JOURNAL OF VIROLOGY, Jan. 1981, p. 7-16
0022-538X/81/010007-10$02.00/0
Vol. 37, No. 1

Simian Virus 40 Early mRNA's in Lytically Infected and Transformed Cells Contain Six 5'-Terminal Caps

CHAIM KAHANA, DAVID GIDONI, DAN CANAANI, AND YORAM GRONER*
Department of Virology, Weizmann Institute of Science, Rehovot, Israel

Late simian virus 40 (SV40) mRNA contains eight different cap structures which we have previously identified and mapped on the viral genome. As reported here, 5'-cap heterogeneity is a common feature to both the early and the late SV40 mRNA's. *methyl*-^{3}H-labeled viral mRNA was purified from cells infected at 41°C with SV40 mutant *tsA*209. Three different cap cores were identified: m^{7}GpppGm, m^{7}GpppCm, and m^{7}GpppAm. An average of three to four m^{6}A residues per mRNA molecule was found. RNase T_2-resistant ^{32}P-labeled early caps from *tsA*209-infected cells were isolated and characterized. Six distinct cap I structures were identified: m^{7}GpppCmpU (30%), m^{7}GpppGmpC (24%), m^{7}GpppAmpG (18%), m^{7}GpppGmpU (13%), m^{7}GpppGmpG (12%), and m^{7}GpppAmpU (3%). A similar 5'-end heterogeneity was observed in early SV40 mRNA from BSC-1 cells infected with wild-type SV40 strain 777 in the presence of cytosine arabinoside and in the SV40 UV-transformed permissive line C-6. Five of these capped dinucleotides are complementary to DNA sequences at 0.66 map unit in a region previously identified by the primer extension method (Reddy et al., J. Virol. **30**:279–296, 1979; Thompson et al., J. Virol. **31**:437–438, 1979) as the 5' end of the early message. DNA sequences upstream from this region contain the TATTTAT (Hogness-Goldberg box), which is missing from upstream of the 5'-cap sites of late SV40 mRNA. Thus, 5'-end heterogeneity is not necessarily related to the presence or the absence of this putative transcriptional "initiation signal." When the possibility that SV40 5' caps represent transcriptional initiation sites is considered, the data also suggest that, on SV40 DNA, eucaryotic RNA polymerase II initiates transcription at multiple nucleotide sequences, including pyrimidines.

In eucaryotic cells, the formation of translatable mRNA is a complex process (for reviews, see references 16 and 53). Simian virus 40 (SV40) serves as a simplified and useful model for exploring the post-transcriptional events which occur between RNA transcription in the cell nucleus and its eventual translation in the cytoplasm (reviewed in reference 45). The genome of SV40 is a closed, double-stranded, circular DNA molecule comprising, as determined by DNA sequencing, 5,226 base pairs (21, 51). It is divided into two distinct transcriptional units, the early region and the late region. The early region occupies approximately 50% of the viral minus strand; it extends from near the origin of DNA replication at 0.67 map unit to 0.165 map unit (45). From early times after infection and throughout the lytic cycle, as well as in SV40-transformed cells, the early region is transcribed into two differently spliced 19S mRNA species which encode the large and small tumor antigens (45). In addition, Alwine et al. (5) have recently described an SV40-associated small RNA from the early region which is induced late in infection, and Mark and Berg (46a) have identified a number of minor early mRNA species which appear at late times after infection. The two major early 19S mRNA's differ in the size of their internal spliced-out segment. The small-t-antigen mRNA spans the entire early region but lacks 66 nucleotides from map position 0.54, whereas the large-T-antigen mRNA loses a 346-nucleotide segment between 0.54 and 0.60 map unit (7, 49). The 5' termini of early cytoplasmic mRNA (19) and early nuclear RNA from *tsA*-infected cells (52) were mapped at 0.66 map unit. Aloni et al. (3) have recently shown that nuclear transcripts of both the late (L) and the early (E) strands are capped and contain m^{6}A and that the nuclear caps are located between 0.56 and 0.72 map unit.

The late region occupies slightly less than half of the SV40 plus strand (0.67 to 0.17 map unit) and encodes the late viral functions. After the onset of DNA synthesis, the early-late switch occurs, and the late region is transcribed into virus-specific messengers that fall into the size classes of 19S and 16S (45). These late mRNA's

are transcribed from noncontiguous sequences of DNA, and each contains a 5′ leader sequence (4, 10, 13, 18, 27, 28, 34, 39, 43). The late 19S and 16S messengers are capped and internally methylated (2, 11, 12, 30, 35, 44). Capping of the 5′-terminal leader sequence with m^7G is a posttranscriptional event (33, 47, 61, 65), but is tightly coupled to the initiation of transcription (24, 25, 31, 56). Recent studies revealed extensive heterogeneity in the 5′ termini of late SV40 mRNA's (8, 12, 28, 36, 42, 50, 60), late polyoma mRNA's (22, 23), and early adenovirus type 2 (Ad2) mRNA's (6, 37, 38). It is conceivable, although not proven, that late SV40 5′ caps are added to the initiating nucleotide of the RNA chain and that each of the multiple 5′-terminal caps of late SV40 mRNA's represents a transcriptional initiation point (12). Characterization of early SV40 cap structures should help in the understanding of their role in transcriptional initiation and in the early-late switch during the SV40 lytic cycle.

In this work, we describe the analysis of 5′-terminal caps found on early SV40 mRNA's in lytically infected and transformed cells. The 5′ termini of the two major early mRNA species were determined recently by means of a primer-directed complementary DNA (cDNA) synthesis followed by either cDNA sequencing (49) or sizing of the extended cDNA (59). These reports differ in the number and the position of the 5′ termini. Reddy et al. (49) have detected two major reverse transcriptase stops at residues 5147 to 5148 and 5152 to 5154 (Reddy et al. [51] numbering system), whereas only a single reverse transcriptase stop at residue 18 (Fiers et al. [21] numbering system or residue 5143 by the Reddy et al. [51] numbering system) was reported by Thompson et al. (59).

We found that early mRNA's in monkey cells either transformed or lytically infected have at least six different capped oligonucleotides at their 5′ termini. 5′-Cap heterogeneity is, therefore, a common feature to both the early and the late SV40 mRNA's. In analogy to late polyoma 5′-terminal caps (22), four of the early SV40 caps can be clustered within a pentanucleotide sequence between residues 5155 and 5150, in good agreement with the reported reverse transcriptase stops (49).

