JOURNAL OF VIROLOGY, Dec. 1991, p. 6677–6685
0022-538X/91/126677-09$02.00/0

Vol. 65, No. 12

Simian Virus 40 Late mRNA Leader Sequences Involved in Augmenting mRNA Accumulation via Multiple Mechanisms, Including Increased Polyadenylation Efficiency

HENRY C. CHIOU, CHRISTINE DABROWSKI,† AND JAMES C. ALWINE*

Department of Microbiology, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6148

Received 14 June 1991/Accepted 29 August 1991

We have examined the contribution of 5′ leader sequences to expression directed by the simian virus 40 (SV40) late promoter. These studies showed that addition of sequences which contain the late leader 3′ splice site to the late promoter led to an increase in the accumulation of mRNA expressed by the promoter. No other sequences within the leader region, between SV40 positions 334 and 560, exhibited a substantial influence on mRNA accumulation. The increase was due, at least in part, to the creation of a spliceable mRNA transcript, since mutation of either the 5′ or 3′ splice site could attenuate the effect. However, sequences at or near the 3′ splice site appeared to play a more important role than did the 5′ splice site in bringing about this increase. In many instances, mutation of the 3′ splice site also led to the accumulation of extended transcripts, whereas mutation of the 5′ splice site did not produce this result in any instance. Analysis of these extended transcripts showed that they retained sequences normally lost upon cleavage and polyadenylation. This finding suggested that mutation of the 3′ splice site sequence led to decreases in the efficiency of polyadenylation. We propose that the SV40 late leader sequences positively contribute to expression of the viral late genes by increasing mRNA accumulation via multiple mechanisms, including the enhancement of pre-mRNA polyadenylation efficiency.

The simian virus 40 (SV40) late gene leader sequences have been shown to be important for the biogenesis of viral late mRNAs. In particular, it has been proposed that splicing within the leader plays a critical role in expression. Inhibition of splicing by deletion of late leader sequences in some cases diminished and in other cases eliminated the accumulation of certain late mRNA species (15–17, 23, 28, 32). Insertion of heterologous intervening sequences into splicing-deficient viral recombinants restored late mRNA production (15, 17, 23). Various studies have suggested that the defect in viral late mRNA accumulation brought about by inhibition of splicing is at the level of posttranscriptional processing, since no link has been found between splicing and the rate of SV40 late gene transcription (16, 17, 23). Suggested mechanisms include splicing-induced enhancement of polyadenylation, mRNA transport, and transcript stability (16, 17, 28, 32).

The ability of intervening sequences to augment mRNA accumulation has been observed in non-SV40 genes as well, including the rat growth hormone (8), rabbit β-globin (9), mouse dihydrofolate reductase (9), and maize alcohol dehydrogenase-1 (10) genes, and in several chimeric genes (20, 24). Thus, it may be an inherent property of the splicing process, or of splice signals, to boost expression of those genes which contain intervening sequences.

The previously reported studies that revealed the significant contribution of splicing to accumulation of SV40 late mRNAs were accomplished by using large deletions through the late leader region or by complete removal of intervening sequences. Therefore, any additional effects, independent of splicing, contributed by sequences included in these large deletions could not be assessed, nor could they be distinguished from the effects caused by inhibition of splicing. However, several studies have suggested that certain sequences within the leader region are involved in regulation of late gene expression via mechanisms unrelated to splicing (1, 4, 19). We therefore initiated these studies in order to define more precisely the sequences within the 5′ leader that influence expression of the viral late genes and to determine whether these leader sequences function via processes other than splicing.

Our results suggest that only the addition of sequences which lead to the creation of a functional intervening sequence significantly increases mRNA accumulation. Introduction of mutations at either the 5′ or 3′ splice site resulted in attenuation or elimination of these increases. However, specific splice site sequences also appeared to influence mRNA accumulation independently of the process of splicing. Specifically, mutation of the 3′ splice site produced a much greater effect than did mutation of the 5′ splice site. We propose that splicing and the presence of splice site sequences lead to an increase in mRNA accumulation via multiple mechanisms. One of these mechanisms is the enhancement of pre-mRNA polyadenylation efficiency due to the presence of 3′ splice site sequences.

MATERIALS AND METHODS

Plasmids. The construction, isolation, and preparation of plasmids were carried out by using standard procedures (3, 29). The promoter and leader sequences of all plasmid constructions were confirmed by sequence analysis. The SV40 nucleotide numbering system described by Buchman et al. (in reference 31), here designated "SV," was adopted to describe all SV40 DNA inserts and fragments used in this study. Plasmid pLS102 is composed of the chloramphenicol acetyltransferase (CAT) gene flanked by the SV40 late gene promoter and polyadenylation site. The promoter is con-

* Corresponding author.
† Present address: Dana-Farber Cancer Institute, Boston, MA 02115.

