JOURNAL OF VIROLOGY, Apr. 1987, p. 962–971
0022-538X/87/040962-10$02.00/0

Vol. 61, No. 4

Structural and Transcriptional Analysis of Human Papillomavirus Type 16 Sequences in Cervical Carcinoma Cell Lines

CARL C. BAKER,[1]* WILLIAM C. PHELPS,[1] VALERIE LINDGREN,[1] MICHAEL J. BRAUN,[2] MATTHEW A. GONDA,[2] AND PETER M. HOWLEY[1]

Laboratory of Tumor Virus Biology, National Cancer Institute, Bethesda, Maryland 20892,[1] and Laboratory of Cell and Molecular Structure, Program Resources, Inc., Frederick Cancer Research Facility, National Cancer Institute, Frederick, Maryland 21701[2]

Received 23 September 1986/Accepted 18 November 1986

We cloned and analyzed the integrated human papillomavirus type 16 (HPV-16) genomes that are present in the human cervical carcinoma cell lines SiHa and CaSki. The single HPV-16 genome in the SiHa line was cloned as a 10-kilobase (kb) *Hin*dIII fragment. Integration of the HPV-16 genome occurred at bases 3132 and 3384 with disruption of the E2 and E4 open reading frames (ORFs). An additional 52-base-pair deletion of HPV-16 sequences fused the E2 and E4 ORFs. The 5′ portion of the disrupted E2 ORF terminated immediately in the contiguous human right-flanking sequences. Heteroduplex analysis of this cloned integrated viral genome with the prototype HPV-16 DNA revealed no other deletions, insertions, or rearrangements. DNA sequence analysis of the E1 ORF, however, revealed the presence of an additional guanine at nucleotide 1138, resulting in the fusion of the E1a and E1b ORFs into a single E1 ORF. Sequence analysis of the human flanking sequences revealed one-half of an *Alu* sequence at the left junction and a sequence highly homologous to the human O repeat in the right-flanking region. Analysis of the three most abundant *Bam*HI clones from the CaSki line showed that these consisted of (i) full-length, 7.9-kb HPV-16 DNA; (ii) a 6.5-kb genome resulting from a 1.4-kb deletion of the long control region; and (iii) a 10.5-kb clone generated by a 2.6-kb tandem repeat of the 3′ early region. These HPV-16 genomes were arranged in the host chromosomes as head-to-tail, tandemly repeated arrays. Transcription analysis revealed expression of the HPV-16 genome in each of these two cervical carcinoma cell lines, albeit at significantly different levels. Preliminary mapping of the viral RNA with subgenomic strand-specific probes indicated that viral transcription appeared to be derived primarily from the E6 and E7 ORFs.

The papillomaviruses are a group of small DNA viruses that are associated with benign squamous epithelial tumors in higher vertebrates. Several animal models have shown that benign papillomavirus lesions can progress to malignant lesions in the presence of a cocarcinogen (14–16, 29). There are over 40 types of human papillomaviruses (HPVs), each of which is usually associated with specific pathologic entities. HPV types 6 (HPV-6), 11, 16, 18, 31, and 33 have been associated with a variety of benign and malignant anogenital lesions (2, 5, 6, 8, 10, 11, 24, 25, 41). HPV-6, -11, and -31 have generally been found in condyloma acuminata and mild cervical dysplasias. Malignant progression of these lesions is apparently rare. The DNAs of HPV-16, -18, and -33 have been detected predominantly in moderate to severe cervical dysplasias, as well as in invasive cervical carcinomas, although mature virus is usually not found in the more malignant lesions. HPV DNA has also been found in metastases of cervical carcinomas (36).

Analysis of the HPV-16 or HPV-18 DNA in malignant cervical lesions has demonstrated that the DNA is usually integrated into the human genome, although in some cases there is also abundant extrachromosomal viral DNA (2, 8, 34, 36). Integration of the viral genome generally results in the disruption of the E2 open reading frame (ORF) (22, 33, 34), which in the bovine papillomavirus has been shown to encode a *trans*-activator of viral transcription (38). Several cell lines derived from cervical carcinomas have been shown to contain transcriptionally active HPV-16 or HPV-18 DNA in an integrated state (2, 27, 33, 34, 44).

* Corresponding author.

We present here the cloning and analysis of the integrated HPV-16 DNA in the SiHa and CaSki cervical carcinoma cell lines and an analysis of HPV-16 transcription in these cell lines. The CaSki and SiHa cells contain greater than 600 and 1 to 2 copies of the HPV-16 genome, respectively (44). In the SiHa cell line, the HPV-16 genome is integrated into the cellular DNA with disruption and partial deletion of the viral E2 ORF. In both of the cell lines, transcription of the HPV-16 genome is limited predominantly to the E6 and E7 ORFs.

MATERIALS AND METHODS

Cell culture. The human cervical carcinoma cell lines SiHa and CaSki were obtained from the American Type Culture Collection (Rockville, Md.) and grown in Dulbecco modified Eagle medium supplemented with 10% fetal calf serum (GIBCO Laboratories, Grand Island, N.Y.).

DNA isolation and Southern blot analysis. DNA was extracted from monolayer cell cultures, digested with restriction enzymes, separated by electrophoresis through 0.5% agarose gels, and transferred to nitrocellulose filters as described previously (17). DNA probes were prepared by labeling gel-purified DNA fragments with ^{32}P by nick translation (28) or random primer labeling (9) to specific activities of 10^8 or 10^9 cpm/μg of DNA, respectively. HPV-16 DNA probes were made from the prototypic clone described by Durst et al. (8). Hybridization was performed with 10^7 to 10^8 cpm of probe in 3× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–10× Denhardt solution–0.1% sodium dodecyl sulfate (SDS)–50 μg of sheared salmon sperm DNA per ml at 60°C for 12 to 18 h. Subsequent washing was in 2×

SSC and 0.1% SDS at 60°C followed by three washes in 0.1× SSC and 0.1% SDS at 60°C. Blots were exposed at −70°C with Kodak Lanex medium screens for 1 to 21 days.

Genomic cloning. CaSki or SiHa cellular DNA cut with *Bam*HI or *Hin*dIII, respectively, was cloned into the *Bam*HI or *Hin*dIII sites of lambda L47.1 following the procedures described by Maniatis et al. (21). Recombinants were plaqued on *Escherichia coli* K802. Plaques were transferred to nitrocellulose membranes and hybridized with ^{32}P-labeled HPV-16 DNA, as described above. The HPV-16-containing cellular inserts were cleaved from lambda L47.1 recombinants with *Bam*HI or *Hin*dIII and cloned into the plasmid vectors pML2 or pML2d for further analysis. Characterization of the clones was carried out by comparison of restriction endonuclease cleavage patterns of the genomic clones and the HPV-16 prototypic clone described by Durst et al. (8) with the published sequence of HPV-16 (35).

Heteroduplex analysis. Insert DNA was cleaved from plasmid sequences with *Bam*HI for the prototypic HPV-16 plasmid or with *Hin*dIII for the SiHa genomic clone and purified by agarose gel electrophoresis (40). Heteroduplexes were prepared by the method described by Davis et al. (7). The stringency of hybridization and spreads was calculated from the equations described by McConaughy et al. (26), assuming that the G+C content of DNA was 54%. The conditions of hybridization and spreading were approximately −25 and −4°C, respectively.

Sequence analysis. DNA sequencing was done by the method described by Maxam and Gilbert (23). DNA was cleaved with the restriction endonucleases shown in Fig. 7 and either 5′ end labeled with polynucleotide kinase or 3′ end labeled with the Klenow fragment of DNA polymerase I. Most sequence analyses were performed on a computer (PC AT; International Business Machines). Protein data base searches were done with the programs described by Lipman and Pearson (19). DNA sequence comparisons against GenBank Release 44.0 (Genetic Sequence Data Bank) were done on a computer (VAX) with the programs described by Wilbur and Lipman (42).

Preparation of RNA. RNA was extracted from subconfluent flasks (150 mm^2) of SiHa or CaSki cells by the procedure described by Chirgwin et al. (3). Cells were washed once with phosphate-buffered saline (GIBCO) and lysed immediately in 4 M guanidinium thiocyanate–0.5% sodium *N*-lauroylsarcosine–25 mM sodium citrate (pH 7.0)–0.1 M 2-mercaptoethanol. High-molecular-weight DNA was sheared by several passages through an 18-gauge needle to reduce the viscosity. The lysate was then layered over a cushion of 5.7 M CsCl and 0.1 M EDTA (pH 7.0) in an SW41 polyallomer tube. RNA was pelleted by centrifugation at 30,000 rpm (Beckman SW41 rotor) for 20 h at 17°C. The RNA pellets were suspended in distilled water and precipitated at −20°C by the addition of one-tenth volume of 3 M sodium acetate (pH 5.2) and two volumes of ethanol. Poly(A)$^+$ RNA was selected by two cycles of oligo(dT)-cellulose affinity chromatography (1).

