Nucleotide Sequence of a Fragment of SV40 DNA That Contains the Origin of DNA Replication and Specifies the 5' Ends of “Early” and “Late” Viral RNA

IV. LOCALIZATION OF THE SV40 DNA COMPLEMENTARY TO THE 5' ENDS OF VIRAL mRNA

(Received for publication, December 29, 1975)

RAVI DHAR, KIRANUR SUBRAMANIAN, JULIAN PAN, AND SHERMAN M. WEISSMAN

From the Department of Human Genetics, Yale University School of Medicine, New Haven, Connecticut 06510

Cytoplasmic mRNA isolated from cells infected with SV40 was isolated by passage over oligo(dT)-cellulose columns. This RNA was annealed to SV40 DNA fragments produced by cleavage with EcoRII endonuclease. The RNA resistant to RNase digestion was analyzed by digestion with ribonucleases and oligonucleotide mapping. The results were compared with oligonucleotides from in vitro transcripts of the fragments and with whole genome SV40 cRNA which had been fractionated by hybridization to the fragments. The 5' ends of “early” and the large “late” SV40 mRNA, transcribed from opposite DNA strands, overlap for a region of 60 to 100 nucleotides. The region of overlap includes a portion of the segment of DNA containing the origin of DNA replication.

The expression of the genome of SV40 virus in productively infected cells is temporally regulated (3-5). Early after infection, cytoplasmic mRNA includes sequences complementary to a portion of one strand (the “early” strand) of the virus DNA. After initiation of viral DNA synthesis, a second class of cytoplasmic mRNA, complementary to the opposite (“late”) strand of SV40 DNA accumulates in the cytoplasm (6-8) and detectable synthesis of structural proteins begins (9, 10). The mechanism that prevents “late” gene expression until viral DNA synthesis has begun is an unsolved puzzle of general interest.

Detailed understanding of the control of SV40 virus gene expression has been difficult to achieve. No promoters have been defined by clear genetic analysis, and, unlike the situation with bacteriophage DNA, no one has demonstrated specific sites on the viral genome for initiation and termination of transcription with a purified mammalian DNA-dependent RNA polymerase. Short-lived intranuclear RNA species complementary to SV40 DNA have been found. Some of these RNAs are of whole genome length or larger and contain self-complementary sequences which may be annealed to form double-stranded RNA (11). Therefore, it is possible that much of the regulation of the information content of cytoplasmic mRNA may occur by specific cleavage, degradation, and transport of RNA species, rather than by regulation of the transcription process itself. Specific cleavage of rRNA, tRNA, and some mRNA is known to occur in prokaryotic systems (12-15). Similar data are lacking for animal cell mRNA precursors although there is abundant suggestive evidence that cleavage of precursor heterogeneous nuclear RNA species is a common path for the synthesis of cytoplasmic mRNA (16).

An additional difficulty in understanding SV40 gene expression is that the “early” proteins are not well known, and some of the peptides that are candidates for these early products (9, 10, 16-18) are rather long to be coded for by the segment of SV40 DNA often defined as the “early” region (19, 20).

We have approached these problems by analysis of the nucleotide sequence of the regions of SV40 DNA which we hope are of greatest biologic interest. As part of this project, we have compared the sequences of RNA transcribed from SV40 DNA I by Escherichia coli DNA dependent RNA polymerase with those of viral specific cytoplasmic mRNA of infected cells. To do this we have taken advantage of the availability of discrete fragments of SV40 DNA either incorporated into adeno-SV40 hybrid viruses (21) or generated by cleavage of intact SV40 DNA with restriction endonucleases. We have located the 3' and 5' ends of “late” mRNA (22-24) and presented the sequence including and spanning the 3' ends. In the present report we describe the location of the 5' ends of “early” and “late” mRNA more precisely. In the accompanying paper (Subramanian et al. (25)) the sequence about the 5' ends is described in detail. Evidence is presented that these terminal sequences are complementary to each other and span a portion of the segment of SV40 DNA where replication is initiated (26-28).
Materials and Methods

Pancreatic RNAase was obtained from Worthington Biochemicals, T1, and U1 RNAase from Sankyo Co., and oligodeoxythymidylic acid cellulose (polymer chain length about 10) from Collaborative Research Associates. Nucleotide triphosphates labeled in position α with 32P (specific activity 30 to 120 Ci/mmol) were obtained from New England Nuclear Corp. DNA-dependent Escherichia coli RNA polymerase holoenzyme was prepared by the method of Lewicki et al. (29).

Preparation of DNA - A small plaque isolate of strain 776 SV40 had been passed in the Vero or BSC-1 continuous lines of African Green Monkey kidney cells at a multiplicity of infection less than 0.02 plaque forming units/cell. To prepare DNA, subconfluent monolayers were infected at a multiplicity of 0.01 plaque forming units/cell. Four days later DNA was extracted from the infected cells by the Hirt procedure (30). Superncoiled SV40 DNA was isolated from the "Hirt" supernatant by centrifugation in a CsCl density gradient containing ethidium bromide. The ethidium bromide was removed from the DNA by extraction with isopropl alcohol saturated with CsCl after which the DNA was dialyzed against 0.01 M SSC (1× SSC: 0.15 M sodium chloride, 0.015 M sodium citrate, pH 7).

