Viral Control of SR Protein Activity

BY

CAMILLA ESTMER NILSSON
Viruses modulate biosynthetic machineries of the host cell for a rapid and efficient virus replication. One important way of modulating protein activity in eukaryotic cells is by reversible phosphorylation. In this thesis we have studied adenovirus and vaccinia virus, two DNA viruses with different replication strategies. Adenovirus replicates and assembles new virions in the nucleus, requiring the host cell transcription and splicing machineries, whereas vaccinia virus replicates in the cytoplasm, only requiring the cellular translation machinery for its replication.

Adenovirus uses alternative RNA splicing to produce its proteins. We have shown that adenovirus takes over the cellular splicing machinery by modulating the activity of the essential cellular SR family of splicing factors. Vaccinia virus, that does not use RNA splicing, was shown to completely inactivate SR proteins as splicing regulatory factors. SR proteins are highly phosphorylated, a modification which is important for their activity as regulators of cellular pre-mRNA splicing. We have found that reversible phosphorylation of SR proteins is one mechanism to regulate alternative RNA splicing. We have demonstrated that adenovirus and vaccinia virus induce SR protein dephosphorylation, which inhibit their activity as splicing repressor and splicing activator proteins. We further showed that the adenovirus E4-ORF4 protein, which binds to the cellular protein phosphatase 2A, induced dephosphorylation of a specific SR protein, ASF/SF2, and that this mechanism was important for regulation of adenovirus alternative RNA splicing.

Inhibition of cellular pre-mRNA splicing results in a block in nuclear- to cytoplasmic transport of cellular mRNAs, ensuring free access of viral mRNAs to the translation machinery. We propose that SR protein dephosphorylation may be a general viral mechanism by which mammalian viruses take control over host cell gene expression.

**Keywords:** adenovirus, ASF/SF2, dephosphorylation, E4-ORF4, L1, MLTU, PP2A, splicing, SR proteins, vaccinia virus

Camilla Estmer Nilsson, Department of Medical Biochemistry and Microbiology, Biomedical Center, Box 582, Uppsala University, SE-751 23 Uppsala, Sweden

© Camilla Estmer Nilsson 2001

ISSN 0282-7476
ISBN 01-554-5124

Printed in Sweden by Eklundshofs Grafiska AB, Uppsala 2001
Till min familj
This thesis is based on the following papers, which in the text will be referred to by their Roman numerals:


III. Camilla Estmer Nilsson*, David Huang* and Göran Akusjärvi. 2001. Functional inactivation of the essential SR family of splicing factors during a vaccinia virus infection. Manuscript. *these authors made equal contributions to the work

Reprints were made with permission from the publishers: Macmillan Magazines Ltd and European Molecular Biology Organization.
TABLE OF CONTENTS

ABBREVIATIONS ..................................................................................................................... 6

INTRODUCTION ......................................................................................................................... 7
  FROM A GENE TO A PROTEIN .................................................................................................. 8

SPLICING .................................................................................................................................. 11
  BIOCHEMISTRY OF SPLICING ................................................................................................. 12
  HOW TO FIND AND DEFINE THE SPlice SITES? .................................................................... 12
  THE SPLICOSOME .................................................................................................................. 13
    Small nuclear ribonucleoprotein particles (snRNPs) ............................................................... 13
  Spliceosome assembly ............................................................................................................ 14
  SR proteins ............................................................................................................................ 16
  SR-related proteins (SRps) ..................................................................................................... 19
  IMPORTANCE OF REVERSIBLE PHOSPHORYLATION IN SPLICING .................................... 19
  REGULATION OF ALTERNATIVE RNA SPLICING ................................................................. 20
    Positive regulation .............................................................................................................. 21
    Negative regulation ........................................................................................................... 22

ADENOVIRUS ........................................................................................................................... 26
  GENERAL BACKGROUND ......................................................................................................... 26
  THE VIRUS LIFE CYCLE ........................................................................................................... 26
  THE E4-ORF4 PROTEIN .......................................................................................................... 29
  CONTROL OF ADENOVIRUS GENE EXPRESSION BY ALTERNATIVE RNA SPLICING .......... 30

VACCinia VIRUS ....................................................................................................................... 34
  GENERAL BACKGROUND ......................................................................................................... 34
  THE VIRUS LIFE CYCLE ........................................................................................................... 34

VIRAL CONTROL OF HOST CELL GENE EXPRESSION ................................................................. 37
  THE CELL CYCLE AND SIGNAL TRANSDUCTION ................................................................ 37
  DNA REPLICATION ................................................................................................................ 38
  TRANSCRIPTION .................................................................................................................... 38
  RNA PROCESSING AND EXPORT OF mRNA ....................................................................... 39
  TRANSLATION ....................................................................................................................... 40
  PROTEIN TRANSPORT ......................................................................................................... 40

PRESENT INVESTIGATION .......................................................................................................... 41
  REGULATION OF ADENOVIRUS ALTERNATIVE RNA SPLICING — INACTIVATION OF SR PROTEINS AS
  SPLICING REGULATORY PROTEINS BY DEPHOSPHORYLATION (PAPER I). ................................. 41
  The adenovirus E4-ORF4 protein induces PP2A dependent SR protein dephosphorylation by
  interacting with a subset of phosphorylated SR proteins (PAPER I and II). ............................. 42
  E4-ORF4 IS A SPlicing ENHANCER PROTEIN — THE IIIA REPRESSOR ELEMENT (3RE) IS NECESSARY
  AND SUFFICIENT FOR E4-ORF4 ACTIVATED SPLICING (PAPER I AND II) ................................. 43
  Strong E4-ORF4 interaction with ASF/SF2 and recruitment of an active PP2A enzyme are
  required for E4-ORF4 mediated activation of IIIa pre-mRNA splicing (PAPER II) ............... 44
  Vaccinia virus inactivates SR proteins by dephosphorylation (PAPER III) ......................... 45

CONCLUSIONS AND DISCUSSION ............................................................................................. 47
  ADENOVIRUS AND VACCinia VIRUS INDUCE SR PROTEIN DEPHOSPHORYLATION — A GENERAL VIRAL
  MECHANISM TO CONTROL HOST CELL GENE EXPRESSION? .................................................. 47
  E4-ORF4 — CELLULAR AND VIRAL FUNCTIONAL HOMOLOGES? ............................................. 48
  E4-ORF4 AND APOPTOSIS ........................................................................................................ 50

ACKNOWLEDGEMENTS ............................................................................................................. 51

REFERENCES ............................................................................................................................ 52
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad-NE</td>
<td>nuclear extract prepared from adenovirus-infected HeLa cells</td>
</tr>
<tr>
<td>ASF/SF2</td>
<td>alternative splicing factor/ splicing factor 2, an SR protein</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>ESE</td>
<td>exonic splicing enhancer</td>
</tr>
<tr>
<td>ESS</td>
<td>exonic splicing silencer</td>
</tr>
<tr>
<td>HeLa-NE</td>
<td>nuclear extract prepared from HeLa cells</td>
</tr>
<tr>
<td>hnRNPs</td>
<td>heterogenous nuclear ribonucleoprotein particles</td>
</tr>
<tr>
<td>ISE</td>
<td>intronic splicing enhancer</td>
</tr>
<tr>
<td>ISS</td>
<td>intronic splicing silencer</td>
</tr>
<tr>
<td>kDa</td>
<td>kilo Dalton, a protein molecular weight unit</td>
</tr>
<tr>
<td>L1</td>
<td>late region one in the MLTU of adenovirus</td>
</tr>
<tr>
<td>MLTU</td>
<td>major late transcription unit in adenovirus</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>OA</td>
<td>okadaic acid</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PP2A</td>
<td>protein phosphatase 2A</td>
</tr>
<tr>
<td>p(Y)tract</td>
<td>polypyrimidine-rich region close to a 3´ splice site</td>
</tr>
<tr>
<td>RS-domain</td>
<td>arginine- and serine-rich domain in the SR family of splicing factors</td>
</tr>
<tr>
<td>snRNPs</td>
<td>small nuclear ribonucleoprotein particles</td>
</tr>
<tr>
<td>SR-Ad</td>
<td>SR proteins purified from adenovirus-infected HeLa cells</td>
</tr>
<tr>
<td>SR-HeLa</td>
<td>SR proteins purified from HeLa cells</td>
</tr>
<tr>
<td>SRrps</td>
<td>SR-related proteins</td>
</tr>
<tr>
<td>SRs</td>
<td>the classical SR proteins</td>
</tr>
<tr>
<td>SR-VV</td>
<td>SR proteins purified from vaccinia virus-infected HeLa cells</td>
</tr>
<tr>
<td>S100</td>
<td>cytoplasmic extract prepared from HeLa cells</td>
</tr>
<tr>
<td>3RE</td>
<td>IIIa repressor element</td>
</tr>
<tr>
<td>3VDE</td>
<td>IIIa viral dependent enhancer</td>
</tr>
</tbody>
</table>
The human body consists of many different cell types, each with unique features and functions. All the information about these cells is stored within the genome, in the DNA, which is made of four different building blocks. The DNA is used as a template for the transcription of a messenger RNA, mRNA, that has similar building blocks as the DNA. In turn, the mRNA is translated into proteins, made of 20 different amino acids, each symbolized by a three “letter” code in the gene. All cells in an organism contain an identical set of genetic material, with about 30,000-40,000 genes encoded in humans (132). However, different regions of the DNA are used as templates to produce the mRNAs that are translated into the proteins needed in a particular cell type. The cell regulates gene expression at multiple levels in order to produce the cell type specific proteins: DNA replication, transcription of RNA, RNA processing and transport, protein synthesis and stability of RNA and protein. This thesis will highlight one of these mechanisms: RNA splicing. Almost all genes in humans are interrupted by so called introns that are removed at the RNA level before the mature mRNA is translated.

We have learned a lot about the cellular mechanisms controlling gene expression by studying viruses. Viruses are parasites, equipped with an RNA or DNA genome, and a protecting shell that consists of lipids and/or proteins. Viruses take over the biosynthetic machineries in the host cell and modify them for their own purpose: to support an efficient replication. Viruses can encode for their own DNA/RNA replication systems, transcription machineries, RNA processing machineries and proteins for transport of viral RNA and proteins. However, for RNA splicing and translation viruses need the cellular spliceosome and ribosome respectively. Both the spliceosome and the ribosome are complex molecular machines, which viruses modulate to suit their own purpose.

In this thesis we have studied adenovirus and vaccinia virus, two DNA viruses with different replication strategies. Adenovirus replicates and assembles new virions in the nucleus and depends on the cellular transcription and splicing machineries, while vaccinia virus replicates in the cytoplasm, only requiring the cellular translation machinery. Adenovirus uses alternative RNA splicing extensively to produce its proteins and we show that adenovirus controls RNA splicing by modulating the activity of the essential SR family of cellular splicing factors (Paper I and II). Vaccinia virus on the other hand, that does not use RNA splicing, was found to completely inactivate SR proteins as splicing factors (Paper III).

One important, perhaps the most important, way of modulating protein activity in all eukaryotic cells is by reversible phosphorylation. Viruses target cellular protein kinases and protein phosphatases that regulate the phosphorylated status of proteins. The SR family of splicing factors is highly phosphorylated, a modification which is crucial for their activity. We have found that reversible phosphorylation of SR proteins is one mechanism to regulate alternative RNA splicing (Paper I). We show that both adenovirus and vaccinia virus induce SR protein dephosphorylation, which inhibit their activity as splicing enhancer and repressor proteins (Paper I and Paper III). We have also found that the adenovirus E4-ORF4 protein, that binds to the cellular protein phosphatase 2A (PP2A), induces dephosphorylation of a specific SR protein and regulates adenovirus alternative RNA splicing (Paper II).
One important feature of the eukaryotic cell is that it has two compartments: the nucleus and the cytoplasm. The DNA is stored in the nucleus where transcription takes place. The precursor RNA (pre-mRNA) is processed in the nucleus before it is transported to the cytoplasm where the translational machinery is located. This RNA processing involves 5’ capping, splicing and 3’ end polyadenylation of the mRNA. This chapter will only briefly summarize the way from transcription of the RNA to protein translation (figure 1) keeping the focus of this thesis on pre-mRNA splicing.

The DNA is transcribed by the multisubunit enzyme RNA polymerase (reviewed in 138). In the eukaryotic cell there are three different RNA polymerases (I, II and III) which synthesize different classes of cellular RNAs. RNA polymerase II transcribes mRNAs and small nuclear RNAs. All genes have a promoter area from which transcription of the gene starts. There are at least two features common to most promoters: the core promoter, which include the transcription start site and the TATA box, and sequences bound by regulatory transcription factors. RNA polymerase II does not by itself recognize the transcription start site, but needs the help of general transcription factors in order to assemble at the promoter.

Transcription is regulated by two major mechanisms: (i) at the level of chromatin structure and (ii) at the level of transcription factor recruitment of the RNA polymerase to the promoter (138). There is a complex interplay between a myriad of activator and repressor proteins, some which bind directly to DNA. These factors function by directly or indirectly recruiting, or disrupting, the basal transcription machinery. The activity of many of these transcription factors are regulated by reversible phosphorylation.

It has been shown that the switch from initiation to elongation of transcription involves phosphorylation of the C-terminal domain (CTD) of RNA polymerase II
which results in the association of many proteins with the CTD (138). A number of studies indicate that the RNA processing reactions, capping, splicing and polyadenylation happens cotranscriptionally, in intimate association with the CTD tail of RNA polymerase II (figure 2) (reviewed in 100).

The 5´ end of the mRNA is capped by addition of an inverted methylated guanosine triphosphate (m7Gppp) (reviewed in 70). The capping enzymes have been shown to bind to the phosphorylated CTD and to carry out the capping reaction before the transcript has reached the size of 30 nucleotides. Two proteins bind to the cap structure in the nucleus and form the cap-binding complex (CBC), which is required for efficient mRNA export from the nucleus. This complex has also been shown to play a role in splicing, by facilitating association of U1 snRNP with the proximal 5´ splice site in a pre-mRNA (141) and for initiation of protein synthesis in the cytoplasm (see below) (100).

A direct connection between the polymerase and splicing of the pre-mRNA is still somewhat controversial although recent evidence suggest a possible link. In vitro splicing reactions is stimulated by the phosphorylated form of RNA polymerase II or its CTD (101) and different types of promoters appear to dispose a specific splicing pattern to transcripts initiated on the promotor (43, 44). Some SR-related proteins have been found to associate with the CTD, called SCAF$s$ (118, 197, 268) and the SR protein ASF/SF2 has been found to bind to the transcriptional cofactor p52 (76). In addition, the recruitment of splicing factors from their storage sites in the nucleus (speckels) to the sites of active transcription has been found to require an intact CTD of RNA polymerase II (168).

**Figure 2.** Transcription by RNA polymerase II connects capping, splicing and polyadenylation (modified from (100)). (SCAF$s$s$ = SR-related CTD associated factors, CE= capping enzyme, CBC= cap binding complex, SR$s$s$= SR protein family, CPSF= cleavage- and polyadenylation-stimulatory factor, CstF= cleavage stimulatory factor, PAP= poly(A) polymerase).

