Regulation of Human Papillomavirus Type 16 mRNA Splicing and Polyadenylation

XIAOMIN ZHAO
Abstract

Human papillomavirus type 16 (HPV-16) is the major causative agent of cervical cancer. The life cycle of this oncogenic DNA tumour virus is strictly associated with the differentiation program of the infected epithelial cells. Expression of the viral capsid genes L1 and L2 can only be detected in the terminally differentiated epithelial cells. The studies here focus on the regulation of HPV-16 late gene expression, which is under tight regulation.

Our experimental system consisted of almost the full length HPV-16 genome driven by a strong CMV promoter. This plasmid and mutants thereof could be transfected into HeLa cells and RNA levels monitored. Using this system, we identified an hnRNP A1-dependent splicing silencer between positions 178 and 226 of the L1 gene. This silencer inhibited the use of the 3’ splice site, located immediately upstream of the L1 AUG. We speculate that this splicing silencer plays an essential role in preventing late gene expression at an early stage of the viral life cycle. We subsequently identified a splicing enhancer located in the first 17 nucleotides of L1 that may be needed to counteract the multiple hnRNP A1 dependent splicing silencers in the L1 coding region. A 55kDa protein specifically bound to this splicing enhancer. We also demonstrated that binding of the cellular factors to the splicing silencer in the L1 coding region had an inhibitory effect on expression from L1 cDNA expression plasmids.

The HPV-16 genome is divided into the early region and the late region, separated by the early poly(A) signal (pAE). pAE is used preferentially early in infection, thereby efficiently blocking late gene expression. We demonstrated that a 57 nucleotide U-rich region of the early 3’ untranslated region (3’eUTR) acted as an enhancing upstream element on the usage of pAE. We demonstrated that this U-rich region specifically interacts with hFip1, Cstf-64, hnRNP C1/C2 and PTB, suggesting that these factors were either enhancing or regulating polyadenylation at the HPV-16 pAE.

In conclusion, two regulatory RNA elements that both act to prevent late gene expression at an early stage in the viral life cycle and in proliferating cells were identified: a splicing silencer in the late region and an upstream u-rich element at the pAE.

Keywords: human papillomavirus type 16, gene regulation, splicing, polyadenylation, 3’ UTR

Xiaomin Zhao, Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden

© Xiaomin Zhao 2005

ISSN 1651-6206
ISBN 91-554-6328-2
urn:nbn:se:uu:diva-5919 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-5919)
To my parents with love
Main references

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals:


II. Xiaomin Zhao, Stefan Schwartz. Multiple splicing enhancers and hnRNP A1-binding silencers control the HPV-16 late 3’splice site *(Submitted)*

III. Xiaomin Zhao, Stefan Schwartz. Exonic splicing silencers in the HPV-16 L1 coding region inhibit gene expression in the absence of splicing *(Manuscript)*


Reprints were made with the permission of the publishers.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPS</td>
<td>branch point sequence</td>
</tr>
<tr>
<td>BPV</td>
<td>bovine papillomavirus</td>
</tr>
<tr>
<td>CBC</td>
<td>cap binding complex</td>
</tr>
<tr>
<td>CPSF</td>
<td>cleavage polyadenylation specificity factor</td>
</tr>
<tr>
<td>CstF</td>
<td>cleavage stimulation factor</td>
</tr>
<tr>
<td>CF I</td>
<td>cleavage factor I</td>
</tr>
<tr>
<td>CF II</td>
<td>cleavage factor II</td>
</tr>
<tr>
<td>CTD</td>
<td>carboxy-terminal domain</td>
</tr>
<tr>
<td>CAT</td>
<td>chloramphenicol acetyltransferase</td>
</tr>
<tr>
<td>CKIs</td>
<td>cyclin-dependent kinase inhibitors</td>
</tr>
<tr>
<td>DSEs</td>
<td>downstream elements</td>
</tr>
<tr>
<td>ESEs</td>
<td>exonic splicing enhancers</td>
</tr>
<tr>
<td>ESSs</td>
<td>exonic splicing silencers</td>
</tr>
<tr>
<td>EV</td>
<td>epidermodysplasia verruciformis</td>
</tr>
<tr>
<td>E6AP</td>
<td>E6-associated protein</td>
</tr>
<tr>
<td>EGFR</td>
<td>epidermal growth factor receptor</td>
</tr>
<tr>
<td>E2BSs</td>
<td>E2 binding sites</td>
</tr>
<tr>
<td>GT</td>
<td>guanylyltransferase</td>
</tr>
<tr>
<td>HPV-16</td>
<td>human papillomavirus type 16</td>
</tr>
<tr>
<td>hTERT</td>
<td>human telomerase reverse transcriptase</td>
</tr>
<tr>
<td>HDACs</td>
<td>histone deacetylases</td>
</tr>
<tr>
<td>hnRNP</td>
<td>heterogeneous nuclear ribonucleoprotein</td>
</tr>
<tr>
<td>HSV-1</td>
<td>herpes simplex virus type 1</td>
</tr>
<tr>
<td>ICP27</td>
<td>infected cell protein 27</td>
</tr>
<tr>
<td>ISEs</td>
<td>intronic splicing enhancers</td>
</tr>
<tr>
<td>ISSs</td>
<td>intronic splicing silencers</td>
</tr>
<tr>
<td>IFs</td>
<td>intermediate filaments</td>
</tr>
<tr>
<td>LCR</td>
<td>long control region</td>
</tr>
<tr>
<td>MHC I</td>
<td>major histocompatibility complex class I</td>
</tr>
<tr>
<td>MT</td>
<td>methyltransferase</td>
</tr>
<tr>
<td>NLS</td>
<td>nuclear localization signals</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>Ori</td>
<td>origin of replication</td>
</tr>
<tr>
<td>pAE</td>
<td>early polyadenylation signal</td>
</tr>
<tr>
<td>pAL</td>
<td>late polyadenylation signal</td>
</tr>
<tr>
<td>PDGF</td>
<td>platelet derived growth factor</td>
</tr>
</tbody>
</table>
PAP  poly(A) polymerase
PABP II  poly A-binding protein II
PTB  polypyrimidine tract binding protein
pol II  RNA polymerase II
RBDs  RNA-binding domains
RTP  RNA triphosphatase
Rb  retinoblastoma
snRNP  small nuclear ribonucleoprotein
SF1  splicing factor 1
SR proteins  serine-arginine-rich proteins
UsnRNPs  U-rich small nuclear ribonucleoproteins
USEs  upstream elements
5´ss  5´ splice site
3´ss  3´ splice site
3´ lateUTR  late 3´ untranslated region
3´ eUTR  early 3´ untranslated region
1. Introduction

Viruses were first reported as pathogens at the end of 19th century, a time when other microbial organisms such as bacteria and fungi were already known. Once James Watson and Francis Crik had elucidated the double-stranded helix model for DNA structure in 1953, the study of virology marched into a new era of molecular virology.

Thousands of different viruses exist intensively in nature and cause many diseases with a variety of symptoms. These viruses share several common properties, including being submicroscopic, infectious and obligate intracellular parasites. The viral genome consists of either DNA or RNA, whereas viroids consist of RNA only and prions contain protein only. Satellite viruses which were discovered around 20 years ago are similar to viruses in that they also contain either DNA or RNA. The distinguishing feature between viruses and their host cell is that viruses themselves lack energy-generating and enzyme systems, making their life cycles strictly dependent on the host cells.

Viruses have been used as valuable tools to unravel biological mechanisms and have been well described at the molecular level. In particular, the use of viruses has greatly assisted in the study of gene regulation. For example, the adenovirus model system enabled the discovery of basic mechanisms of messenger RNA processing, such as splicing (6).

In the study presented here, we investigate how late gene expression in human papillomavirus type 16 (HPV-16) is regulated by RNA processing events, especially by splicing and polyadenylation. Investigation of the molecular mechanisms behind the regulation of HPV late gene expression has a remarkable significance not only in understanding the pathogenicity of the high-risk HPVs but also in identifying novel targets for antiviral therapy.
2. Papillomaviruses

2.1 General background
Papillomaviruses belong to the family of papillomaviridae and are a large group of strictly epitheliotropic DNA tumour viruses (45). Papillomaviruses are ubiquitously distributed in many mammals, including humans, and act as causative agents for induction of warty lesions after the infection of basal cell layers of epithelia (131). Most of these lesions are benign, but some may develop to cancer.

HPVs are rendered the major risk factor for cervical cancer (206), the second most common cancer worldwide in women (173). Approximately 470,000 women in the world develop cervical cancer and about 270,000 deaths occur from this disease per year (65). Cervical cancer is most prominent in developing countries (202). HPV-16 is the most prevalent type in cervical cancer, accounting for 53% of the diagnosed cases (138).

