Identification of proteins binding to the F441 locus of polyomavirus B enhancer that are required for its activity in embryonic carcinoma cells

Sujoy DasGupta, Chittari V. Shivakumar and Gokul C. Das*

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

A point mutation at nucleotide 5233 of the polyomavirus (A2 strain) enables it to overcome growth restriction in undifferentiated embryonal carcinoma cells. We analysed the binding of nuclear proteins from F9 cells to a 38 bp region that spans this site of mutation and encompasses two copies of the bovine papillomavirus core sequence, CCACCC, and characterized this domain by mutational analysis. Our results showed that the F441 mutation creates a sequence motif which binds TEF-1 or a TEF-1-like protein from F9 cells more strongly than its wild-type counterpart and increases its activity by about 10-fold. Another protein identified as CP1 binds with increased affinity in the presence of the F441 mutation to the CAT box-like sequences which is contiguous with the downstream CCACCC box. Point mutations within these two motifs that abolished binding in vitro also impaired the activity of the F441 locus in vivo. As neither the wild-type sequence without the F441 mutation, nor the F441 template without the CAT box has appreciable activity in vivo, interaction between these two elements is required for function. At a higher level of organization, this interaction is probably extended to factors bound to other domains in the A and B enhancer.

Introduction

Polyomavirus (Py) cannot grow in undifferentiated (UD) cell lines, such as the mouse embryonal carcinoma cell lines F9 and PCC4, the stem cells for teratocarcinoma (Swartzendruber et al., 1975). Mutants of this virus with alterations in their enhancer region can overcome this restriction (Amati, 1985). The simplest mutant has an A to G transition at nucleotide (nt) 5233 (Fujimura et al., 1981). This mutation might either destroy a repressor binding site (Hen et al., 1986) or create a binding site for a positive-acting factor (Kovesdi et al., 1987; Kryszke et al., 1987). The latter might enable the virus to overcome the restriction imposed by the lack of certain eukaryotic trans-activators, such as PEA1 in F9 cells.

The F441 mutation in Py creates a sequence motif that has considerable homology with several eukaryotic positive activators, including the GT-IIC motifs of simian virus 40 (SV40) and the binding site of nuclear factor 1 (NF-1) from adenovirus (Davidson et al., 1988; Nagata et al., 1983). In vitro binding experiments have demonstrated that two purified HeLa cell transcription factors, TEF-1 and NF-1, can bind to this sequence with higher affinity than to the wild-type (wt) one (Davidson et al., 1988; Tseng & Fujimura, 1988). It is not known whether these trans-activators are functional in F9 cells.

Two copies of the sequence, CCCACC, that is implicated as a regulatory element in bovine papilloma-virus (BPV) flank the site of mutation (Lusky et al., 1983). The sequence CCACCC also acts as a functional promoter element in the human embryonic zetaglobin gene and binds an Sp1-like factor (Yu et al., 1991). The role of this element in the context of a Py enhancer is not yet known (Bohnlein et al., 1985; Lusky et al., 1983). To understand the function of this locus more fully, we analysed the binding of F9 cell nuclear proteins to a 38 bp region that contains both copies of the sequence CCACCC flanking the site of F441 mutation. We find that the mutation creates an SV40 GT-IIC-like motif that binds TEF-1 or a TEF-1-like factor more strongly than the wt counterpart and increases the binding of another CP1-like protein to a CAT box sequence downstream. These proteins interact to generate enhancer activity in this locus.

