The Adenovirus L4-33K Protein

A Key Regulator of Virus-specific Alternative Splicing

HEIDI TÖRMÄNEN PERSSON
Dissertation presented at Uppsala University to be publicly examined in C10:301, BMC, Husargatan 3, Uppsala, Thursday, November 17, 2011 at 09:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English.

Abstract

Adenoviruses have been extensively studied in the field of gene regulation, since their genes are subjected to a tightly controlled temporal expression during the virus lifetime. The early-to-late shift in adenoviral gene expression distinguishes two completely different programs in gene expression. The adenoviral L4-33K protein, which is the subject of this thesis, was previously implicated to be a key player in the transition from the early to the late phase of infection. Here we show that L4-33K activates late gene expression by functioning as a virus-encoded alternative RNA splicing factor activating splicing of transcripts containing weak 3’ splice sites; a feature common to the viral genes expressed at late times of infection.

The splicing enhancer activity of L4-33K was mapped to a tiny arginine/serine (RS) repeat in the carboxyl-terminal domain of the protein. Also, the subcellular distribution to the nucleus with enrichment in the nuclear membrane and subnuclear redistribution to viral replication centers during a lytic infection was observed to depend on this motif. RS repeats are common features for the cellular splicing factors serine/arginine-rich (SR) proteins, which in turn are regulated by reversible phosphorylation.

We further show that L4-33K is phosphorylated by two cellular protein kinases, the double-stranded DNA-dependent protein kinase (DNA-PK) and protein kinase A (PKA) in vitro. Interestingly, DNA-PK and PKA have opposite effects on the control of the temporally regulated L1 alternative RNA splicing. DNA-PK functions as an inhibitor of the late specific L1-IIIa premRNA splicing whereas PKA functions as an activator of L1-IIIa premRNA splicing.

In summary, this thesis describes L4-33K as an SR protein related viral alternative splicing factor. A tiny RS repeat conveys splicing enhancer activity as well as redistribution of L4-33K to replication centers. Finally, DNA-PK and PKA that phosphorylates L4-33K are suggested to be novel regulatory factors controlling adenovirus alternative splicing.

Keywords: L4-33K, adenovirus, splicing, phosphorylation, localization, replication centers, L4-22K, MLTU, DNA-PK, DNA-dependent protein kinase, PKA, cAMP-dependent protein kinase, transcription sites, SR protein

Heidi Törnänen Persson, Uppsala University, Department of Medical Biochemistry and Microbiology, Box 582, SE-751 23 Uppsala, Sweden.

© Heidi Törnänen Persson 2011

ISSN 1651-6206 0346-5462
ISBN 978-91-554-8178-0
urn:nbn:se:uu:diva-159632 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-159632)
“Enthusiasm is contagious. 
Be a carrier.”

-Susan Rabin
Members of the committee

Opponent
Marie Öhman, Professor
Department of Molecular Biology and Functional Genomics
Stockholm University

Members of the committee
Pernilla Bjerling, Dr.
Department of Medical Biochemistry and Microbiology
Uppsala University

Magnus Essand, Professor
Department of Immunology, Genetics and Pathology
Uppsala University

Tommy Linné, Professor
Department of Biomedical Sciences and Veterinary Public Health
Swedish University of Agricultural Sciences

Suparna Chandra Sanyal, Docent
Department of Cell and Molecular Biology
Uppsala University

Anders Virtanen, Professor
Department of Cell and Molecular Biology
Uppsala University

Cover illustration: Kim Törmänen
List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


* These authors made equal contribution to the work

II  Östberg, S., Törmänen Persson, H., Akusjärvi, G. (2011) Serine 192 in the tiny RS repeat of the adenoviral L4-33K protein is essential for nuclear localization and reorganization of the protein to the periphery of viral replication centers. (Submitted)


Reprint of paper I was made with permission from the publisher, JBC.
Contents

Introduction .......................................................................................................................... 11
  Transcription - decoding the DNA ................................................................................. 11
    The basal transcriptional machinery .............................................................................. 12
Pre-mRNA processing ........................................................................................................ 13
  Pre-mRNA splicing .......................................................................................................... 14
  Coupling of transcription and splicing ............................................................................ 18
  Splicing factors .............................................................................................................. 18
Phosphorylation - one way of regulating protein activity. .............................................. 21
  DNA-PK ......................................................................................................................... 21
  PKA ................................................................................................................................. 22
Adenovirus ......................................................................................................................... 23
  Genome organization ....................................................................................................... 24
  Virus life cycle ................................................................................................................ 24
  The major late transcription unit .................................................................................... 27
  The L4 region .................................................................................................................. 28
  The L1 model system ...................................................................................................... 30

Present investigation - results and discussion .............................................................. 33
  Paper I ............................................................................................................................. 33
    L4-33K, an adenovirus encoded alternative RNA splicing factor ................................. 33
  Paper II ........................................................................................................................... 36
    Serine 192 in the tiny RS repeat of the adenoviral L4-33K protein is essential for nuclear localization and reorganization of the protein to the periphery of viral replication centers ................................................................. 36
  Paper III .......................................................................................................................... 39
    Two cellular protein kinases, DNA-PK and PKA, phosphorylate the adenoviral L4-33K protein and have opposite effects on adenoviral alternative RNA splicing ......................................................................................................................... 39

Concluding remarks and future aims .............................................................................. 45

