JOURNAL OF VIROLOGY, Aug. 1979, p. 437-436
0022-538X/79/08-0437/10$02.00/0

Vol. 31, No. 2

Characterization of the 5'-Terminal Structure of Simian Virus 40 Early mRNA's

JOHN A. THOMPSON,* MICHAEL F. RADONOVICH, AND NORMAN P. SALZMAN

Laboratory of Biology of Viruses, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20014

Received for publication 28 March 1979

RPC-5 reverse-phase chromatography has been used to isolate fragments of simian virus 40 DNA generated by appropriate digestions with restriction endonucleases. Ten specific DNA fragments, mapping successively in a counterclockwise direction from 0.67 to 0.515 on the simian virus 40 genome, were each hybridized to cytoplasmic mRNA obtained during the early phase of simian virus 40 infection. Primer extension methods with reverse transcriptase were used to characterize the 5' ends of two species of viral mRNA which were fractionated on sucrose gradients. Analysis of the complementary DNA products demonstrated the presence of two different spliced structures of simian virus 40 early mRNA's, both of which had the same 5'-end sequences (AUU), located at residues 18 to 20 on the viral genome. The mRNA for small-t contained a segment 588 bases in length (residues 18 to 605) spliced to residue 672. A 66-nucleotide segment rich in adenine-thymine was spliced out of this mRNA. The mRNA for large-T contained a segment 308 bases in length (residues 18 to 325) which is also spliced to residue 672. A 346-base segment was spliced from this mRNA. The results suggest that there are two levels for control of genetic expression. One would be the regulation of initiation of transcription at a common promoter; the other involves post-transcriptional splicing.

It has been found that cells infected with simian virus 40 (SV40) produce two viral proteins both immunoprecipitable with anti-tumor (T) serum. Both of these proteins are found early in the lytic cycle of SV40 infection and are termed small-t (19,000 daltons) and large-T (94,000 daltons) antigens (18). Using in vitro translation of gradient-fractionated cytoplasmic mRNA, it was found that the mRNA which codes for small-t sedimented more rapidly than the mRNA coding for large-T (15, 16a). The general structure of the two early SV40 mRNA's has been determined by enzymatic degradation of hybrids formed between the mRNA molecule and labeled viral DNA (1). These mRNA's, coding for large-T and small-t antigens, were found to be spliced; that is, during their biogenesis, appropriate regions of the primary transcripts have been deleted, which accounts for the relative sizes of the two early viral antigens. Many recent discoveries also have shown that most eucaryotic, as well as viral, mRNA's are generated by splicing of primary transcripts (24, 25). These results suggest that further characterization of mRNA structure is essential to understand the mechanisms of gene regulation.

The present work was undertaken to characterize more specifically the SV40 mRNA's coding for both small-t and large-T. Using primer extension methodology, we have determined the 5'-terminal structure of both of these messages. The sequence around the spliced junctions has been determined on the basis of the present work, and our findings are in excellent agreement with other recent studies (1, 8, 19).

MATERIALS AND METHODS

Isolation of SV40 DNA restriction fragments. SV40 DNA (form I) was extracted from virions (4) and further purified by isopycnic banding in cesium chloride and ethidium bromide (22). The various restriction fragments (Fig. 1) were generated by either digestion of SV DNA I or redigestion of isolated fragments with the appropriate restriction endonucleases (Bethesda Research Lab). The restriction fragments were fractionated by RPC-5 column chromatography (6, 11, 12). Specific restriction fragments were dephosphorylated with alkaline phosphatase (Worthington Biochemicals Corp.) and 5'-end labeled with [γ-^{32}P]ATP (specific activity, 1,000 to 3,000 Ci/mmol; New England Nuclear Corp.) by polynucleotide kinase (New England Biolabs) (14).

Isolation of early cytoplasmic mRNA. Subconfluent monolayers (80 to 90%) of BSC-1 cells were infected for 12 h at 37°C with wild-type SV40, strain 776 (5 to 10 PFU/cell). The cytoplasmic fraction was prepared using 0.5% Nonidet P-40 and was extracted with phenol-chloroform (1:1) containing 1% sodium

dodecyl sulfate. After three extractions with guanidine hydrochloride, the polyadenylic acid-containing mRNA's were selected on polyuridylic acid-Sepharose 4B (Pharmacia Fine Chemicals) (5).

Primer extension methodology. Under conditions which would maximize RNA-DNA hybrids (3), 5 μg of a specific DNA restriction fragment was annealed to total polyadenylic acid-containing mRNA, which contained 1 μg of virus-specific RNA as determined by filter hybridizations (23). For a partial extension reaction, the DNA-RNA hybrids were dissolved in 0.05 M Tris-hydrochloride (pH 8.3), 0.1 M KCl, 0.01 M $MgCl_2$, 0.015 M β-mercaptoethanol, and 0.05 mCi of the specific α-^{32}P-labeled deoxynucleotide triphosphates (specific activity, 350 Ci/mmol; Amersham Corp.) listed in Table 1. Reverse transcriptase, 100 U, (J. W. Beard, Life Sciences, Inc.) was added, and the reaction was incubated for 5 min at 40°C.

After purification of the partial reactions on sucrose gradients, the [^{32}P]DNA-RNA hybrids were dissolved in 0.05 M Tris-hydrochloride (pH 8.3), 0.1 M KCl, 0.01 M $MgCl_2$, 0.015 M β-mercaptoethanol, and 500 μg each of dATP, dCTP, dGTP, and dTTP per ml. For this total extension reaction, 100 U of reverse transcriptase were added, and the reaction mixture was incubated for 30 min at 40°C.

An aliquot of both the partial reaction and the total extension reaction were made 0.3 M in sodium hydroxide and heated for 10 min at 100°C. Both the ^{32}P-labeled primer and complementary DNA (cDNA) products were purified on DEAE-cellulose 52 columns and ethanol precipitated. Each sample was dissolved in 7 M urea–0.25 mM EDTA and loaded onto standard acrylamide-urea gels (14, 21). In separate reactions, a mixture of SV40 DNA restriction fragments from an *Alu*I digest and a *Hin*III digest were dephosphorylated with alkaline phosphatase and ^{32}P-5′-end labeled with polynucleotide kinase (14). These ^{32}P-end-labeled DNA fragments were used as sizing markers in the gel analysis of the cDNA products (2).