Research on papillomaviruses dates back to the beginning of the last century. In 1907 Ciuffo demonstrated that human warts were transmissible from human to human through cell free extracts (32). In 1949, HPV viral particles were visualized with the aid of the electron-microscope (181). The study of the entire differentiation-dependent life cycle of HPV has been hindered by the lack of a permissive culture system to propagate this virus until the organotypic (raft) culture system was used to mimic the entire viral life cycle in vitro (2, 105). This system allows for the investigation of a variety of differentiation-specific manifestations of HPVs, including viral promoter activity, viral mRNA expression, the expression of late genes, and virion morphogenesis. Thus far it has been used for the observation of the viral life cycle of HPV-11 (196), HPV16 (127), HPV18 (131), HPV31 (3) and HPV45 (130).

2.2 Classification
To date, more than 100 different HPV genotypes are known to infect humans, with a subset of more than 30 types which are involved in cervical and genital cancers (117). The heterogeneity of the HPV group and the multitude of detection methods have made it difficult to define an accepted system of classification for each individual type.
HPVs were initially grouped as “cutaneous” and “mucosal” types on the basis of location of the lesion from where that the specific HPV type was first isolated. Cutaneous types induce both benign and malignant warts in skin, HPV-1 causes plantar warts on the sole of feet, whereas HPV-2, -4, -7 induce common warts on hand. However, HPV-2 can also be found in genital warts in children. Flat warts on arms and face are caused by HPV-3, -10, -28. The HPV genotypes 5, 8, 9, 12, 14, 17 and a few others are involved in a hereditary condition named epidermodysplasia verruciformis (EV) and associate with the malignant conversion of the sun-exposed lesions (147, 218). The EV-associated HPV group, for instance, HPV-5 and HPV-8, are frequently detected in benign and malignant skin lesions of organ transplant recipients (13, 180). Extensive studies have been performed on mucosal papillomavirus infections, including HPV-6, -11, -16, -18, -31 and others that primarily infect the genital tract, the respiratory tract, the oral cavity and the larynx.

HPVs can also be divided broadly into “high-risk” and “low-risk” types. The high-risk types are tightly associated with cervical and other anogenital cancers. As mentioned previously, HPV-16 is the most prevalent type found in cervical cancers, followed by HPV-18, HPV-45 and HPV-31 (44, 46). Conversely, the “low-risk” types, such as HPV-6 and HPV-11 are found primarily in benign genital warts.

The advancement of molecular biology techniques has led to a new classification system for papillomaviruses, based on the degree of DNA sequence homology. A novel type was defined when the DNA sequence of its L1 open reading frame (ORF) was proved to have a homology of less than 90% to the corresponding sequence region of the closest related known papillomavirus type(44). Many genomic HPV DNA sequences have been determined and characterized in this way. A newer classification was recently established by the International Committee on the Taxonomy of Viruses (ICTV) (45). In combination with the traditional classification, hundreds of “papillomavirus types” were classified into “family”, “genus”, “species”, “types”, “sub-types”, and “variants” according to their genomic properties.

2.3 HPV viral structure

HPV virions are small, nonenveloped, icosahedral symmetrical structures approximately 55nm in diameter (83). The virion consists of a double-stranded DNA viral genome within a protein coat named capsid. The viral capsid is arranged as 72 pentameric capsomers and consists of 360 L1 major capsid protein molecules and 12 L2 minor capsid protein molecules.
2.4 HPV life cycle

HPV infection initiates in the basal layers of the squamous epithelium after exposure of these cells by microtraumas (Fig. 1.). HPV enters the basal cells by attaching to the three different receptors on the surface of the cells (61, 74, 91). Following entry into the basal epithelial cells and migration to the nucleus, viral genome replication is activated and results in the production and maintenance of approximately 50-100 episomes of viral DNA per cell. This activity is associated with the expression of the viral replication proteins E1 and E2 and oncoproteins E6 and E7. The early promoter in HPV-16 and HPV-31 is referred to as p97, while in HPV-18 it is referred to as p105, both named after their positions in the genome. The E1 and E2 proteins are required for mediating replication by binding to the origin sequence of viral replication and recruiting cellular polymerases and accessory proteins (66, 134). The E2 protein itself acts as an activator and a repressor in controlling viral episomal copy number in undifferentiated basal layers (117). The E6 and E7 proteins can induce cell proliferation by preventing apoptosis of the cell that allows HPV DNA replication and maintenance of HPV in an extrachromosomal form in undifferentiated basal cells (195). As the infected basal cell divides, the viral DNA is distributed between both daughter cells. One of the daughter cells remains proliferation competent and stays in the basal layer while the other daughter cell leaves the basal layer and migrates to the suprabasal layers. Since HPVs rely on cellular enzymes to replicate their genomes, in contrast to uninfected cells, HPV infected cells do not exit the cell cycle as they detach from the basal layers and therefore retain nuclei throughout all differentiation layers of infected epithelia. As the infected cells move to the granular and cornified layers, the differentiation layers of infected epithelia, the viral DNA replicates to several thousand copies and differentiation-specific viral late promoters are activated (182). The late promoters in HPV 16 and HPV 31 are called p670 and p742, respectively. The activation of the late promoters results in the high expression of E4 and E5 early proteins and L1 and L2 capsid proteins. The capsid proteins assemble into virions together with the viral genomic DNA. The mature virions are shed from the cornified layers of the epithelia as the cells are sloughed off. The released virions subsequently infect new individuals. The E4 protein initiates a collapse of the cytoskeleton in the late infection and thus facilitates release of progeny virus (52). The E5 protein is proposed to activate the proliferative capacity of differentiated cells in the late infection (63). The infected cells remain replication competent in differentiated layers under the action of the E6 and E7 proteins (28).
Figure 1. Schematic diagram of the epithelium infected by HPV. A stratified epithelium with various differentiated layers and virion production is indicated on the left. Expression of HPV-16 late mRNA and DNA are represented on the right.

In order to evade the immune surveillance of the host and to allow efficient transmission of the virus throughout a population, production of capsid proteins L1 and L2 are strictly dependent on terminal cell differentiation. This differentiation-dependence permits the infected cell to persist in a non-productive stage in the basal layer for several years. This may ultimately lead to HPV associated cancer. Therefore, investigation of the molecular mechanisms of HPV gene regulation during the viral life cycle is extremely important in order to develop strategies to eliminate oncogenic HPV infection and its associated malignancies. For example, identification of potential targets for drugs that can interfere with the virus replication cycle and the differentiation programs of the host cell (167).

2.5 The HPV genomic organization

HPV genome is a double-stranded, circular DNA approximately 8 kb in size. All HPV genomes can be divided into long control region (LCR), early and late regions (Fig. 2.). The early region and the late region are separated by the early polyadenylation signal (pAE). The LCR spans sequences downstream of the L1 stop codon to the ATG start codon of the E6 gene. It contains the late 3’ untranslated region (3’ lateUTR), the early promoter and the origin of replication (Ori). The 3’ lateUTR contains the late polyadenylation signal (pAL). The early region encodes the early proteins E1, E2, E4, E5, E6 and E7. The late promoter is located in the E7 gene. The differentiation dependent pAE signal is present in the early 3’ untranslated region (3’ eUTR)
region. The late region contains L1 and L2 genes which encode two capsid proteins, the L1 major capsid protein and the L2 minor capsid protein. The LCR and 3’ eUTR are the only non-coding sequences in the genome. The following sections describe the functions of each gene product encoded by different regions of HPV genome (Fig.3.).

**HPV16 Genome**

![Genomic organization of the linearised HPV-16. The diagram shows the ORFs of the early (E) genes and late (L) genes, the LCR, the two major promoters (p97 and p670) and the two poly(A) signal (pAE and pAL). The multiple 5’ splice sites and 3’ splice sites are designated by black and grey ovals, respectively.](image)

**Figure2.** Genomic organization of the linearised HPV-16. The diagram shows the ORFs of the early (E) genes and late (L) genes, the LCR, the two major promoters (p97 and p670) and the two poly(A) signal (pAE and pAL). The multiple 5’ splice sites and 3’ splice sites are designated by black and grey ovals, respectively.

