A promoter within the E6 ORF of human papillomavirus type 16 contributes to the expression of the E7 oncoprotein from a monocistronic mRNA

Jacob A. Glahder, Christina N. Hansen, Jeppe Vinther, Birger S. Madsen and Bodil Norrild

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

Human papillomavirus type 16 (HPV-16) has the capacity to transform human primary keratinocytes. Maintenance of the transformed phenotype requires constitutive expression of the oncoproteins E6 and E7. The low-risk HPV types express E7 from monocistronic mRNA, but for the high-risk types, no mRNA that encodes E7 as the first open reading frame (ORF) has been identified. We recently identified a transcription initiation site within the E6 ORF of HPV-16 at nt 542. In the present study we have characterized the P542 promoter, which putatively controls monocistronic expression of E7. The monocistronic mRNA is not very abundant, but we have shown that an E7–luciferase fusion protein can be expressed in SiHa cells from a monocistronic HPV-16 transcript initiated at nt 542. The monocistronic mRNA expresses E7–luciferase more efficiently than the most abundant in vivo-like mRNA E6*IE7, initiated by P97 and spliced from nt 226 to 409. Furthermore, the translation initiation of E7 is most abundant from the monocistronic mRNA. We have also shown that the P542 promoter is downregulated by the transcription factor activator protein 4 (AP-4) and the differentiation-dependent factor hSkn-1a, both binding downstream of the transcription initiation site. In conclusion, we have found that P542 is a relatively weak promoter compared with P97 and may be downregulated in differentiated epithelial cells.

INTRODUCTION

Almost all cervical cancers are human papillomavirus (HPV) positive (Walboomers et al., 1999; Munoz, 2000; Ishiji, 2000). HPVs are small DNA viruses, with a double-stranded genome of ~8000 bp (International Agency for Research on Cancer, 1995). They are very host- and tissue-specific, infecting only humans, and more than 85 HPV types have been identified and sequenced (McGlennen, 2000). All of the identified HPV types are strictly epitheliotrophic, infecting either cutaneous or mucosal epithelium. The cutaneous subgroup are mostly benign, whereas anogenital infections caused by the mucosal subgroup often are associated with malignant progression (zur Hausen, 1996). Based on their carcinogenic potential, the anogenital HPVs can be divided into two subgroups: low risk, which includes HPV-6 and -11, associated with benign genital lesions; and high-risk, including HPV-16, -18, -31 and -33, associated with invasive carcinomas of the cervix (Phelps et al., 1998; zur Hausen, 1999; de Villiers, 1994). HPV-16 is the most prevalent among the high-risk viruses and is considered the prototype of this group (zur Hausen, 1996).

The oncogenic potential possessed by the high-risk viruses resides in their ability to produce the E6 and E7 oncoproteins (zur Hausen, 2000; Stoppler et al., 1998). The binding of E6 to the cellular tumour-suppressor protein p53 leads to ubiquitin-dependent proteolytic degradation of p53 (Werness et al., 1990; Scheffner et al., 1993). The E7 oncoprotein binds to the nuclear tumour-suppressor protein retinoblastoma (pRb), disrupting the pRb/E2F interaction in early G1 (Dyson et al., 1989; Vousden, 1993).

The transcription of all early mRNAs in the HPV-16 genome, including the E6 and E7 oncogenes, is reported in several studies to be controlled by a single promoter, termed P97, just upstream of the E6 open reading frame (ORF) (Smotkin & Wettstein, 1986; Smotkin et al., 1989; Doorbar et al., 1990; Grassmann et al., 1996). In addition to this constitutively active promoter, a differentiation-dependent promoter has been identified within the E7 ORF in the early region of the HPV-16 genome, termed P670. This promoter is possibly responsible for the expression of an E1^E4 fusion protein, together with the E5 oncoprotein (Higgins et al., 1992; Grassmann et al., 1996).
Furthermore, two novel promoters were recently localized within the L1 ORF together with two transcriptional initiation sites in the HPV-16 long control region (LCR) (Tan et al., 2003). The most abundant mRNAs transcribed from HPV-16 derive from P97 (Smotkin & Wettstein, 1986). Furthermore, they are all bi- or polycistronic, encoding two or more proteins (Doorbar et al., 1990; Rohlfis et al., 1991; Baker & Calef, 1995). The E7 ORF is the second ORF after the E6 ORF in these mRNAs. Eukaryotic mRNAs are usually monocistronic and translational initiation occurs by ribosomal scanning. The presence of an upstream ORF generally inhibits expression from a second downstream ORF. The expression of E7 has been shown to be substantially lower from the polycistronic mRNA than from an E7 monocistronic control construct (Tan et al., 1994a; Stacey et al., 1995). This can be explained by a leaky scanning mechanism, where ribosomes initiate scanning at the 5’ end, but somehow bypass the E6 start codon and continue down to the start codon for E7 (Stacey et al., 2000).

