Multiple *cis*-active elements in the long control region of bovine papillomavirus type 1 (BPV-1)

S.M.Harrison, K.L.Gearing[1], S.-Y.Kim[2], A.J.Kingsman and S.M.Kingsman*

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

ABSTRACT

A 1.0 kb region of the BPV-1 genome (the long control region, LCR), contains controls for transcription and the origin of replication. Transcription directed by the LCR is activated by the viral encoded E2 protein. To define the essential cis acting elements that are required to control transcription we have constructed a series of deletions throughout the LCR. We have identified three important domains in the LCR, two of which respond to E2. We have analysed the ability of small subcloned regions of the E2 responsive domains to act as enhancers in a heterologous assay system. This has led to the identification of five independent E2 responsive elements. We have shown that a fragment of only 38 base pairs is sufficient to respond to activation by E2. We also present evidence to suggest two types of E2 responsiveness that result in strong or weak activation. Strong response is correlated with the presence of the sequence 5'-ACCG/TNNNC/TCGGTGC-3' whereas weak response is correlated with the presence of a related sequence 5'-ACC(N)$_6$GGT-3'. The contribution of these multiple elements to viral transcription is discussed.

INTRODUCTION

Bovine papillomavirus type 1, a member of the papovavirus group, is a small DNA tumour virus that causes benign skin papillomas (warts) in cattle and can be made to induce fibroblastic tumours in some other species e.g., hamsters and rabbits. The virus readily transforms rodent cells in culture but infectious virus has not been obtained from any cultured cell lines (1, 2, 3). Within these transformed cells the viral DNA exists as an autonomous plasmid, maintaining a copy number of 10-200 copies per cell (3).

The BPV-1 genome is 7946 base pairs and a number of significant i.e., longer than 400 base pair open reading frames (ORFs) have been deduced; these are all on the same strand ((4); Figure 1). The genome is transcribed from a single strand to

yield a complex array of spliced transcripts (5, 6, 7, 8), the full array of which may not yet have been identified. Three sets of transcripts in transformed cells initiate within a 1 kb region called either the long control region (LCR) or the upstream regulatory region (URR) (9, 10). The major capsites have been mapped to positions 89, 7185 and 7940, (6, 7, 14, 15) and a set of late transcripts, identified in warts, initiate around 7220 (15). Two additional initiation sites at 2433 and 3080 have also been mapped (6, 7). The LCR also contains the origin of replication (11) and contains cis-active elements that are at present poorly understood but that are required for plasmid maintenance and copy number control (12, 13).

The functions of the various ORFs is being deduced by genetic analysis and by expressing individual ORFs in expression vectors. E5 and E6 are transforming proteins (7, 16, 17), E1 and E7 are required, and E6 may also be important, for replication (18, 19). The E2 ORF is essential for viral gene expression, its gene product activates transcription via sequences that are located within the LCR (20, 21). Both E1 and E2 ORFs have a complex domain structure and may encode both positive and negative regulators of replication and transcription (18, 19, 20, 22). The functions of E3, E4 and E8 ORFs are not known although the HPV1a E4 ORF may encode a late protein involved in virus maturation (23).

To understand both the mechanisms of transformation by BPV-1 and the mechanism of autonomous and copy number controlled replication it is necessary to define the cis active regulatory sequences in the genome and show how they interact with cellular and viral factors. We have used both deletion analysis and subfragment screening to identify essential transcriptional control elements in the BPV-1 LCR. We have discovered a complex series of functionally distinct elements and have also narrowed down an E2 responsive region to a 38 base pair fragment.

MATERIALS AND METHODS
Bacterial strains, mammalian cell lines and media
E. coli strain AKEC28 (C600, thrC, leuB6, thyA, trpC1117, hsdRK, hsdMK) was used for plasmid cloning and preparation. E. coli cultures were grown in Luria broth (24), and ampicillin (Sigma) was added when necessary at 50 µg/ml.
Mouse C127 cells (25) were maintained in DMEM supplemented

with 10% foetal bovine serum, penicillin (100 u/ml) and
streptomycin (100 μg/ml).

DNA manipulations

    Molecular cloning was performed by standard protocols (26).
Restriction enzymes, Klenow fragment of DNA Polymerase 1, T4 DNA
ligase and Bal31 exonuclease were obtained from BRL. Non-
phosphorylated HindIII and XhoI molecular linkers were obtained
from Collaborative Research Inc, BglII linkers were from Celltech
Ltd. DNA fragments were purified from agarose or polyacrylamide
gels as described by Tabak and Flavell (27) and Maniatis et al.
(26), respectively. Bal31 deletion endpoints were sequenced by
the dideoxy chain termination method of Sanger et al. (28).

Mammalian cell transfections

    Mouse C127 cells were transfected by the calcium phosphate
coprecipitation method (29) using 5 μg of CAT plasmid, 5 μg of
BPV ORF expression vector (pMA821) and 5 μg of plasmid pCH110 per
50 mm dish. Four hours after the addition of the precipitate, the
cells were treated with 15% glycerol for two minutes (30), washed
and incubated in complete media for 48 hours. A crude cell
extract was then prepared by sonication (3X 5 second bursts at
low-medium power using a microtip) in 0.25M Tris HCl pH7.8 at
room temperature.

    Plasmid pCH110 contains the SV40 early promoter region
driving the expression of the β-galactosidase gene, and is used
to correct for minor differences in transfection efficiency of
different plasmid preparations. These variations were usually
less than two fold and were not correlated with any individual
plasmid construction. Southern blot analysis of total cellular
DNA isolated 48 hours post transfection indicates that all
plasmids were present at equivalent levels irrespective of
whether the full or deleted BPV LCR was present (data not shown).

