2. Transcriptional regulation of Human Papillomavirus genes

Marcela Lizano-Soberón¹, Imelda Martínez-Ramírez¹ and Adriana Contreras-Paredes¹

¹Unidad de Investigación Biomédica en Cáncer. Instituto Nacional de Cancerología, México
Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México

Abstract. Human papillomaviruses (HPVs) comprise a large family of pathogens of the cutaneous or mucosal squamous epithelia. Despite the knowledge of the genomic sequences of many HPV types, the regulation of viral gene expression is still uncertain. Viral gene expression in infected cells depends on cell differentiation and is tightly regulated at the transcriptional and post-transcriptional levels. All papillomaviruses contain a 400-850 bp long noncoding region (referred to as long control region LCR), which is positioned between the L1 and E6 genes and contains most of the elements that regulate transcription of the viral genome. Only one strand of viral DNA is transcribed and more than one promoter is involved in the transcription initiation process, resulting multiple mRNAs with several open reading frames. The early promoters of HPV-16 (P97) and HPV-18 (P105) have been well characterized, as well as the late differentiation-regulated promoter of HPV-16 (P670). The LCR exhibits recognition sites for positive and negative cellular factors, such as AP-1, Sp-1, YY1 and Oct-1; and binding sites for the viral E2 proteins. The LCR also contains glucocorticoid response
elements (GRE) and an enhancer component that specifically activates transcription in epithelial cells through the synergism between a diversity of transcription factors. Gene regulation occurs at transcriptional and post-transcriptional levels. Papillomavirus are transcribed as bicistronic or polycistronic forms that frequently contain two or more ORFs, and are polyadenylated at the early and late poly(A) sites. This Chapter will focus on the mechanisms of transcriptional regulation of HPV genes.

**Introduction**

Human Papillomaviruses (HPVs) show a great diversity; however their genomic organization is well conserved. These viruses are non-enveloped and contain a double-stranded circular DNA genome of approximately 8 kb in size divided according to its function into three main regions: Early (E), Late (L) and the Long Control Region (LCR or upstream regulatory region URR). Two polyadenylation sites separate these regions. Papillomaviruses have a polycistronic transcription pattern, which yields a series of mRNAs processed by alternative splicing and polyadenylation (1). The early region encode six open reading frames that are expressed early during infection, generating the E1, E2, E4, E5, E6 and E7 proteins, and more recently described, the fusion protein E8-E2C also appear to have a regulatory function in viral transcription and replication. Two L genes are expressed lately producing L1 major and L2 minor capsid proteins. The genetic regulation resides mainly in the non-coding region known as the LCR.

The Early proteins are involved in critical functions such as HPV genome replication and transcription, control of the cell cycle, apoptosis and immune modulation. Some important functions of HPV proteins are described in Table 1.

A differential expression of viral early and late genes occur in oncogenic HPV infections, due to the activation of the viral early promoter in undifferentiated keratinocytes and the viral late promoter in highly differentiated keratinocytes (2, 3, 4). E1, E2, E5, E6 and E7 proteins are expressed along the viral cycle throughout the differentiating epithelia, with reduced expression in the more differentiated layers. Expression of E1, E4, L1 and L2 proteins occur in the upper spinous and more differentiated granular layers of the epithelium (5, 6).

The LCR is the most variable region within Papillomaviruses, consisting of 850 to 1000 base pairs. It carries the origin of replication with binding sites for E1 and E2 viral proteins, and multiple recognition sites for cellular transcription factors, important for transcription initiation through the RNA polymerase II (16). Promoters of HPV-18, HPV-31 and HPV-16 have been well characterized (17, 18).
### Table 1. HPV proteins and functions.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>Viral DNA replication, ATPase, and DNA helicase (7,8).</td>
</tr>
<tr>
<td>E2</td>
<td>Regulation of viral gene transcription and DNA replication. During early stages of viral infection, represses the transcription of the oncogenes E6 and E7 (9).</td>
</tr>
<tr>
<td>E4</td>
<td>Disruption of cytokeratin networks/cell growth arrest (10).</td>
</tr>
<tr>
<td>E5</td>
<td>Modifies the activity of the epidermal growth factor receptor, induces inflammatory cell signaling pathway, contributes to malignant transformation complementing the roles of E6 and E7, enhances angiogenesis and proliferation and may contribute to cancer progression (10,11).</td>
</tr>
<tr>
<td>E6</td>
<td>Transformation and tumorigenesis (binds to p53 amongst other proteins). In addition, a large number of interactions between the E6 and cellular proteins have been reported (12,13.)</td>
</tr>
<tr>
<td>E7</td>
<td>Transformation (binds to pRb amongst other proteins), alter cell cycle control through interactions with histone deacetylases, cyclins and cyclin-dependent kinase inhibitors (14).</td>
</tr>
<tr>
<td>L1</td>
<td>The only Major capsid protein (15).</td>
</tr>
<tr>
<td>L2</td>
<td>Minor capsid protein (15).</td>
</tr>
</tbody>
</table>

