Sequences of bovine papillomavirus type 1 DNA – functional and evolutionary implications

Harri Ahola[1,2], Arne Stenlund[2], Jorge Moreno-Lopez[1] and Ulf Pettersson[2]*

[1]Department of Virology, Faculty of Veterinary Medicine and the National Veterinary Institute, The Biomedical Centre, Box 585, and [2]Departments of Medical Genetics and Microbiology, The Biomedical Centre, Box 589, S-751 23 Uppsala, Sweden

Received 10 February 1983; Revised and Accepted 4 April 1983

ABSTRACT

A Swedish isolate of bovine papillomavirus type 1 (BPV1) was cloned and a 4807 base pairs long sequence was determined, covering the entire late region of the BPV1 genome and strategic parts of the early region. The sequence reveals two large uninterrupted reading frames, designated L1 and L2, which encode tentative polypeptides with molecular weights of 55.5 and 50.0 K. The promoter region for early transcription and the common 3'-end of early transcripts were identified in the sequence by performing S1 nuclease analysis of RNA isolated from a BPV1 transformed cell line.

A detailed comparison between the established sequence and sequences from human papillomavirus type 1a (HPV1a) (1) and another BPV1 isolate (2) was also carried out. Only five differences were found when the sequences of the two BPV1 isolates were compared, two of which are located in reading frame L2. The results revealed furthermore that the genomes of BPV1 and HPV1a appear to be organized in a very similar fashion and the homology between reading frames L1 in BPV1 and HPV1 were particularly striking. A comparison of the established BPV1 sequence with sequences from SV40 and polyomavirus revealed no significant homology.

INTRODUCTION

The papillomaviruses comprise an interesting family of viruses. Although discovered half a century ago (3), molecular studies of the papillomaviruses have in the past been precluded since no tissue culture systems are available which allow their propagation. Molecular cloning has alleviated these shortcomings and molecular studies on a number of different papillomaviruses have been initiated in several laboratories. Papillomaviruses have been isolated from humans and a great variety of animal species, including the cotton-tail rabbit, the deer, the european elk, sheep cows and horses (for review see Lancaster and Olson and references therein (4)). A surprising finding from previous studies is the relative lack of homology between the genomes of papillomaviruses from even comparatively closely related species. The bovine serotypes, for instance, hybridize only under low stringency conditions to DNA from the human serotypes (5, 6, 7). It has moreover been found that

some of the human serotypes show surprisingly weak cross- reactivity when hybridized to each other under stringent conditions (5,8). These observations suggest that the papillomaviruses may be subjected to a rapid genetic drift and raise questions as to the evolution of this virus family. We have studied the late region of bovine papillomavirus type 1 (BPV1) by nucleotide sequence analysis and in this communication we compare our results with those obtained by other investigators for the genomes of human papillomavirus type 1a (HPV1a) (1) and another BPV1 isolate (2,9). The results show that papillomaviruses of different origins have genomes which are organized in a very similar way and that different isolates of BPV1 have very similar sequences. Certain landmarks on the BPV1 genome were furthermore identified by mapping the 5'- and 3'-termini of viral transcripts present in papillomavirus transformed cells.

MATERIALS AND METHODS

Isolation of virus. The original wart material was obtained from a Swedish cow in the Gothenburg region. The virus was purified and DNA was extracted according to already published procedures (10). The isolate was identified as bovine papilloma virus type 1 by its restriction enzyme cleavage pattern.

Cloning of the viral genome. The BPV1 genome was linearized by cleavage with endonuclease Hind-III and was inserted into the corresponding cleavage site of the pBR322 plasmid, using already published procedures (11). Subclones containing TaqI and TaqI/BclI fragments of the BPV1 genome (Fig. 1) were inserted into the ClaI and ClaI/BamHI sites of the vector.

DNA sequence analysis. The method of Maxam and Gilbert (12) was followed.

Nuclease S1 analysis. The method of Favaloro et al. (13) was followed using 5'- and 3'-end labeled probes. Total cytoplasmic RNA was extracted from a BPV1 transformed mouse cell line (Stenlund et al., in preparation) by the method of Brawerman et al. (14).

Computer analysis of the DNA sequence. The computer program "Gene manipulator", designed by Mr. P. Andersson (15; P. Andersson unpublished information) was used.

