JOURNAL OF VIROLOGY, Jan. 1981, p. 244-247
0022-538X/81/010244-04$02.00/0

Vol. 37, No. 1

Nucleotide Sequence at Polyoma VP1 mRNA Splice Sites

E. S. SRIVATSAN, P. L. DEININGER, AND T. FRIEDMANN*

Department of Pediatrics, University of California, San Diego, La Jolla, California 92093

Double-stranded DNA complementary to total cytoplasmic polyadenylated RNA isolated late in infection from polyoma virus-infected mouse 3T6 cells was cloned in *Escherichia coli* by using the large *Hin*dIII-*Bam*HI fragment of pBR322 plasmid DNA. Polyoma-specific DNA inserts were detected by hybridization, and then nucleotide sequences were determined from two clones. The sequence of the spliced region for VP1 mRNA of two clones is: 5'-----TGATCAAGTAAGTG-----
--------TTCTAGGGCTGTA---------3'. Solid arrows indicate splice sites compatible with prototypical mammalian splice sites, and the dashed arrows indicate possible alternative splice sites leading to the same spliced product. A sequence of 897 nucleotides was spliced out of the primary transcript during the processing of the mature VP1 mRNA. Restriction enzyme mapping with four other independently isolated clones indicates that these are the major splicing signals for the VP1 message. The distal splice site is 48 nucleotides upstream from the initiator codon.

The late region of the polyoma genome consists of approximately 2,400 base pairs and encodes three capsid proteins, VP1, VP2, and VP3 (2, 6, 22, 23). It is known that the capsid proteins VP2 and VP3 are encoded at the 5' end of the late region and are synthesized from 19S and 18S mRNA's, respectively (22, 23). The two proteins are read from a single common reading frame (2). The major capsid protein, VP1, is translated from a 16S mRNA consisting of a leader sequence derived from the 5' end and a messenger body derived from the 3' end of the late region. It is read from a reading frame different from that of VP2 and VP3 (22, 23). The three late mRNA's have been mapped on the polyoma genome by S1 mapping and electron microscopy (14, 16), and it has been shown that the VP1 and VP3 messages contain a single intervening sequence, whereas the VP2 message has not been shown to have such a sequence. In the present study, we present the sequence of the spliced region of the VP1 mRNA as determined by the sequence analysis of the virus-specific complementary DNAs (cDNA's) cloned in pBR322.

MATERIALS AND METHODS

The large-plaque, wild-type polyoma strain used in this study corresponds to the strain previously described in detail from this laboratory (3, 8). Restriction enzymes were obtained from New England Biolabs. All other reagents were reagent grade.

Preparation of polyadenylated RNA. Mouse 3T6 cells were grown to 60 to 70% confluence and infected with polyoma virus by using 5 PFU/cell. At 38 h after infection, the cells were lysed with lysis buffer [15 mM Tris-hydrochloride (pH 7.4), 60 mM KCl, 15 mM NaCl, 2 mM EDTA, 0.34 M sucrose, 0.5 mM ethyleneglycol-bis(β-aminoethyl ether)-*N*-*N*-tetraacetic acid, 0.5 mM spermidine, 0.15 M spermine, 15 mM β-mercaptoethanol] containing 0.1% Nonidet P-40 and 0.05% deoxycholate at room temperature; the nuclear and cytoplasmic fractions were prepared by differential centrifugation (18). The cytoplasmic fraction was treated with proteinase K (200 μg/ml) and 0.5% sodium dodecyl sulfate at 37°C for 2 h. Nucleic acids were extracted three times with $CHCl_3$-phenol (1:1) and ethanol precipitated in 0.3 M sodium acetate buffer (pH 5.5) overnight at −20°C. The RNA was dissolved in 10 mM Tris-hydrochloride (pH 7.4)-10 mM $MgCl_2$ buffer, treated with RNase-free DNase (50 μg/ml, 1 h at 37°C), and reprecipitated. This RNA was dissolved in 10 mM Tris-hydrochloride (pH 7.4) buffer containing 0.4 M NaCl and 0.5% sodium dodecyl sulfate and passed through oligodeoxythymidylic acid-cellulose equilibrated with the same buffer. It was eluted with 10 mM Tris-hydrochloride (pH 7.4)-0.5% sodium dodecyl sulfate and repurified on oligodeoxythymidylic acid-cellulose.