### HPV genome integration

Papillomavirus infections are usually transient or result in benign lesions. Nevertheless, sometimes infections with “high risk” or oncogenic HPVs such as HPV-16, HPV-18 and HPV-31 persist, promoting the development of malignant lesions.

A common consequence of cancer associated to persistent infections with high-risk HPVs is the integration of the viral genomes into the cell genome, frequently disrupting the viral E2 regulator gene (19,20). In the normal viral life cycle, HPV genomes exist in a circular or episomal state and are retained in the basal cells of the squamous epithelium at approximately 50 to 100 copies per cell (21). In this state, viral gene expression is mainly regulated by E2 protein (22). High-level expression of E6 and E7 oncogenes is supposed to occur only in suprabasal post-mitotic cells (23,24), where the viral oncogenes induce unprogrammed re-entry into S-phase, activating the host replication machinery. This cell environment is needed for viral genome amplification prior to virion synthesis. In
productive infections, viral oncogenes fail to transform since they are expressed in cells destined to be lost due to the constant renewal of the cervical squamous epithelium. Meanwhile, oncogene expression in basal epithelial cell inhibits differentiation and induces chromosomal instability, representing the potential cells for transformation events (25,26).

In the majority of cervical carcinomas an over-expression of E6 and E7 oncogenes is observed in cells in which truncated viral genomes are integrated into the host DNA. HPV genomic integration correlates with increased viral gene expression and cellular growth advantage, promoting the development of epithelial displasia and eventually cancer (27). Viral integration often occurs in the E1 and E2 region and results in the loss of negative feedback control of oncogene expression. In this situation, interactions of E6 and E7 with cellular suppressor proteins such as p53 and pRb increase, and consequently suppressor functions are inhibited (28,29). Many different interactions of oncogenes E6 and E7 with cellular proteins have been described and the majority of those interactions seem to affect cell cycle regulation (30,31). This is the case for hDlg, which is a member of the membrane-associated guanylate kinase (MAGUK) family of proteins, involved in the inhibition of cell cycle (32). In vitro assays as well as results obtained through cervical cancer cell line analysis, have demonstrated that E6 proteins from high risk HPV, bind to hDlg promoting its degradation via the ubiquitin proteosome pathway (33,34). It has also been found that E6 intratype variants of HPV-18 differentially modulate hDlg degradation, rebounding in levels of the suppressor protein PTEN and activated PKB, which are important proteins in survival signal transduction pathways (35). Through these interactions HPV-18 E6 variants are also able to strongly upregulate other elements implicated in cell proliferation.

Genomes of “Low-risk” or non-oncogenic HPVs, such as HPV-6 and HPV-11, are not found integrated. These viral types can induce some type of dysplasia or benign condyloma acuminata, but it is established that they do not promote cancer (26).

The mechanism of HPV integration is not well understood, since it is not a normal event in viral cell cycle. Therefore integration must represent a chance occurrence; probably through events that increase the frequency of double-strand breaks in host and viral DNA (36).

**HPV transcriptional regulatory region**

The integration of the viral genome into the cellular genome is an essential feature for carcinogenesis. The integrated viral segment contains the regulatory region of the virus, known as “long control region (LCR)” or
upstream regulatory region (URR)” and the sequence encoding E6 and E7 oncogenes. These oncogenes are transcribed from the promoter located in the regulatory region.

