Volume 7 Number 3 1979 Nucleic Acids Research

Comparative study of papovavirus DNA: BKV(MM), BKV(WT) and SV40

Robert C.A.Yang and Ray Wu

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, USA

Received 11 July 1979

ABSTRACT

Extensive physical mapping revealed that approximately 90% of the genomes of BKV(prototype, WT) and BKV (MM strain) are identical or closely related. Nucleotide sequences of the non-homologous regions and a large portion of the homologous regions have been determined for both genomes. The coding sequence of small t antigen of BKV(MM) is 216 nucleotides shorter than that of BKV(WT), even though no differences in biological function of the t antigen was observed. Both genomes contain three similar sets of 44-61 base-pair repeated sequences. However, the DNA sequence of the tandem repeats is totally different between BKV (human cell as host) and SV40 (monkey cell as host). On the other hand, the region between the N-terminus of the T antigen genes and the origin of replication is dominated by a similar set of palindromic sequences in BKV and SV40 DNA. There is also extensive homology between the regions which code for proteins in BKV and SV40, suggesting a close evolutionary relationship.

INTRODUCTION

Human papovavirus BKV was isolated from the urine of an immunosuppressed patient[1]. A variant of BKV, BKV(MM), was later isolated from the brain tumor and urine of a patient with Wiskott-Aldrich syndrome[2]. The genome of BKV is a closed circular duplex DNA of 3.45×10^6 daltons; hybridization analysis has shown that of BKV(MM) to be similar, though 5% smaller[3]. The DNA sequences of portions of BKV[4,6] and BKV(MM) genome have been reported[5,7] and the homology is between 70-90%.

BKV reproduces lytically in human fetal cells. It can transform hamster cells *in vitro* and produce tumors *in vivo* when injected into hamster[8,9]. One consequence of productive infection of human cells by BKV is the induction of tumor antigen (T antigen)[8]. T antigen is important in the initiation of viral DNA synthesis as well as in the induction and maintenance of transformation[10,11]. T antigens induced in virus-infected cells by BKV or SV40 are similar. Both are around 97,000 daltons in size[12]. The small t antigen from both SV40[13,14] and BKV[12] is 17,000 daltons. SV40 T antigen binds[21] around map positions

0.66-0.68 and presumably exerts its biological function in this region of the viral genome.

In SV40 DNA, transcription of early and late mRNA is initiated between map positions 0.65-0.70 in a region which does not code for any known protein. Palindromic sequences have also been found in this region especially at the origin of replication of SV40[18-20] and BKV DNA[4,7].

Though BKV and SV40 appear to be closely related, the extent of DNA homology between the genomes of BKV and SV40 is in question. It varies according to the hybridization technique used with an overall homology of 11-15%[3,15], 50%[16] and 85%[17].

In this communication, we report the DNA sequence of BKV(WT, prototype) and BKV(MM, variant) between map positions 0.53-0.71. The sequence includes palindromes and long repeats of the non-coding region, and includes the coding sequence of the small t antigen. Differences and similarities between the two strains of BKV as well as those between BKV and SV40 DNA are discussed.

MATERIALS AND METHODS

BKV DNA

BKV(prototype, WT) and BKV(MM strain) were both provided by P.M. Howley and K.K. Takemoto. Primary human embryonic kidney cells were purchased from Microbiological Associates (Long Island, N.Y.). The cells were infected with BKV at 0.001 to 0.01 PFU/cell. Purification of viral DNA was as described[22].

Cleavage of Virus DNA with Restriction Endonucleases

Restriction enzymes AluI, BamHI, EcoRI, HaeIII, HhaI, HindIII, HpaII, MboII, KpnI, MboI, PstI, SacI and XbaI were purchased from New England BioLabs. The standard cofactor mixture contained 10 mM Tris-hydrochloride (pH 7.5), 7 mM 2-mercaptoethanol, and 7 mM $MgCl_2$[22-24]. The cofactor mixture for HinIII and EcoRI also included 50 mM NaCl.

^{32}P-labeling of Restricted DNA Fragments

Labeling of DNA at the 5'-terminus with [γ-^{32}P]ATP and T4 polynucleotide kinase, or at the 3'-terminus with [α-^{32}P]ATP and reverse transcriptase (or the large fragment of E. coli DNA polymerase) is according to published procedures[25,26].

Physical Mapping

The details were as described previously[23,24]. Briefly, a terminally ^{32}P-labeled large BKV DNA fragment was digested with a single-cut enzyme and fractionated by gel electrophoresis to give two single-cut ^{32}P-labeled fragments. Each of these fragments was subjected to partial digestion with

the restriction enzyme whose cleavage map was to be determined to produce a series of incompletely digested fragments. The DNA fragments that share the same labeled end appeared as a homologous series of bands on the autoradiogram. The bands were then analyzed for size using DNA markers. The size difference between two bands is equivalent to the distance between two cleavage sites on the BKV genome. Since all partial digestion products share the same ^{32}P-labeled end, their order can be arranged in a unique way to produce a physical map.

Direct DNA Sequencing

The terminally ^{32}P-labeled DNA fragments were strand separated or recut with a second enzyme and then strand separated. The single-stranded end -labeled fragment is sequenced after partial chemical degradation as described by Maxam and Gilbert[27]. For fractionation of DNA fragments, we used 0.4-0.6 mm x 35 cm x 40-80 cm sized polyacrylamide gels (from 3.5% up to 20% gels depending on the length of nucleotide sequence desired, 1/20 to 1/29 cross-linking, and 8-9 M urea). Autoradiography was carried out at -20°C using Kodak XR-5 films and intensifying screens from Picker Corporation (Cleveland, Ohio).

RESULTS AND DISCUSSIONS

Comparison between BKV(WT) and BKV(MM) by physical mapping

A preliminary comparison between BKV(WT) and BKV(MM) was reported earlier in which physical mapping had been carried out using only two restriction enzymes[3], EcoRI and HindIII. Results showed that for the two genomes, the single EcoRI cleavage site as well as HindIII site 1 and site 3 were identical. This suggests that for both circular DNA genomes the region between map positions 0.72-0.18 (clockwise) may be similar. On the remainder of the genome (from HindIII site 1 to site 3, clockwise) an extra HindIII site on BKV(WT) DNA around map position 0.54 suggests that variation may lie within this region. In order to more precisely define the extent of this variation, we have analyzed both genomes by extensive physical mapping using various restriction enzymes.