TABLE 1. *Partial reverse transcriptase reactions on the DNA-RNA hybrids sedimenting at 19S*

Primer	[α-^{32}P]-dNTP[a]	Theoretical no. of bases adding (residue no.)[b]	Theoretical ^{32}P-labeled primer length (no. of bases)[c]
M1	C, T, A	4 (15–12)	58
M3	C, T	6 (69–64)	52
E1	C, T, G	4 (115–112)	35
E2	C, A, G	8 (151–144)	208
H1	C, A, G	6 (351–346)	34
H2	C, A, T	5 (379–375)	129
Ts	C, G, A	8 (503–496)	103
G	A	2 (674–673)	111
G	A, T	3 (674–672)	112
G	A, T, C	23 (674–652)	132

[a] Specific α-^{32}P-labeled deoxynucleotide triphosphate (dNTP) included in the partial reverse transcriptase reaction where A, G, C, and T represent deoxyadenosine, deoxyguanosine, deoxycytosine, and deoxythymidine triphosphate, respectively.

[b] Based on the published nucleotide sequence (7).

[c] Determined by adding the size of the primer to the theoretical number of bases which should be added during the partial reaction.

Endonuclease S_1 digestion. After a partial reverse transcriptase reaction and purification on a sucrose gradient, [^{32}P]DNA-RNA hybrids were ethanol precipitated and dissolved in 3.0 ml containing 0.1 μmol of $ZnCl_2$, 100 μmol of sodium acetate buffer (pH 4.5), 0.3 M NaCl, and 5 μmol of calf thymus DNA (heat denatured). A 0.45-μg amount of nuclease S_1 (Miles Laboratories, Inc.), purified by a modification of the method of Vogt (26), was added, and the reaction was incubated for 30 min at 37°C. S_1-resistant material was precipitated with ethanol and analyzed on acrylamide-urea gels.

RESULTS

Localization of DNA fragments used to probe early SV40 mRNA. Figure 1 shows the physical map of SV40 DNA and the localization of the ten specific DNA restriction fragments which were generated by digestion with appropriate restriction enzyme cleavages and purified by RPC-5 reverse-phase chromatography. These DNA fragments map successively from 0.67 to 0.515 on the SV40 genome. Collectively, these fragments define the DNA segment which is located at the 5′ end of the early region and would be transcribed 768 bases counterclockwise around the viral genome. According to the nucleotide sequence (7), fragments M1, M3, E1, E2, H1, H2, Ts, C1, I, and G contain residues 16 to 69, 70 to 115, 116 to 146, 152 to 351, 352 to 379, 380 to 503, 504 to 598, 599 to 650, 651 to 674, and 675 to 783, respectively. This nomenclature defines the strand of the specific DNA fragment complementary to the viral mRNA based on the cleavage site of the appropriate enzymes (Fig. 1).

Hybridization of specific DNA restriction fragments to early cytoplasmic SV40 mRNA. The specific DNA fragments, generated by restriction enzyme cleavage of SV40 DNA and 5′-end labeled with [γ-^{32}P]ATP, were annealed to total polyadenylic acid-containing cytoplasmic mRNA isolated early during the SV40 lytic cycle. The [^{32}P]DNA-RNA hybrids that were fractionated on sucrose gradients sedimented around 19S, the size of the major species of early SV40 mRNA (Fig. 2). The large amount of ^{32}P-labeled material sedimenting at 4S represents the late DNA strand, which is not expected to hybridize with early mRNA, as well as any unhybridized DNA fragment. When fragment M1 (residues 16 to 69) was annealed to cytoplasmic mRNA and fractionated in a sucrose gradient, two major species of hybrids were ob-

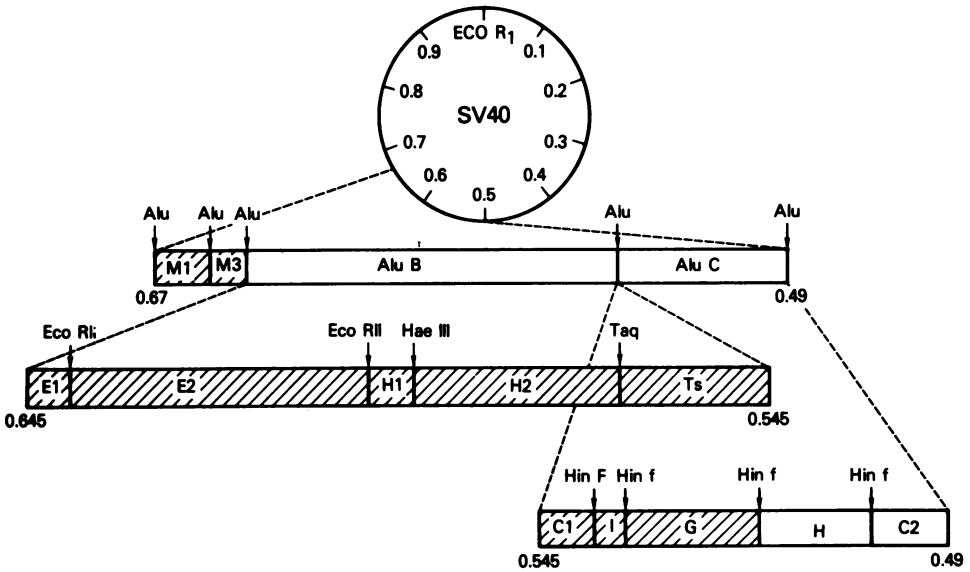

FIG. 1. *Physical map of the SV40 genome and localization of the specific DNA probes. An AluI digest of SV40 DNA generated fragments M1, M3, AluB, and AluC. Redigestion of the AluB fragment with EcoRII, HaeIII, and TaqI generated fragments E1, E2, H1, H2, and Ts. Redigestion of the AluC fragment with HinF generated fragments C1, I, G, H, and C2. RPC-5 reverse-phase chromatography was used to isolate and purify DNA fragments M1, M3, E1, E2, H1, H2, Ts, C1, I, and G containing 54, 46, 31, 200, 28, 124, 95, 52, 24, and 109 base pairs, respectively.*