Gene expression from polycistronic mRNAs seems to be a common feature of the high-risk HPV types (Smotkin et al., 1989). However, in some of the low-risk types, such as HPV-6b and -11, promoters have been identified within the E6 ORF giving rise to mRNAs for monocistronic E7 expression (Smotkin et al., 1989). In HPV-16, we have identified a putative promoter termed P542 located within the E6 ORF (Braunstein et al., 1999) and we have shown in the present study that monocistronic expression of E7 can be obtained from P542. The promoter P542 seems, unlike P97, to be TATA-less. We also searched for an initiator (Inr) element. An Inr element contains a pyrimidine (Y)-rich core sequence, Y(Y)AA(N/T/A)AY, surrounding the transcription initiation site and participates in transcription initiation (Lo & Smale, 1996). A strictly conserved Inr could not be found surrounding nt 542, although the most critical positions for functionality of Inr (−1, +1 and +3) were present. Just downstream of P542 is a motif known as an E box with the sequence CAGCTG. The ubiquitous transcription factor activator protein 4 (AP-4) binds to this motif (Ellenberger et al., 1994; Ou et al., 1994). We have shown that AP-4 binds to the motif in the P542 region and downregulates promoter activity. Others have identified a binding site for hSkn-1a just downstream of the E box and overlapping the E7 start codon (Kukimoto & Kanda, 2001). hSkn-1a is a differentiation-dependent transcription factor (Andersen et al., 1993, 1997), which we found to downregulate the P542 promoter activity.

In summary, we report for the first time on the ability of the P542 promoter to regulate transcription initiation of E7 mRNA at nt 542. This promoter is relatively weak but is able to regulate the E7 protein expression, and the amount of E7 protein expressed is up to 5-fold higher from a monocistronic P542-driven mRNA than the expression from a spliced P97-driven mRNA. In addition, the transcription of E7 is downregulated by the transcription factor AP-4 and the differentiation-dependent transcription factor hSkn-1a.

**METHODS**

**RNA preparation.** Total RNA from CaSki cells was isolated and purified with the RNasey mini kit (Qiagen) according to the manufacturer’s instructions. The isolated RNA was treated with DNase I Amplification Grade (1 U µl⁻¹; Invitrogen).

**Generation of radioactive antisense RNA probes.** The PI and PII probes from nt 489–670 and 489–762 of HPV-16, respectively, were PCR amplified (see Table 1 for primers). The PI fragment was cloned into the pBluescript II SK(+) vector (Stratagene; KpnI and XhoI) and the PII fragment into the pCR 2.1-TOPO vector (Invitrogen; XhoI and HindIII). They were subsequently linearized by BamHI digestion. [α-32P]UTP-labelled RNA was transcribed with T7 RNA polymerase using the Maxiscript kit (Ambion). The probes were gel purified and used in ribonuclease protection assays.

**Ribonuclease protection assay (RPA).** The RPA III kit (Ambion) was used according to the recommendations of the manufacturer. Total RNA (10–20 µg) was mixed with ~100 000 c.p.m. of the probes. The probes were incubated with RNase T1/A diluted 1:50 or 1:25. The protected probe products were precipitated and separated by electrophoresis on a 5% acrylamide gel with 8 M urea. Markers were made using the RNA Century Marker Template Set (Ambion) using [α-32P]UTP.

**Cell culture, transient transfections and functional assay.** The HPV-16-positive human cervical cell line SiHa, the human embryonic kidney cell line HEK and the monkey kidney fibroblast cell line COS were cultivated in Dulbecco’s modified Eagle’s medium (1885 MEM) supplemented with 10% foetal calf serum, 1% L-glutamine and 1% penicillin and streptomycin.

For luciferase reporter assays, SiHa cells were plated into six-well dishes at approximately 2 × 10⁵ cells per dish 24 h before transfection. Transfections were performed using serum-free medium and LipofectAMINE PLUS (Gibco BRL). To each dish was added 0.5 pmol DNA. For co-transfections, 0-2 pmol reporter vector and 0-4 pmol expression vector were added to each dish. At 48 h post-transfection the cells were harvested. To determine the luciferase activity, cells were lysed according to the recommendations of the manufacturer (Promega) and the activity measured in a TD-20/20 luminometer (Turner Designs).

For Western blots, HEK cells were used for transfections. Cells (4 × 10⁶) were plated in 10 cm Petri dishes 24 h before transfection. Cells were thereafter transfected with 8 µg DNA per dish and harvested 48 h post-transfection (see Immunoprecipitation below).

**Plasmid constructs and site-directed mutagenesis.** All probes were derived from a full-length clone of HPV-16 in pBR322. The constructs were cloned into the promoterless pGL3 enhancer vector (pGL3E) (Promega) containing the SV40 enhancer and the firefly luciferase reporter gene.

