

# The primary structure and genetic organization of the bovine papillomavirus type 1 genome

Ellson Y. Chen\*, Peter M. Howley†, Arthur D. Levinson\* & Peter H. Seeburg\*

\* Department of Molecular Biology, Genentech Inc., South San Francisco, California 94080, USA

† Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205, USA

*The complete nucleotide sequence of the double-stranded circular DNA of bovine papillomavirus type 1 (BPV-1) was determined. Analysis of this sequence in conjunction with known transcriptional data for the virus provides a basis for determining the organization of the papillomavirus genome. All the major open reading frames are located on the same DNA strand. The region transcribed in BPV-transformed cells contains open frames in all three translational frames whereas the region transcribed in productively infected bovine fibropapillomas is characterized by two large open reading frames partitioned by a single translational stop codon. The localization of sequences diagnostic of promoters and polyadenylation sites suggests that splicing is involved in the biosynthesis of the viral mRNAs. A sequence comparison indicates the genomic organization of the bovine papillomavirus and that of the simian virus 40-polyomavirus subgroup to be distinct, suggesting that these two groups of viruses are evolutionarily unrelated.*

STUDIES on the papillomaviruses date back to 1933 when the Shope papillomavirus was first described as the causative agent of cutaneous papillomas in cottontail rabbits<sup>1</sup>. The papillomaviruses are widespread in nature and are responsible for a variety of pathological conditions in humans. To date, however, none of the papillomaviruses has been successfully propagated in the laboratory and as a consequence our knowledge of the biology and genetic organization of this group of viruses is limited. In general, papillomaviruses are highly host specific and replicate only in the terminally differentiated epidermal cells of the papillomas (warts) which they cause. There is a subgroup of papillomaviruses, consisting of the bovine papillomaviruses types 1 and 2 (BPV-1 and BPV-2), the deer fibroma virus and the ovine papillomavirus, which are tumorigenic when inoculated into hamsters<sup>2-4</sup>.

BPV-1 is the most extensively studied of the transforming papillomaviruses. It causes cutaneous fibropapillomas in cattle<sup>5</sup> and a naturally occurring fibroblastic tumour, equine sarcoid, in horses<sup>6</sup>. In the laboratory, it is tumorigenic in pika<sup>7</sup>, domestic rabbits<sup>8</sup> and hamsters. The virus is capable of inducing transformation in mouse<sup>9-12</sup>, hamster<sup>13</sup> and bovine conjunctive<sup>14</sup> cells, and the cloned viral DNA can transform susceptible mouse cells<sup>15,16</sup>.

The structure of BPV-1 is similar to that of other papillomaviruses. It is a small icosahedral DNA virus with 72 capsomers whose genome consists of a double-stranded, circular DNA molecule of molecular weight (MW)  $5.0 \times 10^6$  (ref. 17). The virions contain a major capsid protein of MW 53,000 (ref. 18), possibly some minor capsid proteins and the host histones H2A, H2B, H3 and H4. The region transcribed in BPV-1-transformed cells (and by analogy with other DNA tumour viruses such as simian virus 40(SV40) and polyoma, the region transcribed in epidermal cells before the onset of vegetative viral DNA synthesis) is localized to a single strand and extends over 50% of that strand<sup>19</sup>. Additional RNA transcripts not found in transformed cells are present in the epidermal cells of a bovine fibropapilloma. They originate from the remaining portion of the same strand and encode the major capsid protein (L. Engel, C. A. Heilman and P.M.H., in preparation). To date, nothing is known of the virus proteins expressed in either BPV-1-transformed cells, the dermal fibroblasts of a fibropapilloma, or in the proliferating epidermal cells of the basal layer or stratum spinosum of a fibropapilloma.

Despite the lack of a tissue culture system for the *in vitro* propagation of the papillomaviruses, which to date has precluded detailed genetic and molecular study of these viruses, a large amount of taxonomic information has been accumulated

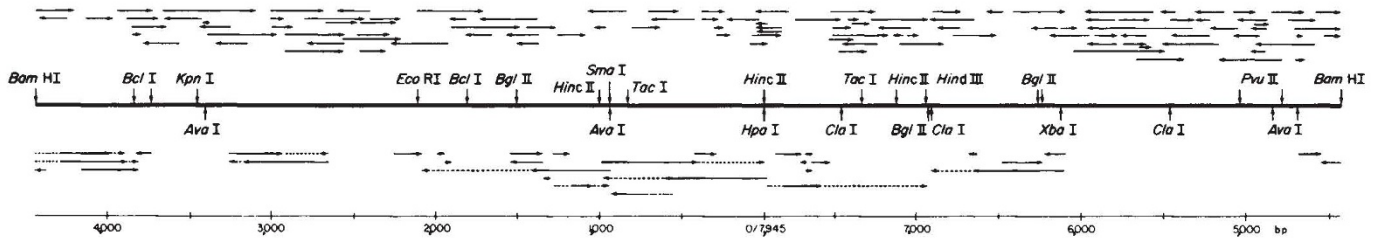
on the human and bovine papillomaviruses. The purified virus has been used to make specific antisera and the viral DNA has been analysed to yield characteristic restriction endonuclease cleavage patterns and to provide a template for the purification of virus-specific RNA or DNA probes. More recently, using recombinant DNA technology, complete papillomavirus genomes have been cloned directly into *Escherichia coli* plasmid vectors for the generation and standardization of specific papillomavirus molecular reagents<sup>16,20,21</sup>. To further our understanding of the functional organization of the BPV-1 genome, we have now determined, and present here, the complete nucleotide sequence of the viral DNA.

