Epidermodysplasia Verruciformis-Associated Human Papillomavirus 8: Genomic Sequence and Comparative Analysis

PAWEL G. FUCHS, THOMAS IFTNER, JÜRGEN WENINGER, AND HERBERT PFISTER*
Institut für Klinische Virologie der Universität Erlangen-Nürnberg, D-8520 Erlangen, Federal Republic of Germany

Received 6 October 1985/Accepted 4 February 1986

Human papillomavirus (HPV) 8 induces skin tumors which are at high risk for malignant conversion. The nucleotide sequence of HPV8 has been determined and compared to sequences of papillomaviruses with different oncogenic potential. The general organization of the HPV8 genome is similar to that of other types. Highly conserved, genus-specific sequences were found in open reading frames (ORFs) E1, E2, and L1. In ORFs E6, E7, and L2, HPV8 is more distantly related, but it was possible to differentiate subgenera in which HPV8 belonged to the HPV1-cottontail rabbit papillomavirus group. Sequences within ORF E4 and part of ORF L2 are rather type specific. HPV8 stands out by several unique features: (i) the considerably reduced size of the noncoding region (397 base pairs), with a seemingly low potential for forming complex secondary structures; (ii) a cluster of putative promoter elements in the 3′ half of ORF E1; (iii) an RNA polymerase III promoter-like sequence close to the C terminus of ORF E2; and (iv) of particular interest, the homology between the putative protein encoded by ORF E4 and the Epstein-Barr virus nuclear antigen 2 protein, which may reflect similar mechanisms in virus-mediated transformation.

Papillomaviruses induce different types of benign tumors of skin and mucosa in humans and in animals (23). Some of these lesions are known to progress into malignant tumors. The papilloma-to-carcinoma sequence has been clearly demonstrated in classical experiments on domestic rabbits infected with the cottontail rabbit papillomavirus (CRPV, Shope virus), but the role of the infecting virus and the molecular mechanisms involved in malignant conversion still remain to be elucidated. Malignant conversion in humans has been observed with the macroscopic skin lesions of epidermodysplasia verruciformis patients, with condylomata acuminata, with laryngeal papillomas, and more recently with human papillomavirus (HPV)-associated dysplasias of the cervix uteri. Epidermodysplasia verruciformis is a lifelong-persisting skin disease (17). Patients suffer from multiple flat warts with different morphology, which, after years, tend to cover the entire body. About 30% of the patients develop carcinoma in situ or squamous cell carcinoma in primary lesions. The benign tumors have been shown to be infected by an exceptionally broad spectrum of specific papillomavirus types (9, 13). Of the 31 distinct HPV types currently recognized, as many as 16 were isolated from biopsies of epidermodysplasia verruciformis. An individual patient may be infected with up to six different HPV types (24). In the cancers, viral DNA persists extrachromosomally in high copy number (22, 24). However, there is usually only one virus type present. HPV5 or HPV8 was identified in about 90% of the cases, thus implicating these types as high-risk viruses (G. Orth, personal communication). A similar differentiation is possible for HPV types infecting the genital mucosa. HPV6, 10, 11, 16, 18, and 31 are prevalent in cervical benign tumors, whereas HPV16 stands out by its persistent occurrence in more than 50% of cervical cancers (H. zur Hausen and A. Schneider, in N. P. Salzman and P. M. Howley, ed., The Papovaviridae: the papillomaviruses, in press). A comparison of low-risk and high-risk viruses is likely to reveal properties which are relevant to differences in the oncogenic potential. There are still no commonly used experimental in vivo or in vitro systems for HPV, but the analysis of DNA sequences provides some insight into the molecular biology of these viruses. DNA sequence data on different papillomaviruses showed a very uniform organization of the viral genomes. Most of the open reading frames (ORFs) are located in similar positions and are of comparable size. Some of them also show considerable homology at the level of nucleotide and amino acid sequences (2, 7, 10, 11, 29, 30). All of the studied papillomavirus genomes also show a noncoding region of roughly 1 kilobase (kb). A functional analysis has been carried out only for bovine papillomavirus (BPV) 1, which induces fibropapillomas in its natural host and is able to transform fibroblasts of different species in vitro. In the nonproductive fibroma part of the bovine wart and in the transformed cells, only about 4 kb are transcribed which are regarded as coding for early viral functions (23). The individual ORFs are therefore designated E1 through E8. Transfection of fibroblasts with deletion and linker insertion mutants as well as cDNA expression vectors revealed the importance of the following ORFs for specific purposes: E1, for extrachromosomal replication; E7, for high-copy maintenance; E2, for trans-activation of transcription; and E6 and E5 for cell transformation (16, 28, 33, 37, 38).

