Proc. Natl. Acad. Sci. USA
Vol. 78, No. 3, pp. 1386–1390, March 1981
Biochemistry

Unspliced functional late 19S mRNAs containing intervening sequences are produced by a late leader mutant of simian virus 40

(mRNA biogenesis/mRNA 5′ termini)

PRABHAT K. GHOSH*, PRANAB ROY†, ALICE BARKAN‡, JANET E. MERTZ‡, SHERMAN M. WEISSMAN†, AND PAUL LEBOWITZ*

Departments of *Internal Medicine and †Human Genetics and Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06510; and ‡McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, Wisconsin 53706

Communicated by Aaron B. Lerner, November 3, 1980

ABSTRACT **dl-805 is a viable deletion mutant of simian virus 40 lacking 187 of the 202 nucleotides coding for the principal leader of the wild-type viral late mRNAs. The structures of the mutant late 16S and 19S RNAs and proteins were examined with the following principal findings. (*i*) The vast majority, if not all, of the 19S RNAs are unspliced over their entire length. These RNAs contain intervening sequences and serve as mRNAs. The vast majority, if not all, of the 16S RNAs, however, retain the characteristic wild-type splice. (*ii*) Mutant 16S and 19S RNAs do not terminate at residue 243 (located six nucleotides upstream from the 5′ end of the deleted DNA segment), the principal 5′ terminus of the wild-type late mRNAs; rather, they begin at a series of sites, mostly upstream, used infrequently or not at all by wild-type virus. (*iii*) Despite the presence of altered late mRNAs, dl-805 late proteins have electrophoretic mobilities similar to those of wild-type late proteins, and mutant virions contain the same relative quantities of late proteins as wild-type virions.**

Although knowledge of the structure of simian virus 40 (SV40) mRNAs is advanced, numerous questions related to both structure and biogenesis remain. In earlier studies on the structure of wild-type viral late mRNAs, we brought out two unexpected findings. The first was the presence in cytoplasmic extracts of infected cells of polyandenylylated 19S RNAs lacking any of three alternative splices present in the 5′ portion of the majority of 19S mRNAs (1). These "unspliced" RNAs constituted up to 10% of total cytoplasmic 19S RNAs and were also identified on polysomes prepared by standard procedures. However, questions remained as to whether they might still be nuclear in origin or spliced in a downstream region. Furthermore, studies with SV40 recombinants suggested that splicing was required for formation of stable mRNAs (2–4). The second surprising finding was marked heterogeneity of the 5′ termini of the late viral RNAs (1, 5). For late 19S RNAs, we detected at least five major 5′ termini and, for late 16S RNAs, two definite termini and probably several others. Although these findings were initially controversial, additional studies have supported 5′-terminal heterogeneity of the late RNAs (6–8).

As a follow-up to our earlier studies, we have been investigating the late RNAs of a series of SV40 late leader deletion mutants (9, 10). We have found that 25% or more of the 19S RNAs produced by several mutants lack a 5′-terminal splice. This report deals with mutant dl-805, which is especially interesting because most, if not all, of its polyadenylylated cytoplasmic and polysomal 19S RNAs are unspliced over their entire lengths. Furthermore, the late RNAs of this mutant lack 5′ termini at residue 243 (325[§]), the principal 5′ terminus of the wild-type late RNAs; instead, they begin principally at a series of genomic sites used infrequently or not at all by wild-type SV40. Despite synthesis of 16S and 19S mRNAs with altered structure, this mutant produces all three late viral proteins and virions that are structurally similar to those of wild-type virus. Unlike the only other unspliced mRNAs reported, histone (14) and adenovirus protein IX (15) mRNAs and one specific human interferon mRNA (16), the unspliced SV40 19S mRNAs contain intervening sequences.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Derivation of dl-805 (17) and the sequence of its late leader template (11, 18) have been described. dl-805 was grown in confluent Vero African green monkey kidney (AGMK) cells infected at 0.1 plaque-forming unit/cell for virion production and 10–20 plaque-forming units/cell for RNA production. For recovery of RNA, cells were harvested 40–44 hr after infection and lysed by 10-min exposures with gentle swirling to either hypotonic or isotonic medium (10 mM and 140 mM NaCl, respectively, in 10 mM Tris·HCl, pH 7.5/3 mM $MgCl_2$) containing 0.5% Nonidet P-40. After lysis, nuclei were pelleted and cytoplasmic RNA was extracted as described (1). Polysomes were isolated from cytoplasmic supernates by addition of sodium deoxycholate to 0.5% and sedimentation in 7.5–45% sucrose gradients in hypotonic buffer at 25,000 rpm for 225 min in an SW27 rotor. Oligo(dT)-cellulose-purified polyadenylylated RNAs were used in all experiments.

