

## Bovine Papillomavirus Type 1 Encodes Two Forms of a Transcriptional Repressor: Structural and Functional Analysis of New Viral cDNAs

JOONHO CHOE,<sup>†</sup> PETER VAILLANCOURT, ARNE STENLUND,<sup>‡</sup> AND MICHAEL BOTCHAN\*

*Department of Molecular Biology, University of California, Berkeley, California 94720*

Received 23 November 1988/Accepted 6 January 1989

Genetic and biochemical evidence has established that the E2 open reading frame (ORF) of bovine papillomavirus type 1 encodes at least two different site-specific DNA-binding proteins, one which activates and the other which represses expression from a viral promoter (P. F. Lambert, B. A. Spalholz, and P. M. Howley, *Cell* 50:69-78, 1987). We have obtained data which show that a second form of the repressor gene is expressed in transformed cells harboring stable viral plasmids. The structural details of this gene have been discerned by cDNA cloning, by RNase protection, and by primer extension analysis of *in vivo* RNA. Moreover, data from *in vitro* transcription experiments support the notion that this form of the E2 repressor is expressed from a novel viral promoter and that a small exon from another ORF is linked to an active repressor domain in E2. Thus, two different forms of the repressor are expressed from different promoters and might be independently regulated either in the cell cycle or in different tissue types. We show by functional *in vivo* assays utilizing a cDNA vector encoding this gene that the *trans*-acting factor has *in vivo* activities similar to those of the known repressor. Our screen of a cDNA library for cDNA clones representing bovine papillomavirus transcripts has also revealed a number of other novel structures defining new donor and acceptor RNA-processing sites. Notably, clones which conceptually can be translated to yield an E7 protein, the viral M gene, and the entire E2 ORF have been characterized. Finally, truncated versions of putative E8 cDNAs were also obtained.

The life cycles of the DNA viruses which infect animal cells can be broadly placed into two categories. Some DNA viruses infect cells and rapidly enter into productive replication with ensuing cell death, while others establish their genomes as episomes and have a long latent period. (We use the word episome to mean an added genetic element which can reproduce itself in the cell in different ways depending on the nature of the element and the cell type. Thus the word episome can refer to an element acquired by infection which is carried by the cell in an integrated state or as a plasmid. For further details, see reference 20.) This latency even persists through many cell divisions if the infected cells are dividing or are induced to proliferate by virally encoded gene products. The viruses which cause warts are interesting members of this latter class because the viral episomes establish as stable nuclear plasmids and must replicate their genomes in coordination with the cell through a large number (54) of cell divisions. Bovine papillomavirus type 1 (BPV-1) in particular causes massive fibropapillomas soon after infection (25), and in these cells the viral genome is carried as a high-copy-number plasmid. Vegetative and exponential replication of this virus, however, occurs only in terminally differentiated keratinocytes. Thus, a highly regulated process coordinating viral gene expression and cell state must play a major role in the maintenance of latency. Along these lines, it is thought that a major viral promoter ( $P_L$ ) is active only in differentiating keratinocytes (2).

We have been interested in the study of BPV-1 plasmid replication with the view that an understanding of how the

replication system is regulated might give some clues into how transcription and replication processes interplay in eucaryotes. The rodent C127 cells provide what we optimistically believe to be a good model for this latent plasmid replication *in vivo*. Viral infection of the cells leads to oncogenic transformation with single-hit kinetics (11), and the viral genome is carried as a stable multicopy nuclear plasmid (26, 27). The facts that the virus transforms with single-hit kinetics and that the copy number of 100 to 200 is detectable after only a limited number of cell doublings imply that an initial amplification of viral DNA occurs before a stable regulated copy number is achieved. This concept is reinforced by transient replication studies (28) which indicate that viral DNA replicates faster than cellular DNA upon initial entry into the cell. After establishment, the viral copy number stays constant.

The genetic analysis of the BPV-1 genome reveals that a large network of viral factors and promoters must interact to achieve the switch from this initiation mode of replication to the homeostasis apparent in the transformed cells. The somewhat surprising feature of this system seems to be its complexity, as at least six different gene products are included in this process. How these factors interact in a kinetic sense and in maintenance has not been elucidated. The E1 open reading frame (E1 ORF) of BPV-1 is the only ORF within which mutation absolutely abolishes transient replication in C127 cells. Within this ORF, genetic analysis indicates that at least two gene products are required *in trans* for establishment of the stable plasmid state. The R gene encoded in part by the 3' portion of the ORF (4, 29) is absolutely required for transient replication and has been postulated to therefore play a direct role in viral replication (4). In contrast, frameshift mutations in the 5' portion of the E1 ORF do not seem to affect transient replication. These mutations define a gene in another complementation group,

\* Corresponding author.

<sup>†</sup> Present address: Department of Biology, Korea Institute of Technology, Taejon 302-338, Korea.

<sup>‡</sup> Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724.

M. This M gene is required for the maintenance of stable plasmids and has the genetic properties of a repressor of replication. Thus, cells harboring low levels of BPV-1 plasmids are immune to amplification of wild type upon superinfection. Yet M<sup>-</sup> mutants replicate as efficiently in these cells as they do in uninfected cells (4, 29). Furthermore, Roberts and Weintraub (39) have shown that a product(s) of BPV-1 likely to include M can repress in *trans* simian virus 40 runaway amplification, if the BPV-1 origins are linked in *cis* to simian virus 40. These results together imply that M is a repressor of amplification or a positive factor for regulated low-level replication which is antagonistic toward amplification or both. Recently, Thorner et al. (49) have identified the M product as a small phosphoprotein whose size is compatible with its gene being encoded mainly by the 5' portion of the E1 ORF.

To define with more precision the structures of these genes encoded by the E1 ORF, we have sought to obtain cDNA copies of their transcripts. In this report, we describe the structures of several new BPV-1 transcripts as deduced from the sequences of cDNAs and from the direct analysis of *in vivo* viral RNAs. One of these cDNAs is indeed a potential candidate for an M messenger. Surprisingly, however, the most abundant RNAs which contain E1 sequences detected in our analysis initiate from a new BPV-1 promoter within the ORF. The major product of this promoter is a spliced RNA which encodes another form of the BPV-1 transcriptional repressor (23).

## MATERIALS AND METHODS

**Cell culture.** C127 cells (11), ID13 cells (26), and HeLa cells were maintained in Dulbecco modified Eagle medium supplemented with 10% fetal calf serum and penicillin-streptomycin.

**cDNA library.** ID13 cells were treated with cycloheximide (35 µg/ml) prior to being harvested. A cytoplasmic extract was obtained by lysis of cells in the detergent Nonidet P-40 and a subsequent pelleting of nuclei via standard protocols. This extract was then processed via the standard guanidium isothiocyanate-cesium chloride method (30). Poly(A) RNA was prepared by passage of cytoplasmic RNA over oligo(dT)-cellulose twice. A cDNA library was constructed by using poly(A) RNA and the λ<sub>590</sub> vector (9). Ten µg of poly(A)<sup>+</sup> RNA was used to synthesize the first strand of cDNA in a 50-µl mixture of 56 mM Tris hydrochloride (pH 7.5)–85 mM KCl–3 mM MgCl<sub>2</sub>–0.5 mM deoxynucleoside triphosphate–2 mM MeHgOH–30 mM β-mercaptoethanol–5 µg of oligo(dT) (Collaborative Research, Inc.)–30 U of RNasin (Promega Biotec)–1,000 U of murine reverse transcriptase (Bethesda Research Laboratories, Inc.). The solution was incubated for 60 min at 37°C. The mixture was then used to make the second strand of cDNA in a 150-µl volume of 85 mM Tris hydrochloride (pH 7.5)–15 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>–30 mM KCl–0.17 mM deoxynucleoside triphosphate–100 µCi of [ $\alpha$ -<sup>32</sup>P]dATP (Amersham Corp., 400 Ci/mmol)–1 U of RNase H (P-L Biochemicals, Inc.)–4.5 U of DNA polymerase I (New England Biolabs, Inc.). Second-strand synthesis continued for 16 h at 16°C. After elongation of the second strand of cDNA, 1 µl of 100 mM ATP and 1 U of T4 DNA ligase were added and incubated for 15 min at room temperature. The mixture was extracted with phenol:chloroform and double-stranded cDNA was precipitated with ethanol. cDNA was treated with T4 DNA polymerase (New England Biolabs) to make blunt ends. Subsequently, *Hind*III linkers were ligated to double-stranded cDNA, using T4 DNA

ligase. The cDNAs were then completely digested with *Hind*III to remove excess linkers from the ends at the cDNAs, and the cDNAs were separated on Bio-Gel A-50M (Bio-Rad Laboratories). cDNAs larger than 500 base pairs were pooled and ligated with λ<sub>590</sub> arms. After ligated λ DNA was packaged by using Gigapack plus (Stratagene), the cDNA library was amplified, using C600 Hfl host bacteria. The cDNA library was screened by the plaque hybridization method (3), using nick-translated BPV-1-specific probes (see Fig. 1). DNA fragments were sequenced by the dideoxy sequencing method (40) after being subcloned into M13mp18 or mp19 vectors.

**Plasmids.** The plasmid CMV E8/E2 contains a 1.2-kilobase (kb) insert of clone N7-3 subcloned into the *Hind*III site of pON260 (43) in the same transcriptional orientation as the cytomegalovirus (CMV) promoter. pON260 contains a 760-base-pair fragment of the CMV promoter and enhancer and an internal β-galactosidase cassette. pON260 is a derivative of pON240. Oligonucleotide-directed (*in vitro*) mutagenesis was carried out as follows to construct the plasmid CMV E8\*/E2. The *Hind*III fragment (1.2 kb) of clone N7-3 was subcloned into M13mp19. The oligonucleotide (5'-CCGTGT TCTAACGCCCTC-3') was synthesized by a DNA synthesizer (Applied Biosystems, Calif.). By using this oligonucleotide and the method of Nakayama and Eckstein (34), codon 7 of the E8/E2 ORF has been changed from leucine (TTA) to a termination codon (TAA). The mutated *Hind*III fragment in M13mp19 was isolated and inserted into pON260, as was the wild-type cDNA. Plasmid pUC C59 is a derivative of pC59 (51). To create pUC C59, the DNA fragment containing the early promoter and the polyadenylation sequences of simian virus 40 and BPV-1 cDNA was subcloned into the *Hind*III and *Sma*I sites of pUC18. Plasmid pMLBPV100 contains the entire BPV genome linearized at the *Bam*HI site cloned into vector pML and has been described elsewhere (27). Plasmids used as probes for the RNase protection experiments and for *in vitro* transcription are described in the figure legends.