Northern blot analysis. Poly(A)$^+$ RNAs were fractionated in 1.2 to 1.4% agarose gels in 2.2 M formamide (18). The RNA was transferred to filters (Gene Screen) by capillary blotting in 25 mM sodium phosphate (pH 6.5) for 12 h. The nylon filter was than wrapped in Saran Wrap while wet and irradiated with UV light (4). Prehybridization and hybridization buffers were composed of 1% crystalline grade bovine serum albumin–1 mM EDTA–0.5 M sodium phosphate (pH 7.2)–7% SDS. Prehybridization (30 min) and hybridization (12 to 20 h) were carried out at 65°C in a hot air shaker with gentle agitation. The filters were then washed twice in 0.5% bovine serum albumin–1 mM EDTA–40 mM sodium phosphate–5% SDS for 5 min with agitation and eight times in 1 mM EDTA–40 mM sodium phosphate–1% SDS for 5 min each by the method described by Church and Gilbert (4). Autoradiography was for 1 to 7 days at −80°C with a Lightning-Plus (Du Pont Cronex; Du Pont Co., Wilmington, Del.) intensifying screen.

Generation of hybridization probes. Strand-specific hybridization probes were generated by the method described by Hu and Messing (13). Single-stranded M13 DNAs with inserts were prepared by precipitation with polyethylene glycol by the method described by Heidecker et al. (12). Briefly, 0.2 to 1 μg of single-stranded DNA was annealed to 1 μl (0.03 A_{260} U/ml) of 17-base hybridization probe primer (P-L Biochemicals, Inc., Milwaukee, Wis.) in 1× *Hae*III buffer (6 mM Tris hydrochloride [pH 7.5], 6 mM NaCl, 6 mM $MgCl_2$, and 1 mM dithiothreitol). The mixture was heated for 10 min at 55°C and cooled slowly to room temperature. [α-^{32}P]dATP (20 μCi; 3,000 Ci/mmol) and 1 μl of a 500 μM mixture of dGTP, dCTP, and dTTP were added. Elongation was accomplished by the addition of 1 μl of DNA polymerase Klenow fragment (0.5 U/μl) and incubation at 15°C for 1 h. Free label was removed by purification through a Sephadex G-50 spin column (21).

RESULTS

Analysis of HPV-16 DNA integrated into the CaSki cell genome. Southern blot analysis of CaSki cellular DNA with HPV-16 probes has shown the presence of approximately 600 copies of integrated HPV-16 DNA per genome (44). Digestion of CaSki DNA with *Bam*HI yielded two major HPV-16-containing fragments of 7.9 and 6.5 kilobases (kb) and several minor fragments including 7.25- and 10.5-kb fragments (data not shown). The production of only a few major HPV-16-containing fragments by digestion with *Bam*HI, a single-cut restriction endonuclease for HPV-16, indicates that most of the HPV-16 genomes are tandemly arranged in a head-to-tail fashion. Representative clones of HPV-16 containing *Bam*HI fragments were obtained by cloning the *Bam*HI-cut CaSki DNA into lambda L47.1 and selecting by hybridization to a HPV-16 probe. Representative clones with insert sizes of 10.5, 7.9, and 6.5 kb were analyzed by detailed restriction analysis. The comparison of these clones with the restriction endonuclease site map of the prototypic HPV-16 DNA (8, 35) is shown in Fig. 1. Clone p894 (7.9-kb insert) was indistinguishable from the prototype HPV-16 DNA clone and represented tandemly repeated full-length copies of the HPV-16 genome with no large deletions, insertions, or rearrangements. Clone p896 had a deletion of 1.4 kb centered on the long control region (LCR). This deletion did not extend to the *Eco*RI site at nucleotide 6818 or the *Taq*I site at nucleotide 505. The deletion included the 3′ end of the L1 ORF, the LCR, and most of the E6 ORF. The third clone, p895, had a tandem repeat of approximately 2.6 kb which contained the E2, E4, and E5 ORFs and portions of the E1 and L2 ORFs. The *Nar*I site at nucleotide 1310 and the *Pst*I site at nucleotide 4759 lie outside the region of duplication.

Cloning of HPV-16 DNA integrated into the SiHa genome. SiHa genomic DNA was digested with no cut (*Hin*dIII) or single-cut (*Bam*HI and *Hin*cII) restriction endonucleases for HPV-16 DNA and hybridized with a HPV-16 DNA probe (Fig. 2A). Digestion with *Hin*dIII revealed only a single HPV-16-containing fragment of 10 kb, indicating that there is

only one site of integration into the cellular genome. Digestion with *Bam*HI yielded two HPV-16-containing fragments measuring 24 kb and greater than 27 kb in size, indicating that the integrated HPV-16 genome retains the *Bam*HI cleavage site. Digestion with *Hinc*II, however, gave only one fragment hybridizing to HPV-16, suggesting that the *Hinc*II site of HPV-16 is either missing or near the site of integration (see below). SiHa cellular DNA cleaved with *Hin*dIII was cloned into lambda L47.1, and the 10-kb HPV-16-containing clone was selected by plaque hybridization with an HPV-16 DNA probe. Four independent clones were obtained. The lambda inserts were subcloned into the *Hin*dIII site of pML2, and the restriction endonuclease patterns were compared with those of the prototype HPV-16 clone (data not shown). All four clones were identical by restriction endonuclease cleavage analysis; the HPV-16 DNA was integrated into cellular DNA within the viral E2 ORF between the *Tth*111-I site (nucleotide 2712) and the *Pst*I site (nucleotide 3697). One clone (p780) was chosen for further analysis.

Heteroduplex analysis. Clone p780 was cut with *Hin*dIII to release the insert from the vector, and the insert was compared with the *Bam*HI cloned insert of the prototype HPV-16 clone (8) by heteroduplex analysis under stringent spreading conditions (T_m, −4°C) (Fig. 3). There was a 0.3-kb deletion of HPV-16 sequences at the point of integration into the host genome (Fig. 3a). The 0.3- and 1.2-kb single-stranded tails in the heteroduplex structure represent the left and right human flanking sequences, respectively (see Fig. 7). There were no other insertions or deletions detectable by heteroduplex analysis between the prototype HPV-16 and the SiHa genomic clone. The stringency of the hybridization conditions indicates that sequence homology of the SiHa clone and the prototype HPV-16 DNA probably exceeds 95% throughout the genome. This is supported by the sequence analysis of the HPV-16 DNA from the SiHa clone which revealed only six point mutations in the 1,168 base pairs (bp) that were sequenced (see below).