A total of 13 restriction endonucleases (see Table 1) were used in this study. For BKV(WT) DNA, the 13 restriction enzymes generated a sum of 106 specific sites. However, the same enzymes created from 98 to 99 sites on BKV(MM) DNA, depending on the particular variants. So far we have identified four BKV(MM) variants from the same virus stock. They are designated BKV(MM)$_{a,b,c\&d}$. Genomic variation among BKV(MM)$_{a,b\&c}$ has been described

TABLE 1

Restriction Endonucleases Used for Comparing the Genomes of the BKV Variants

Restriction Endonuclease	No. of Cleavage Sites on Each Genome		
	BKV(WT)	BKV(MM)$_a$	BKV(MM)$_{b,c}$
AluI	33	30	30
BamHI	1	1	1
EcoRI	1	1	1
HaeIII	22	21	21
HhaI	1	1	1
HindIII	4	3	3
HpaII	0	1	0
KpnI	1	1	1
MboI	12	11	11
MboII	25	23	23
PstI	1	1	1
SacI	3	3	3
XbaI	2	2	2

previously[22,23]. BKV(MM)$_d$, a variant identified recently, is appreciably different from the other three.

The difference in electrophoretic pattern derived from each enzyme digestion was evident between the BKV(WT) and BKV(MM) DNAs (gel pattern not shown). For example, as shown in Figure 1, BKV(WT) DNA contains one extra MboI fragment of 270 base pairs located between MboI site 7 and site 7a. The MboI-A fragment from BKV(WT) DNA is slightly larger than that from BKV(MM) DNA (data not shown), while the remaining 10 MboI fragments are essentially identical. This indicates that over 90% of both genomes are likely to be the same. This conclusion is supported by observations using other restriction enzymes which showed that the sizes of different restricted DNA fragments (as both double- and single-stranded DNA) are the same between BKV(MM) and BKV(WT).

The order of these restriction fragments was determined by the cleavage of single-end ^{32}P-labeled large restriction fragments[22,23]. Comparison of the detailed physical map of BKV(WT) and BKV(MM) genome shows two regions,

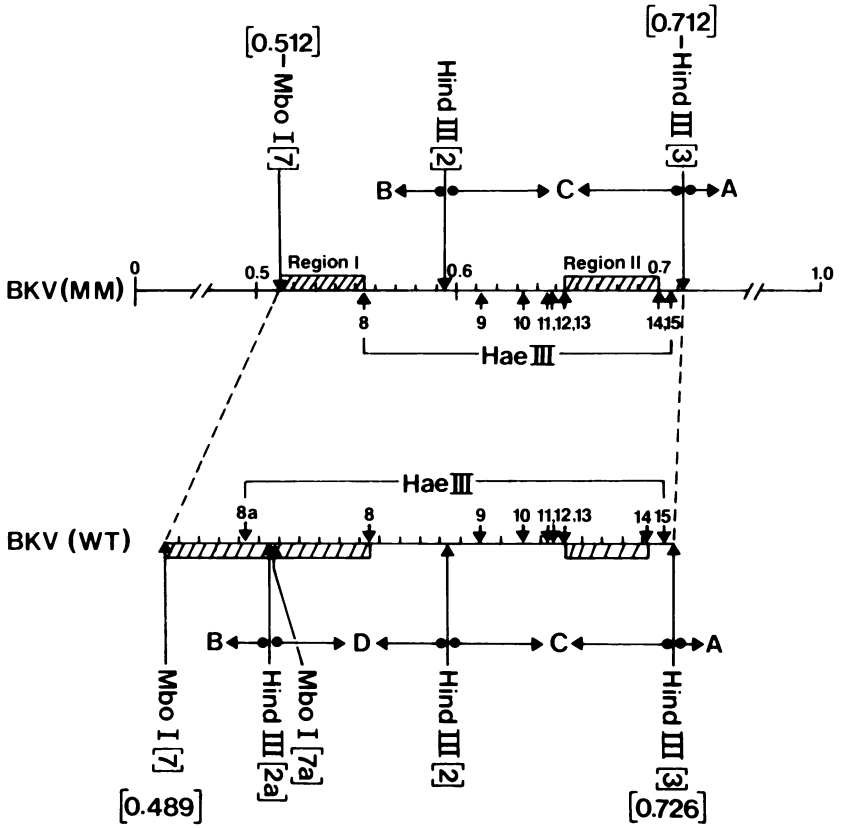


Figure 1. Differences between the genomes of BKV(WT) and BKV(MM)$_c$ as revealed by physical mapping. Extensive physical mapping by various restriction enzymes (see Table 1) of both BKV(WT) and BKV(MM)$_c$ DNA has been carried out. The physical maps include map positions 0.512-0.712 for BKV(MM)$_c$ DNA and 0.489-0.726 for BKV(WT) DNA where differences between the two genomes reside. Cleavage sites of each restriction enzyme are consecutively numbered starting clockwise (from left to right) after the unique EcoRI site at zero map position. The shaded regions between MboI site 7 and HaeIII site 8 (designated as region-I) as well as between HaeIII site 13 and site 14 (as region-II) show differences in the size of restriction fragments of the two genomes. Within these regions, BKV(WT) DNA has three more AluI cleavage sites than BKV(MM)$_c$ (result not shown). Differences between the two genomes were confirmed by direct nucleotide sequence analysis. The base sequence of these regions is shown in Figure 4.

which are different. The two regions of variation (Figure 1) include: region-I (between MboI site 7 and HaeIII site 8) which covers map positions

0.489-0.581 for BKV(WT) and 0.512-0.557 for BKV(MM); and region-II (between HaeIII site 13 and site 14) which covers map positions 0.665-0.714 for BKV(WT) and 0.642-0.701 for BKV(MM). In region-I, BKV(WT) DNA possesses two additional AluI sites, two MboII sites and one each of HaeIII, MboI and HindIII sites as compared to BKV(MM) DNA. Moreover, BKV(WT) DNA is 262 base pairs longer (due to an inserted fragment between HaeIII sites 8a and 8) than BKV(MM) DNA in this region. In region-II, BKV(WT) DNA is 29 base pairs shorter than BKV(MM). The former contains as extra AluI site. The remaining parts of both genomes are identical as determined by this technique and by direct nucleotide sequence analysis (to be described elsewhere).

