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

I. MAPPING OF THE RESTRICTION ENDONUCLEASE SITES WITHIN THE EcoRII-G FRAGMENT AND STRATEGY EMPLOYED FOR ITS SEQUENCE ANALYSIS

(Received for publication, December 29, 1975)

KIRANUR N. SUBRAMANIAN,† RAVI DHAR,§ AND SHERMAN M. WEISSMAN§

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

A fragment (EcoRII-G) of SV40 DNA 316 nucleotides long that contains the origin of SV40 DNA replication and the DNA complementary to the 5’ ends of some “early” and “late” cytoplasmic RNA species was isolated. A cleavage map of the fragment was constructed with four restriction endonucleases, and an approach for the analysis of the nucleotide sequence of the fragment was presented.

Simian vacuolating virus 40 (SV40) is a DNA tumor virus containing a small, circular DNA genome. Because of its small size, the viral DNA contains information to code for only a few proteins. The virus depends almost entirely on the host machinery for its propagation. The relatively small size of the SV40 DNA makes it a convenient model system to study the relationship between structure and function within an eukaryotic environment. Our laboratory has been primarily engaged in attacking this problem using nucleotide sequencing as the tool with the idea of interpreting the activity of a certain segment of the viral DNA as a function of its primary structure, i.e. its nucleotide sequence. We have been interested in elucidating the nucleotide sequences at certain biologically important sites on the SV40 DNA such as the origin of viral DNA replication and the sites of initiation and termination of “early” and “late” viral RNA transcription.

Precise locations of these and other sites on the viral DNA require reference points all around the DNA; these are best obtained by making use of type II restriction endonucleases which make sequence-specific double-stranded breaks in DNA. A cleavage map of type II restriction endonucleases by Nathans and his colleagues (3). Subsequently, more cleavage maps of SV40 DNA have been drawn up by employing other restriction endonucleases such as EcoRII (4), HaeIII (4, 5), HhaI, Hinfl, ² and AluI.²

SV40 DNA replication is known to start from a unique point on the DNA, proceed bidirectionally, and end at a point exactly half-way down the circle (6, 7). The origin of DNA replication has been located by pulse-labeling studies (8) and electron microscopic analysis (9) at 0.67 map unit on the SV40 cleavage map, i.e. within Hind(II + III) Fragment C very near the junction between the -C and -A fragments. Transcriptional maps prepared by means of hybridization work (10) located the 5’ ends of “early” and “late” viral RNAs within Hind(II + III) Fragment A. More precise mapping performed in our laboratory based on nucleotide sequencing (11, 12) has shown that the 5’ ends of “early” and “late” viral RNAs are located very near the junction between Hind(II + III)-A and -C fragments proximal to the origin of DNA replication.

Previous reports from our laboratory (11-13) have presented the nucleotide sequence of a segment of SV40 DNA containing the 3’ ends of “early” and “late” viral RNAs. In this report we present the complete nucleotide sequence of another fragment of SV40 DNA which contains the origin of DNA replication and specifies the 5’ ends of “late” viral mRNAs. This fragment, EcoRII-G, obtained by digesting SV40 DNA with the EcoRII restriction endonuclease, is a little more than 300 nucleotides long and maps between 0.64 and 0.70 map unit (4). EcoRII-G


† W. Fiers, personal communication.
Nucleotide Sequence of SV40 DNA

has been sequenced by a combination of RNA- and DNA-sequencing methods. EcoRII-G contains several interesting symmetrical, palindromic, and repeating sequences which might serve as signals for the initiation of viral DNA replication and transcription. The derivation of the cleavage map of EcoRII-G is presented in this report and the derivation of the nucleotide sequence of EcoRII-G is reported in the two accompanying manuscripts (14, 15).

MATERIALS AND METHODS

Strains – SV40 strains 776 (small plaque) (kindly provided by Dr. K. Danza of the University of Colorado) was used. BSC-1 African green monkey kidney cells were used for the production and propagation of SV40.

Radioactive Materials – α-32P-labeled ribonucleoside triphosphates of specific radioactivity 50 to 120 Ci/mmol and carrier-free 32P-labeled orthophosphoric acid were purchased from New England Nuclear Corp. γ-32P-labeled ATP of high specific radioactivity (1,000 Ci/mmol) was prepared by a modification of the method of Glynn and Chappell (16).

