Identification of a new promoter in the early region of the human papillomavirus type 16 genome


Institute of Molecular Pathology, The Protein Laboratory, University of Copenhagen, Panum Institute, Bldg 6.2, Blegdamsvej 3C, DK-2200 Copenhagen N, Denmark

Transcription of the human papillomavirus type 16 (HPV-16) genome is controlled by several promoters; the $P_{97}$ promoter is considered to be the main one. An additional promoter has been identified within the E7 ORF as well as an antisense promoter just upstream of the L2 ORF. The significance of these promoters for early and late gene expression and their activity related to cell differentiation is not known in detail. Identification of two new, previously undescribed transcription start sites at nt 542 just upstream of the E7 ORF and at nt 611 within the E7 ORF is reported. The promoter responsible for the start site at nt 542 ($P_{542}$) was active in SiHa, HeLa and C33A cells. Very low promoter activity was found upstream of the nt 611 start site. The E7 protein has previously been shown to be synthesized from a polycistronic mRNA encoding both the E6 and E7 proteins under the control of the $P_{97}$ promoter. The data reported in the present paper suggest that promoter $P_{542}$ may control synthesis of the E7 oncoprotein from a monocistronic mRNA.

Introduction

Human papillomaviruses (HPVs) are epitheliotropic DNA viruses with a double-stranded circular DNA genome of approximately 8 kbp. The coding strand of the genome encodes six or seven early and two late proteins, and has a long control region (LCR) of approximately 750 bp (for review, see zur Hausen, 1996; Vouden, 1993). More than 80 different HPV types have been identified and new types are currently being isolated (de Villiers, 1994). The HPV types are divided into malignant and benign types; HPV-16 is considered to be the prototype among the malignant HPVs and HPV-11 is the prototype among the malignant HPVs and HPV-11 is considered to be the benign prototype (Vousden, 1993). The late ORFs encode the two late proteins, and has a long control region (LCR) of approximately 750 bp (for review, see zur Hausen, 1996; Vouden, 1993). More than 80 different HPV types have been identified and new types are currently being isolated (de Villiers, 1994). The HPV types are divided into malignant and benign types; HPV-16 is considered to be the prototype among the malignant HPVs and HPV-11 is considered to be the benign prototype (Vousden, 1993). The early ORFs E1 and E2 encode the proteins necessary for viral DNA replication, episomal maintenance and transcriptional regulation, the E1-E4 ORF encodes a protein which binds to the intermediate filament (Roberts et al., 1997), and the E5 and E7 ORFs encode oncoproteins (Hawley-Nelson et al., 1989; Crook et al., 1989; Münger et al., 1989; Leptak et al., 1991; Conrad et al., 1993). The late ORFs encode the two capsid proteins L1 and L2 (for review, see zur Hausen, 1996).

Regulation of gene expression has been studied in several of the HPV types and the locations of the different promoter regions identified for HPV-16, -18, -31b, -11 and -6b are similar. The viruses have a major promoter in front of the E6 ORF ($P_{97}$ in HPV-16 (Smotkin & Wettstein, 1986), $P_{109}$ in HPV-18 (Thierry et al., 1987), $P_{97}$ in HPV-31b (Hummel et al., 1992)). In HPV-11, the major promoter is in the E7 ORF (Chow et al., 1987). There are, however, promoters at nt 99 in HPV-11 and at nt 80 in HPV-6b which are responsible for E6 protein expression (Zhao et al., 1997; Karlen et al., 1996). Several additional promoters have been identified in the early region of the genome, i.e. in the E7 ORF of HPV-16 (Higgins et al., 1992; Doorbar et al., 1990; Grassmann et al., 1996), HPV-31b (Hummel et al., 1995, Ozbun & Meyers, 1997), HPV-11 (Nasseri et al., 1987; DiLorenzo & Steinberg, 1995) and HPV-6 (Chow et al., 1987; Karlen et al., 1996). The E7-localized promoters of HPV-16, -31b and -11 are differentiation-dependent (Cheng et al., 1995; DiLorenzo & Steinberg, 1995; Grassmann et al., 1996; Ozbun & Meyers, 1997, 1998). A promoter is present in the E6 ORF of HPV-6b as observed by Karlen et al. (1996) and Smotkin et al. (1989). This promoter is thought to be responsible for E7 mRNA in benign HPV types. Promoters have been identified just upstream of or within the E2 ORF such as $P_{3598}$ and $P_{3936}$ of HPV-18 (Karlen & Beard, 1993) and $P_{3279}$ of HPV-6b (Karlen et al., 1996). Late promoters were identified within the E4 ORF of HPV-31b ($P_{3239}$) (Ozbun & Meyers, 1998), just downstream of the E5 ORF of HPV-16 (Geisen & Kahn, 1996; Maki et al., 1996), and between the L2
and L1 genes in HPV-31b (P55069) (Karlen & Beard, 1993). Late mRNAs might also be initiated from promoters upstream of the Pcrf of both HPV-16 and HPV-31b where transcription start sites have been identified at position nt 22 of HPV-16 (Rohlfs et al., 1991) and nt ~7375, nt 7783, nt 7850, nt 49 and nt 77 of HPV-31b (Ozbun & Meyers, 1997, 1998, 1999). An antisense promoter in the nt 4030–4230 region of the HPV-16 genome producing E7 antisense RNA was identified recently (Belaguli et al., 1997). The location of promoters in HPV4s resembles that in bovine papillomavirus type 1 as promoters are identified in the upstream regulatory region and in the E6, E1 and E2 ORFs (Soussa et al., 1990).

