Human Papillomavirus Types 6 and 11 mRNAs from Genital Condylomata Acuminata

LOUISE T. CHOW,^{1*} MOHAMMAD NASSERI,^{1†} STEVEN M. WOLINSKY,^{2‡} AND THOMAS R. BROKER¹

Biochemistry Department¹ and Department of Medicine,² University of Rochester School of Medicine, Rochester, New York 14642

Received 10 March 1987/Accepted 28 April 1987

We have identified and mapped a number of RNA species of human papillomavirus types 6 and 11 from condylomata acuminata by the electron microscopic R-loop technique. Each of the early (E)- and late (L)-region open reading frames (ORFs) deduced from the DNA sequences was represented in one or more transcripts. In addition, RNA species that could encode the modulator of DNA replication and the repressor of transcription, functions recently identified in the genetically similar bovine papillomavirus type 1, were also detected. Some ORFs were 5' proximal in one or more transcripts, whereas others were not 5' proximal in any species, suggesting that internal initiation of translation might be required to gain access to these latter ORFs. Virtually all transcripts had their 5' ends located in the E region and were polyadenylated at one of two sites, i.e., at the end of the E region or at the end of the L region. The great majority of the RNAs were derived from the E region of the genome, with one species approximately 50 to 100 times more abundant than the others. For most of the RNAs, the 5' end mapped near nucleotide 700; minor populations had 5' ends near nucleotide 100 or 1200. By correlating our mapping data with the genomic DNA sequences as well as available RNA structures and cDNA sequences of several papillomaviruses, we predict a number of mRNA splice donor and acceptor sites and suggest that the papillomaviruses have sophisticated usage of ORFs through alternative promoters, mRNA splice sites, and polyadenylation sites.

Papillomaviruses are a large family of related viruses that appear to have evolved from a common ancestor. They infect humans and animals in a host- and tissue-specific manner. The circular double-stranded genomes of about 7,900 base pairs from more than 45 types of human papillomaviruses (HPVs) have been molecularly cloned from various epithelial lesions (for a review, see reference 6). Of particular interest are the mucosa-tropic HPVs that are associated with venereal diseases: HPV types 6 (HPV-6), 11, 16, 18, 31, and 33 (3, 5, 16, 18, 20, 29). Reported clinical cases of genital warts have increased dramatically in recent years and may affect more than 10% of the population. HPV-6 and HPV-11 are generally found in benign condylomata acuminata in which the HPV DNAs exist as monomeric plasmids. HPV-16, HPV-18, HPV-31, and HPV-33 have been shown to be associated with genital malignancies in which HPV sequences integrate or exist as multimeric plasmids (5, 18a, 41, 43). Despite their medical importance, very little is known about the life cycle of these viruses, for lack of a productive tissue culture system.

Although HPV-1, HPV-6b, HPV-11, HPV-16, HPV-18, and HPV-33 have been sequenced and the protein-coding open reading frames (ORFs) have been inferred (10-12, 15, 39, 40), our current knowledge about the functions of these various ORFs is largely extrapolated from genetic studies of bovine papillomavirus type 1 (BPV-1) which transforms certain mouse cell lines in vitro (19) and replicates extrachromosomally as plasmids (26), just as the viruses do in vivo. The BPV-1 genome can be divided into three intervals, i.e., early (E) region, late (L) region, and upstream regula-

tory region (URR; also known as the long control region or the noncoding region). The E region is required for cellular transformation, episomal replication in the transformed cells, and transcriptional control. The L region encodes virion structural proteins. The URR contains an origin of DNA replication, promoters, enhancers, and short 5' exons for certain messages (2, 9, 21; also, see review in reference 6; M. Botchan, personal communication). HPV proteins identified in vivo include HPV-16 E6 and E7 (1, 41), the HPV-1 major capsid protein encoded by the L1 ORF, the minor capsid protein encoded by L2 (23), and a number of related proteins recognized by antibodies generated against the HPV-1 E4 ORF expressed in *Escherichia coli* (5a, 17). The E4 ORF has no identifiable function in the BPV-1 in vitro transformation system (32, 37).

The RNAs of BPV-1 and cottontail rabbit papillomavirus (CRPV) have been studied in detail by S1 mapping, primer extension, electron microscopy of R loops, and cDNA cloning and sequencing (2, 13, 14, 31, 34, 42, 44; Botchan, personal communication). To date, only a few studies have been performed on HPV RNAs in human lesions (27, 28). The transcripts appear to be derived exclusively from the E region, and they usually are present in such small quantities that their structures cannot be determined. Recently, Smotkin and Wettstein (41) have characterized some E-region RNAs from integrated HPV-16 in a cancer, and Chow et al. (9) have identified a number of E- and L-region HPV-1 RNAs from plantar warts. Here we report a comprehensive study, by the electron microscopic R-loop technique, of RNA species in two cases of anogenital condylomata acuminata associated with HPV-6 or HPV-11.

MATERIALS AND METHODS

Symbols for representing ORFs in transcripts. The / indicates overlapping ORFs; the ^ indicates an RNA splice; and the - means that the transcript spans the ORFs indicated.

* Corresponding author.

† Present address: Department of Medical Microbiology and Immunology, Stanford School of Medicine, Stanford, CA 94305.