RESULTS

The nucleotide sequence of the late region of the BPV1 genome. The nucleotide sequence of a 4807 base pairs (bp) long segment of the BPV1 genome was established according to the strategy which is outlined in Fig. 1. Cloned restriction enzyme cleavage fragments were used for the analysis and more

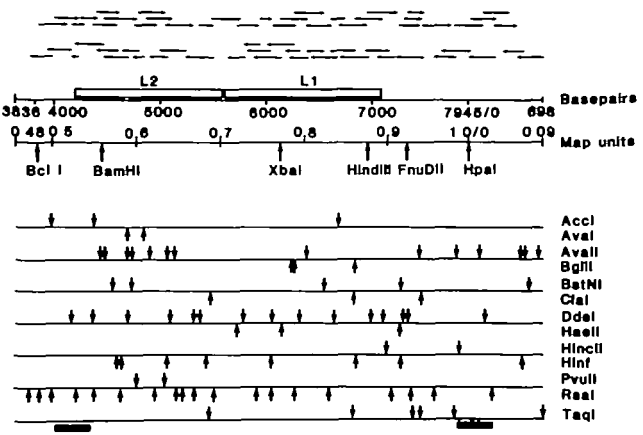

Figure 1: Upper part: A schematic drawing of the region of the BPV1 genome which has been sequenced. Two major open translational reading frames L2 and L1 are indicated. The horizontal arrows show the amount of sequence information that was obtained from different restriction enzyme cleavage sites. The lower panel of the figure shows cleavage sites for selected restriction enzymes. Two fragments which were used for S1 nuclease analysis are also indicated (▬).

than 95% of the sequence was determined by sequencing both the complementary strands. The established sequence (Fig. 2) covers the entire late region of the BPV1 genome and some strategic parts of the early region.

Mapping of the 3'- and 5'-ends of "early" BPV1 RNA. An overlapping set of mRNAs are transcribed from the BPV1 genome in transformed mouse cells (16, 17; Stenlund et al., in preparation). These mRNAs, which most likely correspond to mRNAs expressed early after a lytic papillomavirus infection, have a common 3'-end located at 0.52 map units near the single BamHI cleavage site in the BPV1 genome (Fig. 1). To map the 3'-end more precisely a fragment was 3'-labeled in the AccI cleavage site at position 3965. A 230 nucleotides long fragment was protected after S1 digestion (Fig. 3A) suggesting that the 3'-end is located around nucleotide 4195. This result is in agreement with results reported by Heilman et al. (17).