To create the E7 fusion constructs with luciferase, the start codon of the luciferase gene was removed to ensure that the E7 start codon at nt 562 was the only downstream translational start site. The NcoI restriction site was used just upstream of the luciferase start codon. The resulting single-stranded overhang, containing the luciferase start codon, was removed by mung bean nuclelease, which created a blunt end, followed by digestion with KpnI. Amplification of the first HPV-16 construct was carried out with an antisense primer designed to contain a new NheI site (576b, Table 1) and a sense primer with a KpnI site (499a, Table 1). The construct was blunt-end ligated in...
The luciferase ORF. The following fusion constructs were PCR-amplified with *Pfu* polymerase and cloned in between the *Kpn*I and the new *Nhe*I site. In one assay, the constructs contained the entire E7 ORF (up to nt 858; constructs E7, E6*IE7, E6E7mSD and E6E7wt), whereas in another they contained only the first 15 bp of the E7 ORF (to nt 576) and were designated 's' for short. Primers used for PCR amplifications are listed in Table 1. The pGL3E vector was modified with an insertion of a 103 bp HPV-16 fragment (nt 2428–2530) to avoid read-through from translational start sites in the vector (Braunstein *et al.*, 1999). As well as this modification, the negative control vector did not contain a start codon for luciferase.

To create the two constructs E6E7mSD [nt 28–858; with a mutated splice donor in the E6 ORF (nt 222, 226 and 227, all G→R)] and E6*IE7 (small splice product from P97, nt 28–226^409–858 with nt 226–409 removed), site-directed mutagenesis was performed by PCR. All constructs were sequenced by the Sanger dideoxy method using the Sequenase version 2.0 DNA sequencing kit (USB). Tadahito Kanda kindly provided the hSkn-1a clone in the pCMV1 vector.

**Immunoprecipitation.** The constructs used in this assay were E7, E6*IE7, E6E7mSD and E6E7wt containing the full-length E7 ORF and immunotagged with a FLAG-epitope at the C terminus (made with the antisense primer E7 flag; Table 1). In this assay, the constructs were subcloned from the pG3E vector and inserted into the pcDNA3.1+ expression vector (Invitrogen; *Kpn*I and *Xba*I) downstream of a cytomegalovirus (CMV) promoter controlling transcription. HEK cells were transiently transfected with 8 mg DNA per 10 cm dish, and 48 h post-transfection whole-cell extracts were obtained by adding 300 ml RIPA buffer to each dish. The protein concentrations were determined using the Coomassie Plus Reagent Assay (Pierce). Protein extract (2–5 mg) was precipitated with 15 ml E7 antibody (HPV-16 E7, ED17; 200 mg ml^-21; Santa Cruz). The samples were separated by 12.5% SDS-PAGE and blotted onto PVDF membranes (Bio-Rad), E7 was detected using the same antibody as before in conjunction with a chemiluminescence kit (ECL reagents; Amersham).

**Electrophoretic mobility shift assay (EMSA).** The double-stranded oligonucleotides used in the EMSA described here are listed in Table 1. They were all designed with a 5′ overhang of various lengths, but containing at least one T in each overhang.