‡ Present address: Department of Medicine, Infectious Disease Unit, Northwestern University Medical School, Chicago, IL 60611.

FIG. 1. Graphic representation of RNAs from HPV-6 and HPV-11 associated with condylomata acuminata. The RNAs were mapped by the electron microscope R-loop method and aligned with the DNA sequences of the highly homologous HPV-6 and HPV-11 and the ORFs of HPV-6. The functions of the various ORFs were assigned entirely through analogy to BPV-1. Species a constituted 98 to 99% of the RNA seen in both preparations. Of the remaining 1 to 2% of the R loops, species f, h, k, and l were seen much more often than species b, c, d, e, g, i, j, m, and n in the HPV-6 RNA. In the HPV-11 preparation, species e and f were more abundant than species c, d, g, h, k, i, and n. Compared with species c and d, substantially more frequent HPV-11 R loops had random 5' and 3' ends within the E1 region, some of which extended into the L2 region. We suspect that these were fragmented primary transcripts or partially degraded mRNA or intron sequences. In both preparations, additional molecules of species f appeared to have their 5' ends shortened to variable extent, possibly by degradation. The rare species g could thus be a degradation product instead of a real message. Species b, i, j, and m were not found in the HPV-11 preparation. Beads represent the 5' ends of the RNAs, arrowheads represent the polyadenylation sites, and gaps signify the intron sequences spliced out of the RNAs.

Preparation of RNA. A large vulvar wart (condyloma acuminata) was processed for total RNA as described by Przybyla et al. (35). An anal wart from a male was disrupted in the frozen state with a Braun micro-dismemberator II (24) and was processed by the procedure described by Chirgwin et al. (7). Poly(A)⁺ mRNA was enriched by two passages through an oligo(dT)-cellulose column.

HPV DNA. HPV-11 DNA was excised from the plasmid received from Lutz Gissmann and recloned at its *Bam*HI site (at nucleotide [nt] 7072 in the L1 ORF) into a pBR322 derivative, resulting in a recombinant DNA with a total length of approximately 10,880 base pairs. HPV-6b DNA was received from Gissmann as two *Eco*RI-*Bam*HI fragments. They were ligated, permuted, and recloned at the *Nar*I site (at nt 7694, which was altered to a *Sal*I site) into a modified pFI plasmid (33), as described by Chow et al. (8). The total length of this plasmid was about 13,750 base pairs. Both plasmids were purified in CsCl-ethidium bromide gradients. The HPV-11 recombinant was linearized at its single *Sal*I site in the vector, located 275 base pairs from the cloning site, or 1,134 base pairs from nt 1 of HPV-11. The HPV-6b recombinant was linearized at its sole *Xho*I site in the vector, 180 base pairs from the cloning site, or 390 base pairs from nt 1 of HPV-6. Because of this asymmetrical placement of HPV DNAs in the vectors, R loops could be correlated unambiguously to the physical and genetic maps of the HPVs. The linearized HPV DNAs were extracted with phenol and ether, precipitated from ethanol, and redissolved in water.

R-loop formation and analyses. R loops were prepared with 5 to 10 μ g of poly(A)⁺ RNA and 0.8 μ g of linearized plasmid in 10 μ l of R-loop buffer containing 86% formamide, 0.1 M

HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid) buffer (pH 7.6), 0.01 M EDTA, and 0.4 M NaCl (8). This solution was incubated at 39°C for 18 to 20 h. It was then reacted with 1/7 volume of 40% glyoxal at 7 to 10°C for 10 to 15 min before dilution with 10 volumes of ice-cold water. Glyoxal reacts with G, C, and A and prevents the displaced single-stranded DNA in an R loop from renaturing to its complementary strand, with concurrent displacement of the hybridized mRNA (22). The R loops so treated were stable for many days. Samples of this diluted R-loop preparation were mounted in cytochrome *c* monolayers at room temperature onto electron microscopic grids in the presence of formamide, examined in a Zeiss 10 CA transmission electron microscope, and photographed at $\times 10,000$ magnification. The negatives were projected from a photographic enlarger, and the molecules were measured with a Numonics electronic planimeter. The RNAs were aligned with the HPV-6 or HPV-11 DNA sequences, and ORFs were deduced from the DNA sequences (15, 39).

RESULTS

R-loop mapping of mRNAs. The DNA from the vulvar condyloma was characterized by Southern blot hybridization to contain HPV-11 present mostly as monomeric plasmids in very high copy number (M. Nasser, R. Hirochika, T. R. Broker, and L. T. Chow, Virology, in press). Northern blot (RNA blot) hybridization of the poly(A)⁺ RNA from this lesion indicated the existence of one predominant HPV-11 RNA about 1.2 kilobases in length which was about 50 to 100 times more abundant than the other species detected (Nasser et al., in press). To characterize the HPV-11 RNAs

FIG. 2. Electron micrographs of R loops of HPV-11 RNAs. (A) Species a of Fig. 1; (B) species f; (C) species c. Small and medium arrowheads point to the 5' and 3' ends of the RNAs, respectively; large arrowheads point to the intervening sequences.

