Proc. Nat. Acad. Sci. USA
Vol. 71, No. 2, pp. 371–375, February 1974

Nucleotide Sequences of RNA Transcribed in Infected Cells and by *Escherichia coli* RNA Polymerase from a Segment of Simian Virus 40 DNA

(adenovirus 2-SV40/radioactive labeling)

R. DHAR, S. ZAIN, S. M. WEISSMAN, J. PAN, AND K. SUBRAMANIAN

Departments of Human Genetics, Molecular Biophysics and Biochemistry, and Medicine, Yale University School of Medicine, New Haven, Connecticut 06510

Communicated by Aaron B. Lerner, September 17, 1973

ABSTRACT **The nucleotide sequence of 180 residues of an RNA transcript of DNA of simian virus 40 has been deduced. This sequence adjoins a preferred initiation site for *E. coli* RNA polymerase. Comparison of this sequence with that of complementary RNA of simian virus 40 from infected cells has shown that this site also adjoins the apparent 3′ terminus of some of the cytoplasmic complementary RNA of simian virus 40.**

Simian virus 40 (SV40) is a small virus whose genome consists of a closed circular duplex DNA molecule containing approximately 5000 base pairs. Early after infection with SV40, RNA transcripts from 25 to 40% of one strand (the "E" or "−" strand) accumulate in permissive cells, while later the accumulated RNA includes, in addition, transcripts from most or all of the remaining 60–75% of the other strand (the "L" or "+" strand) (1–5). The recombinant viruses, known as nondefective adenovirus 2–SV40 hybrid viruses, Ad2+ND1 and Ad2+ND3 (6), contain continuous segments of SV40 DNA amounting to 15–18% and 6–7% of the genome, respectively (7). These segments include both DNA transcribed only late in the infectious cycle of SV40 and DNA sequences that are transcribed early after infection (5). When SV40 DNA is cleaved by *Hemophilus influenzae* restriction endonuclease a limited number of discrete fragments are obtained, labeled alphabetically A through K in order of decreasing chain length (9–11). The SV40 DNA segment in Ad2+ND1 consists of 70–90% of fragment "G" and 60–80% of fragment "B", while Ad2+ND3 contains the same proportion of the "G" fragment but only 5–10% of fragment "B", (P. Lebowitz, A. Lewis, Jr., T. Kelly, and P. Nathans, manuscript in preparation; B. S. Zain, *et al.*, unpublished observations). *Escherichia coli* RNA polymerase *in vitro* transcribes preferentially the "E" strand of SV40 DNA (12) and frequently initiates transcription at a single start signal (13, 14) present in the DNA common to Ad2+ND1 and the "G" fragment (14). We have analyzed nucleotide sequences of RNA transcribed and accumulated in infected cells and report here a comparison of these sequences with those derived from *E. coli* RNA polymerase transcripts. The apparent 3′ terminus of a major component of the "L"-strand transcript lies very close to a preferred *E. coli* polymerase initiation site on SV40 DNA.

MATERIALS AND METHODS

Spleen acid ribonuclease, U2 ribonuclease, carboxymethylated pancreatic ribonuclease, and "C" ribonuclease were the generous gifts of Dr. G. Bernardi, the Sankyo Co., Dr. P. Lebowitz, and Dr. K. Levy, respectively. Other materials for sequence analysis were obtained from commercial sources as described (15–17). Oligodeoxythymidylate–cellulose was purchased from Research Associates, Inc. *H. influenza* restriction enzymes were prepared by the method of Smith and Wilcox (18) and *E. coli* RII restriction enzymes by the method of Yoshimori and Boyer (19). The cleavage products of ^{32}P-labeled SV40 DNA were prepared by electrophoresis on 4% acrylamide slab gels (17 and refs. therein).