Preparation of Restriction Endonuclease Fragments - SV40 DNA (12) was digested with EcoRII restriction endonuclease prepared by the procedure of Yoshimori (31). The digestion products were separated by electrophoresis on 4% polyacrylamide slab gels (20×40 cm) and the positions of the bands identified by autoradiography. To elute DNA fragments from gel slices were positioned in 0.1× SSC and soaked for 4 to 5 h. The eluted fragments were then recovered by alcohol precipitation.

The fragment EcoRII-G was either used as such for hybridization against SV40 cDNA, or was further digested with the restriction endonuclease III. Mixtures of fragments were separated by electrophoresis on 4% acrylamide slab gels, eluted, and alcohol-precipitated as described above. Five fragments were obtained by cleavage of EcoRII-G with HaeIII: designated HaeIII EcoRII-G1, -G2, -G3, -G4, and -G5. The synthesis of RNA from intact SV40 DNA was done in mixtures containing 0.18 M KCl, 0.033 M Tris/HCl (pH 7.9), 3.5 mM MgCl2, 0.165 mM GTP, UTP, and ATP, 1 to 3 μg of DNA, and 5 μg of E. coli RNA polymerase holoenzyme in a total volume of 84 μl. One of the four triphosphates was radioactive with 32P in position α (specific activity of 50 to 120 Ci/mmol). Incubation was performed at 37° for 30 min. The transcription was terminated by incubation with 0.18 M of pancreatic DNase for 5 min at 37° followed by phenol extraction of the reaction mixture. The aqueous phase was passed over a Sephadex G-100 column (0.9×20 cm) (32). The transcript eluting with the void volume was precipitated with 2 volumes of alcohol, 0.1 volume of potassium acetate, pH 5.1, and 10 μg of carrier RNA.

Transcription of HaeIII EcoRII-G1, -G2, or -G3 was performed with 0.1 to 0.3 μg of DNA fragment and 3 to 4 μg of DNA-dependent E. coli RNA polymerase. Transcription was done with either native or heat-denatured fragments in the presence or absence of 0.09 M KCl. To transcribe heat-denatured fragments 0.2 to 0.3 μg of DNA was denatured at 97° for 3 min in a volume of 30 μl, chilled immediately, and added to the above reaction mixture in the presence or absence of KCl. The reaction was then allowed to proceed as described above.

Nucleic Acid Hybridization - Nucleic acid hybridizations were performed by the method of Gillespie and Spiegelman (34). DNA fragments (0.2 to 0.3 μg) in 2 ml of 0.01× SSC were denatured at 97° for 5 min. The solution was chilled rapidly and 3 ml of 10× SSC was added. Nitrocellulose filters (0.22 μm) (Millipore Corp.) were pre-soaked in 8× SSC, mounted on a filtration assembly and washed slowly with 10 ml of 6× SSC. The denatured DNA was passed slowly through the filter, washed slowly with 10 ml of 6× SSC and then with 100 ml of 6× SSC rapidly. Filters were then incubated at room temperature and baked in vacuum at 80° for 3 h. Hybridizations were performed in a volume of 0.75 ml of 2× SSC containing 0.1% sodium dodecyl sulfate for 8 h. The filters were rinsed with 2× SSC on both sides and washed with 50 ml of 2× SSC on each side.

Filters were then incubated at room temperature for 40 min with 2 ml of 3× SSC containing either 0.4 μg of pancreatic RNAase or 4 units of T1 RNAase. To 2 ml of the retained RNA was added RNAse digestion mixture, respectively. The filters were then rinsed and washed with 8× SSC as described above and incubated in 2 ml of 0.15 M sodium iodoacetate, 0.1 M sodium acetate in 2× SSC, pH 5.1 at 59° for 40 min. At the end of incubation the were rinsed and washed. The retained RNA was eluted in 1.5 ml of 0.01× SSC containing 100 μg of carrier DNA at 100° and then precipitated with 2 volumes of alcohol and 0.1 volume of potassium acetate, pH 5.1.

Preparation of SV40 mRNA from Infected Cells - Monolayers of the BSC-1, Vero, or CV-1 cells in 50-ml bottles containing approximately 0.5×10⁶ cells per bottle were infected with SV40 (strain 776) at a multiplicity of infection of 75 to 100 plaque forming units/cell. Fifteen to twenty hours later the cells were fed with phosphate-free Eagle's minimal medium containing 2% dialyzed fetal calf serum (35). Ten milliliters of H₃PO₄, was added to each bottle. Cells were harvested 49 to 48 h thereafter, suspended in 20 ml of layered 2× SSC containing 0.01 M Tris/HCl, pH 7.9, 0.01 M NaCl, 3 mM MgCl₂ and disrupted in a French homogenizer. Nuclei were removed by centrifugation at 1000 rpm in an International Centrifuge for 10 min and sodium dodecyl sulfate was added to the cytoplasmic fraction to a final concentration of 5%. Subsequent steps were performed at 0°.

