MOLECULAR AND CELLULAR BIOLOGY, Oct. 1985, p. 2713–2719
0270-7306/85/102713-07$02.00/0

Vol. 5, No. 10

Identification of a Sequence Element on the 3′ Side of AAUAAA Which Is Necessary for Simian Virus 40 Late mRNA 3′-End Processing

MOSHE SADOFSKY,[1] SHEILA CONNELLY,[2] JAMES L. MANLEY,[2] AND JAMES C. ALWINE[1]*

Department of Microbiology and Graduate Group of Genetics, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104,[1] and Department of Biology, Columbia University, New York, New York 10027[2]

Received 24 April 1985/Accepted 22 July 1985

Our previous studies of the 3′-end processing of simian virus 40 late mRNAs indicated the existence of an essential element (or elements) downstream of the AAUAAA signal. We report here the use of transient expression analysis to study a functional element which we located within the sequence AGGUUUUUU, beginning 59 nucleotides downstream of the recognized signal AAUAAA. Deletion of this element resulted in (i) at least a 75% drop in 3′-end processing at the normal site and (ii) appearance of readthrough transcripts with alternate 3′ ends. Some flexibility in the downstream position of this element relative to the AAUAAA was noted by deletion analysis. Using computer sequence comparison, we located homologous regions within downstream sequences of other genes, suggesting a generalized sequence element. In addition, specific complementarity is noted between the downstream element and U4 RNA. The possibility that this complementarity could participate in 3′-end site selection is discussed.

In higher eucaryotes, the 3′ end of most mRNAs appears to be formed by specific cleavage of a larger precursor molecule at the site which is ultimately polyadenylated. The hexanucleotide AAUAAA (35), or close variants (43), has been shown to be an essential signal for 3′-end formation in both viral (12, 32) and cellular (20) systems. However, this sequence alone cannot be sufficient, since it is found within coding regions of mRNAs at positions where no processing occurs. Further, if the hexanucleotide were the only determinant of 3′-end processing, then it would be difficult to explain situations in which specific mRNAs utilize variable 3′-end cleavage sites, sometimes in a tissue- or developmentally specific manner (6, 11, 27). That the AAUAAA hexanucleotide is not sufficient has been suggested in several recent studies (14, 29, 30, 38, 39, 45) which indicate that sequences on the 3′ side of AAUAAA are important in specifying 3′-end cleavage sites.

The formation of the 3′ end of simian virus 40 (SV40) late mRNA has been shown to proceed at wild-type levels in constructs which include the 108 nucleotides immediately downstream of the AAUAAA (9; see Fig. 2). We have previously studied viable viral deletion mutants which deleted sequences within the first 58 nucleotides downstream of the AAUAAA (38). We were limited to this region, since further deletions eliminate virus viability. All of the deletions in this region resulted in (i) a small but notable lowering (ca. 10%) of the efficiency of utilization of the normal site and (ii) the appearance of readthrough transcripts with 3′ ends at other downstream processing sites. Thus the deletion studies indicated that a functional element downstream of both the AAUAAA and the actual 3′-end processing site had been disturbed but not destroyed by the deletions. We concluded from the positions of the deletions that the putative element would most likely be located between 58 and 108 nucleotides beyond the AAUAAA.

In the present studies, we used transient expression analysis to localize a functional downstream element. This approach circumvents the need for virus viability. The data presented, in conjunction with our previous deletion analysis (38), indicate that a significant signal element is contained in the 9 nucleotides (AGGUUUUUU) located 59 to 67 nucleotides downstream of the late AAUAAA. The data also indicate that the spacing between this downstream signal element and the AAUAAA can be varied. We speculate on the generality of this element as a signal for 3′-end selection and note the possibility of interaction of this element with U4 RNA.

MATERIALS AND METHODS

Plasmids. The parent plasmid for the deletion studies, pB0, was constructed from pL4-cat as diagrammed in Fig. 1. Plasmid pL4-cat is a derivative of pL2-cat (21) and contains the SV40 late promoter region between simian virus (SV) nucleotides 5171 and 298 (through the origin; see reference 41 for nucleotide numbering) coupled to the chloramphenicol acetyltransferase (CAT) gene by a *Sal*I linker. The CAT gene is thus under the control of the SV40 late promoter. Plasmid pB0 was deleted of the downstream splicing and polyadenylation sequences. This was accomplished by complete digestion of pL4-cat with *Eco*RI and partial digestion with *Ban*I. We filled in the ends by using the large fragment of *Escherichia coli* DNA polymerase I and then ligated *Bam*HI linkers to the blunt ends. After cleavage of the linkers, the final plasmid was ligated and selected. The terminal nucleotides of the linker completed the filled-in *Ban*I and *Eco*RI sites such that *Ban*I, *Bam*HI, and *Eco*RI sites were clustered (Fig. 1).

Plasmid pBL was derived from pB0 by insertion of the SV40 *Bam*HI to *Bcl*I fragment (SV nucleotides 2533 to 2770; Fig. 2) into the *Bam*HI site of pB0. These sites share homologous overhanging nucleotides. The orientation of the insert was determined by restriction analysis, and that which placed the late polyadenylation site on the CAT RNA was selected as pBL. A series of deletions from the SV40 *Bcl*I site were prepared by digestion with *Bal* 31. The substrate for digestion was a pBR322 plasmid containing a complete

* Corresponding author.

