Volume 4 Number 8 August 1977 Nucleic Acids Research

# 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

M. L. Celma*, R. Dhar**, J. Pan and S. M. Weissman

Department of Human Genetics, Yale University School of Medicine, New Haven, CT 06510, USA

Received 13 June 1977

ABSTRACT

The 16S late mRNA from SV40 directs the synthesis of the major viral structural protein, VP1. We have compared the oligonucleotides in the 16S mRNA with those that would be present in a transcript of the portion of SV40 DNA coding for VP1. The results indicate that a segment of about 200 nucleotides of RNA transcribed from a distant part of SV40 DNA have become linked to the transcript of VP1 codons by a bond resistant to phenol extraction and denaturation in formamide.

## INTRODUCTION

In the preceding report we presented the nucleotide sequence of the portion of SV40 DNA that encodes the amino acids of the major structural protein VP1 (1) of SV40 virus, together with the nucleotide sequence in the DNA immediately preceding the coding portion of the gene. Elsewhere, we have presented the nucleotide sequence of the SV40 DNA extending beyond the nucleotides coding for the carboxy terminal amino acids of VP1 (2). We have also presented evidence that the messenger RNA for VP1 extends to the junction of the SV40 DNA fragments Hind II,III-G[1] and Hind II,III-B (see Fig. 1 of preceding paper) (3), so that there is a total of approximately 70 untranslated nucleotides between the translation termination codon and the beginning of the polyadenylic acid of the VP1 mRNA. We report here the results of the analyses of oligonucleotides from the mRNA from VP1. These results indicate that the mRNA derived from this region of the genome is strongly linked to other RNA derived from a portion of the SV40 DNA 1,000 nucleotides away from the initiator codon for VP1. The transcript of the intervening DNA is not present in the VP1 mRNA.

## MATERIALS AND METHODS

The procedures for virus growth and purification and for growth and labeling of cells in the presence of $^{32}P_i$ have been described before (4). RNA was isolated from the cytoplasm of infected cells and chromatographed on oligo (dT) cellulose. The polyadenylated RNA was then used for futher

analysis. To prepare RNA complementary to SV40 DNA fragments, cytoplasmic polyadenylated RNA was annealed to DNA fragments immobilized on nitrocellulose filters, the filters were treated with $T_1$ RNase and washed. RNA was then eluted at 100° and recovered by alcohol precipitation (see reference 5 for details). The methods for RNA sequence analysis have been described in detail (5, 6 et op. cit.). Briefly, RNA was digested extensively with $T_1$ RNase and the products fractionated two-dimensionally by electrophoresis on Cellogel at pH 3.5 and by homochromatography and DEAE TLC plates. Products were eluted and analyzed with pancreatic or $U_2$ RNase. Since the entire nucleotide sequence of SV40 is now available (2,5,6,7; Reddy, V.B., Thimmappaya, B., Weissman, S.M., unpublished results), this made it possible to correlate individual products with unique longer DNA sequence.

## RESULTS AND DISCUSSION

SV40 late mRNA was prepared, labeled with $^{32}P_i$ as described in the Methods section. This RNA was then annealed to total SV40 DNA in 50% formamide at 37°, eluted and fractionated into 16 and 19S mRNA by electrophoresis in an acrylamide formamide gel. The RNA was recovered by alcohol precipitation and electrophoresed on 2.7% acrylamide gels. The results showed that there was a somewhat broad band of SV40 mRNA at about 16S and a doublet of mRNA moving slightly slower than the 18S mRNA. This latter was presumably 19S mRNA and might include any separate mRNA species coding for the structural proteins VP3 or VP2. The early mRNA is at most only a few per cent of the late mRNA and would not be clearly seen in this type of analysis. Each RNA band was then eluted and recovered by alcohol precipitation. A portion was treated with formamide again and rerun on a 2.7% polyacrylamide gel, alongside of markers (Fig. 1). This demonstrated that the RNA was intact and did not contain any complexed low molecular weight RNA. The remainder of each RNA was then digested with $T_1$ ribonuclease and oligonucleotide maps were prepared. The results for the 16S RNA are shown in Fig. 2. Each oligonucleotide was further analyzed as described in the Methods section.