### 2.5.1 LCR

The HPV LCR is around 850 bp in size. The LCR is the least conserved region in all papillomaviruses sequenced thus far. It contains many cis-regulatory elements necessary for viral transcription and replication. The Ori is the most important cis-element involving the initiation of viral DNA (29). The LCR contains four conserved E2 binding sites (E2BSs), a consensus paralindromic sequence ACCN₆GGT (148). Three of these sites surround TATA boxes of viral early promoter and the viral Ori (192). Regulation of the early promoter is a complicated process involving various different transcription activators and repressors (31, 188). The LCR containing the epithelial cell-specific transcriptional enhancer is flanked by two nuclear matrix attachment regions (MARs) (189). MARs are suggested to modulate transcription by bringing together cis-responsive elements, the nuclear matrix, and its attached enzymatic machineries (183).
2.5.2 The early region

The early region is approximately 4 kb in size and includes the E1, E2, E4, E5, E6, and E7 genes.

**E1 and E2**

E2 gene product is a 45-48 kDa nuclear protein (199). It plays a pivotal role in regulation of viral replication and transcription through binding to the E2BSs in the LCR region (179). In undifferentiated cells, the E2 protein acts as a viral transcriptional trans-activator to stimulate transcription at low concentrations whereas it becomes to be a repressor at high concentrations by blocking the binding of transcriptional factors such as Sp1 and TFIID to their recognition sequences (47, 117). In differentiated cells, E2 loses its repressive function of transcription due to a shift to the late promoter, resulting in a rapid increase in E1 and E2 protein levels and thereby amplification of viral DNA (101, 117). In most human cervical cancer cells, the loss of E2 expression due to the integration of the viral genome has been postulated to contribute to the development of HPV-induced carcinoma (40, 166).

E2 is also required for replication of the viral genome. E2 specifically recognizes E1 and recruits E1 to the viral Ori, forming an E1E2 complex (134). In addition to the recruitment of E1, E2 is also necessary for formation of the entire pre-initiation complex and recruitment of the host cellular DNA replication machinery to the Ori (116).

E1 protein is a 68 kDa nuclear protein that has ATPase, DNA helicase, and DNA-unwinding activities (72, 169). It binds to the viral Ori with low
affinity and low sequence specificity. This affinity of binding can be increased by E2 protein (50, 66). The newly formed E1E2 hexamer complex binds to DNA with high affinity and efficiently unwinds DNA with the assistance from chaperone proteins (115, 169). E1 interacts with subunits of DNA polymerase-α (125) and the component of the Swi/Snf complex (109) to activate DNA replication. E1 exerts a role in both initiation of viral DNA replication and elongation due to its helicase activity at the replication forks (116).

**E6 and E7**

E6 oncoprotein is around 17 kDa in size. It plays an important role in the tumorigenecity of HPV infection. The high-risk E6 proteins are distributed in both the nucleus and the cytoplasm of the host cell.

One of the well-known functions of the high-risk E6 proteins is their ability to transform human cells through targeted degradation of p53. p53 is a well-characterized tumor suppressor that induces cell cycle arrest or apoptosis (102). Some evidence suggests that the low-risk E6 proteins may also bind to p53 with a low affinity (114). E6 binds to p53 with an ubiquitin ligase called E6-associated protein (E6AP). The formation of an E6-E6AP-p53 complex results in the rapid degradation of p53 through the ubiquitin-dependent proteolytic pathway (85), thereby abolishing p53-mediated cell apoptosis and cell cycle regulation (165). E6 also targets p300/CBP, a coactivator of p53, indirectly downregulating p53 activity (216).

Binding of PDZ proteins (99) and activation of human telomerase reverse transcriptase (hTERT) expression (100) are two important p53-independent ways for E6 proteins to immortalize human cells. Proteins of the PDZ family contain a conserved domain that associates with the PSD-95, Discs Large, and ZO-1 proteins (abbreviated: PDZ). Many PDZ proteins play a role in signal transduction. It is clear that the PDZ domain-containing proteins MUPP-1, hDLG, and hSCRIB are targeted for proteasome-mediated degradation by the high-risk E6 proteins (110, 139, 153). The low-risk E6 proteins do not target PDZ domain-containing proteins. hTERT is a catalytic subunit of telomerase that has ability to extend telomeres of eukaryotic chromosomes. Without the activity of telomerase, telomeres are shortened during successive cell divisions and the cells terminate chromosome replication, leading to senescence (98). E6 proteins from the high-risk types activate telomerase via upregulation of hTERT expression at transcriptional level, leading to immortalization of human cells (100, 200).

E7 oncoprotein has a molecular weight of approximately 12 kDa and is found primarily in the nucleus. The central activity of the high-risk E7 proteins is their contribution to transformation of cells through association with the members of the retinoblastoma (Rb) family (58). Rb protein is an inhibitor of cell cycle progression from the G1 to the S phase and its activity is regulated by phosphorylation (79). In HPV-infected cells, binding of E7
proteins to hypophosphorylated Rb results in functional inactivation of Rb protein and permits cell cycle progression into S phase (57). The low-risk E7 proteins have also shown to bind to Rb but with lower affinity (80). Other “pocket” proteins in Rb family include p107 and p130. Both negatively regulating the expression of proteins are necessary for cell cycle progression (137).

Histone deacetylases (HDACs) play an important role in cell cycle regulation (15). The hypophosphorylated form of Rb protein recruits HDACs to inhibit E2F-regulated S-phase specific genes (121). In HPV infected cells, the Rb-HDAC complex is disrupted by involvement of E7 proteins, leading to S phase entry of cell cycle (15). E7 proteins have also been shown to bind HDACs independently of Rb (16). The activation of E2F by E7 proteins in differentiated epithelia is suggested to be a common mechanism for viral replication in both high-risk and low-risk HPVs (16, 118).

In addition, E7 proteins inactivate the inhibitory activities of the cyclin-dependent kinase inhibitors (CKIs), such as p21 and p27, thereby promoting cell cycle progression from G1 into S phase (48, 81).

**E4**

Unlike many of the other early proteins, HPV E4 proteins are abundantly expressed from the differentiation-inducible promoter in late infection (86, 140). The full-length E4 protein is translated from a spliced E1’E4 transcript and forms a 17 kDa E1’E4 fusion protein that contains the first 5 amino acids of E1 and the entire product of E4 ORF (51). The full-length E1’E4 protein is proteolytically cleaved to produce truncated polypeptides in the upper differentiated cell layer (53). The E4 proteins from high-risk HPV distribute extensively in the cytoplasm of the infected cells and form an organized filamentous cytoplasmic network with the keratin intermediate filaments (IFs), inducing a collapse of the cytoskeleton *in vitro* (51). This suggests a function for E4 protein in facilitating release of progeny virus (159). It is reported that high expression level of E4 protein correlates with viral DNA replication and occurs prior to the expression of L1 in the differentiated epithelial cells (54), indicating that E4 protein may promote the HPV DNA replication and regulate late viral function in the HPV productive life cycle (162) (160).

**E5**

E5 gene product is a hydrophobic membrane protein of about 5 kDa in size. The E5 protein is mainly localized to the endosomal membranes, Golgi apparatus (GA), but is also present in the plasma membranes (36). E5 protein in bovine papillomavirus (BPV) mediates the primary cell transformation activity by an interaction with platelet derived growth factor (PDGF) receptors (151), however, the precise role of HPV E5 protein in the viral life cycle is unclear due to its poor expression in mammalian cells. It has been shown
that E5 protein interacts with epidermal growth factor (EGF) receptor (EGFR) and stimulates EGF-mediated signal transduction to the nucleus in HPV infected cell (111, 124). This activity may contribute to a transforming function of HPV E5 protein during the productive stage of the viral life cycle (152). HPV E5 proteins are primarily produced during the late viral life cycle. Recent studies suggested that E5 protein plays a cooperative role with E7 in modulating viral amplification and viral late gene expression in differentiated cells (63, 71). It has recently been described that HPV-16 E5 is able to interfere with major histocompatibility complex class I (MHC I, HLA class I in humans) pathway through reducing the transport of HLA class I to the cell surface (1). Furthermore, HPV-16 E5 can selectively downregulate cell surface HLA class I by retention of HLA-A/B but not HLA-C/E in GA, leading to avoidance of immune clearance of virus-infected cells by host immune system (1). Interestingly, HPV-16 E5 may also downregulate immune recognition of infected cells through inhibition of MHC II maturation (208).