Enzymes – The following restriction endonucleases were used as described in the literature cited within parentheses: HindI + III (17); EcoRII (18); HaeIII (19). AluI and HinfI were kindly donated by Doctors R. Roberts and B. S. Zain of the Cold Spring Harbor Laboratory, respectively. Escherichia coli DNA-dependent RNA polymerase holoenzyme was purified according to the method of Lebowitz et al. (20) and on many occasions was generously provided by Dr. P. Lebowitz. The following enzymes were obtained from commercial sources: HindIII restriction endonuclease (Biola); T1 and T2 ribonucleases (Sankyo Co.); pancreatic RNase (Calbiochem); bacteriophage T4-induced polynucleotide kinase (P-L Biochemicals); bacterial alkaline phosphatase and snake venom phosphodiesterase (Boehringer-Mannheim Corp.).

Preparation of EcoRII-G Fragment of SV40 DNA and Its Subfragments – Form I SV40 DNA, prepared as described elsewhere (3), was digested with the EcoRII restriction endonuclease (usually 200 µg of the DNA was digested with 16 h and 200 µl of the enzyme) in a reaction mixture containing 0.09 M Tris/HCl, pH 7.5, 0.01 M MgCl2 and the products separated by electrophoresis for 2 h at 150 V on slab gels (10 x 40 cm) containing 4% polyacrylamide (4). The fragments were identified by autoradiography, excised from the gel with a scalpel, extracted by homogenization of the gel bands in a solution containing 0.015 M NaCl and 0.0015 M sodium citrate (pH 1), and recovered by ethanol precipitation.

One of the EcoRII fragments, EcoRIII-G, was recleaved by digestion with HaeIII, HindIII, AluI, or HinfI restriction endonucleases in reaction mixtures containing 6 µM Tris/HCl, pH 7.5, 10 mM MgCl2, 6 mM 2-mercaptoethanol, and 50 mM NaCl, and the products separated by electrophoresis on composite polyacrylamide slab gels (consisting of 10% gel in the bottom one-third and 5% gel in the top two-thirds of the slab) and extracted as described above. DNA fragments were usually stored as dry precipitates at −20°C.

RESULTS AND DISCUSSION

Choice of DNA Fragment for Sequence Analysis of Origin of DNA Replication in SV40 – The origin of DNA replication has been located at approximately 0.67 map unit in the physical map of SV40 DNA by two independent methods (8, 9). We had shown earlier that SV40 DNA is cleaved into 10 fragments by the EcoRII restriction endonuclease; these fragments have been mapped and ordered (4). One of these fragments, EcoRII-G, was chosen for sequence analysis in this study because its location between 0.64 and 0.70 map unit indicated that it would contain the replication origin. We had also shown (4) that EcoRII-G was cleaved into smaller subfragments by HindII + III and HaeIII restriction endonucleases. The first step in the sequence analysis of EcoRII-G was the production of subfragments by means of these and other restriction endonucleases. During the sequence analysis of subfragments of EcoRII-G, sequences recognized and cleaved by certain restriction endonucleases were noted (and predicted in two instances described below) within subfragments produced by another restriction endonuclease. These were confirmed later by restriction endonuclease cleavage at these locations.

Mapping of HindIII and HinfI Cleavage Sites within EcoRII-G – EcoRII-G has been shown to overlap HindII + III Fragments A and C (4); the restriction cleavage site at this junction has been found to be a HindIII site (21). HindII cleavage of EcoRII-G resulted in the production of two subfragments of approximate length 230 and 80 base pairs which were also produced by the cleavage of HindII + III and A fragments, respectively, with the EcoRII restriction endonuclease (4). The HindIII site within EcoRII-G was thus localized at about 80 nucleotides from the left-hand end of EcoRII-G (4). The 230- and 80-long HindIII subfragments are denoted HindIII EcoRII-G1 and -G2, respectively.

HinfI cleaves EcoRII-G into two subfragments of approximate mate length 265 and 45 base pairs (Fig. 1). The smaller but not the larger HindIII subfragment of EcoRII-G was cleaved by HinfI, and produced the 45-long HinfI fragment of EcoRII-G (data not shown). This result places the HinfI site at a distance of 45 nucleotides from the left-hand end of EcoRII-G. The 265- and 45-long HinfI subfragments are denoted HinfI EcoRII-G1 and -G2, respectively.