**S1 nuclease and RNase protection analysis of RNA.** For S1 nuclease, RNase protection, and primer extension analysis of *in vivo* RNA, poly(A)-selected cytoplasmic RNA was prepared as described previously (5). Unless otherwise indicated, cells were treated with 30 µg of cycloheximide per ml for 4 h prior to being harvested. S1 nuclease analysis was carried out as described elsewhere (13). Preparation of uniformly labeled SP6 probes and RNase protection analysis were performed as described previously (33), with the following modifications. Hybridizations were carried out in 40 mM PIPES (piperazine-*N,N'*-bis(2-ethanesulfonic acid) (pH 6.7)–0.4 M NaCl–1 mM EDTA in a final volume of 30 µl. Following denaturation for 5 min at 85°C, hybridizations were carried out at 68°C for 3 h. RNase digestions were performed at 25°C for 30 min unless otherwise indicated. All digestion products were fractionated on 5% denaturing polyacrylamide gels. Cordycepin A sequencing ladders were made by preparing uniformly labeled SP6 probes as described previously (33) in the presence of 125 µM dATP.

***In vitro* transcription and primer extension.** Primer extension was performed as described elsewhere (46). Whole-cell extracts of HeLa cells were prepared as described previously (31), and *in vitro* transcription was carried out by the procedure of Dynan and Tjian (12). All reactions contained 25 µl of an optimal amount of HeLa cell extract in TM.1 buffer (50 mM Tris [pH 7.9]–0.1 mM EDTA–12.5 mM MgCl<sub>2</sub>–100 mM KCl–2 mM dithiothreitol–17% glycerol), 2% polyvinyl alcohol, 250 µM each of the four deoxynucle-

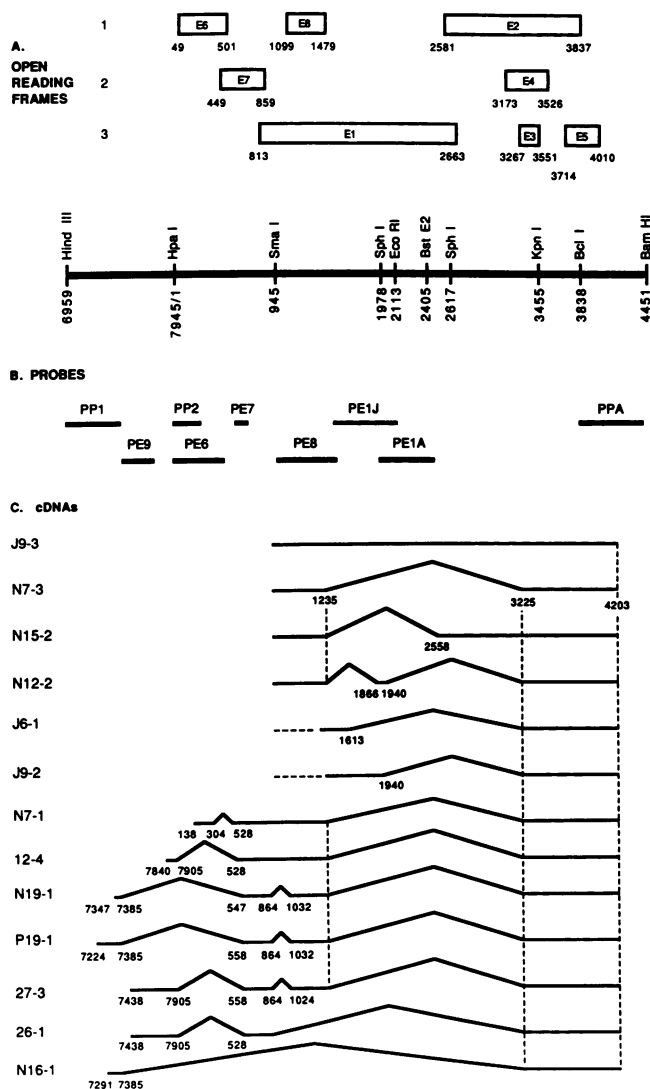


FIG. 1. Structure of BPV-1 early ORFs and cDNAs. (A) Map of the BPV-1 69% (*Hind*III-*Bam*HI) fragment. Early ORFs are represented by open boxes designated E1 to E8 and are aligned according to reading frame. (B) Probes used for screening the cDNA library. The DNA fragments shown were subcloned into pBR322 or pML2, were nick translated, and were used to probe the cDNA library. The nucleotide positions of each of the probes are as follows: PP1, 6958 (*Hind*III) to 7476 (*Cla*I); PE9, 7476 to 7760 (*Rsa*I); PP2, 1 (*Hpa*I) to 240 (*Rsa*I); PE6, 1 to 470 (*Rsa*I); PE7, 697 (*Taq*I) to 767 (*Taq*I); PE-8, 1010 (*Hind*II) to 1515 (*Bgl*II); PE1J, 1515 to 2113 (*Eco*RI); PE1A, 1978 (*Sph*I) to 2450 (*Rsa*I); PPA, 3838 (*Bcl*I) to 4451 (*Bam*HI). (C) Structures of the cDNAs. cDNAs are arranged according to the major groups defined by the hybridization profiles as described in the text. The structures of all cDNAs were completely determined by nucleotide sequencing. Exons (—) and intervening sequences (△) are shown. Numbers beneath the lines indicate the ends and splice junctions of the cDNAs. The top four cDNAs (J9-3, N7-3, N15-2, and N12-2) have 5' ends at nucleotides 1070, 1012, 930, and 948, respectively, and are likely to be transcribed from the P3 promoter. cDNAs J6-1 and J9-2 have 5' ends at nucleotides 1379 and 1410, respectively. As no 5' ends have been mapped in this region thus far, we assume that these are truncated cDNAs representing transcripts which initiate at the P3 promoter. This assumption is depicted in the diagram by the dashed horizontal lines which extend to the P3 cap site. The 5' ends of cDNAs N7-1 through N16-1 are labeled, and their probable promoters are described in the text.

otides, 1.0  $\mu$ g of supercoiled plasmid DNA, and H<sub>2</sub>O to a final volume of 50  $\mu$ l. Where indicated, 4  $\mu$ g of  $\alpha$ -amanitin per ml was added to the reaction mixture prior to the addition of DNA. Reactions were incubated for 30 min at 30°C and were terminated by the addition of 90  $\mu$ l of stop solution containing 1% sodium dodecyl sulfate–20 mM EDTA–200 mM NaCl–250  $\mu$ g of tRNA per ml. Reactions were extracted once with phenol:chloroform, were precipitated with ammonium acetate, and were suspended in hybridization buffer for primer extension reactions. Primer extension products were fractionated on an 8% denaturing polyacrylamide gel.

**CAT (chloramphenicol acetyltransferase) assay.** HeLa cells were transfected by the calcium phosphate method (15) with modifications. Each 60-mm plate of HeLa cells received 10  $\mu$ g with the total amount of DNA adjusted by the addition of carrier DNA. Unless otherwise indicated, cotransfection included 2  $\mu$ g of upstream regulatory region (URR) CAT expression plasmid to provide the CAT activity. The DNA-calcium precipitate was applied to the cells for 12 h followed by 15% glycerol shock for 3 min. Cells were harvested 36 h after transfection. CAT assay was carried out as previously described (14). The amount of all extracts used for the assays was 50  $\mu$ g of protein, and the assays were run for 1 h at 37°C.

**RESULTS**

**Identification of BPV-1 cDNA clones.** A cDNA library was constructed from cytoplasmic poly(A) RNA extracted from virus-infected ID13 cells. Prior to RNA extraction, the cells were treated with cycloheximide to increase our chances of netting viral RNAs with short half-lives (22). cDNAs were synthesized via modifications of the standard protocols (35), using oligo(dT) as a primer for first-strand synthesis, and the library size, which was measured to be  $1.5 \times 10^6$ , was subsequently amplified once (see Materials and Methods).

To screen this library for cDNAs which might represent E1 transcripts, the series of hybridization probes shown in Fig. 1 below a physical map of the early region of BPV-1 were constructed. The location of the probes with respect to the ORFs of the early region are shown and in the legend to Fig. 1B, the precise BPV-1 nucleotide positions are given. Previous RNA mapping experiments indicated that a BPV mRNA exists which contains an exon encoded by a 5' portion of the E1 ORF (47). As outlined in the introduction, we were interested in obtaining cDNAs which might encode for the M gene and thus RNAs with coding sequences from this 5' E1 region were sought. Toward this end, we used the PE8 probe in a primary screen of the library. Approximately 0.01 to 0.05% of the lambda plaques hybridized to this probe. The abundance of such inserts was approximately two- to threefold lower than those frequencies estimated with the PE6 probe.

Forty-five PE8-positive plaques were picked, and the inserts were then analyzed in detail. The first characterization was to hybridize the cDNAs with each of the probes shown in Fig. 1B. The inserts were thus classified into major groups according to their hybridization profiles. The structures of nine representatives of each class of cDNA characterized by this screen are shown in Fig. 1C. They are named clones J9-3 through N19-1; clones P19-1 through N16-1 were obtained by another screen described below. To analyze the structures of these clones in detail, the inserts were subcloned into M13 and pBR322 vectors. Restriction analyses and Southern blots were used to add information about the various families. The DNA sequences of the representative members shown in Fig. 1 were completely determined.

The most abundant such family of cDNAs obtained by this primary screen is represented by clone N7-3 (Fig. 1C). Twenty-four such clones were obtained. These cDNAs hybridized only to the PE8 and PPA probes. Clones of the N7-3 family had 5' ends clustered between BPV-1 nucleotides 900 to 1000. These clones terminated, as did all others in our study, at a 3' position of 4203 followed by a poly(A) tail. cDNA clones with the hybridization profiles of N7-1 were obtained eight times while those of the J9-3 family were obtained seven times. All others were obtained as unique members.