## Sequencing strategy

About 95% of the sequence was obtained by a shotgun approach<sup>22,23</sup> using computer-assisted reassembly from data generated by sequencing with chain terminators<sup>24</sup> single-stranded recombinant phage M13mp7 (ref. 25) DNA templates which contained random overlapping segments from both strands of BPV-1 (isolate 307) DNA. Random segments were generated by partial DNase I digestion in the presence of Mn<sup>2+</sup> (ref. 26) from recombinant pBR322 DNA carrying the complete BPV-1 DNA integrated at its single *Bam*HI site. Suitable fragments were isolated by gel electrophoresis and cloned into phage M13mp7 RF-DNA<sup>25</sup>. Colourless plaques containing recombinant phage were grown and aliquots of the phage DNAs spotted on two nitrocellulose filters and probed with radioactively labelled BPV-1 and pBR322 DNA to select templates for sequencing. The shotgun approach yielded a total of 22,500 base pairs (bp) of sequence information which covered ~95% of the genome.

To complete the sequence defined fragments which spanned the remaining gaps were subcloned and sequenced using the recently developed phage vectors M13mp8 and mp9 (ref. 27). These vectors expedite selective sequencing in an elegant manner by accepting restriction endonuclease-generated DNA fragments with a wide variety of heterologous termini. The regions of BPV-1 DNA selected for sequencing by this approach covered a total of 4,800 bp. Figure 1 shows the origin and strandedness of all the sequenced fragments of BPV-1 DNA relative to a linearized map of the viral genome. The entire nucleotide sequence (7,945 bp) of BPV-1 DNA is shown in Fig. 2.

The BPV-1 genome has been opened at the single *Hpa*I site (0/1.0 map units, previously 0.88 map units) and numbering proceeds in the 5' to 3' direction of transcription, initiating with



**Fig. 1** Distribution and strandedness of sequenced DNA templates relative to a map of the BPV-1 genome. BPV-1 DNA is shown as a solid black line and is linearized at the unique *Bam*HI site used for generating pBPV-1 DNA. The recognition sites of several restriction endonucleases are shown as landmarks. The arrows above the map show the length and 5' to 3' direction of single-stranded portions of the viral DNA cloned into phage M13mp7 DNA. To generate these templates, plasmid pBPV-1 DNA was digested at room temperature with DNase I (ref. 26) in 0.5 ml 50 mM Tris-HCl, pH 7.5, 30 mM NaCl, 1 mM MnCl<sub>2</sub>, containing 500 µg DNA, 50 ng bovine serum albumin and 20 ng pancreatic DNase I (Sigma). Aliquots of 160 µl were withdrawn after 16, 20 and 24 min, digestion was stopped by the addition of EDTA to 5 mM and aliquots were combined, extracted with phenol and the DNA precipitated with ethanol. To maximize blunt ends, DNA was incubated with *Escherichia coli* DNA polymerase I (10 U of large fragment; Boehringer) for 30 min at 15 °C in 300 µl of 20 mM Tris-HCl, pH 7.5, 0.1 mM EDTA, 6 mM MgCl<sub>2</sub>, 50 mM NaCl, 1 mM dithiothreitol and 50 mM each of the four deoxyribonucleotide triphosphates (dCTP was set to a specific activity of 1,000 c.p.m. pmol<sup>-1</sup> using [ $\alpha$ -<sup>32</sup>P]dCTP; Amersham). DNA was recovered by phenol extraction and ethanol precipitation and size-separated on a 10% polyacrylamide gel. DNA 300–1,000 bp long was recovered by electroelution followed by ethanol-precipitation from the eluate. Recovered DNA (~10% of the DNA loaded on the gel) was ligated to endo *Hinc*II-cleaved M13mp7 RF-DNA (100 ng) and the ligation mixture was used to transform competent *E. coli* JM 103 cells. Transformation and plating of transformed cells were performed as described in ref. 25. Template DNAs prepared<sup>25</sup> from 164 recombinant (colourless) plaques were dissolved in water in a final volume of 15 µl. Aliquots of 1 µl added to 1 µl 0.5 M NaOH, 1 M NaCl were spotted onto each of two nitrocellulose filters and filters were processed for hybridization<sup>39</sup> with radiolabelled probes to pBR322 DNA and BPV-1 DNA<sup>23</sup>. Filter-spotted DNAs which gave a signal with the BPV-1 DNA probe and some of those which hybridized with both probes were selected for sequencing. The arrows below the genomic map indicate the templates whose sequences filled gaps in, and confirmed, data obtained from the shotgun approach. These templates were generated by cloning defined restriction fragments with heterologous and homologous termini into phage vectors M13mp8 and mp9 RF-DNA<sup>27</sup>. Typically, the desired fragment was obtained after polyacrylamide gel electrophoretic separation of 1 µg of pBPV-1 DNA cleaved with the appropriate restriction enzyme(s) and isolation from the gel. Fragments (~50 ng) were ligated with T4 DNA ligase to singly or doubly cleaved M13mp8 or mp9 RF-DNA (20 ng). The small arrows represent a selection of sequenced endo *Alu*I-generated fragments obtained by cloning gel-isolated *Alu*I-cleaved BPV-1 DNA into the *Sma*I site of phage M13mp8 RF-DNA.

