Genomic Structure of Human Polyoma Virus JC: Nucleotide Sequence of the Region Containing Replication Origin and Small-T-Antigen Gene

TATSUO MIYAMURA, HIROYUKI JIKUYA, EIICHI SOEDA, and KUNITO YOSHIIKE*  
Department of Enteroviruses, National Institute of Health, Tokyo 141, and Department of Molecular Genetics, National Institute of Genetics, Misima 411, Japan

Received 5 August 1982/Accepted 23 September 1982

The nucleotide sequence of the region of human polyoma virus JC DNA between 0.5 and 0.7 map units from a unique EcoRI cleavage site was determined and compared with those of the corresponding regions of another human polyoma virus, BK, and simian virus 40 DNAs. Within this region consisting of 945 base pairs, we located the origin of DNA replication near 0.7 map units, the entire coding region for small T antigen, and the splice junctions for large-T-antigen mRNA. The deduced amino acid sequences for small T antigen and the part of large T antigen markedly resembled those of polyoma virus BK and simian virus 40. The results strongly suggest that polyoma virus JC has the same organization of early genome as polyoma virus BK and simian virus 40 on the physical map, with the EcoRI site as a reference point.

Human polyoma virus JC (JCV) has been isolated from brain tissues of patients with progressive multifocal leukoencephalopathy, a rare demyelinating disease, for which JCV is probably responsible as an etiological agent (9, 19). Serological surveys have revealed that JCV infection is common among human populations (2, 18). Like other polyoma viruses, JCV has tumorigenic and transforming capacities (7, 20, 33, 35). Therefore, studies of JCV are both etiologically and biologically important.

Biological and biochemical studies on JCV, however, have been greatly hampered because JCV can grow efficiently only in primary human fetal glia cell cultures rich in spongioblasts (17) and because viral DNA from the infected cells is usually heterogeneous in size (12, 16). Although JCV has been successfully adapted to growth in more readily available cells, such as human amnion cells (31) and human embryonic kidney (HEK) cells (14), the viral DNA thus obtained is heterogeneous. A JCV mutant growing well in HEK cells has been found to contain two classes of defective DNAs, which probably complement each other for viral growth (38).

Molecular cloning of JCV DNA in a bacterial vector (10) from DNA obtained by a low multiplicity of infection of primary human fetal glia cells (12) has now made detailed analyses of the JCV genome possible. In the present study, we determined the nucleotide sequence of the region near 0.7 map units from EcoRI site in an attempt to correlate the physical map of JCV DNA with the genomic organization. From the similarity to simian virus 40 (SV40) and human polyoma virus BK (BKV) DNAs, we could locate the replication origin and the entire coding region for the small T antigen. Also found in this region were the signal sequences conserved near the splicing sites of eucaryotic mRNA.

MATERIALS AND METHODS

JCV DNA. A restriction endonuclease cleavage map of JCV (Mad-1) DNA (10, 12, 22) is shown in Fig. 1. A recombinant plasmid, pJC(1-4), which has a JCV DNA insert at the BamHI site of plasmid pBR322 (10), was used. DNA was extracted and purified as described previously (38).

Enzymes. Restriction endonucleases BamHI, HaeIII, HindIII, HpaI, PstI, and PvuII were generously provided by Takara Shuzo Co., Ltd., Kyoto, Japan, and were used as recommended by the manufacturer. Bacterial alkaline phosphatase and T4 poly- nucleotide kinase were purchased from Boehringer Mannheim Biochemicals, Indianapolis, Ind.

DNA sequencing. Nucleotide sequences were determined by the method of Maxam and Gilbert (13) with a strategy illustrated in Fig. 1. pJC(1-4) DNA digested with HindIII, BamHI, HpaI, and PstI was subjected to preparative electrophoresis in 5.5% polyacrylamide gels. After elution, HindIII-B and the shorter HpaI- cleaved HindIII-A were further cleaved each into two subfragments with HaeIII and PvuII, respectively. Double-stranded fragments isolated from gels were labeled at their 5' ends with [γ-32P]ATP (specific activity, 2,000 Ci/mmole, Amersham International) and T4 polynucleotide kinase. The labeled DNA was denatured by heat and separated into each strand on

73
5% polyacrylamide-0.1% bisacrylamide gels. Fragments labeled at one end were submitted to the four sets of partial degradation reaction of A > C, G, C, and C + T. The degradation products were run on 0.5-mm-thick 12% polyacrylamide-0.62% bisacrylamide-7 M urea gels. Autoradiography was done at −70°C with an intensifying screen (Kodak X-Omatic).

RESULTS

Like SV40 and BKV DNAs, JCV DNA has a unique EcoRI site that can be used as a reference point (Fig. 1). Cross-hybridization under less stringent conditions between restriction endonuclease cleavage fragments of these three polyoma viruses has shown that their physical maps could be colinearly aligned by using the EcoRI site as the O map position (11). To locate the replication origin precisely on the physical map of JCV DNA, we determined the nucleotide sequence of the area presumably containing the replication origin and coding for 5′ end of the early mRNAs and compared it with the corresponding areas of BKV and SV40 DNAs, whose entire sequences and relation to the genetic maps have been established (5, 21, 24, 37). Figure 2 shows the nucleotide sequence of 945 base pairs covering the area from 0.5 to 0.7 map units.

