JOURNAL OF VIROLOGY, May 1980, p. 416-430
0022-538X/80/05-0416/15$02.00/0

Vol. 34, No. 2

BK Virus DNA Sequence Coding for the t and T Antigens and Evaluation of Methods for Determining Sequence Homology

ROBERT C. A. YANG, ALEX YOUNG, AND RAY WU*

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York 14853

The DNA sequence of the early region of the human papovavirus BK (MM strain) was determined. A potential initiation signal for translation is located at nucleotides 3,047 to 3,045 or map position 0.614. Extending counterclockwise from this AUG signal there is only one open reading frame, which can code for a putative t antigen of 100 amino acids in length. If the early mRNA of BKV is spliced, then the regions between nucleotides 3,047 to 2,808 and 2,725 to 884 can code for a T antigen 694 amino acids in length. The sequences of the deduced T antigens in BK virus share 71% amino acid homology with those in simian virus 40, whereas the coding sequences of the two viruses share 70% DNA homology. Comparison of DNA sequences and evaluation of homology measurements between these two viruses are discussed.

The human papovavirus, BK virus (BKV), was first isolated by Gardner et al. (11) from the urine of an immunosuppressed renal allograft recipient. A variant of this virus, BKV(MM), was later isolated from the urine and brain tumor of a patient with Wiskott-Aldrich syndrome (40). In addition to its ability to reproduce lytically in human fetal cells, BKV can transform normal hamster cells and human embryonic kidney cells in vitro and produce tumors in vivo when injected into hamsters (20, 29, 31, 35, 42). One consequence of lytic infection or transformation of cells by BKV is the induction of nuclear tumor antigen (35). Tumor antigen (T antigen) is an important factor in the initiation of viral DNA synthesis as well as in the induction and maintenance of transformation (8, 14, 18). Tumor antigens induced in virus-infected cells by BKV or simian virus 40 (SV40) are antigenically related (35, 39). On the basis of ion-exchange chromatography, seven pairs out of 21 BKV and 20 SV40 tryptic polypeptides were shown to be identical (37).

Cells infected or transformed by SV40 or BKV synthesize two forms of tumor antigens, large T and small t (4, 6, 9, 28, 30, 32, 36, 51, 54). These T antigens (95,000 and 17,000 daltons, respectively) are coded for by two different early mRNA molecules, which are transcribed counterclockwise from about map positions 0.65 to 0.17 (9, 32). These viral mRNA's differ in their size and splicing pattern (1, 2, 9, 12, 32). Thus, in SV40 and probably in BKV, and T antigen is encoded by two noncontiguous segments of DNA in the early region. Furthermore, large T and small t antigens share common amino-terminal sequences (6, 9, 28, 32, 54). SV40 and BKV antigens are closely related not only by size but also by function, since BKV can complement the early mutant *ts*A58 of SV40 (22).

The reported values for the DNA sequence homology between the genomes of SV40 and BKV vary greatly according to the hybridization technique used. Under stringent hybridization conditions, an overall sequence homology of 11 to 20% has been reported (17, 19, 45). Under less stringent conditions, a value of up to 50% homology was obtained (26). A value of 85% homology was estimated by electron microscopic visualization of heteroduplexes formed at different effective temperatures (25). An improved nitrocellulose filter hybridization technique at different formamide concentrations also showed very extensive homology (16). Direct DNA sequence analysis of portions of the BKV(WT) and BKV(MM) genome (5, 50, 52, 54) and, recently, the entire genome (53) showed 60 to 70% homology with SV40 DNA. This communication reports the complete BKV(MM) sequence of the early region and the predicted amino acid sequences for both the small t and the large T antigens.

MATERIALS AND METHODS

Cells and viruses. Human embryonic kidney cells were purchased from Microbiological Associates (Long Island, N.Y.). The cells were propagated two or three times in our laboratory in Dulbecco-modified Eagle medium supplemented with 2 g of glucose per liter, 10% fetal calf serum, and 10% tryptose phosphate broth. Plaque-purified BKV (prototype or wild type, WT) and BKV (strain MM) were provided by P. W.

Howley and K. K. Takemoto. For preparation of BKV DNA, cells were infected with approximately 0.1 PFU of virus per cell (17). Virus stocks were made by infecting the cells at a multiplicity of 0.001 to 0.01 PFU/cell.

Cleavage of BKV DNA with restriction endonuclease. Detailed procedures of endonuclease cleavage have been described previously (48, 49). Briefly, BKV DNA was purified from infected cells by Hirt extraction (15) and subsequently by isopycnic centrifugation in cesium chloride-ethidium bromide. The superhelical (form I) BKV DNA thus obtained was cleaved with restriction enzymes.

Restriction endonucleases *Xba*I, *Pst*I, *Sac*I, *Hin*dIII, *Mbo*I, *Hae*III, and *Mbo*II were purchased from New England BioLabs (Beverly, Mass.). The standard cofactor mixture for all these restriction enzymes contained 10 mM Tris-hydrochloride (pH 7.5), 7 mM 2-mercaptoethanol, and 7 mM $MgCl_2$ (49). For complete digestion, 0.3 U of enzyme per μg of DNA was used. Incubation took place at 37°C for 3 to 15 h.

^{32}P labeling of restricted BKV DNA fragments. Labeling of DNA (following dephosphorylation by bacterial alkaline phosphatase) at the 5′ terminus with [γ-^{32}P]ATP and T4 polynucleotide kinase, or at the 3′ terminus with α-^{32}P-labeled deoxyribonucleoside triphosphates and reverse transcriptase (or the Klenow fragment of *Escherichia coli* DNA polymerase), wяs according to published procedures (46, 54).

[γ-^{32}P]ATP (specific activity, 3,000 Ci/mmol) and α-^{32}P-labeled deoxyribonucleoside triphosphates (specific activity, 400 Ci/mmol) were obtained from Amersham Corp. (Arlington Heights, Ill.). Bacterial alkaline phosphatase was purchased from Worthington Biochemical Corp. (Freehold, N.J.). T4 polynucleotide kinase was purchased from New England BioLabs. Klenow fragment of *E. coli* DNA was obtained from Boehringer Mannheim GmbH (West Germany). Reverse transcriptase was kindly supplied by the Office of Program Resources and Logistics, Viral Cancer Program, National Institutes of Health, Bethesda, Md.

Direct DNA sequencing. To obtain ^{32}P label at only one end, the duplex DNA fragments labeled at 5′ or 3′ ends were either strand separated on polyacrylamide gels (23), or cleaved first with an appropriate restriction enzyme and then strand separated. Strand separations have the advantage of eliminating nick-labeled ends that occasionally arise in the polynucleotide kinase-catalyzed 5′-end-labeled DNA fragments of more than 300 bases long. The single-stranded end-labeled fragments were subjected to direct chemical degradation procedures as described by Maxam and Gilbert (23). The four specific degradation mixtures (see legend of Fig. 2) were fractionated on denaturing polyacrylamide gels. For reading from the first nucleotide that bears the ^{32}P label up to nucleotide 40, we used gels of 20% polyacrylamide (and 0.66% methylene bisacrylamide) (0.4 to 0.6 mm by 35 cm by 40 cm) in 7 M urea. For obtaining sequence from nucleotide 30 upwards for approximately 600 bases, we used gels of 3.5% to 8% polyacrylamide (and 1/20 as much methylene bisacrylamide) (0.4 to 0.6 mm by 35 cm by 80 cm) in 8 to 9 M urea. Autoradiography was carried out at −20°C using Kodak XR-5 films and an intensifying screen from Picker Corp. (Cleveland, Ohio).

RESULTS AND DISCUSSION

The general approach used in obtaining the nucleotide sequence was, first, to cleave the BKV DNA into a number of specific fragments of workable sizes using various restriction endonucleases, and then to order these DNA fragments on the circular genome to construct a detailed physical map. For sequence analysis, the fragments were labeled with ^{32}P at either 3′ or 5′ ends (as described in Materials and Methods) and purified by gel electrophoresis. The terminally labeled fragments were either strand separated, or cleaved with a second restriction enzyme, followed by strand separation. These procedures result in defined fragments, each of which is single-end labeled. Sequence analysis was carried out by the direct chemical degradation procedures (23), using thin (33) and long (54) slab gels (0.4 to 0.6 mm by 80 cm by 35 cm) of low-percent (3.5 to 8%) polyacrylamide for obtaining maximum amounts of sequence information.

A detailed physical map of BKV DNA has been constructed recently (48, 49). It contains a total of approximately 100 cleavage sites, which were derived from 13 restriction enzymes. The location (by nucleotide number) of each site is summarized in Table 1. With cleavage by these enzymes, all of the restricted DNA fragments are shorter than 300 nucleotide pairs and thus are of a size to permit complete sequence analysis by the chemical method (23).

A general plan for sequencing the entire early gene region, between map positions 0.15 and 0.65, is represented in Fig. 1. The locations of restriction enzyme sites are given on the top portion of this figure. The specific DNA fragments designated *a* through *x*, which were terminally labeled and sequenced, are indicated in the middle part of the figure. Further information on these fragments is provided in detail in Table 2. Some other fragments (not shown) were also used for sequence analysis. The coding regions for t antigen and T antigen are depicted, including the direction of transcription.