In separate experiments radioactive polyadenylated RNA was prepared by the same method and annealed to the restriction endonuclease fragments Eco RII-A, Eco RII-F, Hind II,III-K, and Hind II,III-G, of SV40 DNA. The complementary DNA was treated with nuclease to remove tails of unhybridized RNA before it was eluted from the DNA (see Methods). Therefore, unlike the experiment with mRNA isolated by gel electrophoresis, RNA covalently linked to VP1 sequences but not transcribed from adjacent DNA would have been removed. The location of the restriction fragments relative to the VP1 codons

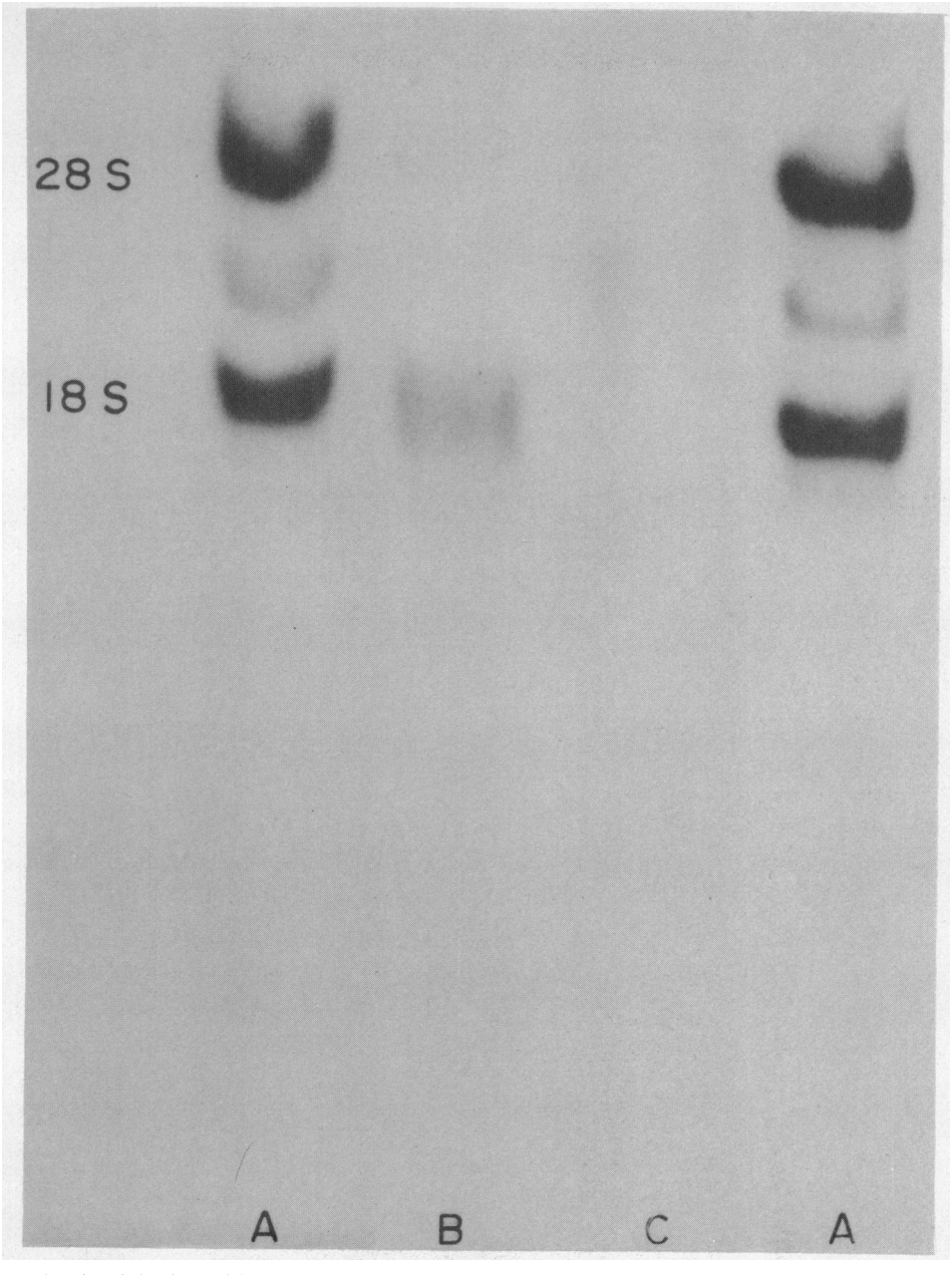


Fig. 1 Autoradiograph of acrylamide gel electrophorectic pattern of purified SV40 16S and 19S mRNA. Five roller bottles of the VERO line of continuous African Green Monkey kidney cells were infected with SV40 at a multiplicity of infection or approximately 12 plaque forming units per cell. Twenty-four houre after infection 30 mCi of $^{32}P$ was added in phosphate free medium with 5% dialyzed serum and incubation was continued an additional 17 hours. Cells were disrupted in buffer containing 0.5% NP40 and RNA was extracted from the cytoplasm by a phenol-sodium dodecyl sulfate chloroform procedure (4). Polyadenylated RNA was prepared with an oligo (dT) cellulose column and annealed to 40 $\mu$g immobilized SV40 DNA at 37° for 6 hr. in a buffer containing 50% formamide, 0.5M NaCl, 0.1M TRIS (pH 7.5), 0.5% sodium dodecyl sulfate. The filters containing the SV40 DNA were then washed with 50ml of the same buffer and RNA was eluted in 4.5ml 90% formamide, 0.01M TRIS (pH 7.5) 0.1% sodium dodecyl sulfate at 65° for 10 min. RNA was recovered by alcohol precipitation and resuspended in 0.3 ml formamide and .01M sodium phosphate buffer (pH 6.9), and electrophoresed through a 4% acrylamide gel in 98% formamide. A total of $3.9 \times 10^7$ cpm of polyadenylated RNA was recovered. $1.7 \times 10^6$ cpm of SV40 cRNA was eluted from the SV40 DNA filters. $7 \times 10^4$ cpm of 19S mRNA and $9.5 \times 10^4$ cpm of 16S RNA were recovered from the formamide acrylamide gel.
Aliquots of the 16S and 19S mRNA were suspended in formamide and electrophoresed in a 2.7% acrylamide gel alongside of radioactive 18 and 28S ribosomal RNA marker
(a) 18 and 28S ribosomal RNA;
(b) 16S mRNA;
(c) 19S mRNA.

