JOURNAL OF VIROLOGY, July 1987, p. 2240–2244
0022-538X/87/072240-05$02.00/0

Vol. 61, No. 7

Promoters and Processing Sites within the Transforming Region of Bovine Papillomavirus Type 1

HARRI AHOLA,[1,2] ARNE STENLUND[2]† JORGE MORENO-LÓPEZ,[1] AND ULF PETTERSSON[2]*

Department of Veterinary Microbiology, Section of Virology,[1] and Department of Medical Genetics, Uppsala University,[2] Biomedical Center, S-75123 Uppsala, Sweden

Received 12 January 1987/Accepted 31 March 1987

The mRNAs present in bovine papillomavirus type 1 (BPV-1)-transformed C127 cells were studied by primer extension. The results show that two internal promoters are present in the E region of BPV-1 in addition to the previously identified promoter at coordinate 1 (H. Ahola, A. Stenlund, J. Moreno-López, and U. Pettersson, Nucleic Acids Res. 11:2639–2650, 1983). One, located at coordinate 31, generated a set of mRNAs with heterogeneous 5′ ends, which may encode the major transforming protein of BPV-1, the E5 protein. The second promoter, which is located at coordinate 39, generates colinear mRNAs which encode either the E4 protein or a truncated form of the E2 protein. Unlike the cottontail rabbit papillomavirus (O. Danos, E. Georges, G. Orth, and M. Yaniv, J. Virol. 53:735–741, 1985), BPV-1 appears to lack a separate promoter for expression of the E7 protein. The major splice sites in the transforming region (E region) of the BPV-1 genome were also identified by nucleotide sequence analysis.

A large number of papillomaviruses have been isolated from warts in humans and various other animal species, which appear to have many properties in common, including capsid morphology and genome organization (for a review, see reference 12). Certain types of papillomaviruses transform C127 and NIH 3T3 mouse cells in vitro (9, 16). Most of our knowledge about papillomavirus replication and gene expression stems from studies of bovine papillomavirus type 1 (BPV-1), a virus which has become the prototype for the papillomavirus family. A subgenomic fragment constituting 69% of the BPV-1 genome, extending between coordinates 87 and 56, is sufficient for transformation of mouse C127 cells (9). An interesting feature of the BPV-1 genome is that it replicates as a multicopy plasmid in transformed mouse and hamster cells (8, 19). The molecular structure of mRNAs expressed in transformed cells has been studied by many methods including cDNA cloning (20) and electron microscopy and S1 nuclease mapping (17). These studies identified several viral mRNA species, many of which were initiated at a promoter located near coordinate 1 (The BPV-1 genome is commonly subdivided into 100 map units; coordinate 0 is located at the single cleavage site for endonuclease *Hpa*I.) This promoter, which includes TATA and CAAT boxes, is followed by a cap site at nucleotide (nt) 89 (coordinate 1.1) (1). However, several other mRNA species were identified which might be initiated internally in the transforming region (17, 20). In the present study we used primer extension to identify two additional promoters in the transforming region (E region) of BPV-1. Our results also confirm previously identified locations of several splice donor and acceptor sites inside the E region.

MATERIALS AND METHODS

Cells and virus. BPV-1-transformed C127 cells (1a, 17) were propagated in Dulbecco modified medium with 10% newborn calf serum. The cells were treated with 15 μg of cycloheximide per ml 2 h before RNA extraction. Cytoplasmic RNA was isolated by the method of Brawermann et al. (3). All RNA preparations were fractionated by oligo(dT)-cellulose chromatography.

Oligonucleotides. Oligonucleotides were kindly supplied by H. Hultberg, KabiGen AB.

cDNA synthesis. Avian myeloblastosis virus polymerase, purchased from Life Sciences, Inc., was used in all reactions. The protocol of Virtanen and Pettersson (18) was followed. The amount of RNA used in each reaction was 200 to 400 μg together with 10 pmol of primer.

Hybridization selection. The extension products were hybridized to nitrocellulose filters containing 80 μg of cloned immobilized BPV-1 DNA. Hybridization was done at 65°C overnight in 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.5% sodium dodecyl sulfate–6× Denhardt solution. The filters were washed four times for 15 min each at 50°C in 0.1× SSC. The hybridized cDNA was eluted by heating the filters to 100°C for 3 min in 1 ml of 10 mM Tris hydrochloride (pH 7.5). Material from three consecutive elutions was pooled for sequence analysis.