CAT and β-galactosidase assays

    Assays for chloramphenicol acetyltransferase (CAT) activity
were performed by the method of Gorman et al. (31). Crude cell
extracts were incubated with ^{14}C-chloramphenicol (Cm) (Amersham,
50 mCi/mmol) in 0.25M Tris pH7.8 and 0.5 mM acetyl CoA (Sigma)
for one hour at 37°C. Conversion of ^{14}C-Cm to its acetylated
derivatives was monitored by thin layer chromatography and
quantitated by scintillation counting.

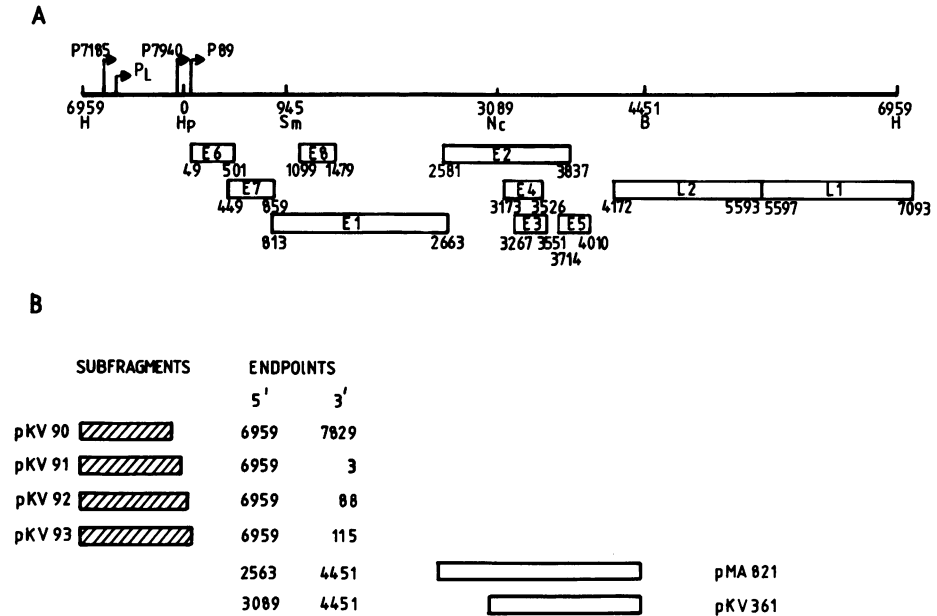   β-galactosidase assays were performed as previously
described (32).

A

B

Figure 1: The BPV-1 genome and location of subfragments.
A- represents the full length BPV-1 genome linearised at the
unique HindIII site. Major open reading frames are indicated
(open boxes). Numbers at the 5' end of each ORF refer to the
first nucleotide of the ORF, those at the 3' end indicate the
nucleotide preceding the stop codon. Three identified early
promoters active in transformed mouse cells located within the
LCR   P7185, P7940 and P89 and direction of transcription are
indicated.   PL represents the major late promoter active in
fibropapillomas. Restriction sites are H = HindIII; Hp = HpaI;
Sm = SmaI; Nc = NcoI; B = BamHI.
B- Hatched boxes represent BPV-1 LCR fragments inserted in the
"sense" orientation into the promoterless CAT plasmid pSVØCAT to
create plasmids pKV9Ø to 93. Open boxes represent BPV-1 coding
region inserted in the "sense" orientation downstream of the RSV
LTR in the vector pMA8Ø2 to create plasmids pMA821 and pKV361.
The positions of 5' and 3' endpoints of each fragment are
indicated.

Plasmids and plasmid construction

Standard plasmids: pCGBPV9 (33) was used as a source of BPV-1
HindIII-BamHI early region, pSVØCAT (31), pRSV-β-globin (35),
pA$_{10}$CAT (36), pAT153 (37) and Sp46 (34) have been described
previously. Plasmid pCH11Ø was obtained from Pharmacia.

Construction of BPV-1 LCR promoter-CAT fusion plasmids

The BPV-1 69% HindIII-BamHI transforming fragment (corresponding to base pairs 6959-4451 of the BPV-1 genome according to the numbering system of Chen et al., (4) see Figure 1) was inserted between the HindIII and BamHI sites of pAT153 to create an intermediate plasmid, pMA800. A HindIII molecular linker (GCAAGCTTGC) was inserted into the unique HpaI site, and the resultant HindIII fragment containing the LCR was isolated and inserted in the "sense" orientation into the HindIII site of pSVØCAT to create plasmid pKV91 (see Figure 1). Three derivatives of pMA800 containing HindIII sites in the region of the HpaI site were also created. pMA800 was linearised at the unique SmaI site (see Figure 1) and digested with exonuclease Bal31 at 37°C for 2-10 minutes. The ends of the deleted molecules were filled in with the Klenow fragment of DNA Polymerase I and ligated in the presence of excess HindIII linkers. Deletion endpoints were determined by DNA sequencing. HindIII fragments with 3' endpoints at base pairs 7829, 88 and 115, relative to the unique HpaI site, were inserted in the "sense" orientation into pSVØCAT to create plasmids pKV90, pKV92 and pKV93. The general structure of all these plasmids is the same as pKV93 shown in Figure 2A; the LCR fragments are shown in Figure 1.