The remaining 3 kb of the viral genome harbors two ORFs, L(leader)1 and L2, which code for the major structural protein and a minor capsid component, respectively (26).

The highest degree of homology between different papillomaviruses was observed within ORFs E1, E2, and L1, which code for proteins fulfilling basic functions in the viral life cycle. Heterogeneity in other genome regions is likely to reflect differences in both cell and tissue tropism and in oncogenic potential.

In this paper we present a comparative analysis of the nucleotide sequence of HPV8 which deserves special attention because of its association with human skin cancer.

MATERIALS AND METHODS
General procedures for HPV8 DNA. Cloning and preliminary characterization of HPV8 DNA has been described (25).
Isolation of plasmid DNA and subgenomic fragments, restriction mapping, and other standard manipulations were performed after Maniatis et al. (18).

**DNA sequencing.** A large set of overlapping DNA fragments was analyzed, providing multiple sequences on different regions of both DNA strands. DNA fragments with 5' protruding or blunt ends were labeled with [γ-32P]ATP (19), and those with 3' protruding ends were labeled with [α-32P]dATP (39). Mixtures of restriction fragments typically served as the substrate for end labeling. The labeled DNA fragments were usually tested with different restriction enzymes to optimize the end-separating cleavages. The fragments chosen to be sequenced were selectively cut out of the gel (agarose or acrylamide, of appropriate concentration), minced briefly in a glass homogenizer, and extracted with three 2-h changes of elution buffer (0.5 M NaCl, 5 mM EDTA, 0.1 M Tris hydrochloride [pH 8.0]). The pooled extracts were filtered through small glass filters inserted into punctured Eppendorf vials and then were ethanol precipitated.

Water-dissolved DNA samples were centrifuged to remove insoluble solids and were ethanol precipitated again. The pelleted DNAs were then submitted directly to the partial chemical cleavage procedure described by Maxam and Gilbert (19).

A set of 20, 8, and 6% urea gels was routinely run for each sequencing sample. In special cases the LKB Multiphor electrophoresis device (LKB, Bromma, Sweden) was used for long-range high-resolution sequencing.

**DNA sequence analysis.** DNA sequence assembly, analysis, and comparisons of different kinds were computer assisted. The version II of the University of Wisconsin Genetic Computer Group program (8) was used with a VAX 11/780 computer facility. The analysis were supported by EMBL, NIH (National Institutes of Health) and NBRF (Georgetown University) DNA/protein data banks.

**RESULTS**

**General properties of the HPV8 genome.** The nucleotide sequence of HPV8 is presented in Fig. 1. The genome consisted of 7,654 base pairs (bp). G+C content was 42.96%. Fig. 2 gives an overview of the genome and shows the distribution of translational termination codons in six possible reading frames. All major ORFs were located on one DNA strand except for an ORF at position 3,882 to 3,208 which lacked an ATG initiation codon. On the putative sense strand, seven ORFs could be identified, which were designated E1 to E7 and L1 to L2 according to the currently accepted nomenclature for other papillomaviruses (5). The characteristics of these ORFs are shown in Table 1. With the exception of E4, they had ATG codons close to their 5' ends (Fig. 2) which could serve as start points of translation.