The cytoplasmic fraction was extracted with an equal volume of phenol:chloroform (1:1) containing 1% isomyl alcohol, and the aqueous phase was re-extracted three times with phenol alone. The RNA was precipitated from the aqueous phase with 0.1 volume of 20% potassium acetate, pH 5.1, and 2 volumes of ethanol. The precipitated RNA was taken up in 10 ml of 0.5 M KCl, 0.01 M Tris/HCl, pH 7.5, and applied to an oligodeoxythymidylic cellulose column previously washed with this buffer. The column was then washed with 50 ml of this buffer after which the poly(A)-terminal material was eluted with 0.3 M NaCl, 0.02 M Tris/HCl, pH 7.5. Carrier RNA (100 μg) was added to the eluate and the RNA precipitated with 0.1 volume of potassium acetate (20%), pH 5.1 and 2 volumes of ethanol. This RNA was pelleted, redissolved in 0.75 ml of 2× SSC containing 0.1% sodium dodecyl sulfate and annealed at 67° for 8 to 10 h of the EcoRII-G fragment or one of its subfragments immobilized on a Millipore filter.

RNA Analyses - Radioactive RNA was digested with either T1 or pancreatic RNAase. The resulting oligonucleotides were fractionated by electrophoresis on Cellogel strips at pH 3.5 followed by "chromatography" on DEAE thin layer chromatography plates (36, 37). The individual oligonucleotides were further analyzed by digestion with pancreatic T1, or U1 RNAase and the resulting products fractionated by electrophoresis on DEAE-paper at pH 3.5 (36). Analyses were performed on oligonucleotides separately labeled with each of the four triphosphates. The details of the methods used in this work have been presented elsewhere (33, 37, 38).

Results

Transcription of Form I SV40 DNA - When 32P-labeled RNA, prepared by transcription of SV40 DNA, was annealed to EcoRII-G fragment DNA, eluted, digested with T1 RNAase, and mapped, a characteristic and reproducible set of prominent oligonucleotides was obtained (Fig. 1). On the basis of the known preference of Escherichia coli DNA-dependent RNA polymerase for transcription of the "early" (E) strand of SV40 DNA (38) and the sequence derived by partial digestion of the transcripts with T1 RNAase (Subramanian et al. (25)), it was possible to tabulate oligonucleotides derived from the transcript of the "early" strand of the DNA. In fingerprints of some preparations, "late" (L) oligonucleotides were seen, probably because transcription occurred symmetrically from nicked DNA. However, in these preparations oligonucleotides from the "early" strand were generally much darker than those from the "late" strand.

Transcription of HaeIII EcoRII-G Subfragments - Microgram amounts of each of the five fragments formed by cleavage of EcoRII-G with Hemophilus aegyptius restriction endonuclease III were prepared (Fig. 2). Each of the fragments HaeIII EcoRII-G1, -G2, and -G3 was used as a template for E. coli polymerase. HaeIII EcoRII-G2 was further cleaved by HindIll (I + III) into two fragments designated G2A and G2B (Fig. 2). Oligonucleotide maps of the T1 RNAase digests of the transcript of these three subfragments were reproducible with several preparations of enzymes and DNA (Figs. 3, 4, and 5). The oligonucleotides present in T1 RNAase digests of the tran-
scripts from HaeIII EcoRII-G - G2, and -G3 were identified as "E" or "L" strand oligonucleotides (Table I) by comparison with Fig. 1. The assignments were confirmed by analysis of the nucleotide sequence of the fragment.

When the DNA subfragments HaeIII EcoRII-G, -G2, and -G3 were denatured and transcribed in the absence of KCl, the relative yields of different oligonucleotides varied considerably. This was partly because transcription did not always begin at the 3' end of the DNA strand. Nevertheless, all products in "early" strand RNA complementary to EcoRII-G were detected in the transcripts of the subfragments. The transcripts of these subfragments have been subjected to complete nucleic acid sequence analysis, the details of which are given in the accompanying report.

Cytoplasmic mRNA Annealed to EcoRII-G Fragment - Radioactive 32P-labeled cytoplasmic RNA was prepared from cells harvested 40 to 50 h after infection with SV40. RNA was passed through oligo(dT)-cellulose columns to isolate the polyadenylic acid-linked molecules and then through a Sephadex G-100 column so as to remove RNA smaller than about 150 nucleotides. Some preparations were further centrifuged through 5 to 20% sucrose gradients to remove RNA smaller than about 12 S (29). The high molecular weight RNA was annealed to the EcoRII-G fragment. Two-dimensional maps of the T and pancreatic RNase digests of this RNA were prepared and each oligonucleotide was further analyzed by digestion with pancreatic RNase. A total of nine in vivo RNA preparations complementary to EcoRII-G were examined and a consistent pattern emerged (Fig. 6). A relatively low yield of SV40 mRNA was obtained in these annealing experiments compared with the yield in analogous experiments with the principal species of "late" SV40 mRNA (RNA complementary to Hind(II + III)-D, -E, -F, -K, -J, and -G).