Sequence analysis. cDNA selected by hybridization to BPV-1 DNA was fractionated in 6 or 10% polyacrylamide gels. Bands identified by autoradiography were eluted by isotachophoresis (11) and sequenced by the procedure of Maxam and Gilbert (10).

RESULTS

Analysis of viral mRNAs by primer extension. Previous studies have shown that several BPV-1-specific mRNAs expressed in transformed C127 cells contain a common 3′ exon located between nt 3224 and 4203 (Fig. 1). To study the exon structure of these mRNAs, primer extension experiments were performed. A primer derived from the BPV-1 sequence between nt 3271 and 3290 was used in the first experiment. Poly(A)-selected mRNA from BPV-1-transformed cells treated with 15 μg of cycloheximide per ml before extraction (7) was used as a template in the reverse transcriptase reactions. The extension products were purified by hybridization to filters containing cloned BPV-1 DNA, and the eluted cDNA was fractionated on a 6% polyacrylamide gel together with size markers. A large

* Corresponding author.
† Present address: Department of Molecular Biology, University of California, Berkeley, CA 94720.

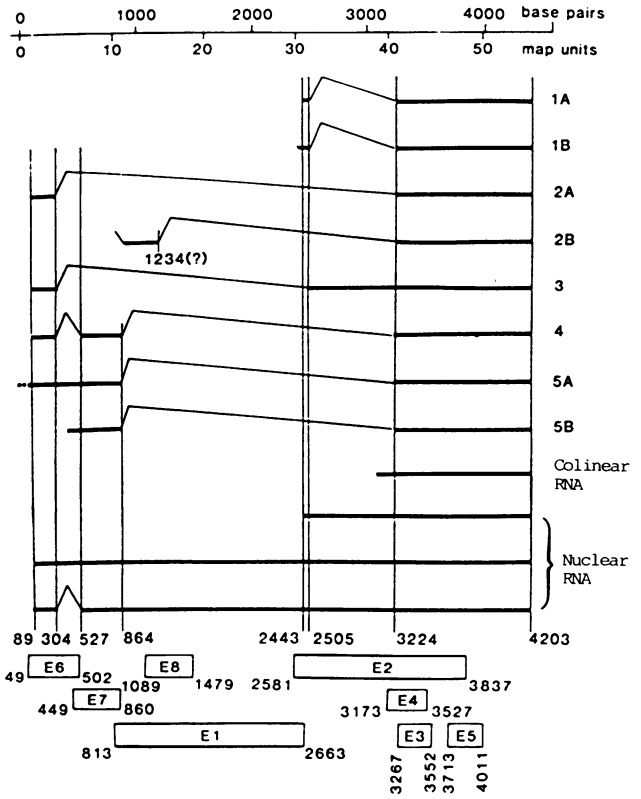

FIG. 1. Schematic drawing of mRNAs expressed in BPV-1-transformed C127 cells (17, 20). ORFs present in the transforming region of BPV-1 are also shown.

number of bands were observed, and those designated *a*, *b*, *c*, and *d* in Fig. 2, lane A, were eluted and sequenced. The extension product designated *a* had an approximate size of 130 nt, although it was very heterogeneous. Sequence analysis showed that these cDNAs contained a short 5′ exon whose 5′ end mapped around nt 2440. The sequence also revealed a splice junction which joined nt 3225 and 2505. Because extension product *a* contained several cDNA species (Fig. 2, lane A), it was subdivided into three portions which were eluted and sequenced separately. The results show that the cDNA species had identical exon structures and only differed in length at the 5′ end (see below). The results suggest that extension product *a* corresponds to mRNA species 1A (Fig. 1). The 5′ end of this mRNA is apparently heterogeneous, mapping in a region centered around nt 2440.

The nt sequence of extension product *b*, which was approximately 210 nt long, showed that it corresponds to a colinear mRNA whose structure is described below.

