Volume 13 Number 15 1985 Nucleic Acids Research

Exon mutations that affect the choice of splice sites used in processing the SV40 late transcripts

Madhu B.Somasekhar* and Janet E.Mertz

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

Received 9 April 1985; Revised and Accepted 16 July 1985

ABSTRACT

The spliced species of late SV40 RNAs present in the cytoplasm of cells infected with various wild-type and mutant strains of SV40 that differ in their leader regions were determined using a novel modification of the primer extension method and the S1 nuclease mapping technique. These data indicated that mutations within the first exon of the late RNAs can affect dramatically the utilization of downstream donor and acceptor splice sites. In one instance, a ten base pair insertion within the predominant first exon increased utilization of an infrequently utilized donor splice site such that the small alteration became part of an intervening sequence, thereby suggesting a novel mechanism for regulation of gene expression. In addition, our method enabled detection of a previously unidentified spliced species, representing less than one percent of the SV40 late 19S RNA present in cells infected with wild-type virus, that may be an intermediate in the synthesis of a known doubly spliced 16S RNA species of SV40.

INTRODUCTION

Most eukaryotic genes are interrupted by intervening sequences that are removed from the primary transcripts by RNA splicing. The primary transcripts may contain numerous intervening sequences and numerous bonafide or cryptic splice sites (1). Nevertheless, in most instances splicing occurs predominantly between specific pairs of 5' and 3' splice sites. The specific sequences in the RNA that determine which splice sites pair in processing have not yet been identified. Two current hypotheses state that certain sequences present predominantly within the intervening sequences might dictate selection of donor and acceptor splice sites. From the analysis of 139 splice junctions, Mount (2) derived a 9 base consensus sequence near the donor splice site. Lerner et al (3) proposed that sequences flanking the splice junctions base pair with sequences at the 5' end of the small nuclear RNA, U1. Padgett et al (4) and Moore and Sharp (5) showed using cell free systems an involvement of U1 SnRNP in the formation of a multifunctional RNA processing complex.

The late region of the simian virus 40 (SV40) genome encodes several spliced species of mRNAs. Transcription initiates at numerous sites within a 300 base pair region of the late promoter (6-10a). These primary transcripts are processed into 16S and 19S size classes of RNAs using many, but not all, of the theoretically possible pairings of three donor and three acceptor splice sites that they encode (reviewed in refs. 11 and 12; see Figs. 1 and 2). The 16S RNAs are fairly homogeneous in structure, with a majority of them having their 5' end at nucleotide residue 325, the major cap (MC) site, and being the product of a single splicing event that joins nucleotide residue 526 to 1463. The 19S RNAs consist of unspliced and three subclasses of molecules spliced using each of the three possible donor splice sites. In addition, their 5' ends map to numerous locations throughout the three hundred base pair promoter region (13).

The reason for the heterogeneity in the structure of the leader region of the late SV40 RNAs is not yet known. Several groups of workers have analyzed in detail the structures of the viral RNAs synthesized by wild-type (WT) and viable mutants of SV40 having deletions in the late leader region (10,13,14). Based upon these data, it has been hypothesized that the precise sequences at the 5' end of a primary late strand transcript of SV40 might influence the choice of splice sites used in its processing (10,13). Unfortunately, most of the mutants studied were not well suited for examining the effect of sequence alterations within an exon on processing of the primary transcripts because they (i) possessed fairly large deletions that included the acceptor splice site at residue 435 or (ii) had small alterations, but contained either duplications encompassing the major cap site at nucleotide residue 325 or additional, undefined sequence changes.

To examine the possibility that sequences within an exon might control the processing of primary transcripts, we have determined by the S1 nuclease mapping technique and a novel modification of the primer extension method the relative abundances of the various spliced species of late SV40 RNAs synthesized in cells infected with several wild-type and mutant strains of SV40 that have small alterations in their late leader regions. Our primary conclusion from this study is that alterations within the region of the SV40 genome that encodes the 5' exons of the late RNAs can affect dramatically the utilization of downstream donor and acceptor splice sites. Additional findings include (i) further support for the hypothesis that the precise sequence of the 5' end of an initial transcript plays an important role in determining which splice sites are used in its processing, and

(ii) the identification of an RNA species lacking residues 295-434 that may be an intermediate in the synthesis of a previously reported doubly spliced SV40 late 16S RNA species.

MATERIALS AND METHODS

Cells, viral mutants, and viral RNAs

CV-1P cells, an established line of African green monkey kidney cells, were grown in Dulbecco's modified Eagle's medium supplemented with 5% fetal calf serum as described previously (15). Mutant S+8, obtained from M. Fromm and P. Berg, contains a 10 base pair insertion between nucleotide residues 348 and 349 (16). The isolation and description of the various wild-type strains of SV40 and deletion mutants dl-1746, dl-1747, and dl-1748 are presented elsewhere (10a). Plasmid pSV1-16SIVS, a cloned mutant of SV40 that lacks inclusively nucleotide residues 527-1462 which encode precisely the intervening sequence of the SV40 major late 16S mRNA (see Fig. 1), was a generous gift of J. Sklar and P. Berg; the SV40 late region BglI-EcoRI DNA fragment of this plasmid was recombined with a BglI-EcoRI DNA fragment of pBR322 DNA (residues 1163-4361) by P. J. Good of our laboratory and then used to make the 16S cDNA probe depicted in Fig. 1. The isolation and purification of cytoplasmic RNA from SV40-infected CV-1P cells were performed as described previously (10a). All of the WT- and mutant-infected cells yielded similar amounts of total late SV40 RNA when harvested at 42 h after infection.

S1 mapping of the late 16S RNAs

To prepare probe for analysis of 16S RNA with S1 nuclease, a mutant of SV40 derived from a 16S cDNA clone was cleaved with EcoRI restriction endonuclease, treated with calf intestinal alkaline phosphatase, 5' end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase as described by Maxam and Gilbert (17), and then digested with TaqI restriction endonuclease. The resulting DNA fragments were denatured in 30% DMSO and strand separated by electrophoresis in a 5% polyacrylamide gel (acrylamide:bis-acrylamide, 40:1) (17). The approximately 950 base single-stranded fragment (see Fig. 1 for structure) was located by autoradiography, electroeluted from the gel, and purified by ethanol precipitation.