"Late" Strand Sequences of Cytoplasmic mRNA - Some, but not all, of the T and pancreatic RNase products expected from the transcript of the "late" strand of EcoRII-G were present in the oligonucleotide maps of polyadenylic acid-linked cytoplasmic mRNA isolated from cells late in the infectious cycle. Analysis of the two fragments obtained by cleavage of EcoRII-G with

---

Fig. 1. A, autoradiograph of a two-dimensional separation of oligonucleotides produced by T-RNase digestion of SV40 complementary RNA annealed to EcoRII-G. Electrophoresis was from left to right on Cellugel in 7 M urea at pH 3.5 and chromatography from bottom upwards on DEAE-cellulose thin layer plates using the "Homo II" solution (37). The RNA was synthesized on SV40 DNA I template using Escherichia coli RNA polymerase in the presence of [α-32P]GTP (specific activity 80 Ci/mmol) for 30 min at 37° and then annealed to EcoRII-G as described under "Materials and Methods." B, schematic sketch of autoradiograph, with numbers assigned to the oligonucleotides derived from the RNA transcript of the "early" strand of the DNA fragment EcoRII-G. Paint oligonucleotides derived from the "late" strand transcript or SV40 sequences outside EcoRII-G are not numbered.

---

Fig. 2. A, schematic diagram of the SV40 genome. The entire SV40 chromosome, opened at the EcoRI cleavage site. The origin of DNA replication and the 5' ends of the mRNAs are located as described in the text. The capital letters refer to the Hemophilus influenzae d(II + III) restriction endonuclease digestion products of the DNA. The decimal numbers refer to the fractional genome length of cleavage sites from the EcoRI cut. → indicates the 5' end of mRNA; ↑ indicates the 3' end of mRNA; † indicates our estimate of the DNA segment corresponding to a particular function. B, an expanded diagram of the region of SV40 DNA complementary to the 5' end of "early" and "late" mRNA. EcoRI, HindIII, HaeIII refer to the cleavage sites for the restriction endonucleases from Escherichia coli with the RI1 plasmid, from H. influenzae strain d, and from H. negeitius, respectively. G1, G2, G3, G4, and G5 are the segments obtained by further cleavage of EcoRII-G with HaeIII. The subfragments G2A and G2B are obtained by cleavage of HaeIII EcoRII-G2 with HindIII restriction endonuclease. The numbers below the horizontal line indicate the number of base pairs in each segment of DNA. The HindIII cleavage site separates the DNA fragments Hind-A and Hind-C. EcoRII and EcoRII-N are the EcoRI fragments adjacent to EcoRII-G.
TABLE I

<table>
<thead>
<tr>
<th>T1 Oligonucleotide No.</th>
<th>Pancreatic RNase digestion products of T1 oligonucleotide(α-32P labeled precursors)</th>
<th>Pancreatic RNase digestion products of T1 oligonucleotides from cytoplasmic RNA annealed to Eco RI-G</th>
<th>Sequence deduced</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CTP</td>
<td>ATTP</td>
<td>GTP</td>
</tr>
<tr>
<td>1</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>17</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>21</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>23</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>26</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>29</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>30</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>101</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>102</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>103</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>105</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>106</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>107</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>109</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>114</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>118</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>120</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>121</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Analysis of T RNase digestion products of RNA complementary to the SV40 DNA fragment EcoRII-G.

a.) Presence (+) or absence (−) of oligonucleotide in cytoplasmic mRNA (see Figs. 1, 3, 4, 5, 6, 7).
b.) Presence (+) or absence (−) of the T oligonucleotide in E. coli RNA polymerase transsects of the "early" strand of EcoRII-G.
c.) Presence (+) or absence (−) of the T oligonucleotides in transcripts of the "late" strand of EcoRII-G (see Figs. 3, 4, and 5).
d., e., f.) Presence (+) of the T oligonucleotides in transcripts of HaeIII-EcoRII-G, G-2, and G-3 respectively, regardless of strandedness.

