Linker Scan Analysis of the Early Regulatory Region of Human Papovavirus BK

KAREN L. DEYERLE AND SURESH SUBRAMANI*

Department of Biology, University of California at San Diego, La Jolla, California 92093

Received 15 March 1988/Accepted 7 June 1988

BK virus (BKV) is a human papovavirus which latently infects a majority of the world population. Reactivation of this virus is associated with acute hemorrhagic cystitis, and BKV DNA has been found in human tumor tissue. BKV is one of many highly homologous papovaviruses, including simian virus 40 and JC virus, which display distinct host and cell-type specificities, transformation potentials, and pathologies. These differences are thought to be determined, in part, by the noncoding regulatory region of each virus, which contains the origin of replication and regulatory elements for both early and late gene expression. We have used linker scan mutants to map functional elements of a truncated BKV early promoter and enhancer and have studied the stereospecific requirements of these elements. We have also identified protein-binding regions through DNase protection studies. Our results show that a minimum of four elements are necessary for efficient early transcription, at least three of which correspond to DNase-protected domains. These protein-binding elements map to the TATA box and two nuclear factor 1 consensus sequences, one located within the enhancer repeat unit and the other located to the late side of the enhancer. The sequence of the fourth element is similar to the transcription factor Sp1 consensus sequence. Additional DNase-protected sites are centered over AP-1 and Sp1 consensus sequences. Finally, we find that the functional elements of the BKV early promoter and enhancer lack strict stereospecific requirements for efficient transcription.

The human papovavirus BK, first isolated in 1971, infects 70 to 80% of the world population during early childhood (17, 49). Following the initial inapparent infection, the virus enters a latent state, with viral DNA persisting in an episomal form in the kidney (21). Subsequent immunosuppression of the host can lead to the reactivation of the virus and the excretion of viral particles, a condition strongly associated with the development of hemorrhagic cystitis (2). BK virus (BKV) is also known to be highly oncogenic in rodents, where it induces a wide variety of extraneural tumors (48). Examination of extraneural human tumor tissue revealed the presence of BKV DNA in 27% of the 57 tumors studied (4, 48). The possibility therefore exists that this virus is involved in the development of several different types of human pathology.

The 5,196-base-pair (bp) BKV genome is organized and expressed in a manner similar to that of other papovaviruses (49). The coding regions for the two early (T and t antigens) and four late (agno and capsid) proteins are separated by a noncoding regulatory region of approximately 400 bp. This noncoding regulatory region contains the transcriptional control elements for both early and late gene expression, and the viral origin of replication. It consists of a true palindrome, two inverted repeats, a 20-bp AT block, three direct 68-bp repeats (the middle repeat has an 18-bp deletion), and distal flanking sequences (49). Although the noncoding regulatory regions of all papovaviruses bear a superficial resemblance to one another, these regions are known to determine, in part, the various host and cell-type specificities and transformation efficiencies of each virus. The tumorigenic and transformation capacity of BKV, in particular, has been correlated with rearrangements in the viral regulatory region (52, 53).

A previous deletion analysis of the BKV noncoding regulatory region identified several early gene transcriptional

control elements (12). This work defined the BKV enhancer as consisting of the three direct 68-bp repeats plus an element, designated the c element, located adjacent to the late side of the distal 68-bp repeat. It was found that in the presence of the c element, the 68-bp repeat units were functionally redundant. Within the 68-bp repeat, two enhancer motifs were identified, plus a promoter element located in the 18 bp proximal to the AT block. Other regulatory elements were found downstream of this region.

We now present a linker scan analysis of this region, using a BKV construct containing a single 68-bp repeat. This truncated enhancer was studied so that the presence of multiple redundant repeat units would not mask the effects of mutations. Protein-DNA contact points have also been identified by DNase protection studies, using both the wild-type and truncated control regions. We show that efficient transcription from the early regulatory region requires a minimum of four regions. DNase protection assays revealed that at least three of these regions are protein-binding domains. One of these domains is located in the 20-bp AT block, while the other two center over consensus sequences for the nuclear factor 1 (NF1) family of transcription factors (13, 26, 35, 41). DNase protection is also seen in a region containing the transcription factor AP-1 consensus sequence (33). A sequence similar to the Sp1 consensus sequence (3) is shown to be functionally significant, although only minor DNase protection is seen in this region. The previously identified simian virus 40 (SV40) and adenovirus E1a enhancer core motifs (42), contained within the 68-bp repeat unit, do not appear to correlate with the presence of functional regulatory elements.

MATERIALS AND METHODS

Materials. Restriction enzymes, T4 DNA ligase, DNA polymerase, S1 nuclease, nucleotides, and linkers were purchased from Boehringer Mannheim Biochemicals, Bethesda Research Laboratories, Inc., or P-L Biochemicals,

* Corresponding author.

FIG. 1. BKV noncoding regulatory region. A segment of BKV DNA (nt 3362 through 3757), encompassing the early transcriptional regulatory region, was isolated as a *Hind*III-*Bam*HI fragment. This segment contains a palindrome (Pal), two inverted repeats (IR), a 20-bp AT block (AT), three direct repeats, the c element (c), and late flanking sequences. The start sites of early mRNAs are shown as arrows, and the locations of the *Mst*II restriction sites used in the construction of mutants are indicated. The 68-bp repeat closest to the AT block is referred to as proximal, and the 68-bp repeat furthest away as distal. (A) The plasmid pSV2ABK3362 contains the entire 395 bp of the BKV noncoding regulatory region. It was used as the starting plasmid in the construction of mutants pBK68 and the LS series. (B) The LS plasmids were constructed by combining segments of the BKV control region from appropriate ΔB and ΔH unidirectional deletions of pSV2ABK3362 as shown. The LS plasmids contain a truncated enhancer region, in which the 50-bp repeat and one 68-bp repeat are deleted, and *n* base pairs of the remaining 68-bp repeat are replaced by an *Xho*I linker. (C) The plasmid pBK68, containing a single 68-bp repeat, was constructed by complete digestion of pSV2ABK3362 with *Mst*II, followed by self-closure ligation. Similarities to previously identified regulatory elements are indicated. □, Sp1 consensus sequence; ---, NF1-binding site; ▨, SV40 enhancer core; ■, E1a enhancer core.