served, as evidenced by the two peaks around 19S (Fig. 2A). When ^{32}P-labeled DNA restriction fragments M3 (residues 70 to 115), E1 (residues 116 to 146), and E2 (residues 152 to 351) were annealed to total cytoplasmic mRNA, two peaks of hybridized [^{32}P]DNA around 19S were observed, generating the same gradient profile as that found with M1 (Fig. 2A). Regardless of which of these four fragments was hybridized, the slower-sedimenting species of the [^{32}P]DNA-RNA hybrid contained four to five times more ^{32}P-labeled material than the faster-sedimenting peak.

When ^{32}P-labeled restriction fragment H1 (residues 352 to 379) was annealed to cytoplasmic mRNA and fractionated in a sucrose gradient, only one species of hybrid was observed (Fig. 2B). Similar results were also obtained when ^{32}P-labeled fragments H2 (residues 380 to 503) and Ts (residues 504 to 598) were annealed to mRNA (data not shown). This single peak of ^{32}P-containing material corresponded to the faster-sedimenting species in Fig. 2A.

[^{32}P]DNA restriction fragments C1 (residues 599 to 650) and I (residues 651 to 674) did not hybridize to either species of mRNA. Since stable hybrids were formed with M1 (54 base pairs), M3 (46 base pairs), E1 (31 base pairs), and H1 (28 base pairs), failure to obtain hybrids was not the consequence of the small size of C1 (52 base pairs) and I (24 base pairs).

Fragment G (residues 675 to 783) gave two peaks of hybridized [^{32}P]DNA at the same positions and in the same relative amounts as was observed with fragments M1, M3, E1, and E2.

Extension of the specific DNA primers annealed to 19S mRNA. The 19S DNA-RNA hybrids isolated in sucrose gradients provide both the template (19S mRNA) and primer (specific DNA fragment) for reverse transcriptase, which extends the primer in a 5' to 3' direction by the sequential addition of complementary deoxynucleotide triphosphates. The cDNA product generated would have a 3' terminus which would correspond to the 5' end of the viral mRNA molecule. To generate a ^{32}P-labeled primer which could be analyzed, the DNA-RNA hybrid initially was used to prime a partial reverse transcriptase reaction in the presence of only one, two, or three specific α-^{32}P-labeled deoxynucleotide triphosphates. The specific ^{32}P-labeled substrates used in this limited reverse transcriptase reaction were chosen based on a knowledge of the nucleotide sequence of SV40 DNA (7, 20) adjacent to the 3' ends of different primers (Table 1). The [^{32}P]DNA-RNA hybrids generated by the partial reverse transcriptase reaction were purified by sucrose gradient centrifugation and subjected to a total reverse transcriptase reaction in the presence of saturating levels of all four unlabeled deoxynucleotide triphosphates. After extension of the ^{32}P-labeled

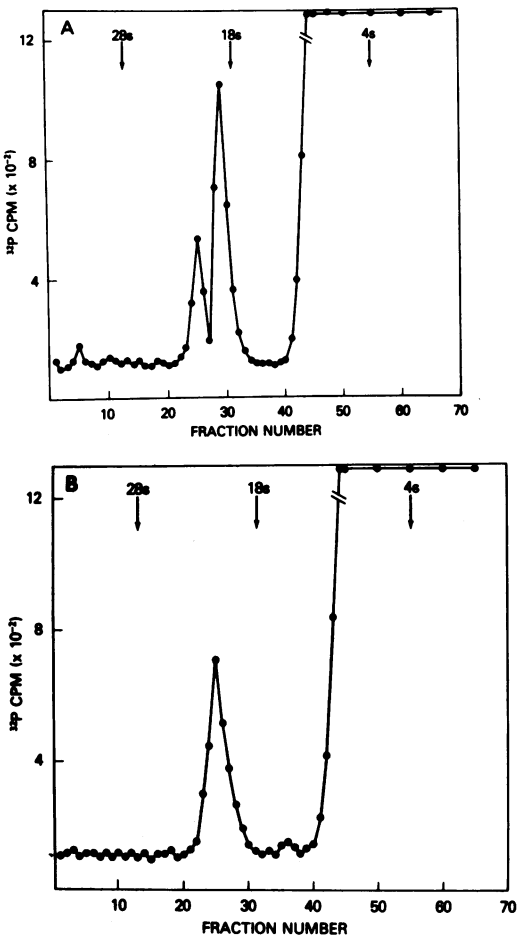

FIG. 2. *Sucrose gradient fractionation of SV40 mRNA-DNA hybrids. Early polyadenylic acid-containing cytoplasmic mRNA was annealed to specific DNA restriction fragments 5'-end labeled with [γ-^{32}P]ATP. The hybridization mixture was layered onto linear sucrose (15 to 40%) gradients in 0.35% sodium dodecyl sulfate-0.1 M NaCl-1 mM EDTA-0.01 M Tris-hydrochloride (pH 7.4) (16). The centrifugation was performed at 20°C for 16 h at 32,000 rpm in a Spinco SW41 rotor. Each fraction was counted for ^{32}P. ^{14}C-labeled rRNA markers were run simultaneously in separate gradients. (A) M1 fragment hybridized to the mRNA. (B) H1 fragment hybridized to the mRNA.*

primers, the primers and cDNA products were isolated and analyzed on denaturing polyacrylamide gels. By this procedure we were able to define precisely the location of discontinuities in the structure of the mRNA's.

