JOURNAL OF VIROLOGY, Nov. 1982, p. 610–624
0022-538X/82/110610-15$02.00/0
Copyright © 1982, American Society for Microbiology
Vol. 44, No. 2

Altered Utilization of Splice Sites and 5′ Termini in Late RNAs Produced by Leader Region Mutants of Simian Virus 40

PRABHAT K. GHOSH,[1] MICHAEL PIATAK,[2] JANET E. MERTZ,[3] SHERMAN M. WEISSMAN,[2] AND PAUL LEBOWITZ[1]*

Departments of Internal Medicine[1] and Human Genetics,[2] Yale University School of Medicine, New Haven, Connecticut 06510, and McArdle Laboratories for Cancer Research, Department of Oncology, University of Wisconsin Medical School, Madison, Wisconsin 53706[3]

Received 21 May 1982/Accepted 9 August 1982

We compared the 5′ termini and splices of the late 16S and 19S RNAs synthesized by wild-type simian virus 40 and five mutants containing deletions in their late leader region. All mutants produced more unspliced 19S RNA than did wild-type virus, and in two mutants, unspliced 19S RNA constituted more than 60% of the total 19S species. The other three mutants each utilized predominantly a different one of the three spliced species of 19S mRNA. All mutants also produced decreased quantities of 16S mRNA, indicating that they may be defective for splicing both late RNAs. None of the 5′ termini of the 16S and 19S RNAs made by the five mutants predominated as in those made by the wild type. Some of the mutant 5′ termini were the same as those used by the wild type, whereas others were different. Although present, the major 5′-end positions used by the wild type were frequently not used as major sites by the mutants. In addition, mutants with very similar deletion endpoints synthesized RNAs with different 5′ ends. Thus, downstream mutations have a pronounced effect on the location of 5′ ends of the late RNAs, and there is no obvious involvement of a measuring function in the placement of 5′ ends. For all mutants and wild-type virus, the 5′ termini used for 16S and 19S RNAs showed major differences, with some degree of correlation found between the 5′ ends and the internal splices of specific mRNA species. A model for the regulation of simian virus 40 late gene expression is presented to explain these findings.

Until 1978, the late region of simian virus 40 (SV40) was thought to code for two discrete mRNAs: (i) 16S mRNA, coding for synthesis of VP-1, the major capsid protein, and (ii) 19S mRNA, serving as the template for synthesis of the minor capsid proteins VP-2 and VP-3. However, studies conducted during the past several years have revealed considerable complexity in these mRNAs, with respect to both their splicing patterns and their 5′ termini (4, 11, 25, 26, 34). The 16S mRNAs all contain an identical "body" region spliced in a unique way to a 5′-terminal leader. However, they fall into three subclasses on the basis of their leader sequences: (i) most of the 16S mRNAs contain leaders colinear with SV40 DNA; (ii) some have spliced leaders; and (iii) a small percentage have leaders possessing a tandem duplication of 93 nucleotides. The 19S RNAs also have an identical body region, but they exhibit even more diversity in their leader segments. One subclass is completely unspliced, whereas three subclasses of 19S mRNA contain splices that join the 5′ end of the 19S body to leader sequences with different 3′ termini (see Fig. 1). Furthermore, within the subclasses of 16S and 19S RNAs, individual members differ in the locations of their 5′ termini. These 5′ termini span a 300-nucleotide region of the viral genome (11) and overlap the region encoding the early-strand transcriptional start sites by approximately 60 nucleotides (8).

The diversity in the structures of the late mRNAs indicates the existence of complex mechanisms for the regulation of late gene expression. Furthermore, even though mutants with deletions in the late leader region are viable (1, 18, 30, 37, 38; J. Mertz, unpublished data), one would expect at least some of the late transcriptional control signals to map in this region. Complicating the picture further is the discovery that a nucleic acid-binding protein, the agnogene polypeptide, is encoded in this region (21, 22; P. Berg, personal communication; D. Nathans, personal communication; J. E. Mertz, A. Murphy, and A. Barkan, J. Virol., in press).

To gain insight into some of the mechanisms controlling late RNA synthesis and processing,

we and others have analyzed the structure of the RNAs synthesized by viable mutants containing deletions in the leader and immediately upstream region of the SV40 genome. We have reported previously (12) that the mutant dl-805, lacking all but 15 nucleotide pairs of the region encoding the major leader, synthesizes 19S RNAs, all or virtually all of which are unspliced over their entire length, contain intervening sequences, and function as mRNAs. Furthermore, this mutant (12), three mutants with deletions in successive segments of the major late leader region (32, 33), and late leader mutants described by others (18, 40) use new genomic sites for the 5' termini of at least some of their major late RNAs.

Here we present a more detailed analysis of the primary structure of the late RNA species produced by five late leader mutants. Our major findings are: (i) all five mutants are defective in splicing late-strand primary transcripts; (ii) a wide variety of 19S RNA splicing patterns are compatible with viral viability; (iii) downstream mutations have dramatic effects on the localizations of late 5' termini; and (iv) some correlation exists between how an RNA is spliced and the location of its 5' end. To explain these diverse and apparently confusing findings, we hypothesize that expression of the late region of SV40 is regulated by two or more overlapping promoter regions, each responsible for controlling the synthesis of one or more of the late virus-coded proteins. The complex control of SV40 late transcription and processing suggests the possibility that expression of other eucaryotic genes is also regulated by extremely complex mechanisms.

MATERIALS AND METHODS

The isolation (30), biological properties (30), and nucleotide sequences of the genomes (1) of the late leader mutants dl-805, dl-806, dl-809, and dl-810 have been reported previously. Mutant dl-1470 was isolated and its late leader region was sequenced by Tom Shenk (unpublished data); this mutant was kindly provided to us by him.

Confluent monolayers of Vero African green monkey kidney cells in roller bottles (750 cm^2) were infected with wild-type SV40 strains 776 and 800 and the late leader mutants at multiplicities of infection of approximately 20 PFU per cell and maintained at 37°C as described previously (11). The cells were harvested with a rubber policeman, washed with normal saline, and lysed by gently swirling for 10 min in 10 mM Tris-hydrochloride (pH 7.5)–10 mM NaCl–3 mM $MgCl_2$–0.5% Nonidet P-40. Nuclei and cytoplasm were separated by low-speed centrifugation, and polyadenylated RNAs were extracted as described previously (11). Extractions performed with the same buffer containing 0.14 M NaCl yielded similar populations of cytoplasmic RNAs (12).

Nucleotide sequences of the 5'-terminal regions of the 16S and 19S RNAs made by wild-type and mutant viruses were determined by "primer extension." This methodology has been described in detail previously (10, 11, 29).

RESULTS

Genomic structure of mutants. Figure 1 depicts schematically the structures of the late leader regions of the two wild-type strains used in these studies (strains 776 and 800) and the mutants derived from each of them (dl-805, dl-806, dl-809, and dl-810 from strain 800 and dl-1470 from strain 776). The two wild-type strains differ in one and only one respect in this region: whereas strain 776 contains a 72-base-pair tandem repeat within the late leader region (nucleotides 25 through 96 and 97 through 168) (7, 35), strain 800 contains an 85-base-pair tandem repeat in which residues 97 through 104 of the sequence in strain 776 are replaced by a second copy of residues 169 through 189 (here designated 169' through 189') (12; L. Trimble and J. Mertz, unpublished data).

Each of our five mutants contains a large deletion in the late leader region. The sizes of the deletions range from 137 nucleotides in dl-806 to 249 nucleotides in dl-1470. dl-805 contains a 187-base-pair deletion which maps totally within the major late leader. The 5' end of this deletion lies six nucleotides downstream from residue 243, the map location of the 5' terminus of most 16S RNAs and several 19S RNAs, whereas its 3' end lies nine nucleotides upstream from residue 444, the donor splice site of all 16S RNAs and some 19S RNAs. As noted previously, although the 16S RNAs made by this mutant are spliced, its 19S RNAs are unspliced (12).