The location of the 5'-end of the mRNAs present in transformed cells has not been determined to date. The reason is that they carry leader sequences which so far have escaped detection. In order to map the approximate location of the 5'-end, mRNA from transformed cells was fractionated using electrophoresis in formaldehyde-containing gels followed by transfer to nitrocellulose filters and hybridization with cloned TaqI fragments. The

```
3837 TGATCACTGC CATTGCCTTT TCTTCATCTG ACTGGTGTAC TATGCCAAAT CTATGGTTTC TATTGTTCTT GGGACTAGTT GCTGCAATGC AACTGCTGCT
3937 ATTACTGTTC TTACTCTTGT TTTTTCTTGT ATACTGGGAT CATTTTGAGT GCTCCTGTAC AGGTCTGCCC TTTTAATGCC TTTACGTCAC TGGCTATTGG
4037 CTGTGTTTTT ACTGTTGTGT GGATTTGATT TGTTTTATAT ACTGTATGAA GTTTTTTCAT TTGTGCTTGT ATTGCTGTTT GTAAGTTTTT TACTAGAGTT
4137 TGTATTCCCC CTGCTCAGAT TTTATATGGT TTAAGCTGCA GCAATAAAAA TGAGTGCACG AAAAAGAGTA AAACGTGCCA GTGCCTATGA CCTGTACAGG
4237 ACATGCAAGC AAGCGGGCAC ATGTCCACCA GATGTGATAC CAAAGGTAGA AGGAGATACT ATAGCAGATA AAATTTTGAA ATTTGGGGGT CTTGCAATCT
4337 ACTTAGGAGG GCTAGGAATA GGAACATGGT CTACTGGAAG GGTTGCTGCA GGTGGATCAC CAAGGTACAC ACCACTCCGA ACAGCAGGGT CCACATCATC
4437 GCTTGCATCA ATAGGATCCA GAGCTGTAAC AGCAGGGACC CGCCCCAGTA TAGGTGCGGG CATTCCTTTA GACACCCTTG AAACTCTTGG GGCCTTGCGT
4537 CCAGGGGTGT ATGAGGACAC TGTGCTACCA GAGGCCCCTG CAATAGTCAC TCCTGATGCT GTTCCTGCAG ATTCAGGGCT TGATGCCCTG TCCATAGGTA
4636 CAGACTCGTC CACGGAGACC CTCATTACTC TGCTAGAGCC TGAGGGTCCC GAGGACATAG CGGTTCTTGA GCTGCAACCC CTGGACCGTC CAACTTGGCA
4737 AGTAAGCAAT GCTGTTCATC AGTCCTCTGC ATACCACGCC CCTCTGCAGC TGCAATCGTC CATTGCAGAA ACATCTGGTT TAGAAAATAT TTTTGTAGGA
4837 GGCTCGGGTT TAGGGGATAC AGGAGGAGAA AACATTGAAC TGACATACTT CGGGTCCCCA CGAACAAGCA CGCCCCGCAG TATTGTCTCT AAATCACGTG
4937 GCATTTTAAA CTGGTTCAGT AAACGGTACT ACACACAGGT GCCCACGGAA GATCCTGAAG TGTTTTCATC CCAAACATTT GCAAACCCAC TGTATGAAGC
5037 AGAACCAGCT GTGCTTAAGG GACCTAGTGG ACGTGTTGGA CTCAGTCAGG TTTATAAACC TGATACACTT ACAACACGTA GCGGGACAGA GGTGGGACCA
5137 CAGCTACATG TCAGGTACTC ATTGAGTACT ATACATGAAG ATGTAGAAGC AATCCCCTAC ACAGTTGATG AAAATACACA GGGACTTGCA TTCGTACCCT
5237 TGCATGAAGA GCAAGCAGGT TTTGAGGAGA TAGAATTAGA TGATTTTAGT GAGACACATA GACTGCTACC TCAGAACACC TCTTCTACAC CTGTTGGTAG
5337 TGGAGTACGA AGAAGCCTCA TTCCAACTCA GGAATTTAGT GCAACACGGC CTACAGGTGT TGTAACCTAT GGCTCACCTG ACACTTACTC TGCTAGCCCA
5437 GTTACTGACC CTGATTCTAC CTCTCCTAGT CTAGTTATCG ATGACACTAC TACTACACCA ATCATTATAA TTGATGGGCA CACAGTTGAT TTGTACAGCA
5537 GTAACTACAC CTTGCATCCC TCCTTGTTGA GGAAACGAAA AAAACGGAAA CATGCCTAAT TTTTTTTGCA GATGGCGTTG TGGCAACAAG GCCAGAAGCT
5637 GTATCTCCCT CCAACCCCTG TAAGCAAGGT GCTTTGCAGT GAAACCTATG TGCAAAGAAA AAGCATTTTT TATCATGCAG AAACGGAGCG CCTGCTAACT
5737 ATAGGACATC CATATTACCC AGTGTCTATC GGGGCCAAAA CTGTTCCTAA GGTCTCTGCA AATCAGTATA GGGTATTTAA AATACAACTA CCTGATCCCA
5837 ATCAATTTGC ACTACCTGAC AGGACTGTTC ACAACCCAAG TAAAGAGCGG CTGGTGTGGG CAGTCATAGG TGTGCAGGTG TCCAGAGGGC AGCCTCTTGG
5937 AGGTACTGTA ACTGGGCACC CCACTTTTAA TGCTTTGCTT GATGCAGAAA ATGTGAATAG AAAAGTCACC ACCCAAACAA CAGATGACAG GAAACAAACA