Nucleotide sequences of selected regions of the dl-805 late RNAs were determined by synthesis of complementary DNAs labeled in the 5′ position, separation of cDNAs on 8% polyacrylamide/7 M urea gels, and sequence analysis of individual cDNAs (1, 19).

RESULTS

Structure of dl-805 Late Leader Template. The late leader template of dl-805 is shown in Fig. 1. It differs from that of strain 776 wild-type SV40, the prototype in our prior studies on the SV40 mRNAs, in three regards: (*i*) It deletes residues

Abbreviation: SV40, simian virus 40.

[§] Nucleotides are numbered according to both a modification of the system of Reddy *et al.* (11) (without parentheses) and the BBB system (12) (in parentheses). Modification of the former involves numbering from 1 to 77 to correspond to original residues 18–94, numbering from 78 to 94 to include a 17-base-pair insert (13) detected after publication of the original sequence, retention of the original numbering of residues 95–5226, and numbering from 5227 to 5243 to correspond to original residues 1–17.

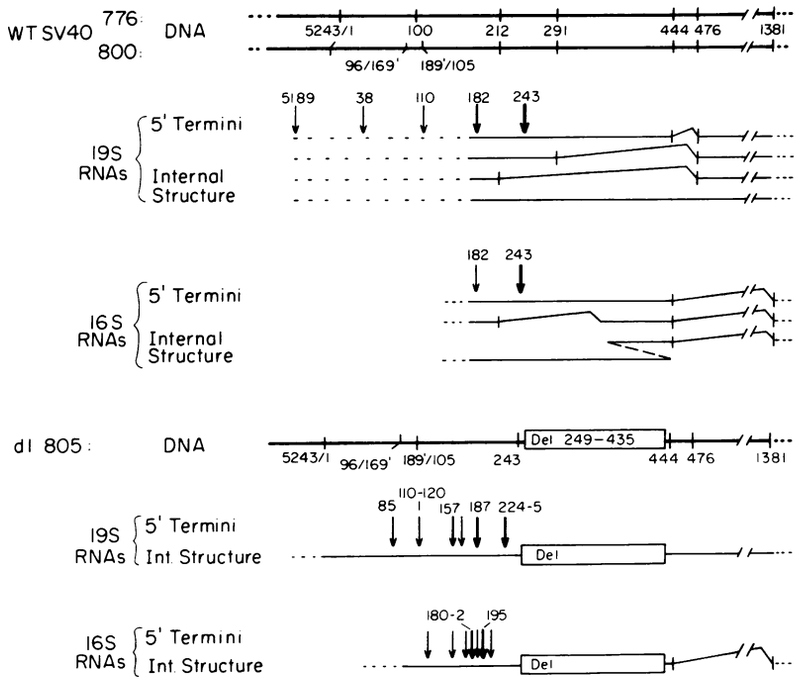

FIG. 1. Comparison of the genomes and late RNAs of wild-type SV40 (strains 776 and WT 800) and the dl-805 mutant. The repeated sequence of residues 169–189 in the mutant and WT 800 is indicated as 169'–189'. Three classes of 16S RNAs and four of 19S RNAs, distinguished on the basis of splice and leader structure (1, 5) and identical for both wild-type strains, are shown. Regions of mRNAs that are spliced out are indicated by peaked lines (⌐⌐). Vertical arrows localize 5' termini of RNAs; thick arrows denote abundant termini and thin ones less abundant termini. Nucleotide numbering follows a modification of the system of Reddy *et al.* (11).