2.5.3 The late region-L1 and L2 genes
The late region is approximately 3 kb in size and contains the L1 and L2 genes encoding the major and minor capsid proteins, respectively (For viral structure see 2.3 “HPV viral structure”). L1 protein has a molecular weight of around 55 kDa whereas L2 protein is approximately 74 kDa in weight. The late HPV genes are transcribed from the late differentiation dependent promoter within the E7 gene (76). L1 protein binds to three different cell surface receptors including α6β4 integrin (61), heparin sulfate (74) and glycosaminoglycans (91) for entry into the cell. Binding of L2 to the cellular surface protein is also required for efficient HPV infection (95). In the cytoplasm of the infected cell, the L2 protein interacts with β-actin to efficiently transport the virus across the cytoplasm to the nucleus. Both L1 and L2 proteins contain nuclear localization signals (NLS) in their C-terminus, which direct them to the nucleus to be assembled into virions after synthesis in the cytoplasm (184, 213). The L1 and L2 proteins are imported into the nucleus in a receptor-mediated manner (141). It is apparent that nuclear import of the L1 and L2 proteins occurs twice during the viral life cycle. The first occurrence is during the release of the HPV genome into the nucleus for replication in the early stage of infection in undifferentiated proliferating cells. The second occurrence is during the terminally differentiated stage of the infected cells when the newly synthesized L1 and L2 proteins in the cytoplasm are transported and assembled into the virions in the nucleus. The L1 protein alone has the capacity to self-assemble into capsids. The L2 protein is proposed to be involved in DNA binding and encapsidation efficiency (214). Recently, it has been reported that the presence of HPV-16 L2 inhibits the transcriptional transactivation function of E2, but not the DNA replication
function of E2 (146). The effect of L2-E2 interaction on regulation of viral life cycle is still under investigation.
3. Regulation of gene expression

In eukaryotic cells, a nascent messenger RNA (mRNA) transcribed from a DNA template in the nucleus is not exported instantly to the translation machinery in the cytoplasm. This so-called pre-mRNA made by RNA polymerase II (pol II) undergoes three major processing events before being exported. These events are capping, splicing and polyadenylation. Furthermore, it is believed that these mRNA processing events influence each other and that they can occur co-transcriptionally (39, 156).

3.1 Capping

The function of 5´ capping of mRNA not only for the identification gene transcription start sites, but also as a major determinant of mRNA maturation, translation and stability. Capping occurs after about 20-30 nucleotides of nascent RNA chains have been synthesized (90). RNA triphosphatase (RTP) removes the γ-triphosphate from the first nucleotide of the pre-mRNA. GMP is transferred from GTP to the resulting diphosphate end by guanylyltransferase (GT) to form GpppN, which is then methylated by methyltransferase (MT) at position N7 of guanine (171). Soon after its formation, the cap structure is recognized by the cap binding complex (CBC). The formed complex has an important effect on the stability of the mRNA (4). The CBC is replaced by the cytoplasmic translation initiation factor eIF-4E after the mRNA has been exported to the cytoplasm through the nuclear pore complex (172).

3.2 Polyadenylation

3.2.1 The mechanism of mRNA polyadenylation

Most eukaryotic mRNAs are processed at their 3´-ends by cleavage and polyadenylation reactions. 3´-end processing is important for mRNA stability, mRNA exporting and initiation of mRNA translation (33).

Pre-mRNA is first cleaved endonucleolytically at the cleavage/polyadenylation site when the cleavage polyadenylation specificity factor complex (CPSF) binds to the canonical AAUAAA element which lies
10-30 nucleotides upstream of the polyadenylation site (171) (Fig.4.). However, this binding is very weak and is greatly stabilized by the participation of the cleavage stimulation factor (CstF) which binds specifically to the GU-rich elements 20-40 nucleotides downstream of the polyadenylation site (33). CPSF, CstF and the pre-mRNA form a stable ternary complex, facilitating the recruitment of other factors to the cleavage site. Cleavage factors I and II (CF I and CF II) are responsible for the cleavage of the RNA, but not the poly(A) addition (186). Poly(A) polymerase (PAP) is also required for the cleavage reaction.

Figure 4. Schematic diagram showing the cleavage and polyadenylation process. The poly(A) signal and GU-rich element are indicated with the small triangle denoting the polyadenylation site. The proteins involved in the process are depicted. See text for details of protein-protein and protein-RNA interaction.

Polyadenylation occurs within the polyadenylation complex containing PAP and poly A-binding protein II (PABP II). PAP associates with the complex through CPSF and adds the poly A to the cleavage site of mRNA. After a short polyA tail is synthesized, PABP II binds to the poly(A) tail. With the combined stimulation of CPSF and PABP II, PAP efficiently synthesizes a full length poly(A) tail in one single binding event (204, 205). Once a complete poly(A) tail with 200-300 nucleotides has been rapidly synthesized, disruption of the polyadenylation complex occurs (7). CPSF is required for
both cleavage and polyadenylation, indicating a tight coupling of these two steps of 3′-end formation.

3.2.2 Regulation of mRNA polyadenylation

A functional polyadenylation signal and efficient usage of the polyadenylation site are required for termination of transcription by RNA pol II. Many eukaryotic gene transcripts possess a multiple of polyadenylation sites (60). Therefore, the efficiency of mRNA polyadenylation is controlled by cis-acting regulatory elements and trans-acting protein factors.

Apart from the poorly conserved GU-rich downstream elements (DSEs) and the canonical poly(A) signal that are essential for the efficiency of polyadenylation, the additional U-rich upstream elements (USEs) located upstream of the hexanucleotides poly(A) signal can also aid in efficient poly(A) site recognition in some viral and cellular genes (22, 132, 198). The USEs use different mechanisms to enhance 3′-end formation. In HIV-1 and the human lamin B2, the USEs directly contact CPSF and enhance the binding of CPSF to the canonical poly(A) sequence (14, 73), whereas the USE of the SV40 late poly(A) signal has been found to interact with the U1 small nuclear ribonucleoprotein (snRNP) protein A (U1A), which may contact CPSF and thus stabilize its binding (122). The human complement C2 USE has been shown to activate cleavage and polyadenylation through a direct interaction with the polyadenylation site and the polypyrimidine tract binding protein (PTB) and the CstF (136).

The usage of a polyadenylation site is dependent on the strength of that polyadenylation site and the concentration of components in the 3′-end processing machinery. In mammalian cells, the best studied example in regulation of polyadenylation site selection is the synthesis of IgM heavy chain (59, 187). During B cell differentiation, a regulated shift in polyadenylation site usage from a strong promoter-distal signal to a weaker promoter-proximal site results in a switch in the production of IgM from the membrane-bound form to the secretory form (70). The shift is mediated by an increase in RNA binding activity of CstF-64 protein during the transition from the resting B cell to the growing lymphoblast (59). In herpes simplex virus type 1 (HSV-1), selection of tandem polyadenylation sites is regulated by the infected cell protein 27 (ICP27), an immediate-early regulatory protein (129, 163).
3.3 Splicing

3.3.1 The mechanism of pre-mRNA splicing

Structure of a pre-mRNA

Most newly synthesized eukaryotic pre-mRNAs contain protein coding sequences (called exons) that are interrupted by non-coding sequences (called introns) (Fig.5). Each intron has three basic consensus sequences necessary for splicing: a 5’ splice site (5’ss), a branch point sequence (BPS) and a 3’ splice site (3’ss) containing a polypyrimidine tract. The 5’ss is defined by the junction of the first exon and the 5’ intron, and marked by the consensus sequence AGGU (R, purine), where the underlined position refers to the highly conserved signal of the 5’ border of the intron. The 3’ss is identified by the consensus sequence YAGU (Y, pyrimidine; N, variable nucleotides), where the underlined position denotes the highly conserved dinucleotide of the 3’ border of introns. The BPS is located 18 - 40 nucleotides upstream of the 3’ss. The consensus sequence of the BPS is YNY-URAC, where A is the adenosine conserved in all genes. The stretch of polypyrimidine tract lies immediately downstream of the conserved adenosine in the BPS and upstream of the 3’ss.

Figure 5. A schematic illustration showing the structure of pre-mRNA (top), and the chemical mechanism of pre-mRNA splicing reaction (bottom).
Two-step chemical reactions of pre-mRNA

Pre-mRNA splicing is a process that removes introns and assembles exons into mature mRNAs that can be translated into protein. The pre-mRNA splicing reaction occurs via a two-step trans-esterification mechanism (Fig. 5.). In the first step, the 2′ hydroxyl group of the conserved A residue in the BPS attacks the phosphodiester bond at the 5′ ss, resulting in the cleavage at the 5′ exon-intron junction. This reaction yields two products: the intron-3′ exon intermediate in which a lariat is formed by binding of the 5′-end of the intron to the conserved A residue in the BPS, and the free 5′ exon. In the second step, the newly formed hydroxyl group at the end of 5′ exon attacks the phosphodiester bond at the 3′ ss, resulting in the ligation of the two exons and release of the lariat-shaped intron (135).