One µg of RNA purified from the cytoplasm of CV-1P cells that had been infected with virus 42 h previously and a molar excess of the EcoRI-TaqI single-stranded probe described above were co-precipitated with ethanol and redissolved in 10 µl of 80% deionized formamide, 40 mM PIPES (pH 6.4),

0.4 M NaCl, 1 mm EDTA. After preincubation at 85° for 5 min, the sample was incubated at 45° for 10-12 h. The unhybridized DNA was degraded by incubation at 37° for 1 h with 100 µl of 250 mM NaCl, 30 mM Na_2OAc (pH 5.5), 1 mM $ZnSO_4$ containing S1 nuclease. The undegraded DNA was precipitated with ethanol and analyzed by electrophoresis in a 0.7 mm thick, 8% polyacrylamide gel containing 7 M urea (18). MspI-digested, 5' end-labeled pBR322 DNA was used as size markers.

Analysis of the 19S RNA by a novel modification of the primer extension method

To prepare primer for the analysis of SV40 late 19S RNA as outlined in Figure 2, SV40 DNA was cleaved with HaeIII restriction endonuclease. The two resulting DNA fragments that were 299 and 300 base pairs in length were purified by electrophoresis, 5' end-labeled as described above, and cleaved with Sau96I restriction endonuclease. The resulting fragments were denatured in 80% formamide and strand separated by electrophoresis in a 8% polyacrylamide gel containing 7 M urea (17). The resulting 72 base fragment that was 5' end-labeled at residue 661 and extended to residue 592 was electroeluted from the gel and purified.

One µg of cytoplasmic RNA obtained from virus-infected CV-1P cells and a molar excess of the HaeIII-Sau96I single-stranded primer were co-precipitated with ethanol and redissolved in 30 µl 80% deionized formamide, 60 mM PIPES (pH 6.4), 0.4 M NaCl, 1 mM EDTA. The primer was hybridized to the RNAs by incubation at 85° for 5 min and then 45° for 5 h. After precipitation with ethanol, cDNAs were synthesized using reverse transcriptase (10 units per reaction, Life Sciences) at 45° for 3/4 h in 30 µl of 50 mM Tris-HCl (pH 8.3), 8 mM MgOAc, 60 mM NaCl, 10 mM dithiothreitol, 50 units of RNAsin (Promega-Biotec) and 1.0 mM each of dATP, dTTP, dGTP, and dCTP. The RNA was degraded by addition of EDTA to 1 mM and NaOH to 0.2 N, followed by incubation at 45° for 1 h. After neutralization of the reaction with HCl, the resulting cDNAs were extracted with phenol:chloroform (1:1), precipitated with ethanol, and then hybridized to single-stranded DNA complementary to the 3' end of the cDNAs (e.g., the BglI-HpaII restriction fragment of SV40, see Fig. 2) by incubation at 100° for 2-3 min in 10 mM Tris-HCl (pH 7.6), 1 mM EDTA and then 68° for 4 h after the addition of NaCl to 0.1 M. The resulting hybrids were then precipitated with ethanol and cleaved with appropriate restriction endonucleases. This procedure created unique 3' ends for all of the cDNAs even though the mRNAs from which they were copied had heterogeneous 5' ends. The lengths of these cleaved cDNAs were then

determined by electrophoresis in 0.3 mm thick, 8% polyacrylamide gels containing 7 M urea (18).

RESULTS

A novel modification of the primer extension method

The late region of the SV40 genome encodes numerous spliced and unspliced species of 19S mRNAs whose 5' ends map to heterogeneous locations throughout a 300 base pair region of the late promoter (see Figs. 1 and 2). Because of this heterogeneity in both 5' ends and splice sites used in processing, no single probe for S1 nuclease mapping can be used to detect all of these species of 19S mRNAs. Previously, these RNAs could only be analyzed in detail by primer extension followed by sequencing of each cDNA. This method suffers the disadvantages of (i) being very time consuming, (ii) requiring large amounts of RNA, and (iii) being only roughly quantitative, with minor species being overlooked. To overcome these problems, a method was needed that eliminates the heterogeneity at the 3' ends of the cDNAs copied from the various species of mRNA molecules. Piatak et al. (14) attempted this by cleavage of the cDNAs with restriction endonucleases that cleave single-stranded DNA. Unfortunately, this procedure is not quantitative because one cannot cleave single-stranded DNAs to completion with restriction endonucleases. In addition, few restriction endonucleases are available that cleave single-stranded DNA.

The method we have developed, outlined in Figure 2, overcomes all of these difficulties. The crucial step is the hybridization of the cDNAs with a DNA complementary to their heterogeneous ends. This simple modification enables one to obtain complete digestions with any restriction endonuclease that has a site within the heterogeneous region. As long as one already knows the precise locations of the splice sites used in processing of the primary transcripts, one can predict the lengths of the cDNAs.

Quantitative analysis of the wild-type SV40 late 19S RNAs

We demonstrate here the utility of this method for analysis of the late 19S RNAs of SV40. Late cytoplasmic RNAs prepared from monkey cells infected with various wild-type strains of SV40 and leader region mutants were hybridized to the 5' end-labeled 70 base pair Sau96I-HaeIII restriction fragment shown in Figure 2. The cDNAs were prepared using reverse transcriptase in 50 mM Tris-HCl, 8 mM MgOAc, 60 mM NaCl, 10 mM dithiothreitol and 1.0 mM each of dATP, dTTP, dGTP, and dCTP. Cleavage at the PvuII site

