JOURNAL OF VIROLOGY, Dec. 1985, p. 1002–1013
0022-538X/85/121002-12$02.00/0
Copyright © 1985, American Society for Microbiology
Vol. 56, No. 3

Sequences Involved in Determining the Locations of the 5′ Ends of the Late RNAs of Simian Virus 40

MADHU B. SOMASEKHAR† AND JANET E. MERTZ*

McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, Wisconsin 53706

Received 25 March 1985/Accepted 23 August 1985

The 5′ ends of the simian virus 40 (SV40) late RNAs are heterogeneous in location, spanning a 300-nucleotide region from residues 28 to 325. To examine whether upstream or downstream measuring functions analogous to the TATA box play roles in positioning the 5′ ends of these RNAs, we determined by S1 and primer extension mapping the locations of the 5′ ends of the late viral RNAs made in monkey cells infected with: (i) three wild-type strains of SV40 that contain tandem duplications of the enhancer region that are 64, 85, and 91, rather than 72, base pairs in length; (ii) four viable mutants that contain alterations in the 21-base-pair tandem repeats; and (iii) four viable mutants that possess small deletions or insertions at or near the major cap site at residue 325. Most of the 5′ ends of the RNAs were identical in location to those seen with wild-type strain 776. The only exceptions were the absence of RNAs whose 5′ ends mapped to within three bases upstream or downstream of a sequence alteration. In addition, the sequences within residues 251 to 277 that function as transcriptional initiation sites in wild-type strain 776 also did so in their second locations in the wild-type strains in which these sequences are duplicated. Differences were noted in the relative abundances of the numerous 5′ ends of the late RNAs, even among the wild-type strains. These findings indicate that many (and likely all) of the approximately two dozen locations of 5′ ends of SV40 late RNAs are each determined largely by sequences within their immediate vicinity. However, sequences somewhat removed from these transcriptional initiation sites may modulate the efficiencies with which they are utilized.

Most eucaryotic genes transcribed by RNA polymerase II contain a TATA box that directs transcription to initiate predominantly at approximately 30 nucleotides downstream from this sequence (see references 6 and 36 for reviews). For example, in the early region of simian virus 40 (SV40), the sequence TATTTAT is situated approximately 21 to 28 bases upstream from the microheterogeneous 5′ termini of the early RNAs (31). Deletion of this AT-rich sequence results in greater heterogeneity in the locations of the 5′ termini but does not affect the amount of RNA made from this promoter (12). Alterations in the distance between the TATA box and the sequences at which transcription normally initiates cause shifts in the locations of the 5′ termini of the RNAs that correspond approximately to the sizes of the alterations (14, 23). Therefore, the TATA box contains a measuring function that positions the site of transcriptional initiation by RNA polymerase II.

Some eucaryotic message-encoding genes do not possess a recognizable TATA box. The late-region promoter of SV40, lacking a TATA box, directs the synthesis of numerous mRNAs whose 5′ termini span a 300-nucleotide region (17, 18). From analysis of the incorporation of β-^{32}P-labeled nucleotide triphosphates into cap structures, Contreras and Fiers (8) and Gidoni et al. (19) have demonstrated that at least some of the 5′ termini of the late SV40 RNAs arise by transcriptional initiation. Except for the studies of Brady et al. (5), which indicated that a sequence approximately 30 nucleotides upstream of the most predominant 5′ terminus of the SV40 late RNAs may act in a manner analogous to a TATA box when mutated, little was known about the signals that determine the locations of the 5′ termini of the late SV40 RNAs when the studies reported here were performed.

* Corresponding author.
† Present address: Department of Biochemistry, University of Wisconsin, Madison, WI 53706.

Another sequence that has been postulated to play a role in determining the site and efficiency of transcriptional initiation is the CAAT box frequently found approximately 80 base pairs upstream from transcriptional initiation sites (20). The SV40 early region contains two putative CAAT boxes (10), both CTAAC, situated approximately 80 and 101 nucleotides upstream from the first 5′ termini of the microheterogeneous early mRNAs. Deletion of these sequences decreases the expression of the SV40 early genes (11). The late region of SV40 contains the sequence CTAAC at residues 251 to 255 and 299 to 303 situated approximately 80 nucleotides upstream of the late RNAs whose 5′ ends map to residues 325 and 392, respectively. It has been postulated that a CAAT sequence, even in the absence of a TATA box, may be involved in directing RNA polymerase II to initiate transcription at a precise distance downstream from it (4).

To test the above hypothesis and to determine whether any other sequences within the SV40 late promoter region direct transcription to initiate at a precise distance from them, we determined the locations of the 5′ ends of the viral RNAs made in monkey cells infected with three wild-type (WT) and eight mutant strains of SV40 that bear fairly small deletions or insertions in various parts of this region of the SV40 genome. Conclusions drawn from this study include the following. (i) Sequences situated within a few base pairs of many (and likely all) of the approximately two dozen transcriptional initiation sites that map within the SV40 late promoter, rather than upstream or downstream measuring functions, play a primary role in determining the precise locations of these sites. (ii) The predominant 5′ ends of the viral RNAs made in cells infected with mutants lacking sequences around residue 325, the major initiation site, map to initiation sites located within 25 base pairs upstream of the deletion. (iii) Although deletions, insertions, and duplications of sequences within the promoter can affect dramatically the relative efficiencies of utilization of the numerous

TABLE 1. Structures relative to WT strain 776 of mutants used in this study

Mutant	Size change (base pairs)	Nucleotide residues deleted (inserted)[a]	Reference
WT Chambon	−8	174–181[b]	3
WT 800	−8 + 21 = +13	179–186 (251′–271′)[b]	15
WT Oxman	−8 + 27 = +19	179–186 (251′–277′)[b]	32
*dl*1746	−71	260–330[c]	This paper
*dl*1747	−60	271–330[c]	This paper
*dl*1748	−59	281–339[c]	This paper
S+8	+10	—[d]	11
XS1	−19 + 8 = −11	35–53[e]	11
XS2	−60 + 8 = −52	35–94[e]	11
XS7	−35 + 8 = −27	73–107[e]	11
XS14	−49 + 8 = −41	59–107[e]	11

[a] The nucleotide numbering system is that of Buchman et al. (7).
[b] Except for differences in the sizes and precise endpoints of their "72" base-pair tandem repeat (Fig. 1), the sequences of the regulatory regions of all four WT strains are identical.
[c] These mutants were derived from WT 800; consequently, they also contain a substitution of nucleotide residues 251′ to 271′ in place of 179 to 186.
[d] —, An insertion of GGTCGACCCG is present between nucleotide residues 348 and 349.
[e] These mutants each contain an insertion of GGTCGAGG at the site of deletion.

initiation sites, such alterations do not result in transcriptional initiations at sites not also utilized in WT-infected cells.

(An account of this work was presented at the DNA Tumour Virus Meeting, Cambridge, England, 1 to 5 August, 1983.)

MATERIALS AND METHODS

Cell lines. CV-1P and MA-134 are established lines of African green monkey kidney cells. They were grown as described previously (25) in medium supplemented with 5% fetal bovine serum.

Viruses and viral DNAs. The structures of the various WT and mutant strains of SV40 used in this study are summarized in Table 1. WT strain 776 and WT Oxman (32) were obtained from S. Weissman. WT 800 is a plaque-purified derivative of SV40 strain Rh911 (25). WT Chambon was reconstructed from pSV-1, a recombinant plasmid that contains the larger *Bam*HI-*Hpa*II fragment of SV40 cloned into pBR322 (3), by substitution of the SV40 *Bcl*I-*Kpn*I fragment of pSV-1 grown in a *dam*⁻ strain of *Escherichia coli* for the corresponding region of WT 800.

SV40 mutants deleted in and around the major late transcriptional initiation site at residue 325 were constructed by in vitro mutagenesis. Briefly, WT 800 DNA, covalently linked to pBR322 DNA via their unique *Bam*HI sites, was linearized by cleavage with restriction endonuclease *Kpn*I, treated briefly with BAL 31 nuclease to remove 10 to 50 base pairs from each end, incubated with T4 DNA polymerase in the presence of all four dNTPs to generate blunt ends, and recyclized with T4 DNA ligase. The resulting mutant viral genomes were then propagated in *E. coli* HB101 cells by standard recombinant DNA techniques. The nucleotide sequence of each mutant was determined essentially as described by Maxam and Gilbert (24). Three mutants, pSV1746, pSV1747, and pSV1748, were selected for the present study.

The SV40 mutants pS+8, pXS1, pXS2, pXS7, and pXS14 were a generous gift of M. Fromm and P. Berg. Their isolation and many of their biological properties have already been described by them (11). Their parental strain, WT 830 (26), is identical in nucleotide sequence to WT strain 776 throughout the late promoter region.