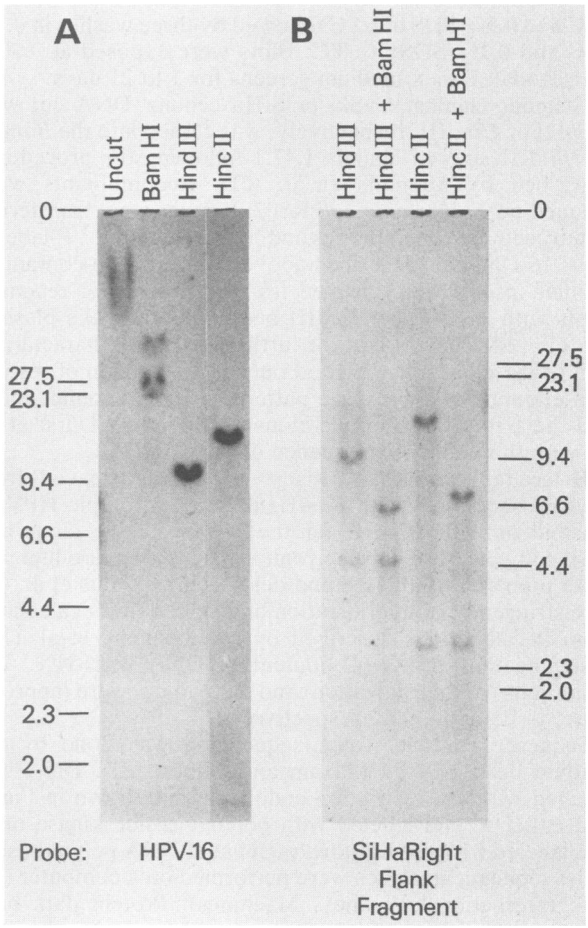


FIG. 2. Genomic southern blots of SiHa cellular DNA. Genomic DNA (5 μg) extracted from the SiHa cell line was digested with the indicated restriction enzymes overnight at 37°C. The DNA was separated by electrophoresis on 1% agarose gels, transferred to nitrocellulose membranes, and hybridized with ^{32}P-labeled HPV-16 (A) or SiHa right-flank (B) DNA probes under stringent conditions for 16 h. The HPV-16 probe was prepared by nick translation (28) of *Bam*HI-cleaved, gel-purified HPV-16 DNA. The SiHa right-flank probe was prepared by the random primer labeling (9) of a gel-purified *Tth*111-I–*Hin*d III fragment (see Fig. 7) derived from an *Ava*II to *Hin*dIII subclone of the SiHa genomic clone. The specific activities of these probes were 10^8 and 10^9 cpm/μg, respectively. Autoradiography was from 3 days to 3 weeks at −80°C with Kodak Lanex medium intensifying screens. The positions of lambda-*Hin*dIII DNA marker fragments are indicated on the sides of the gels. The label O refers to the origin of the gels.

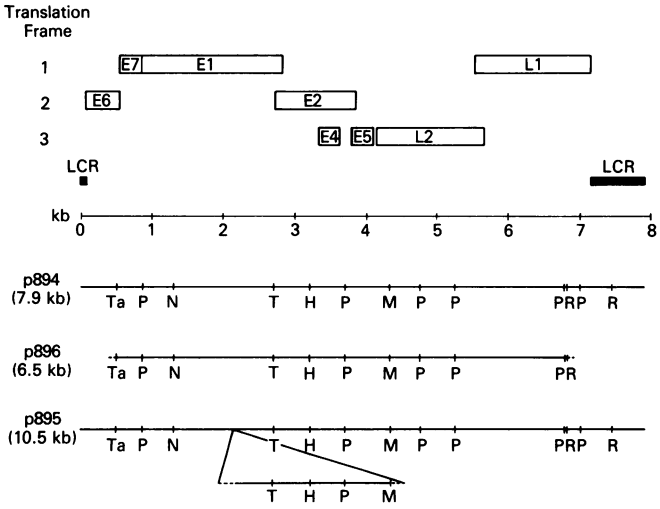


FIG. 1. Restriction map of HPV-16 genomes in the CaSki cell line. The restriction site maps of three HPV-16 genomes from the CaSki cell line subcloned at the *Bam*HI restriction site into the vector pML2d are shown. Restriction enzyme cleavage patterns were determined for each clone and compared with the patterns for the prototype HPV-16 clone (8). The solid lines for each clone represent portions of the HPV-16 genome which are unambiguously present, as indicated by the presence of a restriction enzyme cleavage site and fragment size. The dashed lines represent uncertainty in the endpoints of deletions or insertions. Abbreviations for restriction enzymes: Ta, *Taq*I; P, *Pst*I; N, *Nar*I; T, *Tth*111-I; H, *Hinc*II; M, *Mst*II; R, *Eco*RI. LCR is the long control region between the L1 and E6 ORFs.

DNA sequence analysis of the E1 ORF. The 5′ region of the E1 ORF in p780 was sequenced on both strands between the *Nco*I site (HPV-16 nucleotide 864) and the *Nar*I site (HPV-16 nucleotide 1310). An additional guanine at HPV-16 nucleotide 1138 (Fig. 4A) was the only difference in this region between clone p780 and the prototype HPV-16 clone (8, 35). The sequence of the prototype HPV-16 clone was verified by sequencing by the method described by Maxam and Gilbert (23) from the *Aha*II site at HPV-16 nucleotide 1310 (data not shown). This additional guanine in the SiHa

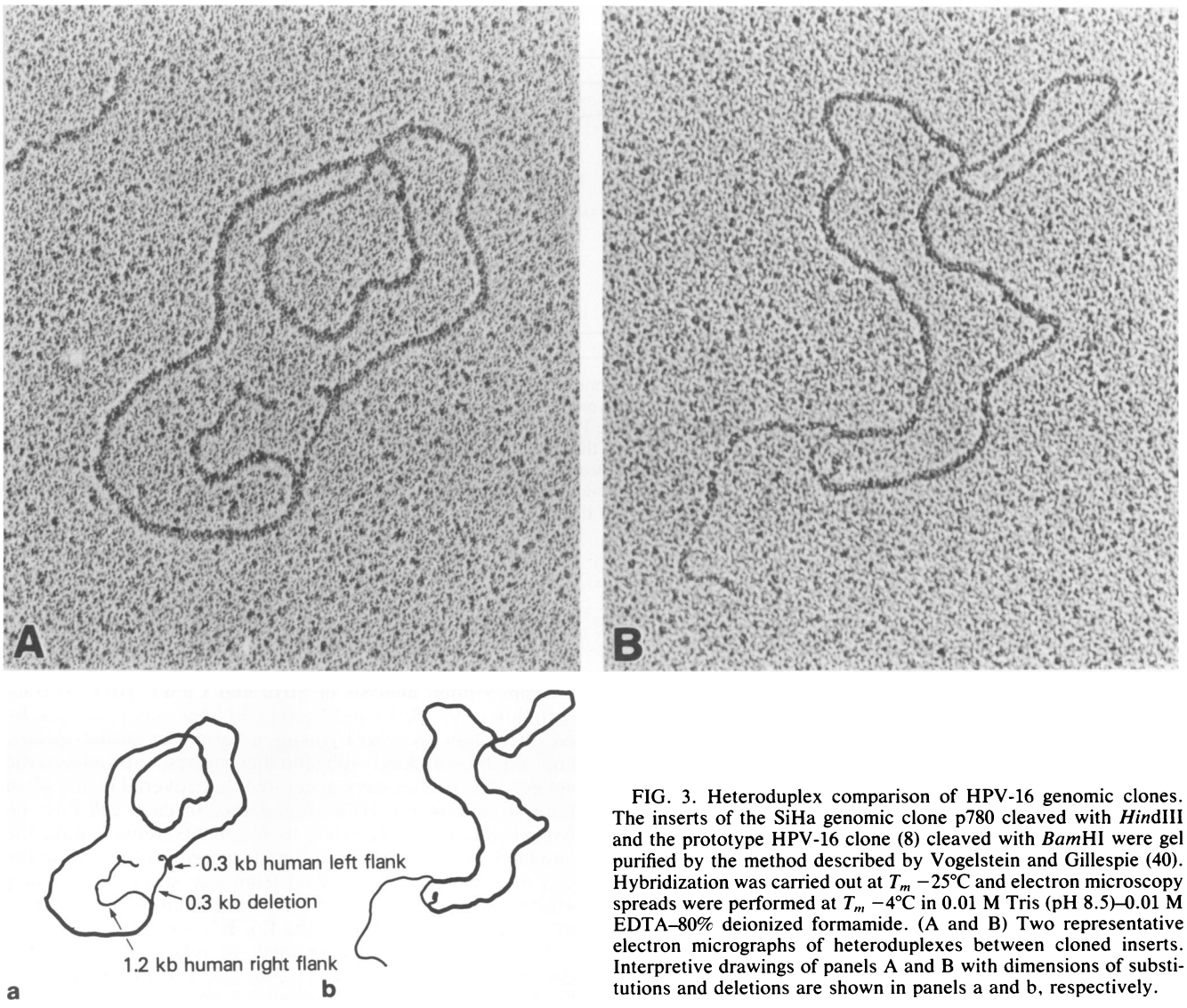


FIG. 3. Heteroduplex comparison of HPV-16 genomic clones. The inserts of the SiHa genomic clone p780 cleaved with *Hin*dIII and the prototype HPV-16 clone (8) cleaved with *Bam*HI were gel purified by the method described by Vogelstein and Gillespie (40). Hybridization was carried out at T_m −25°C and electron microscopy spreads were performed at T_m −4°C in 0.01 M Tris (pH 8.5)–0.01 M EDTA–80% deionized formamide. (A and B) Two representative electron micrographs of heteroduplexes between cloned inserts. Interpretive drawings of panels A and B with dimensions of substitutions and deletions are shown in panels a and b, respectively.

clone resulted in a frame shift which fused the E1a and E1b ORFs into a single long E1 ORF (Fig. 4B). This same change has also been noted in the E1 ORF of another HPV-16 clone isolated from a human cervical carcinoma (22). Thus, the organization of the HPV-16 genome is similar to that of all other papillomaviruses that have been sequenced in that it contains an intact E1 ORF.

