Proteins of the nuclear factor-1 family act as an activator of the late promoter in human polyomavirus BK in vitro

Tushar Chakraborty† and Gokul C. Das*

Department of Molecular Biology, The University of Texas Health Science Center at Tyler, P.O. Box 2003, Tyler, Texas 75710, U.S.A.

The cis-acting elements for the early and late promoters, as well as the enhancer in the prototype strains of human polyomavirus BK (BKV) are located within a 500 bp intergenic region. We previously studied the specificity of protein binding in this region in vitro and showed that the interaction of proteins of the nuclear factor-1 (NF-1) family is crucial for early promoter activity. We have now extended our study to the BKV late promoter. We show that the late promoter activity in HeLa cell extracts is poor compared to the activity of the early promoter. Using a high template to protein ratio, multiple start sites were detected by primer extension analysis. DNase I protection experiments revealed the presence of three NF-1 binding sites in the late side, in addition to those identified previously in the 68 bp repeats and C element. Competition transcription assays using binding sites for NF-1, AP-1, Sp-1 and a complete 68 bp repeat indicated that only the 68 bp repeat and the NF-1 binding site competed significantly with the late promoter activity. A point mutation in the NF-1 binding site, which destroys the ability of the oligonucleotide to bind NF-1, also impaired its capacity to compete with the late promoter. The ability of NF-1 to activate both the early and late promoters suggests that the proteins of this family act as a bidirectional transcriptional activator in this virus.

Introduction

Activation of eukaryotic gene expression is mediated by the binding of multiple sequence-specific DNA-binding proteins to the promoter and enhancer regions (Dynan & Tjian, 1985; Johnson & McKnight, 1989; Mitchell & Tjian, 1989). Cells can modulate precisely the expression of a given gene, as well as maximize its use of a given activator, through the economical use of different combinations of regulatory elements and proteins (Dynan, 1989; Mitchell & Tjian, 1989). The intergenic regulatory region of papovaviruses provides an excellent model for the study of such complex genetic control mechanisms. In human polyomavirus BK (BKV), a region of approximately 500 bp between the early and late genes contains overlapping cis-acting elements for viral replication, early and late promoters, and for the viral enhancer (Fig. 1; Deyerle et al., 1987; Cassil & Subramani, 1988; Chakraborty & Das, 1989; Seif et al., 1979). We have previously identified the control elements of BKV promoters by studying DNA–protein interactions, as well as the transcriptional property of the viral template in a HeLa cell extract. It has been shown that proteins of the nuclear factor-1 (NF-1) family are major activators of early transcription in vitro (Chakraborty & Das, 1989). In the present study, we have addressed the role of these proteins on late transcription in vitro.

The sequence elements that form the BKV late promoter have not yet been fully defined. The late promoters of papovaviruses appear, in general, to be far more complex than the early promoters (Ayer & Dynan, 1988; Brady & Khoury, 1985; Brady et al., 1982, 1984; Cahill & Carmichael, 1989; Hartzell et al., 1984; Omilli et al., 1986; Kim et al., 1987). Even in the well studied simian virus 40 (SV40) and polyoma virus, the regulatory elements that specifically modulate the efficiency of late transcription units and regulate temporal expression remain undefined (Alwine & Picardi, 1986; Alwine et al., 1975; Ayer & Dynan, 1988; Brady & Khoury, 1985; Brady et al., 1982, 1984; Contreras et al., 1982; Hertz & Mertz, 1988; Huang et al., 1990; Mitchell et al., 1987; Somasekhar & Mertz, 1985; Gong & Subramanian, 1988). Both the 72 bp enhancer elements (Hartzell et al., 1984; Hertz & Mertz, 1988; Rio & Tjian, 1984) and the GC-rich sequences (Ernout-Lange et al., 1987; Gidoni et al., 1985; Hertz & Mertz, 1988) have been shown to act as

† Present address: Department of Biochemistry and Molecular Biology, The University of Texas, M. D. Anderson Cancer Research Center, Houston, Texas, U.S.A.
HaelIl adenovirus NF-I binding sites and mutant NF-I binding sites were purified from plasmids pKB 67/88 and pM 26/57, respectively. These pLTG and CAS-9 have been described previously (Chakraborty & Piatigorsky, 1985). For run-off transcription assays, the supercoiled template (10 to 100 µg/ml) and 30 µl of whole HeLa cell extract. After the incubation period, the templates in the reaction mixture were digested with 2 units of RNase-free DNase I (Promega) for 5 min at 30 °C. The RNA was then purified, glyoxylated and analysed on a 1.4% agarose gel as described earlier (Das & Salzman, 1985).