Synthesis of cDNA. Synthesis of double-stranded cDNA was in two steps (20). In step 1, a total reaction volume of 100 μl containing 40 μg of polyadenylated RNA and 10 μg of oligodeoxythymidylic $acid_{12-18}$ in reverse transcriptase buffer (50 mM Tris-hydrochloride [pH 8.3], 8 mM $MgCl_2$, 50 mM NaCl, 10 mM dithiothreitol) was incubated at 41°C in the presence of 1 mM each dATP, dGTP, and dTTP, 20 μCi of [α-^{32}P]dCTP, and 80 U of avian myeloblastosis virus reverse transcriptase (a gift of J. W. Beard, Life Sci-

ences, St. Petersburg, Fla.). After 15 min of incubation, unlabeled dCTP (1 mM) was added, and the incubation was continued for 30 min. The reaction was terminated by the addition of 200 μl of buffer containing 5 mM Tris-hydrochloride (pH 8.3) and 1 mM dithiothreitol and by incubation at 47°C for 30 min. This sample was extracted with $CHCl_3$-phenol (1:1) and ethanol precipitated as above. The RNA-cDNA mixture was treated with 0.25 M NaOH, at 65°C for 30 min, neutralized with 0.25 M sodium acetate (pH 5.5), and ethanol precipitated. The cDNA was further purified by Sephadex G-100 chromatography with 10 mM Tris-hydrochloride (pH 7.4)–1 mM EDTA. In step 2, the single-stranded cDNA synthesized in the above step was incubated with 80 U of avian myeloblastosis virus reverse transcriptase with 20 μCi of [α-^{32}P]dGTP and 1 mM each dATP, dCTP, and dTTP in reverse transcriptase buffer as described in step 1. After the termination of reaction in 5 mM Tris-hydrochloride (pH 8.3) at 47°C, double-stranded cDNA was extracted with $CHCl_3$-phenol, ethanol precipitated, and purified by Sephadex G-100 chromatography.

Cloning of cDNA in pBR322 plasmid. The double-stranded cDNA was digested overnight at 37°C with 10 U each of *Hin*dIII and *Bcl*I in Hin buffer (6.6 mM Tris-hydrochloride [pH 7.4], 6.6 mM $MgCl_2$, 50 mM NaCl, 1 mM dithiothreitol). The *Hin*dIII was inactivated by heating at 65°C for 10 min. The pBR322 plasmid (10 μg) was digested overnight at 37°C with 10 U each of *Hin*dIII and *Bam*HI in Hin buffer, and the restriction fragments were separated on a 1.0% agarose gel in TBE buffer (90 mM Tris-borate [pH 8.3], 2 mM EDTA). The 4-kilobase fragment of pBR322 was electroeluted in 1/5-strength TBE buffer containing 0.1% sodium dodecyl sulfate at 60 mA for 12 h and ligated to the double-stranded cDNA in ligase buffer (67 mM Tris-hydrochloride [pH 7.4], 33 mM NaCl, 12 mM $MgCl_2$, 10 mM dithiothreitol) by using 0.01 U of T4 ligase.

Transformation of SF8 *Escherichia coli* (*recA* strain) with the recombinant DNA was performed by the method of Enea et al. (4). Ampicillin-resistant colonies were transferred onto Whatman 541 filters and hybridized to nick-translated polyoma DNA (15) by the method of Grunstein and Hogness (10). The polyoma-specific clones were grown in Penassay broth (1) and amplified with chloramphenicol (250 μg/ml) at an optical density at 650 nm of 0.6; the plasmid DNA was purified from the bacterial lysate (1) by cesium chloride density gradient centrifugation (19). DNA was digested with restriction enzymes and analyzed on agarose gels as described above.

Preparation of the labeled fragment and sequencing. Purified plasmid DNA (20 μg) was digested with 10 U of *Hin*dIII in Hin buffer at 37°C for 4 h and labeled with 20 μCi each of the four α-^{32}P-labeled deoxynucleoside triphosphates by using 35 U of avian myeloblastosis virus reverse transcriptase at 41°C. This labeled DNA was digested with 10 U of *Sal*I at 37°C for 4 h and applied to a 6% polyacrylamide gel (75 by 18 by 0.05 cm) with cross-linking of acrylamide-bisacrylamide (20:1). Electrophoresis was in TBE buffer at 400 V for 10 h. A cDNA fragment of 480 base pairs predicted to contain the cDNA insert for the VP1 mRNA (25) was eluted overnight at 65°C with ammonium acetate elution buffer (8). The eluted fragment was ethanol precipitated with 20 μg of calf thymus DNA as carrier, subjected to the base-specific modification reactions, cleaved at the modified sites with piperidine, and sequenced by the method of Maxam and Gilbert (17) as modified by Friedmann and Brown (7).