Ordering of HaeIII Subfragments of EcoRII-G – The HaeIII restriction endonuclease was reported earlier to cleave EcoRII-G into three fragments of approximate length 160, 100, and 45 base pairs (4); these fragments are denoted as Bands 8, 9, and 10, respectively, in Fig. 1. The 100-long subfragment was identified as the left-hand end piece and the 160-long subfragment as the right-hand end piece of EcoRII-G by virtue of their production by the cleavage of two HaeIII fragments of SV40 DNA, E, and F1, respectively, with the EcoRII restriction endonuclease (4). The 45-long HaeIII subfragment was identified as an internal constituent of EcoRII-G by virtue of its production in full by the HaeIII cleavage of SV40 DNA (4). The three major HaeIII subfragments are denoted as HaeIII EcoRII-G1, -G2, and -G3, respectively, in the order of decreasing length. HaeIII EcoRII-G2, the left-hand end piece of EcoRII-G, was found to contain the single HindIII and HinfI sites within EcoRII-G.

Analysis of RNA transcripts of SV40 DNA annealed to EcoRII-G with T1 and pancreatic ribonucleases revealed the existence of a few products which could not be accounted for in transcripts of the three major HaeIII subfragments of EcoRII-G (see accompanying papers II and III (14, 15)). Subsequent detailed examination of HaeIII digests of EcoRII-G by the twodimensional separation procedure normally used for the fractionation of smaller oligonucleotides revealed the presence of two very small products nine and six base pairs in length, respectively (Fig. 2). Spots 2 and 3 corresponded to the two strands of the larger of the two fragments and Spot 1 contained

The approximate lengths of subfragments of EcoRII-G were determined by comparison of their electrophoretic mobilities with those of fragments of known size such as the HindII + III (3); EcoRII (4), and HaeIII (4) fragments of SV40 DNA in the same slab gel; determinations were also made by comparison with fragments whose sizes were determined by actual nucleotide sequencing, such as the EcoRII and HaeIII subfragments of the HindII + III-G fragment (11).

Restriction cleavage maps of SV40 DNA are drawn in linear form for convenience (4, 11). According to this notation, HindII + III-A is placed to the left of HindII + III-C; therefore, the reciprocal redigestion analysis reported earlier (4) and described above, would place the HindIII site at a distance of about 80 nucleotides from the left-hand end of EcoRII-G.
FIG. 1. Autoradiogram of a polyacrylamide gel electrophoretic separation of the products of cleavage of EcoRII-G with four restriction enzymes. Electrophoresis was from top to bottom in composite slab gels (consisting of 5% gel in the top two-thirds and 10% gel in the bottom one-third of the slab) and was done as described under "Materials and Methods." Uniformly $^{32}$P-labeled EcoRII-G was added as marker to preparative amounts of EcoRII-G and cleaved with the restriction endonucleases Hintl, Alul, HaeIII, or HindIII; the patterns of the subfragments produced are shown in Panels a, b, c, and d, respectively. Band 1 is undigested EcoRII-G. Bands 2 and 3 are products of cleavage with Hintl. Bands 4, 5, 6, and 7 are products of cleavage with Alul; Band 5 was partial digestion product and gave rise to Bands 6 and 7 by cleavage with more Alul. Bands 8, 9, and 10 denote the three major HaeIII subfragments. Bands 11 and 12 are the two HindIII subfragments.

The two strands of the smaller fragment as shown by analysis described in the accompanying manuscript II (14). These two subfragments were both present in HaeIII digests of an Alul subfragment of EcoRII-G which contained the whole of HaeIII EcoRII-G1 and a portion of HaeIII EcoRII-G3. This indicated that the two very small fragments were located between HaeIII EcoRII-G3 and -G1. Analysis of HaeIII digests of other Alul subfragments of EcoRII-G in a similar fashion failed to reveal the presence of any other very small HaeIII subfragments.

The 9- and 6-long HaeIII subfragments were placed in the order -G3, -G4, -G5, and -G1 on the basis of evidence obtained by labeling the Alul subfragment containing the small HaeIII subfragments at the end generated by the Alul cleavage and determination of the nearby sequences by partial snake venom phosphodiesterase digestion. This method is described in greater detail in the accompanying manuscript III (15). The order was confirmed by analysis of limited T, RNase digest of Alul EcoRII-G1.

Location of Alul sites within EcoRII-G—Alul cleavage of EcoRII-G resulted in the production of four fragments of approximate length 180, 135, 100, and 35 base pairs, shown as Bands 4, 5, 6, and 7, respectively, in Fig. 1. Band 5 was identified as a partial digestion product since it gave Bands 6 and 7 on recleavage with Alul (data not shown) and on the basis of the reciprocal redigestion data presented in Fig. 3.

