JOURNAL OF VIROLOGY, Aug. 1979, p. 484-493
0022-538X/79/08-0484/10$02.00/0

Vol. 31, No. 2

Heterogeneity of the 5′ Terminus of Late mRNA Induced by a Viable Simian Virus 40 Deletion Mutant

GUY HAEGEMAN, HUGO VAN HEUVERSWYN, DIRK GHEYSEN, AND WALTER FIERS*

Laboratory of Molecular Biology, State University of Ghent, B-9000 Ghent, Belgium

Received for publication 21 February 1979

dl-1811 is a viable simian virus 40 deletion mutant which lacks the DNA region corresponding to the major capping site of the late viral RNA. The exact size of the deletion (40 base pairs) was determined by comparison of the mutant DNA sequence with the wild-type simian virus 40 (strain 776) DNA sequence. Although dl-1811 forms somewhat smaller plaques, the amount of viral RNA late after infection was not significantly reduced compared with that of the wild type. Virus-specific, polyadenylate-containing, ^{32}P-labeled late RNA was purified from the cytoplasm and enzymatically degraded to characterize the 5′ terminus. The cap-containing oligonucleotides were isolated, and their structures were analyzed by further digestion. Instead of a single cap structure, we found a variety of capped 5′ termini, with adenosine caps occurring much more frequently than guanosine caps. Nevertheless, there was a remarkable homology between both types of terminal sequences. Conceivably, the minor cap population present in wild-type simian virus 40 late mRNA may correspond to the collection of capped termini identified in the dl-1811 late mRNA. Cellular cytoplasmic RNA shows a similar pattern of cap structures, but the relative abundance is quite different.

Elucidation of the processes involved in initiation and 5′-terminal modification of eucaryotic cellular or viral mRNA's is an intriguing problem. The structure of the 5′-end group, the enzymatic mechanism leading to "cap" formation, and the possible function of capped structures in subsequent processing or in translation of mRNA or in both have been the subject of intensive investigation (for a review, see references 5, 17, 29, 30).

The late mRNA species of the eucaryotic virus simian virus 40 (SV40) also have a capped 5′ terminus (20), whose structure has been described in detail (11, 12, 32). The sequence of the leader fragment, i.e., the segment between the cap structure and the coding sequence, has been characterized both by fingerprinting (3, 13) and by reversed transcription (2, 9). Moreover, the region of the viral genome that specifies the capped terminal sequences derived from the major untranslated leader of the viral late RNA has been identified (14). A number of viable deletion mutants of SV40 have been described which lack different parts of the genome between 0.68 and 0.74 map unit (24, 31). Conceivably, some of these may have lost the segment corresponding to the major capping site of the transcript. As capping and initiation of RNA transcription are presumably closely related or interdependent phenomena (7, 8, 36, 38), it was of interest to determine how a mutant which lacks the normal major capping site of late RNA overcomes this handicap. Therefore, a mutant which has lost the DNA region corresponding to the major 5′-capping site was isolated and characterized by nucleotide sequence analysis. The 5′ termini of the mutant late mRNA molecules were then identified in detail. We found that the mRNA population starts with a variety of different cap structures. The fact that the amount of late mRNA was approximately unchanged means that these alternative sites for RNA initiation (or processing) become activated when the normal major capping site of SV40 wild-type virus is eliminated.

MATERIALS AND METHODS

Virus stocks and cells. SV40 deletion mutant dl-1811 is a spontaneously generated mutant which lacks the *Hha*I restriction site at map position 0.726 and which maintains the *Hpa*II site that partly overlaps the former restriction site in the clockwise direction (see Fig. 2); dl-1811 was selected on this basis. Plaque assays with wild-type SV40 (strain 776) and dl-1811 were performed on monolayers of CV-1 monkey kidney cells. dl-1811 is viable but gives rise to smaller plaques as compared with wild-type plaques.

Nucleotide sequence analysis. The size of the deletion in dl-1811 was estimated by restriction enzyme analysis, and the actual number of deleted base pairs was determined by direct DNA sequence analysis according to the chemical degradation procedure of Maxam and Gilbert (23).

Preparation of [^{32}P]mRNA and analysis of cap structures. Procedures for infecting confluent monolayers of CV-1 cells with dl-1811, labeling with $^{32}P_i$, and isolation and purification of cytoplasmic, virus-specific, polyadenylate-containing RNA are as described previously for wild-type SV40 virus (12). Usually, virus-specific RNA selected by hybridization onto Sepharose-bound SV40 DNA was rehybridized to viral DNA immobilized on amino-cellulose (27). This second hybridization yielded highly purified RNA fractions, as analyzed by RNase T_1 digestion and minifingerprinting (13). Generation of the 5'-terminal capped structures by RNase T_2 hydrolysis and subsequent two-dimensional fractionation was as described previously (12). To obtain a quantitative distribution of the different capped termini, each of the spots on the polyethyleneimine (PEI)-cellulose plate was cut out, and its radioactivity was counted by liquid scintillation. The nucleotide material was recovered from the PEI-cellulose by a microelution technique and further analyzed with *Penicillium* nuclease P_1 or with nucleotide pyrophosphatase (12). Due to incomplete separation of the 5' mononucleotides pC and pA from their methylated derivatives in the second dimension of the PEI miniplate, it was occasionally useful to fractionate the P_1 digests of some cap structures two-dimensionally on cellulose plates (20 by 20 cm; Macherey-Nagel & Co., Düren, Federal Republic of Germany) as described by Nishimura (26). However, because in this system the 5' mononucleotide pC now coincides with the 2'-*O*-methylated pG (pGm), it was necessary in two particular cases (see Results) to confirm the composition by one-dimensional chromatography with 0.2 M lithium formate (pH 3.0) on PEI-cellulose (34).