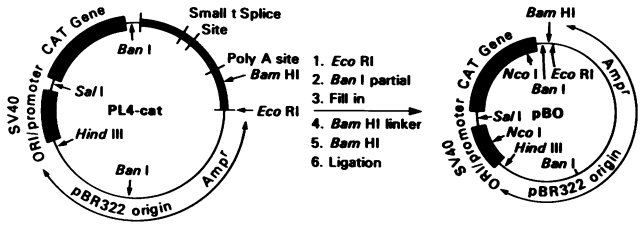

FIG. 1. Diagram of the construction of pB0. The specifics of construction are described in the text. The 3,200-nucleotide plasmid pB0 has the SV40 late promoter driving transcription of the CAT gene. It also contains a functional SV40 origin of replication (ORI). The plasmid contains no splicing or polyadenylation sequences, but instead a cluster of three restriction sites, *Ban*I, *Bam*HI, and *Eco*RI. Sequences from pBR322, an origin and the ampicillin resistance gene, are indicated.

SV40 genome inserted at the *Bam*HI site. This was cleaved with *Bcl*I and progressively digested. The deleted plasmid was closed by the insertion of a *Bgl*II linker at the site of digestion. The endpoints of the deletions were determined by sequencing (28; Fig. 2). Cleavage of the plasmids with *Bam*HI and *Bgl*II yielded deleted fragments which could be inserted into pB0 at the *Bam*HI site. Orientation was determined by restriction analysis, and the plasmids with the late polyadenylation site on the CAT RNA were selected. The relevant portions of the final plasmids (pB1L, pB28L, pB29L, pB33L, and pB35L) are shown in Fig. 2. The control plasmid, pSV0-cat, has no eucaryotic promoter and has been described elsewhere (17). All plasmids were grown and purified by standard procedures (26).

Viral studies. CV-1P cells were infected as previously described (38) with SV40 wild-type strain wt776 or the viable deletion mutant dl1265 (8). Infected samples were harvested 48 h postinfection.

Transfections. Plasmids were transfected into COS cells (16) by the calcium phosphate precipitation procedure (18) and harvested at 48 or 70 h posttransfection.

RNA preparation and analysis. Cytoplasmic or total cellular RNA from either infected or transfected cells was purified as previously described (38). For Northern blot analysis (2, 40), 15-μg samples of cytoplasmic RNA were fractionated on formaldehyde-agarose gels (23) and transferred to nitrocellulose (40). Specific bands were visualized by hybridization with a CAT-specific ^{32}P-labeled RNA probe prepared by transcription of CAT sequences inserted into pSP64 (Promega Biotec; 31). Fifty percent formamide hybridization conditions were used at 65°C for 12 h. For nuclease S1 analysis (4, 38), 15 to 20 μg of total cellular RNA was hybridized to probes (see text and figure legends for specific probes) labeled by replacement synthesis with T4 DNA polymerase (38). Hybridizations were done under 80% formamide conditions at 51°C. Nuclease S1-resistant fragments were fractionated on a 5% polyacrylamide-urea sequencing gel (28).

RESULTS

pBL plasmid constructs. CAT transient expression vectors (18) were constructed for the assessment of 3'-end processing of mRNA. The parent of the pB constructs, pB0, is diagrammed in Fig. 1 and 2A. It contains the SV40 replication origin (Ori) and the SV40 late promoter driving transcription of the CAT gene. When transfected into COS cells, which constitutively produce SV40 T antigen (16), the plasmid will replicate and actively promote transcription from the late promoter owing to *trans*-activation by T antigen (21). Downstream of the CAT gene there is no polyadenylation signal, but instead there is a unique *Bam*HI site. Sequences were inserted into this site and tested for their ability to signal 3'-end processing. The 237-nucleotide *Bam*HI-to-*Bcl*I fragment of SV40 DNA (SV nucleotides 2533 to 2770), containing the late mRNA AAUAAA and the immediate downstream 108 nucleotides, was inserted into the *Bam*HI site forming plasmid pBL (Fig. 2B). This segment efficiently specifies 3'-end processing adjacent to the AAUAAA, as shown previously (9) and below. Also tested were a set of *Bal* 31-generated deletions which progressively remove se-