Digestion of HindIII EcoRII-G2 with Alul generated two fragments 45 and 35 base pairs in length, respectively (Fig. 3). HaeIII EcoRII-G2, which contains the whole of HindIII EcoRII-G2, was cleaved by Alul to produce two fragments of length 65 and 35 base pairs, respectively (Fig. 3). These data

FIG. 2. Autoradiogram of two-dimensional fractionation of a HaeIII digest of EcoRII-G. The HaeIII subfragments in the digest were treated with alkaline phosphatase and were then labeled with [$\gamma$-$^{32}$P]ATP as described under "Materials and Methods." The digest was then fractionated by electrophoresis on Cellogel strips at pH 3.5 from left to right and by chromatography on DEAE-thin layer chromatography plates from bottom to top using a 3% solution of partially hydrolyzed yeast RNA (saturated with urea) as the solvent. Spot 1 contains two strands of a 6-long subfragment; Spots 2 and 3 correspond to the two strands of a 9-long subfragment. The smear in the top is due to unreacted [$\gamma$-$^{32}$P]ATP.
showed that the 35-long fragment (denoted Band 7 in Figs. 1 and 3) was present in the left-hand end of HindIII EcoRII-G2 and HaeIII EcoRII-G2, and therefore of EcoRII-G. Thus, one of the AluI sites was localized at a distance of about 35 nucleotides from the left-hand end of EcoRII-G.

Cleavage of HindIII EcoRII-G1 with AluI (Fig. 3) generated a 180-long subfragment (Band 4) also produced by AluI digestion of whole EcoRII-G (Fig. 1); this AluI subfragment was therefore identified as the right-hand end piece of EcoRII-G and the corresponding AluI site was localized at a distance of about 180 nucleotides from the right-hand end of EcoRII-G. This was confirmed by the production of HaeIII EcoRII-G1 by the HaeIII digestion of the 180-long AluI subfragment (Fig. 3); the location of this AluI site within HaeIII EcoRII-G3 is shown by the fact that HaeIII EcoRII-G3 was cleaved by AluI to produce a 35-long fragment denoted Band 14 in Fig. 3 (the other product of the cleavage, being very small, had probably run out of the gel).

The 100-long AluI subfragment corresponding to Band 6 in Fig. 1 was identified as an internal constituent of EcoRII-G since its cleavage with HaeIII resulted in the production of fragments 65 and 35 base pairs in length which were also generated by the AluI cleavage of HaeIII EcoRII-G2 and -G3, respectively (Fig. 3). This AluI subfragment was also found to contain the single HindIII site located within EcoRII-G (data not shown).

The AluI fragment corresponding to Band 5 was identified earlier as a partial digestion product by virtue of the production of Bands 6 and 7 by extensive digestion with AluI. This was confirmed by the finding that the HaeIII digestion of this fragment yielded HaeIII EcoRII-G2 and a 35-long piece corresponding to a major portion of HaeIII EcoRII-G3 (Fig. 3).

The AluI subfragments corresponding to Bands 4, 5, 6, and 7 (Fig. 1) are denoted AluI EcoRII-G1, -G1', -G2', and -G3, respectively. The ordering of the AluI subfragments of EcoRII-G was confirmed by comparison of the T1 and pancreatic RNase fingerprints of RNA transcripts of the AluI subfragments with those prepared from transcripts of the HaeIII, HindIII, and HinfI subfragments of EcoRII-G and was unequivocally proved by secondary and tertiary analyses of the oligonucleotide products of the respective transcripts. The sequences constructed for HaeIII EcoRII-G2 and -G3 (see accompanying manuscript III (15)) predicted the occurrence of the symmetric nucleotide sequence AGCT in three locations within the DNA, two of them occurring in HaeIII EcoRII-G2 and the other one within HaeIII EcoRII-G3. Two of these (one within HaeIII EcoRII-G2 and the other within HaeIII EcoRII-G3) were later found to be the locations where cleavage by the AluI enzyme occurred readily. The third location occurring within HaeIII EcoRII-G2 formed a part of the HindIII cleavage site (HindIII is known to cleave the sequence AAGCTT (22)); for some unknown reason this location was relatively resistant to cleavage by AluI in our hands. Independent studies in another laboratory have also shown that AGCT is the sequence cleaved by AluI (23).

The above results are summarized in Fig. 4 in which the cleavage maps of EcoRII-G obtained with the HaeIII, HindIII, HinfI, and AluI restriction endonucleases are presented. The nomenclature used for the description of the subfragments is based on a proposal by Smith and Nathans (24).