Only 397 bp of the genome between ORFs L1 and E6 were noncoding. Numerous A-T-rich sequences could be observed within this region. One 35-bp run contained only A and T residues, which were partially arranged as repeats. The rest of the area was surprisingly poor in repeat elements when compared to other papillomaviruses. In contrast, a cluster of six direct and five inverted repeats of considerable length was observed in the DNA sequences coding for ORF E4.

**Transcription control and RNA processing signals.** By using the known consensus sequences, the HPV8 DNA was screened for signals possibly involved in transcription control and RNA processing (Fig. 2). A number of potential RNA polymerase II promoter sequences were disclosed. Two such elements resided at the beginning of the early region: a group of TATA sequences (position 125 to 157), which surrounded the 5' end of E6 (position 136), and a single TATA motif (TATTAG), which appeared beyond the only ATG codon of E6 (position 220). Both presumable promoters were supplemented by upstream CAT boxes. The first (around position 73) was highly reminiscent of homologous CAT-box regions of other papillomaviruses. At least four putative promoter sequences were located within the ORF E1. The sequence at position 2,162 was the most canonical of the four and has numerous equivalents among eucaryotic promoters. Close to the C-terminal part of ORF L1, two additional promoter-like elements were found.

The sequence AATAAA, known to be involved in control of mRNA polyadenylation, was located at the 5' end of E6 and in N-terminal parts of L2 and L1 (around position 136 and at position 4,371 and 6,223, respectively). They were followed by possible poly(A) addition sites (CAA) within a distance range of 6 to 20 nucleotides. Further downstream, motifs like TG dinucleotide clusters, oligo T stretches, or the sequence CACTG provided additional evidence that the signals may be indeed active (1). A possible promoter of RNA polymerase III was discovered between positions 3,959 and 4,031. The HPV8 sequences TGGTTCGGAGG GAG and GTTCAAA which are 53 bp apart corresponded almost perfectly to boxes A and B of the consensus of tRNA gene promoters, as recently given by Sharp et al. (S. Sharp, J. Schack, L. Cooley, D. Johnson-Burke, and D. Soll, Crit. Rev. Biochem, in press). T-rich sequences could be found between 218 and 329 bp downstream of the B box, which may eventually serve as transcription terminators.

The positions of possible splice donor and splice acceptor sequences were determined by screening the HPV8 genome for the consensus sequences established by Mount (20). There were three splice donor sites in good agreement with this consensus at positions 966, 4,088, and 7,510. The quality of acceptor sites was not as easy to evaluate as in the case of donor sites. Figure 2 shows a selection of sites which are close to the consensus sequence. Most of them appeared in genome regions which were homologous to active splice acceptor sites of BPV1, CRPV, or both (6, 21, 35, 37). Early region. The early domain of HPV8 was made up of ORFs E6, E7, E1, E2, and E4 and represented 52% of the genome. There were no equivalents to ORFs E3, E5, and E8 of BPV1. At the beginning of ORF L2, one of the parallel reading frames coded for amino acid sequences which showed homology with translations of ORF E5 from CRPV and HPV1. The relationship to HPV1 was mainly confined to the area upstream of its ATG codon. The HPV8 frame was interrupted with several stop codons and therefore could not give rise to a protein.

The 5' part of the early region in HPV8 contained the ORFs E6 and E7, both moderately conserved among known papillomaviruses (Table 2). A rather close relationship existed between HPV8 and BPV8. A striking feature of E6 and E7 proteins was the precise spacing of the putative C-X-C sequence motif. HPV8 followed this rule, showing, however, an additional repeat in the N-terminal region of E6. The homology observed between E6 of CRPV and the family of ATP synthases (10) seemed to be very poor in the case of HPV8.