Acknowledgements ........................................................................................................... 49

References .......................................................................................................................... 52
Abbreviations

3RE IIIa repressor element
3VDE IIIa virus infection-dependent splicing enhancer
3VDF 3VDE interacting factor
Ad adenovirus
Ad-NE adenovirus infected nuclear extract
AGC kinase PKA, PKG and PKC
AKAP A-kinase anchoring protein
AT adenine and thymine
C catalytic subunit
cAMP cyclic adenosine 3’, 5’ monophosphate
CAR coxsackie and adenovirus receptor
CLK cdc2-like kinase
CRE cAMP-responsive element
CREB CRE binding protein
CREM cAMP response element modulator
CTD carboxyl-terminal domain
C-terminal carboxyl-terminal
DBP single-stranded DNA binding protein, or E2A-72K
DNA-PK DNA-dependent protein kinase
DNA-PKcs catalytic subunit of DNA-PK
DSB double strand break
DSBR double strand break repair
Dscam down syndrome cell adhesion molecule
dsDNA double stranded DNA
dsRNA double stranded RNA
eGFP enhanced green fluorescent protein
eIF eukaryotic initiation factor
ESE exonic splicing enhancer
ESS exonic splicing silencer
GTF general transcription factor
HEK human embryonic kidney
HeLa-NE HeLa cell nuclear extract
hnRNP heterogenous nuclear ribonucleoprotein
IGC interchromatin granule cluster
ISE intronic splicing enhancer
ISS intronic splicing silencer
Introduction

Animals and plants are examples of eukaryotic organisms. Physiologically they contain various set of distinct tissues created by specialized cells. For a specific species, each cell holds the same set of genes but the genes are differentially expressed generating cells with diverse functions. Gene expression is a highly regulated process where some genes never get activated in certain cell types while they are constitutively expressed in others. Also, genes can be differently expressed during the lifetime of a cell.

The final outcome of an expressed gene is usually a protein but can also be other types of molecules needed in the cell, like non-coding RNAs. The production of a protein in the cell was formulated by Francis Crick in 1958, and revised in 1970, in the so-called “Central Dogma of Molecular Biology” (41, 42). This describes how DNA is decoded into RNA that is further translated into protein. Separately these processes are extremely complex and each one is regulated at several levels, which add up to the complexity of the eukaryotic organism.

Eukaryotic cells are targets of virus infection. Viruses consist of an RNA or DNA genome packed into a protein capsid. Some viruses also contain an envelope derived from the host cell plasma membrane. To replicate, virus make use of cellular machineries implicated by the central dogma as intracellular parasites. They reprogram the cell and interfere with the complex processes described above to hinder the host cell gene expression to favor virus gene expression needed to produce new progeny viruses.

Adenovirus gene expression is subjected to a strict temporal regulation. At early times of infection, some genes are active whereas other genes are predominantly expressed at late times of infection. This shift in gene expression is regulated at many levels including pre-mRNA splicing, a phenomenon that will be described in more detail in this thesis.

Transcription - decoding the DNA

The first step in gene expression involves the transfer of information present on DNA into an RNA complement through the process of transcription. Transcription is performed by a DNA-dependent RNA polymerase (RNAP), which is an evolutionary conserved enzyme (40, 214). Common structural and functional features between RNAPs from all three kingdoms of life (bac-
teria, archaea and eukaryotes) suggest that the last universal common ance-
tor harbored an RNAP (75). Three eukaryotic RNAPs exist; RNAP I, II and III. All three can transcribe non-coding RNAs (ribosomal RNA (rRNA), small nuclear RNA (snRNA), micro RNA (miRNA), transfer RNA (tRNA)), however each one can only produce certain types (Table 1) (199). RNAP II is the only RNA polymerase that transcribes a protein-coding RNA, an RNA molecule that acts as the messenger of genetic information, the messenger RNA (mRNA) (Table 1). In eukaryotes, the mRNA is transferred from the nucleus to the cytoplasm where it associates with ribosomes, where after translation into proteins occurs.

Table 1. The RNAP types and their transcribed RNAs.

<table>
<thead>
<tr>
<th>RNAP type</th>
<th>Transcribed RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>rRNA</td>
</tr>
<tr>
<td>II</td>
<td>mRNA, snRNA, miRNA</td>
</tr>
<tr>
<td>III</td>
<td>tRNA, rRNA, snRNA, miRNA, and other small RNAs</td>
</tr>
</tbody>
</table>

The basal transcriptional machinery

The transcriptional machinery recognizes the start site of a gene based on certain sequence elements in the DNA, altogether called the promoter region. There are at least three features common to most promoters for protein-coding genes;
1) the transcription start site, from where the first nucleotide is encoded (+1). Some promoters also contain an initiator site, which is a pyrimidine-rich sequence spanning from nucleotide -3 to +5.
2) the TATA-box; an adenine and thymine (AT) -rich sequence located approximately 25-30 base pairs upstream of the start site of transcription which is responsible of setting up the site of initiation by melting the DNA helix.
3) cis-acting regulatory sequences, which serves as binding sites for transcriptional regulators (enhancers or repressors) (reviewed in (109)).