Construction of BPV-1 LCR promoter deletions

Deletions throughout the LCR present in pKV93 were created (Figure 2). The HindIII site at the 3' end of the LCR was converted to a BglII site by HindIII partial digestion of pKV93 followed by end-fill and insertion of BglII molecular linkers (CAAAAGATCTTTTG) to create plasmid pKll. Deletions in the 5' to 3' direction and in the 3' to 5' direction were created by Bal31 digestion from the unique HindIII and BglII sites of pKll, respectively, followed by end-fill and insertion of XhoI linkers (CCTCGAGG). Internal deletions were created by ligating XhoI-BamHI purified fragments from each series of deletions as indicated (Figure 2). Molecules deleted from the 5' end of the LCR only were repaired by ligating XhoI-BamHI fragments from the 5' to 3' deletion series to the XhoI-BamHI fragment of pKV151 as indicated (pKV151 was created by converting the HindIII site of pKll to an XhoI site by the insertion of XhoI linkers). This simply served to reconstruct the flanking sequence adjacent to the 5' end points of the deletions. All molecules created in
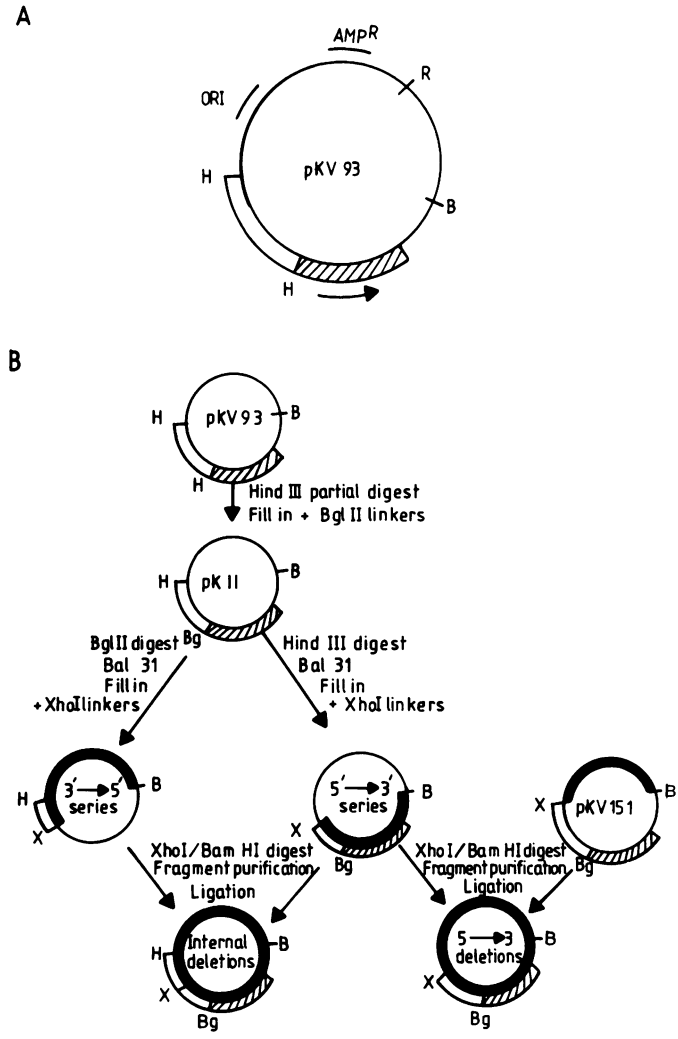
Figure 2: Strategy for construction of BPV-1 LCR promoter deletion CAT fusion plasmids.
Open box represents BPV-1 LCR, hatched box CAT coding region with direction of transcription indicated, thin line represents splice and polyadenylation signals of SV40 and the pBR322 fragment containing the origin of replication and ampicillin resistance gene. Restriction sites are H = HindIII; B = BamHI; X = XhoI; Bg = BglII; R = EcoRI.
A - Structure of pKV93.
B - Construction of deletions. Solid boxes represent fragments of intermediate plasmids used in construction of the final CAT plasmid containing the deleted LCR.

this way were therefore identical to plasmid pK11 except for deletions in the LCR. The endpoints of all deletions are indicated in Figure 4.

Construction of pA$_{10}$CAT derivatives containing BPV-1 LCR restriction fragments

Plasmid pK11 was digested to completion with HindIII, BglII and RsaI to generate fragments 3, 6, 7, 8 and 9 (Figure 5) that together span the LCR. These were end-filled and inserted into the HincII site of plasmid Sp46. The LCR fragments were then removed as BglII/BamHI fragments for insertion in both orientations into the BglII site 5' to the SV40 early promoter in the enhancer negative CAT plasmid pA$_{10}$CAT. Similarly the smaller restriction fragments 4 and 5 were inserted via Sp46 into pA$_{10}$CAT. Fragment 4 is a 190 base pair RsaI-HpaI fragment. Fragment 5 is a 38 base pair RsaI-TaqI fragment (see Figure 5 for coordinates). Fragments 1 and 2 correspond to the LCR fragments in pKV351 and pKV352 of the promoter deletion series.

Construction of RSV-BPV ORF expression vectors

Plasmid pMA802 was constructed by replacing the HindIII-BglII β-globin cassette from pRSV-β-globin (35) with a unique BglII site between the RSV long terminal repeat and the SV40 splice and polyadenylation signals. Plasmid pMA821 was created by inserting a BglII linker 44 base pairs upstream of the first ATG of the E2 ORF (Figure 1b) and inserting the resultant BglII/BamHI fragment, containing intact E2, E3, E4 and E5 ORFs into the BglII expression site of pMA802. Plasmid pKV361 was created by inserting a BamHI molecular linker (CCGGATCCGG) at the NcoI site of pMA800 (position 3089 in the BPV-1 genome) and insertion of the resultant BamHI fragment, containing a truncated E2 ORF and complete E3, E4 and E5 ORFs, into the BglII expression site of pMA802.

RESULTS

Defining the major promoter region

Our first aim was to generate a suitable DNA fragment that exhibited maximum promoter activity, as measured by CAT activity, as a starting point for our analysis. Because the 69% BPV fragment is fully transforming, the 5' limits of the promoter appear to be at or downstream from the HindIII site at base pair 6959 (2). To define a suitable 3' boundary a series of fragments was tested for promoter activity using the standard assay plasmid

TABLE 1

Deletion mapping of the 3' end of the BPV-1 LCR promoter and E2 responsive region (see Figures 1 and 3).