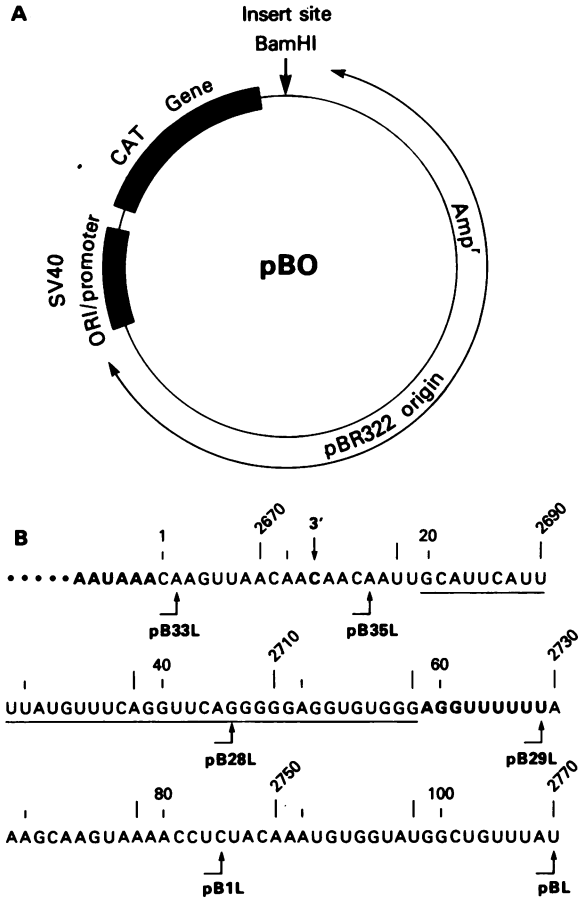

FIG. 2. (A) Map of pB0 indicating the *Bam*HI site where sequences were inserted to be tested for their ability to signal 3'-end processing. (B) The sequences tested for ability to signal 3'-end processing. Each insert began at the SV40 *Bam*HI site (SV nucleotide 2533) and included sequences through the AAUAAA to an endpoint indicated by the name of the specific plasmid construct. The sequences in uppercase letters are the SV40 sequences between the AAUAAA and the SV40 *Bcl*I site (SV nucleotide 2770). The AAUAAA, the site of 3'-end processing, and the downstream element described in the text are shown in bold type. The underlined sequence is the region deleted in viral mutant dl1263. The lowercase letters indicate the plasmid sequences immediately downstream from the inserts. Because of the method of construction, the two plasmid nucleotides in parentheses are absent in pBL but not in the other plasmids.

quences from the 3' end of the inserted SV40 fragment. These deletions are diagrammed in Fig. 2B.

A significant element between 59 and 67 nucleotides downstream of AAUAAA. Cytoplasmic RNA was prepared from COS cells transfected with each of the pB plasmids as well as the control plasmid pSV0-cat (17; see above). The RNA was analyzed by Northern analysis (2, 40) with an RNA hybridization probe complementary to the CAT coding region. The results are shown in Fig. 3A. Despite stringent washing, cross-hybridizing cellular RNAs were detected in the RNA from untransfected COS cells (lane COS), as well as in RNA from the same cells transfected with pSV0-cat (lane pSV0). However, specific bands were apparent in the RNA from cells transfected with the pB plasmids. An RNA with its 3' end generated by utilization of the inserted SV40 late site would be approximately 1,100 nucleotides long. Length variability is expected because of the heterogeneity of 5'-end

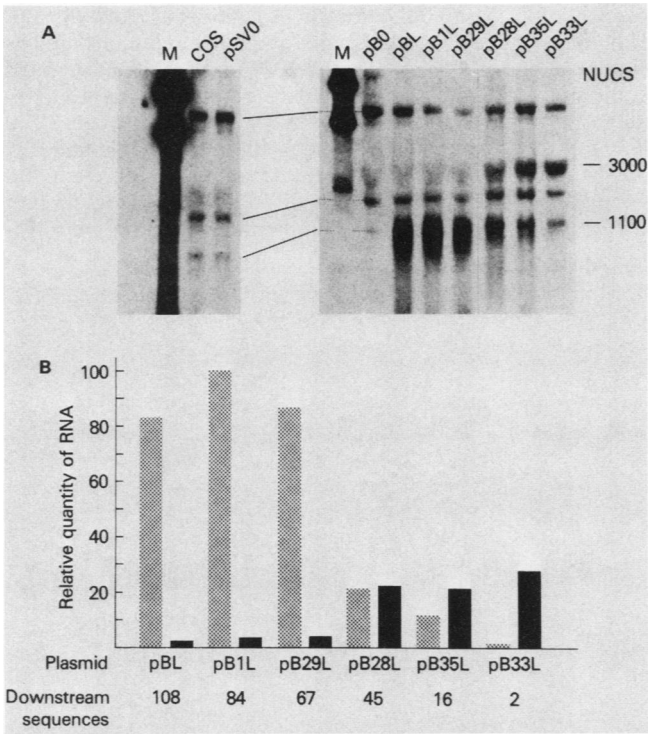

FIG. 3. (A) Northern blot analysis of CAT-specific RNA transcribed from the plasmid constructs after transfection into COS cells. The lines indicate identical bands of the two gels; these bands correspond to cross-hybridizing cellular RNAs discussed in the text. The CAT-specific RNAs are indicated by size in nucleotides (1,100 and 3,000 NUCS). Lanes, COS, RNA from untransfected COS cells; pSV0, RNA from COS cells transfected with the control plasmid pSV0-cat; M, size markers; pB0, pBL, pB1L, pB29L, pB28L, pB35L, and pB33L, RNAs from COS cells transfected with these plasmids, respectively. (B) Relative quantitation of the RNA data. RNAs in each lane were quantitated by laser densitometer scanning. The scans were normalized to the intensity of the largest cross-hybridizing cellular band (this band migrates at the position of the 28S ribosomal RNA). The intensity of the 1,100-nucleotide cellular band determined from lane pB0 was subtracted from the 1,100-nucleotide signal from the other lanes. Stippled and solid bars represent the relative quantities of the 1,100- and 3,000-nucleotide CAT-specific RNAs, respectively. Neither the efficiency of processing nor the stability of the 3,000-nucleotide RNA are known; thus, direct comparisons of the quantities of the two RNAs are not warranted. The length of SV40 sequence remaining downstream of the AAUAAA is presented for each construct.