the first of the six nucleotides of the *Hpa*I site. In previous studies on BPV-1 RNA transcripts in transformed cells<sup>19</sup>, map positions were based on the unique *Hind*III site and the non-coding strand was numbered.

### Open reading frames

The analysis of the nucleotide sequence for each of the possible translational reading frames revealed that all the significant open reading frames are located on one DNA strand. Approximately 90% of this strand contains long open reading frames. The only significant region which has no open frames lies between 0.90 and 0.01 map units (from base ~7,150 to ~80). Figure 3 shows, for all the potential reading frames, the translational maps of both strands of the BPV-1 genome with the open reading frames greater than ~500 bases long. There are several small open reading frames of up to 450–500 bases located on the opposite strand; however there is no evidence that transcripts complementary to that strand are made in transformed or productively infected cells (ref. 19; L. Engel, C. A. Heilman and P.M.H., unpublished).

### The transforming region

Using the quantitative assay described by Dvoretzky *et al.*<sup>11</sup> with the cloned BPV-1 DNAs, it has been possible to localize the region necessary for BPV-mediated transformation to a specific segment making up 69% of the genome<sup>15,16</sup>. This region is bounded by the *Hind*III (base 6,958) and *Bam*HI (base 4,450) sites. Attempts to reduce the size of this region by cleaving within this segment at the *Hpa*I (base 1), *Kpn*I site (base 3,455), or the *Eco*RI (base 2,113) sites result in a loss of transformation activity.

Recent studies mapping the transcripts present in BPV-1-transformed mouse cells have provided further information concerning the genetic organization of the transforming region. Five distinct transcripts were detected and mapped in BPV-1-transformed cells (see Fig. 4). The 3' end of each of these transcripts maps at approximately 0.53 map units and the 5' ends of the bodies of distinct transcripts map at approximately 0.03, 0.09, 0.34, 0.39 and 0.41 map units<sup>19</sup>. In this study, no splicing was demonstrated, but the experimental conditions were such that a small 5'-leader sequence would not have been detected.

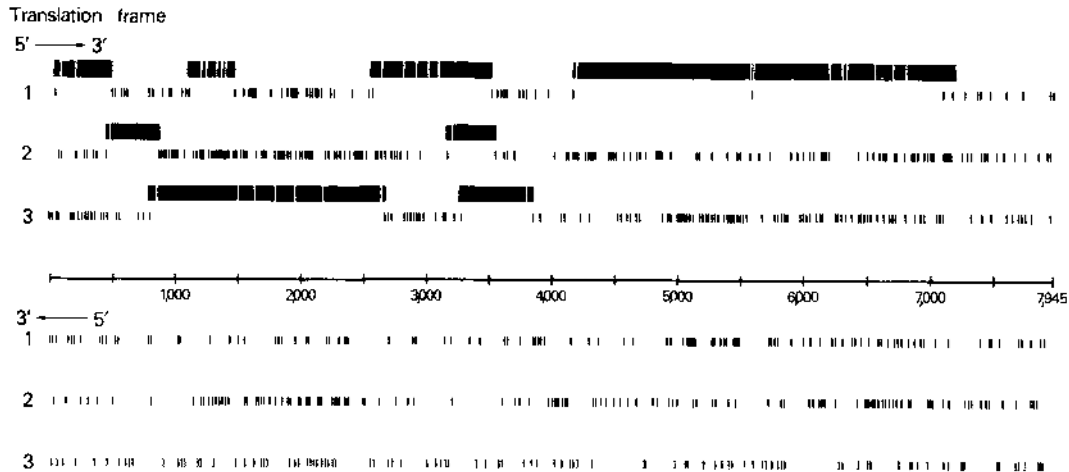
The sequence analysis of this region reveals several open translational frames in the coding strand involving each of the three possible frames (Fig. 3). Within this region there are two TATAAA sequences (bases 7,108 and 58), located just upstream from the 5' end of the body of the largest RNA transcript identified in the transformed cells (Fig. 4). The TATAAA sequence diagnostic of a eukaryotic promoter is located around base -30 of the cap site of viral and cellular mRNAs and is thought to be involved in the proper positioning of the 5' ends of these mRNAs<sup>28</sup>. Which, if either, of the two potential promoter elements is involved in the biogenesis of the viral transcripts in the transformed cells is under investigation. There is a polyadenylation recognition sequence, AATAAA, at base 4,179 (0.53 map units) which corresponds well with the 3' ends of the RNA transcripts identified in BPV-transformed cells. Thus, the sequence analysis is consistent with the transforming region containing a single transcriptional unit consisting of promoters at either 0.90 or 0.01 map units, coding sequences mapping from 0.03 to 0.53 map units and a polyadenylation site at 0.53 map units. The different transcripts detected in transformed cells could therefore be