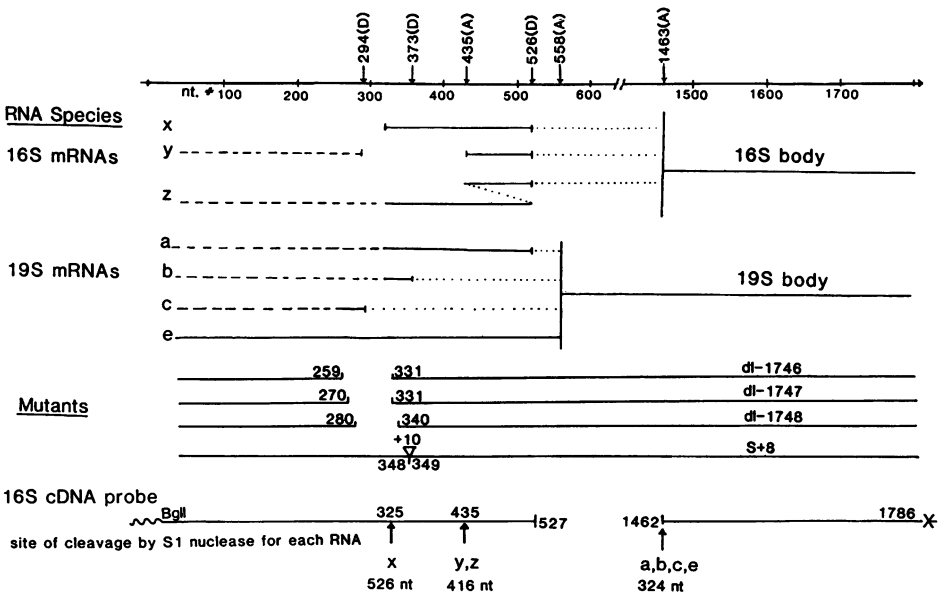

FIG. 1. Structures of the 5' ends of the SV40 late mRNAs, the SV40 mutants studied here, and the probe used in mapping with S1 nuclease. The top line shows in detail the locations of donor and acceptor splice sites used in the synthesis of the SV40 late mRNAs. The next seven lines show the structures of the late 16S and 19S mRNAs species as determined by Ghosh et al. (8). Each RNA species has a 5' leader region segment that is joined covalently to the indicated body. Approximately 70% of the WT late strand viral mRNA molecules have 5' ends that map to residue 325. The remaining 5' ends, represented by dashed lines, map at multiple sites between nucleotide residues 28 and 325. The dotted lines indicate the locations of intervening sequences. The next four lines indicate the positions of the alterations in the mutants studied here. The deletion mutants dl-1746, dl-1747, and dl-1748 lack inclusively nucleotide residues 260-330, 271-330, and 281-339, respectively (10a). Mutant S+8 has an insertion of 10 base pairs between residues 348 and 349 (16). The last line depicts the structure of the probe used in S1 nuclease mapping experiments to distinguish among the various SV40 late RNA species. The approximately 950 base EcoRI-TaqI restriction fragment, containing SV40 nucleotide residues 2 to 526 and 1463 to 1786 of pSV1-16SIVS, was 5' end-labeled with ^{32}P at the EcoRI site. The wavy line indicates the pBR322 sequences present in the probe. Arrows show the points nearest the 5' end of the probe at which S1 nuclease will cleave hybrids between the probe and each of the RNA species indicated above in the figure. The nucleotide numbering system is that of Buchman et al (31).

of cDNAs to the wild-type SV40 late 19S RNAs is predicted to yield DNA fragments 358, 205, 126, and 389 bases in length when they are synthesized, respectively, from the 526-558 spliced, 373-558 spliced, 294-558 spliced and unspliced 19S RNA species (see Fig. 2). When this method of analysis

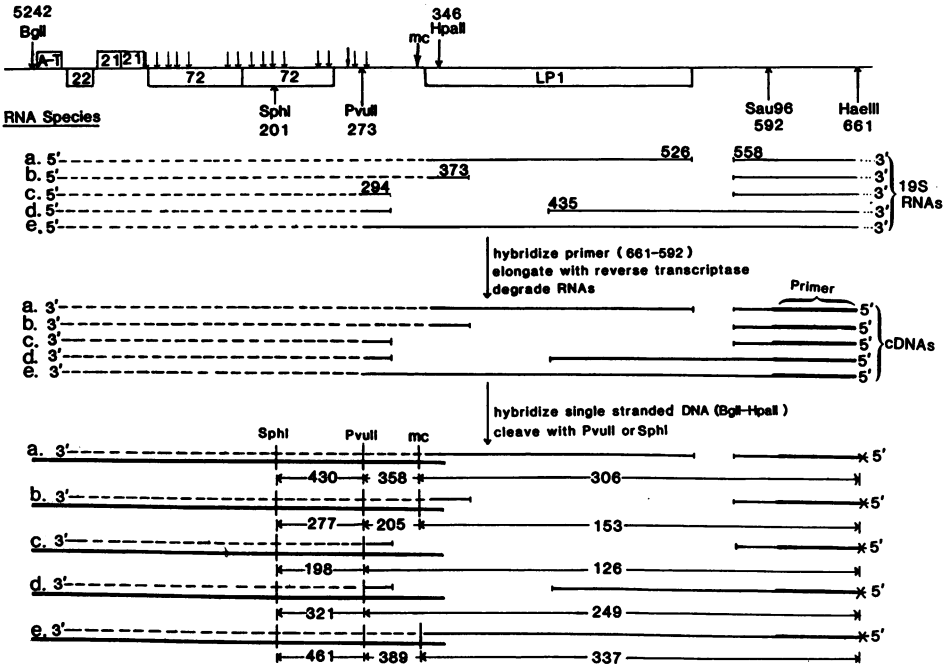

FIG. 2. Schematic representation of the procedure used to quantify the relative amounts of each spliced species of SV40 late 19S RNA. Cytoplasmic RNA was prepared from virus-infected CV-1P cells by lysis with 0.5% Nonidet P-40 at 42 h after infection. The cDNAs were prepared with reverse transcriptase and a 5' end-labeled HaeIII - Sau96I restriction fragment as primer as described in Materials and Methods. After purification, they were hybridized with the late strand of a BglI - HpaII DNA restriction fragment and were digested with restriction endonuclease PvuII or SphI to generate unique 3' ends for all of the cDNAs copied from the various RNA molecules. The resulting clipped cDNAs, whose expected sizes for WT RNAs are indicated by the numbers presented in the bottom section of the diagram, were then separated by electrophoresis in 0.3 mm thick, 8% polyacrylamide gels containing 7 M urea and their amounts quantified by densitometry of autoradiograms of the gels. The numbers next to the restriction sites shown on the top line indicate the last nucleotide residue present after cleavage in the DNAs of relevance to this study. Also shown on the top line are the locations of many of the numerous 5' ends of the late 19S RNAs (unlabeled arrows; ref. 10a), the major cap site (MC; ref. 8) the 72 base pair tandem repeats that encode transcriptional enhancer elements (32), the 21 base pair tandem repeats that form an important part of the early promoter (16), and the A-T rich region that contains a "TATA" box involved in initiation of early strand transcription (33).

was used to determine the relative quantities of the various spliced species of 19S RNAs made in cells infected with any of four wild-type strains of SV40 which differ in the length of their tandemly duplicated enhancer region,