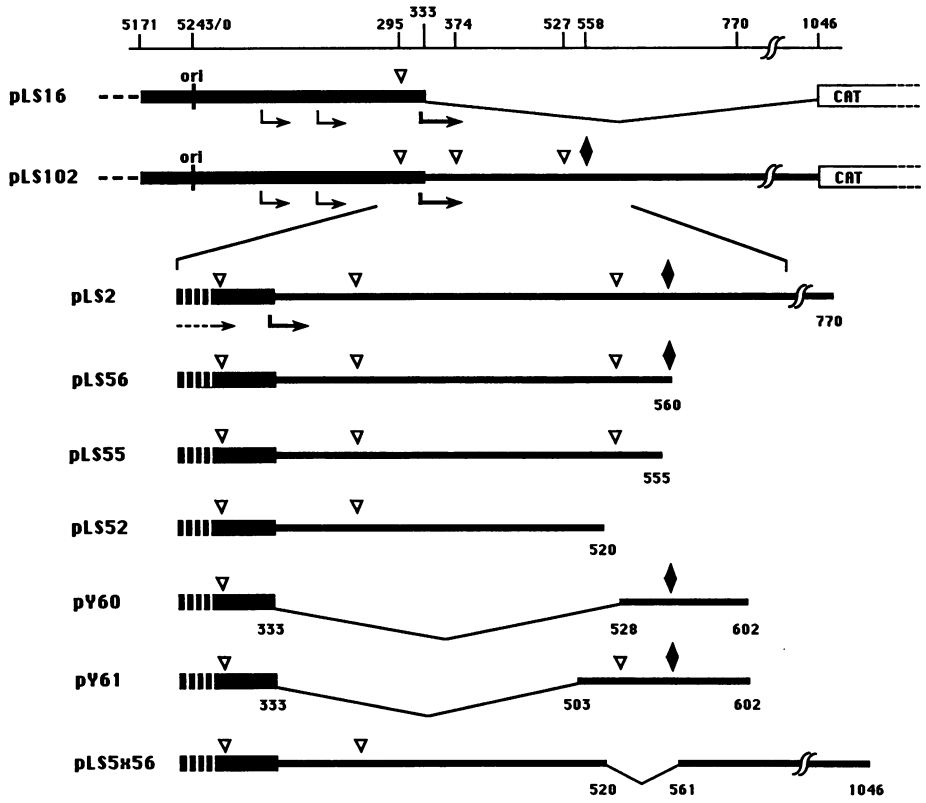

FIG. 1. Schematic representation of the SV40 late promoter and 5′ leader sequences contained in the plasmids used in this study. The top line shows selected SV40 nucleotide position numbers (31) for viral sequences in pLS16 and pLS102, represented on the second and third lines. The thick bar in each plasmid map delineates the basal SV40 late promoter, while the thin bar represents viral sequences downstream of the basal promoter. The thick arrow below the promoter locates the position of the major transcription initiation site. The narrow arrows denote heterogeneous upstream transcription initiation sites. 5′ splice sites are depicted by inverted triangles; the 3′ splice site is represented by the solid diamond. In the following lines the plasmids are shown in slightly magnified detail. The numbers below each plasmid are the SV nucleotide positions at the ends of the bar above. The very thin lines in the bottom three plasmids show the leader sequences that have been deleted in those constructions.

tained within the *Hin*dIII C fragment derived from the viral genome, SV 5171 to 1046 (Fig. 1). This fragment also includes the viral origin of replication, the late gene leader sequence, and the 5′ ends of the VP2 and VP3 coding sequences. The SV40 late polyadenylation site is contained within a fragment spanning SV 2533 to 2770. Plasmid pLS102m is identical to pLS102 except that the AG dinucleotide at the 3′ splice site in the SV40 late leader was changed to GG at SV 556 by oligonucleotide-directed site mutagenesis. In plasmid pLS16, the SV40 *Hin*dIII C sequences downstream of the major late gene transcription start site, from SV 334 to 1046, were removed (Fig. 1). This plasmid thus contains what we consider the basal SV40 late promoter (22). Additional plasmids which contain various portions of the SV40 late leader and late coding region are described in Fig. 1.

Transfections and RNA extraction. Cells were maintained and propagated as previously described (22). COS cells were transfected with appropriate plasmids by calcium phosphate precipitation (3). Briefly, 3×10^5 to 5×10^5 cells were plated onto a 100-mm dish and incubated at 37°C overnight. Three hours prior to transfection, the medium (Dulbecco modified Eagle medium containing 7.5% fetal calf serum) was replaced with 10 to 12 ml of fresh medium. Then 0.5 ml of transfection mix containing 10 μg of DNA, 125 mM $CaCl_2$, 25 mM *N*,*N*-bis(2-hydroxyethyl)-2-aminoethanesulfonic acid, 140 mM NaCl, and 0.75 mM Na_2HPO_4, pH 6.95, was added to the medium, and the cells were incubated at 37°C overnight. At 20 to 24 h posttransfection, the medium containing the transfection mix was removed, the cells were washed extensively with phosphate-buffered saline solution, and 12 ml of fresh medium was added. RNA was harvested after an additional 24 h of incubation (i.e., at 48 h posttransfection).

Total cellular RNA was extracted by washing the cell monolayer three times in an ice-cold solution of 30 mM Tris-HCl (pH 7.5)–0.15 M NaCl–1.5 mM $MgCl_2$ (TSM buffer) and scraping the cells into a 1:1 mix of TSM buffer and 10 mM Tris-HCl (pH 7.5)–1 mM EDTA–0.4 M NaCl–0.2% sodium dodecyl sulfate–7 M urea. The lysed cell suspension was extracted three to four times with an equal volume of 1:1 phenol-chloroform; the nucleic acids were precipitated with ethanol and treated with RNase-free DNase (Boehringer Mannheim).

Northern (RNA) analysis. For Northern analysis, 10 μg of each RNA sample was denatured in 50% formamide–1.1 M formaldehyde, subjected to electrophoresis through a 1.5% agarose gel containing 20 mM 3-[*N*-morpholino]propanesulfonic acid (MOPS), 5 mM sodium acetate, 1 mM EDTA, and 1.1 M formaldehyde, and then transferred to a nylon

membrane (MSI Magnagraph) (3). The membranes were subsequently baked in a vacuum at 85°C for 1.5 to 2 h. Prehybridization and hybridization, carried out at 75°C, and washing of blots were performed as previously described (11). The hybridization probes were ^{32}P-labeled RNAs transcribed by SP6 or T7 RNA polymerase (Bethesda Research Laboratories), using the pGEM system (Promega). These probes were complementary to CAT coding sequences. CAT RNA transcripts generated from the various transfected plasmids were visualized by autoradiography. Quantitations were performed by use of either a Molecular Dynamics scanning densitometer or a Molecular Dynamics PhosphoImager.

S1 nuclease and RNase protection assays. To examine splicing of RNA transcripts, single-stranded ^{32}P-labeled DNA probes were used for S1 nuclease analysis. These probes were generated directly from the same plasmids used in transfections to produce the RNAs to be analyzed. They served as templates for primer-directed DNA synthesis catalyzed by modified T7 DNA polymerase (Sequenase; U.S. Biochemical). The primer used to generate these probes was complementary to the sequence on the noncoding strand of the CAT gene from 98 to 117 nucleotides downstream of the SV40 promoter/leader-CAT gene junction. Two probes were prepared from each template. The first probe was synthesized in the continuous presence of [^{32}P]dATP and thus contained label throughout its length. The second probe was synthesized by adding [^{32}P]dATP for the first 5 min of the synthesis reaction and then adding a large excess of cold nucleotides to complete the reaction. In this case, label was incorporated only into the 5′ 50 to 80 nucleotides of the newly synthesized strand, entirely within CAT coding sequences. The newly synthesized DNA probes were cleaved at the SV40 origin of replication with *BglI* so that the probe spanned the entire SV40 late promoter and leader region from the origin, through the SV40-CAT junction, into the body of the CAT coding sequence. The single-stranded probes were isolated from the plasmid template by denaturing agarose gel electrophoresis.

For analysis of RNA 3′ ends, RNase protection assays were carried out with a ^{32}P-labeled RNA probe generated by T7 RNA polymerase transcription in the same manner as the probes made for Northern analysis. The probe was complementary to the sequence within plasmid pLS102 from 141 nucleotides upstream to 181 nucleotides downstream of the SV40 late polyadenylation site (see Fig. 5A). Both S1 nuclease and RNase protection analyses were carried out according to standard protocols (3). Ten micrograms of sample RNA was used in each S1 reaction, while 20 μg of each sample RNA was used in the RNase protection assays.

