Proc. Natl. Acad. Sci. USA
Vol. 75, No. 5, pp. 2150–2154, May 1978
Biochemistry

BK virus DNA: Cleavage map and sequence analysis

(tumor virus/simian virus 40/leader mRNA sequence)

ROBERT C. A. YANG AND RAY WU*

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

Communicated by Leon A. Heppel, February 9, 1978

ABSTRACT **A detailed physical map of the BK virus (MM strain) genome has been constructed with respect to the cleavage sites of 11 different restriction enzymes. The enzymes cut BKV(MM) DNA at 61 specific sites whose locations have been determined. Preliminary nucleotide sequence was carried out in the region from 0.70–0.75 map positions on BKV(MM) DNA. An 80% homology was found at 0.714–0.744 map positions on BKV(MM) DNA with 0.722–0.752 map positions on simian virus 40 DNA. This region of simian virus 40 DNA codes for the synthesis of the leader sequence of late mRNA.**

The human papovavirus BK (BKV) was first isolated from the urine of renal allograft recipients (3, 4) and later from patients with the Wiskott–Aldrich syndrome (5). Although the association of BKV with human diseases is rare, BKV-specific antibodies are found in sera from most healthy people. A varient of BKV, the MM strain, is very similar to BKV except that its DNA is slightly shorter and it contains three instead of four *Hin*d cleavage sites (6).

BKV(MM) appears to be related to the well-characterized monkey virus, simian virus 40 (SV40). First, the origin of replication (at around 0.67 units) and the origin and direction of transcription are located similarly on the two viral DNA molecules relative to the *Eco*RI sites (P. M. Howley and M. A. Martin, C.-J. Lai and G. Khoury, personal communications). Second, antisera against SV40 directed T-antigen crossreacted with BKV-infected human cells (7). Third, out of 21 tryptic peptides of BKV T-antigen, 7 are similar to those of SV40 T-antigen (8). Fourth, the DNA of the two viruses exhibits sequence homology. Estimates for the extent of homology range from 11 to 92% (9, 10), depending on the exact technique used for the measurement. Thus, only direct nucleotide sequence analysis can reveal the exact homology.

MATERIALS AND METHODS

Cells and Viruses. Human embryonic kidney cells (HEK) were purchased from Microbiological Associates (Long Island, NY). The cells were propagated in Dulbecco's modified Eagle's medium supplemented with 2 g of glucose per liter, 10% fetal calf serum, and 10% tryptose phosphate broth. African green monkey kidney cells CV-1 and TC-7 were obtained and grown as described (11). For preparation of BKV DNA, HEK cells were infected with approximately 0.1 plaque-forming unit of BKV(MM) per cell. Unlabeled and ^{32}P-labeled BKV(MM) DNA and SV40 (small-plaque, from D. Nathans) DNA were prepared as described (11, 12).

Restriction Enzyme Cleavage of BKV(MM) DNA. Restriction enzymes (1) *Bam* I, *Mbo* II, *Eco*RI, *Hae* III, *Hha* I, *Hin*dIII, *Hpa* II, *Kpn* I, *Pst* I, and *Xba* I were purchased from BioLabs. *Mbo* I was a generous gift of R. Roberts. Digestion of DNA by these enzymes and the fractionation of DNA fragments by gel electrophoresis were as described (11, 13).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

RESULTS

Mapping of Cleavage Sites of Restriction Enzymes on BKV(MM) DNA. A detailed physical map of 61 cleavage sites on BKV(MM) DNA was constructed and is summarized in Fig. 1. The cleavage sites of enzymes that cut the DNA only once were determined relative to the known location for the *Eco*RI site and to the three known sites (6) for *Hin*dIII (Fig. 2). For example, for the *Pst* I cleavage, uniformly ^{32}P-labeled BKV(MM) DNA was digested simultaneously with *Pst* I and *Eco*RI. Each of the two resulting fragments was isolated and then digested by *Hin*dIII. The results show that *Hin*dIII-A_1, C, and B_1 fragments are released from one fragment, while *Hin*dIII-A_2 and B_2 are produced from the other fragment (Fig. 2*b*). From this information and the size of each fragment, the exact location of the *Pst* I cleavage site can be deduced. Similarly, the cleavage site of the other single-cut enzymes (*Hha* I, *Kpn* I, and *Bam* I) and of *Xba* I, which has two cleavage sites, were located on the BKV(MM) genome. All the three *Hin*d(II + III) fragments are generated by *Hin*dIII alone.