The human RNAP II consists of 12 subunits (Rbp1 to Rbp12). RNAP II is unique among the DNA-dependent RNA polymerases because its largest subunit, Rbp1, contains a long repetitive carboxyl-terminal domain (CTD) (38). The CTD consists of a tandem heptapeptide repeat (52 times in mammals); Tyr\textsuperscript{1}-Ser\textsuperscript{2}-Pro\textsuperscript{3}-Thr\textsuperscript{4}-Ser\textsuperscript{5}-Pro\textsuperscript{6}-Ser\textsuperscript{7} (6, 38). The CTD can be phosphorylated at several sites in this heptapeptide where phosphorylation of serines at position 2 (Ser\textsuperscript{2}) and 5 (Ser\textsuperscript{5}) is tightly coupled to RNAP II function during transcription initiation and elongation (53, 101). In addition, serine at position 7 (Ser\textsuperscript{7}) also seems to play a role in transcription of protein-coding as well as non-coding genes, however the exact role is not established (32, 49).
RNAP II cannot initiate transcription on its own; it needs the general transcription factors (GTFs) TFIIA, B, D, E, F and H. The GTFs assemble on the promoter in a sequential manner and together with RNAP II and DNA they form the pre-initiation complex (PIC) (150, 166). Transcription is initiated when DNA melts at the TATA-box and the two first ribonucleotides are coupled via a phosphodiester bond. The elongating transcript, or precursor (pre)-mRNA, extrudes from the RNAP II in close proximity to the CTD. Further, the CTD acts as a scaffold, interacting with the pre-mRNA and introducing the processing factors needed during the maturation of pre-mRNA (83).

The phosphorylation pattern of the CTD heptapeptide changes during transcription. Phosphorylation of Ser$^5$ is detected primarily at the promoter region whereas, phosphorylation of Ser$^2$ is seen mostly in coding regions (33, 101). Thus, transition from the initiation to the elongation stage depends on phosphorylation of Ser$^5$ and the processivity of transcription elongation depends on phosphorylation of Ser$^2$. It has also been suggested that RNAP II is recycled for faster transcription initiation. Once RNAP II reaches the termination signal at the end of transcription, it is released from the template at the same time as CTD is dephosphorylated. Dephosphorylation of the CTD allows the RNAP II to recycle and restart initiation resulting in multiple-round transcription (116).

Pre-mRNA processing

The transcribed RNA is not a functional template for protein synthesis until it has undergone certain modifications in specific pre-mRNA processing steps. These processes are co-transcriptional events, which means that they occur during the production of the transcript. Pre-mRNA processing factors assemble on the RNAP II CTD to assist in the conversion of the transcript into a mature mRNA (83). Pre-mRNA processing includes the addition of a 7-methyl guanosine (m$^7$G) cap at the 5’-end, removing non-coding introns, and 3’-end formation by cleavage and polyadenylation to produce a poly(A) tail. The 5’- (m$^7$G) cap and the 3’-poly(A) tail protects the mRNA from degradation by exonucleases, facilitating transport from the nucleus to the cytoplasm and further promotes recognition of the mRNA by the ribosome (35, 58, 61, 169, 194, 196). In higher eukaryotes, 90% of the protein-coding genes are interrupted by one or more intervening non-coding sequences, the introns (197). To make a functional mRNA, introns have to be excised from the transcript and the coding exons joined together before the matured mRNA can be transported to the cytoplasm. This RNA processing step is called pre-mRNA splicing, which presence and complexity distinguishes complex organisms from less complex ones.
Pre-mRNA splicing

In 1977, Phillip Sharp and Richard Roberts discovered the split gene concept while working with adenovirus mRNA structures (17, 34). They were interested in mapping the position of individual mRNAs on the adenovirus chromosome. During this work they discovered that the mRNAs expressed from the adenoviral major late transcription unit (MLTU) lacked some internal sections present in the viral DNA genome. They concluded that the mRNAs contained discontinuous segments derived from multiple places from the viral genome. Soon after this discovery, scientists working in other systems were able to show that the split gene concept was not unique to the adenovirus system (114). They were awarded the Nobel Prize in Physiology or Medicine in 1993.

Splicing is performed by a dynamic and multimegadalton ribonucleoprotein (RNP) complex called the spliceosome. It excises introns with single nucleotide precision and sequentially joins the coding sequences (exons) together, Fig. 1. Splicing of introns is a complex task since a common human gene consists on average of 9 exons with a mean size of 145 nucleotides separated by introns of at least ten times the size (106). There are two classes of introns; the AT-AC and the GU-AG, named after the dinucleotide pairs present at the beginning and the end of the intron. These two types of introns are removed by two different spliceosomes, the minor U12- and the major U2-dependent spliceosome, respectively. The work presented in this thesis is only based on the major U2-dependent spliceosome.