Inc. Bal 31 (slow) was purchased from International Biotechnologies, Inc. [α - 35 S]dATP and [14 C]chloramphenicol were obtained from Amersham Corp., acetyl coenzyme A was from Sigma Chemical Co., and luciferin was from Analytical Luminescence Laboratory.

Plasmid constructions. The construction of the plasmid pSV2ABK3362 has been described previously (12). This plasmid contains 395 bp of BKV(wt) DNA, nucleotides (nt) 3362 through 3757, encompassing the noncoding regulatory region and late flanking sequences, on a *Hind*III-*Bam*HI fragment (Fig. 1A). The location and orientation of this BKV DNA is such that expression of the downstream gene encoding chloramphenicol acetyltransferase (CAT) is driven by the BKV early promoter and enhancer. Transcriptional activity of the BKV region in this plasmid, and all derivatives thereof, can thus be monitored by assaying CAT activity.

The plasmids pBK68, pBK68-50, and pBK68-68 were derived from pSV2ABK3362 by complete or limited digestion of the latter with *Mst*II (Fig. 1A), followed by self-closure ligation of the resulting large DNA fragment. Complete digestion of pSV2ABK3362 with *Mst*II removed the 50-bp element and one 68-bp repeat of the BKV enhancer,

leading to the formation of pBK68. The BKV region in this plasmid contains a single 68-bp repeat in its early regulatory region (Fig. 1C). A limited digest of pSV2ABK3362 with *Mst*II resulted in the formation of plasmids containing two 68-bp repeats (pBK68-68) or one 68-bp and one 50-bp repeat (pBK68-50). In the latter case, the 68-bp repeat was located proximal to the BKV origin of replication.

(i) **Linker scan mutants.** The construction of unidirectional deletion mutants starting with pSV2ABK3362 has been described earlier (12). Plasmids of the ΔB series contain unidirectional deletions beginning at the *Bam*HI site and extending in the early direction, while ΔH series plasmids contain unidirectional deletions beginning at the *Hind*III site and extending in the late direction (Fig. 1B). All of the unidirectional mutants used contained *Xho*I linkers (CCTC GAGG) at the deletion endpoint. The LS (linker scan) series of plasmids was constructed by combining ΔB plasmids with deletion endpoints between BKV nt 3434 and 3522 (plus ΔB3645) with ΔH plasmids with deletion endpoints between BKV nt 3573 and 3659 (Fig. 1B).

The LS plasmids were designed such that approximately 8 bp of BKV sequence would be replaced by an 8-bp *Xho*I linker (CCTCGAGG). To accomplish this, ΔB and ΔH

FIG. 2. Construction of BKV deletion plasmids. The plasmid pSV2ABK3362 contains the entire BKV noncoding regulatory region (nt 3362 through 3757) as a *HindIII*-*Bam*HI fragment. Unidirectional deletions of this region were made by cutting pSV2ABK3362 at either the *HindIII* site (Δ H series) or *Bam*HI site (Δ B series), deleting with *Bal* 31, adding *Xho*I linkers, and recircularizing the resulting DNAs. The linker scan mutants of the LS series were constructed by combining the promoter-containing *Bgl*II-*Xho*I fragments of appropriate pairs of Δ H and Δ B plasmids. The LS plasmids contain a single 68-bp repeat unit in which a small number of base pairs have been replaced by an *Xho*I linker. These plasmids contain sequences from pBR322 (thin lines), including the ampicillin resistance gene (*amp*^R), and sequences from SV40 (thick lines). A_n, polyadenylation signal.

unidirectional mutants were identified which had endpoints separated by 6 to 10 bp of BKV sequence. The large *Bgl*II-*Xho*I Δ B fragment was then ligated to the small *Bgl*II-*Xho*I Δ H fragment (Fig. 2). The BKV regulatory region in the resulting plasmid contained a single 68-bp repeat in which a small number of base pairs had been replaced by the sequence CCTCGAGG (Fig. 1 and 2). The majority of plasmids either were perfect linker scan mutants (the 8-bp linker replacing 8 bp of BKV sequence) or contained small insertions or deletions. A few plasmids were designed to test stereospecific requirements of BKV regulatory elements, and these contained insertions of either half or whole turns of the DNA helix (5 or 10 bp, or multiples thereof). The sequence of all plasmids (Fig. 3) was confirmed by sequencing both strands with the double-stranded DNA dideoxyse-

quencing method (5), using the incorporation of [α -³⁵S]dATP to label the reaction products. Sequencing products were analyzed by electrophoresis on 8% polyacrylamide denaturing (8 M urea) gels.

(ii) **Mutations of the SV40 enhancer core.** Site-directed mutagenesis was used to introduce a 5' TGG-to-GTT 3' mutation at nt 3493 through 3495. The BKV regulatory region of pBK68 was inserted into the vector M13mp19 as a *HindIII*-*Bam*HI fragment, and single-stranded bacteriophage DNA containing this insert was plaque purified. A synthetic 22-nt oligonucleotide containing the mutation was then used following standard procedures (31) to generate phage containing the desired base changes in the BKV region. Mutants were screened for by the dideoxysequencing technique, by using the single-stranded phage DNA as a

template (43). Replicative-form phage DNA containing the mutation was purified, and the BKV regulatory region was excised as a *HindIII-BamHI* fragment. This mutated region was then inserted into the pBK68 in place of the truncated wild-type regulatory region.

CAT and luciferase assays. CV1 African green monkey kidney cells were used for all assays, with cell culture care and transfections carried out as previously described (12). Four independent transfections of each BKV plasmid were done, with each experiment using a cotransfection of 9.5 μ g of BKV plasmid with 500 ng of the plasmid pRSV-L. The latter contains the luciferase coding region driven by the Rous sarcoma virus promoter (11) and was included to correct for transfection efficiency. The Rous sarcoma virus promoter does not compete for limiting transcription factors with the BKV early promoter under the assay conditions used (47). The protein extracts were made 46 to 50 h posttransfection following published procedures, and CAT (12) and luciferase (11) activities were assayed as previously described. The plasmid pBK68 was used as the positive control in each experiment, and the data are presented with the activities of all plasmids normalized to the activity of pBK68.