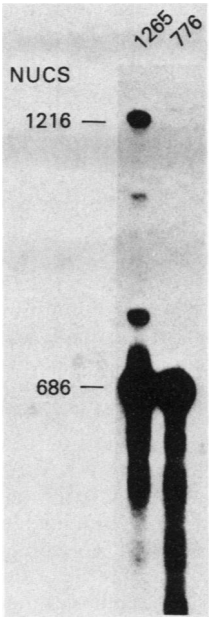

FIG. 4. Nuclease S1 analysis of the 3' ends of SV40 late mRNAs. Total cellular RNA from CV-1P cells was harvested 48 h after infection with SV40 wt776 or dl1265. The hybridization probe was the 1,216-nucleotide *Pst*I fragment (SV nucleotides 1988 through 3204) prepared from the DNA of each virus and labeled by T4 DNA polymerase replacement synthesis (38). RNA samples were hybridized to homologous viral probes and nuclease S1 treated, and protected segments were separated on 5% polyacrylamide-urea sequencing gels. NUCS, Nucleotides.

utilization by the SV40 late promoter (13, 21), as well as some variation in the length of the polyadenylate tail. Plasmids pBL, pB1L, and pB29L (containing 108, 84, and 66 nucleotides downstream of the AAUAAA, respectively) each generate equal quantities of a broad 1,100-nucleotide band indicative of 3'-end processing at the inserted site. This RNA is polyadenylated (not shown). A narrow band representing a cross-hybridizing cellular RNA also migrates close to 1,100 nucleotides (see control lanes COS, pSV0, and pB0). However, this band is easily differentiated from the broad, plasmid-specific 1,100-nucleotide band and can be corrected for in the analysis discussed below. The amount of the broad, plasmid-specific 1,100-nucleotide band drops dramatically in RNA generated from pB28L, which contains 45 downstream nucleotides. Further deletion to 16 and 2 downstream nucleotides (pB35L and pB33L, respectively) showed very low amounts of the 1,100-nucleotide plasmid-specific RNA.

The relative quantities of the 1,100-nucleotide RNA were determined by densitometer scanning. In this analysis the band intensity of the cross-hybridizing cellular RNA (determined from lane pB0) was subtracted as a correction. Figure 3B shows that there is at least a 75% reduction in the abundance of the 1,100-nucleotide RNA when sequences between 45 and 67 nucleotides downstream of the AAUAAA are deleted. Concomitant with the decrease in the 1,100-nucleotide RNA is the increase of a CAT-specific RNA migrating at approximately 3,000 nucleotides (Fig. 3A). This RNA represents a readthrough transcript which appears when the 3'-end processing at the inserted site is impeded or eliminated and reflects alternate 3'-end processing elsewhere in the plasmid. This readthrough phenomenon has been

previously characterized by use of viable SV40 deletion mutants (38; see Fig. 4). The sizes of the various deletions are not large enough to appreciably alter the migration of the readthrough RNA in this gel system. The analysis in Fig. 3B shows that the level of the 3,000-nucleotide RNA increases sharply when processing at the normal site is impaired. The increase in quantity of the readthrough transcript argues that the loss of the 1,100-nucleotide band does not occur because of instability in the precursor RNA, or a transport problem, introduced by the deletions. It is interesting that the readthrough transcript was detected only in pB constructs containing SV40 inserts. We noted no analogous transcript from pB0. At this point, we have no explanation for this other than the possibility that the common sequences on the 5′ side of AAUAAA may have a stabilizing effect.

Overall, these data indicate a significant element of the 3′-end processing signal located between 46 and 67 nucleotides downstream of the AAUAAA. Considering our previous data on viable SV40 deletion mutants (38), we can further define this downstream element. Figure 4 shows nuclease S1 analysis of late RNA from cells lytically infected with wild-type SV40 or viable deletion mutant dl1265. The region deleted in dl1265 is underlined in Fig. 2B. Total cellular RNA was hybridized to DNA probes made from the homologous viral DNAs. The specific probes used (see the legend to Fig. 4) generated a 686-nucleotide band when hybridized to late RNA processed at the proper site; larger bands represent readthrough transcripts which have been characterized previously (38). In Fig. 4 it is apparent that the level of readthrough transcripts is elevated in RNA from cells infected with dl1265, indicating lowered efficiency of normal site utilization (38). However, as previously shown (38) and reiterated in Fig. 4, at least 90% of the late RNA is still processed at the normal site despite the deletion of sequences between 20 and 58 nucleotides downstream of the AAUAAA (Fig. 2A). Considering this and the data of Fig. 3 (comparing pB28L and pB29L), we conclude by deduction that the downstream element we detected may be limited to the sequence AGGUUUUUU between 59 and 67 nucleotides downstream from the AAUAAA (Fig. 2B). The data with dl1265 also suggest that the location of the downstream element need not be fixed, since it functions relatively efficiently in dl1265, where it is 39 nucleotides closer to the AAUAAA than it is in the wild type.

Nuclease S1 analysis of RNA produced by deletion plasmids. To confirm and extend the results shown in Fig. 3, we analyzed total RNA from COS cells transfected with pBL, pB28L, pB29L, and pB0 for 3′-end utilization by nuclease S1 analysis. The probe was prepared from pBL and is described in the figure legend. With this probe, RNAs with 3′ ends at the normal site generate a 299-nucleotide protected fragment. Readthrough transcripts extend to the point of deletion in each plasmid. These bands would be 332 and 354 nucleotides in the cases of pB28L and pB29L, respectively. Figure 5 shows that with the control plasmid pB0 there are no protected bands as expected, since this plasmid contains no insert. For each of the other plasmids, there is a clear band at 299 indicating that the normal cleavage site can be used by all of the plasmids despite the deletions. In agreement with the data of Fig. 3, the deletion of sequences between pB29L and pB28L results in at least a 75% reduction in the utilization of the normal cleavage site. In addition, a 332-nucleotide band appears in the pB28L lane indicative of readthrough transcripts. Such transcripts are predicted for pB28L by Northern analysis (Fig. 3A). Overall, these data agree completely with the conclusions drawn above.