The spliceosome and the catalytic steps of splicing

The spliceosome is formed by the five uridine-rich small nuclear ribonucleoproteins (U snRNP); U1, U2, U4, U5 and U6, each consisting of one specific uridine-rich small nuclear RNA (U snRNA) molecule, seven common Sm-proteins and a variable number of particle specific proteins (27, 201). The spliceosome assembles de novo in a stepwise manner on each new intron to be spliced out. Certain sequence elements in the pre-mRNA provides signals to the spliceosome where the introns (and consequently the exons) are located. The defined general splice sites are located at the exon-intron border (splice donor or 5’ splice site) and the intron-exon border (splice acceptor or 3’ splice site). The branch point is located 18-40 nucleotides upstream of the 3’ splice site and contains an adenine surrounded by more or less conserved sequences, Fig. 1. Between the branch point and the 3’ splice site a 10-15 nucleotides long pyrimidine sequence is positioned (reviewed in (201)).
Figure 1. (Upper panel) Conserved sequences found at the exon-intron (5’ splice site) and the intron-exon (3’ splice site) borders and branch point of U2-type pre-mRNA introns. Exons are indicated by filled boxes and the intron by a thin line, unless specific nucleotide is specified. Y = pyrimidine (C or U), R = purine (G or A) and N = any nucleotide (A, U, G or C). The polypyrimidine tract is indicated by (Y)_{10-15} and the branch point adenine is underlined (A). (Lower panel) Spliceosome assembly showing the U2AF heterodimer and the U snRNPs (U1, U2, U4, U5 and U6) assembling in a stepwise manner on the RNA. Formation of A, B and finally the catalytic C complex, where the transesterification reaction is completed, and the intron lariat and mRNA product are produced.
The spliceosome assembly, Fig. 1, starts when the U1 snRNP recognizes and binds to the 5’ splice site via base pairing of the U1 snRNA. The heterodimer U2 snRNP auxiliary factor (U2AF) recognizes and binds to the poly-pyrimidine tract upstream of the 3’ splice site consequently recruiting the U2 snRNP to the branch point. After an ATP-dependent interaction between the branch point and the U2 snRNA, the spliceosomal A complex is assembled. Following recruitment of the U4/U6.U5 tri-snRNP, the U1 snRNP leaves the complex and major structural changes occur, forming the B complex. The U4/U6 snRNA base paring opens up the RNA enabling the U6 snRNA to interact with the 5’ splice site and the U2 snRNA to form the catalytic core of the active spliceosome, also referred to as the C complex (reviewed in (27, 201)). The biochemical reactions of splicing involve two transesterification steps. The first step is initiated by a nucleophilic attack of the 2’ hydroxyl group of the branch point adenosine on the 3’- 5’ phosphodiester bond at the 5’ splice site, resulting in a free 5’ exon and a lariat structure where the 5’ end of the intron is covalently linked to the branch point adenosine. In the second step, the free hydroxyl group of the 5’ exon performs a nucleophilic attack on the 3’ splice site resulting in the final splicing products; a lariat structured intron and the mature mRNA product with ligated exons, 5’ cap and 3’ poly(A) tail.

The selection of a specific splice site relies in part on binding of trans-acting splicing factors to cis-acting elements in the pre-mRNA. Metazoans usually assemble spliceosomes across exons, because it is easier to define a shorter exon as compared to the very long introns. According to the exon definition model, the downstream 5’ splice site is recognized by the U1 snRNP and the upstream 3’ splice site (pyrimidine and branch point) is bound and recognized by U2AF and U2 snRNP. Serine/arginine-rich (SR) proteins mediate interactions between the sites and the exon is defined. These cross-exon interactions must change into cross-intron interactions before continuation to the later stages of spliceosome assembly can occur (195).

**Constitutive and alternative splicing**

Whereas some exons are constitutively included, meaning that they are present in every mRNA produced from a particular pre-mRNA, others are alternatively included to create variable forms of mRNA from one pre-mRNA. Alternative splicing makes it possible to produce more than one protein (gene product) from a gene by combining exons in variable ways during pre-mRNA splicing. The most extreme example of producing alternatively spliced mRNA isoforms comes from the *Drosophila melanogaster* gene Down syndrome cell adhesion molecule (Dscam), which can generate approximately 38,000 different mRNA isoforms from one pre-mRNA (173).

Alternative splicing is a way of reducing the amount of genetic material, avoiding deleterious mutations, promoting high diversity of proteins and allowing for an additional level of gene regulation. Unicellular eukaryotic
organisms, like yeast, splice their pre-mRNAs in a constitutive way while higher multicellular organisms use alternative splicing. Deep sequencing analysis of human transcriptomes reveal that more than 90% of all human genes codes for RNA which can undergo alternative splicing (154, 197). Regulation of alternative splicing is tissue specific, and varies within a cell depending on e.g. the phase of the cell cycle, differentiated stage, and foreign stimuli like virus or bacterial infections. The basic mechanisms responsible for the production of alternatively spliced mRNAs are alternative 5’ splice site and 3’ splice site selection, cassette-exon inclusion or exon skipping, and intron retention, Fig. 2 (reviewed in (147)). One or several of the different features can be used together or separately to splice a particular pre-mRNA.

![Figure 2](image.png)

**Figure 2.** Four types of alternative splicing; 5’ splice site and 3’ splice site selection, cassette-exon (exon skipping or exon inclusion), and intron retention. To the left, the pre-mRNA where boxes represent exons, thick lines introns, and thin hooked lines the region, which is spliced out. The structure of the mRNAs synthesized after the alternative splicing events are shown to the right.

**Splicing regulation at different levels**

Alternative splicing is regulated at many levels, from the pre-mRNA sequence level to the level of chromatin structure. The first level is directly at the pre-mRNA sequence level, where splicing factors bind and recruit the splicing machinery or hinder the recruitment by sterically blocking splice site usage. The next level of regulation deals with the coupling of splicing to RNAP II transcription of the pre-mRNA template and/or kinetics of transcription and RNA processing. It has been shown that if the rate of transcription slows down exon inclusion is favored e.g. UV-light induced hyperphosphorylation of RNAP II CTD is suggested to inhibit transcription elongation.
and thus promote exon inclusion (47, 139). Also, the spliceosomal component levels, intron size and competition between splice sites can determine the outcome of exon usage (reviewed in (147)). Further, the last level of alternative splicing regulation occurs via the chromatin structure and histone marks, where histone modifications, like trimethylation of histone 3 at lysine 36, mark exons to be used (100, 118).