RESULTS

Characterization of the deleted region of the dl-1811 genome. Preliminary characterization of dl-1811 DNA with various restriction enzymes (*HhaI*, *HpaII*, *HindII* + III, and *BglI*) had indicated that the deletion was approximately 25 to 30 base pairs long and was located in *HindII* + III fragment C between the *BglI* and *HpaII* sites (map positions 0.660 and 0.726, respectively). Further analyses indicated that the size of the *PvuI-BglI* fragment (0.712 to 0.660 map unit) was identical to that of wild type, as was its *EcoRII* digestion pattern. Hence, on the basis of these restriction enzyme analyses, we can conclude that the deleted region is localized in the area between the *HpaII* site (0.726 map unit) and the *PvuII* site (0.712 map unit). Because the DNA was resistant to cleavage by *HhaI* at position 0.725 but was still sensitive to *HpaII* cleavage, it was digested by the latter enzyme to linear (form III) molecules. These were 5'-terminally labeled by means of T_4 polynucleotide kinase and [γ-^{32}P]ATP and digested with *BglI* into two singly labeled subfragments, the smaller of which contained the deletion. After separation by gel electrophoresis, the smaller subfragment was partially degraded as described by Maxam and Gilbert (23). The degradation pattern is shown in Fig. 1; it permits reading of the nucleotide sequence counterclockwise from the *HpaII* cleavage site and reveals a deletion of 40 base pairs, including the segment which specifies the 5' terminus of the major untranslated leader fragment of SV40 late mRNA (14). By analysis of the complementary strand starting from the labeled *PvuII* site at position 0.712 and going in the clockwise direction, at least 100 nucleotides could be read from the gel, and these results confirmed the position of the dl-1811 deletion. This deletion and the 5'-terminal segment of the major wild-type mRNA leader are illustrated in Fig. 2 (it may be noted that the major 5'-capping site is the same for both 16S and 19S mRNA's [9, 10, 12, 14]).

Isolation of viral mRNA. Although one of the phenotypic characteristics of dl-1811 is the small size of its plaques relative to wild-type SV40 (strain 776) plaques, we did not observe a decreased production of viral late mRNA at 48 h postinfection, i.e., at the time we harvested the RNA. Mutant SV40-specific mRNA usually represented 5 to 10% of the total cytoplasmic polyadenylate-containing RNA of the infected cells, and this amount was not substantially different from the wild-type viral [^{32}P]mRNA fraction. Preliminary experiments indicated an extensive heterogeneity of 5'-terminal structures (data not shown). Therefore, to avoid residual cellular mRNA contamination, viral late RNA was purified by two successive hybridizations with viral DNA immobilized on different matrices: the first hybridization was to Sepharose-bound SV40 DNA, and the second was to cellulose-bound viral DNA. At most, 15 to 20% of the material selected by the first hybridization step did not rehybridize to the cellulose-bound SV40 DNA and was of nonviral origin, as evidenced by fingerprinting of RNase T_1 digests of nonbound material. Additional radioactivity was released from the matrix by a number of successive washes; the wash with the so-called clean-up buffer, containing low salt and 50% formamide (27), consisted already almost exclusively of viral RNA sequences with only minute amounts of foreign RNA. The RNA fraction recovered by the final elution represented 50% of the input ^{32}P radioactivity and was considered to be completely SV40 specific, as T_1 oligonucleotides of nonviral RNA origin were not detectable in the T_1 fingerprints. These patterns further showed that the relative abundance of late 19S RNA was normal (10 to 20%) and that the amount of early RNA was very small (<5%), in agreement with previously published results on wild-type mRNA (19, 37).

Isolation and identification of capped ter-

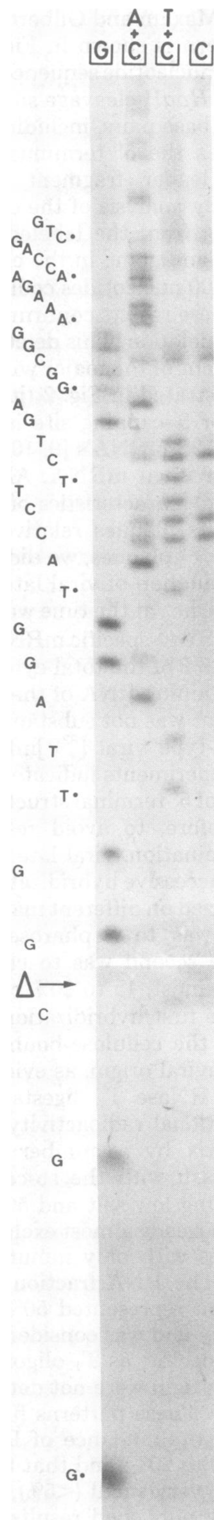

FIG. 1. Chemical degradation pattern of the BglI-HpaII restriction fragment of dl-1811 DNA. The fragment was degraded as described by Maxam and Gilbert (23). The sequence of the strand that codes for late mRNA (l strand) can be read continuously from the HpaII cleavage site (map position 0.726) in the counterclockwise direction (see Fig. 2). The position of the deletion is indicated by Δ.

mini. The mutant late mRNA was digested with RNase T_2 and bacterial alkaline phosphatase, and the hydrolysate was separated two-dimensionally on PEI-cellulose thin-layer plates. A typical degradation pattern is shown in Fig. 3; besides some tailing from the $^{32}P_i$, a great variety of products could be seen on that part of the plate where the terminal cap structures were assumed to occur. All of these were eluted and upon further analysis with P_1 nuclease were found to contain P_1 cap cores (see below). Hence, the multiplicity observed was not due, e.g., to incomplete T_2 hydrolysis (which would mainly yield internal oligonucleotides), but actually represented a number of different 5'-terminal structures, all of which were derived from the highly purified mutant late mRNA. The ^{32}P radioactivity present in the entire cap population amounted to about 0.1% of the input radioactivity, and this agrees with the proportional cap content present in wild-type SV40 RNA (12).