249–435 (331–517), 187 of the 202 nucleotides comprising the template for the principal wild-type virus late leader [residues 243–444 (325–526)]; (*ii*) it lacks residues 97–104 (179–186); and (*iii*) it contains a duplication of residues 169–189 (251–271), designated 169'–189', inserted between residues 96–105. The immediate parent of dl-805, WT 800, also contains this duplication and deletes residues 97–104 (17). The combination of this duplication and deletion in WT 800 and dl-805 leads to tandem duplication of 85 nucleotides in their late leader templates rather than 72 nucleotides as in strain 776 SV40 (11).

Splices of Late 16S and 19S RNAs. Fig. 2 compares the cDNAs transcribed on the 5' ends of 19S RNAs obtained from cytoplasm of infected cells lysed under isotonic and hypotonic conditions (Fig. 2*a*) and from cytoplasm and polysomes of cells lysed under hypotonic conditions (Fig. 2*b*). For these syntheses, a primer complementary to the 5' end of the body of 19S RNA was used. The cDNA patterns are all complex, as noted previously for cDNAs synthesized on wild-type 19S RNAs (1). However, except for minor quantitative differences, the patterns of cDNA copies on RNAs obtained under isotonic and hypotonic conditions are almost identical (Fig. 2*a*). Because extraction under isotonic conditions reduces or even eliminates leakage of nuclear RNAs into cytoplasm (20), this study indicates that our RNA preparations consist virtually exclusively of cytoplasmic RNAs and not nuclear contaminants. Comparison of cDNAs copied from whole cytoplasmic and polysomal RNAs (Fig. 2*b*) reveals two differences: relatively smaller quantities of a number of cDNAs copied from polysomal RNAs and the presence of variable quantities of cDNAs at the electrophoretic origin copied from antilate RNA (data not presented) in the cytoplasmic, but not the polysomal, cDNA pattern. These findings suggest that RNAs extracted from our polysomal preparations are specific polysomal constituents and not adsorbed nonspecifically from cytoplasm.

To investigate the nature of internal splices and determine the genomic localizations of the 5' termini of the dl-805 RNAs, the 10 numbered cDNAs from Fig. 2*b* were subjected to nucleotide sequence analysis. All contained sequences colinear with SV40 DNA to their 3' termini (Fig. 3). Together, the 10 cDNAs analyzed constitute 80–90% of the total dl-805 cDNAs. Thus, the vast majority, if not all, 19S RNAs present in whole cytoplasm and on polysomes of dl-805-infected cells are unspliced in the 5'-terminal region.

These results raised two important questions. Might the 16S RNAs of dl-805 also lack their characteristic 5'-terminal splice? Might the mutant 19S RNAs be spliced in a region downstream from their 5' termini? To answer the first question, we synthesized cDNA on late lytic cytoplasmic RNA with a primer complementary to the 5' end of the 16S body [primer spanning residues 1382–1413 (1464–1495); see Fig. 1] and then performed sequence analyses on the resultant cDNAs. In answering the second query, we focused attention on the region between the 5' ends of the bodies of the 16S and 19S RNAs [from nucleotides 476–1381 (558–1463)] because this region contains a number of sequences that resemble the so-called consensus splicing sequence (21, 22) and because splicing between the initiation codons for proteins VP2 and VP3 [residues 480–834 (562–917)], causing a lesion in VP2, might be compatible with completely viable virus owing to the scarcity of this protein in the viral capsid. We thus looked for the presence of new splices in cDNAs synthesized with the residue 1382–1413 primer and also synthesized cDNAs with two additional primers spanning residues 1035–1154 (1117–1236) and 675–711 (757–793).