Most of the early region of HPV8 was occupied by ORFs E1 and E2. It is well recognized that the C-terminal moieties of E1 as well as N and C termini of E2 are highly conserved among different papillomaviruses, and this also held true for HPV8 (Table 2). A relationship has been noted between the C termini of the putative E2 proteins of various papil-
FIG. 1. The nucleotide sequence of HPV8 DNA sense strand (5' → 3'). The first A of an HpaI recognition site in the noncoding region was arbitrarily chosen as base 1 for numbering purposes.
Promoter elements are presented above the drawing. Symbols: ○, promoter sequence; □, polyadenylation signal; ♦, splice donor; ◇, splice acceptor; RNA polymerase possible reading frames (vertical bars above the lines). Major ORFs. Some chosen restriction sites and control signals present. Positions of termination codons (vertical bars below the lines) are shown. Figure 2. Genome organization of HPV8. The lower panel shows the distribution of termination codons. The upper panel shows the lengths representing gene lengths. 69,000 bp. 7,000 bp. 6,000 bp. 5,000 bp. 4,000 bp. 3,000 bp. 2,000 bp. 1,000 bp. (d) 0.1 cm.
TABLE 1. ORFs in HPV8 genome

<table>
<thead>
<tr>
<th>ORF</th>
<th>Sequence position (nucleotides)</th>
<th>Size (amino acid residues)</th>
<th>Coding capacity (amino acid residues)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total From 1st methionine</td>
</tr>
<tr>
<td>E6</td>
<td>136–660</td>
<td>525</td>
<td>175 155</td>
</tr>
<tr>
<td>E7</td>
<td>572–961</td>
<td>390</td>
<td>130 103</td>
</tr>
<tr>
<td>E1</td>
<td>783–2,759</td>
<td>1,977</td>
<td>659 603</td>
</tr>
<tr>
<td>E2</td>
<td>2,680–4,197</td>
<td>1,518</td>
<td>506 498</td>
</tr>
<tr>
<td>E4</td>
<td>3,266–3,952</td>
<td>687</td>
<td>229 no ATG</td>
</tr>
<tr>
<td>L2</td>
<td>4,269–5,834</td>
<td>1,566</td>
<td>522 518</td>
</tr>
<tr>
<td>L1</td>
<td>5,758–7,392</td>
<td>1,635</td>
<td>545 542</td>
</tr>
</tbody>
</table>

TABLE 2. Overall homology between ORFs of HPV8 and other papillomaviruses

<table>
<thead>
<tr>
<th>Papillomavirus</th>
<th>Homology* with HPV8 ORF:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E6</td>
</tr>
<tr>
<td>HPV1A</td>
<td>0.315</td>
</tr>
<tr>
<td>CRPV</td>
<td>0.219</td>
</tr>
<tr>
<td>HPV16</td>
<td>0.207</td>
</tr>
<tr>
<td>HPV6B</td>
<td>0.194</td>
</tr>
<tr>
<td>BPV1</td>
<td>0.175</td>
</tr>
</tbody>
</table>

* Comparisons of the ORFs from HPV8 with those from other papillomaviruses. The homology was analyzed with program GAP (gap weight = 2.0, gap length weight = 0.15) and given as ratio value by the method of Devereux et al. (8). The ratio of two identical sequences aligned without gaps = 1.0.

The nonhomologous central part of E2 coincided with the position of ORF E4. This reading frame showed almost no homology with its counterparts in other papillomaviruses (Table 2). The overall homology of E4 with HPV8 ORF L1 was 38%, whereas in two further areas, the three proteins shared unique sequences. It should be noted that most of the gaps which were introduced in E4 for optimal alignment with EBNA2 coincided with regions of high variability between both EBNA proteins. Further screening of the available data banks revealed no homologies of comparable quality.

Late region. The late region of HPV8 contains two large ORFs, L2 and L1. The putative polypeptide of L1 follows the rule of well-conserved sequences observed among all papillomaviruses (Table 2). Much more complicated relations could be observed within ORF L2 (Fig. 4). Besides

repeats, whereas in two further areas, the three proteins shared unique sequences. It should be noted that most of the gaps which were introduced in E4 for optimal alignment with EBNA2 coincided with regions of high variability between both EBNA proteins. Further screening of the available data banks revealed no homologies of comparable quality.