DNase protection experiments. The BKV plasmids chosen for study were first digested with *BamHI* and then were labeled by filling in the resultant single-stranded ends with the large fragment of DNA polymerase I and [α - 32 P]dCTP, plus unlabeled deoxynucleotides. The BKV regulatory region of each plasmid was then excised as a *HindIII-BamHI* fragment; these DNAs, ranging in size from 277 to 395 bp, were used as a substrate in subsequent footprinting reactions. Whole HeLa cell extract, partially purified by elution with 0.4 M KCl from a heparin-agarose column, was generously provided by the laboratory of Michael Karin. Footprinting reactions were carried out following published procedures (28), and the reaction products were electrophoresed on 8% acrylamide denaturing gels.

RESULTS

Previous studies have shown that three direct repeats in the BKV enhancer are functionally redundant; mutations in one repeat are thus masked by the presence of compensating elements in the other two repeats (12). Further, it has been shown that two repeats plus late flanking sequences retain full early transcriptional activity, while reduction to one repeat decreases activity (12). A detailed analysis of the early transcriptional regulatory elements in BKV therefore required study of a control region containing a single repeat, as the presence of two or three repeats would permit elements to compensate for one another.

The plasmid pBK68 contains such a truncated BKV early regulatory region (nt 3362 through 3499 and 3617 through 3757 [Fig. 1]). Linker scan and small insertion and deletion mutants of this region were constructed (Fig. 1 and 2); the position and sequence of each mutant is shown below the BKV sequence which it replaced in Fig. 3. Transcriptional activity was tested using a transient expression assay system in CV1 cells; expression from the BKV early regulatory region has been shown to retain its tissue specificity across species lines (50). All activities are the averages of four independent transfections, corrected for transfection efficiency, and normalized to the activity of the control plasmid pBK68.

Overview of BKV early regulatory elements. Earlier inspection of the sequence of the BKV early regulatory region had

revealed similarities to known transcriptional regulatory elements defined in other viruses (Fig. 1 and 3). These include the TATA box, contained within the 20-bp AT block (nt 3432 through 3451), an Sp1 consensus sequence (nt 3459 through 3464) (29), and similarities to the SV40 enhancer core (nt 3489 through 3500) and the adenovirus E1a enhancer core (nt 3503 through 3518) (42). Partially purified NF1 has been shown to bind to BKV(Dun) DNA in this region (nt 3477 through 3491) (39). Although we use the term NF1 with reference to our work, we do not distinguish between this factor and the multiplicity of proteins known to bind to this and similar sequences (13, 26, 41). Finally, deletion analysis has revealed the presence of an element, the c element, located in the 30 bp adjacent to the late side of the distal 68-bp repeat (12). The functional role of each of these was assessed by making several mutations in each area and testing them for transcriptional activity (Fig. 3 and 4).

Functional importance identified by linker scan mutants. Two mutations, LS11 and LS14, tested the effect of the TATA box on promoter function by removing all but 3 bp of the 20-bp AT block, replacing it with unrelated sequences. Both of these mutants had a similar effect on CAT activity, reducing it approximately twofold compared with that of the control. This result was unexpected, since other studies on this and related viruses had suggested that the TATA box sequence lacked effect on the overall rate of transcription (12, 16).

The GC box (Sp1 consensus sequence) was extensively substituted, being affected by eight different mutations (LS mutants 2, 16, 19, 28, 29, 30, 37, and 38). The majority of these showed a fivefold decrease in CAT activity from control (Fig. 4). These data strongly suggest a functional role for the GC box in transcriptional activity from the BKV promoter, consistent with earlier studies which mapped a promoter element between nt 3454 and 3472 (12).

The NF1-binding site, consisting of the symmetrical consensus sequence 5' TGG(N₆₋₇)GCCAA 3' (39) was extensively mutated. The majority of mutants introducing base substitutions in this minimal consensus sequence (LS mutants 15, 17, 18, 33, and 35) decreased activity threefold below that of the control. Alteration of the variable spacer sequence between the inverted repeats (LS43) had less than a twofold (66%) effect on transcriptional activity. Mutants LS4 and LS6, which contain the same transversion mutation at nt 3478 and altered 5'-flanking sequences (LS4 = 5' CTCCGA 3'; LS6 = 5' CCTCGAG 3'), had various CAT activities (17 and 88%, respectively), which reflected their effect on protein binding to this region (see below). These results are compatible with a model in which NF1 plays a role in mediating early BKV transcription by binding to this region.

Adjacent to and overlapping with the NF1 site in BKV is a sequence similar to the SV40 enhancer core sequence (42). This sequence has been shown to be both a functional motif in SV40 enhancer activity (56) and a binding site for transcription factor(s) (6, 23, 25). It was of interest to determine if this motif had functional significance in the context of the BKV enhancer. We used site-directed mutagenesis to introduce the mutation 5' GTTTTT 3' in place of the wild-type 5' TGGTTT 3' (nt 3490 through 3495). When present in the SV40 72-bp repeat, this mutation has been shown to cause a 5- to 10-fold reduction in enhancer activity in five different cell lines, including HeLa and CV1 cells (38). Within the context of the BKV enhancer, when assayed in CV1 cells, this same mutation resulted in a twofold decrease in activity.

These data argue against a dominant functional role for the SV40 enhancer core in the BKV 68-bp repeat.

A sequence similar to the adenovirus E1a enhancer core sequence is located in the distal portion of each repeat unit (Fig. 1 and 3) (42). Four of the six mutations in this region (LS22, LS39, LS40, and LS42) had essentially wild-type CAT activity (73, 98, 110, and 101%, respectively). These data suggest that the adenovirus E1a enhancer core is nonfunctional in the BKV enhancer in CV1 cells.

The c element, previously shown to enhance transcription two- to fourfold, is located in a region with approximate boundaries of nt 3641 through 3675 (Fig. 1 and 3) (12). With one exception (LS34), mutations affecting this region consistently produced a threefold decrease in CAT activity (LS26, LS27, and LS31).

