Proc. Natl. Acad. Sci. USA
Vol. 74, No. 11, pp. 4982–4985, November 1977
Biochemistry

Sequence arrangement of the 5′ ends of simian virus 40 16S and 19S mRNAs

(electron microscopy/RNA·DNA hybridization)

MING-TA HSU AND JOHN FORD

The Rockefeller University, New York, New York 10021

Communicated by James E. Darnell, Jr., September 10, 1977

ABSTRACT **Electron microscopic examination of molecular hybrids between simian virus 40 DNA, that had been cleaved with *Eco*RI and then digested with exonuclease III, and either 16S or 19S mRNA produced late during the viral infection cycle indicated that each of these mRNAs contained a 5′-terminal "leader" sequence that was encoded in the viral genome at about map position 0.71–0.75. Hybridization of each of these mRNAs to viral DNA immobilized on nitrocellulose filters supported the electron microscopic observations.**

A number of different late-produced mRNAs from adenovirus type 2 (Ad2) contain a segment at their extreme 5′ ends that derives from Ad2 DNA remote in the viral genome from the main mRNA coding regions (the "leader" sequence) (1–3). The "leader" is about 150–200 nucleotides long and consists of three short sequences transcribed from different regions of the genome. Possibly by intramolecular RNA rearrangements after transcription, the "leader" is formed and joined to the main portion of a number of different mRNA molecules.

The discovery of the "leader" sequence in Ad2 late mRNAs prompted us to determine whether the presence of such sequences in mRNAs might be a more widespread phenomenon. To examine this possibility we have searched for "leader" sequences in the mRNAs of the small DNA tumor virus, simian virus 40 (SV40).

MATERIALS AND METHODS

DNA. SV40 DNA form I was released from CV-1 cells infected with the SVS strain (4) by Hirt extraction (5) and the SV40 DNA was purified by ethidium bromide/CsCl equilibrium banding (6).

Cytoplasmic RNA. CV-1 cells were infected at 10 plaque-forming units per cell with SV40. For the biochemical experiments 6×10^7 cells were labeled with 2.5 mCi of [^{3}H]uridine (New England Nuclear) per 10 ml of medium 46–48 hr after infection. The RNA for both the biochemical and electron microscopic experiments was extracted 48 hr after infection. The RNA was extracted by the procedure of Spradling *et al.* (7). The cells were washed once with ice-cold buffer (20 mM Tris, pH 7.4/140 mM NaCl/10 mM KCl/1 mM sodium phosphate). Then they were scraped from the monolayer into 10 ml of the buffer and centrifuged at 2000 rpm for 2 min at 4°. Cells (6×10^7) were resuspended in 10 ml of the following solution: 30 mM Tris, pH 8.5/0.1 M NaCl/10 mM $CaCl_2$/25 μg of polyvinyl sulfate (Sigma) per ml/35 μg of spermine (Calbiochem) per ml/1% diethylpyrocarbonate (Naftone). The solution was then made 0.5% in Nonidet P-40 (Shell). The mixture was shaken on a Vortex mixer for 20 sec and then centrifuged at 2000 rpm for 3 min at 4° to remove the nuclei. The cytoplasmic RNA was extracted as described (8).