MATERIALS AND METHODS

Cells and viruses. Two monkey cell lines were used in this study: BSC-1 cells and the SV40 UV-transformed permissive C-6 line established by Gluzman et al. (29). Cells were grown in Dulbecco-modified Eagle medium containing 10% calf serum. The wild-type SV40 strain 777 used was produced by infecting BSC-1 cells at a multiplicity of 25 PFU per 10^7 cells with a single-plaque isolate that had been subjected to two sequential plaque purification procedures. The SV40 *tsA*209 mutant (14) was obtained from R. G. Martin.

Labeling and isolation of SV40 mRNA. Confluent cultures of BSC-1 cells were infected with SV40 at a multiplicity of 50 PFU per cell. At 4 to 5 h postinfection, 20 μg of cytosine arabinoside (araC) per ml was added to the cultures. BSC-1 cells were infected with *tsA*209 at a multiplicity of 10 to 20 PFU per cell. After 2 h at 37°C, infected cells were incubated and labeled at 41°C. For the labeling of RNA with ^{32}P at 24 h postinfection, cells were incubated for 6 to 8 h with 1 mCi (1 Ci = 3.7×10^{10} becquerels) of [^{32}P]-phosphate per ml in otherwise-phosphate-free medium. When araC was omitted, cells were labeled with ^{32}P at 5 h postinfection, and RNA was extracted 8 h later. Transformed C-6 cells were labeled at confluency with [^{32}P]phosphate, as were the infected cells. Labeling of SV40 *tsA*209-infected BSC-1 cells with [*methyl*-^3H]methionine (5 mCi per 10^7 cells) was at 41°C for 6 to 8 h in methionine-free medium. Labeled RNA was extracted and polyadenylic acid-containing RNA was isolated as previously described (30). SV40-specific mRNA was isolated by hybridization to SV40 DNA immobilized on Sepharose as described elsewhere (30).

Preparation of single-stranded SV40 DNA fragments and RNA hybridization. SV40 form 1 DNA was digested by a combination of *Bgl*I and *Hpa*I restriction endonucleases (New England BioLabs). Its resulting fragments were separated into single strands and transferred to nitrocellulose filters as described by Kaufmann et al. (40). Nitrocellulose strips were incubated with a particular ^{32}P-labeled RNA in 0.5 ml of hybridization buffer (0.1 M HEPES [*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid] NaOH [pH 7.5], 0.75 M NaCl, 0.5% sodium dodecyl sulfate, 1 mM EDTA, and 50% 4× recrystallized formamide). Hybridization was for 48 to 72 h at 37°C. After hybridization, strips were washed with 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and treated with 23 μg of RNase A per ml in 2× SSC for 60 min at 24°C. SV40 mRNA was also hybridized to a restriction fragment of SV40 DNA bound to Sepharose. SV40 DNA fragment C (0.57 to 0.67 map unit), obtained by digesting SV40 DNA with restriction endonucleases *Bgl*I, *Bcl*I, and *Taq*I, was covalently bound to Sepharose as previously described. Preselected early SV40 mRNA's were hybridized to the Sepharose-bound fragment, washed, and treated with RNase T_1, and protected RNA was eluted and analyzed as outlined by Canaani et al. (12).

Enzymatic treatments and products analysis. Combined digestion with nuclease P1 (Yamasa Shoya Co.) and alkaline phosphatase (Worthington Biochemicals Corp.), as well as digestion with RNase T_2 (Calbiochem) followed by alkaline phosphatase, was carried out as described by Groner and Hurwitz (33) and Groner et al. (32). Digestion with nucleotide pyrophosphatase (Sigma Chemical Co.) was performed as described by Groner et al. (31). Paper electrophoresis in pyridinium acetate (pH 3.5), two-dimensional analysis of caps, and thin-layer chromatography were carried out as described previously (30–33).

RESULTS

Caps and internal methylation of early SV40 mRNA's. *methyl*-^{3}H-labeled SV40 mRNA's were isolated from the cytoplasm of *tsA*209-infected cells as described above. Viral mRNA was digested with nuclease P1 (which cleaves phosphodiester linkages in polynucleotides, including those containing 2'-*O*-methylated residues to yield 5'-nucleotides), followed by incubation with alkaline phosphatase, and the products were analyzed by paper electrophoresis. As shown in Fig. 1A, about 60% of the radioactivity in the digest migrated toward the cathode in a spot corresponding to adenosine and was further identified by paper chromatography as m^{6}A (Fig. 1B). The remaining radioactivity migrated toward the anode as two peaks in positions identical to those of the cap markers m^{7}GpppAm plus m^{7}GpppCm and m^{7}GpppGm. The A cap (m^{7}GpppAm) and the G cap (m^{7}GpppGm) are common viral caps (for a review, see reference 58). For example, 95% of the late SV40 caps are A caps (12, 30, 35), whereas the C cap has been reported to occur in cellular mRNA's only (15, 26, 48, 61). Figure 1B shows that, in fact, early SV40 mRNA's contain a substantial amount of C cap (m^{7}GpppCm). In this experiment, viral RNA digested with nuclease P1 plus alkaline phosphatase was analyzed by paper chromatography, using a solvent system that separates the three cap species. Material comigrating with the m^{7}GpppCm marker in Fig. 1B (fractions 24 to 27) was eluted, digested with nucleotide pyrophosphatase (which cleaves the pyrophosphate bridge in cap structures), and subjected to alkaline phosphatase treatment and paper electrophoresis (Fig. 1C). Radioactivity which previously coincided with the m^{7}GpppCm marker now migrated as two spots in a ratio of about 1:1, corresponding to 7-methylguanosine and cytosine, confirming the presence of the C cap in early SV40 mRNA.

Methyl-labeled caps comigrating with the m^{7}GpppGm and m^{7}GpppAm markers in Fig. 1B were similarly analyzed; digestion with nucleotide pyrophosphatase plus alkaline phosphatase specifically released m^{7}G plus Gm or Am, respectively, from each cap (data not shown). Further confirmation of the identities of the early caps was obtained by analysis of ^{32}P-labeled RNA as described below. The relative amounts (percentages) of the three *methyl*-^{3}H-labeled early cap cores analyzed in Fig. 1B were as follows: m^{7}GpppGm, 46; m^{7}GpppCm, 32; and m^{7}GpppAm, 21. These values are in good agreement with those found with ^{32}P-labeled caps (see Table 2).

From the data shown in Fig. 1A and B, it is evident that in addition to cap methylations the early SV40 mRNA's contain internal m^{6}A residues. An average number of three to four m^{6}A residues per early mRNA molecule was determined from the m^{6}A/cap ratio in Fig. 1A and B, taking into account that SV40 early cap cores contain two methyl groups (Table 1). Thus, the early mRNA's contain, per molecule, one m^{6}A residue more than does the late message (11).