Reverse transcriptase reactions using M1, M3, E1, and E2 (residues 16 to 351) DNA fragments as primers. Figure 3A shows the autoradiogram of the electrophoretic pattern of ^{32}P-labeled primers and the cDNA products obtained with the specific DNA fragments. The size of the [^{32}P]DNA primers generated by the partial reverse transcriptase reaction were found to be 52, 35, and 208 nucleotides for M3, E1, and E2, respectively (Fig. 3A). These results agreed with the theoretical size anticipated based on the nucleotide sequence (Table 1). Although each of these specific fragments hybridized to both species of 19S mRNA, each primer was extended to yield only one major species of cDNA (Fig. 3A). The chain lengths of the extended products (cDNA) using M3, E1, and E2 were 98, 129, and 330 nucleotides, respectively. Taken together, these results suggest that the 3' end of the cDNA terminates around residues 17 to 20. This would explain why the DNA primer M1 (residues 16 to 69) would not incorporate any ^{32}P-labeled deoxynucleotide triphosphates during a partial reverse transcriptase reaction (data not shown). Even in the presence of all four ^{32}P-labeled deoxynucleotide triphosphates, the M1 DNA fragment hybridized to 19S mRNA would not generate ^{32}P-labeled primer with reverse transcriptase.

Reverse transcriptase reactions using H1, H2, and Ts (residues 352 to 598) DNA fragments as primers. Figure 3A also shows the ^{32}P-labeled primers and cDNA products obtained when the specific DNA fragments, which hybridized to only the fast-sedimenting species of 19S mRNA, were used as a primer for the reverse transcriptase reactions. The sizes of the [^{32}P]DNA primer generated by a partial reverse transcriptase reaction for H1, H2, and Ts were found to be 34, 128, and 103 nucleotides, respectively (Fig. 3A). These results are in close agreement with the theoretical size expected, based on the nucleotide sequence (Table 1). As expected, since each of these fragments hybridized to only the fast-sedimenting species of 19S mRNA, each ^{32}P-labeled primer yielded only one major size of cDNA product after total primer extension techniques. The chain lengths of the cDNA product using H1, H2, and Ts were found to be 362, 485, and 582 nucleotides, respectively. These results place the 3' end of the cDNA product at residues 16 to 18 on the viral genome, which agrees with the results obtained using M3, E1, and E2 DNA fragments as probes.

Reverse transcriptase reactions using the G (residues 675 to 783) DNA fragment as a primer. Figure 3B shows the ^{32}P-labeled primers and cDNA extention products obtained using the restriction fragment G, which hybridized to both the fast- and slow-sedimenting species of 19S mRNA. When only [α-^{32}P]dATP was added to the partial reverse transcriptase reac-