The poly(A) tail at the 3´ end of the mRNA has been proposed to be important not only for termination of transcription but also for mRNA stability, transport of processed mRNA from the nucleus to the cytoplasm and promoting mRNA translation. Polyadenylation is a two-step process with an endonucleolytic cleavage followed by poly(A) synthesis (reviewed in 279). The cleavage/polyadenylation factors have also been clearly associated with the CTD during transcriptional elongation (100). Deletion of the CTD from RNA polymerase II inhibits
polyadenylation of reporter transcripts in transfected cells. Furthermore, the CTD stimulates 3′ cleavage in vitro. Polyadenylation and splicing appears to be connected in the recognition of the 3′ terminal exon (100). As is the case in the natural context, the presence of a 3′ splice site activates polyadenylation, while a strong 5′ splice site positioned downstream or immediately upstream of the poly(A) site inhibits polyadenylation. Thus, U1 snRNP binding inhibits polyadenylation either via inhibition of poly(A) polymerase (89) or of the 3′ cleavage (247). When the 3′ splice site upstream of the poly(A) site is weak, as is the case in many alternatively processed mRNAs, additional interactions mediated by splicing factors bound to exonic enhancer sequences may also contribute to polyadenylation.

After splicing the mRNAs are transported to the cytoplasm as ribonucleoprotein complexes (RNPs). It appears that each major class of RNAs (tRNA, rRNA, UsnRNA and mRNA) uses distinct export pathways (reviewed in 48). Nucleo-cytoplasmic transport occurs through the nuclear pore complex (NPC) which provide docking sites for transport complexes. Nuclear localisation signals (NLS) and nuclear export signals (NES) direct protein import or export through interactions with transport receptors. During transcription the pre-mRNA is bound not only by splicing factors but also heterogenous nuclear ribonucleoproteins (hnRNPs) (reviewed in 274). These are thought to induce appropriate processing and folding of the pre-mRNAs. The hnRNPs remain bound after splicing is completed. During mRNA export to the cytoplasm certain hnRNPs are removed at the NPC and stay in the nucleus (hnRNP C), while others (hnRNP A1 and K), follow the mRNA to the cytoplasm and shuttles back to the nucleus. It has been found that splicing of the pre-mRNA also leaves a protein complex bound 20-24 nucleotides upstream of the splice sites in the mRNAs. This complex consists of DEK, SRm160, Y14 and RNPS1 (115, 136, 137). Importantly, this complex also bind mRNA export factors such as Aly/REF and TAP (283) and thereby provide a link between RNA splicing and transport of the mRNA.

After reaching the cytoplasm the mRNA is translated into a protein by the ribosome, a complex multisubunit molecular machine consisting of RNAs and proteins. Eukaryotic initiation factors (eIFs) mediate the recruitment of the mRNA to the ribosome (reviewed in 79). The activity of eIFs is regulated in multiple ways: by transcription, reversible phosphorylation, binding to inhibitory proteins, and proteolytic cleavage. The initiation of translation is the rate limiting step in translation and involves binding of eIF4E to the 5′ cap, recruitment of the RNA helicase eIF4A to the 5′ region and bridging of the mRNA with the ribosome by eIF4G that also induces circularization of the mRNA via interaction with poly(A)-binding protein (PAB). The 40S subunit assembles with the mRNA as a ternary complex together with eIF2 and the initiator tRNA. Once loaded on the mRNA the 40S ribosomal subunit scans the mRNA for the first start codon, where it assembles with the 60S ribosomal subunit and translation can start. Multiple initiations can occur on a single mRNA creating a beads-on-a-string-like structure that can be visualized by electron microscopy.

The mRNAs in the cytoplasm are sooner or later subjected to degradation (reviewed in 87). The poly(A) tail is gradually deadenylated, which in turn causes decapping at the 5′ end followed by a 5′- to 3′-exonucleolytic degradation. RNA stability can also be regulated through AU-rich elements present in the 3′ untranslated region of several short-lived mRNAs. Premature stop codons (PTCs) have also been shown to cause degradation of mRNAs.
Most eukaryotic genes contain intronic sequences that are removed by splicing in order to assemble the coding regions (exons) into a functional mRNA. The complexity of splicing increases enormously from yeast to mammals. Only 250 out of the 6000 genes in the yeast *S. cerevisae* contain introns, while more than 99% of the genes in humans are thought to contain introns. In yeast, most genes that have introns only have one small intron near the beginning of the transcript. In contrast, most of the mammalian genes consists of intronic sequences. The genetic complexity in mammals is also increased by the usage of alternative splicing of exons (reviewed in 82, 147). One extreme example is the *Dscam* gene in *D. melanogaster*, which produces a pre-mRNA that can be alternatively spliced into over 38,000 different mRNAs.

Introns are removed before the mRNA is transported to the cytoplasm. Unspliced mRNAs are usually not transported to the cytoplasm. Although the basis for the nuclear retention of unspliced transcripts is unclear, splicing seems to deposit a specific complex of proteins on the mRNA that targets the mRNA to transport (136, 137, 150). Splicing is not always compulsory for efficient cytoplasmic accumulation of mRNAs. Thus, there are examples of genes that are intronless and still are efficiently transported to the cytoplasm, for example mRNAs encoding for the histone proteins (116), α-interferons (183) and c-jun (95) and several viral mRNAs.

Pre-mRNA splicing takes place in a large macromolecular complex, the spliceosome, composed of five small nuclear ribonucleoprotein particles (snRNPs) and 50-100 polypeptides many of which are not associated with snRNPs (non-snRNP proteins) (reviewed in 20). The RNA components of snRNPs align the pre-mRNA splice sites and probably mediate splicing catalysis. The bunch of proteins required for splicing mediate the recognition and pairing of the splice sites and the structural reorganisations during spliceosome assembly and catalysis. The spliceosome is restricted to the nucleus in eukaryotic cells and not found in organelles (mitochondria or chloroplasts) or prokaryotes.

From an evolutionary point of view, splicing have many advantages for the eukaryotic cell. It ensure cells a certain grade of combinatorial freedom in using their genetic information. Exons can be rearranged in the genome during the course of evolution to create new genes. This might be a faster way of adapting to environmental changes than from accumulation of mutations in an existing gene. In contrast, bacterial genes are not interrupted by introns. Instead, bacteria create an evolutionary sufficient rate of mutation by replicating very fast. Bacteria also contain mobile introns (called group I and group II introns) which can "self-splice" by having an intrinsic catalytic activity (reviewed in 78). It is thought that the eukaryotic introns might be remnants of group II introns. Group II introns resemble spliceosomal introns and can be found in mitochondria and chloroplasts.
Biochemistry of splicing
An intron is spliced out by two transesterification reactions without the need for an external supply of energy (figure 3). After the first cleavage at the 5´ splice site, the 2´ OH of the branchpoint adenosine attacks the intronic 5´ phosphate to form a 2´-5´ phosphodiester bond, which results in a lariat structure and a free exon. In the second reaction the 3´ OH of the first exon attacks the 5´ phosphate of the second exon yielding the spliced exon product and the intron lariat (reviewed in 20).

![Figure 3](image)

**Figure 3.** The catalytic steps in splicing (from (20)).

How to find and define the splice sites?
Splicing is a simple process from a chemical point of view. However, each splicing reaction takes place with a high degree of accuracy to ensure that coding information is not lost or altered. The consensus sequence elements at the splice sites directs, in part, the assembly of the spliceosome (figure 4). In mammals the 5´ splice site signal is AG/GURAGU (where / shows the exon-intron boundary). Three distinct sequence elements are found at the 3´ splice site: (i) the branchpoint sequence (YNYURAC), located 18-40 nucleotides from the 3´ splice site, (ii) a polypyrimidine tract (p(Y)tract) and (iii) the actual 3´ splice site (YAG/N). In yeast the sequences are more strongly conserved (figure 4) (20).

![Figure 4](image)

**Figure 4.** Splice site signals in mammals and yeast (from (20)). **Bold letters** represent highly conserved nucleotides (>90%). * shows the adenosine residue forming phosphodiester bond with the 5´ intronic phosphate. Y= pyrimidines (U or C), R= purines (G or A), N= any nucleotide.

A major question to be answered is how the correct splice sites are found. Splice site signals in mammals are short and often degenerate. The specificity in splicing vertebrate genes is not determined only by the splice site signals but also by exonic or intronic splicing enhancer or splicing repressor elements. The initial
recognition of splice sites involves multiple relatively weak RNA-RNA and RNA-protein interactions that commit the pre-mRNA to splicing.

In mammals, the 5′ and 3′ splice sites at internal exons are thought to be recognized by a process called "exon definition" (4). A general finding is that the mammalian exons are usually very short (50-300 bases). In the exon definition model, the initial recognition of splice sites is thought to occur by interactions between splicing factors binding to the 3′ and 5′ splice sites across the exon (see figure 7). This model is supported by the observation that mutations in a 5′ splice site can affect the splicing efficiency of the upstream 3′ splice site. The assembly of the spliceosome may also occur by the interaction between the splice sites over long distances. Strong candidate proteins for these types of interactions, both in the exon definition model and over the introns, are the SR proteins.

However, terminal exons appear to be defined by alternative mechanisms. The cap nucleotide has been shown to promote recognition of the first 5′ splice site in a pre-mRNA and the polyadenylation signal has been shown to promote the use of the last 3′ splice site (140). As discussed in the previous chapter, these processing events all seem to occur co-transcriptionally. Thus, in addition to delivering factors to the splice sites, the RNA polymerase II can possibly play a more direct role in the processing reactions. Perhaps splicing factors simultaneously binding to the CTD of RNA polymerase II and the splice site are brought along with the polymerase to the next splice site for efficient pairing (figure 2).

It has been estimated that around 15% of human genetic diseases are caused by mutations that destroy functional splice sites or generate new ones (reviewed in 199). The majority of these mutations are within the conserved sequences at the 3′ and 5′ splice sites. However, there are also examples of mutations in exons that destroy cis-acting elements required for correct splice site choice. One example is the survival of motor neuron gene (SMN). Loss of SMN protein expression correlates with the development of spinal muscular atrophy (SMA). A mutation in exon 7 appears to destroy an exonic splicing enhancer and cause exon 7 skipping, which results in a nonfunctional SMN protein (148, 149, 172).

The spliceosome

Small nuclear ribonucleoprotein particles (snRNPs)

SnRNPs are defined as tight complexes of several proteins and a short RNA molecule (usually 60-300 nucleotides) (reviewed in 267). They exist both in the nucleus and in the cytoplasm of eukaryotic cells. Those that are in the nucleus are divided into two families: the snRNPs in the nucleoplasm, which are required for mRNA formation and the snoRNPs in the nucleolus that are involved in ribosomal RNA formation. Mammalian cells contain about 200 distinct kinds of snRNPs. The snRNPs involved in splicing are very abundant, exceeding 10⁶ copies per cell. These are the (griddle-rich) U1, U2, U4, U5 and U6 snRNPs. The secondary structure and the sequence of the UsnRNPs are conserved from yeast to human. Each UsnRNP particle consists of a UsnRNA molecule complexed with a set of seven Sm or Sm-like proteins and several particle-specific proteins. Via interactions between their RNA and protein components with the pre-mRNA, the UsnRNPs mediate the recognition and pairing of the 5′ and 3′ splice sites during spliceosome assembly. There are also a group of less abundant snRNPs (U11, U12, U4atac and U6atac) that together with U5 snRNP are subunits of the so-called minor spliceosome (U12-type) (20). The major class (U2-
spliceosome is universal in eukaryotes and splices introns containing the canonical GT-AG sequence at the splice sites, whereas the minor class (U12-type) spliceosome, splices introns with AT-AC (and GT-AG) termini. The proportion of nuclear introns that are spliced by the U12 spliceosome seem to be very small, about one in a thousand. This thesis will focus on pre-mRNA splicing with the major U2-type spliceosome.

All UsnRNAs except for U6 are transcribed by RNA polymerase II and are exported to the cytoplasm, where they assemble in a step-wise manner with the Sm proteins, to form the UsnRNP Sm core structure (reviewed in 257). The snRNPs are also modified in several ways, for example, their cap is methylated. The methylated cap and the Sm core are required for the snRNPs to be transported back to the nucleus. In the nucleus snRNPs are assembled with their individually specific snRNP proteins. U6 snRNA, on the other hand, is transcribed by RNA polymerase III and is assembled in the nucleus. The La protein helps U6 snRNA to assemble with the Sm-like proteins. Then, U6 and U4 snRNPs pairs to from the U4/U6 snRNP.

Spliceosome assembly
Spliceosome assembly occurs in a stepwise fashion (figure 5) (reviewed in 230). It is initiated by binding of U1 snRNP to the 5´ splice site, SF1/mBBP (splicing factor 1/mammalian branch point binding protein) to the branchpoint sequence and U2 auxiliary factor (U2AF) to the polypyrimidine tract (p(Y)tract) and the 3´ AG to form the E (early) complex. Subsequently, U2 snRNP binds to the branchpoint to form complex A, followed by the association of the U4/U6-U5 tri-snRNP to form complex B. Next the spliceosome rearranges to form the catalytically active C complexes. SR proteins are required for spliceosome assembly at several steps. They help to commit a pre-mRNA to splicing by recruiting splicing factors to the pre-mRNA. SR proteins and their function will be further discussed in the next chapter.

The recognition of the 5´ splice site by U1 snRNP, which is ATP independent, involves base pairing between the 5´ splice site and the well conserved 10 nucleotides at the 5´ end of U1 snRNA. U1 snRNP association with the 5´ splice site is critical for splicing in vitro. However, increasing the concentration of SR proteins can overcome the requirement for U1 snRNP (45, 239). In this case the 5´ splice site is recognized by U6 snRNA and other spliceosomal components before the first step of splicing (46, 238).

U2 snRNP has been found to be present already in the E complex, but stable binding to the branch point requires ATP (49). Many proteins are devoted to help U2 snRNP to recognize the branch point at the 3´ splice site. U2AF is required for targeting of U2 snRNP to the branch point (213). Human U2AF is a heterodimer consisting of a 65 kDa and a 35 kDa subunit (see section "SR-related proteins") (272, 273, 276). The U2AF^{65kDa} binds to the p(Y)tract (74, 272) and the U2AF^{35kDa} contacts the 3´ splice site (160, 260, 285). The binding affinity of U2AF to the p(Y)tract depends on the length and the pyrimidine content of the p(Y)tract, which is important in regulation of alternative splicing. U2AF^{35kDa} interacts with SR proteins which stabilize binding of U2AF to the p(Y)tract (see figure 7). For further discussion about U2AF, see "Regulation of alternative RNA splicing".
SF1/mBBP cooperates with U2AF to recognize the branch point (1, 5). SF1/mBBP is displaced when U2 snRNP binds to the branch point. The U2 snRNP associated factors, SF3a and SF3b, are also required for stable association of U2 snRNP to the branch point (3, 14, 15, 128). Another protein binding to the branch point is p14, that interacts with SF3b and U2 snRNP (203, 258).