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

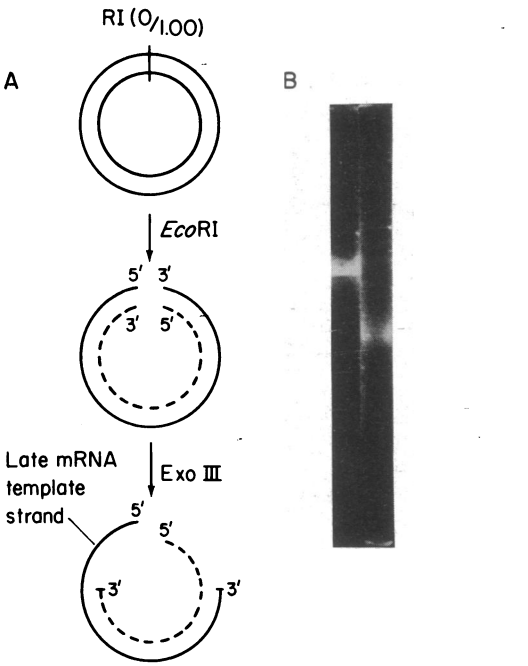

FIG. 1. (*A*) Schematic representation of preparation of SV40 DNA by cleavage with *Eco*RI and then digestion with Exo III. The two DNA strands are arbitrarily depicted as a solid line and an interrupted line. Note the specific preservation of the DNA template for the mRNA in the 5′ direction from the *Eco*RI cut site. (*B*) Agarose (2%) gel electropherograms of SV40 DNA that had been first cleaved with *Eco*RI and then digested with Exo III. The gel on the left shows SV40 DNA after *Eco*RI cleavage and Exo III digestion. The gel on the right shows SV40 DNA that has been digested with *Eco*RI and Exo III and then further digested with S1. The SV40 DNA band in the gel on the right corresponds to 0.40 genome lengths (2000 nucleotides). This result indicates a 0.3 or 30% DNA digestion by Exo III, $1 - (0.3 \times 2) = 0.4$.

Liquid Hybridization. SV40 DNA was first cleaved with *Eco*RI endonuclease (Miles) (9, 10) in 0.1 M Tris, pH 7.4/10 mM $MgCl_2$. The *Eco*RI enzyme was inactivated by heating to 65° for 3 min. The sample was then diluted with an equal volume of 0.1 M Tris, pH 8.5. After the sample was chilled on ice for 15 min, 1.5 units of *Escherichia coli* exonuclease III (Exo III, Biolab) was added per μg of DNA (11). After incubation on ice for an additional 15 min, the sample was incubated at 37° for 25–30 min. The reaction was stopped by the addition of EDTA to 10 mM. The extent of Exo III digestion was monitored by electron microscopy and/or S1 nuclease digestion and gel electrophoresis. For S1 nuclease digestion, 0.1 μg of DNA was incubated with 2 units of nuclease S1 (12). The digestion products were analyzed by gel electrophoresis in 2% agarose;

Abbreviations: SV40, simian virus 40; *Eco*RI → Exo III DNA, SV40 DNA that is first cleaved with *Eco*RI endonuclease and then digested with *Escherichia coli* exonuclease III; Ad2, adenovirus type 2.