Samples of Ad2+ND1 virus and Ad2+ND3 DNA were the generous gifts of Dr. A. Lewis. SV40 virus strain 777 from Dr. M. Oxman, or a small-plaque isolate of SV40 provided by Dr. D. Nathans, were passed at a multiplicity of infection of 0.5–1 infectious unit per cell in the BSC-1 line of African green monkey kidney cells. Ad2+ND1 was grown in either BSC-1 or Vero green monkey kidney cells or in the KB line of human cells and was never carried more than two passages removed from the primary stock.

Details of cell and virus propagation, purification, and DNA extraction (14), preparation of *E. coli* RNA polymerase (16), transcription of DNA (17), and nucleic acid hybridization, and sequence analysis (15–17) have been described. Briefly, in a typical reaction, 2–3 μg of RNA polymerase holoenzyme was incubated with template DNA (0.1 μg of fragment or 1–3 μg of SV40 DNA) in a reaction mixture containing 90 mM KCl, 33 mM Tris·HCl (pH 7.9), 6 mM mercaptoethanol, 3.3 mM $MgCl_2$, 0.17 mM XTP in a volume of 125 μl. The reaction was allowed to proceed at 37° for 30 min. The reaction mixture was then extracted with an equal volume of water-saturated phenol and passed through a Sephadex G-100 column to remove nucleoside triphosphates. The eluted RNA was precipitated in the presence of 100 μg of carrier nonradioactive tRNA by the addition of two volumes of ethanol. This RNA was either analyzed directly or annealed to DNA immobilized on nitrocellulose filters, eluted, and reprecipitated before analysis.

For sequence analysis, oligonucleotide maps of T1 RNase and pancreatic RNase digests were prepared separately from RNA synthesized in the presence of each one of the four labeled triphosphates. The radioactive oligonucleotides were further analyzed by digestion with pancreatic RNase, U2 RNase, and T1 RNase, as appropriate, and the products were further analyzed by electrophoresis on DEAE–paper at pH 3.5 and 1.7 and by alkaline hydrolysis. The order of the oligonucleotides in the RNA was determined by limited digestion of the RNA with carboxymethylated pancreatic RNase,

Abbreviations: SV40, simian virus 40; Ad2, adenovirus-2; cRNA, complementary RNA.