Mutants dl-809 and dl-810 possess similar deletions that start at a common 5' site (residue 215) but terminate five nucleotides apart (at residues 384 and 389, respectively). We chose to investigate the primary structures of the RNAs of these two mutants for four reasons: (i) despite their similar genomic structures, dl-810 is far more defective for growth on monkey cells than dl-809 (1, 30); (ii) both mutants lack residue 243, the map location of the most abundant 5' terminus of the late wild-type RNAs; (iii) the deletions in these mutants remove the donor splice site at residue 291 and come close to the donor splice site at residue 212; and (iv) mutants possessing similar deletions should synthesize late RNAs whose 5' ends map at similar or identical sites if measuring functions for determining the locations of the 5' ends of the RNAs exist.

The deletions in dl-806 and dl-1470 also begin at a common 5'-terminal site (residue 137) and extend for different distances, i.e., to residues 374 and 390, respectively. This pair of mutants was also of interest for four reasons: (i) dl-806 grows at a slow rate, whereas dl-1470, with the

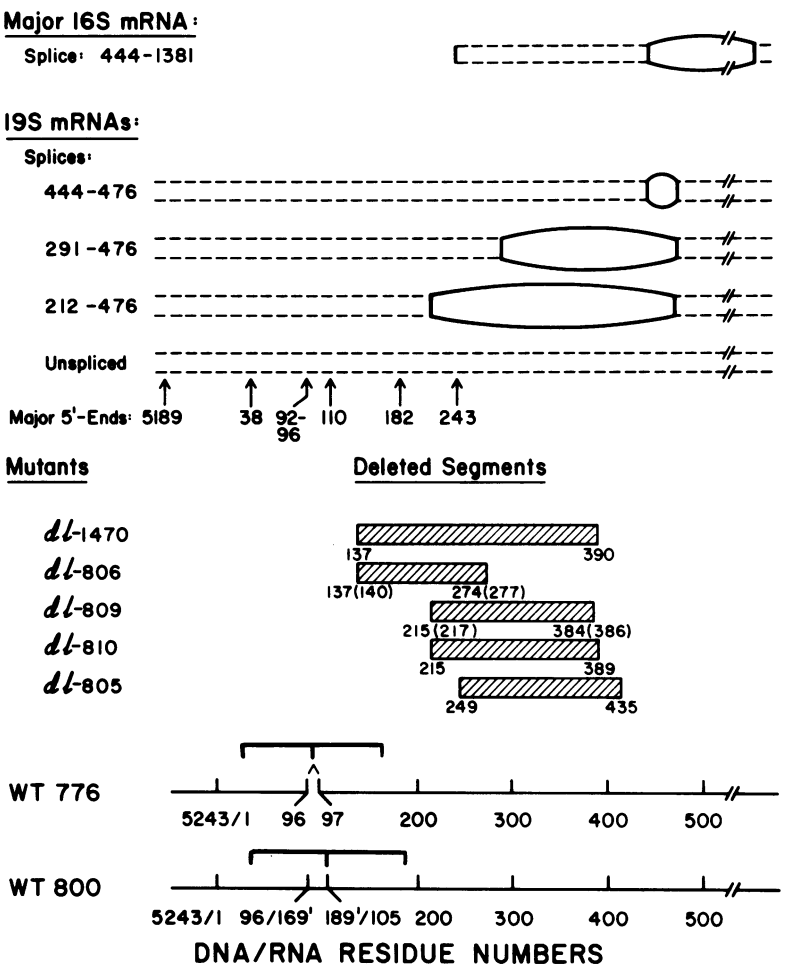

FIG. 1. Structures of wild-type SV40 late RNAs and late leader deletion mutants. Brackets indicate the positions of the 72- and 85-base-pair tandem repeats in the late leader regions of wild-type strains 776 and 800, respectively. Locations of splices are shown by square-ended ovals. Nucleotides are numbered according to a modification (33) of the system of Reddy et al. (35). Nucleotide numbers in the system of Buchman et al. (3) can be calculated from those used here by adding 82 to nucleotides 1 through 5,161 and subtracting 5,161 from nucleotides 5,162 through 5,243.

larger deletion, grows more rapidly; (ii) both lack not only residue 243, but also residue 182, the map location of the second most abundant 5′ terminus of the late RNAs; (iii) both lack the donor splice site at residue 212, and dl-1470 also lacks the donor splice site at residue 291 whereas the deletion in dl-806 comes to within 16 nucleotides of this site; and (iv) these mutants also presented the opportunity to look for involvement of a measuring function in the placement of the 5′ ends of the late RNAs.

Splices in wild-type and mutant 19S RNAs. In previous studies (11, 34), we determined the primary structures of the 5′-terminal regions of wild-type SV40 late RNAs by the method of primer extension. This procedure involves (i) annealing a viral DNA fragment labeled with ^{32}P at the 5′ terminus to the 5′ end of the body region of a specific RNA, (ii) extending the primer in a 5′ → 3′ direction with reverse transcriptase, (iii) separating the resultant cDNAs on an 8% polyacrylamide–7 M urea gel, and (iv) sequencing the individual cDNAs by the method of Maxam and Gilbert (29). Here we have used this procedure to analyze the 5′-terminal primary structures of the late RNAs made by our five mutants. Since the 5′ ends of the body regions of all of the late 19S RNAs lie at residue 476 (Fig. 1), all studies on the 19S RNAs were performed using an *Ava*II restriction fragment of wild-type SV40 DNA spanning residues 479 through 509 as a primer.

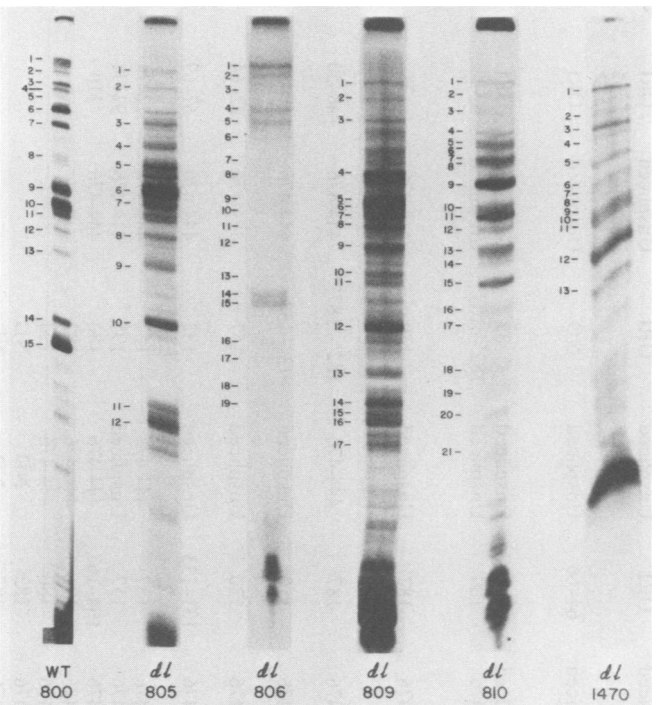

FIG. 2. cDNAs transcribed from 5' ends of wild-type and mutant 19S RNAs. Methods for the synthesis and separation of the cDNAs are given in the text. The 3' termini and splices of the numbered cDNAs are given in Table 1. Late RNA extracted from cells infected with wild-type strain 776 gave a cDNA pattern essentially identical to that shown for wild-type strain 800.

Gel electrophoretic patterns of the cDNAs synthesized from the 5' ends of polyadenylated cytoplasmic late 19S RNA obtained from African green monkey kidney cells infected with wild-type SV40 and each of the five deletion mutants are shown in Fig. 2. The pattern for wild-type strain 800 is similar if not identical to that published previously for wild-type strain 776 (11). As with wild-type virus, a multiplicity of cDNAs are also synthesized on the 5' termini of the 19S RNAs of each of the mutants. For both wild-type and mutant viruses, the cDNA patterns were qualitatively reproducible; however, for as yet unknown reasons, relatively small differences in the relative abundances of specific cDNAs were observed in certain experiments (data not shown). No obvious similarities in the patterns of cDNAs among mutants or between mutants and wild-type virus can be deduced from Fig. 2.

Table 1 lists the splices in all of the abundant species of wild-type and mutant 19S RNAs, as determined from sequence analysis of the numbered cDNAs shown in Fig. 2. By combining this information with estimates of the relative abundances of each cDNA (Fig. 3), as measured by densitometric tracings of the autoradiographs shown in Fig. 2, we determined the approximate percentages of each subclass of 19S RNA produced by each of the mutants and by the wild-type virus (Table 2). The major findings derived from the data presented in Fig. 3 and Table 2 are as follows.