**Table 1** Segments of dyad symmetry located in the non-coding region of the BPV-1 genome

Sequence	Location	Internal segment size
GTGCA TTGAATTA	7,240–7,253	88
CACGCTCAACTGAAT	7,152–7,139	
AGAAACTAAGC	7,331–7,341	222
TCTTTGAATCG	7,108–7,098	
GCTTTT TTTT	7,435–7,444	336
CGAAAAATAAA	7,098–7,089	
TTTTTTGGAGGATTA	7,568–7,581	486
AAAAACGTCGGAAT	7,081–7,098	
ATGCAGCATTTA	7,872–7,882	265
TACGTCGTAAT	7,606–7,596	
GGAAAAAATA	37–47	50
CCCTTTTTTAT	7,931–7,921	







**Fig. 3** Open reading frames in BPV-1 DNA. The entire length (7,945 nucleotides) of BPV-1 DNA opened at the unique *Hpa*I site is depicted in the middle of the figure. The rows of small vertical lines shown above and below the genomic map indicate the positions of the translational stop codons TAA, TAG and TGA found in the translational reading frame of the two DNA strands. Horizontal black bars indicate regions of 500 nucleotides or more which lack a termination codon. White lines superimposed on these bars indicate the positions of the translational initiation codon ATG. Note that the largest bar contains one stop codon.

generated by differential splicing using a common 5'-end leader sequence. There is a potential cap site (ACA) located 31 bases downstream from the TATAAA sequence at position 58. A potential splice donor sequence (AG<sup>1</sup>GTGCAT) is located at base 183, which if used in the biogenesis of the BPV-1-transformed mRNAs, could result in a leader sequence of approximately 100 bases which might be spliced to each of the five bodies mapped in the transforming region. There is also an ACA cap sequence 25 bases downstream from the potential TATAAA promoter at base 7,108. Interestingly, a potential splice donor sequence (AG<sup>1</sup>GTAAGT) occurs at base 7,141, which would result in a leader sequence of only six bases that could be spliced to the bodies of the viral mRNA species in transformed cells. Potential splice acceptor sites (that is CAG or TAG preceded by sequences rich in pyrimidines and devoid of an AG dinucleotide) are located in the regions of the 5' ends of each of the bodies of the RNA transcripts from the transforming region of the genome. The sequence data are therefore consistent with the experimentally unproved prediction that the viral RNAs expressed in transformed cells are derived from a single transcriptional unit and are generated by differential splicing.

Interestingly, the end of the transforming region of BPV-1 was shown to contain a transcriptional activator ('enhancer') element. Lusky and Botchan (unpublished) found that a DNA segment which spans the polyadenylation site at 0.53 map units can substitute for the 72 bp tandem repeats of SV40 DNA. The relevance of this enhancer for the expression of BPV genes remains to be elucidated.

### Capsid protein-encoding region

The 32% of the BPV-1 genome mapping between the *Bam*HI and *Hind*III sites (bases 4,450–6950) is not expressed in BPV-1-transformed cells<sup>19</sup> and is not required for BPV-1 DNA-mediated transformation<sup>15</sup>. This region is spanned by a single open reading frame which extends from nucleotides 4,171 to 7,095, with a termination codon located in frame at base, 5,593 (see Fig. 3). Analysis of the viral RNA transcripts present in bovine fibropapillomas which contain epidermal cells in which productive replication of BPV-1 virus is occurring reveals species of approximately 1,800 and 4,000 bases which are transcribed from this region (L. Engel, C. A. Heilman and P.M.H., unpublished). This same region has been shown to encode the major capsid protein (MW 53,000) of BPV-1 by filter selection of these mRNA species and translation *in vitro* (C. A. Heilman *et al.*, unpublished). The 3' ends of these