^{32}P-labeled early SV40 mRNA from lytically infected and transformed cells. Isolation and hybridization analysis of The analysis of *methyl*-^{3}H-labeled viral RNA described above showed that three different cap cores exist in the early SV40 message from *tsA*209-infected cells. To further identify the early SV40 5'-terminal caps in lytically infected and transformed cells, ^{32}P-labeled cytoplasmic mRNA's were isolated from three different sources: (i) BSC-1 cells infected with wild-type SV40 strain 777 in the presence of araC; (ii) BSC-1 cells infected at the nonpermissive temperature (41°C) with SV40 mutant *tsA*209; and (iii) the SV40 UV-transformed permissive cell line C-6. SV40-specific mRNA's were isolated by preparative hybridization to SV40-Sepharose. The purified viral mRNA's were analyzed by hybridization to blots containing separated strands of SV40 DNA fragments obtained by digestion with restriction endonucleases *Bgl*I and *Hpa*I (Fig. 2). The majority of the polyadenylic acid-containing viral RNA hybridized with the E strand of the early region; i.e., B_E = 0.67 to 0.37 map unit, and C_E = 0.37 to 0.17 map unit. The presence of a low amount of late transcripts was evident (Fig. 2) (i.e., A_L = 0.76 to 0.17 map unit), although the amount was less than 5%. Anti-early, as well as anti-late, RNA (complementary to B_L and A_E) could also be detected. Such low levels of late RNA were previously reported in araC-treated infected cells (55, 64), in *tsA*-infected cells at a restrictive temperature (41), and by transcription complexes isolated from these cells (9, 20, 57).

Multiple RNase T_2-resistant caps of early SV40 mRNA from lytically infected cells. ^{32}P-labeled viral mRNA from *tsA*209-infected cells was prepared and analyzed as described below. The RNA was extensively digested with RNase T_2 and then treated with RNase A and alkaline phosphatase. RNase T_2 cleaves phosphodiester linkages in polynucleotides, except those containing 2'-*O*-methylated residues, to yield 3'-phosphate nucleotides. Therefore, RNase T_2 digestion followed by phosphatase treatment should generate 5'-cap structures of the types m^{7}GpppXmpY (cap I) and m^{7}GpppXmpYmpZ (cap II). Digested RNA was fractionated by two-dimensional electrophoresis

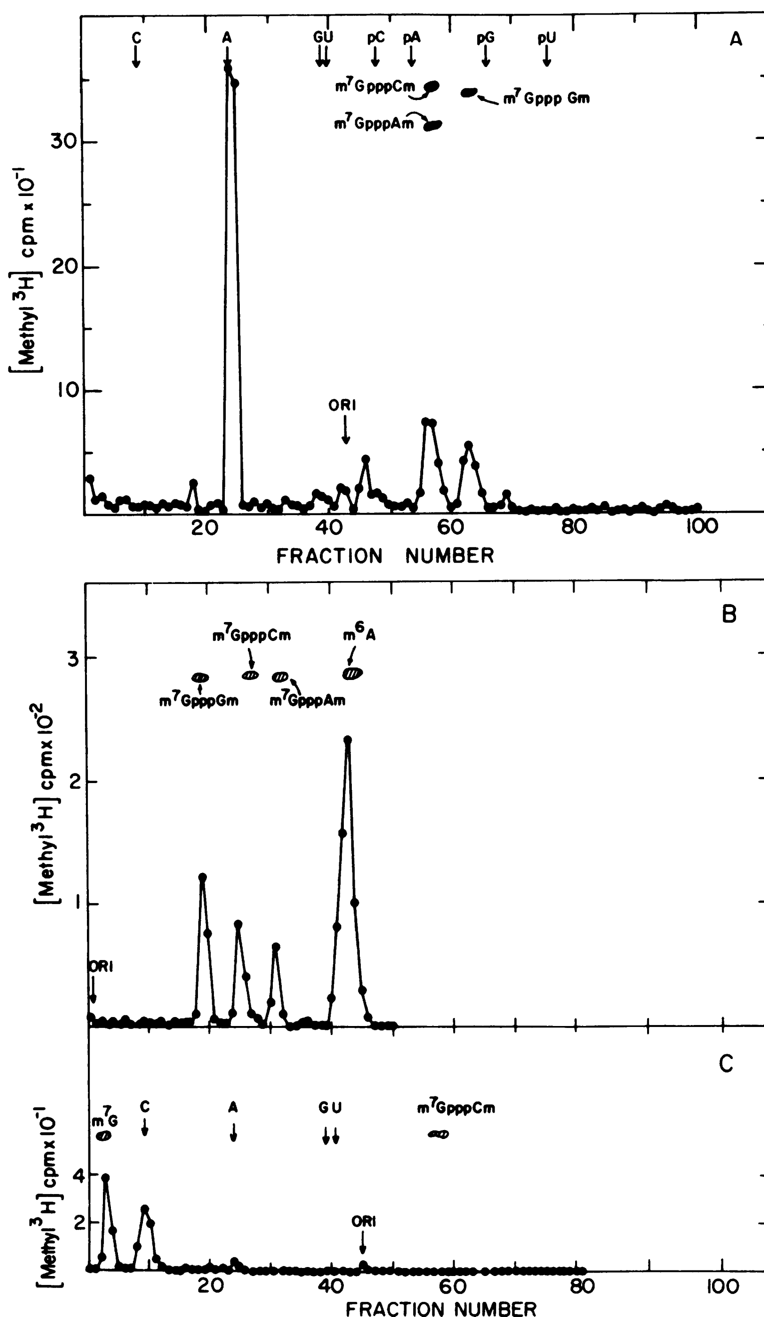

FIG. 1. *Analysis of enzymatically digested methyl-^{3}H-labeled early SV40 mRNA. Purified mRNA was digested with nuclease P1 followed by alkaline phosphatase and analyzed by paper electrophoresis (pH 3.5) (A) and paper chromatography in isobutyric acid–0.5 M NH_4OH (10:6)(B) as described in the text. (C) Material comigrating with the m^{7}GpppCm marker in (B) was eluted from the paper, digested with nucleotide pyrophosphatase plus alkaline phosphatase, and analyzed by paper electrophoresis. The arrows indicate the positions of radioactive markers. Cap markers and the m^{7}G marker are spots detected under UV light.*

and chromatography. A typical fingerprint of early SV40 caps is shown in Fig. 3. Five spots, numbered 1 through 5, were observed. In addition, a few minor spots appeared when films were exposed for a longer period of time (spots 6 and 7). (From their positions on the fingerprint, we assumed that spots 6 and 7 represented cap II structures of the major cap I, but due to