DISCUSSION

The data presented above suggest that 3′-end cleavage of SV40 late mRNAs involves not only the known hexanucleotide AAUAAA but also a sequence element AGGUUUUUU positioned between 59 and 67 nucleotides downstream of the AAUAAA (Fig. 2). Deletion of this region resulted in at least a 75% drop in 3′-end cleavage at the normal site. Corresponding with this loss was an increase in readthrough transcripts which was predicted, and expected, from previous results (38). The data also suggest that the position of this element downstream of the AAUAAA can be somewhat variable. For example, in dl1265 the element is 39 nucleotides closer to the AAUAAA and, in this case, normal 3′ end utilization is diminished only 10% compared with the wild type.

The U series of small RNAs (36) has been implicated in RNA processing mechanisms. For 3′-end processing specifically, both the U1 and U4 RNAs have been suggested as participating in site selection or utilization or both (3, 33). Having identified a significant sequence element downstream of the AAUAAA of the SV40 late RNAs, we used computer sequence comparisons to determine any complementarity with U-series small RNAs. Comparing the SV40 late RNA downstream sequences (the AAUAAA and 125 downstream nucleotides) with the U series RNAs, we found potential for an extensive base-paired structure (Fig. 6) with U4 RNA (37). Much of this complementarity may be phys-

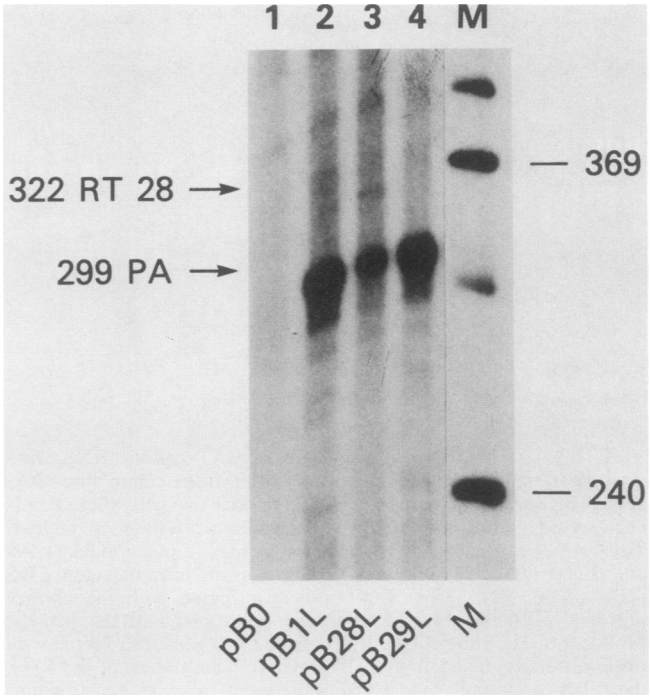

FIG. 5. Nuclease S1 analysis of RNA produced by the deleted plasmids. Total RNA from COS cells transfected for 48 h with pB0, pBL, pB28L, and pB29L were analyzed for 3′-end utilization. The probe was pBL which had been cleaved with *Nco*I. The plasmid has two *Nco*I sites (Fig. 1), one within the promoter region and the other 159 nucleotides upstream from the *Bam*HI insertion site. The larger of the two fragments was isolated and ^{32}P labeled by T4 DNA polymerase replacement synthesis (38). An RNA ending at the normal 3′ end will protect a 299-nucleotide fragment. Extended transcripts will end at the point of the deletions in pB28L and pB29L (332 and 354 nucleotides, respectively).

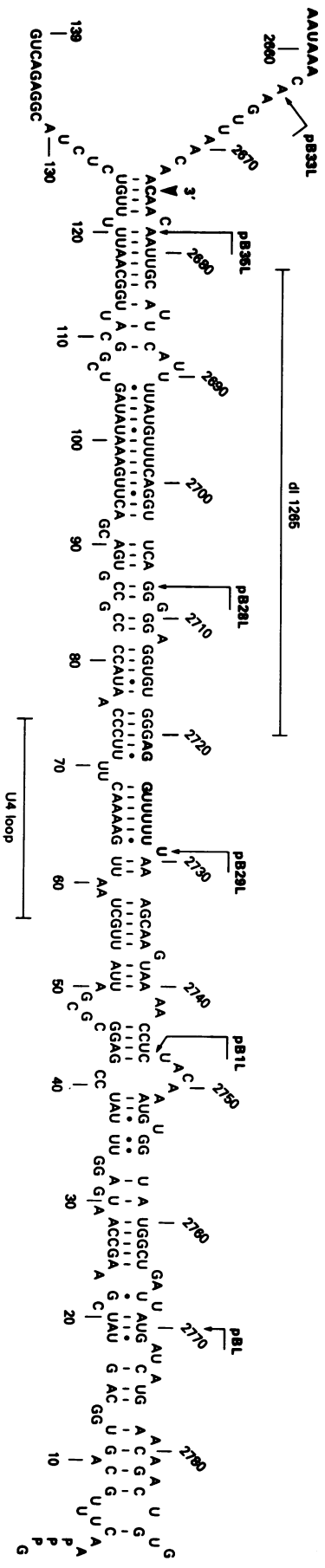

FIG. 6. Potential base pairing between the SV40 late RNA primary transcript and U4 RNA. The 125 nucleotides shown on the top strand are SV40 sequences in the primary transcript starting with the AAUAAA. The AAUAAA, the normal cleavage site, and the downstream element are shown in bold print. The endpoints of the various deletions used in this study, as well as the region deleted in dl1265, are indicated. The bottom strand is the entire 139-nucleotide sequence of U4 RNA (37). The single-strand loop of the U4 RNA secondary structure is indicated (19, 22).

iologically irrelevant, since the U4 RNA exists in a determined secondary structure within a ribonucleoprotein complex (19, 22). However, it should be pointed out that the 3′ ends of early RNAs are homologous through this region, suggesting that the 3′ end of early RNA may be able to assume a structure similar to that of U4 RNA. What this might mean to SV40 biology is unknown. Despite the extensive secondary structure of U4 RNA, a prominent single-stranded loop is formed (22); the bases included in this loop are indicated in Fig. 6. The loop can potentially form base pairs in the region of the AGGUUUUUU downstream element and its surrounding nucleotides (allowing rG-U base pairing). Thus, a theoretical mechanism can be suggested in which U4 RNA could form base pairs through its loop nucleotides with a precursor RNA and then search upstream for an AAUAAA signal.