Coupling of transcription and splicing

It is known that splicing events occur co-transcriptionally and that the splicing machinery is physically linked to the transcriptional apparatus through association of splicing factors with the CTD of the RNAP II (156). In 1997, the group of Bentley first suggested a coupling between RNAP II transcription and pre-mRNA splicing by showing that a deletion of the CTD from RNAP II impaired processing of the RNA (126). They also showed that other post-transcriptional processes like 5’-end capping and 3’-end cleavage and polyadenylation was dependent on the presence of the CTD (125, 126). It is now known that the major function of the CTD is to serve as a binding scaffold for capping, splicing, and polyadenylation factors, and in this way physically couple pre-mRNA processing to transcription (158). This means that the CTD plays a major role in coordinating a number of nuclear RNA processing events with RNA chain synthesis and the translocation of RNAP II along a gene.

As well as the phosphorylation status of the CTD can influence the initiation and elongation of transcription it can also influence the pattern of alternative splicing (13). How the spliceosome is assembled co-transcriptionally is not known. However, splicing is a much faster process when coupled to transcription. Exons seem to be tethered to the RNAP II during elongation and this might be one way of controlling the re-cycling of the spliceosomes, generating a more efficient progression (reviewed in (156)).

Splicing factors

There are several models existing to describe how the spliceosome finds the place to perform its catalytic activity. The most studied splicing regulators are the members of the SR protein and heterogeneous nuclear ribonucleoprotein (hnRNP) families. They have been reported as splicing activators and repressors, respectively. However, the reverse function has also been reported (91, 124). These regulators are known to bind to cis-acting elements either located in exons, called exonic splicing enhancers (ESEs) or silencers (ESSs), or in introns, called intronic splicing enhancers (ISEs) or silencers (ISSs), and thus either to promote or inhibit splicing. They are the main trans-acting factors in the yin-yang model of alternative splicing (60, 147),
where the number of positively acting sites or negatively acting sites determines the splice site usage.

The SR proteins

SR proteins contain one or two RNA recognition motifs and a carboxyl terminal (C-terminal) arginine- and serine-rich domain (RS domain). Constitutive splicing requires SR proteins, but they can also regulate alternative splicing. They recognize and bind enhancer sequences to recruit spliceosomal components like U2AF, U1-70K or tri-snRNP to the splice sites (69). In addition, SR proteins can inhibit the binding of splicing repressor proteins to their silencer elements by binding to enhancer elements in close proximity (184). SR proteins together with the U1 snRNP have also been shown to associate with RNAP II, suggesting an involvement in the coupling of transcription to splicing via the RNAP II CTD (45). Further, they can act as bridges bringing splice sites in close proximity. The RS domain functions as a protein-interaction domain, responsible for the contact with other proteins like SR proteins or SR-related proteins and also with spliceosomal components (20, 72, 99, 189). Furthermore, the subcellular localization of SR proteins is controlled by the RS domain, which acts as a nuclear localization signal (28, 93, 105).

SR proteins are functionally regulated by post-translational modification such as phosphorylation (92). The activity and the subcellular distribution of SR proteins are coordinated by SR-specific protein kinase (SRPK) and cdc2-like kinase (CLK) families of kinases that phosphorylate serines in the RS domain (69). Phosphorylation and dephosphorylation regulate the activity of SR proteins e.g. phosphorylated SR proteins promote splicing by recognizing splice sites. Furthermore, they become dephosphorylated, which drives splicing catalysis within the activated spliceosome. The level of phosphorylation is important for their functions. Thus, hyper- or hypophosphorylation of the RS domain make the SR proteins unable to support splicing (reviewed in (72, 177)). Further, SR proteins are involved in other cellular processes like RNA export, non-sense mediated mRNA decay, and translation. SR proteins have been named according to various features like function, molecular weight and even identification of a recognizing antibody, which have concluded in a non-unified nomenclature. Recently, the nomenclature for the twelve SR proteins in human was revised according to the SR protein human genome organization names, SR splicing factor (SRSF), the new names are listed in Table 2 (122).
Table 2. SR protein/gene symbols.

<table>
<thead>
<tr>
<th>New protein/gene symbol</th>
<th>Old names</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRSF1</td>
<td>ASF, SF2, SRp30</td>
</tr>
<tr>
<td>SRSF2</td>
<td>SC35, PR264, SRp30b</td>
</tr>
<tr>
<td>SRSF3</td>
<td>SRp20</td>
</tr>
<tr>
<td>SRSF4</td>
<td>SRp75</td>
</tr>
<tr>
<td>SRSF5</td>
<td>SRp40, HRS</td>
</tr>
<tr>
<td>SRSF6</td>
<td>SRp55, B52</td>
</tr>
<tr>
<td>SRSF7</td>
<td>9G8</td>
</tr>
<tr>
<td>SRSF8</td>
<td>SRp46</td>
</tr>
<tr>
<td>SRSF9</td>
<td>SRp30c</td>
</tr>
<tr>
<td>SRSF10</td>
<td>TASR1, SRp38, SRp40</td>
</tr>
<tr>
<td>SRSF11</td>
<td>p54, SRp54</td>
</tr>
<tr>
<td>SRSF12</td>
<td>SRp35</td>
</tr>
</tbody>
</table>