The origin of JCV DNA replication was located near the HindIII-A/C junction (Fig. 1). Since the origin of SV40 DNA replication has been confined within a narrow region around the BglII cleavage site (8, 25), the similar nucleotide sequence to that region was searched for in the JCV DNA sequence. The replication origin of JCV DNA was assigned to the region of nucleotides 36 to 121 (Fig. 2), which shows extensive homology with the replication origin of both SV40 (30) and BKV DNA (3). The structures of replication origins are compared among the three polyoma viruses in Fig. 3.

The nucleotide sequence of the replication origin can be divided into three parts with different structural features, AT stretch (nucleotides from 36 to 50; boxed in Fig. 2), two pairs of the inverted repeat sequences (nucleotides 54 to 102), and a palindromic sequence with mirror image form (nucleotides 105 to 121; boxed by dotted line in Fig. 2 and 3). Goldberg-Hogness box structure (24) was detected at the right end of the AT stretch (nucleotides 43 to 50 in Fig. 2 and 3). The region between AT stretch and the palindrome was rich in GC, with intrastrand symmetry which could form two pairs of double hairpin loops, as indicated by the arrows in Fig. 3.

The coding region for small T antigen was found to initiate from ATG at nucleotides 182 to 184 and terminate at TAA at nucleotides 698 to 700 (Fig. 2). This ATG was the first initiation codon in the longest open reading frame downstream from the replication origin. The topology of the initiation codon is similar to that of BKV and SV40 (4, 34). Consequently, JCV small-T-
FIG. 2. Nucleotide sequence of JCV DNA: the region containing replication origin and small-T-antigen gene. Numbering is given from near the HindIII-A/C junction (nucleotide 1) to near the BamHI site (nucleotide 945). We identified the orientation of the strands from the similarity to BKV and SV40 DNAs. The upper strand, read from left to right, is believed to have the same polarity as early mRNA. The AT stretch within the replication origin is indicated in the solid line box (nucleotides 36 to 50). The dotted line box indicates the palindrome (nucleotides 105 to 121). ATG at nucleotides 182 to 184 is considered to be the initiation codon for T antigen mRNA, and TAA at nucleotides 698 to 700 is considered to be the termination codon. The arrows refer to the splice junctions for large-T-antigen mRNA. The arrow between nucleotides 424 and 425 is for the donor site, and the other, between nucleotides 768 and 769, is for the acceptor site (see Fig. 6).
antigen gene can code for 172 amino acids, like those of BKV and SV40 encoding 172 and 174 amino acids, respectively. Furthermore, the deduced amino acid sequences of small T antigens showed extensive homology among the three polyoma viruses (Fig. 4).

From the homology with the signal sequences proposed for splicing of large-T-antigen mRNA in SV40, BKV, and mouse polyoma virus (23), we identified the sequence 5'TCAGAG/GTTGTT3' at nucleotides 419 to 430 as a donor site and the sequence 5'TTTTTTTAG/GTGGCCA3' at nucleotides 760 to 774 as an acceptor site of large-T-antigen mRNA of JCV (Fig. 5). In the nucleotide sequence downstream from the acceptor site of JCV, one reading frame was left open and the other two were closed. The amino acid sequence predicted from the open reading frame downstream from the acceptor site was similar to those from BKV and SV40 T antigen genes in the corresponding regions (Fig. 6).

**DISCUSSION**

The complete nucleotide sequences and the genomic organization for BKV, SV40, and mouse polyoma virus have been established, and several recognition sequences for regulatory functions have been mapped on the nucleotide sequences (5, 21, 24, 28, 37). From the homology determined by DNA-DNA hybridization, the cleavage map of JCV DNA has been aligned, with reference to EcoRI site, to those of BKV and SV40 DNAs (11), but little is known about the genetic map and genomic organization of JCV. In the present study, we attempted to locate the origin of DNA replication and the proximal part of the early region of JCV genome by comparing JCV sequences to those of BKV and SV40 to correlate the physical map to the genomic organization. Apparently, the similarity of the sequences for replication origin and of the amino acid sequences for putative small T antigen between JCV and other primate polyoma viruses proves that JCV has early genome organization similar to that of BKV and SV40 on the physical map of DNA with the conserved single EcoRI site in each genome as a reference point.