Methods

Plasmids and DNA. Oligonucleotides corresponding to sequences of the Py enhancer were synthesized with BamHI cohesive ends and cloned at the BamHI site of the vector pBLCAT2 for an in vivo assay (Luckow & Schutz, 1987). To measure enhancer activity in the normal promoter context, these oligonucleotides were cloned in the vector pAOPCATT by blunt-end ligation at the Smal site (Tseng & Fujimura, 1988). The orientation of the insert was determined by direct nucleotide sequencing. DNAs for transfection experiments were prepared by using an ion-exchange column (Qiagen) resulting in a preparation with 70% or more of supercoiled DNA.
**DNA binding gel electrophoresis assay.** The F441 oligonucleotide and its mutants were cleaved by BamHI from their corresponding clones in pBlCAT2. The F441 insert was dephosphorylated and labelled at the 5' ends with T4 polynucleotide kinase and [γ-32P]ATP. The probe (0.5 to 1.0 ng, 20000 c.p.m.) was incubated with 5 μg of nuclear extract from F9 cells after a 10 min preincubation with 1 μg of salmon sperm DNA as a non-specific competitor (Dignam et al., 1983). The binding reaction was done in 25 mM-HEPES pH 7.6, 60 mM-KCl, 1 mM-DTT, 0.1 mM-EDTA and 10% glycerol at room temperature for 30 min. DNA-protein complexes were separated from protein-free DNA by electrophoresis in a 6% polyacrylamide gel at 250 V for 2 h using low ionic strength buffer containing 22 mM-Tris, 22 mM-boric acid and 0.5 mM-disodium EDTA, pH 8.3 (Garner & Revzin, 1981; Chakraborty & Das, 1989).

**Methylation interference assay.** The end-labelled F441 insert was digested with BstNI, and the larger fragment was purified and used as a probe. The probe was methylated by dimethylsulphate, complexed with nuclear extract and run on a preparative 6% polyacrylamide gel containing 22 mM-Tris, 22 mM-boric acid and 0.5 mM-disodium EDTA, pH 8.3 (Garner & Revzin, 1981; Chakraborty & Das, 1989).

**Chloramphenicol acetyltransferase (CAT) assay.** F9 (UD) cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum. Subconfluent cultures were grown in culture dishes coated with 0.1% gelatin. The cells were transfected with a fixed quantity of the plasmid DNA containing the bacterial CAT gene using the calcium phosphate precipitation procedure (Gorman et al., 1982). After 30 h, the cells were harvested and washed with vigorous vortexing in 40 mM-Tris-HCl pH 7.9, containing 1 mM-EDTA. The cells were disrupted by sonication and the extracts were heated for 10 min at 65 °C and cleared by centrifugation in a microfuge. The protein concentration of each extract was determined in triplicate as described (Gorman et al., 1982). In earlier experiments, we used an internal control, bacterial β-galactosidase, under the control of the Rous sarcoma virus promoter.

The separated forms of acetylated chloramphenicol on silica gel thin-layer plates were detected by autoradiography, and the spots were cut out and counted. As is frequently observed in such transient assays, absolute values varied from experiment to experiment. To establish a pattern of variation, plates in duplicate or in triplicate were transfected with the same quantity of plasmid DNA and the CAT activity in each extract was measured independently. To ensure reproducibility, the transfection experiment was repeated at least twice with independently prepared DNA samples. CAT activity is reported as the mean of two or more observations.

**Results**

**Binding of nuclear factors from embryonal carcinoma cells to the F441 oligonucleotide**

To investigate binding of nuclear proteins to the F441 locus, we synthesized chemically double-stranded oligo-
Fig. 2. (a) Gel electrophoresis DNA binding assay. Nuclear extracts were prepared from F9 (UD) cells and complexed with end-labelled probes (about 0·2 ng; 30000 c.p.m.), as indicated at the top of each lane. The complexes were analysed in 6% polyacrylamide gels at low ionic strength. Five complexes (C1 to C5) identified in the autoradiogram are marked. (b) Competition DNA binding assay using labelled F441 oligonucleotide as the probe and F9 (UD) extract. The unlabelled competitors used are indicated at the top of the lanes together with the number of molar excesses of each competitor. SS stands for non-specific competitor salmon sperm DNA. (c) Competition DNA binding assay using F441 oligonucleotide as the probe. Mutants with base substitutions in the F441 template (F441-1, -2 and -3) and core -1 and -2 sequences were used as competitors. Experiments were carried out under the same conditions as in (a). (d) Gel electrophoresis DNA binding assay. Probes used are indicated at the top of each lane. The probe F441-dll is of smaller size than the others and was run in a separate gel. Other experimental conditions are the same as in (a).
Fig. 3. Methylation interference assay. End-labelled F441 oligonucleotide was digested with BstNI and the larger fragment, which was capable of forming the complex C1 and C2, was used as the probe. The methylated probe was complexed with nuclear extract from F9 (UD) cells and the complexes were separated from the free probe by gel electrophoresis. Complex C1 was purified by electroelution, treated with piperidine and run on a polyacrylamide-urea gel. The free probe treated similarly was run in parallel as a control. A + G, sequencing ladder. The nucleotide sequence of the probe is shown with points of contact with nuclear factors marked by open and filled circles.