RESULTS

The late leader region enhances mRNA expression from the SV40 late promoter. Plasmid pLS16 contains what we have designated the basal SV40 late promoter, directing expression of the CAT gene. The promoter ends at SV 333, 8 nucleotides downstream of the major late transcription initiation site (Fig. 1), and contains most of the constitutive and T-antigen-inducible late promoter elements identified previously (2, 4–6, 18, 21, 22). COS cells were transfected with plasmid pLS16, or with pLS102, which contains the basal promoter plus the 713 nucleotides which immediately follow in the viral genome, from SV 334 to 1046. These additional sequences contain the viral late leader and portions of the late coding region. Total cell RNA was harvested and analyzed by Northern blot hybridization to determine the amount of steady-state, properly processed CAT mRNA expressed by each plasmid. Cells transfected with plasmid pLS102 accumulated approximately fourfold more CAT mRNA than did cells transfected with pLS16 (Table 1 and Fig. 2). This finding suggested that sequences downstream of the major transcription initiation site at SV 325 function to augment expression directed by the late promoter.

TABLE 1. Quantitation of expected-size and extended CAT RNAs produced from SV40 late leader-containing plasmids

Plasmid	Expected-size transcripts[a]	Extended vs expected-size transcripts[b]
Late leader plasmids		
pLS16	1.00	0.25
pLS102	**4.18**	0.30
pLS2	**4.25**	0.76
pLS56	**9.89**	0.08
pLS55	1.90	**0.97**
pLS52	1.29	0.30
pY60	**6.54**	0.07
pY61	**6.66**	0.12
pLS5x56	1.43	0.25
3′ splice site mutations		
pLS102m	1.46	**2.28**
pY60m	0.39	0.97[c]
pY61m	0.78	**1.94**
5′ splice site mutations		
pLS16d	0.93	0.43[d]
pY60d	**3.46**	0.18[d]
pY61d	**2.86**	0.46[d]

[a] Ratio of the amount of CAT RNA of the expected size produced from each plasmid in transient expression assays to the amount of CAT RNA produced from pLS16. Numbers in bold denote a greater than twofold increase in RNA accumulation compared with that observed with pLS16.

[b] Ratio of CAT RNA of greater than expected size to that of the expected size produced from each plasmid in transient expression assays. Numbers in bold denote ratios that are more than threefold greater than the ratio observed with pLS102.

[c] This ratio is skewed upward by the low amount of expected-size CAT RNA produced from this plasmid.

[d] Represents one experiment done in duplicate. All other numbers are averages of three or more assays.

A variety of other plasmids, which contain only portions of the sequences downstream of the basal promoter that are included in pLS102, were used to identify the precise sequences that elicited the increase in mRNA accumulation (Fig. 1). Quantitation of CAT mRNA produced from these plasmids was restricted to the species, observed by Northern analysis, that corresponded in size to a transcript of the predicted length for a properly processed message. A number of plasmids also produced increased amounts of large CAT RNA species, which we discuss below.

The leader sequences included in plasmids pLS52 and pLS55, extending to SV 520 and 555, respectively, either produced no effect or at most induced a very modest increase over the amount of RNA present in cells transfected with pLS16 (Table 1 and Fig. 2). A significant increase in mRNA accumulation, however, was observed when the region of the viral leader between SV 555 and 560, which contains a 3′ splice site at SV 558, was included. Plasmid pLS56, which differs from pLS55 solely by these five nucleotides, exhibited a 5-fold increase in CAT mRNA accumulation compared with pLS55 and a 10-fold increase with respect to the basal promoter in pLS16. Thus, addition of the sequence between SV 555 and 560, which contains a 3′ splice site, led to a significant enhancement of mRNA expression

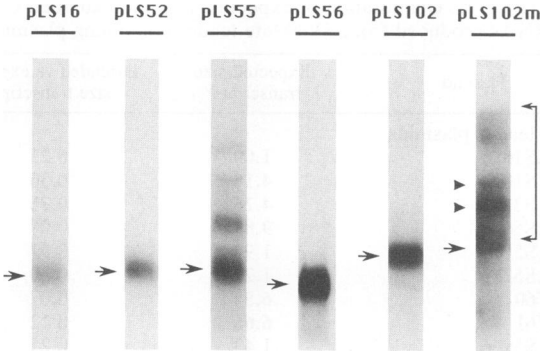

FIG. 2. Representative Northern analysis of RNA harvested from cells transfected with pLS16 and various late leader-containing plasmids. The arrow at the left of each lane designates the RNA species of the predicted length for normally initiated and polyadenylated message from that plasmid. Splicing was also taken into account when transcript size was predicted. The arrowheads show the positions of the 3.3- and 4.3-kb extended transcript species produced from pLS102m. The bracket shows the region scanned within the last lane to quantitate extended transcripts produced from pLS102m. For each plasmid, a region of identical surface area, located at the same relative position with respect to the expected-size transcript, was used for quantitation of extended transcripts. In this set of RNAs, the fold increases in the amount of expected-size CAT mRNA found in cells transfected with the various plasmids compared with that found in pLS16-transfected cells were 1.9 for pLS52, 2.5 for pLS55, 8.3 for pLS56, 3.8 for pLS102, and 2.2 for pLS102m. Table 1 provides the average increases obtained over three or more experiments, each performed in duplicate.

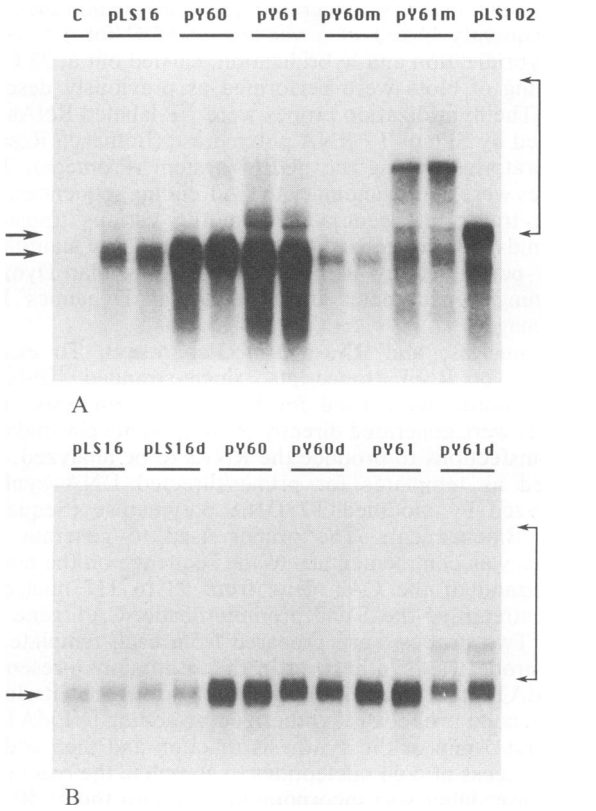

FIG. 3. (A) Representative Northern analysis showing the effects of 3′ splice site mutations on RNA accumulation. (B) Representative Northern analysis showing the effects of 5′ splice site mutations on RNA accumulation. In each panel, the arrows at the left show the positions of the expected-size CAT RNA species produced from each plasmid. In panel A, the upper arrow shows the location of the expected-size RNA from pLS102, while the lower arrow is applicable for all other plasmids. The bracket at the right of each panel shows the area in each lane (except for the pLS102 lane in panel A) scanned to quantitate extended transcripts.

from the SV40 late promoter. Cells transfected with pLS56 also consistently exhibited at least a twofold greater amount of CAT mRNA than did cells transfected with either pLS102 or pLS2, which includes sequences to SV 770 (Table 1). This finding suggested the presence of an element in pLS102 and pLS2, between SV 560 and 770, which can exert a negative influence on mRNA accumulation.