further, we have carried out R-loop analyses of the poly(A)⁺ RNA. R loops were prepared as described in Materials and Methods and were found with moderate frequency in the preparation. Indeed, more than 98% of them were of one species (more than 70 molecules were measured), with a single splice connecting a 5' 120-nt-long exon from the region near the boundary of the E7 and E1 ORFs (nt 700 to 820) to a 3' 1.1-kilobase exon from the region with overlapping E2 and E4 ORFs (nt 3300 to 4400) (Fig. 1, species a, and Fig. 2A). They were polyadenylated at the end of the E region. Of the remaining 1 to 2% of the R loops, practically all were also polyadenylated at the same site at the end of E region. The number of measurements for these infrequent species reported below represents a slight overestimation in the relative abundance of species with fewer measurements (Fig. 1 legend). In one rare species (seen only twice), the same short 5' leader segment was spliced to an exon which extended from the beginning of the E2 ORF near nt 2700 to nt 4400 (Fig. 1, species g). Two other minor species seen more frequently (17 and 19 molecules were measured, respectively) had a 700-base-long 5' exon (nt 100 to 800) which was spliced to either of the two above-mentioned 3' exons (Fig. 1, species e and f; Fig. 2B and 3B). Three molecules were seen which spliced a short 5' exon of 100 nt located entirely within the E1 region (nt 1200 to 1300) to the abundant 3' exon from nt 3300 to 4400 (Fig. 1, species h). A small number of unspliced RNAs had a 5' end around either nt 100 or nt 700 and a 3' end near nt 4400, where they were polyadenylated (Fig. 1, species c and d, and Fig. 2C). The shorter of the two could potentially encode a protein encompassing the entire E1 ORF. The longer one could encode E6. Alternatively, either or both could be unprocessed primary transcripts of the E-region RNAs. Additional RNAs in the E1 region are described in the legend to Fig. 1. Two molecules were seen which had a 5' end near the beginning of L2 and extended continuously through L1; they were polyadenylated near nt 7400 (Fig. 1, species n). Another species, seen only twice, had three exons as depicted in Fig. 1, species l (also see below). This latter species could potentially express the L1 ORF encoding the major capsid protein, by analogy to BPV-1 (2) and HPV-1 (8, 9). These data are summarized in Fig. 1.

An anal condyloma of a male was shown by Southern blot hybridization to contain HPV-6 DNA present as monomeric

episomes with an average of 20 copies per cell in the tissue (data not shown). Poly(A)⁺ RNA from this condyloma was also used to form R loops. The results were quite similar to those described for HPV-11 mRNA, with some differences in relative abundance of various minor species and with additional species detected (Fig. 1 legend). The major species (35 measurements; Fig. 1, species a) was identical to that observed in HPV-11 RNA and constituted more than 98% of the RNAs. However, hardly any long E1 transcripts were seen; instead, a new rare spliced species was found (three measurements). Its 5' exon spanned nt 700 to 1300, where it was spliced to the 3' exon from nt 3,300 to 4,400 (Fig. 1, species b, and Fig. 3A). Species h (Fig. 1), consisting of a 5' exon from nt 1200 to 1300 and a 3' exon from nt 3300 to 4000, was seen more frequently (11 measurements) than in the HPV-11 RNA preparation. Examples of species e and h of HPV-6 are presented in Fig. 3B and C, respectively. The putative L1 mRNA, species l of Fig. 1, was also observed much more frequently (17 measurements; Fig. 4A). The three exons were mapped to nt 720 to 820, 3300 to 3530, and 5800 to 7400. Rarely was the 5' proximal exon located from nt 100 to 800 (Fig. 1, species m). In addition, two new species were found which had only the 5' intron of the putative L1 mRNA removed, and their 5' exon was derived from either nt 700 to 800 (14 measurements; species k of Fig. 1 and Fig. 4B) or from nt 1200 to 1300 (seen once; species j of Fig. 1). Species i of Fig. 1 was seen once. It could potentially encode E5a or E5b protein. These data are summarized in Fig. 1.

Coding potentials of mRNAs. (i) Species a. The predominant species, in which mRNA splicing connected the short overlapping E7/E1 ORFs to the overlapping E2/E4 ORFs (species a of Fig. 1), has recently been identified in a cDNA library constructed from the same HPV-11 poly(A)⁺ RNA preparation studied in our work. Sequence determination of the cDNA has located the splice donor and acceptor sites at nt 847 and 3325, respectively (Nasseri et al., in press) (Table 1). The splice fuses the E1 and E4 ORFs such that 5 of 90 amino acids encoded are derived from the E1 ORF. The HPV-6 and HPV-11 DNA sequences are identical around these sites (15, 39). An HPV-1 RNA of similar splice pattern was found in plantar warts, and it was also the major RNA species detected by R-loop mapping (9). Putative splice donor and acceptor sequences are conserved in HPV-1 and

FIG. 3. Electron micrographs of R loops of HPV-6 RNAs. (A) Species b of Fig. 1; (B) species e; (C) species h. Small and medium arrowheads point to the 5' and 3' ends of the RNAs, respectively; large arrowheads point to the intervening sequences.