The cDNAs that are believed to be generated by common promoters are grouped together in Fig. 1. As discussed below, we believe it is likely that abundant clones J9-3 and N7-3, as well as the relatively rarer clones N15-2 and N12-2, represent RNAs produced from a previously unreported BPV-1 promoter near the 5' end of the E1 ORF. The 5' ends of these clones are not labeled in the figure but are given in the legend. The abundant cDNA N7-1 probably represents RNAs initiating from the previously characterized P2 (P<sub>89</sub>) promoter, as it has a 5' end only 49 nucleotides downstream from this cap site. Two of the clones obtained via this primary screen contained BPV-1 sequences with exons from the URR; 12-4 and N19-1. N19-1 is 161 nucleotides short of the previously characterized P<sub>7186</sub> (P1) promoter, and 12-4 may represent a severely truncated transcript from this P1 promoter or an as-yet-unidentified promoter in the URR.

To obtain other clones containing sequences from the URR, two separate primary screens of the library were done, using the PP1 and PE9 probes. With both probes, two million bacteriophage were independently screened. P19-1 was obtained (Fig. 1C), which has a structure very similar to that of N19-1, although a different acceptor links the position 7385 URR donor to the downstream coding sequences. Three other clones, 27-3 through N16-1, were also characterized. Two clones, 27-3 and 26-1, both have 5' ends at precisely the same nucleotides. This position 7438 is at the end of a T8 stretch in the coding strand and may represent a hot spot for cleavage with RNase H in the preparation for second-strand cDNA synthesis, thus creating artificially the chance for common 5' ends in a family of cDNAs.

We will discuss in another section of this article the potential coding capacities of the various cDNAs described here. However, at this point it is interesting to point out that with the possible exception of cDNA clone J9-3, no cDNAs specifically isolated could encode for the R gene and only this clone contained a nucleotide sequence between positions 1940 and 2558. This is particularly intriguing as mutational data mentioned above shows that mutations at BPV-1 position 2113 inactivate transient as well as stable plasmid replication. To directly screen for cDNAs which might contain sequences from this region, we probed the cDNA library with the probe PE1J (Fig. 1B). From a screen of more than  $5 \times 10^6$  phage, no BPV-1 cDNAs were found that did not fall into the families defined by J9-3, J6-1, or J9-2.

To summarize this structural information briefly, a large number of new donor and acceptor sites within BPV-1 have been established by these cDNAs, and the sequences found at these sites are shown in Fig. 2. Initially, we did not expect to find such a large number of different types of cDNAs which spanned the 5' end of the E1 ORF. However, recently published Northern (RNA) blot analysis of BPV-1 RNAs with probes similar to those used here (6, 46) substantiate this complexity in that a large number of transcripts ranging from 1.2 to 2 kb in length were detected. Clearly, more detailed genetic analysis will be necessary to correlate the

| POSITION | DONOR                           | POSITION | ACCEPTOR                     |
|----------|---------------------------------|----------|------------------------------|
| 304      | <b>GTAA</b> CAG <b>GTG</b> AGGA | 528      | TTCA <b>CTGC</b> AG GACCGTT  |
| 864      | <b>GATA</b> AG <b>GTAC</b> CA   | 547      | CTGATTT <b>TAA</b> G TCCATGT |
| 1235     | <b>CTCT</b> CAG <b>GTAC</b> AGG | 558      | TCAT <b>GTCC</b> AG GCACCC   |
| 1613     | <b>CTTAAA</b> <b>GTAA</b> AG    | 1024     | CAAT <b>CA</b> TCAG TCTTCA   |
| 1940     | <b>AAC</b> TATG <b>GTAA</b> GC  | 1032     | CAGTCTT <b>TC</b> AG GGAATC  |
| 7385     | <b>ATGCGG</b> <b>GTAA</b> GT    | 1866     | TTT <b>GGCTC</b> AG GATCTCA  |
| 7905     | <b>GAA</b> CCG <b>GTAA</b> GT   | 2558     | TTTT <b>TGTAA</b> G GTTATGG  |
|          |                                 | 3225     | TCTGATTT <b>TA</b> G AGATCC  |

FIG. 2. Positions and surrounding DNA sequences of splice junctions deduced from the viral cDNAs sequenced in this study. Consensus sequences are in boldface.

physical RNA map with functional genes. The majority of cDNAs found use a common splice donor in the E1 ORF at nucleotide 1235. The position of this donor was first established by electron microscope analysis of BPV-1 RNAs (47). However, the complexity of the genetic organization of the virus in this region is illustrated by the findings reported here that this donor can join exons to either the most commonly used acceptor in the virus at position 3225 (2, 51) or to acceptors at nucleotide 1866 or 2558.

**Identification of a promoter within the E1 ORF.** The abundant cDNAs described in the previous section all have their 5' ends near the beginning of the E1 ORF. Nevertheless, it is impossible to prove via cDNA cloning where the bona fide 5' ends of the actual RNAs are by this sort of analysis. Even with model systems such as globin 9S mRNA, methods which enrich for full-length clones generally yield cDNA which are truncated at their 5' ends (36). We had anticipated that the RNAs represented by the cDNA clone N7-3 would be initiated by a promoter close to the beginning of the E1 ORF. Alternately, these RNAs could be initiated upstream and a small exon generated by splicing might have been lost in cDNA synthesis. This assumption was based on the facts that the cDNA inserts in the library were large and that sequence analysis of cDNAs from known abundant BPV-1 RNAs terminated close to known start sites (data not shown). Thus, to determine where the likely initiation sites are, it was necessary to analyze *in vivo* RNA by a variety of nuclease protections and primer extension methods.

A series of probes either 5' end labeled or uniformly labeled were used, and their structures are displayed in Fig. 3. First, a DNA fragment from nucleotides 576 to 1010 was 5' end labeled on the noncoding strand and was annealed to cytoplasmic poly(A) RNA from ID13 or C127 cells and was digested with S1 nuclease (Fig. 3). Two protected fragments are clearly seen in the lane marked ID13 and are missing in the C127 control panel; they are labeled P3a and P3b. From the lengths of these fragments measured relative to the sequence ladder and size markers run in parallel, the 5' ends map to BPV-1 positions 896 and 886. The BPV-1 sequence at these positions does not show splice acceptor sites, so we tentatively assumed that these positions demarked bona fide 5' ends. To substantiate this conclusion, synthetic primers extending from BPV-1 nucleotides 925 to 945 or from 1038 to 1058 (Fig. 4) were synthesized, 5' end labeled, and annealed to ID13 cytoplasmic poly(A)-selected RNA, and primer extension analysis was performed. The extension products were fractionated by gel electrophoresis and separated alongside a sequence ladder of BPV DNA obtained by

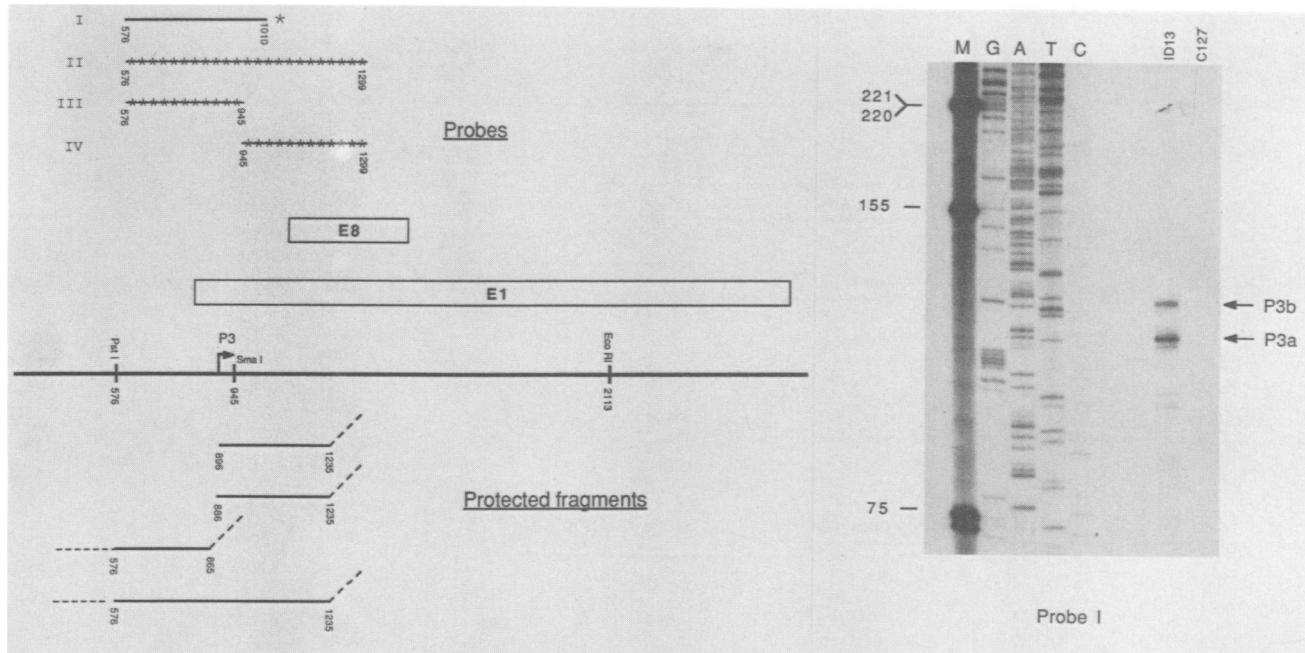


FIG. 3. S1 nuclease mapping of P3 cap sites. Poly(A) cytoplasmic RNA from  $2 \times 10^7$  ID13 cells or C127 cells was hybridized to the 5'-end-labeled noncoding strand of a DNA fragment from positions 576 to 1010 (probe I) and digested with S1 nuclease. The hybrids were fractionated on a 5% denaturing polyacrylamide gel. Lanes: M, 5'-end-labeled fragments from a *Bst*NI-*Eco*RI digest of pBR322; G, A, T, and C, size markers generated by dideoxy sequencing of a single-stranded M13 bacteriophage template. Size markers (in nucleotides) are shown at the left.

dideoxy sequencing, using this same primer annealed to BPV DNA. The lane marked *in vivo* in the autoradiogram shown in Fig. 4 maps two discrete extension products to nucleotides 886 and 896, thus corroborating the S1 analysis.