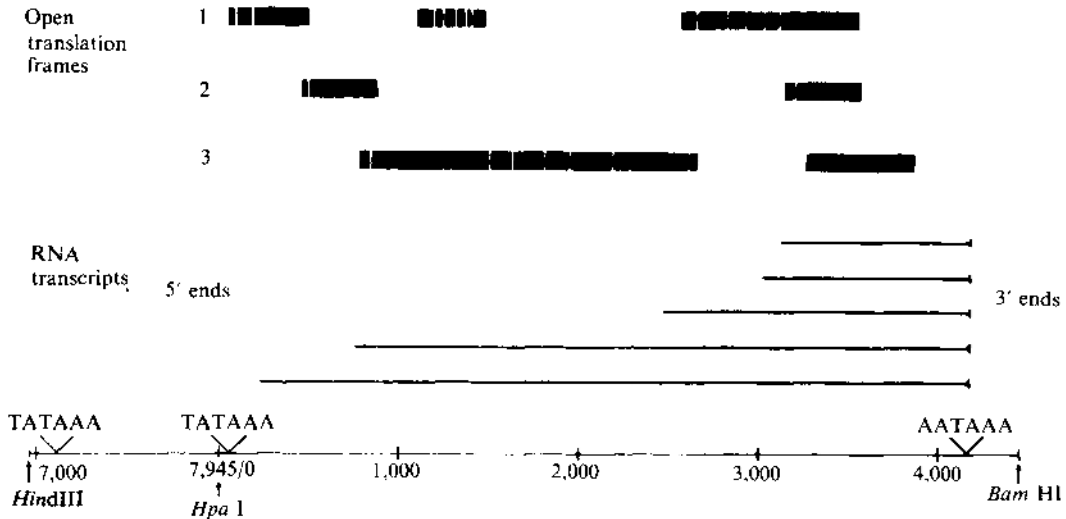
transcripts map to 0.91 map units, where two polyadenylation sites (at nucleotides 7,083 and 7,146) are found. The precise structure and 5' ends of the productive mRNAs are not yet known, thus the amino acid sequence of the 53,000 MW major capsid protein, which would be encoded by ~1,300 bases of the open 3,000 bases in this region, cannot yet be predicted (see Fig. 3). Certain minor capsid proteins such as VP2 and/or VP3 may also be encoded in this region, as may non-structural proteins which are not expressed in BPV-transformed cells.

Given that papillomas in cattle represent an economic problem of some consequence, the identification of genes encoding potential capsid proteins revealed by the DNA sequence may permit the development of a vaccine against bovine papillomas through the design of either synthetic peptides<sup>29,30</sup> or production of the capsid protein(s) in a heterologous host<sup>31</sup>.

### Noncoding region

There is a region of approximately 1,000 bases of the BPV-1 genome between nucleotides 7,146 and 49 in which there are no significant open reading frames. This region is bound on one side by the 3' ends of the productive messenger RNAs (positioned by the polyadenylation sites at base 7,146) and on the other side by an open reading frame in the transforming segment beginning at base 49. By analogy with other well studied virus systems, this noncoding region probably harbours important virus transcription regulatory sequences as well as the sequences comprising the origin of DNA replication.

Within this 1,000 base region no significant repeated sequence units are found analogous to the 72 and 68 bp tandem repeated sequences present in the noncoding regions of SV40<sup>32,33</sup> and BK<sup>34,35</sup> genomes respectively. Yet certain sequence features in the noncoding region of the BPV genome are shared with those origins, suggesting that this region may contain the origin of vegetative DNA replication which functions in epidermal cells. Common sequence features of the papovavirus origins of DNA replication include an A+T-rich region, palindromic sequences and inverted repeated sequences. The noncoding region of the BPV-1 genome contains several A+T-rich regions including one 21 base run (7,077–7,097), a 12 base run (7,879–7,890) and three 10 base runs (7,154–7,163 7,395–7,404 and 7,683–7,692) which contain only A–T residues. Several near-perfect palindromes are located in this region, including CTGAATTTTAATTC (bases 7,392–7,405), TTTGGAGGATT (bases 7,570–7,580) and ATGCTGTA (bases 7,612–7,620). In addition, several areas of dyad symmetries with non-homologous loops are present,



**Fig. 4** The composite structure of the BPV-1 transforming segment. A map of the BPV-1 transforming segment (*Hind*III to *Bam*HI subgenomic fragment) is shown in linear form on the bottom line. The bodies of the polyadenylated RNA transcripts found in BPV-1-transformed mouse cells<sup>19</sup> are indicated. The AATAAA sequence (bases 4,170-4,184), which presumably is the polyadenylation recognition site for this set of RNAs, is positioned on the map. Potential TATAAA promoter elements located at bases 7,108-7,113 and 58-63 are indicated. The horizontal bars above the RNA transcripts represent potential coding regions for BPV-1 proteins in each of the three reading frames (see also Fig. 3).

including AAAGCGAAAGCGACACGCTTT (bases 7,419-7,439). Six segments of dyad symmetry are located in this region with various-sized internal non-homologous loops (see Table 1). Although these sequence features are intriguing, functional data are required to support the contention that the origin of replication is, in fact, in this region.