Since the structural features of the region around the HindIII-A/C junction of JCV DNA (at nucleotides 36 to 121) were strikingly similar to those of the replication origins of BKV and SV40 (3, 30), this area was regarded as the replication origin of JCV DNA (Fig. 3). The area is composed of a long AT stretch, two inverted repeat sequences, and a palindrome. At the right end of the AT stretch, the three primate polyoma viruses had sequences similar to the Goldberg-Hogness box sequence of 5'TATAAATA3' (24). By analogy to SV40 (1), sequence 5'TATAAATA3' of JCV (nucleotides 43 to 50) is probably related to the early promoter region. The AT-rich sequence was followed by the GC-rich region which could form double hairpin loops (Fig. 3). Although it is unclear whether this structure is actually formed in vivo, it is
possible that the higher structures of DNA near the replication origin affect the expression of the viral genome. In the first hairpin loop indicated by an arrow in Fig. 3, 22 out of 23 base pairs were identical to each other among JCV, BKV, and SV40. The composition of the second hairpin structure, on the other hand, was less homologous, but exactly the same sequence of 5'AAATGTTAGGG3' was found. Another highly homologous sequence among the three polyoma virus DNAs was found in 17-base pair palindromes. Of 17 nucleotides, 15 were identical. Although the general features of replication origin were shared by mouse polyoma virus (6, 29), the strong homology of nucleotide sequences was detected only among the three primate polyoma viruses.

There was a stretch of 60 nucleotides between the palindrome and the initiation codon for T antigens (Fig. 3). Although the homology in this noncoding region among the three primate polyoma viruses was weaker than that in the replication origin, they had repeats of 5'TnGC3' and 5'GCaNaN3', which could form hairpin loops (indicated by arrows in Fig. 3), in this region. The structure of the corresponding region in mouse polyoma virus DNA (27) is distinct from those of the primate polyoma viruses.

It seems reasonable to consider the ATG codon at nucleotides 182 to 184 to be an initiation signal for translation of T antigen messengers for the following reasons: (i) this was the first initiation codon in the only open reading frame in this region; (ii) its topology was almost the same as that of its BKV and SV40 counterparts; (iii) the deduced amino acid sequence for JCV small T antigen resembled those of BKV and SV40 (Fig. 4).

Thus, the small-T-antigen gene of JCV can code for 172 amino acids (Fig. 4). In the N-terminal regions, where the small T antigen shares its amino acids with the large T antigen,
JCV and BKV shared 72 of 81 amino acids (89%). The homology between JCV and SV40 was 83% (67 of 81), and that between BKV and SV40 was also 83%. Amino acids shared by the three viruses in the proximal half of the small T antigen were as many as 63 (78%). The rest of the small-T-antigen coding region seemed to be diverged somewhat more from BKV and SV40 than the former region shared by the large T antigen. In this region, 63 of 91 amino acids (69%) were common to JCV and BKV, 49 (54%) to JCV and SV40, and 54 (59%) to BKV and SV40. The three viruses shared only 43 of 91 amino acids (47%) but had a common stretch of five amino acids, Arg-Asp-Leu-Lys-Leu, followed by stop codon TAA. These data are consistent with the earlier reports showing the relatedness of T antigens of the three primate polyoma viruses by immunological cross-reactivity (32, 35) or 35S methionine-labeled tryptic peptide patterns (26).

The signal sequences for splicing of mRNA are conserved among eucaryotic species, and the consensus sequences have been proposed (15). The latter half of the small-T-antigen gene is spliced out in large-T-antigen mRNAs of papovaviruses. The splicing signals for JCV large-T-antigen mRNA agree well with those of polyoma viruses (23) (Fig. 5) and the consensus sequences (15). The amino acid sequence encoded downstream from the splicing site resembled those of BKV and SV40 (Fig. 6). Like BKV, SV40, and mouse polyoma virus (24), JCV had a stretch of aromatic amino acids sequence in the region just downstream from the splicing signal.

It has not been determined yet whether JCV has a virus-coded middle T antigen like that of mouse polyoma virus. Downstream from the splice junction (acceptor site), mouse polyoma virus DNA has two open reading frames, which are used for encoding large and middle T antigens, respectively (28). In the corresponding region of JCV DNA, the two frames not used for encoding large T antigen were short (Fig. 2), and the amino acid sequences deduced from these two frames (data not shown) had no homology with that of mouse polyoma virus middle T antigen (28). The topology of coding frames of JCV in this area resembled those of SV40 and BKV. We conclude, therefore, that JCV cannot code for the protein like mouse polyoma virus middle T antigen.

The present study showed that, in the region containing the replication origin and small-T-antigen gene, the structure of JCV DNA resembles those of the other polyoma viruses. The homology in the nucleotide or amino acid sequence was higher between the primate polyoma viruses than between JCV and mouse polyoma virus (28) and was the highest between JCV and BKV. Despite the similarity between JCV and BKV in this region, the two viruses have different pathogenicity in humans and different host ranges for cultured cells. The sequences responsible for the biological difference between the two viruses remain to be studied.

A host range mutant of JCV, JC-HEK, can grow well in HEK cell cultures (14), whereas wild-type JCV (Mad-1) can grow efficiently only in primary human fetal glia cell cultures (17). Interestingly, mutant JC-HEK has now the same host range as BKV. It is possible that comparison of the nucleotide sequence between JCV and JC-HEK, and between JCV and BKV, will reveal the structure that determines the host range of these viruses. A study of the structure of JC-HEK DNA is now in progress.

ACKNOWLEDGMENTS

This work was supported in part by grants-in-aid for cancer research from the Ministry of Education, Science, and Culture, Japan.

We thank Peter M. Howley, National Cancer Institute, Bethesda, Md., for his generous gift of cloned JCV DNA.

LITERATURE CITED