nucleotides corresponding to the sequences from nt 5216 to 5253, without (wt oligo) or with (F441 oligo) an A to G transition at nt 5233 (Fig. 1). The cloned inserts were purified and used as probes in the DNA binding assay.

Results presented in Fig. 2(a) show that the wt template formed three major complexes, C1, C2 and C5. The F441 template formed in addition to these two other complexes, a major one (C4) and a minor one (C3). The wt template formed these complexes only at low levels. The affinity of complexes C1 and C2 was also increased in the presence of the F441 mutation but less than that of C3 and C4. The intensity of the band corresponding to the C5 complex was the same for both templates, and as shown below C5 appeared to be a non-specific complex or a degradation product. These results show that the F441 mutation increases the affinity of protein binding to this sequence over that of the wt template. This difference could have important implications for the biological activity of this locus.

To determine the specificity of complex formation, the complexes formed by the F441 template were subjected to competition by F441, wt and non-specific salmon sperm DNA competitors (Fig. 2b). The results show that complexes C3 and C4 were highly specific for the F441 template as their formation was completely abolished even at a 50-fold molar excess of the F441 competitor, whereas the wt template did not compete even at a 150-fold molar excess. Complexes C1 and C2 were specific for both templates because they disappeared gradually with increasing concentrations of both wt and F441 competitors. Once again, the F441 template was a better competitor. Salmon sperm DNA did not prevent the formation of any of these complexes, providing additional support for the view that these were specific complexes. Complex C5 remained virtually unaffected with increasing concentrations of both wt and F441 competitors. Thus, these in vitro binding experiments reveal the formation of multiple protein–DNA complexes, of which C3 and C4 are highly specific for only the F441 template, whereas C1 and C2 complexes are specific for both templates.

Sequence requirement for the formation of an F441-specific complex

Our next objective was to identify sequence requirements for the formation of F441-specific complexes. It was shown earlier that the A to G mutation at nt 5233 is crucial for activity of the F441 mutant, and substitution of other nucleotides fails to produce the same effect (Tseng et al., 1988). This strict requirement for G might be due to its location in the middle of the motif TGG which is highly conserved among several positive-acting DNA sequences. Not all of the elements sharing TGG are, however, similar in activity, cell specificity or capacity to bind nuclear factors. Information for this differential function might be contained in sequences downstream of the TGG motif.

We introduced base substitution mutations into the F441 template downstream of the TGG motif as shown in Fig. 1. Two different protein binding experiments were performed. In one experiment, the complexes formed by the F441 template were competed for by the various mutant templates. In the other, the pattern of complex formation was directly investigated using the mutant templates as probes. In the first set of experiments, mutants F441-1 and F441-2 failed to compete with complexes C3 and C4 indicating that the mutated sites in
these templates were important for complex formation (Fig. 2c). The F441-3 template, however, abolished formation of these complexes. Therefore, the CC to AA mutation in the CCACCC box is not involved in the formation of the complexes C3 and C4.