Experiments carried out with plasmids pLS5x56, pY60, and pY61 confirmed that the region around the SV40 late leader 3′ splice site was responsible for the increased mRNA accumulation. In the case of pY60 and pY61, only the 75 or 100 bp immediately surrounding the 3′ splice site were joined directly to the basal promoter. Conversely, plasmid pLS5x56 contains all of the sequences found in pLS102 with the exception of a 40-bp deletion between SV 521 and 560 (Fig. 1). Cells transfected with pLS5x56 exhibited CAT mRNA levels similar to those in pLS16-transfected cells (Table 1), much like pLS52 and pLS55, whereas both pY60 and pY61 generated dramatically higher amounts of CAT mRNA (Table 1 and Fig. 3). We believe that the 6- to 7-fold increases in the amount of CAT RNA seen with pY60 and pY61 are not substantially different from the 10-fold increase obtained with pLS56; thus, these three plasmids are similar in their abilities to produce CAT mRNA. These results therefore showed that of all the downstream sequences added to the basal late promoter in pLS102, the area immediately surrounding the 3′ splice site at SV 558 is necessary and sufficient to produce the observed increase in CAT mRNA. No sequences between the major transcription initiation site at SV 325 and SV 555 exhibited a significant influence on the accumulation of mRNA under our experimental conditions. Sequences downstream of SV 560 also did not contribute to the increase in CAT mRNA, although there appears to be an element that negatively influences mRNA accumulation located between SV 560 and 770.

Increased mRNA accumulation due to presence of functional splice sites. Our results pointed to the 3′ splice site as the sequence responsible for increased accumulation of CAT mRNA. To confirm this, we examined the effect of mutating that site. Plasmid pLS102m differs from pLS102 by a single nucleotide, at SV 556. This mutation changes the conserved AG dinucleotide at the 3′ splice site to GG and prevents splicing from occurring (see below). The amount of CAT mRNA produced in COS cells transfected with either pLS102 or pLS102m was quantitated by Northern analysis. Introduction of the point mutation led to a decrease in CAT mRNA of the expected size to nearly the same level as that produced by the basal promoter (Table 1 and Fig. 2). Introduction of this same point mutation, at SV 556, into plasmids pY60 and pY61, to create pY60m and pY61m, respectively, again abolished the increase in mRNA accumulation produced by addition of the 3′ splice site (Fig. 3A).

These experiments demonstrated that the mechanism(s) responsible for the increase in mRNA accumulation requires the 3′ splice site sequence. This sequence could function in two possible roles. The most likely possibility, and the one

predicted by previous studies, is that addition of the 3′ splice site created a spliceable transcript, which thus led to increased amounts of mRNA. The SV40 late leader region between SV 325 and 1046 contains two 5′ splice sites, at SV 374 and 527, and the 3′ splice site at SV 558 (Fig. 1). There is also an additional 5′ splice site within the late promoter, at SV 295, which is used in some late gene transcripts that are initiated upstream of the major transcription start site at SV 325 (31). Thus, addition of appropriate leader sequences could lead to splicing between one or more of the three 5′ splice sites and the 3′ splice site.

A second possible role for the 3′ splice site could be as a *cis*-acting sequence element which enhances mRNA accumulation without requiring that splicing take place. Such a sequence element might affect expression by a wide range of mechanisms, from promoterlike effects on transcription to posttranscriptional effects on processes such as RNA turnover.

To distinguish between these possibilities, we examined the effect of 5′ splice site mutations on CAT mRNA levels. Plasmids pLS16d, pY60d, and pY61d were generated by creating a four-nucleotide deletion, from SV 295 to 298. This deletion removes the only 5′ splice site, at SV 295, in pY60d RNA. One additional 5′ splice site, at SV 527, remains in pY61d RNA. However, splicing between this 5′ splice site and the 3′ splice site at SV 558 has been reported to occur infrequently when the majority of leader sequences upstream of SV 527 are deleted (13). The deletion at SV 295 abolished splicing to levels at the limits of detection in pY60d and pY61d RNA (see below). pLS16d RNA was not expected to splice regardless of the SV 295 deletion, since it does not contain a 3′ splice site. This plasmid was used as a control to assess the effect of this mutation on promoter activity.

Deletion of SV 295 to 298 did not affect CAT mRNA production from pLS16d (Table 1 and Fig. 3B). Thus, this mutation did not alter late promoter activity. The same mutation in pY60d and pY61d led to an approximately twofold decrease in mRNA accumulation compared with pY60 and pY61, respectively (Table 1 and Fig. 3B). Although deletion of the 5′ splice site did not elicit as dramatic a decrease as did mutation of the 3′ splice site, the results still suggest that the ability to splice augments mRNA accumulation. The failure to observe a greater effect when the 5′ splice site was removed in pY60d and pY61d could have been due, in part, to a small amount of cryptic splicing that may have been occurring (see below).

Deletion of the SV 295 5′ splice site from pLS102, to create pLS102d, had no effect on mRNA accumulation (data not shown). In this case, the 5′ splice site deletion was not expected to inhibit splicing since pLS102d transcripts still contain the 5′ site at SV 374. Thus, mutation of the SV 295 5′ splice site affected mRNA accumulation only when the mutation inhibited splicing. This finding is consistent with the hypothesis that the increased accumulation of CAT mRNA upon addition of the 3′ splice site was due, at least in part, to the creation of a spliceable transcript.

S1 analysis of splicing in pY60 and pY61. RNA harvested from cells transfected with plasmids pY60 and pY61 and their derivatives were analyzed to determine whether addition of the leader sequences containing the 3′ splice site triggered splicing from the SV 295 5′ splice site to the 3′ site at SV 558. This analysis was complicated by the heterogeneity of 5′ ends in transcripts generated from the late promoter (31). Indeed, S1 analysis of pLS16 CAT mRNA, which cannot splice, showed major transcription initiation sites at approximately SV 170, 260, 313, and 320, plus several minor start sites (Fig. 4A and D). These 5′ termini have been observed in previous studies of SV40 late mRNA (13, 14, 26). The initiation site calculated to be at SV 320 could in reality be the known major site at SV 325. S1 analysis of pY60 and pY61 RNA produced a number of protected fragments almost exactly 75 and 100 nucleotides larger, respectively, than those observed with pLS16 RNA (Fig. 4D). This finding suggested that transcription of pY60 and pY61 initiated at the same sites used in pLS16.