Two sequencing gel patterns derived from fragments *g* and *h* (see Fig. 1 and Table 2) are shown in Fig. 2a and b, respectively, as typical examples. Each gel pattern gives the nucleotide sequence derived from a single-end-labeled *Mbo*I-C DNA fragment. The 40-cm-long 8% polyacrylamide gel (see the legend to Fig. 2) was capable of resolving approximately 280 nucleotides with four loadings of the partially digested DNA sample. An 80-cm-long 4% polyacrylamide gel can resolve up to 600 nucleotides with three loadings. The sequence data obtained from this and other gels are summarized in Fig. 3 and 4,

TABLE 1. *Location of the cleavage sites of 17 restriction endonucleases on the BKV(MM) DNA*[a]

Endonuclease	Cleavage site no.	Base no.	Endonuclease	Cleavage site no.	Base no.	Endonuclease	Cleavage site no.	Base no.
*Alu*I (AGCT)	1	442		13(T)	4,347	*Kpn*I (GGTACC)	1	4,457
	2	490		14	4,589	*Mbo*I (GATC)	1	356
	3	511		15	4,716		2	586
	4	542	*Hae*III (GGCC)	1	307		3	603
	5	873		2	477		4	1,035
	6	1,205		3	527		5	1,564
	7	1,346		4	566		6	2,245
	8	1,523		5	903		7	2,539
	9	1,571		6	1,212		8	4,061
	10	1,581		7	1,317		9	4,190
	11	1,655		8	2,755		10	4,406
	12	1,829		9	2,993		11	4,857
	13	1,925		10	3,097	*Mbo*II GAAGA TCTTC	1(G)	710
	14	2,319		11	3,155			
	15	2,459		12	3,161		2	972
	16	2,516		13	3,187		3(G)	1,005
	17	2,813		14	3,478		4	1,050
	18	2,844		15	3,506		5	1,068
	19	2,949		16	3,929		6	1,542
	20	2,982		17	4,082		7	1,703
	21	3,010		18	4,422		8	1,887
	22	3,200		19	4,475		9	2,172
	23	3,523		20	4,643		10	2,181
	24	3,536		21	4,688		11	2,226
	25	3,581	*Hha*I (GCGC)	1	3,518		12	2,562
	26	3,610	*Hin*dIII (AAGCTT)	1	872		13	2,619
	27	3,919		2	2,948		14	2,640
	28	4,444		3	3,535		15	2,658
	29	4,724	*Hin*fI (GANTC)	1(A)	613		16	2,703
	30	4,891		2(A)	698		17	2,891
*Bam*HI (GGATCC)	1	4,857		3(A)	969		18(G)	2,966
*Bgl*II (AGATCT)	1	602		4(T)	1,023		19	3,022
*Eco*RI (GAATTC)	1	4,963		5(G)	1,047		20(G)	3,636
*Eco*RII (CC$\genfrac{}{}{0pt}{}{A}{T}$GG)	1	283		6(A)	1,062		21(G)	4,293
	2	703		7(A)	1,149		22(G)	4,511
	3	746		8(T)	1,518		23(G)	4,682
	4(T)	1,392		9(A)	1,755	*Pst*I (CTGCAG)	1	1,651
	5(T)	1,892		10(A)	1,868	*Pvu*II (CAGCTG)	1	510
	6	2,690		11(A)	2,604		2	3,522
	7	2,976		12(C)	2,694	*Sac*I (GAGCTC)	1	1,828
	8(T)	3,559		13(T)	3,019		2	3,009
	9	3,738		14(C)	3,690		3	3,580
	10(T)	3,968		15(C)	3,702	*Xba*I (TCTAGA)	1	1,178
	11(T)	4,151		16(C)	4,316		2	4,790
	12	4,231						

[a] The cleavage sites of each enzyme are consecutively numbered clockwise after the single *Eco*RI site. Numbering of nucleotides along the entire BKV sequence is as described (see the legend of Fig. 3). Location of each enzyme site is represented by the first nucleotide number from the 5′ end of the enzyme recognition sequence. Each specific restriction sequence (5′ to 3′ and from left to right) is indicated with each enzyme in this table. The nucleotide locations of the vast majority of these sites on the BKV(WT) DNA (see references 10 and 51) are similar to those in the BKV(MM) DNA.

together with the predicted protein sequences. In most regions, the DNA sequence was confirmed by determining the sequence of the complementary strand (see Table 2). Furthermore, approximately 90% of the sequence data were obtained from experiments which have been repeated.

BKV DNAs of various strains, such as BKV(WT), BKV(MM)$_a$, BKV(MM)$_b$, BKV(MM)$_c$, and BKV(MM)$_d$, were also extensively analyzed by both restriction enzyme mapping and nucleotide sequence determination (51). We found that genomic heterogeneity (less than 10% of the BKV genome) is mainly located within the noncoding region (map positions 0.62 to 0.72), whereas the sequence of the majority of the coding region is conserved. Particularly, the early coding region (map positions 0.17 to 0.62) is essentially identical for BKV(MM)$_a$, BKV(MM)$_b$, and BKV(MM)$_c$ and is presented in this communication.

BKV coding sequence for small t antigen. All the nucleotides along the BKV DNA sequence were consecutively numbered starting from the unique *Eco*RI endonuclease recognition site (see the legend of Fig. 3 for details).

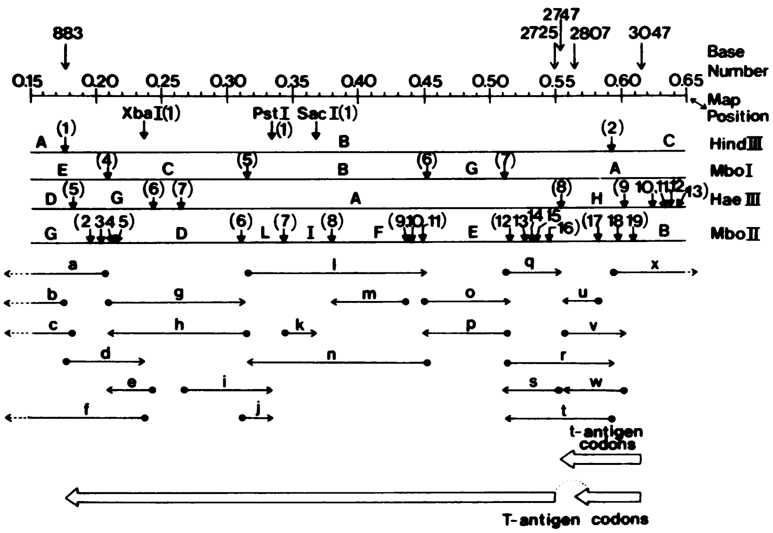

FIG. 1. *Restriction enzyme cleavage sites on BKV(MM) DNA between map positions 0.15 and 0.65, and specific DNA fragments used for sequence analysis. The restriction sites, indicated with downward short arrow, of each enzyme are numbered consecutively (in parentheses) starting clockwise (or from left to right) after the unique EcoRI site at zero map position. Each single-end ^{32}P-labeled fragment is represented as a horizontal arrow plus a dot. The dot indicates the position of the ^{32}P-labeled terminus (either 3′ or 5′ end), and the arrowhead shows the direction of sequence analysis using the method of Maxam and Gilbert (23). Details concerning these DNA fragments, designated a, b, c..., are given in Table 2. The sequence data within the coding region are summarized in Fig. 3 and 4. The genome locations of T and t antigens are represented at the bottom part of the figure.*

Since BKV is closely related to SV40, well-established findings about the latter can be used as a guide to interpret DNA and protein structure and function relationships of BKV. It is known that the SV40 genome extending counterclockwise from map positions 0.65 to 0.17 codes for t and T antigens. Since BKV complements the early mutant *ts*A58 of SV40 (22), and the sizes of BKV and SV40 T antigens are similar, it may be assumed that the early region of BKV also codes for the t and T antigens. Prediction of the amino acid sequence of the t antigen was made by reading the 5′-strand DNA sequence of BKV(MM), which has the same polarity as the early mRNA into protein sequence, namely, the triplet ATG corresponds to the initiation codon AUG, while the triplets TAA, TAG, and TGA correspond to the termination codons UAA, UAG, and UGA, respectively.

The initiation codon ATG appears first at nucleotides 3,047 to 3,045 (map position 0.614) (Fig. 3). Starting from this codon, there is a unique through-reading frame with 100 sense codons followed by a TAA codon at nucleotides 2,747 to 2,745. This putative protein of 100 amino acids in length was considered to be BKV(MM) small t antigen as reported earlier from this laboratory (54). This conclusion was mainly based on the fact that the actual protein sequence of t antigen of SV40 (of monkey origin), a related virus of BKV (of human origin), shared striking similarity with that predicted from BKV (54). Specifically, the first 10 amino acids at the N-terminus are identical in both viruses (see Fig. 3). The sequence between amino acids 11 and 74 is highly homologous. However, there is one difference: BKV(MM) t antigen is 74 amino acids shorter than that of SV40. The percent homology by amino acids is 74%, and DNA homology is about 73%. The DNA sequence of BKV(WT) at the similar region has also been obtained (6, 51). The t-antigen sequence predicted from BKV(WT) is similar to that of BKV(MM) except that the t antigen of BKV(WT) is 172 acids in length, similar to that of SV40. If a t antigen were produced in BKV(MM), it would have to function as a truncated version of the wild-type t antigen. Very recently, Seif et al. (34) reported that a small t antigen could not be detected in cells infected by the BKV(MM). Thus, t antigen in BKV(MM) appears to be nonessential for transformation of cells and for tumor formation in rodents.

BKV coding sequences for large-T antigen. In both BKV and SV40, the majority of the methionine-containing tryptic peptides re-

TABLE 2. *Information on DNA fragments used in sequence analysis*

Frag-ment	Location[a]	Map position	Base sequence obtained[b]
a	*Mbo*I(3)-*Mbo*I(4)[c]	0.122–0.209	746–1,038
b	*Mbo*I(3)-*Hin*dIII(1)[c]	0.122–0.176	660–847
c	*Mbo*II(1)-*Hae*III(5)[c]	0.143–0.182	849–897
d	*Hin*dIII(1)[c]-*Xba*I(1)	0.176–0.237	873–971
e	*Mbo*I(4)-*Hae*III(6)[c]	0.209–0.244	1,053–1,211
f	*Eco*RI(1)-*Xba*I(1)[c]	0.000–0.237	1,010–1,119
g	*Mbo*(4)[c]-*Mbo*I(5)	0.209–0.315	1,035–1,356
h	*Mbo*I(4)-*Mbo*I(5)[c]	0.209–0.315	1,277–1,568
i	*Hae*III(7)[c]-*Pst*I(1)	0.265–0.333	1,320–1,533
j	*Mbo*II(6)[c]-*Pst*I(1)	0.311–0.333	1,545–1,623
k	*Mbo*II(7)[c]-*Sac*I(1)	0.343–0.368	1,720–1,772
l	*Mbo*I(5)[c]-*Mbo*I(6)	0.315–0.452	1,564–1,935
m	*Mbo*II(8)-*Mbo*II(9)[c]	0.380–0.438	2,055–2,161
n	*Mbo*I(5)-*Mbo*I(6)[c]	0.315–0.452	1,895–2,248
o	*Mbo*II(11)[c]-*Mbo*II(12)	0.449–0.516	2,222–2,452
p	*Mbo*II(11)-*Mbo*II(12)[c]	0.449–0.516	2,227–2,550
q	*Mbo*I(7)[c]-*Hae*III(8)	0.512–0.555	2,540–2,661
r	*Mbo*I(7)[c]-*Hin*dIII(2)	0.512–0.594	2,608–2,808
s	*Mbo*I(7)-*Hae*III(8)[c]	0.512–0.555	2,694–2,744
t	*Mbo*I(7)-*Hin*dIII(2)[c]	0.512–0.594	2,795–2,952
u	*Hae*III(8)-*Mbo*II(17)[c]	0.555–0.583	2,767–2,884
v	*Hae*III(8)[c]-*Hae*III(9)	0.555–0.603	2,761–2,928
w	*Hae*III(8)-*Hae*III(9)[c]	0.555–0.603	2,790–2,990
x	*Hin*dIII(2)[c]-*Hin*dIII(3)	0.594–0.712	2,949–3,361

[a] Each DNA fragment is located between two specific restriction enzyme sites. The cleavage sites of each enzyme are numbered (in parentheses) consecutively and clockwise starting after the unique *Eco*RI site.