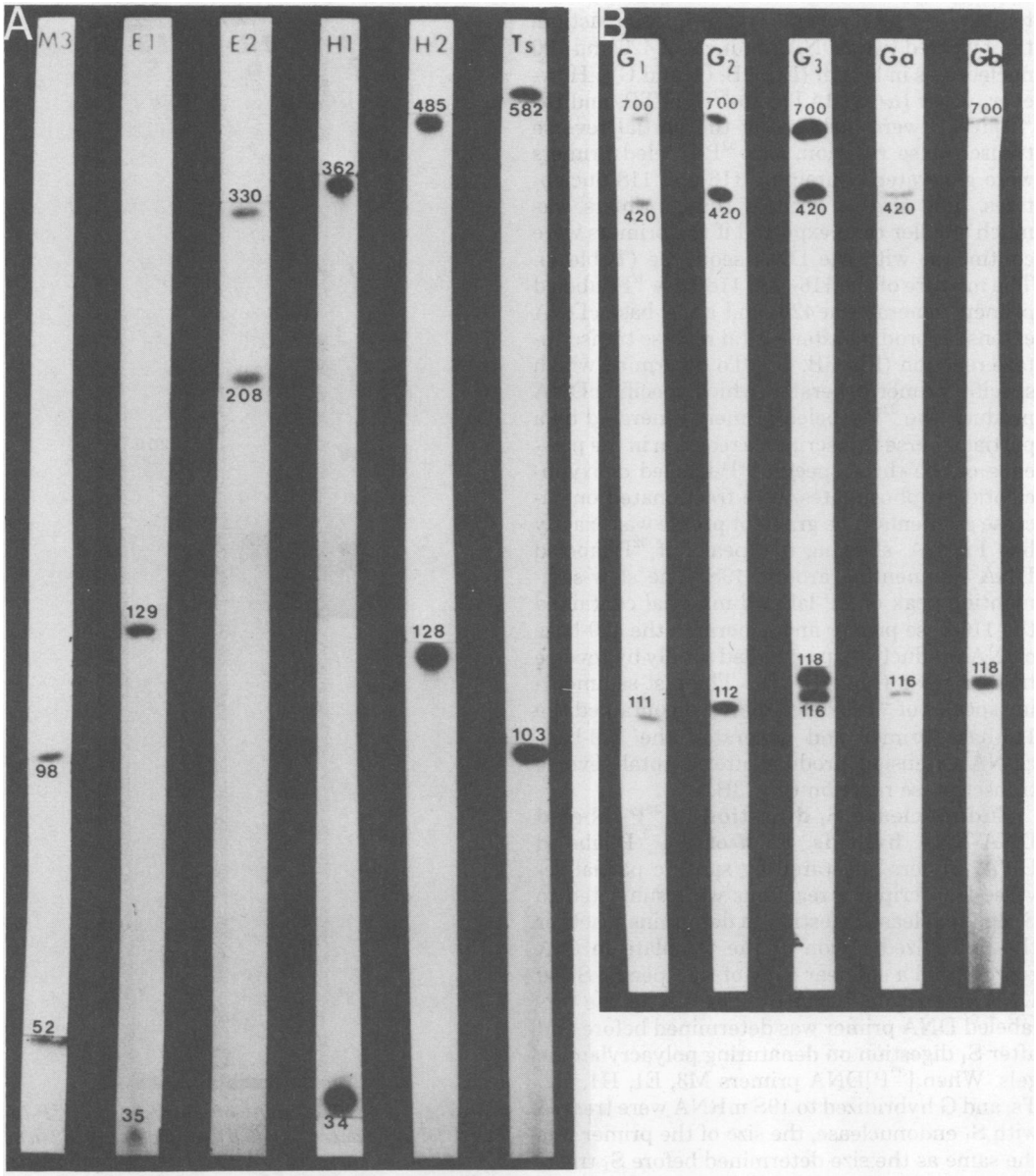

FIG. 3. *Autoradiograph of ^{32}P-labeled cDNA electrophoresed on 5% polyacrylamide gels in 7 M urea. After partial and total reverse transcriptase reactions, the ^{32}P-labeled DNA primers and cDNA products were electrophoresed for about 18 h at 200 V. Autoradiograms from different gels and different exposures of the same gel have been arranged to show the size of the ^{32}P-labeled primers and cDNA products determined as previously described (2). (A) Results obtained when fragments M3, E1, E2, H1, H2, and Ts hybridized to early mRNA were partially extended under the conditions of Table 1, followed by a total extension reaction. (B) Results obtained when the G fragment, hybridized to early mRNA, was partially extended under different conditions (Table 1) followed by a total extension reaction.*

tion, a single ^{32}P-labeled primer was generated which contained 111 nucleotides (Fig. 3B, G_1). When [α-^{32}P]dATP and [α-^{32}P]dTTP were used in the partial reverse transcriptase reaction, again a single ^{32}P-labeled primer was generated which contained 112 nucleotides (Fig. 3B, G_2). The sizes of these two primers are in agreement with the theoretical size expected based on the nucleotide sequence (Table 1). When the 111- and 112-base [^{32}P]DNA primers were fully ex-

tended by a total reverse transcriptase reaction, they yielded two cDNA products of 420 and 700 nucleotides in length (Fig. 3B, G_1 and G_2). However, when [α-^{32}P]dATP, [α-^{32}P]dTTP, and [α-^{32}P]dCTP were included in the partial reverse transcriptase reaction, two ^{32}P-labeled primers were generated containing 116 and 118 nucleotides. The size of both of these primers was much smaller than expected if the primers were continuous with the DNA sequence (Table 1). The mixture of the 116- and 118-base ^{32}P-labeled primers generated a 420- and a 700-base cDNA extension product after a total reverse transcriptase reaction (Fig. 3B, G_3). To determine which specific primer generated which specific cDNA product, the ^{32}P-labeled primers generated by a partial reverse transcriptase reaction in the presence of the three specific ^{32}P-labeled deoxynucleotide triphosphates were fractionated on sucrose gradients. The gradient profile was exactly like Fig. 2A, showing two peaks of ^{32}P-labeled DNA sedimenting around 19S. The slow-sedimenting peak of ^{32}P-labeled material contained the 116-base primer and generated the 420-base cDNA product when extended totally by reverse transcriptase (Fig. 3B, Ga). The fast-sedimenting species of ^{32}P-labeled material contained the 118-base primer and generated the 700-base cDNA extension product after a total reverse transcriptase reaction (Fig. 3B, Gb).

Endonuclease S_1 digestion of ^{32}P-labeled DNA-RNA hybrids. Each of the ^{32}P-labeled DNA primers generated by specific partial reverse transcriptase reactions were subjected to S_1 endonuclease digestion to determine whether the hybridized region of the template mRNA represented a colinear copy of the specific SV40 DNA restriction fragment. The size of the ^{32}P-labeled DNA primer was determined before and after S_1 digestion on denaturing polyacrylamide gels. When [^{32}P]DNA primers M3, E1, H1, H2, Ts, and G hybridized to 19S mRNA were treated with S_1 endonuclease, the size of the primer was the same as the size determined before S_1 treatment (data not shown). When the 200-base pair restriction fragment E2 (residues 152 to 351) was hybridized to 19S mRNA, a ^{32}P-labeled DNA primer of 208 nucleotides was generated by a specific partial reverse transcriptase reaction (Fig. 3A). When this [^{32}P]DNA-RNA hybrid was digested with S_1 endonuclease, two primers were found of different S_1-resistant sizes (Fig. 4, E2a). A small portion of the [^{32}P]DNA was totally resistant to S_1 and yielded a 208-base primer. However, the majority of the ^{32}P-labeled primer was reduced to a primer containing 180 nucleotides. To characterize which species of [^{32}P]-DNA-RNA hybrid was resistant to S_1 digestion,

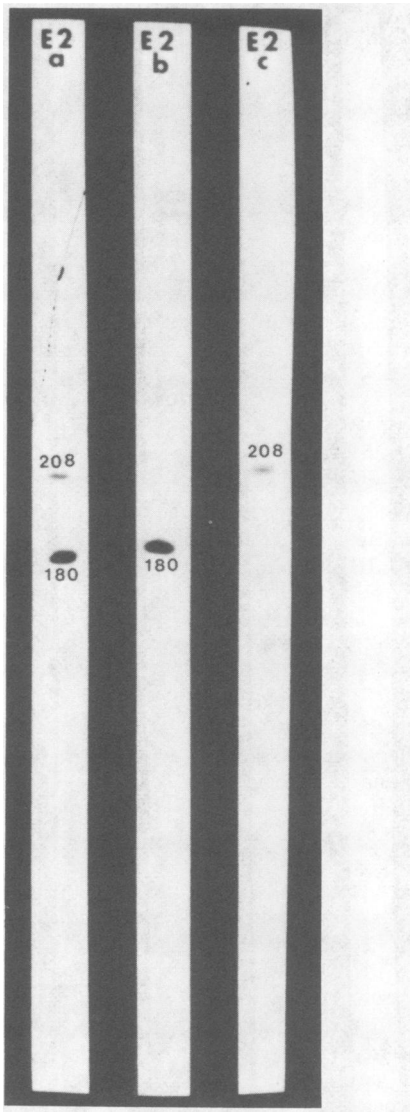

FIG. 4. *Autoradiograph of S_1-resistant, ^{32}P-labeled DNA primers. SV40 DNA fragment E2 was hybridized to early mRNA and ^{32}P labeled by partially extending the primer under the conditions listed in Table 1. After isolation on sucrose gradients, the total [^{32}P]DNA-RNA mixture (a), the slower-sedimenting species of [^{32}P]DNA-RNA (b), or the faster-sedimenting species (c) was treated with S_1 endonuclease. The S_1-resistant material was analyzed on 5% polyacrylamide gels in 7 M urea.*

the ^{32}P-labeled 208-base primer hybridized to 19S mRNA was fractionated on a sucrose gradient. Two peaks of [^{32}P]DNA material sedimenting at 19S, similar to the profile in Fig. 2A, were generated. The size of the ^{32}P-labeled primer in both species of hybrid was 208 nucleo-

tides. After S_1 digestion, the size of the primer in the slow-sedimenting species of hybrid was reduced to 180 nucleotides (Fig. 4, E2b). The size of the primer in the fast-sedimenting species of hybrid remained totally resistant to S_1 endonuclease treatment (Fig. 4, E2c).