The spliceosome and spliceosome assembly

The pre-mRNA splicing reactions are catalyzed by a spliceosome, a multi-component macromolecular complex. The spliceosome is formed by several uridine-rich snRNPs (UsnRNPs) and numerous non-snRNP splicing factors (207). Each UsnRNP particle contains an UsnRNA molecule and several particle-specific proteins (119). The essential components of the major spliceosome that splices the vast majority of pre-mRNA introns include five UsnRNPs (U1, U2, U4, U5, and U6) (120).

Spliceosome assembly occurs in a stepwise fashion (Fig. 6.). In the pathway of spliceosome assembly, four distinct intermediates have been identified in order of appearance on the pre-mRNA. They are called E complex (commitment complex), A complex (pre-spliceosome), B complex (spliceosome), and C complex (active spliceosome). The two-step chemical reaction of the pre-mRNA splicing takes place after assembly of the active complex.

Assembly of the E complex is initiated in an ATP-independent manner by U1 snRNP base pairing with the 5′ ss, Splicing Factor 1 (SF1) base pairing to the BPS, U2AF65 binding to the polypyrimidine tract and U2AF35 to the AG dinucleotides at the 3′ ss (158). Recent studies have shown that U2snRNP is also required for E complex assembly (43). Serine-arginine-rich proteins (SR proteins) play a very important role in recruiting splicing factors to the pre-mRNA during spliceosome assembly. Subsequently, an ATP-dependent process allows the transition from the E complex to the A complex. In the A complex, U2snRNP becomes tightly bound to the BPS through U2AF65 recruiting the 56kDa U2AF65 associated protein (UAP56) (64). In the B complex, the formed U4/U6-U5 triple-snRNP joins the pre-spliceosome in an ATP-dependent manner and results in substantial rearrangements in the pre-spliceosome, making it ready to perform the splicing reactions (104). After the spliceosome complex undergoes a series of rearrangements, the U1 and U4 snRNPs are released from the spliceosome, and
the C complex is formed. In the C complex, both U2 and U6 snRNPs form a bridge that connects the A nucleotide in BPS and the 5′ ss. The final mature spliceosome is assembled and is able to catalyze the first transesterification reaction of splicing (106). Following completion of the second transesterification reaction, the spliceosome is disassembled by relatively unknown mechanisms and its component are recycled for the next round of splicing (142).

**Figure 6.** Spliceosome assembly.

### 3.3.2 Exon definition

The majority of vertebrate internal exons are between 50 and 300 nucleotides in length whereas the introns can be tens of thousands of nucleotides in length. It is therefore difficult to recognize relatively small exons in the context of much larger intronic sequences. A model of exon definition has been proposed whereby initial recognition of an interior exon requires an interaction between a pair of 3′ss and 5′ss across the exon (5). Binding of U1snRNP to the 5′ss enhances the binding of U2snRNP and the other asso-
associated factors to the 3’ss immediate upstream of the exon (107, 203). The recognition of the first exon is promoted by the interaction between the nuclear CBC complex bound on the cap structure and U1snRNP which binds to the cap-proximal 5’ss (112). The polyadenylation signal at the 3’-end of the transcript promotes the use of the last 3’ss. The interactions between splicing components at the 3’ss of the last exon and polyadenylation components at the poly(A) signal facilitate removal of the last intron (37, 143).

Many mammalian pre-mRNAs contain multiple exons and introns. Constitutive splicing is defined as removal of all introns and joining of all exons, generating a single mature mRNA (Fig. 7.). In contrast, alternative splicing is a process in which different isoforms of the mRNA are produced by flexible selection of splice sites (Fig. 7).

### 3.3.3 Alternative splicing

In general, there are several distinct alternative splicing patterns that have been observed: mutual exclusion of exons, alternative 5’ss selection, alternative 3’ss selection, intron retention, and exon exclusion and skipping (9, 23). These alternatively spliced mRNAs produce different proteins required for specific functions. One of the most significant functions of alternative splicing is the contribution to protein diversity in vertebrates. It has been estimated that about 66.8% of all human genes in the human genome undergo alternative splicing (201). For a virus, the advantage of alternative splicing is that it can produce a great number of functional mRNAs from the limited coding capacity of viral genome (157). An equally important effect of alternative splicing is the opportunity for regulation of gene expression. Alternative splicing is regulated by binding of trans-acting factors to 5’ss, 3’ss or to other cis-acting sequences within the pre-mRNA, leading to selective use of splice sites. Many such cis-acting sequences have been identified and are grouped as either enhancer or silencer elements (108).

![Figure 7](image_url)  
*Figure 7.* Schematic drawing representing the constitutive splicing and one of the alternative splicing patterns.
3.3.4 Splicing enhancers & silencers

Exonic splicing enhancers (ESEs) can be both purine-rich and non-purine-rich sequences depending on the different types of consensus sequences they are comprised of (62). The purine-rich ESEs are usually located in the proximity of a suboptimal 3′ss (5). The purine-rich ESEs execute their function of enhancing pre-mRNA splicing through their interactions with serine-arginine-rich (SR) proteins, thereby stimulating the binding of U2AF to an upstream suboptimal 3′ss (11, 78). The purine-rich ESEs can also inhibit splicing of a pre-mRNA when they are present in a regulated intron (69, 93). Therefore, the activities of the purine-rich ESEs are controlled in a position-dependent manner. Non-purine-rich ESEs include exonic AC-rich enhancers and exonic pyrimidine-rich enhancers (209). AC-rich ESEs are involved in regulation of both viral and cellular splicing (38, 212). An exonic pyrimidine-rich enhancer has been shown in β-globin RNA (164). In contrast to ESE motifs, the exonic splicing silencers (ESSs) are less conserved. They are frequently located downstream of an ESE (178, 210). Thus, ESSs may repress the usage of upstream 3′ss through multiple mechanisms. Some ESS elements antagonize the effect of a juxtaposed ESE (77), for example in exons M1 and M2 in IgM (92), while some ESSs function in an enhancer-specific manner, such as the BPV-1 ESS2 (212). It has been shown that an upstream ESE is not required for splicing inhibitory function of the BPV-1 ESS1, indicating that in some cases the function of the ESS is independent of an ESE (210). The silencing function of ESSs is mediated by binding to heterogeneous nuclear ribonucleoprotein (hnRNP) protein family (9).

An intron usually contains intronic splicing enhancers (ISEs) and intronic splicing silencers (ISSs) which enhance and repress the use of the splice sites, respectively. The intronic elements can be proximal or distal to the regulated exon (108). In contrast to the exonic elements, only a few examples of the intronic elements have been identified to date, such as uridine-rich enhancer elements located downstream of alternative 5′ss, the highly conserved hexanucleotide UGCAUG enhancer in the proximal downstream intron sequences (17, 133), and an intronic splicing repression element CUCUCU in regulating c-src NI exon (26). Recently, intronic CA-repeat and CA-rich elements have been identified to be splicing enhancers or silencers, depending on their proximity to their upstream alternative 5′ss (84). The mechanisms of action of intronic elements are still not well understood.

3.3.5 SR proteins

SR proteins contain one or two RNA-binding domains (RBDs) at their N-termini and have varying length arginine/serine-rich domains at their C-termini (referred to the RS domain). About 10 different members of the human SR protein family have been discovered to date. SR proteins are re-
quired for several steps during spliceosome assembly (67). The exon-dependent SR proteins play a crucial role in up-regulating upstream 3′ss selection by recruiting U2AF65/U2AF35 to the polypyrimidine tract/3′ss (197, 217). It has been shown that the recognition of the proximal 5′ss is promoted by binding of SR proteins to the upstream ESEs (12, 170). The exon-dependent SR proteins in constitutive splicing have a similar function to that in alternative splicing. It has been proposed that SR proteins bind to the constitutive exon where they simultaneously recruit U2AF35 to the upstream 3′ss and U1 snRNP to the downstream 5′ss (77, 158). SR proteins can also function in an exon-independent manner. The details of exon-independent models of SR proteins need to be further examined.

SR proteins can also repress splicing, by either binding to an ISS or an ESS. In BPV-1, SR proteins bind ESS1 and suppress the upstream 3′ss usage (211). In adenovirus, SR proteins have been shown to inhibit IIIa premRNA splicing by binding to an intronic repressor element 3RE and preventing recruitment of the U2 snRNP to the spliceosome (93).