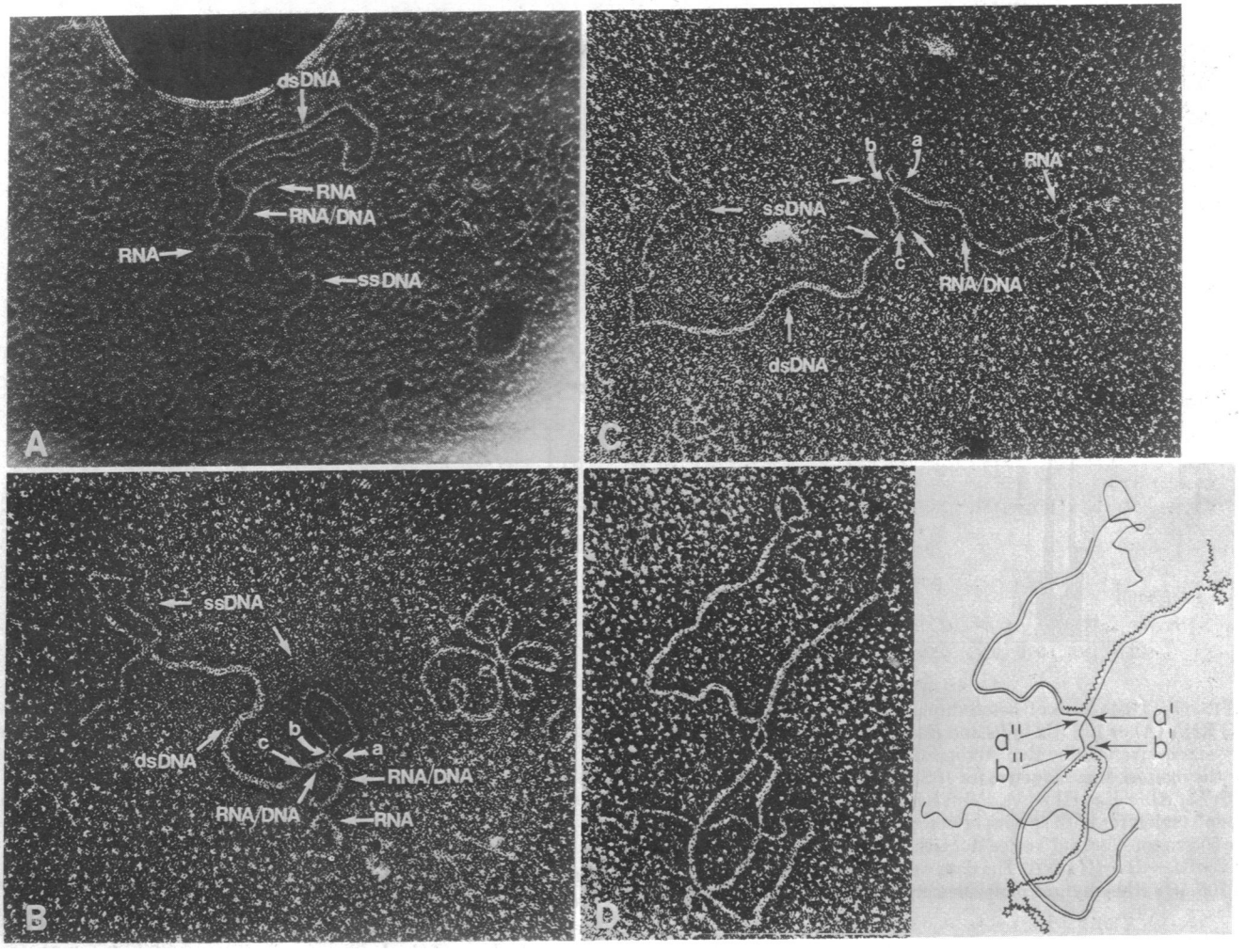

FIG. 2. Heteroduplex molecules between 16S or 19S RNA and SV40 DNA. (*A*) A heteroduplex molecule (ds, double stranded) formed between a 16S RNA molecule and an *Eco*RI → Exo III DNA molecule that has been rendered 10% single stranded (ss) on each end by Exo III digestion. (*B*) A heteroduplex molecule formed between a 16S RNA molecule and an *Eco*RI → Exo III DNA molecule that has been rendered 30% single stranded on each end by Exo III digestion. Point a is the 5′ end of the main coding sequence of 16S RNA in the hybrid with DNA. Length b–c is 16S RNA "leader" in the hybrid with DNA. (*C*) A heteroduplex molecule formed between a 19S molecule and an *Eco*RI → Exo III DNA molecule that has been rendered 30% single stranded. Point a is the 5′ end of main coding sequence of 19S RNA in the hybrid with DNA. Length b–c is 19S RNA "leader" in the hybrid with DNA. (*D*) A complex heteroduplex molecule composed of two *Eco*RI → Exo III DNA molecules (digested as for *C*) that have hybridized two 19S RNAs in a reciprocal manner (see *text*). The molecule is schematically depicted at the right.

they were stained with 0.5 μg of ethidium bromide per ml.

After the Exo III digestion the SV40 DNA was used to hybridize cytoplasmic RNA in 0.1 M Tris, pH 7.4/0.1 M EDTA/0.3M NaCl at 60° for 30 min.

For some experiments the cytoplasmic RNA was fractionated on a sucrose gradient (see *Filter Hybridization*) to enrich for either the 16S or 19S RNA.

Electron Microscopy. For electron microscopy the preparation was made 50% formamide in 0.1 M Tris, pH 8.5/10 mM EDTA was spread either onto a hypophase of 20% formamide in 10 mM Tris, pH 8.5/1 mM EDTA or onto distilled water, and was examined with a Philips 201 electron microscope. The hybrid molecules were photographed and measured with a Numonics digital electronic graphic calculator.

Filter Hybridization. SV40 DNA form I was digested variously with the following restriction endonulceases: *Hha* I (Miles) (13), *Hae* II (Miles) (14), or *Bgl* I (gift from R. Evans) (15) in 6 mM Tris, pH 7.4/6 mM $MgCl_2$/6 mM 2-mercaptoethanol. The digestion with the enzyme *Eco*RI was as described above. Because of the size difference in the DNA fragments generated, the fragments could be separated by two cyles through a 15–30% sucrose gradient in 10 mM Tris, pH 7.4/10 mM EDTA/50 mM NaCl with an SW40 rotor (Beckman) at 39,000 rpm for 14 hr at 22°. The adequacy of fragment separation was tested with agarose gel electrophoresis (see *Liquid Hybridization*). The DNA was denatured in 0.2 M NaOH and adsorbed to nitrocellulose filters at pH 11.3 (16). Filters containing total SV40 DNA were prepared from SV40 linear DNA generated by *Eco*RI cleavage.