For mapping the sites of restriction enzymes that produce a large number of fragments on BKV(MM) DNA, the following procedures were used: (*a*) digestion of uniformly ^{32}P-labeled BKV(MM) DNA to completion with each restriction enzyme, or simultaneously with two enzymes, and determination of the number and size of the DNA fragments from electrophoretic data (2, 13, 14); (*b*) isolation of electrophoretically purified DNA fragments from step *a* and digestion with a second restriction enzyme (e.g., digestion of *Mbo* I-A with *Hin*dIII) to determine the number of cleavage sites and the size of subfragments, and application of overlap analysis to determine the order of the fragments (2, 14); and (*c*) partial digestion of single-end ^{32}P-labeled BKV(MM) DNA fragments with a second enzyme and determining the order of the partial products by size from the common labeled end (11, 15).

Complete Digestion of Uniformly Labeled BKV(MM) DNA. Digestion of uniformly ^{32}P-labeled BKV(MM) DNA with an excess of *Hin*dIII gives three fragments, *Hin*d-A, B, and C (6). We found that *Hin*d-C actually consists of three bands, designated C_a, C_b, and C_c, at approximately equal proportions. The largest fragment (C_a) is from the parent BKV(MM) since it contains one *Hpa* II site. The two shorter fragments (C_b and C_c) appear to come from deletion mutants of BKV(MM) since the *Hpa* II site is lost (unpublished observation). These two

Abbreviations: BKV, BK virus; SV40, simian virus 40. *Bam, Hae, Hha, Hin, Hpa, Kpn, Mbo, Pst,* and *Xba* are restriction enzymes (1) from *Bacillus amyloliquefaciens* H, *Haemophilus aegyptius, H. Haemolyticus, H. influenzae* d, *H. parainfluenzae, Klebsiella pneumoniae, Moraxella bovis, Proteus stuartii,* and *Xanthomonas badrii,* respectively. Each DNA fragment present in a complete digest is assigned a capital letter (2) in order of decreasing size (A, B, C, . . . , where A is largest) with an italicized prefix designating the enzyme used to produce the fragment (e.g., *Hin* III-A). Each fragment produced by simultaneous digestion of two enzymes is designated by the names of enzymes followed by an arabic number (e.g., *Hin* III–*Mbo* I 3).

* To whom correspondence should be addressed.