When the U4/U6-U5 tri-snRNP then joins the complex dynamic rearrangements of RNA-RNA interactions occurs which forms the spliceosome. U5 snRNP makes contact with exon and intron sequences at the 5´ and 3´ splice sites and basepairs with the 5´ end of U1 snRNA (267). These interactions bring the two splice sites close together. U5 snRNP interaction with the 5´ exon is required for the

---

**Figure 5.** Spliceosome assembly.
second catalytic step of splicing (187, 191). U6 snRNP dissociates from U4 snRNP and forms basepairing with U2 snRNP and replaces U1 snRNP as the factor interacting with the 5′ end of the intron (reviewed in 181). It is suggested that competition between U1 snRNP and U6 snRNP for interaction with the 5′ splice site is involved in regulating the transition from an inactive to an active spliceosome (130, 231). Studies in yeast have shown that the U2-U6 helix I is required for splicing catalysis, arguing that it contributes to the catalytic core of the spliceosome (20). During the first catalytic step of splicing the lariat intermediate and the free 5′ exon are generated. Before the second catalytic step, the spliceosome undergoes additional conformational changes, creating new RNA-RNA interactions. The U2-U6 snRNP interaction changes and makes contact with the 5′ splice site in the lariat intermediate and the free 5′ exon is joined together with the 3′ exon (20).

All steps in spliceosome assembly require ATP hydrolysis, except formation of the E-complex. Many proteins with sequence similarity to DNA helicases have been found to associate with the spliceosome. It is believed that the spliceosome consumes ATP to rearrange RNA-RNA interactions during assembly and the catalytic steps of splicing (230).

**SR proteins**

SR proteins are a family of highly conserved nuclear factors that play multiple important roles both in constitutive and regulated splicing of pre-mRNAs in metazoan organisms. Some SR protein functions are redundant, but other functions are unique and specific to a certain family member (reviewed in 66, 83, 153, 234).

SR proteins were independently discovered by two groups taking different approaches. ASF/SF2 (Alternative Splicing Factor/ Splicing factor 2) was purified from HeLa cell nuclear extracts as a factor required to complement splicing deficient cytoplasmic HeLa cell extracts (S100) (77, 127) and to induce splice site switching (75, 126). In another approach, monoclonal antibodies recognizing structures of sites of transcription in oocyte nuclei, were used to identify SR proteins, named after their high serine and arginine content (68, 69, 209, 269). All the “classical” SR proteins (figure 6), range in size of 20-75 kDa and can by themselves complement a splicing deficient S100 extract in vitro (66). SR proteins have been found in plants and all metazoan species examined but not in all eukaryotes. There are two SR proteins found in *S. pombe* (86, 151) while none has been discovered in *S. cerevisiae* (so far).

Some, but not all, SR proteins are essential. SRp55 is essential for development in *D. melanogaster* (B52, (205)), and ASF/SF2 is essential for viability in chicken DT40 cells (254) and in *C.elegans* (rsp-3, (146)). The basis for this is not known. Either the nonessential SR proteins do not participate in the splicing of essential genes, or other SR proteins can functionally substitute for the missing SR protein.
SR proteins have a modular structure. They have one or two N-terminal RNA binding domains (RBDs) and a C-terminal domain rich in arginine- and serine dipeptide repeats (the RS domain). RNA sequences recognized by SR proteins have been studied by SELEX (28, 142, 143, 216, 233, 235, 236) (systematic evolution of ligands by exponential enrichment), which selects high-affinity binding sites from pools of random RNA sequences (58, 245). In general, selected sequences have short consensus binding sites (6-10 nucleotides) without evidence of a secondary structure. SR proteins with two RBDs appear to require both for specific high affinity RNA binding (235). In several cases, sequences identified as binding sites for one SR protein are also recognized by other SR proteins (143, 235, 236). This may in part explain their redundancy in function. Importantly, many reports have shown that high-affinity binding sites are sufficient to function as exonic splicing enhancers (ESEs, see section Regulation of alternative RNA splicing). Examples of such high-affinity binding sites are shown in Table 1.

The RS domain in SR proteins function as a protein-protein interaction domain. SR proteins have been found to interact with eachother, the U2AF^{35kDa} and the U1 snRNP specific protein U1-70K (for example, ASF/SF2 and SC35 in (259)). In contrast, SRp54 interacts with U2AF^{65kDa} but not U1-70K or U2AF^{35kDa} (277). Different protein interactions may have distinct RS domain requirements. For example, ASF/SF2 binds to both U1-70K and RSF1. While the RS domain is sufficient for the interaction with RSF1, both the RS domain and the RBDs are necessary for ASF/SF2 interaction with U1-70K (131, 262). In many respects, the RS domain in different SR proteins appear to have redundant functions. First, when artificially bound to the pre-mRNA, the RS domains of several SR proteins are sufficient to activate enhancer-dependent splicing (84), an activity that probably is mediated by protein-protein interactions. Second, RS domains have been found to be exchangeable between different SR proteins both in vitro and in vivo. For example, the RS domain of TRA2 (a D. melanogaster SR-related protein) can substitute for the ASF/SF2 RS domain in chicken cells lacking endogenous ASF/SF2 (255), and several RS domains can replace the TRA2 RS domain in D. melanogaster (50). Also, exchanging the RS domains of ASF/SF2 and SC35 has no negative effects for the
respective protein to complement splicing deficient S100 extracts in vitro (32). However, the RS domains seem to have distinct properties in directing subcellular localisation of the SR proteins. SR proteins migrate from speckels (subnuclear domains that may function as storage sites for certain splicing factors) to sites of active transcription and ASF/SF2 and SRp20 have been shown to shuttle in and out of the nucleus. The RS domain of ASF/SF2 can convert the nonshuttling protein SRp40 to a protein that shuttles (22). In addition, the RS domain of SRp20 can target an unrelated fusion protein to speckels (21).

<table>
<thead>
<tr>
<th>SR protein</th>
<th>RNA binding site</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRp55</td>
<td>USCGKM</td>
</tr>
<tr>
<td>SRp40</td>
<td>ACDGS</td>
</tr>
<tr>
<td>9G8</td>
<td>GGACGACGA</td>
</tr>
<tr>
<td>ASF/SF2</td>
<td>SRSASGA</td>
</tr>
<tr>
<td>SC35</td>
<td>GRYYCSYR</td>
</tr>
<tr>
<td>SRp20</td>
<td>CCUGGUCC</td>
</tr>
</tbody>
</table>

**Table 1.** SR protein RNA binding sites (reviewed in 83, 234). Examples of ESEs binding respective SR protein identified by functional SELEX (randomized sequences that enhance in vitro splicing). The SRp55 binding sequence was found by conventional SELEX (randomized sequences that bind the SR protein). Sequences for ASF/SF2, SRp40 and SRp55 are from (143), 9G8 and SRp20 from (216) and SC35 from (142). (R=purine, Y=pyrimidine, $S=G/C$, $D=A/G/U$, $M=A/C$, $K=U/G$).

Many papers have demonstrated that SR proteins functions as splicing enhancer proteins by binding to ESEs. These elements have been identified in a number of constitutively and alternatively spliced metazoan exons. SR proteins are proposed to enhance splicing from ESEs by a number of mechanisms (figure 7):

- by recruiting U2AF to the p(Y)tract via an interaction with the RS domain of U2AF$^{35kDa}$ ("U2AF recruitment model")
- by binding to the U1-70K protein and possibly also the 5´ splice site thereby recruiting U1 snRNP to the 5´ splice site (59, 107, 123, 271, 287).
- by recruiting the U4/U6-U5 tri-snRNP to the spliceosome, potentially via interaction with the RS domains of U5-27K and U5-100K (SR-related proteins) (63, 206, 238, 242).
- by bridging splicing factors bound to the 3´ and 5´ splice sites respectively over the exon ("exon definition model"; (4)).

SR proteins are also proposed to promote an interaction between U1 and U2 snRNPs, on the pre-mRNA over the intron, by simultaneously interacting with U1-70K and U2AF$^{35kDa}$ (67, 259). In addition, SR proteins are required for trans-splicing, where the 5´ and 3´ splice sites are present on separate RNA molecules (17, 36).
SR proteins can also repress splicing both by binding to exonic and intronic sequence elements (for further discussion see chapters "Adenovirus" and "Negative regulation of alternative RNA splicing"). It seems that the repressor and enhancer functions of ASF/SF2 are encoded by distinct domains in the protein (Dauksaite and Akusjärvi, submitted). The RNA binding domain 2 (RBD2) of ASF/SF2 are both necessary and sufficient for the splicing repressor function of ASF/SF2. For further discussion of SR protein function in alternative RNA splicing and the U2AF recruitment model, see Regulation of alternative RNA splicing.

![Figure 7. SR proteins different functions. SR proteins recruit splicing factors to the splice sites and are thought to bridge both over the intron and the exon.](image)

**SR-related proteins (SRrps)**

There are also a number of additional RS-domain containing proteins, distinct from the classical SR proteins, that are required for pre-mRNA splicing. These are collectively referred to as SRrps (reviewed in 9). Examples include U2AF^{35kDa} (276), U2AF^{65kDa} (272, 273), U1-70K (229), Sip1 (278), U5-27K (63) and U5-100K (242). Most of the SRrps do not contain an RNA binding domain and eventhough many of the SRrps are essential splicing factors, they can not, as the classical SR proteins, complement splicing deficient S100 extracts.

U2AF^{65kDa} contains three RNA binding domains responsible for binding to the p(Y)tract (273) and an amino-terminal RS domain, which is believed to facilitate binding of U2 snRNP to the branch point (248). The RS domain of U2AF^{35kDa} stabilizes U2AF^{65kDa} binding to the p(Y)tract and is proposed to mediate interactions with SR proteins at the 3´ splice site (259, 286). Although U2AF^{35kDa} does not contain a canonical RNA binding domain, it contacts the 3´ splice site AG (160, 260, 285). The U1-70K protein has one RNA binding domain which tethers the protein to U1 snRNA and an RS domain required for U1 snRNP interaction with SR proteins (25, 123, 185). Another SRrp is the splicing coactivator SRm160/300 that have one RS domain (8, 10). SRm160/300 has been shown to be required for an ESE to promote splicing of a pre-mRNA, by interacting with U2 snRNP and SR proteins over the intron (57).

**Importance of reversible phosphorylation in splicing**

Several studies have shown that reversible protein phosphorylation contributes to spliceosome dynamics at almost every step of the splicing reaction. Generally, it seems that spliceosome assembly requires protein phosphorylation while the catalytic reactions require dephosphorylation (166). Protein phosphatase 1 (PP1) is required for the first step of catalysis, while protein phosphatase 2A (PP2A) is required for the
second step (161, 162, 240). Although, neither of these phosphatases have been shown to be stably associated with the spliceosome.

SR proteins appear to be a key target of reversible phosphorylation in splicing. Both phosphorylation and dephosphorylation of SR proteins are required for splicing in vitro (26, 262, 263), apparently at different steps. Phosphorylated SR proteins are required for the assembly of the spliceosome (162) while dephosphorylation of SR proteins are required for splicing catalysis (263). It has been shown that SR proteins are highly phosphorylated in vivo (208), mainly at serine residues in the RS domain (40). Phosphorylation of the RS domain influences the RNA binding capacity of SR proteins. Phosphorylated SR proteins bind more specifically to the RNA (233, 262). RS domain phosphorylation also regulates protein-protein interactions. For example, it enhances binding to U1-70K which is probably important for 5’ splice site recognition and changes interactions among SR proteins (255, 262). It is speculated that the dephosphorylation of SR proteins helps to weaken the interaction of U1 snRNP to the 5’ splice site in complex B. However, dephosphorylation of U1-70K is also required for splicing catalysis (241). Reversible phosphorylation has also an influence on the subcellular localisation of SR proteins. Phosphorylation of SR proteins causes redistribution of SR proteins from speckles to sites of active transcription in the nucleus (22, 167).

Many types of protein kinases have been identified that can phosphorylate SR proteins: SRPK1 (88), SRPK2 (255b), the CLK-Sty family (40), DNA topoisomerase I (207) and cdc2 (194). Evidence have been presented that CLK-Sty regulates alternative RNA splicing, both of its own pre-mRNA (55) and in two pre-mRNAs involved in the control of sex development in *D. melanogaster* (54).

Phosphorylation of hnRNP proteins may also be important for splicing regulation. Induced phosphorylation of hnRNP A1 results in change of the alternative splicing of the hnRNP A1 pre-mRNA (249).

There are exceptions to the “general rule” of phosphorylation in assembly and dephosphorylation in splicing catalysis. SAP155 is a U2 snRNP protein, that has been found to become phosphorylated during the catalytic steps of splicing (255c). Also, a protein phosphatase 2Cγ (PP2Cγ) has been found to be associated with the spliceosome and required early in assembly prior to complex A formation. PP2Cγ remains associated with the spliceosome and may be involved also in later steps of splicing catalysis (182).

As described earlier, there are many rearrangements of RNA-RNA and RNA-protein interactions occuring during spliceosome assembly and catalysis. A number of spliceosome-associated components have amino acid sequences similar to ATP-dependent helicases and GTPases (91).

**Regulation of alternative RNA splicing**

Alternative splicing is frequent in metazoans from *C. Elegans* and *D. Melanogaster* to humans (reviewed in 82, 94, 147, 225). A low estimate suggests that around 59% of human genes have at least two splice variants (132). Although many sequences within mammalian transcripts match the consensus splice sites, most of them are not used. Instead it seems that many positive and negative cis-acting sequence elements binding various trans-acting factors regulate splice site usage. Positive elements promote splicing at correct splice sites and appropriate times. Negative elements may block splicing at pseudo splice sites and partially or completely repress splicing at inefficient or regulated splice sites.
Alternative splicing is an important mechanism increasing protein diversity by allowing multiple, sometimes functionally distinct, protein isoforms to be encoded by a single gene. Alternative splicing can be tissue specific, developmental specific or induced under stress conditions or in pathological states. Some alternative splicing events appear to be constitutive, with mRNA variants coexisting at a constant ratio in the same cell, whereas others are regulated.

Alternative splicing patterns result from the usage of alternative 5′ splice sites, alternative 3′ splice sites, optional exons, mutually exclusive exons, or retained introns (figure 8). Alternative splicing decisions involve competition among potential splice sites. Thus, splicing patterns can be controlled by any mechanism that changes the relative rates of splice site recognition. Therefore, splicing patterns that look similar can involve fundamentally different pathways. Splice sites with weak signals can be regulated by positive trans-acting factors binding enhancers. At the same time, weak splice sites can efficiently compete with a stronger site when the latter is repressed by negative regulation.

Highly specific alternative splicing factors have not been identified in vertebrate cells, and few have been identified in invertebrates. Instead, it seems that the specificity comes from variations in relative concentrations or activity of competing and cooperating factors together with the strength of binding sites for regulatory and constitutive splicing factors. SR proteins and hnRNP proteins are some of the factors that have been shown to regulate alternative splicing. The following sections will describe basic mechanisms of enhancement and repression of splicing.

Figure 8. Different scenarios of alternative RNA splicing.

Positive regulation
Splicing enhancer elements have been identified in many regulated and constitutively spliced pre-mRNAs (reviewed in 7, 82, 94, 147, 225). Splicing enhancers are position dependent. Changing their location, can change their dependence on particular trans-acting factors or determine whether they activate 5′ or 3′ splice site usage. ESEs are often purine-rich and it is generally believed that SR proteins activate splice sites by binding to ESEs and recruit splicing factors to the nearby 3′ and/or 5′ splice sites. The primary sequences of ESEs are degenerate and not conserved. This may be important because it means that the protein coding capacity of the ESE is not strictly determined by the ESE activity.