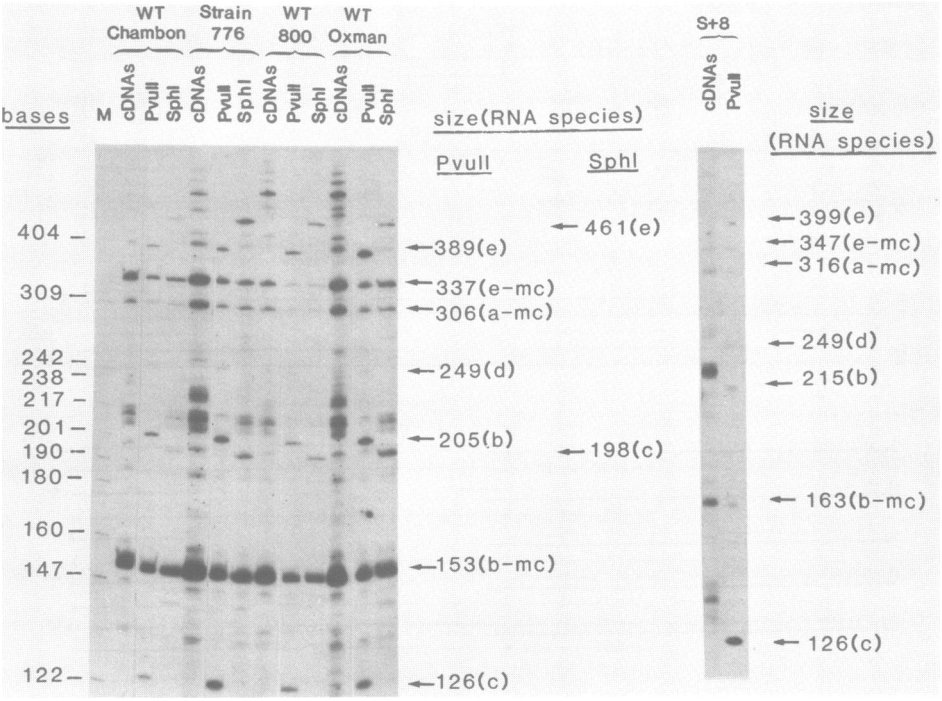

FIG. 3. Autoradiogram of a gel showing the lengths before and after cleavage with PvuII or SphI of the cDNAs synthesized from cytoplasmic RNAs isolated from CV-1P cells infected with various wild-type strains of SV40 and mutant S+8. The cDNA were synthesized as described in Material and Methods with a single-stranded, 5' end-labeled HaeIII - to - Sau96I primer. After degradation of the RNAs, the cDNAs were hybridized to a single-stranded DNA restriction fragment (BglI - HpaII) and digested where indicated with PvuII or SphI prior to electrophoresis in a 7 M urea, 8% polyacrylamide gel. The size markers were MspI-cleaved, 5' end-labeled pBR322 DNA. The prominent cDNAs seen after digestion with PvuII or SphI restriction enzyme are indicated by arrows with letters in parentheses that refer to the 19S RNA species as designated in Fig. 2. "mc" refers to the major transcriptional initiation site; in WT SV40 it maps to nucleotide residue 325 (Fig. 2).

cDNAs 389, 205, and 126 bases in length were readily observed as expected after cleavage with the restriction endonuclease PvuII (Fig. 3).

In addition, cDNA 249 bases in length was also observed in cells infected with each of the four independent isolates of WT SV40. S1 mapping with a 5' end-labeled, nucleotide residues 661 to 363 probe indicated the presence of an RNA species that was continuous with the probe up until the acceptor splice site at residue 435 (Fig. 4A). S1 nuclease analysis with a

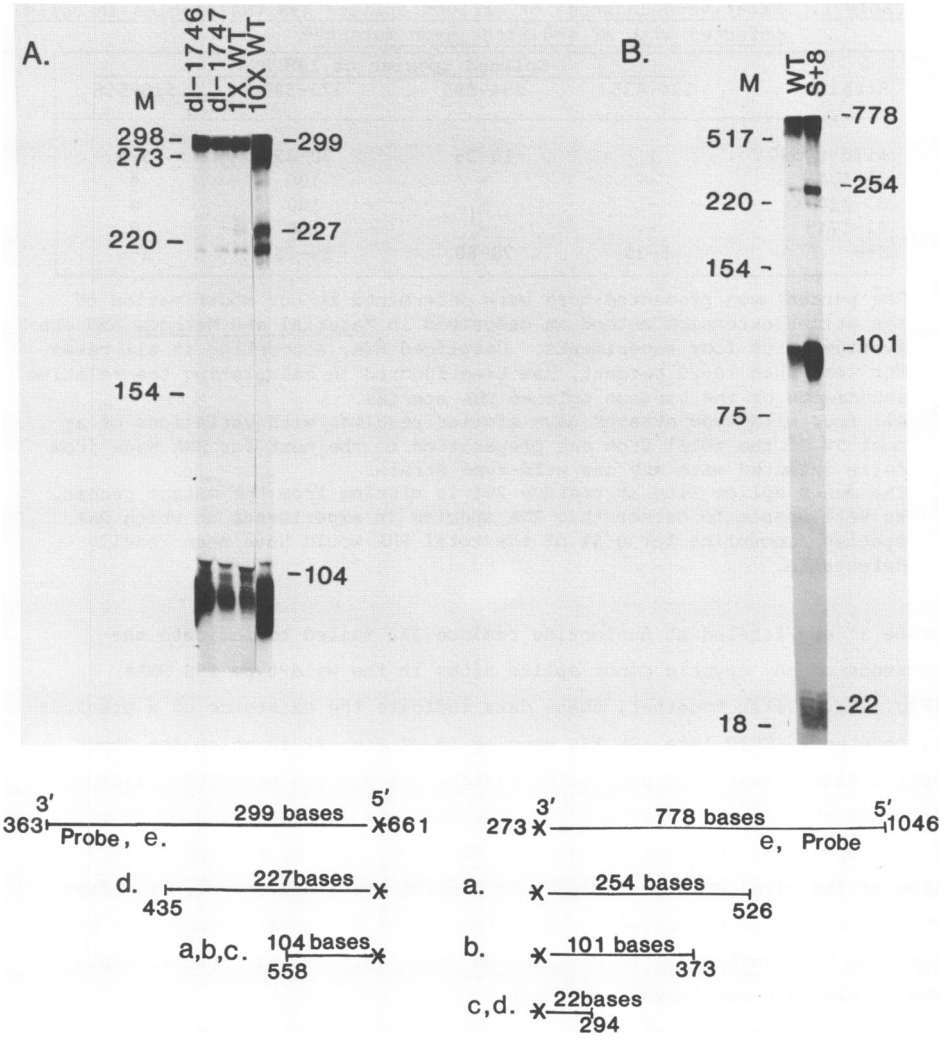

FIG. 4. S1 nuclease mapping of the late 19S viral RNAs made from wild-type strain 776 and several leader region mutants of SV40. Forty two hours after infection with the indicated mutant or WT strain of SV40, cells were harvested and cytoplasmic RNAs were purified and analyzed as described in Materials and Methods with the probes depicted in the lower part of the figure. (A) Mapping of acceptor splice sites with a late 19S-specific 5' end-labeled probe. The band at approximately 214 bases comes from unspliced RNA with a 5' end at residue 448. (B) Mapping of donor splice sites with a 3' end-labeled probe. The lower case letters indicated in the schematic diagrams correspond to the 19S RNA species designated in Fig. 2.