To obtain virus stocks, we excised mutant viral genomes from recombinant DNAs by treatment with the appropriate restriction endonuclease (*Bam*HI or *Eco*RI), recyclized them by incubation with T4 DNA ligase, and transfected them onto freshly confluent monolayers of CV-1P cells as described previously (25). Individual plaques were picked from which virus stocks were then prepared, and their titers were determined as described previously (25). Preparations of viral DNA were prepared from virus-infected MA-134 cells as described previously (25).

Purification of cytoplasmic RNA from virus-infected cells. Freshly confluent monolayers of CV-1P cells were infected with virus at a multiplicity of infection of approximately 20 PFU per cell. After incubation at 37°C for 42 to 46 h, the cells were harvested with a rubber policeman into Tris-buffered saline lacking calcium and magnesium (22), washed, and incubated at 0°C in 0.5% Nonidet P-40–25 mM Tris hydrochloride (pH 7.6)–40 mM KCl–7.5 mM $MgCl_2$. After centrifugation at 2,000 rpm for 5 min to pellet the nuclei, the cytoplasm was made 1.5 M urea–3.0% sodium dodecyl sulfate–5 mM EDTA, extracted twice with phenol-chloroform-isoamyl alcohol (25:25:1) and twice with chloroform-isoamyl alcohol (25:1), and then pelleted through 5.8 M CsCl as described by Ross (34). The RNA in the pellet, essentially free of DNA, was suspended in 10 mM Tris (pH 7.4)–1 mM EDTA and stored at −20°C in 70% ethanol.

Analysis of the 5′ ends of late-strand cytoplasmic RNAs by S1 mapping. Viral DNA of the mutant from which viral RNA was synthesized was cleaved at its unique *Hpa*II site (residue 348 [Fig. 1]), treated with alkaline phosphatase, 5′ end labeled with polynucleotide kinase and [γ-^{32}P]ATP (24), and then cleaved at its unique *Bgl*I site (residue 5239). The strand complementary to the 5′ ends of the viral late RNAs was purified by denaturation in 30% dimethyl sulfoxide followed by electrophoresis in a 7.5% polyacrylamide gel (24). The cytoplasmic viral RNA to be analyzed and a molar excess of an appropriate 5′ end-labeled, single-stranded DNA probe were coprecipitated with ethanol and suspended in 20 μl of 80% deionized formamide–40 mM piperazine-*N*-*N*′-bis(2-ethanesulfonic acid) (pH 6.4)–0.4 M NaCl–1 mM EDTA. After incubation at 45°C for 5 h, the unhybridized DNA was degraded by incubation at 37°C for 1 h in 200 μl of S1 buffer containing 250 mM NaCl, 20 mM Na_2OAc (pH 5.5), 1 mM $ZnSO_4$, and an excess of S1 nuclease. The remaining DNA was precipitated with ethanol and analyzed by electrophoresis in a 0.3-mm-thick, 8% polyacrylamide gel containing 7 M urea (35). Size markers consisted of a sequencing ladder of the probe used for S1 mapping. To determine the relative abundances and precise locations of the 5′ ends mapping farthest upstream and to detect RNA species that had been processed with the splice site at residue 294 (18), we also performed S1 mappings with probes 5′ end labeled at the *Kpn*I (residue 294) and *Pvu*II (residue 272) sites (Fig. 1). S1 nuclease mapping of the 5′ ends of late cytoplasmic and nuclear RNAs gave essentially identical results.

Analysis of cytoplasmic RNA structures by the primer extension method. To distinguish true 5′ ends from discontinuities in the RNA sequence caused by RNA splicing and other artifacts of mapping with S1 nuclease, we verified all results obtained by S1 mapping by primer extension analysis (16). In brief, DNAs complementary to the 5′ ends of the viral RNAs (cDNAs) were synthesized with

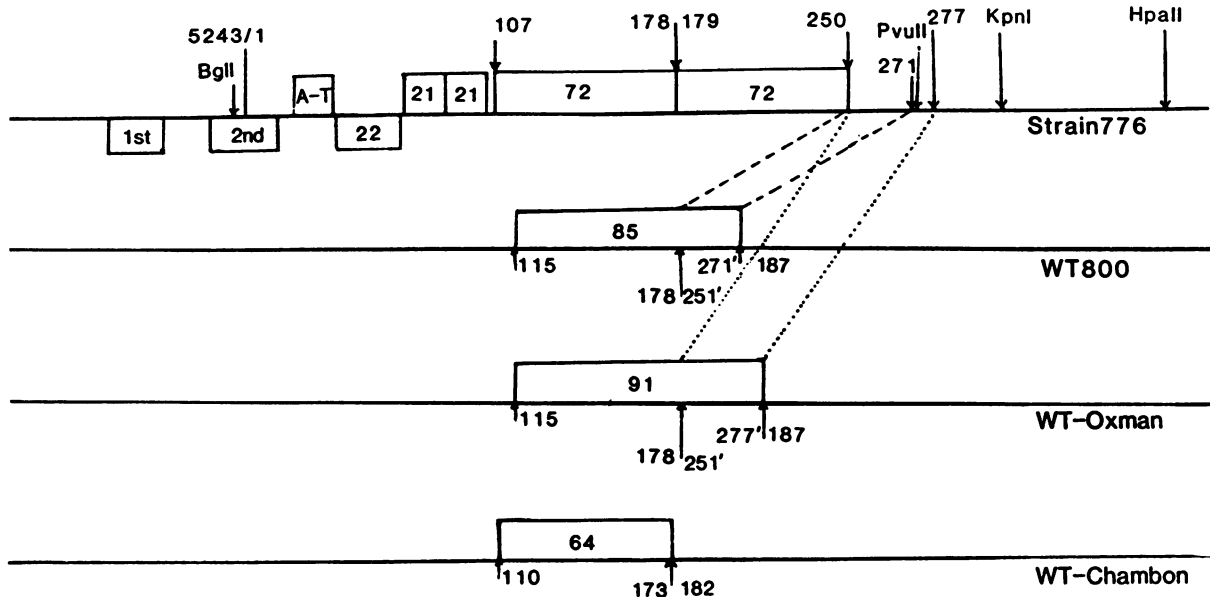


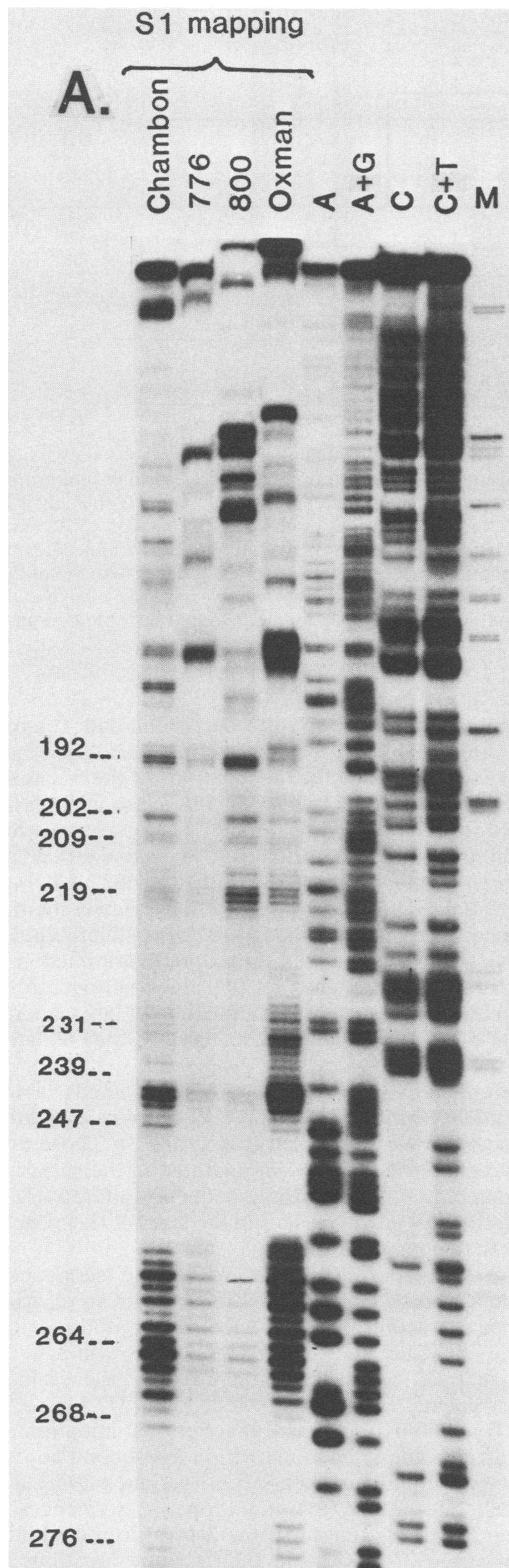
FIG. 1. Maps of the promoter-regulatory regions of four WT strains of SV40. The top line shows in detail the locations in WT strain 776 of a variety of regulatory elements: the first and second T-antigen-binding sites (37); the 17-base-pair A-T region which includes a TATA box directing initiation of early strand transcription (14, 23); the three copies of the 21-base-pair tandem repeats that function as an important element of both the early and late promoters (11); and the "72"-base-pair tandem repeats that encode enhancer elements (27). The sites of cleavage of the restriction endonucleases used in this study are also indicated. The rectangles on the three lower lines indicate the sequences within the enhancer region that are duplicated in tandem in the other three WT strains of SV40 employed in this study. The primed nucleotide residues indicated in WT 800 and WT Oxman are identical in sequence to the corresponding unprimed nucleotide residues in WT strain 776.