TABLE 1. *Average number of m^6A residues per early SV40 mRNA molecule*

Expt	m^6A (cpm)	Caps (cpm)	m^6A/cap	Deduced no. of m^6A residues
1 (Fig. 1A)	705	374	1.88	3.76
2 (Fig. 1B)	630	405	1.55	3.10

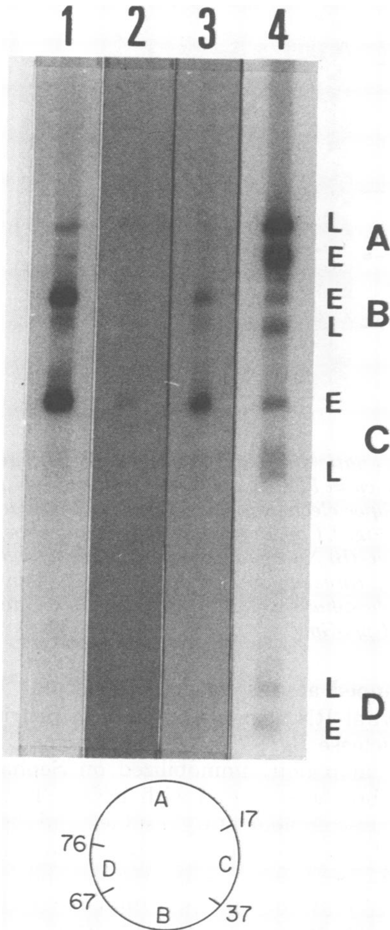

FIG. 2. *Hybridization of ^{32}P-labeled early mRNA to blots containing separated DNA strands of fragments obtained by digestion of SV40 DNA with BglI and HpaI. Conditions for strand separation, fragment transfer, and DNA-RNA hybridization are detailed in the text. Hybridization was carried out with ^{32}P-labeled early SV40 mRNA from: (1) cells infected with tsA209 at 41°C; (2) cells infected with wild-type SV40 in the presence of araC; (3) SV40 (UV-transformed) C-6 cells. (4) Hybridization with SV40 DNA labeled in vitro with [^{32}P]dXTP. Letters mark the strand positions.*

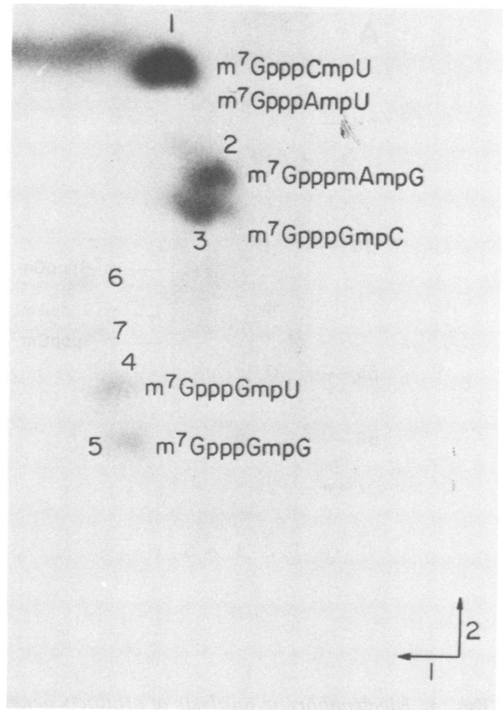

FIG. 3. *Autoradiogram of two-dimensional separation of early SV40 mRNA digested with RNase T_2 and alkaline phosphatase. (First dimension) Cellulose acetate electrophoresis in pyridinium acetate buffer (pH 3.5); (second dimension) chromatography on a polyethyleneimine-cellulose plate (20 by 40 cm) in 1.2 M pyridinium formate buffer (pH 4.3). Approximately 8 × 10^5 cpm (Cherenkov) were used, and plates were autoradiographed at −85°C with preexposed Agfa Curix RP-2 film and Du Pont Cronex Lightning-Plus intensifying screens. The sequences presented were obtained by secondary digestion of material eluted from the various spots, as described in the legend to Fig. 4A and B.*

insufficient radioactivity, they were not analyzed further.) The general pattern of the fingerprint and the relative amounts of the various spots were not altered after an additional selection cycle of the purified SV40 mRNA by hybridization to SV40 DNA. This strongly indicated that all of the spots originated from viral mRNA. The capped oligonucleotides were eluted from the thin-layer plate and digested with nuclease P1, and the resulting products were analyzed by paper electrophoresis (Fig. 4A). Each of the spots released a nuclease P1-resistant core and an additional nucleoside 5'-monophosphate. Nuclease P1 treatment of spot 1 released, in addition to $m^7GpppCm$, small proportions of $m^7GpppAm$. The three viral cap cores were eluted from the paper, and their compositions were analyzed by digestion with nucleotide pyrophosphatase followed by paper electrophoresis. All three caps yielded material migrating with m^7pG, as well as radioactivity which migrated together with pA, pC, and pG, respec-

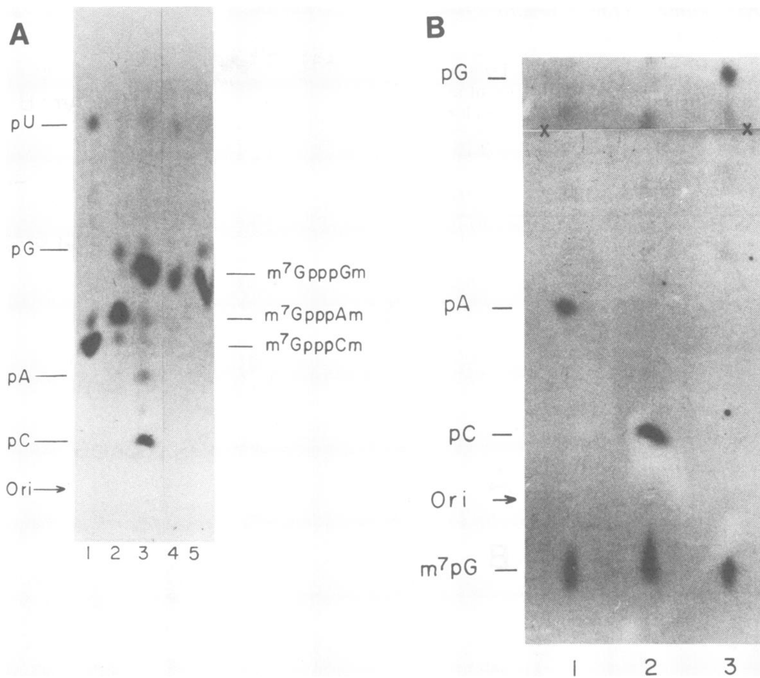

FIG. 4. *Electrophoretic analysis of products generated by enzymatic digestions. (A) Nuclease P1 digestion of the RNase T_2-resistant cap resolved in Fig. 3. Spots 1 to 5 shown in Fig. 3 were eluted with 2 M NH_4HCO_3 (pH 7), desalted, digested with nuclease P1, and analyzed by paper electrophoresis at pH 3.5. Radioactivity was visualized by autoradiography as described in the legend to Fig. 3. Cap and nucleoside 5'-monophosphate markers were included with the samples and located by UV light. (B) Nucleotide pyrophosphatase digestion of nuclease P1-resistant cap cores. Material comigrating with cap cores in paper electrophoresis or thin-layer chromatography was eluted and digested with nucleotide pyrophosphatase, and the products were analyzed by paper electrophoresis. (1) m^7GpppAm; (2) m^7GpppCm; (3) m^7GpppGm.*