The LCR is responsible for carrying out the genetic regulation of the papillomaviruses. Its sequence from 800-1000 bp is extremely variable depending on the viral type; although common regulatory elements, such as AP-1 and SP-1, are conserved (37). The LCR contains binding sites for the E2 protein on the viral DNA replication origin, where E1 and E2 proteins play a very important role. HPV early promoters, such as P105 for HPV-18 and P97 for HPV-16 (38); and P90, P270 and P680 for HPV6 and HPV11, are active throughout the process of viral genome replication (39,40). Those promoters contain the cellular enhancer region whose transcriptional activity depends on cellular transcription factors like, AP-1, NF-1, TEF-1 and Oct-1, whose combination may confer some tissue specificity, (41). Within the papillomavirus genome, late promoters are activated in the upper epithelial or granular layer, through the cell differentiation process. Some of those promoters are located within the E7 gene, such as P670 and P142, for HPV-16 and HPV-31, respectively (42). Additionally, HPV-16 promoter activities have been also detected within the 5’ region of the LCR, also in the L1 region, and within E6 ORF (43,44), whose importance are not yet known although they appear as weak promoters (45).

Within the first portion of the LCR is also the TATA box where transcriptional factors bind in conjunction which polymerase II, carrying out transcription of the viral genome.

LCR transcriptional regulation

Transcriptional regulation by E2 viral protein

The E2 viral protein has approximately 360 amino acids and contains three functional domains: an N-terminal transactivation domain, an internal “hinge” domain, and a DNA binding domain (DBD) located at the C-terminus (Fig. 1). Both the C-terminal and N-terminal domains are relatively well conserved within the papillomaviruses. Based on the results of X-ray crystallography, structural and molecular models of both the N- and C-terminal domains have been constructed (46,47). E2 binds as a dimer at DNA-binding sites through its C-terminal DBD. Some evidence suggests that the transactivation domain mediates linking activity between E2 molecules bound at distant E2-binding sites, and as a consequence DNA loops are formed (48,47).
Figure 1. E2 Protein domains. A region of approximately 200 amino acids at the N-terminus (thick line) is relatively well conserved among the papillomaviruses and encodes the transactivation domain. The 85 C-terminal amino acids comprise the DNA-binding and dimerization domain (DBD).

The papillomavirus E2 protein has demonstrated many functions in the viral cycle, including roles in transcriptional regulation and in viral DNA replication (49). Besides these historical roles, which rely on their aptitude to bind to specific viral DNA sequences, E2 also modulates the host cells through direct protein interactions to several cellular proteins, such as Brd4, cNAP1, Gps2 and p300 (50-54). E2 interacts with topoisomerase II binding protein 1 (TopBP1), which enhances its ability to activate transcription and replication (55). The binding of p53 to E2 alters its transcriptional activity and also inhibits HPV DNA replication (56,57). Finally, E2 binds to APC activators Cdh1 and Cdc20 through which the high-risk E2 proteins might induce genomic instability (58).

On the other hand, E2 regulates replication of the viral genome, due to its interaction with E1 viral protein. It also regulates transcription of viral early genes; E2 can repress E6/E7 promoter from promoter-proximal binding sites (49). E2 also plays important roles in the segregation of HPV genomes between daughter cells after cell division (59,60).

In several HPV types, a spliced RNA product, that expresses a fusion protein consisting of the small E8 domain fused to E2C (61), is produced. The functions of the E8^E2C protein have been mainly investigated in HPV-31. E8^E2C knockout HPV-31 genomes display a strong over-replication of viral genomes in short-term analyses (61). Yet, in stable cell lines, these knockout genomes are not maintained as episomes but are found integrated into the host chromosomes, suggesting that E8^E2C is required for the long-term extrachromosomal maintenance of viral genomes (61). Genetic and biochemical analyses of the 31E8^E2C protein have demonstrated that the 31E8 domain is required for transcriptional repression. This is due to the recruitment of cellular co-repressors, such as the histone deacetylase 3 (HDAC3)/N-CoR complex, by the 31E8 domain (62). E8^E2C activities may vary among different papillomaviruses, since analysis in the context of
Transcriptional regulation of HPV

HPV-16 show that 16E8^E2C is not required for stable maintenance of HPV-16 episomes (63).