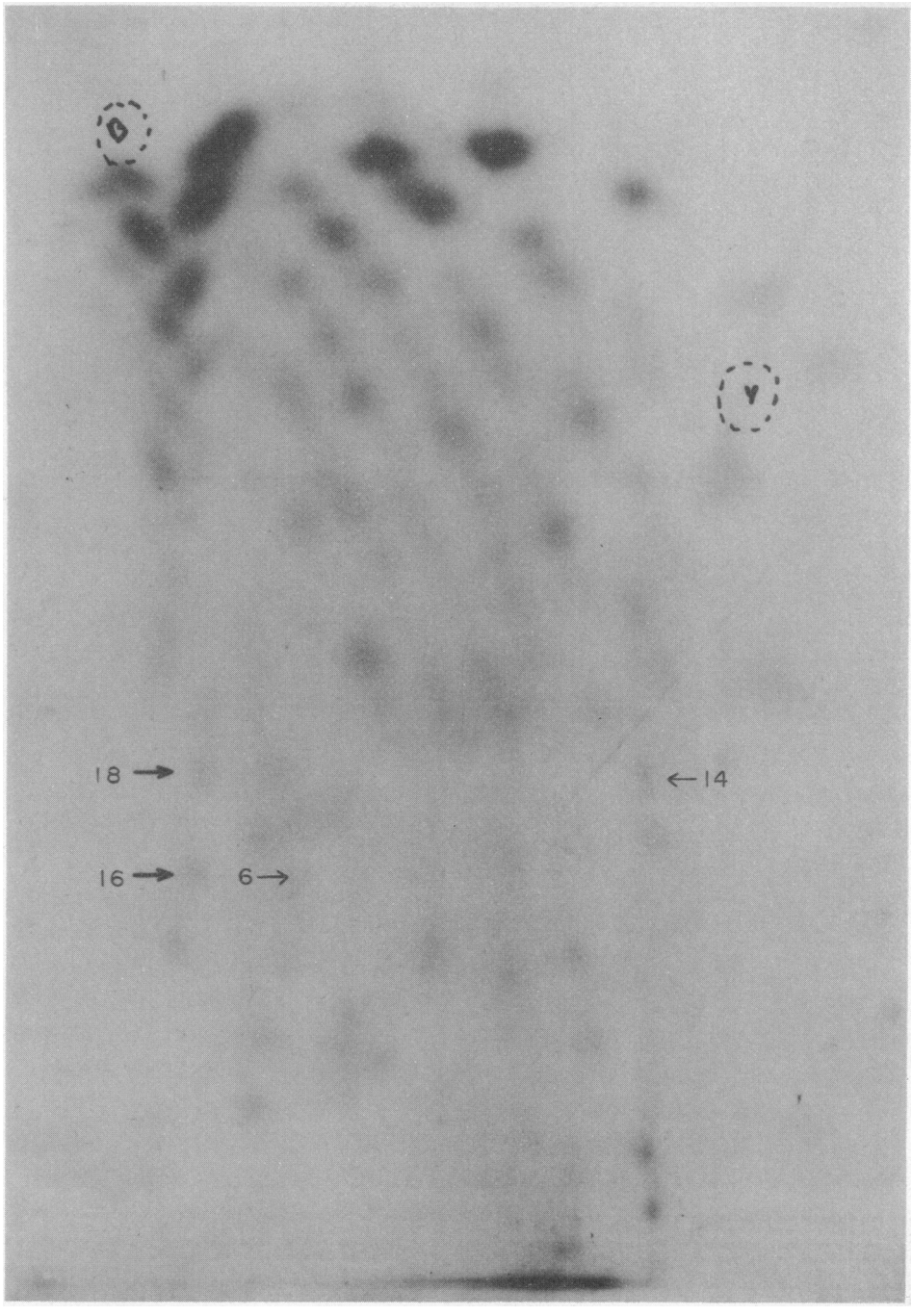


Fig. 2 Oligonucleotide map of SV40 16S mRNA. The 16S mRNA was prepared as described in Methods and in the legend of Fig. 1. The purified RNA was digested with $T_1$ ribonuclease and the resulting oligonucleotides fractionated by electrophoresis from left to right on Cellogel at pH 3.5 and by chromatography from below upwards on DEAE TLC plates. Arrows with numbers refer to oligonucleotides derived form the "agnogene" encoded with Hind II, III-C, L and M, whose transcript is joined to the transcript for the 16S codons to form the cytoplasmic 16S mRNA. Numbers refer to oligonucleotides whose sequence is presented in Table I.