[b] The two numbers represent the stretch of the nucleotide sequence determined from this fragment. The second base of the unique *Eco*RI hexanucleotide recognition sequence (5'-GAATTC) is taken as nucleotide 1. Starting from this point, other nucleotides are numbered clockwise and consecutively.

[c] Representing the 3' or 5' ^{32}P-labeled ends.

leased from the t antigen are in common with those derived from the large-T protein (36). Recently, Paucha et al. determined the partial amino acid sequences at the N-terminal region of both t and T antigens of SV40 (28). They discovered that the two antigens indeed have identical amino-termini. These findings are in perfect agreement with those predicted directly from DNA sequence, assuming that the ATG triplet between nucleotides 5,081 and 5,079 (at map position 0.647) is used as the initiation codon for both antigens (9, 32). In BKV, it is assumed that the same triplet ATG that is used for initiating small-t protein synthesis is also the starting codon for the large-T protein (Fig. 3). BKV and SV40 T antigens are related not only by size (90,000 to 100,000 daltons) and antigenic determinants (21, 39), but also by amino acid composition (37). Approximately 30% (6/19 to 6/18) of the methionine-containing tryptic peptides are identical between the two T antigens. Furthermore, four common methionine-containing tryptic peptides between homogeneous T and t antigens were again found to be identical in BKV and SV40. This result is in general agreement with our data in predicting three such peptides.

In SV40, if the mRNA for the T antigen were a continuous transcript of the entire early region, translation of such a transcript would be interrupted by a large number of termination signals on all three reading frames. However, data obtained from transcriptional mapping revealed that SV40 large-T mRNA is transcribed from two noncontiguous segments of DNA sequences (1, 2, 12). The first 82 NH_2-terminal amino acids in SV40 T antigen are identical to those in t antigen. The SV40 large-T mRNA is spliced from map position 0.60 to 0.534, corresponding to nucleotides 4,835 to 4,489 (Fig. 3). This splice removes a 346-base-long sequence where termination codons from all three reading frames reside. After splicing, the mRNA continues to code for another 626 amino acids before reaching a termination codon at map position 0.174.

At the junctions of splicing in SV40 mRNA for T antigen, an AGGU sequence (nucleotides 4,837 to 4,834) appears as the donor (the junction between an exon and an intron), and an AGAU sequence (nucleotides 4,491 to 4,488) appears as the acceptor site (the junction between an intron and an exon). In BKV(MM) DNA, at the region analogous to the SV40 large-T coding sequence, an AGGU occurs between nucleotides 2,809 and 2,806 (Fig. 3), suggesting that BKV large-T mRNA may be spliced at this point. Given this assumption, BKV T and t antigens share 80 common amino acids from the unique NH_2-terminal end. Also, by analogy with the second part of SV40 T-antigen coding region, another AGGU (instead of AGAU as in SV40) appears between nucleotides 2,727 and 2,724 as the possible acceptor splicing point for BKV large-T mRNA. From nucleotide 2,725 on, BKV large-T mRNA can code for another 614 amino acids as the C-terminal sequence (see Fig. 4). It is assumed that the splicing events for processing the early mRNA in both viruses are most likely similar.

In Fig. 4, approximately 80% amino acid homology is found within the first 525 amino acids of BKV and SV40 large T_2 antigen (the second part of the T protein). Homology between amino acids 526 to 614 is much less. Conservation of most of the first 525 amino acids may be due to functional reasons. It has been reported that the SV40 coding region between map positions 0.54 and 0.21 is essential for expression of the T antigen, whereas deletions at map positions 0.21 and 0.18 give viable mutants (3).

In SV40 DNA, it has been noticed that if the sequences coding for amino acids 551 to 554 were