DNA sequence analysis of the junctional and flanking regions. The entire 315 bp of left human-flanking region and the host-viral junction of the SiHa clone was sequenced from the *Hin*dIII site to the *Pst*I site (HPV-16 nucleotide 3696). A search of GenBank for sequences homologous to the left-flanking region showed that integration occurred within the center of an *Alu* repeat (32) (Fig. 5). Only half of the *Alu* repeat remained. The left viral junction was at HPV-16 nucleotide 3384. The 5′ nucleotides of the E2 and E4 ORFs of HPV-16 were 2725 and 3332, respectively; integration disrupted both the E2 and E4 ORFs. A 52-bp deletion of HPV-16 nucleotides 3459 to 3511 caused a frameshift termination of the E2 ORF and fused the E4 ORF to the 3′ portion of the E2 ORF. Two human-viral hybrid ORFs were generated across the left junction, encoding potential human-E2 (labeled H2) and human-E4-E2 (labeled H4/2) fusion proteins of 123 and 152 amino acids, respectively (Fig. 5). A search for homology with proteins in the National Biomedical Research Foundation protein data base yielded no significant homologies other than with the E2 and E4 proteins of other papillomaviruses.

A 1.2-kb *Ava*II fragment was sequenced which contained the 5′ half of the E2 ORF and the right viral-host junction. The right viral-host junction was at HPV-16 nucleotide 3132. There was a deletion of 251 nucleotides of the HPV-16 E2 ORF between the two viral-host junctions. This deletion included the *Hin*cII site at HPV-16 nucleotide 3210. The 5′ E2 ORF terminated after three codons in the human flanking region (Fig. 6A). No splice consensus sequence was found in this region, and therefore, there was no evidence that the E2 ORF might be fused at the level of transcriptional processing to a host ORF. In addition, no significantly large ORFs were found in the greater than 750 bp of right flanking DNA that was sequenced. The sequence immediately to the right of the junction did not contain any *Alu* repeat sequences, suggest-

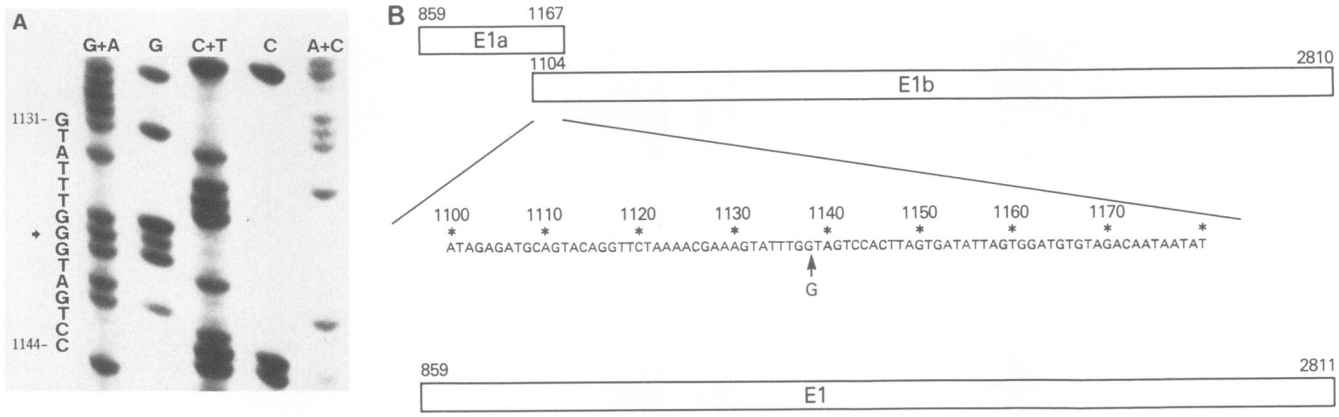


FIG. 4. Sequence analysis of the E1 region of the SiHa genomic clone. (A) The 5′ portion of the E1 ORF between the *Nco*I and *Nar*I sites in the SiHa genomic clone p780 was sequenced on both strands by the method described by Maxam and Gilbert (23). Fragments were labeled with ^{32}P at the 3′ ends with the Klenow fragment of DNA polymerase I (21). The sequencing gel of the region between HPV-16 nucleotides 1131 and 1144 is presented along with the DNA sequence at the left of the figure, and the cleavage specificities are indicated above each lane. The arrow indicates the additional guanine. (B) The sequence of the prototype HPV-16 clone (35) from nucleotides 1100 to 1180 is shown in the center, along with the location of the additional guanine (arrow) found in the SiHa HPV-16 genomic clone p780 at HPV-16 nucleotide 1138. Above the sequence is a schematic representation of the E1 ORFs of the prototype HPV-16 clone. The schematic representation of the E1 ORF of the SiHa genomic clone is shown below the sequence.

ing that cellular sequences were deleted along with the 251 nucleotides of HPV-16 (see below). The sequence of the *Spe*I to *Ava*II portion of the *Ava*II fragment is shown in Fig. 6B. A search of GenBank for homologous sequences revealed a close homology with the human O repeat (39).

Determination of the HPV-16 DNA copy number in the SiHa cell line. Hybridization of Southern blots of SiHa cellular DNA with a unique probe from the human-flanking DNA was used to determine the number of copies of the HPV-16 genome that are integrated into the SiHa genome. Such a probe should give bands of equal intensity for both the normal allele and the allele with the integrated HPV-16 DNA if there is only a single integrated copy. The 850-nucleotide *Tth*111-I to *Hin*dIII fragment from the right human flank (Fig. 7) was chosen to avoid the highly repeated *Alu* and O-repeat elements. The intensity of hybridization to the 10-kb *Hin*dIII and 13.4-kb *Hin*cII HPV-16-containing fragments was, at most, twice the intensity of the hybridization to the 4.6- and 2.7-kb fragments from the normal allele (Fig. 2B). This indicates that there are at most two copies of the HPV-16 genome integrated in the SiHa cell line. Because there was only a single site of integration, it is most likely that there is only one copy of the HPV-16 genome per cell. We cannot rule out that there was a duplication of the region of the human genome containing the integrated HPV-16 genome, however.

Only the 10-kb *Hin*dIII and 13.4-kb *Hin*cII fragments were cut by *Bam*HI, as expected for a single-cut enzyme for HPV-16 DNA. *Bam*HI did not cleave within the vicinity (>18 kb) of the integrated HPV-16 DNA (Fig. 2A). Assuming that there is no restriction site polymorphism, we can conclude from the length of the *Hin*dIII fragment (4.6 kb) of the normal allele that there must be at least a 2.6-kb deletion of human DNA associated with the integration of HPV-16 into the SiHa genome. This is consistent with the presence of only half of an *Alu* repeat at the left host-virus junction and no *Alu* repeat at the right junction. There was insufficient homology between cellular sequences at the host-viral junctions and viral sequences in the E2 ORF to suggest that integration occurred by homologous recombination. Because of the large deletions that occurred, however, homologous recombination cannot be ruled out without sequencing the normal cellular allele.

Transcription analysis of SiHa and CaSki. HPV-16 transcription in the SiHa and CaSki cell lines was examined by Northern blot analysis by using a variety of strand-specific and double-stranded subgenomic probes. Strand-specific subgenomic probes were generated that covered nearly all of both strands of the HPV-16 genome. In each cell line, no hybridization was detected to fragments representing the late ORFs or the LCR. The viral transcripts observed in the cell lines were all transcribed from one strand. The major viral transcripts in both cell lines were composed principally of sequences derived from the E6, E7, and E1 ORFs.

The CaSki cell line contained, at relatively high abundance, four major transcripts: 3.6, 1.5, 1.0, and 0.6 kb (Fig. 8). Minor transcripts of 3.0 and 2.2 kb were detected with longer exposures. Viral transcription in the CaSki cell line included one long transcript (3.6 kb) that originated in the E6 region and that continued through the E1 region into the E2 ORF. It could not be determined from the Northern blot data whether this transcript was spliced. The 1.5- and 1.0-kb transcripts appear to splice the E6-E7 region to the E2 region. The smaller 650-base transcript hybridized only to E2 ORF-specific probes, suggesting that a downstream promoter, possibly analogous to the P_{2443} or P_{3080} promoter of bovine papillomavirus type 1 (BPV-1) (38, 43), might be active in these cells.