In Table 1 we analyzed downstream sequences of other genes for complementarity with the sequences of the U4 RNA loop as well as homology to the AGGUUUUUU element. In most cases, reasonable complementarity (allowing rG-U base pairs) to the U4 loop could be found within 16 and 83 nucleotides downstream of the AAUAAA. This range of location is reasonable considering the dl1265 data discussed above. Within these complementary regions there is a very striking conservation of sequences resembling the AGGUUUUUU element defined by our analysis. A generalized form of this sequence based on the data in Table 1 is RGUUUUYRR.

Recently McLauchlan et al. (30) and Birnstiel et al. (5) have used sequence comparisons to suggest theoretical downstream elements involved in 3′-end cleavage. Respectively, these are denoted as a consensus sequence (YGUGUUYY; 30) or G-T clusters (5). Both groups predict that these elements are located between 26 and 35 nucleotides downstream of the AAUAAA in the case of SV40 late RNAs. Additionally, a recent report by Conway and Wickens (10) used frog oocyte injection techniques to suggest that sequences between 21 and 37 nucleotides downstream of the AAUAAA are important for SV40 late RNA 3′-end formation. Considering these predictions and data, it is interesting to note that this region is completely deleted in the viable viral mutant dl1265 (dl1265 deletes the nucleotides between 20 and 58 nucleotides downstream from the AAUAAA; see the underlined nucleotides in Fig. 2B). We have shown, both previously (38) and above, that dl1265 forms normal late mRNA 3′ ends at efficiencies equal to at least 90% of that of the wild type during lytic infection in the natural host. Clearly, in the dl1265 situation, these sequences can be deleted with little effect on late 3′-end formation. These variations may reflect the alterations in sequence structure caused by deletions; indeed, the data presented in this paper suggest that sequences between 15 and 45 nucleotides downstream of the AAUAAA (sequences between pB35L and pB28L; see Fig. 2B) can direct proper 3′-end utilization, in the absence of the AGGUUUUUU element, at approximately 25% of the wild-type level. Therefore, we suggest that available data imply the possibility of multiple downstream signals (often rich in U's) which may function alternately under different conditions. For example, in the case of the SV40 late genes, the oocyte may emphasize the utilization of slightly different downstream elements in comparison with monkey cells. Such a scenario could help explain situations in which a specific mRNA utilizes variable 3′-end sites in a tissue- or developmentally specific manner (6, 11, 27). With respect to the possibility of U4 RNA involvement in 3′-end processing, it should be pointed out

TABLE 1. Tabulation of homologous sequences from other genes[a]

Gene	Reference	Perfect match[b]																	
		G	G	R	R	R	R	G	U	U	U	U	Y	R	R	U	U	R	G	NDS[c]
SV40 late promoter	41	u	G	G	G	A	G	G	U	U	U	U	U	u	A	a	a	G	c	62
												g								
SV40 early promoter	41	G	u	u	G	u	G	G	U	U	U	U	C	c	A	a	a	c	u	30
Polyomavirus late promoter	41	c	a	A	G	G	G	G	U	c	g	U	C	G	c	c	U	u	c	39
BK virus early promoter	41	a	u	u	u	u	G	G	U	U	U	g	C	A	A	U	U	G	u	24
BK virus late promoter	41	G	G	u	G	G	u	G	U	U	U	U	a	G	G	c	c	u	u	52
												g								
HSV TK	42	u	G	G	G	u	c	G	U	U	U	U	U	c	A	U	a	A	a	20, 33
Hepatitis B surface antigen	39	a	c	u	c	u	c	G	U	U	U	U	U	G	c	c	U	u	c	31*
Adenovirus 2 E1A	41	u	a	A	A	G	G	G	U	a	U	a	U	A	A	U	g	c	G	55
Adenovirus 2 E1B	41	u	u	A	G	G	G	G	U	U	U	U	g	c	G	c	g	c	G	58
													a		g					
Adenovirus 2 E2	41, 29	a	c	A	u	u	A	u	U	U	U	U	C	A	G	U	U	u	u	40*
Adenovirus 2 E3–1	23	G	u	u	G	c	G	c	U	U	U	U	C	u	G	U	g	c	G	47
Adenovirus 2 E3–2	23	u	c	c	A	G	c	u	U	a	U	U	C	A	G	c	a	u	c	40
						c														
Adenovirus 2 E4	23	G	c	A	A	A	G	G	U	U	c	a	C	G	A	U	U	c	u	27
Adenovirus 2 IV A-2	41	a	G	u	c	u	G	G	U	U	U	U	U	u	A	U	U	c	u	44
Adenovirus 2 L1	41	G	c	G	u	u	G	G	U	U	U	U	C	u	u	g	U	A	u	33
													g				a			
Adenovirus 2 L2	41	G	u	A	A	c	u	a	U	U	U	U	U	A	G	a	U	G	G	39
Adenovirus 2 L3	41	G	G	G	u	G	A	u	U	a	U	U	U	A	c	c	c	c	c	35
Adenovirus 2 L4	41	G	c	c	A	c	c	G	U	U	U	U	U	A	c	c	c	A	c	54
Adenovirus 2 L5	24	u	G	u	u	u	A	u	U	U	U	U	C	A	A	U	U	G	c	41
Bovine growth hormone	45	G	G	G	A	G	G	a	U	U	g	g	g	A	A	g	a	c	a	83*
Human delta globin	34	G	u	c	c	A	G	G	U	g	U	g	U	A	A	g	a	A	G	56
Mouse I-A beta light chain	25	a	c	u	c	u	G	G	U	c	U	U	U	G	A	U	U	u	G	18
T24 c-Ha-Ras	7	a	u	c	u	u	G	G	U	U	U	U	C	G	G	c	U	G	a	24
Chicken alpha-1 collagen[d]	1	G	a	G	G	A	A	G	U	U	U	a	a	A	A	c	c	A	c	55, 60
		a	u	u	A	A	A	G	U	U	a	c	U	u	G	c	U	u	a	42
Mouse complement factor 3	44	c	u	u	u	c	A	G	U	g	U	U	U	G	G	U	U	u	G	30
													g							
Mouse metallothionein I	15	G	G	u	G	u	G	G	U	U	U	g	C	A	A	U	a	A	G	37
Rabbit beta globin	14	u	u	G	G	A	A	u	U	U	U	U	U	G	u	g	U	c	u	36*

