Regulation of RNA Processing in Human Papillomavirus Type 16

MARGARET RUSH
Dissertation presented at Uppsala University to be publicly examined in Room B42, BMC, Husargatan 3, Uppsala, Thursday, October 27, 2005 at 10:00 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English.

Abstract

Human papillomavirus type 16 (HPV-16) is the major cause of cervical cancer. HPV-16 gene expression is tightly linked to the differentiation programme of the infected epithelium. Expression of the late genes, L1 and L2, encoding the capsid proteins, is delayed until the more terminally differentiated cells. Successful inhibition of HPV-16 late gene expression early in the viral life cycle is essential for persistence of infection, the highest risk factor for cervical cancer.

The goal of this thesis was to identify regulatory RNA elements and cellular factors that influence RNA processing events, such as alternative splicing and polyadenylation, during late gene expression. For this purpose, transfection of plasmids containing almost the full-length HPV-16 genome into HeLa cells, followed by RNA analysis, was employed. An exonic splicing enhancer (ESE) was identified that firmly supported the use of the E4 3’ splice site. A key regulator of HPV-16 gene expression, the E4 ESE was required for early mRNA splicing and polyadenylation, as well as for inhibition of premature late gene expression. The early polyadenylation signal (pAE) is also an important block of premature late gene expression. An upstream polyadenylation element (USE) was identified in the early 3’ untranslated region that enhanced polyadenylation at pAE, and interacted specifically with the cellular factors CstF-64, hnRNP C1/C2, PTB and hFip1. With the help of adenoviral E4or4, a protein which causes dephosphorylation of SR proteins, we found that overexpression of SRp30c activated HPV-16 late gene expression by an exon skipping mechanism, and that SRp30c may interfere with early mRNA terminal exon definition. This work identified a crucial splicing enhancer, as well as a number of cellular proteins binding to an USE in the early region of HPV-16. Furthermore, the cellular splicing factor SRp30c was shown to play a role in the regulation of HPV-16 late gene expression.

Keywords: RNA processing, human papillomavirus, HPV-16, alternative splicing, polyadenylation, exonic splicing enhancer, 3’UTR, upstream polyadenylation element, SR proteins

Margaret Rush, Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden

© Margaret Rush 2005

ISSN 1651-6206
ISBN 91-554-6356-8
urn:nbn:se:uu:diva-5972 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-5972)
List of papers

This thesis is based on the following papers, referred to in the text by their roman numerals:


Reprints were made with permissions from the publishers: the American Society of Microbiology.
Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARE</td>
<td>AU-rich element</td>
</tr>
<tr>
<td>bps</td>
<td>base pairs</td>
</tr>
<tr>
<td>BPV-1</td>
<td>bovine papillomavirus type 1</td>
</tr>
<tr>
<td>CBC</td>
<td>cap-binding complex</td>
</tr>
<tr>
<td>C/EBPs</td>
<td>CCAAT/enhancer binding proteins</td>
</tr>
<tr>
<td>CF1/2</td>
<td>cleavage factors 1/2</td>
</tr>
<tr>
<td>CPSF</td>
<td>cleavage and polyadenylation specificity factor</td>
</tr>
<tr>
<td>CstF</td>
<td>cleavage stimulatory factor</td>
</tr>
<tr>
<td>C terminus</td>
<td>carboxyl terminus</td>
</tr>
<tr>
<td>CTD</td>
<td>carboxy-terminal domain</td>
</tr>
<tr>
<td>DSE</td>
<td>downstream enhancer</td>
</tr>
<tr>
<td>ESE</td>
<td>exonic splicing enhancer</td>
</tr>
<tr>
<td>ESS</td>
<td>exonic splicing suppressor</td>
</tr>
<tr>
<td>EJC</td>
<td>exon-exon junction complex</td>
</tr>
<tr>
<td>E6-AP</td>
<td>E6-associated protein</td>
</tr>
<tr>
<td>eIFs</td>
<td>eukaryotic initiation factor</td>
</tr>
<tr>
<td>hnRNPs</td>
<td>heterogeneous nuclear ribonucleoproteins</td>
</tr>
<tr>
<td>HDACs</td>
<td>histone deacetylase</td>
</tr>
<tr>
<td>HPV</td>
<td>human papillomavirus</td>
</tr>
<tr>
<td>ISE</td>
<td>intronic splicing enhancer</td>
</tr>
<tr>
<td>ISS</td>
<td>intronic splicing suppressor</td>
</tr>
<tr>
<td>IRES</td>
<td>internal ribosome entry site</td>
</tr>
<tr>
<td>kDA</td>
<td>kilo Daltons</td>
</tr>
<tr>
<td>LCR</td>
<td>long control region</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>mRNP</td>
<td>messenger ribonucleoparticle</td>
</tr>
<tr>
<td>NLS</td>
<td>nuclear localisation signal</td>
</tr>
<tr>
<td>NMD</td>
<td>nonsense-mediated decay</td>
</tr>
<tr>
<td>NPCs</td>
<td>nucleopore complex</td>
</tr>
<tr>
<td>NRE</td>
<td>HPV-16 negative regulatory element</td>
</tr>
<tr>
<td>N terminus</td>
<td>amino terminus</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>ori</td>
<td>origin of replication</td>
</tr>
<tr>
<td>PABPII</td>
<td>poly(A) binding protein II</td>
</tr>
<tr>
<td>pAE</td>
<td>early polyadenylation signal</td>
</tr>
<tr>
<td>pAL</td>
<td>late polyadenylation signal</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>PAP</td>
<td>poly(A) polymerase</td>
</tr>
<tr>
<td>PTB</td>
<td>polypyrimidine tract binding protein</td>
</tr>
<tr>
<td>PTC</td>
<td>premature termination codon</td>
</tr>
<tr>
<td>Rb</td>
<td>retinoblastoma protein</td>
</tr>
<tr>
<td>REF</td>
<td>RNA export factor</td>
</tr>
<tr>
<td>RNAPII</td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td>RRM</td>
<td>RNA recognition motif</td>
</tr>
<tr>
<td>SF1</td>
<td>splicing factor 1</td>
</tr>
<tr>
<td>snRNAs</td>
<td>small nuclear RNAs</td>
</tr>
<tr>
<td>snRNPs</td>
<td>small nuclear ribonucleoprotein particles</td>
</tr>
<tr>
<td>SR proteins</td>
<td>serine/arginine rich proteins</td>
</tr>
<tr>
<td>SV40</td>
<td>simian virus 40</td>
</tr>
<tr>
<td>TAR</td>
<td>trans-activation response element (TAT binding site)</td>
</tr>
<tr>
<td>U2AF</td>
<td>U2 snRNP auxiliary factor</td>
</tr>
<tr>
<td>U snRNPs</td>
<td>uridine snRNPs</td>
</tr>
<tr>
<td>Upfs</td>
<td>up frame shift proteins</td>
</tr>
<tr>
<td>USE</td>
<td>upstream polyadenylation element</td>
</tr>
<tr>
<td>3'UTR</td>
<td>3’ untranslated region</td>
</tr>
<tr>
<td>5'UTR</td>
<td>5’ untranslated region</td>
</tr>
</tbody>
</table>
Introduction

Cancer, a malignant tumour arising from abnormal cell growth, and oncogenesis, the complex multistage process by which tumours develop, have been a focus of research since the beginning of the last century. Much of our recent understanding of the molecular basis for these processes can be attributed to the earlier studies of members of several virus families that can cause cancer in animals. Such viruses have the ability to alter the growth properties and morphology of infected cells, and in some cases these transformed cells can give rise to tumour formation when transplanted into animals. Thus, during the elucidation of viral transformation, much knowledge has been accumulated about cell proliferation, as well as the discovery of the role of tumour suppressor genes and oncogenes in the development of cancer.

Papillomaviruses, small DNA viruses, exemplify such viruses, the study of which has added significantly to our understanding of cancer. These viruses are members of the *Papillomaviridae* family, and were the first oncogenic DNA viruses to be identified. As early as 1933, Richard Shope demonstrated that papillomaviruses cause warts (papillomas) in cottontail rabbits (1). Though research was hampered by the lack of a suitable cell-culture system, papillomaviruses and their human counterparts, the human papillomaviruses (HPVs) continued to be studied intensively. This eventually led to the recognition of HPVs as significant contributors to malignant tumours of the genital tract in humans.

HPVs cause warts but are also the sexually transmitted agents that cause cervical cancer and associated pre-malignant lesions. Although infection with HPV is very common, very few HPV-infected individuals develop cervical cancer. Despite this, cervical cancer is the most common female cancer in the developing world (2), and the second most common among women worldwide with an annual incidence of almost half a million (3). Cervical screening and treatment are costly, and unavailable in most of the developing world. Although promising clinical trials of effective prophylactic vaccines have recently been performed (4,5), the global availability of such a vaccine is still far off.

Most HPV research has focused on the role of HPV in cervical carcinogenesis, but there are considerable gaps in our knowledge of the HPV life cycle, not least in the regulation of the immunogenic HPV capsid proteins. Could activation of the host immune response against capsid proteins, if induced prematurely by anti-viral therapeutic methods, result in clearance of
HPV infection in infected individuals? With the identification of HPV gene regulation steps as possible targets in anti-viral therapy as an eventual goal, HPV-16 gene expression at the level of RNA processing has been investigated in detail in this study.

**Human papillomaviruses and cervical cancer**

Over one hundred different types of human papillomaviruses have been identified to date and all share the common property of infecting epithelia, either cutaneous (e.g. HPV-1 and 2) or mucosal (e.g. HPV-6 and 11) (6). HPV infections can cause benign, self-limiting lesions such as verrucas or genital warts, but can also lead to cancer of the genital tract, or more rarely cancers of the head and neck (7). Mucosal HPV types cause sexually transmitted infections that can be designated *low-risk* or *high-risk* according to their oncogenic potential. Low-risk HPV types such as HPV-11 can lead to the development of genital warts, which usually regress spontaneously. Infection with high-risk HPV types such as HPV-16 and -18 can, however, result in the development of malignant tumours in the genital tract. The majority of high-risk HPV infections are self-limiting, and are cleared by an effective immune response. Despite immune surveillance, in rare cases high-risk types establish a persistent infection that can last several decades, and this is causally linked to the development of cervical cancer (8,9).

HPV infections are very common and the life-time risk of ever contracting HPV is estimated to be 80% (10), but the ultimate development of cervical cancer is rare. Persistent infection with HPV is necessary for the development of cervical cancer but is not sufficient by itself. Another consistent risk factor is the number of sexual partners (10). Cervical cancer arises from pre-existing, non-invasive pre-malignant lesions known as cervical intraepithelial neoplasias (CINs), which are classified according to the progressive abnormal appearance of epithelial cells. The deregulated expression of viral oncoproteins eventually leads to genetic instability and critical alterations of host cell regulatory genes, permitting cancer progression. The long time period between identification of a persistent high-risk infection and development of cervical cancer emphasises that this is a multistep process.

High-risk type HPV DNA is found in 99.7% of cervical squamous cell carcinomas (11). The most commonly associated high-risk types in cervical cancer are HPV-16 followed by HPV-18, -31, and -45. As HPV-16 is associated with 54% of invasive cervical carcinomas (12), this HPV type represents the major cause of cervical cancer, and is also one of the most common cancer-causing infections globally.

**The HPV-16 life cycle**

HPV-16, like all HPV types, is a small, double-stranded DNA virus. It is nonenveloped with an icosahedral capsid, and replicates its genome inside
the nucleus of the infected host cell. HPV-16 exclusively infects mucosal stratified squamous epithelia, and the HPV-16 viral life cycle is strictly linked to the differentiation status of these cells, as illustrated in Figure 1. Through cuts or abrasions virions reach the basal replicating cells of the epithelia. Receptor-mediated entry presumably takes place, which is suggested to be dependent on heparin sulphate (13), but was recently shown not to be the case when normal human keratinocytes were infected by HPV-31b, closely related to HPV-16 (14). A receptor has not been identified, but internalisation occurs through endocytosis of clathrin-coated pits (15,16). Once internalised, the virus is uncoated and an episomal copy number of 5-100 copies per infected cell is established (6). The replication of copy number is achieved by expression of the early non-structural proteins E1 and E2 (17), which are expressed at a low level along with the early oncogenes E6 and E7. The infected basal epithelial cells replicate to produce daughter cells known as transit amplifying cells, which are forced towards the superficial layers and into a differentiating state.

Upon migration of the infected cells towards the uppermost, differentiated layers, a proliferative phase takes place. HPV-16 expresses the non-structural proteins E7 and E6 that function to delay cell-cycle arrest and differentiation, and avoid apoptosis. Both E6 and E7 proteins stimulate cell cycle progression (18), leading to the expression of proteins needed for DNA replication, and normal terminal differentiation is retarded (19).

![Figure 1. A schematic illustration of the differentiating layers of epithelium, upon which the life cycle of HPV is dependent. Daughter cells produced by the replicating basal cells migrate towards the skin/mucosal surface. Once HPV is introduced into an epithelial cell, HPV genes are expressed sequentially during the course of differentiation these cells follow. Whereas the early genes are expresses throughout all layers, the expression of late messenger RNAs and protein, and mature virion production occurs only in the upper-most differentiated layers of the epithelium.](image-url)
In the upper layers the virus must amplify its viral genome and package it into infectious particles. To this end, the virus uses the host cellular machinery to amplify viral copy number, and differentiation-dependent transcription is activated, which in part serves to increase the levels of the early proteins required for replication (20). Thus, when the infected cells terminally differentiate and viral genome amplification is complete, the late structural proteins L1 and L2, which comprise the capsid, are expressed. Based on studies using HPV-33 and immuno-histochemistry staining, L1, the major capsid protein, has been detected after L2, the minor capsid protein (21), allowing the virus genome to be packaged into capsids in the nucleus. Infectious viral particles are released as the superficial layers of the epithelia are discarded.