The SiHa cell line also contained four major transcripts: 4.6, 4.1, 1.6, and 1.4 kb (Fig. 8). In addition, longer autoradiographic exposures revealed a number of low-abundance viral transcripts, particularly when probes that spanned the E1 ORF were used. The two large viral transcripts (4.6 and 4.1 kb) hybridized only with E6-E7-specific probes, suggesting that they are largely comprised of host cellular sequences. All transcripts must use cellular polyadenylation signals because integration placed the normal viral polyadenylation signal upstream of the viral promoter(s). A polyadenylation signal (AATAAA) was found in the right flank sequence (Fig. 6B, nucleotide 320) but could not be

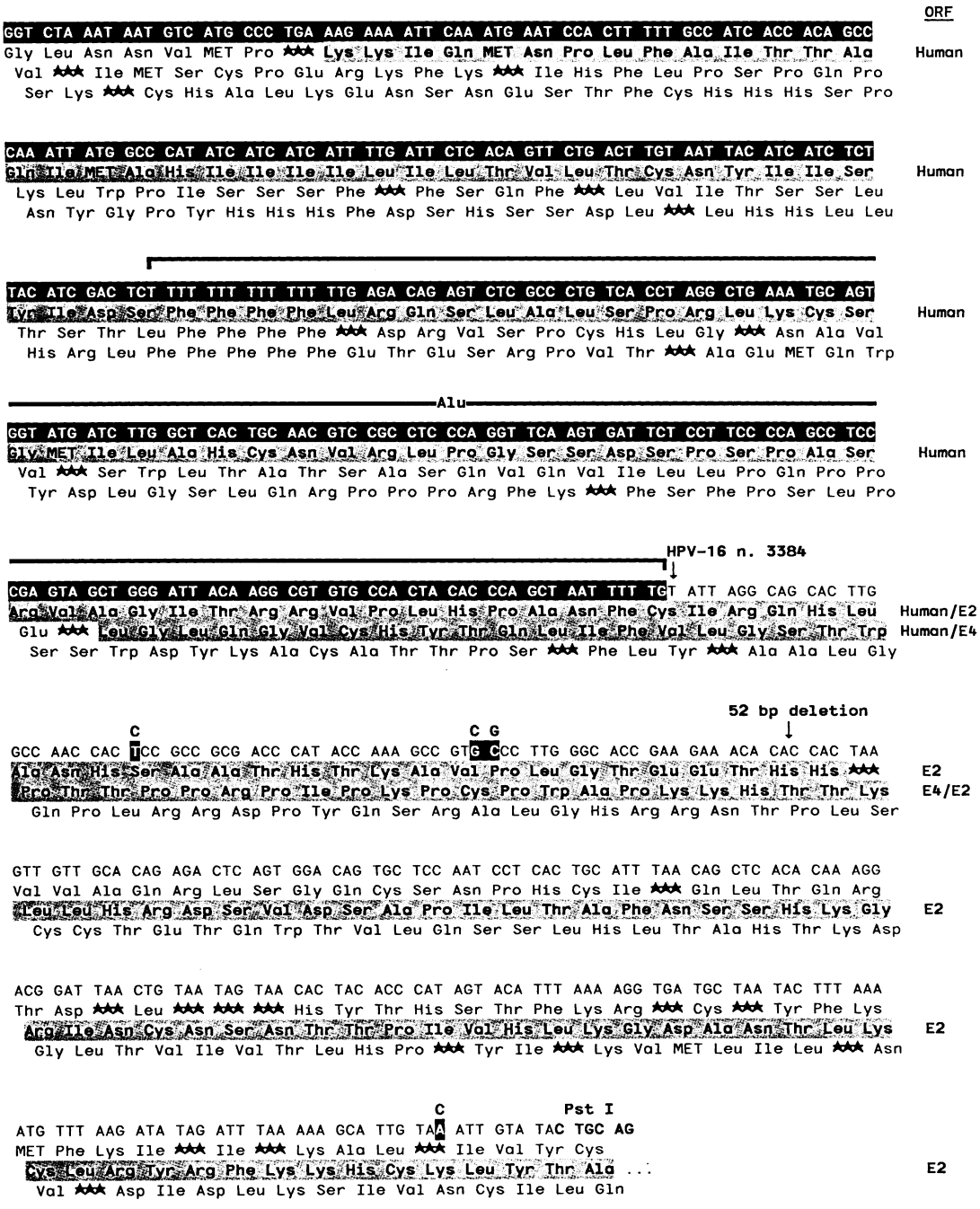
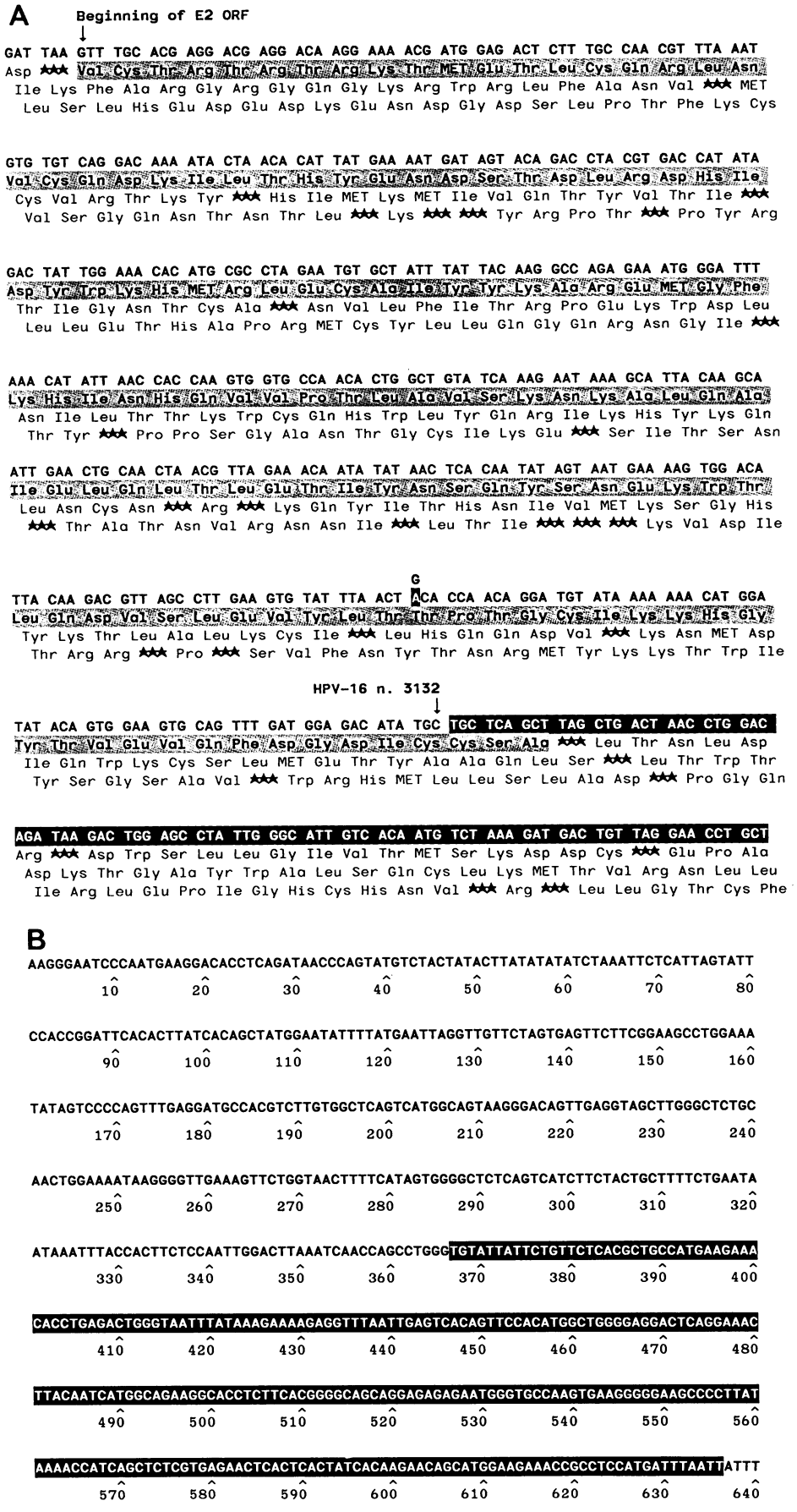