[a] Sequences were aligned for complementarity to the U4 loop sequence. This is similar to analyzing on the basis of homology to the region surrounding the downstream element defined by our deletion studies. Permitting rG-U pairing defines the perfect match (R, purine; Y, pyrimidine). Minor bulges were permitted in several sequences and are printed above the positions at which they occur.

[b] Perfect match with U4 RNA loop, allowing rG-U pairing.

[c] NDS (nucleotides downstream) indicates the number of nucleotides separating the AAUAAA from the position of the underlined U in the putative perfect match. Two NDS numbers reflect the distance from each of two possible AAUAAA sequences in the specific gene. An asterisk indicates that reported studies involving these genes are compatible with the choice of this sequence as a downstream element.

[d] In the case of the chick collagen gene, two separate polyadenylation sites were analyzed.

that the SV40 late RNA sequences indicated as being important in the frog oocyte study (10) are complementary to another single-stranded loop structure of U4 RNA (see U4 nucleotides numbered 96 through 103 in Fig. 6) different from the one discussed above. Again, the possibility of multiple signals is suggested. For example, a putative 3'-end selection mechanism involving U4 RNA may function if complementarity with either or both loops occurs, each situation resulting in a different efficiency of cleavage site utilization.

ACKNOWLEDGMENTS

We thank Jane Picardi for excellent technical assistance.

This work was supported by Public Health Service grants CA28379 awarded to J.C.A. by the National Cancer Institute and GM 28983 awarded to J.L.M. by the National Institute of General Medical Sciences.

LITERATURE CITED

1. **Aho, S., V. Tate, and H. Boedtker.** 1983. Multiple 3' ends of the chicken proα2(I) collagen gene. Nucleic Acids Res. **11**:5443–5450.
2. **Alwine, J. C., D. J. Kemp, and G. R. Stark.** 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl paper and hybridization with DNA probes. Proc. Natl. Acad. Sci. U.S.A. **74**:5350–5354.
3. **Berget, S. M.** 1984. Are U4 small nuclear ribonucleoproteins involved in polyadenylation? Nature (London) **309**:179–182.
4. **Berk, A. J., and P. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. Cell **12**:721–732.
5. **Birnstiel, M. L., M. Busslinger, and K. Strub.** 1985. Transcription termination and 3'-end processing: the end is in sight! Cell **41**:349–359.
6. **Capetanaki, Y. G., J. Ngai, C. N. Flytzanis, and E. Lazarides.** 1983. Tissue-specific expression of two mRNA species transcribed from a single vimentin gene. Cell **35**:411–420.
7. **Capon, D. J., E. Y. Chen, A. D. Levenson, P. H. Seeburg, and D. V. Goeddel.** 1983. Complete nucleotide sequence of the T24 human bladder carcinoma oncogene and its normal homologue. Nature (London) **302**:33–37.
8. **Cole, C. N., T. Landers, S. P. Goff, S. Manteuil-Brutlag, and P. Berg.** 1977. Physical and genetic characterization of deletion mutants of simian virus 40 constructed in vitro. J. Virol. **24**:277–294.
9. **Cole, C. N., and G. M. Santangelo.** 1983. Analysis in Cos-1 cells