To positively identify the fragments protected by spliced transcripts amid the numerous unspliced transcript fragments, a pair of single-stranded, ^{32}P-labeled DNA probes was generated for each of the plasmids used in this study. One probe contained ^{32}P label along its entire length, while an identical probe contained label only within the CAT coding sequences at its 5′ end. The probes were identical in sequence to their parent plasmid from the SV40 origin of replication to the position within the CAT gene 117 nucleotides downstream of the SV40-CAT junction. Both the 5′-labeled and the full-length labeled probes produced S1 nuclease-protected fragments representing unspliced RNA and the downstream exons of spliced species. Only the full-length labeled probe could detect fragments corresponding to upstream exons of spliced CAT RNA. Fragments visible only with the full-length labeled probe would be indicative of upstream exons and thus establish that the RNA was spliced. The predicted sizes of protected fragments representing possible upstream and downstream exons are depicted in Fig. 4B and C.

Analysis of pY60 and pY61 RNA indeed showed that one major band, a fragment of approximately 125 nucleotides, was visible when the full-length labeled probe was used but not when the 5′-labeled probe was used (Fig. 4D). This fragment corresponds in size to the predicted upstream exon beginning at SV 170 and ending at SV 295. Another prominent band seen with both pY60 and pY61 RNA corresponded to a fragment of about 160 nucleotides. This band most likely represents the predicted 161-nucleotide downstream exon. These results demonstrated that some portion of pY60 and pY61 RNAs was spliced from SV 295 to 558. Fragments of 35 nucleotides were beyond the resolving power of the gel system used and thus not detected. Therefore, we could not determine whether RNAs with 5′ ends at SV 260 were spliced. As expected (13), we were unable to detect use of the 5′ splice site at SV 527 in pY61 RNA.

S1 nuclease analysis of pY60m and pY61m RNA did not produce a 125-nucleotide protected fragment (Fig. 4D). This finding indicated that splicing was abolished when the 3′ splice site was mutated in these plasmids. Similar experiments with RNA generated from pY60d and pY61d showed that mutation of the 5′ splice site also inhibited splicing, but those results suggested the possibility that a small amount of cryptic splicing could still be occurring (data not shown). These experiments thus showed that addition of the 3′ splice site in plasmids pY60 and pY61 did lead to splicing of the CAT mRNA produced from these plasmids and that mutation of either the 5′ splice site at SV 295 or the 3′ splice site inhibited splicing.

3′ splice site mutations lead to accumulation of extended transcripts. Our results showed that addition of a functional 3′ splice site to the SV40 late promoter significantly boosted mRNA accumulation. They also suggested that this increase was due, at least in part, to creation of a functional splice within the CAT transcript, since mutation of either the 5′ or 3′ splice site attenuated the increase. Mutation of the 3′

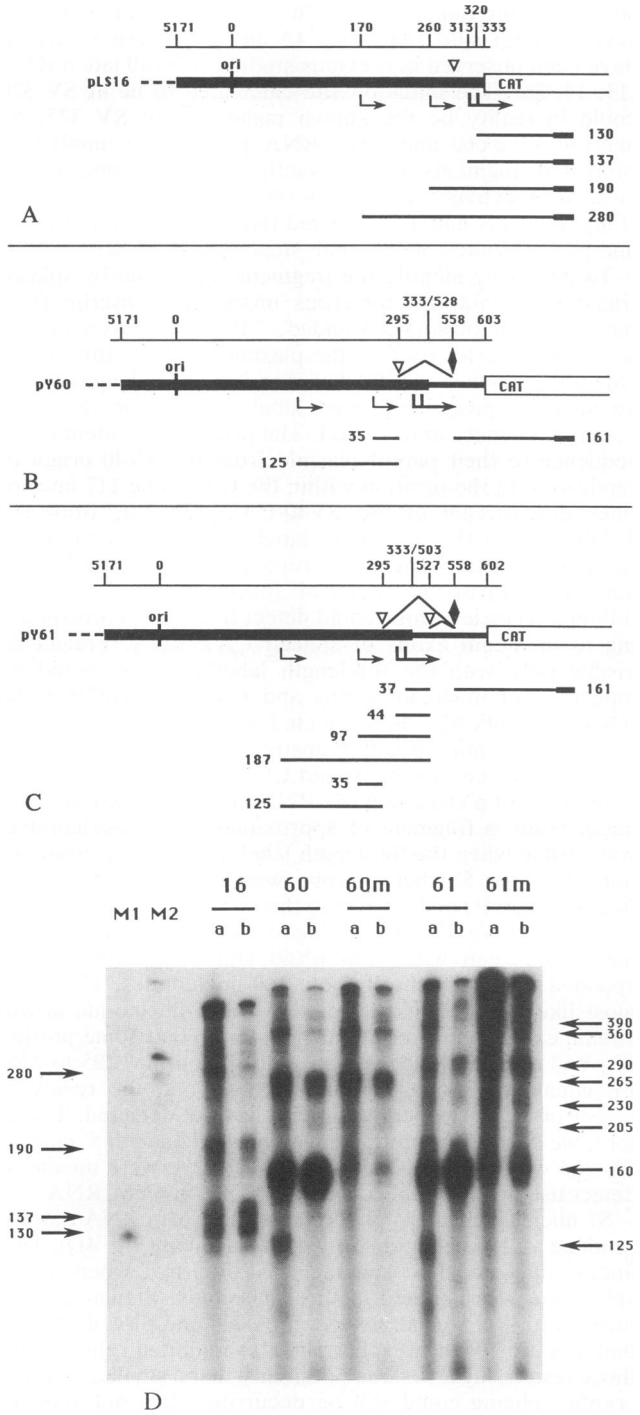

FIG. 4. S1 analysis of splicing. (A) The four major protected fragments observed with pLS16 RNA (in panel D) are shown, with their sizes in nucleotides given at the right. The thicker portion of these lines represents the location of the primer used to generate the S1 probes. The estimated 5′ ends are denoted by the arrows directly underneath the map of pLS16. The two initiation sites near the known major transcription initiation site are depicted by a thick arrow. (B and C) Schemes showing the expected sizes of protected fragments from spliced pY60 and pY61 RNA. Their sizes in nucleotides are shown to the left or right of the fragments. Possible splices for pY60 and pY61 are shown by thin lines immediately above the plasmid maps. (D) S1 analysis of pLS16, pY60, pY60m, pY61, and pY61m RNAs. For each sample, the full-length labeled probe was used in lane a, while the 5′-labeled probe was used in lane b. M1 and M2 are size marker lanes. The arrows at the left show the sizes (in nucleotides) and positions of the four major bands produced by protection with pLS16 RNA. The arrows at the right show the sizes (in nucleotides) and positions of major RNA species protected by pY60, pY60m, pY61, and pY61m RNA.

splice site, however, elicited a much more dramatic effect than did mutation of the 5′ splice site. These observations thus posed the questions of how splicing, or splice sites, positively influences mRNA accumulation and why the 3′ splice site plays a more important role in this increase than does the 5′ splice site.