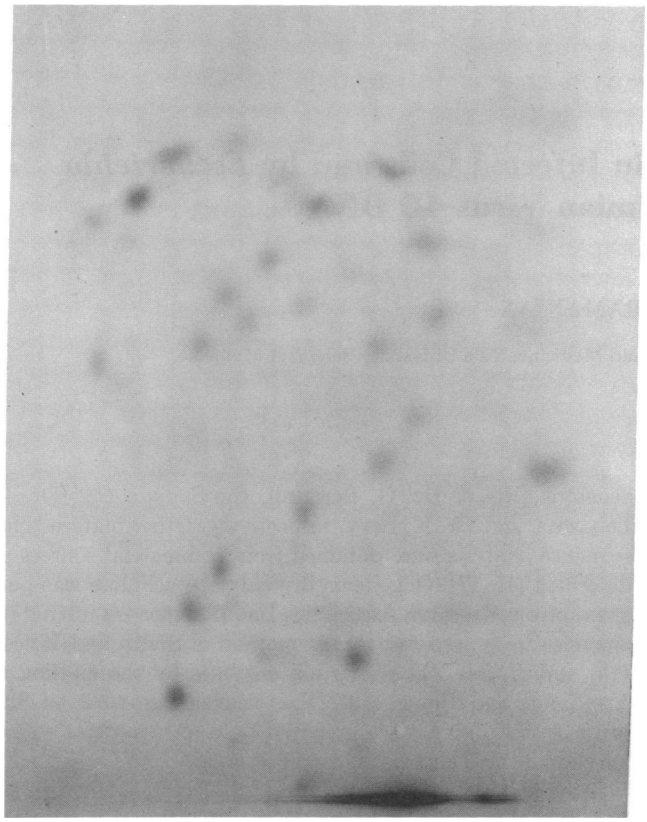

FIG. 1. An oligonucleotide map of T1 RNase digest of the cytoplasmic RNA from infected cells that annealed to the SV40 segment of the Ad2+ND1. Confluent monolayers of BSC-1 in 16 32-oz. bottles, containing approximately 3×10^6 cells per bottle, were infected with SV40 (strain 777) at multiplicity of infection of 75–100 plaque-forming units per cells. Fifteen to 18 hr after infection, the medium was drained and cells were refed with phosphate-free Earle's-minimal essential medium which was prepared by eliminating phosphate from Earle's salt solution in the minimal essential medium formula (30), plus 2% dialyzed fetal-calf serum. [^{32}P]Phosphate was then added to a final concentration of 20 μCi/ml of medium. The cells were harvested at 48 hr after infection, suspended in 20 ml of hypotonic buffer (0.01 M Tris·HCl, pH 7.9–3 mM $MgCl_2$–0.01 M NaCl) at 4°, and disrupted in a Dounce homogenizer. The nuclei were removed by centrifugation at 400 rpm ($800 \times g$) for 10 min, and sodium dodecyl sulfate was added to the cytoplasmic extract to a final concentration of 0.5%. Subsequent steps were performed at room temperature. The RNA from the cytoplasm was extracted with an equal volume of phenol–chloroform (1:1) with 1% isoamyl alcohol. The phenol had been previously saturated with 50 mM Tris·HCl (pH 8.3)–5 mM EDTA–75 mM NaCl–0.5% Na dodecyl sulfate. Three successive extractions with phenol–chloroform (1:1) were done, except now the phenol was water-saturated. One extraction was performed with phenol at 60° (31) and the RNA precipitated with ethanol. The pelleted RNA was taken up in 2 ml of application buffer (0.5 M KCl–0.01 M Tris·HCl, pH 7.5) and applied to a column containing 0.2 g dry weight of oligo(dT)-cellulose (32) previously washed with application buffer. The column was washed with 50 ml of the application buffer at room temperature. The material retained by the column was eluted first with 50 ml/0.01 M Tris·HCl (pH 7.5) and then with distilled water. Carrier tRNA (100 mg) was added to the eluate and the RNA was precipitated with 0.1 volume of potassium acetate and 2 volumes of ethanol. This RNA was pelleted, redissolved in 0.75 ml of 0.30 M NaCl–0.030 M Na citrate containing 0.01% Na dodecyl sulfate, and annealed at 67° for 8 hr to 12 μg of Ad2+ND1 immobilized on a filter (33). The RNA was eluted in the presence of 100 μg of carrier tRNA (14), and precipitated and digested with 25 units in 10 μl of T1 RNase. A two-dimensional fingerprint of the oligonucleotide was prepared. Electrophoresis was from left to right, in pH 3.5 pyridine–acetate buffer, on a strip of Cellogel, and chromatography was from bottom to top with the "homochromo B" solution of Brownlee and Sanger (34).

"C" enzyme, spleen acid RNase, or T1 RNase. The resulting products were fractionated by electrophoresis and homochromotography and analyzed by extensive digestion with pancreatic or T1 RNase followed by electrophoresis on DEAE–paper at pH 1.7.

RESULTS

To correlate SV40 complementary RNA (cRNA) in the infected cell with transcripts from SV40 DNA or its fragments, we prepared radioactive RNA of high specific activity from the cytoplasm of cells in tissue culture which were infected 48–50 hr previously with SV40 virus. This RNA was further fractionated on oligodeoxythymidylate–cellulose columns (20, 32) to obtain the molecules that presumably represent messenger RNA. This fraction of RNA was annealed to Ad2+ND1 DNA to isolate a limited segment of the SV40 message. The RNA eluted from the hybrid was digested with T1 RNase or pancreatic RNase, and oligonucleotide maps were prepared.