Spacing alterations have little effect on BKV early transcription. It has been demonstrated that efficient cooperativity between transcription factors may require their specific location on one or the other side of the DNA helix with respect to each other (30, 51). To assess stereospecific requirements between the different BKV transcriptional control elements, mutants were constructed which altered the spacing between different functional sequences. Sequences corresponding to half (4 to 6 bp), whole (10 to 12 bp), or 1.5 turns (15 to 18 bp) of the DNA helix were introduced between the AT block and the 68-bp repeat, and between the GC box-NF1 site and the c element. These mutations were then tested for their effect on CAT activity.

Increasing the distance between the BKV TATA box and the 68-bp repeat diminished transcriptional activity. This decrease was found to be a function of absolute distance, with the introduction of a whole turn (LS3, 23% activity) being much more deleterious than insertion of a half-turn (LS1, 62% activity). Other data indicate a lack of strict stereospecific requirements between the downstream (GC box, NF1 site) region of the 68-bp repeat and the c element. Insertion of a half-turn at nt 3500 (LS21, 66% activity), a whole turn at nt 3516 (LS10, 76%), and 1.5 turns at nt 3515 (LS12, 79%) had less than a twofold effect on transcriptional activity. These results are surprising in that similar studies in the closely related SV40 have shown strict stereospecific requirements between the 21-bp repeats and the enhancer (51), and to a lesser extent, between the TATA box and the 21-bp repeats (30, 51). Further, footprinting studies have shown that the binding of protein to the BKV NF1 site affects the binding of a protein to the c element site (see below). Since NF1 is thought to bind to only one side of the DNA helix (10, 39), it might be expected that a particular relative alignment of the protein binding to the c element would maximize their interaction. In light of these results, the lack of stereospecific requirements between functional elements in the BKV early regulatory region is all the more surprising.

Protein-binding sites correlate with functionally important regions. DNase protection experiments were performed, using partially purified HeLa whole cell extracts to identify

FIG. 4. CAT activities of the BKV linker mutant plasmids. The CAT activity of each linker scan mutant, normalized to that of the control plasmid pBK68, is shown in a graphical representation. The number of each LS mutant is shown on top of the vertical bar indicating the activity of that mutant. Similarities to transcriptional regulatory elements contained within the BKV control region are indicated above the bar graph. Bold lines beneath the name of the similarity indicate that protein binding to that region in BKV is detected in vitro, while thin lines indicate that no protein binding has been detected. The lines are drawn such that all mutants affecting that region are shown below the line. Ad 5, Adenovirus type 5; NF1/CTF, NF1/CCAAT transcription factor.

protein-binding sites. Initial experiments were done using the entire BKV noncoding regulatory region, nt 3362 through 3757 (Fig. 2A), as the DNA substrate. The results showed that the sequences determined to be of functional importance generally coincided with protein-binding regions (Fig. 5). Since a partially purified cell extract was used in these studies, it is possible that not all protein-binding sites in this region were detected.

Footprint analysis revealed unusual alterations in DNase sensitivity over portions of downstream regions and the inverted repeat (nt 3409 through 3431) that constitutes the bidirectional origin of replication (Fig. 5A, boxes 1, 2, and 3). The significance of this pattern with respect to protein binding is unclear. The major early mRNA start sites, located just inside the early boundary of this inverted repeat (12), fall just outside the unusual pattern of altered DNase sensitivity in this region.

A protein footprint was located within the 20-bp AT block, spanning nt 3434 through 3454 (Fig. 5A, box 4). This region, deleted by two of our mutants (LS11 and LS14), was found to have an effect on activity from the BKV promoter. Further upstream, minor protection from DNase digestion was seen in the region of the GC box (Fig. 5A, box 5). Although this sequence has been shown to be of functional

FIG. 3. BKV linker scan mutants. The nucleotide sequence of the 20-bp AT block, a single 68-bp repeat unit, and 39 bp of late flanking sequences contained within the truncated BKV regulatory region of pBK68 is shown. Similarities to previously identified transcriptional regulatory elements are indicated above the sequence. Numbering follows that of the BKV early region and the proximal 68-bp repeat to the junction with the late flanking sequences, which are numbered according to their position following the distal 68-bp repeat in BKV. The locations of the LS mutations are shown, and the normalized CAT activity of each plasmid is shown in bold type directly over the plasmid name. The LS mutants are arranged in roughly decreasing order of CAT activities from the top to the bottom of the figure. The bar indicates the extent of the BKV nucleotides deleted in each plasmid, and the mutant sequence replacing them is shown below the bar. The insertion (+) or deletion (-) of extra base pairs compared with wild-type sequence is indicated in parentheses after the plasmid name. Ad5, Adenovirus type 5.

FIG. 5. Proteins bind in vitro to the BKV noncoding regulatory region. Whole HeLa cell extracts, partially purified by passage over a heparin-agarose column, were used for footprinting experiments. The DNA substrate shown is the entire BKV noncoding regulatory region, excised from pSV2ABK3362 as a *Hind*III-*Bam*HI fragment, and labeled with [α - 32 P]dCTP at the *Bam*HI site. (A) A schematic representation of the BKV noncoding regulatory region, showing the palindrome (Pal), inverted repeats (IR), 20-bp AT block (AT), three direct repeats, c element (c), and late flanking sequences. Numbered boxes indicate the locations and extent of altered patterns of DNase sensitivity. Regions of clear protein protection (■), and regions with altered DNase sensitivity but lacking in clear protein protection (▨) and homologies to known transcriptional regulatory elements are shown. □, Sp1 consensus sequence; ---, NF1-binding site; ▨, SV40 enhancer core; ■, adenovirus enhancer core. (B) Regions of altered DNase sensitivity on the BKV early coding strand, with the approximate boundaries of

significance, it is unclear if this area constitutes a protein-binding site.

Complete protection from DNase was seen in the regions previously identified by others (39) as NF1-binding sites (Fig. 5A, boxes 6 and 9). In the proximal 68-bp repeat, this footprint spans nt 3472 through 3497 and is virtually identical to that produced by NF1 partially purified from chicken liver nuclei (39) and HeLa cells (36). This footprint was also found in the corresponding region of the distal 68-bp repeat, but not in the 50-bp repeat. An 18-bp deletion in the latter results in the removal of the first nucleotide (T) of the NF1 recognition sequence.