The cytoplasmic RNA from 6×10^7 cells was fractionated on a 15–30% sucrose gradient in 10 mM Tris, pH 7.4/50 mM NaCl/10 mM EDTA/0.2% sodium dodecyl sulfate and centrifuged with an SW40 rotor at 35,000 rpm for 10 hr at 22°. Aliquots of gradient fractions were treated with 0.2 M NaOH at 0° for 90 min. The hybridization, on filters, against either total SV40 DNA or SV40 fragments (1 μg of DNA equivalent/fraction hybridization) was performed at 60° for 24 hr (15). There was a greater than 20-fold excess of DNA to RNA in the hybridization reaction (unpublished observation). The filters were washed and treated with RNase as described (17).

RESULTS

Late during the infectious cycle, SV40 mRNA consists of two major species: 16S, the bulk of the steady state viral mRNA, and 19S, the more rapidly labeled but less abundant species (18).

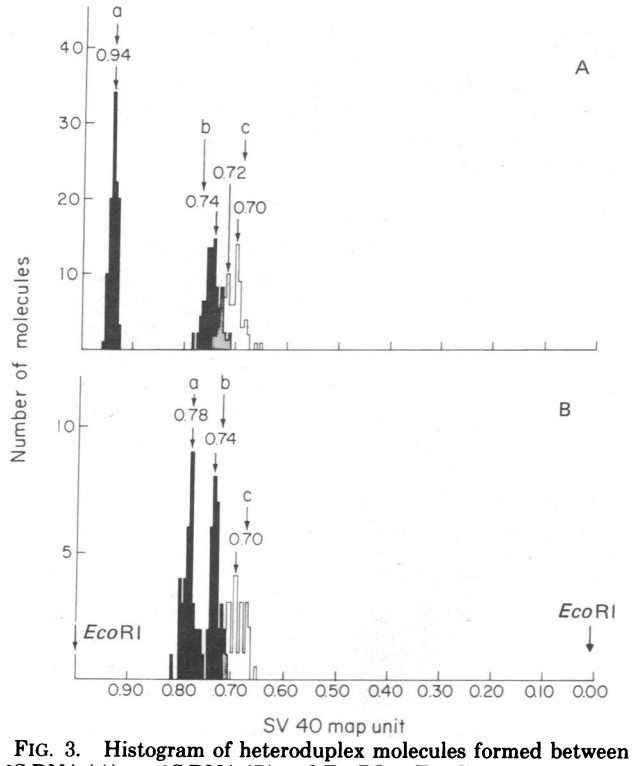

FIG. 3. Histogram of heteroduplex molecules formed between 16S RNA (*A*) or 19S RNA (*B*) and *Eco*RI → Exo III DNA rendered 30% single stranded by Exo III digestion. Point a is the 5′ terminus for the main coding sequence for 16S RNA or 19S RNA hybridized with *Eco*RI → Exo III DNA. Point b is the 5′ terminus for the DNA "loop" region for both species of heteroduplex molecules. Point c is the 5′ terminus for the "leader" for either 16S or 19S RNA hybridized to *Eco*RI → Exo III DNA. The data were compiled from the analysis of 105 16S RNA hybrid molecules and 52 19S RNA hybrid molecules.