The Ad2+ND1 cRNA from cells infected with SV40 gave a discrete set of prominent oligonucleotides (Fig. 1). Comparison of electrophoretic and chromatographic mobility and pancreatic RNase digestion products showed that these oligonucleotides were present in digests of the "G"-fragment transcript (Fig. 2) and therefore derived from the SV40 DNA segment common to "G" and Ad2+ND1. Unless there is extensive sequence heterogeneity in a portion of the region common to "G" fragment and Ad2+ND1, or selective loss of certain sequences in the hybridization procedures such as might result from competition by RNA·DNA hybridization because of overlapping transcription of the two strands of DNA (21, 22), a large part of the cytoplasmic SV40 "L"-strand cRNA has a termination point very close to one end of the "G" fragment. Exactly the same results were obtained when we analyzed RNA extracted from the Vero continuous line of African green monkey kidney cells rather than from the BSC-1 line of this cell type. In all the maps of cRNA from infected cells, there were substantial amounts of radioactive material that did not move from the application line in the second dimension. This material was not further identified.

In an effort to disrupt double-stranded RNA structures that might interfere with DNA·RNA annealing, the RNA solution was heated to 100° for 5 min and quickly cooled immediately before it was annealed to DNA. No change in the oligonucleotide pattern was observed. A similar pattern was observed when RNA primed by "G" fragment and lacking poly(adenylic acid) was annealed to Ad2+ND1 DNA.

Too little radioactivity was present in the RNA from infected cells for complete sequence analysis. However, sequence analysis of RNA transcribed from the "G" fragment has demonstrated that the oligonucleotides derived from RNA complementary to the "L" strand of SV40 DNA and present in digests of the mRNA come from a continuous RNA sequence.

Recently other groups have obtained results indicating that SV40 transcripts from infected cells, complementary to DNA of Ad2+ND3 (23) and the "G" fragment (24), come principally from the "L" strand. Our results confirm their conclusions.

When large amounts of radioactive mRNA from SV40-infected cells were annealed to Ad2+ND1, oligonucleotide maps showed a second, less prominent set of spots. Many of these occurred in the same position in oligonucleotide maps as products in digests of SV40 RNA synthesized by *E. coli* RNA polymerase and complementary both to the "E" strand of the "B" fragment and to Ad2+ND1 DNA. While these might include Ad2+ND3 sequences, it is difficult to confirm this with the limited amount of radioactivity available. The results support the earlier report of a transition point from "L"- to "E"-strand transcription in the SV40 segment present in Ad2+ND1 (5). In collaboration with Patch, Lewis, and Levine, we have verified these conclusions by the use of separated strands of Ad2+ND1 (manuscript in preparation). Apparently even 18–48 hr after infection of BSC-1 or Vero cells with SV40, "E"-strand transcripts were made, poly(adenylic acid) sequences were attached to them, and they were exported from the nucleus and attached to ribosomes. More copies accumulated of the "L"-strand transcripts of the "G" fragment than of "E"-strand transcripts of the "B" fragment, suggesting a difference in transcription rates or stability of the transcripts.

In other experiments, RNA transcribed *in vitro* from SV40 DNA was annealed to Ad2+ND1 and Ad2+ND3 so as to isolate RNA transcribed from the "E" strand of the "G" and "B" fragments. The nucleotide sequence of the RNA complementary to the SV40 segment of Ad2+ND3 was analyzed. Comparison of this partial sequence with that of RNA transcribed directly from the "G" fragment shows that the sequences found in infected cells are complementary to a portion of the "E"-strand transcripts of Ad2+ND1 and Ad2+ND3 (Fig. 3). This result confirms that the strand transcribed by *E. coli* RNA polymerase is the opposite strand of DNA from that transcribed in infected cells, and should make it possible to determine the nucleotide sequence preceding a point at which *E. coli* RNA polymerase starts transcription on SV40 DNA.