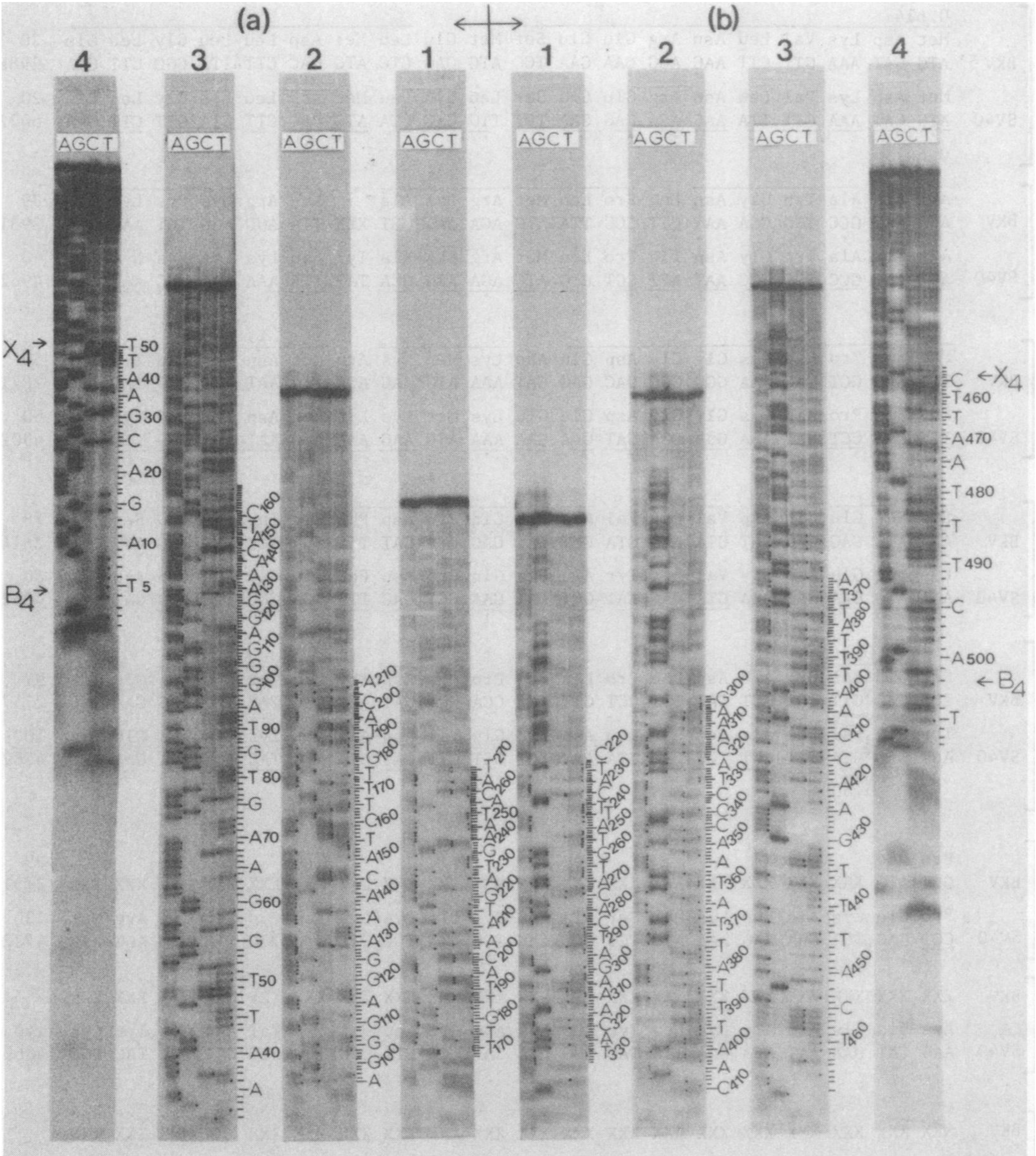

FIG. 2. *Autoradiogram of a typical sequencing gel. Fragments g and h (Fig. 1 and Table 2) are the two single-stranded DNA fragments separated by gel electrophoresis from double-stranded 5'-^{32}P-labeled MboI-C. These single-end-labeled fragments were subjected to partial chemical degradation according to the procedures of Maxam and Gilbert (23). The DNA digests were fractionated in an 8% polyacrylamide gel (0.4 mm by 40 cm by 35 cm) as described by Sanger and Coulson (33). Four specific reactions are indicated at the tops of the lanes: A lanes represent the A- and C-specific cleavage (A > C), G lanes represent the G-specific cleavage, C lanes represent the C-specific reaction, and T lanes represent the C and T reaction (C > T). The sequencing patterns of fragments g and h, shown in parts (a) and (b), respectively, are derived from a single gel. The four loadings in each part are also indicated on top. The position of each nucleotide is marked with a dot next to each band, and the identity of each nucleotide can be clearly read from the figure. Only every fifth nucleotide is written on the figure. Nucleotides 1 to 270 in part (a) correspond to nucleotides 1,554 to 1,285 in Fig. 3, and nucleotides 220 to 508 in part (b) are complementary to nucleotides 1,335 to 1,047 in Fig. 3.*

```
       0.614
       Met Asp Lys Val Leu Asn Arg Glu Glu Ser Met Glu Leu Met Asp Leu Leu Gly Leu Glu   20
BKV 5' ATG GAT AAA GTT CTT AAC AGG GAA GAA TCC ATG GAG CTC ATG GAC CTT TTA GGC CTT GAA   2988

       Met Asp Lys Val Leu Asn Arg Glu Glu Ser Leu Gln Leu Met Asp Leu Leu Gly Leu Glu   20
SV40   ATG GAT AAA GTT TTA AAC AGA GAG GAA TCT TTG CAG CTA ATG GAC CTT CTA GGT CTT GAA   5022
       0.647

       Arg Ala Ala Trp Gly Asn Leu Pro Leu Met Arg Lys Ala     Leu Arg Lys Cys Lys Glu   39
BKV    AGA GCT GCC TGG GGA AAT CTT CCC TTA ATG AGA AAA GCT XXX TTA AGG AAG TGT AAG GAA   2931

       Arg Ser Ala Trp Gly Asn Ile Pro Leu Met Arg Lys Ala Tyr Leu Lys Lys Cys Lys Glu   40
SV40   AGG AGT GCC TGG GGG AAT ATT CCT CTG ATG AGA AAG GCA TAT TTA AAA AAA TGC AAG GAG   4962

       Phe His Pro Asp Lys Gly Gly Asp Glu Asp Lys Met Lys Arg Met Asn Thr Leu Tyr Lys   59
BKV    TTT CAC CCT GAC AAA GGG GGC GAC GAG GAT AAA ATG AAG AGA ATG AAT ACT TTG TAT AAA   2871

       Phe His Pro Asp Lys Gly Gly Asp Glu Glu Lys Met Lys Lys Met Asn Thr Leu Tyr Lys   60
SV40   TTT CAT CCT GAT AAA GGA GGA GAT GAA GAA AAA ATG AAG AAA ATG AAT ACT CTG TAC AAC   4902

       Lys Met Glu Gln Asp Val Lys Val Ala His Gln Pro Asp Phe Gly Thr     Trp Ser Ser   78
BKV    AAA ATG GAG CAG GAT GTA AAG GTA GCT CAT CAG CCT GAT TTT GGA ACC XXX TGG AGT AGC   2814

       Lys Met Glu Asp Gly Val Lys Tyr Ala His Gln Pro Asp Phe Gly Gly Phe Trp Asp Ala   80
SV40   AAA ATG GAA GAT GGA GTA AAA TAT GCT CAT CAA CCT GAC TTT GGA GGC TTC TGG GAT GCA   4842

           0.566
       Ser Glu↓Val Cys Ala Asp Phe Pro Leu Cys Pro     Asp Thr Leu Tyr Cys Lys Glu Trp   97
BKV    TCA GAG↓GTT TGT GCT GAT TTT CCT CTT TGC CCA XXX GAT ACC CTG TAC TGC AAG GAA TGG   2757

       Thr Glu↓Val Phe Ala Ser Ser Leu Asn Pro Gly Val Asp Ala Met Tyr Cys Lys Gln Trp   100
SV40   ACT GAG↓GTA TTT GCT TCT TCC TTA AAT CCT GGT GTT GAT GCA ATG TAC TGC AAA CAA TGG   4782
            0.600

       Pro Met                                                                           99
BKV    CCT ATG XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX   2751

       Pro Glu Cys Ala Lys Lys Met Ser Ala Asn Cys Ile Cys Leu Leu Cys Leu Leu Arg Met   120
SV40   CCT GAG TGT GCA AAG AAA ATG TCT GCT AAC TGC ATA TGC TTG CTG TGC TTA CTG AGG ATG   4722

BKV    XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX
       Lys His Glu Asn Arg Lys Leu Tyr Arg Lys Asp Pro Leu Val Trp Val Asp Cys Tyr Cys   140
SV40   AAG CAT GAA AAT AGA AAA TTA TAC AGG AAA GAT CCA CTT GTG TGG GTT GAT TGC TAC TGC   4662

BKV    XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX
       Phe Asp Cys Phe Arg Met Trp Phe Gly Leu Asp Leu Cys Glu Gly Thr Leu Leu Leu Trp   160
SV40   TTC GAT TGC TTT AGA ATG TGG TTT GGA CTT GAT CTT TGT GAA GGA ACC TTA CTT CTG TGG   4602

                                                              (0.553)
                                                           Pro ↓                         100
BKV    XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX CCC TAA GTA ATT XXX XXX ATT   2736

       Cys Asp Ile Ile Gly Gln Thr Thr Tyr Arg Asp Leu Lys Leu    ↓                      174
SV40   TGT GAC ATA ATT GGA CAA ACT ACC TAC AGA GAT TTA AAG CTC TAA GGT AAA TAT AAA ATT   4542
                                                                (0.546)
                                                                        ↓
BKV    TTT XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX TTT ATA G             2726
                                                                        ↑
SV40   TTT AAG TGT ATA ATG TGT TAA ACT ACT GAT TCT AAT TGT TTG TGT ATT TTA G             4490
                                                                        ↓
```

read in a different reading frame, four consecutive ATG triplets occurred which could initiate the synthesis of a 91-amino acid protein (9, 32). However, no ATG triplets were found in BKV(MM) DNA at the corresponding positions.

Sequence homology between BKV and SV40. BKV(MM) and SV40 share about 70% DNA sequence homology within the coding region and 70% protein sequence homology. The high degree of homology suggests a close evolutionary relationship. Either the two viruses are evolved from a common ancestor, or one virus is evolved from the other. The two longest stretches of perfect amino acid homology in the T antigen of the two viruses are 26 and 28 long. These stretches correspond to a coding sequence of 78 and 84 nucleotides in length, respectively. However, the longest stretches of perfect nucleotide homologies between BKV(MM) and SV40 DNA within these two regions are only 16 and 12 nucleotides long, respectively. This finding is not surprising, since codon degeneracy allows the third letter of the genetic code word to be different without changing the amino acid. However, the high level of protein sequence homology suggests that during evolution there has been a strong pressure to conserve the amino acid sequence of the T antigen even though many point mutations have occurred. Due to the high degree of sequence homology of T antigen in these two viruses, the tertiary structure of T antigen is probably similar. This might explain the strong immunological cross-reactivity.

The percent DNA homology between BKV(WT) and SV40 as determined under different hybridization conditions gave different values. For example, hybridizations under stringent conditions (17, 19, 45) were carried out at around T_m − 25°C. According to the equation of McConaughy et al. (24), T_m (°C) = 81.5 + 16.6 (log M) + 0.41 (% G+C) − 0.72 (% formamide), where M equals the concentration of Na^+. If the guanine plus cytosine (G+C) content of SV40 DNA is taken as 41%, the T_m for SV40 is thus 83°C at 0.13 M Na^+. BKV(WT) has a G+C content of 39% and the T_m is calculated to be 82°C, whereas the early region of BKV(WT) has a G+C content of 35% and the expected T_m of 81°C. Thus, the hybridization temperature employed was around 83°C − 25°C = 58°C (equal to around 60°C at 0.15 M Na^+). The T_m of the 19-base pair-long cohesive end of bacteriophage 186 DNA, which contains 13 G+C pairs (27), is 63°C at 0.13 M NaCl (43). The T_m of the 12-base pair-long cohesive end of λDNA, which includes 10 G+C pairs (47), is 50°C at 0.13 M Na^+ (44). This value is somewhat lower than that calculated from $(G)_{12}$:$(C)_{12}$, which has an estimated T_m of 50°C at 0.02 M Na^+ (7). Thus, perfect homologous stretches of nucleotides between SV40 and BKV of 16 (relatively rare, see Table 3) and 12 base pairs long and 40% G+C are expected to have estimated T_m values of around 50 and 42°C, respectively. Using these values, the hybridization analysis carried out at 58°C is not likely to detect duplex regions formed between SV40 and BKV DNA containing stretches of base-paired regions of 12 to 16 long and 40% G+C. If this type of analysis and comparison is applicable to other hybridization experiments using the stringent condition (such as 60°C at 0.15 M Na^+), then many published values on percent hybridization may be underestimating the extent of sequence homology between two species of DNA molecules. However, two short, perfectly homologous regions (e.g., 6 to 12 bases long) separated by a small number of mis-.