Both junctions between repeat units were clearly shown to be protein-binding regions. Between the proximal 68-bp and 50-bp repeat, nt 3514 through 3533 was protected (Fig. 5A, box 7), while a larger footprint, spanning nt 3558 through 3586, was seen over the 50-bp and distal 68-bp repeat boundary (Fig. 5A, box 8). Although these protected regions were of unequal size, competition experiments demonstrated the equivalence of these two sites. A DNA fragment containing the downstream site (nt 3514 through 3533) could successfully inhibit the upstream footprint between nt 3558 and 3586, and a fragment containing the upstream site could inhibit the downstream region footprint (data not shown). We conclude that the proteins binding to these two regions are closely related, if not identical.

A single area of protection (nt 3636 through 3659) was seen spanning the junction between the distal 68-bp and late flanking sequences (Fig. 5A, box 10). The position of this footprint correlates well with the location of the functional c element, as defined by our mutation analysis here and elsewhere (12). Yet another footprint, spanning approximately 15 to 20 bp, is found centered over the Sp1 consensus sequence located between nt 3673 through 3682 (Fig. 5A, box 11). This weakly protected area can be deleted without effect on the activity of the early promoter/enhancer (12), and thus is not required for BKV early transcription.

Additional footprinting experiments were performed with the 277-bp *Hind*III-*Bam*HI fragment (Fig. 1C) from pBK68 to verify that the pattern of protein binding to the BKV regulatory region was not artificially altered by reduction of the enhancer to a single 68-bp repeat. Corresponding patterns of DNase digestion were found to be identical between the intact and truncated regulatory regions, with the exception that the footprints spanning the repeat junctions were absent in pBK68 (data not shown).

Certain linker scan mutants were chosen for footprinting experiments, to assess the effect of the mutation on protein binding (data not shown). These results show that the LS29 mutation in the GC box (5' GGAGGAGC to CCTCGAGG 3') causes major alterations in the DNase-protection pattern throughout the 68-bp repeat. We therefore believe that the activity of LS29 (66%) is a sequence-specific aberration, not indicative of the functional importance of the GC box. The alteration of the NF1-flanking sequences introduced by LS6 (5' TTACCCAT to CCTCGAGG 3') does not have any apparent effect on the binding of that or any other protein in

each footprint indicated by arrowheads. The number between the arrowheads corresponds to the appropriate box in panel A, indicating the position of that footprint within the BKV noncoding regulatory region. Footprints 6, 9, and 10 show protein protection of the NF1 consensus sequences, while footprints 7 and 8 protect AP-1 consensus sequences. Lanes a are control lanes, whereas lanes b show protection by HeLa cell extract.

this region, commensurate with the wild-type CAT activity of this mutant. In contrast, a similar mutation in LS4 (5' ACCCATG to CTCCGAGG 3'), which decreases CAT activity fivefold, causes diminished protection of both the NF1 site and the c element site. These data correlate well with the functional data obtained with these mutants.

DISCUSSION

We have shown that a minimum of four regions, at least three of which constitute protein-binding sites, are required for maximal transcriptional activity from the BKV early regulatory region. These regions are the TATA box, GC box, a binding site for the NF1 family of proteins, and the c element. In addition, three other protein-binding regions have been identified. One of these is located at the boundary between 68-bp and 50-bp repeats, one at the junction of the 50-bp and 68-bp repeats, and one to the late side of the c element.

Substitutions in the 20-bp AT block of BKV show that this sequence augments transcriptional efficiency from the BKV early promoter. Factors are known to bind to the TATA sequence in eucaryotes (8, 44, 54), and we have shown that this region is protected from DNase digestion in BKV. Although the TATA box is known to contribute to transcriptional efficiency in some genes (14, 55), the activity of the BKV AT block is in contrast with the SV40 TATA box, which does not affect the overall level of transcription (16).

Both this and earlier studies (12) have found that the GC box in the downstream region of the proximal 68-bp repeat is essential for promoter activity. It was therefore surprising that we detected only minor protection from DNase digestion in this region. While not excluding other formal possibilities, we suggest that the relatively crude nature of our HeLa extract hampers detection of low-affinity protein binding to this region. On the basis of sequence homology and footprinting data of the GC box, we propose that an Sp1-related protein binds to this sequence and plays a role in BKV early gene transcriptional regulation.

The downstream region of each BKV 68-bp repeat contains an 8 of 10 match (AGGGAGGAGC) to the SP1 consensus binding sequence (G/T G/A GGCG G/T G/A G/A C/T) (3), and Sp1 has been shown to bind to the closely related sequence TGGGAGGAGC (14). The faint DNase protection that we see over the BKV GC box resembles the weak protection of SV40 GC boxes I and II by partially purified Sp1 (18). Interestingly, low-affinity Sp1 sites (e.g., SV40 GC box 1, herpes simplex virus thymidine kinase GC box 1) are often important in promoter activation (29). Like the BKV GC box, these critical Sp1 sites are usually located proximal to the gene, 40 to 70 bp upstream of the mRNA start site (29). Although most Sp1-responsive promoters contain multiple binding sites (29), a single GC box is sufficient for Sp1 binding and transcriptional activation (27, 32, 55), particularly when positioned in close proximity to the TATA box (28, 32) as is the BKV GC box. Finally, Sp1 may act in conjunction with other transcription factors (NF1 and AP-1), which are also found to bind to the BKV enhancer region (3, 32, 36). Since purified Sp1 does not bind to the GC box in BKV(MM) (36), an Sp1-related (i.e., recognizing a similar sequence as Sp1) protein may be involved in BKV early transcription.

Immediately upstream from the GC box in each of the 68-bp repeats is the consensus sequence TGG(N₆₋₇)GCCAA for the NF1 family of transcription factors (10, 13, 19, 20, 26, 34, 41). We use the term NF1 to refer to all members of this

family and do not distinguish among them. Although NF1 does not bind to all known consensus sequences (20), it has been shown to protect these two regions of BKV(Dun) (34, 39). Significantly, the footprint obtained with the HeLa cell extract coincided with this NF1 footprint, and mutants of this site with impaired transcriptional activity also show decreased DNase protection in this region.