avian myeloblastosis virus reverse transcriptase (10 U/30 μl reaction volume; Life Sciences, Inc.) at 45°C for 0.75 h in 50 mM Tris hydrochloride (pH 8.3)–8 mM MgOAc–60 mM NaCl–10 mM dithiothreitol–50 U of RNasin (Promega-Biotec)–1 mM each dATP, dTTP, dCTP, and dGTP. The RNA was then hydrolyzed by addition of EDTA to 10 mM and NaOH to 0.2 N, followed by incubation at 45°C for 1 h. After neutralization of the reaction with HCl, the resulting cDNAs were extracted with phenol-chloroform, precipitated with ethanol, and sized by electrophoresis in 0.3-mm-thick, 8% polyacrylamide gels containing 7 M urea.

The primers, present in molar excess and hybridized with the RNAs before addition of reverse transcriptase, were short, 5′ end-labeled, strand-separated restriction fragments of SV40 that extended from nucleotide residues 407 to 349, 348 to 295, or 348 to 273, inclusive. To determine the locations of RNA 5′ ends that mapped downstream of the mutations in *dl*1746, *dl*1747, *dl*1748, and S+8, we used primers extending from nucleotide residues 661 to 593 and 1497 to 1463 plus 526 to 522 inclusive (i.e., made from a cDNA clone of SV40 late 16S mRNA [2]) to analyze late viral 19S and 16S RNAs, respectively. The markers used to determine the sizes of the cDNAs were sequence ladders of large restriction fragments that had been 5′ end labeled at the same nucleotide residues as the primers. *Msp*I-digested, 5′ end-labeled pBR322 DNA was also used occasionally as size markers.

Quantitation of data. The nucleotide residue numbers indicated in Table 2 and Fig. 3, 6, and 7 for the locations of the 5′ ends of the RNAs were taken from the primer extension data. They agree well with the locations of 5′ ends reported previously by others (8, 9, 13, 17, 18). We consider their placement to be in error by at most plus or minus two nucleotide residues. This error resulted from the technical uncertainty as to whether reverse transcriptase always polymerizes precisely to the first base in the RNA, combined with the markers from sequence ladders being displaced one-half nucleotide from cDNAs that are identical in sequence. Fortunately, neither of these technical difficulties exists when the patterns of cDNAs made with the same primer from different RNA samples are compared side by side on high-resolution gels. Therefore, the techniques employed here enabled us to determine to the base whether or not the 5′ ends of RNAs from different samples were qualitatively identical even though the precise locations we assigned to them may be slightly in error. In addition, although all of the locations of 5′ ends of RNAs are indicated by single residue numbers, some (e.g., 264, which includes 5′ ends mapping to residues 260 and 262 as well) are microheterogeneous (Fig. 4 and 5 and reference 18). Except for 175, the 5′ ends indicated in the diagrams were also observed in similar locations by the S1 mapping technique, although with less accuracy because of the inability of S1 nuclease to degrade consistently down to the last unpaired nucleotide. Our failure to detect a 5′ end at residue 175 by S1 mapping (Fig. 2B) was probably due to the fact that this RNA species lacks residues 295 to 557, inclusively (15).

The relative amount of each species of RNA was determined by densitometry of autoradiograms of high-resolution gels in which the cDNAs were sized. Because the SV40 late RNAs exhibit heterogeneity within the leader region as to how they are spliced (18), no one primer can detect all of them. Therefore, to estimate the relative amount of each RNA species, we used the data obtained with each of the three primers in aggregate with normalizations made from RNA species that were detectable with more than one of the primers. To rule out the possibility that incomplete reverse transcription might have occurred, resulting in underestimation of the relative abundances of the more upstream 5′ ends, we also determined the abundances of the 5′ ends mapping



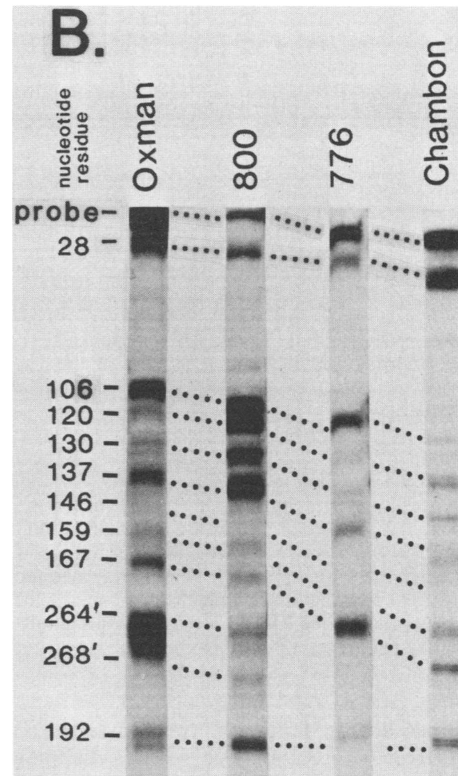


FIG. 2. (A) S1 nuclease mapping of the 5′ ends of late cytoplasmic RNAs made from four WT strains of SV40. CV-1P cells were infected with each of the four WT strains described in Table 1. At 42 h after infection, the cells were harvested and the RNAs were purified and analyzed as described in Materials and Methods. The probes used in the S1 mappings shown here had been 5′ end labeled at nucleotide residue 348; in each case, they had been prepared with viral DNA of the strain from which the viral RNAs had been made. The lanes marked A, A+G, C, and C+T contained partial chemical degradation products (24) of the 5′ end-labeled probe made from strain WT Chambon. Lane M contained *Msp*I-digested, 5′ end-labeled pBR322 DNA. The numbers at the left indicate the locations within the DNA sequencing reactions of selected nucleotide residues that correspond to 5′ ends of some of the RNAs. In most cases, each 5′ end, although specific in location, appears as 2 to 4 consecutive bands displaced one-half nucleotide from the corresponding nucleotide residue in the DNA sequencing reactions. However, a few (e.g., 264, which includes 5′ ends at 260 and 262 as well) are truly microheterogeneous, as shown by primer extension analyses (Fig. 4). The bottom of the gel, including the major 5′ end at residue 325, was omitted from the figure so the minor upstream 5′ ends could be shown clearly. (The low intensities of the bands seen in the lower half of the gel in the lanes marked 776 and 800 were caused by a failure in this particular experiment to recover quantitatively during precipitation and washing with ethanol the smallest of the S1-protected fragments that were obtained with these two samples.) (B) Reproduction of the upper portion of A with the columns separated and the bands labeled to indicate the locations of the 5′ ends of the RNAs made from each WT strain.

upstream of residue 192 relative to those mapping to residues 192, 202, 209, and 219 by the S1 nuclease technique. For any one preparation of RNA, the data obtained with these different methods and probes varied from, at worst, twofold for the least abundant 5′ ends to less than 10% for the 5′ ends which accounted for greater than 5% of the total. However, by far the largest error in quantitation was biological in nature, with variations of as much as threefold being observed frequently with preparations of viral RNA obtained on different days from cells infected with the same mutant or WT strain (Table 2).