Transcription assays using vector pLTG were carried out in the presence of [α-32P]UTP and in the absence of GTP (Sawadogo & Roeder, 1985; Schorpp et al., 1988). The products were treated with RNase T1 for 10 min at 30 °C after transcription. Transcripts were analysed on a 4% acrylamide–urea gel and subjected to autoradiography. Competition experiments were performed by preincubating the competitor DNAs with the template for 15 min at 0 °C, followed by 30 min incorporation at 30 °C.

The 5' ends of the RNAs were mapped by primer extension analysis as described earlier (Chakraborty & Das, 1989; Jones et al., 1985). A 20-mer decyoxynucleotide complementary to the late RNA sequence corresponding to the viral genome (Gardner strain) between nucleotide positions 554 and 573 was synthesized chemically and used as a primer after labelling with [α-32P]ATP and T4 polynucleotide kinase. The RNAs synthesized in vitro were hybridized with the primer for 60 min at 50 °C, elongated with avian myeloblastosis virus (AMV) reverse transcriptase at 42 °C for 30 min and analysed on 8% denaturing acrylamide gels.

DNase I footprinting. Nuclear extracts from HeLa cells were prepared as described (Dignam et al., 1983; Gorski et al., 1986). DNase I footprinting was performed as described before (Chakraborty & Das, 1989; Alwine & Picardi, 1986). The HordII fragment containing the regulatory region was purified from pBK504 which contains a single 68 bp repeat. After radiolabelling with [α-32P]ATP and T4 polynucleotide kinase, the labelled fragment was further digested with Stul to prepare a probe of 280 bp for NF binding domains in the late region. The cDNAs were analysed on a 7% polyacrylamide–urea gel.

**Methods**

*Plasmids and DNA.* The plasmids pBK501, pBK504, pBKDUN, pLTG and CAS-9 have been described previously (Chakraborty & Das, 1989). The plasmid pLTG-BKL was constructed by cloning the HaeIII fragment of BKV (Gardner strain) from nucleotide position 142 to 400 with a PstI linker at the PstI site of the G-less cassette. The adenovirus NF-I binding sites and mutant NF-I binding sites were purified from plasmids pKB 67/88 and pM 26/57, respectively. These plasmids were a gift from T. Kelly, John Hopkins School of Medicine, Baltimore, Md., U.S.A. (Rosenfeld & Kelly, 1986; Rosenfeld et al., 1987). The AP-1 binding sites were purified from plasmid pBL 5 X AP-1, obtained from M. Karin, University of California at San Diego, Ca., U.S.A. All plasmid DNAs were purified before use in transcription assays by centrifugation twice through CsCl/ethidium bromide gradients.

*In vitro transcription assays and primer extension analysis.* In vitro transcription reactions were performed in a HeLa whole cell extract (Manley et al., 1980). For run-off transcription assays, the supercoiled plasmid DNA was restricted with appropriate enzymes to generate termination sites for transcription, and a fixed quantity of this linear DNA was used (Das & Piatigorsky, 1986). Standard transcription reactions (60 µl) were carried out at 30 °C for 60 min in 10 mM-HEPES pH 7.9, 8 mM-Tris·HCl pH 8.6, 50 mM-KCl, 16 mM-NaCl, 6.5 mM-MgCl₂, 0.05 mM-EDTA, 1.6 mM-DTT, 0.16 mM-PMSF, 0.006% Triton X-100, 8.5% glycerol, 0.4 mM each of ATP, GTP and CTP, 0.04 mM-UTP, 20 µCi [α-32P]UTP (500 Ci/mmol) containing DNA template.