Recent studies using chimeric constructions or histone genes have suggested that splicing, or the presence of a 3′ splice site, enhances the efficiency of mRNA polyadenylation (20, 24, 25). Our Northern blot analyses supported this proposal. Inefficient 3′ end processing could lead to accumulation of large RNA transcripts that have not been cleaved at the usual polyadenylation site. Figure 2 shows that pLS102 produced abundant amounts of a single CAT RNA species of approximately the correct size for a normally initiated, properly spliced and polyadenylated message (1.7 kb). Insertion of a point mutation at the 3′ splice site, in pLS102m, not only led to a threefold decrease in CAT mRNA of the expected size for an unspliced, properly polyadenylated transcript (1.9 kb) but also led to the appearance of a number of novel, larger than expected RNA transcripts (which we call extended transcripts). The most abundant CAT RNA species in pLS102m-transfected COS cells was 3.3 kb in size. Another apparent RNA species migrated at a rate equivalent to that of a molecule of about 4.3 kb. However, this band was located just below the 28S rRNA band and may be the result of the compression of several CAT RNA species into one band. Nonetheless, the RNA species represented by this band were 4.3 kb or larger. There was also an increase in the amount of diffuse hybridization occurring in the pLS102m RNA lane above the expected 1.9-kb band. We interpret this result to be indicative of a heterogeneous population of CAT-containing transcripts generated from pLS102m. This heterogeneous population of RNAs ranged in size to greater than 10 kb. Calculation of the ratio of the sum of extended CAT transcripts to that of the expected length (Table 1) showed that pLS102m exhibited a seven- to eightfold-higher ratio than did pLS102. The total amounts of CAT-containing RNA produced by the two plasmids were similar (data not shown). Thus, the introduction of a point mutation at the 3′ splice site led to a decrease in the amount of CAT mRNA of the expected size and a concomitant increase in extended transcripts, without significantly altering the total amount of CAT RNA produced.

It is unlikely that these extended transcripts were the result of aberrant initiation upstream of the usual transcription initiation site. The estimated sizes of the two most abundant extended transcripts produced in pLS102m-transfected cells would place their initiation sites very far upstream. The 3.3-kb RNA would have to initiate within or near the 3′ end of the β-lactamase coding sequence, while the RNA or RNAs migrating at the position of a 4.3-kb transcript would have to initiate close to the SV40 late polyadenylation site or upstream of it within the CAT coding region. It is more likely that these extended RNAs resulted from the addition of sequences at their 3′ ends because of inefficient cleavage and polyadenylation.

Two other plasmids, pLS55 (Fig. 2) and pY61m (Fig. 3A), also exhibited significantly higher ratios of extended versus expected-size CAT RNA than did pLS102 (Table 1). pY61m contains the same point mutation as pLS102m, and pLS55 contains all of the late leader sequences up to but not including the 3′ splice site. Thus, all three plasmids that generated high proportions of extended transcripts relative to the expected-size CAT transcripts contained a mutated 3′ splice site or were lacking that site.

Three plasmids, pY60m, pLS52, and pLS5x56, also either contain a 3′ splice site mutation or are missing that site. However, these plasmids did not generate a high proportion of extended transcripts as did pLS102m, pY61m, and pLS55. If, as predicted, these extended transcripts are the result of inefficient polyadenylation, it would be expected, in most cases, that such large, unpolyadenylated RNAs would rapidly degrade in the nucleus. This is probably what is occurring with pY60m, pLS52, and pLS5x56. Fortunately, in the case of plasmids pLS102m, pY61m, and pLS55, these extended transcripts were able to accumulate. We believe that these transcripts were stabilized by a putative RNA stability element within the late leader, near or at the 5′ splice site at SV 527 (12), and subsequently, in some instances, by cryptic polyadenylation downstream of the normal polyadenylation site (unpublished results).

It is important to note that mutation of the 5′ splice site, in pY61d, did not lead to increased accumulation of extended transcripts as did mutation of the 3′ splice site in pY61m (Fig. 3). This finding would suggest that the mechanism(s) which leads to the appearance of these extended transcripts involves exclusively the 3′ splice site.

Extended transcripts contain the SV40 late polyadenylation site. We directly tested the hypothesis that mutation of the 3′ splice site led to a decrease in the efficiency of processing at the SV40 late polyadenylation site by examining use of that site in pLS102m RNA. RNase protection assays were performed to determine whether sequences just downstream of the cleavage/polyadenylation site were retained in extended transcripts. A ^{32}P-labeled riboprobe complementary to the region of pLS102 and pLS102m from 141 nucleotides upstream of the late polyadenylation site to 181 nucleotides downstream of it was used in these assays (Fig. 5A). The probe also contained an additional 265 nucleotides of noncomplementary sequence. Hybridization of this probe to RNAs polyadenylated at the usual site should produce a protected fragment of 141 nucleotides, while RNAs that contain this region but are not polyadenylated should produce a protected fragment of 322 nucleotides.

RNA harvested from pLS102-transfected COS cells produced a single protected fragment of the correct size for properly processed RNA (Fig. 5B). Two protected fragments, approximately 141 and 322 nucleotides in length, were produced from pLS102m RNA. The intensities of the 141-nucleotide protected fragments in lanes 3 and 4 did not reflect the relative proportions observed in Northern assays, probably because the labeled probe was not present in sufficient excess. Regardless, our results clearly demonstrated that a portion of the transcripts produced in pLS102m-transfected cells still retained but did not utilize the SV40 late polyadenylation site. This finding strongly suggests that one of the consequences of mutating the 3′ splice site is a decrease in the efficiency of polyadenylation, thus leading to an increased proportion of transcripts with extended 3′ ends. The addition of a 3′ splice site could thus serve to increase mRNA accumulation by ensuring that a greater proportion of the RNA transcripts produced from a gene are polyadenylated and thereby become mature, stable mRNA.