Nucleotide sequence analysis

A plan for sequencing BKV(WT) DNA between MboII site 10 (map position 0.418) and MboI site 8 (map position 0.826) is represented in Figure 2. This segment includes the variable regions, the entire non-coding region, and a portion of the coding sequence for the early as well as the late proteins. A similar plan was made for sequencing BKV(MM) DNA at the corresponding region (not shown)[5,7]. Every terminally ^{32}P-labeled double-stranded DNA fragment was strand separated before being subjected to the sequencing procedure of Maxam and Gilbert[27]. Two typical sequencing gel patterns are shown in Figure 3. The sequence information from these gels and other gels not shown is summarized in Figures 4, 5 and 9.

Nucleotide sequence around 'region-II' of variation

The DNA sequences shown in Figure 4 are from BKV(MM)$_c$ between map positions 0.728 and 0.614, and BKV(WT) between 0.740 and 0.637. Both of these cover the entire 'region-II' and the neighboring homologous region. Identical sequences between the two strains of BKV are given as one sequence whereas differences are shown on separate lines. Major differences include a 25-base-long deletion in BKV(MM)$_c$ at position 3459, and a 53-base-long deletion in BKV(WT) at position 3523. In this region, 115 nucleotides are different between BKV(MM)$_c$ and BKV(WT) DNA.

BKV(MM)$_a$ DNA is a closely related variant of BKV(MM)$_c$, and is larger by a 26-base-long sequence, 5' A-A-C-A-T-G-T-C-T-G-T-C-T-G-G-C-T-C-G-T-T-T-C-C-G-G, inserted between nucleotides 3459-3458. A HpaII site (C-C-G-G) is found in BKV(MM)$_a$ DNA but not in BKV(MM)$_c$. BKV(MM)$_b$ DNA[23], estimated to be five base pairs larger than that of BKV(MM)$_c$ has not yet been sequenced. A new variant, designated as BKV(MM)$_d$, was recently identified from our BKV(MM) stock. The DNA sequence of BKV(MM)$_d$ in the similar region was analyzed and shown in Figure 5. This sequence is about 70 and 100 base pairs

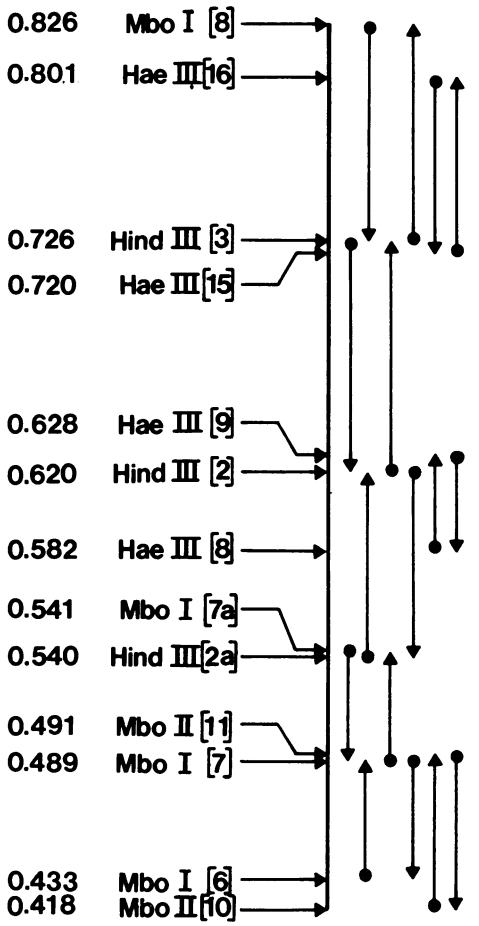
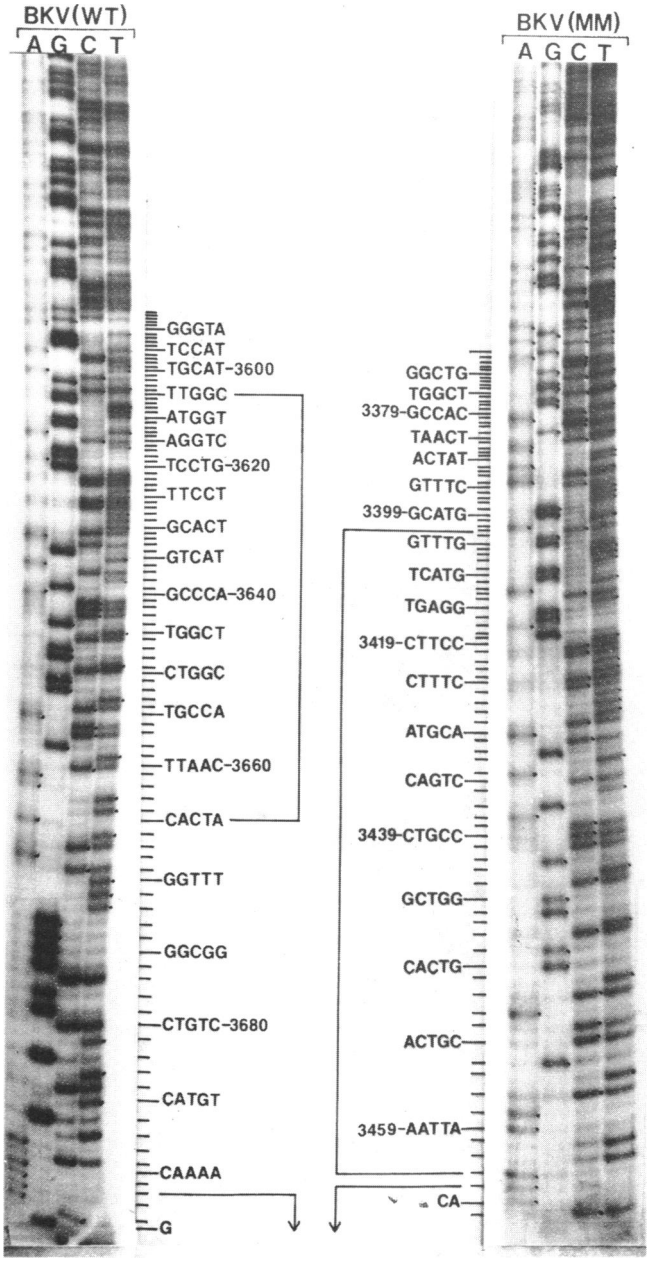