RESULTS

The leader and coding regions of the VP1, VP2, and VP3 mRNA's are illustrated in Fig. 1 (3, 9). By a comparison of the transcriptional studies with the nucleotide sequence analysis, Deininger et al. (2) have predicted the positions of the proximal and distal VP1 splice sites. The proximal site was predicted to be close to, but a few nucleotides downstream from, the single *Bcl*I restriction site. The distal site was predicted to be about 200 nucleotide pairs or 4 map units upstream from the *Hin*dIII site at position 3,943 in Fig. 1. Since the *Hin*dIII site is well within the body of the VP1 mRNA, the cDNA fragment was prepared with *Hin*dIII and *Bcl*I. Ligation of the *Bcl*I end of the *Bcl*I-*Hin*dIII cDNA fragment to the *Bam*HI site of pBR322 caused the ligation site to become insensitive to both *Bcl*I and *Bam*HI. The size of the polyoma insert in plasmid DNA purified from the poly-

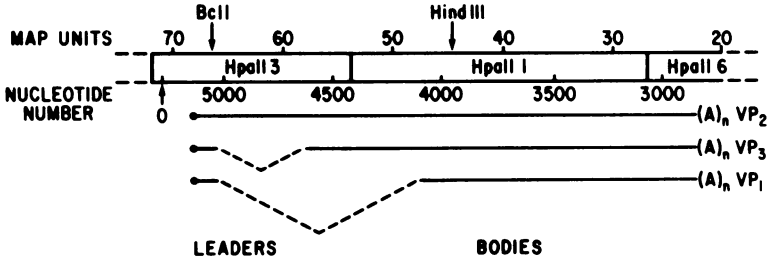

FIG. 1. *Map of the late region of polyoma DNA, indicating the junction points of 5′ capped leader sequences with the main bodies of the mature mRNA's. Map units are from the physical map of the virus from the EcoRI site (9). The nucleotide numbers and the restriction sites are derived from the sequence map of the viral genome (3). The dashed lines indicate the positions of the intervening sequences for the VP1 and VP3 mRNA's.*

oma-containing clones was therefore determined by an analysis of the *TaqI* enzyme products on agarose gels, as illustrated in Fig. 2

Figure 2a shows the *TaqI* fragments of six independently isolated clones compared with pBR322 alone, all visualized by ethidium bromide staining. Two pBR322-specific bands, at 315 and 312 base pairs, are missing and are replaced by a new band estimated to be 475 base pairs long. From the known position of the *TaqI* site in pBR322, it is estimated that this new fragment consists of 275 bases derived from pBR322 and approximately 205 base pairs of polyoma sequence. Analysis of the same fragments by transfer to nitrocellulose filters and hybridization (24) with a nick-translated polyoma probe (Fig. 2b) demonstrates that the new fragment does indeed contain polyoma sequences.

Two of the six clones were subjected to sequence analysis, and a partial sequence of the region which spans the splice junction is shown in Fig. 3. The sequence is compatible with splicing events occurring after positions 5,044 and 4,147 or, alternatively, 5,043 and 4,146, in either case removing an intervening sequence of 897 nucleotides from the primary transcript. The ATG initiation codon for VP1 is 48 nucleotides downstream from the distal splice. The rest of the nucleotide sequence extending from nucleotide 4,075 to the 3′ end of the splice was identical to that of the polyoma DNA.

DISCUSSION

Splicing is a common phenomenon during the synthesis and processing of many eucaryotic transcripts and is probably a late event in RNA processing subsequent to polyadenylation (13). It is known that the RNA primary sequence plays an important role in determining the specificity of splicing, and Seif et al. (21) have described prototypical sequences determined from most well-characterized eucaryotic splice sites. The donor site at the 5′ end of intervening sequences has one of four forms, ↓GTAXG, ↓GTAXXT, ↓GTXAC, or ↓GTXXGT. An optimal sequence might be ↓GTAAGT. The acceptor site at the 3′ end of the intervening sequence is pyrimidine rich and usually takes the form PyPyxPyAG↓. Dipurines other than AA are rarely present within 16 nucleotides of this site. It has also been suggested, but not proven, that RNA conformation may play some role in the splicing event (12).