The U2AF factor

The heterodimeric U2AF factor is an example of an SR-related protein. It is composed of two subunits, U2AF$^{65}$ and U2AF$^{35}$, which have been shown to bind the pyrimidine tract and the AG dinucleotide at the 3’ splice site, respectively (96, 204, 212). U2AF is a vital splicing factor required to recruit U2 snRNP to the branch point during the first steps of spliceosome assembly, even though U2AF$^{35}$ is reported not always to be present in the dimer (77, 168). U2AF$^{65}$ binds preferentially to long pyrimidine tracts (strong splice sites) and U2AF$^{35}$ stabilizes the interaction by binding to the AG dinucleotide. Alternative 3’ splice sites generally have short and/or interrupted pyrimidine tract (weak splice site) and usually requires U2AF$^{35}$ for recognition (204). However, U2AF$^{35}$ is not required for splicing of all weak introns, but rather regulate a specific subset (153).

Subcellular localization of splicing factors

Splicing factors are enriched in nuclear compartments called speckles, which are dynamic structures, found within the nucleoplasm. The speckles, also called the interchromatin granule clusters (IGCs) or splicing factor compartments (SFC), serves as storage and reassembly site for the splicing factors. Active RNAP II transcription and pre-mRNA splicing, however, is carried out in separate compartments called nuclear perichromatin fibrils (177, 186). As mentioned above in the section “The SR proteins”, the RS domain of SR proteins serves as a nuclear localization signal. If the RS domain is deleted, the protein is distributed not only in the nucleus but also in the cytoplasm. Also, for some SR proteins the RS domain is important for subnuclear distribution to speckles, like SRSF2 and SRSF3, while for others like SRSF1 and SRSF5, it is not (28). Some RS domains contain a nuclear retention signal, preventing the protein from exiting to the cytoplasm. In
addition, the RS domain is crucial for the shuttling SR proteins as transport through the nuclear pore is regulated by phosphorylation of the RS domain (29). Phosphorylation also regulates the recruitment of splicing factors from IGCs to the sites of active transcription, the perichromatin fibrils (129, 130). For example, after the cytoplasmic SRPK1-dependent phosphorylation of the SR protein SRSF1, it localizes to nuclear speckles (10, 105), where the CLK/Sty interacts and further phosphorylates SRSF1, which results in a relocalization of the protein to the perichromatin fibrils (37). If intron-containing gene expression or transcription increase rapidly, perhaps because of viral infection, splicing factors are redistributed from storage speckles (IGCs) to the sites of active transcription (reviewed in (186)).

Phosphorylation - one way of regulating protein activity.

Post-translational modifications expand the complexity of the proteome by producing multiple forms of a protein, for instance by attaching biochemically functional groups, such as acetates, phosphates, lipids or carbohydrates. Also, it allows for a fast cellular response to extra- or intracellular stimuli by activating or inactivating proteins already present in the cell. Reversible phosphorylation is an important regulatory post-translational modification caused by the opposing actions of kinases (phosphorylation) and phosphatases (dephosphorylation). The kinase induces a conformational change in the protein structure upon addition of a phosphate group, which leads to changes in the activity of the substrate protein. The importance of a phosphorylation as a regulatory mechanism in signal transduction is reflected by the genome content where about 2% of eukaryotic genes encode for protein kinases (123).

DNA-PK

The double stranded DNA-dependent protein kinase (DNA-PK) is a nuclear serine/threonine protein kinase that belongs to the family of phosphatidylinositol 3-kinase-related kinases (PIKKs) (112). The PIKKs are conserved through evolution with homologues in many organisms from yeast to mammals and are considered as atypical protein kinases in the sense that their kinase domain has a low sequence similarity to other classical eukaryotic protein kinases and that they rather resemble the lipid phosphatidyl inositol 3-kinase family (94, 123). The biological functions of PIKKs are diverse and involve responses to DNA damage, metabolism and cell growth control, regulation of nonsense-mediated mRNA decay and transcriptional regulation (112, 203).
Biochemical studies have shown that DNA-PK is a heterotrimeric enzyme composed of a catalytic subunit (DNA-PKcs) and two regulatory subunits Ku86 and Ku70 (71, 208). DNA-PK preferentially phosphorylates serine or threonine residues followed by a glutamine, (Ser/Thr-Gln) motif (36). Several functions have been ascribed to this complex including DNA repair, transcription regulation and maintenance of telomeres (143).

The most studied function of DNA-PK is its role in the double strand break repair (DSBR) system and non-homologous end joining (NHEJ) pathway (86). To repair double strand breaks (DSB) in the DNA, the Ku heterodimer recognizes the DSB and facilitates the recruitment of DNA-PKcs to the DNA. DNA-PK is activated upon DNA binding and its kinase activity is required for autophosphorylation of the catalytic subunit. Following autophosphorylation, DNA-PK dissociates and the DNA ends become accessible for the rest of the NHEJ components, which join and ligate the DNA ends together (30, 111, 143).