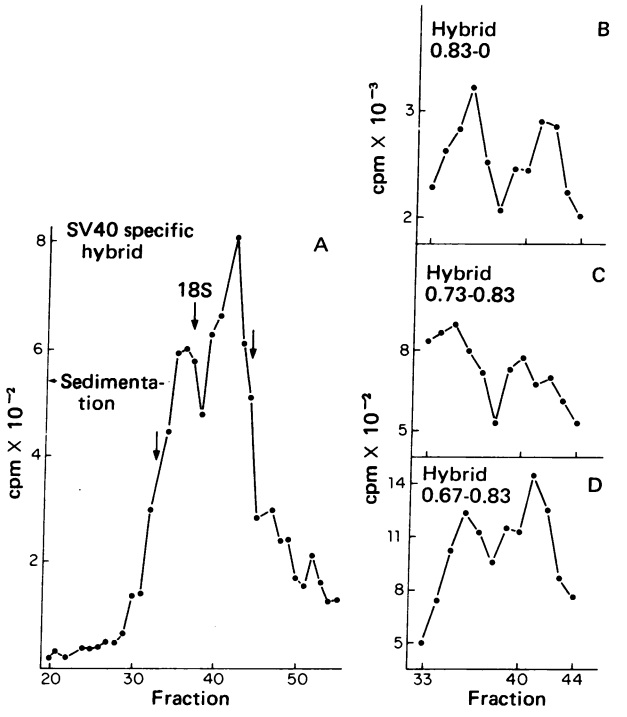

FIG. 4. (*A*) Hybridization profile of cytoplasmic RNA from infected cells to SV40 DNA. The RNA was sedimented in a 15–30% sucrose gradient and then part of each fraction was hybridized to total SV40 DNA immobilized on nitrocellulose filters. Fractions 33–44 (between the unmarked arrows) from the sucrose gradient were hybridized to SV40 DNA fragments covering map positions 0.83–0 (*B*), 0.73–0.83 (*C*), and 0.67–0.83 (*D*). See text.

To determine whether these two mRNAs contain "leader" sequences they were each hydridized to SV40 DNA that had previously been digested with the Endo R·*Eco*RI restriction nuclease and subsequently treated with exonuclease III (*Eco*RI → Exo III DNA). In this way only the complementary DNA strand was available to hybridize with the mRNAs, thus eliminating the competing DNA·DNA re-annealing reaction (Fig. 1*A*). For such DNA molecules to hybridize to all the corresponding mRNA sequences, most DNA molecules would have to be digested beyond the 5′ terminus of the mRNA coding region. This requires that the Exo III reaction be somewhat synchronous. Fig. 1*B* shows an agarose gel profile of SV40 DNA after restriction by the *Eco*RI enzyme followed by Exo III treatment before and after further treatment with S1 nuclease, which removes the single-stranded DNA generated by the Exo III. As shown in the figure, the doubly digested DNA migrated as a relatively sharp band in the gel, indicating that a relatively synchronous reaction with the exonuclease was obtained.

When *Eco*RI → Exo III treated DNA that had been digested so that 10–20% was rendered single stranded on either end was hybridized with 16S mRNA, an RNA·DNA duplex formed extending from map position 0.94 to 0.0/1.0 (the *Eco*RI cut site) (see Fig. 2*A*). This observation agrees well with previously reported findings that showed the 5′ end of the main coding sequence of the 16S mRNA to be encoded at position 0.95 in the viral genome (19). Additionally, the heteroduplex molecule included a 150- to 200-nucleotide length of unhybridized RNA extending in the 5′ direction from the hybrid terminus at map position 0.94. If the *Eco*RI → Exo III DNA was rendered 30% single stranded on either end by the Exo III treatment, the hybrid with 16S mRNA formed a single-stranded DNA loop as shown in Fig. 2*B*. The loop was 0.19 ± 0.01 map units long, beginning at the 5′ terminus of the hybrid to the main coding sequence of 16S RNA (map position 0.94). Counterclockwise from the base of the loop was an additional 0.037 ± 0.01 map unit RNA·DNA hybrid at position 0.71–0.75, suggesting that the 16S mRNA contained a "leader" sequence at its 5′ terminus. Thus, the leader sequence appeared to be encoded approximately 0.2 genome lengths removed from the DNA sequences coding for the 5′ end of the main body of the 16S mRNA molecule.

The 19S mRNA was also analyzed in a manner similar to that described for the 16S mRNA. The 5′ end of the main coding region of the 19S mRNA was localized at map position 0.78 ± 0.015, in agreement with previous R-loop data (19). As found for the 16S mRNA, the 5′ end of the main coding region of the 19S mRNA (at map position 0.78) served as the base of a single-stranded DNA loop of about 0.04 map units. Counterclockwise from the loop was an RNA·DNA hybrid of 0.045 ± 0.02 map units, extending from map position 0.70 to 0.74 (Fig. 2*C*). Thus, there appeared to be a "leader" at the 5′ end of the 19S mRNA that, within the precision of the technique, mapped at the same position in the viral genome as that observed at the 5′ end of the 16S mRNA.