All the T_2-resistant 5'-terminal structures were further hydrolyzed with P_1 nuclease, and the products were separated by two-dimensional chromatography on PEI miniplates (35). The most important results are shown in Fig. 4. The main common spot, near the blue dye marker xylene cyanol, was the P_1 cap core ^{7m}Gppp^{m}Am (the methylation state of the adenosine is further established below) (pXm stands for a nucleotide methylated at the 2'-*O*-position of the sugar residue, $\overset{m}{p}$X indicates that the methylation occurs on the base moiety of the nucleotide, and products modified at both positions are denoted by $\overset{m}{p}$Xm). The common spot on the other miniplates, near the yellow dye marker, was the alternative purine cap core ^{7m}GpppGm. It may be noted that in most (if not all) cases, the major adenosine cap (A-cap) as well as the major guanosine cap (G-cap) were accompanied to a certain extent by a minor product migrating slightly slower in both dimensions. These minor spots were initially thought to originate from incomplete P_1 hydrolysis (14), but we obtained evidence that they corresponded to P_1 cap cores with a ring-opened p^{7m}G residue (data not shown). This reaction, which is greatly enhanced at alkaline pH (21), may have occurred when the T_2-generated termini were eluted from the PEI miniplate with 2 M triethylammonium carbonate (pH 10). Ring opening is responsible for the loss of a fractional positive charge from the

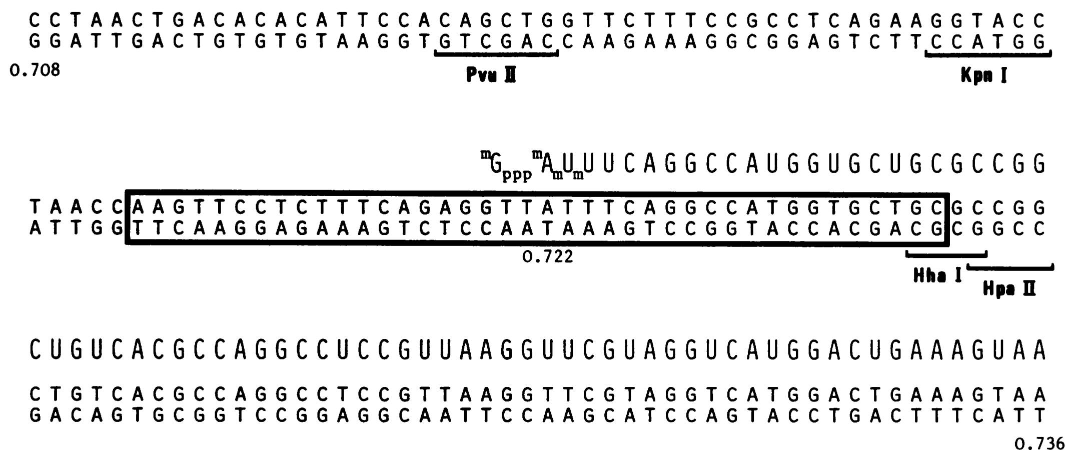

FIG. 2. *Nucleotide sequence of the SV40 DNA region extending from 0.708 to 0.736 map unit. The nucleotide region absent from dl-1811 DNA is enclosed in the box; it comprises 40 base pairs and includes the major capping site of the viral late mRNA at position 0.722. The sequence of the major leader fragment of wild-type mRNA is shown above and colinearly with the DNA sequence, beginning from the capping site (13, 14). The recognition sequences of relevant restriction endonucleases are indicated. Note that the nucleotide sequence of wild-type SV40 DNA in the segment deleted in dl-1811 has been corrected by insertion of an additional T-G dinucleotide (H. Van Heuverswyn and W. Fiers, manuscript in preparation).*

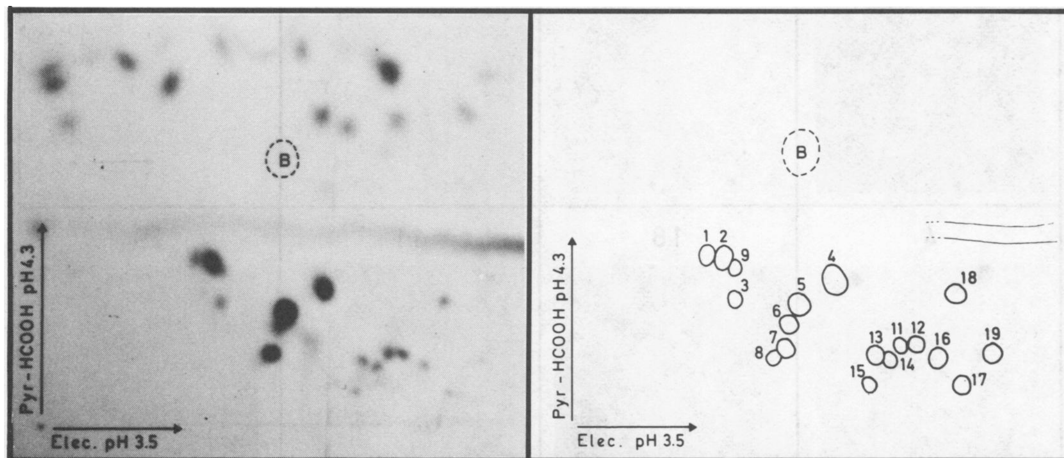

FIG. 3. *Two-dimensional fractionation on PEI-cellulose of a combined RNase T_2 and bacterial alkaline phosphatase digest of ^{32}P-labeled dl-1811 late mRNA. The first dimension was electrophoresis at pH 3.5 in 5 M urea, and the second dimension consisted of chromatography with 1 M pyridine-formic acid, pH 4.3. The capped products are numbered, as indicated in the accompanying diagram, and their identification is summarized in Table 1. B denotes the position of the blue dye marker xylene cyanol.*

base-methylated pG residue; hence, ring-opened cap cores are more tightly bound to the PEI ion exchanger and have retarded mobility in both chromatographic dimensions as compared with the undegraded products.