Figure 2. Specific restriction fragments of BKV(WT) DNA used for direct nucleotide sequence determination. Location of each restriction fragment is indicated both as map positions on the left-hand side and as an arrow plus a dot on the right-hand side. The dot indicates the single ^{32}P-labeled end (either 3' or 5' end), while the arrow shows the direction of nucleotide sequence analysis by the method of Maxam and Gilbert[27].

shorter than those of BKV(WT) and BKV(MM)$_c$, respectively.

Mapping of BKV(MM)$_d$ DNA using several restriction enzymes (e.g. HindIII, HpaI and EcoRI) showed that the physical map is similar to that reported for JC Virus[28], another human papovavirus. Whether the entire genome of JC Virus and BKV(MM)$_d$ is identical remains to be established.


BKV(WT)
A G C T
GGGTA
TCCAT
TGCAT-3600
TTGGC
ATGGT
AGGTC
TCCTG-3620
TTCCT
GCACT
GTCAT
GCCCA-3640
TGGCT
CTGGC
TGCCA
TTAAC-3660
CACTA
GGTTT
GGCGG
CTGTC-3680
CATGT
CAAAA
G
BKV(MM)
A G C T
GGCTG
TGGCT
3379-GCCAC
TAACT
ACTAT
GTTTC
3399-GCATG
GTTTG
TCATG
TGAGG
3419-CTTCC
CTTTC
ATGCA
CAGTC
3439-CTGCC
GCTGG
CACTG
ACTGC
3459-AATTA
CA

Sequence repetition is another unique feature of the DNA within 'region-II'. Stretches of tandem repeats 41-87 bases long are found in all BKV variants. As shown in Figure 6, BKV(MM)$_c$ DNA has three sets of a 61-base long sequence and two sets of a 87-base long sequence. The former sequence is a subset of the latter. Similarly, BKV(WT) DNA has three repeats of 44 and two repeats of 68-bases long; BKV(MM)$_d$ has two repeats of 41 bases long. Thus, though sequences among the BKV variants are related, the sizes and patterns of repeats are different. In contrast, between BKV and SV40 DNA there is almost no DNA sequence homology in this region, even though extensive repeated sequences are found in both genomes. This is somewhat surprising since the overall homology between BKV (human cell as host) and SV40 DNA (monkey cell as host) is 70% (Yang and Wu, manuscript in preparation). It is likely that the tandem repeats in this region are host rather than virus specific. Since virus-host interaction is inevitable, the region of direct repeats of the viral DNA could result from an unequal crossing over during recombination with the host DNA. Homology between a portion of a defective SV40 and its host cell has been reported[30,31]. Unequal crossing over has been suggested[32] and recently shown to be involved in the generation and maintenance of the tandemly repeated ribosomal RNA gene in yeast (Szostak and Wu, unpublished observation). Alternatively, since this part of the genome does not code for any known proteins, mutations may be more frequent in this region. Certain sequences are not required for biological activities of these viruses since viable deletion mu-

Figure 3. A typical gel pattern for direct nucleotide sequence determination showing differences in sequences between BKV(MM)$_c$ and BKV(WT) DNAs. The 5' ^{32}P-labeled HinIII-C duplex fragments derived from BKV(WT) and BKV(MM)$_c$ DNAs were strand separated. The resulting single-stranded single-end ^{32}P-labeled C$_a$ (the slow moving band) and C$_b$ (the fast band) were subjected to four specific chemical degradations[27]. Mixtures of the partially degraded DNA were then fractionated in 4-20% (0.4 or 0.6 mm x 35 cm x 40 or 80 cm, 8 M urea and 1/20 or 1/29 cross-linked) polyacrylamide gel[5]. The gel patterns shown here are from C$_a$ of both BKV(WT) (left-hand panel) and BKV(MM)$_c$ (right-hand panel) between HindIII site 2 and site 3 and labeled at site 3 (see Figures 1 and 2). This particular gel was 8% (1/20 cross-linking, 0.6 mm x 35 cm x 40 cm). The four lanes are indicated as follows: A lanes, the A- and C-specific cleavage (A>C); G lanes, the G-specific cleavage; C lanes, the C-specific reaction; T lanes, the T- and C-specific reaction (T>C). The sequences in homology between BKV(WT) and BKV(MM)$_c$ are marked out. For reading the sequence of the first nucleotide that carried the ^{32}P-label up to about 40 nucleotides, a 20% gel of 1/29 cross-linking was used (gel patterns not shown).