FIG. 5. Sequence of the left flank and host-viral junction in the SiHa cell line. The sequence of the *Hin*dIII to *Pst*I fragment (see Fig. 7) containing the left host-viral junction was determined as described in the legend to Fig. 4. The black highlighting indicates host-derived sequences. The *Alu* repeat is indicated by a heavy bar above the sequence. Sequence differences between the HPV-16 prototype clone and the SiHa HPV-16 clone are highlighted in black, and the corresponding bases in the prototype clone are above the SiHa sequence. The position of a 52-bp deletion within the E2 and E4 ORFs is indicated by an arrow. The translation of the sequence in all three reading frames is shown below the sequence, with asterisks denoting a stop codon. ORFs are shaded and identified at the right of the sequence. n., Nucleotide.

used by the 4.6- and 4.1-kb transcripts because the maximum length of unspliced transcript from viral promoters at nucleotides 90 to 100 (36) would be approximately 3,500 nucleotides. The smaller transcripts (1.6 and 1.4 kb) hybridized to probes derived from both the E6-E7 and the E1 regions. Probes specific for the E2-E5 ORFs failed to hybridize to these or any other transcripts in the SiHa cell line. The precise structure of these 1.4- and 1.6-kb transcripts with respect to viral splice donor and acceptor sites and whether or not they contain host sequences are not known. The lack of hybridization of SiHa transcripts to E2-E4-E5 probes also implies that there is no detectable expression of the hybric human-E2 and human-E4-E2 ORFs (Fig. 5). The pattern of transcription seen in the SiHa line thus differs from that of


A
Beginning of E2 ORF
GAT TAA GTT TGC ACG AGG ACG AGG ACA AGG AAA ACG ATG GAG ACT CTT TGC CAA CGT TTA AAT
Asp *** Val Cys Thr Arg Thr Arg Thr Arg Lys Thr MET Glu Thr Leu Cys Gln Arg Leu Asn
Ile Lys Phe Ala Arg Gly Arg Gly Gln Gly Lys Arg Trp Arg Leu Phe Ala Asn Val *** MET
Leu Ser Leu His Glu Asp Glu Asp Lys Glu Asn Asp Gly Asp Ser Leu Pro Thr Phe Lys Cys
GTG TGT CAG GAC AAA ATA CTA ACA CAT TAT GAA AAT GAT AGT ACA GAC CTA CGT GAC CAT ATA
Val Cys Gln Asp Lys Ile Leu Thr His Tyr Glu Asn Asp Ser Thr Asp Leu Arg Asp His Ile
Cys Val Arg Thr Lys Tyr *** His Ile MET Lys MET Ile Val Gln Thr Tyr Val Thr Ile ***
Val Ser Gly Gln Asn Thr Asn Thr Leu *** Lys *** *** Tyr Arg Pro Thr *** Pro Tyr Arg
GAC TAT TGG AAA CAC ATG CGC CTA GAA TGT GCT ATT TAT TAC AAG GCC AGA GAA ATG GGA TTT
Asp Tyr Trp Lys His MET Arg Leu Glu Cys Ala Ile Tyr Tyr Lys Ala Arg Glu MET Gly Phe
Thr Ile Gly Asn Thr Cys Ala *** Asn Val Leu Phe Ile Thr Arg Pro Glu Lys Trp Asp Leu
Leu Leu Glu Thr His Ala Pro Arg MET Cys Tyr Leu Leu Gln Gly Gln Arg Asn Gly Ile ***
AAA CAT ATT AAC CAC CAA GTG GTG CCA ACA CTG GCT GTA TCA AAG AAT AAA GCA TTA CAA GCA
Lys His Ile Asn His Gln Val Val Pro Thr Leu Ala Val Ser Lys Asn Lys Ala Leu Gln Ala
Asn Ile Leu Thr Thr Lys Trp Cys Gln His Trp Leu Tyr Gln Arg Ile Lys His Tyr Lys Gln
Thr Tyr *** Pro Pro Ser Gly Ala Asn Thr Gly Cys Ile Lys Glu *** Ser Ile Thr Ser Asn
ATT GAA CTG CAA CTA ACG TTA GAA ACA ATA TAT AAC TCA CAA TAT AGT AAT GAA AAG TGG ACA
Ile Glu Leu Gln Leu Thr Leu Glu Thr Ile Tyr Asn Ser Gln Tyr Ser Asn Glu Lys Trp Thr
Leu Asn Cys Asn *** Arg *** Lys Gln Tyr Ile Thr His Asn Ile Val MET Lys Ser Gly His
*** Thr Ala Thr Asn Val Arg Asn Asn Ile *** Leu Thr Ile *** *** *** Lys Val Asp Ile
G
TTA CAA GAC GTT AGC CTT GAA GTG TAT TTA ACT ACA CCA ACA GGA TGT ATA AAA AAA CAT GGA
Leu Gln Asp Val Ser Leu Glu Val Tyr Leu Thr Thr Pro Thr Gly Cys Ile Lys Lys His Gly
Tyr Lys Thr Leu Ala Leu Lys Cys Ile *** Leu His Gln Gln Asp Val *** Lys Asn MET Asp
Thr Arg Arg *** Pro *** Ser Val Phe Asn Tyr Thr Asn Arg MET Tyr Lys Lys Thr Trp Ile
HPV-16 n. 3132
TAT ACA GTG GAA GTG CAG TTT GAT GGA GAC ATA TGC TGC TCA GCT TAG CTG ACT AAC CTG GAC
Tyr Thr Val Glu Val Gln Phe Asp Gly Asp Ile Cys Cys Ser Ala *** Leu Thr Asn Leu Asp
Ile Gln Trp Lys Cys Ser Leu MET Glu Thr Tyr Ala Ala Gln Leu Ser *** Leu Thr Trp Thr
Tyr Ser Gly Ser Ala Val *** Trp Arg His MET Leu Leu Ser Leu Ala Asp *** Pro Gly Gln
AGA TAA GAC TGG AGC CTA TTG GGC ATT GTC ACA ATG TCT AAA GAT GAC TGT TAG GAA CCT GCT
Arg *** Asp Trp Ser Leu Leu Gly Ile Val Thr MET Ser Lys Asp Asp Cys *** Glu Pro Ala
Asp Lys Thr Gly Ala Tyr Trp Ala Leu Ser Gln Cys Leu Lys MET Thr Val Arg Asn Leu Leu
Ile Arg Leu Glu Pro Ile Gly His Cys His Asn Val *** Arg *** Leu Leu Gly Thr Cys Phe
B
AAGGGAATCCCAATGAAGGACACCTCAGATAACCCAGTATGTCTACTATACTTATATATATCTAAATTCTCATTAGTATT
10 20 30 40 50 60 70 80
CCACCGGATTCACACTTATCACAGCTATGGAATATTTTATGAATTAGGTTGTTCTAGTGAGTTCTTCGGAAGCCTGGAAA
90 100 110 120 130 140 150 160
TATAGTCCCCAGTTTGAGGATGCCACGTCTTGTGGCTCAGTCATGGCAGTAAGGGACAGTTGAGGTAGCTTGGGCTCTGC
170 180 190 200 210 220 230 240
AACTGGAAAATAAGGGGTTGAAAGTTCTGGTAACTTTTCATAGTGGGGCTCTCAGTCATCTTCTACTGCTTTTCTGAATA
250 260 270 280 290 300 310 320
ATAAATTTACCACTTCTCCAATTGGACTTAAATCAACCAGCCTGGGTGTATTATTCTGTTCTCACGCTGCCATGAAGAAA
330 340 350 360 370 380 390 400
CACCTGAGACTGGGTAATTTATAAAGAAAAGAGGTTTAATTGAGTCACAGTTCCACATGGCTGGGGAGGACTCAGGAAAC
410 420 430 440 450 460 470 480
TTACAATCATGGCAGAAGGCACCTCTTCACGGGGCAGCAGGAGAGAGAATGGGTGCCAAGTGAAGGGGGAAGCCCCTTAT
490 500 510 520 530 540 550 560
AAAACCATCAGCTCTCGTGAGAACTCACTCACTATCACAAGAACAGCATGGAAGAAACCGCCTCCATGATTTAATTATTT
570 580 590 600 610 620 630 640

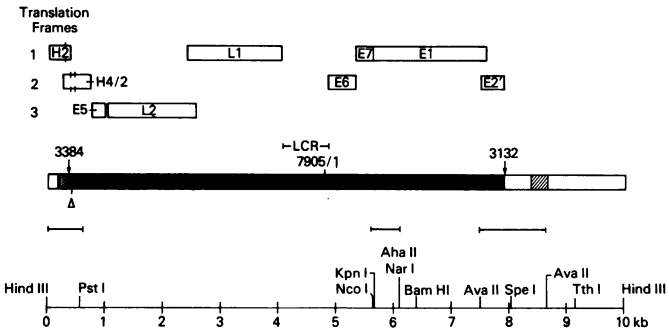


FIG. 7. Schematic diagram of the HPV-16 *Hin*dIII fragment from the SiHa cell line. The solid black region represents the integrated HPV-16 genome. The stippled region to the left of HPV-16 is an *Alu* repeat. The hatched region to the right of HPV-16 is the human O repeat. Open regions represent unique host DNA. The nucleotide positions (HPV-16 numbering system) of the host-viral junctions are labeled above the genomic representation. The position of the 52-bp deletion in the E4-E2 region is indicated by the open triangle. The positions of all major ORFs are shown in the top half of the figure by open boxes. H2 and H4/2 signify human-E2 and human-E4-E2 fusion ORFs, respectively, and the tick marks on the open boxes represent the boundaries of each of these ORFs. The regions that were sequenced are shown in the lower half of the figure by bars. Cleavage sites for the restriction enzymes used for cloning, sequencing, and probe production are indicated above the bottom line.