The positions of the mononucleotides could be predicted from an earlier publication (35), considering that 3′ nucleotides and 5′ nucleotides migrate with about the same mobility in the system used. Methylated nucleotides moved only slightly faster in the first dimension but had considerably greater mobility in the second dimension compared with the corresponding nonmethylated residues. The positions of various nucleotides were established by a number of experiments with ^{32}P-labeled digestion products derived from viral or cellular mRNA's and comparisons with different methylated and nonmethylated reference compounds revealed by UV illumination (optical density references). On the basis of the characteristic mobilities of the degradation products, the structure of a given T_2-resistant cap could be directly deduced from simple visual inspection of the P_1 digestion pat-

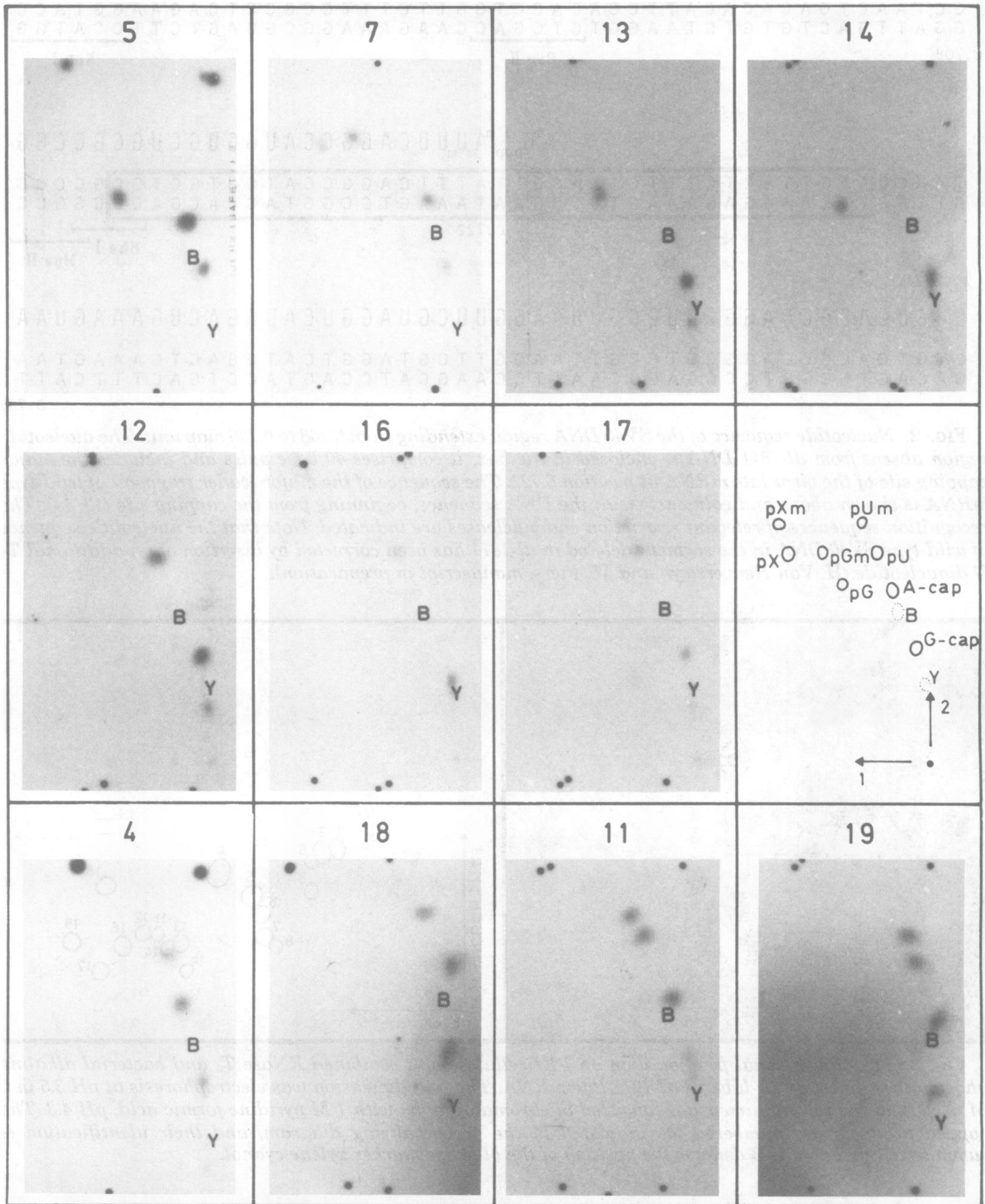

FIG. 4. *Two-dimensional separation on PEI-cellulose miniplates of the P_1 analysis of several T_2-resistant, capped products which were resolved by fractionation in two dimensions (Fig. 3). The separation was as described by Volckaert and Fiers (35). The radioactive spots can be deduced by comparison with the diagram, which shows the locations of different hydrolysis products. pX denotes either pC or pA, which are not always well resolved in this chromatographic system (see text for further identification). B and Y mark the positions of the dye markers xylene cyanol (blue) and methyl orange (yellow), respectively. Small (radioactive) dots (on top and bottom) were used as references for superposition of the autoradiograms and optical density patterns.*

terns. However, as there was poor and sometimes inconclusive resolution between the mononucleotides pC, pA, and their methylated relatives, the P_1 analyses of cap structures containing pC or pA residues were confirmed in different chromatographic systems (see below).