HindIII restriction enzyme followed by transcription of the two fragments with E. coli RNA polymerase showed that some "late" strand oligonucleotides in SV40 mRNA were derived from that portion of EcoRII-G near HindIII-A. The oligonucleotide CAAAG was consistently present. The characteristic oligonucleotide T16 (UUUAAACUUAUCCAUC-3') was not indicated between nucleotides or at the 5' end of a nucleotide, either in the text or in the tables.
UUG\textsuperscript{2} (Figs. 4 and 6 and Table I) offered some difficulty. A product was found that released AAAC, AU, G, U, and G on pancreatic digestion, but its mobility on homochromatography was slightly high for intact T\textsubscript{106}. The oligonucleotide linked to the 5' end of T\textsubscript{106} in the "late" strand transcript within Hind-A is T\textsubscript{106} (AUUCCUCUCUG) (Figs. 4 and 6 and Table I). This oligonucleotide T\textsubscript{106} would be superimposed on T\textsubscript{1a} (Fig. 4 and Table I) (UUUUAACAG), a product of the "early" strand transcript of \textit{HaeIII EcoRII-G2A}. Analysis of the product present in this position in maps of cytoplasmic mRNA were consistent with the sequence of T\textsubscript{106}, since it showed U, AAAC, and AG on pancreatic RNase digestion. The dinucleotide AU, a product from T\textsubscript{106}, was not found, although the low levels of radioactivity meant that it might be there in smaller amounts than the other products. The next two characteristic products of the "late" strand transcript in Hind-A (\textit{HaeIII EcoRII-G2A})
would be T111 (UCCAUUGA) and T101 (CACUCCUUCUAAGA) (Fig. 4 and Table I). Neither T101 nor T101 was detected in cytoplasmic mRNA. The T, RNase products T101 and T101 are not visible consistently in the fingerprints of the transcripts of the EcoRII-G2 subfragment (Fig. 4) as they lie very close to the end of the fragment and the in vitro transcription appears to start internally on this fragment. Most or all of the oligonucleotides present in the in vitro transcript of the HindII + III-C fragment of HaeIII EcoRII-G2B were detected in the cytoplasmic mRNA, i.e., T100 (CUUCUUGG) and T101 (CAAAAG) (Figs. 4 and 6 and Table I). Oligonucleotide T101 was often faint but pancreatic RNase digestion of cytoplasmic RNA complementary to EcoRII-G showed AAAAGC which could only have come from this oligonucleotide or sequences outside EcoRII-G. The T, oligonucleotides of the "late" strand transcript of HaeIII EcoRII-G3, T100 (CCUCACACCUUG), and T101 (CUUCUCACUUCUUG) (Figs. 5 and 6 and Table I) appear to be present in the mRNA. The product T111 (CUCAG) was not detected, but this oligonucleotide was in the crowded region of the chromatogram and could have been obscured by other products of similar composition. All the large T, RNase products of the "late" strand transcript of HaeIII EcoRII-G1 that were detected in in vitro transcripts except for T111 were also present in cytoplasmic mRNA. The most characteristic of these were T111, T110, T101, and T121 (Figs. 3 and 6 and Table I). Oligonucleotide T111 would have been lost in the radioactivity remaining at the bottom of the homochromatography plate (presumably polyadenylic acid and other unidentified material). However, the product AAAAAAAAU was present in pancreatic RNase digests of cytoplasmic RNA and could have come from T111. These results indicate that detectable "late" strand transcript extended on the EcoRII-G fragment no more (and probably less) than 45 nucleotides into HindII + III-A and are consistent with the known polarity of transcription (19, 20, 22).

"Early" Strand Sequences of Cytoplasmic mRNA—Some of the "early" strand products that have been identified by transcription of SV40 DNA I by E. coli RNA polymerase were detected in maps of the in vivo mRNA annealed to EcoRII-G (Fig. 6). The most characteristic "early" strand products of HaeIII EcoRII-G2A are T133, T134, T135, and T136 (Figs. 1, 4, and 6 and Table I). The most characteristic "early" strand transcripts within HaeIII EcoRII-G2B (Fig. 2) are T138 and T139. These were present in maps of the in vivo mRNA annealed to EcoRII-G.

The "early" strand T, products of HaeIII EcoRII-G3 are T126 (CUUUCUAUG), T125 (CUAUCCCAAG), and T124 (CCUCUG), which were detected in in vivo mRNA. However, T124 is rather short, the products of pancreatic RNase digestion of this oligonucleotide are mononucleotides and there was insufficient material to determine the order of the pyrimidines in the in vivo fragment of the mRNA. The fragment was denatured before transcription with [a-32P]UTP as one of the labeled precursors, B, schematic sketch of the autoradiograph with numbers assigned to the oligonucleotides listed in Table I. FIG. 7. Autoradiograph of a two-dimensional separation of oligonucleotides produced by digestion with T, RNase of SV40 complementary cytoplasmic poly(A)-linked RNA prepared early after infection in the presence of cytosine arabinoside and annealed to EcoRII-G. African Green Monkey kidney cells (BSC-1) were infected at a multiplicity of 100 p.f.u./cell. Four Blake bottles were used. Cells were grown for 8 h after infection in Earl's Minimal Essential Medium containing 30 μg/ml of cytosine arabinoside to block DNA replication and then put for 3 h in phosphate-free medium containing cytosine arabinoside and 20 μC of 35P. The cells were harvested 16 h after infection. Cytoplasmic RNA was passed over a Sephadex G-100 column (1 x 100 cm) and two peaks of radioactive material were observed. The pool in the excluded volume was passed over oligo(dT)-cellulose column to isolate the poly(A)-linked mRNA which was then precipitated with alcohol and annealed to EcoRII-G. The fractionation of the oligonucleotides produced by T, RNase digestion of this RNA was performed as in Fig. 1. The fingerprint shows some oligoadenylic acid products from the poly(A) bound to the oligo(dT)-cellulose filters. B, schematic sketch of the autoradiograph. The numbers refer to oligonucleotides listed in Table I. The circles with × are some faint T, oligonucleotides which have not been identified.