To further extend the notion that a stable 5' end maps to these nucleotides, a plasmid containing the BPV-1 *Pst*I fragment from nucleotides 576 to 1299 was assayed *in vitro* for promoter activity in whole-cell extracts of HeLa cells in the presence or absence of  $4 \mu\text{g}$  of  $\alpha$ -amanitin per ml. This is a level of drug which inhibits RNA polymerase II transcription *in vitro*. The *in vitro* products were analyzed via primer extension analysis as described for the *in vivo* RNA. The *in vitro* products produced by this fragment have the same 5' ends as do the *in vivo* RNA, and their synthesis is sensitive to the RNA polymerase II inhibitor (Fig. 4). We will hereafter refer to this region of DNA as the P3 promoter, using the convention from our laboratory first described for the P1 promoter (46). We use this convention rather than naming the promoter after putative cap sites (2) for two reasons. First, promoters usually refer to *cis*-acting regulatory sequences and we assume that the starts at nucleotides 886 and 896 share common *cis* control elements. Second, for very heterogeneous ends (e.g., those around nucleotide 2440), it is extremely cumbersome to label each nucleotide as a promoter. The disadvantage, of course, is that new promoters which may be found would require different names since the order from P1 would be shifted.

To determine the structure of the exons of the mRNAs initiating at the P3 promoter, a series of RNase protection experiments were performed, using the uniformly labeled cRNA probes depicted in Fig. 3. Cytoplasmic poly(A) RNAs extracted from ID13 or C127 cells were hybridized to an SP6 polymerase-generated probe which contained BPV-1 sequences from nucleotides 576 to 1299 (probe II in Fig. 3).

After treatment with RNase A and RNase T<sub>1</sub>, the protected fragments were separated by electrophoresis. The autoradiogram of this gel is shown in Fig. 5A; lane C is the C127 control, and lanes A and B are identical experiments except for the RNase digestion conditions. By the mobilities of the fragments we could tentatively ascribe each band to a particular transcript. The model for these assignments is shown diagrammatically in Fig. 3. The topmost protected band is attributed to full-length probe protection and represents unspliced (within the probe region) RNAs; this band is unlabeled. The next band down represents RNAs initiating upstream of the 576 position and splicing at 1235; RNAs such as cDNA clone N7-1 and 12-4 would generate such bands. The two bands labeled P3b and P3a correspond to fragments produced by transcripts initiating at nucleotides 886 and 896 and splicing at nucleotide 1235; specifically, the length of these bands corresponds to 338 and 348 nucleotides, the length from the P3 promoter to the donor. The most frequent cDNA clone that we obtained in our screen N7-3 would generate such fragments, and these results corroborate those findings. The band labeled 576-865 is ascribed to all RNAs initiating upstream of 576 (e.g., E6 and E6-7 messengers) which use the 865 donor (47, 51). To test these assignments, probe II was halved to generate SP6 probes III and IV. Each of these probes was annealed to cellular RNAs and was analyzed. The predicted fragment lengths were detected with each probe (Fig. 5B). Specifically, the protected fragment from probe IV (nucleotides 945 to 1299) represents the cumulative signal of all exons which cross nucleotide 945 and use the 1235 splice donor (i.e., P3a and P3b as well as N7-1 and 12-4 transcripts). Probe III detects the RNAs which use the 865 donor.

From the intensities of these bands in Fig. 5, it would seem that the ratios of the P3 transcripts to transcripts initiating

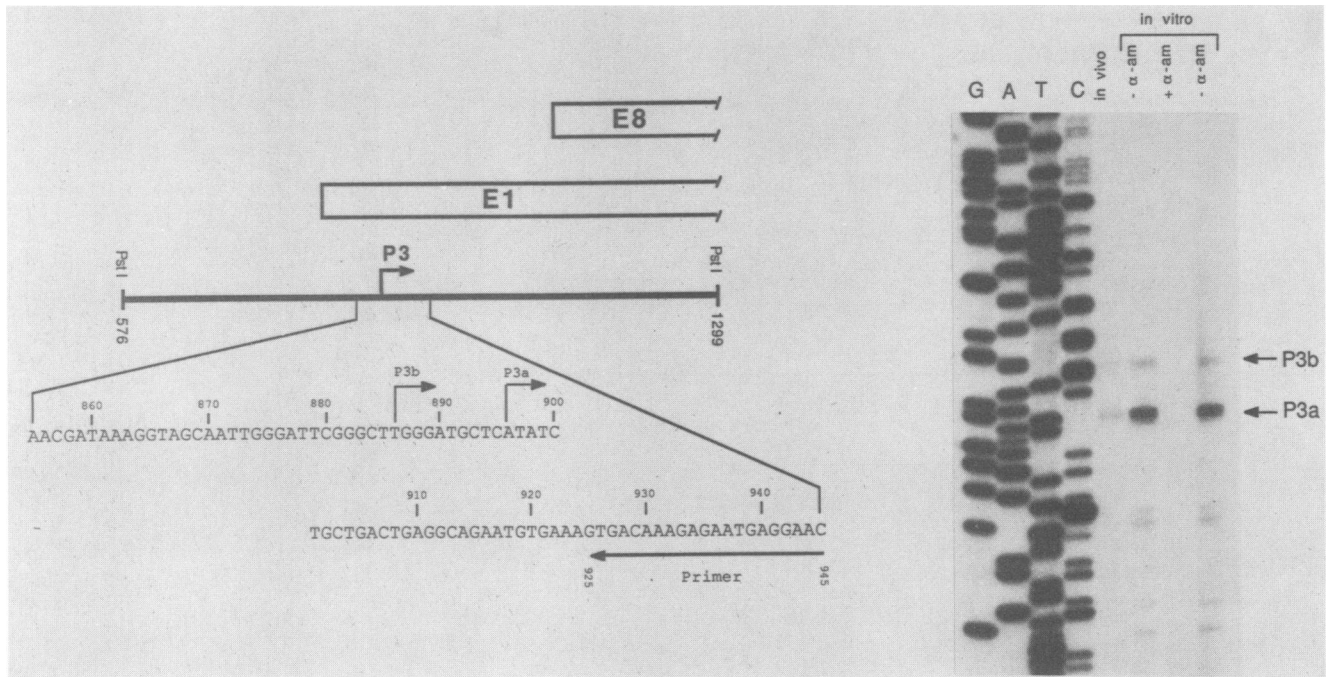


FIG. 4. Primer extension analysis of in vivo- and in vitro-transcribed RNA from the P3 promoter. The BPV *PstI* fragment from positions 576 to 1299, cloned into the *PstI* site of pSP65, was tested in an in vitro transcription assay using Manley extracts of HeLa cells in the presence (+  $\alpha$ -am) or absence (- $\alpha$ -am) of 4  $\mu$ g of  $\alpha$ -amanitin per ml (see Materials and Methods), followed by primer extension using a 5'-end-labeled oligonucleotide containing a noncoding-strand sequence from nucleotides 1038 to 1058. (Two primers were used in this sort of experiment, one from nucleotides 925 to 945 and one from nucleotides 1038 to 1058. The results from the latter are shown.) Primer extension was also performed on poly(A) cytoplasmic RNA from  $2 \times 10^7$  ID13 cells. Lanes G, A, T, and C contain dideoxy sequencing ladders using both the primer and template used for the in vitro transcription experiment. Extension products were fractionated on an 8% denaturing polyacrylamide gel and are labeled as shown on the right.

upstream (such as those at P2) is roughly 1:5. (For example, compare the intensities of bands labeled P3a and P3b to the band at 576-865 in Fig. 5A.) This number agrees reasonably with the results of the cDNA cloning experiments described above. Furthermore, a significant number of transcripts probably initiate at P2 and generate RNAs which use the 1235 donor (e.g., N7-1). This can be estimated from the band labeled 576-1235 in Fig. 5A.

As the RNAs used as templates for cDNA synthesis were prepared from cycloheximide-treated cells, it was of interest to address the question of whether drug treatment changes qualitatively the RNA profile in this region. The results (Fig. 5C) show that no qualitative differences can be discerned, although there is a large increase in signal specifically from the P3 promoter with cycloheximide treatment. (Note that 10-fold more RNA was used in the cycloheximide-minus experiment.)

**Clone N7-3 encodes a repressor of BPV gene expression.** Conceptual translation of the sequences encoded by clone N7-3 predicts a 24,000-dalton protein. This putative polypeptide links 11 amino acids from the E8 ORF to the C-terminal domain of the BPV-1 E2 ORF. Interestingly, it has recently been shown (32) that the carboxy-terminal domain of the E2 protein contains sequences both necessary and sufficient for site-specific binding to BPV-1 DNA (17). Further heightening our interest in the functional significance of clones such as N7-3 and the P3 promoter in general was the work of Lambert et al. (23, 24). In their original work defining functionally a BPV-1 transcriptional repressor, Lambert et al. (23) assayed the repressor by placing a surrogate promoter upstream of BPV-1 sequences included within the

clone pCW1-28 which contains BPV sequences from position 845 continuously to 4203 (Fig. 1). On the basis of the available genetic evidence at that time (deletion and termination mutants within pCW1-28), they tentatively concluded that the enhancer in the surrogate promoter was stimulating transcription from a distal internal promoter which initiated starts more than 2 kb away. The promoter was positioned around BPV nucleotide 3080. Indeed, surrogate promoters placed just upstream of 3080 did produce vectors which functionally produced a repressor. Supporting evidence that a truncated form of E2 was actually expressed by virus-infected cells was found in the literature. RNA mapping experiments which showed that 5' ends could be mapped to nucleotide 3080 (47) and the cloning of a cDNA which could represent such RNAs, p1153, produced by Baker and Howley (2) from infected bovine fibropapillomas, made it seem likely that such a repressor protein species probably exists. However, certain deletions upstream of the putative start site at nucleotide 3080 in the promoted pCW1-28 clone eliminated the ability of this vector to express the repressor (24). These latter mutants could have deleted an upstream exon of the repressor gene encoded by pCW1-28 or a promoter element for the putative p3080 promoter. From these findings and our results, we suspected that two forms of the repressor gene are expressed: one initiated from the P3 promoter, and the other initiated from within the E2 ORF. Accordingly, the surrogate enhancers in pCW1-28 were probably initiating transcripts from P3.