(i) No new splices are present in the mutant 19S RNAs. However, the relative abundances of unspliced 19S RNA and the three subclasses of spliced 19S RNA differ greatly among the five mutants and the wild-type virus.

(ii) Most, if not all, of the dl-805 late 19S RNAs are unspliced. Approximately 20, 25, 40 and 60% of the 19S RNAs produced by dl-809, dl-806, dl-1470, and dl-810, respectively, are also unspliced at their 5' ends. Thus, all of our mutants produce relatively larger quantities of unspliced 19S RNA than wild type. Furthermore, since the mutants synthesize approximately the same absolute amounts of 19S RNA as wild type (A. Barkan and J. Mertz, manuscript in preparation), the absolute amounts of unspliced 19S RNA produced by the mutants are increased. For dl-805, we have shown that the 19S RNA is unspliced over its entire length and resides on polysomes (12). Piatak et al. (33) have also found increased levels of unspliced 19S

TABLE 1. cDNAs transcribed from 5′ ends of wild-type and mutant 19S RNAs

Gel band	Wild-type strain 776		dl-805 (249–435)[a]		dl-806 (137–274)[a]		dl-809 (215–384)[a]		dl-810 (215–389)[a]		dl-1470 (137–390)[a]	
	Splice	Terminus	Splice	Terminus	Splice	Terminus	Splice	Terminus	Splice	Terminus	Splice	Terminus
1	444-476	38	Unspliced	ND[b]	Unspliced, 444-476 (2:1)	ORI[c]	Unspliced	ORI	Unspliced	ORI	Unspliced	ORI
2	Unspliced	110	Unspliced	ND	Unspliced, 444-476 (1:2)	ORI	Unspliced	ORI	Unspliced	ORI	Unspliced	ORI
3	Unspliced	182	Unspliced	38	Unspliced, 444-476 (1:2)	38	Unspliced	64–66	Unspliced	ND	Unspliced	ORI
4	444-476	182	Unspliced	85	Unspliced	110	Unspliced	164–167	Unspliced	85	Unspliced	ORI
5	Unspliced	210–215	Unspliced	110–120	Unspliced, 444-476 (1:2)	110	212-476	180′	Unspliced	182′– 187′	Unspliced	57–59
6	Unspliced	243	Unspliced	157	Unspliced, 444-476 (1:2)	131	212-476	182′	Unspliced	110	444-476	38
7	444-476	243	Unspliced	164–166	Mostly 291-476	21	212-476	187′	444-476	182′– 187′	444-476	48–50
8	Unspliced	305–315	Unspliced	187–189	Mostly 291-476	38	212-476	110	Unspliced	131–137	444-476	57–59
9	212-476	92–96	Unspliced	200–203	Mostly 291-476	55	212-476	120	Unspliced	157	444-476	62–64
10	291-476	182	Unspliced	224–225	Mostly 291-476	64	212-476	131–133	Unspliced	182	444-476	67–69
11	291-476	187	Unspliced	436	291-476	72–75	212-476	137	444-476	157	444-476	72–74
12	291-476	194–198	Unspliced	441	291-476	85	212-476	157	Unspliced	195	444-476	92–94
13	291-476	208–212			291-476	170′– 172′	212-476	164–167	444-476	180	444-476	110
14	291-476	234–236			291-476	176′– 178′	212-476	176	444-476	182		
15	291-476	243			291-476	180′– 182′	212-476	180	444-476	194		
16					291-476	110	212-476	182	ND	ND		
17					291-476	115	212-476	187	ND	ND		
18					291-476	120–122			Unspliced	434–436		
19					291-476	127			Unspliced	440–441		
20									Unspliced	445		
21									Unspliced	453		

[a] Numbers within parentheses indicate deleted residues.
[b] ND, Not determined for technical reasons.
[c] ORI, Termini lying within the origin of replication as determined by chain length of cDNA bands.

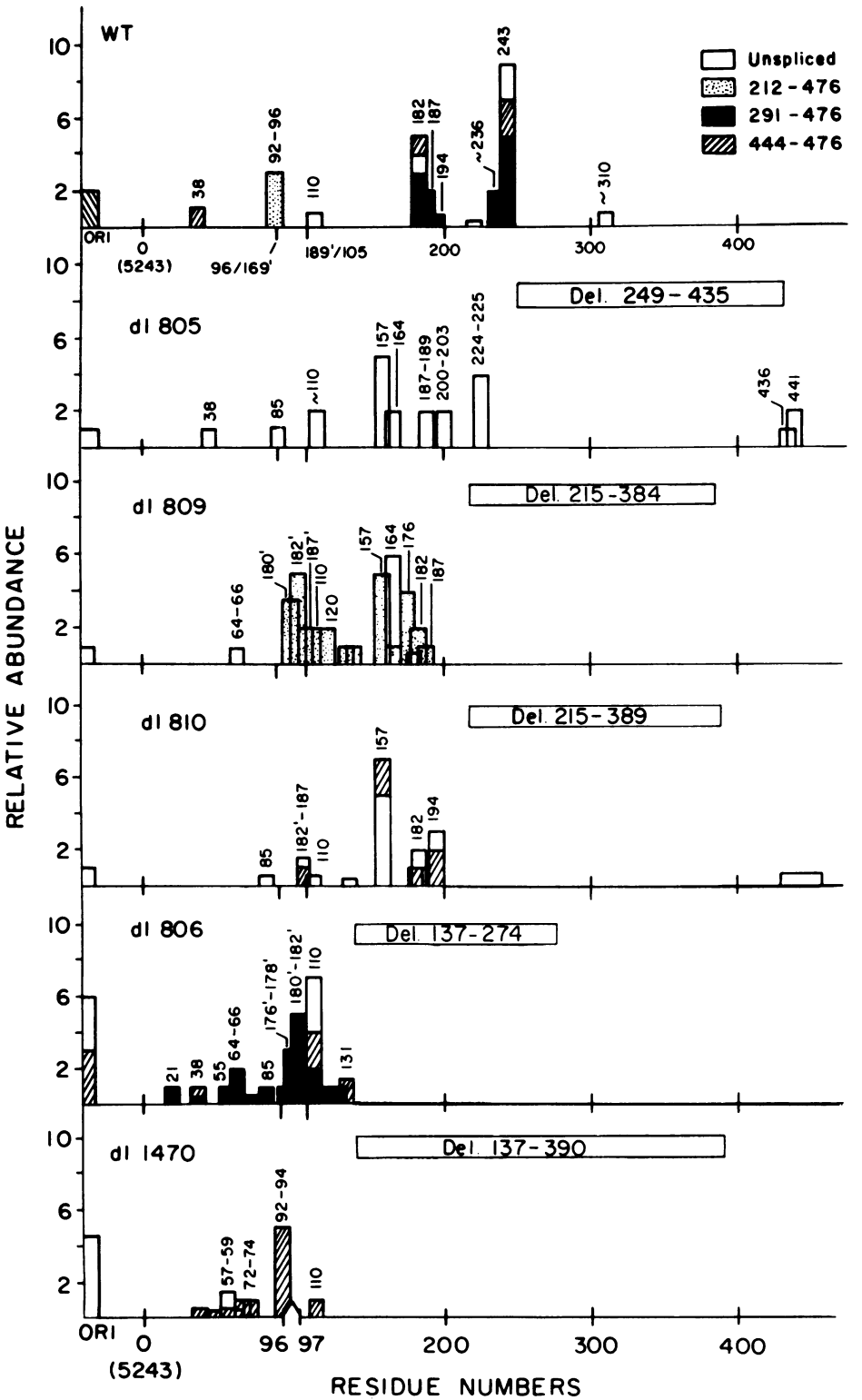

FIG. 3. Relative abundances of 5′ termini and splices of prominent 19S RNAs. The relative abundances of individual 19S RNAs were estimated by densitometry of the gels shown in Fig. 2 autoradiographed for various periods of time. For each virus, the most abundant cDNA(s) was assigned a value of +5, and other cDNAs were assigned values based on relative abundances. Splices and 5′ termini are taken from Table 1. The graph for the wild type was obtained from an analysis of strain 776 19S RNAs. However, similar data were obtained for strain 800 19S RNAs except that the strain 776 RNA terminating at residues 92-96 terminated in strain 800 at residues 178′-182′. ORI, Within origin of replication.