**Figure 2.** A schematic diagram of the HPV-16 genome, with colour coded bubbles above summarising the functions of HPV proteins. Below the genome is a subset of representative mRNA transcripts and their coding potential depicted to the right. Arrows indicate the positions of the viral promoters p97 and p670. Rectangles represent open reading frames, whereas triangles represent 5' (filled) and 3' splice sites (empty). The early (pAE) and late (pAL) polyadenylation signals are indicated.
The HPV-16 genome and gene products

The HPV-16 genome consists of circularized, double-stranded DNA of approximately 8000 base pairs (bps), encoding all gene products from overlapping open reading frames (ORFs). Genes are expressed from polycistronic messenger RNA (mRNA), transcribed from one strand (6). An untranslated Long Control Region (LCR) contains the origin of replication \((ori)\), as well as an epithelial-specific enhancer (22) and other transcription elements, acting on the early promoter at position 97 of the genome (p97). The early region (E) contains the early genes encoding the non-structural proteins. The early transcripts contain an early, untranslated region (UTR) where the early polyadenylation signal is located (early 3'UTR). The late region (L), coding for the major and minor capsid proteins, L1 and L2, is flanked by a 3'UTR containing the late polyadenylation signal (late 3'UTR). A differentiation-specific “late” promoter has been reported at position 670 of the genome in the E7 ORF (p670) (23). Figure 2 shows a diagram of the genome and some representative mRNA transcripts. Differential usage of promoters and polyadenylation signals, as well as alternative splice sites, results in mRNA transcripts encoding different combinations of early proteins, or transcripts encoding some early proteins followed by either of the late proteins (24).

The early proteins

The early proteins function in the viral life cycle to regulate transcription and replication, and also to deregulate the cell cycle. As a result, some early proteins have transformation potential.

\textit{E1 and E2 replication proteins}

The E1 protein is essential for replication of the virus and weakly recognises AT-rich sequences at the origin of replication \((ori)\) upstream of the early promoter. To facilitate binding, E1 interacts with E2 protein to form high-affinity complexes (reviewed in (25)). E2, also essential for replication, binds to sequences surrounding the \(ori\) and localises E1, which, based on work performed on bovine papillomavirus type 1 (BPV-1) E1, forms a DNA-dependent hexameric complex, with ATPase and helicase activity (26). The E2 protein, required for both replication and transcriptional regulation, has a transactivation domain in its N terminus and binds E1 through its C terminus (27). E2 binds DNA as homodimers at multiple high-affinity E2-binding sites in the LCR and acts as a transcriptional activator or repressor depending on E2 levels (28). At high levels E2 interferes with the binding of cellular transcription factors and represses transcription (29). High levels of E1 and E2 produced from the differentiation-dependent late promoter, not repressed by E2, allow viral amplification. E2 also exhibits coop-
erative transcriptional activation with CCAAT/enhancer binding proteins (C/EBPs), transcription factors that regulate promoters of differentiation-related genes (30). E2 has been shown to induce apoptosis, both p53-dependent and independent (31,32), as well as G1 arrest (33).

The E6 and E7 oncoproteins

It is the oncoproteins E6 and E7 from high-risk HPV types that are responsible for the development of cancer (7,34,35). If integration of high-risk HPV DNA occurs and interrupts the E2 gene, overexpression of E6 and E7 due to loss of E2 can result in proliferating cells that accumulate mutations over time, leading to malignant lesions (34). E7 binds to several members of the pocket protein family, most notably the retinoblastoma tumour suppressor protein (Rb) itself, (36-38), which functions to prevent cell cycle entry into S-phase. E7 binds hypo-phosphorylated Rb, preventing its binding to E2F (39), a key transcription factor in the regulation of the cell cycle, and mediates the ubiquitinylated-degradation of Rb (40). In addition, E7 proteins associate with other cell cycle regulators, cyclins A and E and cyclin-dependent kinase inhibitors p21 and p27 (reviewed in (18,35)). E7 also associates with class I histone deacetylases (HDACs) (41), which play a role in repression of transcription, including that of E2F-dependent genes (42). In a recent report, high-risk type HPV-31 E7 protein activated E2F transcription by its association with HDACs in differentiating cells, thereby facilitating HPV replication (43).

The E6 protein binds the tumour suppressor gene product p53 (44) and targets it for ubiquitinylated degradation by recruiting cellular protein E6-associating protein (E6-AP), a ubiquitin ligase (45). E6 is presumed to thereby prevent apoptosis, which would be triggered upon E7-induced entry of the infected cell into S phase. In addition, E6 can also activate telomerase in human keratinocytes (46), and can bind and degrade members of the PDZ family (47-49). The conserved PDZ family of proteins are involved in the development and control of epithelial polarity (50), and certain family members in Drosophila have been designated tumour suppressor proteins (51), suggesting that degradation of these proteins enables E6 protein to stimulate cell proliferation through its C-terminal PDZ-ligand domain. This is independent of its ability to degrade p53 (35).

E1^E4 and E5 proteins

Due to an alternative splicing event, the HPV-16 E1^E4 protein avails of the E1 start codon to initiate translation. E1^E4 protein has been considered a late protein because its expression coincides with vegetative viral DNA replication prior to the late capsid proteins. It is the most abundant viral protein in HPV-infected epithelia, accumulating in the differentiated cell layers (52). HPV-16 E1^E4 has been shown to cause the collapse of the cytoskeletal network (53) and to associate with mitochondria (54). A G2 arrest function was recently
described for HPV-16 E1\(^{\text{E4}}\) in HeLa cells (55) suggesting that E1\(^{\text{E4}}\) has a role early in the life cycle, possibly to counteract E7-mediated proliferation. The E1\(^{\text{E4}}\) protein may possibly have a role in the regulation of late gene expression. This premise is supported by the accumulated late expression of E1\(^{\text{E4}}\) protein just prior to capsid protein expression and by its interaction with E4-DEAD Box protein (56), member of a family of RNA helicases involved in regulating mRNA stability and degradation (57,58). Lack of HPV-31 E1\(^{\text{E4}}\) protein in a genomic context caused a decrease in genome amplification and late gene expression in differentiation studies, further supporting a late function for E1\(^{\text{E4}}\) protein (59).

HPV-16 E5 protein, a hydrophobic protein associated with cellular membranes, while not a major transforming protein of HPV-16, has weak oncocytic potential (60,61). However, in BPVs, E5 is the major transforming protein (62), acting through the platelet-derived growth factor receptor (63). HPV-16 E5 has been shown to interact with the epidermal growth factor receptor (60), affect its activity (60,64), inhibit its degradation (65,66) and also endosomal trafficking (67), resulting in the induction of downstream signal transduction pathways (68,69). E5 is also believed to interfere with the immune response to HPV infection (70,71). Nevertheless, E5 protein has an undetermined function in the viral life cycle. However, in a recent study of the role of high-risk type HPV-31 E5, lack of E5 impaired activation of late-specific events during differentiation (34), implying a role late in the HPV life cycle.

The late proteins

The L1 and L2 proteins are the structural components of the viral capsid. They form an icosahedral structure, consisting of 360 copies of L1 protein organised into 72 pentameric capsomeres, with 12 copies of L2 protein (72). Both proteins have C-terminal nuclear localization signals (NLSs) to direct them into the nucleus for virus assembly (73-75). However, whereas the L1 protein (~55kDa) can self-assemble into virus-like particles (76), the minor capsid protein L2 (~72kDa) is required for infectivity (77). Our current, albeit incomplete, knowledge of where, when and how the papillomavirus capsid is assembled, has been gathered from many studies using mainly artificial virions from various HPV types, such as HPV-33. L2 is thought to be synthesised before L1 (21) and be located to the nucleus by the L2 NLS, where it accumulates in nuclear domains containing the promyelocytic leukaemia (PML) protein, known as nuclear domain 10 (ND 10) (78,79). L1 may assemble into capsomeres in the cytoplasm, which are relocated to the nucleus by the NLS of L1. L2 then induces the localisation of L1 to these nuclear domains (78), and is incorporated into the capsomeres, facilitating formation of the capsid. L2 has been demonstrated to interact with E2 protein (80), and it has been speculated that the DNA binding capacity of E2
protein could enable packaging of the genome into the viral capsid (78). HPV-31b, harbouring a mutant L2, was able to complete the viral life cycle, but a reduction in viral DNA incorporated into the capsid and infectivity were observed (81).

Late gene expression – a link to cervical cancer development?

Expression of the late genes is tightly linked to the differentiation programme of the epithelial cells infected. Whereas early mRNAs can be detected throughout all layers of the infected epithelium, the late mRNAs are only detected in the more differentiated layers (6). This feature of the papillomavirus life cycle evidently contributes to the evasion of the host immune response by the immunogenic capsid proteins and may be a prerequisite for viral persistence. Indeed, high-risk HPV types produce lower amounts of virus and express the late genes in the more terminally differentiated layers than low-risk HPV types (82), which may also contribute to the ability of high-risk types to persist. Persistence of HPV-16 infection is itself the highest risk factor for the development of cervical cancer (8,83). As the late mRNAs or proteins are never found in cervical cancer cells, one can speculate that inhibition of late gene expression is required for the development of cervical cancer.

Differentiation-dependent transcription, as well as co-transcriptional and post-transcriptional processes, regulates the inhibition of late gene expression (84). RNA processing events taking place concurrently with transcription, such as alternative splicing and polyadenylation, provide a strategic means for the regulation of late gene expression. Downstream events such as export, stability and translation in the cytoplasm are intimately linked to the late mRNAs’ nuclear experience. Co-transcriptional regulation of the late mRNAs during RNA processing is a major topic in this thesis. Below is an outline of gene expression in eukaryotes, described from the perspective of RNA processing.

The central role of RNA processing in gene expression

RNA processing refers to a series of modifications that bestow upon mRNAs molecular features required for recognition by the translation machinery. RNA processing can be seen as the hub of gene expression, where many intricate processes converge, and the fate of the mRNA is most often defined. In order to understand these processes that viruses also employ and sometimes hijack, it is important to recognise the central role of RNA processing in eukaryotic expression.

A complicated and highly regulated multi-step process, eukaryotic gene expression involves the conversion of information in DNA coding se-
quences into functional proteins, co-ordinated to occur at the appropriate
time and location. Protein coding units – genes – must be identified within
the genomic DNA and transcribed into specific precursor mRNA, which is
then processed and packaged for export through pores in the nuclear mem-
brane. Once in the cytoplasm, ribosomal machinery translates the mRNA
into a protein, which in turn may need post-translational modifications to
become functional. Decay of both the mRNA and the protein it encodes can
be seen as final steps in the regulation of gene expression.

Transcription, the first step in gene expression, occurs in the nucleus and is
carried out by RNA polymerases. These multi-subunit enzymes are recruited
to the promoter region of the gene, from where transcription of that gene
starts. Of the three types of RNA polymerases in a eukaryotic cell, it is RNA
polymerase II (RNAP II) that transcribes protein-coding genes into mRNA.
The promoter region consists of at least three components, a transcription
start site, the TATA box, and sequences bound by transcriptional regulators,
the first two of which are sufficient for transcription initiation by basal tran-
scription factors (reviewed in (85)). Transcription is highly regulated; for
example transcriptional activators recruit large complexes controlling acces-
sibility of the chromatin structure. Transcription factors in turn regulate the
recruitment of the RNA polymerase to the promoter region (reviewed in
(86)). As we shall see below, the subsequent gene expression steps also
“reach back” to influence transcription itself.

The formation of an export-competent and functional mRNA requires not
only synthesis but also processing in the nucleus. A eukaryotic precursor
mRNA (pre-mRNA) must be capped at its 5’ end, spliced to remove non-
coding sequences, and polyadenylated at its 3’ end. The occurrence of cap-
ping during the initial stages of transcription has been known for a few dec-
dades (87), but as evidence is mounting that other steps also occur concur-
rently with pre-mRNA transcription, it is now generally accepted that
RNAP II transcription and mRNA processing are tightly coupled. This close
association allows RNAP II transcription and mRNA processing events to
be regulated by each other (reviewed in (88)).

When RNAP II has cleared the promoter region, a regulatory region of
the large subunit of RNAP II, known as the carboxy-terminal domain,
(CTD), becomes phosphorylated and elongation of the transcript begins.
The nascent transcript exiting the elongating RNAP II is the site of mRNA
processing events described below. The CTD of RNAP II acts as a loading
platform to recruit mRNA processing factors (reviewed in (89-92)). This
serves to concentrate processing factors near the nascent transcript and al-
lows the rate of transcription elongation to influence processing events such
as splicing (93,94). Truncation of the CTD has been shown to cause defects
in capping, splicing and polyadenylation (95); in addition, isolated frag-
ments of the CTD can activate in vitro splicing (96). Phosphorylation of the CTD can activate or inhibit mRNA processing factors (97) and proteins involved in one mRNA processing step may influence another (98-102).