Table 1. Relative abundances of various spliced 19S RNA species in cells infected with WT and first exon mutants[a]

Strains	Spliced species of 19S RNA			
	294-435	294-558	373-588	526-558
Wild-type[b]	1	10-25	70-85	3-5
dl-1746	-[c]	-	100	d
dl-1747	-	-	100	d
dl-1748	-	-	100	d
S+8	5-10	70-80	10-15	3-5

[a] The percentages presented here were determined by our modification of the primer extension method as described in Material and Methods and are averages from four experiments. Unspliced RNA, accounting in all cases for less than 10-20 percent, has been ignored in calculating the relative abundances of the various spliced RNA species.

[b] All four wild-type strains gave similar results, with variations of at most 5% of the total from one preparation to the next for RNA made from cells infected with any one wild-type strain.

[c] The donor splice site at residue 294 is missing from the mutant genome.

[d] We were unable to detect this RNA species in experiments in which RNA species accounting for 0.5% of the total RNA would have been readily detectable.

probe 3' end-labeled at nucleotide residue 273 failed to indicate the presence of any cryptic donor splice sites in the wild-type 19S RNAs (Fig. 4B). Taken together, these data indicate the existence of a previously undetected SV40 late 19S RNA species (d of Fig. 2) in which the donor splice site situated at nucleotide residue 294 and the acceptor splice site situated at nucleotide residue 435 were utilized in processing of WT primary transcripts. S1 nuclease analysis of the cDNAs made by reverse transcription of the viral RNAs confirmed this finding (data not shown). The fact that this species accounted for less than 1% of the total cytoplasmic SV40 late 19S RNA synthesized in WT-infected cells (Fig. 3 and Table 1) explains why it had been overlooked previously.

The cDNAs to RNAs that have 5' ends mapping downstream of residue 273 do not get cleaved with endonucleases PvuII or SphI (Fig. 2). Therefore, their sizes remain unchanged following treatment with these enzymes. For example, the 19S RNAs with 5' ends mapping to the major cap site at residue 325 are predicted to yield cDNAs 337, 306, and 153 bases in length, respectively, from the unspliced (e), 526-558 spliced (a) and 373-558 spliced (b) 19S RNA species (Fig. 2). The data presented in Figure 3 confirm this prediction. In addition, they show that the cDNAs that have 5' ends mapping between residues 273 to 200 get cleaved with endonuclease PvuII, but not with endonuclease SphI. Densitometric analysis of autoradiograms of gels

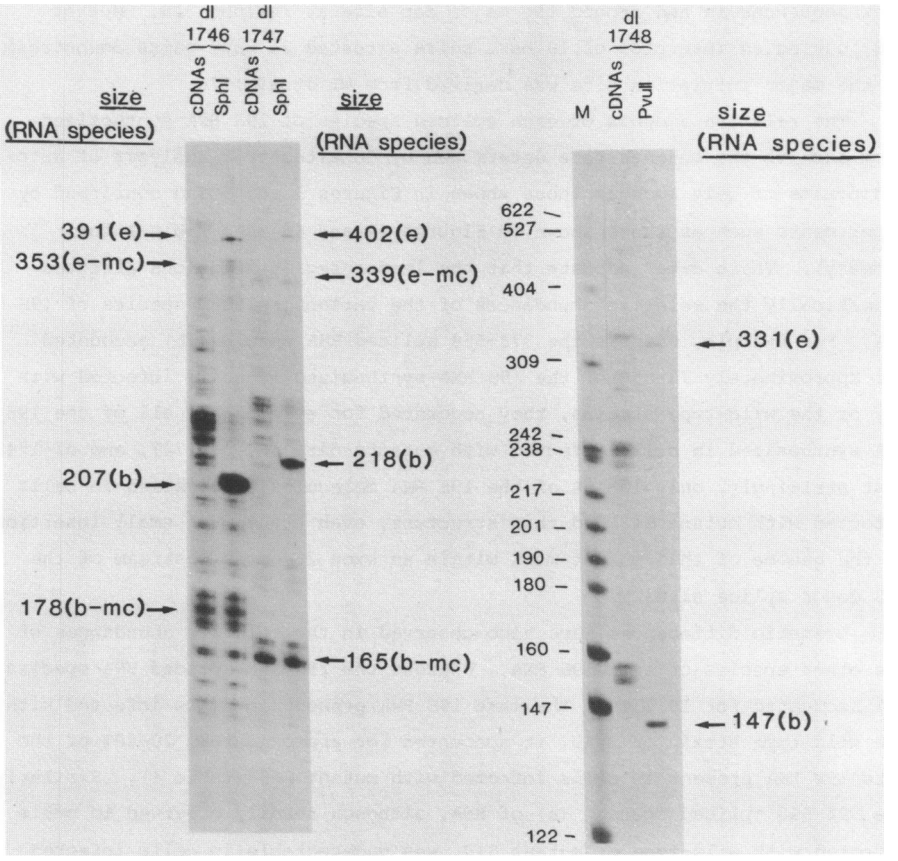

FIG. 5. Autoradiogram showing the lengths before and after cleavage with PvuII or SphI of cDNAs synthesized from SV40 late RNA molecules obtained from the cytoplasm of CV-1P cells infected with the leader region mutants dl-1746, dl-1747, and dl-1748. The cDNAs were synthesized and analyzed as in Fig. 3. The major cap site in these mutants maps to residues 264 (dl-1747; dl-1748) and 239 (dl-1746) rather than to residue 325.

such as those shown in Figure 3 enabled the relative abundances of the various spliced 19S RNA species to be quantified (Table 1).