6037 GGCCTAGATG CTAAGCAACA ACAGATTCTG TTGCTAGGCT GTACCCCTGC TGAAGGGGAA TATTGGCACAA CAGCCCGTCC ATGTGTTACT GATCGTCTAG
6137 AAAATGGCGC CTGCCCTCCT CTTGAATTAA AAAACAAGCA CATAGAAGAT GGGGATATGA TGGAAATTGG GTTTGGTGCA GCCAACTTCA AAGAAATTAA
6237 TGCAAGTAAA TCAGATCTAC CTCTTGACAT TCAAAATGAG ATCTGCTTGT ACCCAGACTA CCTCAAAATG GCTGAGGACG CTGCTGGTAA TAGCATGTTC
```

```
6337 TTTTTTGCAA GGAAAGAACA GGTGTATGTT AGACACATCT GGACCAGAGG GGGCTCGGAG AAAGAAGCCC CTACCACAGA TTTTTATTTA AAGAATAATA
6437 AAGGGGATGC CACCCTTAAA ATACCCAGTG TGCATTTTGG TAGTCCCAGT GGCTCACTAG TCTCAACTGA TAATCAAATT TTTAATCGGC CCTACTGGCT
6537 ATTCCGTGCC CAGGGCATGA ACAATGGAAT TGCATGGAAT AATTTATTGT TTTTAACAGT GGGGGACAAT ACACGTGGTA CTAATCTTAC CATAAGTGTA
6637 GCCTCAGATG GAACCCCACT AACAGAGTAT GATAGCTCAA AATTCAATGT ATACCATAGA CATATGGAAG AATATAAGCT AGCCTTTATA TTAGAGCTAT
6737 GCTCTGTGGA AATCACAGCT CAAACTGTGT CACATCTGCA AGGACTTATG CCCTCTGTGC TTGAAAATTG GGAAATAGGT GTGCAGCCTC CTACCTCATC
6837 GATATTAGAG GACACCTATC GCTATATAGA GTCTCCTGCA ACTAAATGTG CAAGCAATGT AATTCCTGCA AAAGAAGACC CTTATGCAGG GTTTAAGTTT
6937 TGGAACATAG ATCTTAAAGA AAAGCTTTCT TTGGACTTAG ATCAATTTCC CTTGGGAAGA AGATTTTTAG CACAGCAAGG GGCAGGATGT TCAACTGTGA
7037 GAAAACGAAG AATTAGCCAA AAAACTTCCA GTAAGCCTGC AAAAAAAAAA AAAAATAAA AGCTAAGTTT CTATAAATGT TCTGTAAATG TAAAACAGAA
7137 GGTAAGTCAA CTGCACCTAA TAAAAATCAC TTAATAGCAA TGTGCTGTGT CAGTTGTTTA TTGGAACCAC ACCCGGTACA CATCCTGTCC AGCATTTGCA
7237 GTGCGTGCAT TGAATTATTG TGCTGGCTAG ACTTCATGGC GCCTGGCACC GAATCCTGCC TTCTCAGCCA AAATGAATAA TTGCTTTGTT GGCAAGAAAC
7337 TAAGCATCAA TGGGACGCGT GCAAAGCACC GGCGGCGGTA GATGCGGGGT AAGTACTGAA TTTTAATTCG ACCTATCCCG GTAAAGCGAA AGCGACACGC
7437 TTTTTTTTCA CACATAGCGG GACCGAACAC GTTATAAGTA TCGATTAGGT CTATTTTTGT CTCTCTGTCG GAACCAGAAC TGGTAAAAGT TTCCATTGCG
7537 TCTGGGCTTG TCTATCATTG CGTCTCTATG GTTTTTGGAG GATTAGACGG GCCACCAGTA ATGGTGCATA GCGGATGTCT GTACCGCCAT CGGTGCACCG
7637 ATATAGGTTT GGGGCTCCCC AAGGGACTGC TGGGATGACA GCTTCATATT ATATTGAATG GGCGCATAAT CAGCTTAATT GGTGAGGACA AGCTACAAGT
7737 TGTAACCTGA TCTCCACAAA GTACCGTTGC CGGTCGGGGT CAAACCGTCT TCGGTGCTCG AAACCGCCTT AAACTACAGA CAGGTCCCAG CCAAGTAGGC
7837 GGATCAAAAC CTCAAAAAGG CGGGAGCCAA TCAAAATGCA GCATTATATT TTAAGCTCAC CGAAACCGGT AAGTAAAGAC TATGTATTTT TTCCCAGTGA
7937 ATAATTGTT

1    GTTAACAATA ATCACACCAT CACCGTTTTT TCAAGCGGGA AAAAATAGCC AGCTAACTAT AAAAAGCTGC TGACAGACCC CGGTTTTCAC ATGGACCTGA
101  AACCTTTTGC AAGAACCAAT CCATTCTCAG GGTTGGACTG TCTGTGGTGC AGAGAGCCTC TTACAGAAGT TGATGCTTTT AGGTGCATGG TCAAAGACTT
201  TCATGTTGTA ATTCGGGAAG GCTGTAGATA TGGTGCATGT ACCATTTGTC TTGAAAACTG TTTAGCTACT GAAAGAAGAC TTTGGCAAGG TGTTCCAGTA
301  ACAGGTGAGG AAGCTGAATT ATTGCATGGC AAAACACTTG ATAGGCTTTG CATAAGATGC TGCTACTGTG GGGGCAAACT AACAAAAAAT GAAAAACATC
401  GGCATGTGCT TTTTAATGAG CCTTTCTGCA AAACCAGAGC TAACATAATT AGAGGACGCT GCTACGACTG CTGCAGACAT GGTTCAAGGT CCAAATACCC
501  ATAGAAACTT GGATGATTCA CCTGCAGGAC CGTTGCTGAT TTTAAGTCCA TGTGCAGGCA CACCTACCAG GTCTCCTGCA GCACCTGATG CACCTGATTT
601  CAGACTTCCG TGCCATTTCG GCCGTCCTAC TAGGAAGCGA GGTCCCACTA CGCCTCCGCT TTCCTCTCCC GGAAAACTGT GTGCAACAGG GCCACGTC
```