To test this possibility directly, we placed the N7-3 cDNA clone downstream of the CMV promoter to create a putative repressor expression vector called CMV E8/E2. As a con-

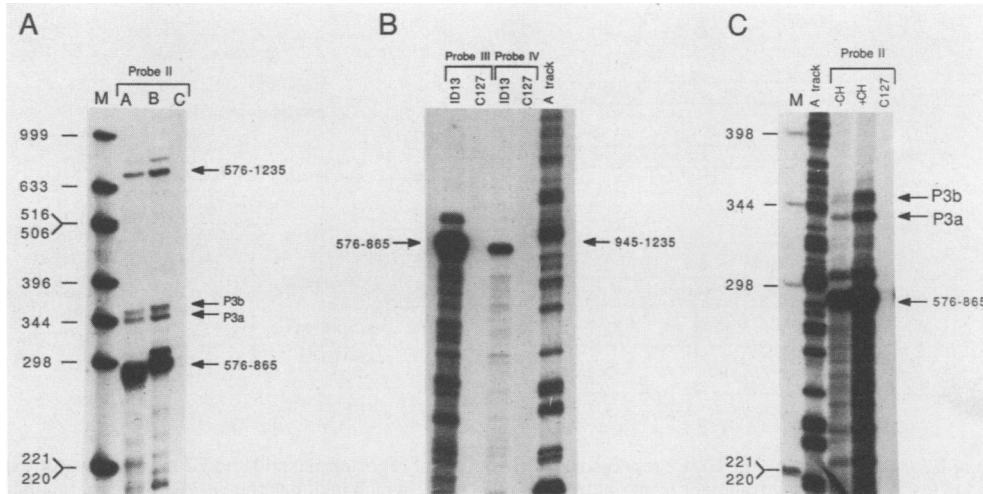


FIG. 5. RNase protection mapping of P3 transcripts. Poly(A) cytoplasmic RNA from  $2 \times 10^7$  ID13 cells or C127 cells was hybridized to uniformly labeled antisense SP6 probes containing sequences depicted in Fig. 3 as probes II, III, or IV and was digested with RNase A and RNase T<sub>1</sub>. Hybrids were fractionated on 5% denaturing polyacrylamide gels. (A) RNase digestion of hybrids made between probe II SP6 RNA and ID13 RNA (lanes A and B) or C127 RNA (lane C) was carried out at 10°C (lanes B and C) or 25°C (lane A). End-labeled DNA markers (lane M) were prepared as described in the legend to Fig. 3. (B) Protected fragments derived from RNase digestion of hybrids formed between probe III or probe IV SP6 RNA and ID13 or C127 RNA, as shown (arrows). Cordycepin A sequencing ladder using as template the BPV *SmaI-PstI* fragment cloned into the *SmaI* and *PstI* sites in the polylinker of pSP65 (probe IV) is shown (lane A track). (C) RNA from  $2 \times 10^7$  ID13 or C127 cells which were treated with 30  $\mu$ g of cycloheximide per ml for 4 h prior to harvest (lanes +CH and C127, respectively) or from  $2 \times 10^8$  ID13 cells which were not cycloheximide (lane -CH) treated were hybridized to probe II SP6 RNA and were digested.

tol, to assess both the *trans* effects of the repressor vector due to its potential protein production, as opposed to promoter competition effects, and to test the notion that the specific protein is an E8/E2 fusion, the clone CMV E8\*/E2 was constructed (Fig. 6). By oligonucleotide-directed mutagenesis, the leucine codon (TTA) at position 1223 was changed to a termination codon (TAA). This single point mutation eliminates only the coding potential of all E8 proteins in the vector and leaves intact possible expression from other frames. (The transversion introduces a silent mutation in the E1 ORF).

We first asked if the CMV E8/E2 vector could repress the transactivation of the E2 product upon the inducible enhancer in the BPV-1 control region (23, 24, 49). Figure 7A shows the reporter plasmid, URR CAT, which provides the *cis*-binding sites for the E2 products, and the plasmid pUC C59, which provides the transactivator for the reporter. pUC C59 is a derivative of the C59 E2 vector kindly provided by P. M. Howley (51). The URR CAT construct can produce the protein CAT, and its activity can be quantitatively assayed as described elsewhere (14). All transfections were carried out by using HeLa cells as recipients, and the results of the CAT assays are shown as raw data in Fig. 7B and

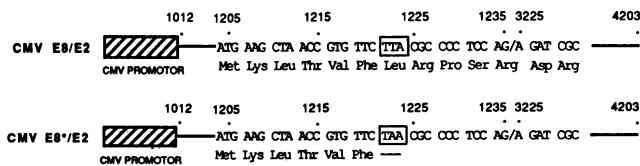


FIG. 6. Structure of plasmids CMV E8/E2 and CMV E8\*/E2. The CMV promoter-enhancer (▨) and sequences representing the cDNA sequences cloned into the vector with *HindIII* linkers (—) are shown. Critical BPV nucleotide and amino acid sequences are given, and the codon which has been altered by site-directed mutagenesis (□) is shown.

quantitatively in Fig. 7C. The table shows the percentage of acetylated chloramphenicol for each particular experiment. For each datum point shown, 2  $\mu$ g of CAT reporter DNA was either transfected alone or with combinations of the various *trans* effector constructs. Interestingly, we found that the CMV E8/E2 vector represses the basal activity of the reporter as shown at the top of Fig. 7B. This effect is measured to be at least fivefold (Fig. 7C, 6.7% conversion in contrast to 1.3% conversion) and clearly is not due to artifactual plasmid competition since the point mutant construct CMV E8\*/E2 has no effect on the reporter. The next three experiments (Fig. 7B) repeat the classic transactivation experiments which show that even at very low doses the E2 vector can evoke a large stimulation of the ability of the BPV URR to promote gene expression (16, 17, 45, 48). In our hands in this particular system, the stimulation is about 10-fold. The results shown illustrate that the E2 stimulation saturates at levels below or at 0.1  $\mu$ g of transactivator vector. We next mixed together reporter CAT vector, saturating levels of transactivator E2 (0.1  $\mu$ g), and increasing levels of the test plasmid CMV E8/E2. As shown, the CMV E8/E2 construct counteracts the *trans*-positive effects of E2 and lowers the basal activity of the reporter. Again, the CMV E8\*/E2 vector has no effect in this type of triple mixing experiment (e.g., BPV URR CAT, pUC C59, and CMV E8\*/E2), illustrating that the negative effects of the repressor vector must be due to the *trans* action of its product.

Overexpression of a protein which can counter the effects of the E2 protein would be expected to interfere with BPV-1 focus-forming capacity (23). We tested this hypothesis by mixing together increasing concentrations of either the CMV E8/E2 or CMV E8\*/E2 vector with a constant amount of wild-type BPV-1 DNA. In our assays, 1  $\mu$ g of BPV DNA induced about 175 foci in a 60-mm culture dish. The repressor vector did indeed inhibit BPV-1 foci in a dose-dependent

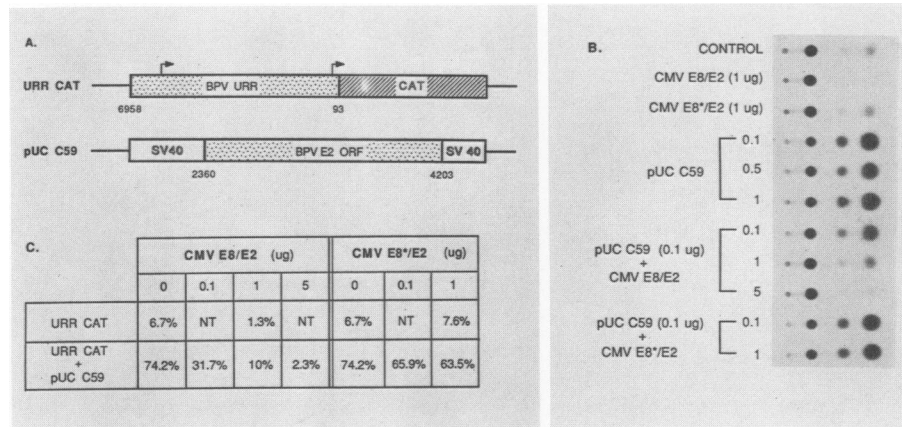


FIG. 7. Inhibition of E2 *trans* activity by the E8/E2 repressor. (A) Structure of reporter plasmid and E2 *trans*-activator expression plasmid. The reporter plasmid URR CAT contains a fragment from nucleotides 6958 to 93, which includes the entire upstream regulatory region as well as the P2 promoter at nucleotide 89, linked to the CAT gene (14). The positions of the P1 and P2 promoters are indicated by arrows. pUC C59 is a derivative of C59 (51; see Materials and Methods). (B) Autoradiogram of CAT assays prepared from extracts of HeLa cells which were transfected with 2  $\mu$ g of URR CAT alone (control) or together with the indicated amounts of expression plasmid. (C) Quantitation of results of CAT assay shown in panel B. Values are expressed as a percent conversion of [<sup>14</sup>C]chloramphenicol to the acetylated forms. NT, Not tested.

way, while the mutant vector had no measurable effects in *trans* (Fig. 8).

## DISCUSSION

BPV-1 has a relatively small genome, yet a large portion of the sequence has not been found expressed as mRNA. Most conspicuously missing from the maps of RNAs produced in either productively infected fibropapillomas or latently infected cultured cells are mRNAs from parts of the highly conserved E1 ORF and from the URR. Together these two regions constitute more than one-third of the 8-kb genome. As *trans*-acting gene products have been defined genetically for the E1 ORF, it is generally assumed that the reason for this lack of detection is the low abundance of these transcripts. Using specific probes from both the E1 ORF and the URR, we have been able to isolate and characterize a variety of new cDNAs and define a new transcription unit in the virus.

The main points of this article concern the functional and structural analysis of the transcripts from the P3 promoter. The major transcript from this promoter encodes a repressor, and from the results presented above we estimate that it is as abundant as the individual transcripts from the upstream promoter P2. This estimate is based on both the frequency of clones such as N7-3 relative to E6 clones in the library and on the direct RNA mapping data shown in Fig. 5. Before we discuss specific results relevant to the rarer cDNAs, two caveats should be mentioned. First, we used cyclohexamide prior to harvesting cellular cytoplasmic RNA to increase our chances of obtaining mRNAs which might be degraded rapidly or expressed rarely. The block to protein synthesis which slows down mRNA degradation increases BPV-1 transcript abundance but potentially augments very weak promoters (e.g., the P1 promoter) by lowering repressor levels. Thus, we may have obtained cDNAs in some cases which are extremely rare in stably transformed cells. We do not, however, know of a situation wherein the drug creates new promoters or causes aberrant splicing. A second limitation concerns the promoter assignments for the rare cDNAs. As the cDNAs were in general truncated by 50 to

300 base pairs, it was relatively straightforward for cDNAs such as N7-3 to be assigned to promoter P3, for we obtained many cDNAs of this type, all of which had 5' ends close to the mapped P3 end. Furthermore, we could map an exon of expected size with RNase protection experiments by using RNAs harvested from non-drug-treated cells. However, for cDNAs obtained only once (e.g., 26-1 and 27-3), the promoter assignment is clearly speculative and is simply based on our current knowledge of BPV-1 promoters. Thus in all cases we ascribed the nearest 5' promoter to a given cDNA.