Confirmatory evidence for the 19S mRNA "leader" is shown in Fig. 2*D*. A complex structure was observed in the electron microscope that arose from the hybridization of two 19S mRNA molecules with two *Eco*RI → Exo III DNA molecules. The "leader" and main coding regions of one 19S RNA molecule are hybridized to different *Eco*RI → Exo III DNA molecules. The other 19S RNA molecule is hybridized to the same two *Eco*RI → Exo III DNA molecules in a reciprocal fashion. The RNA·DNA hybrid regions (map position 0.78–0) shown in Fig. 2*D* lying between each of the two short single-stranded regions (a′–b′ and a″–b″) and the corresponding unhybridized RNA

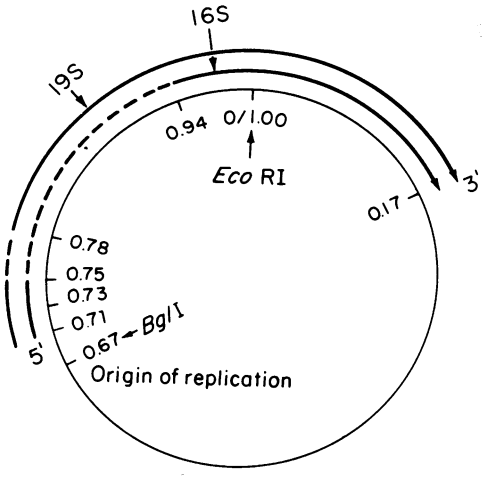

FIG. 5. Map of 16S and 19S RNA derived from the present work. The dashed lines represent regions in the DNA template that separate the RNA "leader" from the RNA main coding sequences. The "leaders" for 19S RNA and 16S RNA are drawn identically, but the accuracy of techniques involved cannot exclude subtle differences between the two.

"tail" were both equal in length to the RNA·DNA hybrid region shown in Fig. 2C between the loop and the unhybridized RNA "tail". These data indicated that the "single-strand" (Fig. 2D, a′–b′ and a″–b″) region was the unhybridized DNA separating the template for the main coding sequence of 19S RNA from that of the 19S RNA leader. This is additional evidence for a "leader" sequence at the 5′ end of the 19S mRNA. A histogram showing the number of hybrid molecules with *Eco*RI → Exo III DNA and the 16S or 19S mRNAs is shown in Fig. 3.

Biochemical evidence that would corroborate the presence of the "leader" sequence at the 5′ ends of the two late mRNAs was sought with [^{3}H]uridine-labeled cytoplasmic RNA from SV40-infected cells. Size-fractionated RNA, including both the 16S and 19S regions, was hybridized to restriction enzyme-generated DNA fragments immobilized on nitrocellulose filters. Three different specific SV40 DNA fragments were used covering genome regions 0.67–0.83 (*Bgl* I cut site to *Hae* II cut site), 0.73–0.83 (*Hha* I generated fragment), and 0.83–0 (*Hae* II cut site to *Eco*RI cut site). If the 16S RNA had its 5′ terminus at map position 0.94, then there should be no 16S RNA hybridized to SV40 DNA counterclockwise from map position 0.94. However, Fig. 4 shows that 16S RNA did hybridize to SV40 DNA from the region 0.67–0.83. About half the number of counts were found in the hybrid to the fragment covering map position 0.67–0.83 as were found in the hybrid to the fragment covering map position 0.83–0. Since 16S RNA has sequences transcribed from the region 0.94–0, the data are consistent with a "leader" of about 0.03 map unit in agreement with the electron microscopy data, which show a "leader" of 0.04 ± 0.01 length.

A smaller amount of the 16S RNA also hybridized to SV40 DNA from the region 0.73–0.83, and the relative number of cpm in the hybrid indicated a duplex region of about 0.015 map unit. This is also consistent with this electron microscopy data, which show that the 3′ end of the 16S RNA "leader" lies at 0.75 ± 0.01.