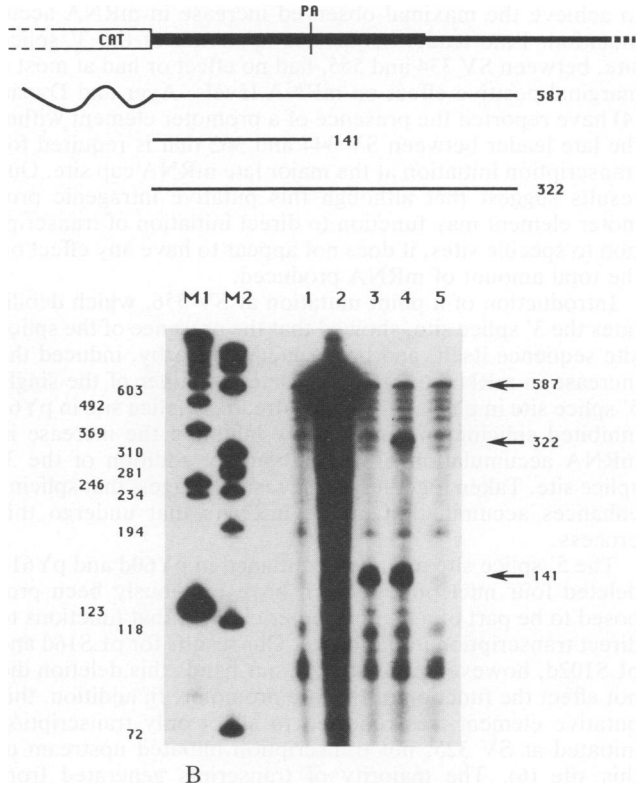

FIG. 5. Analysis of the 3′ ends of pLS102 and pLS102m RNAs. (A) Locations of sequences, relative to the SV40 late polyadenylation site, contained in the riboprobe used in the RNase protection assay. The probe was 587 nucleotides in length. The bottom two lines show the sizes (in nucleotides) and locations of probe fragments protected by normally polyadenylated and unpolyadenylated RNAs. (B) RNase protection assay. Lanes 1 to 5 contain the labeled probe hybridized to various RNAs and digested with RNase A and RNase T_1. Lanes: M1 and M2, size markers, with sizes (in nucleotides) indicated at the left; 1, yeast tRNA; 2, unhybridized, undigested probe; 3, pLS102 RNA; 4, pLS102m RNA; 5, RNA from mock-transfected COS cells. The three arrows at right show the bands representing fragments of 587, 322, and 141 nucleotides.

DISCUSSION

Our studies were initiated to examine the contribution of the SV40 late leader sequences to expression directed by the SV40 late promoter. The results show that the sequence at the 3′ splice site is important for optimized production of properly processed mRNAs. In part this is due to the creation of a functional intervening sequence, since mutations either at the added 3′ splice site or at the 5′ splice site already present in the promoter region attenuated this increase. In addition, the sequence at the 3′ splice site serves to enhance the efficiency of transcript polyadenylation by a mechanism independent of the 5′ splice site. Our results also suggest that there are additional elements, between SV 560 and 770, which, in our plasmid constructions, negatively influence mRNA accumulation. We have not as yet looked into the nature of these negative elements.

Addition of the sequences at or immediately surrounding the 3′ splice site at SV 558 to the late promoter was sufficient

to achieve the maximal observed increase in mRNA accumulation. Late leader sequences upstream of the 3′ splice site, between SV 334 and 555, had no effect or had at most a marginal positive effect on mRNA levels. Ayer and Dynan (4) have reported the presence of a promoter element within the late leader between SV 344 and 362 that is required for transcription initiation at the major late mRNA cap site. Our results suggest that although this putative intragenic promoter element may function to direct initiation of transcription to specific sites, it does not appear to have any effect on the total amount of mRNA produced.

Introduction of a point mutation at SV 556, which debilitates the 3′ splice site, showed that the presence of the splice site sequence itself, and not sequences nearby, induced the increase in mRNA expression. Deletion either of the single 5′ splice site in pY60 or of the upstream 5′ splice site in pY61 inhibited splicing and moderately inhibited the increase in mRNA accumulation brought about by addition of the 3′ splice site. Taken together, these results suggest that splicing enhances accumulation of the mRNAs that undergo this process.

The 5′ splice site mutation contained in pY60d and pY61d deleted four nucleotides which have previously been proposed to be part of a late promoter element that functions to direct transcription initiation (6). Our results for pLS16d and pLS102d, however, show that in our hands this deletion did not affect the function of the late promoter. In addition, this putative element was reported to affect only transcription initiated at SV 325, not transcription initiated upstream of this site (6). The majority of transcripts generated from pY60, pY61, and their derivatives started upstream of SV 325. Therefore, we believe that the 5′ splice site deletion decreased mRNA accumulation via its effect on splicing and not by an effect on transcription initiation.

Although inhibition of splicing by disabling either the 5′ or 3′ splice site led to decreased accumulation of CAT mRNA, the effects of mutations at these sites were not equivalent. In addition, the accumulation of extended transcripts increased significantly only when the 3′ splice site was disabled. These observations suggested to us that increased accumulation of mRNA, attained upon introduction of a spliceable intervening sequence, is likely due to multiple mechanisms. One mechanism, exerting only a modest effect, is dependent either upon the 5′ splice site or upon the presence of a complete and functional intervening sequence. This mechanism could lead to increased mRNA accumulation via some influence of the splicing process itself or by virtue of an intrinsic property of spliced transcripts.

A second, and in this case more significant, mechanism involves the sequence at the 3′ splice site, independent of the 5′ splice site. In many cases, 3′ splice site mutations led to the fortuitous accumulation of extended transcripts, thus providing an opportunity to examine this mechanism. A number of groups have previously proposed that splicing and/or the 3′ splice site can enhance polyadenylation (20, 24, 32). Pandey et al. (25) showed that introduction of an intervening sequence into a normally intronless and unpolyadenylated histone mRNA transcript caused a shift in 3′ end processing from the normal stem-loop structure common to histone mRNAs to a cryptic polyadenylation site nearby. Our studies showed that extended transcripts contained intact cleavage and polyadenylation sites and therefore suggested that mutation of the 3′ splice site leads to decreased efficiency of utilization of this site. These results thus support the conclusions cited above and supply another mechanism whereby addition of the 3′ splice site sequence increases mRNA accumulation. Polyadenylation helps to stabilize mRNA transcripts (7). Thus, an increase in the efficiency of polyadenylation would mean that a higher proportion of mRNA transcripts are stable, thereby leading to a greater accumulation of mRNA. This mechanism appears to be dependent solely upon the presence of a 3′ splice site, since 5′ splice site deletions did not produce the same results. This finding points to the possibility that spliceosome components that interact with the 3′ splice site, or formation of intermediate spliceosome complexes dependent only on the 3′ splice site, are involved in enhancing polyadenylation. Our results are consistent with the exon definition model proposed by Berget and coworkers (24, 27, 30), which postulates that the paired sequence elements that define the ends of an exon, 3′ and 5′ splice sites, or a 3′ splice site and a polyadenylation site communicate with each other and enhance each other's functionality.