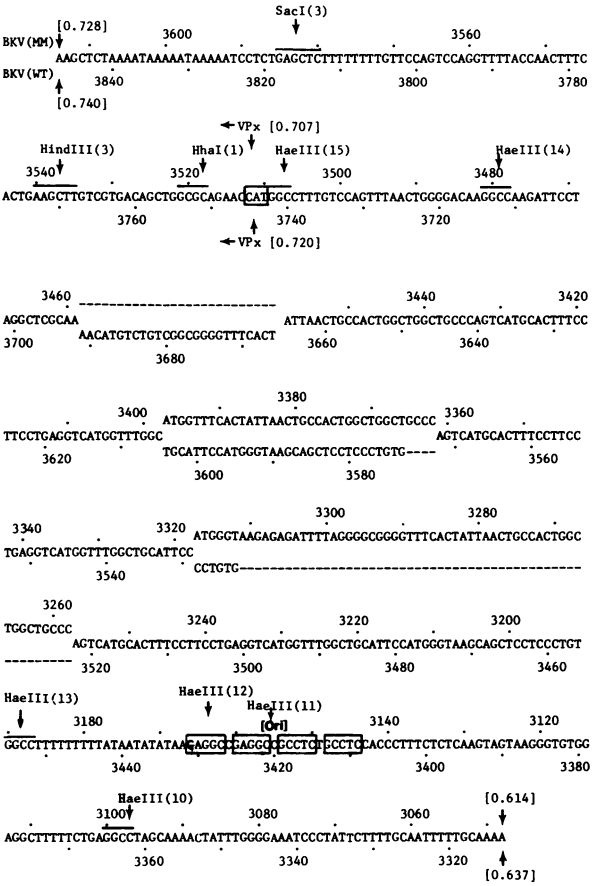


Figure 4. Nucleotide sequences of BKV(WT) and BKV(MM)$_c$ DNA in region-II. The second base from the 5' end of the unique EcoRI hexanucleotide recognition sequence is taken as nucleotide number 1. From this point on the nucleotides are numbered consecutively and clockwise along the circular BKV genome. The entire genomes of BKV (MM)$_c$ and BKV(WT) contain 4963 and 5196 uncleotides, respectively (details will be published elsewhere). The sequences shown here correspond to map positions 0.728-0.614 on BKV(MM)$_c$ DNA and 0.740-0.637 on BKV(WT) DNA. The top sequence represents that of BKV(MM)$_c$ and the bottom one BKV(WT). A single sequence is given in regions where the sequences are identical in the two strains of BKV. The dashed lines indicate deletions in one of the strains.
The sequence coding for a putative 66-amino acid long protein, VPx, initiates at nucleotides 3510-3512 for BKV(MM)$_c$ DNA and 3743-3745 for BKV(WT) DNA on the complementary strand (not shown). The potential origin of DNA replication (Ori) is boxed in. Blocks of tandemly repeated sequences are found between N-terminus of VPx and the origin of replication (see Figure 6).

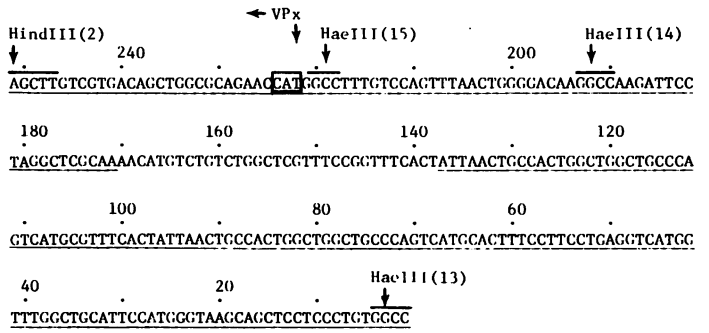


Figure 5. Nucleotide sequence of BKV(MM)$_d$ DNA in 'region-II'. HaeIII site 14 and site 13 are indicated, within which variable region-II resides. The sequences identical to that of BKV(MM)$_c$ are underlined. The 41-base long repeated sequences are marked with brackets. Numbering of the nucleotides in this sequence is arbitrary. The sequence between nucleotides 251-1 shown here corresponds to that in between nucleotides 3539-3187 of BKV(MM)$_c$ (see Figure 4). The latter has an additional 102 bases within this region.

tants of SV40 have been isolated which lack some repeated sequences between 0.68-0.74 map positions[33].

Palindromic sequences in a region between 'region-II' and the T-antigen gene

Viral DNA such as SV40 and BKV is replicated in eucaryotic cells using host cell enzymes[10,11]. It is likely that the recognition sequences used for the initiation of viral DNA replication or transcription may resemble those used within the host cells.

As shown in Figure 7, there are several palindromes (P), perfect palindromes (PP) and a true palindrome (TP) in both SV40[18] and BKV DNA. In addition to similarity in the locations of these palindromes, considerable sequence homology (around 60%) is found between the two viruses. Among these structures a 27-base-long palindrome for SV40[18] and a corresponding 23-base-long sequence for BKV(WT)[4] have been considered as the origin of DNA replication (Ori). We found the 23-base-long palindrome to be identical in BKV(MM)$_a$, BKV(MM)$_b$, BKV(MM)$_c$, BKV(MM)$_d$ and BKV(WT), which suggests that the conservation of this sequence is important for the virus. This palindromic sequence is characterized by the G-C rich sequence, 5' G-C-C-T-C, which is repeated twice on the upper strand and twice on the lower strand as shown in the lower part of Figure 7. Since this sequence contains inverted repeats, each strand can be drawn as a hairpin structure. An alternative

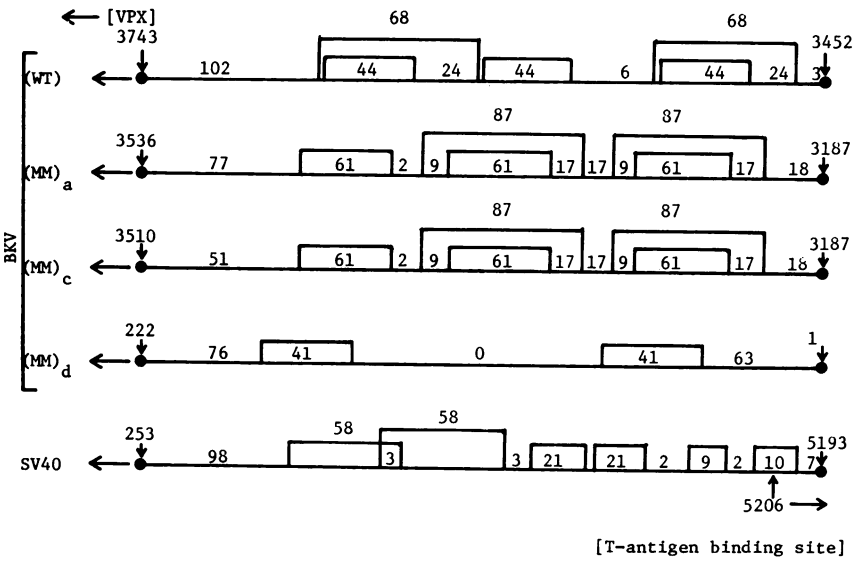


Figure 6. Schematic representation of the non-coding region of the genomes of BKV variants and SV40. The genomes of four BKV strains (human cell as host) and SV40-strain 776 (monkey cell as host) are compared. The exact location of this region is indicated by nucleotide numbers. Numbering of nucleotides in the BKV sequences (including MM_a, MM_c and WT) is as described in Figure 4. The base numbers used for the $BKV(MM)_d$ sequence is arbitrary (see Figure 5). The sequence of SV40 DNA is numbered according to Reddy et al[19]. The sequences in tandem repeats are boxed in. The numbers inside, on top or adjacent to the boxes indicate the number of nucleotides involved. SV40 T-antigen binding site corresponding to nucleotides 5093-5206 is indicated partly in this figure and mostly in Figure 7.