Relative utilization of splice sites in late 19S RNAs synthesized in cells infected with leader region mutants of SV40

Figures 3 and 5 show experiments in which our modification of the primer extension method was used to analyze the 19S RNAs made in monkey cells infected with each of the four leader region mutants depicted in Figure 1. The deletion mutants dl-1746, dl-1747, and dl-1748, derived from WT800, all

lack sequences in and around the major cap site at residue 325. Mutant S+8, having an insertion of 10 base pairs situated 24 base pairs downstream of the major initiation site was derived from WT strain 776.

The relative amounts of each spliced species of 19S RNA synthesized from each of the mutants were determined by densitometric analysis of autoradiograms of gels such as those shown in Figures 3 and 5 and confirmed by experiments such as those shown in Figures 4A and 4B (see Table 1 for summary). These data indicate that the leader region mutations affected dramatically the relative abundances of the various spliced species of 19S RNA. For example, whereas the 373-558 spliced RNA species (b) accounted for approximately 70-85% of the 19S RNA synthesized in cells infected with any of the wild-type strains, they accounted for essentially all of the 19S RNA synthesized in cells infected with mutants dl-1746, dl-1747, and dl-1748. Most strikingly, only 10-15% of the 19S RNA molecules synthesized in cells infected with mutant S+8 had this structure, even though the small insertion in the genome of this mutant maps within an exon 26 bases upstream of the 373 donor splice site.

Dramatic differences were also observed in the relative abundances of the other species of late 19S RNA. Whereas the 294-558 spliced RNA species (c) accounted for 10-25% of the late 19S RNA present in cells infected with the wild-type strain of SV40, it accounted for approximately 70-80% of the late 19S RNA present in cells infected with mutant S+8 (Table 1). Similarly, the 526-558 spliced species (a) of RNA, although readily observed in cells infected with wild-type or mutant S+8, was undetectable in cells infected with mutants dl-1746, dl-1747, and dl-1748. Lastly, the newly discovered 294-435 spliced species of RNA, which accounted for less than one percent of the late 19S RNA present in WT-infected cells, accounted for 5-10% of the 19S RNA present in cells infected with mutant S+8.

Relative utilization of splice sites in 16S RNA

Figure 1 shows the structure of the probe used for determining the relative abundances of the two classes of late 16S RNA molecules found in cells infected with the various wild-type and mutant strains of SV40 studied here. Figure 6 shows an example of one of the S1 nuclease mapping experiments from which the data summarized in Table 2 were obtained. Whereas the doubly spliced 16S RNA species accounted for 15-25% of the 16S RNA present in cells infected with the wild-type strains of SV40, they accounted for 50-60% and none, respectively, of the 16S RNA present in cells infected with mutant S+8 and any of the three deletion mutants. This latter observa-

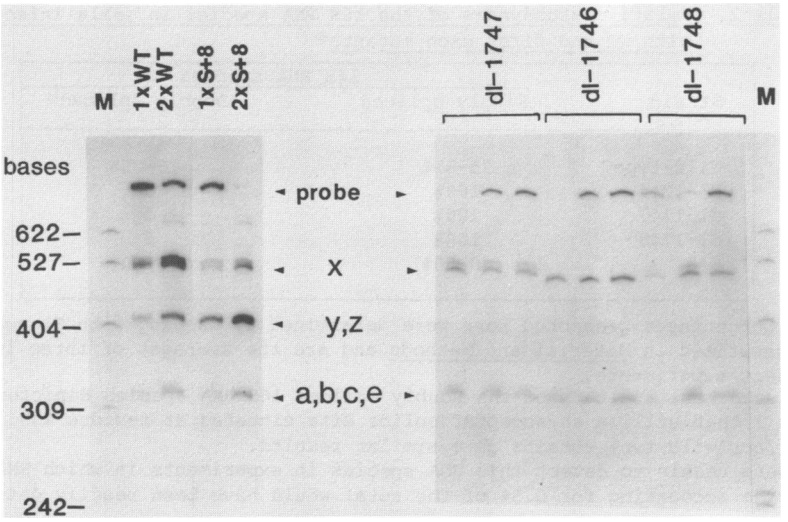

FIG. 6. S1 nuclease mapping of the late 16S viral RNAs made from wild-type strain 776 and several leader region mutants of SV40. CV-1P cells were infected with wild-type strain 776 and each of the mutants depicted in Fig. 1. At 42 h after infection, the cells were harvested and cytoplasmic RNAs were purified and analyzed as described in Material and Methods with the probe depicted in Fig. 1. Lane M contained MspI-digested, 5' end-labeled pBR322 DNA.

tion is particularly intriguing in view of the fact that mutants dl-1746, dl-1747, and dl-1748, although lacking the nucleotides around the 294 donor splice site, might still theoretically have synthesized the doubly spliced 16S RNA in which the acceptor site at residue 435 joins with the donor splice site at residue 526 (16S RNA species z of Fig. 1).

Ratio of 16S to 19S late RNA present in cells infected with the wild-type and mutant strains of SV40

S1 nuclease mapping experiments of the type shown in Figure 6 also enabled us to determine the relative abundances of 16S versus 19S late viral RNA present in our mutant-infected cells. We observed that the 16S/19S molar ratios of the late viral RNAs made by wild-type and S+8 were approximately 4/1. On the other hand, we found that the 16S/19S molar ratios of the late viral RNAs obtained from cells infected with the deletion mutants dl-1746, dl-1747, and dl-1748 were approximately 1-2/1. Therefore, the leader region mutations affected the relative utilization of the two most downstream acceptor splice sites as well. These observations confirm the findings of Villarreal and Berg (19) and Barkan (20) that cells infected

Table 2. Relative abundances of the 16S RNA species in cells infected with WT and first exon mutants[a]

Strain	16S RNA species	
	Singly spliced	Doubly spliced[b]
Wild-type[c]	75-85%	15-25%
dl-1746	100%	d
dl-1747	100%	d
dl-1748	100%	d
S+8	40-50%	50-60%

[a] The percentages presented here were determined by mapping with S1 nuclease as described in Material and Methods and are the averages of three independent experiments.

[b] Included here are both of the doubly spliced 16S RNA species depicted in Fig. 1 that utilize an acceptor splice site situated at residue 435.

[c] All four wild-type strains gave similar results.

[d] We were unable to detect this RNA species in experiments in which RNA species accounting for 0.5% of the total would have been readily detectable.

with mutants containing deletions within the late leader region of SV40 can produce late 19S-to-16S RNA at ratios different from those seen in WT-infected cells.