The understanding of the roles of E2 proteins has been evolving over the years. Binding of E2 to DNA palindromic sequences has been associated with dimerization of the protein. E2 recognizes the consensus sequence, 5’-ACCgNNNNeGgT-3’, with nucleotide positions 4 and 9 allowing some variability (64,65). The sequence of the 4-nucleotide spacer varies by HPV type, and is thought to be critical for determining E2 binding affinity and gene regulation (66). The E2 homodimer binds the DNA by the alpha helices of each monomer by contact with two successive major grooves of the target site (67,47). Four E2 binding sites are conserved in the LCR of most papillomaviruses, assigned with numbers according to distance from the early promoter (68). Each site is differentially regulated by variable binding affinity for the E2 protein, resulting in varying replication and transcriptional effects during the viral life cycle (51).

Regarding LCR activity regulation by E2 during a viral infection, mutation of the promoter-proximal E2 binding site in HPV-31 genome induces transcriptional repression (69). In an HPV-Epstein-Barr virus (EBV) model system the HPV-16 E2 protein represses the transcription of the HPV-16 LCR (70). By other side, it has been suggested that transcriptional activity of the HPV-31 E2 protein is not required during the viral life cycle, in this context being critical the replication function of E2 (71).

HPV E2 is among the first early proteins expressed from the HPV genome involved in viral transcription, replication and maintenance of viral genomes (72,73). E2 is known to regulate the viral DNA replication. During this process the papillomavirus E1 helicase is recruited by E2 to the viral origin of replication, where it assembles into a double hexamer that begins replication of the viral genome (74,75). On the other hand, the role of E2 in the control of HPV gene expression is not clearly known. There are several ways in which the E2 proteins can either activate or repress transcription, depending on the position of the E2-binding sites and the nature of the E2 polypeptides. In one case, binding of the full-length E2 transactivator to a number of sites located upstream from the promoter activates transcription (74). In these cases, the E2 sites behave like conditional enhancer elements. On the other hand, E2 can also repress transcription, when the E2-binding motif are situated close to the transcriptional start site, overlapping with binding sites for essential cellular transcription factors. This mechanism suggests that the role of E2 transcription repressor or activator depends of binding sites within the LCR. The promoter-proximal E2 binding sites are flanked by a TATA box sequence and a Sp-1-binding site. E2 sites are located just upstream from the TATA box and binding of E2 to these sites may sterically hinder the binding of the cellular factor TFIID, and/or
interfere with the assembly of the transcriptional initiation complex (75). In addition to competing with Sp-1 and the TATA box binding protein (TBP) for binding to DNA, E2 can also interact with both of these proteins (76,77). It has been proposed that E2 levels might determine its affinity to promoter E2 binding sites, since at early stages of infection there are low levels of E2 that might increase the binding of this protein to a distal promoter, resulting in transcription activation; whereas at later times of infection higher E2 protein levels might bind to the proximal promoter E2 sites inducing to the repression of transcription. However, other HPV types E2 proteins binds equally well to all of its sites within the genome (78,79).

**HPV transcriptional regulation through cellular factors**

Papillomavirus early gene transcription is regulated by sequences located in the LCR known as cis-acting response elements. Some response elements located in HPV-16 and -18 LCRs are described in table 2. These elements bind a variety of cellular transcription factors. Is worth to mention that both, the composition of a transcriptional control region, as well as the space between different regulatory elements, pressure the transcriptional activity (80,81). Due to its high prevalence, HPV-16 and HPV-18 types have been the main models in the study of HPV transcriptional regulation.

Transcription factors which modulate LCR activity of high-risk HPVs, as SP-1, AP-1, NF-1, Oct-1, YY-1 and the glucocorticoid receptor GRE (Figure 2) (82), are also important in normal cell growth, development and differentiation.

![Figure 2. Schematic representation of the HPV-18 LCR, showing binding sites for cellular and viral regulatory proteins. Four binding sites for the E2 viral protein are shown. The origin of replication is defined by E1 together with the adjacent E2 binding sites. These sites are conserved among mucosal HPV types, as well as the Sp-1 and AP-1 response elements.](image-url)
Table 2. HPV-16 and -18 LCR response elements.