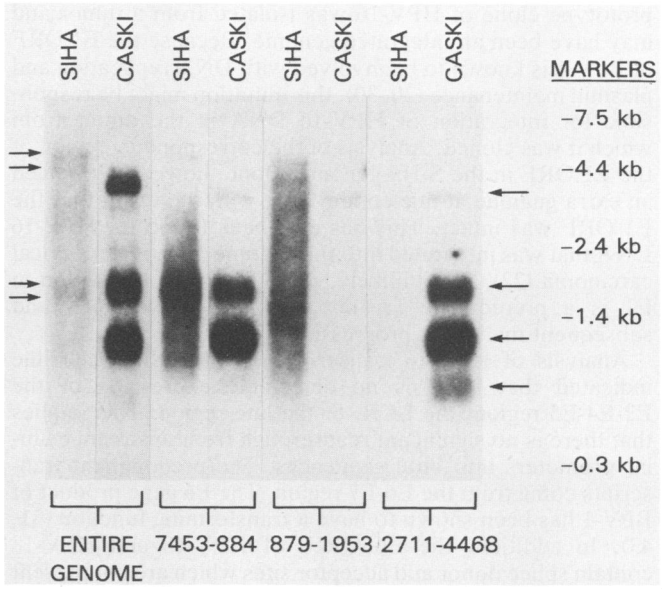


FIG. 8. Northern blot analysis of RNA from the SiHa and CaSki cell lines. Poly(A)-selected RNA from the SiHa and CaSki cell lines was separated by electrophoresis, transferred to a nylon membrane, and hybridized to HPV-16 subgenomic probes as described in the text. The source of the RNA is indicated above each lane, and the hybridization probe is indicated below each lane. The arrows to the left and right of the autoradiogram identify the positions of the major transcripts in the SiHa and CaSki cell lines, respectively. The positions and sizes of the RNAs in the RNA ladder (Bethesda Research Laboratories, Gaithersburg, Md.) used as markers are shown on the far right.

the CaSki cell line in that no E2 sequences are transcribed in the SiHa line. A further difference between the cell lines is that the SiHa cell line contains a considerable number of somewhat poorly delineated viral transcripts that traverse the E1 region, whereas the CaSki cell line does not.

DISCUSSION

HPVs types 6, 11, 16, 18, 31, and 33 have been found associated with human anogenital lesions. A subset of these (HPV-16, -18, and -33) are associated predominantly with severe cervical dysplasia, carcinoma in situ, and invasive cervical carcinoma. HPV-16 DNA is integrated into the genomes of the SiHa and CaSki cervical carcinoma derived cell lines and is transcriptionally active (44). Our analysis of the HPV-16 DNA integrated into the SiHa cellular genome revealed that integration occurred within the E2 ORF of HPV-16. This integration was accompanied by the deletion of a portion of E2 as well as a deletion of a portion of the human genome. Analysis of other HPV-16- and HPV-18-containing cell lines and tumors (22, 33, 34) has also shown that integration usually occurs within the E1 ORF, the E2 ORF, or both. The product of the full E2 ORF of BPV-1 has been shown to transcriptionally activate BPV-1 promoters (37; B. A. Spalholz, C. C. Baker, P. F. Lambert, and P. M. Howley, *in* B. M. Steinberg, J. L. Brandsma, and L. B. Taichman, ed., *Cancer Cells. 5. Papillomaviruses*, in press). Analogous studies with HPV-16 have shown that its E2 gene also encodes a factor that is involved in the activation of viral transcription (W. C. Phelps and P. M. Howley, submitted for publication). Thus, disruption of the E2 ORF by integration would interfere with this transcriptional regulatory function. One can postulate that such an integration and disruption of a gene involved in the control of viral gene expression may play a role in malignant progression because of the deregulation of HPV expression.

The human flanking regions were analyzed for homology with any known DNA sequences. There was no significant homology with any sequenced oncogene. The only significant homologies were with the human *Alu* sequence and the human O repeat (Fig. 7). There were two human ORFs in the left-flanking region that were fused in frame to HPV-16 ORFs to yield human-E2 and human-E4-E2 ORFs (Fig. 5). No E2-E4-containing transcripts were detectable on Northern blots, however, indicating that these fusion ORFs may have no biological significance. The human portions of these chimeric ORFs did not show any significant homology with proteins in the National Biomedical Research Foundation data base.

Sequence analysis of the prototype HPV-16 clone (8) sequenced by Seedorf et al. (35) revealed that the E1 ORF was split into two ORFs, each of which had homology to portions of the E1 ORFs of other papillomaviruses. This

FIG. 6. Sequence of the right viral-host junction and right-flanking region in the SiHa HPV-16 genomic clone. The sequence of the 1.2 kb *Ava*II fragment (see Fig. 7) containing the right host-virus junction of the SiHa genomic clone was determined as described in the legend to Figure 4. (A) Sequence of the *Ava*II to *Spe*I (New England Biolabs, Inc., Beverly, Mass.) fragment. Sequence annotations are as described in the legend to Fig. 5. Black highlighting indicates host-derived sequences. (B) Sequence of the *Spe*I to *Ava*II fragment. The sequence homologous to the human O repeat is highlighted in black.

prototype clone of HPV-16 was isolated from a tumor and may have been an integrated genome. Because the E1 ORF of BPV-1 is known to be involved with DNA replication and plasmid maintenance (20, 30), this mutation might be responsible for integration of HPV-16 DNA in the tumor from which it was cloned. Analysis of the corresponding region of the E1 ORF in the SiHa genomic clone, however, revealed an extra guanine at nucleotide 1138, with the result that the E1 ORF was intact. This has also been found for HPV-16 DNA that was integrated into the genome of another cervical carcinoma (22). It is unlikely, therefore, that a mutation in E1 is a prerequisite for integration of HPV DNA and subsequent malignant progression.

Analysis of HPV-16 transcription in the SiHa cell line indicated that there is no detectable expression of the E2-E4-E5 region, the LCR, or the late region. This implies that there is no significant readthrough from upstream cellular promoters into viral sequences. The predominant transcripts come from the E6-E7 region. The E6 gene product of BPV-1 has been shown to have a transforming function (31, 43). In addition, the E6 ORFs of HPV-16 and HPV-18 contain splice donor and acceptor sites which are not present in HPV-6 or HPV-11 (33, 36). Thus, the HPVs associated with malignant cervical lesions may make E6 proteins that differ from those of the other HPVs. It is tempting to speculate that expression of the E6 gene or an E6-E7 fusion gene product is important in malignant progression. The E6-E7 region is presumably transcribed from viral promoters. Because the E2 *trans*-activating protein is not expressed in the SiHa cells, we presume that the viral promoters are activated directly by cellular transcriptional factors or are enhanced by adjacent cellular *cis*-acting regulatory elements. In addition, the normal early region polyadenylation signals are no longer downstream of the promoters. The HPV-16 sequences must be spliced to RNA from adjacent human DNA, and transcription must use cellular polyadenylation sites. The control of the expression of a potential viral oncogene by cellular regulatory elements may be crucial for the progression of a benign to a malignant lesion.