Based on our previous study, in which we demonstrated translation of the HPV-16 E5 protein from a polytropic mRNA encoding the entire E2 and E5 ORFs (Johnsen et al., 1995), we initiated a search for promoters in the early region of the genome applying a new method for the identification of transcription start sites. We report the identification of two new transcription start sites at nt 542 and nt 611 just upstream of and in the E7 gene, respectively. The promoter region P542 was characterized and shown to be active in both the HPV-positive cell lines HeLa and SiHa, and in the HPV-negative cell line C33A. The potential specificity of this new promoter for E7 is discussed.

Methods

Cell culture. The HPV-negative cervical carcinoma cell line C33A, the HPV-18-positive cervical cell line HeLa and the HPV-16-positive cervical cell line SiHa were maintained in Dulbecco’s modified Eagle’s 1885 medium supplemented with 10% foetal calf serum (FCS), 1% t-glutamine, and 1% penicillin and streptomycin at 37 °C and 5% CO2.

Cell transfections. The day before transfection, 2 × 104 cells were seeded in 3.5 cm Petri (6-well) dishes (NUNC). Just before transfection, the medium was replaced with serum-free DMEM containing 1% t-glutamine, and 1% penicillin and streptomycin. The cells were transfected using Lipofectamine Plus (Gibco BRL) according to the procedure recommended by the manufacturer. Each dish was transfected with 0.5 pmol plasmid DNA, adjusting for differences in the size of the constructs. Three hours after transfection, the medium was replaced with DMEM containing 10% FCS, 1% t-glutamine, and 1% penicillin and streptomycin.

Cloning in the pGL3 vectors. The vectors contain the firefly luciferase reporter gene. To increase the signal of the 97LCR construct relative to the empty vector (pGL3-Basic; Promega), particularly in C33A cells, the pGL3-Basic vector was modified by insertion of a 103 bp HPV-16 fragment (nt 2428–2530) in the KpnI site in front of the DNA fragment of interest. The inserted segment contains seven ATG codons, rendering translation of the luciferase gene from messengers originating from the vector very unlikely. The segment was PCR-amplified with Pfu polymerase (Stratagene) using specific HPV-16 primers containing restriction sites for KpnI and stop codons in both directions in all three reading frames (see Table 1). HPV-16 fragments to be tested for promoter activity were PCR-amplified with Pfu polymerase and specific primers containing restriction enzyme sequences for KpnI (sense) and Nhel (antisense). The fragments were cloned into the KpnI and Nhel sites in the MCS of the pGL3-Basic vector, ensuring that the first downstream ATG codon was the initiation codon for the luciferase gene (Fig. 4).