Fig. 2*c* demonstrates the cDNAs obtained with the residue 1382–1413 primer. The pattern is complex, showing more than 20 discrete cDNAs. The sequences of 14 cDNAs were determined, allowing their subdivision into two groups (Table 1). Seven cDNAs (nos. 5–11) comprise the first group and dem-

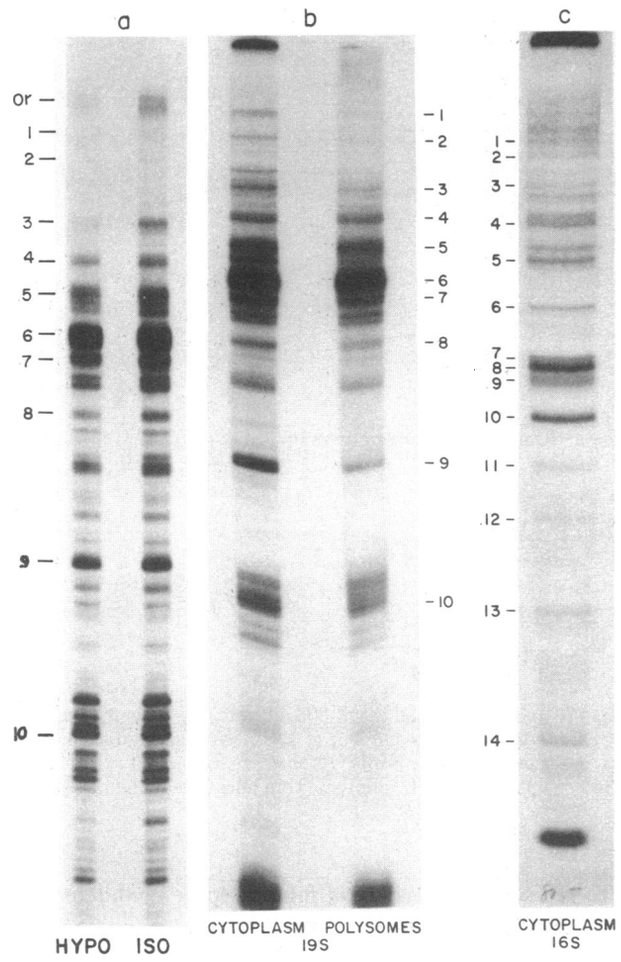

FIG. 2. The 8% polyacrylamide/7 M urea gel electrophoretic patterns of radiolabeled DNAs synthesized on 5′ termini of (*a*) cytoplasmic 19S RNAs from infected cells lysed under isotonic and hypotonic conditions, (*b*) cytoplasmic and polysomal 19S RNAs from cells lysed in hypotonic buffer, and (*c*) cytoplasmic 16S RNAs from cells lysed in hypotonic buffer. Primers spanning residues 479–509 and 1382–1415 (see Fig. 1) were used for cDNA syntheses on 19S and 16S RNAs. The nucleotide sequences of numbered cDNAs in *b* and *c* were determined and results are shown in Table 1 and Fig. 3.

onstrate the wild-type 16S splice linking residues 444–1381 (526–1463); they contain no new splices and differ only in the location of their 3′ termini. This result suggests a multiplicity of dl-805 16S RNAs and contrasts strikingly with our prior results for the wild-type 16S RNAs (5) in which one molecular species comprised 80–90% of total 16S species. cDNA nos. 1–4 and 12–14 do not contain any splice in the 300 nucleotides upstream from the 5′ end of the 16S RNA body and comprise the second subclass of cDNAs. It is not clear whether these cDNAs are derived from 19S species degraded during extraction or from *in vivo* species of mutant 16S RNA. However, in identical studies performed with cDNAs copied on wild-type late RNAs, unspliced species with termini immediately upstream from the primer have not been noted (5), suggesting the possibility that some of these shorter cDNAs are copied from mutant 16S RNAs lacking a 5′-terminal splice.