We next synthesized two core sequences, one a 17 bp sequence without flanking CCACCC boxes, with a BamHI linker (F441-d11) which was cloned as a dimer at the BamHI site in the vector pBLCAT2. The other was an 18 bp sequence as a tandem dimer of a 9 bp core (F441-d12). When used as competitors, both F441-d11 and -d12 competed out the complexes C3 and C4, indicating that a minimal 9 bp domain present in F441-d12 is sufficient for the formation of the complexes C3 and C4. The behaviour of complex C1 was exactly the opposite, since both F441-1 and F441-2 competed efficiently with the formation of this complex, whereas F441-3, F441-d11 and -d12 did not. These results indicate the difference in sequence requirements for formation of complexes C3 and C4 compared to C1.

Similar conclusions were drawn from the second binding experiment where the mutant templates were used as the probes. The results showed that the templates, F441-1 and F441-2, that did not inhibit formation of the complexes C3 and C4 on the F441 template in the previous experiments, were similarly incapable of forming C3 and C4 complexes. F441-3 and F441-d11 could not form the complex C1. Unlike F441-3, F441-d11 formed only the complex C4 and not C3. Formation of C3 may require flanking sequences not present in F441-d11. However, formation of C4 might facilitate formation of C3, which could explain why F441-d11 competed for the C3 complex. The sequence requirement for the formation of C2 is less clear at present. Under our assay conditions, all mutant templates either formed this complex or competed with its formation to some extent.

**Defining the binding site of proteins forming complex C1**

Based on our *in vitro* and *in vivo* data (see below), it appears that the formation of C1 also helps to determine the biological activity of the oligonucleotide. Since factor binding to other enhancer elements present in the wt sequence might cooperate with the newly created motif to generate high level activity, we sought to determine
(a) Vector pBLCAT2

![Diagram of vector pBLCAT2 with BamHI site and TK, CAT genes.]

Conversion (%): 0.4 0.5 3.7 0.6 0.7 1.3
Fold activity: 1.0 1.2 10.3 1.1 1.4 2.7

(b) Vector pAOPCATT

![Diagram of vector pAOPCATT with SmaI site and A, OP, CAT genes.]

Conversion (%): 2.1 13.2 65.6
Fold activity: 1.0 4.4 19.2
the binding site of C1 in the methylation interference assay. An end-labelled probe was generated by digesting the labelled wt probe with BstNI. The larger end-labelled fragment formed complexes C1 and C2 (data not shown). Methylation interference experiments with C1 identified important contact points in the GC-rich region at the 3' end of the oligonucleotide (Fig. 3). The points of contact, as shown in Fig. 3, suggest that the protein forming the C1 complex binds to the CCACCCAATC sequence and makes a close contact with two guanine residues within the canonical CCAAT box. In addition, it comes in close contact with another guanine residue at nt 5243, four nucleotides upstream from the first C of the CCAAT box. It is interesting to note that this essential G residue is mutated in F441-3, which explains why F441-3 could not form a C1 complex.

Identification of proteins forming specific complexes with the F441 oligonucleotides

To identify proteins bound to the mutated F441 core and to the CAT box-like sequences, we performed protein binding studies with individual core sequences as probes and known regulatory protein-binding sites as competitors. As the F441 core sequence resembles the SV40 GT-IIC motif which binds the TEF-1 protein, we first performed protein binding studies with the SV40 motif as the probe. We demonstrated that nuclear proteins from the F9 cell extract formed a single complex with this probe (Fig. 4a). This complex formation is competed out by a self-competitor or by sequences from the adenovirus major late promoter which bind CP1, a member of the family of CAT box-binding proteins. However, other sequences with homology with this site, such as the binding site of NF-1 from adenovirus and the fibronectin gene and that of another CAT box-binding protein, CP2, in the γ-fibrinogen gene, did not abolish complex formation. This result demonstrated that CP1 or a CP1-like protein binds to the CAT box and is involved in generating enhancer activity as shown below.

Functional activity of the F441 oligonucleotide

To assess the functional significance of the protein binding data, we tested the ability of the oligonucleotides to potentiate expression of a reporter gene from a heterologous thymidine kinase (TK) promoter in F9 cells. This transfection experiment showed that the F441 oligonucleotide in isolation stimulated activity of the linked TK promoter in either orientation by eight- to 12-fold over the backbone. In contrast, the corresponding wt oligonucleotide was only weakly active (Fig. 5a).