Figure 2: The nucleotide sequence of a 4807 base pairs long segment of BPV1 DNA. The sequence is indexed according to Chen et al. (2).

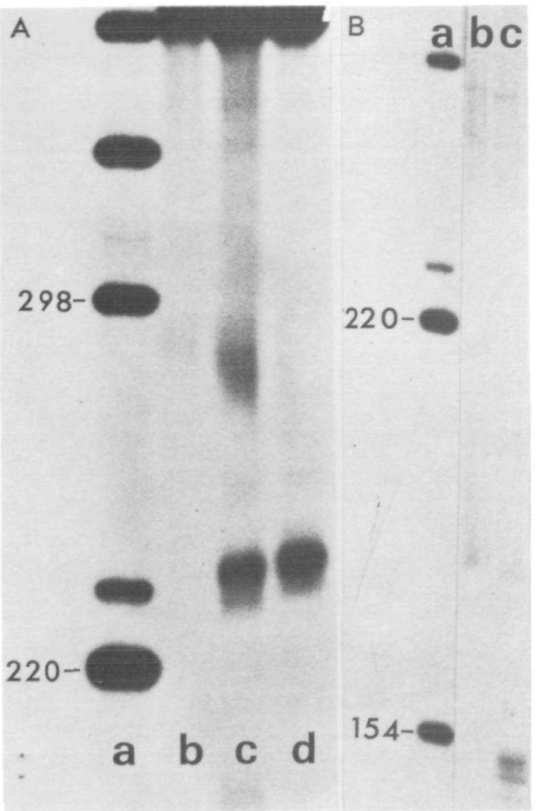

Figure 3A: Identification of the common 3'-end of mRNAs from BPV1 transformed cells. A probe, 3'-labeled in the cleavage site for AccI at nucleotide 3965 was used for S1 nuclease analysis. Lane (a) shows ^{32}P-labeled marker fragments derived from the pBR322 plasmid. Lanes (c) and (d) show fragments protected by cytoplasmic and nuclear RNA from a BPV1 transformed cell line, respectively. Lane b shows results obtained in the absence of RNA.
Figure 3B: Identification of the 5'-end of mRNAs from BPV1 transformed cells. A fragment, 5'-labeled in the RsaI cleavage site at position 239 was used as probe for hybridization and S1 nuclease analysis. The figure shows, marker fragments (a), fragments protected by RNA from a BPV1 transformed cell line (c) and results obtained in the absence of RNA (b).

results revealed that the 5' end is located within the TaqI fragment which traverses the HpaI cleavage site (Fig. 1). For a more detailed analysis of the 5'-end a fragment was 5'-labeled at a cleavage site for endonuclease RsaI, located at position 239 in the sequence depicted in Fig. 2. A 150 bp long fragment was protected after S1 cleavage (Fig. 3B) which suggests that the cap-site for "early" transcription is located close to nucleotide 90 in the BPV1 sequence.

DISCUSSION

Features of the established sequence. The established sequence is depicted in Fig. 2 and Fig. 1 shows the major open translational reading frames which can be identified in the sequence. Position zero is defined as the single HpaI clevage site which is present in the BPV1 sequence. The sequence is compara-

tively AT rich having 44.6% GC base pairs. It has previously been determined that the sequence which is depicted in Fig. 1 is transcribed in the rightwards direction (16,17). Two large open translational reading frames are present in the sequence which are designated L1 and L2 (2). Both reading frames L1 and L2 contain ATG triplets near their 5'ends and they comprise 495 and 469 codons respectively, counted from the first ATG. The predicted amino acid compositions of the L1 and L2 proteins are shown in Table 1. Five so-called TATA-motifs are found in the sequence which are located in the following positions: 58-63 (TATAAA), 4072-4077 (TATATA), 5089-5094 (TATAAA), 6859-6864 (TATATA) and 7108-7113 (TATAAA). The TATA-box around position 60 precedes the cap-site for the "early" BPV1 transcripts, as determined in the present study with 25-30 nucleotides and is thus likely to represent the promotor region for early BPV1 transcription. Three hexanucleotide sequences, AAUAAA, are found in the established sequence at positions: 4179-4184, 7091-7096 and 7155-7160. The hexanucleotide around position 4180 maps about 10 nucleotides before the common poly(A) addition site for the "early" BPV1 transcripts (Fig. 3A). The hexanucleotide between 7155 and 7160 follows immediately after reading frame L1 and is therefore likely to represent the polyadenylation signal for late mRNAs.