is shown in Fig. 1 of the accompanying report (1). We have published oligonucleotide maps of SV40 polyadenylated RNA annealed to Eco RII-F (3), a fragment that spans the 3' end of the 16S mRNA and is contiguous to Eco RII-A. Oligonucleotides in the map of mRNA complementary to Eco RII-F included all the $T_1$ products from the transcript of the late strand of Hind II,III-G and no oligonucleotide from the late strand of the adjacent fragment Hind II,III-B (3) nor any other prominent oligonucleotides except polyadenylic acid. This indicated that polyadenylic acid but no guanylic acid residues were linked to the 3' end of the transcript of the late strandof Hind II,III-G present in the cytoplasm of infected cells. A representative oligonucleotide map from cytoplasmic polyadenylated RNA annealed to Eco RII-A is shown in Fig. 3. The fragment spans most of Hind II,III-K, all of Hind II,III-F and Hind II, III-J and 2/3rds of Hind II,III-G. Comparison of the nucleotide sequence of the VP1 gene showed that the large oligonucleotides present in the RNA complementary to Eco RII-A were all predicted from the sequence. Most of the oligonucleotides present in the purified 16S mRNA were also accounted for by the nucleotide sequence and detected in mRNA complementary to Eco RII-A, Eco RII-F or Hind II,III-K. However, there was a definite group of oligonucleotides present in equal molar amounts with the oligonucleotides in the coding region of VP1 mRNA that could not be accounted for, either by coding sequences or by nearby sequences within the DNA, and that were not present in mRNA complementary to Eco RII-F, Eco RII-A, or Hind II,III-K. Some of these products are numbered in Fig. 2 and their sequences are presented in Table I. Comparison of these sequences with the known sequences of SV40 DNA showed that the only source for these oligonucleotides within the SV40 DNA was from a portion of the DNA fragment Hind II,III-C and the adjacent small DNA fragments Hind II,III-L and M (Fig. 4,5).

The sequence of this portion of SV40 DNA has been published previously and the oligonucleotides present in VP1 mRNA not derived from contiguous regions of DNA could all be accounted for by a transcript of a continuous stretch of approximately 200 oligonucleotides beginning within Hind II,III-C and extending across Hind II,III-L,M and N, but not into Hind II,III-D. We have previously pointed out that the transcript of this portion of SV40 is present abundantly in polyadenylated RNA in the cytoplasm of SV40 RNA infected cells (6) and that there is a gap on the late mRNA just beyond the transcript of Hind II,III-L, M. The four longest $T_1$ RNase oligonucleotides produced from the transcript of this region were all present in SV40 16S mRNA. The smaller products overlapped products from the coding region of the VP1