Cloning of the selected short DNA fragments into the KpnI and Nhel sites of the unmodified pGL3-Enhancer vector (Promega) was also done by PCR amplification as described above. All the DNA fragments were designed without any ATG codons between the putative cap site and the 3’-end of the fragment (Fig. 5). The mutations were designed based on a transcription factor search using Tess-String-Based Search: Basic Query Results Transfac v.3.2 and MatInspector v.2.2 on the HPV-16 sequence. The program listed transcription factors predicted to bind in the HPV-16 genome. All constructs were sequenced to verify that both the 3’- and 5’-ends were correctly inserted into the vector. The point mutations introduced in the three mutated nt 489–556 fragments were also verified by sequencing.

Luciferase assay. Two days after transfection, the cells were washed with PBS and lysed for 90 min in 300 µl Cell Culture Lysis reagent (Promega) per dish. Cell lysate (20 µl) was mixed with 50 µl luciferin substrate (Promega) and measured in a TD-20/20 luminometer (Turner Designs) with 3 s delay and 12 s integrate period. Each plasmid construct was tested in triplicate in at least two independent experiments with two different DNA preparations (a minimum of 12 transfections). The promoter activities of the pGL3-Basic constructs were all calculated relative to the activity of the 97LCR construct which was included in every experiment. The activities of the pGL3-Enhancer constructs were calculated relative to the empty pGL3-Basic vector.

In vitro transcription. The plasmid pBRHPV-16 [pBR322 containing the full-length HPV-16 genome linearized with BamHI (nt 6153)] was digested with PstI (nt 7010), Nhel (nt 281), SpeI (nt 1463) and BstXI (nt 2900). The three HPV-16 genome fragments covering the early ORFs were purified from an agarose gel. Each fragment (1-5 µg) was incubated with 0.4 mM ATP, 0.4 mM UTP, 0.4 mM CTP, 0.016 mM GTP, 10 µCi [γ-32P]GTP, 3 mM MgCl2 and 8 units HeLa nuclear extract (Promega) at 30 °C for 1 h. The RNA was extracted with phenol–chloroform, the pellet was resuspended in nuclease-free water and the products were analysed on a 6% urea–polyacrylamide gel. Markers were made by T7 transcription of the pALTER-1 vector (Promega) linearized with EcoRV (nt 642), and the pALTER-Ex2 vector (Promega) linearized with HindIII (nt 164) and EcoNI (nt 1259) in the presence of [γ-32P]GTP.

CapFinder PCR cDNA synthesis and cloning. The CapFinder PCR cDNA Synthesis kit (Clontech) is designed to amplify full-length cDNA from mRNA with the included oligo(dT) primer and a special 5’-primer (Chenchik et al., 1998). The 5’-primer has at its 3’-end three guanosines which are able to bind to the small (dC) stretch synthesized by the MMLV reverse transcriptase. The kit was used with 2 µg total RNA from the HPV-16 positive cell line CaSk i and specific HPV-16 antisplice primers instead of the included oligo(dT) primer (see Table 1). The best result was achieved by using Pfu polymerase for the secondary amplification steps. Otherwise, the standard protocol was followed. The specific HPV-16 antisplice primers contained a Nhel site. The cDNA was digested with Nhel and cloned to Xhol and Smal in the MCS of pBluescript II KS (+) (Stratagene), blue/white selected and sequenced.

Sequencing. For sequencing, the T7 Sequenase DNA polymerase kit version 2.0 (Amersham) was used according to the manufacturer’s protocol. Products were separated on 6 % urea–polyacrylamide gels and exposed to a Bio Max MR X-ray film (Kodak).

Total RNA purification. Total RNA was prepared using Ultraspec (BioTecx Laboratories) according to the instructions given by the manufacturer.
Table 1. The names, coordinates and sequences of the primers used for PCR of HPV-16 DNA fragments and for CapFinder PCR

The underlined bases represent restriction enzyme recognition sites for KpnI (GGTACC) or NheI (GCTAGC), both written in the 5’ to 3’ direction. Bold letters represent point mutations introduced into the wild-type HPV-16 DNA sequence.