DNA-PK has also been suggested to have a direct role in transcription. It phosphorylates transcription factors like Sp1, Oct-1, c-Myc, c-Jun and p53, thereby regulating their functions (110, 172, 180). Further, DNA-PK has been shown to interact with RNAP II and also to phosphorylate the CTD, and thereby regulating the initiation step of transcription (121, 157, 203). Interestingly, DNA-PK has been shown to induce phosphorylation of Ser\(^2\) and Ser\(^7\) in the heptapeptide repeat (192). The biological relevance of this phosphorylation is unknown, but Ser\(^7\) phosphorylation seems to be enriched on introns, which could suggest a link to pre-mRNA splicing (97).

**PKA**

The cyclic adenosine 3′, 5′ monophosphate (cAMP) dependent protein kinase A (PKA) is one of the best-characterized protein kinases. It is a member of the AGC kinase family (PKA, cGMP-dependent protein kinase G (PKG), and phospholipid-dependent protein kinase C (PKC)) and transduce the signal of many extracellular factors (155). PKA is activated by the intracellular second messenger cAMP, which is produced from ATP by adenylyl cyclases as a result of extracellular stimuli, such as hormones and neurotransmitters, binding to G protein-coupled receptors in the cell membrane (127). Active PKA phosphorylates serine and threonine residues in a subset of various sequence motifs in its substrate, where the most common one is Arg-Arg-X-Ser/Thr (174). PKA phosphorylation has a variety of downstream effects, as its substrates are involved in numerous processes, among them cell differentiation, metabolism, gene expression and subcellular distribution of proteins (26, 59, 68, 70).

In the absence of cAMP, PKA is an inactive tetramer composed of a regulatory subunit (R) dimer and two catalytic subunits (C). The PKA subunits are encoded by four R and C subunit genes (RI\(\alpha\), RII\(\alpha\), RI\(\beta\), RII\(\beta\), Ca, C\(\beta\),
PKA is activated by the binding of two cAMP molecules to each regulatory subunit, releasing the C subunits to phosphorylate substrates in its vicinity. Specificity in the cAMP/PKA signal transduction pathway is gained both by tissue-specific expression of the different variants, as well as specific targeting of the R subunits to A-kinase anchoring proteins (AKAPs) in the cytosol and of the C subunit to C subunit-binding proteins both in the cytosol and nucleus. Upon activation, a proportion of the C subunits translocate to the nucleus, a redistribution that takes about 15-20 minutes making it the rate-limiting step in the cAMP-dependent transcription activation pathway. PKA regulates gene expression by phosphorylating several transcription factors, including the cAMP-responsive element (CRE) binding protein (CREB), cAMP response element modulator (CREM), and NF-κB. In addition to gene transcription, nuclear PKA regulates pre-mRNA splicing through phosphorylation of splicing factors, such as SRSF1 and by interaction with the AKAP and splicing factor SFRS17A.

**Adenovirus**

Adenoviruses belong to the family *Adenoviridae*, which host range includes all vertebrates from fish to humans. Evolutionary, adenoviruses have most probably co-speciated with their host and the most recent ancestor is likely to have existed before the divergence of bony fish to other vertebrates. Adenoviruses have not yet been discovered in invertebrates, even though they share a structural overall similarity to the bacteriophage PRD1, suggesting a pre-dated eukaryotic ancestor.

The first human adenovirus serotype was isolated from adenoids of military recruits in the beginning of the 1950s. Since then, approximately 50 serotypes have been discovered and divided into seven subgroups, A to G, based on their ability to agglutinate red blood cells, oncogenic potential, genome homology and immunological characteristics. The most studied adenoviruses are adenovirus type 2 (ad2) and 5 (ad5), both belonging to subgroup C, and subgroup A type 12 (ad12). Ad12 is oncogenic in rodents but no human adenovirus has so far been shown to cause any human cancers. Despite this, adenovirus belongs to the small DNA tumor virus group, along with papilloma- and polyomavirus and has been proven to be important for the study of cell transformation, cell cycle control, gene regulation and tumor formation. In recent years, the adenovirus has also been exploited as a potential vector in gene therapy.
Genome organization

The human adenovirus genome consists of linear double stranded DNA (dsDNA) of around 34-36 kbp in length (Fig. 3). Each end of the genome contains inverted terminal repeats (ITR), which function as origins for viral DNA replication. Covalently attached to the 5’ end of each strand is the terminal protein (TP). The genome is divided into transcriptional units, located on both strands, which products are subjected to alternative splicing and polyadenylation to generate multiple mRNAs that are translated into approximately 30-40 proteins, dependent on the serotype. The genome encodes for early (E1A, E1B, E2, E3 and E4), intermediate (IX and IVa2) and late (L1-L5) genes, which are transcribed by the cellular RNAP II machinery. In addition, two genes are transcribed by RNAP III, generating the highly structured virus-associated (VA) RNAs, VA RNA I and VA RNA II.