The nt sequence of extension product *c*, which was approximately 280 nt long, showed that it contained a 5′ leader transcribed from the promoter region at coordinate 1. A splice junction which joined nt 304 and 3225 was also identified in the sequence. No further splices were identified, and based on the size of product *c*, it was estimated that its 5′ end coincides with the previously identified cap site at nt 89. Judged from the structure and size of extension product *c*, it corresponds to mRNA 2A (Fig. 1). The nt sequence of extension product *d*, which was approximately 600 nt long, revealed a splice junction between nt 864 and 3225. Because of the length of the cDNA, it was impossible to read the sequence all the way to the 5′ end. However, judged from its size and the identified splice junction, it corresponds to type 4 mRNA (Fig. 1).

Most of the other extension products which are visible in Fig. 2, lane A, were also subjected to sequence analysis. However, the amount of labeled cDNA recovered was insufficient for sequence determinations to be made.

Identification of a cap site at coordinate 31. Previous studies have identified a prominent mRNA species in BPV-1-transformed C127 cells, designated type 1A by Stenlund et al. (17). This mRNA consists of two exons, one of which is located between nt 3224 and 4203 (Fig. 1). The second exon is only about 60 nt long and maps around nt 2500 (17, 20). Based on previous studies, it was impossible to conclude whether the short 5′ exon contained the capped 5′ end or was linked to another exon. To address this question, an oligonucleotide derived from the sequence between nt 2484 and 2503 was used for primer extension. Extension products

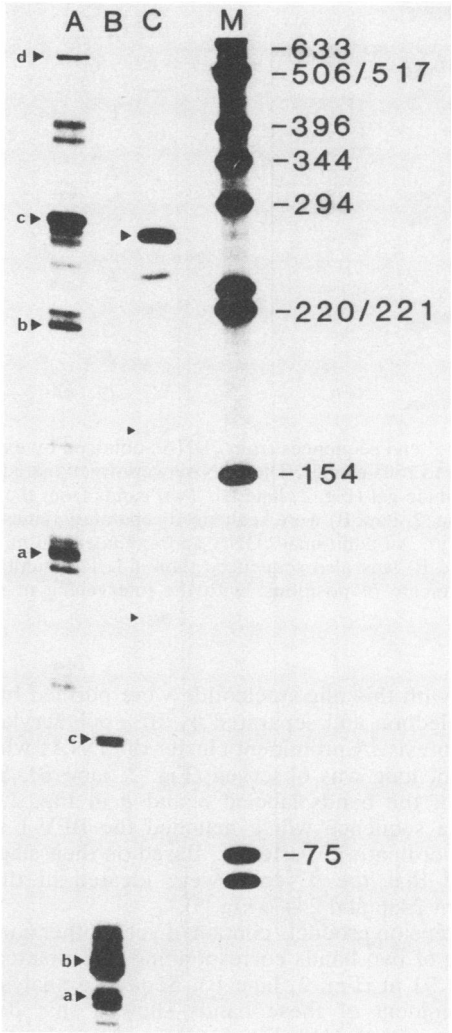

FIG. 2. Analysis by polyacrylamide gel electrophoresis of cDNAs obtained by primer extension. cDNAs were obtained after extension of the nt 3271 to 3290 primer (lane A), the nt 2484 to 2503 primer (lane B), and the nt 563 to 582 primer (lane C). Lane M, ^{32}P-labeled marker fragments. The cDNAs which were subjected to sequence analysis are indicated by arrowheads. The numbers in the margin indicate lengths in nucleotides.