In the "U2AF recruitment model" SR proteins binding to an ESE recruit the U2AF$^{65kDa}$ to the p(Y)tract by interacting with the RS domain of U2AF$^{35kDa}$ (figure 7).
This seems to be most important for weak 3’ splice sites which bind U2AF poorly. In these cases the RS domain of the SR proteins are required for activation. However, there are pre-mRNAs where ESE function do not correlate with increased binding of U2AF to the p(Y)tract. For example, introns with strong p(Y)tracts that bind efficiently to U2AF$^{65kDa}$, do not require the U2AF$^{35kDa}$ for activation (284). This suggests that SR proteins binding to an ESE may act through other mechanisms. Perhaps by competing with repressing factors, such as hnRNP proteins (see “Negative regulation”). This activity does not require the RS domain of the SR proteins (284). Collectively, available data suggests that SR proteins both may have RS domain dependent and independent functions in activating splicing through ESEs (reviewed in 94).

Much of our understanding of ESE function has been derived from studies on alternative splicing of key factors involved in sex determination in D. melanogaster (reviewed in 147). A well characterized example is the ESE present in the D. melanogaster double-sex gene (dsx). In males, dsx exon 4 is skipped, while in females, exon 4 is included. Exon 4 inclusion requires a complex ESE in exon 4 that consists of six 13-nucleotide repeats, called the dsx repeat element (dsxRE). Each repeat is recognized by the SR protein 9G8, the D. melanogaster SR protein RBP1 and the splicing regulators TRA and TRA2 (152). All four proteins bind cooperatively to the splicing enhancer and activates the upstream weak dsx 3´ splice site.

ESEs can also activate the usage of a downstream 5´ splice site. For example, the D. melanogaster fruitless gene (fru), has an ESE consisting of three nearly perfect copies of the dsxRE repeat unit, immediately upstream of the female specific 5´ splice site (214). TRA, TRA2 and RBP1 are also important in female specific fru 5´ splice selection splicing, although the mechanistic details are not yet known (97).

There are also intronic splicing enhancers (ISEs), mostly pyrimidine-rich and located close to the 5´ splice site. In c-src the inclusion of the N1 exon is enhanced in neuronal cells by a complex of hnRNP proteins binding to an ISE downstream of the 5´ splice site (figure 9) (37, 164, 165, 170).

Negative regulation

Splicing can also be negatively regulated, by intronic and exonic splicing silencer elements (ISSs and ESSs). A summary of possible events is shown below (for some of the examples see figure 9).

Ways to inhibit splicing:
- SR protein binding to ISSs inhibiting the usage of 3´ and 5´ splice sites. Blocks recruitment of general splicing factors to the splice sites (Ad-L1 3RE (114), CFTR exon 9 (195) and BPV-1? (281)).
- SR protein binding an ESS and inhibiting the 3´ splice site usage. Sequesters SR proteins from the ESEs (BPV-1 (282)).
- ISSs and ESSs binding hnRNPs inhibiting splicing probably by changing the conformation of the RNA. Are thought to “hide” the exon by binding simultaneously to bordering introns, bridging between the introns (hnRNPA1, PTB).
- Factors binding to p(Y)tract, inhibiting U2AF recruitment (SXL, PTB).
- Decoy splice sites recruiting splicing factors to inappropriate splice sites (IgM (113), Caspase-2 (41), NRS in RSV (159)).
- Antagonism between activators and repressors (SR proteins and hnRNPA1).
One repressor protein that has been extensively studied is the *D. melanogaster* sex-lethal protein (SXL) (147). SXL is an hnRNP like protein produced exclusively in female flies and induces female specific alternative splicing of at least three pre-mRNAs in *D. melanogaster*. SXL binds specifically to poly(U) in the p(Y)tract, thereby outcompeting U2AF binding. For example, SXL autoregulates its own expression by promoting skipping of the male-specific exon 3 in its own pre-mRNA (*sxl*).

The human polypyrimidine tract binding protein (PTB or hnRNP I) is a general splicing repressor that, like SXL, binds U-rich elements often close to the 3′ splice sites in introns (reviewed in 252). It is ubiquitously expressed in mammalian tissues and contain four RNA recognition motifs (RRMs). PTB has been implicated as a regulatory protein involved in the control of tissue-specific alternative splicing of several genes, for example: α-tropomyosin (81), FGF-R2 (27), c-src (figure 9) (38), GABA, receptor (275) and α-actinin (228). The mechanism for PTB repression of splicing is not exactly clear. A common feature is that PTB binding sites are clustered near the branch point of the alternatively spliced exon and suggests that PTB may block the binding of U2AF to the p(Y)tract, similar to the SXL protein. However, in many cases multiple PTB binding sites located also at other places (in the downstream intron) are essential for the regulation. The repression is therefore probably not caused by direct competition with general splicing factors. Instead, PTB may act as an antagonist of the exon definition model.

Another hnRNP protein that inhibits splicing by binding to ESSs or ISSs is the hnRNP A1 protein. For example, hnRNP A1 promotes skipping of exon 7B in its own pre-mRNA probably by binding to the introns bordering exon 7B (6).

As mentioned above, SR proteins can also repress splicing, both by binding to an ISS or an ESS. In CFTR (cystic fibrosis transmembrane regulator) SR proteins inhibit the 5′ splice site of exon 9 by binding to an ISS (195). In adenovirus, SR proteins bind to the intronic 3RE and inhibit IIIa 3′ splice site usage (see Conclusions and discussion, figure 18) (114). Moving the 3RE to the downstream exon resulted in IIIa 3′ splice site activation. This show that SR protein function is dependent on where on the pre-mRNA they bind (114). However, in bovine papilloma virus 1 (BPV-1) SR proteins binding to the exon inhibit upstream 3′ splice site activation. The late mRNAs in BPV-1 has two ESEs (ESE1 and ESE2) and an ESS located close to ESE1 (figure 9) (281, 282). All three elements bind SR proteins but only the ESEs enhance the use of the weak upstream 3′ splice site. The ESS inhibits the use of the 3′ splice site probably by sequestering SR proteins from binding to the other elements or by interfering with the normal bridging activities of the SR proteins at the ESEs. ESE2 is closer to the downstream alternative 3′ splice site and potentially ESE2 works as the adenovirus 3RE.

In some cases, splicing have been shown to be inhibited by decoy splice sites, that fool the splicing machinery by recruiting splicing factors to nonproductive splice sites. One example of this is in the Ig-M pre-mRNA. The M2 exon contains an ESS that binds U2 snRNP in an ATP dependent manner, to a decoy branchpoint (figure 9). This is belived to form a nonproductive inhibitory complex that hides the authentic 3′ splice site (113).

The relative concentration of antagonizing factors is probably important for splice site selection. Variations in the level of SR protein and hnRNP A/B protein expression have been reported in different cell types (62, 92, 112). Under limiting U1 snRNP concentrations, U1 snRNP binds preferentially to the strongest splice site. Thus, a weak 5′ splice site will not be selected, even if it is closer to the 3′ splice site.
Higher levels of ASF/SF2 promote full occupancy of U1 snRNP to all 5' splice sites and under these conditions the 5' splice site closest to the 3' splice site is selected (225). HnRNP A1 can antagonize this activity of ASF/SF2, and thus promote distal splice site usage. HnRNP A1 and ASF/SF2 seem to compete for binding to the pre-mRNA (60). ASF/SF2 enhances U1 snRNP binding, while hnRNPA1 interferes with U1snRNP binding such that the 5' occupancy is lowered (94). The molar ratio of hnRNP A1 and ASF/SF2 varies over a range of at least 100-fold in different tissues, well over what is expected to be required to induce a complete shift between competing 5' splice sites in vitro (92).

SR proteins also antagonize each other in many pre-mRNAs. For example, the β-tropomyosin gene in chicken encodes for two mutually exclusive exons. Exon 6A is specific to fibroblast and smooth muscle cells, while exon 6B is specific to skeletal muscle cells. A pyrimidine-rich element (S4) in the intron downstream of exon 6A is essential for recognition of the exon 6A 5' splice site. ASF/SF2 binds to S4 and stimulates inclusion of exon 6A. SC35 antagonizes the stimulatory effect of ASF/SF2. The ratio between SC35 and ASF/SF2 is at least 2-fold higher in skeletal muscle compared to HeLa cells, resulting in exon 6A skipping in HeLa cells in vitro (72).

Figure 9. Different types of regulation of alternative splicing. See text for details. (Pictures are modified from references (281) (BPV-1), (113) (IgM) and (252) (c-src)).
It has been suggested that many, possibly all, exons are under global repressive influence mediated by many intronic sequences (61). Thus, splice site usage is decided both by splice site strength and the repressor activity within the pre-mRNA. There are examples of alternatively spliced exons that are repressed in most tissues, but "derepressed" in specific celltypes. One example is the c-src N1 exon (figure 9) (reviewed in 252). The N1 exon is both positively regulated in neurons and negatively regulated in non-neuronal cells. In non-neuronal cells the inclusion of the N1 exon is repressed by PTB binding to the upstream intron (30, 31). In neuronal tissues the inclusion of exon N1 is derepressed by displacement of PTB from binding to the intron by an ATP dependent mechanism (37, 38). As mentioned before, there is also a downstream control sequence (DCS), that enhances N1 exon inclusion. The DCS appears to regulate N1 exon splicing by binding hnRNP proteins (37, 164, 165, 169, 170). A neuronal specific form of PTB (nPTB) also enhances N1 exon inclusion by binding to the DCS in neuronal cells and competing with PTB binding (157).
ADENOVIRUS

Here I will give a brief summary of the adenovirus lifecycle and explain functions of the E4-ORF4 protein and regulation of adenovirus alternative RNA splicing in more detail. Many of the mechanisms of adenoviral proteins interfering with the host cell are described more in the chapter "Viral control of host cell gene expression". This chapter is based on two main references, when no specific references are given in the text (64, 221).

General background
Adenovirus was found as a viral agent in tonsils and adenoidal tissue from military recruits with febrile illness, hence the name "adeno"-virus. Some adenovirus serotypes can cause tumors in rodents, but so far, adenovirus has not been associated with tumors in humans. Adenoviruses are widespread in nature infecting mostly mammals and birds. New serotypes are frequently described. Adenoviruses are transmitted by direct contact and most commonly cause respiratory tract infections, diarrhoea, pneumonia, croup and bronchitis. In recent years, there has been a lot of interest in developing adenovirus as a vector to express foreign genes for therapeutic purposes in vaccination, gene therapy and in cancer therapy.

Adenovirus is a nonenveloped virus with a linear double stranded DNA genome of 30-38 kb in length. It encodes for 30-40 proteins of which around 15 are components of the virion. The virion has a regular icosahedral structure of 70-90 nm in diameter and consists of 240 hexons and 12 pentons, where each penton base projects a fiber. The genome is condensed with three histone-like proteins (V, VII and μ) and the terminal protein (TP) covalently bound to the DNA ends.

The virus life cycle
By convention the replication cycle is divided into an early and a late phase which are separated by the onset of viral DNA replication. The infectious cycle is completed approximately 30 hours post infection, resulting in the production of about $10^5$–$10^6$ virus particles per cell under optimal growth conditions.

Adenovirus enters its host cell by interaction of the fiber and penton base proteins with a range of cellular receptors, including the primary receptor CAR (coxsackievirus-adenovirus receptor), MHC class I, and members of the integrin family. After binding to the cell, the virus is phagocytosed and inside the endosome some virion proteins are degraded. During the disassembly process the vacuole membrane is ruptured and the DNA is transported and injected into the nucleus. It is believed that the viral core proteins (V and VII) are removed from the DNA before or at the time the viral DNA enters the nucleus.
In the nucleus the viral DNA is converted to a virus-DNA-cell histone-complex that is used as a template for viral gene expression.

The early region 1A (E1A) is the first unit to be transcribed during infection. E1A encodes for two virus-specific transcription factors required for activation of early viral gene expression, but has also the capacity to regulate a variety of cellular genes. Adenovirus genes do not contain a common promotor sequence element indicating that E1A regulates the viral promotors by different mechanisms. E1A can both activate and repress transcription by interacting with transcriptional coactivators/repressor proteins and the TATA-box binding protein (TBP). The E1A proteins induce the S-phase in the host cell by interacting with the cellular transcription factors p300/CBP and pRb. p300/CBP activates genes involved in cell differentiation and E1A binding to pRb releases transcription factor E2F, which becomes available for activation of genes involved in DNA-synthesis. Furthermore, E2F is important for activation of the viral E2 promotor. E1A also causes accumulation of the tumor suppressor protein p53 and therefore E1A by itself causes p53-dependent apoptosis. However, adenovirus also encodes for multiple proteins that block apoptosis, thereby facilitating viral growth.

The E1B region encodes for two major proteins, both of which inhibit apoptosis. The E1B-19K is a homolog of the cellular anti-apoptotic Bcl-2 protein and inhibits apoptosis by heterodimerizing with pro-apoptotic factors (Bax, Bak), thereby preventing induction of apoptosis. The E1B-55K protein is a multifunctional phosphoprotein essential for efficient virus replication. E1B-55K, in complex with E4-ORF6, regulates both p53 activity and nucleocytoplasmic transport of viral and...
cellular mRNA. E1B-55K binds to p53 and inhibits p53 transactivation of target genes, probably by localizing p53 to the cytoplasm and, together with E4-ORF6, promoting degradation of p53.

The E3 region encodes a cluster of proteins that counteract the immune responses in the host. The E3-gp19K protein blocks transport of MHC class I from the ER to the cell surface. Three E3 proteins inhibit TNF-induced apoptosis: the RID (receptor of internalisation and degradation consisting of E3-10.4K and E4-14.5K heterodimers) and E3-14.7K.

The viral proteins required for viral DNA replication are encoded by the E2 region: the single stranded DNA binding protein (DBP), the DNA polymerase and the preterminal protein (pTP). DBP binds to the displaced strands of the template DNA. pTP is the primer protein required for initiation of viral DNA replication. pTP is cleaved during virus assembly to TP which remains covalently bound to both ends of the linear viral DNA. Three cellular proteins are also required for efficient viral replication: transcription factors, NF-1 and Oct-1, and topoisomerase NF-II. Adenovirus replicates very efficiently in human cells and at 48 h postinfection up to $10^9$ new DNA molecules have been synthesized, an amount of DNA that almost equals the cellular DNA content (reviewed in 51).

After initiation of viral DNA replication, the structural proteins are expressed from the MLTU. The major late promoter (MLP) is active also during the early phase of the infection but the transcripts terminate prematurely prior to the L2 poly(A) site. At late time points of infection, the transcription from MLP is enhanced and produces a 29,000 nucleotides long pre-mRNA, which by alternative RNA splicing and polyadenylation gives rise to at least 18 mRNAs. The increase in MLP activity is probably due to the enormous increase in DNA templates, rather than an increase in transcription initiation from a few templates (139). However, viral and cellular proteins have also been shown to activate transcription of MLP: namely the viral proteins IVa2 and DBP (33, 244), and the cellular transcription factors SP1 and MAZ (196). At the late phase of infection the transport of cellular mRNAs to the cytoplasm is blocked despite a continuous transcription. Only viral mRNAs are efficiently transported and translated at the late stage of virus infection. Adenovirus uses two mechanisms to ensure an efficient translation of viral mRNAs late during infection. First, all mRNAs from the MLTU contain a common 5’ sequence, the tripartite leader, that allows the late viral mRNAs to be translated in a cap-independent manner by ribosome "shunting". Second, the virus associated RNAs (VA RNAs), prevent the translational shutoff in the host cell by inhibiting the protein kinase PKR that otherwise would inhibit eIF2α by phosphorylation.