Extension of the residue 1035–1154 and 654–711 primers has also revealed cDNAs lacking splices in the 300 and 200 upstream nucleotides, respectively, ruling out any splices in the dl-805 19S RNAs from residues 476–1381 (558–1463). Furthermore, by using dl-805 polyadenylylated late cytoplasmic RNA prepared by us, R. Dhar and G. Khoury (personal communication) have been unable to detect any splices in the 19S RNA body by analysis of radiolabeled cDNAs protected from nuclease S1 and exonuclease 7 digestion. Taken together, these results indicate that most, if not all, cytoplasmic and polysomal 19S RNAs produced by dl-805 are unspliced over their entire length.

5′ Termini of 16S and 19S RNAs. Table 1 indicates the 3′ termini of the 16S and 19S cDNAs whose sequences we have determined. For reasons presented earlier (1, 5, 23), we believe the 3′ termini of the major cDNAs mark the 5′ termini of *in vivo* RNA species. Figure 1 compares the genomic sites used by wild-type SV40 and dl-805 for the 5′ termini of their principal late mRNAs. It is apparent that the 5′ termini of wild-type and mutant 16S and 19S RNAs differ almost totally; the principal wild-type terminus at residue 243 is not used at all by the mutant, and the only sites used in common are residues 180–182 and 110–120 for 16S and 19S RNAs, respectively.

Late Proteins. In view of the altered structure of the 19S, and to a lesser extent the 16S, RNAs of dl-805, we compared the late proteins present in mutant and wild-type virions. Gel electrophoresis of late proteins isolated from purified mutant and wild-type virions revealed no major differences in either the sizes or relative quantities of the VP1, VP2, or VP3 component proteins (results not shown). Thus, these unspliced mutant polysomal mRNAs probably function properly within the cell for synthesis of virion proteins.

DISCUSSION

The existence and significance of unspliced, polyadenylylated RNAs in the cytoplasm of SV40-infected cells have been the subjects of debate. On the one hand, we found that as much as 10% of wild-type late 19S cytoplasmic RNAs lacked a splice at the 5′ terminus and retained intervening sequences (1). In contrast, experiments involving expression of recombined SV40 16S RNA genes lacking intervening sequences suggested that intervening sequences and their removal from RNA by splicing are required for the cytoplasmic accumulation of mRNAs (2–4).

In view of this controversy, the most provocative finding of our present studies is that most, if not all, cytoplasmic and polysomal late 19S RNAs of mutant dl-805 are unspliced over their entire length. Given the absence of any detectable spliced 19S RNAs in mutant infected cells, it is difficult to escape the conclusion that the unspliced 19S RNAs serve as mRNAs for synthesis of the mutant VP2 and VP3 capsid proteins. Furthermore, because at most 1–2% of early cytoplasmic RNAs extracted under hypotonic conditions from SV40-infected or SV40-transformed cells are unspliced, we also suspect that the unspliced 19S late RNAs previously described in cytoplasmic extracts and on polysomes of cells infected with wild-type virus are bona fide mRNAs and not nuclear contaminants.

Unspliced mRNAs are not restricted to the SV40 system. Unspliced mRNAs coding for cellular histones have been recognized for some time (14) and recent studies indicate that adenovirus protein IX mRNA (15) and one human interferon mRNA (16) are also unspliced. Thus, mRNAs may be divided into two classes: those in which splicing is essential for cytoplasmic accumulation and those in which cytoplasmic accumulation can take place in the absence of splicing. SV40 16S mRNA is a prime example of the former, and it is noteworthy that dl-805 continues to make spliced 16S mRNAs while accumulating unspliced 19S mRNAs. However, the amount of 16S mRNAs made by this mutant is somewhat reduced, suggesting impairment in its biogenesis as well. In systems giving rise to mRNAs of the first class, intervening sequences, as noted, appear to be essential for the cytoplasmic accumulation of the spliced mRNA species. Whether they perform this function by specifying their own