<table>
<thead>
<tr>
<th>Name</th>
<th>HPV-16 sequence</th>
<th>Primer sequence (5’ to 3’ direction)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Forward (sense primers)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7103a</td>
<td>7103–7120</td>
<td>GCATGGACGGTACCCCAACACCTCATCTACCG</td>
</tr>
<tr>
<td>28a</td>
<td>28–45</td>
<td>GCATGGACGGTACCGGCTGAAACCTGAGCCG</td>
</tr>
<tr>
<td>201a</td>
<td>201–222</td>
<td>GCAAGCAACAGTTACTGGAGC</td>
</tr>
<tr>
<td>272a</td>
<td>272–291</td>
<td>GCATGGACGGTACCAGGAAAATCCATATGCTGATG</td>
</tr>
<tr>
<td>448a</td>
<td>448–463</td>
<td>GCATGGACGGTACGCAAGACATCTGGAC</td>
</tr>
<tr>
<td>489a</td>
<td>489–504</td>
<td>GCATGGACGGTACCGGTGTTGAGGACG</td>
</tr>
<tr>
<td>549a</td>
<td>549–564</td>
<td>GCATGGACGGTACCGGTAACTGATATCATG</td>
</tr>
<tr>
<td>615a</td>
<td>615–632</td>
<td>GCATGGACGGTACCGAACATGGATCTTACTG</td>
</tr>
<tr>
<td>2428a</td>
<td>2428–2445</td>
<td>GCATGGACGGTACCTAGGTAGGTAGGATGCTACAGTGC</td>
</tr>
<tr>
<td><strong>Reverse (antisense primers)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>102b</td>
<td>102–85</td>
<td>GCATGGACGGTACCCGCTTCTTTTGTGC</td>
</tr>
<tr>
<td>141b</td>
<td>141–124</td>
<td>GCATGGACGGTACCGTCAATCTCTGGTGC</td>
</tr>
<tr>
<td>222b</td>
<td>222–205</td>
<td>GCATGGACGGTACCACACATACATGGATATGTCCTC</td>
</tr>
<tr>
<td>293b</td>
<td>293–272</td>
<td>GCATGGACGGTACCACACATACATGGATATGTCCTC</td>
</tr>
<tr>
<td>448b</td>
<td>448–427</td>
<td>GCATGGACGGTACCACACATACATGGATATGTCCTC</td>
</tr>
<tr>
<td>496b</td>
<td>496–478</td>
<td>GCATGGACGGTACCACACATACATGGATATGTCCTC</td>
</tr>
<tr>
<td>556b</td>
<td>556–537</td>
<td>GCATGGACGGTACCACACATACATGGATATGTCCTC</td>
</tr>
<tr>
<td>565b</td>
<td>565–549</td>
<td>GCATGGACGGTACCACACATACATGGATATGTCCTC</td>
</tr>
<tr>
<td>622b</td>
<td>622–606</td>
<td>GCATGGACGGTACCACACATACATGGATATGTCCTC</td>
</tr>
<tr>
<td>676b</td>
<td>676–658</td>
<td>GCATGGACGGTACCACACATACATGGATATGTCCTC</td>
</tr>
<tr>
<td>712b</td>
<td>712–694</td>
<td>GCATGGACGGTACCACACATACATGGATATGTCCTC</td>
</tr>
<tr>
<td>2530b</td>
<td>2530–2514</td>
<td>GCATGGACGGTACCACACATACATGGATATGTCCTC</td>
</tr>
<tr>
<td><strong>Mutant oligonucleotides</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mut1a</td>
<td>489–556</td>
<td>CCGGGCTCGGTGAGGGCGGTACACGTAATGTCTTGTGCAGA</td>
</tr>
<tr>
<td>Mut1b</td>
<td>556–489</td>
<td>CTAAGCCAGCTGGTCTCTACCTGAGGTGATCTCGCA</td>
</tr>
<tr>
<td>Mut2a</td>
<td>489–556</td>
<td>CCGGGCTCGGTGAGGGCGGTACACGTAATGTCTTGTGCAGA</td>
</tr>
<tr>
<td>Mut2b</td>
<td>556–489</td>
<td>CTAAGCCAGCTGGTCTCTACCTGAGGTGATCTCGCA</td>
</tr>
<tr>
<td>Mut1 + 2a</td>
<td>489–556</td>
<td>CCGGGCTCGGTGAGGGCGGTACACGTAATGTCTTGTGCAGA</td>
</tr>
<tr>
<td>Mut1 + 2b</td>
<td>556–489</td>
<td>CTAAGCCAGCTGGTCTCTACCTGAGGTGATCTCGCA</td>
</tr>
<tr>
<td><strong>CapFinder</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CapSwitch II</td>
<td>--</td>
<td>AAGCAGTGTTATCAACCGAGATCGGCGG</td>
</tr>
</tbody>
</table>

- **Plasmid purification.** Plasmids were purified with the Qiagen Plasmid Midi kit following the procedure of the manufacturer. The concentration of the DNA was determined spectrophotometrically (Uvikon 922; Kontron Instruments) and verified by agarose gel electrophoresis.