Strategy for Analysis of Sequence of EcoRII-G—One of the methods used for the sequence analysis of EcoRII-G was preparation of RNA transcripts which were separately labeled with each one of the four a-32P-labeled ribonucleoside triphosphates and subsequent analysis of the RNA by digestion with T1 or pancreatic RNase. The sequences of the oligonucleotide products were deduced by methods presented in detail in the accompanying manuscript II (14), and were then assembled as described in the accompanying manuscript III (15) to yield the sequences of the RNA transcripts, and in turn, of the DNA used as the template.

We analyzed two types of RNA preparations: (a) oligonucleotides produced by T1 or pancreatic RNase digestion of that portion of SV40 cRNA which was complementary to the fragment EcoRII-G, and (b) oligonucleotides prepared by T1 or pancreatic RNase digestion of transcripts of subfragments of EcoRII-G produced by various restriction endonucleases. Most of the analyses were done on transcripts of the three major
Nucleotide Sequence of SV40 DNA

Fig. 4. A, a physical map of SV40 DNA showing the sites of cleavage by the EcoRI restriction endonuclease (4). A, B, C, etc., denote the EcoRI-generated fragments of SV40 DNA and the numbers indicate the map positions of the EcoRI cleavage sites. B, HaeIII subfragments. HaeIII EcoRII-G1, -G2, and -G3; transcripts of other fragments, notably AluI EcoRII-G1 and HindIII EcoRII-G1, were also analyzed in detail. In order to circumvent the possibility that the ends of DNA fragments may not be transcribed very well by Escherichia coli DNA-dependent RNA polymerase (which was employed to prepare the transcripts described above) transcripts prepared from subfragments of EcoRII-G overlapping the HaeIII (and all other) restriction cleavage sites were analyzed as described above. The subfragments analyzed in this manner are listed in Table I.

We also determined the sequences of radioactive pyrimidine tracts prepared by depurination of subfragments HaeIII EcoRII-G1, -G2, and -G3, followed by the introduction of labeled phosphate into the 5' end of the tracts. The sequences of the largest pyrimidine tracts of the DNA fragments were used to supplement the analysis of the larger T, and pancreatic RNase digestion products of the RNA transcripts. In this way it was possible to put together the sequence of HaeIII EcoRII-G2 by making use of overlapping sequences between large T, and pancreatic RNase digestion products belonging to the same strand and complementarity between oligonucleotides pertaining to opposite strands of the DNA. The sequences obtained in this way were confirmed by two other independent methods described below and in greater detail in the accompanying manuscript III (15).

Transcripts of HaeIII EcoRII-G1, -G2, and -G3, HindIII EcoRII-G1 and AluI EcoRII-G1 and -G1 (Table I) were digested with T, RNase under conditions that resulted in the production of partial digestion products, each of which contained more than one T, product. The constituent T, products of each limited T, RNase digestion product were released by extensive digestion with T, RNase and were identified by further analysis with pancreatic or JJz RNases. The sequences of the partial T, RNase digestion products were constructed from those of the constituent T, products.

Sequences adjoining each restriction cleavage site within EcoRII-G were also determined by direct DNA sequencing (15). Double-stranded fragments containing the whole or a part of EcoRII-G were labeled at their 5' ends by use of γ-32P-labeled ATP and polynucleotide kinase. The fragments were cleaved with another restriction endonuclease to separate the two labeled ends. The resulting subfragments of EcoRII-G were digested with small amounts of pancreatic DNase I and snake venom phosphodiesterase resulting in the production of single-stranded partial digestion products labeled at their 5'
The sequence near the labeled end was deduced by determination of mobility shifts between each successive partial digestion product. The sequences were confirmed by determination of the electrophoretic mobility on DEAE-paper of the shorter oligonucleotide products. A sequence of up to 20 nucleotides adjoining the labeled end could be determined by this method. This method provided a confirmation of a portion of the EcoRII-G sequence obtained by other methods described above and was useful in determining the strand orientation and order of the two very small HaeIII subfragments HaeIII EcoRII-G3 and -G5 (described in greater detail in the accompanying manuscript III (15)). A list of subfragments analyzed by this method is presented in Table II.

### REFERENCES

Nucleotide Sequence of SV40 DNA

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. I. Mapping of the restriction endonuclease sites within the EcoRII-G fragment and strategy employed for its sequence analysis.
K N Subramanian, R Dhar and S M Weissman


Access the most updated version of this article at http://www.jbc.org/content/252/1/333.citation

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/252/1/333.citation.full.html#ref-list-1