### Table 1. Primers used for PCR of HPV-16 DNA fragments and for EMSA probes

<table>
<thead>
<tr>
<th>Name</th>
<th>HPV-16 sequence</th>
<th>Primer sequence (5′→3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Forward (sense) primers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>489a</td>
<td>489–504</td>
<td>GCATGGACGGTGACGGGCGGTCGTTGGACCGG</td>
</tr>
<tr>
<td>409a SPLICE</td>
<td>211–226; 409–428</td>
<td>GTCATGCAGGACGGTGACGGGCGG</td>
</tr>
<tr>
<td>28a</td>
<td>28–45</td>
<td>GCATGGACCGGTTACCCAGGCGG</td>
</tr>
<tr>
<td>mSDs</td>
<td>217–240</td>
<td>GCACATTGTTATATGACTTTG</td>
</tr>
<tr>
<td><strong>Reverse (antisense) primers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>576b</td>
<td>576–557</td>
<td>GCTAGCTGTATCTCAGTACAGTATA</td>
</tr>
<tr>
<td>858b mSTOP</td>
<td>888–834</td>
<td>GCATGGACGCTAGCAGTGGTTTTCGAGAAGAAGAGG</td>
</tr>
<tr>
<td>226b SPLICE</td>
<td>422–409; 226–205</td>
<td>GACAGCTTAATACAGCCCTACGTCGAGTACTGTTG</td>
</tr>
<tr>
<td>mSDb</td>
<td>240–217</td>
<td>GCAAAGGTCATATAATACAGCCCTACGTCGAGTACTGTTG</td>
</tr>
<tr>
<td>E7 flag</td>
<td>858–837</td>
<td>AGTCATGTTTGGGCTGAAGAACAGATG</td>
</tr>
<tr>
<td><strong>EMSA oligonucleotides</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AP-1</td>
<td>–</td>
<td>ATCGGCTGTAGTACAGC</td>
</tr>
<tr>
<td>WTI</td>
<td>519–541</td>
<td>AACTAGCTAGTACGCGCCTT</td>
</tr>
<tr>
<td>MutI</td>
<td>519–541</td>
<td>GGTGACATCATCAAGAACAGT</td>
</tr>
<tr>
<td>WTII</td>
<td>540–561</td>
<td>ACGTCTAGTACGTCATTGACATC</td>
</tr>
<tr>
<td>MutII</td>
<td>540–561</td>
<td>GGTGACATCATCAAGAACAGT</td>
</tr>
<tr>
<td>AP-4</td>
<td>–</td>
<td>TCTCTTTGGGCTGAAGAACAGAT</td>
</tr>
<tr>
<td><strong>RPA primers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pla and PlLa</td>
<td>489–524</td>
<td>GCATGGACGGTGACGGGCGGTCGTTGGACCGG</td>
</tr>
<tr>
<td>Plb</td>
<td>670–647</td>
<td>GCATGGACGGTGACGGGCGGTCGTTGGACCGG</td>
</tr>
<tr>
<td>PlbII</td>
<td>762–740</td>
<td>GCATGGACGGTGACGGGCGGTCGTTGGACCGG</td>
</tr>
</tbody>
</table>
Tris/HCl, pH 7-5, 5 mM MgCl₂, 60 mM KCl, 0-5 mM EDTA, 5% Ficoll-400, 2-5% glycerol, 1 mM DTT, 0-1 mM PMSF) were incubated with 2 μg poly(dI-dC) (Amersham), 11 μg HeLaScribe nuclear extract (Promega), 15 μg SiHa nuclear extract (made as previously described; Dignam et al., 1983) or ~2 μl in vitro-translated AP-4. Competition assays were performed with ~30-fold excess of DNA unless otherwise indicated. Samples were run on 5-6% non-denaturing polyacrylamide gels at 125–200 V for 3 h.

In vitro translation of the transcription factor AP-4. The full-length AP-4 clone (a kind gift from Richard B. Gaynor) under the control of a T7 promoter in the pcDNA3.1+ vector was used to synthesize the protein using the TNT-coupled transcription–translation system as recommended by the manufacturer (Promega).

Statistical method. To calculate possible differences between the different luciferase constructs, a Student’s t-test was used. The activities measured followed the normal distribution. The level of significance is indicated in the figure legends.

RESULTS

Verification of a transcription initiation site at nt 542 by RNase protection assay

The transcription initiation site at nt 542 was previously identified by the cap-finder 5′-RACE (rapid amplification of cDNA ends) technique and by primer extension (Braunstein et al., 1999). To verify further this transcription initiation site, we carried out RPAs with antisense RNA probes spanning nt 489–670 (PI) and nt 489–762 (PII) of HPV-16 (Fig. 1a). Several fragments were protected when the two probes were hybridized with CαSk1 RNA but not with yeast RNA (Fig. 1b). Fragments representing the abundant mRNA spliced from nt 226–409 and possible unspliced mRNA (181 bp), as well as the mRNA spliced up to nt 526 (145 bp) and a fragment of 130 bp, were protected by PI. This 130 bp fragment corresponds to transcriptional initiation at nt 542. The expected bandshift was seen when the PII probe was used (Fig. 1b, dotted lines). An RPA with a probe that covered the entire early region of HPV-16 also confirmed our observation and verified the identification of the transcription initiation site at nt 542 (M. Rosenstierne, J. Vinther, C. N. Hansen, M. Prydsøe & B. Norrild, unpublished results).