<table>
<thead>
<tr>
<th>Virus type</th>
<th>Binding site (Transcription factor)</th>
<th>Sequences</th>
<th>Nucleotide Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPV16</td>
<td>AP-1</td>
<td>TTAATCA</td>
<td>7631-7637</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGAGTCA</td>
<td>7648-7654</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTGTCA</td>
<td>7810-7816</td>
</tr>
<tr>
<td></td>
<td>NF-1</td>
<td>TTGGC</td>
<td>7473-7477</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTGGC</td>
<td>7558-7554</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTGGC</td>
<td>7591-7587</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTGGC</td>
<td>7679-7675</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTGGC</td>
<td>7711-7715</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTGGC</td>
<td>7742-7746</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTGGC</td>
<td>7770-7766</td>
</tr>
<tr>
<td></td>
<td>GRE</td>
<td>TGTACATTGTGTCA</td>
<td>7641-7655</td>
</tr>
<tr>
<td></td>
<td>Oct-1</td>
<td>AATTGCAT</td>
<td>7732-7739</td>
</tr>
<tr>
<td></td>
<td>Sp-1</td>
<td>GGGCGT</td>
<td>28-33</td>
</tr>
<tr>
<td></td>
<td>KRF</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>HPV18</td>
<td>AP-1</td>
<td>TTAGTCA</td>
<td>7608-7614</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTAGTCA</td>
<td>7798-7792</td>
</tr>
<tr>
<td></td>
<td>NF-1</td>
<td>CTGGCN5TGCAAA</td>
<td>7513-7528</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTGGCN8TTGGC</td>
<td>7569-7586</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTGGC</td>
<td>7731-7735</td>
</tr>
<tr>
<td></td>
<td>Oct-1</td>
<td>GCCTTGCAT</td>
<td>7644-765</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AATTGCAT</td>
<td>7721-7723</td>
</tr>
<tr>
<td></td>
<td>KRF-1</td>
<td>TGCATAACTATATCC</td>
<td>7647-7669</td>
</tr>
<tr>
<td></td>
<td>GRE</td>
<td>ACTCCCTA</td>
<td>7839-7853</td>
</tr>
<tr>
<td></td>
<td>Sp-1</td>
<td>AGCACATACATACTA</td>
<td>35-40</td>
</tr>
</tbody>
</table>

Positions of consensus binding sequences for each transcription factor on HPV-16 and -18 LCR.

**Sp-1**

Sp-1 is a zinc finger transcription factor that binds to GC-rich motifs of many promoters. This factor has three isoforms and is involved in many
cellular processes such as cell differentiation, cell growth, apoptosis, immune responses, response to DNA damage, and chromatin remodeling. Post-translational modifications significantly affect Sp-1 activity. In conjunction with AP-1, Sp-1 plays a critical role in transcriptional regulation of mucosal high- and low-risk HPVs. Sp-1 binding sites in HPV-16 and HPV-18 are generally found close to the transcriptional initiation site. Mutation of this site in HPV-18 leads to a strong reduction of the transcriptional activity in the URR, reinforcing its importance in transcriptional activation of E6/E7 expression (82). Sp-1 performs its transcriptional activation in cooperation with other cis elements found within the viral enhancer.

**AP-1**

AP-1 is a transcription factor constituted by heterodimers of Jun and Fos family proteins by hydrophobic interactions of “Leucine-Zippers” (83). AP-1 controls the expression of genes involved in cellular proliferation, differentiation and transformation (84). JunB and Fra2, components of AP-1, are highly expressed in keratinocyte and modulate its differentiation through regulation of several genes such as involucrin, loricrin, cytokeratins or integrins (85).

HPV-18 URR contains two AP-1 elements, one located within the constitutive enhancer region and other within the promoter-proximal region, playing a key role in activation of the HPV-18 E6/E7 promoter (82).

**Glucocorticoid Response Element (GRE)**

Steroid hormones may act as cofactors for HPV in cervical cancer development. Glucocorticoids in particular, such as the synthetic dexamethasone, have demonstrated to enhance the transformation rate of high-risk HPV-transfected cells (86,87).