The P1 promoter is weakly expressed from stable plasmids in ID13 cells (2, 46), and several of the cDNAs characterized here are likely to be expressed from this promoter (Fig. 1; see Fig. 9). In particular, P19-1 can be conceptually translated to yield an E7 protein. This 102-amino-acid peptide would utilize an ATG codon just two codons upstream of the nucleotide 7385 donor sequence. Given the central role that E7 plays in human papillomaviruses (38, 42; Banks and Crawford, personal communication), it is possible that such a protein might be involved in some stage of the BPV-1 life cycle. Recently we have found that certain BPV-1 mutations in the P1 promoter reduce transformation frequencies of the DNAs (G. L. Bream and M. Botchan, unpublished results) and that these mutants can be complemented by the expression of the P19-1 cDNA. Whether this effect is due to only the direct effect of E7 or to the effect of the genes on copy number stability is not known. The cDNA 27-3, which uses the same E7 acceptor site as P19-1, cannot encode such a protein since in-frame terminators exist 5' to the E7 ORF in this cDNA and no in-frame ATGs are present in E7. The cDNA clone 27-3, however, has the capacity for a small 100-amino-acid 5' E1 protein, since this cDNA contains the conserved E1 ATG at position 838. This putative polypeptide is not likely to be the M protein, as mutations within the intron (864 to 1024) affect the M complementation group. Neither N19-1 nor P19-1 can code for this putative E1 gene since the splice acceptor at nucleotide 1032 (as opposed to the acceptor at nucleotide 1024 in 27-3) leads to codon usage out of the E1 ORF.



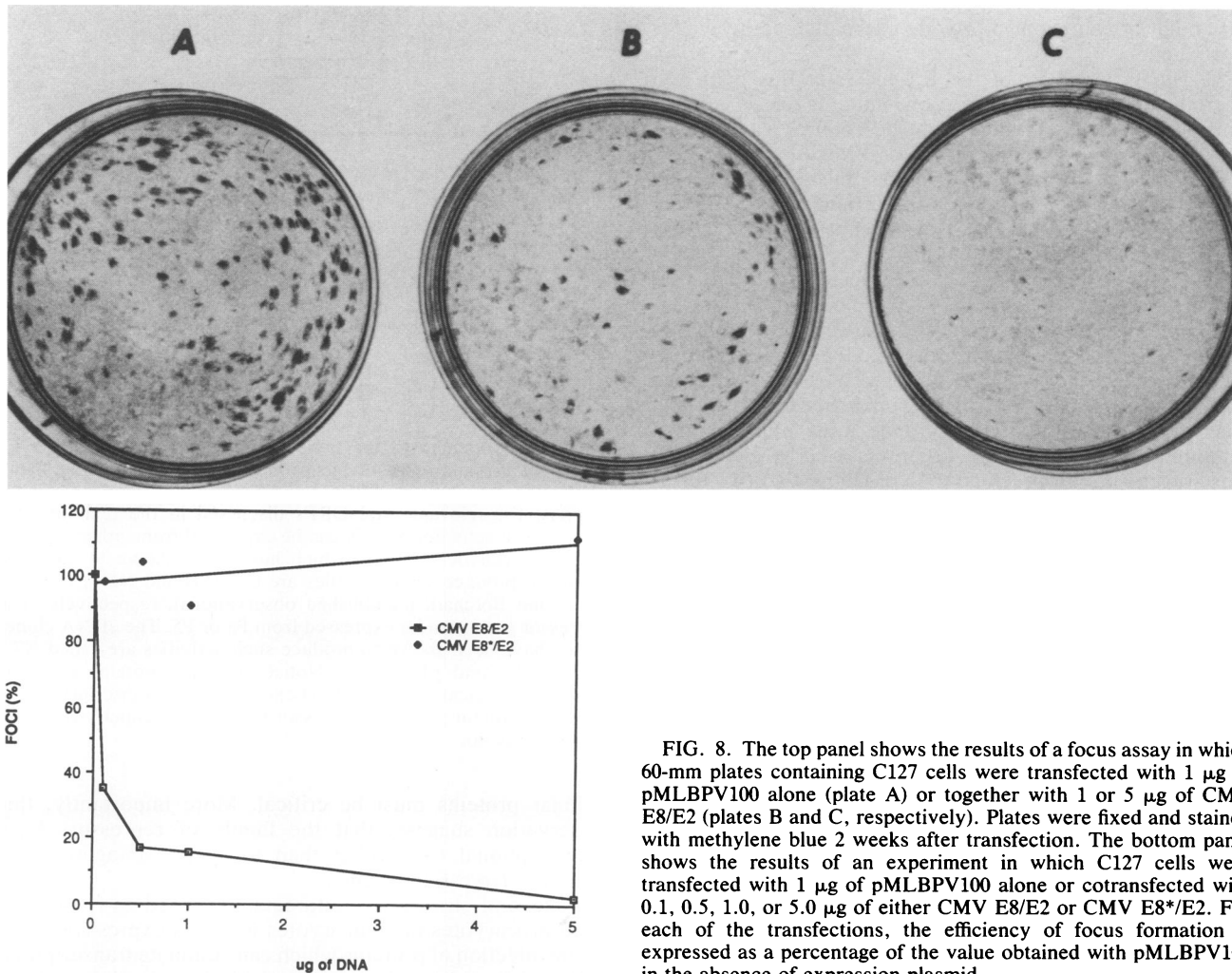


FIG. 8. The top panel shows the results of a focus assay in which 60-mm plates containing C127 cells were transfected with 1  $\mu$ g of pMLBPV100 alone (plate A) or together with 1 or 5  $\mu$ g of CMV E8/E2 (plates B and C, respectively). Plates were fixed and stained with methylene blue 2 weeks after transfection. The bottom panel shows the results of an experiment in which C127 cells were transfected with 1  $\mu$ g of pMLBPV100 alone or cotransfected with 0.1, 0.5, 1.0, or 5.0  $\mu$ g of either CMV E8/E2 or CMV E8\*/E2. For each of the transfections, the efficiency of focus formation is expressed as a percentage of the value obtained with pMLBPV100 in the absence of expression plasmid.

Our characterization of several new donors and acceptors within the E1 region raises the possibility that other genes might overlap E1. In particular, recent genetic studies suggest that the E8 ORF encodes a *trans*-acting gene product (M. Lusky and J. Choe et al., manuscript in preparation). Point mutations in the E8 ORF either 5' or 3' to the splice donor at 1235 lead to BPV DNAs which integrate in focus assays. If one assumes that both J6-1 and J9-2 are truncated cDNAs whose RNAs initiated from P3, either of these products could code for an E8 protein. Indeed, these E8 mutations can be complemented by a vector which carries a surrogate promoter linked to clone J6-1 (Fig. 1) if the E8 ATG start region is reconstructed.

One of the major objectives of our cDNA screen was to find clones which could represent putative mRNAs for the M gene product. We had anticipated that either a small exon would be spliced in frame into the E1 ORF or that a promoter would be found just 5' to the ORF. This is because the limited genetic studies on the M gene define coding sequences up to the 5' end of the ORF (29). Chow et al. (8) have characterized an abundant 5' end at an analogous position in human papillomavirus types 6 and 11 RNAs extracted from infected tissues. One of the RNA structures defined by Chow et al. (8) which has this 5' end position (their species b) has an exon over the E1 ORF region which overlaps with the M-coding sequences in BPV-1 and resem-

bles the BPV-1 species (2b) characterized by Stenlund et al. (47). While our data cannot exclude the existence of such an analogous promoter at the end of the E1 ORF, our data do show that such a message must be at least 10- to 20-fold less abundant than other RNA species which can be detected from this region and can code for M. (See, for example, Fig. 5A; such putative structures with ends just at the end of the E1 ORF would produce bands just above the P3a and P3b exon bands.) The abundant cDNA family represented by clone N7-1 can code for the M gene if one allows for a single RNA species to have the coding potential for two separate proteins. Indeed, there is ample precedent for such potential polycistronic RNA in the papillomavirus literature. For example, the RNA likely to encode the E2 protein in the cottontail rabbit papillomavirus system has the E6 ORF 5' to the transactivator sequences (50; M. S. Barbosa and F. O. Wettstein, personal communication). Recently, Thorner et al. (49) have identified the BPV-1 M protein, using serological techniques. Various animal cell expression vectors which have the N7-1 cDNA inserted can indeed transiently produce a nuclear phosphoprotein which comigrates with protein detected in virus-infected cells (L. Thorner and M. Botchan, unpublished results). Thus, we conclude that at least one way to make M would be to initiate from P2 and to generate an RNA with the structure of N7-1. Other less abundant mRNAs such as those represented by cDNA clone

12-4 could, of course, also provide alternative pathways for M production.

The genetic analysis of the R gene is at this point too uncertain to make clear predictions about a role in replication for a P3 mRNA which might be colinear with the E1 ORF, such as one that might be represented by clone J9-3. On the one hand, we point out that all published mutations in the E1 ORF between the donor site at nucleotide 1235 and the *Bg/III* site at nucleotide 1515 affect not only E1 but also E8. Thus, the possibility exists that a species such as J9-3 could encode for an R factor, if one speculates that it is E8 ORF functions that are affected by these mutants. However, on the other hand, recent findings suggest that stop codons in this region of E1 that do not affect E8 codons lead to mutants that are carried only as integrated genomes in transformed cells (M. Lentz and M. Botchan, unpublished observations). This would suggest that the R gene must utilize a start methionine in the E1 ORF 5' to any carried by clone J9-3. These results do not preclude the possibility that two forms of the R message exist, one required for stable plasmid maintenance and represented by clone J9-3, and another expressed transiently upon initial infection. Thus, while the genomic organization of the M gene seems to be clarified substantially by the present data, the hunt for the structures of the R message(s) continues.