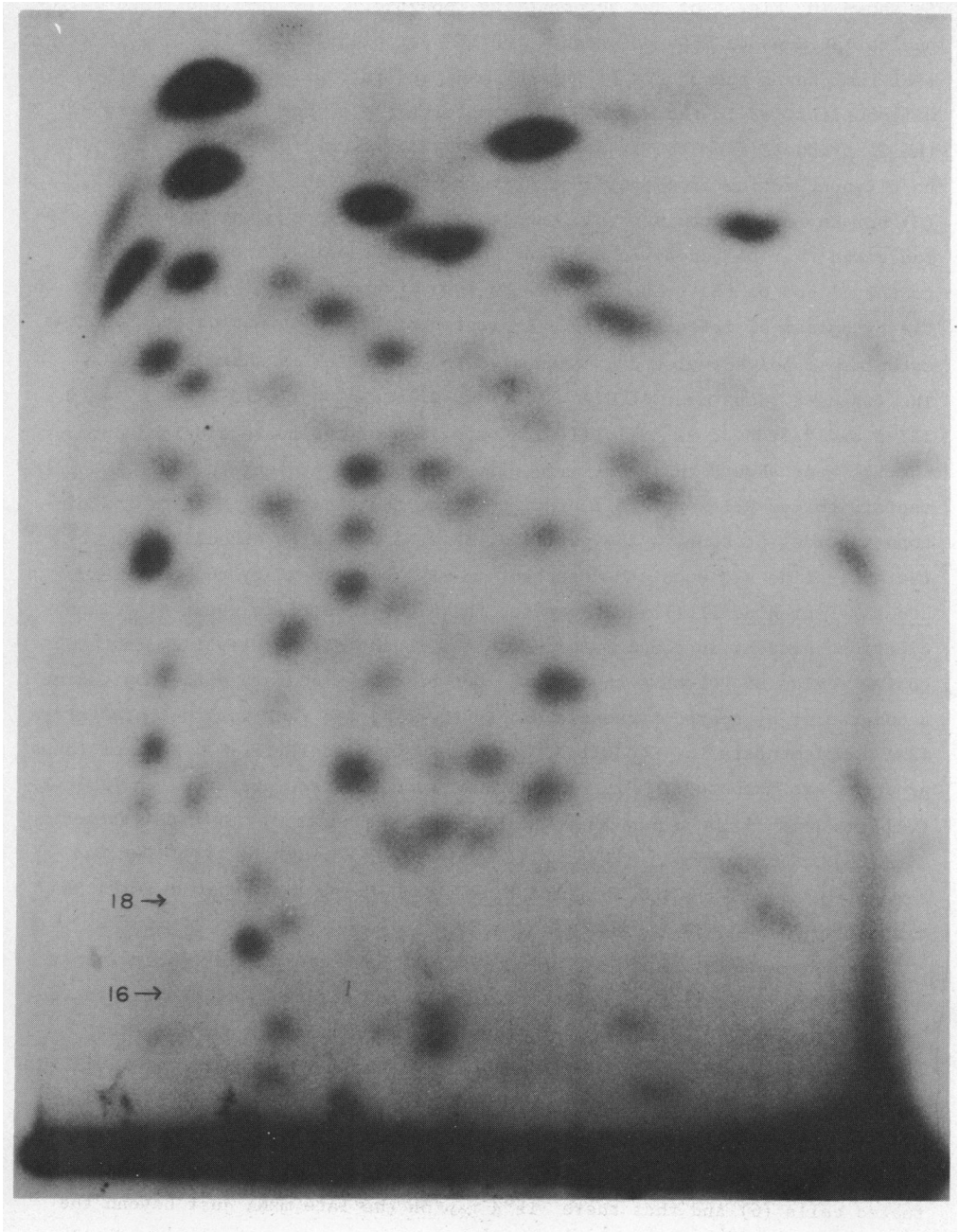


Fig. 3 Oligonucleotide map of SV40 cytoplasmic mRNA complementary to the DNA fragment Eco RII-A. The location of Eco RII-A on the SV40 genome is shown in Fig. 1 of the accompanying manuscript. The oligonucleotides shown were all identified as products of the transcript of the continuous segment of SV40 DNA containing the codons for VP1. Of note is the fact that the oligonucleotides indicated by the arrows in Fig. 2 are not present in the present autoradiogram. Arrows in Fig. 3 indicate the positions where oligonucleotides 16 and 18 (see Table I and Fig. 2) would have appeared.

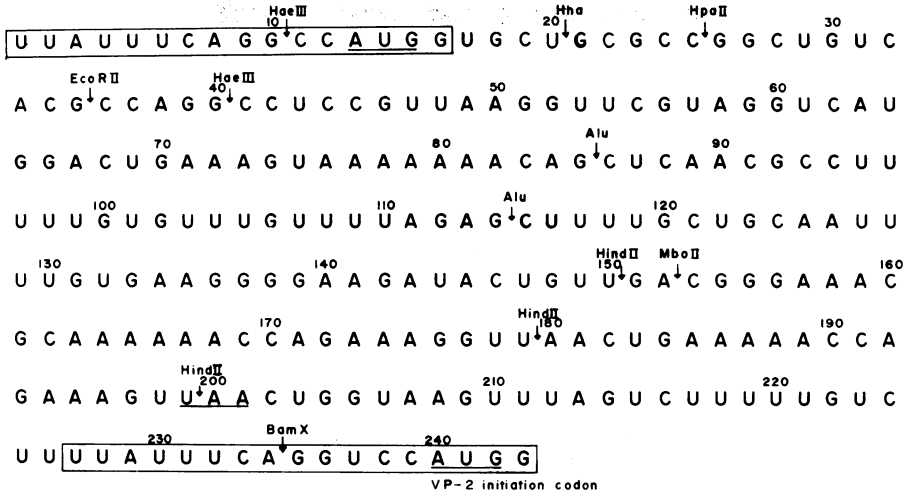


Fig. 4 Nucleotide sequence of the late strand transcript of an SV40 DNA segment contained within Hind II, III-C, L, M, and D. The portion of this transcript between about postion 10 and 205 is joined to the transcript of the VP1 codons in the 16S mRNA. Boxes enclose similar sequences about the AUG in the "agnogene" and the AUG initiation codon for VP2 (minor structural protein of SV40). Arrows show sequence corresponding to restriction endonuclease cleavage sites.