HPV-11, HPV-16 and HPV-18 have active glucocorticoid response elements (GREs), since glucocorticoids activate transcription from their early promoters. In the proximal promoter region of HPV-18 LCR it is found AGCACATACATATACCT sequence, which confers dexamethasone and progesterone response. The LCR HPV-31 contains at least one functional GRE, although sequence analysis identified three potential GREs (88).
**Oct-1 and NF-1**

The Oct-1 transcription factor was among the first identified members of the POU transcription factor family. Members of this family contain the POU domain, a 160-amino acid region necessary for DNA binding to the octameric sequence ATGCAAAT (89). An Oct-1-like site located near a NF-1 element at the enhancer of HPV-16 has been identified, that even being aberrant it can activate the promoter of E6/E7. Functional differences among high-risk HPV genomic variants have been described, suggesting repercussions on their oncogenic potential. The LCR of HPV-18 variants have distinct transcriptional activities in different Cervical Cancer cell lines (90). It has been proposed that these differences may be due to variations in transcription factor binding sites KRF-1/OCT-1 (positions 7651, 7658) and OCT-1/TEF-1 (positions 7726, 7730).

Neurofibromin 1 (NF-1) is a Ras-GAP protein, which negatively regulates Ras signaling. Spliced transcript variants encoding different isoforms have been described for this gene (91,92). However, the precise cellular function of neurofibromin has yet to be clarified. Using proteomic strategies, a set of neurofibromin-associating cellular proteins was identified (92). The consensus NF-1 transcription factor-binding site is (5´- TTGGC-3´). In HeLa cells, NF-1 synergizes with Oct-1 bound to a juxtaposed site. NF-1 has an activating function in HPV its role in HPV-18 is less clear (82).

**YY-1**

The transcription factor YY-1 is a ubiquitously distributed zinc finger protein that regulates important cellular processes such as growth, differentiation and development and plays an important role in tumorigenesis (93) and HPV infection (94). It is involved in repressing (95) and activating (96) a diverse number of cellular and viral promoters. YY-1 may direct histone deacetylases and histone acetyltransferases to a promoter in order to activate or repress. In HPV-18 (97) and HPV-11 (98) YY-1 represses major early promoters when bound to a distal upstream site. In HPV-8 YY-1 appears both to repress and to activate E6 promoter (99); while in HPV-16, YY-1 repress the E6 and E7 transcription when bound upstream of the P97 start site (100).

**Post-transcriptional regulation**

HPV gene expression depends on cell differentiation and is regulated at the transcriptional and post-transcriptional levels. Papillomavirus transcripts are
transcribed as a bicistronic or polycistronic from genes frequently containing two or more ORFs, polyadenylated at an early and a late poly (A) sites.

Post-transcriptional regulation of gene expression can involve regulation at various steps in mRNA biogenesis, as splicing (101), polyadenylation (102), nuclear export (103), RNA stability (104), and translation (105). For example, addition of the 7 methyl guanosine cap stabilizes the RNA, and this is essential for export to the cytoplasm and for translation on polysomes; splicing removes introns from the pre-mRNA, joining exonic coding regions. Alternative splicing in viral RNA is a means of production of multiple proteins. Splicing is also important for nuclear export of mRNA to the cytoplasm and for appropriate translation (106).

How keratinocyte differentiation regulates HPV gene expression remains to be determined, but it is believed that infection of basal cells leads to the expression of the early region of the viral genome, producing the six non-structural viral proteins (E1, E2, E4, E5, E6 and E7). In more differentiated keratinocytes the two structural viral capsid proteins are expressed (L1 and L2) from the late region of the viral genome.

The study of the complexity of papillomavirus gene expression has been a hard task and limited knowledge exists concerning the proteins translated from each viral transcript. Much of the information has been obtained from HPV-16 transcript mapping.