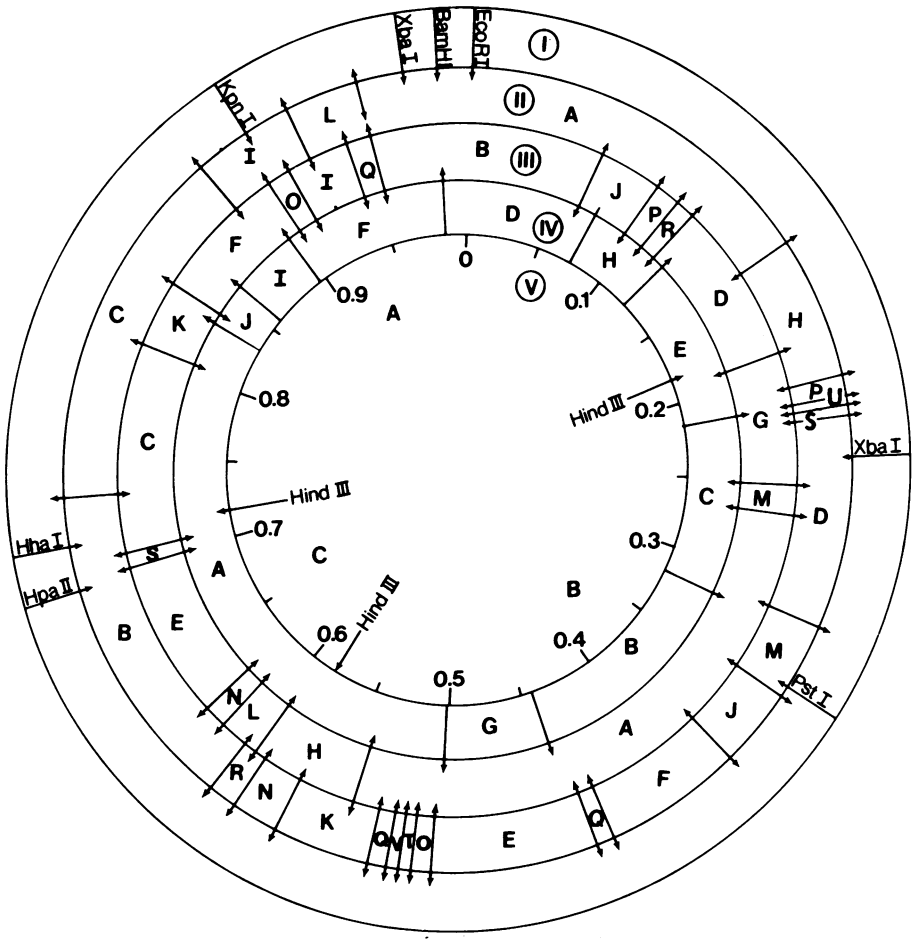

FIG. 1. A detailed physical map of BKV(MM) DNA. The exact location of all the 61 cleavage sites produced by 11 restriction enzymes are assigned on this genome. The *Mbo* II, *Hae* III, *Mbo* I, and *Hin*dIII sites are shown in circles II, III, IV, and V, respectively. The unique *Eco*RI site is taken as the reference zero site.

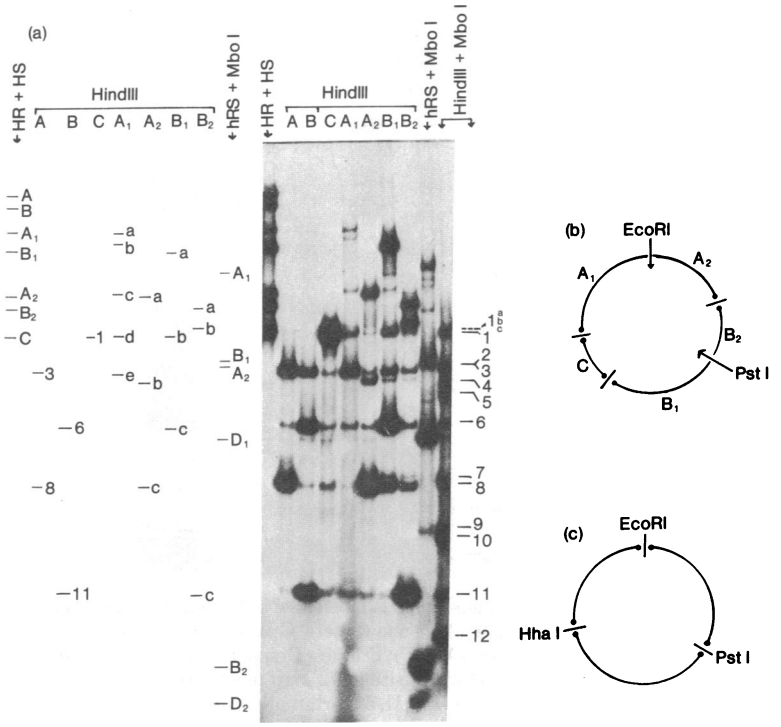

FIG. 2. (*a*) Partial cleavage pattern of the single-end ^{32}P-labeled *Hin*dIII DNA fragments digested with *Mbo* I. The single-end ^{32}P-labeled *Hin*d-A_1, A_2, B_1, and B_2 were obtained by digestion with *Eco*RI and *Pst* I, respectively (see *b* and *c*). Each single-end labeled fragment was then partially digested with *Mbo* I (lanes A_1, A_2, B_1, and B_2). Electrophoresis of the partial DNA digests (1.5×10^4 cpm) was on a 4% polyacrylamide gel. (Lanes A, B, and C) Complete digestion of 5′-^{32}P-labeled *Hin*d-A, *Hin*d-B, and *Hin*d-C fragments, respectively, with *Mbo* I to give the end-labeled DNA subfragments (e.g., *Hin*dIII–*Mbo* I 3 and 8 from *Hin*d-A). (Lane HR+HS) Mixture of two double digests, *Hin*dIII–*Eco*RI (HR) and *Hin*dIII–*Pst* I (HS) digests of BKV(MM) DNA. (Lane hRS+*Mbo* I; also see *c*) 5′-^{32}P-Labeled DNA from a *Hha* I–*Eco*RI–*Pst* I digest was incubated with *Mbo* I to give six single-end labeled bands. (*b*) Circular *Hin*dIII cleavage map of BKV(MM) DNA. The 5′-^{32}P-labeled ends are represented by a dot and a line (●—); the cleavage sites of *Eco*RI and *Pst* I are marked by arrows. (*c*) Circular map of BKV(MM) DNA with sites of three single-cut restriction enzymes.