arrangement of this sequence is shown in Figure 8 in which the double-stranded structure folds back to form a four-stranded structure. This structure may be stabilized by an additional hydrogen bond contributed by each of seven G-C pairs. This unusual structure may be specifically recognized by the host cell polymerase for the initiation of viral DNA replication.

There is a 19-base-long palindromic sequence composed of a 12 and a 6 base-long perfect palindromes immediately before the initiation codon for BKV(MM) large-T and small-t antigens. Similarly, a 22-base-long palindromic structure is found in SV40[18] (Figure 7). Each of these palindromes can be drawn as a long hairpin structure adjacent to the initiation codon[5] (ATG)

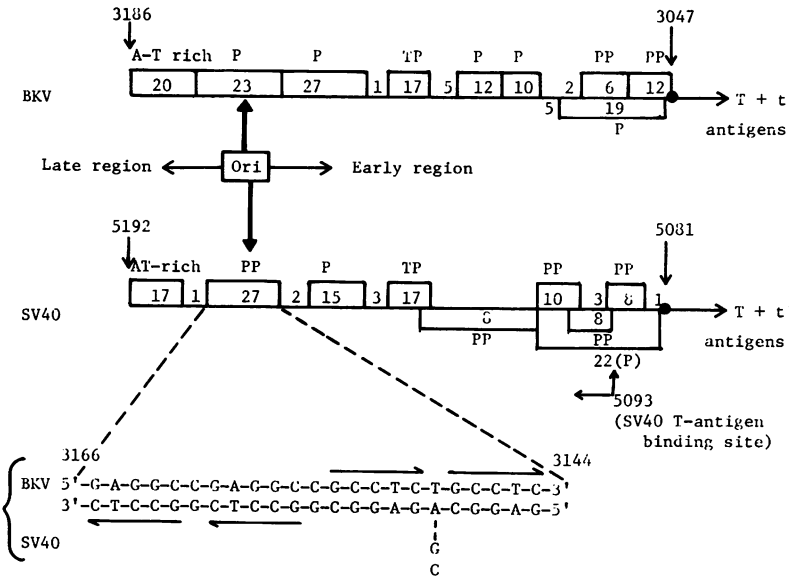


Figure 7. Schematic representation of early non-coding region of the genomes of BKV and SV40. The top structure reprresents the DNA sequence of BKV(WT, MM_a, MM_c and MM_d, all with the same sequence in this region), while the lower structure represents the SV40 sequence. BKV(MM)$_c$ nucleotide numbers are used for the BKV sequence in comparing with the SV40 sequence whose numbers are adopted from Reddy et al [19]. The location of the origin of DNA replication (Ori) as well as T- and t-antigen genes are indicated. Abbreviations placed near the boxes which represent the palindromic sequences are as follows: P, palindrome (sequence with a two-fold axis of symmetry); PP, perfect palindrome; TP, true palindrome (word palindrome on the same strand). A-T rich clusters are also indicated. DNA sequence homology in this region is approximately 60% between BKV and SV40. A typical palindromic structure for both BKV and SV40 at or around the origin of replication is represented at the bottom. Four sets of the sequence 5' G-C-C-T-C, shown as tandem and inverted repeats, are marked with arrows.

and having an 8 base-paired stem. This strucutre may serve as a recognition signal for the synthesis or translation of the T antigen mRNA.

It has been shown that SV40 t antigen binds specifically to this region[21]. Replication of the viral DNA appears to depend on this specific binding; the same is probably true in BKV.

Nucleotide sequence of 'region-I' and the t-antigen gene

The DNA sequence from map positions 0.614 to 0.534 in BKV(MM) and from 0.637-0.510 in BKV(WT) is shown in Figure 9. This sequence includes 'region-

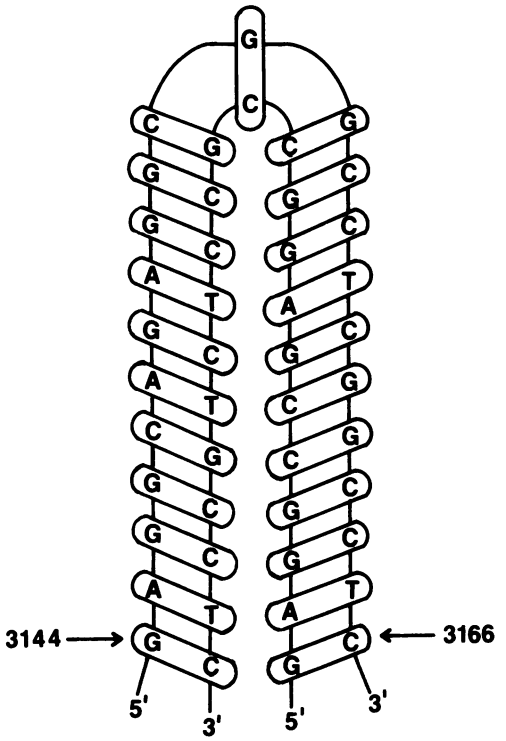


Figure 8. A proposed four-stranded structure of the BKV DNA at the origin of replication (Ori). The double-stranded 23-base-pair-long sequence at the region of BKV DNA replication (shown in Figure 7) is written as a four-stranded structure or has been proposed for the polyoma 'Ori' sequences [25]. This structure has 10 base-pairs on each side and three on top as a loop. The additional hydrogen bonds between G-C pairs in this structure is indicated by dotted lines. Since each complete turn of the double helix contains 10 base pairs, this four-stranded structure may exist as a compact structure.