Analysis of the promoter activities upstream of the E7 ORF

In vivo, the E7 protein might be expressed from at least three splice variants derived from P97 (Smotkin et al., 1989; Grassmann et al., 1996). These mRNAs are either unspliced, called E6E7, or spliced within the E6 ORF, called E6*IE7 (large splice) and E6*IE7 (small splice), of which the latter is the most abundant (Bohm et al., 1993; Doorbar et al., 1990). To analyse the promoter activity of P542 in relation to P97, we constructed reporter plasmids that carry the E7 ORF fused to the N terminus of and in-frame with, the reporter gene firefly luciferase in the pG13E vector. This vector contains the SV40 enhancer, which in these experiments was used instead of the HPV-16 LCR to make it possible to compare the P97 core sequence with the P542 core sequence. We generated three constructs by PCR to match the physiologically produced mRNAs (Fig. 2a). All constructs started at nt 28 and included the core of P97, but excluded a putatively transcriptionally repressing YY1 site and a papillomavirus silencing motif at nt 7893–7911 (O’Connor et al., 1996, 1998). Furthermore, we designed a monocistronic reporter construct starting at nt 489 that included the P542 promoter but excluded the transcriptional start site at nt 480 (Fig. 2a) (Grassmann et al., 1996; Braunstein et al., 1999). To investigate E7 expression from these constructs, SiHa cells were transfected and luciferase activity, here reflecting both transcription and translation efficiency, was measured. In these experiments the spliced construct (E6*IE7) displayed ~0.7-fold activity compared with the monocistronic E7 construct, whereas the wild-type E6E7wt and the splice-donor mutant E6E7mSD showed ~3- and 4.5-fold the activity of the monocistronic E7, respectively (Fig. 2a).

The entire E7 ORF fused to the N terminus of the luciferase gene could influence the enzymatic activity of the fusion protein and we therefore made reporter constructs that only contained the first five amino acids of the E7 ORF (3′ end at nt 576), fused to the N terminus of the luciferase gene (Fig. 2b). The activities of these constructs were much higher than in the previous assay and the differences among the constructs were more pronounced. Here, the activity from the spliced E6*IE7s construct was ~0.2-fold the activity of the monocistronic E7s construct. The wild-type E6E7wts and the splice donor mutant E6E7mSDs were ~2.4-fold and ~3.4-fold the activity of the E7s construct (Fig. 2b).

In summary, the monocistronic E7s expressed significantly higher levels of luciferase than E6*IE7. The E6E7mSD construct expressed most E7–luciferase, followed by E6E7wt in both experiments. The E6*IE7 constructs showed the lowest E7–luciferase expression.

Translational efficiency of the different early mRNAs encoding E7

The luciferase activities measured from the constructs made in the pG13E vector are dependent on both the efficiency of transcription from the different promoters and the translation initiation of the different mRNAs. To investigate only the efficiency of translational initiation from the different E7 mRNAs, the constructs shown in Fig. 2(a) were cloned downstream of a complete CMV promoter in the pCDNA3.1+ vector. The relative amount of E7 protein obtained from these constructs indicates translation initiation and mRNA stability, as the transcription from the CMV promoter should be identical for the different constructs and much higher than the HPV promoter-driven transcription. All constructs had a FLAG immunotag added at the C terminus of the E7 protein, which increased the molecular mass of E7 to approximately 18 kDa (Fig. 3a). This epitope was used as a backup
to detect E7 if the E7-specific antibody was inefficient. Immunoprecipitation of E7 and subsequent Western blot analysis was performed and bands were quantified relative to the most abundant expression, which was obtained from the monocistronic E7 construct (Fig. 3). The translational efficiency from the spliced E6*IE7 was only 0.2-fold that of the monocistronic E7 (Fig. 3b). The protein expression from the E6E7wt and the E6E7mSD constructs was much lower, but could be detected after a longer exposure (Fig. 3c).

In summary, these experiments indicate that the monocistronic E7 construct and the spliced E6*IE7 construct are much more efficiently translated than the other constructs containing P97.

**Binding of cellular factors to the promoter P542 region**

A promoter is dependent on *cis* regulatory elements for functionality. Previously it has been shown that point
mutations in the P542 region affect the promoter activity of P542 (Braunstein et al., 1999). These sequences could be part of transcription factor binding sites and we therefore investigated the binding of cellular factors to the promoter region. EMSA experiments with two different doublestranded probes from the P542 region were conducted.

**Fig. 2.** Luciferase reporter assay. (a) The DNA constructs containing the entire E7 ORF cloned in frame with the luciferase gene into the modified pGl3E vector. (b) The DNA constructs contain only the first 15 bp of the E7 ORF fused to luciferase. The negative control is a modified and mutated empty pGl3E vector. In both (a) and (b), the activity from the monocistronic construct with a 5' end at nt 489 was set to 1. Activities from the other constructs are relative to this. Measurements were performed on two different DNA batches for each construct with a total of 12 measurements. Standard deviations are marked for each column. *Significantly different, \( P < 0.02. \)
(Fig. 4). The first probe, WTI, covered sequences upstream of the transcription initiation site, while the second probe, WTII, contained downstream sequences. The two probes only overlapped by four nucleotides (Fig. 4a). When SiHa nuclear extract was added, four complexes appeared with each wild-type probe (Fig. 4b, black arrowheads). Based on a computer search for possible transcription factor binding sites in the two regions, two probes containing point mutations critical for putative protein binding sites were designed (Fig. 4a, asterisks). The mutations in the first region covered by probe MutI (Fig. 4b, lane 4) did not influence complex formation, whereas mutations in the second region covered by probe MutII (Fig. 4b, lane 6) abolished one of the complexes (grey arrowhead) indicating that a protein-binding site had been affected.