The activity of our mutants in this region is consistent with their expected effect on NF1 binding. All five mutations of one of the two inverted repeats of the minimal NF1 recognition sequence caused decreased activity, as did mutations of the flanking sequences. Close contacts have been observed between NF1 and the DNA backbone phosphate groups located outside the minimal recognition sequence (34), and the flanking sequences are known to modulate binding affinity (10, 34, 35, 46). Significantly, the activity of our flanking sequence mutants reflects their effect on protein binding to this region. Finally, base changes in the center of symmetry of the NF1 site caused a decrease in CAT activity. This region is not conserved in the NF1 sites of adenovirus variants (9), and studies have shown that NF1 binding will tolerate base substitutions (19, 20), but not spacing changes (37, 39), in this region. It has been found that alteration of the central 5 bp of the NF1 consensus sequence decreases binding of the protein twofold (46), in good agreement with our functional data.

This mutation analysis and footprinting data strongly support a role for a member of the NF1 family of proteins as a transcriptional regulatory factor in the BKV early promoter. The presence of the NF1 consensus sequence in the 5'-flanking sequences of many genes (e.g., human immunoglobulin M, *c-myc*, chicken lysozyme, mouse mammary tumor virus long terminal repeat, human cytomegalovirus major immediate early gene [19, 20, 37]) suggests a transcriptional role for this protein. In the context of the mouse mammary tumor virus long terminal repeat, NF1 has been found to contribute quantitatively to transcriptional expression without affecting the mRNA start site selection (37). The location of the BKV NF1 site, within the 68-bp repeat, is of interest for two reasons. First, NF1 was originally isolated on the basis of its ability to stimulate DNA replication from the adenovirus origin of replication (35). In the proximal 68-bp repeat of BKV, the NF1 site is in close proximity to the BKV origin of replication; NF1 may therefore function both in replication and transcriptional regulation in this virus. Second, the 68-bp repeat region of BKV has been shown to possess the properties of an enhancer. Although the NF1 site itself, in either single or multiple copies, lacks enhancer function (24, 39), it may contribute to enhancer function when present in combination with binding sites for other factors, as it is in the BKV 68-bp repeat. It is interesting to note that the repeat regions of two human papovaviruses, BK and JC, which are thought to contain overlapping promoter and enhancer functions, both contain NF1 sites (15). The closely related SV40, with clearly separable promoter and enhancer regions, lacks NF1 sites (39).

The SV40 enhancer core sequence (GTGGAAAG) has been shown to be the dominant functional motif of the 72-bp repeat (40, 45). Homologies to this sequence are found in several enhancers, where it appears to be of functional significance in a variety of cell types (25, 38, 40, 45). Protein factors have been shown to bind to this sequence (6, 23, 25), and core mutations decrease protein binding and drastically impair enhancer function (6, 22, 56). Mutation of the BKV 68-bp repeat core sequence produced only a two-fold de-

crease in CAT activity, inconsistent with a dominant functional role for this motif in the BKV enhancer. It is possible that in BKV the presence of NF1, which protects the core sequence as well as its minimal binding sequence from DNase digestion, sterically hinders the binding of any protein recognizing the core sequence.

Mutations of the adjacent adenovirus E1a enhancer core sequence were found to have little, if any, effect on transcriptional efficiency from this region. This region was also found to be devoid of any protein protection in footprinting experiments. We conclude that, in CV1 cells, both the SV40 and adenovirus enhancer core homologies lack significant function in BKV early transcriptional regulation. This does not rule out their use in other cell types, as the function of enhancer motifs and the presence of transcription factors are known to be cell type specific (7, 38, 40, 45).

Both functional and footprinting data indicate the existence of an enhancer element, the c element, directly adjacent to and slightly overlapping with the late side of the 68-bp repeat. We have compared the protected sequence to known transcription factor consensus sequences, and two of these present a reasonable match. The first, AP-1, has the consensus sequence C/G TGACT C/A A (33), with a 5 of 8 match to the BKV sequence. The asymmetry of the c element footprint around the AP-1 consensus sequence makes it unlikely that pure AP-1 is binding to this site (1, 33), and purified AP-1 does not bind to this region *in vitro* (36). The second is a 1-bp mismatch from the NF1 protein family consensus sequence TGG(N₆₋₇)GCCAA, centrally located within the protected region (10, 13, 19, 20, 26, 34, 41). A protein purified by affinity chromatography on a column containing the BKV NF1 site (GATCTGGAATGCAG CCAA) has been shown to bind to this sequence in BK(MM) (36). Mutants which destroy one inverted repeat of this binding site, or disrupt the normal flanking sequences, show decreased CAT activity, consistent with our data from corresponding mutants affecting the downstream NF1 site. The wild-type activity of LS34 bears further investigation.

It is interesting that none of our mutants decrease CAT activity more than fivefold. The multiplicity of proteins shown to bind to the BKV and other viral enhancers (7, 38) may include functionally redundant factors. BKV early transcriptional activation appears to require a minimum of two elements acting in concert, consistent with our observations that neither the GC box or c element alone is capable of promoting transcription (12).

Finally, a footprint located to the late side of the c element (Fig. 5, box 11) centers on an Sp1 consensus sequence (TCGGCGGGGT). This sequence can be deleted without effect on early transcription (12). We have yet to assess the functional role of the protein-binding sequences located at the junctions of the repeat units. It is noteworthy that both junction footprints (Fig. 5, boxes 7 and 8) are approximately centered over homologies (a match of 7 of 8 bp) to the AP-1 consensus sequence, and AP-1 has been shown to bind to these regions in BK(Dunlop) (36). BKV mutants identical to pBK68 but containing two 68-bp repeats (pBK68-68) or a 68-bp and a 50-bp repeat (pBK68-50) in their control region both show higher transcriptional activity (157 and 233%, respectively) than pBK68 (data not shown). Both of these mutants contain one junction sequence, lacking in pBK68, and their increased CAT activity may be due to the presence of an extra enhancer element.

In conclusion, we have shown that transcription from the BKV early regulatory region requires a minimum of four distinct regions, at least three of which constitute protein-

binding domains. Our data implicate an NF1 family protein, an Sp1-related protein, and an AT-block recognizing factor in binding to these areas and stimulating transcription from the BKV early promoter. A fourth protein-protected domain may bind the transcription factor AP-1. It will be interesting to compare the type and arrangement of transcriptional factors used by BKV(wt) to that of BKV variants and other papovaviruses. Such studies will help elucidate the role of transcription factors in determining the cell-type specificities of these viruses.