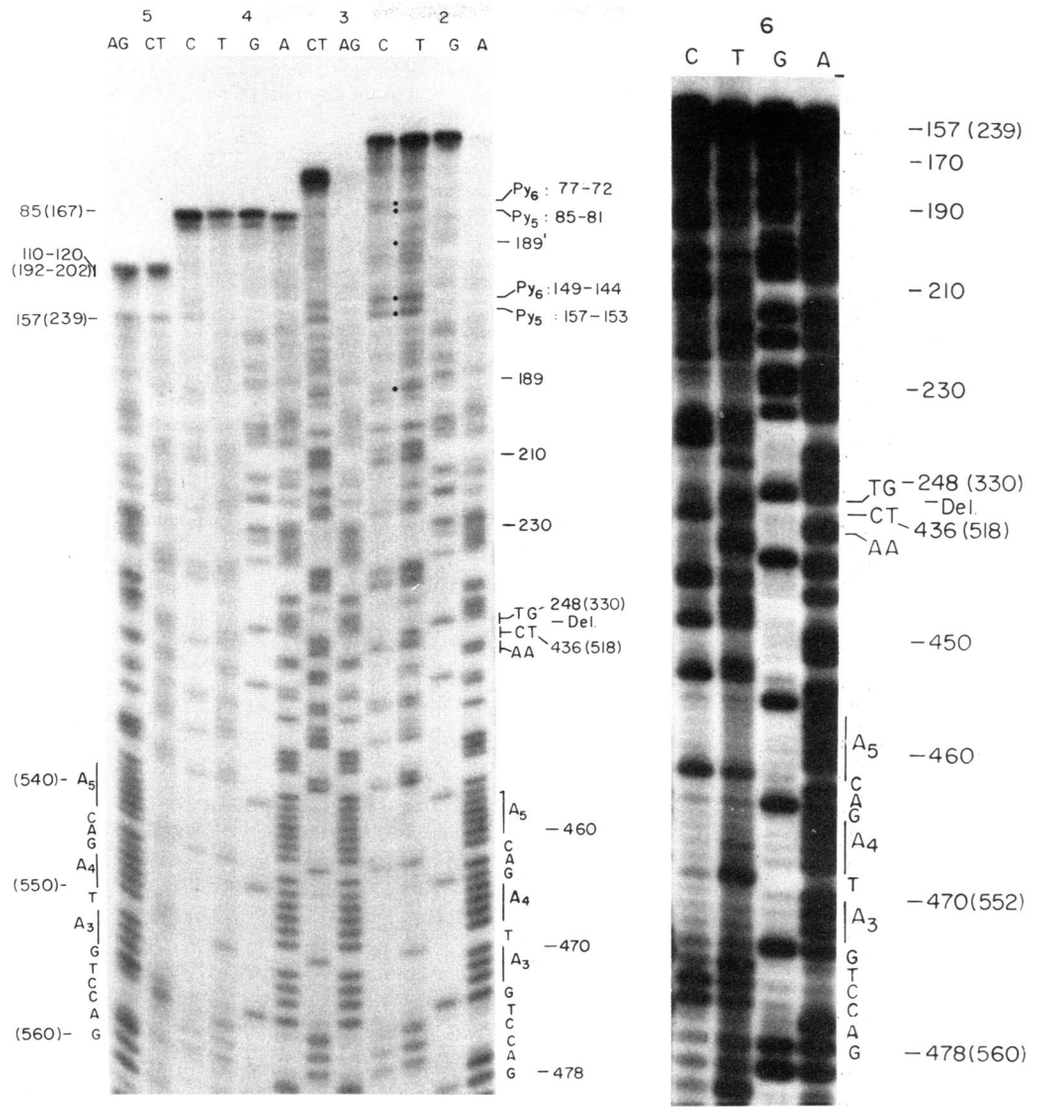

FIG. 3. Maxam–Gilbert (19) sequence analyses of DNA nos. 3–6 of Fig. 2*b* complementary to the 5′ termini of the dl-805 late cytoplasmic 19S RNAs. Sequences are read from the bottom of gels upward, corresponding to a 5′-to-3′ direction on cDNAs and a 3′-to-5′ direction on template RNAs. The full SV40 DNA sequence is presented in Reddy *et al.* (11). Residue numbers not in parentheses are modified from those of Reddy *et al.* and conform to those in Fig. 1. BBB system numbers (12) are given in parentheses.