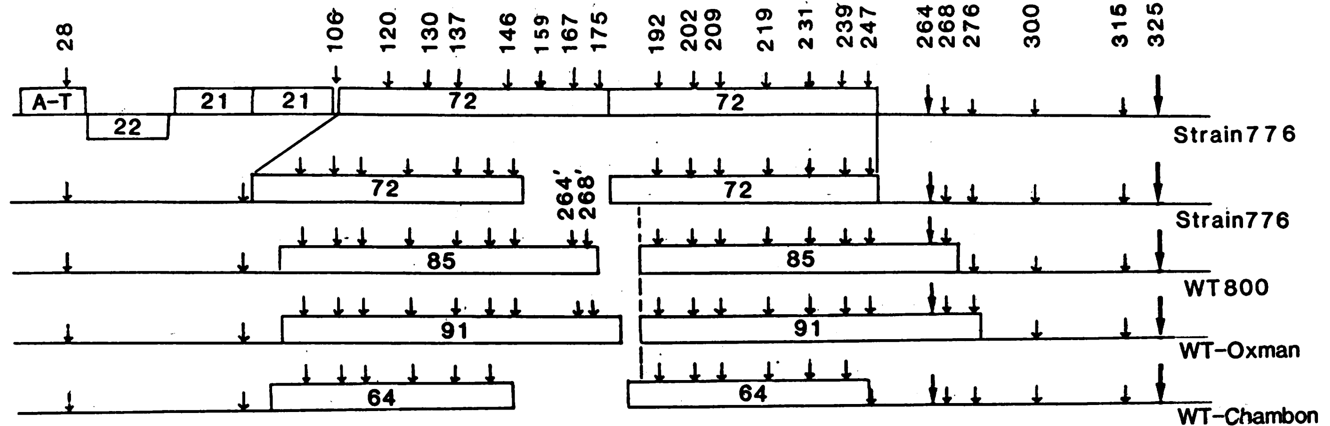


FIG. 3. Summary of the locations of the 5′ ends of the late viral RNAs made from four WT strains of SV40. The results shown here were determined from the complete analysis by both the S1 nuclease and primer extension methods with several different probes and primers (see Materials and Methods for details), respectively, of RNA samples obtained from three or more independent infections with each WT strain of SV40. The arrows situated at numerous sites along each horizontal line indicate the map locations at which the 5′ ends of RNAs made from that WT strain of SV40 were found. The sizes of the arrows indicate the approximate percentages of the late viral RNA molecules that had 5′ ends mapping to these locations, with large, intermediate, and small arrows denoting greater than 50, 7 to 15, and less than 5% of the total, respectively. The tandem pairs of rectangles indicate the locations and extents of the duplicated enhancer regions diagrammed in detail in Fig. 1; the two copies present in each WT strain are contiguous on the viral genome but have been separated here so that the common nucleotide sequences of the strains line up with each other.

Last, it should be noted that a few percent of the SV40 late RNAs made from WT SV40 have 5′ ends mapping downstream of nucleotide residue 325 (18). For all of the RNAs analyzed here, except those made from mutants *dl*1746, *dl*1747, *dl*1748, and S+8, we ignored the presence of these minor species.

RESULTS

Promoter-regulatory regions of four WT strains of SV40. The promoter-regulatory regions of four WT strains of SV40 are shown diagrammatically in Fig. 1 and 3. The promoter-regulatory regions of these four WT strains of SV40 are identical in sequence except for the sizes of their "72" base-pair tandem repeats (i.e., enhancer regions). Strain 776, the prototypical WT strain of SV40, has the 72 base pairs from residues 107 to 178 repeated in tandem at residues 179 to 250 (7). Strains 800 and Oxman contain tandem repeats of 85 and 91 base pairs, respectively. In these two WT strains, residues 179 to 186 of strain 776, duplicated from residues 107 to 114, are missing, and residues 251 to 271 and 251 to 277, respectively, (designated here as 251′ to 271′ and 251′ to 277′) are duplicated in their stead. Thus, strains 800 and Oxman have net insertions relative to strain 776 of 13 and 19 base pairs, respectively, at the junction of the 72-base-pair tandem repeats (Table 1). Strain Chambon, missing residues 174 to 181 of strain 776, and therefore having a 64-base-pair tandem repeat, lacks 8 base pairs relative to strain 776. In summary, although three of these WT strains differ substantially in sequence from the prototypical WT strain of SV40, their differences occur only in the precise sequences they duplicate and, therefore, they are unlikely to possess any *cis*- or *trans*-acting defects.

5′ Termini of late RNAs synthesized from the four WT strains. The locations and relative abundances of 5′ termini of the late RNAs synthesized in monkey cells infected with each of the four WT strains described above were determined by the S1 and primer extension procedures used in combination as described in Materials and Methods. Figure 2 shows an autoradiogram of one of the numerous S1 mapping gels that were used to derive the data summarized in Fig. 3 and Table 2.

The results obtained for the locations of the 5′ ends of the late RNAs made from WT strain 776 agree well with those reported previously by others (13, 18), except that, in addition to those noted previously, we also identified 5′ ends mapping to nucleotide residues 219, 209, 202, 167, 146, 137, 130, and 106 (Fig. 3). Since each of these newly identified 5′ ends accounted for less than 2% of the total late viral RNA, it is easy to understand why they were overlooked in earlier studies. Furthermore, the fact that they had been observed previously in RNAs made from a variety of late leader region mutants of SV40 (15) also indicates that their existence is real.

Data concerning the locations of the 5′ ends of the late RNAs made from the other three WT strains indicated that their locations were identical (Fig. 2 and 3). Those mapping downstream of residue 186 co-migrated in the gel (Fig. 2A); those mapping upstream of the sequence alteration differed in migration by the differences in the sizes of their duplicated enhancer regions (Fig. 2B).

These data also show that nucleotide sequences that function as transcriptional initiation sites do so regardless of their location within the 300-nucleotide promoter region. For example, all of the nucleotide sequences to which 5′ ends of RNAs mapped within the first copy of the 72-base-pair tandem repeats (i.e., residues 120, 130, 137, 146, 159, 167, and 175) also functioned as transcriptional initiation sites at their corresponding locations within the second copy of the 72-base-pairs tandem repeats (i.e., residues 192, 202, 209, 219, 231, 239, and 247). In addition, the sequences within residues 251 to 277 that function as transcriptional initiation sites in strain 776 did so in their second locations in WT strains 800 and Oxman as well. The only exception to this rule was lack of RNA made from WT Oxman whose 5′ end mapped to residue 276′ (Table 2). However, this exception is likely due to either (i) the fact that the nucleotide sequence differs within two base pairs downstream of the unutilized site from that found downstream of the residue 276 transcrip-

TABLE 2. Effect of nearby sequence alterations and position on utilization of specific transcriptional initiation sites

Mutant or WT strain	Sequence surrounding initiation sites[a]	% Viral late RNAs with 5′ ends mapping to the site indicated[b]			
WT 800	264 268 276 A C A C A T T C C A C A G C T G G T T C T T T C C G	10-15;	0.6-1.0;	0.2-0.5	
(WT-Oxman)	"	(10-14;	0.3-0.6;	0.3-0.7)	
dl-1748	264 268 276 A C A C A T T C C A C A G C T G G T T C T/G C T G C	50-60;	5-8;	1-2	
dl-1747	264 268 A C A C A T T C C A C/G G C C A T G G T G C T G C G	50-60;	5-8		
WT 800	264′ 268′ A C A C A T T C C A C A G/A C T A A T T G A G A T G	0.4-0.9;	0.5-1.0		
WT-Oxman	264′ 268′ 276′ 192 A C A C A T T C C A C A G C T G G T/G A C T A A T T	3-5;	0.2-0.6;	ND[c];	0.2-0.6
WT 800	239 247 264 A C T T T C C A C A C C C T A A C T G A C A C A C A	0.6-1.0;	0.1-0.3;	10-15	
dl-1746	239 247 A C T T T C C A C A C C C T A A C T G A C/G G C C A	45-50;	10-15		
Strain 776	167 175 192 A C T T T C C A C A C C T G G T T G C T G A C T A A	3-5;	0.3-0.7;	0.3-0.8	
Strain 776	239 247 264 A C T T T C C A C A C C/C T A A C T G A C A C A C A	0.2-0.6;	0.3-0.7;	11-16	
WT-Chambon	167 192 A C T T T C C/T T G C T G A C T A A T T G A G A T G	0.2-0.6;		0.5-1.5	
WT 830	106 120 G G C G G G A C T A T G G T T G C T G A C T A A T T	3-6;			0.3-1.0
(Strain 776)	"	(3-6;			0.2-0.6)
XS7	106* 120 G T T G G T C G A G G/G G T T G C T G A C T A A T T	ND;			0.3-1.0
XS14	106* 120 C G G G G T C G A G G/G G T T G C T G A C T A A T T	ND;			0.5-1.5
Strain 776	175 192 T T T C C A C A C C/T G G T T G C T G A C T A A T T	0.3-0.7;			0.3-0.8
WT-Chambon	167 192 C T G G G G A C T T T C C/T T G C T G A C T A A T T	0.2-0.6;			0.5-1.5
WT 800	264′ 268′ 192 A C T G A C A C A C A T T C C A C A/G A C T A A T T	0.4-0.9;	0.5-1.0;		0.2-0.6
WT-Oxman	264′ 268′ 276′ 192 A C A C A T T C C A C A G C T G G T/G A C T A A T T	3-5;	0.2-0.6;	ND;	0.2-0.6
WT 830	28 A A A A A T T A G T C A G C C A T G G G G C G G A G	1-2			
XS1	28 A A A A A T T A G T C/G G T C G A G G G G C G G A A	3-5			
XS2	28 A A A A A T T A G T C/G G T C G A G G G G G G C G G	3-5			

[a] Numbers above specific letters indicate, in nucleotide residue numbers, sites of transcriptional initiation determined as described in Materials and Methods; they are accurate to within ±2. In addition, initiation at some of these sites exhibits microheterogeneity.