FIG. 3. *Nucleotide sequence of BKV(MM) DNA coding for small-t and T_1 (the first portion of large T) antigens and the predicted amino acid sequences. Only the 5'-strand sequence with the same polarity (counterclockwise) as the early mRNA of the BKV(MM) genome is given. Approximately 70% of the sequence was derived from sequence analysis of both strands. The remainder was obtained from a single strand in at least two separate experiments. The unique reading frame is presented with three-letter codons, and the deduced amino acid sequence is placed above the nucleotide sequence. Both nucleotide and amino acid sequences of SV40 are also presented for comparison. Homologous nucleotides are underlined under the SV40 DNA sequence, and homologous amino acids are marked above the BKV protein sequence. For BKV DNA, the second base from the 5' end of the unique EcoRI hexanucleotide recognition sequence (5'-GAATTC) is taken as nucleotide 1. From this base on, all the other nucleotides are numbered consecutively and clockwise along the circular BKV genome. The numbering system of the complete SV40 (strain 776) DNA sequence is adopted from Reddy et al. (32). Numbers for both DNA and protein sequences are given at the right-hand margins. Every tenth nucleotide is also marked with a dot underneath. Regions marked with XXX represent gaps where matching of sequences shows that nucleotides are absent in one virus. The splice sequences are underlined, together with the splice points (arrows). The indicated splicing point between nucleotides 4,835 and 4,836 corresponds to the end of SV40 T_1 gene. The partial amino acid sequence of the first 19 residues from the unique NH_2 terminus of SV40 t and T_1 antigens was determined by Paucha et al. (28).*

```
        0.549
        Val Pro Thr Tyr Gly Thr Glu Glu Trp Glu Ser Trp Trp Ser Ser Phe Asn Glu Lys Trp   20
BKV 5'  GTG CCA ACC TAT GGA ACA GAA GAG TGG GAG TCC TGG TGG AGT TCC TTT AAT GAA AAA TGG 2666
        Ile Pro Thr Tyr Gly Thr Asp Glu Trp Glu Gln Trp Trp Asn Ala Phe Asn Glu           18
SV40    ATT CCA ACC TAT GGA ACT GAT GAA TGG GAG CAG TGG TGG AAT GCC TTT AAT GAG XXX XXX 4436
        0.534

        Asp Glu Asp Leu Phe Cys His Gln Asp Met Phe Ala Ser Asp Glu Glu Ala Thr Ala Asp   40
BKV     GAT GAA GAT TTA TTT TGC CAT GAA GAT ATG TTT GCC AGT GAT GAA GAA GCA ACA GCA GAT 2606
            Glu Asn Leu Phe Cys Ser Glu Glu Met Pro Ser Ser Asp Asp Glu Ala Thr Ala Asp   37
SV40    XXX GAA AAC CTG TTT TGC TCA GAA GAA ATG CCA TCT AGT GAT GAT GAG GCT ACT GCT GAC 4379

        Ser Gln His Ser Thr Pro Pro Lys Lys Lys Arg Lys Val Glu Asp Pro Lys Asp Phe Pro   60
BKV     TCT CAA CAC TCA ACA CCA CCC AAA AAA AAA AGA AAG GTA GAA GAC CCT AAA GAC TTT CCC 2546
        Ser Gln His Ser Thr Pro Pro Lys Lys Lys Arg Lys Val Glu Asp Pro Lys Asp Phe Pro   57
SV40    TCT CAA CAT TCT ACT CCT CCA AAA AAG AAG AGA AAG GTA GAA GAC CCC AAG GAC TTT CCT 4319

        Ser Asp Leu His Gln Phe Leu Ser Gln Ala Val Phe Ser Asn Arg Thr Leu Ala Cys Phe   80
BKV     TCT GAT CTA CAC CAG TTT CTT AGT CAA GCT GTA TTT AGT AAT AGA ACC CTT GCC TGC TTT 2486
        Ser Glu Leu Leu Ser Phe Leu Ser His Ala Val Phe Ser Asn Arg Thr Leu Ala Cys Phe   77
SV40    TCA GAA TTG CTA AGT TTT TTG AGT CAT GCT GTG TTT AGT AAT AGA ACT CTT GCT TGC TTT 4259

        Ala Val Tyr Thr Thr Lys Glu Lys Ala Gln Ile Leu Tyr Lys Lys Leu Met Glu Lys Tyr  100
BKV     GCT GTG TAT ACT ACT AAA GAA AAA GCT CAA ATT CTG TAT AAA AAA CTT ATG GAA AAA TAT 2426
        Ala Ile Tyr Thr Thr Lys Glu Lys Ala Ala Leu Leu Tyr Lys Lys Ile Met Glu Lys Tyr   97
SV40    GCT ATT TAC ACC ACA AAG GAA AAA GCT GCA CTG CTA TAC AAG AAA ATT ATG GAA AAA TAT 4199

        Ser Val Thr Phe Ile Ser Arg His Met Cys Ala Gly His Asn Ile Ile Phe Phe Leu Thr  120
BKV     TCT GTA ACT TTT ATT AGT AGA CAC ATG TGT GCT GGG CAT AAT ATT ATA TTC TTT TTA ACT 2366
        Ser Val Thr Phe Ile Ser Arg His Asn Ser Tyr Asn His Asn Ile Leu Phe Phe Leu Thr  117
SV40    TCT GTA ACC TTT ATA AGT AGG CAT AAC AGT TAT AAT CAT AAC ATA CTG TTT TTT CTT ACT 4139

        Pro His Arg His Arg Val Ser Ala Ile Asn Asn Phe Cys Gln Lys Leu Cys Thr Phe Ser  140
BKV     CCA CAT AGA CAT AGA GTT TCT GCA ATT AAT AAT TTC TGT CAA AAG CTG TGT ACC TTT AGT 2306
        Pro His Arg His Arg Val Ser Ala Ile Asn Asn Tyr Ala Gln Lys Leu Cys Thr Phe Ser  137
SV40    CCA CAC AGG CAT AGA GTG TCT GCT ATT AAT AAC TAT GCT CAA AAA TTG TGT ACC TTT AGC 4079

        Phe Leu Ile Cys Lys Gly Val Asn Lys Glu Tyr Leu Leu Tyr Ser Ala Leu Thr Arg Asp  160
BKV     TTT TTA ATT TGT AAG GGT GTT AAT AAG GAA TAC TTA CTA TAT AGT GCC TTA ACT AGA GAT 2246
        Phe Leu Ile Cys Lys Gly Val Asn Lys Glu Tyr Leu Met Tyr Ser Ala Leu Thr Arg Asp  157
SV40    TTT TTA ATT TGT AAA GGG GTT AAT AAG GAA TAT TTG ATG TAT AGT GCC TTG ACT AGA GAT 4019

        Pro Tyr His Thr Ile Glu Glu Ser Ile Gln Gly Gly Leu Lys Glu His Asp Phe Asn Pro  180
BKV     CCA TAC CAT ACT ATA GAA GAA AGC ATT CAA GGG GGC TTA AAG GAG CAT GAT TTT AAC CCA 2186
        Pro Phe Ser Val Ile Glu Glu Ser Leu Pro Gly Gly Leu Lys Glu His Asp Phe Asn Pro  177
SV40    CCA TTT TCT ←GTT ATT GAG GAA AGT TTG CCA GGT GGG TTA AAG GAG CAT GAT TTT AAT CCA 3959

        Glu Glu Pro Glu Glu Thr Lys Gln Val Ser Trp Lys Leu Ile Thr Glu Tyr Ala Val Glu  200
BKV     GAA GAG CCT GAA GAA ACA AAG CAG GTG TCT TGG AAA TTA ATT ACT GAG TAT GCA GTA GAG 2126
        Glu Glu Ala Glu Glu Thr Lys Gln Val Ser Trp Lys Leu Val Thr Glu Tyr Ala Met Glu  197
SV40    GAA GAA GCA GAG GAA ACT AAA CAA GTG TCC TGG AAG CTT GTA ACA GAG TAT GCA ATG GAA 3899
```

FIG. 4. *Nucleotide sequence of BKV(MM) DNA coding for T_2 (the second portion of T) antigen and the predicted amino acid sequence. The DNA sequence from map positions 0.549 counterclockwise to 0.178 of BKV(MM) genome is given with the predicted amino acid sequence atop. The corresponding SV40 DNA and protein sequences are also included. See the legend of Fig. 3 for details.*

```
      Thr Lys Cys Glu Asp Val Phe Leu Leu Leu Gly Met Tyr Leu Glu Phe Gln Tyr Asn Val  220
BKV   ACA AAG TGT GAG GAT GTG TTT TTA TTA TTA GGT ATG TAT TTA GAA TTT CAA TAC AAT GTA  2066
      Thr Lys Cys Asp Asp Val Leu Leu Leu Leu Gly Met Tyr Leu Glu Phe Gln Tyr Ser Phe  217
SV40  ACA AAA TGT GAT GAT GTG TTG TTA TTG CTT GGG ATG TAC TTG GAA TTT CAG TAC AGT TTT  3839

      Glu Glu Cys Lys Lys Cys Gln Lys Lys Asp Gln Pro Tyr His Phe Lys Tyr His Glu Lys  240
BKV   GAG GAG TGT AAA AAG TGT CAG AAA AAA GAC CAG CCT TAT CAC TTT AAG TAT CAT GAA AAG 2006
      Glu Met Cys Leu Lys Cys Ile Lys Lys Glu Gln Pro Ser His Tyr Lys Tyr His Glu Lys  237
SV40  GAA ATG TGT TTA AAA TGT ATT AAA AAA GAA CAG CCC AGC CAC TAT AAG TAC CAT GAA AAG  3779

      His Phe Ala Asn Ala Ile Ile Phe Ala Glu Ser Lys Asn Lys Lys Val Ile Cys Gln Gln  260
BKV   CAC TTT GCA AAT GCT ATT ATT TTT GCA GAA AGT AAA AAC AAA AAA GTT ATT TGT CAG CAA  1946
      His Tyr Ala Asn Ala Ala Ile Phe Ala Asp Ser Lys Asn Gln Lys Thr Ile Cys Gln Gln  257
SV40  CAT TAT GCA AAT GCT GCT ATA TTT GCT GAC AGC AAA AAC CAA AAA ACC ATA TGC CAA CAG  3719

      Ala Val Asp Thr Val Leu Ala Lys Lys Arg Val Asp Thr Leu His Met Thr Arg Glu Glu  280
BKV   GCA GTA GAT ACA GTT TTA GCT AAA AAA AGA GTA GAT ACC CTT CAT ATG ACC AGG GAA GAA  1886
      Ala Val Asp Thr Val Leu Ala Lys Lys Arg Val Asp Ser Leu Gln Leu Thr Arg Glu Gln  277
SV40  GCT GTT GAT ACT GTT TTA GCT AAA AAG CGG GTT GAT AGC CTA CAA TTA ACT AGA GAA CAA  3659

      Met Leu Thr Glu Arg Phe Asn His Ile Leu Asp Lys Met Asp Leu Ile Phe Gly Ala His  300
BKV   ATG CTA ACA GAA AGA TTC AAT CAT ATA TTA GAT AAA ATG GAT TTA ATA TTT GGA GCT CAT  1826
      Met Leu Thr Asn Arg Phe Asn Asp Leu Leu Asp Arg Met Asp Ile Met Phe Gly Ser Thr  297
SV40  ATG TTA ACA AAC AGA TTT AAT GAT CTT TTG GAT AGG ATG GAT ATA ATG TTT GGT TCT ACA  3599

      Gly Asn Ala Val Leu Glu Gln Tyr Met Ala Gly Val Ala Trp Leu His Cys Leu Leu Pro  320
BKV   GGA AAT GCT GTA CTA GAA CAA TAT ATG GCA GGT GTT GCT TGG CTG CAC TGT TTG CTA CCT  1766
      Gly Ser Ala Asp Ile Glu Glu Trp Met Ala Gly Val Ala Trp Leu His Cys Leu Leu Pro  317
SV40  GGC TCT GCT GAC ATA GAA GAA TGG ATG GCT GGA GTT GCT TGG CTA CAC TGT TTG TTG CCC  3539

      Lys Met Asp Ser Val Ile Phe Asp Phe Leu His Cys Ile Val Phe Asn Val Pro Lys Arg  340
BKV   AAA ATG GAT TCT GTA ATA TTT GAT TTT TTG CAC TGT ATT GTT TTC AAT GTA CCT AAA AGA  1706
      Lys Met Asp Ser Val Val Tyr Asp Phe Leu Lys Cys Met Val Tyr Asn Ile Pro Lys Lys  337
SV40  AAA ATG GAT TCA GTG GTG TAT GAC TTT TTA AAA TGC ATG GTG TAC AAC ATT CCT AAA AAA  3479

      Arg Tyr Trp Leu Phe Lys Gly Pro Ile Asp Ser Gly Lys Thr Thr Leu Ala Ala Gly Leu  360
BKV   AGA TAC TGG TTA TTT AAA GGT CCC ATT GAT AGT GGA AAA ACA ACA CTA GCT GCA GGG TTG  1646
      Arg Tyr Trp Leu Phe Lys Gly Pro Ile Asp Ser Gly Lys Thr Thr Leu Ala Ala Ala Leu  357
SV40  AGA TAC TGG CTG TTT AAA GGA CCA ATT GAT AGT GGT AAA ACT ACA TTA GCA GCT GCT TTG  3419

      Leu Asp Leu Cys Arg Gly Lys Ala Leu Asn Val Asn Leu Pro Met Glu Arg Leu Thr Phe  380
BKV   TTA GAT TTG TGT AGA GGT AAA GCC TTA AAT GTA AAC CTA CCC ATG GAA AGG CTA ACC TTT  1586
      Leu Glu Leu Cys Gly Gly Lys Ala Leu Asn Val Asn Leu Pro Leu Asp Arg Leu Asn Phe  377
SV40  CTT GAA TTA TGT GGG GGG AAA GCT TTA AAT GTT AAT TTG CCC TTG GAC AGG CTG AAC TTT  3359

      Glu Leu Gly Val Ala Ile Asp Gln Tyr Met Val Val Phe Glu Asp Val Lys Gly Thr Gly  400
BKV   GAG CTA GGT GTA GCT ATA GAT CAG TAC ATG GTT GTT TTT GAA GAT GTA AAA GGG ACA GGA  1526
      Glu Leu Gly Val Ala Ile Asp Gln Phe Leu Val Val Phe Glu Asp Val Lys Gly Thr Gly  397
SV40  GAG CTA GGA GTA GCT ATT GAC CAG TTT TTA GTA GTT TTT GAG GAT GTA AAG GGC ACT GGA  3299
```