To test the specificity of the different complexes, a series of competition experiments was carried out. In the presence of unlabelled WTI probe, two of the four complexes were eliminated (Fig. 4c, lane 4, grey arrowheads). This was also the result when the WTII probe was competed with an unlabelled WTII probe (Fig. 4c, lane 8). The P542 region contains two weak binding sites for the activator protein 1 (AP-1) located in each of the two probe regions (WTI, nt 525–532; WTII, nt 555–562). Therefore, a probe containing a consensus AP-1 site (TGAGTCA; Rösl et al., 1997) was used as a competitor to eliminate possible complex formation between AP-1 and the two wild-type probes. The AP-1 competition probe abolished one of the specific complexes binding the WTI probe (Fig. 4c, lane 5, white arrowhead) and another of the specific complexes binding the WTII probe (Fig. 4c, lane 9, white arrowhead). This indicates binding of AP-1 in both regions. Finally, the mutated HPV-16 probes were used as competitors to eliminate all specific complex formations not affected by the mutations. When the MutI probe was used as competitor, the lower of the specific complex bands in the WTI region was eliminated (Fig. 4c, lane 6, white arrowhead), indicating that the binding site represented by the upper of the two specific shifts was mutated. In the WTII region, the upper of the two specific complex bands was eliminated (Fig. 4c, lane 10), indicating that the binding site for the lower band complex was mutated (indicated by a dot).

**Binding of the transcription factor AP-4 to the promoter P542 region**

The so-called E box, which has been found in several promoters, is a sequence motif with the consensus CANNTG, where the two central nucleotides can vary (Ellenberger et al., 1994; Ou et al., 1994; Hu et al., 1990). The motif binds the basic helix–loop–helix (bHLH) class of proteins (Murre et al., 1989a, b; Ellenberger et al., 1994; Lawrenz-Smith & Thomas, 1995). As well as the bHLH proteins, the bHLH leucine zipper proteins can also bind the E box, including the ubiquitous factor AP-4. This protein binds to the consensus motif CAGCTG (Ou et al., 1994), which is the one present in the HPV-16 P542 region (Fig. 4a). To investigate the binding of AP-4 to the promoter region, the AP-4 protein was *in vitro*-translated from an expression vector containing the full-length clone (kindly provided by Richard B. Gaynor). Although the protein is 37 kDa, it migrated in an SDS gel corresponding to a size of 46 kDa (Fig. 5a, lanes 4–6), as previously reported (Mermod et al., 1988). EMSA analysis with *in vitro*-translated AP-4 protein showed a complex with the WTII probe and the complex migrated approximately to the same position as the second and third band formed with SiHa nuclear extract (Fig. 4b, lane 5), which in this specific experiment co-migrated (Fig. 5b, lane 3). The MutII probe showed no AP-4 complex at the same position (Fig. 5b, lane 6). When increasing amounts of AP-4 were added to the WTII probe, saturation was reached between 4 and 8 µl (Fig. 5c, lanes 5 and 6).
Fig. 4. EMSA analysis of the P542 region. The HPV-16 fragment spanning nt 519–563 was analysed by EMSA. (a) The HPV-16 region from nt 511–570 is shown, where the two EMSA probes are situated upstream (WTI, nt 519–543) and downstream (WTII, nt 540–563) of the transcription initiation site. The initiation site P542 is indicated by an arrowhead and each point mutation introduced is marked with an asterisk. The E box is shown in light grey and the hSkn-1a binding site in dark grey. (b) EMSA using both probes. Lane 1 shows the WTI probe without nuclear extract. A consensus AP-1 probe was used as a positive control (lane 2). Lanes 3 and 4 show the WTI and MutI probes, respectively, with black arrowheads indicating the four bands. Lane 5 shows the WTII probe with the four bands indicated by arrowheads. One of the bands disappears when the MutII probe is used, as indicated by the grey arrowhead (lane 6). (c) Competition EMSAs using both probes. The controls are the same as in (b). When the WTI probe (lane 3) is competed with unlabelled WTI (lane 4), the two middle bands disappear (grey arrowheads). When AP-1 is used as a competitor (lane 5), one band disappears (white arrowhead). The same band disappears when the MutI probe is used as a competitor (lane 6). When the WTII probe (lane 7) is competed with unlabelled WTII (lane 8), the two middle bands disappear (grey arrowheads). When AP-1 (lane 9) or MutII (lane 10) are used as competitors, one band disappears (white arrowheads). The dot indicates a band corresponding to a protein binding in the mutated region. SiHa nuclear extract was used in both EMSAs. n.e., Nuclear extract.
**Fig. 5.** For legend see page 3438.
carrying the AP-4 consensus binding site were performed. Although a computer search indicated exclusive binding of AP-4 to this probe, several complexes appeared in the EMSA analysis when SiHa nuclear extract was added (Fig. 5d, lane 2). Competition with the homologous unlabelled probe made one band disappear, indicating that this band represented specific binding (Fig. 5d, lane 3, black arrowhead). When the WTII probe was used as the competitor, the same band disappeared, indicating that this was the AP-4 complex (Fig. 5d, lane 4). The reverse assay was carried out (lanes 5–7). Both competitor probes abolished the two specific complexes formed with WTII (Fig. 5d, lane 5, grey arrowheads). Thus, this analysis, together with the previous EMSA experiments, indicates that AP-4 binds to the promoter region downstream of P542.