Plasmid	CAT Activity (% Cm acetylated)		
	-pMA821	+pMA821	+pKV361
pKV90	<0.1	<0.1	-
pKV91	<0.1	1.5	-
pKV92	<0.1	5.0	-
pKV93	<0.1	27.0	<0.1
Mock	-	<0.1	-

pSV0CAT (31). The fragments, shown in Figure 1, share a common 5' end at base pair 6959 but differ at their 3' ends which terminate at base pairs 7829 (pKV90), 3 (pKV91), 88 (pKV92) and 115 (pKV93). Each of the plasmids was introduced into mouse C127 cells either alone or together with plasmid pMA821. This expresses the viral E2 protein which is essential for promoter activity (20, 21). The strength of the promoter fragments was assessed by measuring CAT activity in cell extracts prepared 48 hours after transfection (Figure 3; Table 1). There is a five fold decrease in promoter activity when the fragment is deleted from 115 (pKV93) to 88 (pKV92). Deletion to the HpaI site at base pair 3 reduces the activity even further and there is minimal CAT activity (less than 0.1%) with pKV90. These data indicate that retention of the integrity of the major capsite at base pair 89 is essential for full promoter activity in this assay. We therefore chose pKV93 as the starting plasmid for our analysis.

This preliminary study also shows that pKV93 is totally dependent upon pMA821 for promoter activity as no CAT activity was detected in the absence of pMA821. Also, when a derivative of pMA821 that is deleted of the first 483 bp of the E2 ORF, (pKV361) was transfected with pK11 no CAT activity was observed. This indicates that expression of the intact E2 ORF is required for transcriptional activation, although cooperative effects with E3, E4 and E5 ORFs cannot be excluded.

Identification of multiple cis active elements

To localise the essential cis active elements within the BPV-1 promoter region a series of one-sided and internal deletions were constructed within the LCR from pKV93. The
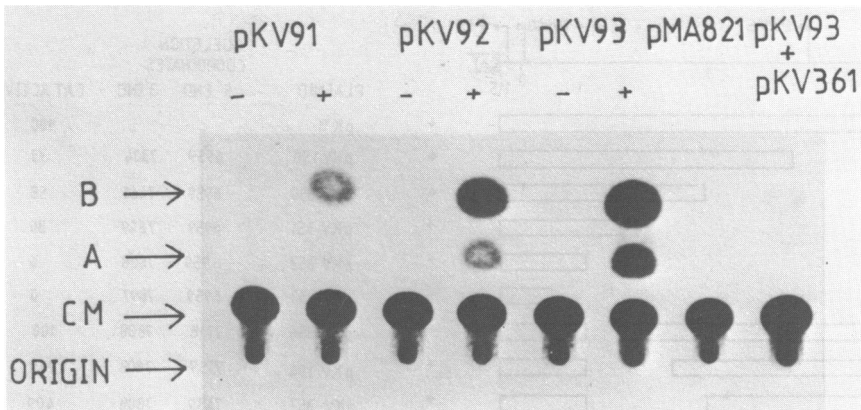
Figure 3: Autoradiograph of CAT assay illustrating deletion mapping of the 3' end of the BPV-1 LCR. Each plasmid, pKV91-93, was transfected into C127 cells in the absence (-) or presence (+) of pMA821. pKV93 was also transfected in the presence of pKV361. A transfection containing only pMA821 was included as a negative control. CAT assays were performed as described. A and B represent the monoacetylated derivative of Cm. CAT activities of each have been quantified and are presented in Table 1.

deleted derivatives were assayed for promoter activity in the presence of pMA821. The data are summarised in Figure 4. To aid a simple interpretation we have scored any CAT values over 10% of wild type levels as positive and below 10% as negative; the real values are also presented.

Deletions from the 5' end point at 6959 showed some reduction in CAT activity to between 33 to 80% of wild type levels but when an end point at 7808 was reached in pKV352, there was a dramatic drop to 6% of full activity. These data indicate that a key element is located between nucleotides 7649 and 7808. For the sake of discussion at this point, we refer to this as domain I. Analysis of internal deletions revealed that other regions of the LCR were also significant. The results with pKV356 show that removal of nucleotides 7718 to 7808 has no effect. This deletion retains part of domain I and taken alone might suggest that this domain can be further delimited to 7649 to 7718. The results with deletion pKV194, however, show that deletion pKV356 cannot be used to delimit domain I. The fragment in pKV194 has completely lost domain I and has the same 3' end point as pKV352 which lacks activity but in this case there is full promoter activity. The levels of CAT are also reproducibly
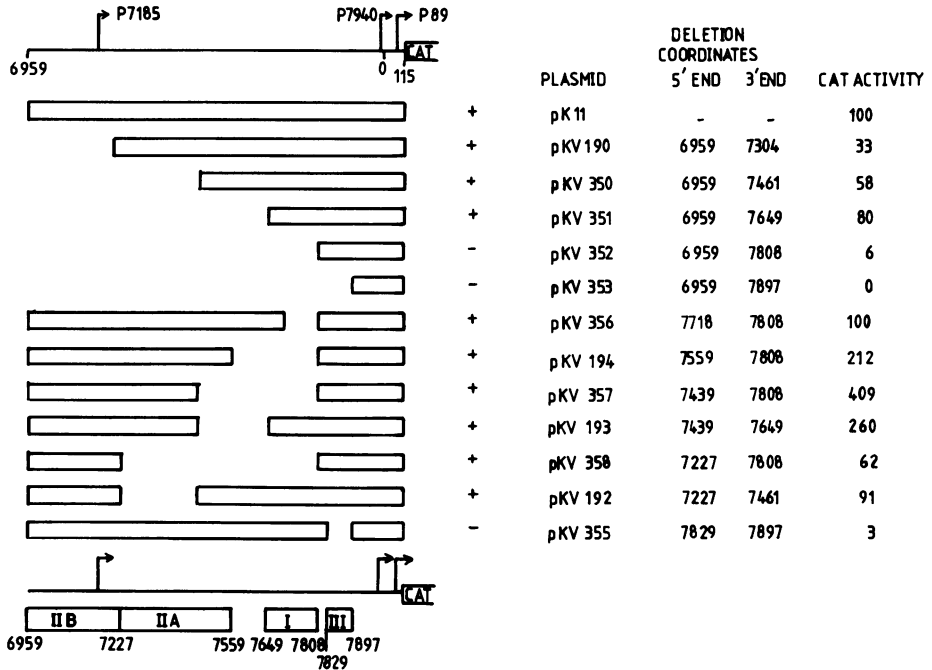
	PLASMID	DELETION COORDINATES 5' END	3' END	CAT ACTIVITY
+	pK 11	-	-	100
+	pKV 190	6959	7304	33
+	pKV 350	6959	7461	58
+	pKV 351	6959	7649	80
-	pKV 352	6959	7808	6
-	pKV 353	6959	7897	0
+	pKV 356	7718	7808	100
+	pKV 194	7559	7808	212
+	pKV 357	7439	7808	409
+	pKV 193	7439	7649	260
+	pKV 358	7227	7808	62
+	pKV 192	7227	7461	91
-	pKV 355	7829	7897	3