The first mRNA processing factors to be recruited to the CTD are those of the mRNA capping complex (103,104), the association of which is controlled by phosphorylation of the CTD (105,106). When the first 22 to 25 nucleotides of the mRNA have been transcribed, a cap structure, 7-methylguanylate is added onto the 5' end (107). The 5' cap confers stability to the mRNA by protecting it against 5'-3' exonucleases and plays a role in splicing, export and translation of the mRNA (107,108). The 4E subunit of the eukaryotic translation initiation factor eIF4F and the cap-binding complex (CBC) bind to the 5’ cap (109,110) and facilitate the recognition of the capped mRNA by the translation machinery (109). The CBC is an example of a factor assembling early on the nascent transcript that is linked to the subsequent steps of splicing, translation and mRNA decay (107,108,110-112).

Splicing of non-coding, intronic sequences can occur as soon as the splicing signals are transcribed. Recent experiments indicate that efficient co-transcriptional splicing is dependent on the phosphorylation of the CTD in vivo (113). The outcome of splicing is coupled to the regulation of transcription at different levels (reviewed in (114,115)). Changes in the accessibility of competing splice sites and regulatory complexes caused by different rates of transcription constitute a link between alternative splicing and transcription. That the elongation rate modulates alternative splicing has been demonstrated in vivo both in mammalian cells and yeast (93,94). Choice of promoter and phosphorylation status of the recruited RNAP II CTD will affect mRNA processing factor recruitment. For example, different RNAP II promoters can result in different alternative splicing patterns (116-118). Transcriptional activators influencing initiation and elongation affect alternative splicing patterns differentially (119) and the effect of elongation on splicing seems dependent on the strength of the splicing signal (120). Splicing itself affects not only downstream processing events such as translation (121) and export (122), but also decay (123) and localisation (124).

Another RNA processing event that has been linked to splicing is the co-transcriptional modification of RNA editing. The most important form of editing in higher eukaryotes is base conversion, where typically adenosines in nuclear double-stranded RNA are converted to inosines by a family of adenosine deaminases. As inosine is recognized as guanosine in the translation machinery, a missense codon occurs rather than a nonsense codon. Consequently, in addition to amino acid change, editing can alter splicing signals and therefore induce alternative splicing (125), and may be important in nervous system development (126) and disease (127).
All functional eukaryotic mRNAs are processed at their 3’ end by a two-step process of cleavage and polyadenylation (128). Firstly, the pre-mRNA is recognized and cleaved at a site 20 to 30 nucleotides downstream of a highly conserved poly(A) site AAUAAA, involving several factors. Secondly, poly(A) polymerase (PAP) adds a tail of adenosines to the 3’OH end, which is not encoded by the gene. RNAP II transcribes past the polyadenylation site and it seems that the degradation of the RNA downstream of the cleavage site by a recently reported 5’-3’ RNA exonuclease (129,130), is linked to termination of transcription. Cleavage/polyadenylation factors are already recruited to the CTD at the promoter, and increasingly along the gene towards 3’ end of the pre-mRNA (131).

Figure 3. The central role of RNA processing in the regulation of gene expression. The CTD of RNA polymerase II loads onto the emerging transcript RNA processing factors, which dictate the fate of the mRNA in processes such as packaging, export and nonsense mediated decay. Nuclear RNA processing in turn determines the fate of the mRNA in the cytoplasm. RNAP II, RNA polymerase II, NPC, nuclear pore complex, CBC, cap-binding complex, SR proteins, serine/arginine rich proteins, hnRNPs, heterogeneous ribonucleoproteins.
During transcription, RNA-binding proteins assemble on the emerging nascent transcript to form a mature export-competent mRNA, in the form of a densely packed ribonucleoprotein complex, known as a messenger ribonucleoprotein particle (mRNP). These factors “label” the mRNP for processing, nuclear export, localization, stability and translation. Transcription and the mRNA processing events described above are linked to the packaging of mRNPs so that only correctly processed and packaged mRNAs enter the cytoplasm (reviewed in (132-134)). As well as proteins directly implicated in processing, a diverse group of nuclear RNA-binding proteins known as heterogenous nuclear ribonucleoproteins (hnRNPs) also bind to the emerging newly transcribed transcript. The hnRNPs are involved in multiple processes and the choice and arrangement of hnRNPs will influence the fate of the mRNA. Cytoplasmic roles for hnRNPs have been implicated in processes such as translation, localization and stability, as well as shuttling activity (reviewed in (135,136)).

An important complex of proteins sequentially deposited on spliced transcripts, affecting downstream processing, is the exon-exon junction complex (EJC). This is found ~20-24 nucleotides upstream of the 5’ splice site, after splicing has taken place (137,138). The complex includes the Y14:Mago heterodimer, eIF4AIII, splicing co-activators RNPS1 and SRm160, export factors Aly, UAP56 and TAP/NFX1:p15 as well as the nonsense-mediated mRNA decay (NMD) factors Up frame-shift proteins Upf2 and Upf3b ((139) and references therein). Through the interactions of its member proteins, the EJC complex is functionally linked to export (140-142) and also degradation of aberrant mRNAs by NMD (123,143,144), an mRNA surveillance mechanism described below.

Export competent mRNPs are transported from the nucleus to the cytoplasm by interacting with export complexes at docking sites in the nucleopor complexes (NPCs) of the nuclear membrane. Export factors for mRNA are loaded on transcripts during elongation, linking the export of mRNPs with transcription (133). A central player in the export of mRNA to the cytoplasm is the conserved TREX (Transcription/Export) complex, which contains the THO complex and mRNA export proteins UAP56 and Aly (145), suggested to be recruited during splicing in mammals (146). These three components have been reported as present in the spliceosome (147), linking splicing to export. UAP56 and Aly, also found in the EJC, were recently seen in vivo to co-localise with the spliceosome at transcription sites and this was dependent on splicing (148). The mRNA export adaptor Aly, a member of the RNA export factor (REF) family, binds directly to UA56, a putative ATPase/RNA helicase, and recruits nuclear export factor TAP/NXF1, that in turn recruits NXT/p15. This TAP complex binds to nucleoporins, facilitating mRNP transport through the NPC (reviewed in (132,133,149,150)).
Incompletely or incorrectly processed mRNPs are retained in the nucleus and degraded. This can occur at many stages of mRNA processing (reviewed in (151)). Degradation of mRNA is initiated at the poly(A) tail by deadenylases. Deadenylation can trigger decapping by a two-subunit decapping enzyme, exposing the mRNA to degradation by a 5’-3’ exonuclease, or alternatively deadenylation is followed by 3’-5’ degradation (reviewed in (152)). The nuclear exosome, a complex of 10 or more 3’-5’ exonucleases, targets the body of mRNPs, whereas 5’-3’ degradation is carried out at the unprotected 5’ end by Xrn2 in the nucleus, an enzyme recently linked also to transcription termination (130), or Xrn1 in the cytoplasm. General mRNA degradation or turnover in the cytoplasm is a critical step in the regulation of gene expression, and is mediated by the same enzymes as in the decay of specific aberrant mRNAs. Processing bodies in the cytoplasm have been shown to contain the decapping complex and other proteins involved in mRNA turnover and may represent special cellular compartments of mRNA turnover by decapping and subsequent 5’-3’ degradation (reviewed in (153)).

Rapid degradation of mRNAs can be triggered by the mRNA surveillance mechanism of NMD, which recognizes and targets mRNAs containing premature termination codons (PTCs) as a result of mutations or errors in transcription or processing. This prevents the translation of a truncated protein potentially harmful to the cell. The recognition of PTCs is dependent on translation and although NMD is conserved, the recognition of PTCs and the decay pathway triggered vary among species (reviewed in (154)). In mammals, NMD will occur if the nonsense/stop codon is more than 50-55 nucleotides upstream of an EJC deposited on an exon-exon boundary. If translating ribosomes stall at a stop codon upstream of an EJC, the Upf3b NMD factor, present as part of the EJC, recruits Upf1, an RNA helicase. Upf1 associates with translation termination factors and is presumably the link between surveillance and translation termination (reviewed in (155,156)). NMD-triggered decay in mammals can occur by rapid deadenylation followed by 3’-5’ degradation (157), but decapping can also be activated (158,159). Given that one third of all alternatively spliced transcripts harbour a nonsense codon that can trigger NMD (160), it is not surprising that a recent microarray study implicated NMD to be widely involved in regulating normal gene expression (161).

The ribosomal translation machinery, a multi-component complex of RNA and proteins, translates the mRNA sequence into an amino-acid peptide chain in the cytoplasm. Similar to transcription, translation has three stages – initiation, elongation and termination (reviewed in (162)). In terms of translational regulation of gene expression, initiation is the rate-limiting step and is most subject to control (163). Initiation starts with the 40S ribosomal subunit binding to the cap, aided by initiation factors (eIFs). The complex of
40S and factors scan the 5’ untranslated region (5’UTR) for the initiation codon. When this is located, the eIFs are released and the 60S ribosomal subunit is recruited before elongation can begin. The mRNA itself contains many features that regulate its translation, not least the 5’ cap and poly(A) tail, which can modulate translation efficiency. Translation can be affected by the initiation codon sequence surroundings, secondary structure features of the 5’UTR, upstream start codons or small ORFs and the presence of internal ribosomal entry sites (IRES). Both the 5’UTR and 3’UTR may contain binding sites for regulatory factors determining translation efficiency (reviewed in (164)). The stability of mRNAs in the cytoplasm can also be controlled by elements in the 3’ UTR, which will be discussed further below.

**Alternative splicing**

Certain RNA processing steps in the formation of mature mRNA, namely splicing and polyadenylation, are of special interest in this thesis. Below is a more detailed description of these, with a view to highlighting the processes in particular that HPV depends on to regulate late gene expression co-transcriptionally.

Alternative splicing is of paramount importance for viruses to maximize the protein coding capacity of their relatively compact genomes, and was first observed in another small DNA virus, adenovirus (165,166). Alternative splicing enables viruses to tightly regulate the temporal expression of genes during different phases of infection. The use of alternative splicing to control the timely expression of the late capsid proteins in HPV-16 is a central theme of this thesis, and therefore the mechanism of splicing and its regulation will be described below in more detail.

**The importance of being “spliced”**

One cannot underestimate the importance of being “spliced”! With the successful mapping of the human genome, it is now known that humans have an unexpectedly low number of genes – 32,000 protein-coding genes in comparison to the nematode *Caenorhabditis elegans* with 19,000 and the fruit fly *Drosophila melanaster* with 14,000 (167-169). What then lies behind the biological complexity of humans if not the number of genes? The answer is proteomic diversity due to alternative splicing. Alternative splicing, the removal of non-coding introns to selectively join exons, generates multiple mRNAs encoding distinct proteins from a single pre-mRNA, thereby increasing the coding capacity of genes. This is often particular to certain tissues, stages of development and even disease states. Alternative
splicing can result in the loss or addition of coding sequences, a frame shift, or the presence of a premature stop codon. Regulatory elements controlling translation, stability or localisation can also be added or removed by alternative splicing. In fact, 80% of alternative splicing leads to changes in the encoded protein (170). Between 60-74% of human genes are estimated to be alternatively spliced (167,171), yet relatively few reports of how splicing regulates the expression of human genes have been made. Given the much cited estimation that 15% of all point mutations linked to human genetic diseases deregulate alternative splicing (172-174), and additional links between splicing and cancer (reviewed in (175)), the genome-wide study of alternative splicing regulation should become an area of major medical interest.

Pre-mRNA Splicing

Alternative splicing is a central processing event in the regulation of gene expression, first identified almost thirty years ago (165,166). As many mRNAs contain multiple splicing possibilities and some contain thousands, the potential effect on protein diversity can be enormous, and can be crucial in many cellular and developmental processes. Most exons are constitutively spliced, meaning that they are always included, whereas regulated exons can be included or skipped depending on the outcome of competing splice sites and surrounding regulatory sequences. Alternative 3’ ends of transcripts through alternative polyadenylation sites also play a role in determining the outcome of combined splicing events (Figure 4).

Figure 4. Various modes of alternative splicing. A(n) represents sites of polyadenylation.
Splicing excises intronic sequences, which in humans can be thousands of base pairs long. Exon sequences are generally shorter, between 50 and 300 nucleotides long, and can sometimes nestle in introns tens of thousands of base pairs long (167). In spite of the obvious difficulties this creates, splicing must be precisely carried out to ensure that coding sequences joined together are in frame for translation.

**Splice signals**

Short conserved sequences required for excision of an intron from pre-mRNA are located on the intron (Figure 5). There are two major classes of introns, one less common intron type known as U12-dependent (AT-AC) (176), and the more common U2-dependent type, which will be discussed here. The 5’ splice site signals the 5’ end of the intron, consisting of the invariable dinucleotide GU surrounded by a less conserved sequence. The 3’ end of the intron is defined by three conserved sequence elements, the branch point sequence, followed by a polypyrimidine tract and the 3’ splice site (AG). The branch point sequence, found 18 to 40 nucleotides upstream of the AG dinucleotide, contains an invariable adenosine residue crucial for the splicing reaction, surrounded by a conserved sequence, whereas the polypyrimidine tract is characteristically of variable length and sequence. Interrupting purines often weaken this uridine-rich element, the length of which correlates with the ability of the splicing factor U2 auxiliary factor (U2AF) to be recruited. U2AF in turn facilitates the recognition of the 3’ splice site, consisting of the invariable AG dinucleotide in a surrounding sequence. These splicing signals act as binding sites for splicing factors, allowing the splicing machinery to assemble onto the pre-mRNA.