TABLE 2. Wild-type and mutant 19S RNA splicing patterns

Splice	% Total 19S RNA in:					
	Wild type[a]	dl-805	dl-806	dl-809	dl-810	dl-1470
Unspliced	5–15	100	24	20	60	35
212-476	5–10	<1	0[b]	80	<1	0[b]
291-476	40–50	0[b]	60	0[b]	0[b]	0[b]
444-476	25–40	<1	16	<1	40	65

[a] Similar results were obtained for wild-type strains 776 and 800.
[b] Donor splice site is deleted in mutant.

RNA in cells infected with three late leader mutants and have shown that they are also present on polysomes (although for one mutant the level of unspliced 19S RNA on polysomes was much less than in whole cytoplasm). Since all of our mutants, including dl-805, synthesize VP-2 and VP-3 (Barkan and Mertz, in preparation), we believe that their unspliced 19S RNAs can probably function as mRNAs.

(iii) Although all of the mutants studied retain the donor splice site at residue 444, they differ greatly in their utilization of it. Mutants dl-805 and dl-809 make no 19S RNA containing the 444-476 splice, whereas mutant dl-1470 uses it more frequently than wild type. The finding that some mutants fail to make any 19S RNA having this splice seems remarkable since these same mutants do use the donor splice site at residue 444 to make 16S mRNA (see below).

(iv) Of the three mutants that retain the donor splice site at residue 212, two (dl-805 and dl-810) do not make detectable amounts of RNA containing the 212-476 splice, whereas one (dl-809) uses this splice in the vast majority (approximately 80%) of its 19S RNA even though it is a very minor splice in wild type. The disparate behavior of dl-809 and dl-810 is particularly noteworthy since the deletions in both mutants have the same 5′ endpoint at residue 215 and nearby 3′ endpoints.

(v) dl-806, the only mutant retaining the donor splice site at residue 291, uses the 291-476 splice even more frequently than wild type and, seemingly, at the expense of the 444-476 splice.

(vi) Mutants dl-810 and dl-1470 delete both the 212 and the 291 donor splice sites and have deletions with similar 3′ ends. Nevertheless, the former makes primarily unspliced 19S RNA, whereas the latter makes 19S RNA primarily with the 444-476 splice.

These striking differences in the relative abundances of the various species of 19S RNA produced by wild-type SV40 and the five mutants studied here indicate that several factors, including specific sequences within the late leader region and the precise primary and, consequently, secondary structures of the initial transcripts, play important direct or indirect roles in regulating the processing of the late 19S RNAs.

Splices in wild-type and mutant 16S RNAs. The primary structures of the 5′-terminal regions of the 16S RNAs were also determined by primer extension. In this case, cDNAs were synthesized by the extension of primers binding just downstream from the 5′ end of the body of 16S RNA. The DNA primer used for the analysis of the dl-810 16S species was a restriction fragment spanning residues 1,427 through 1,457; a fragment spanning residues 1,382 through 1,415 was used for the remaining analyses (see Fig. 1).

Figure 4 shows the gel electrophoretic patterns of cDNAs obtained from the polyadenylated cytoplasmic late 16S RNAs synthesized in monkey cells infected with wild-type SV40 and each of our mutants. As with the 19S RNAs, the cDNA patterns of the 16S RNAs synthesized by each mutant are very different from one another and from the pattern seen with wild-type 16S RNAs. Whereas a single band predominates in the wild type, the cDNA pattern for each of the mutants is composed of numerous bands. Densitometric tracings of these cDNA patterns indicated that whereas the major species of 16S RNA made by the wild type constitutes 80 to 90% of the total, the most abundant 16S RNA species made by any of the mutants constitutes no more than 20 to 25% of the total.

The results of nucleotide sequence analysis of each of the numbered cDNAs shown in Fig. 4 are summarized in Table 3. The salient points regarding the primary structures of the 16S RNAs made by wild-type SV40, as deduced from the cDNA data, have been reported previously (34) and are as follows: (i) all of the RNAs contain a splice linking residues 444 and 1,381; (ii) the sequence of the leader of the principal RNA species is colinear with the viral DNA sequence; (iii) several of the minor species (corresponding to cDNAs 8, 9, and 11 of the wild type) contain a second splice within their leaders, linking residues 212 and 351; and (iv) the RNA corresponding to cDNA 12 contains a tandem duplication of residues 351 through 444.

Essentially all of the cDNAs made from the

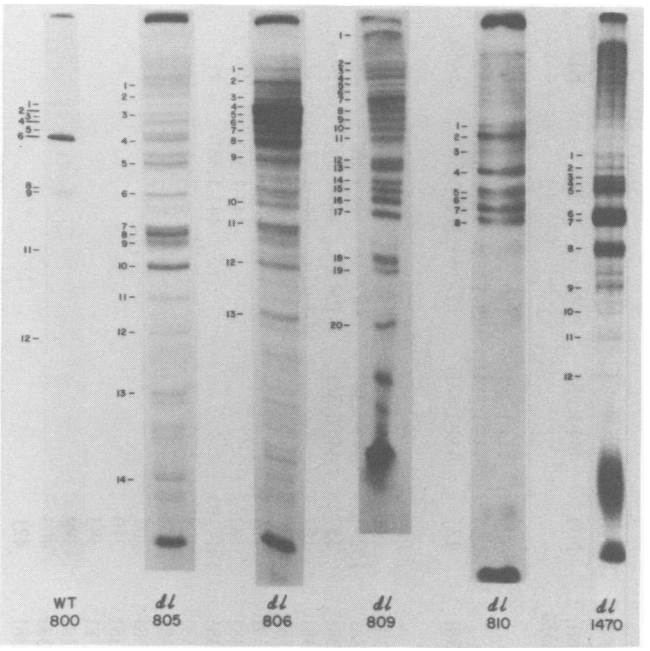

FIG. 4. cDNAs transcribed from 5' ends of wild-type and mutant 16S RNAs. Methods for the synthesis and separation of the cDNAs are given in the text. The 3' termini and splices of the numbered cDNAs are given in Table 3. Late RNA extracted from cells infected with wild-type strain 776 gave a cDNA pattern indistinguishable from that shown for wild-type strain 800.

16S RNAs of mutants dl-806, dl-809, dl-810, and dl-1470 also contain the 444-1,381 splice. cDNAs 5 through 10 of mutant dl-805 also have this splice; however, the remainder appear to lack any splice 5' to residue 1,381. The significance of the unspliced cDNAs of dl-805 is unclear and has been discussed previously (12). Although they could be derived from 19S RNAs by either premature termination of reverse transcription or RNA degradation, the possibility still exists that they are derived from processing intermediates or unspliced 16S RNA species.

Neither the 212-351 splice nor the duplication of residues 351 through 444 was seen in any of the 16S RNAs made by the mutants. The former was expected since none of the mutants studied contains the genomic sequences for both ends of this minor splice. In addition, only mutant dl-806 contains the sequences around residue 351 that are presumably needed to make the duplication. As with the 19S RNAs, no new splices, duplications, or sequence rearrangements were found in any of the mutant 16S RNAs.

5' Termini of mutant late RNAs. The residue locations of the 3' ends of the cDNAs numbered in Fig. 2 and 4 are presented in Tables 1 and 3, respectively. For both early and late SV40 RNAs, a good correspondence exists between the 3' ends of cDNAs as determined by primer extension (8, 11) and the sequences of 5'-terminal capped dinucleotides (4, 16, 17, 23; Y. Groner, personal communication). Furthermore, various control studies, including reverse transcription of SV40 cRNA (8) and the finding here that late RNAs produced by mutants with similar sequences yield cDNAs with different 3' ends, have led us to conclude that premature termination of reverse transcription or degradation of RNAs cannot account for the termini of the cDNAs we have observed. Hence, we believe that the 3' termini of all major cDNAs and probably those of most minor cDNAs correspond within one or two nucleotides to the 5' termini of in vivo species of viral RNA.

The relative abundances of the various sites serving as 5' termini for the 16S and 19S late RNAs produced by wild-type SV40 and our five mutants are summarized graphically in Fig. 3 and 5, respectively. Points of note include the following.