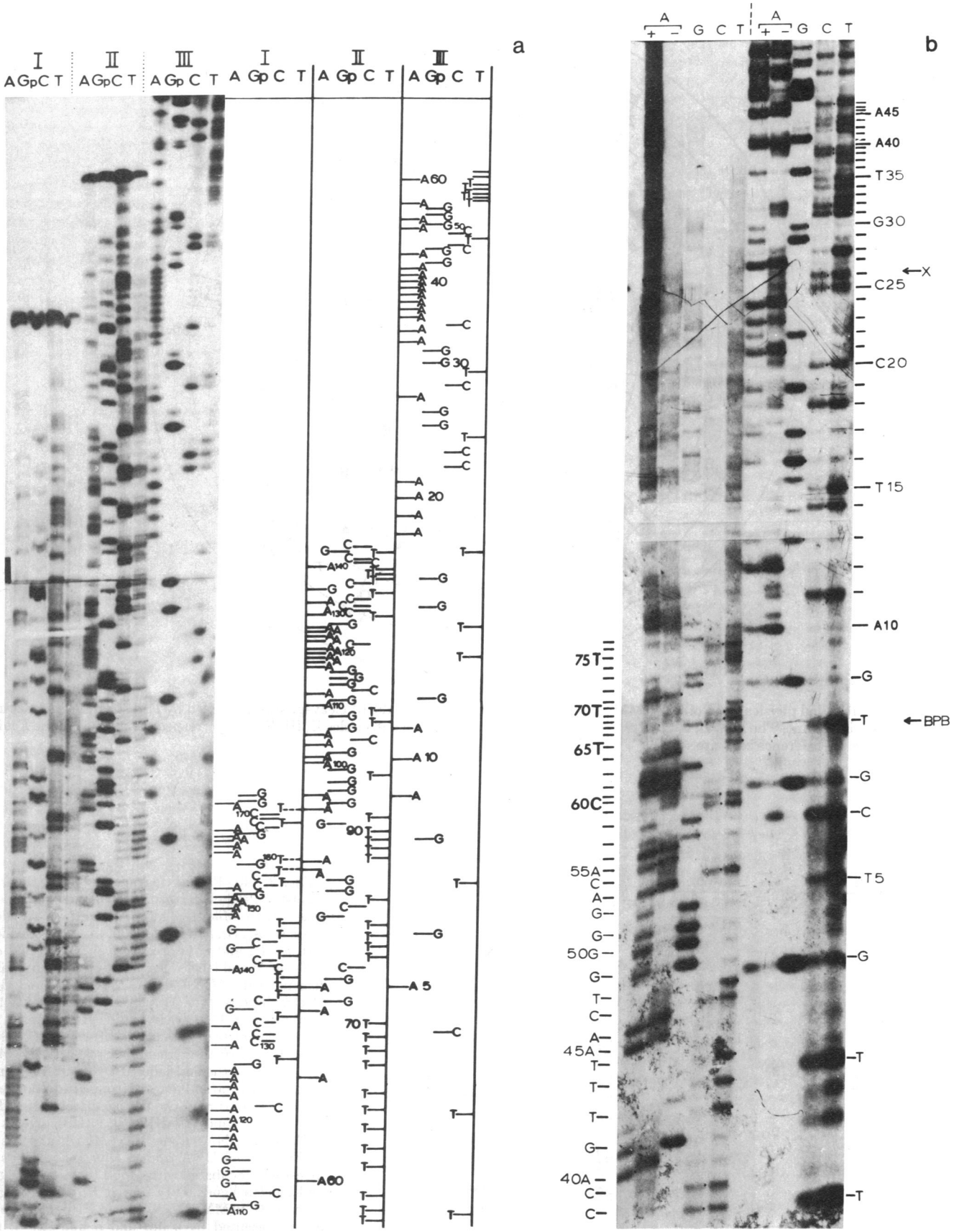

FIG. 3. DNA sequence analysis of BKV(MM) and SV40 DNA fragments. Sequences of single-end ^{32}P-labeled DNA fragments were determined by the chemical degradation procedure (21). The autoradiograms of the polyacrylamide gels (*a–c*) are shown here. Gp lanes, the G-specific reaction; A^+ lanes, the A- and G-specific cleavage (A > G); A^- or A lanes, A- and C-specific reaction (A > C); C lanes, C-specific reaction; T lanes, T + C. (*a*) BKV(MM) *Hin*dIII-A fragment cleaved with *Hae* III, 5′-terminally labeled at the *Hin*dIII site (0.72 map position). Nucleotides 1–170 correspond to nucleotides 91–260 on Fig. 4. (*b*) BKV(MM) *Hin*dIII-C_a fragment cleaved with *Hpa* II, 5′-terminally labeled at the *Hin*dIII site.

(*Continued on following page.*)

deletion mutants are different from the original isolate of BKV (6, 9) because the latter contains more DNA than BKV(MM).

Digestion of BKV(MM) DNA with *Mbo* I results in 10 fragments (A–J), varying in length from 1659 to 125 base pairs, which can be fractionated into 9 discrete bands in a 3.5% polyacrylamide gel. Incubation of BKV(MM) DNA with *Hae* III produces 19 fragments (A–S), which migrate as 15 bands in a 3.5% gel. The relative molecular weights of these fragments were estimated by direct comparison with the known sizes of SV40 DNA *Hae* III fragments (14) fractionated in the same gel.

Digestion of Individual *Hin*dIII Fragments with *Mbo* I. Uniformly labeled *Hin*d-A gives seven subfragments upon digestion with *Mbo* I. *Hin*d-B is cleaved by *Mbo* I into five subfragments while *Hin*d-C is not cleaved by *Mbo* I. In reciprocal digestion, each *Mbo* I fragment is digested by *Hin*dIII. By overlap analysis, a preliminary order of *Mbo* I fragments can be shown as follows: -A-(J, I, D, H, F)-E-(C, B, G).