**The mechanics of splicing and spliceosome assembly**

Splicing, the removal of introns and catalysis of the ligation of exons, is carried out by the spliceosome, comprised of a set of small nuclear ribonucleoprotein particles (snRNPs) and possibly up to 300 additional proteins (177). Numerous studies investigating the components of the spliceosome at different stages in the splicing reaction have revealed the dynamic nature of its make up and the presence of accessory splicing factors such as serine/arginine (SR) proteins, hnRNP proteins and constituents of the export, decay and polyadenylation machinery (147,177-179).

Five very abundant nuclear uridine-rich snRNPs (U snRNPs), each containing a small stable RNA (snRNA) bound by core components Sm or Sm-like proteins, and some proteins specific to each, direct the splicing reaction. U1, U2, U4 and U5 snRNPs are transcribed by RNAPII and are assembled with protein components and modified in the cytoplasm before returning to the nucleus. However, U6 is transcribed by RNAPIII and U6 snRNP is thought to be assembled in the nucleus (reviewed in (180)).
Figure 5. Splice site elements and the mechanics of splicing. (A) The consensus splicing signals and their location on the pre-mRNA. Y represents U or C, R represents G or A, and N represents any nucleotide. (B) Splicing consists of a two-step reaction, resulting in the ligated exons and a lariat formed intron.

Assembly of the spliceosome is promoted by the base-pairing of snRNAs with the pre-mRNA at the splicing signals on the intron described above and occurs in a step-wise fashion (Figure 6) (181). This begins with the recruitment of U1 snRNP by base pairing of the U1 snRNA to the 5’ splice site. Splicing Factor 1 (SF1) binds to the branchpoint (182), while the 65kDa subunit of U2AF (U2AF65) binds the polypyrimidine tract (183). In most cases the 35kDa subunit of U2AF (U2AF35) binds the 3’ splice site to help recruit U2AF65 to the polypyrimidine tract (184). U2 snRNP base-pairs with the branchpoint, facilitated by the binding of U2AF (185). The U4/U6-U5 triple snRNP joins the spliceosome before it undergoes extensive rearrangement to produce an active splicing-competent complex, across the intron. The activity of a group of DexD/H box RNA helicases, perhaps to rearrange base-pairing between snRNAs and the pre-mRNA, makes spliceosome assembly ATP dependent (181). Catalysis of the splicing reaction may be carried out by an RNA component of the spliceosome (186,187). It has been speculated that U6 RNA may perform this function as it coordinates a metal ion necessary for spliceosome catalysis (188).

The splicing reaction consists of two trans-esterification reactions (Figure 5). The 2’ hydroxyl group of the invariable branch point adenosine attacks the phosphate at the 5’ splice site. The 5’ exon is cleaved from the intron,
which is ligated to the branchpoint adenosine, forming a lariat structure with the 3’ exon. Then the 3’ hydroxyl group of the 5’ exon attacks the phosphate group on the 3’ end of the intron. Thus, the two exons are ligated and the lariat-formed intron excised.

Figure 6. A simplified overview of splicing complex assembly. Splice site elements on the pre-mRNA are bound by splicing factors as indicated. The early commitment complex is facilitated by the interactions of SR proteins. U2 snRNP replaces SF1 in the A complex, and splicing takes place upon rearrangement of the complex, following the arrival of U5:U4/U6 snRNPs.

Regulation of splice site choice – defining the exon

How does the spliceosome recognize the boundaries of intron/exons to be spliced? The sequences of splice site signals are often weakly conserved, and are necessary but not enough to define accurate splicing amid a myriad of sequences resembling splice sites that are never recognized by the spliceosome. Consequently, how are relatively small exons defined within very large introns such as in the human genome? It has been proposed that spliceosomal proteins on 3’ and 5’ splice sites flanking an exon interact to define the exon early in splicing, in a process known as exon definition (189,190). In addition, interaction between the 5’ cap on pre-mRNA and the proximal 5’ splice site defines the first exon (111). Polyadenylation at the 3’
end of pre-mRNA in turn promotes the use of the upstream 3’ splice site and removal of the last intron (191,192) through interaction of splicing and polyadenylation machinery, thereby defining the terminal exon (an exon flanked by a 3’ splice site and a poly(A) site at the 5’ end of the pre-mRNA).

The need for modulating alternative splicing in a cell-specific or temporal fashion, or for example in the tightly ordered expression of viral genes during infection, demands extra complex signals to recruit the spliceosome. Splice site choice defined by exon recognition can be regulated by specific exonic or intronic sequences known as splicing enhancers and suppressors. These promote or suppress the recognition of suboptimal 5’ and 3’ splice sites, thus regulating gene expression (193,194). These cis-acting elements, described in more detail below, provide binding sites for proteins regulating splice site choice, most notably the well-studied family of splicing regulators, the SR proteins.

SR proteins

The serine-arginine (SR) family of proteins is a family of highly conserved proteins with roles in both constitutive splicing and alternative splicing (195). SR proteins are believed to recruit splicing factors to form initial complexes during spliceosome assembly. Stabilisation of the interaction of U1 snRNP with the 5’ splice site (196) by SR proteins has been reported, as well as the recruitment of the U4/U6-U5 triple snRNP by SR proteins (197). Furthermore, they may help pair 5’ and 3’ splice sites through their simultaneous interaction with U1-70K, a U1 snRNP component, and U2AF35 (196,198).

SR protein structure

The SR proteins are a growing family of structurally related phosphoproteins, characterised by the presence of 1 or 2 RNA recognition motifs (RRMs) and a distinctive arginine/serine dipeptide-enriched C-terminal domain, known as an RS domain. All SR proteins contain one type of RRM that contains binding motifs common to many RNA binding proteins, whereas a subset has an additional RRM, which is recognized by an invariant amino acid sequence. The SR proteins fall into two categories based on domain structure: SRp20, SC35, 9G8, SRp46, SRp54 have one RRM, whereas SRp30c, ASF/SF2, SRp40, SRp55, and SRp75 have two RRMs. The RRMs are required for sequence specific binding of SR proteins to RNA. Much work has gone into trying to identify specific binding sites for each SR protein, which appear to have some distinct binding properties but can also be redundant in their ability to activate specific splicing events ((199) and references therein).

The RS domains, which can be phosphorylated at multiple sites, are of varying length and sequence and are accepted to be involved in protein-
protein interactions. Proteins containing RS domains have been shown to interact, where the RS domain was necessary (196,198). Interesting experiments in which the RS domain of different SR proteins were tethered artificially to pre-mRNA, showed that RS domains could activate splicing and this activity depended on the amount of RS dipeptides present (200,201). In addition, many publications document the role of the RS domain in the regulation of alternative splicing by SR proteins (reviewed in (193,195,202)). However, recent publications describe the interaction of the RS domain of the splicing factor U2AF35 with the branchpoint, and the RS domain of an enhancer-bound SR protein with the 5’ splice site on pre-mRNA (203,204). This argues that RS domains may contact the pre-mRNA at different stages of spliceosome assembly to stabilize splicing.

**Regulation of alternative splicing by SR proteins**

The regulation of alternative splicing by SR proteins is carried out mainly through exonic regulatory RNA elements that act as binding sites for SR proteins. Enhancers are often adjacent to introns that have weak splicing signals, the splicing of which SR proteins promote. For example, a splicing enhancer downstream of a 3’ splice site in the ED1 exon of the fibronectin gene allows more efficient binding of the U2snRNP through the binding of SR proteins (205). Several studies support a model in which enhancer-bound SR proteins activate the use of upstream weak 3’ splice sites by promoting recruitment of U2AF65 to compensate for a poor polypyrimidine tract (206,207). More recent studies tried to elucidate if SR proteins recruited U2AF65 to the polypyrimidine tract by interacting with the RS domain of U2AF35 (reviewed in (193)), and it seems that the recruitment of U2AF35 are required for maximal enhancer-dependent splicing (208).

Alternatively, SR proteins can act to promote the use of a downstream 5’ splice site by binding directly to an upstream enhancer (209,210), or can also bind as part of a large complex, as seen in the well-characterised regulation of alternative splicing in the *Drosophila* sex determination cascade (reviewed in (211)). SR proteins can also counteract the antagonistic role of hnRNPs recognizing exonic splicing silencers in alternative splicing. A well-documented example is the opposing roles of SR protein ASF/SF2 and hnRNP A1 in the selection of 5’ splice sites. Excess ASF/SF2 results in selection of a 5’ splice site proximal to the 3’ splice site, whereas hnRNP A1 favours the distal 5’ splice site, both *in vitro* and *in vivo* (212,213). This led to cross-linking experiments showing that ASF/SF2 enhances the binding of U1 snRNP to 5’ splice sites, but hnRNP A1, however, reduces binding of U1 snRNP, perhaps by competing for binding (214). Further work revealed that the relative concentration of these factors vary in many different cell types (215), indicating that the ratio between positive and negative regulatory factors may determine alternative splicing in cell-specific situations.
The importance of alternative splicing in determining organism complexity has led to investigations to determine if SR proteins may be vital in physiological processes where splice site choice is biological significant. A recent functional study showed that a heart-specific knockout of ASF/SF2 switched “off” a specific isoform of an important transcript in the maturation of the heart, resulting in excitation-contraction coupling defects (216). As well as their role as key regulators in tissue-specific alternative splicing, SR proteins may also act as splicing regulators in response to extracellular signals. SRp38 was shown to act as a splicing suppressor in M phase of the cell cycle (217) and also in response to heat shock (218), establishing a connection between acute stimuli and the control of gene expression during RNA processing.

**SR proteins – not just splicing regulators**

Phosphorylation of the RS domain, required for SR protein participation in spliceosome assembly, regulates the activity of SR proteins, both in protein-protein interaction and in the ability of SR proteins to regulate alternative splicing (219-221). Whereas all SR proteins are found in the nucleus, distributed both in the nucleoplasm and in interchromatin granules or speckles (222), a subset of SR proteins can shuttle in and out of the nucleus (223), and their subcellular localisation is modulated by phosphorylation (224,225). Therefore, phosphorylation could have an impact on the role of SR proteins as factors linking between splicing and export. Initially SRp20 and 9G8 were shown to promote the export of intronless mRNA (226). Following studies showing that REF proteins were not essential in export (227), SRp20 and 9G8, along with ASF/SF2, were suggested as alternative adapters for the export factor TAP/NXF1 (226,228). Huang et al recently demonstrated that the dephosphorylation of export adaptor SR proteins was critical for their role in TAP dependent export (229). Consequently, it has been speculated that the dephosphorylation of SR proteins in the mRNP may then label the mRNP as spliced and ready for export (230). In addition, a recent report showed that the overexpression of certain SR proteins enhanced the sensitivity of PTC containing pre-mRNAs to NMD and that the RS domain was required for this (231). A further cytoplasmic role for SR proteins might be in translation. Sanford et al recently described the association of the shuttling SR protein ASF/SF2 with polysomes, and demonstrated that MS2-tethered ASF/SF2 could stimulate translation of reporter mRNA in *Xenopus* oocytes. Overexpression of wildtype ASF/SF2 also promoted translation of mRNAs containing ASF/SF2 binding-sites in a luciferase assay in HeLa cells (232). These were interesting findings, suggesting that not only could shuttling SR proteins promote a certain splicing variant but could also stimulate its translation.
Additional splicing signals

Exonic splicing elements

As mentioned above, extra information is required to dictate the order of exons to be included during alternative splicing. Exonic splicing enhancers (ESEs) determine if the exon on which they are located on should be included, whereas exonic splicing silencers (ESSs) can result in the skipping or exclusion of that exon, or suppress the splicing of an adjacent intron (Figure 7). Both elements have been reported in cellular and viral genes (194). Most reported ESEs bind SR proteins and the most common class of enhancers are purine-rich enhancers. These usually contain a core motif of alternate As and Gs given the sequence \((\text{GAR})_n\), and enhancers that regulate the use of 3’ splice sites include those described in the human fibronectin ED1 exon (205) and the cardiac troponin T gene (233). An example of a purine-rich ESE that regulates 5’ splice site usage was reported in exon 5 of the caldesmon gene (234), whereas Bourgeois et al described a purine-rich enhancer in adenovirus E1A pre-mRNA that enhanced the use of both upstream 3’ and downstream 5’ splice sites (209). ESEs can also function to antagonize ESSs; this is the case in exon M2 of the IgM gene, where removal of an ESS allowed splicing without an ESE sequence (235).

Non-purine-rich ESEs include the newly described AC-rich enhancers (ACEs), initially characterised using \textit{in vivo} selection experiments (236). ACEs have since been shown to enhance splicing at an upstream 3’ splice site in an NS2-specific exon of parvovirus MVM (237), at a differentiation-specific 3’ splice site in BPV-1 (238) and in exon 4 of the calcitonin/calcitonin gene-related peptide (CRGP) gene (239). A particular family of proteins does not seem to bind ACEs. One reported ACE-binding cellular factor is YB-1 protein, demonstrated to bind to an ACE in exon v4 of the human CD44 gene (240).