A search for homology between the origins for SV 40 and polyoma virus DNA replication and the non-coding region of the established BPV1 sequence yielded negative results; a run of eight T residues present both in the SV40 and the BPV1 sequence is the longest region of homology.

Functional implications of the structural organization of the established sequence. Since the established sequence covers the entire late region and also extends into the early region it would be expected to cover most of the controlling elements for early as well as late BPV1 transcription. Other investigators have shown that early and late genes in the BPV1 genome are located on the same strand (16,17). It is thus conceivable that the same promoter controls early and late BPV1 transcription and that late RNA simply is produced through anti-termination. Alternatively the TATA-motif around nucleotide 4075 could represent the late promoter.

It is noteworthy that the promoter region for early transcription is located 58-64 nucleotides downstream from the HpaI cleavage site. It has previously been shown that the HpaI/BamHI fragment which covers the entire early region including the TATA-box is unable to transform cells in tissue culture (18). Since the HindIII/BamHI fragment, in contrast, is positive for transformation this may indicate that either an enhancer sequence element or

Table 1

Predicted amino acid compositions of the L1 and L2 polypeptides[1)]

Amino acid	Polypeptide L1	L2	VP1[2)]
Lys	35	14	35
His	10	9	9
Arg	23	25	19
Asp	24	26	42 (Asx)
Asn	26	8	
Thr	36	50	34
Ser	32	49	52
Glu	28	28	64 (Glx)
Gln	25	14	
Pro	31	38	25
Gly	32	42	64
Ala	33	33	35
Cys	8	2	6
Val	29	31	28
Ile	24	26	18
Leu	45	43	36
Phe	20	10	14
Tyr	18	17	12
Trp	8	3	4
Met	8	1	4
	495	469	501
Molecular weight	55.5K	50.0K	53.5K[3)]

1) Only triplets that follow the putative initiation ATGs were considered. The entire reading frame L1 contains 499 codons and the entire L2 contains 474 codons.

2) Data from Meinke and Meinke (19).

3) Estimated by SDS polyacrylamide gel electrophoresis

the replication origin is located between the HindIII and the HpaI cleavage sites (nucleotides 6958-7945/0).

A sequence comparison between two BPV1 isolates. While this work was in progress Chen et al.(2) reported the complete nucleotide sequence of another BPV1 isolate. Their isolate (307) was collected from a cow in Wisconsin approximately 30 years ago and the virus was kept frozen until the genome was cloned recently (Olson, personal communication). Our isolate was collected from a Swedish cow in 1981. The two isolates are thus clearly of different origin and it is therefore interesting to determine the degree of nucleotide conservation between the two genomes. A total of five differences, three of which are transitions, were noticed (Table 2) when the 4807 nucleotides long sequence, established in the present study, was compared to the revised BPV1 sequence (9). Two of these changes, one of which is an insertion of a single base pair, appear to be located in non-coding regions (nucleotides 4022 and 7761 in Fig.2) whereas two of the differences, which both represent point mutations, are found in reading frame L2. One of the latter two differences

Table 2

SEQUENCE DIFFERENCES BETWEEN TWO ISOLATES OF BPV1[1)]

POSITION	NUCLEOTIDE		AMINO ACID		READING FRAME
	strain "Wisconsin"	strain "Sweden"	strain "Wisconsin"	strain "Sweden"	
4022	A	G	-	-	Non coding
4922	C	T	Ala	Val	L2
5340	T	A	No change	No change	L2
7761	-	C inserted	No change	No change	Non coding
138	T	C	No change	No change	E6[2)]

[1)]The comparison includes nucleotides 3837 to 698 (4807 in total) in the sequence of Chen et al. (2).
[2)]As defined by Danos et al. (9).

leads to an amino acid substitution whereas one change is silent (Table 2). Finally one mutation (nucleotide 138) is located in reading frame E6 of the early region (9). This mutation occurs in the third position of a codon for Asp and is silent. The sequences are thus remarkably well conserved.