[b] Each range of numbers indicates the range of percentages obtained from analyses, as described in Materials and Methods, of 3 to 7 preparations of viral RNA obtained on different days from cells infected with the same mutant or WT strain of SV40. For sequences in which more than one initiation site is indicated, our findings concerning the relative utilization of these sites are presented in the same order as they appear in the sequences.

[c] ND, Not detected in experiments in which 5′ ends accounting for 0.1% of the total would have been seen readily.

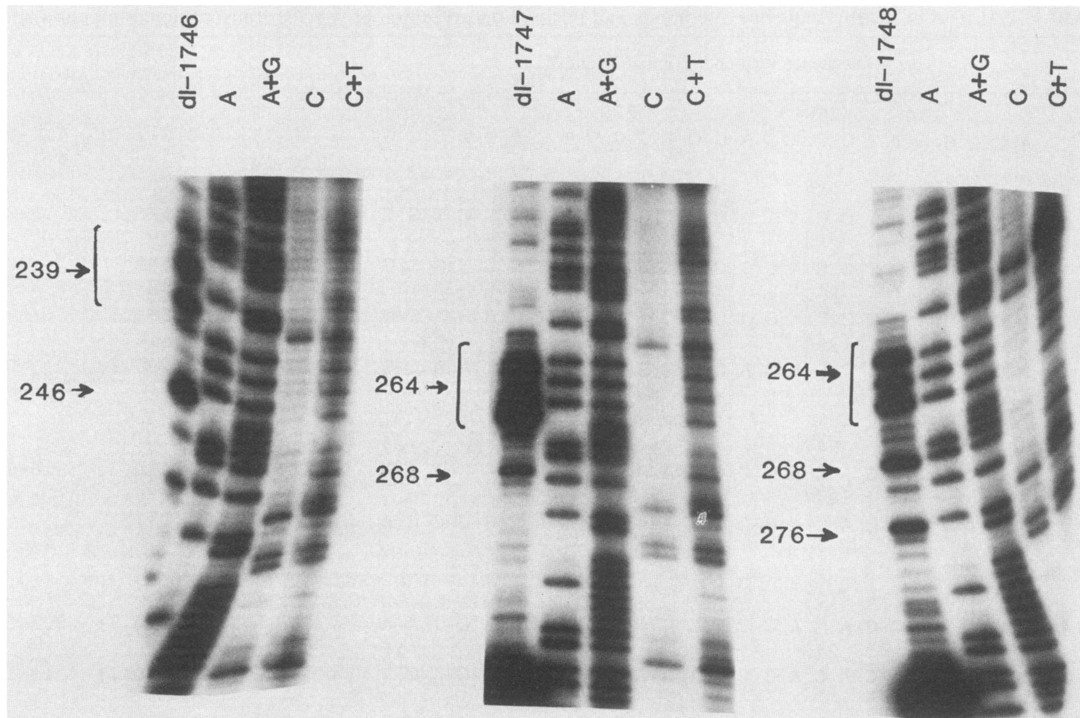


FIG. 4. High-resolution gels showing the cDNAs synthesized from the RNAs made from mutants *dl*1747, *dl*1747, and *dl*1748 whose 3′ ends map near the endpoints of their deletions. The cDNAs were synthesized, as described in Materials and Methods, with a single-stranded, 5′ end-labeled primer that extended from nucleotide residues 407 to 349, inclusive. The DNA sequencing reactions were those of the SV40 mutant *dl*1741, isolated in our laboratory and lacking nucleotide residues 151 to 291 inclusive (M. B. Somasekhar and J. E. Mertz, unpublished data), that had been 5′ end labeled at residue 348.

tional initiation site (Table 2) or (ii) the possibility that there actually were RNAs with 5′ ends mapping to residue 276, but their relative abundance was so minor that we failed to detect them. Taken together, these data indicate that many (and likely all) sites within the late SV40 promoter region that function as transcriptional initiation signals are determined, at least in part, by the specific nucleotide sequences situated within the immediate vicinity of each site.

Experiments such as those presented in Fig. 2 and 5 also provided information concerning the relative abundances of the numerous 5′ ends of the late RNAs made from the various WT strains (see Fig. 3 and Table 2 for summaries). For example, RNAs with 5′ ends mapping to residues 325 and 264 accounted for approximately 60 to 65 and 10 to 15%, respectively, of the total RNA made from each of the WT strains. On the other hand, RNAs with 5′ ends mapping to residues 28, 106, 120, 130, 137, 146, 159, 167, 175, 192, 202, 209, 219, 231, 239, and 247 accounted in each case for less than 5% of the total late viral RNA. Also noteworthy is the finding that, whereas 3 to 5% of the late viral RNA molecules present in cells infected with WT Oxman were reproducibly found to have 5′ ends mapping to residue 264′, less than 1% of the WT 800 RNA molecules had 5′ ends mapping to this residue (Table 2).

These experiments also yielded some information as to the number of nucleotides within the immediate vicinity of the initiation site that make up the initiation signal. For example, with only five nucleotides upstream of residue 192 being common to all four WT strains, this amount of sequence appeared to be sufficient to specify utilization of that initiation site (Table 2). However, although WT 800 and WT Oxman differ in their nucleotide sequences by only the six base pairs 272′ to 277′, late viral RNA molecules with 5′ ends mapping to residue 264′ were present in cells infected with the former at only one-fifth the relative abundance as cells infected with the latter. Even more confusing is the observation that the relative abundances of 5′ ends mapping to residues 106, 120, 130, and 146 were consistently greater in RNA samples prepared from WT 800-infected cells than in ones prepared from cells infected with any of the other WT strains (Fig. 2). Possible reasons for these findings are considered below.

5′ Termini of late RNAs synthesized from mutants with alterations in the 3′ end of the SV40 late promoter region. To look for measuring functions that might determine the locations of the 5′ ends of late RNAs that map near the 3′ end of the late promoter region, we examined the structures of viral RNAs obtained from CV-1P cells infected with mutants that contain deletions (*dl*1746, *dl*1747, and *dl*1748) or an insertion (S+8) in this region (Table 1). Figures 4 and 5 (left panel) show autoradiograms of two of the numerous gels that were analyzed to determine the locations and relative abundances of the 5′ ends of the RNAs made from these mutants. Figure 6 and Table 2 summarize the date obtained for the three deletion mutants.

The cDNAs made from late RNA of mutant S+8 were identical to those made from late RNA of its parental strain, 776, except they were all 10 bases larger because of the insertion of 10 base pairs that exists in the mutant between nucleotide residues 348 and 349 (Fig. 5). Those mapping downstream of the insertion were identical in size (data not shown). Therefore, the 5′ ends of the RNAs synthesized from this mutant were identical in nucleotide sequence locations to those of its parental WT strain.

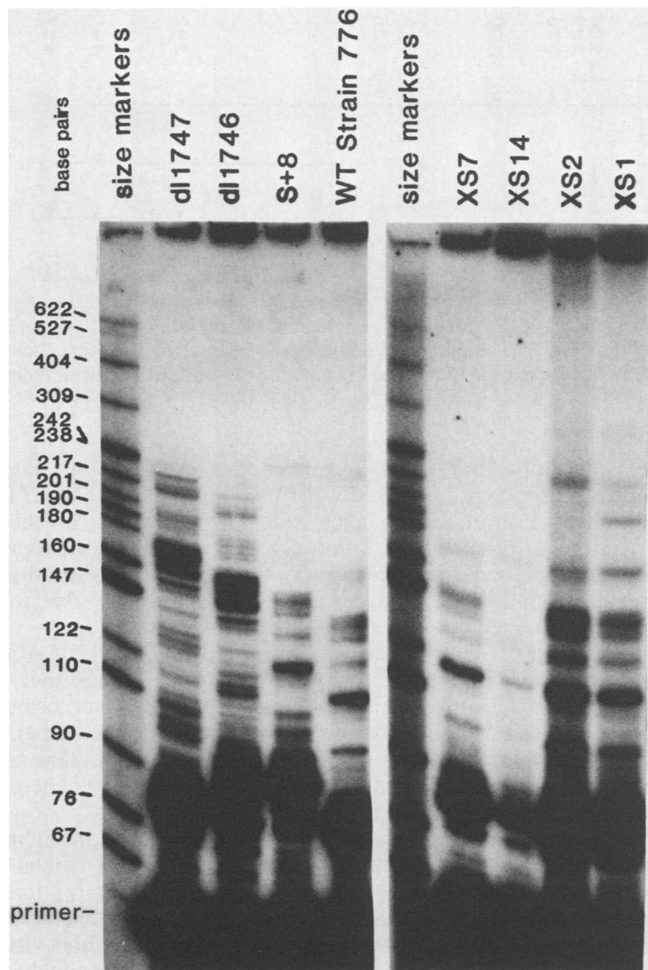


FIG. 5. Left panel, autoradiogram of a low-resolution gel of the cDNAs synthesized from SV40 late cytoplasmic RNAs isolated from CV-1P cells infected with downstream mutants *dl*1747, *dl*1746, and S+8. The cDNAs were synthesized, as described in Materials and Methods, with a single-stranded, 5′ end-labeled primer that extended from nucleotide residues 407 to 349, inclusive. The size markers were *Msp*I-cleaved, 5′ end-labeled pBR322 DNA. Right panel, autoradiogram of a low-resolution gel of the cDNAs synthesized from SV40 late cytoplasmic RNAs isolated from CV-1P cells infected with upstream mutants XS7, XS14, XS2, and XS1. The cDNAs were synthesized, as described in Materials and Methods, with a single-stranded, 5′ end-labeled primer that extended from nucleotide residues 407 to 349, inclusive. The size markers were *Msp*I-cleaved, 5′ end-labeled pBR322 DNA.