ESSs inhibit the use of upstream 3’ splice sites and in some cases antagonize the activity of ESEs (241). Although not as well studied as ESEs, ESSs may be more common than previously thought and play a critical role in alternative exon inclusion (242). While ESEs usually contain SR protein binding sites, ESSs have less similarity to each other but are commonly bound by members of the hnRNP family including hnRNP A1 (243), hnRNP H (244) and hnRNP L (245). PTB, also a member of the hnRNP family, can act through ESSs (246,247), though this protein more often negatively regulates alternative splicing through intronic sequence elements (248).
Intronic splicing elements

Intronic splicing enhancers and silencers are not as well documented as their exonic counterparts. One example of an intronic splicing enhancer (ISE) is a U-rich sequence found downstream of a 5’ splice site, demonstrated to enhance splicing in the FGFR 2 transcript (249) and also in the Drosophila Msl2 intron (250). Both ISEs bind TIA-1 protein, which stimulates binding of U1 snRNP to the 5’ splice site (250). Another type of ISE is the sequence UGCAUG, which enhances splicing in c-src, fibronectin and calcitonin transcripts (reviewed in (211)). A further class of intronic enhancers are CA-rich sequences. Hui et al recently reported that dinucleotide CA repeats act as intronic splicing enhancers and that their stimulatory activity depends on the number of repeats, and can be correlated to the binding of hnRNP L (251). First demonstrated in the human endothelial nitric oxide synthase gene (252), this work was extended to other transcripts and confirmed that intronic CA repeats activate an upstream 5’ splice site at a close distance, and that their action was intron-specific and sufficient to enhance splicing (251).

Intronic splicing suppressors (ISSs) that block the use of downstream 3’ splice sites have been reported. A purine-rich intronic sequence near the branchpoint of a 3’ splice site in adenovirus L1 mRNA acts to block splicing when bound by SR proteins, which later in infection become dephosphorylated; as a result, splicing at this site is activated (221,253). The hnRNP A1 transcript contains several intronic splicing elements, including an ISS, through which SRp30c suppresses the use of a 3’ splice site (254). In the same transcript, hnRNP A1 itself regulates exon exclusion by binding the introns on either side of exon 7b (255). In exon 3 of the HIV tat gene, hnRNP A1 suppresses a 3’ splice site by binding at a branchpoint sequence and blocking U2 snRNP binding (256). PTB has been implicated in the repression of splicing through intronic elements, typically at the polypyrimidine tract of 3’ splice sites, including those in the SM exon of the alpha-actinin gene, alpha-tropomysin and the PTB transcript itself (reviewed in (257)).
Regulation of mRNA 3’-end formation

Due to the confirmed significance of alternative splicing in genetic diversity, 3’-end formation of mRNA is somewhat underestimated in RNA processing. Nevertheless, alternative 3’ end formation occurs in over 50% of human genes (258), and is very significant in the regulation of alternative splicing itself (259). Additionally, as new coding sequences or regulatory sequences within alternative 3’ UTRs can be introduced, alternative 3’ ends increase the diversity of transcripts. In relation to the coupling of gene expression steps, the CTD of the RNAP II permits the interdependence of 3’-end formation with other RNA processing steps, as polyadenylation factors are co-localised with other processing factors on the CTD already at the promoter (260,261). Furthermore, mutations in the RNAP II CTD abolish polyadenylation (95). The 3’-end formation of mRNA is functionally linked to transcription termination (262) and mRNA export (263,264). In the cytoplasm the poly(A) tail confers mRNA stability and enhances translation by interactions between poly(A) tail binding proteins and translation initiation factors bound to the 5’ cap (265,266). In summary, 3’-end formation can have a profound effect on the sequence and fate of a transcript and therefore gene expression.

Poly(A) tails allow viral transcripts to be transported and then translated by cellular machinery. Polyadenylation site selection is also very critical for the temporal expression of viral genes. In fact, regulation of polyadenylation has been described mostly in viruses, emphasizing the importance of polyadenylation site switching in the life cycle of many viruses. In the case of a small virus such as HPV-16, regulation of 3’ end formation, or polyadenylation as it is referred to in the detailed discussion below, plays a crucial role in gene expression.

The polyadenylation machinery

The two-step process of site-specific endonucleolytic cleavage and polyadenylation, and the efficiency with which these processes are performed, is determined by sequences on the pre-mRNA (reviewed in (107,267,268). The core polyadenylation signal consists of the conserved hexanucleotide AAUAAA motif 20 to 30 nucleotides upstream of where the RNA is cleaved and adenosine residues are added. The cleavage site itself and a less conserved downstream element (DSE) consisting of a GU or U rich stretch, 20 to 70 nucleotides downstream of the cleavage site (269,270), complete the core polyadenylation signal. The site of cleavage is determined by the distance between the AAUAAA element and the DSE (271) and can commonly be defined by a CA dinucleotide (268). The multimeric ‘Cleavage and polyadenylation specificity factor’ (CPSF) loosely recognizes and binds AAUAAA. Poly(A) sites containing sequences deviating from the consen-
sus AAUAAA hexanucleotide are also recognized, as the original estimation of the presence of AAUAAA in 90% of mRNAs was recently suggested to be somewhat lower (reviewed in (272)). The binding of CPSF is greatly enhanced by cooperative interaction with ‘cleavage stimulatory factor’ (CstF), which binds to the DSE. This complex then associates with cleavage factors CF1 and CF2 (273). Subsequently, the pre-mRNA is cleaved by an endoribonuclease activity, the identity of which may be confirmed soon, as a subunit of CPSF called CPSF-73, has recently been proposed as a candidate (274). This is partly due to its metal-binding properties and the requirement of Zn$^{2+}$ for 3’ cleavage (274). The CTD of RNAP II and poly(A) polymerase (PAP) are also involved in cleavage, whereas the addition of the poly(A) tail involves CPSF, PAP and poly(A) binding protein (PAPBII). The free 3’ end is polyadenylated to 200 to 250 A residues by PAP, which can be regulated by phosphorylation (275). PAPBII is required for rapid elongation and length determination of the poly(A) tail (276).

The polyadenylation machinery, illustrated in Figure 8, is complex for such a relatively simple process and most of the polyadenylation factors have several subunits. CPSF contains subunits -160, -100, -73 and -30 kDa, in addition to the recently-identified component named hFip1, which interacts with PAP (277). CstF consists of three subunits of 77, 64 and 55 kDa. The key RNA-binding factors in polyadenylation are the CPSF subunit, CPSF-160 (278), and the CstF-64 subunit of CstF (279). The binding of CstF-64 to the DSE seems to be a major point of control in poly(A) site choice. However, an interesting recent study suggests that CF1 may be involved in the recognition of non-AAUAAA poly(A) sites, because it was sufficient to direct polyadenylation from a non-AAUAAA poly(A) site in vitro by recruiting hFip1 and PAP (280).

Figure 8. Assembly of the basal polyadenylation machinery at the core polyadenylation signal on the pre-mRNA. CPSF subunits are in blue, whereas CstF subunits are in green. An arrow indicates the cleavage site.
In addition to the core polyadenylation signal, additional sequence elements can positively or negatively regulate polyadenylation efficiency. Examples of these are auxiliary upstream polyadenylation enhancers (USEs), often U-rich in sequence and commonly found in viral poly(A) sites. As these are of particular interest to this thesis, USEs will be more fully described below.

Figure 9. Alternative 3’ end formation, using HPV-16 as an example. (A) The late 3’ UTR of HPV-16 contains three hexanucleotide polyadenylation signals arranged in tandem array; however, only one signal is used \textit{in vitro} and \textit{in vivo} (285). (B) The HPV-16 exon containing the E4 and E5 ORFs can be viewed as a composite exon, in which the 5’ end can be formed by the 5’ splice site in E4 or by the early polyadenylation signal, pAE, resulting in a terminal exon (C) Due to the fast rate of transcription, it is possible that the E4 3’ splice site and the downstream L1 3’ splice site could compete for splicing with an upstream 5’ splice site. This would result in alternative terminal exons. However, this event is most likely to be strongly inhibited.

Regulation of polyadenylation

As described above and in Figure 9 using HPV-16 as an example, regulation of polyadenylation site selection provides a choice of 3’ ends on the mRNA. Most commonly, the choice is whether to include or exclude a 3’ terminal exon (an exon flanked by a 3’ splice site and a poly(A) site) in a multiexon transcript, ie. alternate 3’ terminal ends. Alternatively, a polyadenylation site can be downstream of a 5’ splice site, with which it competes to form the 5’ end of an exon, or multiple polyadenylation sites can be arranged one after another in the same 3’UTR. Ultimately it is the strength of the sequences stabilising the cooperative binding of the polyadenylation machin-
ery that regulates the efficiency of polyadenylation. Core poly(A) signals that deviate from consensus sequences are often weak and regulated; given the number of polypeptides and interactions involved in polyadenylation, there is much scope for regulation. Efficiency of polyadenylation, or recognition of a polyadenylation site, can depend on the concentration or activity of basal polyadenylation factors in the cell as well as the recruitment of additional factors to polyadenylation enhancer elements, or indeed interaction of repressor proteins. Recent *in vitro* experiments showed that elevated levels of CF1 could suppress poly(A) site cleavage (281). In addition, secondary RNA structure of the poly(A) region can be significant (282-284).

**Positive regulation of polyadenylation by additional sequence elements**

One example of positive polyadenylation regulation is that achieved by USEs, mentioned above. These are U-rich regions found upstream of the AAUAAA element and enhance polyadenylation by a variety of mechanisms. Many have been identified in viral poly(A) sites, such as in SV40 virus (286), adenovirus (287), and HIV-1 (288). HIV-1 3’ end processing provides a good opportunity to examine polyadenylation regulation and the role of USEs, because the long terminal repeats of retroviruses contain a promoter proximal poly(A) site that must be ignored and a promoter distal poly(A) site to be recognized. HIV-1 achieves this by a USE unique to the distal poly(A) site (288) that binds CPSF and stabilizes its binding to AAUAAA (289). The distal site is also favoured by the presence of a 3’ splice site upstream and the presence of the RNA stem-loop structure of the Tat binding site (TAR) between the USE and the distal poly(A) site (290). Repression of the 5’ poly(A) site is achieved by inhibitory proximity to the 5’ cap and promoter region (291) and by the binding of U1 snRNP to a downstream 5’ splice site (292). In addition, the poly(A) signals lie within a stem-loop structure that causes inefficient binding of polyadenylation factors, but the USE of the 3’ poly(A) site restored efficient binding (283).

USEs often bind peripheral modulators of polyadenylation. For example, the SV40 late poly(A) site USE binds the U1A protein that stabilizes the interaction of CPSF by contacting the CSPF-160 subunit directly (293,294). The adenovirus L3 poly(A) site has a U-rich USE that binds hnRNP C1/C2, leading to enhanced CstF-64 binding (287). A cellular example in which a USE enhances polyadenylation occurs in the poly(A) site of the complement C2 gene, where the USE binds PTB and enhances CstF binding (295). In the cellular gene lamin B2, an USE stabilises CPSF binding (296).

Another class of positive polyadenylation elements emerging is the G-rich element. The downstream enhancer of SV40 L poly(A), which contains both upstream and downstream elements to enhance polyadenylation, is G-rich and binds a hnRNP H protein (297,298). A more recent study revealed that G-rich sequences resembling the binding site of hnRNP H proteins were present downstream of 34% of mammalian poly(A) sites examined, and all
elements tested bound hnRNP H, suggesting that downstream G-rich elements may play a wide-spread role in polyadenylation regulation (299).

**Negative regulation of polyadenylation**

As mentioned above, a 5’ splice site downstream of a poly(A) site can inhibit polyadenylation, mediated by the binding of U1 snRNP as in the proximal poly(A) site of HIV-1 (292). In undifferentiated epithelial cells the late mRNAs of BPV-1 are also repressed by the binding of U1 snRNP to a 5’ splice site, but in this case to a cryptic 5’ splice site upstream of the late poly(A) site (300).

An example of differentiation-dependent poly(A) site usage and negative regulation of polyadenylation occurs in immunoglobulin genes in B cells (reviewed in (268)). Here a distal poly(A) site required for the membrane-bound form of the immunoglobulin heavy chain is used in preference to the proximal poly(A) site required for the secretory form. In immature B cells, the use of an internal 5’ splice site removes the proximal poly(A) site, but in terminally differentiated plasma cells this secretory poly(A) site predominates. Interestingly, differentiated plasma cells had a lower ratio of hnRNP F to hnRNP H than in immature B cells, and hnRNP F, stably expressed in plasma cells, decreased the use of the secretory poly(A) site. That hnRNP F has been proposed to block CstF binding downstream of the secretory poly(A) site argues that changing the levels of a regulating factor can negatively regulate polyadenylation, (301). The secretory poly(A) site may later be positively regulated by the presence of an upstream AU-rich region as well as two GU-rich downstream elements, the more distal of which requires a stem-loop structure to function (302,303).

**Role of the 3’ untranslated region (3’UTR)**

The 3’UTR of an mRNA is a hot-bed of regulation, and can harbour elements controlling many steps of gene expression, including polyadenylation (287) or the half-life and translation of the mRNA (164). Due to its multi-fold role in gene expression, defective 3’UTRs are also associated with a variety of diseases such as mantle cell lymphoma (304) and myotonic dystrophy (305).