of processing and polyadenylation signals by using derivatives of the herpes simplex virus type 1 thymidine kinase gene. Mol. Cell. Biol. 3:267–279.
10. **Conway, L., and M. Wickens.** 1985. A sequence downstream of AAUAAA is required for formation of SV40 late mRNA 3'-ends in frog oocytes. Proc. Natl. Acad. Sci. U.S.A. **82**:3949–3953.
11. **Early, P., J. Rogers, M. Davis, K. Calame, M. Bond, R. Wall, and L. Hood.** 1980. Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. Cell **20**:313–319.
12. **Fitzgerald, M., and T. Shenk.** 1981. The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell **24**:251–260.
13. **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 SV40. J. Mol. Biol. **126**:813–846.
14. **Gil, A., and N. J. Proudfoot.** 1984. A sequence downstream of the AAUAAA is required for rabbit β-globin mRNA 3'-end formation. Nature (London) **312**:473–474.
15. **Glanville, N., D. M. Durnam, and R. D. Palmiter.** 1981. Structure of mouse metallothionein-I gene and its mRNA. Nature (London) **292**:267–269.
16. **Gluzman, Y.** 1981. SV40-transformed simian cells support the replication of early SV40 mutants. Cell **23**:175–182.
17. **Gorman, C. M., L. F. Moffat, and B. H. Howard.** 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. **2**:1044–1051.
18. **Gorman, C. M., R. Pabmanabhan, and B. H. Howard.** 1983. High efficiency DNA mediated transfection of primate cells. Science **221**:551–553.
19. **Hashimoto, C., and J. A. Steitz.** 1984. U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. Nucleic Acids Res. **12**:3283–3293.
20. **Higgs, D. R., S. E. Y. Goodburn, J. Lamb, J. B. Clegg, D. J. Weatherall, and N. J. Proudfoot.** 1983. α-thalassemia caused by a polyadenylation signal mutation. Nature (London) **306**:398–400.
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. **Krol, A., and C. Branlant.** 1981. Primary and secondary structures of chicken, rat, and man nuclear U4 RNAs. Homologies with U1 and U5 RNAs. Nucleic Acids Res. **9**:2699–2716.
23. **Lehrach, H., D. Diamond, J. M. Wozney, and H. Boedtker.** 1977. RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry **16**:4743–4751.
24. **Le Moullec, J. M., G. Akusjärvi, P. Stålhandske, U. Pettersson, B. Chambraud, P. Gilardi, M. Nasri, and M. Perricaudet.** 1983. Polyadenylic acid addition sites in the adenovirus type 2 major late transcription unit. J. Virol. **48**:127–134.
25. **Malissen, M., T. Hunkapiller, and L. Hood.** 1983. Nucleotide sequence of a light chain gene of the mouse I-A subregion $A\beta^d$. Science **221**:750–754.
26. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
27. **Mather, E. L., K. J. Nelson, J. Haimovich, and R. P. Perry.** 1984. Mode of regulation of immunoglobulin μ and δ-chain expression varies during B-lymphocyte maturation. Cell **36**:329–338.
28. **Maxam, A. E., and W. S. Gilbert.** 1980. Sequencing end-labeled DNA with base specific chemical cleavages. Methods Enzymol. **65**:499–560.
29. **McDevitt, M. A., M. J. Imperiale, H. Ali, and J. R. Nevins.** 1984. Requirement of a downstream sequence for generation of a poly(A) addition site. Cell **37**:993–999.
30. **McLauchlan, J., D. Gaffney, J. L. Whitton, and J. B. Clements.** 1985. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. Nucleic Acids Res. **13**:1347–1368.
31. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.
32. **Montell, C., E. F. Fisher, M. H. Caruthers, and A. J. Berk.** 1983. Inhibition of cleavage but not polyadenylation by a point mutation in the mRNA 3' consensus sequence AAUAAA. Nature (London) **305**:600–605.
33. **Moore, C. L., and P. A. Sharp.** 1984. Site-specific polyadenylation in a cell-free reaction. Cell **36**:581–591.
34. **Poncz, M., E. Schwartz, M. Ballantine, and S. Surrey.** 1983. Nucleotide sequence analysis of the δβ-globin gene region in humans. J. Biol. Chem. **258**:11599–11609.
35. **Proudfoot, N., and G. G. Brownlee.** 1976. 3' non-coding region sequences in eukaryotic messenger RNA. Nature (London) **263**:211–214.
36. **Reddy, R., and H. Busch.** 1983. Small nuclear RNAs and RNA processing. Prog. Nucleic Acid Res. Mol. Biol. **30**:127–161.
37. **Reddy, R., D. Henning, and H. Busch.** 1981. The primary nucleotide sequence of U4 RNA. J. Biol. Chem. **255**:3532–3538.
38. **Sadofsky, M., and J. C. Alwine.** 1984. Sequences on the 3' side of hexanucleotide AAUAAA affect efficiency of cleavage at the polyadenylation site. Mol. Cell. Biol. **4**:1460–1468.
39. **Simonsen, C. C., and A. D. Levinson.** 1983. Analysis of processing and polyadenylation signals of the hepatitis B virus surface antigen gene by using simian virus 40-hepatitis B virus chimeric plasmids. Mol. Cell. Biol. **3**:2250–2258.
40. **Thomas, P.** 1983. Hybridization of denatured RNA transferred or dotted to nitrocellulose paper. Methods Enzymol. **100**:255–266.
41. **Tooze, J. (ed.).** 1981. Molecular biology of tumor viruses. Part 2. DNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
42. **Wagner, M. J., J. A. Sharp, and W. C. Summers.** 1981. Nucleotide sequence of the thymidine kinase gene of herpes simplex virus type 1. Proc. Natl. Acad. Sci. U.S.A. **78**:1441–1445.
43. **Wickens, M., and P. Stephenson.** 1984. Role of the conserved AAUAAA sequence: four point mutants prevent messenger RNA 3' end formation. Science **226**:1045–1051.
44. **Wiebauer, K., H. Domdey, H. Diggelmann, and G. Fey.** 1982. Isolation and analysis of four genomic DNA clones encoding the third component of mouse complement. Proc. Natl. Acad. Sci. U.S.A. **79**:7077–7081.
45. **Woychik, R. P., R. H. Lyons, L. Post, and F. M. Rottman.** 1984. Requirement for the 3'-flanking region of the bovine growth hormone gene for accurate polyadenylation. Proc. Natl. Acad. Sci. U.S.A. **81**:3944–3948.