The relative abundances of the 5′ ends of the RNAs obtained from cells infected with mutant S+8 were also determined (data not shown). Except for a slight reduction in the amount initiated at residue 325, all were similar to those of WT strain 776.

Mutants *dl*1746, *dl*1747, and *dl*1748 are particularly interesting because they lack the major initiation site at residue 325. Mutant *dl*1746 lacks the second most abundantly used initiation site, located at residue 264, as well. Nevertheless, the nucleotide sequence locations of the 5′ ends of the RNAs made from these mutants were identical to those made from their parental strain, WT 800 (Fig. 6). The lengths of the cDNAs shown in Fig. 5, that were made from RNAs of mutants *dl*1746 and *dl*1747, can be compared with each other. As expected, they differed in those mapping upstream of the mutations by 11 bases. They should not be compared directly with those of S+8 and strain 776 because (i) their deletions map upstream rather than downstream of the primer used in the cDNA syntheses; (ii) as with their parental strain, they contain 85- rather than 72-base-pair duplications of the enhancer region; and (iii) they lack the 5′ splice site situated at residue 294.

As with the WT strains, these deletion mutants also enabled us to determine how many base pairs of sequence information surrounding transcriptional initiation sites was sufficient for their utilization. The data presented in Fig. 4 and 6 and Table 2 show that the existence of as few as two nucleotides downstream of residue 268 before the deletion was sufficient for the synthesis of RNA from mutant *dl*1747 with 5′ ends mapping to residue 268. Similarly, at most four downstream nucleotides were needed for the synthesis from *dl*1748 of RNA with 5′ ends mapping to residue 276.

Last, it should be noted that these three deletion mutants exhibit dramatic differences from the WT in the relative abundances of the 5′ ends of the RNAs made from them. The 5′ ends mapping to residues 264, 268, and 264′ accounted for only approximately 12, 1, and 1%, respectively, of the total RNA made from WT 800, but they accounted for approximately 60, 8, and 7%, respectively, of those made from *dl*1747 and *dl*1748 (Table 2; Fig. 4, 5, and 6). In cells infected with mutant *dl*1746, which lacks residues 264 as well as 325, RNAs with 5′ ends mapping to residues 239, 247, and 264′ accounted for approximately 50, 10, and 8%, respectively, of the total RNA. Surprisingly, approximately 50% of the RNAs made from this latter mutant have 5′ ends mapping to residue 239, which accounted for less than 2% of the total in both of the other two deletion mutants and WT 800.

5′ Termini of late RNAs synthesized from mutants with alterations in the 5′ end of the SV40 late promoter region. To identify sequences involved in determining the locations of the 5′ ends of late RNAs that map near the 5′ end of the late promoter region, we also examined the structures of the late SV40 RNAs obtained from cells infected with mutants XS1, XS2, XS7, and XS14, that contain alterations in this latter region (Tables 1 and 2). Figure 5 (right panel) shows an autoradiogram of one of the many gels that we analyzed to determine the locations and relative abundances of the 5′ ends of the RNAs made from these mutants. Figure 7 and Table 2 summarize the results. Once again, we found that the nucleotide residue locations of the 5′ ends of the RNAs made from these mutants were identical to those of the parental WT strain, with those mapping downstream of their alterations producing cDNAs that co-migrated in the gel. As few as six nucleotides of sequence downstream of residue 28 was sufficient for initiation of transcription to occur at that site in mutants XS1 and XS2 (Table 2). Twelve nucleotides of sequence upstream of residue 120 was sufficient for initiation to occur at that site in mutants XS7 and XS14. However, initiation was not detected in the latter two mutants two bases upstream of their deletion endpoints at the location that corresponds to residue 106 in the WT. Residue 178 also functioned at most very inefficiently as an initiation site in WT strain 776, even though all of the nucleotides directly downstream of it are identical to those downstream of residue 106 (Table 2). Therefore, at least some of the nucleotides mapping at or upstream of this initiation site are an essential part of the initiation signal.

The only significant differences that were found in the relative abundances of the various 5′ ends were in those mapping to residues 28 and 106. Whereas approximately 3 to 5% of the RNAs made from the WT and mutants XS1 and

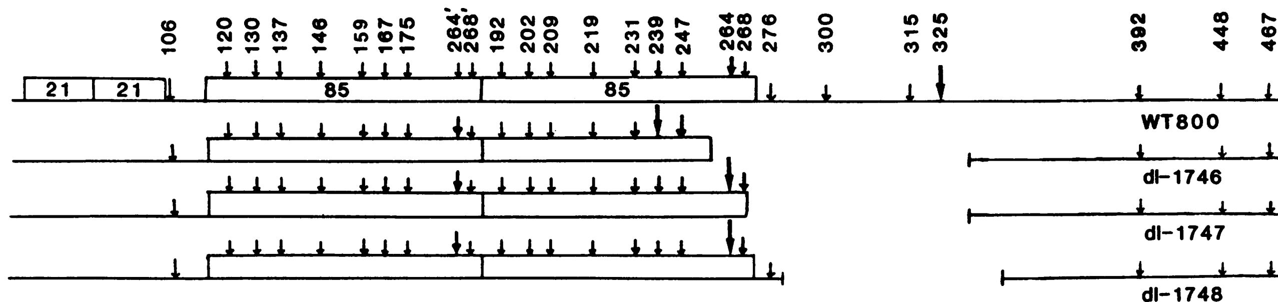


FIG. 6. Summary of the locations of the 5′ ends of the late viral RNAs made from the downstream deletion mutants *dl*1746, *dl*1747, and *dl*1748. The results shown here were determined as described in the legend to Fig. 3. The mutant genomes have been aligned so that their common nucleotide sequences line up with those of their parental strain, WT 800; the gaps in their lines indicate the regions missing from their genomes.

XS2 had 5′ ends mapping to both residues 28 and 106, approximately 6 to 9% of the 5′ ends mapped to residue 28 in mutants XS7 and XS14, which lack the sequences around residue 106 (Table 2).

DISCUSSION

Sequences involved in determining the locations of 5′ ends of late RNAs. The primary aim of the studies reported here was to determine whether the late leader region of SV40 contained TATA-boxlike elements that determine the precise locations of the 5′ ends of the late RNAs. If such elements existed, we would have seen shifts in the locations of at least some of the 5′ ends of the late RNAs made from the WT and mutant strains of SV40 analyzed here. No shifts were found, i.e., the locations of the 5′ ends of the late RNAs both downstream and upstream from altered sequences were qualitatively always identical to those of the parental WT strains. Based on a preliminary analysis of several mutants containing deletions in the 5′ end of the SV40 late promoter, Contreras et al. (9) have noted that the locations of at least some of the minor 5′ ends are not altered by upstream mutations. In studies performed concurrently with those reported here, Piatak et al. (29) have published an analysis of 11 mutants containing duplications of or deletions near the major initiation site. Their extensive study demonstrated definitively that initiation at residue 325 is determined largely by sequences situated within 20 nucleotides of that site. Although the studies of Brady et al. (5) indicated the possible existence of a TATA-boxlike sequence involved in initiation of transcription at residue 325, the failure of Piatak et al. (29) to observe dramatic changes with their mutants in the locations of the major initiation sites suggests, as discussed by them and confirmed recently by Nandi et al. (28), that the results obtained by Brady et al. (5) were probably due to their having generated a TATA-boxlike sequence.

The data presented here also indicated that the CTAAC sequences are not involved in determining the precise locations of transcriptional initiation within the late promoter of SV40. This finding extends the conclusion of Grosveld et al. (20), who observed for the rabbit β-globin gene that a CCAAT box was not involved in specifying precisely where transcription initiated.