HPV-16 DNA (Table 1). S1 nuclease mapping of HPV-16 mRNAs from a cancer has identified a splice donor and an acceptor very near the predicted sites (41). Concordantly, the E4 protein of HPV-1 has recently been shown to be by far the most abundant protein in plantar warts (5a, 17). The equivalent splice donor and acceptor sites have been reported in BPV-1 and CRPV transcripts (14, 42, 44). However, in the animal papillomaviruses, the 5' exon contains the E6 and E7 ORFs or the E7 ORF in addition to the beginning of the E1 ORF. As a result, internal reinitiation would be required to translate the E1-E4 fusion protein. In each of the papillomaviruses examined, the 5' exon would contribute five or fewer amino acids (Table 1). Curiously, the E4 ORF has no identifiable function in the BPV-1 transformation of mouse cell lines (32, 37).

(ii) **Species b.** A splice donor at nt 1235 in BPV-1 is linked, in some transcripts, to the same acceptor as that used for the E1-E4 RNA (42). This donor splice site is equivalent to the one near nt 1300 which we found in HPV-6 and HPV-11 RNAs. Inspection of HPV-1, HPV-6b, HPV-11, HPV-16, BPV-1, and CRPV sequences revealed a highly conserved potential splice donor (5'..CAG/GTA..3') at these positions (Table 2). If the long exon of RNA species b (Fig. 1) preceding this donor site (nt 700 to 1300) is translated in the E1 ORF, the splice to the same acceptor site used for the E1-E4 mRNA would result in a shift of the reading frame and termination of translation after a relatively few amino acids. This would produce a truncated E1 protein. The amino acids deduced from the 3' exon varied a great deal among the papillomaviruses, suggesting that the function of this protein might not depend on the sequence of its C terminus. Termination of translation and polyadenylation of the mRNA are apparently the purposes of this splicing event. Genetic evidence in the BPV-1 system has suggested the existence of such a truncated E1 protein (E1-M) with a role in modulation of copy number during episomal replication (4, 30, 36).

FIG. 4. Electron micrographs of R loops of HPV-6 RNAs. (A) Species l of Fig. 1; (B) species k. Small and medium arrowheads point to the 5' and 3' ends of the RNAs, respectively; large arrowheads point to the intervening sequences.

TABLE 1. Splice sites for E1-E4 mRNAs^a

Papillo-mavirus	ORF	Donor	Acceptor	ORF
BPV-1	E1	.. <u>ATG</u> GCA AAC GAT AAA G/GT AGC AAT .. 864	.. ATT TTA G\AG ATC GCC .. 3225	E4
CRPV	E1	.. <u>ATG</u> GCT GAA G/GT ACA GAC .. 1371	.. TCA CCA G\CT CCC CCC .. 3714	E4
HPV-11	E1	.. <u>ATG</u> GCG GAC GAT TCA G/GT ACA GAA .. 847	.. TAT CTA G\CA CTG TAC .. 3325	E4
HPV-6	E1	.. <u>ATG</u> GCG GAC GAT TCA G/GT ACA GAA .. 847	.. TAT CTA G\CA CTA CAC .. 3325	E4
HPV-16	E1	.. <u>ATG</u> GCT GAT CCT GCA G/GT ACC AAT .. 880	.. TTA GCA G\CA ACG AAG .. 3360	E4
HPV-1	E1	.. <u>ATG</u> GCA GAT AAT AAA G/GT ACT GAA .. 827	.. CCA CTA G\CT CCC CAA .. 3200	E4

^a DNA sequences spanning the putative or established mRNA splicing donor and acceptor sites are shown with the position of the last nucleotide of the donor exon and first nucleotide of the acceptor exon listed. /, Splice donor site; \, splice acceptor site. ORFs for each upstream and downstream exon are indicated. Underscored ATGs are or are hypothesized to be initiation codons. ↓, Alternative possible splice acceptor.

(iii) **Species h.** The short exon in E1 from nt 1200 to 1300 (species h of Fig. 1) could potentially be translated in a frame other than the E1 ORF starting from a methionine conserved in all papillomaviruses followed by 10 or 12 additional amino acids. This frame is designated E8 in BPV-1 but is closed in other viruses. The splice to the same acceptor as used for the major E1-E4 species would shift the utilized frame to the E2 ORF, resulting in a fusion protein consisting predominantly of the carboxyl terminal 45% of the E2 protein (Table 3). This interpretation is consistent with the genetic evidence of Lambert et al. (25) that the carboxyl-terminal portion of the BPV-1 E2 ORF (E2-C) appears to be functionally distinct from the full-length E2 product. Unlike the E2 protein, which is an enhancer activator, E2-C suppresses transformation, presumably through repression of the transcriptional enhancer.

To summarize these analyses, the 5' exons in the E1

region that are derived from two putative promoters could potentially be translated in two different reading frames, whereas the 3' exon could be translated in all three reading frames to produce three proteins. This remarkable complexity is no doubt responsible for the absolutely conserved pattern of all the papillomaviruses, in which the E1, E2, and E4 ORFs are distributed one each into the three translation phases.