*mRNA stability regulated by 3’UTR*

The half-life or stability of an mRNA can have a significant effect on the expression of specific genes, in particular cytokines, growth factors, oncogenes and cell-cycle regulators. Therefore mRNA stability can be critical for cell growth and differentiation, as well as in response to external stimuli (306). In addition to a direct effect on stability, the 3’UTR can also influence mRNA half-life indirectly by influencing polyadenylation and translation efficiency. Stability elements provide an important means of controlling
gene expression in the cell and have been identified in both cellular (307) and viral genes (308). The most common class of mRNA stability elements are the AU-rich elements or AREs, which are found in the 3'UTR of many short-lived mRNAs and can promote rapid deadenylation-dependent mRNA decay (309,310), and possibly also decapping (311). The common motif of AREs is AUUUA and is found in the 3'UTR of mRNAs such as c-fos (312), GM-CSF (307), and interleukin-3 (313). Many ARE binding proteins have been characterised but AREs are often bound by members of the Hu protein family (314). For example, the binding of HuR protein stabilises certain mRNAs (315,316) and destabilises others (317).

An example of a stability determining element other than ARE has been reported in alpha-globin mRNA. In this transcript a cytosine-rich element in the 3'UTR binds a stable complex that protects against deadenylation-dependent decay and endonucleolytic cleavage (318).

Control of polyadenylation by 3’UTR
In addition to the core polyadenylation signals in the 3’UTR discussed above, other elements can regulate specific polyadenylation events. Interestingly, during Xenopus development, polyadenylation in the cytoplasm triggers translation of certain short poly(A) tailed mRNAs harbouring cytoplasmic polyadenylation elements in the 3’UTR (319). Alternatively, the 3’UTR can harbour elements triggering deadenylation of similar early oocyte mRNAs as a means of reducing translational efficiency (319).

Translational control by 3’UTR
The 3’UTR is thought to regulate translation (320), primarily based on the widely accepted assumption that the interaction of PABP with eIF4G causes the circularisation of mRNA, though the latter has not been directly demonstrated in vivo (the reader is directed to a provocative review (321)). Elements in the 3’UTR, nevertheless, are believed to regulate translation initiation, ribosome binding, or post-initiation translation. Evidence for the role of PABP in translation was demonstrated when parts of PABP stimulated translation of reporter mRNAs when tethered to their 3’UTRs (322). The AREs have also been reported to regulate translation of neurofilament-M mRNA through the binding of the Hu B protein (323), and tumour necrosis factor-alpha mRNA by binding of the TIAR protein (324). In rabbit 15-loxygenase mRNA, an element in the 3’UTR binds hnRNP proteins, thus blocking the binding of the 60S ribosomal unit to the small subunit during translation initiation (325). The imperfect base pairing of micro-RNA to elements in the 3’UTR (326) can also trigger translational repression.
Viral regulation of RNA processing

In order to be recognised by the host translation machinery, viral pre-mRNAs must also be processed and exported. Many viruses have developed strategies to interfere with host RNA processing. For example, adenovirus inhibits the export of cellular mRNAs into the cytoplasm during the late stage of infection, whereas herpes simplex virus type 1 inhibits transcription of the majority of cellular mRNAs by RNAPII (327). Many viruses circumnavigate the restrictions on cellular mRNA export in order to transport unspliced and partially spliced mRNAs to the nucleus. The best-studied example is the role of the viral protein Rev, encoded by HIV-1, in the export of intron-containing HIV-1 mRNAs. Rev binds to the highly structured Rev-responsive element on late viral mRNAs and promotes their nuclear export (328).

Figure 10. Viruses can code for proteins that interfere with RNA processing. A simplified diagram of splicing events in the L1 transcription unit of adenovirus shows the inhibition of IIIA splicing early in infection. SR proteins bind the 3RE splicing repressor element (253). Late in infection, adenoviral protein E4orf4 forms a complex with PP2A and SR proteins (221), causing the dephosphorylation of a subset of SR proteins (331). Splicing inhibition of IIIA is relieved, and splicing further enhanced by the action of a virus-dependent intronic enhancer called 3VDE (330).

Of interest to this thesis is an example of a virally encoded protein that interferes with alternative splicing by inducing dephosphorylation of SR proteins. The adenovirus E4orf4 protein binds the cellular protein phosphatase 2A (PP2A) (329), and regulates adenovirus L1 alternative splicing (221) (Figure 10). The adenovirus L1 unit can give rise to two late transcripts, 52.55K and IIIA, using a common 5’ splice site and two alternative 3’ splice sites. The use of the distal 3’ splice site and therefore the splicing of IIIA is under control of a repressor element 3RE (253) and a virus-infection de-
ependent enhancer 3VDE (330). Early in infection, SR proteins bind to 3RE and repress IIIA splicing (253), by blocking U2 snRNP recruitment. As a result, 52,55K mRNA is produced exclusively. Late in infection, E4orf4 induces the dephosphorylation of SR proteins, causing a switch to IIIA splicing (221). E4orf4 was shown to interact with the SR proteins ASF/SF2 and SRp30c, and interaction with both PP2A and ASF/SF2 were necessary to relieve repression of IIIA splicing (331).

Regulation of HPV-16 late gene expression

Persistence of HPV-16 infection is itself the highest risk factor for the development of cervical cancer (8,83) and requires successful inhibition of late gene expression early in the life cycle to avoid the immune response. Therefore it is of great interest to understand the key events that regulate late gene expression.

HPV-16 gene expression is regulated by differentiation-dependent transcription from two major promoters. Early and late transcripts are generated by alternative splicing and differential polyadenylation at the early and late poly(A) sites, pAE and pAL, respectively. HPV transcripts can have multiple ORFs, and thus alternative splicing and translation regulate the expression of genes. As described above, late gene expression is tightly linked to the differentiation of the infected epithelia and is regulated by co-transcriptional processes. In particular, regulation of splicing efficiency and choice of polyadenylation signal or splice site may contribute to the inhibition of late mRNA production. Cis-acting elements involved in these processes are dependent on cellular host factors, which may also be regulated throughout differentiation. As well as differentiation-dependent transcription from the late specific promoter (23), the late genes are presumably regulated by a complex combination of positive and antagonistic elements and their trans-acting factors, contributing to stability, export and translation of the late transcripts (84). Regulation of late gene expression must be multi-factorial and takes place at many locations throughout the whole genome, as transcription of all HPV-16 mRNAs, including the late transcripts, initiates in the early region.

Role of the late coding region in HPV-16 late gene regulation

Initial studies demonstrated that HPV-16 L1 and L2 proteins could not be expressed using a strong constitutive promoter and an efficient poly(A) site in undifferentiated HeLa cells (332,333). Nevertheless, L1 protein expression, but not L2 expression (334), could be overcome by the inclusion of retroviral nuclear export elements (332). These data hinted at post-transcriptional regulation of the late genes, and inhibition of expression already in the nucleus. In addition, the L2 coding region reduced mRNA and
protein levels from a reporter gene (333) and suppression of L2 transcript translation was also reported (335). Regions that inhibited expression were identified in the coding regions of L1 and L2 (332,333), and were further mapped. Sequences inhibitory to L1 expression were found in the first 514 nts of the L1 coding region, and could be inactivated by mutagenesis, allowing production of L1 protein (336). In the L2 coding sequence, inhibitory elements were identified in the 5’ and middle region of the gene that had a negative effect on mRNA levels and utilization (334).

Sequence-specific mutagenesis of L1 and L2 coding regions (334,336) resulted most likely in the disruption of regulatory RNA elements embedded in the coding sequences, explaining the difficulty in producing L1 and L2 protein in initial studies (332,333). Further work has subsequently identified mechanisms for parts of the HPV-16 late coding region inhibitory sequences, such as the identification of an ESS at the 5’ end of L1 (337), and a regulatory polyadenylation element in the 5’ end of the L2 coding region (338), both described below.

The use of rare codons can decrease translation efficiency (339). In BPV-1, it was reported that the use of rare codons in the BPV-1 L1 gene and also tRNA levels prevented L1 production in undifferentiated cells (340). Upon replacement of rare codons with more commonly used codons, BPV-1 L1 was translated more efficiently in in vivo transfection and in vitro translation experiments (340). Leder et al reported that codon-optimised HPV-16 L1 and L2 proteins were translated with greater efficiency than the wildtype sequences, but did not examine mRNA levels (341). However, it does not seem that HPV-16 L1 protein production is repressed by rare codon usage, because a codon-optimised HPV-16 L1 sequence that avoided rare codons, was translated with the same efficiency as the wildtype L1 sequence in in vitro translation experiments (336). Moreover, as wildtype L1 mRNA levels were reduced in the nucleus and absent in the cytoplasm in transfection experiments, it was clear that a nuclear experience decreased mRNA levels, contributing to inefficient translation (336). The subsequent identification of an exonic splicing suppressor at the 5’ end of L1 (337) indicates that at least in vivo, inhibition of late mRNA transcripts by splicing mechanisms is very likely to contribute to apparent translation repression.

Interestingly, several studies demonstrate that the codon optimisation of early genes increases protein production, suggesting that coding region regulatory RNA elements may be widely used in HPV to regulate gene expression (342-344). In addition, inhibitory sequences in the L1 coding region appeared to be conserved over a wide range of HPV types (336).

Finally, due to the polycistronic nature of HPV transcripts, HPV-16 gene expression is likely to be translationally regulated. Upstream ORFs can block downstream translation. In low risk genital HPV types, the E1^E4 ORF blocks translation of the downstream L1 ORF (345), suggesting another level in which the expression of L1 capsid protein may be inhibited.
Role of alternative splicing in HPV late gene regulation

The animal papillomavirus BPV-1 has been a model for the role of alternative splicing in HPV gene expression. Here, alternative 3’ splice site selection and subsequently late mRNA production is controlled by at least five separate splicing elements. A suboptimal 3’ splice site is used in the processing of most early BPV-1 mRNAs, from which a switch to a downstream late-specific 3’ splice site allows the production of L1 mRNA, by a second splicing event at a downstream 5’ splice site (346). Two enhancers SE1 and SE2 and a suppressor element ESS control the use of the proximal 3’ splice site (246,347), whereas the distal 3’ splice site is controlled by an AC-rich enhancer SE4 and a suppressor element ESS2 (238). Cellular factors shown to bind to these elements include SR proteins, in particular ASF/SF2, as well as U2AF65 and PTB (246,348). In conclusion, in BPV-1, splicing elements and their cellular binding partners partly regulate the switch from early to late gene expression.

Despite a different arrangement of splicing signals, it seems that HPV-16 late gene expression is also regulated in part by splicing elements and cellular factors. In the late region of the genome, an hnRNP A1-dependent splicing silencer has been identified downstream of the L1 3’ splice site (337). The strong inhibitory effect of the ESS on L1 mRNA levels in undifferentiated cells suggested that it is presumably active early in the life cycle, to prevent premature L1 expression due to direct splicing into the L1 3’ splice site. In addition to the identified ESS located between nts 178 and 226 in the L1 gene, other regulatory splicing elements may be present in the first 514 nts of the L1 gene, because in earlier experiments the first 129 nts alone had a strong inhibitory effect (336). In support of this, according to recent unpublished data, the L1 coding region harbours multiple enhancers and silencer sequences throughout the length of the gene (Zhao et al, unpublished data). Regulation of the splicing of L1 mRNA is also likely to be controlled by several regulatory RNA elements upstream of the L1 gene and throughout the genome.

Alternative splicing also allows the abundant E1^E4, considered a late protein, to be translated using the E1 AUG. However, it is not only the late genes that are controlled by alternative splicing in HPV-16. In high-risk HPV types 16 and 18, alternative splicing may play a role in the regulation of the oncogenic E6 and E7 proteins. Bicistronic and polycistronic mRNAs encode E6 and E7, and alternative splicing gives rise to truncated forms of E6, known as E6*. HPV-18 E6*I rescued p53 from E6-mediated degradation by binding to full-length E6 and E6-AP (349,350) and, interestingly, expression of E6*I varied throughout the cell cycle, whereas E6 expression was constant (351), suggesting that E6*I could influence E6 function in concert with the cell cycle. Alternative splicing of the E6*I mRNA is thought to enhance E7 protein expression, because a premature stop codon
upstream of the E7 AUG would allow for translation termination and re-initiation (352,353). The 5’ splice site at position 880 is also required for E7 expression (354).

Role of polyadenylation in HPV late gene regulation
All HPVs have an early polyadenylation signal that is used throughout the virus life cycle. A partial switch to a late polyadenylation signal by mRNA transcripts initiating at the late promoter is thought to contribute to late gene expression. This is achieved by less efficient polyadenylation, and therefore readthrough, of the early polyadenylation signal upon differentiation. HPV-31, the closest relative to HPV-16 phylogenetically, has been a model for the study of polyadenylation in HPVs. In HPV-31, a significant level of readthrough at pAE was observed in a luciferase reporter assay, and upon differentiation this readthrough increased by 50% (355). pAE was identified as a weak poly(A) site containing three weak CstF-64 binding sites and, interestingly, a decrease of 40% in CstF-64 protein levels was seen in keratinocytes upon differentiation (355). A further study identified the most 5’ 800 nts of the L2 coding region as necessary for efficient polyadenylation at pAE, and that removal of this region caused a dramatic decrease in genome replication in a transient replication assay (356).