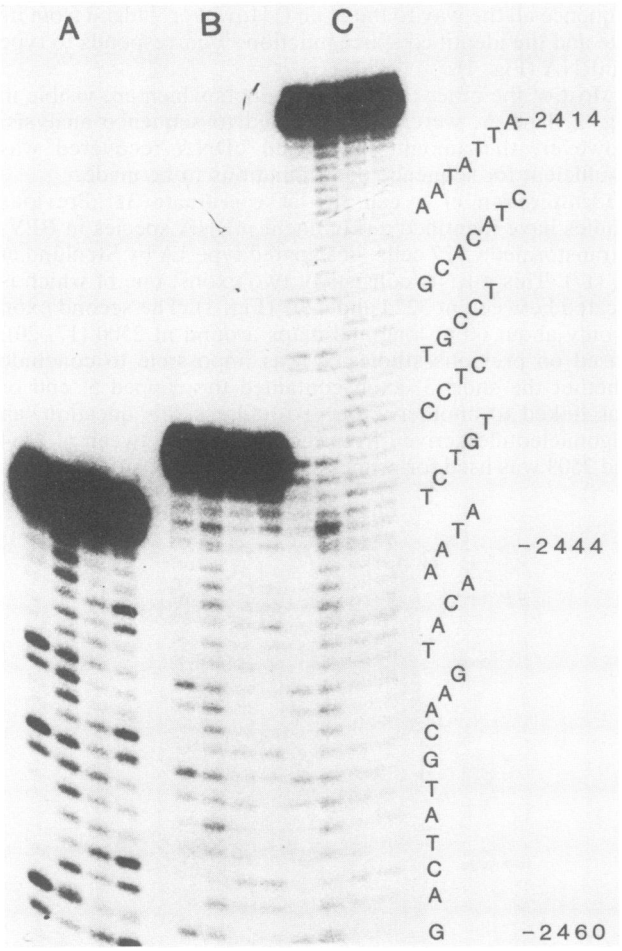

FIG. 3. 5′ end sequences from cDNAs obtained by extension of the nt 2484 to 2503 primer. The cDNAs were fractionated on a 10% polyacrylamide gel (Fig. 2, lane B). Two bands from the cluster (a and b in Fig. 2, lane B) were sequenced separately (lanes A and B, respectively). An additional cDNA approximately 90 nt long (c in Fig. 2, lane B) was also sequenced (lane C). The numbers in the margins indicate nt positions, with the intervening nt sequences shown.

obtained with this oligonucleotide were purified by hybridization selection and separated by 10% polyacrylamide gel electrophoresis. A prominent cluster of cDNAs which were 55 to 65 nt long was observed (Fig. 2, lane B). Sequence analysis of the bands labeled *a* and *b* in Fig. 2, lane B, revealed a sequence which matched the BPV-1 sequence around coordinate 31 perfectly. Based on their sizes, it was calculated that the 5′ ends were located in the region between nt 2436 and 2447 (Fig. 3).

The extension products contained yet another component, consisting of two bands corresponding to estimated lengths of 92 and 93 nt (Fig. 2, lane B). Sequence analysis of the most prominent of these bands showed that its 5′ end mapped around nt 2410 (Fig. 3). No additional splices were detected in the sequence.

Taken together the results show that the type 1 mRNAs consist of several mRNA species whose 5′ ends are located between nt 2410 and 2447, suggesting that a promoter is located in this part of the BPV-1 genome.

Identification of a cap site at coordinate 38. Sequence analysis of extension product b in Fig. 2, lane A, showed that it represents a colinear RNA species whose 5′ end mapped around nt 3080 (data not shown). The structure of this mRNA species was also studied by the nuclease S2 protection technique with a *Pst*I-*Hin*fI fragment (5′ end at nt 3239) as a probe. A major protected fragment with the 5′ end located in the region between nt 3078 and 3080 was identified. A second protected fragment was also observed, whose 5′ end mapped between nt 3070 and 3072. The latter fragment presumably corresponds to an extension product in Fig. 2, lane A, which migrated at a slightly slower rate than did extension product b. It is noteworthy that no putative splice acceptor sites are present in BPV-1 sequence at the positions where the 5′ ends were mapped. We therefore conclude that extension product b (Fig. 2, lane A) is a cDNA copy of a colinear mRNA whose promoter is located around coordinate 39. Moreover, the results show that two cap sites exist.

Further analysis of mRNAs transcribed from the promoter at coordinate 1. A primer matching the BPV-1 sequence between nt 563 and 582 was used for further analysis of mRNAs transcribed from the promoter at coordinate 1. The reason for performing this analysis was that studies of mRNAs in cottontail rabbit papillomavirus-transformed cells have identified a separate promoter located before the E7 open reading frame (ORF) (5). Moreover, Yang et al. (20) have identified cDNA clones whose 5′ ends map near the beginning of the E7 ORF (mRNA type 4B [Fig. 1]). A prominent 270-nt cDNA was identified among the extension products after hybridization selection (Fig. 2, lane C) whose sequence revealed a splice junction joining nt 304 and 528. The structure of this cDNA together with its size suggested that it corresponds to type 4 mRNA (Fig. 1).