FIG. 8. Autoradiograph of a two-dimensional separation of oligonucleotides produced by digestion with T, RNase of SV40 mRNA prepared "late" after infection and annealed to EcoRII-I fragment. The preparation of the SV40 mRNA was the same as in Fig. 6. B, schematic sketch of the autoradiograph. FIG. 9. Autoradiograph of a two-dimensional fractionation of oligonucleotides produced by T, RNase digestion of SV40 cRNA which had been annealed to EcoRII-I. The transcript was prepared and analyzed as in Fig. 1, with [α-32P]ATP as one of the labeled precursors. B, schematic sketch of the autoradiograph. FIG. 10. Autoradiograph of a two-dimensional separation of oligonucleotides produced by T, RNase digestion of transcripts of the EcoRII-I fragment. The fragment was denatured before transcription as described under "Materials and Methods," and [α-32P]GTP was used as a labeled precursor. B, schematic sketch of the autoradiograph.
vivo RNA, so that it was uncertain whether or not the RNA extends through this sequence. None of the "early" strand T, products of HaeIII EcoRII-G1 was detected in the SV40 mRNA annealed to EcoRII-G (compare Figs. 1, 3, and 6 and Table I). The "early" strand oligonucleotide T, (CCCCUAACUCCG) which lies within HaeIII EcoRII-G1 but very close to HaeIII Eco-G3 is absent in in vivo RNA (Fig. 6) and all the "early" strand T, oligonucleotides preceding T, (Subramanian et al. (25)) are also absent. These results show that the "early" mRNA appears to extend 60 but not more than 90 nucleotides into HindII (I + III) -C fragment.

The in vivo mRNA annealed to EcoRII-G (Fig. 6), prepared "late" after infection, contained some large T, oligonucleotides which could not be identified as coming from either of the two strands. Products T, (identical to spot T, of Figs. 1 and 6) and T, (Fig. 6, Table II) were identified as coming from the "early" strand of EcoRII-I, the region of SV40 DNA adjacent to EcoRII-G (Fig. 2). The other T, oligonucleotides T, T, T, T, T, and T, have not been identified yet. Most of the in vivo T, fingerprints have picked up some background and some material which remains at the bottom of the homochromatography plate whose structure is not known.

It was possible that the 5' end of mRNA made before DNA replication would differ from the 5' end of "early" strand mRNA made after the "late" strand message was transcribed. To investigate this we prepared cytoplasmic mRNA from cells infected with SV40 in the presence of cytosine arabinoside so as to block DNA replication and harvested the cells 8 h after infection. The details of the isolation of cytoplasmic mRNA are given under "Materials and Methods." This RNA was annealed to EcoRII-G, eluted, and mapped (Fig. 7). The resulting T, RNase fingerprint showed exactly the same "early" strand products as were detected "late" in the course of infection in the absence of cytosine arabinoside. Comparison of data in Figs. 1, 5, and 7 and Table I shows that the 5' end of the "early" strand transcript lies very close to the HaeIII EcoRII-G1, -G4 junction.

The sequence of the HaeIII EcoRII-G4 "early" strand transcript (Subramanian et al. (25)) (CCGCCUCCCG) is difficult to detect in the in vivo experiments. It appears that (CCUCG) is absent in the "early" RNA but we cannot exclude the possibility that it is present in relatively low levels, because of non-uniformity of the 5' end of the RNA after hybridization. The sequence of HaeIII EcoRII-G5 "early" strand is (CCGAGG) (Subramanian et al. (25)). There were several sources for G and AG, but CCG was not visible in the SV40 mRNA obtained early after infection. These results indicate the 5' end of the "early" RNA extends approximately to the end of HaeIII EcoRII-G5 (G4, G5) junction.

Comparison of in Vivo mRNA Annealed to EcoRII-I Fragment with in Vitro Transcripts — To confirm the result that "late" mRNA extended not more than 45 nucleotides into the Hind-A fragment, we annealed cytoplasmic mRNA prepared "late" after infection or E. coli RNA polymerase transcribed of SV40 DNA to EcoRII-I DNA fragment on filter and prepared T, oligonucleotide maps of the protected RNA (Figs. 8 and 9, respectively). Also, the EcoRII-I fragment was denatured and transcribed to give T, oligonucleotides (Fig. 10) from both the strands. The T, products were further digested with pancreatic RNase and their primary products were compared (Table II). All the "early" strand T, oligonucleotides present in Fig. 9 were identified in the in vivo RNA annealed to EcoRII-I (Fig. 8). The in vivo mRNA annealed to EcoRII-I fragment was not clean, partly because the EcoRII-I isolated directly from SV40 is relatively impure. However, we have determined the entire nucleotide sequence of EcoRII-I and could distinguish a number of unique T, RNase products from the transcript of the "early" strand of this DNA fragment. The most characteristic "early" strand sequences of EcoRII-I (Figs. 8, 9, and 10 and Table II) are T, T, T, T, T, T, T, T, T, and T. All these were detected in the in vivo mRNA annealed to EcoRII-I fragment. None of the "late" strand sequences (Fig. 10) obtained by transcription of EcoRII-I was detected in vivo fingerprints. The most characteristic of these were T, T, T, T, T, T, T, T, T, and T. Oligonucleotides T, T, T, and T come from the transcript of the 80 nucleotides of EcoRII-I closest to EcoRII-G. These should have been seen readily if they were present in amounts approaching the abundance of products of "early" strand transcripts. These results support the conclusion that the detectable 5' end of "late" mRNA does not extend into EcoRII-I.