(i) The multiple 16S RNAs of wild-type strains 776 and 800 start at identical loci and are present in the same relative abundances. The 19S RNA species of both wild-type strains also start at identical loci, with one exception and one uncertainty: (a) the 19S RNA giving rise to cDNA 9 (Fig. 3) is located in the region of residues 92 to 96 in strain 776 but appears to be shifted down-

TABLE 3. cDNAs transcribed from 5′ ends of wild-type and mutant 16S RNAs

Gel band	Wild-type strains 776 and 800		dl-805 (249–435)[a]		dl-806 (137–274)[a]		dl-809 (215–384)[a]		dl-810 (215–389)[a]		dl-1470 (137–390)[a]	
	Splice	Terminus	Splice	Terminus	Splice	Terminus	Splice	Terminus	Splice	Terminus	Splice	Terminus
1	444-1,381	ND[b]	Unspliced	ND	Unspliced, 444-1,381 (1:1)	ND	Unspliced, 444-1,381 (1:1)	ND	444-1,381	182′	ND	ND
2	ND	ND	Unspliced	ND	444-1,381	ND	ND	ND	444-1,381	110	ND	ND
3	ND	ND	Unspliced	1,160–1,180	444-1,381	85–90	ND	ND	444-1,381	129–131	444-1,381	45–50
4	ND	ND	Mixture, mostly unspliced	1,224–1,228	444-1,381	182′	444-1,381	131	444-1,381	157	444-1,381	55–60
5	ND	ND	444-1,381	129–131	444-1,381	110–116	444-1,381	141	444-1,381	182	444-1,381	65–70
6	444-1,381	243	444-1,381	157	444-1,381	129–131	444-1,381	152–153	444-1,381	187	444-1,381	85
7	444-1,381	157	444-1,381	178–180	444-1,381	278–282	444-1,381	157	444-1,381	194	444-1,381	92–94
8	444-1,381	182	444-1,381	182	444-1,381	300–302	444-1,381	165		208–210	444-1,381	110
9	444-1,381	180–181	444-1,381	187	444-1,381	313–315	444-1,381	170			444-1,381	129–131
10	444-1,381	344	444-1,381	195–197	444-1,381	353–354	444-1,381	177–178			444-1,381	395
11	444-1,381	365–366	444-1,381	208–210	444-1,381	365–366	444-1,381	182			444-1,381	401
12	444-1,381	383–385	Unspliced	1,343–1,345	444-1,381	383–385	444-1,381	191–193			444-1,381	412
13	444-1,381	399	Unspliced	1,357	444-1,381	399–401	444-1,381	194				
14			Unspliced	1,357			444-1,381	200				
15							444-1,381	203				
16							444-1,381	210				
17							444-1,381	385				
18							444-1,381	399				
19							444-1,381	405				
20							444-1,381	421				

[a] Numbers within parentheses indicate deleted residues.
[b] ND, Not determined for technical reasons.

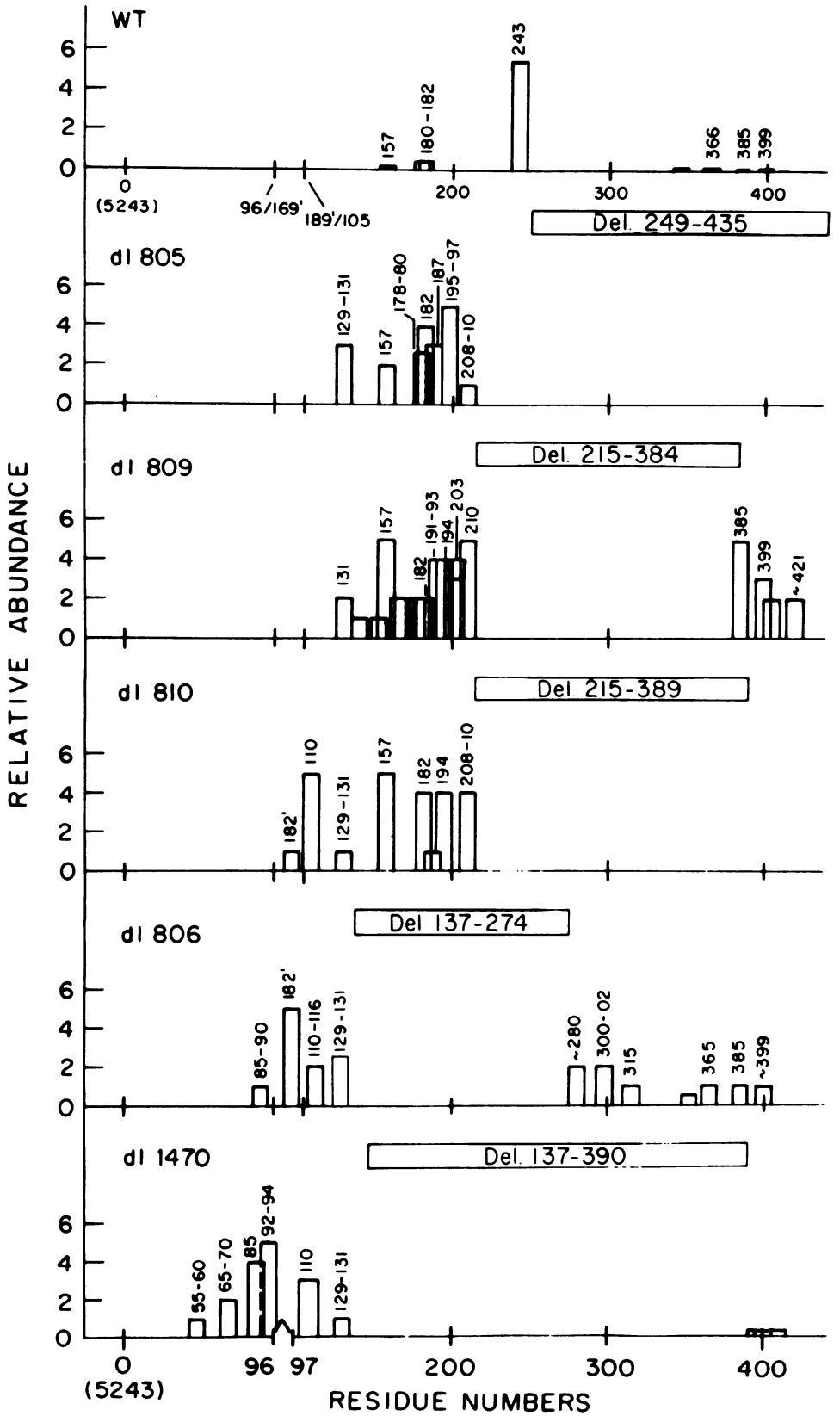

FIG. 5. Relative abundances of 5′ termini of prominent 16S RNAs. The relative abundances of individual 16S RNAs were estimated by densitometry, as described in the legend to Fig. 3, of the gels shown in Fig. 4. The 5′ termini are taken from Table 3. The graph for the wild type was obtained from an analysis of strain 776 16S RNAs. However, similar data were obtained for strain 800 16S RNAs.

stream to the region of residues 178′ to 182′ in strain 800; and (b) it is not clear whether the strain 800 19S RNA giving rise to cDNA 1 starts at residue 38, as it does in strain 776. In addition, the relative amounts of the individual species of 19S RNA in the two wild-type strains are similar. Thus, the deletion of residues 97 through 104 and the insertion of a second copy of residues 169 through 189 at this site in strain 800 does not alter late 5′ termini downstream from this site, but does alter the location of one and possibly two 5′ termini upstream.

(ii) As noted previously for wild-type SV40 (11), the 5′ termini of the late RNAs made by the mutants also map at numerous sites throughout a 300-nucleotide region of the viral genome. The mutants use even more sites than the wild-type virus as 5′ termini, with some sites being the same and others different. Similar observations have been made previously by our group (31–33) and others (18, 40).

(iii) Wild-type strain 800 and mutants dl-806, dl-809, and dl-810 use sites between residues 169′ and 189′ as 5′ termini of their late RNAs. Especially prominent are starts about residues 182′ and 187′. Thus, this sequence of 21 base pairs can be utilized in its upstream location for the formation of 5′ termini in a manner similar to its utilization in a downstream location. Only dl-805 among the mutants we have studied fails to utilize sequences from residues 169′ through 189′ as 5′ termini while making use of the identical sequences at residues 178-180, 182, and 187-189 for 5′ termination.