To further locate and orient these sequences in SV40 and Ad2+ND1, we made use of *E. coli* RII restriction endonuclease. After digestion of SV40 DNA with this enzyme, we obtained about 14 DNA fragments. The largest fragment contained *H. influenza* restriction enzyme fragments J, F, and about $^2/_3$ of G; a second RII fragment (the fifth largest) contained the rest of G (G_2) and part of *H. influenza* fragment "B". The relation between the *H. influenza* and RII restriction fragments was confirmed by cleavage of each RII fragment with *H. influenza* enzyme and vice versa, and by comparison of oligonucleotide maps of SV40 RNA annealed to each fragment. After cleavage of *H. influenza* "G" fragment with RII enzyme, the two products "G1" and "G2" were isolated by acrylamide gel electrophoresis and transcribed. Maps of the transcripts showed that "G2" contained sequences 1–112 of Fig. 3 and "G1" contained sequences 120–180 and the remaining sequences shown in Fig. 3. Therefore, the *E. coli* polymerase initiation site lies very near the junction of the "G" and "B" fragments in SV40, and transcription proceeds away from "B". This result confirms the results of Khoury *et al.* (8).

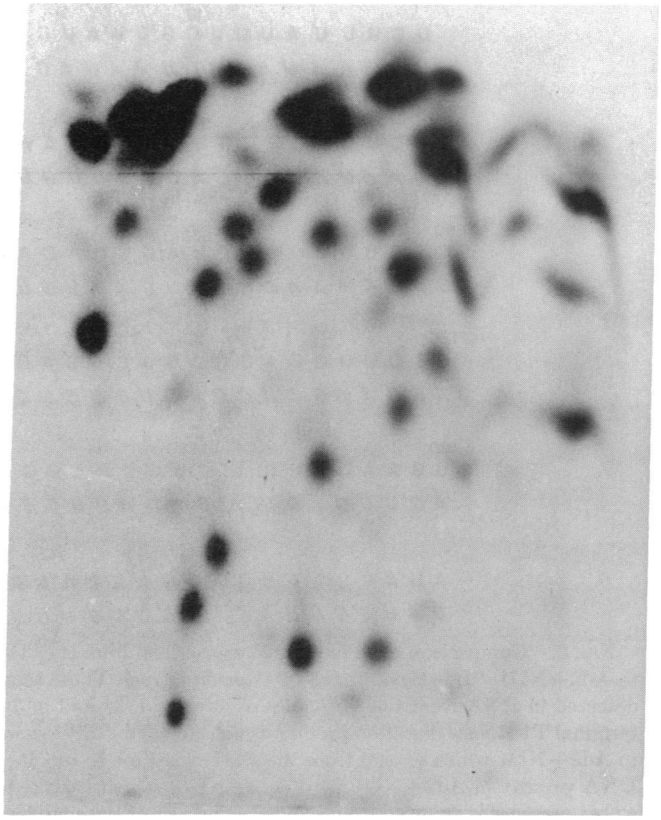

FIG. 2. A two-dimensional fingerprint of a T1 RNase digest of the *E. coli* RNA polymerase transcript of the "G" fragment. Transcription of the DNA fragment was performed in the presence of 0.15 mM [α-^{32}P]GTP as described in the *text*. Electrophoresis and chromatography of a T1 RNase digest were performed exactly as described in the legend to Fig. 1.

DISCUSSION

The proximity of the 3′ terminus of a major component of the SV40 RNA accumulated in intact cells that anneals to Ad2+ND1 DNA and the site of initiation of *E. coli* RNA polymerase on SV40 DNA *in vitro* (Fig. 4) would appear to be more than coincidental and suggests that there may be common structural features or interaction of the *E. coli* RNA polymerase initiation sites and sites where transcription terminates or where post-transcriptional cleavage of the RNA occurs.