The results of this study may be indicative of a widespread interdependence between many aspects of mRNA metabolism. Splicing, polyadenylation, transport, and processes involved in RNA stability may all work in an interrelated, balanced manner to posttranscriptionally regulate mRNA accumulation.

ACKNOWLEDGMENTS

A number of plasmids used in this study were derived from constructions made by J. Keller, S. Carswell, and M. Sadofsky; we thank them for their contributions to this project. We also thank S. Carswell, N. Schek, and S. Zeichner for helpful discussions and M. Chiou for assistance with the illustrations.

This work was supported by Public Health Service grants CA28379 awarded by the National Cancer Institute and GM 45773 awarded by the National Institute of General Medical Sciences to J.C.A. H.C.C. is supported by Public Health Service award GM12919 from the National Institute of General Medical Sciences.

REFERENCES

1. **Alwine, J. C.** 1982. Evidence for simian virus 40 late transcriptional control: mixed infections of wild-type simian virus 40 and a late leader deletion mutant exhibit *trans* effects on late viral RNA synthesis. J. Virol. **42**:798–803.
2. **Alwine, J. C., and J. Picardi.** 1986. Activity of simian virus 40 late promoter elements in the absence of large T antigen: evidence for repression of late gene expression. J. Virol. **60**: 400–404.
3. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1987. Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
4. **Ayer, D. E., and W. S. Dynan.** 1988. Simian virus 40 major late promoter: a novel tripartite structure that includes intragenic sequences. Mol. Cell. Biol. **8**:2021–2033.
5. **Brady, J., and G. Khoury.** 1985. *trans*-activation of the simian virus 40 late transcription unit by T antigen. Mol. Cell. Biol. **5**:1391–1399.
6. **Brady, J., M. Radonovich, M. Vodkin, V. Natarajan, M. Thoren, G. Das, J. Janik, and N. P. Salzman.** 1982. Site-specific base substitution and deletion mutations that enhance or suppress transcription of the SV40 major late RNA. Cell **31**:625–633.
7. **Brawerman, G.** 1987. Determinants of messenger RNA stability. Cell **48**:5–6.
8. **Brinster, R. L., J. M. Allwn, R. R. Behringer, R. E. Gelinas, and R. D. Palmiter.** 1988. Introns increase transcriptional efficiency in transgenic mice. Proc. Natl. Acad. Sci. USA **85**:836–840.
9. **Buchman, A. R., and P. Berg.** 1988. Comparison of intron-dependent and intron-independent gene expression. Mol. Cell. Biol. **8**:4395–4405.
10. **Callis, J., M. Fromm, and V. Walbot.** 1987. Introns increase gene expression in cultured maize cells. Genes Dev. **1**:1183–1200.

11. **Carswell, S., J. Resnick, and J. C. Alwine.** 1986. Construction and characterization of CV-1P cell lines which constitutively express the simian virus 40 agnoprotein: alteration of plaquing phenotype of viral agnogene mutants. J. Virol. **60**:415–422.
12. **Chiou, H., and J. Alwine.** Unpublished data.
13. **Ghosh, P. K., M. Piatak, J. E. Mertz, S. M. Weissman, and P. Lebowitz.** 1982. Altered utilization of splice sites and 5′ termini in late RNAs produced by leader region mutants of simian virus 40. J. Virol. **44**:610–624.
14. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. M. Weissman.** 1978. Heterogeneity and 5′-terminal structures of the late RNAs of simian virus 40. J. Mol. Biol. **126**:813–846.
15. **Gruss, P., and G. Khoury.** 1980. Rescue of a splicing defective mutant by insertion of an heterologous intron. Nature (London) **286**:634–637.
16. **Gruss, P., C.-J. Lai, R. Dhar, and G. Khoury.** 1979. Splicing as a requirement for biogenesis of functional 16S mRNA of simian virus 40. Proc. Natl. Acad. Sci. USA **76**:4317–4321.
17. **Hamer, D. H., and P. Leder.** 1979. SV40 recombinants carrying a functional RNA splice junction and polyadenylation site from the chromosomal mouse β(maj) globin gene. Cell **17**:737–747.
18. **Hartzell, S. W., B. J. Byrne, and K. N. Subramanian.** 1984. Mapping of the late promoter of simian virus 40. Proc. Natl. Acad. Sci. USA **81**:23–27.
19. **Hay, N., H. Skolnik-David, and Y. Aloni.** 1982. Attenuation in the control of SV40 gene expression. Cell **29**:183–193.
20. **Huang, M. T. F., and C. M. Gorman.** 1990. Intervening sequences increase efficiency of RNA 3′ processing and accumulation of cytoplasmic RNA. Nucleic Acids Res. **18**:937–947.
21. **Keller, J. M., and J. C. Alwine.** 1984. Activation of the SV40 late promoter: direct effects of T antigen in the absence of viral DNA replication. Cell **36**:381–389.
22. **Keller, J. M., and J. C. Alwine.** 1985. Analysis of an activatable promoter: sequences in the simian virus 40 late promoter required for T-antigen-mediated transactivation. Mol. Cell. Biol. **5**:1859–1869.
23. **Lai, C.-J., and G. Khoury.** 1979. Deletion mutants of simian virus 40 defective in biosynthesis of late viral mRNA. Proc. Natl. Acad. Sci. USA **76**:71–75.
24. **Niwa, M., S. D. Rose, and S. M. Berget.** 1990. In vitro polyadenylation is stimulated by the presence of an upstream intron. Genes Dev. **4**:1552–1559.
25. **Pandey, N. B., N. Chodchoy, T.-J. Liu, and W. F. Marzluff.** 1990. Introns in histone genes alter the distribution of 3′ ends. Nucleic Acids Res. **18**:3161–3170.
26. **Piatak, M., K. N. Subramanian, P. Roy, and S. M. Weissman.** 1981. Late messenger RNA production by viable simian virus 40 mutants with deletions in the leader region. J. Mol. Biol. **153**:589–618.
27. **Robberson, B. L., G. J. Cote, and S. M. Berget.** 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol. Cell. Biol. **10**:84–94.
28. **Ryu, W.-S., and J. E. Mertz.** 1989. Simian virus 40 late transcripts lacking excisable intervening sequences are defective in both stability in the nucleus and transport to the cytoplasm. J. Virol. **63**:4386–4394.
29. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
30. **Talerico, M., and S. M. Berget.** 1990. Effect of 5′ splice site mutations on splicing of the preceding intron. Mol. Cell. Biol. **10**:6299–6305.
31. **Tooze, J. (ed.).** 1980. DNA tumor viruses. The molecular biology of tumor viruses, part 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
32. **Villarreal, L. P., and R. T. White.** 1983. A splice junction deletion deficient in the transport of RNA does not polyadenylate nuclear RNA. Mol. Cell. Biol. **3**:1381–1388.