DISCUSSION

Our earlier results established that the 3' terminus of abundant "late" SV40 cytoplasmic mRNA lies almost exactly at the junction of the HindII (II + III) restriction fragments G and B (22-24). There is a large excess of "late" strand transcript from Hind-G over the amount of "early" strand transcript from Hind-D late in the infectious cycle and it is possible that overlapping "early" strand transcripts extending into Hind-G were obscured by the competing RNA-RNA hybridization reaction occurring in solution. With this reservation however we have also put the 3' end of "early" strand transcripts in HindII + III-G within 100 nucleotides of the junction between HindII (II + III)-G and -B, but in the present experiments the amounts of "early" strand transcript and "late" strand transcript about the junction of the HindII (II + III)-A and -C fragments were approximately equal as judged by the radioactivity recovered in individual oligonucleotides transcribed from either strand. This provided a favorable experimental situation to demonstrate the transcripts. In fact, RNA-RNA hybridization did not appear to decrease substantially the recovery of either strand of the RNA, since oligonucleotides from "late" strand RNA of EcoRII-G1 region where there was no complementary "early" strand transcribed sequences were not substantially more prominent than "late" strand products from EcoRII-G2 and -G3. We had previously concluded that the transition from "early" to "late" strand transcript lay near the HindII (II + III)-A and -C junction (23, 24).

The present results demonstrate that the 5' end of "early" transcript extends between 60 and 80 nucleotides into HindII (II + III)-C while the 5' end of "late" transcripts extends to near the Hind-A-C junction but not more than 45 nucleotides into HindII (II + III)-A (Fig. 2). One possibility to be considered is that the "late" RNA has leaked out from the nuclei and is not actually a cytoplasmic component. However, we have found anti- "early" strand transcripts of the HindII (II + III)-B segment of SV40 in the nucleus but not in the cytoplasm of infected cells. Therefore, if nuclear leakage of "late" strand RNA occurred it would have to be selective for some but not all nuclear RNA and the species of RNA that did leak would preferably include polyadenylated RNA. Difficulties in assigning precise locations for the 5' ends are the low yields of SV40 mRNA from this region of the genome and the possibility that modified bases or sugars, or other unusual structures at the 5' ends of the mRNA, might cause either shifts in mobility or streaking that would interfere with recognition of the precise termini. No specific search was conducted for modi-
the function of the protein coded for by this region is not
including portions proximal to Hind(I1 + 111)-D (39). Either
for the "late" proteins lie beyond this region. Replication of
SV40 DNA proceeds bidirectionally from what is apparently a
needed for replication in cultured cells or the initiation codons
about one-third of the Hind(I1 + 111)-C region of the genome,
sors have been polyadenylated and entered the cytoplasm. It is
remarkable in this regard that viable deletion mutants of
SV40 are known that replicate without helper and lack up to
5'-terminal sequences removed even after precur-
"late" and "early" RNA could extend 5 to 12 nucleotides fur-
therefore, if specific DNA sequences are required to initiate
transcription of RNA does or does not share a common initia-
tion mechanism with the initiation of DNA replication. For
example, if specific DNA sequences are required to initiate
replication, it might be difficult to locate these sequences
within translated portions of mRNA. Lai and Nathans (41)
have isolated a deletion mutant of SV40 that replicates in the
presence of helper virus thought to supply some "early" func-
tion, and that appears to lack the entire Hind(I1 + 111)-A
junction (26, 28, 40). We previously
reported that there are symmetric sequences in this region of
DNA replication begins. Another, not necessarily alternative,
necessity for symmetry would be a requirement for specific se-
quencies at the 5' ends of both "early" and "late" RNA, together
with the demonstrated overlap of these transcripts. While it is
curious that the origin of DNA replication lies near the 5' end
of the "early" mRNA, there is no direct evidence as to whether
transcription of RNA does or does not share a common initia-
tion mechanism with the initiation of DNA replication. For
example, if specific DNA sequences are required to initiate
replication, it might be difficult to locate these sequences
within translated portions of mRNA. Lai and Nathans (41)
have isolated a deletion mutant of SV40 that replicates in the
presence of helper virus thought to supply some "early" func-
tion, and that appears to lack the entire Hind(I1 + 111)-A
A. coli RNA polymerase transcripts of the
plasmic mRNA (Figs. 8, 10).
and cause apparent disappearance of this terminal T, RNase
digestion products derived from the "late"
strand transcript are marked (C).
and "late" transcripts (Fig. 2). If the 5'-terminal one or two nucleotides
of the transcript were guanylic acid and were methylated
at the 2'-hydroxyl position, this would change the composition
and cause apparent disappearance of this terminal T, RNase
product. This would mean that the 5' ends of the transcript
of "late" and "early" RNA could extend 5 to 12 nucleotides
further from the 3' ends.
Possibly the translated forms of "early" and "late" mRNA
may have 5'-terminal sequences removed even after precursors
have been polyadenylated and entered the cytoplasm. It is
remarkable in this regard that viable deletion mutants of
SV40 are known that replicate without helper and lack up to
about one-third of the Hind(II + III)-C region of the genome,
including portions proximal to Hind(II + III)-D (39). Either
the function of the protein coded for by this region is not
needed for replication in cultured cells or the initiation codons
for the "late" proteins lie beyond this region. Replication of
SV40 DNA proceeds bidirectionally from what is apparently a
unique origin that lies in the Hind(II + III)-C fragment close
to the Hind(II + III)-A junction (26, 28, 40). We previously
reported that there are symmetric sequences in this region of
the DNA (23, 24). One possible justification for symmetry
would be that these are the precise sites where bidirectional
DNA replication begins. Another, not necessarily alternative,
necessity for symmetry would be a requirement for specific se-
quencies at the 5' ends of both "early" and "late" RNA, together
with the demonstrated overlap of these transcripts. While it is
curious that the origin of DNA replication lies near the 5' end
of the "early" mRNA, there is no direct evidence as to whether
transcription of RNA does or does not share a common initia-
tion mechanism with the initiation of DNA replication. For
example, if specific DNA sequences are required to initiate
replication, it might be difficult to locate these sequences
within translated portions of mRNA. Lai and Nathans (41)
A. coli RNA polymerase transcripts of the "early" strand of
EcoR II-1 (Fig. 3).
and "late" strand transcript are marked (D).
Pancreatic RNase digestion products derived from the "early"
strand transcript are marked (C).
Pancreatic RNase digestion products derived from the "late"
strand transcript are marked (D).
Pancreatic RNase digestion products of T1 oligonucleo-
tides from cytoplasmic mRNA annealed to Eco RII-1.
Pancreatic RNase digestion products of T1 oligonucleo-
tides (T2 labeled precursors)