**Results**

**Transcriptional analysis of the HPV-16 early genome**

To search for promoters in the early region of the HPV-16 genome, three fragments of DNA (nt 7010–281, nt 281–1463 and nt 1463–2900) covering the LCR and part of the early region of the genome were analysed for initiation of transcription in a HeLa cell nuclear extract. The DNA fragments were obtained by restriction enzyme digestion of the HPV-16 genome as described in the Methods. A linearized control DNA construct with a CMV promoter (Promega) produced a transcript of the predicted size (363 bp) (Fig. 1, lane 1). The HeLa nuclear extract was free of endogenous transcriptionally active DNA as proven in a reaction free of externally added DNA (Fig. 1, lane 2). The DNA fragment nt 281–1463, which
Fig. 1. In vitro transcription of HPV-16 DNA using HeLa nuclear extract. Transcription was monitored by incorporation of $\alpha$-32P]GTP in transcripts initiated from promoters present in the HPV-16 genome. The positive CMV control fragment supplied with the HeLa nuclear extract gave the expected fragment of 363 nt (lane 1). The negative control without DNA is shown in lane 2. The DNA fragment covering nt 281–1463 gave a transcript of approximately 1000 nt (lane 3). The DNA fragment nt 201–622 made by PCR gave a transcript of 120–140 nt (lane 4). The marker was made by in vitro transcription with $\alpha$-32P]GTP of digested pALTER-1 and pALTER-Ex2 vectors and the fragment sizes are marked to the left of lane 1 and in the lane marked M. The transcripts were assayed on a 6% polyacrylamide gel.

covers part of the E6, the E7 and the 5′-end of the E1 ORFs, showed a product of approximately 1000 nt indicating a transcriptional start site(s) in the region of nt 500 of the HPV-16 genome (Fig. 1, lane 3). The fragment nt 1463–2900 did not appear to have any transcriptional start sites (data not shown). This region covers the 3′-end of the E1 ORF and the beginning of the E2 ORF.

To confirm the presence of a transcriptional start site in the vicinity of nt 500, a second in vitro transcription experiment was conducted using a Pfu PCR-amplified fragment of DNA. The fragment nt 201–622 gave rise to a transcript of approximately 120–140 nt in two separate experiments (Fig. 1, lane 4). This transcript corresponded to a transcriptional start site (on the sense strand) at nt 480–500, which is in agreement with the previous in vitro transcription experiment (Fig. 1, lane 3).

Identification of transcriptional start sites by CapFinder

The transcriptional start points in the E6/E7 region of the genome were further analysed with the aid of the CapFinder technique (Clontech) using total RNA isolated from CaSki cells. The method is similar to the RACE technique. Instead of the 3′ oligo(dT) primer supplied by the manufacturer, we used specific antisense HPV-16 primers: 712b, 565b and 293b (Table 1). We performed three separate Pfu PCR amplifications...
of the cDNA with the CapSwitch II primer as the upstream primer and each of the three downstream primers. The PCR amplifications identified several DNA fragments corresponding to possible transcription initiation sites when analysed by agarose gel electrophoresis. The fragments amplified with primer 712b (corresponding to initiation sites nt 542 and nt 611), 565b (corresponding to nt 482) and primer 293b (corresponding to nt 97) are shown in Fig. 2(a). Sequence analysis of the cloned fragments amplified with the 293b primer showed the main transcriptional start site at nt 97 (Fig. 2b). The cap sites at nt 482 and nt 542 were found in PCR fragments amplified with the 565b primer and the 712b primer (Fig. 2c, d). As for the nt 97 cap site, these transcripts started with an adenine. The transcriptional start site at nt 542 is situated immediately upstream of the E7 ORF. Three cDNA clones revealed this transcriptional start site, which has not been shown previously. A new transcriptional start site in the E7 ORF at nt 611 was identified and the first nucleotide was a cytosine (Fig. 2e).