Bidirectional elements capable of stimulating SV40 late promoter activity. Some recent studies, however, suggest the SV40 late promoter has a complex tripartite nature with a combination of initiation sites with both upstream and downstream elements (Ayer & Dynan, 1988). The BKV late promoter, which lacks a TATA-like sequence, has not been subjected to such detailed analysis. However, it has been shown that late RNA initiation sites in transfected CV-1 or COS cells are heterogeneous, and at least some elements of the late promoter are contained within the 68 bp triplicate in the early promoter enhancer region (Cassil & Subramani, 1988). We have mapped the sites of RNA initiation and the sites of protein binding in the late region in vitro and show that NF-I also acts as an activator of the late genes.

**Results**

In vitro transcription of BKV late promoter

To define elements of the late promoter, we first developed an in vitro transcription system capable of accurate late RNA initiation. The recombinant plasmid, pBK501, containing the entire viral genome (Gardner strain) cloned in pBR322 (Gardner et al., 1971; Watanabe & Yoshike, 1982) was digested with BamHI to provide a termination site for run-off transcription assays. This template was expected to generate early...
NF-1 proteins activate BKV late promoter

Fig. 2. Comparison of early and late promoter activity of BKV (a) and SV40 (b) in vitro. Supercoiled plasmid DNAs, pBK501 for (a) and pSVK104 for (b) were digested with BamHI and a fixed quantity of the linearized template was transcribed in a HeLa cell extract as described in Methods. Lane M, DNA size markers (bp); lanes 1 and 2, template concentrations of 20 and 30 µg/ml per whole cell extract, respectively. E and L indicate the expected positions of early and late transcripts of the size indicated within parentheses.

Fig. 3. Primer extension analysis of BKV (Gardner strain) late RNA. A 20-mer late primer was end-labelled with [γ-32P]ATP and T4 polynucleotide kinase, hybridized with the RNA synthesized in vitro from supercoiled templates and extended by AMV reverse transcriptase. The RNA initiation sites are indicated by arrows. The initiation sites (*) mark the ones found in vitro, but not yet reported in vivo (Cassil & Subramani, 1988). Bands indicated by a bracket were obtained with pBR322 DNA alone and are thought to be non-specific. (a) The titration of late promoter activity at different template concentrations. Lane 2, 50; lane 3, 100; lane 4, 125; lane 5, 150 and lane 6, 200 µg template/ml. Lane 1 contains 200 µg template/ml in the presence of α-amanitin. The primer sequence is 5' GGTGAAGACAGTGTAGACGG 3'. (b) Competition transcription assays. The reaction mixtures were preincubated with a 50-fold molar excess of wild-type (lane 1) and mutant NF-1 (lane 2) binding sites and transcribed at a 150 µg/ml DNA to extract ratio. The transcripts were analysed by primer extension under identical conditions to those in (a). Lane M, DNA size markers (bp). The sequence of the NF-1 binding site is 5' TTTTGCTGAAGCCATATGAG 3'. The mutant NF-1 binding site contains a G→A transition at the underlined G residue.

RNAs of about 3560 nucleotides (nt) and late RNAs of approximately 1400 nt in length. However, when it was titrated with a range from 10 to 100 µg of template DNA per ml of whole cell extract under standard assay conditions, we could not detect late RNAs in the expected size class. The early RNAs, however, were synthesized optimally in assays containing 25 to 30 µg of template/ml of whole cell extract (Fig. 2a). Another band of about 1850 nt was present in the autoradiogram and was produced by 18S RNA present in the HeLa cell extract. Under similar assay conditions one plasmid template containing SV40 promoters, pSVK104, produced both early RNA species (2150 nt) and a late RNA species (3130 nt) which initiates at nucleotide 185 (Fig. 2b). In addition, two other transcripts of about 2400 and 950 nt were observed which originated from pBR322 DNA, the former from a TATA box-like sequence at nt position 2612 (Das & Salzman, 1985; Brady et al., 1984). No RNA from the major late initiation site at map position 325 was observed as the recombinant plasmid did not contain this site. When early RNA transcription for this template attained its optimum level, late RNA was synthesized to approximately 25 to 30% the level of early RNA.