The E4 region encodes a minimum of seven polypeptides called open reading frames (ORFs) (65, 243, 251) and is required for lytic virus growth, viral DNA replication and viral late nuclear and cytoplasmic mRNA accumulation (13, 90, 117, 256). E4-ORF3 and E4-ORF6 appear to have redundant activities, since either of the two proteins is sufficient to support growth of an E4-deletion mutant virus (13, 98, 105, 117). However, the mechanisms by which E4-ORF6 and E4-ORF3 support virus growth seem to differ. E4-ORF6 forms a complex with the E1B-55K protein which has been shown to cause degradation of p53 and selective transport of late viral mRNAs from nucleus to cytoplasm (reviewed in 53). E4-ORF3 also binds to E1B-55K, but the importance of this interaction is still unclear (53). Both E4-ORF3 and E4-ORF6 induce DNA replication, although the exact mechanisms are not known.

Since virus capsid assembly takes place in the nucleus, the viral structural proteins have to be imported to the nucleus. Virion assembly and encapsidation of the
viral dsDNA genome is facilitated by the L1 52,55K protein that is present in capsid assembly intermediates but not mature virions. The E3-11.6K, also called the adenovirus death protein (ADP), is thought to promote virus release, by disrupting the cytoskeleton and making the cell more susceptible to lysis.

The E4-ORF4 protein
The E4-ORF4 protein is not essential for virus growth (90), and an E4-mutant virus only expressing E4-ORF4 is actually less viable than a mutant virus lacking the entire E4 region (180). E4-ORF4 binds to the cellular protein phosphatase 2A (PP2A), via an interaction with the B subunit (122). PP2A is one of the major serine/threonine protein phosphatases in the cell, and plays a central role in multiple cellular processes: metabolism, transcription, RNA splicing, translation, cell cycle progression, signal transduction, development and transformation (reviewed in 108, 163, 250). PP2A is a heterotrimer composed of a 36 kDa catalytic C subunit, a 63 kDa regulatory A subunit, and a variable B subunit. The A and C subunits, that form the core enzyme, exist in two isoforms, while the multiple B subunits belongs to three unrelated gene families, B/B55/PR55, B′/B56/PR61 and B′′/PR72/PR130. The mechanisms by which PP2A can regulate such diverse cellular functions is not well understood. Different holoenzymes might dephosphorylate distinct substrates in distinct cellular compartments. Studies in yeast have suggested that different B subunits perform nonredundant functions (280). E4-ORF4 has been shown to interact with B subunits belonging to both the B and B′ families (224). It is possible that an interaction of E4-ORF4 and various PP2A subpopulations contribute to its multiple functions, all of which appears to require an active PP2A. Here follows a summary of E4-ORF4-PP2A functions known so far (see also figure 11).

The E4 promotor is, as the other early adenovirus promoters, activated by E1A during the viral infection. E1A induces the activity of cellular transcription factors, for example AP-1 (a complex of junB and c-fos proteins), ATF, E4F and E2F, that transactivates the viral promoters. E1A activation leads to an increased phosphorylation of several of the transcription factors, important for viral transcription. During virus infection the E4-ORF4-PP2A complex induces hypophosphorylation of E1A and c-fos and reduce the level of AP-1 (180). This hypophosphorylation seems to be specific since neither the E1B-55K or junB proteins were affected (180). E4-ORF4-PP2A has also been found to counteract E1A-mediated activation of the E4 promotor (11) and the E2 promotor (154). Although the direct targets of the phosphatase activity are not known, it appears likely that the E4-ORF4-PP2A complex induces hypophosphorylation of the E4F and E2F transcription factors respectively, thereby inhibiting their transactivation capacity.

E4-ORF4 also induces p53-independent apoptosis selectively in transformed cells (135, 156, 222). This activity appears to require the B55α subunit of PP2A (155, 223, 224). In cells expressing p53, E4-ORF4 does not alter p53 expression (222) however, the anti-apoptotic proteins Bcl-2, Bcl-XL and E1B-19K were found to inhibit E4-ORF4-induced apoptosis (121). Mutational analysis of the E4-ORF4 protein has shown that amino acids 64 to 102 are critical for PP2A-B55α interaction with E4-ORF4 (155, 223). A good correlation between E4-ORF4 binding to PP2A-B55α, the recruitment of an active phosphatase and the ability to induce apoptosis have been found. Many apoptotic pathways involve activation of caspases at various stages of the apoptotic process (reviewed in 47). Caspases are cystein proteases that cleave proteins after aspartic acid residues and thereby cause degradation and
inactivation of many proteins. Some reports indicates that apoptosis can be caspase independent (85, 173, 204, 261). In E4-ORF4 induced apoptosis caspase activation may not be crucial, although it appears to contribute to the process in some cell lines. Thus, activation was detected in H1299 and 293 cells but not in CHO cells (135, 145). It is currently not known which substrates the PP2A-E4-ORF4 complex targets in the apoptotic pathway.

E4-ORF4 has been shown to induce PP2A-dependent G2 to M phase arrest in *S.cerevisae* and in 293 cells (125). The E4-ORF4-PP2A complex was found to interact with, and inhibit the anaphase-promoting complex/cyclosome (APC/C) in yeast.

In addition, E4-ORF4 has been found to interact with the Src family of kinases and to downregulate Src-dependent signalling which is suggested to play a role in E4-ORF4 induced apoptosis (134). E4-ORF4 itself seems to be phosphorylated at tyrosines in cells expressing activated Src. It is not clear if E4-ORF4 association with Src kinases is facilitated by binding to PP2A, as is the case for polyoma middle T-antigen (See Py-MT in "Viral control of host cell gene expression") (134).

---

**Figure 11.** Examples of E4-ORF4 functions.

---

**Control of adenovirus gene expression by alternative RNA splicing**

Almost all pre-mRNAs encoded by the adenovirus genome are alternatively spliced. The exceptions are the genes encoding for polypeptides IX and IVa2. The pattern of alternatively spliced mRNAs expressed from the different transcription units changes from the early to late phase of infection, indicating that alternative splicing is temporally regulated. During infection there is a general tendency, that at later time points, shorter mRNAs are produced by splicing out larger introns. More and more evidence support the hypothesis that viral control of SR protein function is important for the proper expression of viral proteins during lytic virus growth. Most studies have focused on adenovirus alternative splicing of the E1A and MLTU mRNAs.
The E1A pre-mRNA gives rise to three major mRNAs, the 13S, the 12S and the 9S mRNA that are spliced using a common 3’ splice site and alternative 5’ splice sites (see figure 12) (reviewed in 106). During the early phase of infection the 13S is the most abundant mRNA, while the 9S mRNA is the predominant E1A mRNA expressed late in infection. Many reports indicate that the shift in E1A alternative splicing is regulated by SR proteins. High concentrations of ASF/SF2 and SC35 favours splicing of the 13S mRNA both in vitro and in vivo (23, 99, 253). In contrast, SRp40, SRp55, SRp75, SRp20 and 9G8 favour 12S mRNA splicing (12, 219, 270), while high concentrations of hnRNPA1 and SRp54 activate 9S mRNA splicing (23, 277). Extracts prepared from late adenovirus infected cells have been shown to activate 9S mRNA splicing (73). This phenomenon has been explained by nuclear splicing factors and SR proteins being sequestered late during infection by the enormous increase in late mRNA expression (73, 99, 133).

Figure 12. Alternative splicing pattern of the E1A transcription unit. (Splicing to the minor 10S and 11S mRNAs are not shown).

The MLTU generates approximately 18 mRNAs by alternative 3’ splice site usage (figure 10) (reviewed in 106). These mRNAs are grouped into five families (L1-L5) where mRNAs within each group share the polyadenylation site. As mentioned above all mRNAs from the MLTU have a common 201 nucleotide long tripartite leader sequence at their 5’ end. In addition a fraction of the MLTU mRNAs have a fourth exon, the so called i-leader exon. In early infected cells the i-leader is preferentially included, while the majority of the late mRNAs contain the classical tripartite leader. The i-leader encodes a 16 kDa protein, which function is unknown, but appears to be dispensable for virus growth in tissue culture cells (226, 232). The E4-ORF3 and E4-ORF6 proteins have been shown to regulate tripartite leader splicing (189, 190, 193). They do this by opposite mechanisms: E4-ORF6 stimulates i-leader exon skipping while E4-ORF3 facilitates i-leader exon inclusion (189). Interestingly, E4-ORF3 and E4-ORF6 can also regulate exon inclusion/exon skipping on nonviral pre-mRNAs (188, 189).

During the infection it appears that the specificity in 3’ splice site recognition changes. Generally, 3’ splice sites with long consensus-type p(Y)tracts are repressed while 3’ splice sites with short atypical p(Y)tracts, which bind U2AF65kDa inefficiently, are enhanced in splicing (176, 177).

The L1 gene can be alternatively spliced using a common 5’ splice site and two competing 3’ splice sites, generating the 52,55K mRNA (proximal 3’ splice site) or the IIIa mRNA (distal 3’ splice site), respectively (figure 13) (106). During a lytic adenovirus infection, the splicing pattern of L1 is regulated such that the 52,55K
mRNA is produced both early and late after infection, whereas the IIIa mRNA is produced exclusively late (2, 186, 220). The shift in L1 alternative splicing can be reproduced in vitro (129). The activation of IIIa splicing results from an enhanced efficiency of IIIa 3´ splice site recognition, combined with a decrease in 52,55K splicing (129). Recent results have shown that IIIa splicing is regulated by two cis-acting elements located at the IIIa 3´ splice site: the IIIa repressor element (3RE) and the IIIa virus-infection dependent splicing enhancer (3VDE) (figure 13 and 18) (114, 178). SR proteins purified from HeLa cells (SR-HeLa) inhibit the usage of the IIIa 3´ splice site by binding to the 3RE (114). SR proteins binding to the 3RE block IIIa splicing by inhibiting recruitment of U2 snRNP to the branch point. Several of the classical SR proteins (SRp75, SRp40, SRp55 and SRp30s) were shown to individually repress IIIa splicing, with SRp30 and SRp55 being most effective. The absence of IIIa splicing in early virus-infected cells can therefore, in part, be explained by a direct inhibitory effect of SR proteins on IIIa 3´ splice site usage (see figure 18).

![Figure 13. L1 alternative splicing. 3RE= IIIa repressor element, 3VDE= IIIa virus dependent splicing enhancer. * the branch point of IIIa 3´ splice site. Sequences underlined are putative SR protein binding sites.](image)

The 3VDE is the minimal element required for activation of IIIa splicing in nuclear extracts from adenovirus infected HeLa cells (Ad-NE) (178). It is not yet clear how the 3VDE activates splicing. Current evidence suggest that U2AF binds less efficiently to the 3VDE in Ad-NE compared to HeLa-NE ((178), Lützelberger and Akusjärvi, unpublished). Thus, available data suggest that splicing activation by 3VDE may operate via a novel U2AF independent pathway. Potentially a novel factor replaces U2AF as the p(Y)tract binding factor and helps to recruit U2 snRNP to the IIIa branch point (see figure 18). Interestingly, there are specific differences in the proteins binding to the 3VDE in Ad-NE compared to HeLa-NE (Lützelberger and Akusjärvi, unpublished). These may be important for the activation of IIIa splicing by 3VDE.
ASF/SF2 overexpression blocks the early to late shift in viral mRNA expression resulting in the retention of the early splicing patterns of the E1A, E1B and L1 mRNAs, also at late time points of infection (171). ASF/SF2 overexpression also reduced the efficiency of viral DNA replication, major late mRNA accumulation, and new virus particle formation (171).

In addition, viral DNA replication, per se, appears to have a positive effect on the early to late shift in alternative splicing of viral mRNAs (133). However, the shift was not complete when late viral protein synthesis was blocked, suggesting that late viral protein synthesis also play a role. The requirement for late viral protein synthesis differed between E1A, E1B and L1, with L1 having the highest dependence for late protein synthesis (133). The shift to IIIa 3’ splice site usage was incomplete when protein synthesis was blocked. The L1 mRNA profile resembled that seen in an infection using a recombinant adenovirus overexpressing ASF/SF2 (171); i.e, a large fraction of the L1 mRNAs retained the i-leader sequence.
VACCINIA VIRUS

This chapter refers to four main references (64, 111, 174, 175), when no other reference is given in the text.

General background
Vaccinia virus belongs to the Poxviridae family of viruses that replicate in the cytoplasm of vertebrate and invertebrate cells. The most notorious member, variola virus, causes smallpox and has luckily been eradicated as a human pathogen. Vaccinia virus has no natural host, although it has been isolated from buffalo. Vaccinia virus was the first virus that was seen microscopically, grown in tissue culture, purified and chemically analyzed. Today, recombinant vaccinia viruses are tested as vectors for vaccination against a variety of infectious agents.

Vaccinia virus has a brickshaped virion, and a dsDNA genome of about 200 kbp. The genome encodes for over 200 potential proteins. Both strands of the genome are transcribed, but in contrast to smaller DNA viruses, overlapping open reading frames (ORFs) are rare. By convention the genes are named based on their location relative to the HindIII restriction map of the genome. Open reading frames are numbered in the order they appear from left to right. The letters “L” and “R” indicate whereas the ORFs are encoded in the left- or rightwards direction. The majority of essential genes map within the central region of the genome, which is highly conserved among poxviruses. The life cycle is divided into three phases: the early, intermediate and late phases. There is a temporal regulation of vaccinia virus transcription. Early genes encode proteins that are important for activation of intermediate genes, which in turn are required for transcription of late genes.

The virion is a complex structure consisting of a lipoprotein bilayer surrounding a biconcave core, and is composed of more than 100 polypeptides. The virus encodes for an eukaryotic-like multisubunit RNA polymerase and all of the additional enzymes required for capping, methylation and polyadenylation of its mRNAs. All of the proteins necessary for transcription and processing of the early genes are packaged within the virus particle.

The virus life cycle
The infectious cycle (figure 14) is rapid and new infectious virus particles are detected already in 4-6 hours post infection, with a maximum yield of virus produced approximately 24 hours post infection (10⁴ particles per cell). Entry of vaccinia virus into the cytoplasm involves fusion with cell membrane and subsequent virus internalization. The virus-encoded growth factor (VGF or C11R) has been found to bind to the EGF receptor and is believed to be, in part, responsible for virus entry.

Immediately after the virus has entered the cell it starts to uncoat by a two-step process. In the first step of uncoating the phospholipids and about 50% of the virion proteins are released. This step of uncoating is not blocked by inhibitors of either RNA or protein synthesis. The first step of uncoating releases the factors required for early transcription and RNA processing. Induction of early gene expression induces a second uncoating reaction in which a nucleoprotein complex containing the genome is released from the core. The second step of uncoating which begins 0.5 to 2 hours post infection requires RNA and protein synthesis and is prevented by UV irradiation of the virus. It is believed that the second step of uncoating is necessary for viral DNA replication.
Early transcription is detected within 20 minutes post infection, with the maximum level of expression reached at 1-2 hours. The early proteins made are the RNA polymerase subunits, enzymes required for viral DNA replication, factors needed for transcription of intermediate genes and proteins that counteract the host immune system. The vaccinia virus RNA polymerase is, like the eukaryotic RNA polymerase II, a complex multisubunit enzyme built from the products of at least eight viral genes. The viral RNA polymerase transcribes all classes of vaccinia virus genes. Like the cellular RNA polymerase, the vaccinia virus RNA polymerase recognizes the viral promoters in cooperating with additional regulatory proteins. For example, transcription from the early promoters require two viral proteins: the vaccinia early transcription factor (VETF, consisting of subunits A7L and D6R) and the protein H4L (RAP94).