It is remarkable that certain T_2-generated caps with different mobilities in the electropho-

retic (first) dimension (Fig. 3) displayed P_1 digestion patterns identical to each other on the PEI miniplates (products 12 and 16, 13 and 14, 4 and 18, and 11 and 19 in Fig. 4). Although the actual reason for this phenomenon is not known, the apparent heterogeneity did not originate from differences in composition (as revealed by repeated P_1 analysis), nor did it arise from different methylation states (as shown by nucleotide pyrophosphatase treatment of the cap cores involved) or from a ring-opened $p^{7m}G$ residue (which would have resulted in a considerable lag of the affected product in the chromatographic [second] dimension of the fractionation system). Possibly, a different spatial configuration of the cap structure (16) or the effects of salt, which greatly alters electrophoretic mobility on cellulose acetate, may have been responsible for the phenomenon.

5′-Terminal sequences containing methylated or unmethylated pC or pA residues in the penultimate position (e.g., product 7 in Fig. 4) were further characterized on a two-dimensional chromatography system in which pC and pA were clearly distinguishable from the corresponding methylated derivatives. Figure 5 shows the separation pattern of the P_1 hydrolysates of compounds 2, 3, and 6 (see Fig. 3). Again, the compositions of the capped products could be directly determined by visual inspection of the plate and by referring to added optical density marker nucleotides. However, because in this system pGm migrated with pC, it was useful (though not essential) to prove by a complementary experiment the authenticity of the pGm residue. Such a separation was performed by thin-layer chromatography at pH 3.0 on PEI-cellulose plates (34) for products 7 and 15, the only two capped termini that contained a pGm residue in combination with either pC or pA, as revealed by the preceding analysis on PEI miniplates.

To establish the methylation state of the adenosine residue of the A-cap cores, the mutant late mRNA was totally degraded with P_1 nuclease. Two-dimensional fractionation of the products yielded the purine cap cores in a ratio of A-cap to G-cap of about 8:2, together repre-

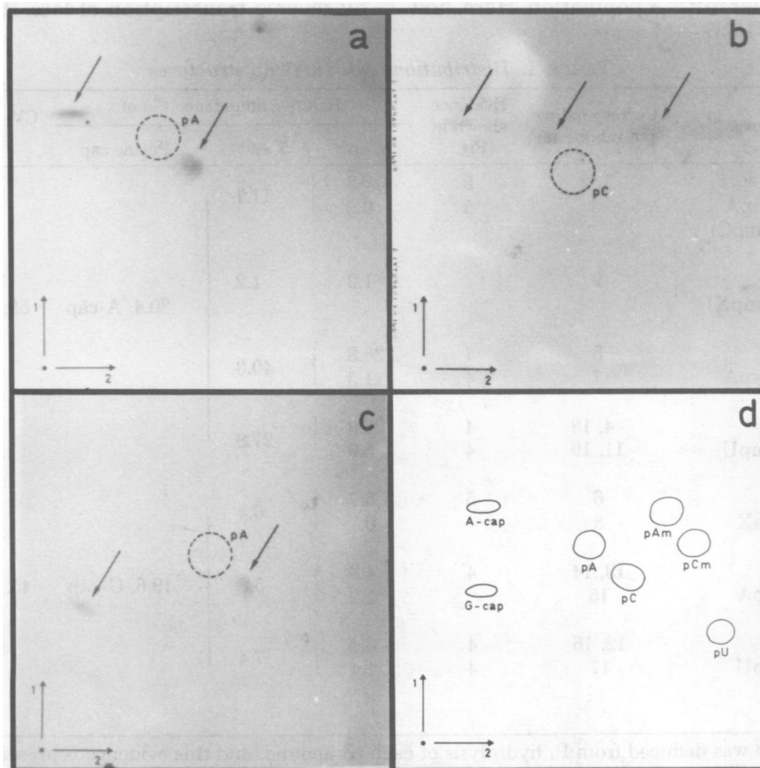

FIG. 5. *Two-dimensional chromatographic separation on cellulose plates (20 by 20 cm) of the P_1 analysis of capped products 2 (a), 3 (b), and 6 (c) (see Fig. 3). The first dimension was developed with isobutyric acid–0.5 M ammonia, 5:3 (vol/vol); the second dimension was run with isopropanol-concentrated HCl-water, 70:15:15 (vol/vol), as described by Nishimura (26). The identities of the radioactive spots were deduced from their unique positions compared with internal optical density references, which are indicated by dashed circles. The accompanying diagram (d) shows the locations of the relevant products.*

senting 0.09% of the total radioactivity of the RNA. The products were recovered and further analyzed by means of nucleotide pyrophosphatase. The G-cap revealed only pGm as the penultimate nucleotide and no nonmethylated pG residues; the A-cap showed exclusively N^6-methyl-pAm (p^{6m}Am) in this position (data not shown). Hence, we concluded that all the A-cap cores present in the T_2-resistant termini contained a doubly methylated pA residue, as previously reported for wild-type viral RNA (12).

This last analysis completed the structural determination of all the different capped termini present in dl-1811 late mRNA. The structures as well as their relative abundances are summarized in Table 1. Each cap II structure is classified next to the cap I terminus from which it is presumably derived, and their joint radioactivity is considered to represent one particular 5′ end. There appeared to be a great preference for A-cap structures to G-cap termini, in agreement with the ratio 8:2 deduced by a direct analysis of the P_1-generated cap cores (see above). Also included in Table 1 is the relative distribution of the different T_2-generated cap structures from the CV-1 cellular mRNA population. Here, however, the relative amount of G-caps was much higher, viz. 45%.