We investigated the initiation of late transcription by primer extension analysis using relatively high DNA/protein ratios. A 20-mer synthetic oligonucleotide from nucleotide positions 554 to 573 in the BKV genome was hybridized to in vitro synthesized RNAs and extended with reverse transcriptase (Fig. 3a). The results showed that late RNAs were synthesized from multiple start sites within the range of 100 to 200 µg of template per ml of whole cell extract. These transcripts were not detected at
a low concentration of α-amanitin or when a supercoiled pBR322 plasmid DNA was used as a template (except the cluster of bands shown within the bracket), indicating that they are RNA polymerase II transcripts. Some of the start sites were located within the 68 bp repeat region and the adjacent C element (Fig. 3a). These sites matched the in vivo late RNA start sites in COS cells reported earlier (Cassil & Subramani, 1988). Other start sites were mapped further downstream from the C element. The major late RNA synthesis in vitro started at about nt 473 in the late region, which is downstream from both the 68 bp repeat and the C element (Fig. 1, Fig. 4b). Such a difference in the selection of start sites in vitro has been observed with other papovaviruses (Jat et al., 1982; Brady et al., 1982). The possibility that this major in vitro product results from a premature termination generated by the secondary structure of the primer-extended RNA is less likely since the level of transcription is modulated differentially by wild-type and mutant NF-1 binding sites when used as competitors in the transcription assay (Fig. 3b).

It is worth noting that the concentration of template required for BKV late transcription in a HeLa cell extract was unusually high compared to other papovaviruses. This might reflect titration of a repressor for the late promoter in the HeLa cell extract, or might indicate that a special conformation of the nucleoprotein complex, which supports late transcription, is generated at this DNA/protein ratio. At this concentration (100 μg/ml), run-off assays became unsuitable for detection of transcripts because of high background.

Multiple NF-1 binding domains are present in the late region of BK virus

The distinguishing feature of the intergenic regulatory region of BKV, when compared to other members of papovaviruses, is the presence of multiple copies of NF-1 binding sites in all strains of BKV (Markowitz & Dynan, 1988). We previously mapped three NF-1 binding sites in the BKV (Gardner strain) regulatory region (Chakraborty & Das, 1989) (Fig. 4b). Careful examination of the late region downstream from the C element suggested that there were other potential NF-1 binding sites. Hence, the DNA–protein interaction in this region was critically investigated by a DNase I footprinting assay using HeLa cell nuclear extracts. The DNase I cleavage pattern with a HindIII–StuI fragment (Fig. 4b) labelled at the HindIII site identified five footprints over TGG(N)6→GCCA-like sequences (Fig. 4a). Two of the
upper region footprints corresponded to the already
detected. The one adjacent to the NF-1 binding site in
the C element indicated high affinity binding to NF-1.
The two lower footprints were comparatively weak, and
these sites required 10-fold more protein for a similar
degree of protection than other NF-1 sites in BKV. All
these protective effects were abolished when an
increased amount of NF-1-binding oligonucleotide was
added (Fig. 4a, lanes 5 and 6). The concentration of
competitor NF-1 oligonucleotide required to eliminate
DNase I footprints over the two terminal sites was also
about 10 times less than required for all other NF-1
binding sites. The two weak binding sites in the late
region were tandemly located without any intervening
spacer sequence. Simultaneous occupancy of both sites
by NF-1 may be difficult because of steric hindrance.
The three new NF-1 binding sites identified occur in
both the Dunlop and Gardner strains of BKV.