If a separate promoter is used for transcription of the E7 ORF in BPV-1, cDNAs that are 140 to 170 nt long would be expected to exist. A few minor cDNAs in this size range were observed after fractionation (Fig. 2, lane C). However, sequence analysis showed that they represented prematurely terminated extension products of type 4 mRNA (data not shown).

DISCUSSION

Previous studies of BPV-1 gene expression in transformed C127 cells have revealed a complex pattern of viral mRNAs which apparently are generated through differential splicing (17, 20). A major cap site was identified at nt 89 by Ahola et al. (1). This cap site, which is preceded by TATA and CAAT motifs, is used by at least four differentially spliced mRNAs (Fig. 1). The results of the present study (Fig. 4) show that at least two internal promoters are utilized for transcription of viral mRNA in BPV-1-transformed C127 cells. Previous studies (17, 20) have identified an abundant mRNA approximately 1 kilobase long in BPV-1-transformed C127 cells (type 1 mRNA [Fig. 1]). However, these studies failed to determine the precise structure of the leader, and the possibility remained that the 5′ exon might be connected to yet another leader sequence, derived from the cap site at coordinate 1. Our present results show that the type 1 mRNAs have heterogeneous 5′ ends mapping between nt 2410 and 2447, suggesting that these ends represent cap sites. Based on our results, it can be concluded that a promoter region should be located around nt 2410. We have previously reported that a TATA-like sequence, TAATATT, is located between nt 2414 and 2420 in BPV-1 (17). Hence, this sequence motif is likely to serve as a TATA box for

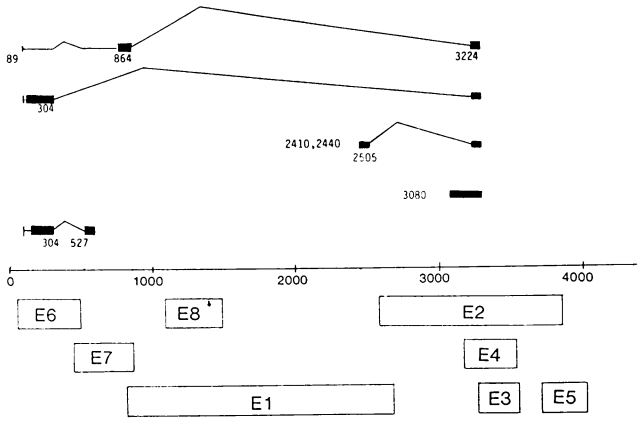

FIG. 4. Summary of results obtained from primer extension experiments. The closed boxes indicate sequences obtained by extension of the different primers. The locations of splice donor and acceptor sites and 5′ ends (nt 2410, 2440, and 3080) are indicated.

transcription of the type 1A mRNAs. The heterogeneous cap sites might be related to the aberrant sequence of this TATA box. It is noteworthy that this region is extremely well conserved between BPV-1, European elk papillomavirus, and deer papillomavirus (1), further supporting the notion that it contains a promoter. Using electron microscopic heteroduplex mapping, Stenlund et al. (17) observed two classes of type 1 mRNAs, one of which contained significantly longer 5′ leaders. In the present study we identified two minor cDNAs which contained leaders that were 30 to 40 nt longer than those of the major type 1 mRNAs. Their transcription is apparently not dependent on the TATA box at nt 2414 to 2420 since their cap sites are located upstream of this position.

The function of the type 1 mRNAs is difficult to predict. The leader sequence contains a putative initiation codon. However, if translation initiates at this AUG, a product 16 amino acids long would be translated, terminating soon after the splice junction. The 3′ exon covers approximately half of the E2 ORF, a major fraction of the E4 ORF, and the entire E3 ORF. However, no AUGs are present in these ORFs after the splice acceptor site at nt 3225. The 3′ exon also includes the E5 ORF, which contains an internal AUG. This ORF was recently shown to encode an important transforming protein (6, 14, 15, 21). Consequently, an interesting possibility is that the type 1 mRNAs encode the E5 protein. If this is true, the mRNA contains an unusually long 5′ noncoding region. It is interesting that no equivalent to the type 1 mRNA has been detected in tumors induced by cottontail rabbit papillomavirus, a papillomavirus which seems to lack an ORF that is equivalent to E5 (5).