(iv) Only mutant dl-805 among our mutants retains residue 243, the principal 5′ terminus of wild-type 16S and 19S RNAs. However, this site is not used as a 5′ terminus in this mutant. Consequently, downstream sequences play an important role in the generation of 5′ ends.

(v) Although lacking residue 243, mutants dl-809 and dl-810 retain residue 182, the second most abundant 5′ terminus of the wild-type late RNAs. Nevertheless, this site is used minimally by these mutants even though their deletions start 33 nucleotides downstream. Similarly, dl-805, with its residue 243 site nonfunctional, uses residue 182 at a very low level. These findings also point to an important role for downstream sequences in 5′-end determination.

(vi) In all of our mutants, there is an upstream shift in the sites used as 5′ termini of the late 16S and 19S RNAs. It is uncertain whether this shift is a consequence of or is unrelated to the deletion or underutilization of residues 243 and 182 as 5′-terminal sites. Sites used frequently by the mutants as 5′ termini include residues 194-197, 164, 157, 129-131, 110, 182′, and 187′. Of these, only residues 157 and 110 are used detectably by wild-type virus. Thus, the upstream shift in 5′ termini may be to sites used infrequently by wild type or to entirely new sites.

(viii) For wild-type SV40 and for each of our mutants, the sites used as 5′ termini of the 16S RNAs and of each of the 19S RNA species show very little similarity qualitatively or quantitatively (compare Fig. 3 and 5). For example, most late wild-type RNA with a 5′ end at residue 243 is processed to 16S RNAs. On the other hand, most, if not all, wild-type RNAs with 5′ ends at residues 38, 92-96, 187, and 110 are processed to 19S RNAs with the 444-476, 212-476, and 291-476 splices and to unspliced 19S RNA, respectively. Our mutants behave in an analogous fashion. For example, in dl-805 the residue 110, 164, 200-203, and 224-225 sites are used exclusively for unspliced 19S RNAs, whereas the residue 129-131, 178-180, 182, 195-197, and 208-210 sites are used predominantly, if not exclusively, for 16S RNAs. Furthermore, even when RNAs of different viruses have the same 5′ end, they are likely to be processed differently. For example, RNA with its 5′ end at residue 110 is processed almost exclusively to unspliced 19S RNA in wild type and in dl-805, to 19S RNA with the 212-476 splice in dl-809, and to 16S RNA in dl-810. In dl-806, this same RNA gets processed to form either 16S RNA or 19S RNAs containing either the 291-476 or the 444-476 splice.

DISCUSSION

The studies presented here were designed to investigate some of the mechanisms controlling SV40 late RNA synthesis and processing. Previous studies (12, 18, 31–33, 40) and the data in this paper all indicate that the regulation of SV40 late mRNA biogenesis is complex.

Factors involved in the biogenesis of 5′ termini of late RNAs. From analyses of the incorporation of β-^{32}P-nucleotide triphosphates into cap structures, Contreras and Fiers (5) and Gidoni et al. (13) have demonstrated that at least some, and probably most, of the 5′ termini of the late SV40 RNAs arise by transcription initiation. Recently, we have found that certain 5′ termini of the late RNAs are sites of active in vitro transcription initiation (Ghosh, Weissman, and Lebowitz, unpublished data). In contrast, transcription is initiated weakly or not at all at other 5′-terminal sites. These results suggest that a factor not present in crude transcriptional extracts may be required for high-level initiation from certain sites or that certain 5′ ends sometimes arise by the processing of primary transcripts initiated upstream.

As noted, residue 243 is the most abundant 5′ terminus of the late wild-type RNAs (11, 16, 34). The fact that mutants lacking this site are viable (1, 18, 30, 37, 38) indicates that other sites can

also serve as the 5' ends of functional late mRNAs. In the mutants studied here, the absence of or failure to utilize residue 243 for 5' termination resulted almost exclusively in the utilization of numerous upstream sites as 5' termini. Many of these upstream sites (e.g., residues 194, 182, 157, and 110) also serve as 5' termini of wild-type late RNAs, albeit relatively infrequently compared with residue 243. However, in the mutants, these sites become abundant 5' termini, although they never dominate as residue 243 does in the wild type. This finding indicates that these sites are of roughly equal strength in the mutants and weaker than residue 243 for 5'-end determination. Of note is that sites downstream from residue 243 are also used as 5' termini in certain mutants that do not use residue 243 (for example, see dl-806 and dl-809 in Fig. 5), and, in some mutants, downstream 5' termini may be quite prominent (32).

The upstream progression in the location of the 5' termini of the late RNAs made by most late leader mutants lacking a functional 243 site contrasts sharply with the downstream progression of early-strand 5' termini when the principal 5'-terminal sites are deleted in cells transformed by mutant viruses (2, 9). The basis for this difference appears to lie in the structures of their promoter-regulatory regions. The control region for SV40 early-strand transcription contains a TATA box embedded in a stretch of 17 consecutive A-T base pairs. This TATA box is involved in positioning the major early transcription initiation sites 21 to 26 nucleotides downstream (2, 9). Thus, when the site used for the principal 5' termini of the early mRNAs is deleted, the major new start sites are driven by the A-T-rich sequence further downstream. In contrast, there is no A-T-rich sequence upstream from residue 243, and, hence, there is no strong measuring function moving the 5' termini downstream when residue 243 is deleted or nonfunctional.

In view of the measuring function associated with the early TATA box, we examined our data for the possible existence of a simple downstream or upstream measuring function that might be involved in placement of late-strand 5' termini. None was found. In fact, little correlation appears to exist between either the length or the location of deleted sequences and the pattern of 5' termini. For example, mutants dl-809 and dl-810, which have deletions differing by only five base pairs at their 3' ends, produce late 16S and 19S RNAs with different locations and different relative abundances of 5' termini (Fig. 3 and 5).

A third contrast with SV40 early transcription is our finding that downstream sequences play an important role in the regulation of late RNA synthesis. This effect is demonstrated clearly with mutant dl-805, in which the deletion lies entirely downstream from all but the most infrequently used wild-type 5' termini. Possible hypotheses to explain this phenomenon include the following. (i) Downstream sequences play a direct role in regulating transcription initiation in a manner analogous to that demonstrated for the transcription of low-molecular-weight RNAs by RNA polymerase III (6, 24, 36). However, if this were true, one might expect some mutants in the leader region to be nonviable or a measuring function to be operative. Instead, mutants lacking almost the entire late leader region, including all of both copies of the 72 base pair tandem repeat, (e.g., dl-1737, which lacks residues 20 through 379) synthesize late-strand RNAs in normal or nearly normal amounts when T antigen is supplied to compensate for their defect in early RNA synthesis (G. Z. Hertz, L. Trimple, and J. E. Mertz, unpublished data). (ii) Mutants with deletions in the late leader region produce RNAs, some of which have reduced stability due to the shortened leader region. However, Piatak et al. (33) have shown by chase experiments with actinomycin D that the relative abundances of different late mutant RNA species do not change significantly with time. (iii) Initiation sites may be determined by secondary structure features in DNA, e.g., denatured regions or regions that are "single-stranded." (iv) If the 5' termini of RNAs are sometimes determined by processing, then mutations in downstream sequences might alter the processing either directly or indirectly by changing the primary or secondary structures of the initial transcripts. (v) Mutations in the leader region may only indirectly affect transcription initiation as a result of feedback regulatory systems. For example, the primary problem with dl-805 may be a translational defect in some, but not all, of its late mRNAs, caused by their having mutated leader regions. Possibly, SV40 possesses feedback control systems that enable it to compensate partially for such defects by changing the relative frequencies of usage of different 5' ends. This latter hypothesis will be discussed in detail elsewhere (Barkan and Mertz, in preparation).

Regulation of splicing of late RNAs. The data presented here confirm our previous results showing that SV40 late transcription results in the production of three spliced species of 19S RNA as well as unspliced 19S RNA. Moreover, we have identified viable mutants in which (i) essentially all of the 19S RNA is unspliced (dl-805), (ii) virtually all of the 19S RNA has the 212-476 splice (dl-809), and (iii) the vast majority of the 19S RNA has the 291-476 splice (dl-806). Thus, SV40 can maintain viability despite great variability in the relative quantities of the four classes of 19S RNA. In addition, the relative

rates of synthesis of VP-2 and VP-3 are similar in cells infected with each of these mutants (Barkan and Mertz, manuscript in preparation). Therefore, each of the classes of 19S RNA may be capable of functioning in the synthesis of both VP-2 and VP-3. Consequently, the reason for the existence of four classes of 19S RNA remains obscure.