During brief incubation periods at reduced temperatures, *E. coli* RNA polymerase transcription of SV40 DNA begins at nucleotide 3, proceeds through nucleotide 135, but does not reach nucleotide 166 of the sequence in Fig. 3 (14). There is no conclusive feature of the sequence to explain this, although it is not impossible that pairing between nucleotides 152 and 159 and nucleotides 164 and 170 may play a role. Although "G" fragment includes a preferred start for *E. coli* RNA polymerase, this start is not active when the isolated fragment is used as template since very little "E"-strand transcript is made. Further sequences may be necessary for this start to function or there may be interference between the polymerase molecules on the *in vivo* "L"-strand termination and the *in vitro* "E"-strand initiation site.

SV40 DNA synthesis terminates, on the average, in a localized region of the molecule variously described as near the junction of the "G" and "B" fragment (8) or 0.17 map units from the site of the R1 restriction endonuclease cut in SV40

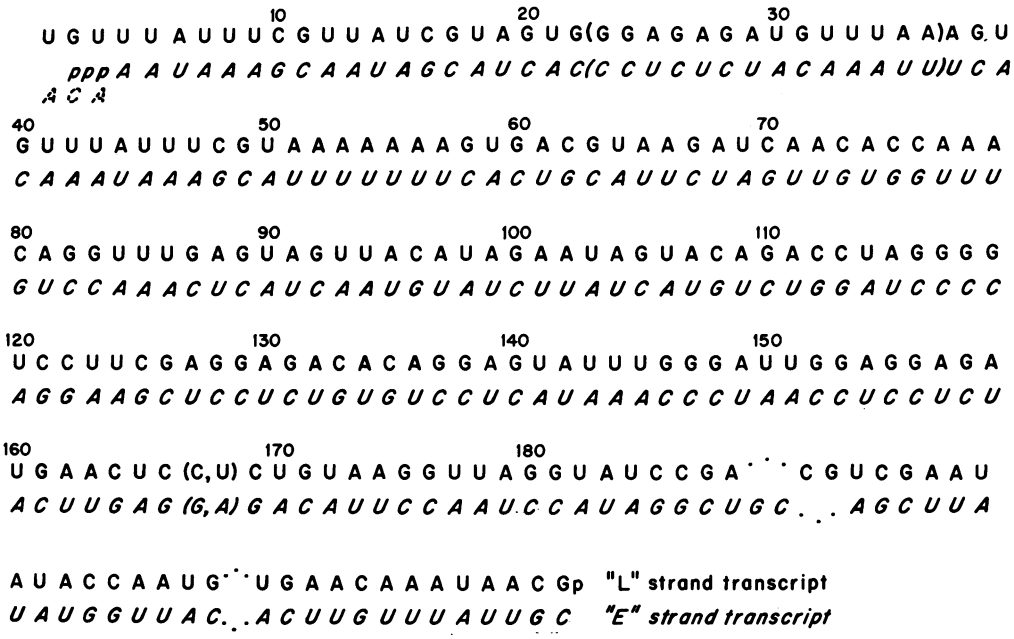

FIG. 3. Comparison of the sequence transcribed from the "G" fragment of SV40 and those transcribed from intact SV40 and annealed to Ad2+ND1. *"L"-strand transcript (standard type)*: Those sequences derived by analysis of transcripts of "G" fragment which are also detected in RNA from the ribosomes of infected cells and in "G"-fragment transcripts that had been annealed to Ad2+ND1. The terminal T1 RNase digestion product is not included. *"E"-strand transcript (italics)*: Those sequences present in SV40 transcripts annealed to Ad2+ND1 which extend from the start signal for *E. coli* RNA polymerase (14) to the end of the "L" strand of the segment of SV40 DNA present in Ad2+ND1 (and Ad2+ND3) which is linked by its 3′ hydroxyl to adenovirus 2 DNA. *Dotted letters* are nucleotides whose presence is deduced from sequence analysis of the complementary-strand transcript. *Parentheses* enclose sequences whose order is not determined; (· · ·) separates sequences that may be joined directly or separated by a few nucleotides, but for which the full sequence of the junction is not known. The nucleotide sequences proposed are based on independent analyses of sequences from "E"- and "L"-strand transcripts, except for sequences of the largest T1 RNase digestion products, which are based in part on complementary-strand sequence data. "E"-strand transcripts were obtained by hybridization of transcripts of intact SV40 DNA to Ad2+ND3 DNA, and RNA containing principally "L"-strand transcripts was obtained by use of "G"-fragment DNA as a template (17).