DISCUSSION

Sequences within the first exon affect the choice of splice sites utilized in the synthesis of the SV40 late RNAs

The primary aim of the studies reported here was to examine how mutations in an exon can affect the utilization of downstream donor and acceptor splice sites. Our data showing that the 373-558 spliced species of RNA is dramatically underrepresented in cells infected with mutant S+8 (Table 1) demonstrate that a small alteration within an exon can affect drastically the utilization of a 5' distal donor splice site.

In further support of this conclusion was our failure to detect RNA species lacking residues 526-558 in the late RNAs made from the deletion mutants dl-1746, dl-1747, and dl-1748. This observation, taken together with our finding that these deletion mutants also underproduce 16S RNA, indicates that sequences mapping between residues 270 and 340 within the first exon are required for efficient utilization of a donor splice site (526) situated greater than 180 nucleotides away.

Our results concerning the structures of the late RNAs present in WT-infected cells provide additional support for this latter conclusion and

also indicate that these observations are not an artifact of the mutations. Both we (Fig. 6 and Table 2) and others (21) can detect readily in WT-infected cells the 16S doubly spliced RNA species lacking both residues 295-434 and 527-1642. Data presented here (Figs. 3 and 4A) show the existence of an RNA species lacking only residues 295-434. One would, theoretically, expect a doubly spliced 19S RNA species lacking residues 294-435 and 526-558 to exist also. Nevertheless, we were unable to detect it. These findings indicate that sequences missing from both the deletion mutants and the 294-435 spliced RNAs are required for efficient splicing of residue 526 to 558.

Utilization of other splice sites was also affected by the mutations studied here. For example, mutant S+8 overutilized the residue 435 acceptor splice site in both 19S and 16S RNA synthesis (Tables 1 and 2). In contrast, the deletion mutants all failed to utilize this splice site. Therefore, sequences mapping within residues 270 and 373 are important for selection of the acceptor splice site at residue 435.

Data presented here also demonstrate that the small alteration in S+8, not only reduces the utilization of the downstream donor splice site at residue 373, but also increases the utilization of the upstream donor splice site at residue 294 such that the alteration becomes part of the excised intervening sequence. This finding suggests a novel mechanism by which expression of genes can be regulated without altering the structure of the product being synthesized.

What determines selection of the donor splice site utilized in synthesis of the SV40 late 19S RNAs?

Data presented here and elsewhere (10a) show clearly that the 5' ends of the late mRNAs made from the mutants studied here differ quantitatively, but not qualitatively, from those made from WT. For example, whereas the major transcription initiation site in the wild-type 19S RNAs mapped to residue 325, the major 5' end in the 19S RNAs of mutants dl-1746, dl-1747, dl-1748, and S+8 mapped to residues 239, 264, 264, and 167, respectively (Figs. 3 and 5). Most of the late 19S RNAs made from WT SV40 and the three deletion mutants was spliced using the 373 donor splice site (Table 1). In contrast, most of the 19S RNA made from mutant S+8 was spliced using the donor splice site at residue 294. Therefore, one alternative explanation for the changes observed in splicing of the transcripts made from this latter mutant is that the small insertion in the viral genome affects transcription initiation and/or termination, thereby leading to primary transcripts

that differ considerably in sequence from the majority of those made from WT SV40 and that, consequently, are processed differently. Regardless of the reason, these findings provide further support for the hypothesis that sequences at or near the 5' end of a primary transcript are responsible in part for determining which of the three potential donor splice sites is utilized in RNA processing.

Another plausible explanation is that the leader-encoded protein LP1 (22), also called agnoprotein (23,24) which is not made by mutant S+8 because of the frameshift in the translational reading frame, influences directly or indirectly the choice of splice sites used in processing. However, this is unlikely since the mutant dl-1748, which cannot synthesize LP1 because it is missing its translation initiation codon at nucleotide residues 335-337, synthesizes RNAs similar in structure to those made by dl-1747 (Table 1), which theoretically can still encode LP1.

The data presented here also indicate that mutants lacking the 294 donor splice site utilize exclusively the 373 splice site in the synthesis of 19S RNA (Table 1). Similarly, mutants lacking the 373 donor splice site utilize exclusively the 294 donor splice site (M.B.S., unpublished). In no case is the donor splice site at residue 526 used efficiently for the synthesis of 19S RNA. Therefore, although most transcripts are processed utilizing the donor splice site at residue 526, this donor splice site pairs almost exclusively with the acceptor splice site at residue 1463 to make 16S RNA, skipping over the nearby acceptor splice site at residue 558 that some models (e.g., 25) for splice site pairing would predict should have been preferentially used.

The relative efficiencies of utilization of splice sites for synthesis of the 16S and 19S RNAs are affected similarly by sequence alterations within the late leader region

We (Table 1) have shown that 10-25% of WT 19S RNA molecules are synthesized utilizing the donor splice site at residue 294. Mutant S+8, which exhibited an increased utilization relative to wild-type SV40 of the residue 294 donor splice site in the synthesis of 19S RNA (Table 1) also showed an increased utilization of the residue 435 acceptor splice site in the synthesis of 16S RNA (Table 2). Similarly, mutants dl-1746, dl-1747 and dl-1748, which lack sequences around donor splice site 294 needed for synthesis of the 294-435 spliced species of 19S RNA, failed to utilize the acceptor splice site at residue 435 for the synthesis of 16S RNA even though this site could theoretically combine with the 526 donor splice site. Therefore, the selec-

tion of splice sites utilized in the synthesis of both the 16S and 19S RNAs are affected similarly by these alterations. Possible reasons for this finding are: (i) changes in the sequences at or near the 5' ends of the initial transcripts affect the selection of splice sites used in the synthesis of the RNAs; (ii) intermediates in synthesis of the 19S and 16S RNAs are affected similarly in wild-type and the mutants by as yet undefined factors that govern the selection of donor and acceptor splice sites; and (iii) sequences between residues 270 and 373 play a role in some aspect of the splicing reaction.