gene and were more difficult to distinguish. However, when total SV40 polyadenylated RNA was annealed to a DNA fragment spanning this region, all products of the 200 base long RNA sequence were present in molar ratios and products beyond either end of the region were less prominent (Fig. 5, Table I). It therefore seems likely that the entire transcript of this region is attached to the 16S mRNA. The sequence of this portion of Hind II,III-C, L and M is also of interest, in that it contains an AUG codon followed by 59 sense triplets, followed by a termination codon. It is unknown whether this sequence is translated or what other functions it may have, and we have referred to this as an "agnogene". The present results indicate that the transcript of this region of DNA could potentially code for a small peptide and is linked to the 16S mRNA for VP1, while none of the transcript of the intervening 1,000 nucleotides of SV40 DNA is present in the mRNA for VP1. One of us (M.L. Celma) has noted that the SV40 16S mRNA isolated by formamide gel electrophoresis contains material with the chromatographic mobility of six methyl adenylic acid and also "capped" structures. S. Lavi had found that the methylated capped structures from SV40 late mRNA annealed to the Hind II,III-C fragment. This is also consistent with the present results and indicates that RNA transcribed from a portion of Hind II,III-C fragment becomes firmly associated with the RNA transcribed from the Hind II,III-K, J, F and G fragments in spite of the large distance between the fragments. After the present results were obtained we learned of evidence from R. Roberts, T. Broker and their colleagues and from P. Sharpe and his colleagues (8), in-