3.3.6 hnRNP proteins

hnRNP proteins are a large and structurally diverse family of proteins. Over 20 major proteins have been identified in human cells and are termed hnRNP A1 to U (56). Some hnRNP proteins, such as hnRNP A1 and hnRNP C1, are abundant in the nucleus of vertebrate cells (55). Many of hnRNP proteins shuttle between the nucleus and the cytoplasm, and have multiple functions, including regulation of mRNA splicing (hnRNP A1) (25), mRNA polyadenylation (PTB) (136) and mRNA export (hnRNP L) (154). hnRNP A/B, PTB and hnRNP H are the best-characterized mediators of splicing inhibition (27, 175, 203). These proteins have been shown to repress splicing by binding to UAGG, UCUU, or polyG motifs, respectively (18, 21, 26). A recent study revealed a novel role for hnRNP L in enhancing the removal of CA-rich intron in the eNOS gene and mediating an ESS1-dependent exonic repression in CD45 (84, 161). The precise mechanism by which splicing is silenced through these proteins is not entirely known.

3.4 Connecting RNA processing and transcription

It is known that many pre-mRNA processing events including 5′-end capping, splicing and 3′-end polyadenylation occur while the nascent RNA is being synthesized by RNA pol II (123, 155). The carboxy-terminal domain (CTD) of pol II plays a critical role in acting to coordinate each stage of these processes (Fig.8.). The phosphorylation form of the CTD plays an important role in functioning of capping enzymes (128), the assembly of spliceosome (82) and the binding of the polyadenylation complex (39).
It has been demonstrated that elongation affects capping, splicing and polyadenylation but in addition, it has been shown that each of these events also feeds back on the mRNA transcription process (123). Furthermore, coupling among all of the pre-mRNA processing steps is also essential for gene expression (123). In summary, the CTD of the pol II is a key player that connects all stages of pol II transcription machinery and RNA processing machinery into the mRNA “factory”. The activities of the machineries of gene expression are complex networks that maximize the efficiency and specificity of each reaction.

![Diagram of mRNA processing](image)

**Figure 8.** Schematic illustration representing cotranscriptional processing. Three different shaped symbols represent the proteins that interact with RNA pol II, and participate in capping, splicing, and cleavage/polyadenylation processing. RNA pol II is involved in different steps of pre-mRNA processing. Capping enzymes and 3’-end cleavage/polyadenylation factors are recruited at the initiation of transcription. As RNA pol II traverses the gene, splicing factors interact with the transcriptional machinery.

### 3.5 Five cellular proteins with key roles in this thesis

In Paper I-II, the identified splicing regulatory elements in L1 region are hnRNP A1 dependent. In Paper-IV, four polyadenylation factors have been reported to bind to the 3’ eUTR sequence of HPV-16. Therefore, it is of interest to describe more background about these factors in mRNA processing.
3.5.1 hnRNPC1/C2
hnRNPC1/C2 proteins are two abundant nuclear restricted pre-mRNA binding proteins approximately 41 kDa and 43 kDa in size, respectively. The difference between the C1 and C2 protein is that the C2 protein contains an extra 13 amino acids in-frame (19). hnRNPC1/C2 proteins bind to a simple RNA binding motif of penta-U nucleotides (75). The proteins are involved in pre-mRNA packaging, spliceosome assembly and in nuclear retention of unspliced hnRNA (56).

3.5.2 hFip1
It was recently reported that the 66 kDa hFip1 is a subunit of CPSF and is capable of initiating PAP activity alone by binding to a U-rich upstream sequence element on a pre-mRNA (94). hFip1 has been suggested to be involved in CPSF-mediated stimulation of PAP activity by forming a ternary complex with CPSF160 and PAP in vitro (94).

3.5.3 PTB
PTB is an about 55 kDa protein that is a member of the hnRNP family. The most conserved RNA binding site for PTB is UCUU flanked by pyrimidines (150). PTB acts as a splicing silencer that blocks the binding of U2AF65 to the polypyrimidine tract, resulting in inhibition of pre-mRNA splicing (174). Another major role for PTB in mRNA processing is its positive and negative effects on mRNA polyadenylation processing (24). PTB can compete with CstF-64 for binding to pyrimidine rich DSEs, resulting in down-regulation of polyadenylation. PTB can also interact with USEs in the C2 complement gene, which up-regulates the efficiency of polyadenylation (136).

3.5.4 CstF-64
CstF-64, a subunit of the CstF proteins, contains the N-terminal RNA binding domain for recognition of the GU-rich or U-rich sequences (185). CstF-64 plays a key role in regulation of mRNA polyadenylation by recognition of DSEs (see 3.2.1 “The mechanism of mRNA polyadenylation” and 3.2.2 “Regulation of mRNA polyadenylation”). Furthermore, it has been reported that sequences containing consecutive Us bind more tightly to CstF-64 (149).

3.5.5 hnRNPA1
hnRNPA1, a member of the hnRNP protein family, is a nucleo-cytoplasmic shuttling protein of 34kDa in size. It is one of the most abundant nuclear
proteins in mammalian cells. In structure, hnRNP A1 contains two RBDs at the N-terminus and a glycine-rich domain at the C-terminus for protein-protein interaction (18). The A1 protein binds to nascent pre-mRNAs and plays an important role in the inhibition of splicing (8, 20). The specific binding site of hnRNA A1 is UAGGGA/U (18). Several models for the action of hnRNP A1 have been proposed. Binding of hnRNP A1 to a high affinity site on HIV-1 pre-mRNAs has been shown to nucleate additional hnRNP A1 molecules to fill up low-affinity sites in a cooperative manner and SR proteins are unable to block the binding of hnRNP A1 to its binding sites, splicing silencing ensues. (215). hnRNP A1 could bind to several ISSs in the intron. This bridging interaction may cause looping out exon and result in exon skipping (10). In a situation where enhancers and silencers partially overlap, the splicing efficiency is determined by direct competition of hnRNP A1 with the splicing stimulation factors (for example, SR proteins) (126).
4. Regulation of HPV gene expression

4.1 Transcriptional regulation

Regulation of HPV gene expression at the transcriptional level is controlled through the use of different promoters, which are activated at different stages of cell differentiation. It was shown the late transcripts in HPV-31b are mainly initiated from the differentiation-dependent promoter (87). E2 is a sequence specific transcriptional regulator found in all HPVs. It has been reported that E2 causes transcriptional repression or activation through binding different E2BSs in the LCR (47). In addition, cellular factors are also important for HPV transcriptional regulation. The epithelial cell-specific enhancer in LCR contains several binding sites for cellular transcriptional factors (for example, NF-1, TEF-2, AP-1 and Oct1). All these cellular factors regulate the promoter activity (30). Apart from the transcriptional regulation, the transcribed HPV polycistronic messages are also regulated through the use of the polyadenylation signals and extensive splicing during transcription.

4.2 Polyadenylation and splicing in HPV

As mentioned previously, the HPV genome consists of an early region and late region separated by the pAE. The late region is followed by the pAL. In the early stage of the viral life cycle, all viral transcripts are polyadenylated at the pAE, generating early mRNAs. In the late stage of infection, both the pAE and the pAL are used for production of the early and the late mRNAs. In response to differentiation of the HPV infected cell, the use of the pAE is down regulated, resulting in read-through into the late region and polyadenylation of the late transcripts at the pAL. As in the case of the IgM heavy chain gene and the HSV-1 gene, it is likely that both the pAE and the pAL of HPV are under the control of multiple regulatory elements.

Interestingly, there are no obvious U or G/U rich elements downstream of the pAE in HPV-16. Instead of DSE, a 57 nucleotides of U-rich sequence in the 3’ eUTR has been found to act as an enhancing USE on the pAE (See Paper-IV). The USE was found to interact with PTB and CstF-64 (See Paper-IV), as they do on the C2 complement mRNA, suggesting that these factors may regulate polyadenylation (136). The USE also interacts specifi-
cally with the hFip1 protein and the hnRNPC1/C2 protein that may modulate the activity of the USE (See Paper-IV). Recently, it has been shown that six consecutive hnRNP H-binding GGG motifs in the L2 coding region of HPV-16 are necessary for early polyadenylation (145). These RNA elements have been mapped to a position 174 nucleotides downstream of the pAE (145). In HPV-31, a close relative to HPV-16, the activity of the early polyadenylation is promoted by a major RNA element within the first 800 nucleotides of the L2 coding region (193). Three weak binding sites for CstF-64 were identified in the early polyadenylation region, suggesting that these binding sites may modulate read-through activity (194). The exact effect of these binding sites on the early polyadenylation is still unknown.