ACKNOWLEDGMENTS

We thank Michael Karin for advice and the gift of HeLa cell extract and Aaron Cassill for assistance with footprinting experiments. We also thank Mark Hannink, Ingrid Akerblom, Pam Mellon, and other colleagues for useful comments and discussions.

This work was supported by a Biotechnology grant from the University of California and Public Health Service grant CA44997 from the National Institutes of Health to S.S. This work is a publication of the Center for Molecular Genetics.

LITERATURE CITED

1. Angel, P., M. Imagawa, R. Chiu, B. Stein, R. J. Imbra, H. J. Rahmsdorf, C. Jonat, P. Herrlich, and M. Karin. 1987. Phorbol ester-inducible genes contain a common *cis* element recognized by a TPA-modulated *trans*-acting factor. *Cell* 49:729-739.
2. Arthur, R. R., K. V. Shah, S. J. Baust, G. W. Santos, and R. Saral. 1986. Association of BK viruria with hemorrhagic cystitis in recipients of bone marrow transplants. *N. Engl. J. Med.* 315: 230-234.
3. Briggs, M. R., J. T. Kadonaga, S. P. Bell, and R. Tjian. 1986. Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. *Science* 234:47-52.
4. Caputo, A., A. Corallini, M. P. Grossi, L. Carra, P. G. Balboni, M. Negrini, G. Milanesi, G. Federspil, and G. Barbanti-Brodano. 1983. Episomal DNA of a BK virus variant in a human insulinoma. *J. Med. Virol.* 12:37-49.
5. Chen, E. Y., and P. H. Seeburg. 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. *DNA* 4: 165-170.
6. Chiu, R., M. Imagawa, R. J. Imbra, J. R. Bockoven, and M. Karin. 1987. Multiple *cis*- and *trans*-acting elements mediate the transcriptional response to phorbol esters. *Nature (London)* 329:648-651.
7. Davidson, I., C. Fromental, P. Augereau, A. Wildeman, M. Zenke, and P. Chambon. 1986. Cell-type specific protein binding to the enhancer of simian virus 40 in nuclear extracts. *Nature (London)* 323:544-548.
8. Davison, B. L., J. Egly, E. R. Mulvihill, and P. Chambon. 1983. Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences. *Nature (London)* 301:680-686.
9. de Vries, E., W. van Driel, M. Tromp, J. van Boom, and P. C. van der Vliet. 1985. Adenovirus DNA replication *in vitro*: site-directed mutagenesis of the nuclear factor I binding site of the Ad2 origin. *Nucleic Acids Res.* 13:4935-4952.
10. de Vries, E., W. van Driel, S. J. L. van den Heuvel, and P. C. van der Vliet. 1987. Contactpoint analysis of the HeLa nuclear factor I recognition site reveals symmetrical binding at one side of the DNA helix. *EMBO J.* 6:161-168.
11. deWet, J. R., K. V. Wood, M. DeLuca, D. R. Helinski, and S. Subramani. 1987. Firefly luciferase gene: structure and expression in mammalian cells. *Mol. Cell. Biol.* 7:725-737.
12. Deyerle, K. D., J. A. Cassill, and S. Subramani. 1987. Analysis of the early regulatory region of the human papovavirus BK. *Virology* 158:181-193.
13. Dorn, A., J. Bollekens, A. Staub, C. Benoist, and D. Mathis. 1987. A multiplicity of CCAAT box-binding proteins. *Cell* 50: 863-872.
14. Evans, T., T. DeChiara, and A. Efstratiadis. 1988. A promoter of the rat insulin-like growth factor II gene consists of minimal