FIG. 4—*Continued.*

```
      Ala Glu Ser Lys Asp Leu Pro Ser Gly His Gly Ile Asn Asn Leu Asp Ser Leu Arg Asp  420
BKV   GCT GAA TCA AAG GAT TTG CCT TCA GGA CAT GGA ATA AAC AAT TTA GAC AGT TTG AGA GAT  1466
      Gly Glu Ser Arg Asp Leu Pro Ser Gly Gln Gly Ile Asn Asn Leu Asp Asn Leu Arg Asp  417
SV40  GGG GAG TCC AGA GAT TTG CCT TCA GGT CAG GGA ATT AAT AAC CTG GAC AAT TTA AGG GAT  3239

      Tyr Leu Asp Gly Ser Val Lys Val Asn Leu Glu Lys Lys His Leu Asn Lys Arg Thr Gln  440
BKV   TAT TTA GAT GGA AGT GTT AAG GTA AAT TTA GAA AAG AAA CAT TTA AAC AAA AGA ACC CAA  1406
      Tyr Leu Asp Gly Ser Val Lys Val Asn Leu Glu Lys Lys His Leu Asn Lys Arg Thr Gln  437
SV40  TAT TTG GAT GGC AGT GTT AAG GTA AAC TTA GAA AAG AAA CAC CTA AAT AAA AGA ACT CAA  3179

      Ile Phe Pro Pro Gly Leu Val Thr Met Asn Glu Tyr Pro Val Pro Lys Thr Leu Gln Ala  460
BKV   ATA TTT CCA CCA GGC TTG GTT ACA ATG AAT GAG TAT CCT GTC CCT AAA ACC CTG CAA GCT  1346
      Ile Phe Pro Pro Gly Ile Val Thr Met Asn Glu Tyr Ser Val Pro Lys Thr Leu Gln Ala  457
SV40  ATA TTT CCC CCT GGA ATA GTC ACC ATG AAT GAG TAC AGT GTG CCT AAA ACA CTG CAG GCC  3119

      Arg Phe Val Arg Gln Ile Asp Phe Arg Pro Lys Ile Tyr Leu Arg Lys Ser Leu Gln Asn  480
BKV   AGA TTT GTA AGA CAA ATA GAT TTT AGG CCC AAA ATA TAT TTA AGA AAA TCC TTA CAA AAC  1286
      Arg Phe Val Lys Gln Ile Asp Phe Arg Pro Lys Asp Tyr Leu Lys His Cys Leu Glu Arg  477
SV40  AGA TTT GTA AAA CAA ATA GAT TTT AGG CCC AAA GAT TAT TTA AAG CAT TGC CTG GAA CGC  3059

      Ser Glu Phe Leu Leu Glu Lys Arg Ile Leu Gln Ser Gly Met Thr Leu Leu Leu Leu Leu  500
BKV   TCA GAG TTC TTA CTT GAA AAA AGA ATT TTA CAA AGT GGA ATG ACC TTG TTG CTA CTG CTA  1226
      Ser Glu Phe Leu Leu Glu Lys Arg Ile Ile Gln Ser Gly Ile Ala Leu Leu Leu Met Leu  497
SV40  AGT GAG TTT TTG TTA GAA AAG AGA ATA ATT CAA AGT GGC ATT GCT TTG CTT CTT ATG TTA  2999

      Ile Trp Phe Arg Pro Val Ala Asp Phe Ala Thr Asp Ile Gln Ser Arg Ile Val Glu Trp  520
BKV   ATT TGG TTT AGG CCT GTA GCT GAT TTT GCA ACT GAT ATA CAA TCT AGA ATT GTT GAA TGG  1166
      Ile Trp Tyr Arg Pro Val Ala Glu Phe Ala Gln Ser Ile Gln Ser Arg Ile Val Glu Trp  517
SV40  ATT TGG TAC AGA CCT GTG GCT GAG TTT GCT CAA AGT ATT CAG AGC AGA ATT GTG GAG TGG  2939

      Lys Glu Arg Leu Asp Ser Glu Ile Ser Met Tyr Thr Phe Ser Arg Met Lys Try Asn Ile  540
BKV   AAG GAA AGG CTG GAT TCT GAG ATA AGT ATG TAT ACT TTT TCA AGG ATG AAA TAT AAT ATA  1106
      Lys Glu Arg Leu Asp Lys Glu Phe Ser Leu Ser Val Tyr Gln Lys Met Lys Phe Asn Val  537
SV40  AAA GAG AGA TTG GAC AAA GAG TTT AGT TTG TCA GTG TAT CAA AAA ATG AAG TTT AAT GTG  2879

      Cys Met Gly Lys Cys Ile Leu Asp Ile Thr Arg Glu Glu Asp Ser Glu Thr Glu Asp Ser  560
BKV   TGC ATG GGG AAA TGT ATT CTT GAT ATT ACA AGA GAA GAG GAT TCA GAA ACT GAA GAC TCT  1046
      Ala Met Gly Ile Gly Val Leu Asp Trp Leu Arg Asn Ser Asp Asp Asp Asp Glu Asp Ser  557
SV40  GCT ATG GGA ATT GGA GTT TTA GAT TGG CTA AGA AAC AGT GAT GAT GAT GAT GAA GAC AGC  2819

          Gly His Gly Ser Ser Thr Glu Ser Gln Ser Gln Cys Ser Ser Gln Val Ser          577
BKV   XXX GGA CAT GGA TCA AGC ACT GAA TCC CAA TCA CAA TGC TCT TCC CAA GTC TCA XXX XXX  995
      Gln Glu Asn Ala Asp Lys Asn Glu Asp Gly Gly Glu Lys Asn Met Glu Asp Ser Gly His  577
SV40  CAG GAA AAT GCT GAT AAA AAT GAA GAT GGT GGG GAG AAG AAC ATG GAA GAC TCA GGG CAT  2759

      Asp Thr                                         Ser Ala Pro Ala Glu Asp Ser Gln  587
BKV   GAT ACT XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX TCA GCC CCT GCT GAA GAT TCC CAA  965
      Glu Thr Gly Ile Asp Ser Gln Ser Gln Gly Ser Phe Gln Ala Pro     Gln Ser Ser Gln  596
SV40  GAA ACA GGC ATT GAT TCA CAG TCC CAA GGC TCA TTT CAG GCC CCT XXX CAG TCC TCA CAG  2702
```

FIG. 4—*Continued.*

```
      Arg Ser Asp Pro His Ser Gln Glu Leu His Leu Cys Lys Gly Phe Gln Cys Phe Lys Arg  607
BKV   AGG TCA GAC CCC CAT AGT CAA GAG TTG CAT TTG TGT AAA GGC TTT CAG TGT TTT AAA AGG  905
           •           •               •           •           •               •
      Ser Val His Asp His Asn Gln Pro Tyr His Ile Cys Arg Gly Phe Thr Cys Phe Lys Lys  616
SV40  TCT GTT CAT GAT CAT AAT CAG CCA TAC CAC ATT TGT AGA GGT TTT ACT TGC TTT AAA AAA  2642
       •           •               •           •           •               •

                                          (0.178)
      Pro Lys Thr Pro Pro Pro Lys                  614
BKV   CCT AAA ACA CCA CCC CCA AAA XXX XXX XXX TAA  881
           •           •
      Pro Pro Thr Pro Pro Pro Glu Pro Glu Thr      626
SV40  CCT CCC ACA CCT CCC CCT GAA CCT GAA ACA TAA  2609
       •           •               •           •
                                              (0.174)
```

FIG. 4—*Continued.*

matched bases might give higher T_m values than either region taken separately (7). Tinoco et al. (41) have reported a method to calculate the stability of secondary structures of RNA molecules consisting of base-paired regions interrupted by mismatches. They found that a 55-base fragment RNA of known sequence from R17 virus can form a structure with 20 base pairs and gives a free energy [ΔG(25°C)] value of −21.8 kcal, or approximately 0.8 kcal per base pair (1 kcal = 4.186 kJ). In the absence of ΔG values for each deoxynucleotide base pair and for internal loops in DNA, we made use of the same values assigned for RNA (41), realizing that the calculation would only provide relative information. A calculation of the ΔG(25°C) for the BKV/SV40 DNA heteroduplex, divided into regions of 130 base pairs in length, gave values as shown in Table 3 (last column). The average ΔG value per base pair varies from +0.3 to −1.7 kcal. In general, where the ΔG values are more positive than −0.4 kcal, the sequence homology is low (e.g., 34 to 41% at map position 0.15 to 0.225, and 34 to 49% at map positions 0.675 to 0.725; it may be noted that any two random DNA sequences will show a 25% homology by chance). Where the ΔG values are more negative than −0.8 kcal, sequence homology is usually greater than 70%. Although the correlation is good between percent homology and ΔG values, such correlation is not apparent when either parameter is compared to the low homology values (e.g., the lowest value is 4% for map positions 0.375 to 0.45, and the highest value is 32% for map positions at 0.75 to 0.825) as determined by the conventional hybridization methods (17).