Using the above model and published transcriptional data, Deininger et al. (2) have suggested that there may be a common proximal splice for both VP1 and VP3 RNAs at position

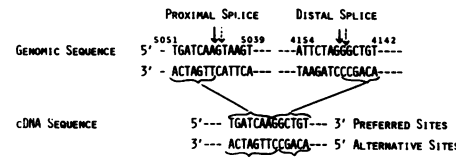

FIG. 3. *Proximal and distal splice sites for the VP1 cDNA. The sequence shown extends from the BclI site at map position 5,051 to 5 bases past the distal splice site at map position 4,147. The solid arrows indicate the positions for the preferred splice sites (see text), and the dashed arrows give alternative positions that might result in an identical sequence in the spliced mRNA.*

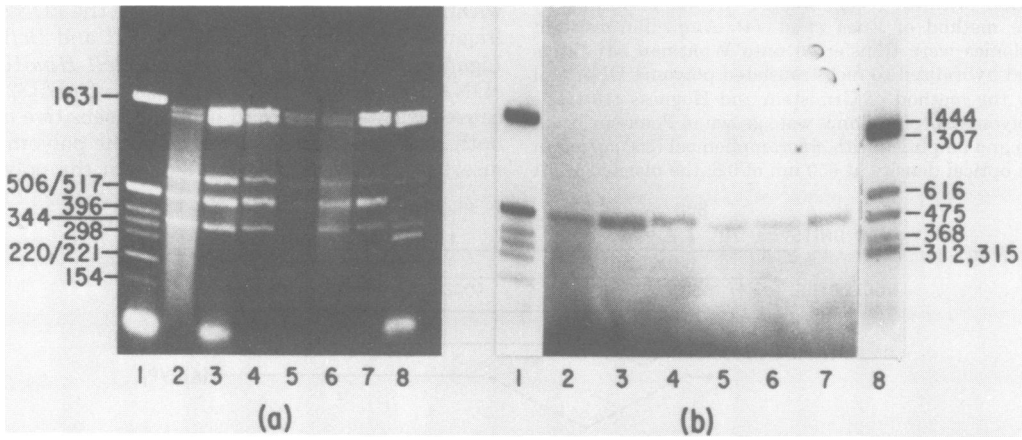

FIG. 2. *Size of the polyoma DNA inserts in recombinants. Purified plasmid DNA (2 μg) was digested with 2 U of TaqI in Hin buffer at 65°C overnight and analyzed on 1.5% agarose gel in TBE buffer at 50 mA for 2.5 h. (a) The gel was stained with ethidium bromide (5 μg/ml) and visualized under UV light. (b) The DNA fragments separated on the same gel as in (a) were transferred onto nitrocellulose filters and hybridized either to nick-translated polyoma DNA in the case of the recombinants or to pBR322 DNA in the case of the markers (23). Lanes: (1) pBR322 DNA digested with HinfI; (2–7) cloned plasmid DNAs digested with TaqI; (8) pBR322 DNA digested with TaqI. Numbers on the sides of the gel represent the size of the DNA fragments.*

251 in the late region of the polyoma genome within the sequence 5′ - TCAA ↓GTAAGT - - - -3′. This predicted splice site is in agreement with the sequence determined from the cDNA clones. The acceptor site, with the sequence 5′ ----- AATTCTAGGG-----3′ is also compatible with rules suggested by Seif et al. (21). Since the four other, nonsequenced clones give rise to cDNA fragments of similar size (Fig. 2b), the splice sequence determined in the present analysis probably results from the major splicing event in the synthesis of VP1 mRNA's. The VP1 leaders with different 5′ termini (5) therefore seem to have the common 3′ terminus 5′-GAUCAA linked to the body of the message starting with the sequence 5′-GGCUGU. This study was not designed to examine the nature of the leader-to-leader splices recently described for the polyoma late region mRNA's (11, 14).

From the primary sequence of the DNA, transcriptional data, and the deduced amino acid sequence, the VP3 splice sites can be predicted by using the model of Seif et al. We expect the proximal splice site to be the same as that for VP1, and the distal splice might then be ↓AGGATT....3′ 52 nucleotides upstream from the initiation codon for the VP3 protein, splicing out 310 nucleotides from the viral genome.

We are currently characterizing other cDNA clones with the aim of characterizing the splice sites for VP3 and possibly for VP2, as well as for the early gene transcripts.

ACKNOWLEDGMENTS

We thank Douglas Jolly of our laboratory for many helpful discussions. Thanks are due to Abby Esty for her excellent technical assistance and Dorothy Miller for her administrative assistance in the preparation of the manuscript.

P.L.D. was supported by Public Health Service training grant GM07199 from the National Institutes of Health, and E.S.S. was supported by a training grant from the Kroc Foundation and by Public Health Service research grant CA24288 from the National Cancer Institute.