Vaccinia virus also encodes for its own apparatus for capping and polyadenylation of the viral mRNAs. The capping enzyme (heterodimer of D1R and D12R) binds directly to the RNA polymerase and adds the cap when the transcripts are about 30 nucleotides in length, which is similar to the cellular capping enzyme. However, transcription termination is different in vaccinia virus compared to eukaryotes. Transcription of vaccinia virus early genes terminates at discrete sites, 20-50 bp downstream of a T-rich sequence. The viral capping enzyme remains bound to the polymerase, recognizes the termination signal (TTTTTAT) and induces the polymerase to terminate transcription. Thus, the 3′ ends of the viral mRNAs correspond to sites of transcription termination. This is in contrast to eukaryotes, where transcription is terminated at random sites, 200-2000 bp after the poly(A) signal (201). The J3L protein (VP39) is also important for the methylation of the cap and acts as a factor stimulating transcription elongation. In addition, J3L is a subunit of the poly(A) polymerase. The larger subunit, E1L (VP55), catalyzes the addition of As to the 3′ end of the transcript. J3R binds to poly(A) and stimulates E1L activity.

Activation of intermediate genes requires the expression of two factors, VITF-1 (E4L) and VITF-2 (viral intermediate transcription factors). VITF-1 is a viral protein while VITF-2 is a cellular transcription factor, which is relocated from the nucleus to the cytoplasm in infected cells. The function of VITF-2 in uninfected cells is still unclear.

Vaccinia virus also encodes for all proteins needed for viral DNA replication. It replicates in the cytoplasm at discrete foci, sometimes called "viral factories". DNA replication begins within a few hours of infection when also the late phase of the infectious cycle starts.

Activation of late gene expression requires three intermediate gene products: A1L, A2L and G8R. The proteins made at late times of infection include most of the structural proteins and many of the enzymes encapsidated into new virus particles, including the early transcription factors. Most of the late genes are encoded in clusters in the central part of the genome. The late mRNAs have unusual features with a 5′ p(A) leader and heterogenous 3′ ends. The late mRNA can self-anneal with the early RNAs to form ribonuclease resistant hybrids. The late mRNAs do not have a defined 3′ end and mRNAs encoding the same polypeptide can vary manyfold in length. Currently, there are no logical explanation for the failure of transcription termination in late viral mRNA production. The capped 5′ p(A) leader sequences may provide a binding site for translation initiation factors and the ribosomal 40S subunit. During a vaccinia virus infection cellular transcription, DNA replication and RNA processing is inhibited, but the mechanisms behind these events are not known.
Virion assembly is a complex process involving the cleavage of structural proteins from high-molecular precursors. The initial assembly results in formation of the immature virion, a spherical particle, that is believed to have a Golgi derived double membrane wrapped around the assembling particle. The virus particle then matures into the intracellular mature virion which is released only upon cell lysis. The particle can also acquire a second double membrane from the Golgi and then be released from the cell by fusion with the cell membrane or transferred directly into a neighbouring cell through cell-cell contacts.

Figure 14. The vaccinia virus life cycle. Modified from (64).
VIRAL CONTROL OF HOST CELL GENE EXPRESSION

In this chapter I will briefly describe how viruses regulate host cell gene expression. Most importantly, I will describe viral proteins that interact with the cellular protein phosphatase 2A and viral proteins that affect RNA processing. The main reference is (64).

The cell cycle and signal transduction

Highly differentiated cells have slow rates of RNA and protein synthesis and do not express many of the components required for DNA replication. Thus, they are unable to support growth of most viruses, which prefer a rapid rate of DNA, RNA and protein synthesis. Herpesvirus provides its own replication machinery and can reproduce in highly differentiated cells, such as neuronal cells. Papovaviruses and adenoviruses, stimulate resting cells to enter the S-phase by disrupting the cellular control of cell growth and division. They do this by interfering with two important regulators in the cell: the tumor suppressor proteins pRb and p53.

pRb controls cell proliferation by binding and inhibiting the activity of the cellular transcription factor E2F. E2F stimulates transcription of cellular genes whose products are involved in DNA synthesis or control of cell cycle progression. The adenovirus E1A protein, the SV40 large T-antigen (LT) and the HPV E7 protein bind to Rb and displace it from E2F which becomes activated. As a side-effect, unscheduled DNA synthesis activates p53. p53, which is a transcription factor, activates or represses transcription of several genes, leading to either cell cycle arrest or apoptosis. The E1A, SV40 LT and HPV E7 proteins cause indirectly p53 stabilisation. However, viruses have evolved to survive apoptosis and encode proteins that interfere at multiple steps in the apoptotic pathway. The HPV E6 protein and the adenoviral proteins E4-ORF6 and E1B-55K bind to p53 and cause p53 degradation. SV40 LT also binds to p53 but is believed to sequester p53 in an inactive complex, facilitated by the SV40 small T-antigen (ST).

Promoting cell growth and proliferation include a myriad of signal transduction pathways in the cell, which are regulated by reversible phosphorylation. One cellular target that is important for this is the protein phosphatase 2A (PP2A, see "Adenovirus" chapter, section E4-ORF4). Mouse polyoma virus small and middle T (Py-ST and Py-MT) as well as SV40 ST form stable complexes with the PP2A core enzyme, by replacing the B subunit and alter PP2A substrate specificity and localisation (figure 15) (reviewed in 108, 212). The viral antigens target PP2A to overcome its negative role in some signalling pathways. SV40 ST enhances SV40 LT induced proliferation by inhibiting PP2A activity towards most substrates (212, 265). This activates kinases important in signal transduction that will promote entry into S phase. For example, SV40 ST enhances AP-1 driven transcription probably by activation of the MAP kinase pathway (227). SV40 ST has also been shown to inhibit PP2A induced dephosphorylation of p53 in vitro (217). In transient transfection assays SV40 ST enhances the transactivation capacity of p53 (264). However, SV40 ST does not induce p53-mediated apoptosis (264). Although SV40 ST stimulates cell cycle progression, it arrests cells before they enter the mitosis phase (G2/M) in a PP2A dependent manner (212).

Binding of both Py-ST and Py-MT to PP2A inhibit the activity of PP2A to all tested substrates. However, there are mechanistic differences (179). Thus, Py-ST promotes cell cycle progression in a PP2A dependent manner, while PP2A binding to Py-MT is not sufficient to promote cell proliferation (179). Py-MT is a multimeric
adaptor protein which binds several cellular signalling molecules. Apart from PP2A, Py-MT also interacts with the cellular c-Src family of kinases (42, 102, 124). Py-MT activation of the c-Src results in activation of the ras/MAP kinase pathway and cell proliferation. PP2A is required for Py-MT to interact with c-Src, but the catalytical activity of PP2A seems not to be required (24, 80, 192). A virus expressing Py-ST, but not Py-MT and Py-LT, does not transform cells. Py-ST cooperates with Py-MT to block p53-induced growth arrest and apoptosis (202). In addition, Py-ST cooperates with Py-LT to induce synthesis of specific cyclins and stimulate proliferation of resting cells (218).

The PP2A core enzyme has been shown to inhibit HIV-1 transcription and virus production (211). The HIV encoded proteins Vpr and NCp7 stimulate PP2A activity in vitro (246). In addition, Vpr induces cell cycle arrest at the G2/M transition by interacting with the B55 subunit of PP2A (figure 15) (56, 104). Vpr targets the PP2A complex to the nucleus which results in dephosphorylation and inactivation of Cdc25 which regulates cyclin activity (104).

**Figure 15.** Viral proteins binding to PP2A. See the text for details. *The adenovirus E4-ORF4 protein and its interaction with PP2A is described in the “Adenovirus” chapter.*

DNA replication
DNA viruses that do not rely on components of the cellular replication system often inhibit cellular DNA synthesis, probably to eliminate the competition for substrates from cellular DNA replication. The larger viruses; herpesvirus and poxvirus, efficiently inhibit synthesis of cellular DNA during infection. There are indications that a viral DNase causes the degradation of the cellular DNA during a poxvirus infection (52). Adenovirus also inhibits cellular DNA synthesis although the virus induces resting cells to re-enter the cell cycle. However, very little is known about the mechanisms that shut down cellular DNA synthesis.

Transcription
Cellular transcription is also often inhibited in virus-infected cells. This results in a more selective production of viral mRNAs and furthermore reduces the competition from cellular mRNAs during translation. Retroviruses and most DNA viruses depend on the cellular transcriptional machinery for gene expression and therefore does not inactivate this machinery. The larger DNA viruses selectively express viral genes. For example, herpes simplex virus type 1 (HSV-1) inhibits transcription of the majority of cellular genes transcribed by RNA polymerase II. The selective transcription of HSV-1 genes depends on several viral immediate early gene products, including ICP4 and ICP27. Poxviruses which encode for its own viral transcription system, induce a rapid
inhibition of synthesis of all classes of cellular RNAs in infected cells. In vaccinia virus infected cells, RNA polymerase II is inactivated by translocation to the cytoplasm. The majority of RNA viruses also inhibit cellular mRNA synthesis during infection. RNA viruses encode for an RNA dependent RNA polymerase and do not need the cellular RNA polymerase for gene expression. For example, a poliovirus encoded protease cleaves TBP (a general transcription factor required for RNA polymerase II transcription). This results in an inhibition of TBP binding to DNA and as a consequence initiation of transcription.

RNA processing and export of mRNA

All viruses use the cellular protein-synthesis machinery to translate their mRNAs. Viruses that rely on the cellular transcription and RNA processing machineries in the nucleus cannot completely inactivate these systems. However, viruses have evolved several mechanisms to inhibit cellular gene expression by selectively inhibiting cellular RNA processing. Several viruses have been shown to encode for proteins that inhibit RNA splicing and/or transport of cellular RNAs. Most cellular pre-mRNAs undergoes splicing prior to being exported to the cytoplasm. Thus inhibition of RNA splicing is an efficient way of keeping cellular mRNAs in the nucleus.

The influenza virus NS1 protein inhibits both polyadenylation and splicing of cellular pre-mRNAs. For example, dimers of NS1 binds to and inhibits U6 snRNA and U6atac snRNAs. NS1 thereby increases the nuclear concentrations of cellular pre-mRNAs, which the viral cap-dependent endonuclease can use to cleave primers required for viral mRNA synthesis.

The HSV protein ICP27 (IE63) enhances transport of unspliced viral RNA during infection (reviewed in 215). It has been shown that ICP27 modulates the polyadenylation machinery and inhibits splicing of cellular pre-mRNAs (93), although the exact mechanisms are not yet known. In general, herpesvirus genes lacks introns and inhibition of splicing could uncouple the splicing and mRNA export pathways, facilitating nuclear export of viral intronless genes. This allows the viral mRNAs to bypass the normal requirement for splicing for mRNA export. ICP27 appears to inhibit splicing prior to the first catalytic step (19). It has been suggested that ICP27 induces SR protein dephosphorylation (215). ICP27 also interacts with the cellular splicing factor SAP145 and the SF2/ASF-associated protein p32, interactions which have been suggested to mediate ICP27 inhibition of splicing (18, 19).

In retroviruses, the transport of partially spliced and unspliced viral mRNAs are essential for viral protein production and formation of virus genomes for virus assembly (96). HIV encodes for a transport protein, Rev, which interacts with the Rev-responsive-element (RRE) present in viral transcripts. Binding of Rev to an RNA results in nuclear to cytoplasmic transport of the RNA. Rev has also been suggested to inhibit splicing by repressing the formation of the spliceosome in RRE-containing transcripts (119, 120). Rev interaction with p32 could potentially be involved in the inhibition of splicing (237). Rev inhibition of splicing has also been suggested to involve recruitment of ASF/SF2 to the Rev/RRE complex (200). In addition, the equine infectious anemia virus (EIAV) ERev protein has been shown to induce alternative splicing of its own pre-mRNA, probably by simultaneously recruiting ERev and ASF/SF2 to the EIAV RRE (39).

Adenovirus disrupt cellular gene expression by inhibiting export of cellular mRNAs from the nucleus to the cytoplasm (reviewed in 53). During the late phase of infection the majority of mRNAs in the cytoplasm are of viral origin. The preferential
export of viral late mRNAs in adenovirus infected cells requires the E1B-55K/E4-ORF6 protein complex. It is not yet known how the complex induces selective transport of viral mRNAs. One hypothesis is that E1B-55K binds up a limited cellular factor needed for cellular mRNA export and makes it available to viral mRNA export. A candidate protein is the E1B-AP5 protein, a nuclear RNA binding protein belonging to the hnRNP family. It has been speculated that E1B-AP5 bridges between viral mRNAs and the E1B-E4-ORF6 complex, which serves as an adaptor protein that connects newly synthesized mRNAs to an active shuttling machinery (71). Evidence has also been presented suggesting that E1B-55K binds directly to RNA (103).

**Translation**
As part of the antiviral defense, the cell responds to a virus infection by inhibiting protein synthesis in an attempt to limit virus production. The infected cell produces interferons, that diffuse to neighbouring cells and enhance the transcription of a number of genes that are involved in the antiviral defense, among them encoding the cellular RNA-activated protein kinase (PKR). PKR becomes activated when it binds to double stranded (dsRNA) which is a hallmark of a virus infected cell. PKR binding to dsRNA causes PKR dimerization and auto-phosphorylation which then in turn phosphorylates many substrates. A key substrate is the eukaryotic translation initiation factor 2α (eIF2α), which can not reload from eIF2a-GDP to eIF2a-GTP when it is phosphorylated, and therefore causes inactivation of translation initiation. Many viruses encode for proteins that inhibit PKR activity by different mechanisms: by inhibiting dsRNA binding (adenovirus VA-RNA, vaccinia virus E3L, influenza virus NS1, reovirus σ3, HIV TAR RNA), by inhibiting PKR dimerisation (influenza virus p58ψψpK, hepatitis C virus NS5A), by functioning as a pseudosubstrate (vaccinia virus K3L, HIV Tat) or by blocking eIF2α dephosphorylation by recruitment of the cellular protein phosphatase 1 (herpes simplex virus γ134.5).

Many viruses also modify the eIF4F complex, which is required for cap recognition by the 40S ribosomal subunit. Inactivation of this complex ensures that capped cellular mRNAs are not translated while viruses which have evolved other mechanisms to initiate translation, still are able to translate the viral mRNAs. Adenovirus and influenza virus induce dephosphorylation of an eIF4F component which results in a reduced binding of eIF4F to the cap structure. A poliovirus infection inactivates a protein component of eIF4F by proteolytic cleavage. During an adenovirus infection, the tripartite leader, which is attached to the late mRNAs, mediates translation by ribosome "shunting”. Poliovirus uses an IRES (internal ribosome entry site) to recruit the 40S ribosome to the viral mRNA.