(iv) **Species l.** The splice acceptor for the L1 mRNA of BPV-1 has been identified as nt 5609 (2). Inspection of the sequences of several papillomaviral DNAs at the beginning of the L1 ORF, where we mapped the acceptor sites for HPV-6, HPV-11, and HPV-1 (8, 9), again revealed highly conserved sequences 5'..CAG/ATG..3'. We hypothesize that they are the splice acceptor sites (Table 4). The additional AUG codons in the upstream exons in the BPV-1 mRNA are apparently nonfunctional (2). As for HPV's,

TABLE 2. Splice sites for E1-M mRNAs^a

Papillo-mavirus	ORF	Donor	Acceptor	n ^b	
BPV-1	E1	.. CCC CTC CAG/GTA CAG GGG .. 1235	.. TGA TTT TAG\AGA TCG CCC .. 3225	(30)	.. TGA ..
CRPV	E1	.. TCG TTA CAG/GTA GAA TCG .. 1751	.. TGT CAC CAG\CTC CCC CCA .. 3714	(18)	.. TGA ..
HPV-11	E1	.. GCA ACG CAG/GTA GAG AAA .. 1272	.. TGT ATC TAG\CAC TGT ACG .. 3325	(15)	.. TGA ..
HPV-6	E1	.. GGA ACG CAG/GTA GAG AAA .. 1278	.. TGT ATC TAG\CAC TAC ACA .. 3325	(15)	.. TGA ..
HPV-16	E1	.. ATG TTA CAG/GTA GAA GGG .. 1301	.. GTT TAG CAG\CAA CGA AGT .. 3360	(9)	.. TGA ..
HPV-1	E1	.. TCG ACA CAG/GTA GAT CAA .. 1231	.. TTC CAC TAG\CTC CCC AAG .. 3200	(51)	.. TGA ..

^a The downstream exon is closed in each of the viruses, and the amino acids encoded are very dissimilar and vary from 6 to 20 in length before termination. DNA sequences spanning the putative or established mRNA splicing donor and acceptor sites are shown with the position of the last nucleotide of the donor exon and first nucleotide of the acceptor exon listed. /, Splice donor site; \, splice acceptor site. ↓, Alternative possible splice acceptor.

^b n, Additional nucleotides not shown.

TABLE 3. Splice sites for E2-C mRNAs^a

Papillo- mavirus	ORF	Donor																Acceptor				ORF
BPV-1	E8	.. <u>ATG</u> AAG CTA ACC GTG TTC TTA CGC CCC TCC AG/G TAC AGG . . . . GAT TTT AG/A GAT CGC CCA . .																3225				E2
CRPV	— ^b	.. <u>ATG</u> AAG CTG CTT CTG GTG CTG AGT CTG TAC AG/G TAG AAT . . . . GTC ACC AG/C TCC CCC CAG . .																3714				E2
HPV-11	—	. <u>ATG</u> GCT ATT CTG AAG TGG AAG CTG CAA CGC AG/G TAG AGA . . . . GTA TCT AG/C ACT GTA CGA . .																3325				E2
HPV-6	—	.. <u>ATG</u> GCT ATT CTG AAG TGG AAG CTG GAA CGG GAA CGC AG/G TAG AGA . . . . GTA TCT AG/C ACT ACA CAA . .																3325				E2
HPV-16	—	.. <u>ATG</u> GCA ATA CTG AAG TGG AAA CTC AGC AGA TGT TAC AG/G TAG AAG . . . . TTT AGC AG/C AAC GAA GTA . .																3360				E2
HPV-1	—	.. <u>ATG</u> AAA CTG AAA ATA TTG ATG AAT CGA CAC AG/G TAG ATC . . . . TCC ACT AG/C TCC CCA AGG . .																3200				E2

^a The amino acids encoded by the 5' exons show some homology. Underscored ATGs are hypothesized to be initiation codons. DNA sequences spanning the putative or established mRNA splicing donor and acceptor sites are shown with the position of the last nucleotide of the donor exons and first nucleotide of the acceptor exons listed. /, Splice donor site; \, splice acceptor site; ↓, Alternative possible splice acceptor.

^b —, Reading frame closed immediately after donor splice site.

TABLE 4. Acceptor sites for L1 mRNAs^a

Papillo- mavirus	Acceptor											
BPV-1	. . TTT TTT TTG CAG\ATG GCG TTG TGG CAA . 5609											
CRPV	. . TTT TAC TTG CAG\ATG GCA GTG TGG CTG . . 5828											
HPV-11	. . TAT TTT TTA CAG\ATG TGG CGG CCT AGC . . 5771											
HPV-6	. . TAT TTT TTT CAG\ATG TGG CGG CCT AGC . . 5789											
HPV-16	. . ATT TTT TTT CAG\ATG TCT CTT TGG CTG . . 5637											
HPV-1	. . AAT GTT TTT CAG\ATG GCT GTC TGG TTA . . 5431											

^a DNA sequences spanning the putative or established mRNA splicing acceptor sites are shown with the first nucleotide of the acceptor exons listed. Underscored ATGs are or are hypothesized to be initiation codons. \, Splice acceptor site.

processing of a fusion protein would be required to produce the major capsid protein of about 55 kilodaltons if translation begins in either of the two 5' exons. However, we prefer the alternative possibility that the methionine encoded at the acceptor splice site is the initiation codon for the L1 protein, for the upstream leaders are quite different in length and position for BPV-1, HPV-1, HPV-6, and HPV-11.