Figure 4: Analysis of the cis-acting sequences essential for BPV-1 LCR promoter activity.
Solid line indicates BPV-1 LCR fragment (base pairs 6959-115) present in pK11. Position of LCR early promoters and CAT gene are indicated. Open bars represent extent of LCR in each deletion. Each deletion derivative has a plasmid name, deletion endpoints (coordinates indicate the last base pair not deleted) and corresponding normalised CAT activity as indicated. Below is a schematic representation of domains referred to in the text.

two fold higher than wild type levels. Similar results are obtained with the larger deletion in pKV357. We interpret these data as suggesting the presence of a second functional region which we refer to as domain II which can completely compensate for the loss of domain I. The presence of this second domain may therefore explain the positive result in pKV356. The two to four fold higher levels of CAT activity when domain II is functioning may reflect a higher activity of domain II as compared to domain I. Alternatively, the deletions may remove a negative element. Further deletions into the upstream region as in pKV358 which also lacks domain I still retains promoter activity. This activity is not as high as wild type. These data can be

interpreted either as indicating that the deletion disrupts the full activity of domain II or that there are additional weaker compensating elements in this region. At this point we locate domain II between 6959 and 7559 but divide it into two subdomains A and B to reflect the differences in promoter activity. The internal deletions also reveal a third important region. This is shown by the deletion in pKV355 which removes base pairs 7829 to 7897. This deletion retains both domain I and domain II but lacks promoter activity. The deletion removes 67 base pairs downstream from domain I and this region is referred to as domain III.

These deletion studies indicate a complexity in the BPV-1 promoter. They suggest that there are at least two functionally equivalent regions, domains I and II either of which can provide full promoter activity. These are not sufficient as they require domain III to be present. Domain III cannot function alone e.g., pKV352 has domain III but lacks domains I and II. The reproducible modulations in promoter activity with the different deletions suggests that the domains may comprise multiple elements. The tentative coordinates of the domains are indicated in Figure 4.

Positive location of enhancer elements

Our deletion analysis has identified three functional domains in the BPV-1 promoter, but some regions may be missed in such an analysis because of compensation by functionally equivalent elements. The juxtaposition of the different domains may also influence the results. We, therefore, sought a positive identification of some of the upstream control elements by assaying the activity of subfragments of the LCR in $pA_{10}CAT$ (36). This plasmid retains the SV40 capsite but lacks the SV40 enhancer and can therefore be used to assay for enhancer elements independent of their ability to direct transcription initiation. The locations of the fragments used in this analysis and the respective CAT activities relative to the background level in $pA_{10}CAT$, taken as 1.0, are shown in Figure 5. All the fragments were tested in both orientations and in all cases where CAT activity was obtained, it was absolutely dependent upon the presence of E2 and therefore data without pMA821 is not presented.

Fragment 1 is derived from pKV351 and contains domains I and III. Fragment 2 is derived from pKV352 and has domain III.