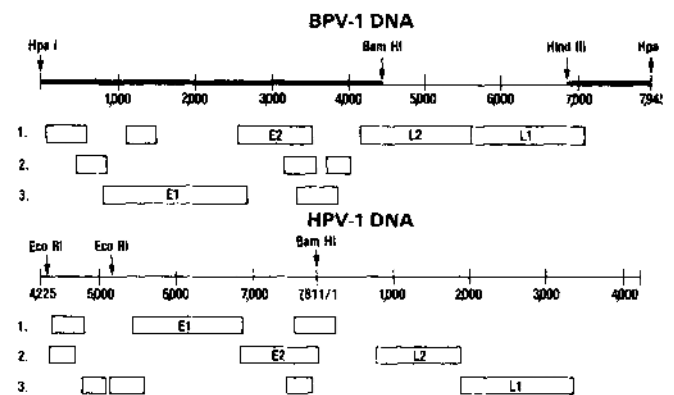
**Conclusions**

The absence of a tissue culture system to propagate the papillomaviruses has precluded their analysis by standard virological procedures. Consequently, although these viruses have been studied since 1933, virtually nothing is known of the genetic organization of the papillomavirus DNAs or of the elements controlling viral gene expression. A number of laboratories have successfully cloned the genomes of various papillomaviruses<sup>16,20,21</sup>, and the recent advances in DNA sequencing technology have made the rapid sequencing of an 8 kilobase genome feasible. The entire sequence of the 7,945 base DNA of BPV-1 presented here, in conjunction with the transcriptional data known for this virus<sup>19</sup>, provide a basis for deducing the genetic organization of the papillomavirus genome.

Only one strand of the genome contains significant open reading frames and transcripts from only this strand can be detected in transformed cells and in productively infected fibropapillomas of cattle. This contrasts with the SV40-polyoma subgroup of viruses and the adenoviruses, where sequences from both DNA strands are transcribed in productively infected cells<sup>17</sup>.

The complete genome of a human papillomavirus (HPV-1a) has also recently been sequenced and found to contain 7,811 bp (ref. 36). As is the case with BPV-1 all the genetic information appears to be encoded on a single strand of the genome. These two genomes have been shown previously to be genetically related and up to ~25% nucleotide sequence homology has been detected and mapped within each genome<sup>37</sup>. Using these homologous regions it is possible to align the two genomes (Fig. 5). The first base of the BPV-1 genome corresponds approximately to base 4,225 in the HPV-1a genome. The open reading frames predicted from each of the sequences in each of the three possible translation frames for the coding strands are depicted below the linearized and aligned genomes. Previous homology studies<sup>37,38</sup> and a comparative analysis of the primary sequence data define two regions which are highly conserved between these two genomes and which map to the

sequences corresponding to the open reading frames designated L1 and E1. We have used the nomenclature proposed by Dano *et al.*<sup>36</sup> and designated the two large reading frames spanning the region not essential for transformation as L1 and L2, and the two largest open reading frames in the transforming region of BPV-1 as E1 and E2. A comparison of the genetic organization of these genomes reveals marked similarities. The size and position of each of the major open reading frames are quite similar. In BPV-1, L2 and L1 are located in the same open reading frame and are partitioned by a single stop codon, and in HPV-1, the 3' end of the L2 reading frame overlaps the 5' end of L1 by five bases and is located in a different reading frame. In the regions corresponding to the transforming segment of BPV-1, there is a noncoding segment ~1,000 base long, 5' to a series of small open reading frames preceding the



**Fig. 5** Alignment of BPV-1 and HPV-1A genomes. The linearized genomes were aligned (HPV-1a base, 4,225 corresponding to BPV-1 base 1) using regions of conserved nucleotides between the DNAs. Open reading frames for each of the three potential translation frames are depicted beneath the genomic maps. The 69% transforming segment of the BPV-1 genome is indicated by a heavy bar. The two large open reading frames spanning the region not essential for transformation and located in the corresponding region in HPV-1a are labelled L1 and L2 in agreement with the proposed nomenclature<sup>36</sup>. Similarly, the two largest open reading frames in the transforming segment of the BPV-1 genome and the corresponding open reading frames in HPV-1a are designated E1 and E2.



large E1 reading frame which is highly conserved. The 5' end of the E2 reading frame overlaps the 3' end of the E1 reading frame in each genome and there is some conservation of amino acids near the amino terminus in the putative proteins encoded by the E2 reading frames. A more thorough analysis of these two genomes is currently the subject of a collaborative study.