Improved methods for determining homology have been reported recently. Newell et al. (25) used four different effective hybridization temperatures, $T_m - 35$°C, $T_m - 28$°C, $T_m - 20$°C, and $T_m - 13$°C, for heteroduplex analysis by electron microscopy, and Howley et al. (16) used similar temperature ranges for hybridization by blotting on nitrocellulose paper according to Southern (38). These recent methods not only include lower hybridization temperatures, but the methods in principal allow an estimate of the T_m of each region of the SV40/BKV DNA heteroduplex. Using these methods, much higher percent homology values were obtained. Homology values obtained from heteroduplex analysis (25) are in fairly close agreement with those obtained by direct DNA sequence analysis of BKV DNA (53), particularly when we used a value of 0.5°C (instead of 1.4°C) of T_m lowering per each percent mismatch (see Table 3).

Selective usage of codons in BKV early genes. The codons used in the early coding region of BKV(MM) DNA are documented in Table 4. The codon usage of BKV(WT) DNA in the corresponding region is essentially similar. The selective use of codons in SV40 DNA (9, 32) and other eucaryotic and procaryotic genes (13) has been extensively discussed. In those cases, it has been found that there is a remarkable deficiency of the dinucleotide sequence CG in sense codons. Such nonrandom utilization of code words has also been demonstrated in BKV DNA. As shown in Table 4, codons of the NCG type (N, any nucleotide) for serine, proline, threonine, and alanine are completely absent. There is also a shortage of CGN codons, all for arginine. Other selective examples of BKV early genes include the strong preference of AAA over AAG for lysine and UUU over UUC for phenylalanine. For all the NNpu (pu, purine) type of codons, selection is overwhelmingly in favor of adenine over guanine. For the NNpy (py, pyrimidine) type of codons, selection is in favor of uracil over cytosine. This is especially striking in the strong preference for NUU (used 82 times) over NUC (used 8 times). This preference has been observed in SV40 DNA by Reddy et al. (32) and Fiers et al. (9). Presumably, the nonrandom nature of codon selection is related to

TABLE 3. *DNA sequence homology between BKV and SV40*

Map position of SV40/BKV DNA	% Homology				% Homologous segment with length[e]			ΔG per base pair (kcal)[f]
	Sequence analysis[a]	Heteroduplex analysis		Hybridization analysis[d]	≧8	≧12	≧16	
		0.5°C/1% mismatch[b]	1.4°C/1% mismatch[c]					
0.00–0.025	79	76	91	16	45	22	0	−1.2
0.025–0.05	76	82	94	16	29	11	0	−1.2
0.05–0.075	78	88	96	25	46	26	0	−1.3
0.075–0.10	79	88	96	25	54	0	0	−1.2
0.10–0.125	80	88	96	25	40	10	0	−1.1
0.125–0.15	78	86	95	25	39	14	14	−1.3
0.15–0.175	34	(32)[g]	(76)	25	0	0	0	−0.08
0.175–0.20	43	(28)	(75)	11	0	0	0	−0.40
0.20–0.225	41	34	76	11	6	0	0	−0.04
0.225–0.25	63	40	79	11	13	0	0	−0.5
0.25–0.275	72	58	85	11	46	12	12	−0.7
0.275–0.30	79	64	87	11	51	32	0	−0.9
0.30–0.325	78	68	89	11	36	9	0	−1.2
0.325–0.35	73	68	89	6	31	0	0	−0.6
0.35–0.375	75	68	89	6	11	11	0	−1.0
0.375–0.40	76	70	90	4	28	0	0	−0.8
0.40–0.425	73	70	90	4	36	16	15	−0.7
0.425–0.45	77	70	90	7	40	40	0	−1.1
0.45–0.475	76	66	88	7	41	10	16	−0.9
0.475–0.50	77	66	88	7	41	24	13	−0.9
0.50–0.525	71	70	90	7	18	10	0	−0.9
0.525–0.55	67	56	84	7	13	0	0	−0.7
0.55–0.575	68	52	84	7	25	18	0	−0.7
0.575–0.60	59	54	84	7	30	0	0	−0.6
0.60–0.625	73	56	84	7	38	0	0	−0.9
0.625–0.65	77	56	84	18	37	10	0	−1.1
0.65–0.675	54	52	84	18	32	25	14	−1.0
0.675–0.70	34	(20)	(72)	18	0	0	0	+0.3
0.70–0.725	39	20	72	18	0	0	0	−0.2
0.725–0.75	76	66	88	18	33	12	12	−0.9
0.75–0.775	64	66	88	32	26	19	0	−0.8
0.775–0.80	84	64	87	32	55	24	12	−1.7
0.80–0.825	79	70	90	32	37	12	12	−1.2
0.825–0.85	79	72	90	14	16	10	0	−1.0
0.85–0.875	69	43	83	14	23	0	0	−0.8
0.875–0.90	74	56	83	14	23	0	0	−0.8
0.90–0.925	58	56	83	14	12	0	0	−0.5
0.925–0.95	84	62	86	20	62	62	52	−1.6
0.95–0.975	84	62	86	20	64	56	18	−1.6
0.975–1.0	64	62	86	16	6	0	0	−0.7

[a] Result based on direct DNA sequence analysis of BKV(WT) DNA (53) and SV40 DNA (9, 32).

[b] Result based on heteroduplex analysis as shown in footnote *c*, except that a value of 0.5°C of T_m lowering per 1% mismatch was used to calculate the percent homology. This value was empirically chosen for giving the best fit between heteroduplex analysis and sequence analysis.

[c] Result based on heteroduplex analysis by electron microscopy (25). For example, from map positions 0 to 0.025 (130 base pairs), it can be estimated from the data of Newell et al. (25) that the percent duplex at four different effective temperatures corresponds to a lowering of the T_m by 12°C. By using the value of 1.4°C of T_m lowering 1% sequence mismatch, a value of 9% mismatch was obtained. Thus, the homology can be calculated to be 91% for this section of BKV/SV40 heteroduplex DNA.

[d] Data taken from Khoury et al. (19).

[e] For each section of 130 base pairs, the percent of base pairs that gives perfect homologous stretches of at least 8, 12, or 16 long is presented. For example, for map positions 0 to 0.025, the stretches of perfect homology include one each of 15, 14, 11, 10, and 8 base pairs. The sum of these stretches amounts to 58 long, and after dividing it by 130, it gives 45% of nucleotides equal or larger than 8 in a stretch. When homologous stretches are joined by one or more mismatched base pairs, the stability of the heteroduplex in each region can be estimated as shown by the Δ*G* in the last column.

[f] 1 kcal = 4.186 kJ. Estimation of the free energy [ΔG(25°C)] for the BKV/SV40 heteroduplex was based on the value of Tinoco et al. (41) reported for RNA. The value of Δ*G* for every 130-base pair-long section was first calculated and then expressed as Δ*G* per base pair. For a region with 100% homology and 40% G+C, an average Δ*G* of −2.5 kcal per base pair is expected. For a region with 90% homology and one mismatch for every nine base pairs, an average Δ*G* value of −2.0 kcal is expected.

[g] Values in parentheses are less reliable.

TABLE 4. *Codon usage for BKV(MM) t and T antigens*

Codon		Codon usage															
		U				C				A				G			
		Region	Sequence	BKV	SV40	Region	Sequence	BKV	SV40	Region	Sequence	BKV	SV40	Region	Sequence	BKV	SV40
U	U	Phe	UUU	33	35	Ser	UCU	10	10	Tyr	UAU	16	15	Cys	UGU	15	8
	C	Phe	UUC	5	2	Ser	UCC	7	4	Tyr	UAC	5	10	Cys	UGC	4	7
	A	Leu	UUA	28	20	Ser	UCA	12	9	End	UAA	1	1	End	UGA	0	0
	G	Leu	UUG	12	22	Ser	UCG	0	0	End	UAG	0	0	Trp	UGG	11	12
C	U	Leu	CUU	10	10	Pro	CCU	13	15	His	CAU	16	15	Arg	CGU	0	0
	C	Leu	CUC	1	1	Pro	CCC	8	6	His	CAC	8	5	Arg	CGC	0	1
	A	Leu	CUA	11	9	Pro	CCA	9	9	Gln	CAA	19	14	Arg	CGA	0	0
	G	Leu	CUG	6	9	Pro	CCG	0	0	Gln	CAG	9	15	Arg	CGG	0	1
A	U	Ile	AUU	27	20	Thr	ACU	13	15	Asn	AAU	18	21	Ser	AGU	15	18
	C	Ile	AUC	0	0	Thr	ACC	10	7	Asn	AAC	7	13	Ser	AGC	3	6
	A	Ile	AUA	13	10	Thr	ACA	13	10	Lys	AAA	44	39	Arg	AGA	22	18
	G	Met	AUG	23	23	Thr	ACG	0	0	Lys	AAG	18	24	Arg	AGG	10	7
G	U	Val	GUU	12	12	Ala	GCU	18	26	Asp	GAU	34	34	Gly	GGU	6	7
	C	Val	GUC	2	1	Ala	GCC	6	6	Asp	GAC	9	14	Gly	GGC	5	7
	A	Val	GUA	18	9	Ala	GCA	10	7	Glu	GAA	38	37	Gly	GGA	13	13
	G	Val	GUG	4	12	Ala	GCG	0	0	Glu	GAG	19	20	Gly	GGG	6	9

various factors such as the tRNA composition, the mRNA structures, and their interaction energy.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant CA-14989, awarded by the National Cancer Institute, and by grant VC-216 from the American Cancer Society.

We thank T. Kelly for helpful suggestions in the heteroduplex analysis calculations for Table 3.