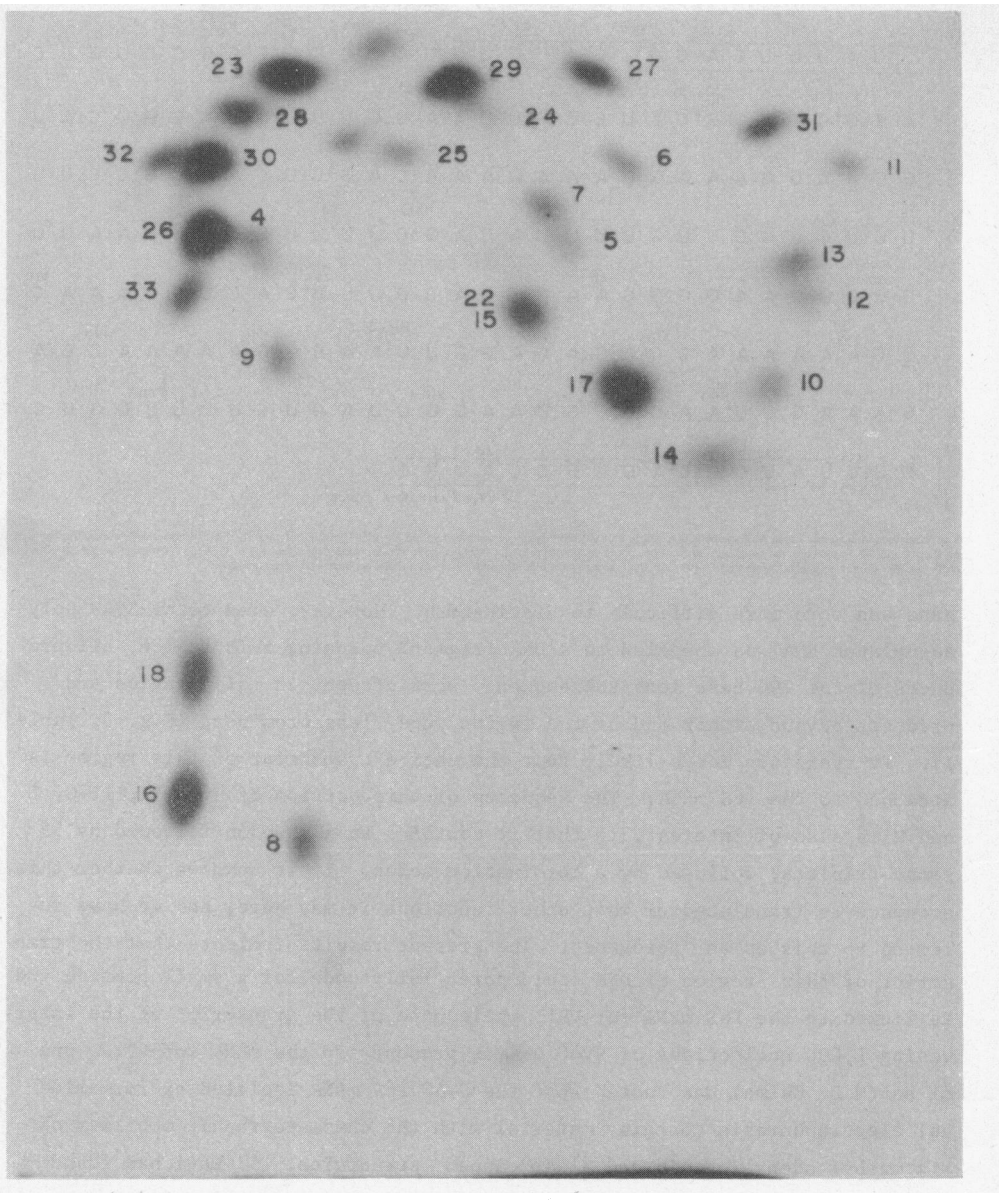


Fig. 5 Oligonucleotide map of cytoplasmic polyadenylated RNA from SV40 infected cells complementary to the DNA fragment Hae III-F and G. The location of the DNA fragments is shown in Fig. 1 of the accompanying manuscript. Numbers refer to oligonucleotides whose sequence is presented in Table I. Of note is the fact that although the Hae III-F and G fragments comprise a total of 12% of SV40 DNA, the most prominent oligonucleotides could all correspond with the continuous transcript of the "agnogene" whose sequence is given in Fig. 4. The oligonucleotides immediately preceding and following the segment of transcript from nucleotide 11 to nucleotide 205 of Fig. 4 are not detectable in the map of in vivo RNA, suggesting that there is a gap following and perhaps preceding the transcript of the "agnogene". The $T_1$ RNase product CCAUG was present in the transcript in molar yield but could also have arisen from transcript of upstream sequences within Hind II, III-C nearer Hind II, III-A.

dicating that in an Adenovirus mRNA, 5' sequences of the mRNA may be derived by transcription of regions of DNA widely removed from the DNA complementary to most of the mRNA. The mechanism of such a translocation or of RNA is not immediately apparent. There is no strong similarity nor complementarity be-

TABLE I

Nucleotide sequence of oligonucleotides in $T_1$ RNase digest of SV40 mRNA complementary to the SV40 DNA fragments Hae III-F and G.(1)

| Oligonucleotide number(2) | Sequence(3) |
|---|---|
| 1 | UUAUUUCAG |
| 2 | CCAUG |
| 3 | UCACG |
| 4 | CCUCCG |
| 6 | UUCG |
| 7 | UCAUG |
| 8 | UAAAAAAACAG |
| 9 | CUCAACG |
| 10 | CCUUUUUG |
| 11 | UUUG |
| 12 | UUUUAG |
| 13 | CUUUUG |
| 14 | CAAUUUUG |
| 15 | AUACUG |
| 16 | CAAAAACCAG |
| 18 | AAAAACCAG |
| 19 | UUUAG (4) |
| 20 | UCUUUUUG(4) |
| 21 | UCUUUUAUUUCAG(4) |
| 22 | UCCAUG |
| 23 | G |
| 24 | UAG |
| 25 | ACUG |
| 26 | AAAG |
| 27 | UG |
| 28 | AG |
| 29 | CUG |
| 30 | AAG |
| 31 | UUG |
| 32 | ACG |
| 33 | AAACG |
| 34 | UAAG |

1) See Figure 1 of accompanying report.

2) Numbers refer to oligonucleotides shown in Figure 5

3) Sequence was deduced from results of further enzyme digestion of the oligonucleotides as described in Methods and from comparison with the DNA sequence.

4) These oligonucleotides lie within the "gap" in late mRNA (6) and are not seen in oligonucleotide maps of polyadenylated RNA annealed to SV40 DNA fragments.

tween the SV40 DNA preceding or encoding the 5' end of VP1 and the sequences in SV40 DNA within Hind C, L, and M or within the adjacent regions. Homology or possibilities for strong base pairing between these regions of DNA or between the transcripts in these regions are not immediately evident from inspection of the sequences (1,6). Aloni (9) has previously presented evidence that 16S mRNA of SV40 infected cells may be formed in the cytoplasm by cleavage of a larger RNA. Further work is needed to characterize whether the RNA formed in the enucleated cells has the same initial sequence as the 16S RNA that has been synthesized in intact cells and whether this RNA is able to direct the synthesis of VP1. At any rate, the entire series of codons for VP1, including the initiator AUG, is represented in a continuous segment of SV40 DNA and translocation of RNA is not necessary to constitute an mRNA en-