Having been unable to identify any sequences that might contain measuring functions, we confirm the novel mechanism of Piatak et al. (29) that the transcriptional initiation sites of some RNA polymerase II promoters are determined largely by sequences within their immediate vicinities. In addition, the studies presented here generalize their conclusion by demonstrating that many (and likely all) of the approximately two dozen locations of 5′ ends of late RNAs of SV40 are each determined largely by this novel mechanism.

What are the specific nucleotide sequences within the SV40 late promoter that specify sites of transcriptional initiation? If we assume that the 5′ ends of the SV40 late RNAs reflect transcriptional initiation events, the data presented here

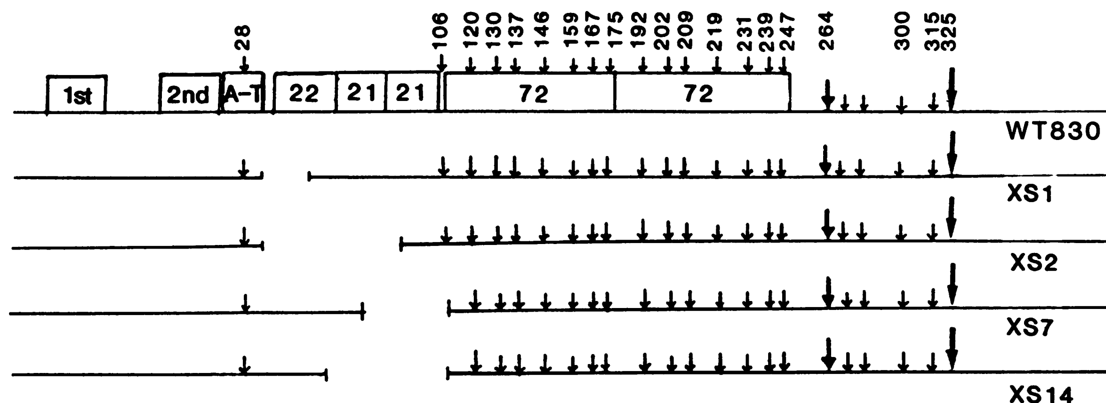


FIG. 7. Summary of the locations of the 5′ ends of the late viral RNAs made from upstream deletion mutants XS1, XS2, XS7, and XS14. The results shown here were determined as described in the legend to Fig. 3. The mutant genomes have been aligned so that their common nucleotide sequences line up with those of their parental strain, WT 830; the gaps in their lines indicate the regions missing from their genomes. The size markers consisted of *Msp*I-digested, 5′ end-labeled pBR322 DNA.

indicate that at most five bases adjacent to the 5′ ends of the RNAs is sufficient for utilization of these sites. For example, the data in Fig. 4 and Table 2 show clearly that, although the deletions in mutants *dl*1747 and *dl*1748 are situated as few as two to five base pairs downstream from the map location of 5′ ends of RNAs, RNAs made from these mutants with 5′ ends mapping to these sites were, nevertheless, observed. In fact, the only instance in which we failed to observe a 5′ end at an expected location was with RNA obtained from WT Oxman-infected cells in which a sequence alteration begins two base pairs downstream from residue 276′. These data indicate that the nucleotides situated within a few base pairs upstream and downstream of the map location of 5′ ends of SV40 late RNAs are sufficient to specify their utilization as transcriptional initiation sites.

We analyzed in detail the sequences surrounding each of the initiation sites present both (i) near novel joints in the various WT strains and mutants studied here (Table 2) and (ii) throughout the late promoter region of WT strain 776 (data not shown). Based on an analysis of cap structures present in SV40 late mRNAs, Haegeman et al. (21) had reported previously that most initiations begin with a purine residue. Consistent with their observation, we found that most of the 5′ ends of the SV40 late RNAs mapped to purines. Although some minor 5′ ends may have mapped to pyrimidines (residues 175 and 315, for example), we cannot state this definitively because of possible errors of as much as ±2 nucleotide residue numbers in our designated locations of initiation sites.

The only other nucleotide sequence preference we detected was the presence of TGG approximately 10 to 15 base pairs downstream of initiation sites that were utilized with fairly high efficiency (e.g., the ones at residues 264 and 325). This latter finding might be the reason for the relatively inefficient utilization of the residue 264′ initiation site in WT 800 compared with WT Oxman and the residue 167 initiation site in WT Chambon compared with WT strain 776 (Table 2). It could also explain, at least in part, why the deletion in mutant *dl*1747 did not decrease relative to *dl*1748 the efficiency of utilization of the residue 264 initiation site.

Although we failed to identify a larger consensus sequence, one cannot yet state definitively that one does not exist because (i) we may not have precisely aligned the initiation sites with each other because of errors in assigning their exact locations and (ii) there may exist within the SV40 late promoter two or more classes of initiation site (e.g., strong and weak ones), each with its own consensus sequence. Characterizations of sets of mutants containing insertions of related, synthetically defined oligonucleotides will be necessary to identify definitively sequences that define initiation sites within the complex SV40 late promoter.

What determines the relative abundances of the 5′ ends of the SV40 late RNAs? Our data and those reported previously by others (15, 18, 29, 30) show that the relative abundances of RNAs with 5′ ends mapping to each transcriptional initiation site within the late promoter region of SV40 can range from less than 1 to greater than 60%. Also shown here is the fact that the relative abundances of RNAs with 5′ ends mapping to specific sites within the late promoter region (e.g., residue 239 in WT versus *dl*1746) sometimes vary by as much as 50-fold (see above; Table 2; Fig. 3 and 6).

The first of these observations can be rationalized by proposing that the late leader region of SV40 encodes numerous transcriptional initiation signals, each of which is utilized at a relative efficiency that is determined in part by its relative intrinsic strength as a promoter. The 5′ ends at residues 264 and 325 are fairly strong transcriptional initiation signals, whereas those mapping at other locations are rather weak. This hypothesis explains why duplication of the sequences around residues 264 in WT Oxman (Table 2) and 325 in mutants KMTD-2 and KMTD-3 (29) results in efficient utilization of these transcriptional initiation signals at their second sites as well.

This hypothesis is not sufficient to explain all of the changes in the relative abundances of the various 5′ ends that have been observed with mutants containing alterations within the late leader region. These data indicate that the sequences within the SV40 late promoter region situated downstream of the transcriptional initiation signal, as well as its intrinsic strength, determine the percentage of RNAs with 5′ ends that map to it.

Several hypotheses can explain this observation. One is that the precise sequences at the 5′ end of an initial transcript affect its half-life and, consequently, the rate at which it accumulates in infected cells. Although this possibility has not been excluded definitively, the findings of Piatak et al. (30), showing by chase experiments with actinomycin D that the relative abundances of different late mutant RNA species do not change significantly with time, indicate that this explanation is unlikely to be valid.

An alternative hypothesis invokes the existence of two transcriptional control elements, one mapping at or near each end of the late promoter region. We propose that at early times after infection the upstream control element directs RNA polymerase to enter within or near the 21-base-pair tandem repeat region. At late times after infection, this upstream control element becomes inactivated, possibly via changes in chromatin structure induced directly or indirectly by T antigen. Therefore, RNA polymerase enters at the downstream site (which we propose may be situated somewhere downstream of residue 325) and scans the promoter region starting from here until it interacts with a transcriptional initiation site. The probability that a given transcriptional initiation site will trap a given RNA polymerase molecule will then depend on both its relative intrinsic strength as an initiation signal and its location relative to the site at which polymerase entered, with promoters situated closest to the site of entry being preferentially utilized.

This second hypothesis is consistent with most of the data we and others have obtained with late leader region mutants of SV40. For example, we found, as predicted, that translocation of initiation sites to positions near the 3′ end of the late promoter region results in their being utilized more frequently (Fig. 6; Table 2). Second, deletion of the downstream control element, as occurs in mutant *dl*805, results in a dramatic underrepresentation of RNAs with 5′ ends mapping to downstream initiation sites that are still present in the mutant genome (see reference 15 for data). Lastly, Alwine (1) has reported that mRNAs made from a late leader region mutant are underrepresented in mixed infections with WT virus. Possibly, if the leader region mutant were defective in the proposed downstream control element, WT viral DNA might outcompete mutant DNA for binding of a limited supply of RNA polymerase II molecules or other transcriptional factors at late times after infection.

A third hypothesis to explain the altered representation seen with late leader region mutants of RNAs with 5′ ends mapping to various sites within the late promoter invokes the existence of a regulated transcriptional terminator which controls termination of transcription at approximately residues 416 to 432. We propose that this transcriptional termi-

nator may determine not only when during the lytic cycle of infection late strand mRNAs are synthesized but also, depending on the site utilized for transcriptional initiation, whether newly initiated RNA polymerase molecules will terminate transcription prematurely or proceed beyond the leader region to give rise eventually to large transcripts that can be processed into stable mRNA molecules. This hypothesis, which can also explain the data presented here and elsewhere concerned with late leader region mutants, implies that the stable cytoplasmic RNAs whose 5′ ends we and others have mapped may only be indicative of the subset of transcriptional initiation events that resulted in the synthesis of mRNA molecules. The details of this model and data that support it directly will be presented elsewhere (J. E. Mertz, manuscript in preparation).