Fragment 2 had no activity in either orientation whereas fragment 1 was fully active. These data confirm the results of the deletion analysis and indicate that domain I defines a region that responds to E2 trans-activation and that domain III does not. Fragment 1 was significantly orientation-dependent being three fold more active in the "sense" direction. To localise further the E2 responsive region in domain I fragments 2-6 were analysed. Fragment 3 which extends from base pair 7760 to 115 was active which localises a region essential for E2 responsiveness to between 7760 and 7808. This is confirmed by fragment 4 which spans 7760 to base pair 3. Both fragments 3 and 4 also showed orientation dependence with stronger activity but this time in the "antisense" direction. However, the even smaller 38 base pair fragment 5 was still active but was now orientation-independent. These data indicate that there is a bidirectional E2 responsive region between 7760 to 7797 and that adjacent regions impose an orientation dependence. We have called this E2 responsive region element A. The next upstream fragment also derived from domain I, fragment 6 was also active and displayed absolute orientation-dependence; there was no activity in the "sense" orientation. This, therefore, defines a second E2 responsive region and orientation determinant which we have called element B. The activity of domain I identified in the deletion series can, therefore, be explained by the presence of two E2 responsive regions. Fragment 7 which contains part of domain II has some activity in both orientations, this is relatively weak but reproducibly above base level and therefore defines a third E2 responsive region called element C. Fragment 8 from within domain II has strong activity in both orientations although marginally more effective in the "antisense" orientation and this has been called element D. Fragment 9 which is in domain IIB, has very weak activity which is absolutely orientation-dependent in the sense direction and therefore this must contain a fifth element, element E. The activity of domain II identified in the deletions can, therefore, be explained by the presence of elements C, D and E. The modulation of promoter activity in the different deletions through this region probably reflects the disruption of the stronger element D and the contributions of the weaker flanking elements C and E. The tentative location of these elements, their strengths, orientation dependence, and relation to domains I, II and III are shown in Figure 5.
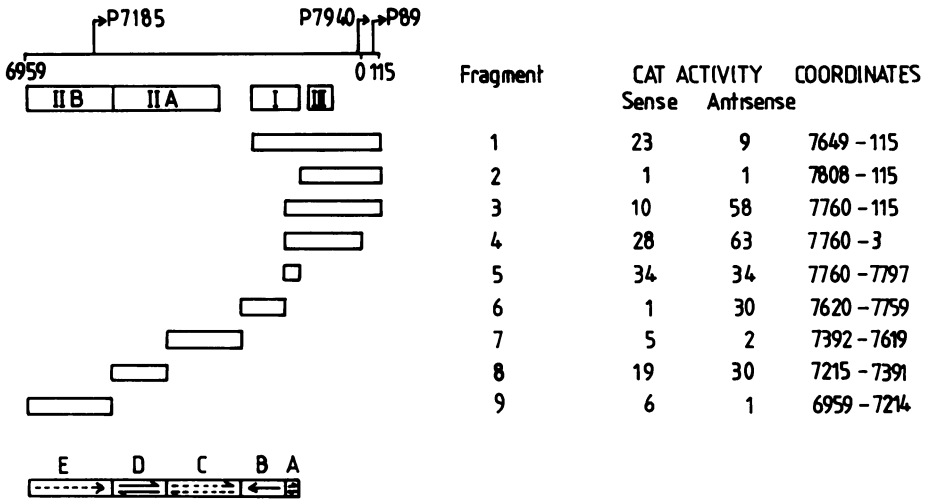
P7185   P7940 P89

6959   0 115

II B | II A | I | III

Fragment	CAT ACTIVITY		COORDINATES
	Sense	Antisense	
1	23	9	7649 - 115
2	1	1	7808 - 115
3	10	58	7760 - 115
4	28	63	7760 - 3
5	34	34	7760 - 7797
6	1	30	7620 - 7759
7	5	2	7392 - 7619
8	19	30	7215 - 7391
9	6	1	6959 - 7214

E   D   C   B A

Figure 5: Localisation of E2 responsive domains within the BPV-1 LCR. Solid line indicates BPV-1 LCR fragment (base pairs 6959-115) present in pK11. Position of LCR early promoters are indicated. Immediately below is a schematic representation of elements referred to in the text and Fig. 4. Below this open bars indicate restriction fragments cloned in both orientations into the enhancerless CAT plasmid $pA_{10}CAT$. Fragment endpoints and numbers used in the text are indicated. Each construct was transfected in the presence and absence of pMA821. Only in the presence of pMA821 was CAT activity detectable and included in the figure. CAT activity is expressed relative to that obtained by transfecting $pA_{10}CAT$ in the presence of pMA821. Below the relative strengths and orientation dependence of E2-dependent enhancer fragments identified are indicated by the direction and intensities of the boxed arrows.

DISCUSSION

Using one-sided and internal deletions of the BPV-1 LCR we have identified three functional domains. We have analysed these domains by testing the ability of subfragments of the LCR to activate transcription in the enhancer assay plasmid $pA_{10}CAT$. One of the domains, domain III, is most likely to contain upstream elements involved in binding cellular transcription factors such as Spl and CAAT transcription factor (40, 15), and start elements (TATA boxes) associated with the P7940 and P89 promoters as it is absolutely essential for promoter function but it has no inherent enhancer activity. The other two domains, I and II, are absolutely dependent upon E2 for their function.
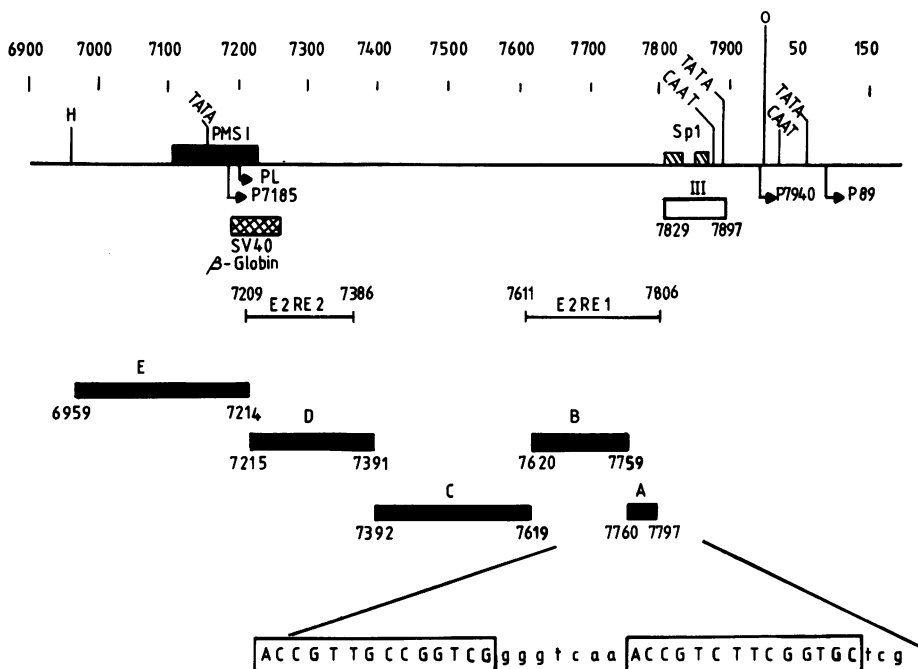
Previous studies have only identified one (9) or two E2