The papillomaviruses (subgroup A) and the SV40-polyoma virus (subgroup B or miopapovavirus genus) of viruses together form the papovavirus family of viruses<sup>39</sup>. The term papovavirus derived from the first two letters of each of the known members at the time it was proposed: papillomavirus, polyomavirus and vacuolating virus (SV40)<sup>40</sup>. Although these two groups of viruses are morphologically similar, the papillomaviruses are larger than the SV40-polyomaviruses and comparative studies have suggested that these two virus groups are unrelated. In non-stringent hybridization conditions capable of detecting conserved genomic regions with as much as 33% base mismatch, no homology could be detected between papillomavirus DNAs and those of SV40 or BK DNA<sup>37</sup>. Highly conserved capsid antigenic determinants detectable in all known members of the SV40-polyomavirus group are not present in papillomaviruses; and conversely, a highly conserved cross-reactive antigenic determinant found in all known papillomaviruses is not present in the SV40-polyomavirus group<sup>41</sup>. The sequence analysis presented here shows convincingly, however, that the SV40-polyomavirus group and the papillomaviruses are evolutionarily distinct. The genetic organization of the papillomaviruses is different from that of the polyomaviruses and there is no sequence homology between the coding and non-coding regions of SV40 and BPV-1 DNAs.

There are only limited data available concerning BPV-1 transcription, but it is clear that cellular controls regulate BPV-1 gene expression. In mouse cells, only a portion of the viral genome is expressed. No capsid proteins are synthesized, nor are transcripts complementary to the sequences which encode these proteins observed. In a bovine fibropapilloma, capsid proteins are not produced in the mesenchymal fibroblastic component or in the epidermal cells of the basal layer or the lower stratum spinosum. Only in the terminally differentiated epidermal cells of the granular layer are the capsid proteins detectable. It seems likely, therefore, that the regulation of viral gene expression in the fibropapilloma is at the level of transcription. Two additional transcripts are present in bovine fibropapillomas which are not present in transformed cells. These transcripts are complementary to the sequences which encode the BPV-1 capsid proteins and must therefore be expressed in the terminally differentiated epidermal cells in which capsid protein production occurs. Thus, expression of viral productive functions, such as capsid protein synthesis, are linked to the state of epidermal cell differentiation.

There are three TATAAA sequences (bases 58, 5,089 and 7,108) and two TATATA sequences (bases 4,072 and 6,859) in the coding strand of the BPV-1 genome. These sequences have been shown to be important in localizing the 5' ends of transcripts *in vivo* and *in vitro* when part of a eukaryotic promoter. The TATAAA sequence located at base 58 maps near the 5' end of the largest RNA transcript detected in BPV-1-transformed cells. It seems possible that this TATAAA sequence and the surrounding region define the viral promoter active in transformed cells.

From the DNA sequence it is not clear what transcriptional promoter(s) function in the biosynthesis of the transcripts for the productive messages for the capsid proteins. These mRNAs are complementary to sequences located between bases 4,450 and 7,945. There is a TATATA sequence, which might function as part of the promoter located, at bases 4,072-4,077, just 5' to a large open reading frame starting approximately at base 4,150. Alternatively, the promoter active in the transformed cells may function for the biosynthesis of the productive transcripts and the control of the gene expression may be a result of differential splicing. There are additional TATAAA sites on this strand at bases 5,089-5,094 and 7,108-7,113, and a TATATA sequence at bases 6,859-6,864, any of which may be part of the productive promoter. The definition of the productive transcriptional unit must await precise mapping of the productive transcripts and their 5' ends.

Due to the ability of molecularly cloned BPV-1 DNA to transform mouse cells<sup>15,16</sup>, the possibility of exploiting BPV-1 DNA as a novel eukaryotic cloning and expression vehicle was suggested. The utility of this approach has been demonstrated using the rat preproinsulin gene. It has recently been reported that cells transfected with plasmid DNA containing both BPV-1 and rat preproinsulin sequences can be transformed, and that such transformants harbour the recombinant DNA in an episomal state, directing the expression of proinsulin<sup>42</sup>. The availability of the complete nucleotide sequence should greatly facilitate efforts to enhance the utility of BPV-1 DNA as a cloning/expression vehicle, and will provide the framework required to complete our understanding of the molecular biology of the BPV life cycle.

We thank R. Najarian for help in establishing the BPV-1 template bank used for the shotgun sequencing, L. Engel for assistance in the analysis of the sequence, C. Heilman and L. Engel for making available their unpublished data concerning the transcription of the BPV-1 genome in bovine fibropapillomas, M. Lusky and M. Botchan for communicating their unpublished results on the presence and location of an enhancer sequence in BPV-1 DNA, and S. Bock in helping to prepare the manuscript.