I' which is different in the two genomes. Sequence analysis revealed a deletion in BKV(MM) of a 262-base-long sequence between nucleotides 2745 and 3006 of BKV(WT). This deletion results in the shortening of the putative t antigen from 172 amino acids in BKV(WT) to 100 amino acids in $BKV(MM)_c$, as shown in Figure 10. Both BKV(WT) and BKV(MM) are able to transform hamster cells[9]. Therefore, it must be that either a shortened t antigen in $BKV(MM)_c$ can serve the same biological function as in BKV(WT), or splicing of small t-antigen mRNA can occur to produce a longer t antigen in BKV(MM).

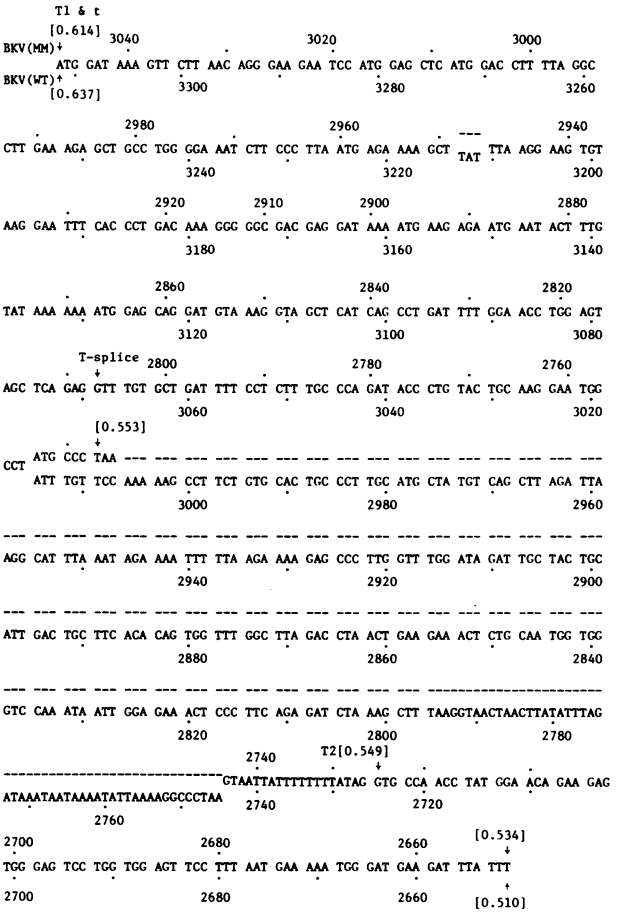


Figure 9. Nucleotide sequences of 'region-I' and t antigen gene in BKV(MM) and BKV(WT). The DNA sequences shown here which correspond to map positions 0.614-0.534 on BKV(MM) DNA and 0.637-0.510 on BKV(WT) DNA have the same polarity as the mRNA sequences derived from this part of the genomes. The top sequence of each set is specific for BKV(MM) and the bottom sequence for BKV(WT). Where no variation occurs, a single sequence is given. Initiation points for the synthesis of t, T1 (the first part of T) and T2 (the second part of T) antigen are indicated. Potential splicing points (involving A-G-G-T sequence) of T1 and T2 are shown by arrows and horizontal lines on top of the sequences. 'Region-I' starts from HaeIII site 8 downstream to MboI site 7 (not shown) which is located 109 bases beyond the end of this sequence. The BKV(WT) sequence shown here has been compated to that published recently by Dhar et al [6]. The result of the comparison is documented in Table 2.

The coding sequence for the first 98 amino acids of t antigen in BKV(MM) is identical to that in BKV(WT) except that nucleotides 3211-3213 of BKV(WT) are deleted in BKV(MM) DNA. The predicted amino acid sequences of the t antigen in BKV(MM) and BKV(WT) are shown in Figure 10. Minor discrepancies between our data and those from Dhar et al.[6] for the coding sequence of the t antigen and the derived amino acid sequence in BKV(WT) are shown in Table 2. Some of these differencs may have arisen from mutations in BKV(WT) during repeated passages over the past four years.

```
(I)

BKV(MM)                                    10
       Met Asp Lys Val Leu Asn Arg Glu Glu Ser Met Glu Leu Met Asp Leu Leu Gly Leu
BKV(WT)                                    10


20                                         30                  ---
Glu Arg Ala Ala Trp Gly Asn Leu Pro Leu Met Arg Lys Ala     Leu Arg Lys Cys Lys
20                                         30                  Tyr

    40                                         50
Glu Phe His Pro Asp Lys Gly Gly Asp Glu Asp Lys Met Lys Arg Met Asn Thr Leu Tyr
40                                         50


    60                                         70
Lys Lys Met Glu Gln Asp Val Lys Val Ala His Gln Pro Asp Phe Gly Thr Trp Ser Ser
60                                         70

    80                                         90
Ser Glu Val Cys Ala Asp Phe Pro Leu Cys Pro Asp Thr Leu Tyr Cys Lys Glu Trp Pro
80                                          90

    100
Met Pro --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Ile Cys Ser Lys Lys Pro Ser Val His Cys Pro Cys Met Leu Cys Gln Leu Arg Leu Arg
100                                        110

--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
His Leu Asn Arg Lys Phe Leu Arg Lys Glu Pro Leu Val Trp Ile Asp Cys Tyr Cys Ile
120                                        130


--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Asp Cys Phe Thr Gln Trp Phe Gly Leu Asp Leu Thr Glu Glu Thr Leu Gln Trp Trp Val
140                                        150


--- --- --- --- --- --- --- --- --- --- --- --- ---
Gln Ile Ile Gly Glu Thr Pro Phe Arg Asp Leu Lys Leu
160                                        170

(II)

BKV(MM)                                    10
       Val Pro Thr Tyr Gly Thr Glu Glu Trp Glu Ser Trp Trp Ser Ser Phe Asn Glu Lys
BKV(WT)                                    10

20
Trp Asp Glu Asp Leu Phe
20
```

Figure 10. Predicted amino acid sequences of small t and part of large T antigens of BKV(MM) and BKV(WT). The amino acid sequences shown here correspond to the nucleotide sequences of Figure 9. Part (I) gives the entire t-antigen of BKV(MM) and the first part of T-antigen of BKV(MM) and BKV(WT). Part (II) gives the beginning sequence of the second part of T antigen that is a unique sequence for both strains of BKV.