(v) **Other species.** A number of species could potentially be translated into proteins encoded by the E6, E1, and L2 ORFs because each was the 5'-most exon in one or more mRNA species (Fig. 1). However, no RNAs were consistently seen which appeared to have initiated at the beginning of E7, nor was there any RNA which had a splice within E6 as found with HPV-16 and HPV-18 (38, 41) or a splice connecting E6 to E7 as in BPV-1 (42, 44). Apparently, sequence divergence has destroyed the equivalent splice sites in HPV-6 and HPV-11 (38). Therefore, the E7 protein would have to be translated from some extremely rare species which we did not detect or from a polycistronic message which could be translated into both E6 and E7 (and perhaps other) proteins. Similarly, full-length E2 protein might be translated through internal initiation from an RNA which also contained E6 and E7 ORFs upstream (species f of Fig. 1). Virtually no RNA was seen with a splice to the E5 ORF. Internal reinitiation is most likely the mechanism for their translation. The BPV-1 E5 protein seems indeed to be translated from messages which also contain E2 and E4 ORFs (Y.-C. Yang, C. C. Baker, and P. M. Howley, personal communication). Species j, k, and m of Fig. 1 could either be intermediates for what we believe to be the mature L1 mRNA (species l of Fig. 1) or be the message for E5a, E5b, or L2. The existence of unspliced L2 mRNA (species n of Fig. 1) was less certain, as the molecules were extremely rare and could be a result of degradation from species j, k, or m of Fig. 1.

DISCUSSION

We have mapped a number of HPV-6 and HPV-11 RNAs extracted from anogenital condylomata. The structures of many of these RNAs were quite similar to those reported for BPV-1 (2, 42, 44) and for HPV-1 (9), but additional species

were identified in this study. One of the major differences between BPV-1 and the HPVs concerned the locations of the predominant 5' ends of the transcripts. In BPV-1, they are in the URR, just upstream from the E6 ORF and near nt 2400 and 3000. In contrast, in human warts, the major promoter is located in E7, close to the boundary with E1; minor promoters are just upstream of the E6 ORF and in the E1 ORF (Fig. 1). In this study, no transcripts were found to have 5' exons located in the URR or near nt 2400 or 3000, although a very minor fraction of HPV-1 RNAs from plantar warts appeared to originate from the URR (9). If additional 5' exons derived from the URR existed in the HPV-6 and HPV-11 RNAs, they must have been very short to have escaped detection. However, analysis of the E1-E4 cDNAs made from a sample of the same HPV-11 RNA preparation (Nasseri et al., in press) did not reveal any additional 5' exons and placed the 5' end at nt 716, in excellent agreement with the data presented here. In addition, Smotkin and Wettstein (41) uncovered no evidence of any URR exons in the HPV-16 E-region RNAs in the cancer they studied. The various RNA species we observed were quite similar in the two condylomata, which was not unexpected considering that the two viruses share 85% homology in their DNA sequences. The transcripts were polyadenylated at two different sites, at the ends of the E and L regions (Fig. 1).

There is no conventional AAUAAA polyadenylation-cleavage signal at the end of the E region of HPV-6 or HPV-11. However, the transcripts were polyadenylated there. Based on the HPV-11 cDNA sequence analysis, we believe that the hexanucleotide AGUAAA is a component of the variant signal used (Nasseri et al., in press). This signal is also used efficiently in monkey COS cells transfected with the homologous HPV-6b DNA, because only E-region mRNAs that are polyadenylated at similar locations were detected (8). The AAUAAA signal at the end of the L1 ORF near nt 7400 for both viruses was clearly used. There are other AAUAAA sequences in E1, E2, and L1 in the sequenced HPV-6b and HPV-11 DNAs. They did not appear to be functional (if they were also present in the HPVs associated with the lesions used in this study), because we saw no RNA ending specifically at those sites.

It was rather curious that, despite the high copy number of HPV-11 viral DNA in the lesions, more than 99.9% of the viral RNAs was polyadenylated at the end of the E region. Hardly any RNA encoding viral capsid proteins from the L region were seen. Nonetheless, this is in agreement with the general observation that mucosa-tropic HPVs produce very few virus particles. It is intriguing that the abundant E1-E4 mRNA and the rare L1 RNA had the same 5' end and thus presumably originated from the same putative promoter located in the E7 ORF. These results suggest that the regulation of late gene expression resides in the ability of the RNA polymerase II to transcribe through just one or both polyadenylation sites and in the utilization of one of the two sites. Additional control of gene expression by mRNA processing or transport is also possible.

In situ hybridization of serial thin sections of the HPV-6-associated anal condyloma tissue adjacent to the portion used for the RNA extraction leading to R-loop mapping showed only occasional and rather focal HPV-6 replication and transcription activity (data not shown). Accordingly, we believe that the electron microscope R-loop technique is extremely sensitive for mapping rare RNA species.