**Promoter activity in keratinocyte cell lines**

All DNA constructs were made by cloning of PCR-amplified DNA fragments of the HPV-16 genome into two different promoter-free luciferase reporter vectors, pGL3-Basic and pGL3-Enhancer containing the simian virus 40 (SV-40) enhancer. The pGL3-Basic was modified by insertion of a ‘nonsense’ segment, as described in the Methods, which reduced the luciferase activity caused by transcription initiation in the vector itself to about one-third of the activity of the empty vector. This was the negative control in Figs 4, 5 and 6. The activity of the pGL3-Control plasmid, which has an SV-40 promoter and enhancer, when measured relative to the 97LCR construct, was shown to be dependent on the cell line used (Fig. 3). The pGL3-Control plasmid was approximately threefold more active than the 97LCR construct in SiHa and HeLa cells, but in C33A cells the ratio was 18-fold (Fig. 3). The pGL3-Control plasmid was, however, 17-fold more active than the empty pGL3-Enhancer vector in all three cell lines (also shown in Fig. 3). The cell lines C33A, SiHa and HeLa were selected for the present study.

**Transcriptional activity of the promoters**

To characterize the responsible promoters and gain information on their activity, a series of reporter constructs was designed carrying different upstream regions in front of each identified cap site. The HPV-16 DNA constructs designed for each cap site and cloned into the pGL3-Basic vector are shown in Fig. 4. Two series of constructs were tested. One began at nt 7103 with the entire LCR and the early region including the cap site; these were named after the cap site followed by LCR. The second began at nt 272 of the HPV-16 genome; these were named after the cap site followed by L.

The constructs were transiently transfected into C33A, SiHa and HeLa cells and the luciferase activity, reflecting the transcriptional activity, was measured relative to the activity of 97LCR. In the HPV-negative C33A cells, the promoter activity of both 482LCR and 542LCR was about 60% (Fig. 4). The activity was similar for the L constructs. The 611LCR construct had approximately 20% activity. An analogous construct made for the known start site at nt 670 (670LCR) also showed 20% activity as illustrated.

When the same constructs were analysed after transfection into SiHa or HeLa cells, only the 482LCR and 542LCR constructs were active with a relative luciferase activity of 20–40% (Fig. 4). To gain information about the impact of upstream ATG codons on translational activity, the 3′-end of the 97LCR construct was extended to nt 141, thus including one ATG codon upstream of the luciferase start codon, and to nt 222 where four ATG codons were present. The presence of four ATG codons reduced the luciferase activity threefold (Fig. 4).

**Promoter activity in the presence of the SV-40 enhancer**

To characterize the promoter regions, DNA fragments of approximately 60–70 bp containing the identified cap sites were cloned into the pGL3-Enhancer vector with the SV-40 enhancer. All activities were measured relative to the empty
vector. The activity of the \( P_{97} \) promoter in the absence of the LCR (28–102E construct) was 150–200% of the activity of the empty vector in all three cell lines tested (Fig. 5). Analysis of the \( P_{97} \) promoter activity using the 272–496E construct and the truncated fragment in 448–496E showed that the upstream sequences were necessary for activity. The promoter activity of the 272–496E construct in C33A and HeLa cells was approximately 200% and in SiHa cells was approximately 600%, whereas the truncated fragment was inactive in all three cell lines. The upstream sequence in 272–448E was as active as the full-length fragment in 272–496E in C33A and SiHa cells, but in HeLa cells, the 272–496E construct was less active than the 272–448E construct.

The \( P_{542} \) promoter in the 489–556E construct showed an activity of approximately 350% in C33A, SiHa and HeLa cells. When the \( P_{482} \) promoter region was present (272–556E), the SiHa cells showed an activity of 700%. The activity of this fragment in C33A and HeLa cells was approximately 350% and thus equivalent to the activity of the 489–556E construct. The 549–622E construct with the transcription initiation site at nt 611 showed no activity in the cell lines tested (Fig. 5). The fragment might be too short but if more of the upstream sequence was added, the sequence would overlap the \( P_{542} \) region. The \( P_{670} \) promoter in the 615–676E construct showed the same activity (200%) as the \( P_{97} \) construct (28–102E) in SiHa and HeLa cells but it was less active in C33A cells (Fig. 5).