Partial Digestion of Single-End ^{32}P-Labeled *Hin*dIII Fragments with *Mbo* I. 5′-End ^{32}P-labeled *Hin*dIII fragments A and B were digested with *Eco*RI or *Pst* I to generate single-end labeled *Hin*d-A_1, A_2, B_1, and B_2 (Fig. 2*b*). Each of these fragments was subjected to partial cleavage with *Mbo* I to produce a series of incompletely digested fragments. The DNA bands (sharing the same labeled end) resulting from partial digestion were then analyzed for size. The size difference between two adjacent bands was calculated and compared to the known size of DNA fragments produced by complete digestion. As illustrated in Fig. 2*a*, the partial *Mbo* I cleavage of *Hin*dIII-A_1 produces subfragments a, b, c, d, and e. The smallest fragment (e) is equivalent to fragment 3 in a *Hin*dIII + *Mbo* I double digest of BKV(MM) DNA (data not shown). Fragment d is longer than fragment e by 2.4% genome length of the BKV(MM) DNA, which is equivalent in size to *Hin*dIII–*Mbo* I 12. Thus, fragment 12 is immediately adjacent to fragment 3. By extending this analysis, the order of the component *Mbo* I fragments included in *Hin*d-A_1 can be deduced as *Hin*dIII–*Mbo* I 3-12-10-5-4B or *Mbo* I-(A-J-I-F-D) (Fig. 1). In a similar manner, partial digestion of *Hin*d-A_2 by *Mbo* I defines the order of *Mbo* I-(E-H-D), and *Hin*d-B_1 corresponds to *Mbo* I-(A-G-B). These data completely define the physical map for *Mbo* I (Fig. 1).

Mapping of *Hae* III Cleavage Sites. First, each uniformly labeled *Hin*dIII fragment was digested with *Hae* III to completion. In reciprocal digestion, each *Hae* III fragment was digested to completion with *Hin*dIII. Next, each uniformly labeled *Mbo* I fragment (A–J) was digested with *Hae* III. In reciprocal digestion, each *Hae* III fragment was cleaved with *Mbo* I. Results from these digestions (data not shown) give extensive information for ordering and mapping of most of the *Hae* III fragments with respect to *Mbo* I fragments. The order of several small *Hae* III fragments, such as *Hae* III-P and R, was obtained by partial digestion of certain single-end ^{32}P-labeled *Hin*dIII and *Mbo* I fragments with *Hae* III. The physical map for the *Hae* III cleavage sites is shown in Fig. 1.