Figure 6: A summary of cis-acting sequences within the LCR. The whole LCR is represented with associated RNA cap sites at base pairs 7185, ~7200, 7940 and 89, and potential CAAT and TATA boxes. Open box is domain III identified in this study. Regions of the LCR with homology to sequences involved in the interaction with cellular transcription factors are indicated (hatched boxes). E2 responsive regions localised by Spalholz et al. (10) are labelled E2RE1 and E2RE2. Those localised in this study (A-E) are shown as solid boxes and the sequence of the shortest, element A, is shown with E2 response sequences outlined.

responsive regions designated by Spalholz et al. as E2RE1 and E2RE2 (10). E2RE1 has been shown to function as a 195 base pair fragment (nucleotides 7611 to 7806) and E2RE2 has been tentatively located to a 177 base pair region (nucleotides 7209 to 7386 (Figure 6). We have now shown that E2RE1, in fact, contains two distinct E2 responsive elements A and B. Element D is contained within E2RE2 and elements C and E have not been previously identified. This finding of at least five independent E2 responsive regions indicates that the BPV LCR has a more complex array of transcriptional controls than previously thought.

E2 responsiveness appears to be mediated by the direct

TABLE 2

The ACC(N)$_6$GGT motifs implicated in E2 response.

MOTIF		SEQUENCE	COORDINATES	BINDING	ELEMENT	ACTIVATION
1	P	ACCACACCCGGTAC	7203-7216	-	E	WEAK
2	F	ACCGCCGCCGGTGC	7376-7363	+	D	STRONG
3	P	ACCTATCCCGGTAA	7408-7421	-	C	WEAK
4	P	ACCAGTTCTGGTTC	7521-7508	-	C	WEAK
5	P	ACCAGTAATGGTGC	7591-7604	-	C	WEAK
6	F[#]	ACCGCCATCGGTGC	7620-7633	X	B	STRONG
7	F	ACCTATATCGGTGC	7645-7632	+	B	STRONG
8	P	ACCGTTGCCGGTCG	7760-7773	+	A	STRONG
9	F	ACCGTCTTCGGTGC	7781-7794	X	A	STRONG
10	P	ACCGGTTTCGGTGA	7907-7894	*	-	NONE

The sequence and location of each motif, together with its presence (+) or
absence (-) in the high-affinity binding restriction fragments identified
by Androphy et al. (39). X refers to motifs destroyed in generating the
restriction fragments. The location of each motif within enhancer
elements identified in this study, and the corresponding activity of
elements are indicated alongside.
* Some binding was observed with high concentrations of E2 (39).
[#]Footprinted by Moskaluk and Bastia (38).
P = Partial E2 response sequence.
F = Full E2 response sequence.

binding of the E2 protein to the DNA (38, 39). Androphy et al.
(39) identified three restriction fragments within the LCR (base
pairs 7336-7406, 7624-7683 and 7767-7822) that bind E2 very
efficiently.   These contain the consensus sequence 5'-
ACCG/TNNNC/TCGGTGC-3' which we call the "full E2 response
sequence".   There are, however, seven other closely related
sequences scattered through the LCR (Table 2) which we have
called "partial E2 response sequences" that have the general
consensus sequence 5'-ACC(N)$_6$GGT-3'. One restriction fragment
that contained the full response sequences 6 and 7 was subjected
to DNaseI footprinting. A 23 base pair protected region was
identified that contained response sequence 6 but response
sequence 7 on the opposite strand was not protected (38). The
significance of these sequences identified by in vitro protein
binding is therefore not yet fully clear. Our functional
analysis shows that the three elements which exhibits the

strongest enhancer activity, elements A, B and D all contain a
copy of the full response sequence. Element B has two copies and
element A has one copy and an additional partial response
sequence. Element A is only 38 base pairs and is the smallest E2
responsive region identified to date.   The E2 response sequences
of element A (boxed in Figure 6) are separated by 7 base pairs
and there is only 3 additional base pairs at the 3' end. This
implies that flanking sequences are not required for maximum
activation and support the importance of the E2 response
sequences for transcriptional activation.

     Our analysis also implies that the partial motifs may be
significant in determining the response to E2. The elements
that we have shown to have weak activity, elements C and E
contain three and one partial motif, respectively. Fragments
containing these functional E2 responsive elements did not,
however, appear to bind E2 in vitro (38, 39). This may reflect
either a very weak binding affinity, or the E2 protein may be
modified in vivo to effect binding. There is one case where we
have failed to detect E2 dependent enhancer activity with a
fragment that contains an E2 response sequence. Fragment 2
(Figure 5) has a partial response sequence but is negative in our
assay. E2 protein binds to this region in vitro but only when
high concentrations are used (39).

     The significance of the orientation dependence of some of
the E2 motifs is, at present, unclear. Orientation dependence is
clearly influenced by context. It is possible that there are
sequences that limit the range of activity of the different
E2-response elements so that different transcript start sites can
be independently regulated. It is also possible that the various
E2 response regions interact differentially with the different
promoter elements that determine the 5' cap sites. They may also
contain additional sequences that interact with cellular factors.
Element D contains a region of homology with the SV40 late
promoter (42), the "octamer" element found in viral and cellular
enhancers (43) and to a β-globin upstream element (41).