The newly identified RNA species lacking residues 295-434 may be an intermediate in the synthesis of 16S RNA

It has been suggested previously (26-29) that SV40 late 16S RNA may be synthesized from a 19S precursor. The RNA species lacking nucleotide residues 295-434 accounted for 5-10% of the 19S RNA made from mutant S+8, whereas it accounted for at most 1% of the 19S RNA made from WT SV40 (Table 1). Preliminary results from pulse and pulse-chase experiments indicate that monkey cells infected with mutant S+8, although unaffected in their rate of synthesis of SV40 late strand transcripts, process these transcripts more slowly than WT-infected cells (M.B.S., unpublished). If the RNA species lacking residues 295-434 were an intermediate in the synthesis of 16S RNA, one would predict as was observed that it would be overrepresented in cells infected with a mutant such as S+8 whose RNAs are processed slowly. These findings are reminiscent of those made by Ross and Knecht (30) who demonstrated the existence of at least two intermediate size classes of RNA involved in the synthesis of β-globin mRNA.

Applications of a novel modification of the primer extension method

The data presented here show that our modification of the primer extension method is both less labor intensive and more quantitative for the determination of the relative amounts of various spliced species of RNAs because it involves neither sequencing numerous cDNAs nor the summation of various cDNAs copied from RNAs with heterogeneous 5' termini. This method has also enabled us to detect a previously unidentified spliced species of late RNA which represents less than one percent of the viral 19S RNA present in WT-infected cells. We have also found late 19S RNA species synthesized from other late leader region mutants of SV40 in which cryptic donor and acceptor splice sites at nucleotide residues 321 and 499, respectively, are utilized (M.B.S. and J.E.M., manuscript in preparation). Additional advantages of this method are: (i) it is fairly easy to perform; (ii) it does not require

large amounts of RNA; and (iii) large numbers of samples can be analyzed simultaneously.

In the experiments presented here we utilized a strand separated restriction fragment for the hybridization to cDNA molecules obtained after the reverse transcriptase reaction (see Material and Methods). Recently, we have found that phage M13 cloned DNA of SV40 can be used satisfactorily for this purpose. Lastly, although we have not yet done so, it should also be possible to use synthetic oligonucleotides complementary to the sequence to be cleaved. This latter method would enable one to be readily in molar excess and to obtain great specificity as to the site of cleavage of the cDNAs.

ACKNOWLEDGMENTS

We wish to thank Mike Fromm for pS+8 and Peter Good, David Reisman, Jeff Ross, Richard Spritz, and Howard Temin for helpful comments on the manuscript. This work was supported by Public Health Service research grants CA-07175 and CA-22443 from the National Cancer Institute and MV-201 from the American Cancer Society.

*Present address: Department of Biochemistry, University of Wisconsin, Madison, WI 53706, USA

REFERENCES

1. Wieringa, B., Meyer, F., Reiser, J., and Weissmann, C. (1983) Nature (London) 301, 38-43.
2. Mount, S.M. (1982) Nucl. Acids Res. 10, 459-472.
3. Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L., and Steitz, J.A. (1980) Nature 283, 220-224.
4. Padgett, R.A., Mount, S.M., Steitz, J.A., and Sharp, P.A. (1983) Cell 35, 101-107.
5. Moore, C.L., and Sharp, P.A. (1984) Cell 36, 581-591.
6. Contreras, R., and Fiers, W. (1981) Nucl. Acids Res. 9, 215-236.
7. Contreras, R., Gheysen, D., Knowland, J., van de Voorde, A., and Fiers, W. (1982) Nature (London) 300, 500-505.
8. Ghosh, P.K., Reddy, V.B., Swinscoe, J., Lebowitz, P., and Weissman, S.M. (1978) J. Mol. Biol. 126, 813-846.
9. Gidoni, D., Kahana, C., Canaani, D., and Groner, Y. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 2174-2178.
10. Piatak, M., Ghosh, P.K., Norkin, L.C., and Weissman, S.M. (1983) J. Virol. 48, 503-520.
10a. Somasekhar, M.B., and Mertz, J.E. (1985) J. Virol., in press.
11. Lebowitz, P., and Weissman, S.M. (1979) Curr. Top. Microbiol. Immunol. 87, 43-172.
12. Tooze, J., Ed. (1981) Molecular Biology of Tumor Viruses, Part 2, DNA Tumor Viruses, 2nd ed., revised, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
13. Ghosh, P. K., Piatak, M., Mertz, J.E., Weissman, S.M., and Lebowitz, P. (1982) J. Virol. 44, 610-624.

14. Piatak, M., Subramanian, K.N., Roy, P., and Weissman, S.M. (1981) J. Mol. Biol. 153, 589-618.
15. Mertz, J.E., and Berg, P. (1974) Virology 62, 112-124.
16. Fromm, M., and Berg, P. (1982) J. Mol. Appl. Genet. 1, 457-481.
17. Maxam, A.M., and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
18. Sanger, F. and Coulson, A. R. (1978) FEBS Lett. 87, 107-110.
19. Villarreal, L.P., White, R.T., and Berg, P. (1979) J. Virol. 29, 209-219.
20. Barkan, A. (1983) Ph.D. Thesis. University of Wisconsin, Madison, Wisconsin.
21. Ghosh, P.K., Reddy, V.B., Swinscoe, J., Choudary, P.V., Lebowitz, P., and Weissman, S.M. (1978) J. Biol. Chem. 253, 3643-3647.
22. Mertz, J.E., Murphy, A., and Barkan, A. (1983) J. Virol. 45, 36-46.
23. Jackson, V., and Chalkley, R. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 6081-6085.
24. Jay, G., Nomura, S., Anderson, C.W., and Khoury, G. (1981) Nature (London) 291, 346-349.
25. Lang, K.M., and Spritz, R.A. (1983) Science 220, 1351-1355.
26. Weinberg, R.A., Ben-Ishai, Z., and Newbold, J.E. (1974) J. Virol. 13, 1263-1273.
27. Aloni, Y., Shani, M., and Reuveni, Y. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 2587-2591.
28. Khoury, G., Carter, B.J., Ferdinand, F.-J., Howley, P.M., Brown, M., and Martin, M.A. (1976) J. Virol. 17, 832-840.
29. May, E., Maizel, J.V., and Salzman, N.P. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 496-500.
30. Ross, J., and Knecht, D.A. (1978) J. Mol. Biol. 119, 1-20.
31. Buchman, A.R., Burnett, L., and Berg, P. (1981) in Molecular Biology of Tumor Viruses, Tooze, J., Ed., Part 2, DNA Tumor Viruses, 2nd ed., revised, pp. 799-841, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
32. Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P., and Chambon, P. (1981) Nucl. Acids Res. 9, 6047-6068.
33. Mathis, D.J., and Chambon, P. (1981) Nature 290, 310-315.