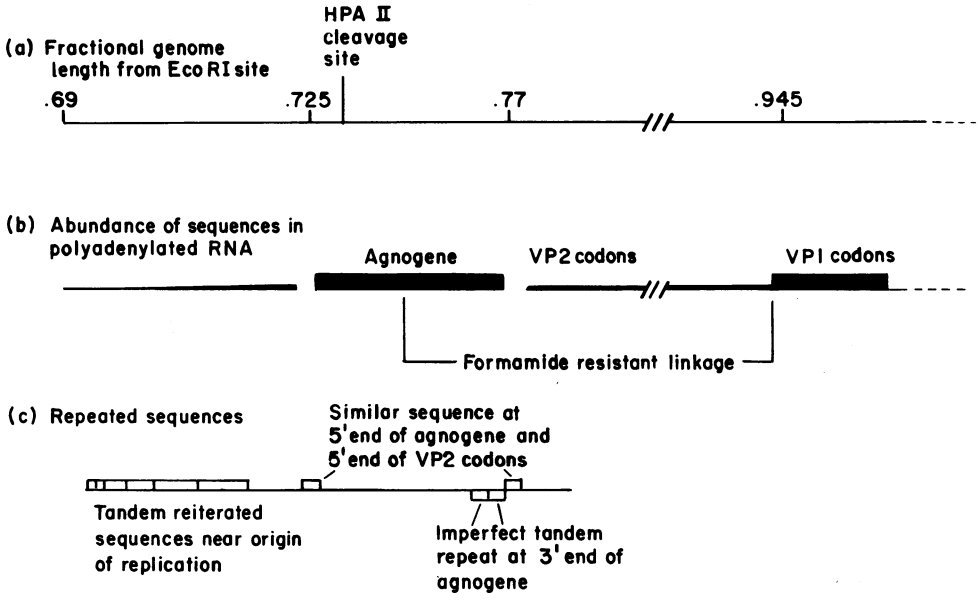


Fig. 6 Schematic representation of the location of the SV40 DNA codons for N terminal portions of the major structural protein VP1 and the minor structural protein VP2, as well as the segment of DNA whose transcipt is joined to that of the VP1 codons.

(a) represents the location of the segments of DNA in fractional genome length relative to the single Eco RI site in SV40 DNA. Hpa II refers to restriction endonuclease Hemophilus parahemolyticus II that makes a single cleavage site on SV40 DNA. This cleavage site lies within the "agnogene".

(b) represents the abundant sequences in polyadenylated RNA in the cytoplasm of cells infected with SV40. The thickness of the line corresponds to our visual estimate of the relative abundance of the oligonucleotides complementary to different segments of DNA.

(c) represents location of repeated sequences (Subramanian, et al., 9) present in the segment of the SV40 viral genome that spans Hind II, III-C and Hind II, III-D.

coding the protein.

The nucleotide sequence about the initial part of the RNA segment transcribed from Hind II,III-C attached to VP1 shows a marked resemblance to the nucleotide sequence about the initiation codon for the minor structural protein VP2 of SV40 (Fig. 4, 6). The resemblance in sequence raises the possibility that similar reactions may occur preceding the AUG of the "agnogene" (Fig. 6). Since two determinants of translation are thought to be the nucleotides shortly preceding the AUG in the message RNA and the presence or absence of capped structures, it seems possible that the run of 59 triplets derived from transcripts of Hind II,III-C, L, M and N and attached to the 16S RNA may itself be translated. If so, the translation product would be a peptide of under 7,000 molecular weight and the search for it would require somewhat different methods from those previously used in examining SV40 infected cells. Also, ifthis sequence is translated it is curious to note that its translation product would contain five arginines, several of which are coded for by CG codons rarely used elsewhere in SV40 mRNA.

This work was supported by grants from the American Cancer Society and the National Institutes of Health.

Present Address:

*University of Madrid, 144 Velasquez, Madrid 28, Spain

**National Institutes of Health, Bethesda, MD 20014, USA

REFERENCES

1 Pan, J., Reddy, V.B., and Weissman, S.M., accompanying manuscript.
2 Zain, B.S., Thimmappaya, B., and Weissman, S.M., Submitted for publication.
3 Dhar, R., Subramanian, K.N., Zain, S., Pan, J., and Weissman, S.M., (1974) Cold Spring Har. Symp. Quant. Biol. 39, 153-160.
4 Dhar, R., Zain, B.S., Weissman, S.M., Pan, J., and Subramanian, K.N. (1974) Proc. Natl. Acad. Sci. USA 71, 371-375.
5 Thimmappaya, B., Dhar, R., Zain, B.S., and Weissman, S.M. Submitted for publication.
6 Dhar, R., Subramanian, K.N., Pan, J., and Weissman, S.M. (1977) Proc. Natl. Acad. Sci. USA 74, 827-831.
7 Volckaert, G., Contreras, R., Soeda, E., Van de Voorde, A., and Fiers W. (1977) J. Mol. Biol. 110, 467-510.
8 Sharpe, P., in press, Proc. Natl. Acad. Sci. USA.
9 Aloni, Y., Shani, M., Reuveni, Y. (1975) Proc. Natl. Acad. Sci. USA 72, 2587-2591.
10 Subramanian, K.N., Reddy, V.B., and Weissman, S.M. (1977) Cell, in press.

ABBREVIATIONS

Restriction Endonucleases:

| | |
|---|---|
| Hpa II | Hemophilus parahemolyticus II |
| Eco RI | Enzymes from plasmid-bearing E. coli. |
| Hind II,III | Hemophilus influenzae strain d enzyme. |
| Alu I | Arthrobacter luteus enzyme. |
| Bam X | Bacillus amyloliquefaciens. |
| Mbo I | Moraxella bovis enzyme. |
| Pst | Providentia stuartii. |
| Hae III | H. aegyptius enzyme. |
| Hinf | H. influenzae strain f. |