DISCUSSION

Previously, it has been shown that late after wild-type SV40 infection of CV-1 cells, two cap structures are most prominent in the virus-specified RNA. They have been characterized as ^{7m}Gppp^{m}AmpU and ^{7m}Gppp^{m}AmpUmpU (12), and they are (mainly) derived from late mRNA, as only minute amounts of viral early RNA sequences occur late in infection (19, 37). The 16S species of late viral RNA has been fingerprinted, and the oligonucleotide content was found to be in perfect agreement with the nucleotide sequence of the late region of the SV40 genome (3, 13). A cap-containing RNase T_1 oligonucleotide, corresponding to the 5′-terminal sequence of the major untranslated leader segment of the viral late mRNA, was localized at SV40 map position 0.722 (14). The major leader constitutes an uninterrupted sequence of 202 nucleotides that is spliced onto the coding portion of the 16S mRNA (2, 3, 9, 13, 32). Recently, by reverse transcription of late 16S mRNA, al-

TABLE 1. *Distribution of dl-1811 cap structures*

Cap structure[a]	Corresponding spot number(s)[b]	Evidence shown in Fig.:	Relative abundance (%) of: Cap	5′ end[c]	Purine cap	CV-1 mRNA cap distribution (%)
^{m}Gppp^{m}AmpC	1, 2	5	10.2	11.1		9.9
^{m}Gppp^{m}AmpCmpA	3	5	0.9			0.5
(^{m}Gppp^{m}AmpCmpC)						1.0
^{m}Gppp^{m}AmpA	9		1.2	1.2		1.8
(^{m}Gppp^{m}AmpAmpX)					80.4, A-cap	55, A-cap: 1.0
^{m}Gppp^{m}AmpG	5	4	28.8	40.3		21.5
^{m}Gppp^{m}AmpGmpA	7	4	11.5			3.0
^{m}Gppp^{m}AmpU	4, 18	4	22.8	27.8		11.0
^{m}Gppp^{m}AmpUmpU	11, 19	4	5.0			5.6
^{m}GpppGmpC	6	5	5.7	6.4		24.9
^{m}GpppGmpCmpX	8		0.7			1.6
^{m}GpppGmpG	13, 14	4	4.6	5.8	19.6, G-cap	45, G-cap: 6.1
^{m}GpppGmpGmpA	15		1.2			1.3
^{m}GpppGmpU	12, 16	4	6.0	7.4		3.4
^{m}GpppGmpUmpU	17	4	1.4			1.0
(Others)						6.4

[a] Composition was deduced from P_1 hydrolysis of each compound, and this evidence is presented in Fig. 4 or 5. pX stands for either pC or pA. Structures in parentheses were observed in cellular mRNA but were not detected among the dl-1811 cap structures.

[b] See Fig. 3.

[c] Since cap I and cap II structures (30) with similar terminal sequences were assumed to be derived from the same 5′ terminus, the sum of their relative abundances was thought to represent a defined primary 5′ end.

ternative leaders, some of which are gapped, have been detected in minor quantities (28). Also, a great diversity of gapped and ungapped leader sequences has been observed after reverse transcription of 19S late mRNA, but the major 5′ terminus starts apparently at the same position as the major 16S mRNA (10), which agrees with our previous findings (14). Most of the alternative minor leader sequences, both in 16S and in 19S mRNA's, do not begin at position 0.722 and therefore may have cap structures different from the ones at the 5′ terminus of the major species. Also, characterization of the late mRNA's by the S_1 mapping technique has suggested the occurrence of additional 5′ termini (18, 33). It is of interest to note that the late RNA of polyoma virus (which is distantly related to SV40) displays a diversity of 5′-terminal cap structures which are similar in each of the different late mRNA species (6). This situation is consistent with a mechanism in which RNA transcription can start at multiple initiation points, producing 5′-terminally heterogeneous precursors which afterwards are spliced and processed to yield the mature viral mRNA's. A common precursor for SV40 16S and 19S RNAs has indeed been postulated (1, 4). However, it cannot at present be ruled out that 5′-terminal cap structures correspond to sites of processing of a larger primary transcript.

SV40 mutant dl-1811 lacks part of the DNA sequence corresponding to a fraction of the major leader, including the major capping site. Although dl-1811 produces smaller plaques than those of the wild-type virus, late mRNA production is not diminished. Assuming that the various mutant late mRNA species correspond to the minor mRNA's also synthesized in wild-type-infected cells (see below), there must exist some compensating regulatory mechanism such that these alternative species now account for the bulk of the late mRNA.

Analysis of the 5′ terminus of dl-1811 late mRNA revealed instead of a nearly unique cap structure a variety of capped products. They were all virus specific, because the mRNA was stringently selected by double hybridization and was highly purified, as ascertained by T_1 fingerprint analysis. As early sequences are relatively scarce late in infection (19, 37) and indeed were (almost) not detectable in the fingerprints, the variety of 5′ termini can be attributed solely to the late viral mRNA's. The relative abundance of the different terminal products in the viral RNA population clearly deviates from the overall distribution of capped structures present in the cellular cytoplasmic mRNA population (Table 1). Only purine caps have been found among the viral RNA termini, and in these the A-cap is much preferred to the G-cap, whereas in the cellular cap population there is a more nearly equimolar distribution of both purine caps. Possibly, mRNA's with A-caps are more efficiently translated, as it has been observed that vaccinia virus mRNA's with A-caps seem to be favored over G-capped molecules for ribosome binding (25).

Considering that in wild-type SV40 late mRNA the major cap II (^{7m}Gppp^{m}AmpUmpU) definitely originates from the same position on the genome as the major cap I (^{7m}Gppp^{m}AmpU) (14), it is likely that also in the mutant each pair of cap I and corresponding cap II is derived from the same 5′ end. It may be noted that the well-characterized eucaryotic α and β globin mRNA's also comprise mixtures of cap I and cap II structures (15, 22). In dl-1811 only the minor cap sequence ^{7m}Gppp^{m}AmpA lacks a related cap II derivative; the latter may have been present in insufficient amounts in the mutant mRNA to be detected by the methods used. The relative abundance of terminal sequences is decidedly not random; one type of 5′ end, namely, ^{7m}Gppp^{m}AmpG, seems to be particularly favored over the others, comprising 40% of the terminus population. This cap structure is also prominent, although not the most abundant, in the cellular mRNA.