TABLE 2

Differences in the Nucleotide and the Amino Acid Sequences between Two Strains of BKV(WT)

Location of Nucleotides (amino acids)	Data* of Yang et al.†	Dhar et al.§
3165-3163 (50)	GA<u>T</u> (Asp)	GA<u>C</u> (Asp)
3015-3013 (100)	<u>ATT</u> (<u>Ile</u>)	<u>GAG</u> (<u>Glu</u>)
2952-2950 (121)	T<u>TA</u> (<u>Leu</u>)	T<u>CA</u> (<u>Ser</u>)

* The nucleotides and amino acids which are different from the two sources are underlined.

† Figures 9 and 10 of this paper.

§ Data report by Dhar et al.[6]

Sequence comparison between BKV(MM), BKV(WT) and SV40

Of the 1181 nucleotides in BKV(WT) and the 964 in BKV(MM)$_c$ presented in this communication, only 385 nucleotides are different. The majority of the differences are due to delections in BKV(WT) or BKV(MM), since only 34 nucleotides are attributable to direct base changes. It may be concluded that between BKV(MM)$_c$ and BKV(WT) DNA, point mutations in this region can account for 34 base changes and a minimum of 3 deletion events can account for the remaining differences. DNA sequence analysis of another 1000 nucleotides in the coding regions (data not shown) revealed no differences. Thus, there appears to be a strong evolutionary pressure to resist mutations in the coding regions.

ACKNOWLEDGEMENTS

This work was supported by research grants CA 14989 and GM 24904 from the National Institute of Health, U.S. Public Health Service. We thank A. Young and S. Hanson for valuable assistance.

REFERENCES

1. Gardner, S.D., Field, A.M., Coleman, D.V. and Hulme, B. (1971) Lancet 1, 1253-1257.
2. Takemoto, K.K., Rabson, A.S., Mullarkey, M.F., Blaese, R., Garon, C.F. and Nelson, D. (1974) J. Natl. Cancer Inst. 53, 1205-1207.

3. Howley, D.M., Khoury, G., Bryne, J.C., Takemoto, K.K. and Martin, M.A. (1975) Virology 16, 959-973.
4. Dhar, R., Lai, C. and Khoury, G. (1978) Cell 13, 345-358.
5. Yang, R.C.A. and Wu, R. (1979) Virol. 92, 340-352.
6. Dhar, R., Seif, I. and Khoury, G. (1979) Proc. Natl. Acad. Sci. USA 76, 565-569.
7. Yang, R.C.A. and Wu, R. (1979) Proc. Natl. Acad. Sci. USA 76, 1179-1183.
8. Shah, K.V., Daniel, R.W. and Warszawski, R.M. (1973) J. Infect. Dis. 128 784-787.
9. Major, E.O. and di Mayorca, G. (1973) Proc. Natl. Acad. Sci. USA, 70, 3210-3212.
10. Fareed, G.C. and Davoli, D. (1977) Ann. Rev. Biochem. 46, 471-522.
11. Kelly, T.K. and Nathans, D. (1977) Adv. Virus Res. 21, 85-173.
12. Simmons, D.T., Takemoto, K.K. and Martin, M.A. (1977) J. Virol. 24, 319-325; Proc. Natl. Acad. Sci. USA (1978) 75, 1131-1135.
13. Prives, C., Gilboa, E., Revel, M. and Winocour, E. (1977) Proc. Natl. Acad. Sci. USA 74, 457-461.
14. Paucha, E., Mellor, A., Harvey, R., Smith, A., Hewick, R. and Waterfield, M. (1978) Proc. Natl. Acad. Sci. USA 75, 2165-2169.
15. Wold, W.S.M., Mackey, J.K., Brackman, K.H., Takemori, N., Rigden, J. and Green (1978) Proc. Natl. Acad. Sci. USA 75, 454-458.
16. Osborn, J.E., Robertson, S.M., Pagett, B.L., Walker, D.L. and Weisblum, B. (1976) J. Virol. 19, 675-684.
17. Newell, N., Lai, C-J., Khoury, G. and Kelly, T.J., Jr. (1978) J. Virol. 25, 193-201.
18. Subramanian, K.N., Reddy, V.B. and Weissman, S.M. (1977) Cell 10, 497-507.
19. Reddy, V.B., Thimmappaya, B., Dhar, R., Subramanian, K.N., Zain, B.S., Pan, J., Ghosh, P.K., Celma, M.L. and Weissman, S.M. (1978) Science 200, 494-502.
20. Fiers, W., Contreras, R., Haageman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herrewoghe, J., Volckaert, G. and Ysebaert, M. (1978) Nature 273, 113-120.
21. Tjian, R. (1978) Cell 13, 165-179.
22. Yang, R.C.A. and Wu, R. (1978) Proc. Natl. Acad. Sci. USA 75, 2150-2154.
23. Yang, R.C.A. and Wu, R. (1978) J. Virol. 27, 700-712.
24. Yang, R.C.A. and Wu, R. (1978) J. Virol. 28, 851-864.
25. Wu, R., Jay, E. and Roychoudhury, R. (1976) Methods in Cancer Res. 12, 87-176.
26. Bahl, C.P., Wu, R., Stawinsky, J. and Narang, S.A. (1977) Proc. Natl. Acad. Sci. USA 74, 966-970.
27. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
28. Martin, J.D., Frisque, R.J., Padgett, B.L. and Walker, D.L. (1979) J. Virol. 29, 846-855.
29. Soeda, E., Arrand, J.R., Smolar, N. and Griffin, B.E. (1979) Manunscript in press.
30. Lavi, S. and Winocour, E. (1972) J. Virol. 9, 309-316.
31. Rosenberg, M., Segal, S., Kuff, E.L. and Singer, M.F. (1977) Cell 11, 845-858.
32. Smith, G.P. (1973) Cold Spring Harbor Symp. Quant. Biol. 38, 507-513.
33. Shenk, T.E., Carbon, J. and Berg, D. (1976) J. Virol. 18, 664-671.