tively, in yields of 1:1 by ^{32}P counts (Fig. 4B). The results obtained by nuclease P1 and nucleotide pyrophosphatase digestions were consistent with the presence of six distinct cap I structures in early SV40 mRNA's. Their relative abundance and comparison with late SV40 caps, as well as with BSC-1 and C-6 cellular mRNA caps, are summarized in Table 2. In contrast to late SV40 mRNA, in which 95% of the cap cores were A caps, the distribution of the early cap cores was 50% G caps and only 20% A caps. The finding of relatively high proportions of C caps was rather unexpected and will be discussed later. A comparison between the viral and cellular RNase T_2-resistant caps (cap I) revealed large differences in the distribution of early and late SV40 caps and differences between them and the cellular population (Table 2). On the other hand, the two monkey cell lines, BSC-1 and CV-1-derived C-6, showed similar patterns.

As mentioned above, other investigators have located the 5' termini of early cytoplasmic mRNA at 0.66 map unit (19, 49, 52, 59). To determine whether the multiple cap I structures are mapped at this region, preselected ^{32}P-labeled viral RNA was hybridized to restriction endonuclease *Bgl*I-*Taq*I-*Bcl*I fragment C (0.57 to 0.67 map unit) immobilized on Sepharose. The hybrid was treated with RNase T_1 to remove non-complementary residues, and the protected RNA was eluted, digested, and fingerprinted as before. The two major caps, m^7GpppCmpU and m^7GpppGmpC, were clearly visible on the autoradiogram (not shown), whereas the appearance of the other caps was questionable, possibly because of insufficient radioactivity. Thus, at present, we cannot distinguish between the following two alternatives: (i) the other 5' caps were not well protected against RNase T_1 by the DNA fragment spanning 0.57 to 0.67 map unit because of their close proximity to the *Bgl*I site at 0.67 map unit (see Fig. 5, residue 5156), or (ii) the other 5' caps were encoded by DNA sequences not included in this region of the genome.

Early 5' caps in SV40 UV-transformed cells and in araC-treated lytically infected cells. Assuming that the 5'-terminal caps rep-

TABLE 2. *Analysis and distribution of SV40 and cellular cap I structures*

Spot no. (Fig. 3)	Nuclease P1 products		Deduced cap I structure	Relative abundance (%) in mRNA			
				SV40		Cellular	
	Cap core[a]	Nucleotides[a]		Early[b]	Late[c]	BSC-1[d]	C-6[d]
1	m^{7}GpppCm	pU	m^{7}GpppCmpU	30	<2	10	10
	m^{7}GpppAm		m^{7}GpppAmpU	3	66	1–2	
2	m^{7}GpppAm	pG	m^{7}GpppAmpG	18	8	13	17
3	m^{7}GpppGm	pC	m^{7}GpppGmpC	24	5	30	32
4	m^{7}GpppGm	pU	m^{7}GpppGmpU	13	<2	7	10
5	m^{7}GpppGm	pG	m^{7}GpppGmpG	12	<2	15	16
			m^{7}GpppmAmpC[e]	<1	21	10	9
			m^{7}GpppCmpC[e]	<1	<2	5	4
			Others			9	2

[a] Evidence is shown in Fig. 3 and 4.
[b] Percentages are the mean values of four different experiments with *tsA*209-infected cells as described in the text.
[c] From Canaani et al. (12).
[d] The analysis was performed as described in the text for early SV40 caps.
[e] Not detected in early SV40 mRNA.

resent transcriptional initiation points, early caps of SV40 *tsA*209 may be different from those of wild-type SV40 or from viral mRNA in transformed cells. To explore this possibility, ^{32}P-labeled early SV40 mRNA was prepared from the SV40 UV-transformed C-6 line and from araC-treated wild-type-infected cells, as outlined above. Labeled RNA was digested and fingerprinted as described above. The six early caps identified in *tsA*209 were present in both mRNA preparations in approximately the same proportions as summarized for *tsA*209 in Table 2. Therefore, we concluded that early SV40 caps are similar in lytically infected and transformed monkey cells.

Early SV40 caps were also analyzed in viral RNA from lytically infected cells not treated with araC for the following reasons. Thompson et al. (59) have recently identified a single reverse transcriptase stop on early mRNA at residue 18 of the viral genome (Fiers et al. [21] numbering system) in the sequence 5′-ATT. This 5′ end lies 4 to 11 nucleotides downstream from the two principal reverse transcriptase stops previously described by Reddy and coworkers (49). The two reports differ in the methods applied for analysis of the extended primer (DNA sequencing [Reddy et al.] versus sizing of the extended cDNA [Thompson et al.]). In addition, araC was not used by Thompson et al. to inhibit DNA replication; accordingly, they prepared cytoplasmic RNA 12 h postinfection. Our analysis revealed a low proportion (3%) of m^{7}GpppAmpU caps which may fit the ATT terminus. Therefore, we have considered the possibility that, in addition to the inhibition of DNA replication, araC induces a secondary effect on transcription (1) or the initiation of transcription and therefore may affect the 5′-terminal caps. To this end, ^{32}P-labeled viral RNA was prepared from lytically infected cultures at 10 h postinfection, and early SV40 caps were analyzed. The amount of m^{7}GpppAmpU was still low, and no other significant changes in the relative abundance of either the viral or the host caps were observed.

DISCUSSION

In this study, early virus-specific mRNA was isolated from lytically infected and transformed cells. The mRNA 5′-terminal caps and methylation of internal m^{6}A were examined.

Multiple caps and 5′ heterogeneity. Six distinct early cap I structures were identified. Since only two early mRNA's exist which have one (59) or two (49) principal 5′ termini, these data imply 5′-terminal heterogeneity of the early message. Neither the SV40-associated small RNA (5) nor the additional early mRNA species described by Mark and Berg (46a) have been detected before DNA replication. Therefore, we feel that no significant amounts of these RNAs were labeled in our experiments.

m^{7}GpppCmpU: the major early cap. In contrast to late SV40 caps that contain 95% A caps (i.e., contain methylated adenosine at the penultimate position), of the early caps, 50% are G caps, 30% are C caps, and only 20% are A caps.