The fact that dl-805 is viable yet produces no detectable spliced 19S mRNA indicates that splicing is not essential for this mRNA species to be functional (12). Similarly, Triesman et al. (39) have recently shown that a polyoma mutant containing a gene for middle T antigen without an intron makes a functional middle T mRNA, and Gruss et al. (14) have demonstrated that an SV40 recombinant containing a rat preproinsulin gene lacking an intron makes a functional preproinsulin mRNA. In contrast, mutants of SV40 lacking precisely the intervening sequence for late 16S mRNA are unable to synthesize stable 16S mRNA (15, 18, 19; P. Berg, personal communication). Possibly, some function coded within this region, rather than the act of splicing, is essential for the accumulation of 16S mRNA.

The ratio of 16S to 19S RNAs produced by wild-type SV40 is about 5:1 to 10:1 (41). Villareal et al. (40) have reported that some late leader mutants produce late RNAs with reduced 16S-to-19S ratios. Experiments to be reported elsewhere (Barkan and Mertz, in preparation) have also shown reduced ratios of 16S to 19S RNAs in all of our 800-series mutants. The fact that dl-805 demonstrates both the greatest reduction in 16S RNA (an approximately fourfold decrease relative to 19S RNA) and the greatest concentration of unspliced 19S RNA suggests that these may be related phenomena. Furthermore, given the dependence of 16S RNA accumulation on splicing, it seems reasonable to propose that the reduction of 16S RNA levels and enhancement of unspliced 19S RNA levels are the results of a generalized late RNA splicing defect from which all of our mutants suffer. Possibly, dl-805, which deletes most of the major late leader region (residues 243 through 444), and our other mutants, which delete smaller portions of this region, lack sequences within the major late leader that are involved in splicing both late RNAs.

One hypothesis for control of late RNA splicing is that it is mediated in part by a 61-amino-acid nucleic acid-binding protein encoded by nucleotides 253 through 438 in the major late leader (21, 22; Mertz et al., in press). However, this possibility appears to be unlikely since none of the mutants studied here can make this protein although they vary widely in the relative concentrations of the various late RNA species that they synthesize. Furthermore, mutants with small deletions or alterations in the template for this protein make late RNAs similar in structure to wild-type late RNAs (33; G. Khoury, personal communication).

In addition to a generalized late splicing defect in all of our mutants, the data presented here suggest the involvement of additional specific factors in the regulation of the splicing of mutant late RNAs. For example, mutant dl-805 fails to use the donor splice site at residue 212 even though its deletion is not nearby; on the other hand, it uses the donor splice site at residue 444 for making 16S, but not 19S, RNA with the 3′ end of its deletion only nine nucleotides away. As noted above, dl-809, with a deletion starting at residue 215, makes 19S RNAs, about 80% of which carry the 212-476 splice. In contrast, the wild type produces 19S RNAs of which only 5 to 10% contain this splice, and dl-810, with a deletion also starting at residue 215 but extending five nucleotides farther in the 3′ direction, does not use this donor splice site at all. Table 4 compares the sequences around residue 212 in the RNAs of the wild type, dl-809, and dl-810. Residues 211 through 215 (AGGUA) are identical in wild-type and dl-809 RNAs and complementary to a sequence at the 5′ end of U1 RNA (3′ → 5′ UUCAU^{m}) (27). On the other hand, the sequence in dl-810 (AGGUU) is complementary to the U1 sequence in only the first four of the indicated positions. Since U1 RNA may facilitate splicing by base pairing with regions near splice sites, a reduced degree of complementarity over these five positions may play a role in the utilization of this splice in dl-809 and the wild type. However, complementarity in this region is not sufficient to ensure splicing at the 212 site, since mutant dl-805, as noted, fails to utilize this splice site. Possibly, sequences downstream from residues 211 through 215 are responsible, either directly or indirectly, for the enhanced utilization of residue 212 for splicing in dl-809.

A working model for regulation of late RNA synthesis. Any model for the control of late gene expression in SV40 not only must be consistent with the mass of confusing data presented here and elsewhere, but also must explain why even wild-type SV40 synthesizes late RNAs possessing a variety of splices and 5′ termini. We propose that the late gene region of SV40 consists of genes that overlap, not only in their coding regions, but in their promoter-regulatory regions as well. In other words, we hypothesize that the beginning of the late gene region contains two or more overlapping promoter-regulatory regions, each responsible for controlling the synthesis of one or more of the late virus-coded proteins. Assuming that the 5′ termini of the mRNAs are located where transcription begins, each promoter could control the synthesis of its own protein by the following series of events: (i)

TABLE 4. Influence of sequences around residue 212 on formation of the 212-476 splice in 19S RNA

Source of RNA	Sequence around residue 212 donor splice site	Relative abundance of 212-476 splice (%)
	211 215	
Wild type	5' CCUCAGAAGGUACCUAAC 3'	5–10
dl-810	CCUCAGAAGGUUGACGGG	<1
dl-809	CCUCAGAAGGUACUGUU	80
U1	3' GGUCCAUUCAU ^{m}A ^{m}p p pG^{m} 5'	

each promoter-regulatory region determines how frequently and precisely where transcription begins for the synthesis of its mRNAs; (ii) the precise primary and, consequently, secondary structures of the resultant initial transcripts influence the choice of splice sites; and (iii) the precise sequence of the leader region of the resultant mRNA determines which downstream translational initiation codons are used for protein synthesis.

The data presented here on the relationship between 5' termini and splices of the RNAs made by both the wild type and the mutants provide strong support for the second event. The finding that the same 5' ends are sometimes seen with several of the spliced species could be due to the 5' end of the initial transcript influencing, only in a probabilistic manner, which of several potential conformations the RNA is most likely to assume and, consequently, how it will be processed. Our model would further predict that deletion mutants such as those studied here would exhibit numerous alterations in late gene expression occurring at several levels of control. First, genomic alterations of the sizes examined here probably affect the functions of several promoter-regulatory regions, leading directly to changes in both relative abundances and locations of the 5' termini of RNAs. Second, the deletions change the primary sequences as well as the 5' ends of the initial transcripts; consequently, the preferred secondary structures and splice sites used are likely to vary. Third, the deletions, both directly and indirectly through altered processing, affect the precise sequences of the leader regions of their mRNAs; consequently, these mRNAs may exhibit altered translational efficiencies or initiation codon usage or both.

The complexities of SV40 late gene expression and the mechanisms required for its control brought out by our studies suggest the likelihood that the expression of certain other eucaryotic genes may be governed by equally complex regulatory mechanisms.

ACKNOWLEDGMENTS

We thank Tom Shenk for generously providing mutant dl-1470, Charlene Ivory, and Donna Mitchell for excellent technical assistance, and Howard Temin and Jeff Ross for critical review of this manuscript.

This work was supported by grants CA-07175, CA-16038, CA-22443, and CA-32799 from the National Cancer Institute, grant IN-31-T7 from the American Cancer Society, and an award from the Swebilius trust of the Yale Comprehensive Cancer Center to P.K.G.