It seems, therefore, that a weak pAE is necessary for the HPV viral life cycle. In support of this, a strong poly(A) site in place of the HPV-31 weak pAE significantly reduced replication levels in a transient replication assay (356). Indeed, sequence alignment of HPV sequences revealed that GU-rich DSEs, determinants of poly(A) site strength, were generally not present at the reported optimal distance of 20-70 nts downstream of the hexanucleotide signal at pAE (338). Öberg et al reported that the 5’ region of HPV-16 L2 coding region was necessary for full efficiency of the early poly (A) site (338), and that mutation of triple G motifs in the L2 coding region correlated with reduced polyadenylation efficiency at pAE. Similarly to HPV-31 (355), CstF-64 was seen to bind to the HPV-16 L2 RNA sequence required for polyadenylation, but in the case of HPV-16, it was the cellular factor hnRNP H that interacted with the triple G motifs and binding correlated with increased efficiency of polyadenylation at pAE (338). The detection of hnRNP H in basal and suprabasal layers of cervical epithelium but not in the upper, more differentiated layers (338), suggests that downregulation of hnRNP H may play a role in readthrough of pAE and a switch to HPV-16 late gene expression.

The HPV-16 late 3’UTR contains three hexanucleotide polyadenylation processing signals in tandem, of which the second, referred to here as pAL, was shown to function in vitro and in vivo in HeLa cells (285) (Figure 9). An inhibitory sequence in the 3’UTR, described in the next section below, did not affect polyadenylation efficiency in vitro (285). However, subsequent identification of cellular factors binding to this 3’UTR sequence, now
termed the HPV-16 negative regulatory element (NRE), suggest that regulation of the HPV-16 pAL may be part of its function. Data describing the affinity purification of U1A and U1 snRNA with the 5’ part of the NRE containing four putative 5’ splice sites, and not when these sites were mutated (357), is reminiscent of pAL regulation in BPV-1. The binding of U1 snRNP to a 5’ splice site upstream of the BPV-1 pAL inhibited polyadenylation (300,358). However, as this was mediated by interaction between PAP and U1-70K (359), a U1 snRNP component not affinity-purified with the HPV-16 NRE (357), the mechanism by which the NRE could regulate HPV-16 late polyadenylation must differ from that of BPV-1. In HPV-31, in contrast to the use of heterogeneous cleavage sites in the early region in undifferentiated cells, HPV-31 late transcript 3’ ends were more homogeneous and located within a narrow region (355). The identification of a single, high-affinity CstF-64 binding site indicated that pAL in HPV-31 is a strong poly(A) site (355), suggesting that the closely related HPV-16 pAL is likely to be efficient.

The HPV-16 pAL has a GU-rich sequence resembling a reported consensus DSE sequence (360), allowing its preferential use over that of non-functional hexanucleotide signals in the late 3’UTR (285). This emphasises the importance of this DSE in supporting HPV-16 pAL processing. However, regulation of 3’ end processing at pAE, rather than at pAL, is likely to be more important for both prevention of premature late gene expression and later induction of downstream late gene expression. In support of this, pAE is a weak processing signal lacking a GU-rich DSE and requires part of the L2 coding sequence for full efficiency (338), indicating the importance of further regulation. In contrast, pAL is a strong processing signal with a consensus DSE (285). Evidence from studies using HPV-31 and BPV-1 early poly(A) sites indicates that tight regulation of papillomavirus pAEs directs polyadenylation to the early 3’UTR, despite inactivation or deletion of the hexanucleotide signal (356,361,362). Presumably several regulatory mechanisms to down-regulate pAE efficiency are required to permit transcription to continue to pAL.

### Role of the UTR in HPV gene regulation

The early 3’UTR is the only non-coding region in the HPV-16 genome apart from the long control region, and with its strategic position on the majority of mRNAs produced, is sure to have a regulatory role in the expression of early and late genes. The 3’UTR contains the early polyadenylation signal pAE, which is efficiently used early in infection to block expression of the downstream late genes. As described above, 3’UTR sequences often harbour elements regulating stability, export or translation (363). One report suggested that the expression of the HPV-16 E6 and E7 oncogenes are stabilised by integration events disrupting the early 3’UTR, and that the 3’UTR contains an instability element, which reduced the half-life of β-globin tran-
scripts greater than 7 fold (364). However, Zhao et al saw no effect on protein expression levels of a reporter gene when the HPV-16 early 3’UTR was placed downstream, whereas the late 3’UTR caused a substantial reduction in expression levels (365).

The 3’-end of the late transcripts extends into the late 3’ UTR, which contains a 79 nucleotide regulatory element, the NRE (366). This regulatory element is GU-rich, contains four putative 5’ splice sites and is thought to influence nuclear export and to destabilise the late transcripts (366,367). The NRE was previously shown to interact with several RNA-processing factors, including U2AF (368), CstF 64, HuR (367), a U1 snRNP-like splicing complex (357) and ASF/SF2 indirectly in a complex with U2AF65 (369). Interestingly, the closely related HPV-31 has a bipartite inhibitory element in the late 3’UTR that also binds U2AF65, CstF-64 and HuR (370). Upon differentiation of epithelial cells harbouring episomal HPV-16, U2AF65 levels were seen to decrease, whereas CstF-64 levels increased and HuR was relocalised to the cytoplasm (367). ASF/SF2 levels and its phosphorylation appeared to increase upon differentiation of epithelial cells harbouring HPV-16 episomally, in comparison with differentiated epithelial cells containing integrated HPV-16 (369). These cellular factors may play different roles in the fate of the late transcripts upon differentiation, aiding expression of the late genes.
Aims

- To investigate the role of the early region in the regulation of HPV-16 late gene expression.

- To investigate the regulation of splice sites in the early region that are required for late mRNA production.

- To identify cellular factors that activate HPV-16 late gene expression.
Present Investigation and Discussion

The common aim of Papers I and II presented in this report was to investigate regulatory RNA elements in the early region of HPV-16 that influence late gene expression. This involved identifying specific hotbeds of regulation in the early region that influence RNA processing events, such as alternative splicing and polyadenylation, during late gene expression. Paper I highlighted the E4 exon as a source of regulatory sequences influencing late mRNA splicing and, importantly, identified an ESE that may play a key role in the prevention of premature late gene expression, as well as a necessary role in early polyadenylation. The location of the early UTR also implied involvement in regulation of polyadenylation of the early mRNAs at pAE and, as a consequence, the inhibition of the late mRNAs. Paper II analysed the effect of the early 3’UTR on mRNA expression levels and reported that a U-rich region of the early 3’UTR had a stimulatory effect on early polyadenylation and bound polyadenylation factors and that its deletion induced late gene expression.

In Paper III we investigated if SR proteins were involved in late gene expression by employing the adenovirus protein E4orf4. We saw that E4orf4, which causes dephosphorylation of a subset of SR proteins (331), could induce HPV-16 late gene expression, and this was partially due to its interaction with SRp30c. A possible role for SRp30c in the regulation of alternative splicing in HPV-16 was implied, as SRp30c interfered with terminal exon definition of the early mRNAs and could induce L1 mRNA production by an exon skipping mechanism.

Methodology to study HPV-16 RNA processing in this thesis

Splicing events that have been detected in HPV-16 infected epithelium (371) have been faithfully reproduced in HPV-16 subgenomic expression plasmids in HeLa cells (337). An epithelial cell-line derived from squamous cell carcinoma of the cervix, HeLa cells represent undifferentiated epithelial cells. The inability to express the capsid proteins L1 and L2 in a genomic context after a strong constitutive promoter in HeLa cells (337) confirms that, indeed, like the lower layers of infected epithelia, this undifferentiated cellular environment does not support late gene expression. The studies in this thesis were based on an experimental set-up (337), whereby the expression of HPV-16 mRNAs could be analysed using transfection of HPV-16 mammalian expression plasmids into HeLa cells and Northern blotting of
extracted RNA (337). By modifying regions of the HPV-16 genome, one can identify regulatory sequences that are necessary for late gene expression by examining late mRNA levels. Indeed, the parent plasmid called pBEL which consists of the HPV-16 genome lacking the E6 and E7 genes, placed after the strong immediate-early cytomegalovirus (CMV) promoter, did not express detectable levels of late mRNA, whereas mutational inactivation of a splicing silencer in the 5’ L1 sequence of the same plasmid (pBELM) produced high levels of L1 mRNA (337).

**Paper I: A splicing enhancer in the E4 coding region of HPV-16 is required for early mRNA splicing and polyadenylation, as well as inhibition of premature late gene expression.**

A series of deletions were made in pBEL to investigate the importance of the early region in late gene expression. The effect of these deletions on late mRNA levels was examined by transfection, followed by RNA extraction and Northern blotting, using a late gene-specific probe. A region of 817 nucleotides was identified upstream of the E4 3’ splice site at 3358, between positions 3396 and 4212, that caused a dramatic increase in late gene expression when deleted. The lack of this region activated “induced” late gene expression, with splicing occurring directly from the 5’ splice site at 880 to the 3’ splice site in L1 at 5639, with or without the presence of the early polyadenylation signal. This deletion also caused a complete switch from the production of early to late mRNAs, detected by probing specifically for early or late mRNAs. The sequence necessary for early mRNA production was mapped to 255 nucleotides, between the E4 3’ splice site at position 3358 and the 5’ splice site at 3632 upstream of HPV-16 pAE. Optimisation of the suboptimal E4 3’ splice site and polypyrimidine tract at 3358 by PCR mutagenesis overcame the absence of this sequence and allowed production of early mRNAs. We concluded that an exonic splicing enhancer, supporting the E4 3’ splice site at 3358 and polyadenylation at pAE, was located upstream of position 3652. Further deletions in pBELM-based plasmids mapped the element to an AC-rich 65 nucleotide sequence about 100 nucleotides downstream of the E4 3’ splice site at 3358.

To confirm that a functional enhancer element was present downstream of the E4 3’ splice site, the sequence between 3462 and 3547 was inserted back into a plasmid with the original deletion of 255 nucleotides, between the E4 3’ splice site at 3358 and the 5’ splice site at 3632. The insertion restored the production of early mRNA to this plasmid and confirmed the presence of an enhancer supporting the E4 3’ splice site in the above inserted sequence.

A sequence adjacent to and downstream of the enhancer (positions 3547 to 3682) was identified that increased late gene expression upon deletion, and suggested the presence of a sequence inhibitory to splicing at the 5’ splice site. This sequence had a lesser effect on L1 mRNA levels than the
mutational inactivation of the splicing silencer in L1 (pBELM) but combined with the L1 mutant sequence, resulted in higher L1 mRNA levels than pBELM alone. The 5’ splice site at 3632 was shown to be suboptimal by replacing it with the complimentary U1snRNA sequence. Such optimization of the 5’ splice site decreased early mRNA levels as expected and, combined with the L1 mutant sequence, resulted in higher L1 mRNA levels than pBELM alone. These data demonstrated that use of the weak 5’ splice site at 3632 is presumably regulated, possibly by an inhibitory element contained in the above deleted sequence.

The exonic enhancer supported the use of the E4 3’ splice site and its inactivation resulted in loss of polyadenylation at HPV-16 pAE and subsequently early mRNA production. This established that the enhancer strengthened the coupling of the 3’ splice site to pAE, and indicated that the enhancer-supported 3’ splice site may prevent downstream splicing at the L1 3’ splice site by successful competition and subsequent polyadenylation of transcripts at pAE. The E4 ESE constitutes a key regulator of temporal HPV-16 gene expression and must therefore play an important role in the prevention of premature late gene expression early in the viral life cycle.

**Paper II: A 57 nucleotide upstream early polyadenylation element in human papillomavirus type 16 interacts with hFip1, CstF-64, hnRNP C1/C2 and PTB.**

Here the role of the early UTR was investigated in expression of both early and late genes. Initially, we used simplified versions of our CMV-driven pBEL experimental system. One plasmid named pBearly, lacked the E6, E7 and late genes, whereas a second plasmid, p97Bearly, lacked only the late genes. The major mRNA produced from p97Bearly contained a splicing event from the 5’ splice site at 226 to the 3’ splice site at 3358 in E4, while the major mRNA expressed from pBearly was spliced from the 5’ splice site at 880 to the 3’ splice site at 3358 in E4. While characterizing the mRNAs expressed from these plasmids using Northern blot and RT-PCR, a novel 3’ splice site was identified at position 742. This could potentially be employed to produce E1 and E4 mRNAs that would be more efficiently translated into E1 and E4 proteins. On these mRNAs the E1 and E4 open reading frames would be preceded only by the suboptimal E6 AUG start codon.

Choosing pBearly to study the role of the 3’UTR, a modest increase in early mRNA levels was seen upon deletion of the entire early 3’UTR or the 3’ U-rich half. To test this effect on the late mRNA, the same deletions were made in both pBEL and pBELM. Inactivation of the early polyadenylation signal pAE, or removal of the 57-nucleotide U-rich region followed by pAE, both resulted in induction of late gene expression in pBEL and increased spliced L1 levels in pBELM. Deletion of the UTR, or the U-rich region alone, induced late gene expression but did so to a lesser extent than inacti-
vation of pAE. We concluded that the 3’ U-rich half of the UTR acted as an upstream sequence element to enhance polyadenylation at pAE.

Northern blots revealed that early mRNAs were still produced despite inactivation of pAE. Cloning and sequencing of the 3’RACE product identified a faster migrating E1^E4 mRNA as being cleaved and polyadenylated upstream of the absent pAE at position 3820/3821. In addition, several other cryptic sites were activated upstream of pAE upon removal of pAE.

To investigate the interaction of the 3’UTR with cellular factors, UV cross-linking and competition experiments were performed. Four factors were demonstrated to interact specifically with the 57-nucleotide USE. These were CstF64, hnRNP C1/C2, PTB and hFip1. One can speculate a role for these proteins in the regulation of the early polyadenylation signal.