We next studied the effect of various base substitution mutations within the F441 backbone. We found that the mutants F441-1 and F441-2 resulted in a drastic decrease of the enhancer activity to near wt levels, whereas mutant F441-3 had a moderate effect. The CAT activity of the two core sequences, F441-d11 and F441-d12, without the flanking CCACCC sequences increased to only about 1.2- and 1.9-fold above the backbone. Both of these core sequences were capable of forming complexes C3 and C4. This suggests that the short sequence CTGGAATGT, present as a dimer in F441-d12, can form the major complexes specific for the F441 template, but generates very little enhancer activity compared to the F441 template.

To test the effect of mutations in the context of a natural promoter, we used the vector pAOPCATT developed by Tseng et al. (1988). In this vector the PvuII 4 (enhancer B) fragment that contains the F441 mutation has been deleted from the regulatory region and replaced by a SmaI linker. Since the domain under investigation is part of this large fragment, we reinserted the F441 oligonucleotide at the SmaI site and evaluated its performance relative to the entire PvuII 4 fragment. As a control experiment, we first measured the activity of the whole B enhancer (PvuII 4 fragment) in F9 cells with and without the F441 mutation relative to the pAOPCATT vector (Fig. 5b, left panel). Although it was not possible

Fig. 5. Enhancer activity of the synthetic oligonucleotide in F9 (UD) cells in the context of TK promoter (a) and of Py promoter (pAOPCATT) (b). The details of the CAT assay are described in Methods. A representative autoradiogram is shown in the figure with percentage of chloramphenicol acetylated indicated at the bottom of each lane. The fold activity represents the mean value from duplicate or triplicate plates in a particular experiment. Designations of the samples are indicated at the top.
to compare the activity for pAOPCATT F441 because the reaction passed the linear range, the B enhancer without the mutation was approximately fourfold stronger than the vector; in a separate experiment, the activity of the enhancer with the mutation was about 10-fold more than the vector itself (Fig. 5b, right panel). We observed that in different sets of experiments the mutant B enhancer was generally two- to fourfold more active than the wt under our assay conditions.

Insertion of the wt oligonucleotide in the pAOPCATT vector caused no appreciable change in the enhancer activity, but the F441 oligonucleotide increased the total activity by about fivefold. Activity of the two core sequences (F441-d11 and F441-d12) was two- and threefold greater than that of the wt oligonucleotide. Thus the whole PvuII 4 fragment with the mutation increases the activity by at least twofold over that of the F441 oligonucleotide. These data indicate that although protein binding to the motif created by the F441 mutation plays an important role in generating the enhancer activity, other elements in the PvuII 4 fragment are required to achieve the full level.

Discussion

In the present study we characterized, functionally and biochemically, the locus of the simplest mutation with an A to G transition at nt 5233 that helps Py overcome the block in F9 cells. We have identified two motifs that bind proteins more strongly in the presence of the F441 mutation. One is an SV40 GT-IIC-like motif created by the point mutation itself which binds a TEF-1-like protein (Kovesdi et al., 1987; Lusky et al., 1983), and the other is a CAT box-like sequence contiguous with the downstream CCACCC box. A CDP-like protein binds to this CAT box. A cooperative interaction between these two proteins is probably required to generate enhancer activity in this locus.

Our observation on increased activity of the F441 mutant and the binding of an activator to a site spanning the point mutation agrees, in general, with those published from other laboratories (Kovesdi et al., 1987; Satake et al., 1988; Tseng & Fujimura, 1988). The relative activity of the F441 oligonucleotide (about 10-fold) reported here was, however, higher than that observed by another group (Satake et al., 1988). There are several differences between our methods and those of the other group. We used a different vector (pBLCAT2) for assay of enhancer activity containing a TK promoter instead of an SV40 promoter in pA10CAT2. Furthermore, the oligonucleotide we tested contained both copies of the BPV core sequence CCACCC, whereas that used by Satake et al. (1988) contained the upstream CCACCC box only. We showed that the last two C residues of the downstream CCACCC sequence are important for binding of another factor from F9 cells mentioned above, which has a modest effect on enhancer activity.