**HPV RNA splicing**

RNA splicing regulates viral gene expression at the post-transcriptional level, giving rise to a bunch of possibilities in HPV coding potential of its pre-mRNAs. Extensive RNA mapping in HPV infected cells and raft tissues have led to the establishment of transcription maps from various HPV genomes (107,108), which provide important information for further HPV studies (Fig. 3). One splice donor and two alternative splice acceptor sites within the E6 transcript are unique to high-risk HPV types (109); accordingly, low-risk HPV types do not undergo splicing, being essentially monocistronic. Transcripts in HPV-16 could be spliced though three alternative 3’ splice sites, positioned in the first intron at nt 409, 526 or 742 (110); and three other alternative 3’ splice acceptor sites at nt 2582, 2709, 3358 in the second intron (111,112,113). All combinations can potentially produce more than 14 species of mRNA transcripts with various coding potentials. All the early region transcripts contain a downstream 3’exon that includes E2C/E4/E5a/E5b ORFs, while all the late region transcripts contain a 3’exon that encodes the L1 major capsid protein. Consequently, most mRNAs contain multiple ORFs, being some of them polycistronic.
Figure 3. Genome structure and transcription map of Early genes of HPV-16. A) Blue boxes represent open reading frames of early and late genes. Formation of E4 requires RNA splicing (upper dashed lines). LCR is the long control region. B) Linear form of the virus genome with promoters (arrows) and early (A_E) and late (A_L) polyadenylation sites. C) Some early RNA species derived from alternative promoter usage and alternative splicing. Exons (blue thick lines) and introns (thin lines) are exemplified on each species of the RNA, with the indicated splice site nucleotide positions.

Five small Nuclear Ribonucleoproteins (snRNP) named U, participate in viral RNA splicing removing an intron and synchronizing the binding of individual exons and splicing factors. Classical intronic splicing signals have three cis-elements consisting in a 5’splice site (5’ss), a branch site, and a 3’ splice site (3’ss) (110). The set up of the exon-intron boundary is the first step in the correct recognition of an intron 5’ss and an intron 3’ss, which involves the interaction of the 5’ss with U1 snRNP of the branch site, with U2 snRNP, and the 3’ss with other heterodimers of U2 auxiliary factors. This process requires many cellular splicing factors, such as SR proteins (serine/arginine-rich), and exonic splicing enhancers (ESEs). Moreover, regulation of HPV-16 splicing late pre-mRNA also appears to utilize splicing silencers (ESS). The presence of ESS immediately upstream of the
nt 3632 5’ss (114) may be necessary to counteract U1 recognition of the nt 3632 5’ss and enhance recognition of the early polyadenylation site (pA) at nt 4215 by the cellular polyadenylation machinery (115). Meanwhile, a 79 nt RNA regulatory element, the late regulatory element (LRE), involved in the regulation of L1 and L2 expression in response to epithelial differentiation is sited at the 3’ end of the L1 gene and extends in the late 3’ untranslated region (3’ UTR). This element represses late gene expression in differentiating epithelia and may be active in differentiated cells (116).

Alternative splicing of HPV RNA regulates the final concentration of E6 and E7 proteins of the high-risk papillomaviruses. Splicing in the E6 ORF can give rise to the smaller species E6*I, E6*II, and E6*III, affecting the amount of full length E6 protein (20). Some authors propose that this kind of splicing also regulates translation of the E7 ORF by enhancing translation initiation at the AUG in the E7 ORF (117,110,118,119)). Splicing of E6/E7 RNA has been shown to provide more E7 RNA templates and to promote E7 translation, whereas a lack of RNA splicing was found to lower the levels of E7 proteins. Nevertheless, the excess of an efficient splicing of E6/E7 transcripts could prevent the expression of enough amounts of full-length E6. Even there is no evidence of how does keratinocyte differentiation regulates viral RNA splicing, it is possible that the splicing process could be precisely controlled by cellular factors differentially expressed according to the cell differentiation status. As a result, high risk HPVs appear to limit the activities of the full-length E6 protein, under certain conditions, although it is not known whether this occurs in stage of infection dependant manner.