**Downregulation of P542-mediated E7 expression by AP-4 and hSkn-1a**

The mutations introduced into the P542 region have previously been shown to repress the promoter activity in luciferase reporter assays in SiHa cells (Braunstein et al., 1999). Thus, one would expect the mutations to eliminate the binding of activating proteins or induce binding of repressing proteins. To test whether AP-4 had an effect on the promoter activity, co-transfections of the E7s reporter construct (Fig. 2b) and an AP-4 expression vector were carried out in COS cells (Fig. 6). An ~12-fold decrease in luciferase activity was seen in this experiment. In similar assays performed in SiHa cells, the decrease in luciferase activity was approximately 2- to 6-fold, depending on the amount of expression plasmid used (data not shown). COS cells were also co-transfected with a construct expressing the differentiation-dependent transcription factor hSkn-1a (kindly provided by Tadahito Kanda) using the same luciferase reporter construct (Fig. 6). HSkn-1a has previously been shown to bind to HPV-16 (nt 560–569) and upregulate the differentiation-dependent promoter P670 (Kukimoto & Kanda, 2001). Since the binding region overlaps the E7 start codon, the factor was expected to affect the expression from P542. This factor, as with AP-4, inhibited P542 activity, decreasing the luciferase activity by approximately 7-fold.

**DISCUSSION**

The high-risk HPV types are believed to express the E6 and E7 oncoproteins from polycistronic mRNAs produced by one main promoter. However, our laboratory has previously identified a putative promoter in HPV-16, P542, located just

![Image](https://example.com/figure6.png)

**Fig. 6.** Analysis of the P542 promoter activity exposed to AP-4 and hSkn-1a. COS cells were co-transfected with the HPV-16 E7s fragment fused with luciferase and twice the amount of either an AP-4 or an hSkn-1a expression vector and the luciferase activity was measured. The values from the negative control (a modified and mutated empty pG3E) were subtracted from the other values. The light and dark bars show the results from co-transfections with empty vector and vector expressing protein, respectively. Solid bars represent co-transfections with AP-4. Hatched bars represent co-transfections with hSkn-1a. The results using the empty vector were set to 100% in each assay and the other results were relative to these. Measurements were performed three to nine times for each construct.
upstream of the E7 ORF (Braunstein et al., 1999). Promoters upstream of the E7 ORF have also been identified in low-risk HPV types such as HPV-6b and -11. These promoters, which control the major transcripts in these HPV types, are capable of regulating monocistronic expression of E7 (Smotkin et al., 1989).

The transcriptional start site at nt 542 was identified by the cap-finder 5′-RACE and primer-extension methods (Braunstein et al., 1999) and we have further verified this promoter using RPAs. Our results suggest that HPV-16 E7 can be expressed from a monocistronic mRNA driven by P542 in a significant amount comparable with the E7 expressed from the P97 transcripts (Fig. 2). These constructs use the SV40 enhancer instead of the HPV enhancer to facilitate higher expression from weak promoters. This only influences promoter activity and does not create additional cryptic transcriptional start sites (Rosenstierne et al., 2003). The RPA experiments also show that in CaSki cells the P542 mRNA is expressed at a much lower level than the P97-derived mRNAs. The relatively high level of E7–luciferase expression from the E7s construct compared with a P97 construct (Fig. 2b) is thus most likely dependent on higher translation levels from the monocistronic mRNA than from the P97 mRNAs. This notion is supported by the results of our CMV promoter-driven constructs, demonstrating that the P542 mRNA is much more translationally active than the P97 mRNAs (Fig. 3c).

Splicing within the E6 ORF was previously believed to facilitate efficient translation of E7 (Smotkin et al., 1989). However, Stacey et al. (1995, 2000) showed that the three mRNAs E6*IE7, E6*IIE7 and the unspliced E6E7 produced equal amounts of E7, whereas a synthetic monocistronic construct initiating at nt 553 was 10-fold more efficient. In our study, we have found that the monocistronic mRNA initiated from P542 leads to the highest level of E7 protein, whereas both the spliced and unspliced polycistronic mRNAs lead to less E7 protein. For efficient translation, a leader sequence of more than 10 bp is the most optimal (Kozak, 1989). The leader sequence in HPV-16 between P542 and the start codon for E7 is 20 bp, which is short but most likely translationally favourable. It is therefore likely that the probably scarce mRNA with a 5′ end at nt 542 compensates for low abundance by more efficient translation, probably because of a short but optimal leader length.