Role of NF-1 as an activator for the late promoter

We previously showed that binding of NF-1 is required
for efficient early promoter activity in vitro (Chakraborty
& Das, 1989). The role of NF-1 binding on late RNA
synthesis was explored using the G-less cassette vector
originally developed by Sawadogo & Roeder (1985). A
258 bp regulatory region was cloned in the late
orientation into the vector pLTG upstream of the TATA
box of the adenovirus major late promoter, which was
linked to a downstream 400 nt G-less cassette (Schorpp
et al., 1988). Transcription of this plasmid (pLTG-BKL)
gave rise to a transcript of 400 nt that was resistant to
RNase T1. In the absence of any upstream element
(vector alone) or in the presence of the BKV promoter
without the adenovirus TATA box (pLTG-BK3'), no transcripts were detected (data not shown). Another
plasmid, CAS-9, with a shorter G-less cassette (190 nt)
was included in the assay as an internal control. This
promoter did not compete for limiting factors with the
BKV late promoter under our assay conditions (results
not shown). Densitometric scanning determined the
ratio of the transcripts originating from the BKV late
promoter to the 190 nt internal standard.

Both the 68 bp repeat and the adenovirus NF-1
binding site in 100-fold molar excess reduced the late
transcription by about 75\% (Fig. 5). In contrast, a
mutant NF-1 binding site which is unable to bind NF-1
(Rosenfeld et al., 1987), did not compete for late
transcription. Similarly, when the binding sites for AP-1
or Sp-1 (GC box) were used as competitors, there were
no detectable effects on transcription. The results of this
experiment suggest that NF-1 activates RNA synthesis
from the late promoter. This result, together with our
previous studies on the early promoter (Chakraborty
& Das, 1989), clearly establishes the role of the proteins of
the NF-1 family from HeLa cells in bidirectional
activation of BKV promoters.

To extend this study in the natural context, pBK501
template was transcribed using excess wild-type or the
mutant NF-1 binding site from adenovirus as competi-
tors. The RNAs were then analysed by primer extension
methods as described previously. This experiment
showed that late RNA is synthesized at a reduced level in
the presence of the wild-type NF-1 binding site in
comparison with the mutant site (Fig. 3b). Although we
were unable to quantify the levels of synthesis because of
multiple initiation sites, these results support the notion
that efficient late transcription requires NF-1.

Discussion

Differential and temporal control of late gene expression
in papovaviruses is important, but control mechanisms
have not yet been adequately defined. Studies on SV40 indicate the participation of multiple cis-acting elements, which include the T antigen binding sites (Alwine & Picardi, 1986; Alwine et al., 1975; Brady & Khoury, 1985; Keller & Alwine, 1984, 1985; Mitchell et al., 1989), the SP-1 binding sites (Ernoult-Lange et al., 1987; Gidoni et al., 1985; Hertz & Mertz, 1988), the enhancer elements (May et al., 1987; Mermod et al., 1988; Mitchell et al., 1987), as well as the sequence to the late side of the enhancer (Kim et al., 1987). Late genes of the human papovaviruses, BKV and JCV, have not yet been subjected to such critical analysis. In this study, we used a HeLa cell extract to map the late RNA start sites and detected three NF-1 binding sites in the late region which were previously unknown. Competition transcription assays indicated that NF-1 binding to the regulatory region was important for efficient late transcription. The ability of NF-1 to activate both early (Chakraborty & Das, 1989) and late promoter identifies this protein as a bidirectional activator of transcription.

Late promoters of papovaviruses arc, in general, weakly transcribed in HeLa cell extracts when compared to their early promoters (Nandi et al., 1985; Brady et al., 1982). The early and late promoters of SV40 are activated simultaneously in HeLa cell extracts. When the BKV early promoter had optimal activity, late RNAs were not detected, indicating that the late promoter is weaker than its counterpart in SV40. This may be due to the presence of a repressor protein(s) in human cell lines which has to be titrated out to initiate late transcription (Grinnel et al., 1988). The requirement for a high concentration of template for late transcription is consistent with the presence of a repressor. It is noteworthy that late transcription in vivo predominates after the onset of DNA replication when the template copy number is very high (Padgett, 1980). Alternatively, efficient late promoter activity might require the involvement of a cellular positive activator(s) which is induced or modified by the viral early gene products (Gallo et al., 1990; Saff er et al., 1990). This phenomenon is absent in the cell-free system. Our results are compatible with those reporting that mouse polyomavirus late promoter also requires a high template concentration for a detectable level of transcription in vitro (Jat et al., 1982).