Late region. The late region of HPV8 contains two large ORFs, L2 and L1. The putative polypeptide of L1 follows the rule of well-conserved sequences observed among all papillomaviruses (Table 2). Much more complicated relations could be observed within ORF L2 (Fig. 4). Besides

repeats, whereas in two further areas, the three proteins shared unique sequences. It should be noted that most of the gaps which were introduced in E4 for optimal alignment with EBNA2 coincided with regions of high variability between both EBNA proteins. Further screening of the available data banks revealed no homologies of comparable quality.

Late region. The late region of HPV8 contains two large ORFs, L2 and L1. The putative polypeptide of L1 follows the rule of well-conserved sequences observed among all papillomaviruses (Table 2). Much more complicated relations could be observed within ORF L2 (Fig. 4). Besides

repeats, whereas in two further areas, the three proteins shared unique sequences. It should be noted that most of the gaps which were introduced in E4 for optimal alignment with EBNA2 coincided with regions of high variability between both EBNA proteins. Further screening of the available data banks revealed no homologies of comparable quality.
generally conserved N and C termini consisting of about 80 and 10 amino acids, respectively, there were four patches of genus-specific sequences in the middle of L2 proteins. Almost no relatedness appeared within the C-terminal third of the ORF, which must be therefore regarded as type-specific. The homology pattern between coordinates 80 and 400 (Fig. 4) suggests subdivision of the analyzed papillomaviruses into three groups: (i) HPV8, HPV1, and CRPV; (ii) HPV6 and HPV16; and (iii) BPV1 and deer papillomavirus (DPV1).

**DISCUSSION**

Sequencing data on HPV8 revealed a viral genome showing a number of specific features projected onto a basic structure, common to all papillomaviruses. With its 7,654-bp genome, HPV8 was relatively small. This was due to the considerably reduced size of the noncoding part of the genome (397 bp as compared to 650 to 900 bp in other papillomaviruses). The small size seemed to be a consistent feature of HPV8. The DNA length of an HPV8 variant was 7,600 bp, as determined by electron microscopy (14). By using restriction enzyme analyses and partial sequencing, we found a noncoding region of comparable size in a third independent isolate (unpublished data). In contrast to other papillomaviruses, the noncoding region of HPV8 seemed to have only a limited potential to form complex secondary structures. Nevertheless, a subclone carrying the short noncoding region of HPV8 persisted extrachromosomally over at least 40 generations in C127 and HeLa cells (P. G. Fuchs, unpublished data). This finding points to a functional integrity of the origin of replication, supposed to be located within the noncoding region of papillomaviruses (15, 36). In this context, it may be noteworthy that HPV8 sequences at position 173 to 193 show similarity with plasmid maintenance sequences 1 and 2 of BPV1 (15). In contrast to the noncoding part of the genome, all ORFs of HPV8 were larger than their counterparts in other papillomaviruses.

A number of interesting features could be detected when looking at the transcription control signals of HPV8. Most surprisingly, a cluster of promoter elements was found within ORF E1, which may define a specific transcription unit for the downstream part of the early region. A single promoter consensus sequence of similar location is known in HPV16 (30). There were also some indications for a functional promoter in E1 of BPV1, which is, however, without manifestation at the sequence level (2, 35). Conserved promoter elements at the beginning of ORF E6 were shown to represent the major early promoter of BPV1 and CRPV. There was a homologous CAAT motif with 18 nucleotides at position 59 in front of the first methionine of HPV8 E6. No canonical TATA box consensus sequence could be detected, but TATA-like elements appeared between positions 124 and 147 in front of the first ATG codon of E6 which are likely to function as an early promoter of HPV8. Sequences which were shown to represent active promoters of the adenovirus E1A gene, of SV40 early genes, and of Rous sarcoma virus large terminal repeat (31) were found 20 nucleotides downstream of the only HPV8 E6 ATG codon. Such a promoter element could give rise to a message coding for ORF E7, for example. Two promoter consensus sequences without a precedent in other papillomaviruses were located close together within ORF L1. It is interesting to note that the distance from L1 promoters to the beginning of E6 is roughly comparable to the same distance in the case of the second noncoding region promoter of BPV (2). One might speculate that the HPV8 promoter elements are pushed into L1 as a result of the short noncoding region.