- control elements. *J. Mol. Biol.* **199**:61–81.
15. Frisque, R. J., G. L. Bream, and M. T. Cannella. 1984. Human polyomavirus JC virus genome. *J. Virol.* **51**:458–469.
 16. Fromm, M., and P. Berg. 1982. Deletion mapping of DNA regions required for SV40 early region promoter function *in vivo*. *J. Mol. Appl. Genet.* **1**:457–481.
 17. Gardner, S. D., A. M. Field, D. V. Coleman, and B. Hulme. 1971. New human papovavirus (B.K.) isolated from urine after renal transplantation. *Lancet* **i**:1253–1257.
 18. Gidoni, D., W. S. Dynan, and R. Tjian. 1984. Multiple specific contacts between a mammalian transcription factor and its cognate promoters. *Nature (London)* **312**:409–413.
 19. Gronostajski, R. M., A. Adhya, K. Nagata, R. A. Guggenheimer, and J. Hurwitz. 1985. Site-specific DNA binding of nuclear factor 1: analyses of cellular binding sites. *Mol. Cell. Biol.* **5**:964–971.
 20. Hennighausen, L., and B. Fleckenstein. 1986. Nuclear factor 1 interacts with five DNA elements in the promoter region of the human cytomegalovirus major immediate early gene. *EMBO J.* **5**:1367–1371.
 21. Heritage, L., P. M. Chesters, and D. J. McCance. 1981. The persistence of papovavirus BK DNA sequences in normal human renal tissue. *J. Med. Virol.* **8**:143–150.
 22. Herr, W., and J. Clarke. 1986. The SV40 enhancer is composed of multiple functional elements that can compensate for one another. *Cell* **45**:461–470.
 23. Imagawa, M., R. Chiu, and M. Karin. 1987. Transcription factor AP2 mediates induction by two different signal-transduction pathways: protein kinase c and cAMP. *Cell* **51**:251–260.
 24. Jeang, K., D. R. Rawlins, P. J. Rosenfeld, J. H. Shero, T. J. Kelly, and G. S. Hayward. 1987. Multiple tandemly repeated binding sites for cellular nuclear factor I that surround the major immediate-early promoters of simian and human cytomegalovirus. *J. Virol.* **61**:1559–1570.
 25. Johnson, P. F., W. H. Landschulz, B. J. Graves, and S. McKnight. 1987. Identification of a rat liver nuclear protein that binds to the enhancer core element of three animal viruses. *Genes Dev.* **1**:133–146.
 26. Jones, K. A., J. T. Kadonaga, P. J. Rosenfeld, T. J. Kelly, and R. Tjian. 1987. A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. *Cell* **48**:79–89.
 27. Jones, K. A., and R. Tjian. 1985. Sp1 binds to promoter sequences and activates herpes simplex virus 'immediate early' gene transcription *in vitro*. *Nature (London)* **317**:179–182.
 28. Jones, K. A., K. R. Yamamoto, and R. Tjian. 1985. Two distinct transcription factors bind to the HSV thymidine kinase promoter *in vitro*. *Cell* **42**:559–572.
 29. Kadonaga, J. T., K. A. Jones, and R. Tjian. 1986. Promoter-specific activation of RNA polymerase II transcription by Sp1. *Trends Biochem. Sci.* **11**:20–23.
 30. Khalili, K., G. Khoury, and J. Brady. 1986. Spacing between simian virus 40 early transcriptional control sequences is important for regulation of early RNA synthesis and gene expression. *J. Virol.* **60**:935–942.
 31. Kunkel, T. A. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* **82**:488–492.
 32. Lee, W., A. Haslinger, M. Karin, and R. Tjian. 1987. Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40. *Nature (London)* **325**:368–372.
 33. Lee, W., P. Mitchell, and R. Tjian. 1987. Purified transcription factor Ap-1 interacts with TPA-inducible enhancer elements. *Cell* **49**:741–752.
 34. Leegwater, P. A. J., P. C. van der Vliet, R. A. W. Rupp, J. Nowock, and A. E. Sippel. 1986. Functional homology between the sequence-specific DNA-binding proteins nuclear factor I from HeLa cells and the TGGCA protein from chicken liver. *EMBO J.* **5**:381–386.
 35. Leegwater, P. A. J., W. van Driel, and P. C. van der Vliet. 1985. Recognition site of nuclear factor I, a sequence-specific DNA-binding protein from HeLa cells that stimulates adenovirus DNA replication. *EMBO J.* **4**:1515–1521.
 36. Markowitz, R.-B., and W. S. Dynan. 1988. Binding of cellular proteins to the regulatory region of BK virus DNA. *J. Virol.* **62**:3388–3398.
 37. Miksicek, R., U. Borgmeyer, and J. Nowock. 1987. Interaction of the TGGCA-binding protein with upstream sequences is required for efficient transcription of mouse mammary tumor virus. *EMBO J.* **6**:1355–1360.
 38. Nomiyama, H., C. Fromental, J. H. Xiao, and P. Chambon. 1987. Cell-specific activity of the constituent elements of the simian virus 40 enhancer. *Proc. Natl. Acad. Sci. USA* **84**:7881–7885.
 39. Nowock, J., U. Borgmeyer, A. W. Püschel, R. A. W. Rupp, and A. E. Sippel. 1985. The TGGCA protein binds to the MMTV-LTR, the adenovirus origin of replication, and the BK virus enhancer. *Nucleic Acids Res.* **13**:2045–2061.
 40. Ondek, B., A. Shepard, and W. Herr. 1987. Discrete elements within the SV40 enhancer region display different cell-specific enhancer activities. *EMBO J.* **6**:1017–1025.
 41. Raymondjean, M., S. Cereghini, and M. Yaniv. 1988. Several distinct "CCAAT" box binding proteins coexist in eukaryotic cells. *Proc. Natl. Acad. Sci. USA* **85**:757–761.
 42. Rosenthal, N., M. Kress, P. Gruss, and G. Khoury. 1983. BK viral enhancer element and a human cellular homolog. *Science* **222**:749–755.
 43. Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe. 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143**:161–178.
 44. Sawadogo, M., and R. G. Roeder. 1985. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Cell* **43**:165–175.
 45. Schirm, S., J. Jiricny, and W. Schaffner. 1987. The SV40 enhancer can be dissected into multiple segments, each with a different cell type specificity. *Genes Dev.* **1**:65–74.
 46. Schneider, R., I. Gander, U. Müller, R. Mertz, and E. L. Winnacker. 1986. A sensitive and rapid gel retention assay for nuclear factor I and other DNA-binding proteins in crude nuclear extracts. *Nucleic Acids Res.* **14**:1303–1317.
 47. Scholer, H. R., and P. Gruss. 1984. Specific interaction between enhancer-containing molecules and cellular components. *Cell* **36**:403–411.
 48. Schreier, A. A., and J. Gruber. 1986. Oncogenic human polyomaviruses. *JNCI* **6**:1255–1258.
 49. Seif, I., G. Khoury, and R. Dhar. 1979. The genome of human papovavirus BKV. *Cell* **18**:963–977.
 50. Small, J. A., G. Khoury, G. Jay, P. M. Howley, and G. A. Scangos. 1986. Early regions of JC virus and BK virus induce distinct and tissue-specific tumors in transgenic mice. *Proc. Natl. Acad. Sci. USA* **83**:8288–8292.
 51. Takahashi, K., M. Vigneron, H. Matthes, A. Wildeman, M. Zenke, and P. Chambon. 1986. Requirement of stereospecific alignments for initiation from the simian virus 40 early promoter. *Nature (London)* **319**:121–126.
 52. Watanabe, S., and K. Yoshiike. 1982. Change of DNA near the origin of replication enhances the transforming capacity of human papovavirus BK. *J. Virol.* **42**:978–985.
 53. Watanabe, S., and K. Yoshiike. 1985. Decreasing the number of 68-base-pair tandem repeats in the BK virus transcriptional control region reduces plaque size and enhances transforming capacity. *J. Virol.* **55**:823–825.
 54. Wu, C. 1985. An exonuclease protection assay reveals heat-shock element and TATA box DNA-binding proteins in crude nuclear extracts. *Nature (London)* **317**:84–87.
 55. Wu, L., D. S. E. Rosser, M. C. Schmidt, and A. Berk. 1987. A TATA box implicated in E1a transcriptional activation of a simple adenovirus 2 promoter. *Nature (London)* **326**:512–515.
 56. Zenke, M., T. Grundstöm, H. Matthes, M. Wintzerith, C. Schatz, A. Wildeman, and P. Chambon. 1986. Multiple sequence motifs are involved in SV40 enhancer function. *EMBO J.* **5**:387–397.