(25). The "G" end-fragment of the Ad2+ND1 SV40 segment lies 0.11 map units from the R1 cut (26). The proposed sequence from the end of Ad2+ND1 to the start for *E. coli* RNA polymerase on the "L" strand of SV40 covers approximately 0.04 map units. These estimates would place the site of termination of DNA synthesis somewhere in the 100 nucleotides of the "G" fragment nearest the "B" fragment and within or very near the sequence shown in Fig. 3. It is not clear whether this termination site is a specific sequence or just the region of DNA most distant from a replication initiation site.

There is a repeat of nucleotides located at position 1–12 and again at positions 39–50 from the 5′ end of the "L"-strand sequence in the proposed sequence. We have observed a further partial repeat of a somewhat similar sequence lying within the "B" fragment very close to the "G" fragment. These repeats of relatively short sequences are reminiscent of the structures proposed for satellite DNA of animal cells (27, 28). The marked paucity of CpG sequences here, and in the entire easily transcribed portions of the Ad2+ND1 SV40 DNA segment is even more striking than that reported earlier for whole SV40 DNA (29). Finally, the presence of termina-

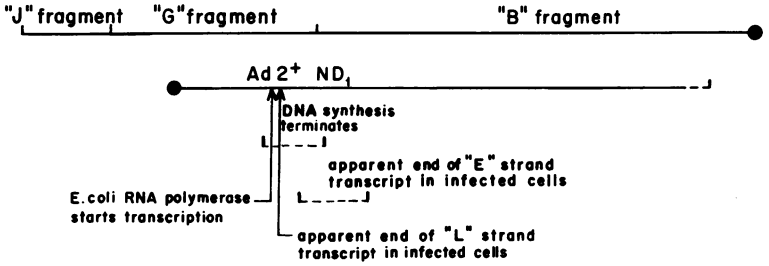

FIG. 4. Schematic diagram of the regions of SV40 DNA including the Ad2+ND1 segment of SV40. The location of the site where DNA synthesis terminates is based on data of Nathans and Danna (24) and Fareed *et al.* (25). The location of the "B" fragment is based on comparison of oligonucleotide maps of transcripts of that fragment with those of SV40 transcripts annealed to Ad2+ND1. The *upper line* represents a portion of the "L" strand of SV40 DNA containing "J", "G", and "B" fragments. The *lower horizontal line* represents the "E" strand of the portion of SV40 DNA included in Ad2+ND1. The *solid dot* at one end of each line represents the 5′-phosphoryl terminus of each DNA strand. The polarity of the strands is consistent with observations on the direction of transcription of SV40 DNA linear forms produced by cleavage with R1 restriction endonuclease (26) and with the results of DNA·RNA hybridization studies (8), as well as the results reported here.

tion triplets in each possible reading phase of the "L"- and "E"-strand sequences indicates that there are over 50 untranslated 3'-terminal nucleotides preceding the poly(adenylic acid) in an SV40 late RNA, provided UGA is a terminator in animal cells.

We acknowledge the excellent technical assistance of Miss Su-Fang-Su and Frances DeNoto. We thank Dr. R. A. Weinberg for introducing us to the method of RNA extraction from infected cells and to Drs. R. M. Herriot and H. O. Smith for sending us the strain of *H. influenzae*, Dr. D. Nathans for sending us samples of fragments of SV40 produced by cleavage with restriction endonuclease from *H. influenzae*, and Jeffrey Sklar for giving us samples of the restriction enzyme. We also thank Drs. C. T. Patch, A. S. Levine, and A. Lewis, Jr. for their critical discussions during the investigation and also for reviewing the manuscript. This work was supported by the American Cancer Institute, Grant #ACS-VC-1D and by the National Cancer Institute, Grant # USPHS 5-RO1-CA-10, 137-06.