Inhibitory RNA elements in the 3´ lateUTR have been found in BPV-1 (68), HPV-1 (177, 191), HPV-16 (97), and HPV-31 (42). These negative elements appear to inhibit viral late gene expression through different mechanisms. In the case of HPV-16, an inhibitory element has been mapped to 79 nucleotides overlapping the 3´-end of the L1 coding region and extending into the 3´ lateUTR (96, 97). The 5´ portion of this element contains four weak 5´ss and a predicted stem-loop structure, while the 3´ portion is very GU-rich. Several factors including CstF-64, U2AF65 and HuR proteins were found to bind this element (49, 103). Recently, it has been shown that all four weak 5´ss contribute to the partly repressive effect of the inhibitory element (41).

In HPV, as mentioned previously, multiple splice sites have been identified on the genome. To prevent late gene expression at an early stage, alternative splicing can be used, whereby all mRNAs using the 5´ss in the early region must be spliced to the early 3´ss that are also located in the early region. If the late 3´ss was used, there would consequently be premature expression of late mRNAs in the viral life cycle. Therefore, expression of HPV late genes must be under regulation of many RNA processing events including polyadenylation and alternative splicing. With this goal in mind, the following two projects to study the regulation of HPV-16 late gene expression were initiated (See section 6. Present investigation).
5. Aims of the study

The overall aim of this study was to understand the regulation of late gene expression in HPV-16. The specific aims were:

- to set up a system to study the regulation of HPV-16 gene expression.
- to determine the function of previously identified RNA elements in HPV-16 L1.
- to identify and characterise RNA binding proteins that interact specifically with the cis-acting regulatory elements in L1.
- to study the role of the HPV-16 eUTR region in HPV-16 late gene expression.
- to identify RNA binding proteins interacting with the HPV-16 eUTR region.
6. Present investigation


One of the typical characteristics of HPV-16 is that the production of the late mRNAs encoding L1 and L2 proteins is tightly associated with the terminal differentiation of epithelial cells. This indicates that HPV-16 late gene expression is regulated by certain elements. We have previously found that the L1 and L2 coding regions contain negative cis-acting regulatory elements that inhibited the downstream chloramphenicol acetyltransferase (CAT) report gene expression (34, 176, 190). We further mapped the inhibitory RNA sequences in L1 and L2. The results showed that the first 514 nucleotides of the L1 gene contain the major inhibitory RNA elements and that L2 gene has two inhibitory regions (35, 144). Altering the RNA sequences without changing the amino acid sequences resulted in the production of high levels of L1 and L2 mRNAs and proteins upon transfection of human cells (HeLa cells) with CMV promoter driven L1 and L2 cDNAs (35, 144). We speculated that the RNA elements in the HPV-16 late coding region may play an important regulatory role in controlling the viral late gene expression during the viral life cycle (167, 168). Based on these data, it is of interest to initiate the following project on the functional analysis of the inhibitory elements in the L1 coding region.

Paper-I

With an aim to determine the function of the inhibitory RNA elements in L1, we first generated the pBEL plasmid, which contains the CMV promoter followed by HPV-16 E1, E2, E4 and E5 early genes, the pAE signal and the late region encoding L1 and L2 followed by the pAL signal. This plasmid contains the strong CMV promoter in place of the weak HPV promoters and also lacks E6 and E7 genes. Upon transfection of pBEL into HeLa cells, only mRNAs polyadenylated at the pAE are produced. To induce expression of the late mRNAs from pBEL, we generated the pBEL-pAE plasmid by removing 59 nucleotides of the 3’ eUTR and the canonical pAE signal. Using pBEL and pBEL-pAE, pBELM and pBELM-pAE were generated by inactivating the first 514 nucleotides of the L1 coding region in pBEL and pBEL-pAE by introducing point mutations that altered the RNA sequence but not the protein sequence of L1. Using Northern Blotting and RT-PCR,
we found that the pBEL-pAE plasmid produced primarily partially spliced late L2/L1 mRNA species, whereas pBEL failed to produce late mRNAs. pBELM and pBELM-pAE produced high levels of spliced L1 mRNA. These results suggested that the inhibitory sequences in the 5’-end of L1 encode RNA elements that inhibit splicing of the late mRNAs. In addition, RT-PCR analysis revealed that the mutations in L1 activated the production of a novel late L1 mRNA splicing event that spliced from the major 5’ss (at genomic position 880) in the early region to the L1 3’ss, suggesting that one role of the splicing silencer is to prevent late gene expression at an early stage of the viral life cycle. We also found that the 3’ss immediately upstream of L1 is inefficiently utilized as a result of a suboptimal PPT. We found and mapped a splicing silencer in the 5’-end of the L1 coding region that inhibited splicing in vitro and in vivo. This silencer was identified to specifically bind a 35 kDa cellular protein. By immunoprecipitation it was shown that this protein is hnRNP A1. This result was further confirmed by binding of the recombinant his-tagged hnRNP A1 to this splicing silencer. In conclusion, the role of the hnRNP A1 dependent splicing silencer is to inhibit use of the L1 3’ss, perhaps to prevent the production of late mRNAs at an early stage of the viral life cycle. In addition, it may determine the balance between the partially spliced L2/L1 and spliced L1 mRNAs.

Paper-II

In agreement with the conclusion drawn from Paper-I and the previous work identifying that the first 514 nucleotides of the L1 region contain the inhibitory RNA elements, our next goal was to further investigate the regulatory elements in the L1 coding region. We introduced a series of deletions within the first 514 nucleotides of the L1 sequence in the CMV driven, subgenomic HPV-16 expression plasmid named pBELDPU. By using Northern Blotting analysis we found that a positive splicing element was located in the first 17 nucleotides of L1 and that a negative splicing element was present in the first 22 nucleotides of L1. The splicing enhancing function of this splice enhancer was also shown by the in vitro splicing assay. We also identified a 55 kDa cellular protein that interacted with the 17 nucleotide splicing enhancer, suggesting that this protein was required for the enhancement of splicing. However, the characterization of this 55 kDa protein remains yet to be determined.

To investigate whether the L1 sequences downstream of nucleotide position 17 inhibit splicing, this downstream sequence was inserted in anti-sense orientation in the p1-3 derived HPV-16 plasmid, generating p3A. In contrast to the p1-3, p3A expressed primarily spliced mRNAs, indicating that multiple splicing inhibitory elements are present in the HPV-16 L1 coding region. Next, we wanted to test if the multiple splicing inhibitory elements spanning the entire L1 coding region interact with trans-acting cellular factors. The
entire L1 sequence was split into four parts and inserted into a pT7 vector driven by a T7 promoter, respectively. By using UV cross-linking assay, we found that all L1 sequences interacted with a 35 kDa protein. To confirm that the 35 kDa protein was hnRNP A1, recombinant GST-hnRNP A1 was separately UV cross-linked to these RNAs. We found that hnRNP A1 interacted with multiple non-overlapping HPV-16 L1 sequences with splicing inhibitory activity. In conclusion, in addition to the identified hnRNP A1-dependent splicing silencer in L1 (Paper-I), multiple hnRNP A1 binding splicing silencers exist in the L1 gene that may counteract the effect of the 17 nucleotide splicing enhancer. We proposed a model of splicing regulation of HPV-16 late gene expression (Fig.9.). In undifferentiated cells, the abundant hnRNP A1 proteins may occupy all of the binding sites on the L1 coding region and inhibit the function of the ESE. This could result in inhibition of the late 3’ss in HPV-16. In the terminally differentiated cells, lower levels of hnRNP A1 are thought to activate HPV late mRNA splicing.

Undifferentiated cells:

Differentiated cells:

Figure 9. A model of splicing regulation of HPV-16 late gene expression in different layers of epithelia.
Paper-III

In the previous project, we identified an hnRNP A1-dependent splicing silencer and a 55 kDa protein binding splicing enhancer in the HPV-16 L1 gene. We also found that hnRNP A1 interacts with the multiple splicing silencers in the L1 coding region. We have previously reported that the HPV-16 L1 gene contains sequences that strongly inhibit expression of L1 from L1 cDNAs. A mutant L1 cDNA in which the negative elements were inactivated produced high levels of L1. We asked the question: “can splicing silencers inhibit gene expression in the absence of splicing?”

To gain an insight into the link between inhibition of L1 mRNA splicing and inhibition of L1 cDNA expression, we first compared the inhibitory effect on gene expression with the inhibitory effect on splicing by three hybrids between wt and mutant L1 sequences. To do so, the first 514 nucleotides of L1 were divided into three regions (1, 2 and 3). Hybrids between wt and mutant fragment 1, 2 and 3 were generated and inserted into pBEL-pAEPL containing a polylinker between the L1 ATG and the BamHI site 514 nucleotides further down. Northern blotting analysis revealed that the strongest inhibitory effect on splicing was encoded in the first 177 nucleotides of L1.