Mapping of *Mbo* II Cleavage Sites. Complete digestion of uniformly ^{32}P-labeled BKV(MM) DNA with *Mbo* II produces 22 fragments. Digestion of ^{32}P-labeled *Hin*dIII fragments or *Mbo* I fragments by *Mbo* II was used in the construction of a preliminary map for *Mbo* II. The final physical map (Fig. 1) was constructed with additional information from partial digestion of certain single-end ^{32}P-labeled *Hin*dIII and *Mbo* I fragments with *Mbo* II.

DNA Sequence Analysis of BKV(MM) DNA. A "leader" RNA sequence has been found at the 5′ end of the 16S late mRNA which codes for VPI protein in SV40 (16, 17). This leader RNA, about 200 nucleotides long, which is coded by SV40 DNA at around 0.72–0.76 map positions (17), appears to be spliced to the 16S late mRNA. It is of interest to determine whether a similar leader RNA sequence exists in BKV(MM). Since such spliced leader sequences have also been found in adenovirus mRNAs (18–20), it seems likely that they constitute an essential signal that may be expected to be conserved in evolution. Comparison of DNA sequences of SV40 and BKV in the region around 0.70–0.76 map position may give insight about the minimum length of the leader sequence. The BKV(MM) DNA sequence in the region of 0.72–0.76 map positions was determined by labeling the 5′ ends of *Hin*dIII-A with ^{32}P, digesting the fragment with *Hae* III, and determining the sequence (21) from the single 5′ labeled end at 0.72 map position. A sequencing gel is shown in Fig. 3 and the sequence of 170 nucleotides (91–260) deduced therefrom in Fig. 4. Comparison with the SV40 DNA sequence in the same region (0.70–0.76 map positions, Fig. 4) shows 80% homology (solid underline) in a stretch of 103 nucleotides [nucleotides 114–216 in BKV(MM) DNA in Fig. 4]. The longest stretch of common sequence is 17 nucleotides long [157–173 on BKV(MM) DNA]. There is only a 30% homology from nucleotides 217 to 260.

The 5′-end-labeled *Hin*dIII-C_a fragment of BKV(MM) DNA was digested with *Hpa* II, to produce a single-end labeled fragment spanning 0.70–72 map units, and analyzed. Extensive homology (93%) was found between nucleotides 58–85 of

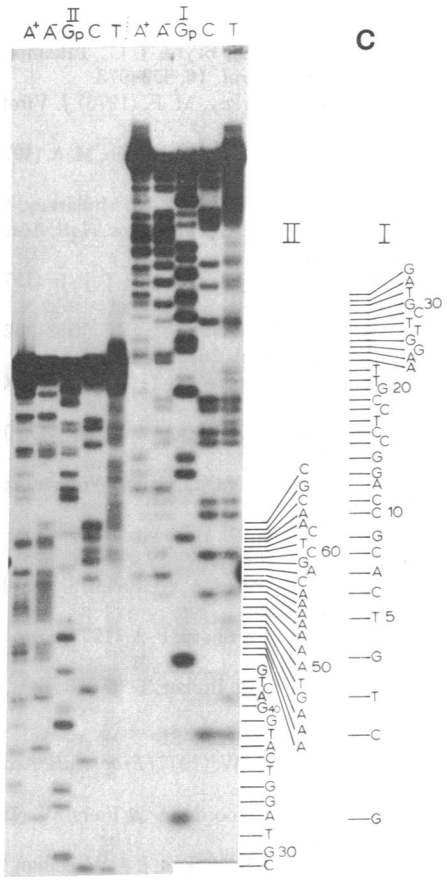

FIG. 3 (*continued*). The complementary sequence is shown on Fig. 4 as nucleotides 1–75. (*c*) A representative SV40 sequence is shown for the *Hin*d(II+III)-C *Hpa*(I+II)-D fragment, 5′-terminally labeled at the *Hpa* II site (nucleotide 5 corresponds to nucleotide 81 on Fig. 4). Other SV40 fragments whose sequences have been determined (gel patterns not shown) include *Hpa*-B *Alu*-R, *Alu*-R *Hpa*-B, *Hpa*-D *Alu*-R, *Hin*d-L, *Hin*d-M, and *Alu*-K (22). The sequence analysis of this region (0.71–0.76) was completed in March 1977. Over 99% of our data are in agreement with those of others (23, 24).