excision from the unspliced precursor or by some other means is not known. mRNAs that do not require splicing for cytoplasmic accumulation may be further subdivided into two subclasses: those which contain intervening sequences and those which do not. Both histone and adenovirus protein IX mRNAs fall into the latter subclass. At present, the dl-805 19S mRNAs and a fraction of the wild-type 19S mRNAs are striking examples of mRNAs that retain intervening sequences. In fact, their existence raises the question of whether transcripts of other eukaryotic genes containing intervening sequences also reach the cytoplasm unspliced. In the unspliced dl-805 and wild-type SV40 19S mRNAs, the function of intervening sequences, if any, must be other than to specify splicing that is obligatorily coupled to production of stable cytoplasmic mRNAs.

The biochemical basis for failure of splicing of the 19S RNAs of dl-805 is unknown. The deletion in this mutant removes the important 19S splice site at residue 291 (373), but the residue 212 (294) and 444 (526) splice sites are intact, and other mutants lacking the 291 splice site make 19S RNAs with the residue 444–476 or 212–476 splice (Fig. 1). Extension of the dl-805 deletion to within nine nucleotides of residue 444 may be important in abolition of splicing of the 19S RNAs, but this explanation cannot by itself account for lack of splicing because mutant dl-1613, with a deletion spanning residues 420–440 (502–522), makes 19S RNAs with the three wild-type splices (10).

We have pointed out that neither 16S nor 19S mutant RNAs terminate at residue 243, the predominant 5′ terminus of the wild-type late RNAs. This finding is probably related to the fact that the deletion in this mutant extends to within six nucleotides of the 243 site. It seems likely either that adjacent downstream sequences are required for use of this site as a 5′ terminus or that the downstream sequences present in dl-805 in some way interfere with its use. Probably as a consequence of its inability to use residue 243, dl-805 uses a variety of upstream sites for the 5′ termini of its late RNAs. However, residue 182, the second most abundant 5′ terminus of the wild-type late RNAs, is used minimally whereas several other sites that are used rarely

Table 1. Splices and 3′ termini of cDNAs copied on dl-805 late cytoplasmic RNAs

	16S RNAs		19S RNAs	
cDNAs	Splice	Terminus	Splice	Terminus
1	Unspliced	—	Unspliced	—
2	Unspliced	—	Unspliced	—
3	Unspliced	1160–1180 (1242–1262)	Unspliced	?38 (120)
4	Mixture: mostly unspliced	1224–1228 (1306–1310)	Unspliced	85 (167)
5	444–1381 (526–1463)	129–131 (211–213)	Unspliced	110–120 (192–202)
6	444–1381	157 (239)	Unspliced	157 (239)
7	444–1381	178–180 (260–262)	Unspliced	164–166 (246–248)
8	444–1381	180–182 (262–264)	Unspliced	187 (269)
9	444–1381	185–187 (267–269)	Unspliced	224–225 (306–307)
10	444–1381	195 (277)	Unspliced	440–441 (522–523)
11	444–1381	208–210 (290–292)		
12	Unspliced	1343 (1425)		
13	Unspliced	1357 (1439)		
14	Unspliced	1371 (1453)		

or not at all by wild-type virus are used by dl-805. Clearly, mechanisms exist that allow mutants lacking templates for the principal wild-type 5′ termini to use new sites for their 5′ termini (8–10, 24).

Although dl-805 19S mRNAs are unspliced and both 16S and 19S mRNAs contain a different population of 5′ termini than wild-type late mRNAs, the protein composition of mutant and wild-type virions appears similar. Furthermore, early experiments have provided evidence that *in vivo* rates of synthesis of mutant and wild-type late proteins are similar, suggesting comparable template efficiencies for mutant and wild-type late mRNAs. Thus, the basis for the reduced growth rate of dl-805 remains elusive (17, 18).