In summary, we have identified and mapped a number of RNA species from condylomata associated with HPV-6 and HPV-11. While some of the species are undoubtedly genuine

messages, others might be processing intermediates. Our data show that the majority of the RNAs have their 5' ends located in the E7 ORF. Minor species have their 5' ends situated just upstream of the E6 ORF, within the E1 ORF and just upstream from the L2 ORF. There are two polyadenylation sites, one at the end of the E region and the other at the end of the L region. Correlation of the RNA splice patterns and DNA sequences suggests that some ORFs can be translated only through internal reinitiation within polycistronic messages and that the viruses have sophisticated utilization of coding regions through alternative promoters and mRNA splicing.

ACKNOWLEDGMENTS

This research was sponsored by Public Health Service grant CA36200 from the National Institutes of Health and grant 1587 from the Council for Tobacco Research-USA to L.T.C. and T.R.B. and by Public Health Service institutional postdoctoral fellowship T32CA09363 to M.N. S.M.W. was a Wilmot Cancer Research Fellow supported by The James P. Wilmot Foundation and by American Cancer Society institutional research award IN-18.

We thank Richard Reid and Robert Graves for providing the condylomata and Lutz Gissmann for providing the cloned HPV-6 and HPV-11 DNAs and the sequence of HPV-11 DNA before publication.

LITERATURE CITED

1. Androphy, E. J., N. L. Hubbert, J. T. Schiller, and D. R. Lowy. 1987. Identification of the HPV-16 E6 protein from transformed mouse cells and human cervical carcinoma cell lines. *EMBO J.* 6:989-992.
2. Baker, C. C., and P. M. Howley. 1987. Differential promoter utilization by the bovine papillomavirus in transformed cells and productively infected wart tissues. *EMBO J.* 6:1027-1035.
3. Beaudenon, S., D. Kremsdorf, O. Croissant, S. Jablonska, S. Wain-Hobson, and G. Orth. 1986. A novel type of human papillomavirus associated with genital neoplasias. *Nature (London)* 321:246-249.
4. Berg, L., M. Lusk, A. Stenlund, and M. R. Botchan. 1986. Repression of bovine papilloma virus replication is mediated by a virally encoded trans-acting factor. *Cell* 46:753-762.
5. Boshart, M., L. Gissmann, H. Ikenberg, A. Kleinheinz, W. Scheurle, and H. zur Hausen. 1984. A new type of papillomavirus DNA and its presence in genital cancer biopsies and in cell lines derived from cervical cancer. *EMBO J.* 3:1151-1157.
- 5a. Breitburd, F., O. Croissant, and G. Orth. 1987. Expression of human papillomavirus type-1 E4 gene products in warts. *Cancer Cells* 5:115-122.
6. Broker, T. R., and M. Botchan. 1986. Papillomaviruses: retrospectives and perspectives. *Cancer Cells* 4:17-36.
7. Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 18:5294-5299.
8. Chow, L. T., A. J. Pelletier, R. Galli, U. Brinckmann, M. Chin, D. Arvan, D. Campanelli, S. Cheng, and T. R. Broker. 1986. Transcription of human papilloma virus types 1 and 6. *Cancer Cells* 4:603-614.
9. Chow, L. T., S. S. Reilly, T. R. Broker, and L. B. Taichman. 1987. Identification and mapping of human papillomavirus type 1 RNA transcripts recovered from plantar warts and infected epithelial cell culture. *J. Virol.* 61:1913-1918.
10. Cole, S. T., and O. Danos. 1987. Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. Phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products. *J. Mol. Biol.* 193:599-608.
11. Cole, S. T., and R. E. Streeck. 1986. Genome organization and nucleotide sequence of human papillomavirus type 33, which is associated with cervical cancer. *J. Virol.* 58:991-995.