Another interesting finding was that a set of colinear RNAs appear to exist which are initiated around nt 3080. Judging from the extension products, these mRNAs are fairly abundant. Stenlund et al. (17) have previously mapped colinear mRNAs to this region by S1 nuclease analysis, and Yang et al. (20) identified a class of colinear cDNA clones which also started near this position. However, based on the previous studies, the possibility that these RNAs were connected to an additional short leader sequence could not be excluded. Our present results demonstrate that a cap site is indeed located around nt 3080, suggesting that a promoter region is located immediately upstream of this position. However, although a search for promoterlike sequence motifs did not reveal any such elements, it is known that certain promoters of animal viruses, such as the late simian virus 40 promoter (4) and the promoter for the IVa2 mRNA of adenovirus (2), lack these elements. The colinear mRNAs with cap sites at nt 3080 may encode the E4 protein. The E4 ORF starts at nt 3173 and contains an ATG between nt 3191 and 3193. An alternative possibility is that they encode a truncated E2 protein. The E2 ORF, which starts at nt 2581, contains an internal ATG between nt 3091 and 3093. An additional question to be addressed is whether a separate promoter controls transcription of mRNAs encoding the E7 protein. Danos et al. (5) have previously shown that several TATA motifs are located near the 5′ end of the E7 ORF in the cottontail rabbit papillomavirus genome and at least some of these motifs appear to function as promoters. Moreover, Yang et al. (20) have identified BPV-1 cDNAs which start close to the 5′ end of the E7 ORF. However, our results suggest that BPV-1 transcription differs from cottontail rabbit papillomavirus transcription in this respect, since no major cap sites were identified in this region. This was not unexpected since no TATA-like motifs are present in this region of the BPV-1 genome. Based on the structure of the type 4 mRNA (Fig. 1), it can be predicted that the E7 ORF is translated into an E6-E7 fusion protein in BPV-1-transformed cells. Thus, there may be no need for a separate promoter for expression of the E7 protein in BPV-1.

Rösl et al. (13) have studied DNase-hypersensitive regions in the BPV-1 genome. Major hypersensitive regions were identified in the noncoding region and close to coordinates 30, 39, and 49. Since hypersensitive sites often reflect transcriptional control regions, it is interesting that two of these sites are located very close to the internal promoters that were identified in the present study.

Our study shows that major splice donor sites are located at nt 304, 864, and 2505 and that major splice acceptor sites are located at nt 527 and 3224, confirming previous results obtained by S1 nuclease mapping and by cDNA cloning (17, 20).

Taken together our results demonstrate that BPV-1 gene expression is very complex. At least three different promoters control BPV-1 gene expression in transformed cells. Recent results obtained in the laboratories of P. Howley and M. Botchan (P. Howley and A. Stenlund, personal communications) suggest that additional promoters are located in the noncoding region. It thus seems likely that further studies of BPV-1 gene expression will unravel additional complexities. It is already clear based on available data that the BPV-1 genome is considerably more complex than the genomes of the other members of the papovavirus family.

ACKNOWLEDGMENTS

We thank Hans Hultberg for the generous gift of the oligonucleotides.

This investigation was supported by grants from the Swedish Cancer Society, the Swedish Council for Forestry and Agricultural Research, and the Swedish National Board for Technical Development.

LITERATURE CITED

1. **Ahola, H., P. Bergman, A.-C. Ström, J. Moreno-López, and U. Pettersson.** 1986. Organization and expression of the transforming region from the European elk papillomavirus (EEPV). Gene **50**:195–205.

1a. **Ahola, H., A. Stenlund, J. Moreno-López, and U. Pettersson.** 1983. Sequences of bovine papillomavirus type 1 DNA—functional and evolutionary implications. Nucleic Acids Res. **11**:2639–2650.