**Protein transport**
Some viruses encode for proteins that interfere with the transport of specific cellular proteins to their normal compartments, mostly for the purpose of hiding the infected cell from being recognized by the immune system. The MHC class I presents viral polypeptides on the cell surface in order to activate the immune response. The adenovirus E3-gp19K protein blocks transport of MHC class I from the ER to the cell surface. Human cytomegalovirus and HIV encode for proteins that target MHC class I for rapid degradation by the cellular proteosome. Poliovirus and other enteroviruses cause a total disruption of the ER, which efficiently inhibits protein trafficking to the cell surface.
Regulation of adenovirus alternative RNA splicing — inactivation of SR proteins as splicing regulatory proteins by dephosphorylation (paper I)

The adenovirus L1 region is regulated at the level of alternative RNA splicing, producing either the 52,55K mRNA or the IIIa mRNA (see "Adenovirus" chapter). IIIa mRNA splicing is restricted to the late phase of the infection. It was previously demonstrated that IIIa splicing is repressed in nuclear extracts prepared from uninfected cells (HeLa-NE) by the SR family of splicing factors that bind to the intronic repressor element 3RE (114). Removing the 3RE element from a mini-52,55K-IIIa tandem transcript, mimiced the early to late shift in L1 alternative splicing in an in vitro splicing assay. This result indicated that SR protein inhibition of IIIa splicing is important for L1 alternative splicing. The steady-state amount of the classical SR proteins does not change during virus infection. We therefore analyzed whether SR proteins were post translationally modified late during virus infection.

For this experiment, we purified SR proteins from late adenovirus-infected cells (SR-Ad) and compared their biological activity with SR proteins simultaneously prepared from uninfected HeLa cells (SR-HeLa). The results showed that SR-Ad was much less active, compared to SR-HeLa, in in vitro splicing reactions: both in activation of 52,55K pre-mRNA splicing in splicing deficient S100 extracts (HeLa-S100), and repression of IIIa pre-mRNA splicing in nuclear extracts (HeLa-NE) (Paper I, fig.1a and b). It was also found that SR-Ad bound less efficiently to the IIIa pre-mRNA, compared to SR-HeLa (Paper I, fig.1c). The reduced capacity of SR-Ad to bind RNA did not result from a copurification of late adenovirus RNA, because pre-incubation of SR-Ad with micrococcal nuclease gave the same result in the crosslinking assay. This result indicated that SR-Ads are not sequestered from repressing IIIa splicing by binding to late RNA, the mechanism that was previously suggested to control the alternative splicing of the E1A pre-mRNA (99).

As described earlier, SR proteins are highly phosphorylated in vivo, mainly at serine residues in the RS-domain, a modification which is important for their function as splicing enhancer proteins. To investigate whether the decrease in activity and RNA binding of SR-Ad correlated with a change in phosphorylation, 32P-labelled SR-HeLa and SR-Ad proteins were analyzed by two-dimensional gel electrophoresis (paper I, fig. 2). The result demonstrated that SR-HeLa had more 32P incorporated compared to SR-Ad and furthermore, that the isoelectric point of a large fraction of the SR-Ad proteins was shifted towards the basic side. Collectively, the results suggested that SR-Ad were underphosphorylated. One exception was the SRp20 protein, which appeared to behave similarly in the two SR preparations. Interestingly, SRp20 also showed a low activity as a repressor protein of IIIa pre-mRNA splicing in vitro (114).

In conclusion, these results showed that SR proteins from adenovirus infected HeLa cells are inactivated as splicing enhancer or splicing repressor proteins by a virus-induced dephosphorylation. This work lead to the first description of a regulated system where changes in the phosphorylated status of SR proteins control the alternative RNA splice site choice.
The adenovirus E4-ORF4 protein induces PP2A dependent SR protein dephosphorylation by interacting with a subset of phosphorylated SR proteins (paper I and II)

The results presented in paper I raised the question, how does adenovirus induce SR protein dephosphorylation. The E4-ORF4 protein was a good candidate protein because it was previously shown to bind to PP2A and induce dephosphorylation of specific transcription factors (11, 122, 180). We therefore tested whether E4-ORF4 also could induce SR protein dephosphorylation. The results showed that E4-ORF4 in the presence of HeLa-NE, induced SR protein dephosphorylation (see figure 16). E4-ORF4 by itself did not induce dephosphorylation, the HeLa-NE was required apparently to recruit the actual phosphatase to the reaction. Addition of okadaic acid (OA), which inhibits PP2A activity (20 mM OA), also inhibited the effect of E4-ORF4 on SR protein dephosphorylation, again suggesting the importance of this phosphatase. Incubating SR proteins in HeLa-NE induced phosphate release that was not PP2A dependent. This result was not unexpected, considering the number of different phosphatases that are present in the nuclear extract and that potentially could dephosphorylate SR proteins. Importantly, addition of E4-ORF4 to nuclear extracts induced a specific increase in SR protein dephosphorylation.

![Figure 16](image)

**Figure 16.** E4-ORF4 causes SR protein dephosphorylation. (paper I, fig. 3a).

In paper I we showed that all SR proteins, except SRp20, were underphosphorylated during an adenovirus infection. In paper II we determined whether E4-ORF4 targets all SR proteins. In our model, E4-ORF4 makes simultaneous interactions with the SR proteins and PP2A. Therefore, we determined the binding specificity of E4-ORF4 interaction with SR proteins both by in vitro binding assays and by co-immunoprecipitations from extracts of transiently cotransfected cells. Surprisingly, E4-ORF4 was shown to bind efficiently to only two of the SR proteins tested: ASF/SF2 and SRp30c (paper II, fig. 1). Interestingly, testing ASF/SF2 mutant proteins showed that the RS domain of ASF/SF2 was dispensable for the interaction between E4-ORF4 and ASF/SF2 (paper II, fig. 2). The observation that the RRM s of ASF/SF2 are required for interaction with E4-ORF4 was not an indirect effect, mediated by E4-ORF4 and ASF/SF2 simultaneously binding to contaminating RNA in the extract. Two results suggested this: (i) E4-ORF4 did not interact with
other SR proteins, although they are also expected to bind to RNA, (ii) RNase treatment of the in vitro translated ASF/SF2 protein or extracts used for the immunoprecipitations, did not disrupt the interaction with E4-ORF4.

Since E4-ORF4 induces SR protein dephosphorylation, one would predict that E4-ORF4 interacts preferentially with hyperphosphorylated SR proteins. To test this hypothesis, we compared the capacity of E4-ORF4 to interact with SR-HeLa and SR-Ad. Equal amounts of SR-Ad and SR-HeLa were analyzed by a western blot assay (paper II, fig. 3). Probing filters with the SR protein specific monoclonal antibody 104 (mAb104), that recognizes phospho-epitopes in the RS domain, provided additional support for our conclusion that SR-Ad are hypophosphorylated (paper II, fig. 3A). (i), SR-HeLa were efficiently recognized by mAb104, while SR-Ad were not. (ii), the SR-Ad proteins recognized by mAb104 migrated slightly faster than the corresponding proteins in SR-HeLa. (iii), incubating SR-Ad in a small amount of HeLa-NE under conditions supporting phosphorylation restored the phospho-epitopes recognized by mAb104. Also, the mobility of the re-phosphorylated SR-Ad proteins was reduced such that SR-Ad migration was identical to SR-HeLa. Reprobing this filter with mAb96, that recognizes RRM1 in ASF/SF2, showed even more clearly the difference in migration pattern between ASF/SF2 in SR-HeLa and SR-Ad (paper II, fig. 3B).

A far western blot assay detecting E4-ORF4 protein interacting with SR proteins, showed that E4-ORF4 interacts efficiently only with the 30 kDa fraction in SR-HeLa (paper II, fig. 3C). This finding supported the binding studies described above, since ASF/SF2 and SRp30c both belong to the SRp30 fraction. However, E4-ORF4 did not interact efficiently with SR-Ad. Interestingly, in vitro phosphorylation of SR-Ad in HeLa-NE restored E4-ORF4 interaction, suggesting that E4-ORF4 preferentially interacted with the hyperphosphorylated form of SR proteins.

In conclusion, these experiments suggest that E4-ORF4 induces SR protein dephosphorylation by interacting with a subset of phosphorylated SR proteins. Still, the observation that E4-ORF4 interacts preferentially with hyperphosphorylated SR proteins seems paradoxical. Thus, E4-ORF4 interacts efficiently with SF2-ΔRS, which lacks the RS domain encoding the phosphorylated serines. However, results have been presented suggesting that ASF/SF2 also is phosphorylated at residues in the RRM (40, 198). Potentially E4-ORF4 induces RS domain dephosphorylation, but the actual binding to ASF/SF2 is mediated through phosphorylated residues in the RRM. Alternatively, RS domain dephosphorylation may change the structure of ASF/SF2 such that E4-ORF4 cannot bind. Finally, nearly all SR proteins (except SRp20) were shown to be dephosphorylated in SR proteins prepared from adenovirus-infected HeLa cells (paper I). In paper II, E4-ORF4 was found to interact specifically with ASF/SF2. Thus, there are apparently other mechanisms during an adenovirus infection that cause SR protein dephosphorylation.

E4-ORF4 is a splicing enhancer protein — the IIIa repressor element (3RE) is necessary and sufficient for E4-ORF4 activated splicing (Paper I and II)

The E4-ORF4-PP2A complex was previously suggested to regulate transcription by inducing dephosphorylation of specific transcription factors (11, 122, 154, 180). Since E4-ORF4 was shown to induce SR protein dephosphorylation, we tested whether E4-ORF4 also controls RNA splicing. The results showed that preincubation of SR-HeLa with a recombinant E4-ORF4 protein in HeLa-NE alleviated the repressive effect of SR-HeLa on IIIa pre-mRNA splicing (paper I, fig. 3b). As previously shown, SR
proteins inhibit IIIa splicing by blocking U2 snRNP recruitment to the spliceosome and thus A complex formation (114). Accordingly, addition of E4-ORF4 to splicing extracts alleviated the inhibitory effect of SR-HeLa on A complex formation (paper I, fig. 3c). Collectively, these results suggested that E4-ORF4 could convert the splicing properties of SR-HeLa to that of SR-Ad. This finding argues against the hypothesis that the reduced splicing repressor activity of SR-Ad was due to copurification of an unknown splicing repressor protein with SR-Ad. Preincubation of E4-ORF4 with HeLa-NE alone also had a minor activating effect on A complex formation, suggesting that the endogenous pool of SR proteins in the nuclear extract is enough to repress complex A formation on the IIIa 3’ splice site.

To further investigate the role of E4-ORF4 in IIIa pre-mRNA splicing (paper I) we reproduced the effect of E4-ORF4 in a transient transfection assay. The results showed that E4-ORF4 expression activated IIIa splicing both on the single IIIa pre-mRNA and the mini-52,55K-IIIa-tandem pre-mRNA (paper I, fig. 4). Deletion of the 3RE from the IIIa pre-mRNA enhanced basal IIIa splicing, confirming the previous in vitro splicing data (114). However, E4-ORF4 did not activate IIIa splicing in a transcript lacking the 3RE. This result suggested that the effect of E4-ORF4 on IIIa splicing was 3RE dependent.

As described previously, IIIa splicing is also regulated by a second viral element, the 3VDE (178). Therefore in paper II we examined whether the 3RE was the only element required for E4-ORF4 induced activation of splicing. Inserting either the 3RE or the 3VDE into a β-globin minigene reduced β-globin pre-mRNA splicing. This result was in agreement with previous in vitro splicing data; the 3RE was shown to be the minimal sequence element required for SR protein mediated inhibition of splicing (114), and the 3VDE was shown to only function as a splicing enhancer in the context of virus-infected cells (178). However, cotransfection of E4-ORF4 activated the glob (3RE) splicing, but only marginally the glob (3VDE) splicing. Collectively, these data show that the 3RE is the primary element both necessary and sufficient for E4-ORF4 enhanced splicing.

**Strong E4-ORF4 interaction with ASF/SF2 and recruitment of an active PP2A enzyme are required for E4-ORF4 mediated activation of IIIa pre-mRNA splicing (paper II)**

In paper II we examined more thoroughly how E4-ORF4 activated IIIa splicing. Our hypothesis was that E4-ORF4 functions as a specificity factor that targets ASF/SF2 for PP2A dephosphorylation, thereby reducing its activity as a IIIa splicing repressor protein. In a previous study, Shtrichman et al. 1999 (223) analyzed the capacity of different E4-ORF4 mutant proteins to bind to PP2A and to recruit a functional phosphatase activity. In paper II we used this collection of E4-ORF4 mutant proteins to analyze their capacity to interact with ASF/SF2 in vitro and to activate IIIa pre-mRNA splicing in a transient transfection assay. The results were compared with the data presented on PP2A recruitment (223). The summary of the three different assays is presented in figure 17. The E4-ORF4 mutants have point mutations (one to three amino acids) at conserved regions in E4-ORF4 (figure 17). The in vitro binding assays showed that two regions were important for E4-ORF4 interaction with ASF/SF2; the N-terminal part and amino acids 61 to 95. Probably, the interaction between E4-ORF4 and ASF/SF2 depends on E4-ORF4 protein conformation, since a simple amino acid sequence determinant for binding was not identified.
Collectively the results suggested that both strong binding to ASF/SF2 and an efficient recruitment of a functional phosphatase is necessary for E4-ORF4 activation of IIIa splicing (figure 17). E4-ORF4 interaction with only one component was not enough for E4-ORF4 induced activation of IIIa splicing. R3-9 was the only mutant that did not conform to this general rule. Although it bound strongly to ASF/SF2 and recruited PP2A efficiently it did not, for unknown reasons, activate IIIa splicing. Potentially, the interaction of R3-9 with ASF/SF2 and PP2A is nonproductive, and fail to present ASF/SF2 correctly for dephosphorylation. However, these experiments do not exclude the possibility that E4-ORF4 also targets another splicing factor required for E4-ORF4 function in splicing. Although E4-ORF4 also interacted with SRp30c, all the E4-ORF4 mutant proteins bound with that same efficiency to SRp30c (data not shown). Thus, SRp30c probably does not make a significant contribution to the E4-ORF4-induced activation of IIIa splicing.

In conclusion, we showed that reduced interaction of E4-ORF4 to ASF/SF2 or the PP2A phosphatase complex did not activate IIIa splicing in transient transfection experiments, suggesting that an interaction between all three proteins is required for E4-ORF4 to induce SR protein dephosphorylation.