The finding that early mRNA contains a high proportion of C caps was surprising. In fact, m^{7}GpppCmpU was the most abundant early cap (Table 2). The existence of both U and C caps in total cellular mRNA is well documented (15, 26, 48, 61; Table 2). However, in late mRNAs of Ad2 (67), polyoma virus (23), and SV40 (12, 30, 35), only purines were found as the penultimate nucleotides, and A caps were favored over G caps. In early Ad2, a Um terminus was detected recently (38), although in this case too, a preference of purines versus pyrimidines was evident (6, 38). In late polyoma and early Ad2 mRNA's, few of the 5'-terminal caps are clustered and encoded in one oligonucleotide of four to six residues. Again, none of the potential pyrimidines in this region are used as cap sites (6, 22). Thus, the fact that the major early SV40 cap (m^{7}GpppCmpU) is a pyrimidine dinucleotide is conspicuous and may be important in the expression of the early SV40 region.

During the past years, in vivo and in vitro transcription experiments (23, 46, 56, 63, 67), as well as direct biochemical approaches (25, 31, 62), have indicated that caps may be attached to the triphosphate termini resulting from transcriptional initiation events and that this occurs very early after initiation. Transcriptional initiation in eucaryotes by RNA polymerase II is thought to occur at purine nucleotides; however, we have recently shown (D. Gidoni, C. Kahana, and Y. Groner, manuscript in preparation) that SV40 multiple 5' caps represent transcriptional initiation sites, which implies that eucaryotic RNA polymerase II can initiate transcription with pyrimidines, reminiscent of recently reported procaryotic initiation with CTP (17, 54, 66).

Localization of the early caps. Using the primer extension method, Reddy et al. (49) have recently demonstrated two major, equally abundant, reverse transcriptase stops on early mRNA at residues 5147 to 5148 and 5152 to 5154 (Reddy et al. [51] numbering system) and one minor stop just short of these sites. A comparison of the nucleotide sequence at these regions with the early capped dinucleotides revealed that they are complementary to residues in this region and can be aligned as a cluster within 10 nucleotides (Fig. 5). The reverse transcriptase stop ATT, identified by Thompson et al. (59), which may be related to the minor AU caps is also in this region. Alternatively, this minor m^{7}GpppAmpU may represent a late cap derived from the low proportion of late SV40 mRNA found in this experiment (Fig. 2). Another minor cap, m^{7}GpppGmpU, has a complementary site 15 residues downstream and may be compatible to the Reddy et al. (49) minor terminus. A point

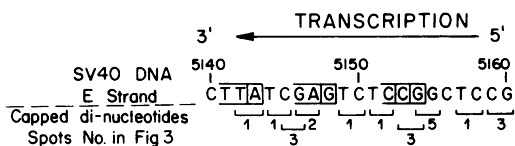

FIG. 5. *Alignment of early capped dinucleotides on SV40 DNA sequences in the region corresponding to the reverse transcriptase stops determined by Reddy et al. (49) and Thompson et al. (59). The numbering of the DNA residues was that of Reddy et al. (51). Dinucleotide numbers identify the corresponding spot in Fig. 3. Residues within brackets respresent 5' termini determined by the primer extension method: CCG and GAG, Reddy et al. (49); TTA, Thompson et al. (59).*

of interest is the correlation between the number of CU residues in this region (Fig. 5) and the relatively high proportion of m^{7}GpppCmpU caps, reminiscent of the situation in late SV40 mRNA with the most abundant AU cap (12, 28, 35, 50). As mentioned earlier, multiple 5' termini which are transcribed from adjacent nucleotides on the viral genome have been noted before in polyoma late mRNA's (22) and Ad2 early mRNA's (6). However, a different situation occurs in late SV40 mRNA, in which 5'-capped termini are not restricted to a short oligonucleotide (12, 35) and terminal heterogeneity of the leader is more extensive. In fact, 5' ends of late SV40 mRNA were reported at between 0.59 and 0.75 map unit (28, 42, 50, 60). Thus, although both early and late SV40 mRNA's display 5' heterogeneity, they still may differ in the mechanism of transcriptional initiation (for example, the Hogness-Goldberg box is present in the SV40 early, but not late, transcriptional unit).

Multiple 5' caps and Hogness-Goldberg box. Several different eucaryotic genes contain upstream from their 5'-cap sites an A-T-rich sequence, TATTTAT (Hogness-Goldberg box), which conceivably could serve as the transcriptional initiation signal (67). As noted by Baker et al. (6) and Flavell et al. (22), such an A-T-rich signal is missing, or it is present in an inappropriate location, upstream from the coding regions of the viral mRNA's that have been reported to contain multiple 5' caps, i.e., late polyoma virus (23), late SV40 (12), and early Ad2 region EII (6). However, as Lebowitz and Weissmann (45) have pointed out, in early SV40, a TATTTAT sequence lies 21 and 27 nucleotides upstream from the two principal 5' termini identified by Reddy et al. (49) and to which caps were assigned (Fig. 5). Moreover, Gluzman et al. (29a) recently obtained strong indications that either this A-T-rich region or sequences adjacent to it play a role in transcriptional initiation in vivo. Nevertheless, as this report shows, the

early mRNA's contain multiple 5'-terminal caps, as do the late SV40 and polyoma virus and early Ad2 mRNA's. Thus, the 5'-heterogeneity phenomenon is not necessarily related to the presence or the absence of a Hogness-Goldberg box upstream from the 5'-cap sites.

ACKNOWLEDGMENTS

We thank M. Revel for his encouragement and support and A. Mukamel for her excellent technical assistance.

Support for this work was provided by funds from the U.S.-Israel Binational Foundation, Jerusalem, Israel. Y.G. is an incumbent of the Helena Rubinstein Career Development Chair.