2. **Baker, C. C., and E. B. Ziff.** 1981. Promoters and heterogeneous 5′ termini of the messenger RNAs of adenovirus serotype 2. J. Mol. Biol. **149**:189–221.
3. **Brawermann, G. J., Mendecki, and S. Y. Lee.** 1972. A procedure for the isolation of mammalian messenger ribonucleic acid. Biochemistry **11**:637–641.
4. **Contreras, R., D. Gheysen, J. Knowland, A. van de Voorde, and W. Fiers.** 1982. Evidence for the direct involvement of DNA replication origin in synthesis of late SV40 RNA. Nature (London) **300**:500–506.
5. **Danos, O., E. Georges, G. Orth, and M. Yaniv.** 1985. Fine structure of the cottontail rabbit papillomavirus mRNAs expressed in the transplantable VX2 carcinoma. J. Virol. **53**: 735–741.
6. **DiMaio, D., D. Guralski, and J. T. Schiller.** 1986. Translation of open reading frame E5 of bovine papillomavirus is required for its transforming activity. Proc. Natl. Acad. Sci. USA **83**: 1797–1801.
7. **Kleiner, G., W. Dietrich, and H. Pfister.** 1986. Differential regulation of papillomavirus expression in transformed fibroblasts and carcinoma cell lines. EMBO J. **5**:1945–1950.
8. **Law, M.-F., D. R. Lowy, I. Dvoretzky, and P. M. Howley.** 1981. Mouse cells transformed by bovine papillomavirus contain only extrachromosomal viral DNA sequences. Proc. Natl. Acad. Sci. USA **78**:2727–2731.
9. **Lowy, D. R., I. Dvoretzky, R. Shober, M.-F. Law, I. Engel, and P. M. Howley.** 1980. In vitro tumorigenic transformation by a defined subgenomic fragment of bovine papillomavirus DNA. Nature (London) **287**:72–74.
10. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
11. **Öfverstedt, L.-G., K. Hammarström, N. Balgobin, S. Hjertén, U. Pettersson, and J. Chattopadhyaya.** 1984. Rapid and quantitative recovery of DNA fragments from gels by displacement electrophoresis (isotachophoresis). Biochim. Biophys. Acta **782**:120–126.
12. **Pfister, H.** 1984. Biology and biochemistry of papillomaviruses. Rev. Physiol. Biochem. Pharmacol. **99**:111–181.
13. **Rösl, F., W. Waldeck, H. Zentgraf, and G. Sauer.** 1986. Properties of intracellular bovine papillomavirus chromatin. J. Virol. **58**:500–507.
14. **Schiller, J. T., W. C. Vass, K. H. Vousden, and D. R. Lowy.** 1986. E5 open reading frame of bovine papillomavirus type 1 encodes a transforming gene. J. Virol. **57**:1–6.
15. **Schlegel, R., M. Wade-Glass, M. S. Rabson, and Y.-C. Yang.** 1986. The E5 transforming gene of bovine papillomavirus encodes a small hydrophobic polypeptide. Science **233**:464–467.
16. **Stenlund, A., J. Moreno-López, H. Ahola, and U. Pettersson.** 1983. European elk papillomavirus: characterization of the genome, induction of tumors in animals, and transformation in vitro. J. Virol. **48**:370–376.
17. **Stenlund, A., J. Zabielski, H. Ahola, J. Moreno-López, and U. Pettersson.** 1985. Messenger RNAs from the transforming region of bovine papillomavirus type 1. J. Mol. Biol. **182**:541–554.
18. **Virtanen, A., and U. Pettersson.** 1985. Organization of early region 1B of human adenovirus type 2: identification of four differentially spliced mRNAs. J. Virol. **54**:383–391.
19. **Waldeck, W., F. Rösl, and H. Zentgraf.** 1984. Origin of replication in episomal bovine papilloma virus type 1 DNA isolated from transformed cells. EMBO J. **3**:2173–2178.
20. **Yang, Y-C., H. Okayama, and P. M. Howley.** 1985. Bovine papillomavirus contains multiple transforming genes. Proc. Natl. Acad. Sci. USA **82**: 1030–1034.
21. **Yang, Y.-C., B. A. Spalholz, M. S. Rabson, and P. M. Howley.** 1985. Dissociation of transforming and trans-activation functions for bovine papillomavirus type 1. Nature (London) **318**: 575–577.