**Mutational analyses of \( P_{542} \)**

A computer search of the HPV-16 sequence showed that the fragment nt 489–556 had potential binding sites for several transcription factors. Point mutations were introduced into a region that had abundant possibilities for consensus sequences as follows. Mut1 was mutated at positions nt 532 and 535 (A to C) and Mut2 was mutated at positions nt 552 (A to C) and nt 553 (G to T). Mut1 + 2 was mutated in all four positions. The three mutated nt 489–556 fragments were cloned into the pGL3-Enhancer vector and analysed in all three cell lines. The mutation Mut1 reduced promoter activity strongly only in SiHa cells; this was also the case for Mut2. The Mut1 + 2 abolished activity in all three cell lines (Fig. 6). These data indicate that the activity of \( P_{542} \) was dependent on binding of transcription factors to specific sequences.
**Fig. 5.** The activity of constructs in which fragments with only one or two transcription start sites were inserted in the pGL3-Enhancer vector. The map location of the potential promoters relative to the E6 and E7 ORFs is shown in the top left. The DNA fragments tested are illustrated below the map on the left; those with the same initiation site are filled in with the same pattern. Start site nt 97, 28–102E; start site nt 482, 272–496E and 448–496E; 272–448E contains the upstream fragment; start site nt 542, 272–556E and 489–556E; start site nt 611, 549–622E; start site nt 670, 615–676E. Luciferase activity measured relative to the pGL3-Enhancer vector in all three cell lines is shown on the right and each column is the mean of at least 12 measurements. The negative control is the modified pGL3-Basic vector. SD is marked for each column. Filled-in bars, SiHa cells; open bars, HeLa cells; hatched bars, C33A cells.

**Discussion**

CaSki cells have approximately 600 integrated copies of the HPV-16 genome linked head-to-tail and this line was therefore selected as the source of HPV-16 RNA (Smotkin & Wettstein, 1986; Smits et al., 1991). Our investigations have revealed two novel putative intragenic promoters in HPV-16, which have not been described previously. These putative promoters are located upstream of and in the E7 ORF, respectively. Transcriptional start sites in the E6/E7 region have been previously reported for HPV-16, but the corresponding promoter regions have not been characterized (Doorbar et al., 1990; Higgins et al., 1992; Grassmann et al., 1996). The major cap site at nt 97 was first discovered in CaSkI cells (Smotkin & Wettstein, 1986) and later also in the HPV-16-positive cell lines W12 (Doorbar et al., 1990) and HPKII (Rohlfs et al., 1991).

Our *in vitro* transcription experiments identified transcriptional start sites corresponding to approximately nt 100 (data not shown) and nt 500. However, CapFinder experiments using total RNA from CaSki cells confirmed both the presence of the P_{97} promoter and the existence of additional transcriptional start sites at nt 482, nt 542 and nt 611. The nt 482 site is probably identical to the nt 480 transcriptional start site observed in raft cultures of the HPV-16-positive cell line KG with both primer extension and RACE cDNA analysis (Grassmann et al., 1996). The promoter corresponding to the nt 480 transcriptional start site was not further characterized. We therefore included this region of the genome in the promoter analysis using the luciferase reporter gene and the cell lines SiHa, HeLa and C33A. Our data support the proposal that within fragments nt 272–496 (P_{482}) and nt 489–556 (P_{542}) there are two individual, transcriptionally active promoters.