12. Danos, O., L. W. Engel, E. Y. Chen, M. Yaniv, and P. M. Howley. 1983. Comparative analysis of the human type 1a and bovine type 1 papillomavirus genomes. *J. Virol.* **46**:557-566.
13. 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.
14. Danos, O., R. C. Mulligan, and M. Yaniv. 1986. Production of spliced DNA copies of the cottontail rabbit papillomavirus genome in a retroviral vector. *Ciba Found. Symp.* **120**:68-82.
15. Dartmann, K., E. Schwarz, L. Gissmann, and H. zur Hausen. 1986. The nucleotide sequence and genome organization of human papilloma virus type 11. *Virology* **151**:124-130.
16. de Villiers, E.-M., L. Gissmann, and H. zur Hausen. 1981. Molecular cloning of viral DNA from human genital warts. *J. Virol.* **40**:932-935.
17. Doorbar, J., D. Campbell, R. J. A. Grand, and P. H. Gallimore. 1986. Identification of the human papilloma virus-1a E4 gene products. *EMBO J.* **5**:355-362.
18. Dürst, M., L. Gissmann, H. Ikenberg, and H. zur Hausen. 1983. A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. *Proc. Natl. Acad. Sci. USA* **80**:3812-3815.
- 18a. Dürst, M., A. Kleinheinz, M. Hotz, and L. Gissmann. 1985. The physical state of human papillomavirus type 16 DNA in benign and malignant genital tumours. *J. Gen. Virol.* **66**:1515-1522.
19. Dvoretzky, I., R. Shober, S. K. Chattopadhyay, and D. R. Lowy. 1980. A quantitative in vitro focus assay for bovine papilloma virus. *Virology* **103**:369-375.
20. Gissmann, L., V. Diehl, H.-J. Schultz-Coulon, and H. zur Hausen. 1982. Molecular cloning and characterization of human papilloma virus DNA derived from a laryngeal papilloma. *J. Virol.* **44**:393-400.
21. Hirochika, H., T. R. Broker, and L. T. Chow. 1987. Enhancers and *trans*-acting E2 transcriptional factors of papillomaviruses. *J. Virol.* **61**:2599-2606.
22. Kaback, D. B., L. M. Angerer, and N. Davidson. 1979. Improved methods for the formation and stabilization of R-loops. *Nucleic Acids Res.* **6**:2499-2517.
23. Komly, C. A., F. Breitburd, O. Croissant, and R. D. Strebeck. 1986. The L2 open reading frame of human papillomavirus type 1a encodes a minor structural protein carrying type-specific antigens. *J. Virol.* **60**:813-816.
24. Krieg, P., E. Amtmann, and G. Sauer. 1983. The simultaneous extraction of high-molecular-weight DNA and of RNA from solid tumors. *Anal. Biochem.* **134**:288-294.
25. Lambert, P. F., B. A. Spalholz, and P. M. Howley. 1987. Evidence that bovine papillomavirus type 1 may encode a negative transcriptional regulatory factor. *Cancer Cells* **5**:15-22.
26. 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.
27. Lehn, H., T.-M. Ernst, and G. Sauer. 1984. Transcription of episomal papillomavirus DNA in human condylomata acuminata and Buschke-Löwenstein tumours. *J. Gen. Virol.* **65**:2003-2010.
28. Lehn, H., P. Krieg, and G. Sauer. 1985. Papillomavirus genomes in human cervical tumors: analysis of their transcriptional activity. *Proc. Natl. Acad. Sci. USA* **82**:5540-5544.
29. Lörinicz, A. T., W. D. Lancaster, and G. F. Temple. 1986. Cloning and characterization of the DNA of a new human papillomavirus from a woman with dysplasia of the uterine cervix. *J. Virol.* **58**:225-229.
30. Lusky, M., and M. R. Botchan. 1986. A bovine papillomavirus type 1-encoded modulator function is dispensable for transient viral replication but is required for establishment of the stable plasmid state. *J. Virol.* **60**:729-742.
31. Nasser, M., and F. O. Wettstein. 1984. Differences exist between viral transcripts in cottontail rabbit papillomavirus-induced benign and malignant tumors as well as non-virus-producing and virus-producing tumors. *J. Virol.* **51**:706-712.
32. Neary, K., B. H. Horwitz, and D. DiMaio. 1987. Mutational analysis of open reading frame E4 of bovine papillomavirus type 1. *J. Virol.* **61**:1248-1252.
33. Pelham, H. R. B. 1982. A regulatory upstream promoter element in the *Drosophila* hsp70 heat shock gene. *Cell* **30**:517-528.
34. Phelps, W. C., S. L. Leary, and A. J. Faras. 1985. Shope papillomavirus transcription in benign and malignant rabbit tumors. *Virology* **146**:120-129.
35. Przybyla, A. E., R. J. MacDonald, J. D. Harding, R. L. Pictet, and W. J. Rutter. 1979. Accumulation of the predominant pancreatic mRNAs during embryonic development. *J. Biol. Chem.* **254**:2154-2159.
36. Roberts, J. M., and H. Weintraub. 1986. Negative control of DNA replication in composite SV40-bovine papilloma virus plasmids. *Cell* **46**:741-752.
37. 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.
38. Schneider-Gädick, A., and E. Schwarz. 1986. Different human cervical carcinoma cell lines show similar transcription patterns of human papillomavirus type 18 early genes. *EMBO J.* **5**:2285-2292.
39. Schwarz, E., M. Dürst, C. Demankowski, O. Lattermann, R. Zech, E. Wolfspurger, S. Suhai, and H. zur Hausen. 1983. DNA sequence and genome organization of genital human papillomavirus type 6b. *EMBO J.* **2**:2341-2348.
40. Seedorf, K., G. Krämer, M. Dürst, S. Suhai, and W. Röwekamp. 1985. Human papillomavirus type 16 DNA sequence. *Virology* **145**:181-185.
41. Smotkin, D., and F. O. Wettstein. 1986. Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein. *Proc. Natl. Acad. Sci. USA* **83**:4680-4684.
42. Stenlund, A., J. Zabielski, H. Ahola, J. Moreno-Lopez, and U. Pettersson. 1985. Messenger RNAs from the transforming region of bovine papilloma virus type 1. *J. Mol. Biol.* **182**:541-554.
43. Tsunokawa, Y., N. Takebe, S. Nozawa, T. Kasamatsu, L. Gissmann, H. zur Hausen, M. Terada, and T. Sugimura. 1986. Presence of human papillomavirus type-16 and type-18 DNA sequences and their expression in cervical cancers and cell lines from Japanese patients. *Int. J. Cancer* **37**:499-503.
44. Yang, Y.-C., H. Okayama, and P. M. Howley. 1985. Bovine papillomavirus contains multiple transforming genes. *Proc. Natl. Acad. Sci. USA* **82**:1030-1034.