LITERATURE CITED

1. **Aloni, Y., R. Dhar, O. Laub, M. Horowitz, and G. Khoury.** 1977. Novel mechanism for RNA maturation-leader sequences of simian virus 40 mRNA are not transcribed adjacent to the coding sequences. Proc. Natl. Acad. Sci. U.S.A. **74:**3686–3690.
2. **Berk, A. J., and P. A. Sharp.** 1978. Spliced early messenger RNA's of simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **75:**1274–1278.
3. **Cole, C. N., T. Landers, S. P. Goff, S. Manteuil-Brutlag, and P. Berg.** 1977. Physical and genetic characterization of deletion mutants of simian virus 40 constructed in vitro. J. Virol. **24:**277–294.
4. **Crawford, L. V., C. N. Cole, A. E. Smith, E. Paucha, P. Tegtmeyer, K. Rundell, and P. Berg.** 1978. Organization and expression of early genes of simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **75:**117–121.
5. **Dhar, R., C. J. Lai, and G. Khoury.** 1978. Nucleotide sequence of the DNA replication origin for human papovavirus BKV: sequence and structural homology with SV40. Cell **13:**345–358.
6. **Dhar, R., I. Seif, and G. Khoury.** 1979. Nucleotide sequence of the BK virus DNA segment encoding small t antigen. Proc. Natl. Acad. Sci. U.S.A. **76:**565–569.
7. **Dodgson, J. B., and R. D. Wells.** 1977. Synthesis and thermal melting behavior of oligomer polymer-complexes containing defined lengths of mismatched dA-dG and dG-dG nucleotides. Biochemistry **16:**2367–2374.
8. **Fareed, G. C., and D. Davoli.** 1977. Molecular biology of papovaviruses. Annu. Rev. Biochem. **46:**471–522.
9. **Fiers, W., R. Contreras, G. Haegeman, R. Rogiers, A. Van de Voorde, H. Van Heuverswyn, J. Van Herreweghe, G. Volckaert, and M. Ysebaert.** 1978. Complete nucleotide sequence of SV40 DNA. Nature (London) **273:**113–120.
10. **Freund, J., G. diMayorca, and K. N. Subramanian.** 1979. Mapping and ordering of fragments of BK virus DNA produced by restriction endonucleases. J. Virol. **29:**915–925.
11. **Gardner, S. D., A. M. Field, D. V. Coleman, and B. Hulme.** 1971. New human papovavirus (B.K.) isolated from urine of renal transpla.ntation. Lancet **i:**1253–1257.
12. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. M. Weissman.** 1978. Heterogeneity and 5′-terminal structures of the late RNAs of simian virus 40. J. Mol. Biol. **126:**813–846.
13. **Grantham, R.** 1978. Viral, prokaryote and eukaryote genes contrasted by mRNA sequence indexes. FEBS Lett. **95:**1–11.
14. **Griffin, B. E., and M. Fried.** 1976. Structural mapping of an oncogenic virus (polyoma viral DNA). Methods Cancer Res. **12:**49–86.
15. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26:**365–369.
16. **Howley, P. M., M. A. Israel, M. Law, and M. A. Martin.** 1979. A rapid method for detecting and mapping homology between heterologous DNAs. J. Biol. Chem. **254:**4876–4883.
17. **Howley, P. M., G. Khoury, J. C. Byrne, K. K. Takemoto, and M. A. Martin.** 1975. Physical map of the BK virus genome. J. Virol. **16:**959–973.
18. **Kelly, T. J., Jr., and D. Nathans.** 1977. The genome of simian virus 40. Adv. Virus Res. **21:**85–173.
19. **Khoury, G., P. M. Howley, C. Garon, M. F. Mullarkey, K. K. Takemoto, and M. A. Martin.** 1975. Homology and relationship between the genomes of papoviruses, BK virus and simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **72:**2563–2567.
20. **Major, E. O., and G. DiMayorca.** 1973. Malignant transformation of BHK$_{21}$ clone 13 cells by BK virus 40. Proc. Natl. Acad. Sci. U.S.A. **76:**3210–3212.
21. **Martinis, J., and C. M. Croce.** 1978. Somatic cell hybrids producing antibodies specific for the tumor antigen of simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **75:**2320–2323.
22. **Mason, D. H., Jr., and K. K. Takemoto.** 1976. Complementation between BK human papovavirus and a simian virus 40 tsA mutant. J. Virol. **17:**1060–1062.
23. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74:**560–564.
24. **McConaughy, B. L., C. B. Laird, and B. J. McCarthy.**

1969. Nucleic acid reassociation in formamide. Biochemistry 8:3289–3295.

25. **Newell, N., C. Lai, G. Khoury, and T. J. Kelly, Jr.** 1978. An electron microscope study of the base sequence homology between simian virus 40 and human papovavirus BK. J. Virol. 25:193–201.

26. **Osborn, J. E., S. M. Robertson, B. L. Padgett, D. L. Walker, and B. Weisblum.** 1976. Comparison of JC and BK human papovavirus with simian virus 40: DNA homology studies. J. Virol. 19:675–684.

27. **Padmanabhan, R., and R. Wu.** 1972. Nucleotide sequence analysis of DNA. IV. Complete nucleotide sequence of the left-hand cohesive end of coliphage 186 DNA. J. Mol. Biol. 65:447–467.

28. **Paucha, E., A. Mellor, R. Harvey, A. Smith, R. Hewick, and M. Waterfield.** 1978. SV40 large and small T-antigens have identical amino termini in mapping at 0.65 map units. Proc. Natl. Acad. Sci. U.S.A. 75:2165–2169.

29. **Portolani, M., G. Barbanti-Brodano, and M. La Placa.** 1975. Malignant transformation of hamster kidney cells by BK virus. J. Virol. 15:420–422.

30. **Prives, C., E. Gilboa, M. Revel, and E. Wincour.** 1977. Cell-free translation of simian virus early messenger RNA coding for viral T-antigen. Proc. Natl. Acad. Sci. U.S.A. 74:457–461.

31. **Purchio, A., and G. C. Fareed.** 1979. Transformation of human embryonic kidney cells by human papovavirus BK. J. Virol. 29:763–769.

32. **Reddy, V. B., B. Thimmappaya, R. Dhar, K. N. Subramanian, S. B. Zain, J. Pan, P. K. Ghosh, M. L. Celma, and S. M. Weissman.** 1978. The genome of simian virus 40. Science 200:494–502.

33. **Sanger, F., and A. R. Coulson.** 1978. The use of thin acrylamide gels for DNA sequencing. FEBS Lett. 87: 107–110.

34. **Seif, I., G. Khoury, and R. Dhar.** 1979. The genome of human papovavirus BKV. Cell 18:963–177.

35. **Shah, K. V., R. W. Daniel, and J. Strandberg.** 1975. Sarcoma in a hamster inoculated with BK virus, a human papovavirus. J. Nat. Cancer Inst. 54:945–949.

36. **Simmons, D. T., and M. A. Martin.** 1978. Common methionine-tryptic peptides near the amino-terminal end of primate papovavirus tumor antigens. Proc. Natl. Acad. Sci. U.S.A. 75:1131–1135.

37. **Simmons, D. T., K. K. Takemoto, and M. A. Martin.** 1977. Relationship between the methionine tryptic peptide of simian virus 40 and BK virus tumor antigens. J. Virol. 24:319–325.

38. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503–517.

39. **Takemoto, K. K., and M. F. Mullarkey.** 1973. Human papovavirus, BK strain: biological studies including antigenic relationship to simian virus 40. J. Virol. 12:625–631.

40. **Takemoto, K. K., A. S. Rabson, M. F. Mullarkey, R. M. Blaese, C. F. Garon, and D. Nelson.** 1974. Isolation of papovavirus from brain tumor and urine of a patient with Wiskott-Aldrich syndrome. J. Nat. Cancer Inst. 53:1205–1207.

41. **Tinoco, I., Jr., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure of ribonucleic acids. Nature (London) New Biol. 246:40–41.

42. **Van der Noordaa, J.** 1976. Infectivity, oncogenicity and transforming ability of BK virus—a human papovavirus. J. Gen. Virol. 30:371–373.

43. **Wang, J. C.** 1967. Cyclization of coliphage 186 DNA. J. Mol. Biol. 28:403–411.

44. **Wang, J. C., and N. Davidson.** 1966. Thermodynamic and kinetic studies on the interconversion between linear and circular forms of phage lambda DNA. J. Mol. Biol. 15:111–123.

45. **Wold, W. S. M., J. K. Mackey, K. M. Brackmann, N. Takemori, P. Ridgen, and M. Green.** 1978. Analysis of human tumors and human malignant cell lines for BK virus-specific DNA sequences. Proc. Natl. Acad. Sci. U.S.A. 75:454–458.

46. **Wu, R., E. Jay, and R. Roychoudhury.** 1976. Nucleotide sequence analysis of DNA. Methods Cancer Res. 12:87–176.

47. **Wu, R., and E. Taylor.** 1971. Nucleotide sequence analysis of DNA. II. Complete nucleotide sequence of the cohesive ends of bacteriophage lambda DNA. J. Mol. Biol. 57:491–511.

48. **Yang, R. C. A., and R. Wu.** 1978. Cleavage map of BK virus DNA with restriction endonucleases *Mbo*I and *Hae*III. J. Virol. 27:700–712.

49. **Yang, R. C. A., and R. Wu.** 1978. Physical mapping of BK virus DNA with *Sac*I, *Mbo*II, and *Alu*I restriction endonucleases. J. Virol. 28:851–864.

50. **Yang, R. C. A., and R. Wu.** 1978. BK virus DNA: cleavage map and sequence analysis. Proc. Natl. Acad. Sci. U.S.A. 75:2150–2154.

51. **Yang, R. C. A., and R. Wu.** 1979. Comparative study of papovavirus DNA: BKV(MM), BKV(WT) and SV40. Nucleic Acids Res. 7:651–668.

52. **Yang, R. C. A., and R. Wu.** 1979. BK virus DNA sequence: extent of homology with simian virus 40 DNA. Proc. Natl. Acad. Sci. U.S.A. 76:1179–1183.

53. **Yang, R. C. A., and R. Wu.** 1979. BK virus DNA: complete nucleotide sequence of a human tumor virus. Science 206:456–462.

54. **Yang, R. C. A., and R. Wu.** 1979. BK virus DNA sequence coding for the amino-terminus of the T-antigen. Virology 92:340–352.