A sequence comparison between the late regions of BPV1 and HPV1a. Danos et al. (1) have recently reported the complete nucleotide sequence for the HPV1a genome. A comparison between the sequence established in this study and the corresponding HPV1a sequence reveals striking similarities. Two long open reading frames L1 and L2 are present in the HPV1a sequence as well, although these are located in different phases of the sequence (9).

The region which covers reading frame L1 is particularly well conserved, having a sequence homology of 55% when the alignment of the two sequences was optimized. A comparison at the amino acid level reveals even more striking similarities; 253 or 51% of the residues are identical whereas no less than 72% of the residues are homologous if the amino acids are grouped with regard to their functional properties. Several identical hexa-, hepta-, octa-, and nonapeptides are present in the two sequences as shown in Fig.3A. When only residues which are known to be of strategic importance in protein sequences were considered, the comparison reveals additional similarities; most of the cysteine residues are located in precisely equivalent positions which also is the case for proline, and particularly tryptophane residues as shown in Fig. 4B. Silent nucleotide changes in the third codon position are frequent in reading frame L1 (120 of 253 homologous amino acids or 47%).

The homology between reading frames L2 in the BPV1 and HPV1a genomes is less striking both at the amino acid and nucleotide sequence levels. A

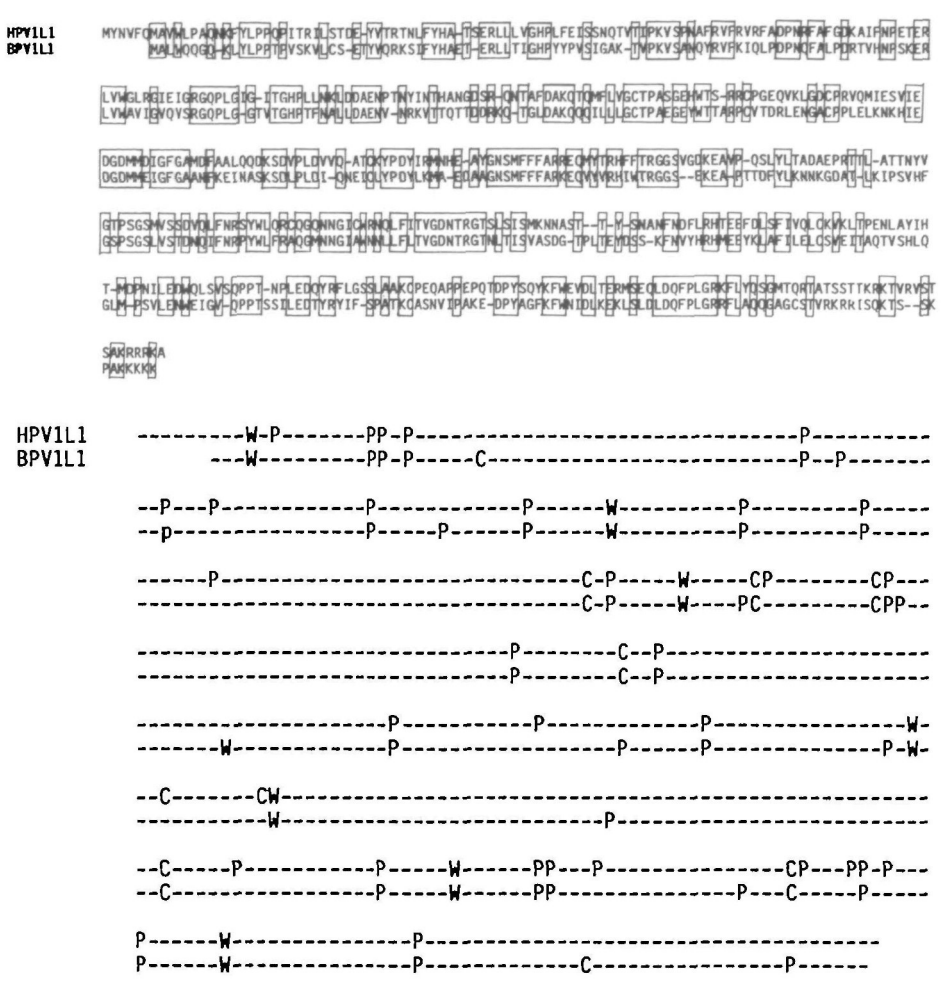

Figure 4A: A comparison between the predicted L1 proteins from HPV1a and BPV1. Homologous amino acid sequences are boxed.
Figure 4B: A comparison of the amino acid sequences of the predicted L1 proteins from HPV1a and BPV1. Only proline, tryptophane, and cysteine residues are indicated. The remaining residues are indicated with dashes.

comparison of the amino acid sequences is shown in Fig. 5. The homology was estimated to be 32% for the amino acid sequence and 34% for the nucleotide sequence. The L2 proteins are best conserved in the N-terminal part (Fig. 5) and identical penta-, hepta and octa peptides were noticed.