Interpretation of the hybridization data for 19S RNA is made more difficult because of the proximity on the DNA template of the 3′ end of the "leader" at about map position 0.74 and the 5′ end of the main coding sequence at about map position 0.78. However, for 19S RNA the ratio of cpm hybridizing to the DNA region from 0.67 → 0.83 to those hybridizing to the DNA region from 0.83 → 0 should be 0.05, 0.17, or 0.29, if there is no "leader". The ratio for the three points defining the 19S RNA peak is 0.41 ± 0.04. The excess RNA hybridization to DNA from the region 0.67–0.83 is consistent with "leader" sequence for 19S RNA. Other investigations located a 5′-methyl-labeled structure from 16S mRNA remote from the main coding sequence of 16S RNA (20, 21).

DISCUSSION

The data presented here provide evidence for the presence of a common "leader" sequence in both the 16S and 19S late SV40 mRNAs. Fig. 5 depicts the arrangement of the nucleotide sequences within each of these two mRNAs. The "leader" sequence is encoded in the viral genome at position 0.71–0.75, with a gap of 0.04 map units between it and the beginning of the main coding sequences for the 19S mRNA at position 0.78; for the 16S mRNA, there is a gap of 0.19 map units between the position coding for the "leader" and the beginning of the main mRNA coding sequences at map position 0.94.

The "leader" sequence in SV40 differs from that previously described for late Ad2 RNAs in that for the SV40 late messengers the "leader" is a 150 to 200-nucleotide RNA segment that is continuously transcribed. In Ad2, there are three noncontiguous regions of the viral genome that are transcribed which eventually become juxtaposed in the mature mRNA molecules to form the 150- to 200-nucleotide "leader" (2).

We gratefully appreciate the advice and guidance of J. E. Darnell. We recognize the kind help of W. Jelinek in revising the manuscript text. J. Cozzitorto gave diligent assistance to the project. This study was supported by National Cancer Institute Grant CA19073-02 and National Institutes of Health Training Grant CA09256-01.

1. Chow, L. T., Gelinas, R. E., broker, T. R. & Roberts, R. J. (1977) *Cell*, **12**, 1–8.
2. Berget, S. M., Moore, C. & Sharp, P. A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3171–3175.
3. Klessig, D. F. (1977) *Cell*, **12**, 9–21.
4. Takemoto, K. K., Kirschenstein, R. L. & Habel, K. (1966) *J. Bacteriol.* **92**, 990–994.
5. Hirt, B. (1967) *J. Mol. Biol.* **26**, 365–369.
6. Radloff, R., Bauer, W. & Vinograd, J. (1967) *Proc. Natl. Acad. Sci. USA* **57**, 1514–1521.
7. Spradling, A., Hui, H. & Penman, D. (1975) *Cell* **4**, 131–137.
8. Soeiro, R. & Darnell, J. E. (1969) *J. Mol. Biol.* **44**, 551–562.
9. Morrow, J. F. & Berg, P. (1973) *Proc. Natl. Acad. Sci. USA* **69**, 3365–3369.
10. Mulder, C. & Delius, H. (1973) *Proc. Natl. Acad. Sci. USA* **69**, 3215–3219.
11. Richardson, C. C., Inman, R. B. & Kornberg, A. (1964) *J. Mol. Biol.* **9**, 46–69.
12. Shenk, T. E., Rhodes, C., Rigby, P. W. J. & Berg, P. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 989–993.
13. Roberts, R. J., Myers, P. A., Morrison, A. & Murray, K. J. (1976) *J. Mol. Biol.* **103**, 199–208.
14. Roberts, R. J., Breitmeyer, J. B., Tabachnik, N. F. & Myers, P. A. (1975) *J. Mol. Biol.* **91**, 121–123.
15. Wilson, G. A. & Young, F. E. (1976) in *Microbiology*, ed. Schlessinger, D. (American Society for Microbiology, Washington, DC), pp 350–357.
16. Melli, M., Ginell, E. & Corneo, G. (1975) *J. Mol. Biol.* **93**, 23–38.
17. Pagoulatos, G. N. & Darnell, J. E. (1970) *J. Mol. Biol.* **54**, 517–535.
18. Aloni, Y., Shani, M. & Reuveni, Y. (1975) *Proc. Natl. Acad. Sci. USA* **74**, 496–500.
19. May, E., Maizel, J. V. & Salzman, N. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 496–500.
20. Aloni, Y., Dhar, R., Laub, O., Horowitz, M. & Khoury, G. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3686–3690.
21. Lavi, S. & Groner, Y. (1977) *Proc. Natl. Acad. Sci. USA* **74**, in press.