of the G+T-rich block are more important for efficient processing and polyadenylation than the sequences suggested by McLauchlan et al. (35). Deletion of the distal portion of this G+T-rich block, leaving the GGTGTTGG, resulted in an extremely low level of use of the normal processing and polyadenylation site. A variant of their consensus sequence (TATGTCCCT) is also found in the polyomavirus late transcription unit. However, the polyomavirus late polyadenylation signal functions inefficiently when attached to the tk coding region (12) and in polyomavirus, in which late mRNA is processed from giant nuclear precursors containing up to seven tandem transcripts of the polyomavirus genome (26). Furthermore, many of the sequences tabulated by McLauchlan et al. (35) display only moderate homology to their YGGTGGY consensus sequence. Many of these sequences do contain a G+T-rich element downstream from the AATAAA. We suggest that such a G+T-rich sequence is the element which, in combination with AATAAA, constitutes an efficient processing and polyadenylation signal.

In summary, we have shown that efficient processing and

\[
\text{FIG. 5. Possible secondary structure in tk mRNA including AAUAAA and the G+T-rich block needed for efficient processing and polyadenylation. The site of processing and polyadenylation is indicated by the arrow.}
\]
polyadenylation of HSV tk mRNA requires two distinct sequence elements—the highly conserved AATAAA and the distal portion of a G+T-rich segment located 20 to 38 bp beyond the AATAAA. This second element is not highly conserved among different genes, and in many, nothing resembling this G+T-rich element can be found. We suggest that the efficiency with which a given polyadenylation signal functions and the choice between multiple polyadenylation signals when multiple AATAAAs occur in one transcription unit are governed by interactions between the AATAAAAs, downstream elements such as the G+U-rich element in HSV tk heterogenous nuclear RNA, and other factors. Furthermore, it is likely that secondary structures formed by the interactions of these elements and factors form the substrates for the processing and polyadenylation reactions.

Kaufman and Sharp (24) reported that polyadenylation signals may be involved in mediating cell-cycle dependent changes in the levels of dihydrofolate reductase mRNA. Similarly, different polyadenylation sites are used in immunoglobulin genes at different stages of differentiation (2, 3, 28), allowing the formation of membrane-bound immunoglobulins at early stages of maturation and secreted immunoglobulin in more mature cells. In the major late transcription unit of adenovirus type 2, there are five processing and polyadenylation sites. Entirely different patterns of site selection occur early and late during the infection cycle (40). Since the AUAAs moiety is almost universal in mRNAs from higher eucaryotes, other elements and factors essential for processing and polyadenylation probably show much greater variation, thus permitting regulation of gene expression by choice of polyadenylation sites. It is likely that differential expression of protein and RNA factors involved in processing and polyadenylation are central to a role for processing and polyadenylation in gene regulation.