LITERATURE CITED

1. **Acheson, N. H.** 1976. Transcription during productive infection with polyoma and simian virus 40. Cell 8:1-12.
2. **Aloni, Y.** 1975. Transcription of SV40. IV. Methylated SV40 mRNAs. FEBS Lett. 54:363-367.
3. **Aloni, Y., R. Dhar, and G. Khoury.** 1979. Methylation of nuclear simian virus 40 RNAs. J. Virol. 32:52-60.
4. **Aloni, Y., R. Dhar, O. Laub, M. Horowitz, and G. Khoury.** 1978. Novel mechanism for RNA maturation: the leader sequences of simian virus 40 are not transcribed adjacent to the coding sequences. Proc. Natl. Acad. Sci. U.S.A. 74:3686-3690.
5. **Alwine, J. C., R. Dhar, and G. Khoury.** 1980. A small RNA induced late in simian virus 40 infection can associate with early viral mRNAs. Proc. Natl. Acad. Sci. U.S.A. 77:1379-1383.
6. **Baker, C. C., J. Herisse, G. Courtois, F. Galibert, and E. Ziff.** 1979. Messenger RNA for the Ad-2 binding protein: DNA sequences encoding the first leader and heterogeneity at the mRNA 5' end. Cell 18:569-580.
7. **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.
8. **Bina-Stein, M., M. Thoren, N. Salzman, and J. A. Thompson.** 1979. 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.
9. **Birkenmeier, E. H., N. Chiu, M. F. Radonovich, E. May, and N. P. Salzman.** 1979. Regulation of simian virus 40 early and late gene transcription without viral DNA replication. J. Virol. 29:983-989.
10. **Bratosin, S., M. Horowitz, O. Laub, and Y. Aloni.** 1978. Electron-microscopic evidence for splicing of SV40 late mRNAs. Cell 13:783-790.
11. **Canaani, D., C. Kahana, S. Lavi, and Y. Groner.** 1979. Identification and mapping of N^6-methyladenosine containing sequences in simian virus 40 RNA. Nucleic Acids Res. 6:2879-2899.
12. **Canaani, D., C. Kahana, A. Mukamel, and Y. Groner.** 1979. Sequence heterogeneity at the 5'-termini of late simian virus 40 19S and 16S mRNAs. Proc. Natl. Acad. Sci. U.S.A. 76:3078-3082.
13. **Celma, M. L., R. Dhar, J. Pan, and S. M. Weissmann.** 1977. Comparison of the nucleotide sequence of the messenger RNA for the major structural protein of SV40 with the DNA sequence encoding the amino acids of the protein. Nucleic Acids Res. 4:2549-2559.
14. **Chou, J. Y., and R. G. Martin.** 1974. Complementation analysis of simian virus 40 mutants. J. Virol. 13:1101-1109.
15. **Cory, S., and J. Adams.** 1975. The modified 5'-terminal sequences in messenger RNA of mouse myeloma cells. J. Mol. Biol. 99:519-547.
16. **Darnell, J. E.** 1979. Transcription units for mRNA production in eukaryotic cells and their DNA viruses. Prog. Nucleic Acid Res. Mol. Biol. 22:327-353.
17. **de Boer, H., and M. Nomura.** 1979. *In vivo* transcription of rRNA operons in *Escherichia coli* initiates with purine nucleoside triphosphates at the first promoter and with CTP at the second promoter. J. Biol. Chem. 254:5609-5612.
18. **Dhar, R., K. N. Subramanian, J. Pan, and S. M. Weissmann.** 1977. Structure of a large segment of the genome of simian virus 40 that does not encode known proteins. Proc. Natl. Acad. Sci. U.S.A. 74:827-831.
19. **Dhar, R., K. N. Subramanian, J. Pan, and S. M. Weissmann.** 1977. Nucleotide sequence of a fragment of SV40 DNA that contains the origin of DNA replication and specifies the 5' ends of early and late viral RNA. J. Biol. Chem. 252:368-376.
20. **Ferdinand, F., M. Brown, and G. Khoury.** 1977. Characterization of early simian virus 40 transcriptional complexes: late transcription in the absence of detectable DNA replication. Proc. Natl. Acad. Sci. U.S.A. 74:5443-5447.
21. **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.
22. **Flavell, A. J., A. Cowie, J. R. Arrand, and R. Kamen.** 1980. Localization of three major capped 5' ends of polyoma virus late mRNA's within a single tetranucleotide sequence in the viral genome. J. Virol. 33:902-908.
23. **Flavell, A. J., A. Cowie, S. Legon, and R. Kamen.** 1979. Multiple 5'-terminal cap structures in late polyoma virus RNA. Cell 16:357-371.
24. **Fraser, N. W., P. H. Sehgal, and J. E. Darnell.** 1979. Multiple discrete sites for premature RNA chain termination late in adenovirus-2 infection: enhancement by 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB). Proc. Natl. Acad. Sci. U.S.A. 76:2571-2575.
25. **Furuichi, Y.** 1978. Pretranscriptional capping in the biosynthesis of cytoplasmic polyhedrosis virus mRNA. Proc. Natl. Acad. Sci. U.S.A. 75:1086-1090.
26. **Furuichi, Y., M. Morgan, A. Shatkin, W. Jelinek, M. Salditt-Georgieff, and J. Darnell.** 1975. Methylated blocked 5'-termini in HeLa cell mRNA. Proc. Natl. Acad. Sci. U.S.A. 72:1904-1908.
27. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. Choudary, P. Lebowitz, and S. M. Weissmann.** 1978. The 5'-terminal leader sequence of late mRNA from cells infected with simian virus 40. J. Biol. Chem. 253:3643-3647.
28. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. M. Weissmann.** 1978. Heterogeneity and 5'-terminal structures of the late RNAs of simian virus 40. J. Mol. Biol. 126:813-846.
29. **Gluzman, Y., J. Davison, M. Oren, and E. Winocour.** 1977. Properties of permissive monkey cells transformed by UV-irradiated simian virus 40. J. Virol. 22:256-266.

29a. **Gluzman, Y., J. F. Sambrook, and R. J. Frisque.** 1980. Expression of early genes of origin-defective mutants of SV40. Proc. Natl. Acad. Sci. U.S.A. 77:3898-3902.

30. **Groner, Y., P. Carmi, and Y. Aloni.** 1977. Capping structures of simian virus 40 19S and 16S mRNAs. Nucleic Acids Res. 4:3958-3968.
31. **Groner, Y., E. Gilboa, and H. Aviv.** 1978. Methylation and capping of RNA polymerase II primary transcripts by HeLa nuclear homogenates. Biochemistry 17:977-982.
32. **Groner, Y., H. Grosfeld, and U. Z. Littauer.** 1976. 5' Capping structures of artemia salina mRNA and the translational inhibition by cap analogs. Eur. J. Biochem. 71:281-293.
33. **Groner, Y., and J. Hurwitz.** 1975. Synthesis of RNA containing a methylated blocked 5'-terminus by HeLa nuclear homogenates. Proc. Natl. Acad. Sci. U.S.A. 72:

2930-2934.