A peculiar feature of HPV8 was the occurrence of RNA polymerase III (PolIII) promoterlike sequences within ORF E2. Adenoviruses and EVB are known to code for small RNAs originating from PolIII transcription units. Virus-associated RNAs from adenovirus type 2 were shown to be involved in regulation of protein biosynthesis (27). Preliminary attempts to demonstrate PolIII transcripts in vitro with HeLa cell-derived extract and cloned DNAs of HPV 1, 6, 8, 16, and 18 yielded negative results. However, this approach was subject to many experimental restrictions. Further experimental data are necessary to evaluate the relevance of polymerase III promoter sequences in HPV8.

The typically located AATAAA signal at the beginning of ORF L2 is likely to control the polyadenylation of early transcripts. The putative polyadenylation signal for late messages appeared intermingled with the presumable major early promoter. This unusual position could again be seen in connection with the short noncoding region of HPV8. There are no examples so far for termination at the beginning of L1, and the role of the corresponding signal of HPV8 remains to be established.

As far as RNA processing is concerned, consistent splice donor and acceptor sites within ORFs E1 and E4, respectively, could account for splicing events comparable to BPV1 and CRPV (6, 21, 35, 37). There are no obvious signals in HPV8, which would allow for the E6-E7 fusion protein as in the case of BPV1 (37).

Patterns of homology between putative proteins from HPV8 and other papillomaviruses followed the typical schema. ORFs E1, E2, and L1 revealed rather genus-specific sequences, reflecting their involvement in basic viral functions such as replication, gene expression, and capsid formation. The homology within E1 extended to described similarities with the large T antigen of polyomaviruses (3), suggesting that ORF E1 of HPV8 also has nucleotide binding properties. ORFs E6, E7, and L2 revealed only a scantly framework of genus-specific elements. However, the areas of low homology were of special interest since they were likely to code for type-specific functions. Based on the relationship of putative E6, E7, and L2 proteins, the sequenced papillomaviruses could be classified in three groups, HPV1, HPV8, and (less clearly) CRPV, HPV6, and HPV16, and finally, BPV1 and DPV. This grouping obviously correlated with the tissue specificity of the viruses. HPV1, HPV8, and CRPV infect the skin, HPV6 and HPV16 are specific for mucosa, and BPV1 induces fibropapillomas (23). The correlation might suggest that E6, E7, and L2 code for functions which are important for virus-target cell interaction. The greatest divergence in coding regions of papillomaviruses occurred within ORF E4 and in the C-terminal third of ORF L2. This indicates that these regions are subject to heavy selective pressure, which may be exerted by the host’s immune system or by intracellular control systems.

Certainly the most intriguing question to ask with regard to HPV8 is whether there are any special sequence properties potentially connected with the frequent malignant conversion of skin lesions induced by this virus. Genetic analysis of BPV1 demonstrated the involvement of ORFs E6 and E7 in transformation of mouse fibroblasts (28, 38). These findings cannot necessarily be extrapolated to transformation of epithelial cells and malignant conversion in vivo. Indeed, there is no E5 in HPV8. Further, as pointed out above, the relationship within ORF E6 correlates with tropism rather than with the assumed oncogenic potential. In the latter case, one would expect HPV8, HPV16, and CRPV to form a separate group. The homology between the puta-
tive protein of ORF E4 and the EBNA2 antigen of EBV can be of particular interest. EBV regularly persists in human anaplastic nasopharyngeal carcinomas, which represent epithelial cell-derived malignancies (40). The EBNA2 protein has been so far shown to be necessary for immortalization of B lymphocytes by EBV (32). It will be interesting to unravel the activities of E4 and EBNA2 proteins in epithelial cells and to see whether the sequence homologies reflect similar biological properties with regard to cell transformation.

ACKNOWLEDGMENTS

We gratefully acknowledge the cooperation of P. J. Farrell in running and interpreting some special computer programs. We are also indebted to E. Kieff for most encouraging discussions.

This work was supported by Deutsche Forschungsgemeinschaft (SFB 118).

LITERATURE CITED