Discussion: Paper I and II

The 65 nucleotide exonic splicing enhancer, identified 100 nucleotides downstream of the E4 3’ splice site at 3358, is present on many early mRNAs as well as the late mRNAs. The E4 ESE, therefore, holds a strategic position from where it can regulate gene expression, perhaps in a differentiation-dependent manner throughout the life cycle. However, the E4 3’ splice site is required throughout the life cycle to express E1^E4 protein, as well as for late gene expression later in infection. In addition, the expression of E1^E4 must be coordinated with its different functions during the life cycle as well as with that of expression of upstream early genes E2 and E1. The enhancer allows the efficient use of the E4 3’ splice site, which may be subject to modulation throughout the life cycle, but presumably blocks late gene expression by coupling the E4 3’ splice site to the use of pAE. This coupling and subsequent block is supplemented by the masking of the 5’ splice site in E4, utilized in the splicing of L1 mRNA, which must remain unrecognized early in infection. The optimization of this 5’ splice site had a weak effect on late mRNA levels in comparison with the disruption of a splicing silencer in the L1 coding region. These data pointed towards the overall inefficiency of the E4 5’ splice site and the presence of regulatory sequences in the region, such as the inhibitory sequence identified here, upstream of the 5’ splice site.

The ESE is AC-rich and as such resembles a class of AC-rich enhancers (ACE), whose role has been previously described in splicing regulation (237,240,372,373). Identification of cellular factors interacting with the E4 AC-rich enhancer could shed more light on how regulation of HPV-16 gene expression is coordinated, as well as pinpoint proteins to target in antiviral therapy. It is possible that factors interacting with the E4 ESE function by recruiting U2AF65 to the weak polypyrimidine tract of the E4 3’ splice site (Figure 11), supporting the processing of the E4 terminal exon, as interaction between PAP and U2AF65, resulting in enhanced U2AF65 binding has been reported (374).
Paper II revealed a variety of interesting findings in regard to co-transcriptional gene regulation of HPV-16. A novel 3’ splice site was identified in the early region of HPV-16. This, in combination with the 5’ splice site at 226 and downstream 5’ splice site at 880, would produce an E1\(^\text{E4}\) mRNA that could be more efficiently translated than previously identified E1\(^\text{E4}\) mRNAs due to the presence of only one suboptimal start codon upstream. If splicing at 880 did not take place in the above situation, a more efficiently translated E1 mRNA would similarly be produced. This novel splice site, the use of which removes the E7 AUG, has subsequently been identified in HPV-16 infected cells (375).

It is critical for HPV-16 to have a strong pAE in order to avoid premature late gene expression early in the life cycle due to read-through. The early polyadenylation signal must be regulated by many different elements. This is suggested by the activation of cryptic polyadenylation sites upstream of pAE, upon removal or mutation of pAE. Similar observations have been made in BPV-1 (362) and HPV-31 (356). Despite the absence of pAE, strong regulatory sequences that are most likely downstream of pAE direct polyadenylation to the vicinity. Like the E4 ESE, the early 3’UTR is also present on all early mRNA transcripts. The early 3’UTR USE, coupling of the 3’ splice site at 3358 with pAE, through support of the E4 ESE, and downstream polyadenylation enhancing elements in the L2 coding region (338) all work in concert to ensure polyadenylation of early mRNAs and the subsequent block of late gene expression.

**Paper III: Adenovirus E4orf4 and SRp30c induce HPV-16 late gene expression.**

In order to investigate the role of the cellular splicing factors, SR proteins, in the regulation of HPV-16 gene expression, we employed a virally en-
coded protein known to induce dephosphorylation of SR proteins. Adenoviral E4orf4 protein, in a complex with the cellular phosphatase PP2A, dephosphorylates SR proteins, thereby contributing to activation of alternative splicing in the late phase of an adenovirus infection (221). Adenoviral E4orf4 was co-transfected with pBEL and pBELM into HeLa cells and induced HPV-16 L1 and L2 late gene expression. HPV-16 late mRNA induction required the interaction of E4orf4 with PP2A and ASF/SF2, but overexpression of polyomavirus small T-antigen that also binds PP2A did not induce HPV-16 late gene expression. E4orf4 induced both L1 and L2 mRNA, suggesting that E4orf4 activated late gene expression partially by inhibiting polyadenylation at pAE.

E4orf4 is known to interact with a subset of SR proteins, namely SRp30c and ASF/SF2 (331). To ascertain if these proteins were involved in the induction of HPV-16 late gene expression by E4orf4, SRp30c and ASF/SF2 were overexpressed in co-transfection experiments with pBEL and pBELM. SRp30c but not ASF/SF2 induced late gene expression in HPV-16. Furthermore, SRp30c induced primarily L1 and not L2 or early mRNAs, in contrast to E4orf4. RT-PCR experiments indicated that L1 mRNA production occurred by exon skipping of an internal exon in the L1 mRNA.

In addition, SRp30c caused a specific decrease in E1\(^\alpha\)E4 mRNA concurrently with an increase in late mRNA in pBEL. Therefore, it seems that due to disruption of early HPV-16 mRNA terminal exon definition between the E4 3’ splice site at 3358 and the pAE by SRp30c, L1 mRNA was produced, primarily by an exon skipping mechanism. SRp30c influenced the fate of the E4 exon, perhaps through the E4 3’ splice site. However, the ESE supporting the use of the E4 3’ splice site at 3358, reported in Paper I, was not required for SRp30c-induced late mRNA production. Interestingly, optimisation of the E4 5’ splice site at 3632 reduced interference with the terminal exon definition by SRp30c. We concluded that overexpression of SRp30c interferes with the definition of the terminal exon in HPV-16 early mRNAs. Either directly or by antagonising a cellular factor regulating HPV-16 gene expression, SRp30c may be involved in the downregulation of early mRNA and activation of L1 mRNA production. Another possible role for SRp30c is the modulation of the E4 3’ splice site to allow for the expression of the early genes E6, E7, E1 and E2, upstream of E4.
Concluding Remarks

The work in this thesis identifies an exonic splicing enhancer that supports the E4 3’ splice site at 3358, the most commonly used 3’ splice site in HPV-16. Efficient splicing at the E4 3’ splice site prevents downstream splicing to the late region, and therefore the E4 ESE constitutes a block in premature late gene expression. In addition, regulation of early polyadenylation by the early 3’ UTR was examined and an upstream polyadenylation element was identified that moderately enhances polyadenylation at pAE, a further block to late gene expression. A step towards determining the cellular factors involved in late gene expression, SRp30c was found to induce L1 mRNA production and interfere with terminal exon definition. RNA processing has a central role not only in eukaryotic gene expression, but also in HPV gene expression. The studies in this thesis support the concept that the regulation of RNA processing events, such as splicing and polyadenylation, is critical for HPV-16 gene expression.

The presence of the E4 ESE and also an hnRNP A1 dependent ESS (337) are exciting recent discoveries relating to the co-transcriptional regulation of HPV-16. In a matter of years, the intricate balance of early and late mRNA production may be fully elucidated in a similar fashion to BPV-1 (194), based on splicing elements and their cellular interacting factors. Already we know that the L1 coding region itself and the E4 exon are strategic sources of splicing regulatory elements, and that strict inhibition of the L1 3’ splice site by multiple RNA elements and efficient use of the E4 3’ splice site are important for the prevention of premature L1 mRNA production (Figure 12). It will be exciting to identify the cellular factors interacting with these strategic RNA elements, and analyse their differentiation status directly using immuno-histochemistry methods.

Using the raft cell culture technique, in which primary keratinocytes stably transfected with HPV genomes can differentiate and support the viral life cycle (376), it would be intriguing to disrupt the E4 ESE and examine the effect on HPV-16 gene expression patterns. If L1 mRNA production occurred in the lower layers of the raft, this would pave the way for antiviral therapy using the E4 ESE. Premature production of capsid proteins by disrupting the E4 ESE could stimulate a specific immune response in the HPV-16 infected individual and permit the infection to be cleared.

Therapeutic strategies against HPV-16 infection targeting the E4 ESE could have two approaches, the first being an anti-sense technique. In a bid
to develop therapy for diseases caused by aberrant alternative splicing, anti-sense strategies using modified oligonucleotides have recently been developed (reviewed in (377)). In the case of HPV-16 infection, one could employ 2'-O-methyl oligoribonucleotides complimentary to the E4 ESE, resulting in the skipping of the E4 exon and presumably use of the downstream L1 3’ splice site. Alternatively, using a chimeric compound consisting of an RS domain tethered to an antisense oligonucleotide complimentary to the L1 exon, one could also enhance splicing at the L1 3’ splice site.

Using a second approach, depending on the identity of the cellular factors directly or indirectly carrying out the functions of the E4 ESE, one could target cellular factors that regulate specific HPV-16 splicing events. Targeting an E4 ESE-binding factor could result in a switch to the use of the L1 3’ splice site. One can speculate that SR proteins may interact with the E4 ESE. Interestingly, drugs that disrupt the kinase activity of topoisomerase I, required for ESE-dependent splicing (378), affect the phosphorylation status of SR proteins and spliceosome assembly (379). In a recent study, a large-scale small molecule screen revealed that certain small compounds could inhibit enhancer-dependent splicing reactions through interaction with SR proteins (380). When the gaps in our knowledge of HPV-16 late gene splicing regulation are filled, such small molecules may provide a means to inhibit specific splicing reactions in infected mucosa. SRp30c could be a target for small compounds to enhance premature L1 mRNA splicing, when further studies clarify the mechanism of activation of the late genes by SRp30c.

Figure 12. HPV-16 late gene expression is controlled at the level of RNA processing. The presence of regulatory RNA elements at strategic positions along the genome, identified by work in this thesis and by others, inhibits production of the late mRNAs early in the HPV-16 life cycle.
Acknowledgements

This work was carried out at the Department of Medical Biochemistry and Microbiology, in the Biomedical Centre at Uppsala University, Sweden. Financial support was provided by the Swedish Research Council (Medicine), the Swedish Cancer Society and the Rektors Wallenbergsfund.

I would like to extend my gratitude to all those who helped to make this thesis possible.

To my supervisor Stefan Schwartz, a very motivated and driven scientist, for asking that first question to an Irish exchange student “What are you planning to do for the rest of your life?” You opened the door to the exciting world of research, which in a short time has enhanced my life enormously. You were on stand-by 24 hours a day for the latest developments. I am very grateful for your drive, which propelled us all forward in the lab, and I wish you great success in your future scientific endeavours.

To present and former members of the Schwartz group, thanks to each and every one of you so much for your guidance and support throughout my studies. A very special thank you and so much more goes to Xiaomin and Daniel, my closest colleagues, “lab family” and truly dependable and caring friends. We shared all we learnt, as well as our hopes and fears. Thank you both for help daily, and for your collaborations, commiserations and congratulations! Those crazy, late nights at the lab… what a journey we have had together, I couldn’t have chosen better companions myself (Thanks again, Stefan!).

“Ett stort tack” to Göran Akusjärvi, Catharina Svensson and Göran Magnusson and their groups for all their help, both practical and intellectual. The support, encouragement and guidance I received from you all, especially at the Friday meetings, has been very significant during my PhD studies.

To all members of the virology corridor, both new and old, for the friendly and productive atmosphere. Thank you all for your scientific help, as well as advice about everything under the sun, lab-laughers, spex-planning, fikas, garden parties, fun-days, training classes, tjej kväller.
TACK to everyone at IMBIM for the sharing of knowledge, smooth running of day-to-day services and problem solving, as well as my daily fix of friendly faces and chats. My sincere gratitude goes to the administration staff, present and former, who time after time solved all my paper work with extreme efficiency, and gave me kind words when I really needed them, especially towards the end.

To all fellow PhD friends in IMBIM, the BMC and also at Rudbeck, who have always been so encouraging and supportive.

Special thanks to my very dear friends and fellow PhD students, Sarah, Patrick, Máire, Ingrid and Marie, who in particular have really been there for me during my studies. Thank you sincerely for all your advice and support, sharing of your own experiences and all our discussions about our futures in research. It has helped me enormously, as has your unfailing belief in me.

To my opponent Professor Jørgen Kjems and members of the thesis committee for their much appreciated participation.

To the lecturers in the School of Biological Sciences at the Dublin Institute of Technology for providing a great entry into the scientific world and also for such kind reassurance during visits to Sweden.

To Aidan for excellent proofreading, as well as faithful and firm friendship.

This book is not big enough to thank EVERYONE who has contributed to the most fantastic time I have had while in Sweden. Thank you for enriching my life and for making me part of yours.

To old childhood, climbing or college friends, now spread out around the world, who have been cheering me on. Thank you for visits, phone-calls, and letters and for generally not forgetting about me. I will be at your doortop or on the end of your rope soon!

Finally, words are not enough to thank my wonderful family and dearest friends in the world – Suzanne, Colm, Aideen, Rebecca, Steffen and Shane. What would I do without your constant love and encouragement? You have been beside me at the helm of this, my toughest undertaking yet, and all other adventures. Go raibh mile maith agaibh!
References


is a novel mammalian splicing regulator acting through intron sequences adjacent to a 5′ splice site. *Mol Cell Biol*, **20**, 6287-6299.


16 L2 coding region induces production of high levels of L2 in human cells. J. Virol., 77, 11674-11684.


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 75

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.”)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-5972