The correlation of transcription patterns with the genomic structure is more difficult for the CaSki cell line than for the SiHa cell line because of the many integrated copies of HPV-16 DNA in the CaSki cell line. The three most abundant HPV-16 genomic structures (Fig. 1) are present in large tandem arrays. The junctional copies are present in very low abundance and have not yet been isolated or analyzed. Similar to HPV-16 transcription in the SiHa cell line, however, there is no evidence of transcription of the late region or of transcription of the LCR in CaSki cells. Three of the four transcripts contain the E6-E7 region and are presumably transcribed from viral LCR promoters. Unlike the SiHa cells, only the largest transcript contains the E1 region, and all transcripts contain the E2 region. The smallest transcript contains only the E2 region and is probably transcribed from the equivalent of the P_{2443} or P_{3080} promoter in BPV-1 (38, 43). Results of the analyses of Smotkin and Wettstein (36) have suggested that the CaSki transcripts do not utilize the viral polyadenylation site, implying that viral sequences are either spliced to cellular sequences or are fused to cellular sequences at the viral-host junctions. This suggests that only the terminal HPV-16 copies in the large tandem repeats may be transcriptionally active and supports the hypothesis that HPV-16 promoters are under altered regulation, perhaps because of adjacent *cis*-acting regulatory elements. It is possible that the key event in malignant progression is the integration of HPV-16 near such a cellular regulatory element and that it is necessary to destroy the HPV-16 *trans*-acting transcriptional regulatory element(s) by integration within the E2 region of HPV-16.

ACKNOWLEDGMENTS

We thank Jean Noe for technical assistance, Warren Jelinek for supplying the lambda L47.1, Susan Haynes for advice and help with the genomic cloning, and George Michaels for running the GenBank searches.

W.C.P. was supported by National Research Service Award postdoctoral fellowship CA 07713. This study was supported in part by grant 1760 from the Council for Tobacco Research to V.L. M.J.B. and M.A.G. were supported by Public Health Service contract NOI-CO-23910 with Program Resources, Inc., from the National Cancer Institute.

LITERATURE CITED

1. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**:1408–1412.
2. **Boshart, M., L. Gissmann, H. Ikenberg, A. Kleinheinz, W. Scheurlen, 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.
3. **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.
4. **Church, G. M., and W. Gilbert.** 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA **81**:1991–1995.
5. **Crum, C. P., H. Ikenberg, R. M. Richart, and L. Gissmann.** 1984. Human papillomavirus type 16 and early cervical neoplasia. N. Engl. J. Med. **310**:880–883.
6. **Crum, C. P., M. Mitao, R. U. Levine, and S. Silverstein.** 1985. Cervical papillomaviruses segregate within morphologically distinct precancerous lesions. J. Virol. **54**:675–681.
7. **Davis, R. W., M. Simon, and N. Davidson.** 1971. Electron microscopic heteroduplex methods for mapping regions of base homology in nucleic acids. Methods Enzymol. **21**:413–428.
8. **Durst, 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.
9. **Feinberg, A. P., and B. Vogelstein.** 1984. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137**:266–267.
10. **Gissmann, L.** 1984. Papillomaviruses and their association with cancer in animals and in man. Cancer Surv. **3**:161–181.
11. **Gissmann, L., and E. Schwarz.** 1986. Persistence and expression of human papillomavirus DNA in genital cancer. CIBA Found. Symp. **120**:190–207.
12. **Heidecker, G., J. Messing, and B. Gronenborn.** 1980. A versatile primer for DNA sequencing in the M13mp2 cloning system. Gene **10**:69–73.
13. **Hu, N. T., and J. Messing.** 1982. The making of strand-specific M13 probes. Gene **17**:271–277.
14. **Jarrett, W. F. H., P. E. McNeil, W. T. R. Grimshaw, I. E. Selman, and W. I. M. McIntyre.** 1978. High incidence area of cattle cancer with a possible interaction between an environmental carcinogen and a papilloma virus. Nature (London) **274**:215–217.
15. **Jarrett, W. F. H., J. Murphy, B. W. O'Neil, and H. M. Laird.** 1978. Virus-induced papillomas of the alimentary tract of cattle. Int. J. Cancer **22**:323–328.
16. **Kidd, J. G., and P. Rous.** 1937. Effect of the papillomavirus (Shope) upon tar warts of rabbits. Proc. Soc. Exp. Biol. Med. **37**:518–520.
17. **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.

18. **Lehrach, H., D. Diamond, J. M. Wozney, and H. Boedtker.** 1977. RNA molecular weight determination by gel electrophoresis under denaturing conditions, a critical re-examination. Biochemistry **16**:4743–4751.
19. **Lipman, D., and W. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
20. **Lusky, M., and M. R. Botchan.** 1985. Genetic analysis of bovine papillomavirus type 1 trans-acting replication factors. J. Virol. **53**:955–965.
21. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Matsukura, T., T. Kanda, A. Furuno, H. Yoshikawa, T. Kawana, and K. Yoshiike.** 1986. Cloning of monomeric human papillomavirus type 16 DNA integrated within cell DNA from a cervical carcinoma. J. Virol. **58**:979–982.
23. **Maxam, A., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560–564.
24. **McCance, D. J., M. J. Campion, P. K. Clarkson, P. M. Chesters, D. Jenkins, and A. Singer.** 1985. Prevalence of human papillomavirus type 16 DNA sequences in cervical intraepithelial neoplasia and invasive carcinoma of the cervix. Br. J. Obstet. Gynaecol. **92**:1101–1105.
25. **McCance, D. J., P. K. Clarkson, J. L. Dyson, and A. Singer.** 1985. Human papillomavirus types 6 and 16 in multifocal intraepithelial neoplasias of the female lower genital tract. Br. J. Obstet. Gynaecol. **92**:1093–1100.
26. **McConaughy, B. L., C. D. Laird, and B. J. McCarthy.** 1969. Nucleic acid reassociation in formamide. Biochemistry **8**:3289–3295.
27. **Pater, M. M., and A. Pater.** 1985. Human papillomavirus types 16 and 18 sequences in carcinoma cell lines of the cervix. Virology **145**:3313–318.
28. **Rigby, P. D., M. D. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
29. **Rous, P., and J. G. Kidd.** 1936. The carcinogenic effect of a virus upon tarred skin. Science **83**:468–469.
30. **Sarver, N., M. S. Rabson, Y. C. Yang, J. C. Byrne, and P. M. Howley.** 1984. Localization and analysis of bovine papillomavirus type 1 transforming functions. J. Virol. **52**:377–388.
31. **Schiller, J. T., W. C. Vass, and D. R. Lowy.** 1984. Identification of a second transforming region in bovine papillomavirus DNA. Proc. Natl. Acad. Sci. USA **81**:7880–7884.
32. **Schmid, C. W., and W. R. Jelinek.** 1982. The Alu family of dispersed repetitive sequences. Science **216**:1065–1070.
33. **Schneider-Gadicke, 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.
34. **Schwarz, E., U. K. Freese, L. Gissmann, W. Mayer, B. Roggenbuck, A. Stremlau, and H. zur Hausen.** 1985. Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. Nature (London) **314**:111–114.
35. **Seedorf, K., G. Krammer, M. Durst, S. Suhai, and W. Rowekamp.** 1985. Human papillomavirus type 16 DNA sequence. Virology **145**:181–185.
36. **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.
37. **Spalholz, B. A., Y. C. Yang, and P. M. Howley.** 1985. Transactivation of a bovine papilloma virus transcriptional regulatory element by the E2 gene product. Cell **42**:183–191.
38. **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.
39. **Sun, L., K. E. Paulson, C. W. Schmid, L. Kadyk, and L. Leinwand.** 1984. Non-Alu family interspersed repeats in human DNA and their transcriptional activity. Nucleic Acids Res. **12**:2669–2690.
40. **Vogelstein, B., and D. Gillespie.** 1979. Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. USA **76**:615–619.
41. **Wagner, D., H. Ikenberg, N. Boehm, and L. Gissmann.** 1984. Identification of human papillomavirus in cervical swabs by deoxyribonucleic acid in situ hybridization. Obstet. Gynecol. **64**:767–772.
42. **Wilbur, W., and D. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein databases. Proc. Natl. Acad. Sci. USA **80**:726–730.
43. **Yang, Y.-C., H. Okayama, and P. M. Howley.** 1985. Bovine papillomavirus contains multiple transforming genes. Proc. Natl. Acad. Sci. USA **82**:1030–1034.
44. **Yee, C., I. Krishnan-Hewlett, C. C. Baker, R. Schlegel, and P. M. Howley.** 1985. Presence and expression of human papillomavirus sequences in human cervical carcinoma cell lines. Am. J. Pathol. **119**:361–366.