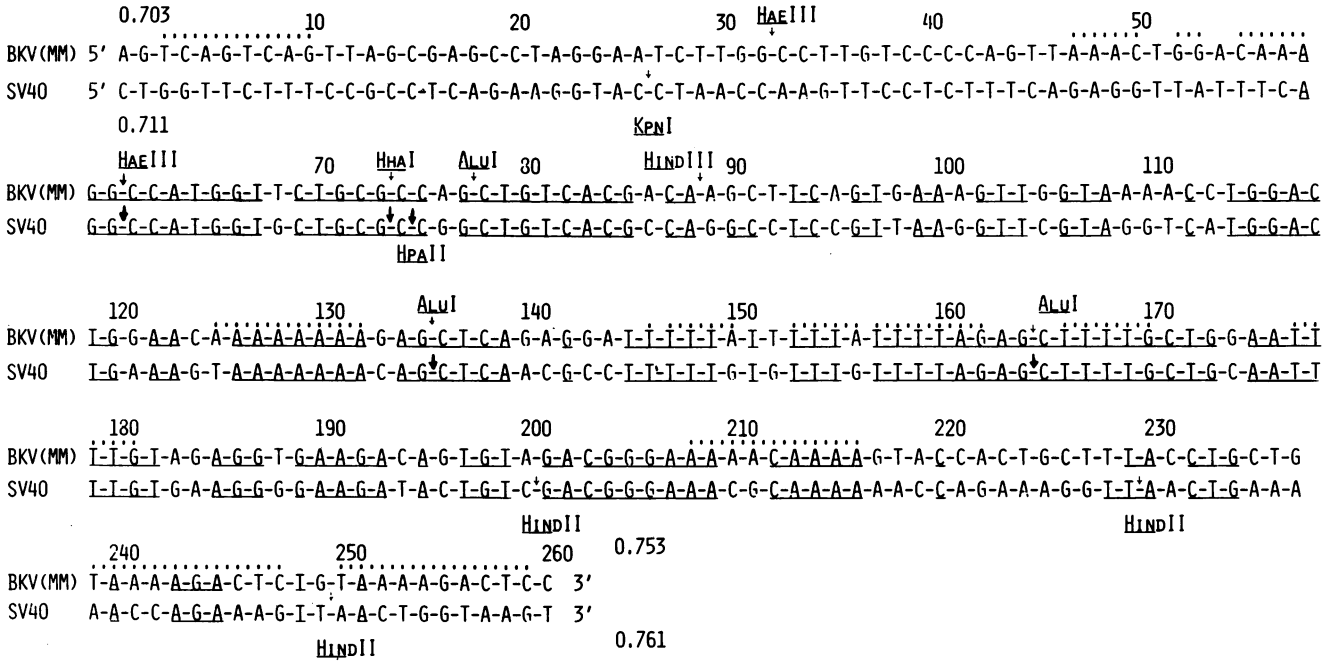

FIG. 4. Nucleotide sequences of the restriction fragments of BKV(MM) and SV40 DNA. A portion of the sequences are deducted from the autoradiograms shown in Fig. 3. Other portions of the sequences are deduced from other autoradiograms not shown. The sequences of certain regions are fully confirmed by the complementary sequence, e.g., 0.72–0.76 map positions on SV40 DNA. Sequences of other regions are determined from one DNA strand only, with at least 98% confidence. Nucleotides with solid underlines represent common sequences between BKV(MM) and SV40 DNA. Some sequences, marked with dotted lines, represent repeats or palindromes.

BKV(MM) DNA (Fig. 4) with those of SV40 DNA. However, almost no homology was found between nucleotides 1–57 of BKV(MM) DNA with that of SV40 in the corresponding region.