The analysis of the promoter constructs in the pGL3-Basic vector where the entire upstream region of the genome was present including the LCR, showed activity of the promoter constructs 482LCR and 542LCR in all three cell lines. It should be noted, however, that the HPV-negative cell line C33A was less prone for transcription of HPV promoter-driven sequences than the HPV-positive cell lines SiHa and HeLa. This was
supported by several observations. Firstly, the activity of the modified pGL3-Basic vector relative to that of the 97LCR showed a high background activity of 10% in C33A cells (Fig. 4). Secondly, the activity of the pGL3-Control vector, which was driven by the SV-40 promoter and enhancer relative to the 97LCR, indicated that the 97LCR was more active in the HPV-positive cells than in the C33A cells (Fig. 3). Thirdly, the LCR constructs were 10–19 times more active than the 482L and 542L constructs in HPV-positive cells demonstrating a strong influence of the HPV enhancer region in these cells (Fig. 4). All three observations indicate that cells already adapted to HPV give the most reliable results for HPV-16 promoter activity. This is in agreement with the observation that the activity of the HPV-31b LCR correlates to the level of AP1, which is reported to be low in C33A cells (Kyo et al., 1997).

The presence of ATG codons upstream of the reporter gene start codon in the LCR constructs also influenced the transcription of the reporter gene to some extent. The presence of one ATG codon (construct 97LCR + 1ATG) did not affect the promoter activity, but four ATG codons (construct 97LCR + 4ATG) did reduce the transcription initiation to approximately one-third in all three cell lines (Fig. 4). This was in accordance with the observations of Stacey et al. (1995), who found that the presence of four upstream ATG codons present in a polycistronic messenger for the HPV-16 E7 gene left a residual activity of only 10% compared to that of a monocistronic E7 messenger. This observation revealed that the activities measured for the 482LCR and 542LCR constructs are most likely due to promoters other than P97, as at least four or more ATG codons are present upstream of the two transcription initiation sites. The promoter activity of the short DNA fragments nt 272–496 (with the nt 482 cap site) and nt 489–556 (with the nt 542 cap site) inserted into the pGL3-Enhancer vector confirmed the presence of two new promoters which were active in all three cell lines (Fig. 5). The 28–102E construct with P97 showed a surprisingly low activity, which could not be increased if the construct was made longer (1–102E construct; data not shown). This might be because of the presence of a recently identified inhibitory element from nt 7903–15 (O’Connor et al., 1998).

The construct 272–496E with the nt 482 cap site was twice as active in SiHa cells as in C33A and HeLa cells and it should be noted that the truncated construct 272–448E was at least as active (Fig. 5). It is therefore likely that there is an additional promoter present in this region.

As one of our goals was to characterize this P94 promoter, the 489–556E construct was mutated in regions where consensus-like sequences for transcription factors were predicted by a computer search. The two mutations Mut1 and Mut2, as well as the combination of these, reduced or completely abolished activity thus indicating that transcription factors are bound in this region of the promoter, although they are still not characterized. The data thus indicate that the transcription start site at nt 542 identifies a separate TATA-less promoter. In HPV-11, a promoter located in the E6 ORF (P11) was reported to be responsible for E7 protein expression (Cheng et al., 1995) and the HPV-16 P94 could very well govern expression of the E7 protein, since the first ATG codon
following the 542 cap site is the E7 start codon positioned at nt 562. Considering the consensus sequence for initiation of translation in higher eukaryotes (for review, see Kozak, 1989), the ATG at nt 562 is regarded to be in a quite strong initiator context. The transcription initiation site identified at nt 611 within the E7 ORF has not been shown before, but a potential cap site at nt 626 has been suggested (Doorbar et al., 1990). The 611LCR construct, as well as the nt 549–622 fragment, had no significant activity in the cell lines tested but the fragment might be either too short or the promoter might be differentiation-dependent, as reported by others for P\textsubscript{670} (Grassmann et al., 1996).

In conclusion, our data show the presence of a potential new E7-specific promoter, P\textsubscript{542}, from which an E7 transcript might be initiated. The previously identified P\textsubscript{180} is also a potential E7 promoter where the transcript holds two minicistrons upstream of the E7 ATG codon at nt 562 (Grassmann et al., 1996). If the E7 protein is translated from both, transcripts will need to be further analysed.

The work was supported by scholarships from The Danish Cancer Society to T.H.B. and B.G. Equipment and materials for the study were made available with support from The Danish Medical Research Council, Lawyer G. C. Kremer and wife Foundation, The Danish Cancer Society, The Biotechnology Program, The Novo Foundation, Bent Bogh and wife Inge Bogh Foundation and The A. P. Møller and Chastine McNyKinney Møllers Foundation.

**References**


Received 23 June 1999; Accepted 12 August 1999