Meinke and Meinke (19) have determined the amino acid composition of

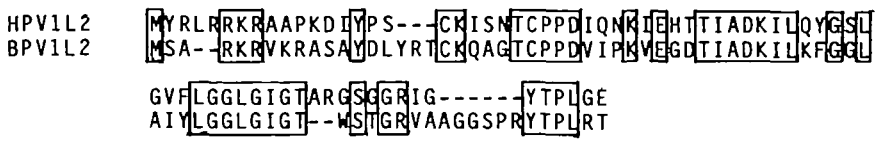

Figure 5: A comparison between the predicted amino acid sequences for the L2 proteins from HPV1a and BPV1. Homologous amino acids are boxed. Only the N-terminal part is shown since no significant homology was detected beyond the first 80 amino acids.

the VP1 protein from BPV1 and Table 1 shows a comparison between the L1, L2 and VP1 proteins. The estimated molecular weight of the VP1 protein (53.5K) agrees well with the predicted molecular weight of the L1 protein (55.5K). The amino acid composition of the VP1 protein is somewhat more related to that of the L1 protein than to that of the L2 protein. The predicted compositions of both the L1 and L2 proteins deviate, however, from that of the VP1 protein.

The homology between the non-coding parts of the BPV1 and HPV1a sequences was found to be even less than for the L2 regions.

ACKNOWLEDGEMENTS

We are greatly indepted to Marianne Gustafson and Ingegerd Schiller for excellent secretarial help. 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.

*To whom correspondence should be sent

REFERENCES

1. Danos, O., Katinka, M & Yaniv, M. (1982) EMBO J. 1,231-236.
2. Chen, E., Howley, P., Levinson, A. and Seeburg, P. (1982) Nature 299,529-534.
3. Shope, R.E. and Hurst, W. (1933) J.Exp.Med. 58, 607-624.
4. Lancaster, W. and Olson, C. (1982) Microbiol. Rev. 46, 191-207.
5. Heilman, C., Law, M.-F., Israel, M. and Howley, P. (1980) J. Virol. 36,395-407.
6. Howley, P., Law, M.-F., Heilman, C., Engel, L., Alonso, M., Israel,M. and Lowy, D. (1980) In "Viruses in Naturally Occurring Cancers" M. Essex, G. Todaro and H. zur Hausen, editors. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. pp. 233-247.
7. Law, M.-F., Lancaster, W. and Howley, P. (1979) J. Virol. 32, 199-207.
8. Kremsdorf, D., Jablonska, S., Favre, M., Orth, G. (1982) J.Virol. 43, 436-447.

9. Danos, O., Engel, L., Ellson, C., Yaniv, M. and Howley, P. (1983) J. Virol. in press.
10. Moreno-Lopez, J., Pettersson, U., Dinter, Z. and Philipson, L. (1981) Virology 112,589-595.
11. Stenlund, A., Perricaudet, M., Tiollais, P. and Pettersson, U. (1980) Gene 10, 47-52.
12. Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65,499-560.
13. Favaloro, J., Treisman, R. and Kamen, R. (1980) Methods Enzymol. 65, 718-749.
14. Brawerman, G., Mendecki, J. and Lee, S.Y. (1972) Biochemistry 11, 637-641.
15. Akusjärvi, G., Mathews, M.B., Andersson, P., Vennström, B. and Pettersson, U. (1980) Proc.Natl.Acad.Sci. 77, 2424-2428.
16. Amtmann, E. and Sauer, G. (1982) J.virol. 43, 59-66.
17. Heilman, C.., Engel, L., Lowy, D. and Howley, P. (1982) Virology 119, 22-34.
18. Lowy, D., Dvoretzky, I., Shober, R., Law, M.-F., Engel,L. and Howley, P. (1980) Nature 287, 72-74.
19. Meinke, W. & Meinke, G. (1981) J. Gen. Virol. 52,15-24.