LITERATURE CITED

1. **Clewell, D. B., and D. R. Helinski.** 1969. Super-coiled circular DNA-protein complex in *Escherichia coli*: purification and induced conversion to an open circular DNA form. Proc. Natl. Acad. Sci. U.S.A. **62**:1159-1166.
2. **Deininger, P. L., A. Esty, P. LaPorte, and T. Friedmann.** 1979. Nucleotide sequence and genetic organization of the polyoma late region: features common to the polyoma early region and SV40. Cell **18**:771-779.
3. **Deininger, P. L., A. Esty, P. La Porte, H. Hsu, and T. Friedmann.** 1980. The nucleotide sequence and restriction enzyme sites of polyoma genome. Nucleic Acids Res. **8**:855-860.
4. **Enea, V., G. F. Vovis, and N. D. Zinder.** 1975. Genetic studies with heteroduplex DNA of bacteriophage f1, asymmetric segregation, base correction and implications for the mechanism of genetic recombination. J. Mol. Biol. **96**:495-509.
5. **Flavell, A. J., A. Cowie, S. Legon, and R. Kamen.** 1979. Multiple 5′ terminal cap structures in late polyoma virus RNA. Cell **16**:357-371.
6. **Fried, M., and B. Griffin.** 1977. Organization of the genomes of polyoma virus and SV40. Adv. Cancer Res. **24**:67-114.
7. **Friedmann, T., and D. Brown.** 1978. Base specific reactions useful for DNA sequencing. Nucleic Acids Res. **5**:615-622.
8. **Friedmann, T., A. Esty, P. LaPorte, and P. L. Deininger.** 1979. The nucleotide sequence and genome organization of the polyoma early region: extensive nucleotide and amino acid homology with SV40. Cell **17**: 715-724.
9. **Griffin, B. E., M. Fried, and A. Cowie.** 1974. Polyoma DNA: a physical map. Proc. Natl. Acad. Sci. U.S.A. **71**: 2077-2081.
10. **Grunstein, M., and D. S. Hogness.** 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. U.S.A. **72**:3961-3965.
11. **Kamen, R., J. Favaloro, and J. Parker.** 1980. Topography of the three late mRNA's of polyoma virus which encode the virion proteins. J. Virol. **33**:637-651.
12. **Khoury, G., P. Gruss, R. Dhar, and C. J. Lai.** 1979. Processing and expression of early SV40 mRNA: a role for RNA conformation in splicing. Cell **18**:85-92.
13. **Lai, C. J., R. Dhar, and G. Khoury.** 1978. Mapping the spliced and unspliced late lytic SV40 RNAs. Cell **14**: 971-982.
14. **Legon, S., A. J. Flavell, A. Cowie, and R. Kamen.** 1979. Amplification in the leader sequence of late polyoma virus mRNAs. Cell **16**:373-388.
15. **Maniatis, T., A. Jeffrey, and D. G. Kleid.** 1975. Nucleotide sequence of the rightward operator of phage λ. Proc. Natl. Acad. Sci. U.S.A. **72**:1184-1188.
16. **Manor, H., M. Wu, N. Baran, and N. Davidson.** 1979. Electron microscopic mapping of RNA transcribed from the late region of polyoma viral DNA. J. Virol. **32**:293-303.
17. **Maxam, A., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560-569.
18. **Mulligan, R. C., B. H. Howard, and P. Berg.** 1979. Synthesis of rabbit β-globin in culture monkey kidney cells following infection with a SV40 β-globin recombinant genome. Nature (London) **277**:108-114.
19. **Redloff, R., W. Bauer, and J. Vinograd.** 1967. A dye-buoyant-density method for the detection and isolation of closed circular duplex DNA: the cloned circular DNA in HeLa cells. Proc. Natl. Acad. Sci. U.S.A. **57**:1514-1521.
20. **Rougeon, F., and B. Mach.** 1977. Cloning and amplification of rabbit α and β-globin gene sequences into *Escherichia coli* plasmids. J. Biol. Chem. **252**:2209-2217.
21. **Seif, I., G. Khoury, and R. Dhar.** 1979. BKV splice sequences based on analysis of preferred donor and acceptor sites. Nucleic Acids Res. **6**:3387-3398.
22. **Siddell, S., and A. E. Smith.** 1978. Polyoma virus has three late mRNA's: one for each virion protein. J. Virol. **27**:427-435.
23. **Smith, A. E., R. Kamen, W. F. Manzel, H. Shure, and T. Wheeler.** 1976. Location of the sequences coding for capsid proteins VP1 and VP2 on polyoma virus DNA. Cell **9**:481-487.
24. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503-517.
25. **Sutcliffe, J. E.** 1978. pBR322 restriction map derived from the DNA sequence: accurate DNA size markers up to 4362 nucleotide base pairs long. Nucleic Acids Res. **5**:2721-2728.