The similarity between the A-cap and G-cap terminal sequences is remarkable: in general, out of 32 different possibilities, the trinucleotides Pu-C-A, Pu-G-A, and Pu-U-U serve as the effective 5′-terminal ends of the late viral mRNA and are presumably involved in initiation (or processing) of RNA transcription (with the 5′-terminal purine more often being pA than pG). Even the cap II formation seems to demonstrate a certain parallelism between similar A-cap and G-cap structures (e.g., A-G-A and G-G-A, A-U-U and G-U-U), and this phenomenon may reflect a kind of sequence dependency for the consecutive 2′-*O*-methylation reaction; indeed, cap II formation exceeds 20% for the type Pu-G-A, varies around 18% for the type Pu-U-U, and drops to about 10% for the Pu-C-A termini. A difference in the degree of cap II formation has been observed between 16S and 19S species of wild-type SV40 late mRNA, possibly related to the cytoplasmic half-life (12). We believe, therefore, that our present results about cap diversity, obtained by analysis of total unfractionated late RNA, will also in general remain applicable to the isolated species, except for the degree of cap II formation. We may note here also that the pattern of caps remains indeed identical for the different late polyoma virus mRNA's (6). More-

over, nonspliced dl-1811 nuclear RNA, which may be a direct precursor to both types of cytoplasmic RNA, exhibits a similar variety of A-cap structures to those described for the total cytoplasmic viral late mRNA (our unpublished data; not enough radioactivity was available to make a comment about the nuclear G-cap population).

Although cap structure diversity is restricted to a limited number of appropriate 5′-terminal sequences, the specificity of initiation of transcription (or, alternatively, the specificity of processing) is not well understood. Evidently, several trinucleotide groups distributed throughout the *Hin*dII + III restriction fragment C of SV40 DNA can be correlated with possible 5′ termini. The minimun length into which the entire cap population can be fitted extends over about 40 nucleotides. In polyoma virus RNA, it has been suggested that all virus-specific cap structures may be derived from a region of about 30 nucleotides in length (6). However, as characterization of the leader segments of SV40 RNA has provided evidence for extensive heterogeneity (10, 18, 28, 33), it seems rather unlikely that in this case putative initiation (or processing) would be restricted to such a limited area.

Finally, it should be noted that in wild-type SV40 RNA the major cap structures reported so far represent about 70% or more of the entire T_2-resistant material (12). As the origin of the remaining structures was unknown, we could not exclude the possibility that they were due to cellular mRNA contamination; however, we can now conclude that they correspond remarkably well to the alternative caps reported here for the dl-1811 late mRNA. Indeed, in addition to the major spots 1 and 2 (see Fig. 5 in reference 12) derived from the major wild-type capping site, which is deleted in dl-1811, we also observed spots 3 and 4, which were shown by enzymic analysis to correspond to ^{7m}Gppp^{m}AmpG and ^{7m}Gppp^{m}AmpC, respectively (unpublished data). Now, the latter in addition to ^{7m}GpppAmU constitute the predominant caps in dl-1811 late mRNA (Table 1). Presumably, the autoradiographs should be overexposed to reveal the very minor G-cap population and relevant cap II structures in wild-type virus late mRNA. The appearance of multiple alternative cap structures in wild-type late mRNA, although in minor quantities, should not be overlooked and may be of important biological significance. The selection of one major capping site out of a number of possibilities along the DNA chain indicates a well-defined regulatory mechanism possibly operating at the level of initiation of RNA transcription. Why one particular site should be preferred over all those available and what molecular regulation mechanism is responsible for this preference is not clear. However, the elimination of the major site does not decrease the total viral late mRNA production, which demonstrates that ultimately the alternative (minor) initiation sites are together about as effective in capping (and, presumably, in RNA initiation) as the wild-type major site. As mentioned above, a number of secondary (minor) leader segments have been reported in both 16S and 19S RNAs of SV40 (10, 18, 28). These have been characterized by reverse transcription, a method which allows identification of, e.g., spliced-out regions, but which does not permit unambiguous identification of the 5′ terminus. It is likely that the 5′ end of most of these minor mRNA's corresponds to one of the series of caps which we have identified here. We are now determining the exact positions on the genome whence these different cap structures originate.

ACKNOWLEDGMENTS

We thank J. Van der Heyden for culturing cells and virus stocks, A. Van de Voorde for helpful discussions, and W. Kuziel for help with the manuscript.

This research was supported by a grant from the Algemene Spaar- en Lijfrentekas. One of us (G.H.) acknowledges a fellowship from the Belgian Nationaal Fonds voor Wetenschappelijk Onderzoek.

LITERATURE CITED

1. **Aloni, Y., M. Shani, and Y. Reuveni.** 1975. RNAs of simian virus 40 in productively infected monkey cells: kinetics of formation and decay in enucleate cells. Proc. Natl. Acad. Sci. U.S.A. **72:**2587–2591.
2. **Bina-Stein, M., M. Thoren, N. Salzman, and J. 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.
3. **Celma, M., R. Dhar, J. Pan, and S. Weissman.** 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.
4. **Chiu, N., M. Radonovich, M. Thoren, and N. Salzman.** 1978. Selective degradation of newly synthesized nonmessenger simian virus 40 transcripts. J. Virol. **28:**590–599.
5. **Filipowicz, W.** 1978. Functions of the 5′-terminal ^{m7}G cap in eukaryotic mRNA. FEBS Lett. **96:**1–11.
6. **Flavell, A., S. Legon, A. Cowie, and R. Kamen.** 1979. Multiple 5′-terminal cap structures in late polyoma virus RNA. Cell **16:**357–371.
7. **Furuichi, Y.** 1978. Pretranscriptional capping in the biosynthesis of cytoplasmic polyhedrosis virus mRNA. Proc. Natl. Acad. Sci. U.S.A. **75:**1086–1090.
8. **Furuichi, Y., S. Muthukrishnan, J. Tomasz, and A. Shatkin.** 1976. Mechanism of formation of reovirus mRNA 5′-terminal blocked and methylated sequence ^{m7}GpppGmpC. J. Biol. Chem. **251:**5043–5053.
9. **Ghosh, P., V. Reddy, J. Swinscoe, P. Choudary, P. Lebowitz, and S. Weissman.** 1978. The 5′-terminal leader sequence of late 16S mRNA from cells infected with simian virus 40. J. Biol. Chem. **253:**3643–3647.
10. **Ghosh, P., V. Reddy, J. Swinscoe, P. Lebowitz, and S. Weissman.** 1979. The heterogeneity and 5′-terminal

structures of the late RNAs of simian virus 40. J. Mol. Biol. **126**:813–846.