Note Added in Proof. We have recently observed that, although some wild-type VP1 is made, the predominant virus-encoded protein seen in both dl-805 virions and infected cell lysates has a slightly faster mobility than VP1 in $NaDodSO_4$/polyacrylamide gels. The reason for this difference is not yet clear.

We thank Charlene Ivory for excellent technical assistance. This research was supported by grants from the National Cancer Institute, American Cancer Society, Leonard Eckstein Fund for Leukemia Research, and the Swebilius Trust of the Yale Comprehensive Cancer Center.

1. Ghosh, P. K., Reddy, V. B., Swinscoe, J., Lebowitz, P. & Weissman, S. M. (1978) *J. Mol. Biol.* **126**, 813–846.
2. Gruss, P., Lai, C. J., Dhar, R. & Khoury, G. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4317–4321.
3. Hamer, D. H. & Leder, P. (1979) *Nature (London)* **281**, 35–39.
4. Hamer, D. H., Smith, K. D., Boyer, S. H. & Leder, P. (1979) *Cell* **17**, 725–735.
5. Reddy, V. B., Ghosh, P. K., Lebowitz, P. & Weissman, S. M. (1978) *Nucleic Acids Res.* **5**, 4195–4214.
6. Lai, C. J., Dhar, R. & Khoury, G. (1978) *Cell* **14**, 971–982.
7. Canaani, D., Kahana, C., Mukamel, A. & Groner, Y. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3078–3082.
8. Villarreal, L. P., White, R. T. & Berg, P. (1979) *J. Virol.* **29**, 209–219.
9. Piatak, M., Ghosh, P. K., Reddy, V. B., Lebowitz, P. & Weissman, S. M. (1979) in *Extrachromosomal DNA*. ICN-UCLA Symposium on Molecular and Cellular Biology, eds. Cummings, D., Borst, P., Dawid, I., Weissman, S. M. & Fox, C. F (Academic, New York), pp. 199–215.
10. Piatak, M., Olive, M., Subramanian, K. N., Ghosh, P. K., Reddy, V. B., Lebowitz, P. & Weissman, S. M. (1980) in *Analysis of Gene Structure and Expression*, eds. Dean, D. H., Johnson, L. F., Kimball, P. C. & Perlman, P. S. (Ohio State Univ. Press, Columbus, OH), pp. 151–173.
11. Reddy, V. B., Thimmappaya, B., Dhar, R., Subramanian, K. N., Zain, B. S., Pan, J., Ghosh, P. K., Celma, M. & Weissman, S. M. (1978) *Science* **200**, 494–502.
12. Tooze, J., ed. (1980) *DNA Tumor Viruses. Molecular Biology of Tumor Viruses* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), Part 2.
13. van Heuverswyn, H. & Fiers, W. (1979) *Eur. J. Biochem.* **100**, 51–60.
14. Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H. O. & Birnstiel, M. L. (1978) *Cell* **14**, 655–671.
15. Alestrom, P., Akusjarvi, G., Perricaudet, M., Mathews, M. B., Klessig, D. F. & Petterson, U. (1980) *Cell* **19**, 671–681.
16. Nagata, S., Mantei, N. & Weissmann, C. (1980) *Nature (London)* **287**, 401–408.
17. Mertz, J. E. & Berg, P. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 4879–4883.
18. Barkan, A. & Mertz, J. E. (1981) *J. Virol.* **37**, 730–737.
19. Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
20. Levis, R. & Penman, S. (1977) *Cell* **11**, 105–113.
21. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4853–4857.
22. Lerner, M. A., Boyle, T. A., Monnet, S. M., Wolin, S. L. & Steitz, J. (1980) *Nature (London)* **283**, 220–224.
23. Reddy, V. B., Ghosh, P. K., Lebowitz, P., Piatak, M. & Weissman, S. M. (1979) *J. Virol.* **31**, 279–296.
24. Haegeman, G., van Heuverswyn, H., Gheysen, D. & Fiers, W. (1979) *J. Virol.* **31**, 484–493.