The unspliced mRNA with the full-length E6 ORF in this study expressed E7–luciferase efficiently. The transcripts from this wild-type construct were not further investigated, but probably lead to a mixture of both spliced and unspliced mRNAs. The relatively high expression from the E6E7mSDs and E6E7wts constructs (Fig. 2b) could be explained by the presence of a cis regulatory element in the nt 226–409 (intron) region. In fact, the region spanning nt 272–332 has been shown to be a cis regulatory element (Rosenstierne et al., 2003), which also may affect the P97 or P542 activities. Still, we cannot exclude the possibility that the high activity of expression relates to the spliced nature of both mRNAs. The E6E7mSD construct might use a cryptic splice donor and thereby reduce the leader length of the E7–luciferase mRNA.

To characterize further the regulation of P542, we performed EMSAs and EMSA competition experiments. We found that the transcription factor AP-4 binds to the P542 region. In addition, possibly one or two AP-1 transcription factors also bind to this region. The region only contains very weak AP-1 binding sites compared with the AP-1 consensus binding site, but it has previously been shown that weak AP-1 sites found close to AP-4 binding sites can be functionally relevant (Mermod et al., 1988).

Previously, our laboratory has shown that point mutations in the P542 region downregulate promoter activity (Braunstein et al., 1999). Here we have shown that the binding of AP-4 is prevented by these mutations and that AP-4 binding represses the P542 promoter activity. The AP-4 binding site is located immediately downstream of the transcription initiation site and this might repress transcriptional initiation by blocking the assembly of the TFIID complex, although E boxes downstream of transcriptional initiation sites recently have been shown to activate transcription (Leach et al., 2003). Another possibility is that addition of excess amounts of AP-4 protein competes with endogenous activating transcription factors binding in the same region. Furthermore, we have shown that co-transfections with a vector expressing the differentiation-dependent transcription factor hSkn-1a decreased the activity of P542 several fold. This indicates that P542 might be silenced in differentiated cells, where hSkn-1a is upregulated (Andersen et al., 1993, 1997).

The main HPV-16 promoter P97 has a consensus TATA box (TATAAA) approximately 30 bp upstream of the transcription start site. In contrast, there is no obvious TATA box upstream of P542. Some TATA-less promoters contain upstream Sp1 sites, which facilitate the assembly of the TFIID complex (Weis & Reinberg, 1997). In fact, an Sp1 site is located upstream of P542 (at nt 491–496), only deviating in one position from the consensus binding site (GGGCGG; Gloss & Bernard, 1990). Only 3 bp downstream from this site (nt 500–511) there is a possible E2 binding site, although it deviates from the consensus (ACCGNgCGGT; Tan et al., 1994b) in two of eight sites.

A nuclear matrix attachment region element is situated in the region from nt 105 to 560 (Stunkel et al., 2000). This element was shown to inhibit transcription when the virus genome was episomal but activate transcription of integrated viral genomes. Virus integration, as well as a complex interplay between binding of several transcription factors, could therefore have an effect on the activity of P542.

In conclusion, we have shown that: (i) P542 is a weak promoter; (ii) the amount of E7 protein expressed from a monocistronic P542-driven mRNA is higher than from any
of the P97-driven mRNAs investigated in our study; and (iii) the transcription factors AP-4 and hSkn-1a function as repressors of E7 expression from P542, possibly in cooperation or competition with other cellular factors. Thus, the TATA-less promoter P542 could potentially influence the regulation of the oncoprotein E7 expression and be of importance in the viral life cycle and the increase in E7 expression found in HPV-16-induced cancers.

ACKNOWLEDGEMENTS

We thank Dr Richard B. Gaynor, Simmons Cancer Center, Division of Molecular Virology, Department of Internal Medicine, University of Texas Southwestern Medical School, Dallas, Texas, for providing us with the full-length AP-4 clone in the pEx10x vector. Furthermore, we thank Dr Tadahito Kanda, Division of Molecular Genetics, National Institute of Infectious Diseases, Toyama, Siniiku-ku, Tokyo, Japan, for donating the full-length hSkn-1a clone in the pCMV vector. This work was funded by a grant from The Danish Cancer Society, together with grants from the following foundations: Therese Marie Hansen, Købmand Sven Hansen, Lægevidenskabens Fremme, Ragnar Ibsen’s Legat, Novo, Fabrikant Einar Willumsen and Den Owensenske Fond.

REFERENCES