11. **Groner, Y., P. Carmi, and Y. Aloni.** 1977. Capping structures of simian virus 40 19S and 16S mRNAs. Nucleic Acids Res. **4**:3959–3968.
12. **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.
13. **Haegeman, G., and W. Fiers.** 1978. Evidence for splicing of SV40 16S mRNA. Nature (London) **273**:70–73.
14. **Haegeman, G., and W. Fiers.** 1978. Localization of the 5′ terminus of late SV40 mRNA. Nucleic Acids Res. **5**: 2359–2371.
15. **Hecke, W., R. Fenton, T. Wood, C. Merkel, and J. Lingrel.** 1977. Methylated nucleosides in globin mRNA from mouse nucleated erythroid cells. J. Biol. Chem. **252**:1764–1770.
16. **Kim, C., and R. Sarma.** 1977. Spatial configuration of mRNA 5′ terminus. Nature (London) **270**:223–227.
17. **Kozak, M.** 1978. How do eucaryotic ribosomes select initiation regions in messenger RNA? Cell **15**:1109–1123.
18. **Lai, C., R. Dhar, and G. Khoury.** 1978. Mapping the spliced and unspliced late lytic SV40 RNAs. Cell **14**: 971–982.
19. **Laub, O., and Y. Aloni.** 1975. Transcription of simian virus 40. V. Regulation of simian virus 40 gene expression. J. Virol. **16**:1171–1183.
20. **Lavi, S., and A. Shatkin.** 1975. Methylated simian virus 40-specific RNA from nuclei and cytoplasm of infected BSC-1 cells. Proc. Natl. Acad. Sci. U.S.A. **72**:2012–2016.
21. **Lawley, P., and P. Brookes.** 1963. Further studies on the alkylation of nucleic acids and their constituent nucleotides. Biochem. J. **89**:127–138.
22. **Lockard, R.** 1978. Different cap 1:cap 2 ratios in rabbit α- and β-globin mRNA. Nature (London) **275**:153–154.
23. **Maxam, A., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560–564.
24. **Mertz, J., and P. Berg.** 1974. Viable deletion mutants of SV40: selective isolation by means of a restriction endonuclease from *Haemophilus parainfluenzae*. Proc. Natl. Acad. Sci. U.S.A. **71**:4879–4883.
25. **Muthukrishnan, S., B. Moss, J. Cooper, and F. Maxwell.** 1978. Influence of 5′-terminal cap structure on the initiation of translation of vaccinia virus mRNA. J. Biol. Chem. **253**:1710–1715.
26. **Nishimura, S.** 1972. Minor components in transfer RNA: their characterization, location and function. Prog. Nucleic Acid Res. Mol. Biol. **72**:50–58.
27. **Noyes, B., and G. Stark.** 1975. Nucleic acid hybridization using DNA covalently coupled to cellulose. Cell **5**:301–310.
28. **Reddy, V., P. Ghosh, P. Lebowitz, and S. Weissman.** 1978. Gaps and duplicated sequences in the leaders of SV40 16S RNA. Nucleic Acid Res. **5**:4195–4213.
29. **Revel, M., and Y. Groner.** 1978. Post-transcriptional and translational controls of gene expression in eukaryotes. Annu. Rev. Biochem. **47**:1079–1126.
30. **Rottmann, F.** 1978. Methylation and polyadenylation of heterogeneous nuclear and messenger RNA, p. 45–73. *In* B. Clark (ed.), Biochemistry of nucleic acids II, International review of biochemistry, vol. 17. University Park Press, Baltimore.
31. **Shenk, T., J. Carbon, and P. Berg.** 1976. Construction and analysis of viable deletion mutants of simian virus 40. J. Virol. **18**:664–671.
32. **Van de Voorde, A., R. Contreras, G. Haegeman, R. Rogiers, H. Van Heuverswyn, J. Van Herreweghe, G. Volckaert, M. Ysebaert, and W. Fiers.** 1977. Structural organisation of the SV40 genome, p. 17–30. *In* P. May, M. Monier, and R. Weil (ed.), Early proteins of oncogenic DNA viruses, vol. 69. Institut National de la Santé et de la Recherche Médicale, Paris.
33. **Villarreal, L., R. 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.
34. **Volckaert, G., and W. Fiers.** 1977. A micromethod for base analysis of ^{32}P-labeled oligoribonucleotides. Anal. Biochem. **83**:222–227.
35. **Volckaert, G., and W. Fiers.** 1977. Micro thin-layer techniques for rapid sequence analysis of ^{32}P-labeled RNA: double digestion and pancreatic ribonuclease analysis. Anal. Biochem. **83**:228–239.
36. **Wei, C., and B. Moss.** 1977. 5′-terminal capping of RNA by guanylyltransferase from HeLa cell nuclei. Proc. Natl. Acad. Sci. U.S.A. **74**:3758–3761.
37. **Weinberg, R., Z. Ben-Ishai, and J. Newbold.** 1974. Simian virus 40 transcription in productively infected and transformed cells. J. Virol. **13**:1263–1273.
38. **Winicov, I., and R. Perry.** 1976. Synthesis, methylation and capping of nuclear RNA by a subcellular system. Biochemistry **15**:5039–5046.