Received 14 April; accepted 30 June 1982.

- Shope, R. E. *J. exp. Med.* **58**, 607-624 (1933).
- Friedman, J. C. *et al. C. r. heb. Séanc. Acad. Sci. Paris* **257**, 3238-3231 (1963).
- Gibbs, E. P. J., Smale, C. J. & Lawman, M. J. *P. J. comp. Path.* **85**, 327-334 (1975).
- Koller, L. D. & Olson, C. *J. invest. Derm.* **58**, 366-368 (1972).
- Lancaster, W. D. & Olson, C. *Virology* **89**, 372-379 (1978).
- Olson, C. & Cook, R. H. *Proc. Soc. exp. Biol. Med.* **77**, 281-284 (1951).
- Puget, A., Favre, M. & Orth, G. *C. r. heb. Séanc. Acad. Sci. Paris* **280**, 2813-2816 (1975).
- Brietburd, F., Favre, M., Zoorob, R., Fortin, D. & Orth, G. *Int. J. Cancer* **27**, 693-702 (1981).
- Thomas, M., Boiron, M., Tanzer, J., Levy, J. P. & Bernard, J. *Nature* **202**, 709-710 (1964).
- Boiron, M., Levy, J. P., Thomas, M., Friedman, J. C. & Bernard, J. *Nature* **201**, 423-424 (1964).
- Dvoretzky, I., Shober, R. & Lowry, D. R. *Virology* **103**, 369-375 (1980).
- Meischke, H. R. C. *J. gen. Virol.* **43**, 473-487 (1979).
- Geraldes, A. *Nature* **226**, 81-82 (1970).
- Black, P. H., Hartley, J. W., Rowe, W. P. & Huebner, R. J. *Nature* **199**, 1016 (1963).
- Lowy, D. R. *et al. Nature* **287**, 72-74 (1980).
- Howley, P. M. *et al. in Viruses in Naturally Occurring Cancer* (eds Essex, M., Todaro, G. & zur Hausen, H.) 233-247 (Cold Spring Harbor Laboratory, New York, 1980).
- Toozé, J. in *Molecular Biology of Tumor Viruses*, Part II (Cold Spring Harbor Laboratory, New York, 1980).
- Meinke, W. & Meinke, G. C. *J. gen. Virol.* **52**, 15-24 (1981).
- Heilman, C. A., Engel, L., Lowy, D. R. & Howley, P. M. *Virology* **119**, 22-34 (1982).
- Heilman, C. A., Law, M.-F., Israel, M. A. & Howley, P. M. *J. Virol.* **36**, 395-407 (1980).
- Danos, O., Katinka, M. & Yaniv, M. *Eur. J. Biochem.* **109**, 457-461 (1980).
- Anderson, S. *Nucleic Acids Res.* **9**, 3015-3027 (1981).
- Messing, J. & Seeburg, P. H. in *ICN-UCLA Symp. molec. Cell Biol.* **23**, 659-669 (1981).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Messing, J., Crea, R. & Seeburg, P. H. *Nucleic Acids Res.* **9**, 304-321 (1981).
- Melgar, E. & Goldthwait, D. A. *J. Biol. Chem.* **243**, 4409-4416 (1968).
- Messing, J. & Vieira, J. *Gene* (submitted).
- Mathis, D. J. & Chambon, P. *Nature* **290**, 310-315 (1981).
- Lerner, R. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 3403-3409 (1981).
- Dressman, G. R. *et al. Nature* **295**, 158-160 (1982).
- Kleid, D. G. *et al. Science* **214**, 1125-1129 (1981).
- Rcdy, V. B. *et al. Science* **200**, 494-502 (1978).
- Fiers, W. *et al. Nature* **273**, 113-120 (1978).
- Scif, I., Khoury, D. & Dhar, R. *Cell* **18**, 963-977 (1979).
- Yang, R. C. A. & Wu, R. *Science* **206**, 456-462 (1979).
- Danos, O., Katinka, M. & Yaniv, M. *EMBO J.* **1**, 232-237 (1982).
- Law, M.-F., Lancaster, W. D. & Howley, P. M. *J. Virol.* **32**, 199, 207 (1979).
- Croissant, O., Testaniere, V. & Orth, G. *C. r. heb. Séanc. Acad. Sci., Paris* (in the press).
- Melnick, J. L. *et al. Internirology* **3**, 106-120 (1974).
- Melnick, J. L. *Science* **135**, 1128-1130 (1962).
- Jenson, A. B. *et al. J. natn. Cancer Inst.* **64**, 495-500 (1980).
- Sarver, N., Gruss, P., Law, M.-F., Khoury, G. & Howley, P. M. *Molec. cell. Biol.* **1**, 486-496 (1981).
- Sanger, F. & Coulson, A. R. *FEBS Lett.* **87**, 107-110 (1980).