LITERATURE CITED

1. **Barkan, A., and J. E. Mertz.** 1981. DNA sequence analysis of simian virus 40 mutants with deletions mapping in the leader region of the late viral mRNAs: mutants with deletions similar in size and position exhibit varied phenotypes. J. Virol. **37**:730–737.
2. **Benoist, C., and P. Chambon.** 1981. In vivo sequence requirements of the SV40 early promoter region. Nature (London) **290**:304–310.
3. **Buchman, A. R., L. Burnett, and P. Berg.** 1980. The SV40 nucleotide sequence, p. 799–829. *In* J. Tooze (ed.), DNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
4. **Canaani, D., C. Kahana, A. Mukamel, and Y. Groner.** 1979. Sequence heterogeneity at the 5' termini of late simian virus 40 19S and 16S mRNAs. Proc. Natl. Acad. Sci. U.S.A. **76**:3078–3082.
5. **Contreras, R. and W. Fiers.** 1981. Initiation of transcription by RNA polymerase II in permeable, SV40-infected and noninfected CV1 cells; evidence for multiple promoters for SV40 late transcription. Nucleic Acids Res. **9**:215–236.
6. **Engelke, D. R., S. Ng, B. S. Shastry, and R. G. Roeder.** 1980. Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. Cell **19**:717–728.
7. **Fiers, W., R. Contreras, G. Haegeman, R. Rogiers, A. van de Voorde, H. van Heuverswyn, J. van Herreweghe, G. Volckaert, and M. Ysebart.** 1978. Complete nucleotide sequence of SV40 DNA. Nature (London) **273**:113–120.
8. **Ghosh, P. K., and P. Lebowitz.** 1981. Simian virus 40 early mRNAs contain multiple 5'-termini upstream and downstream from a Hogness-Goldberg sequence; a shift in 5'-termini during the lytic cycle is mediated by large T antigen. J. Virol. **40**:224–240.
9. **Ghosh, P. K., P. Lebowitz, R. Frisque, and Y. Gluzman.** 1981. Identification of a promoter component involved in positioning the 5' termini of simian virus 40 early mRNAs. Proc. Natl. Acad. Sci. U.S.A. **78**:100–104.
10. **Ghosh, P. K., V. B. Reddy, M. Piatak, P. Lebowitz, and S. M. Weissman.** 1980. Determination of messenger RNA sequences by primer-directed synthesis and sequencing of their cDNA transcripts. Methods Enzymol. **65**:580–595.
11. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. M. Weissman.** 1978. The heterogeneity and 5'-terminal structures of the late RNAs of simian virus 40. J. Mol. Biol. **126**:813–846.
12. **Ghosh, P. K., P. Roy, A. Barkan, J. E. Mertz, S. M. Weissman, and P. Lebowitz.** 1981. Unspliced functional late 19S mRNAs containing intervening sequences are produced by a late leader mutant of SV40. Proc. Natl. Acad. Sci. U.S.A. **78**:1386–1390.
13. **Gidoni, D., C. Kahana, D. Canaani, and Y. Groner.** 1981.

Specific in vitro initiation of transcription of simian virus 40 early and late genes occurs at the various cap nucleotides including cytidine. Proc. Natl. Acad. Sci. U.S.A. **78**:2174–2178.

14. **Gruss, P., A. Efstratiadis, S. Karathanasis, M. Konig, and G. Khoury.** 1981. Synthesis of stable unspliced mRNA from an intronless simian virus 40-rat preproinsulin gene recombinant. Proc. Natl. Acad. Sci. U.S.A. **78**:6091–6095.
15. **Gruss, P. A., C. J. Lai, R. Dhar, and G. Khoury.** 1979. Splicing as a requirement for biogenesis of functional 16S mRNA of SV40. Proc. Natl. Acad. Sci. U.S.A. **76**:4317–4321.
16. **Haegeman, G., and W. Fiers.** 1978. Characterization of the 5'-terminal capped structures of late simian virus 40-specific mRNA. J. Virol. **25**:824–830.
17. **Haegeman, G., and W. Fiers.** 1980. Characterization of the 5'-terminal cap structures of early simian virus 40 mRNA. J. Virol. **35**:955–961.
18. **Haegeman, G., H. van Heuverswyn, D. Gheysen, and W. Fiers.** 1979. Heterogeneity of the 5' terminus of late mRNA induced by a viable simian virus 40 deletion mutant. J. Virol. **31**:484–493.
19. **Hamer, D. H., and P. Leder.** 1979. Expression of the chromosomal mouse globin gene cloned in SV40. Nature (London) **281**:35–39.
20. **Hamer, D. H., K. D. Smith, S. H. Boyer, and P. Leder.** 1979. SV40 recombinants carrying rabbit β-globin gene coding sequences. Cell **17**:725–735.
21. **Jackson, V., and R. Chalkley.** 1981. Use of whole cell fixation to visualize replicating and maturing simian virus 40: identification of a new viral gene product. Proc. Natl. Acad. Sci. U.S.A. **78**:6081–6085.
22. **Jay, G., S. Nomura, C. W. Anderson, and G. Khoury.** 1981. Identification of the SV40 agnogene product: a DNA binding protein. Nature (London) **291**:346–349.
23. **Kahana, C., D. Gidoni, D. Canaani, and Y. Groner.** 1981. Simian virus 40 early mRNAs in lytically infected and transformed cells contain six 5'-terminal caps. J. Virol. **37**:7–16.
24. **Kressman, A., H. Hofstetter, E. DiCapua, R. Grosschedl, and M. L. Birnstiel.** 1979. A tRNA gene of *Xenopus laevis* contains at least two sites promoting transcription. Nucleic Acids Res. **7**:1749–1763.
25. **Lai, C. J., R. Dhar, and G. Khoury.** 1978. Mapping the spliced and unspliced late lytic SV40 RNAs. Cell **14**:971–982.
26. **Lebowitz, P., and S. M. Weissman.** 1979. Organization and transcription of the simian virus 40 genome. Curr. Top. Microbiol. Immunol. **87**:43–172.
27. **Lerner, M. A., T. A. Boyle, S. M. Monnet, S. L. Wolin, and J. Steitz.** 1980. Are snRNPs involved in splicing? Nature (London) **283**:220–224.
28. **Manley, J. L., A. Fire, A. Cano, P. A. Sharp, and M. L. Gefter.** 1980. DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Natl. Acad. Sci. U.S.A. **77**:3855–3859.
29. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol **65**:499–580.
30. **Mertz, J. E., and P. Berg.** 1974. Viable deletion mutants of simian virus 40: selective isolation by means of a restriction endonuclease from *Hemophilus parainfluenzae*. Proc. Natl. Acad. Sci. U.S.A. **71**:4897–4883.
31. **Piatak, M., P. K. Ghosh, P. Lebowitz, V. B. Reddy, and S. M. Weissman.** 1979. Complex structures and new surprises in SV40 mRNA. ICN-UCLA Symp. Mol. Cell. Biol. **15**:199–216.
32. **Piatak, M., M. Olive, K. N. Subramanian, P. K. Ghosh, V. B. Reddy, P. Lebowitz, and S. M. Weissman.** 1980. The structure and expression of the late region of SV40 DNA, p. 151–173. *In* D. H. Dean, L. F. Johnson, P. C. Kimball, and P. S. Perlman (ed.), Analysis of gene structure and expression. Ohio State University Press, Columbus.
33. **Piatak, M., K. N. Subramanian, P. Roy, and S. M. Weissman.** 1981. Late mRNA production by viable SV40 mutants with deletions in the leader region. J. Mol Biol. **153**:589–618.
34. **Reddy, V. B., P. K. Ghosh, P. Lebowitz, and S. M. Weissman.** 1978. Gaps and duplicated sequences in the leaders of SV40 16S RNA. Nucleic Acids Res. **5**:4195–4214.
35. **Reddy, V. B., B. Thimmappaya, R. Dhar, K. N. Subramanian, B. S. Zain, J. Pan, P. K. Ghosh, M. Celma, and S. M. Weissman.** 1978. The genome of SV40. Science **200**:494–502.
36. **Sakonju, S., D. F. Bognehagen, and D. D. Brown.** 1980. A control region in the center of the 5S RNA genes directs specific initiation of transcription. I. The 5'-border of this region. Cell **19**:13–25.
37. **Shenk, T. E., J. Carbon, and P. Berg.** 1976. Construction and analysis of viable deletion mutants of simian virus 40. J. Virol. **18**:664–671.
38. **Subramanian, K. N.** 1979. Segments of simian virus 40 DNA spanning most of the leader sequence of the major late viral messenger RNA are dispensable. Proc. Natl. Acad. Sci. U.S.A. **76**:2556–2560.
39. **Treisman, R., U. Novak, S. Favaloro, and R. Kamen.** 1981. Transformation of rat cells by an altered polyoma virus genome expressing only the middle-T protein. Nature (London) **292**:595–600.
40. **Villarreal, L. P., R. T. White, and P. Berg.** 1979. Mutational alterations within the simian virus 40 leader segment generate altered 16S and 19S mRNAs. J. Virol. **29**:209–219.
41. **Weinberg, R. A., Z. Ben-Ishai, and J. E. Newbold.** 1978. Simian virus 40 transcription in productively infected and transformed cells. J. Virol. **13**:1263–1273.