**Vaccinia virus inactivates SR proteins by dephosphorylation (Paper III)**

In paper III we investigated whether control of SR protein activity was a strategy used by viruses other than adenovirus to limit host cell gene expression. Since RNA splicing is required for most cellular mRNAs to be exported to the cytoplasm, a virus-induced suppression of RNA splicing could be an important mechanism by which viruses inhibit host cell gene expression. We selected vaccinia virus as our first
candidate virus causing SR protein inactivation, since vaccinia virus genes lack introns and its gene expression therefore should not dependent on the host cell splicing machinery. For this experiment, we purified SR proteins from vaccinia virus-infected HeLa cells (SR-VV) and compared their activity as splicing regulatory proteins with that of SR-HeLa and SR-Ad. The results show that SR-VV were completely inactivated as splicing repressor and splicing activator proteins in in vitro splicing reactions using the IIIa and the 52,55K pre-mRNAs, respectively (paper III, fig.1 and 2). SR-Ad had an intermediate activity compared to SR-HeLa and SR-VV. Based on our finding that SR-Ad were inactivated as splicing regulatory proteins by a virus-induced dephosphorylation, we hypothesized that the SR-VV would similarly be inactivated by dephosphorylation. To test this we compared SR-HeLa, SR-Ad and SR-VV in a western blot assay using the phospho-epitope specific mAb104. The results showed that SR-VV was dephosphorylated to a similar extent as SR-Ad (paper III, fig. 3). Importantly, incubation of SR-Ad and SR-VV in a small amount of HeLa nuclear extract restored the phospho-epitopes recognized by mAb104 to both SR protein preparations. In addition, this in vitro rephosphorylation assay restored the IIIa splicing repressor activity to SR-Ad and SR-VV (paper III, fig 4), further emphasizing our conclusion that both SR-Ad and SR-VV are inactivated as splicing proteins by dephosphorylation. Compared to the splicing activity of SR-HeLa, SR-VV did not regain full activity to repress IIIa splicing, as did SR-Ad. Based on our current work we can not explain this finding. Potentially SR-VV are subjected to another post-translational modification. A modification that is not reversed under the rephosphorylation conditions.

How does vaccinia virus cause SR protein dephosphorylation? Vaccinia virus encodes, in fact, a dual specificity phosphatase (VH1), that is required for vaccinia virus gene expression, and is vital for virus growth (144). VH1 has been shown to induce dephosphorylation of a specific transcription factor, STAT1, and in this way inhibit the interferon-gamma response pathway (184). We therefore tested if VH1 could induce SR protein dephosphorylation. The results show that a recombinant VH1 protein could dephosphorylate SR-HeLa in an in vitro dephosphorylation assay system (data not shown). Since preincubation of E4-ORF4 with SR-HeLa in HeLa-NE induced SR protein dephosphorylation and activated IIIa pre-mRNA splicing, we tested VH1 in the same assay. However, even at low concentrations, VH1 repressed IIIa splicing (data not shown). Potentially VH1 is more efficient compared to E4-ORF4 in dephosphorylation of SR proteins and hence all the SR proteins, essential for splicing, are inactivated in VH1 treated extracts. Alternatively, VH1 may dephosphorylate (an)other important splicing factor(s). Further experiments are required to determine whether VH1 is the vaccinia virus protein causing SR protein dephosphorylation during virus growth.

Based on the experiments presented in paper I and III we propose that SR protein inactivation may be an important strategy used by viruses to ensure an efficient virus production. Inhibition of SR proteins ensures that host cell pre-mRNA splicing is inhibited, and hence the mRNAs are not transported to the cytoplasm.
Adenovirus and vaccinia virus induce SR protein dephosphorylation — a general viral mechanism to control host cell gene expression?

We have demonstrated that both adenovirus and vaccinia virus modulate the activity of the essential SR family of splicing factors (Paper I and III). SR proteins are highly phosphorylated, a modification important for their activity in splicing. We show that adenovirus and vaccinia virus induce SR protein dephosphorylation, a modification that inhibits their activity as splicing repressor and splicing enhancer proteins.

Vaccinia virus replicates in the cytoplasm and encodes for genes that do not contain introns. Thus, vaccinia virus is not dependent on the cellular splicing machinery for its replication. We found that vaccinia virus completely inactivates SR proteins as splicing regulatory proteins. In adenovirus the situation is more complex. Adenovirus replicates in the nucleus and most adenovirus genes contain introns (although few compared to cellular genes). Therefore, adenovirus is dependent on the cellular splicing machinery. Adenovirus also causes SR protein inactivation, which might seem contradictory to the requirement of a functional splicing system during an adenovirus infection. However, adenovirus appears to redirect the splicing machinery such that splicing of late viral mRNAs are enhanced. Three observations support this hypothesis. (i) in general, splicing of 3’ splice sites with short atypical p(Y)tracts, that bind U2AF65kDa inefficiently, are selectively enhanced in Ad-NE (176). (ii), efficient activation of IIIa splicing in Ad-NE requires the 3VDE element (178). U2 snRNP binds to the IIIa branch point late in infection. However, U2AF binds less efficiently, which point to a viral mediated recruitment of U2 snRNP. (iii), SR-Ad were not completely inactivated as splicing regulatory proteins as was the case for SR-VV (paper III). It is likely that some SR protein activity is still needed for splicing of adenovirus late mRNAs. In conclusion, both alleviation of the repressive effect of SR proteins binding to the 3RE and activation of IIIa splicing through the 3VDE is important for regulation of L1 alternative splicing in an adenovirus infection (see figure 18) (178).

Interestingly, SR-Ad could be reactivated as splicing repressor proteins by \textit{in vitro} phosphorylation. In contrast, SR-VV were only partially reactivated as splicing...
repressor proteins. Potentially, SR-VV are subjected to an additional modification. Methylation and acetylation have also been found to be important post-translational modifications of many proteins regulating gene expression, for example transcription factors and histones. Thus, it could be worthwhile to investigate whether these modifications also are important for SR protein activity in splicing.

Both adenovirus and vaccinia virus infections inhibit cellular mRNA production. In adenovirus, viral mRNAs are selectively transported to the cytoplasm by the help of the E4-ORF6/E1B-55K protein complex. However, the exact mechanisms behind the selective inhibition of host cell mRNA production are still unclear for both viruses. We propose that an SR protein inactivation through a virus-induced dephosphorylation contributes to the selective expression of viral mRNAs in adenovirus and vaccinia virus infected cells.

As described in the section "Viral control of host cell gene expression", additional viruses have been shown to inhibit cytoplasmic accumulation of cellular mRNAs during the course of infection. Most cellular mRNAs have to be spliced before they are transported to the cytoplasm. Thus a virus-induced inhibition of RNA splicing will ensure that cellular pre-mRNAs stays in the nucleus. It is possible that modulation of SR protein activity is a general mechanism by which viruses control host cell gene expression. Future studies should be aimed at determining whether additional virus families inactivate the SR family of splicing factors.

E4-ORF4 — cellular and viral functional homologues?
We have shown that the adenovirus E4-ORF4 protein interacts with a subset of SR proteins, induces PP2A dephosphorylation of SR proteins and regulates adenovirus alternative RNA splicing (figure 19) (Paper I and II). In future studies, E4-ORF4 could potentially be used as a tool to investigate ASF/SF2 specific functions in splicing regulation. As previously mentioned, the repressor and enhancer functions of ASF/SF2 are encoded by distinct domains in the protein (V. Dauksaite and G. Akusjärvi, submitted). Thus, it could be worthwhile to investigate whether E4-ORF4 also inactivates ASF/SF2 as a splicing enhancer protein. In addition, the mapping of the phosphorylated residues in ASF/SF2 targeted by the E4-ORF4-PP2A complex will be important. E4-ORF4 also interacted with SRp30c, although SRp30c did not seem to contribute to the E4-ORF4 induced activation of IIIa splicing. The functional importance of the E4-ORF4 and SRp30c interaction should be further characterized. Another issue to be addressed is whether the E4-ORF4/PP2A complex targets additional splicing factors controlling virus and cell specific splicing events.

In an adenovirus infection all SR proteins, except SRp20, were shown to be dephosphorylated (Paper I). Since E4-ORF4 appears to specifically target ASF/SF2 and SRp30c, E4-ORF4 can not account for all SR protein dephosphorylations. Thus, there should exist additional factors that regulate SR protein dephosphorylation during an infection. Two other adenovirus proteins have been found to regulate splicing: E4-ORF3 and E4-ORF6 (189, 193). Preliminary studies suggest that E4-ORF6 also activates IIIa splicing in a 3RE dependent manner (S. Fan, T. Punga and G. Akusjärvi, unpublished results). However, it is not yet known if this activation operates via SR protein dephosphorylation. The E1B-55K protein appears to enhance the activity of E4-ORF6 as an enhancer protein of IIIa splicing in vitro (S. Fan, T. Punga and G. Akusjärvi, unpublished results). It is therefore possible that the E4-ORF6/E1B-55K complex regulates RNA transport by directly controlling RNA splicing.
As described earlier, other viruses also encode for proteins that target PP2A. Thus, they could also have a role as regulators of RNA splicing. Interestingly, Py-ST activates IIIa splicing both in vitro and transient transfection experiments (S. Fan, E. Boija and G. Akusjärvi, manuscript in preparation). However, the mechanism of Py-ST activation of IIIa splicing appears to differ from that of E4-ORF4 activation. First, Py-ST activates IIIa splicing by a mechanism requiring the 3VDE. Second, Py-ST activation of IIIa splicing appears to be PP2A independent (S. Fan, E. Boija and G. Akusjärvi, manuscript in preparation). Another viral protein that would be particularly interesting to test as a homolog to the E4-ORF4 protein is the HIV Vpr protein. Both Vpr and E4-ORF4 proteins bind to the B55 subunit of PP2A, activate PP2A activity and induce growth arrest of G2/M phase (104, 125). Although SV40 ST, in contrast to Vpr and E4-ORF4, binds to the core enzyme of PP2A and inhibits PP2A activity against many substrates, SV40 also induces cell cycle arrest at G2/M. Potentially, all three proteins acts by the same mechanism in causing G2/M arrest. Future studies should be aimed at investigating if also SV40 ST and Vpr has a role as regulators of splicing.

There is a cellular protein that like E4-ORF4 inhibits ASF/SF2 as a splicing regulatory protein. We have shown that the ASF/SF2-associated protein p32 regulates splicing by inhibiting ASF/SF2 RNA binding and phosphorylation (198). Thus, both E4-ORF4 and p32 induce accumulation of hypophosphorylated ASF/SF2. However, they do this by different mechanisms. E4-ORF4 induces SR protein dephosphorylation, while p32 inhibits SR protein phosphorylation by binding to and sequestering ASF/SF2. Another similarity between p32 and E4-ORF4 is that p32 also
interacts with SRp30c. Importantly, p32 does not inhibit SRp30c as a repressor of IIIa splicing. Potentially, the same could be true for E4-ORF4 interaction with SRp30c. The functional difference between ASF/SF2 and SRp30c should be further characterized. Interestingly, as with PP2A, there are many viral proteins that interact with p32. In addition to the previously mentioned HIV-Rev and HPV-ICP27 proteins (18, 237), the Ad-pV (158), HIV-Tat (266), EBV-EBNA 1 (35) and HSV-Orf-P (16) proteins all interact with p32. Although the significance of these interactions are currently unknown, they could potentially influence RNA splicing.

The viral factor(s) causing SR protein dephosphorylation in vaccinia virus infected cells is(are) currently not known. As described in "Present investigation", one candidate protein is the virus encoded phosphatase VH1. In future studies it should be investigated whether there actually is an interaction between VH1 and SR proteins. Furthermore, purifying SR proteins from VH1 mutant vaccinia virus infected cells may also yield interesting results. However, such an experiment could be difficult since VH1 is required for viral gene expression. An alternative strategy to study the effect of VH1 on SR protein dephosphorylation would be to investigate SR protein localisation in the nucleus. As previously mentioned, SR protein phosphorylation have been shown to cause redistribution of SR proteins from speckels in the nucleus (22, 167).

**E4-ORF4 and apoptosis**

E4-ORF4 induces p53-independent apoptosis in transformed cells (135, 156, 222) by binding to the B55α subunit of PP2A (155, 223, 224). It is not known what substrates are important for E4-ORF4-induced apoptosis, or what functions of E4-ORF4 contribute to induction of apoptosis. E4-ORF4-induced apoptosis is not required for the viral infection, and is possibly a byproduct of its other activities. E4-ORF4-induced apoptosis is anyway not harmful for the virus, since the virus encodes for several proteins that inhibit apoptosis: the E1B proteins and E4-ORF6. It has been suggested that E4-ORF4, together with the adenovirus death protein (E3-11.6K), could be responsible for lysis of the infected cell at the end of the infectious cycle (210).

E4-ORF4 induces apoptosis either by up/down-regulating the expression of specific proteins that cause apoptosis or by regulating the phosphorylation levels of kinases or other proteins important in signal transduction during apoptosis. E4-ORF4 might regulate protein expression by influencing transcription factor activity or inducing alternative splicing of genes involved in the apoptotic pathway. Recent studies seem to favour the latter mechanism. Many genes involved in apoptosis are alternatively spliced producing products with opposite effects in apoptosis, for example Bcl-X, caspase-9, caspase-2 and CED-4 (109). Interestingly, E4-ORF4 has been shown to induce alternative splicing pattern of caspase-2 in 293 cells (R. Aravalli and G. Akusjärvi, unpublished). In addition, SR proteins have been found to regulate specific splicing patterns of the caspase-2 (110) and activation of Fas-receptor mediated apoptosis results in SR protein dephosphorylation (29).

E4-ORF4 binds to several B subunits of PP2A (224). Potentially the interaction of E4-ORF4 with the various PP2A subpopulations could contribute to E4-ORF4s multiple functions. Future studies should be aimed at investigating which B subunit(s) that is(are) involved in E4-ORF4s activity in splicing and transcription. This may provide a more direct link between these events and E4-ORF4-induced apoptosis.
ACKNOWLEDGEMENTS

This work was carried out at the Department of Biochemistry and Microbiology, section Microbiology at the Uppsala University.

Jag skulle vilja ge ETT STORT TACK OCH EN JÄTTEKRAM till alla som på något sätt bidragit till att denna avhandling blev till:


To all former and present members of the ”adenogroup”. Thank you for your contribution to a stimulating and entertaining environment both in the lab and at other places. THANK YOU for all your help!!! Former: Edyta Bajak, Shaoan Fan, David Huang, Arvydas Kanopka, Mattias Männervik, Magnus Molin, Oliver Mühlemann, Svend Petersen-Mahrt, Anne-Christine Ström, Baigong Yue and Karin Öhman-Forslund. Present: Raj Aravelli, Saideh Berenjian, Elisabet Boija, Vita Dauksaite, Cecilia Johansson, Fredrik Granberg, Peter Kreivi, Anette Lindgren, Martin Lützelberger, Tanel Punga, Josef Seibt, Anders Sundqvist, Catharina Svensson, Regine Wagner, Bin Wang, Hongxin Zhao and Christina Öhrmalm. Thank you for broadening my knowledge and for always being there, helping with all kinds of things. Thank you very much for all your ideas, it contributed a lot! Many hugs to the ”lunch-group” for discussions about everything from science to life-outside-BMC. Thank you, my roommates, for sharing the office and the computer with me (perhaps not so much ”sharing” lately, but anyway).

To Göran Magnusson, Stefan Schwarz and their groups. Thank you for all your help and your contribution to the stimulating and entertaining environment in the lab!

Till alla övriga anställda på IMBIM, bland andra Erika Enström, Tony Grundin, Ylva Jansson, Kerstin Lidholt, Olav Nordli and Maud Pettersson. Tack för er ovärderliga hjälp!


Till mina föräldrar: Tack för att ni alltid har stöttat mig i vått och torrt. Ni är helt otroliga! Tack även hela ”tjocka släkten” för att ni finns där.

Slutligen till min Per. Du är det bästa och roligaste jag vet. Därför är det ju fantastiskt att det snart kommer att finnas en till av dig…. Tack för all service den senaste tiden, även för kontrollläsningen av avhandlingen!
REFERENCES


61


Bemöda dig icke om att veta allt, ty då blir du okunnig om allt.
Demokritos