In summary, the data presented here provide strong support for the hypothesis that sequences within the immediate vicinity of many (and likely all) of the numerous transcriptional initiation sites situated within the late promoter region of SV40 play a primary role in determining the precise locations of the 5′ ends of the SV40 late RNAs. Although sequences somewhat removed from these sites appear to modulate the relative efficiencies with which they are utilized, the mechanism by which this occurs is not yet clear. It will be interesting to see whether the promoters of cellular genes such as that encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase that also contain numerous transcriptional initiation signals and lack TATA boxes (33) are functionally similar to the SV40 late promoter.

ACKNOWLEDGMENTS

We thank Sherman Weissman for WT strains 776 and Oxman, Christophe Benoist for pSV-1, Mike Fromm for pS+8, pXS1, pXS2, pXS7, and pXS14, and Jeff Ross, Marv Wickens, and Peter Good for helpful comments on the manuscript.

This work was supported by Public Health Service research grants CA-07175 and CA-2443 from the National Cancer Institute and MV-201 from the American Cancer Society.

LITERATURE CITED

1. **Alwine, J. C.** 1982. Evidence for simian virus 40 late transcriptional control: mixed infections of wild-type simian virus 40 and a late leader deletion mutant exhibit *trans* effects on late viral RNA synthesis. J. Virol. **42**:798–803.
2. **Barkan, A., and J. E. Mertz.** 1984. The number of ribosomes on simian virus 40 late 16S mRNA is determined in part by the nucleotide sequence of its leader. Mol. Cell. Biol. **4**:813–816.
3. **Benoist, C., and P. Chambon.** 1981. *In vivo* sequence requirements of the SV40 early promoter region. Nature (London) **290**:304–310.
4. **Benoist, C., K. O'Hare, R. Breathnach, and P. Chambon.** 1980. The ovalbumin gene-sequence of putative control regions. Nucleic Acids Res. **8**:127–142.
5. **Brady, J., M. Radonovich, M. Vodkin, V. Natarajan, M. Thoren, G. Das, J. Janik, and N. P. Salzman.** 1982. Site-specific base substitution and deletion mutations that enhance or suppress transcription of the SV40 major late RNA. Cell **31**:625–633.
6. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
7. **Buchman, A. R., L. Burnett, and P. Berg.** 1981. The SV40 nucleotide sequence, appendix A. *In* J. Tooze (ed.), DNA tumor viruses, molecular biology of tumor viruses, 2nd ed., part 2, revised, p. 799–841. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
8. **Contreras, R., and W. Fiers.** 1981. Initiation of transcription by RNA polymerase II in permeable, SV40-infected or noninfected, CV1 cells; evidence for multiple promoters of SV40 late transcription. Nucleic Acids Res. **9**:215–236.
9. **Contreras, R., D. Gheysen, J. Knowland, A. van de Voorde, and W. Fiers.** 1982. Evidence for the direct involvement of DNA replication origin in synthesis of late SV40 RNA. Nature (London) **300**:500–505.
10. **Fiers, W., R. Contreras, G. Haegeman, R. Rogiers, A. Van de Voorde, H. Van Heuverswyn, J. Van Herreweghe, G. Volckaert, and M. Ysebaert.** 1978. Complete nucleotide sequence of SV40 DNA. Nature (London) **273**:113–120.
11. **Fromm, M., and P. Berg.** 1982. Deletion mapping of DNA regions required for SV40 early region promoter function *in vivo*. J. Mol. Appl. Genet. **1**:457–481.
12. **Fromm, M., and P. Berg.** 1983. Transcription *in vivo* from SV40 early promoter deletion mutants without repression by large T antigen. J. Mol. Appl. Genet. **2**:127–135.
13. **Gheysen, D., A. van de Voorde, R. Contreras, J. Vanderheyden, F. Duerinck, and W. Fiers.** 1983. Simian virus 40 mutants carrying extensive deletions in the 72-base-pair repeat region. J. Virol. **47**:1–14.
14. **Ghosh, P. K., P. Lebowitz, R. J. 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. USA **78**:100–104.
15. **Ghosh, P. K., M. Piatak, J. E. Mertz, S. M. Weissman, and P. Lebowitz.** 1982. Altered utilization of splice sites and 5′ termini in late RNAs produced by leader region mutants of simian virus 40. J. Virol. **44**:610–624.
16. **Ghosh, P. K., V. B. Reddy, M. Piatak, P. Lebowitz, and S. M. Weissman.** 1980. Determination of RNA sequences by primer directed synthesis and sequencing of their cDNA transcripts. Methods Enzymol. **65**:580–595.
17. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. V. Choudary, P. Lebowitz, and S. M. Weissman.** 1978. The 5′-terminal leader sequence of late 16 S mRNA from cells infected with simian virus 40. J. Biol. Chem. **253**:3643–3647.
18. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. M. Weissman.** 1978. Heterogeneity and 5′-terminal structures of the late RNAs of simian virus 40. J. Mol. Biol. **126**:813–846.
19. **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. USA **78**:2174–2178.
20. **Grosveld, G. C., E. de Boer, C. K. Shewmaker, and R. A. Flavell.** 1982. DNA sequences necessary for transcription of the rabbit β-globin gene *in vivo*. Nature (London) **295**:120–126.
21. **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.
22. **Kimura, G., and R. Dulbecco.** 1972. Isolation and characterization of temperature-sensitive mutants of simian virus 40. Virology **49**:394–403.
23. **Mathis, D. J., and P. Chambon.** 1981. The SV40 early region TATA box is required for accurate *in vitro* initiation of transcription. Nature (London) **290**:310–315.
24. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
25. **Mertz, J. E., and P. Berg.** 1974. Defective simian virus 40 genomes: isolation and growth of individual clones. Virology **62**:112–124.
26. **Mertz, J. E., J. Carbon, M. Herzberg, R. W. Davis, and P. Berg.** 1975. Isolation and characterization of individual clones of simian virus 40 mutants containing deletions, duplications and insertions in their DNA. Cold Spring Harbor Symp. Quant. Biol. **39**:69–84.
27. **Moreau, P., R. Hen, B. Wasylyk, R. Everett, M. P. Gaub, and P. Chambon.** 1981. The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res. **9**:6047–6068.
28. **Nandi, A., G. Das, and N. P. Salzman.** 1985. Characterization of a surrogate TATA box promoter that regulates in vitro transcription of the simian virus 40 major late gene. Mol. Cell. Biol.

5:591–594.
29. **Piatak, M., P. K. Ghosh, L. C. Norkin, and S. M. Weissman.** 1983. Sequences locating the 5' ends of the major simian virus 40 late mRNA forms. J. Virol. **48**:503–520.
30. **Piatak, M., K. N. Subramanian, P. Roy, and S. M. Weissman.** 1981. Late messenger RNA production by viable simian virus 40 mutants with deletions in the leader region. J. Mol. Biol. **153**:589–618.
31. **Reddy, V. B., P. K. Ghosh, P. Lebowitz, M. Piatak, and S. M. Weissman.** 1979. Simian virus 40 early mRNA's. I. Genomic localization of 3' and 5' termini and two major splices in mRNA from transformed and lytically infected cells. J. Virol. **30**: 279–296.
32. **Reddy, V. B., B. Thimmappaya, R. Dhar, K. N. Subramanian, B. S. Zain, J. Pan, P. K. Ghosh, M. L. Celma, and S. M. Weissman.** 1978. The genome of simian virus 40. Science **200**:494–502.
33. **Reynolds, G. A., S. K. Basu, T. F. Osborne, D. J. Chin, G. Gil, M. S. Brown, J. L. Goldstein, and K. L. Luskey.** 1984. HMG CoA reductase: a negatively regulated gene with unusual promoter and 5' untranslated regions. Cell **38**:275–285.
34. **Ross, J.** 1976. A precursor of globin messenger RNA. J. Mol. Biol. **106**:403–420.
35. **Sanger, F., and A. R. Coulson.** 1978. The use of thin acrylamide gels for DNA sequencing. FEBS Lett. **87**:107–110.
36. **Shenk, T.** 1981. Transcriptional control regions: nucleotide sequence requirements for initiation by RNA polymerase II and III. Curr. Top. Microbiol. Immunol. **93**:25–46.
37. **Tjian, R.** 1978. The binding site on SV40 DNA for a T antigen-related protein. Cell **13**:165–179.