**DNA methylation in papillomavirus gene expression**

As a consequence of HPV integration into the host DNA, the viral genome becomes a potential target for cellular control mechanisms acting on the integration site. Thus, DNA methylation is suggested to be another mechanism involved in HPV transcriptional regulation. The introduction of methyl groups to cytosine residues at CpG dinucleotides often leads to a transcriptional silencing; therefore DNA methylation regulates cellular and viral gene expression. When the methylation is held in a promoter region, it prevents transcription factors from binding to their recognition sites. Alternatively, methylated CpG dinucleotides interact with methyl CpG binding proteins, which recruit histone deacetylase in association with chromatin. Histone deacetylation then leads to changes in the chromatin configuration inhibiting transcription.

Significance of the methylation in papillomavirus gene expression and carcinogenesis is not fully studied; even some authors showed that CpG
methylation within the binding sites of HPV16 E2 protein, disturb DNA-binding activity of E2 (120,121), affecting viral gene expression. By other side, in the host genome context CpG methylation is found to be more common in cervical cancer than in premalignant lesions (122), being a frequent event throughout the host genome. It has been reported that in HPV-16 and HPV-18 genomes, methylation occurs more often in LCR regions and part of L1 ORF region, which is not transcribed when the viral genome is integrated (123,124). Some authors propose that CpG methylation in integrated HPV DNA may prevent most viral DNA from transcription activation, leaving only few transcription units per cell (125).

**Papillomavirus RNA polyadenylation**

Viral gene expression, as well as eukaryotic gene expression, is highly regulated by RNA polyadenylation. Cleavage of the nascent transcript in the pre-mRNA and the addition of a poly(A) tail with 150-200 adenylate residues are coupled events. RNA polyadenylation is an important feature of post-transcriptional regulation of gene expression, since it confers stability to the nascent RNA and favours ribosome recruitment improving translation. At least five cellular factors are involved in RNA polyadenylation, including poly(A) polymerase (PAP), cleavage and polyadenylation specificity factor CPSF, cleavage stimulation factor (CstF), and cleavage factors I and II (126). The late transcription unit overlaps at part of the early transcription unit and contains two polyadenylation sites (early and late) (Fig. 3) which potentially are in competition with each other and are mutually exclusive events. Early to late switch in papillomavirus gene expression could be due to changes in either splicing or polyadenylation or both. In addition, negative cis-elements (NRE), spanning from the 3’ end of the L1 ORF into the late 3’ UTR regulate HPV-16 late transcripts (127). The NRE binds multiple RNA processing factors, including U2AF65, CstF-64, HuR, and SF2/ASF (128-131). Even the exact function of the HPV-16 NRE is not fully understood; it is known that two NRE-binding proteins (U1A and U2F65) may inhibit polyadenylation. The NRE may also compete with downstream CstF binding polyadenylation element for CstF, leading to a less stable polyadenylation complex and reduced polyadenylation efficiency (132).

Polyadenylation and splicing are both co-transcriptional processes that can modulate each other. A 65-nt AC rich splicing enhancer located approximately 100-nts downstream of the splice site 3358 3’ of HPV-16 stimulates splicing and up-regulates polyadenylation of the early transcripts at the early poly(A) signal (114), mainly leading to the expression of E1^E4 early transcript. By other side, L1 mRNA splicing of the nt 3632 5’ splice
site to the 5639 3’ site is supposed to compete with early polyadenylation since these two events are mutually exclusive.

**Conclusion**

Papillomavirus transcription is restricted to epithelial cells of human origin, more specifically to keratinocytes. These viruses encode multiple proteins in a very compact genome, taking advantage of alternative splicing to differentially express the proteins in a temporal manner, linked with the cell differentiation program. Viral transcription is controlled by response elements present in the long control region. Tissue specificity is modulated by epithelial cellular factors acting in the LCR transcriptional enhancer. HPV gene expression relies on a complex relationship of differential promoter use, selection of the appropriate polyadenylation site, generation of alternative splice isoforms of polycistronic RNAs transcribed from the genome, and translation efficiency of the mRNAs generated. All these processes are controlled by cellular DNA or RNA-binding factors, some of which are known to be expressed in a differentiation stage-specific manner. Significant progress has been made in the comprehension of the mechanisms of transcriptional and post-transcriptional regulation of viral gene expression. However, this knowledge must be strengthened to profoundly understand papillomavirus life cycle, host interactions and viral oncogenesis.

**References**