     In conclusion, we have demonstrated a marked complexity in
the BPV LCR showing multiple elements capable of functioning
independently but which vary in their efficiencies. Our data
would support a notion that efficiency is correlated with the
strength of E2 binding to the E2 response sequence. We have
shown that E2 responsiveness can be conferred by a 38 base pair

fragment that contains only two E2 response sequences. It is likely that finer structural analysis will reveal additional elements and that there will be a complex interplay between these elements and both viral and cellular proteins. Unravelling these complexities will provide insight into the mechanism of transcriptional control by BPV-1.

Present addresses: [1]NERC, Institute of Virology, Mansfield Road, Oxford, UK and [2]The Whitehead Institute, Boston, MA, USA

*To whom correspondence should be addressed

REFERENCES
1. Law, M.-F., Lowy, D.R., Dvoretzky, I. and Howley, P.M. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**, 2727-2731.
2. Lowy, D.R., Dvoretzky, I., Shober, R., Law, M.-F., Engel, I. and Howley, P.M. (1980) Nature **287**, 72-74.
3. Lancaster, W.D. (1981) Virology **108**, 251-255.
4. Chen, E.Y., Howley, P.M., Levenson, A.D. and Seeburg, P.H. (1982) Nature **299**, 529-534.
5. Heilman, C.A., Engel, L., Lowy, D.R. and Howley, P.M. (1982) Virology **119**, 22-34.
6. Stenlund, A., Zabielski, J., Ahola, H., Moreno-Lopez, J. and Pettersson, U. (1985) J. Mol. Biol. **182**, 541-554.
7. Yang, Y.-C., Okayama, H. and Howley, P.M. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 1030-1034.
8. Engel, L.W., Heilman, C.A. and Howley, P.M. (1983) J. Virol. **47**, 516-529.
9. Haugen, T.H., Cripe, T.P., Ginder, G.D., Karin, M. and Turek, L.P. (1987) EMBO J. **6**, 145-152.
10. Spalholz, B.A., Lambert, P.F., Yee, L. and Howley, P.M. (1987) J. Virol. **61**, 2128-2137.
11. Waldeck, W., Rosl, R. and Zenfgraf, H. (1984) EMBO J. **3**, 2173-2178.
12. Roberts, J.M. and Weintraub, H. (1986) Cell **46**, 741-752.
13. Lusky, M. and Botchan, M. (1984) Cell **36**, 391-401.
14. Ahola, H., Stenlund, A., Moreno-Lopez, J. and Pettersson, U. (1987) J. Virol. **61**, 2240-2244.
15. Baker, C.C. and Howley, P.M. (1987) EMBO J. **6**, 1027-1035.
16. Schiller, J.T., Vass, W.C., Vousden, K.H. and Lowy, D.R. (1986) J. Virol. **57**, 1-6.
17. Schiller, J.T., Vass, W.C. and Lowy, D.R. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 7880-7884.
18. Lusky, M. and Botchan, M. (1985) J. Virol. **53**, 955-965.
19. Berg, L., Singh, K. and Botchan, M. (1986) Mol. Cell. Biol. **6**, 859-869.

20. Spalholz, B.A., Yang, Y.-C. and Howley, P.M. (1985) Cell **42**, 183-191.
21. Yang, Y.-C., Spalholz, B.A., Rabson, M.S. and Howley, P.M. (1985) Nature **318**, 575-577.
22. Lambert, P.F., Spalholz, B.A. and Howley, P.M. (1987) Cell **50**, 69-78.
23. Doorbar, J., Campbell, D., Grand, R.J.H. and Gallimore, P.H. (1986) EMBO J. **5**, 355-362.
24. Miller, J.H. (1972) Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, N.Y.
25. Lowy, D.R., Rands, E. and Scolnick, E.M. (1978) J. Virol. **26**, 291-298.
26. Maniatis, T., Eritsch, E.F. and Sambrook, J. (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, N.Y.
27. Tabak, H.F. and Flavell, R.A. (1978) Nucl. Acids Res. **5**, 2321-2332.
28. Sanger, F., Nicklen, S. and Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463-5467.
29. Graham, F.L. and Van der Eb, A.J. (1973) Virology **52**, 456-467.
30. Frost, E. and Williams, J. (1978) Virology **91**, 39-50.
31. Gorman, C.M. Moffat, L.F. and Howard, B.H. (1982) Mol. Cell. Biol. **2**, 1044-1051.
32. Hall, C., Jacob, E., Ringold, G. and Lee, F. (1983) J. Molec. Appl. Genet. **2**, 101-109.
33. Matthias, P.D., Bernard, H.U., Scott, A., Brady, G., Hashimoto-Gotoh, T. and Schutz, G. (1983) EMBO J. **2**, 1487-1492.
34. Ogden, J.E., Stanway, C., Kim, S.-Y., Mellor, J., Kingsman, A.J. and Kingsman, S.M. (1986) Mol. Cell. Biol. **6**, 4335-4343.
35. Gorman, C.M., Padmanabhan, R. and Howard, B.H. (1983) Science **221**, 551-553.
36. Laimins, L.A., Khoury, G., Gorman, C., Haward, B. and Gruss, P. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 6453-6457.
37. Twigg, A.J. and Sherratt, D. (1980) Nature **283**, 216-218.
38. Moskaluk, C. and Bastia, D. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 1215-1218.
39. Androphy, E.J., Lowy, D.R. and Schiller, J.T. (1987) Nature **325**, 70-73.
40. McKnight, S. and Tjian, R. (1986) Cell **46**, 795-805.
41. Myers, R.M., Tilly, K. and Maniatis, T. (1986) Science **232**, 613-618.
42. Brady, J., Radonovich, M., Thoren, M., Das, G. and Salzman, N.P. (1984) Mol. Cell. Biol. **4**, 133-141.
43. Maniatis, T., Goodbourn, S. and Fischer, J.A. (1987) Science **236**, 1237-1245.