DISCUSSION

In SV40, viable mutants with short deletions (32–190 base pairs) around the *Hpa* II site have been isolated and shown to have reduced growth rates (25, 26). The deleted regions are within 0.72–0.76 map positions of that which codes for the leader RNA sequence (16, 17). Thus, at least a part of this region (0.72–0.76) is important in supporting the normal growth of SV40. If the common DNA sequence found in BKV(MM) at this region is also involved in coding for a leader mRNA sequence, it can be narrowed down to nucleotides 58–216 on the SV40 DNA (Fig. 4). It is likely that only part of this sequence constitutes specific recognition signals and is absolutely required. For example, nucleotides 58–85 and 153–182 show the highest homology to BKV(MM) sequence and therefore may be more important. Further work is needed to pinpoint the minimal DNA sequence required to code for a fully functional leader RNA.

We thank Miss J. Tomassini, Mrs. B. Hicks, and Mr. A. Young for valuable assistance. This work was supported by Research Grants CA-14989, awarded by the National Cancer Institute, Department of Health, Education, and Welfare, and VC-216 from the American Cancer Society. This is paper III in a series on "Nucleotide Sequence Analysis of Tumor Virus DNA." Paper II is by Tu *et al.* (27).

1. Roberts, R. (1976) *Crit. Rev. Biochem.* **4**, 123–164.
2. Danna, K. J., Sack, G. H. & Nathans, D. (1973) *J. Mol. Biol.* **78**, 363–376.
3. Gardner, S. D., Field, A. M., Coleman, D. V. & Hulme, B. (1971) *Lancet* **i**, 1253–1257.
4. Coleman, D. V., Gardner, S. D. & Field, A. M. (1973) *Br. Med. J.* **3**, 371–375.
5. Takemoto, K. K., Rabson, A. S., Mullarkey, M. F., Blaese, R. M., Garon, C. F. & Nelson, D. (1974) *J. Natl. Cancer Inst.* **53**, 1205–1207.
6. Howley, P. M., Khoury, G., Bryne, J. C., Takemoto, K. K. & Martin, M. A. (1975) *J. Virol.* **16**, 959–973.
7. Takemoto, K. K. & Mullarkey, M. F. (1973) *J. Virol.* **12**, 625–631.
8. Simmons, D. T., Takemoto, K. K. & Martin, M. A. (1977) *J. Virol.* **24**, 319–325.
9. Khoury, G., Howley, P. M., Garon, C., Mullarkey, M. F., Takemoto, K. K. & Martin, M. A. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 2563–2567.
10. Newell, N., Lai, C., Khoury, G. & Kelly, T. J., Jr. (1978) *J. Virol.* **25**, 193–201.
11. Jay, E. & Wu, R. (1976) *Biochemistry* **15**, 3612–3620.
12. Hirt, B. (1967) *J. Mol. Biol.* **26**, 365–369.
13. Wu, R., Jay, E. & Roychoudhury, R. (1976) *Methods Cancer Res.* **12**, 87–176.
14. Yang, R. C. A., Van de Voorde, A. & Fiers, W. (1976) *Eur. J. Biochem.* **61**, 101–117.
15. Smith, H. O. & Birnstiel, M. L. (1976) *Nucleic Acids Res.* **3**, 2387–2398.
16. Aloni, Y., Dhar, R., Laub, O., Horowitz, M. & Khoury, G. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3686–3690.
17. Celma, M. L., Dhar, R., Pan, J. & Weissman, S. M. (1977) *Nucleic Acids Res.* **4**, 2549–2559.
18. Berget, S. M., Moore, C. & Sharp, P. A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3171–3175.
19. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. (1977) *Cell* **12**, 1–8.
20. Klessig, D. F. (1977) *Cell* **12**, 9–21.
21. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
22. Yang, R. C. A., Van de Voorde, A. & Fiers, W. (1976) *Eur. J. Biochem.* **61**, 119–138.
23. Dhar, R., Subramanian, R. N., Pan, J. & Weissman, S. M. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 827–831.
24. Heuverswyn, A. H., Van de Voorde, A. & Fiers, W. (1977) *Nucleic Acids Res.* **4**, 1015–1024.
25. Shenk, T. E., Carbon, J. & Berg, P. (1976) *J. Virol.* **18**, 664–671.
26. Mertz, J. E. & Berg, P. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 4879–4883.
27. Tu, C. D., Roychoudhury, R. & Wu, R. (1976) *Biochem. Biophys. Res. Commun.* **72**, 355–362.