34. **Haegeman, G., and W. Fiers.** 1978. Evidence for splicing of SV40 16S mRNA. Nature (London) **273**:70-73.
35. **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.
36. **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.
37. **Hashimoto, S., and M. Green.** 1979. Methylated 5'-terminal caps of adenovirus type 2 early mRNA: evidence for at least six 5'-termini. Virology **94**:254-272.
38. **Hashimoto, S., M. H. Pursley, W. S. M. Wold, and M. Green.** 1980. Characterization of distinct 5'-terminal cap structures of adenovirus type 2 early messenger ribonucleic acid and KB cell messenger ribonucleic acid. Biochemistry **19**:294-300.
39. **Hsu, M., and J. Ford.** 1978. Sequence arrangements of the 5' ends of simian virus 40 16S and 19S mRNAs. Proc. Natl. Acad. Sci. U.S.A. **74**:4982-4985.
40. **Kaufmann, G., R. Bar-Shavitt, and M. L. DePamphilis.** 1978. Okazoki pieces grow opposite to the replication form direction during simian virus 40 DNA replication. Nucleic Acids Res. **5**:2535-2545.
41. **Khoury, G., and E. May.** 1977. Regulation of early and late simian virus 40 transcription: overproduction of early viral RNA in the absence of a functional T-antigen. J. Virol. **23**:167-176.
42. **Lai, C.-J., R. Dhar, and G. Khoury.** 1978. Mapping the spliced and unspliced late lytic SV40 RNAs. Cell **14**: 971-982.
43. **Lavi, S., and Y. Groner.** 1978. 5'-Terminal sequences and coding regions of late simian virus 40 mRNAs are derived from non-contiguous segments of the viral genome. Proc. Natl. Acad. Sci. U.S.A. **74**:5323-5327.
44. **Lavi, S., and A. G. Shatkin.** 1975. Methylated SV40-specific RNA from nuclei and cytoplasm of infected BSC-1 cells. Proc. Natl. Acad. Sci. U.S.A. **72**:2012-2016.
45. **Lebowitz, P., and S. M. Weissmann.** 1979. Organization and transcription of the simian virus 40 genome. Curr. Top. Microbiol. Immunol. **87**:43-172.
46. **Manley, J. L., P. A. Sharp, and M. L. Gefter.** 1979. RNA synthesis in isolated nuclei: *in vitro* initiation of adenovirus 2 major late mRNA precursor. Proc. Natl. Acad. Sci. U.S.A. **76**:160-164.

46a. **Mark, D. F., and P. Berg.** 1980. A third splice site in SV40 early mRNA. Cold Spring Harbor Quant. Biol. **44**:55-62.

47. **Mizumoto, K., and F. Lipman.** 1979. Transmethylation and trans-guanylation in 5'-RNA capping system isolated from rat liver nuclei. Proc. Natl. Acad. Sci. U.S.A. **76**:4961-4965.
48. **Perry, R. P., D. E. Kelly, K. Friderichi, and F. Rottman.** 1975. The methylated constituents of L cell messenger RNA: evidence for an unusual cluster at the 5'-terminus. Cell **4**:387-394.
49. **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.
50. **Reddy, V. B., P. K. Ghosh, P. Lebowtiz, and S. M. Weissmann.** 1978. Gaps and duplication sequences in the leaders of SV40 16S RNA. Nucleic Acids Res. **5**: 4195-4214.
51. **Reddy, V. B., B. Thimmappaya, R. Dhar, K. N. Subramanian, B. S. Zain, J. Pan, P. K. Ghosh, M. L. Celma, and S. M. Weissmann.** 1978. The genome of simian virus 40. Science **200**:494-502.
52. **Reed, S. I., and C. J. Alwine.** 1977. An analyses by electron microscopy of early simian virus 40 RNA from a *tsA* mutant. Cell **11**:523-531.
53. **Revel, M., and Y. Groner.** 1978. Post-transcriptional and translational controls of gene expression in eukaryotes. Annu. Rev. Biochem. **47**:1079-1126.
54. **Rosenberg, M., D. Court, H. Shimotoke, and C. Brady.** 1978. Structure and function of an intercistronic regulatory region in bacteriophage lambda, p. 345-371. *In* J. H. Miller and S. Reznikoff (ed.), The operon. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
55. **Rosenthal, L. J., and M. Brown.** 1977. The control of SV40 transcription during a lytic infection: late RNA synthesis in the presence of inhibitors of DNA replication. Nucleic Acids Res. **4**:551-565.
56. **Salditt-Georgieff, M., M. Harpold, S. Chen-King, and J. E. Darnell.** 1980. The addition of 5'-cap structures occur early in hnRNA synthesis and prematurely terminated molecules are capped. Cell **19**:69-78.
57. **Salzman, N. P., N. Birkenmeier, N. Chiu, E. May, M. F. Radonovich, and M. Shani.** 1977. Regulation of transcription during the SV40 lytic cycle. INSERM Colloq. **69**:243-252.
58. **Shatkin, A. J.** 1976. Capping of eukaryotic mRNAs. Cell **9**:645-653.
59. **Thompson, J. A., M. F. Radonovich, and N. P. Salzman.** 1979. Characterization of the 5'-terminal structure of simian virus 40 early mRNA's. J. Virol. **31**:437-438.
60. **Villarreal, L. P., R. T. White, and P. Berg.** 1979. Mutational alterations within the simian virus 40 leader segment generate altered 16S and 19S mRNA's. J. Virol. **29**:209-219.
61. **Wei, C.-M., A. Gershowitz, and B. Moss.** 1975. Methylated nucleotides block 5'-terminus of HeLa cell messenger RNA. Cell **41**:379-386.
62. **Wei, C.-M., and B. Moss.** 1977. 5'-Termini capping of RNA by guanylyltransferase from HeLa cell nuclei. Proc. Natl. Acad. Sci. U.S.A. **74**:3758-3761.
63. **Weil, P. A., D. S. Luse, J. Segall, and R. G. Roeder.** 1979. Selective and accurate initiation of transcription at the Ad-2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA. Cell **18**:469-484.
64. **Weil, R., C. Salomon, E. May, and P. May.** 1974. A simplifying concept in tumor virology: virus-specific pleotropic effectors. Cold Spring Harbor Symp. Quant. Biol. **37**:381-395.
65. **Winicov, I., and R. Perry.** 1976. Synthesis, methylation and capping of nuclear RNA by a subcellular system. Biochemistry **15**:5039-5046.
66. **Young, R. A., and J. A. Steitz.** 1979. Tandem promoters direct *E. coli* ribosomal RNA synthesis. Cell **17**:225-234.
67. **Ziff, E. B., and R. M. Evans.** 1978. Coincidence of the promoter and capped 5' terminus of RNA from adenovirus 2 major late transcription unit. Cell **15**:1463-1475.