Direct RNA mapping experiments and in vitro transcription data presented above establish the existence of the P3 promoter. The major mRNA product of this promoter expressed in cultured mouse cells is likely to be represented by cDNA clone N7-3. About one-half of the cDNA clones isolated from our primary screen of the library with 5' E1 probes yielded structures essentially identical to N7-3. The abundance of this particular 5' exon in total BPV RNA is substantiated by the SP6 protection experiment shown in Fig. 5. As indicated in a summary diagram in Fig. 9, the product of mRNAs represented by N7-3 is an E8/E2 fusion protein and has been shown here to be a repressor of the transactivator functions of the E2 proteins. Recently Chin et al. (7) have provided conclusive evidence for a similar gene encoding repressor activity in human papillomavirus type 11.

The basic assay that we relied on was to show that the E8/E2 gene could counteract the positive effects of the E2 gene product in expression experiments. Lambert et al. (23) have discussed various models which one can envision to account for such an antagonism of functions, given that both forms of the repressor share along with the transactivator a common carboxy-terminal DNA-binding domain. Thus, if one postulates that the amino terminus of E2 contains an essential domain for transcriptional activation, the expression of the carboxy-terminal domain alone might directly create a competitive situation for binding to DNA sites or might create poisoned protein aggregates with the positive effector which would inhibit its activity. Interestingly, we show here that the E8/E2 gene product can act directly to negatively regulate the basal gene expression from the URR CAT construct. Thus, the repressor proteins are likely to interfere directly with the transcription apparatus, perhaps by making abortive interactions with other site-specific DNA-binding proteins or by simply blocking contacts between the polymerizing apparatus and the DNA (see, for example, Keleher et al. [21]). This later finding regarding the effects of the E8/E2 gene on basal level expression does not preclude the models suggested by Lambert et al. (24) vis à vis the antagonism between E2 and the repressor but suggests, as one would certainly predict, that interactions with

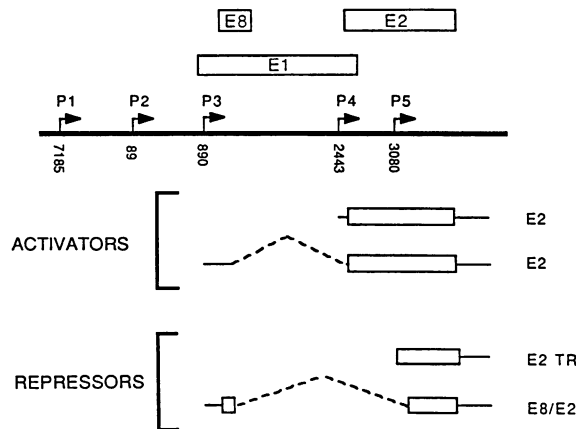


FIG. 9. Summary of the transcription units expressing the E2 transactivator and repressor genes in BPV-1. The top of the figure shows the early region of BPV-1 and demarks the known promoter regions; the relevant viral ORFs discussed in the text are also shown. The activator protein can be expressed from either P4 or P3 and the "generic" cDNAs which have been shown to have the ability to produce such activities are C59 (51) and N15-2 (Vaillancourt and Botchan, unpublished observations), respectively. The repressor genes can be expressed from P3 or P5. The cDNA clones which have been shown to produce such activities are called N7-3 (this study) and p1153 (23). Notice that the proteins encoding sequences (indicated by the open boxes above the transcript) predict different structures for the repressor forms and identical structures for the activator.

cellular proteins must be critical. More importantly, this observation suggests that the family of repressors have transcriptional roles other than to simply antagonize the action of large E2 proteins.

A general strategy of autoregulation used by viruses in both procaryotes and eucaryotes involves expression early in the infection of proteins which can stimulate transcription, followed by expression of proteins which modulate down this activation. This strategy is even used in acutely lytic systems such as are found in adenoviruses and simian virus 40 polyomaviruses, and it makes perfect sense that viruses with long latencies, such as BPV-1, should have such systems. If one assumes that a single bovine wart with a mass of 1.5 kg is the result of one viral infection, this leads to the calculation that the virus must maintain its latency in a transformed cell through at least 43 cell doublings. A complicated circuitry of factors and *cis*-regulatory sites might be presupposed to exist to ensure continued growth of such a massive structure. Thus, the problem of maintaining homeostasis with regard to cell division is a much more important problem in vivo than is usually addressed in focus formation and expansion to a cell line in culture in which routinely only 20 to 25 cell divisions occur before analysis. We can formally suggest two discrete but not exclusive reasons for why the two different forms of the repressor gene are arranged the way they are in BPV-1. On the one hand, the proteins predicted by their gene structures are different and thus they may have different functions. Both proteins have a common 204-amino-acid carboxy terminus and diverge at an amino-terminal domain defined by the acceptor site at BPV-1 nucleotide 3225 (Fig. 9). The amino terminus of the E8/E2 protein has 11 amino acids from an upstream exon while E2TR has an additional 35 amino acids which are encoded by a continuation of the E2 ORF 5' to the acceptor. Another difference between the proteins may involve their

putative interactions with cellular factors mediated by their amino-terminal domains. We might point out that overexpression of either of the gene products for repressor function through surrogate promoters may obscure differences in function between the two proteins. For example, one protein may have higher affinities for a particular binding site than does another, and overexpression might cancel such differences. These points focus on structural differences in the proteins. Alternatively, as the repressor genes are linked to separate promoters, they are perhaps independently regulated. Thus, one or the other gene may function as the important negative regulator in a particular cell type (or time in the cell cycle). Thus, much more work must be focused on how these promoters for the repressors are expressed and where these proteins are expressed in situ, and more specific assays (e.g., with different promoters) may be called for to distinguish the putative differences between the genes. Interestingly, the recent work of Hubbert et al. (19) shows via serological methods that both forms of the repressor protein coexist in unsynchronized cultures of virus-transformed cells. These workers have shown that the E2TR (or long form of the repressor protein) is more abundant than is the E8/E2 form in the C127 cell system. Finally, it is possible that the putative different functions of the genes alluded to above cannot be addressed in short-term cultures of heterologous mouse cells. In this system, either repressor may suffice to maintain appropriate levels of gene expression.

We were surprised to find cDNA clone N15-2 because it encodes for a function which is antagonistic to the function of the E8/E2 repressor gene regulated by the action of the P3 promoter. As indicated in Fig. 9, the potential promoter indicated for clone N15-2 is the P3 promoter; this clone can code for the entire E2 ORF since the splice acceptor at nucleotide 2558 is just upstream of the likely ATG initiator codon in E2. Moreover, recent unpublished work with N15-2 shows that it can functionally make an E2 transactivator when linked to surrogate promoters (P. Vaillancourt, J. Choe, and M. Botchan, manuscript in preparation). However, by far the bulk of the RNAs initiated from P3 must splice from the common donor at nucleotide 1235 to the acceptor at 3225 rather than to the acceptor at nucleotide 2558. Thus, splicing as well as promoter utilization may contribute to the balances between positive and negative factors. Similarly, the vast majority of starts from P4 splice to yield a putative E5 message (1, 18, 47, 51), and it is questionable whether the bulk of the E2 protein is indeed expressed from this promoter. Stenlund et al. (47) characterized a BPV RNA (their species 3) which links the internal E6 donor to an acceptor just upstream of the E2 ORF, and we may speculate, on the basis of the relative abundance of this transcript, that the P2 promoter may be the strongest contributor to the production of the transactivator E2. Again, however, we would like to emphasize the extremely low abundance of RNAs such as N15-2 and point out that genetic studies assessing the role of the acceptor at nucleotide 2558 will be needed to begin to test critically the complexities implied by Fig. 9. In summary, then, the structurally different transcriptional repressor proteins seem to be expressed from two different promoters while the putatively unique activator may be expressed from three different promoters. While on the surface it may seem bewilderingly complex to have so many different ways to express similar activities, given the complexity of the life cycle of the virus mentioned above and the range of cell types it can infect, this is perhaps expected. Recent findings on the expression patterns of a single cellular locus which

codes for a family of transcription factors shows a similar level of complexity (41).

If the plethora of promoters and *trans* factors discussed above does not as yet describe a transcriptional regulatory system for the BPV-1 genome, some of the complexities concerning BPV-1 E2 genetics may be clarified by the discovery of the E8/E2 gene. Lusky and Botchan (28) showed that an E2 mutant, *dl211*, was severely crippled in a transient replication assay. Moreover, in side-by-side assays, the mutant BAL15 replicated transiently as efficiently as did wild-type BPV-1 even though it is deleted for all of the E2 ORF (Lusky and Botchan [29]). It seems likely to us that *dl211*, although destroying the positive transactivation function of the large E2 transactivator and deleting the promoter for the negative repressor E2TR, left intact the E8/E2 gene. Thus, expression of only the negative regulator in the absence of a positive activator may shut the replication genes off too severely. This hypothesis would therefore include the speculation that in any short-term assay, the complete absence of both negative and positive genes might be roughly equivalent to that of the wild type if the initial expression of the gene(s) being assayed is E2 independent and only becomes E2 regulated at a later stage of infection. Similar ideas may be relevant to the genetic data from other groups. For example, DiMaio and Settleman (10) have defined a *ts* mutant in the E2 ORF wherein the lesion was engineered 5' to the nucleotide 3225 acceptor in the E2 ORF. This temperature-sensitive mutation potentially inactivates both the E2 transactivator and the E2TR repressor, yet it leaves intact the E8/E2 gene. This situation may accentuate some of the phenotypes they measured; specifically, it would increase the apparent requirement for the positive E2 product itself. Finally, Hermonat and Howley (18) showed that certain E2 mutants sustained in the ORF 3' to the acceptor at nucleotide 3225 could not be complemented by the C59 E2 expression vector (Fig. 9). C59 cannot provide an E8/E2 product and thus might not be expected to complement any E2 mutants which affect this member of the E2 family. Furthermore, it is conceivable that the C59 cDNA may not provide E2 itself in precisely the right way. Perhaps E2 provided by P2 or P3 is needed to achieve most-efficient virus transformation and plasmid establishment.

#### ACKNOWLEDGMENTS

This research was supported by Public Health Service grants CA 42414 and CA 30490 from the National Institutes of Health. J.C. was supported by a postdoctoral fellowship from the Cancer Research Coordinating Committee of the University of California.