In order to compare the inhibitory activity of the hybrids on the expression of a reporter gene in the absence of splicing, the three same hybrid sequences described above and the entire first 514 nucleotides of wt or mutant L1 were inserted in frame with the CAT reporter gene. We found that the first 177 nucleotides of L1 reduced mRNA levels to the highest extent. This result demonstrated that there was a good correlation between inhibition of CAT expression and inhibition of splicing. Further investigation demonstrated that the wt 1-129 mRNA specifically interacts with a 35 kDa and a 65 kDa factor, whereas the mutant 1-129 L1 sequence did not. This 35 kDa protein was identified as hnRNP A1. However, the 65 kDa protein remains to be identified. In conclusion, inhibition of splicing correlates well with inhibition of gene expression in the absence of splicing. Binding of the 35 kDa hnRNP A1 and the 65 kDa protein to wt L1 RNA sequence is necessary for inhibition of gene expression.

6.2 Project 2. (Paper-IV)

UTR sequences are often involved in various steps of the mRNA processing pathway including mRNA transport, mRNA half life and mRNA translation. HPV-16 genome contains a 3´ eUTR and a 3´ lateUTR. The pAE signal in the 3´ eUTR region plays a key role in the switch from early to late mRNAs. Previous reports has suggested that the 3´ eUTR in HPV-16 contains mRNA instability elements that regulate the expression levels of the early mRNAs.
Here, we wish to further determine the effect of HPV-16 eUTR on the expression levels of early and late mRNAs.

**Paper IV**

Firstly, we generated a plasmid called pBearly expressing the early region of HPV-16 under the control of the CMV promoter. In order to investigate how deletions in the eUTR of HPV-16 affect early mRNA levels, we removed the entire eUTR, or the 3′- half (U-rich sequences) of the 3′ eUTR in the pBearly plasmid. Using Northern blotting with different probes which annealed with different sequences in the plasmids. It was shown that deletion of either the entire UTR or the U-rich alone resulted in a modest increase in early mRNA levels. We concluded that the 3′ eUTR had a modest effect on the steady state levels of the early HPV-16 mRNAs in HeLa cells. Next, we wanted to investigate if the 3′ eUTR sequence could affect the function of the pAE signal and regulate late gene expression. To test this, the various 3′ UTR deletions including the mutational inactivation of the pAE signal were inserted into the previously described pBEL and pBELM subgenomic HPV-16 plasmids, respectively. The effects on polyadenylation of the 3′ eUTR deletions could be measured by probing the late mRNAs using the L1 probe. It was shown that inactivation of the pAE signal resulted in the induction of the late gene expression and also indicated that the 3′ eUTR affect expression levels of the late mRNAs. Furthermore, we found that the 57 nucleotide U-rich region of the 3′ eUTR acts as an USE that enhances polyadenylation at the pAE. This U-rich USE has been found primarily on viral mRNAs encoded by various unrelated viruses such as retroviruses, as well as on some cellular mRNAs such as C2 complement.

Four proteins were identified that interact specifically with the U-rich region in the eUTR. They were hFip1, hnRNP C1/C2, CstF-64 and PTB, suggesting a role for these proteins in the regulation of the pAE in HPV-16. The exact functions of these proteins in regulating the polyadenylation of HPV-16 mRNAs remain to be determined.
7. Future perspectives

**Does the level of hnRNP A1 affect the HPV-16 late gene splicing?**
We recently identified the multiple hnRNP A1 dependent splicing silencers along the L1 coding region may counteract a splicing enhancer located in the first 17 nucleotides of L1. The splicing enhancer specifically bound a 55 kDa cellular protein. It would be of interest to determine the function of these RNA elements and factors in the viral life cycle. It has been shown that a virus infection of a cell causes a redistribution of hnRNP A1 from the nucleus to the cytoplasm in the infected cell (113). We wish to investigate if down-regulation of hnRNP A1 expression would activate HPV-16 late mRNA splicing. By using raft culture system, analysis of infectious molecular clones of HPV-16 in which the regulatory elements in L1 gene have been destroyed would provide strong evidence on the importance of these RNA elements in the viral life cycle.

**What are the functions of the identified cellular factors in the polyadenylation of the early HPV-16 mRNAs?**
We have found an U-rich USE element in the 3´eUTR of HPV-16. This USE interacts with hFip1, CstF-64, hnRNP C1/C2 and PTB. It would be of interest to determine the exact functions of these cellular factors in the early polyadenylation of HPV-16 mRNAs by using in vitro cleavage/polyadenylation assay. It would be of interest to study the role of the HPV-16 USE in response to cell differentiation to determine its role in the viral life cycle.

**Are the RNA regulatory elements identified in HPV-16 L1 conserved in other HPV types?**
The study could interpret whether the functional difference of these regulatory elements are conserved in the highly diverse group of HPVs. By using the same experimental system as used in HPV-16 described previously, we wish to extend the experiments to other HPV types, primarily HPV-1, HPV-6 and HPV-18. Similar plasmids could be used to study RNA elements in the late region of various HPV types. This investigation is important since different HPV types have different pathogenic properties that may be explained by differences in the gene expression programme of the viruses.
8. Acknowledgements

I would like to express my sincere gratitude to all the people who supported me in various ways to make this thesis completion. In particular,

I am greatly indebted to my supervisor Stefan Schwartz for accepting me to be his student, for sharing his broad knowledge in the field of molecular virology, for his great patience and most skillful guidance, for his encouragement and believing in me during my PhD study.

I extend my gratitude to group leaders Göran Akusjärvi (my examiner), Göran Magnusson, Catharina Svensson and Peter Kreivi for their generous help and valuable advise during my study career.

All former and present members in the lab: Brian, Lisa, Anette, Karin, Anna, Margaret, Daniel and Monika. I’m grateful for being able to work with them all.

Especially, many thanks to Brian for giving me great help to initiate my study.

I specially thank to Daniel, for his generous helping anytime when I need it. Also to Margaret, for the frank discussions and polishing my English. I can not be thankful enough for all the benefits I got from them. It was such a wonderful time we spent in the lab. I look forward to meeting them in China!

All the people on the Friday’s seminars: Saideh, Christina, Cecillia, Ning, Ellinor, Lena, Xiaofu, Josef and Heidi for creating a nice environment for working. I will never forget their kindness.

All the people in the B:11 corridor, for their nice help.

Allison, for correcting English of my thesis.

Maud, Liisa, Erica, Barbro and Ulla for always being helpful. Kerstin and Olav, for their invaluable help in computer and instrumental services.
Ylva and Tony, for providing the conveniences for the experiments.

All my chinese friends in Uppsala, for sharing the pleasant time together.

Yu Li, for the nice talkings. Wenru Guo, for the long-time friendship. Lei & Tim, for lots of good time throughout these years.

I owe special thanks to Uncle Zuo and Aunty Li for their strong support all along.

Finally, my dearest parents, for their endless love, support and understanding, for always being my best friends in my life. Thanks a lot for everything!
9. References

1. Ashrafi, G. H., M. R. Haghshenas, B. Marchetti, P. M. O'Brien, and M.
downregulates surface HLA class I. Int J Cancer 113:276-83.
   Wilbanks, and L. A. Laimins. 1991. Amplification of human papillomavi-
rus genomes in vitro is dependent on epithelial differentiation. J Virol
65:2254-60.
Cell 81:179-83.
270:2411-4.
6. Berget, S. M. 1977. Spliced segments at the 5' terminus of adenovirus 2 late
mRNA. Proc Natl Acad Sci U S A 74:3171-3175.
   Stoltzfus. 2001. RNA splicing at human immunodeficiency virus type 1 3'
splice site A2 is regulated by binding of hnRNP A/B proteins to an exonic
high-affinity hnRNP A1-binding sites and by intron elements that repress
Identification of a bidirectional splicing enhancer: differential involvement of SR
proteins in 5' or 3' splice site activation. Mol Cell Biol 19:7347-56.
2001. Human papillomavirus infection and skin cancer risk in organ trans-
polyadenylation factor by the upstream sequence element of the human lamin
15. Brehm, A., E. A. Miska, D. J. McCance, J. L. Reid, A. J. Bannister, and
T. Kouzarides. 1998. Retinoblastoma protein recruits histone deacetylase to
Bannister, and T. Kouzarides. 1999. The E7 oncoprotein associates with

45
Mi2 and histone deacetylase activity to promote cell growth. Embo J 18:2449-58.


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 67

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine”.)