<table>
<thead>
<tr>
<th>T1 Oligonucleotide</th>
<th>Pancreatic RNase digestion products of T1 oligonucleotides (T2 labeled precursors)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>a</td>
</tr>
<tr>
<td>152</td>
<td>+</td>
</tr>
<tr>
<td>155</td>
<td>+</td>
</tr>
<tr>
<td>160</td>
<td>+</td>
</tr>
<tr>
<td>161</td>
<td>+</td>
</tr>
<tr>
<td>162</td>
<td>+</td>
</tr>
<tr>
<td>163</td>
<td>+</td>
</tr>
<tr>
<td>164</td>
<td>+</td>
</tr>
<tr>
<td>165</td>
<td>+</td>
</tr>
<tr>
<td>166</td>
<td>+</td>
</tr>
<tr>
<td>167</td>
<td>+</td>
</tr>
<tr>
<td>168</td>
<td>+</td>
</tr>
<tr>
<td>200</td>
<td>-</td>
</tr>
<tr>
<td>201</td>
<td>-</td>
</tr>
<tr>
<td>202</td>
<td>-</td>
</tr>
<tr>
<td>203</td>
<td>-</td>
</tr>
<tr>
<td>204</td>
<td>-</td>
</tr>
<tr>
<td>205</td>
<td>-</td>
</tr>
<tr>
<td>206</td>
<td>-</td>
</tr>
<tr>
<td>207</td>
<td>-</td>
</tr>
<tr>
<td>208</td>
<td>-</td>
</tr>
</tbody>
</table>

Analysis of T1 RNase digestion products of RNA complementary
to the SV40 DNA fragment EcoR II-1.

a.) Presence (+) or absence (-) of oligonucleotides in cyto-
plasmic mRNA (Figs. 8, 10).
b.) Presence (+) or absence (-) of the oligonucleotides in
E. coli RNA polymerase transcripts of the "early" strand of
EcoR II-1 (Fig. 9).
c.) T1 RNase digestion products derived from the "early"
strand transcript are marked (C).
d.) T1 RNase digestion products derived from the "late"
strand transcript are marked (D).
transcription of "late" mRNA. Ketner and Campbell (42) in their studies of the biotin operon of E. coli have suggested that the initiation of bidirectional DNA replication may require divergent transcription of the replication origin. This model would be consistent with the SV40 replicon where the overlapping divergent transcription does in fact include the most probable origin of DNA replication.

Khoury et al. (43) have re-evaluated earlier experiments localizing the 5’ and 3’ ends of SV40 "early" and "late" RNA. Their present results agree within experimental error with our previous localization of the 3’ ends of the RNA (6, 22) and are not inconsistent with our localization of the 5’ ends of the largest "early" and "late" mRNA.

Acknowledgments - We are grateful for the excellent technical assistance of Cheryl Christner, Adeline Tucker, Katherine Minahan, Richard Milardo, Alexandra Breslin, Nancy Bartol, and Sharlene Ivory. The authors express particular thanks to Dr. Paul Lebowitz for his careful reading of the manuscript and helpful comments.

Note Added in Proof - At least a part of the late viral RNA complementary to Hind(II + III)-C may not be covalently bound to the rest of the late RNA.

REFERENCES

Nucleotide sequence of a fragment of SV40 DNA that contains the origin of DNA replication and specifies the 5' ends of "early" and "late" viral RNA. IV. Localization of the SV40 DNA complementary to the 5' ends of viral mRNA. R Dhar, K N Subramanian, J Pan and S M Weissman


Access the most updated version of this article at http://www.jbc.org/content/252/1/368

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/252/1/368.full.html#ref-list-1