DISCUSSION

Two species of cytoplasmic spliced viral 19S mRNA have been detected during the early phase of SV40 infection (1). One is 2,200 nucleotides in length and composed of two colinear parts, 330 and 1,900 nucleotides, spliced together. The other is slightly larger (2,500 nucleotides) and composed a 630-nucleotide segment spliced to the same 1,900-nucleotide segment as that found in the smaller species (1). Furthermore, it has been shown that by using in vitro translation of gradient-fractionated mRNA from SV40-infected cells, a partial resolution of these two species of early mRNA could be effected (15, 16a). The faster-sedimenting species of 19S mRNA coded for the small-t antigen, whereas the slower-sedimenting species coded for the large-T antigen (15). Hybridization of specific viral DNA restriction fragments with 19S mRNA allowed a clear resolution of these two species of 19S mRNA. The faster-sedimenting species of DNA-RNA hybrid (Fig. 2A) defines the mRNA for small-t. The slower-sedimenting species represents the mRNA for large-T. The fact that the resolution of these mRNA's was better than that obtained with translation experiments (15, 16a) could be due to the use of specific probes located around spliced regions at the 5' ends of these mRNA's. Regardless, the excellent resolution has allowed independent studies of the mRNA for small-t and large-T using the same DNA restriction fragment as a probe.

Paucha and Smith (17) have presented evidence that at least the first 25 amino acids are common to both large-T and small-t antigens. It has also been shown that the gene for both small-t and large-T antigens start at nucleotide 80, the AUG sequence (27). These results suggest that residues 80 to 155 are common to both species of mRNA. This is confirmed by the fact that when DNA restriction fragments M3 (residues 70 to 115) and E1 (residues 116 to 146) are used as primers, both species of 19S mRNA give identical cDNA products. This demonstrates that M3 and E1 hybridize in a colinear fashion (S_1 data) to both species of 19S mRNA. E2 (residues 152 to 351) also hybridized to both species of 19S mRNA but was colinear only with the mRNA for small-t (Fig. 4). Only 180 nucleotides of the E2 fragment were colinear with the mRNA for large-T. These results show the common region between the mRNAs for large-T and small-t extend from residues 70 to 323. This suggests that the first 81 amino acids of large-T and small-t are identical, a figure slightly lower than the 100 amino acids predicted by Fiers et al. (7).

DNA restriction fragments H1 (residues 352 to 379), H2 (residues 380 to 503), and Ts (residues 504 to 598) hybridized only to the mRNA for small-t. Since each of these fragments is a colinear transcript, the small-t mRNA contains at least residues 70 to 598. However, residues 602 to 604 contain the termination codon TAA and surely must be contained in the small-t mRNA sequence. Since C1 (residues 599 to 650) did not hybridize to either species of mRNA, the spliced region for small-t mRNA must occur very soon after the termination codon.

Independent primer extension studies using M3, E1, E2, H1, H2, and Ts DNA restriction fragments hybridized to 19S early viral mRNA have collectively located the 5' end of the cDNA extension products. These results (Fig. 3A) suggest that both species of mRNA have identical 5' termini located at residue 18 on the viral genome 62 nucleotides before the AUG initiation codon, which begins at residue 80. The fact that fragment M1 (residues 16 to 69) could not be extended by reverse transcriptase is consistent with these results since this primer overlaps the 5' termini by two residues. The sequence of residues 18 to 20 on the viral genome is ATT, which would correspond to an AUU sequence in the mRNA. The principal sequence adjacent to capped structures in SV40 late mRNA is AUU (9, 10, 13). Hence, the SV40 early mRNA may have the same principal sequence. Furthermore, termination of cDNA synthesis at ATT (residues 18 to 20) by reverse transcriptase suggests that these results accurately reflect the 5'-terminal structure of the 19S early viral mRNA.

Studies with the G fragment (residues 675 to 783) provided the most specific data concerning the position of the region where these initial transcripts at the 5' ends of the mRNA are spliced to the main body of the mRNA. Figure 5 summarizes the data obtained from reverse transcriptase reactions using the G fragment hybridized to early mRNA. Since G hybridized to both species of 19S mRNA and since neither C1 nor I (residues 599 to 674) hybridized to either species, we anticipated that a total reverse transcriptase reaction would generate two species of cDNA products, one reflecting the small-t mRNA and the other reflecting the large-T mRNA. When the G fragment was used to prime a partial reverse transcriptase reaction in the