The core sequence created by the F441 mutation has only one base pair mismatch with the SV40 GT-IIC motif: a C is substituted by a G at the first position. A TEF-1 protein purified from HeLa cells binds to this motif to generate enhancer activity (Davidson et al., 1988; Xiao et al., 1991). Our DNA binding assays with cross-competitors demonstrated that the factors binding to these sequences from F9 cells are indistinguishable from TEF-1. The isolated GT-IIC motif is active as a tandem dimer but not as a spaced dimer (Davidson et al., 1988). Although the activity is not high, it is interesting to note that in the pAOPCATT backbone, F441-d12 with a tandem dimer of the motif is more active than F441-d11 (equivalent to a spaced dimer), characteristic of a GT-IIC motif (Davidson et al., 1988; Fromenthal et al., 1988). A 2 bp mismatch in the dimer sequence with the SV40 GT-IIC motif might alter the affinity of the protein binding to this sequence and reduce the corresponding functional activity.

Although the mutated sequence together with the flanking CCACCC box closely resembles the NF-1 binding site, and a purified HeLa cell NF-1 also can bind to this sequence in vitro (Tseng & Fujimura, 1988), it is unlikely to be involved in mediating the F441-specific function. The protein forming the F441-specific complex is highly abundant in F9 cells, whereas NF-1 is present in these cells at a very low level (our unpublished observation; McQuillan et al., 1991; Speck & Baltimore, 1987). The second half of the putative binding site is not required for the formation of the high affinity C3/C4 complex and mutation in the spacer region is also deleterious for its activity in vivo, in contrast to what is observed for NF-1 (Rosenfeld et al., 1987). However, the second half of the putative NF-1-like motif is contiguous with the CAT box-like sequence CCACCCAT and binds CDP, a member of the family of CAT box-binding proteins (Chodosh et al., 1987; Cohen et al., 1986). However, the method by which these two proteins interact and generate activity remains to be investigated.

It is not known whether the upstream CCACCC sequence is functionally important and binds any factor. However, F441-d11, which does not contain either of these elements, is less active than F441-3 in which only the downstream copy is inactivated by mutation (Fig. 5). This sequence, together with the F441 core sequences, was protected against DNase I digestion at higher concentrations of proteins (Kovesdi et al., 1987). When the activities of F441-d11 and F441-3 are compared with that of the F441 oligonucleotide, the integrity of the CCACCCAT motif together with an intact GT-IIC-like
motif appears important. This suggests that a cooperative interaction between TEF-1 and CP1-like proteins primarily generates the activity of this locus. Several other functionally important protein-binding sites in the B enhancer were identified earlier (Ostapchuk et al., 1989; Tseng et al., 1988). These results, taken together, suggest that the activity of the whole PvuII 4 fragment (B enhancer) involves an interaction of multiple proteins with the one bound to the site of the F441 mutation. This interaction is again probably extended to proteins bound to the A enhancer region at higher levels of organization of the enhancer (Fromenthal et al., 1988; Ondek et al., 1988). Understanding the mechanisms involved in enhancer function of many Py mutants will require identification and a detailed biochemical characterization of the regulatory proteins present in F9 (UD) cells. The present work is an important step in that direction.

We are grateful to Dr F. Fujimura, Nichols Institute, California and to Dr B. Luckow, Germany, for providing us with the plasmids, and to Dr D. Dean, Washington University School of Medicine, St Louis, Missouri, for sharing the synthetic probes. We thank Dr A. Johnson for reading the manuscript. This work was supported by Public Health Service Grant, CA 47611 from the National Institutes of Health.

References


(Received 14 July 1992; Accepted 11 November 1992)