We thank Virginia Staubach-Summa for typing the manuscript, P. M. Howley for providing the E2 expression clone C59, and E. S. Mocarski for the pON260 vector.

#### LITERATURE CITED

- Ahola, H., A. Stenlund, J. Moreno-López, and U. Pettersson. 1987. Promoters and processing sites within the transforming region of bovine papillomavirus type 1. *J. Virol.* **61**:2240-2244.
- Baker, C. C., and P. M. Howley. 1987. Differential promoter utilization by the bovine papillomavirus in transformed cells and in productively infected wart tissues. *EMBO J.* **6**:1027-1035.
- Benton, W. D., and R. W. Davis. 1977. Screening  $\lambda$ gt recombinant clones by hybridization to single plaques in situ. *Science* **196**:180-182.
- Berg, L., M. Lusky, A. Stenlund, and M. R. Botchan. 1986. Repression of BPV replication is mediated by a virally-encoded *trans*-acting factor. *Cell* **46**:753-762.
- Brawerman, G., J. Mendecki, and S. Y. Lee. 1972. A procedure for the isolation of mammalian messenger ribonucleic acid.

- Biochemistry 11:637-641.
6. Burnett, S., J. Moreno-López, and U. Pettersson. 1987. Messenger RNAs from the E1 region of bovine papilloma virus type 1 detected in virus-infected bovine cells. *Nucleic Acids Res.* 15:8607-8620.
  7. Chin, M. T., R. Hirochika, H. Hirochika, T. R. Broker, and L. T. Chow. 1988. Regulation of human papillomavirus type 11 enhancer and E6 promoter by activating and repressing proteins from the E2 open reading frame: functional and biochemical studies. *J. Virol.* 62:2994-3002.
  8. Chow, L. T., M. Nasser, S. M. Wolinsky, and T. R. Broker. 1987. Human papillomavirus types 6 and 11 mRNAs from genital condylomata acuminata. *J. Virol.* 61:2581-2588.
  9. Davis, R. W., D. Botstein, and J. R. Roth. (ed.). 1980. *Advanced bacterial genetics*, p. 245. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
  10. DiMaio, D., and J. Settleman. 1988. Bovine papillomavirus mutant temperature sensitive for transformation, replication, and transactivation. *EMBO J.* 7:1197-1204.
  11. Dvoretzky, I., R. Shober, S. K. Chattopadhyay, and D. Lowy. 1980. A quantitative (*in vitro*) focus assay for bovine papillomavirus. *Virology* 103:369-375.
  12. Dynan, W. S., and R. Tjian. 1983. Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. *Cell* 32:669-680.
  13. Favalaro, J., R. Treisman, and R. Kamen. 1980. Transcription maps of polyoma virus-specific RNA: analysis by two-dimensional nuclease S1 gel mapping. *Methods Enzymol.* 65:718-749.
  14. Gorman, C. M., L. F. Moffat, and B. H. Howard. 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* 2:1044-1051.
  15. Graham, F. L., and A. J. van der Eb. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* 52:456-467.
  16. Haugen, T. H., T. P. Cripe, G. D. Ginder, M. Karin, and L. P. Turek. 1987. (*Trans*)activation of an upstream early gene promoter of bovine papillomavirus-1 by a product of the viral E2 gene. *EMBO J.* 6:145-152.
  17. Hawley-Nelson, P., E. J. Androphy, D. R. Lowy, and J. T. Schiller. 1988. The specific DNA recognition sequence of the bovine papillomavirus E2 protein is an E2-dependent enhancer. *EMBO J.* 7:525-531.
  18. Hermonat, P. L., and P. M. Howley. 1987. Mutational analysis of the 3' open reading frames and the splice junction at nucleotide 3225 of bovine papillomavirus type 1. *J. Virol.* 61:3889-3895.
  19. Hubbert, N. L., J. T. Schiller, D. R. Lowy, and E. J. Androphy. 1988. Bovine papillomavirus-transformed cells contain multiple E2 proteins. *Proc. Natl. Acad. Sci. USA* 85:5864-5868.
  20. Jacob, F., and E. L. Wollman. 1961. *Sexuality and the genetics of bacteria*. Academic Press, Inc., New York.
  21. Keleher, C. A., C. Goutte, and A. D. Johnson. 1988. The yeast cell-type-specific repressor  $\alpha 2$  acts cooperatively with a non-cell-type-specific protein. *Cell* 53:927-936.
  22. Kleiner, E., W. Dietrich, and H. Pfister. 1986. Differential regulation of papillomavirus early gene expression in transformed fibroblasts and carcinoma cell lines. *EMBO J.* 5:1945-1950.
  23. Lambert, P. F., B. A. Spalholz, and P. M. Howley. 1987. A transcriptional repressor encoded by BPV-1 shares a common carboxy-terminal domain with the E2 transactivator. *Cell* 50:69-78.
  24. Lambert, P. F., B. A. Spalholz, and P. M. Howley. 1987. Evidence that bovine papillomavirus type 1 may encode a negative transcriptional regulatory factor, p. 15-22. *In* B. M. Steinberg, J. L. Brandsman, and L. B. Taichman (ed.), *Cancer cells 5: papillomaviruses*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
  25. Lancaster, W. D., and C. Olson. 1982. Animal papillomaviruses. *Microbiol. Rev.* 46:191-207.
  26. Law, M.-F., D. R. Lowy, J. Dvoretzky, and P. M. Howley. 1981. Mouse cells transformed by bovine papilloma virus contain only extrachromosomal viral DNA sequences. *Proc. Natl. Acad. Sci. USA* 78:2727-2731.
  27. Lusky, M., and M. R. Botchan. 1984. Characterization of the bovine papillomavirus plasmid maintenance sequences. *Cell* 36:391-401.
  28. Lusky, M., and M. R. Botchan. 1986. Transient replication of bovine papillomavirus type 1 plasmids: *cis* and *trans* requirements. *Proc. Natl. Acad. Sci. USA* 83:3609-3613.
  29. Lusky, M., and M. R. Botchan. 1986. A modulator function for BPV replication is required for the establishment of the viral plasmid state. *J. Virol.* 60:729-742.
  30. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: a laboratory manual*, p. 191-198. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
  31. Manley, J. L., A. Fire, A. Cano, P. A. Sharp, and M. L. Gelfer. 1980. DNA-dependent transcription of adenovirus genes in a soluble whole cell extract. *Proc. Natl. Acad. Sci. USA* 77:3855-3859.
  32. McBride, A., R. Schlegel, and P. M. Howley. 1988. The carboxy-terminal domain shared by the bovine papillomavirus E2 trans-actin and repressor proteins contains a specific DNA binding activity. *EMBO J.* 7:533-539.
  33. Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green. 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* 12:7035-7056.
  34. Nakayama, K., and F. Eckstein. 1986. Inhibition of restriction endonuclease NciI cleavage by phosphorothioate groups and its application to oligonucleotide directed mutagenesis. *Nucleic Acids Res.* 14:9679-9698.
  35. Nathans, J., and D. Hogness. 1983. Isolation, sequence analysis, and intron-exon arrangement of the gene encoding bovine rhodopsin. *Cell* 34:807-814.
  36. Okayama, H., and P. Berg. 1982. High-efficiency cloning of full-length cDNA. *Mol. Cell. Biol.* 2:161-170.
  37. Pfister, H. 1987. Papillomaviruses: general description, taxonomy, and classification, p. 1-38. *In* N. P. Salzman and P. M. Howley (ed.), *The papovaviridae*, vol. 2. The papillomaviruses. Plenum Publishing Corp., New York.
  38. Phelps, W. C., C. L. Yee, K. Munger, and P. M. Howley. 1988. The human papillomavirus type 16 E7 gene encodes transactivation and transformation functions similar to those of adenovirus E1A. *Cell* 53:539-547.
  39. Roberts, J. M., and H. Weintraub. 1986. Negative control of DNA replication in composite SV40-bovine papillomavirus plasmids. *Cell* 46:741-752.
  40. Sanger, F., and A. R. Coulson. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* 94:441-448.
  41. Santoro, C., N. Mermod, P. C. Andrews, and R. Tjian. 1988. A family of human CCAAT-box-binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. *Nature (London)* 334:218-224.
  42. Smotkin, D., and F. O. Wettstein. 1986. Transcription of human papillomavirus type 15 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein. *Proc. Natl. Acad. Sci. USA* 83:4680-4684.
  43. Spaete, R. R., and E. S. Mocarski. 1985. Regulation of cytomegalovirus gene expression:  $\alpha$  and  $\beta$  promoters are *trans* activated by viral functions in permissive human fibroblasts. *J. Virol.* 56:135-143.
  44. Spalholz, B. A., P. F. Lambert, C. Yee, and P. M. Howley. 1987. Bovine papillomavirus transcriptional regulation: localization of the E2-responsive elements of the long control region. *J. Virol.* 61:2128-2137.
  45. Spalholz, B. A., Y.-C. Yang, and P. M. Howley. 1985. Transactivation of a bovine papilloma virus transcriptional regulatory element by the E2 gene product. *Cell* 42:183-191.
  46. Stenlund, A., G. L. Bream, and M. R. Botchan. 1987. A promoter with an internal regulatory domain is part of the origin of replication in BPV-1. *Science* 236:1666-1671.
  47. Stenlund, A., J. Zabielski, H. Ahola, J. Moreno-Lopez, and U. Pettersson. 1985. Messenger RNAs from the transforming region of bovine papilloma virus type 1. *J. Mol. Biol.* 182:541-554.

48. **Thiery, F., and M. Yan.** 1987. The BPV1-E2 *trans*-acting protein can be either an activator or a repressor of the HPV18 regulatory region. *EMBO J.* **6**:3391–3397.
49. **Thorner, L., N. Bucay, J. Choe, and M. R. Botchan.** 1988. The product of the bovine papillomavirus type 1 modulator gene (M) is a phosphoprotein. *J. Virol.* **62**:2474–2482.
50. **Wettstein, F. O., M. S. Barbosa, and M. Nasseri.** 1987. Identification of the major cottontail rabbit papillomavirus late RNA cap site and mapping and quantitation of an E2 and minor E6 coding mRNA in papillomas and carcinomas. *Virology* **159**: 321–328.
51. **Yang, Y.-C., H. Okayama, and P. M. Howley.** 1985. Bovine papillomavirus contains multiple transforming genes. *Mol. Cell. Biol.* **2**:1030–1034.