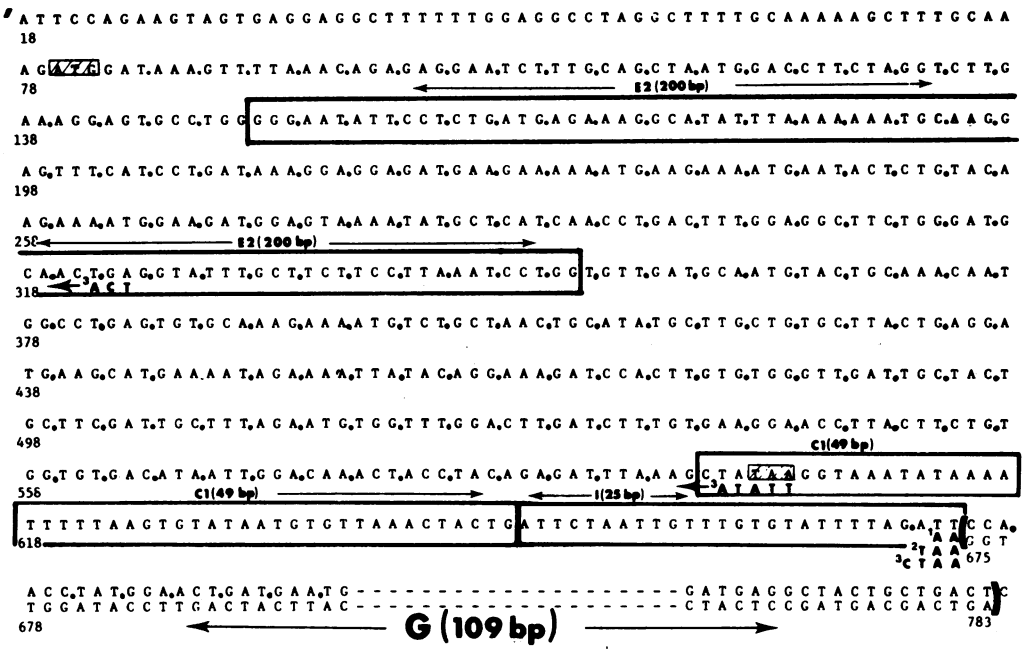

FIG. 5. *Sequence of the strand of SV40 DNA with the same polarity as the early mRNA. The thymidine residues in the DNA sequence would contain uridine bases in the mRNA sequence. The numbering system is the same as determined by Fiers et al. (7), and the reading frame for translation is indicated by dots (7). The sequence starts at residue 18, the position of termination of the total reverse transcriptase reaction. The position of the G fragment (109 base pairs [bp]) hybridized to residues 675 to 783 is included. The actual complementary residues incorporated in the partial extension of G are included when (1) [α-^{32}P]dATP, (2) [α-^{32}P]dATP and [α-^{32}P]dTTP, or (3) [α-^{32}P]dATP, [α-^{32}P]dTTP, and [α-^{32}P]dCTP are included in the reaction.*

presence of [α-^{32}P]dATP only two nucleotides were added (residues 673 and 674). When both [α-^{32}P]dATP and -dTTP were included, three nucleotides were added (residues 672 to 674). Both of these conditions led to the generation of a single primer that gave rise to two species of cDNA products, 420 and 700 nucleotides, after a total reverse transcriptase reaction (Fig. 3B). When three ^{32}P-labeled deoxynucleotide triphosphates (adenosine, thymidine, and cytosine) were included, we expected to add 23 ^{32}P-labeled nucleotides up to a C residue (residue 651), where a guanosine would be incorporated. Instead, two different ^{32}P-labeled primers were generated, one which added nine nucleotides and was associated with the small-t mRNA and another which added seven nucleotides and was associated with the large-T mRNA. The 118-base ^{32}P-labeled primer, which was associated with the small-t mRNA, generated the 700-base cDNA product by total primer extension methodology. The 116-base ^{32}P-labeled primer associated with the large-T mRNA generated the 420-nucleotide cDNA product. From the size of the total reverse transcriptase products, the specific α-^{32}P-labeled deoxynucleotide triphosphates incorporated, and the size of the cDNA products generated by the other specific DNA fragments, we can position the sequence around the spliced junction of each species of 19S mRNA (Fig. 5).

Figure 6 shows the sequences of both the large-T and small-t mRNA which have been determined by the present methodology. In many eucaryotic genomes the dinucleotide sequence AG has been found at both the donor site (3′ end of the segment on the 5′ side of the splice) and the acceptor site (5′ end of the segment on the 3′ side of the splice) (24, 25). Furthermore, we previously have determined the presence of the AG sequence at the acceptor site in the spliced junction of the late 16S SV40 mRNA (2).

Very recently, abundant SV40 mRNA species containing splices from residues 325 to 672 and 605 to 672 were identified in both transformed and infected cells by means of reverse transcriptase-catalyzed cDNA synthesis and cDNA sequencing (19). The sequences of the spliced junction of the early SV40 mRNA for large-T and small-t were determined to be AACUG/AG/AUUCCA and CUAUA/AG/AUUCCA, respec-

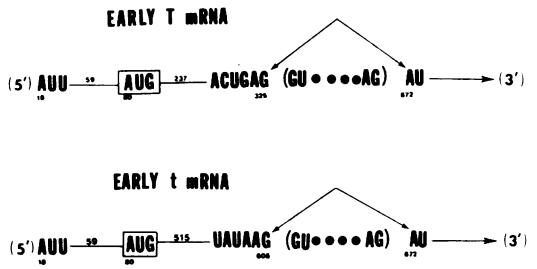

FIG. 6. *Structure of the early SV40 mRNA's determined by the primer extension methodology. Numbers below the sequence indicate the residues from SV40 DNA sequencing data (7). Numbers on the lines connecting the residues represent the length of the sequences which are joined together. The residues in parentheses represent the segments of DNA which have been deleted from the mature species of mRNA.*

tively (8, 19). This is consistent with the results determined here (Fig. 6), which characterize the overall structure of the 5' ends of the two spliced SV40 mRNAs detected early during infection. The mRNA for large-T has a 308-base segment spliced to residue 672, and the mRNA for small-t has a 588-base segment spliced to the same residue. The numbers of nucleotides spliced out of the mRNA's for both large-T and small-t are 346 and 66, respectively. The proposed spliced sequence for the small-t mRNA would include the termination codon (residues 602 to 604) for the proper translation of the small-t antigen. For the large-T mRNA, the sequence around the junction allows synthesis of large-T, which would contain the predicted amino acid sequence following the common N-terminal segment (7).