As with other papovaviruses, the 5' ends of RNA synthesized in vitro were heterogeneous; two of these sites are used in vivo in CV1 and COS cells (Cassil & Subramani, 1998). The major late RNA initiation site around nt 473 is located downstream from six NF-1 binding sites. One of these sites is located about 20 bp downstream from a TTAAAn sequence which resembles the TATA box (Fig. 2). A similar TATA-like sequence is also found 30 nt upstream of the in vitro start site of the major late RNA of the mouse polyomavirus (Jat et al., 1982). The significance of this TATA sequence remains to be determined.

Among the three additional NF-1 binding sites, the one proximal to the C element had a 10-fold higher affinity than the other two, which were tandemly arranged as TGG(N),GCCATGG(N),GCCA without any intervening spacer. Although the minimal consensus recognition sequence for NF-1 binding in BKV is TGG(N),CCA, the footprints protected approximately 24 nt on both strands, including the flanking sequences (Chakraborty & Das, 1989). We suggest that steric hindrance prevents simultaneous NF-1 binding at both of these sites. A cluster of NF-1 binding sites has also been observed in the regulatory region of simian and human cytomegaloviruses (Jeang et al., 1987).

Since both BKV and JCV, but not the other papovaviruses, contain NF-1 binding sites, it is tempting to speculate that one or more species of NF-1 are predominantly used by human cells as an activator for both DNA replication and transcription (Nagata et al., 1983; Santoro et al., 1988). The predominant role of NF-1 in BKV is apparently to modulate transcription, since these sites are not included in the minimum replication origin (Vecchio et al., 1989). Although a number of protein families can bind to CCAAT-like sequences (Chodosh et al., 1988; Dorn et al., 1987; Johnson & Mc Knight, 1989; Jones et al., 1987; Knight et al., 1987; Maity et al., 1988; Nagata et al., 1983; Santoro et al., 1988), we tentatively identified the HeLa nuclear protein that interacts with the BKV regulatory region as CTF/NF-1 by three criteria: (i) competitive binding with the NF-1 site from the adenovirus origin (Chakraborty & Das, 1989), (ii) the requirement for TGG(N),CCA dual symmetry for efficient binding and (iii) the footprint pattern and transcriptional activity (Chakraborty & Das, 1989; De Vries et al., 1987).

The binding of AP-1 and Sp-1 does not seem to play an important role in vitro for late transcription although these factors might play a more important role in vivo. The early promoter is inducible in vivo by the phorbol ester 12-O-tetradecanoyl phorbol-13-acetate which is mediated by AP-1 binding (Moens et al., 1990; our unpublished results) and mutating the Sp-1 binding site also causes a significant decrease in its activity (Dey erle & Subramani, 1988). The activation by a single copy of the Sp-1 binding site might require the binding of other proteins to the TATA box, which is absent in the late promoter, or a specific nucleoprotein structure which is not generated in vitro. This site is not conserved among all strains of BKV, indicating that it might not be crucial for promoter function.

Based on our earlier work and the present study, it appears that NF-1 can act as a bidirectional activator for
both the early and late promoters in BKV, with binding sites similar to six copies of GC boxes found in the SV40 promoter. This finding is in agreement with a report of bidirectional stimulation of RNA synthesis in vitro from the adenovirus major late promoter by cloned synthetic and natural NF-1 binding sites (Gronostajski et al., 1988). It is not yet known how different NF-1 binding sites contribute to early and late transcription. All the NF-1 binding sites in BKV have a twofold symmetry and are arranged in the same polarity in both the Gardner and Dunlop strains. This arrangement implies a cooperative interaction among them and bidirectionality in the BKV regulatory region.

The authors are grateful to Dr Sujay DasGupta for synthesizing the oligonucleotides, Drs Michael Karin and Thomas Kelly for the plasmids and Drs C. Zweib and A. Johnson for critical reading of this manuscript. The excellent secretarial assistance of Miss Darlene Bean is also gratefully acknowledged.

References


(Received 3 December 1990; Accepted 28 March 1991)