1. Lindstrom, D. M. & Dulbecco, R. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 1517–1525.
2. Khoury, G. & Martin, M. M. (1972) *Nature New Biol.* **238**, 4–6.
3. Khoury, G., Byene, J. C. & Martin, M. M. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 1925–1928.
4. Sambrook, J., Sharp, P. A. & Keller, W. (1972) *J. Mol. Biol.* **70**, 57–71.
5. Patch, C. T., Lewis, A. M., Jr. & Levine, A. S. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 3375–3379.
6. Lewis, A. M., Jr., Levin, M. J., Wiese, W. H., Crumpacker, C. S. & Henry, P. H. (1969) *Proc. Nat. Acad. Sci. USA* **63**, 1128–1135.
7. Kelly, T. J., Jr. & Lewis, A. M., Jr. (1973) *J. Virol.*, in press.
8. Khoury, G., Martin, M. & Nathans, D. (1973) *J. Mol. Biol.*, in press.
9. Danna, K. & Nathans, D. (1971) *Proc. Nat. Acad. Sci. USA* **68**, 2913–2917.
10. Nathans, D. & Danna, K. (1972) *Nature New Biol.* **236**, 200–202.
11. Danna, K. (1973) Dissertation submitted to the Johns Hopkins University.
12. Westphal, H. (1970) *J. Mol. Biol.* **50**, 407–420.
13. Westphal, H., Delius, H. & Mulder, C. (1973) *Lepetit Collequim Biol. Med.*, in press.
14. Zain, B. S., Dhar, R., Weissman, S. M., Lebowitz, P. & Lewis, A. M., Jr. (1973) *J. Virol.* **11**, 682–693.
15. Sanger, F., Brownlee, G. G. & Barrell, B. G. (1965) *J. Mol. Biol.* **13**, 373–398.
16. Lebowitz, P., Radding, C. & Weissman, S. M. (1971) *J. Biol. Chem.* **246**, 5120–5139.
17. Marotta, C. A., Lebowitz, P., Dhar, R. & Weissman, S. M. (1973) in *Methods in Eynzmology* (Academic Press, New York), Vol. X, in press.
18. Smith, H. & Wilcox, X. (1970) *J. Mol. Biol.* **51**, 379–383.
19. Yoshimori, X. (1973) Dissertation to University of California.
20. Aviv, H. & Leder, P. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 1408–1412.
21. Aloni, Y. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 2404–2409.
22. Fried, A. H. (1972) *J. Virol.* **10**, 1236–1239.
23. Patch, C. T., Lewis, A. M., Jr. & Levine, A. J., personal communication.
24. Danna, K. J. & Nathans, D. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 3097–3100.
25. Fareed, G., Claude, F., Garen, A. & Salzman, N. P. (1972) *J. Virol.* **10**, 484–491.
26. Morrow, J. F. & Berg, P. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 3365–3369.
27. Southern, X. (1970) *Nature* **227**, 794–798.
28. Yunis, J. J. & Yasmineh, W. G. (1971) *Nature* **174**, 1200–1209.
29. Subak-Sharp, H., Shephard, W. M. & Hay, J. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31**, 583–591.
30. Eagle, H. (1959) *Science* **130**, 432–437.
31. Penman, S. (1966) *J. Mol. Biol.* **17**, 117–130.
32. Gilham, P. (1964) *J. Amer. Chem.* **86**, 4982–4985.
33. Gillespie, D. & Spiegelman, S. (1965) *J. Mol. Biol.* **12**, 829–842.
34. Brownlee, G. G. & Sanger, F. (1969) *Eur. J. Biochem.* **11**, 395–401.