Recent evidence also has determined two principal reverse transcriptase stops at the 5' termini of early SV40 mRNA, both occurring with approximately equal frequency (19). The most distal of these stops would be located at residues 7 to 9, and the second, at residues 13 to 14. Several additional minor stops, between 0.62 and 0.65 map unit, were also observed. These data raise the possibility of additional 5' termini of early lytic and transformed cell SV40 mRNA (19). This elaborate sequencing was performed on a large pool of early mRNA and did not permit correlation of splice sequence to 5' terminus. The results presented here directly correlate the structure and sedimentation of the respective mRNA's and suggest that both species of mRNA have identical 5' ends, as evidenced by a single reverse transcriptase stop. The contrasting results could equally be the consequence of different times of the reverse transcriptase reaction. Previous studies used a 3-h reaction (19), whereas we have used a 30-min reaction. The fact that our results show that both species of mRNA have identical 5' ends and different spliced regions suggests that the expression of the early SV40 genes is controlled at two different levels. One would be the regulation of the initiation of transcription of a single promoter and the other would be post-transcriptional splicing of mRNA to control the abundance of the different viral translation products.

ACKNOWLEDGMENTS

We thank Marilyn Thoren for competent technical assistance. We also thank Minou Bina-Stein for valuable discussions during this study and Claude Garon for assistance during the preparation of this manuscript.

LITERATURE CITED

1. **Berk, A. J., and P. A. Sharp.** 1978. Spliced early mRNAs of simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **75:** 1274–1278.
2. **Bina-Stein, M., M. Thoren, N. Salzman, and J. A. Thompson.** 1978. Rapid sequence determination of late simian virus 40 16S mRNA leader by using inhibitors of reverse transcriptase. Proc. Natl. Acad. Sci. U.S.A. **76:** 731–735.
3. **Casey, J., and N. Davidson.** 1977. Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. Nucleic Acids Res. **4:**1539–1552.
4. **Crawford, L. A., and P. H. Black.** 1964. The nucleic acid of simian virus 40. Virology **24:**388–392.
5. **Deely, R. G., J. I. Gordon, A. T. H. Burns, K. P. Mullinix, M. Bina-Stein, and R. F. Goldberger.** 1977. Primary activation of the vitellogenin gene in the rooster. J. Biol. Chem. **252:**8310–8319.
6. **Eshaghpour, H., and D. M. Crothers.** 1978. Preparative separation of the complementary strands of DNA restriction fragments by alkaline RPC-5 chromatography. Nucleic Acids Res. **5:**13–21.
7. **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.
8. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. M. Weissman.** 1978. The heterogeneity and 5' terminal structures of the late RNAs of simian virus 40. J. Mol. Biol. **126:**813–846.
9. **Groner, Y., P. Carmi, and Y. Aloni.** 1977. Capping structures of simian virus 40 19S and 16S mRNAs. Nucleic Acids Res. **4:**3959–3968.
10. **Haegeman, G., and W. Fiers.** 1978. Characterization of the 5'-terminal capped structures of late simian virus 40-specific mRNA. J. Virol. **25:**824–830.
11. **Hardies, S. C., and R. D. Wells.** 1976. Preparative fractionation of DNA restriction fragments by reversed phase column chromatography. Proc. Natl. Acad. Sci. U.S.A. **73:**3117–3121.
12. **Landy, A., C. Foeller, R. Reszelbach, and B. Dudock.** 1976. Preparative fractionation of DNA restriction fragments by high pressure column chromatography on RPC-5. Nucleic Acids Res. **3:**2575–2592.
13. **Lavi, S., and A. J. Shatkin.** 1975. Methylated simian virus 40-specific RNA from nuclei and cytoplasm of infected BSC-1 cells. Proc. Natl. Acad. Scid. U.S.A. **72:** 2012–2016.
14. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74:**560–564.

15. **May, E., M. Kress, and P. May.** 1978. Characterization of two SV40 early mRNAs and evidence for a nuclear "presliced" RNA species. Nucleic Acids Res. **5**:3083-3099.
16. **McCarty, K. S., Jr., R. T. Vollmer, and K. S. McCarty.** 1974. Improved computer program data for the resolution and fractionation of macromolecules by isokinetic sucrose density gradient sedimentation. Anal. Biochem. **61**:165-183.
16a. **Paucha, E., R. Harvey, R. Smith, and A. E. Smith.** 1977. The cell-free synthesis of SV40 T-antigens. EMBO-INSERM Symp. **69**:189-198.
17. **Paucha, E., and A. E. Smith.** 1978. The sequences between 0.59 and 0.54 map units on SV40 code for the unique region of small-t antigen. Cell **15**:1011-1020.
18. **Prives, C., E. Gilboa, M. Revel, and E. Winocour.** 1977. Cell-free translation of simian virus 40 early messenger RNA coding for viral T-antigen. Proc. Natl. Acad. Sci. U.S.A. **74**:457-461.
19. **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.
20. **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.
21. **Sanger, R., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463-5467.
22. **Sebring, E. D., T. J. Kelly, Jr., M. M. Thoren, and N. P. Salzman.** 1971. Structure of replicating simian virus 40 deoxyribonucleic acid molecules. J. Virol. **8**:478-490.
23. **Shani, M., E. Birkenmeier, E. May, and N. P. Salzman.** 1977. Properties of simian virus 40 transcriptional intermediates isolated from nuclei of permissive cells. J. Virol. **23**:20-28.
24. **Tilghman, S. M., D. C. Tiemeier, J. G. Seidman, B. M. Peterlin, M. Sullivan, J. V. Maizel, and P. Leder.** 1978. Intervening sequence of DNA identified in the structural portion of a mouse β-globin gene. Proc. Natl. Acad. Sci. U.S.A. **75**:725-729.
25. **Tonegawa, S., A. M. Maxam, R. Tizard, O. Bernard, and W. Gilbert.** 1978. Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain. Proc. Natl. Acad. Sci. U.S.A. **75**:1485-1489.
26. **Vogt, V. M.** 1973. Purification and further properties of single-strand-specific nuclease from *Aspergillus oryzae*. Eur. J. Biochem. **33**:192-200.
27. **Volckaert, G., A. Van de Voorde, and W. Fiers.** 1978. Nucleotide sequence of the simian virus 40 small-t gene. Proc. Natl. Acad. Sci. U.S.A. **75**:2160-2164.