![Figure 3. The Adenovirus type 2 and 5 genome (kindly provided by G. Akusjärvi). Filled arrows indicated early genes and open arrows indicate late genes. The major late transcription unit (MLTU) encodes five families of mRNAs (L1-L5) with co-terminal ends. To the left in the MLTU, the position of the constitutive tri-partite leader exons (1, 2, and 3) and the i-leader exon (i) are indicated.](image-url)

Virus life cycle

**Entry**

The virus particle consists of a non-enveloped icosahedral capsid of 70 – 90 nm in diameter. The capsid has protruding fiber structures placed at the
12 vertices of the virus capsid. These fibers are used during entry when the virus attaches to the host cell receptor. Depending on the adenovirus serotype, it makes a contact with the cellular coxsackie and adenovirus receptor (CAR) or other receptor like CD46 (16, 62). Adenovirus entry is a two-step process where binding of the fiber to its primary receptor is followed by the penton base interacting with the integrin family of secondary receptors in order for the virus to enter via receptor-mediated endocytosis (200). Within the endosome the virus particle becomes partly dismantled due to cleavage by the viral L3 protease and the lower pH in the endosomal milieu. The virus is subsequently released into the cytosol by disruption of the endosomal membrane (39, 74). The capsid is transported on microtubules, to the nuclear membrane where the linear viral dsDNA genome is imported through the nuclear pore complex (73).

**Early phase**

When the genome has entered the nucleus, the immediate early E1A unit is transcribed. The E1A proteins are trans-acting transcriptional regulators of the adenoviral early genes (E1B, E2, E3 and E4) (19, 144). Some additional viral promoters are also known to be stimulated by E1A, like the major late promoter (MLP) and the recently proposed L4 promoter (L4P) (18, 135). For successful replication of the viral genome, the cell is reprogrammed by the E1A proteins to enter the S-phase. E1A binds to the cellular retinoblastoma tumor suppressor protein (RB) releasing the cellular E2F transcription factor, which turns on a subset of cellular genes involved in S-phase activation (89). In addition, E1A confers an epigenetic reprogramming to restrict the acetylation of histone 3 at lysine 18 to a limited set of genes, thereby stimulating entry to S-phase (54). These events also trigger the activation of the tumor suppressor p53, which would prevent S-phase entry and also promote apoptosis if not bound and inactivated by the adenoviral E1B-55K protein (170). In addition, a p53-independent activation of apoptosis, through repression of NF-κB mediated transcription, is prevented by the E1B-19K protein, which inhibits Bak and Bax protein dimerization (43). Consequently, the E1A and the E1B proteins are the viral oncogenes that are necessary for adenovirus-induced transformation.

The E4 region codes for proteins involved in efficient viral DNA synthesis (21, 82, 198), viral late mRNA metabolism (splice site selection (50, 149), stabilization, export (22, 23)), disruption of promyelocytic leukaemia protein nuclear domains (PML-ND) (113), inhibition of cellular mRNA export, degradation of host cell proteins (115, 163), and also, histone modifications (185). The E4-ORF4 protein binds to the cellular serine/threonine specific protein phosphatase 2A (PP2A) and induces hypophosphorylation of cellular proteins and of interest for this thesis, regulating adenovirus alternative RNA splicing (reviewed in (2)). The E4-ORF3 and E4-ORF6 proteins
also serve a function in preventing genome concatenation by targeting DNA-PK in its role as a double strand break repair factor (21, 82, 198).

Mammals have innate and adaptive defenses that protect them from virus infections. The viral E3 unit encodes for multiple proteins that counteract the antiviral defense of the host. The E3 proteins rescue infected cells from cytotoxic T lymphocytes (E3-gp19K), prevent tumor necrosis factor cytolysis (E3-14.7K, E3-10.4K/14.5K), and down-regulate the epidermal growth factor receptor (E3-10.4K/14.5K). However, they are not essential for virus growth in tissue culture cells. Upon infection, the innate immune response produces interferons, which induce the production of the double stranded RNA (dsRNA)-activated protein kinase (PKR). dsRNA produced as a by-product of the symmetrical transcription of the viral genome, activates PKR, which inhibits protein translation by phosphorylating the eukaryotic initiation factor 2α (eIF2α) (67). To protect the virus protein production, the highly structured VA RNAs bind to PKR and repress its inhibitory function on translation (63, 190). In addition, these non-coding VA RNAs are targeting the RNA interference (RNAi)/microRNA pathways by saturating the system with virus-derived small RNAs (7, 117, 207).

**DNA replication and replication centers**

Proteins encoded from the E2 transcription unit are required for replication of the viral genome. E2 codes for three proteins in two transcription units; the single-stranded DNA binding protein (DBP) is encoded from the E2A unit, and the viral DNA polymerase (Ad-Pol) along with the precursor terminal protein (pTP) are encoded from the E2B unit. The viral DNA polymerase forms a complex with the pTP, which serves as a protein primer for initiation of DNA replication and initiates replication at either end of the duplex adenovirus DNA molecule (141). pTP becomes covalently attached to the produced 5’ termini of the linear genome and is later processed by the viral protease to form a smaller terminal protein (TP) that stays attached to the genome through the rest of the viral life cycle. The DBP binds to the displaced DNA strand and confer processivity to the viral DNA polymerase (193). In addition, some cellular proteins like NF-I, NF-III/Oct-1, pL and the topoisomerase NFII are needed (95, 138, 142). Replication is localized to the nucleus to defined sites called replication centers. By immunostaining with an antibody directed against DBP these centers can be visualized as small spots increasing in amount over the course of the infection to finally spreading throughout the whole nucleus (160, 161).

**Late phase**

The onset of viral DNA replication is the restriction point that distinguishes the early phase from the late phase of gene expression. Infection causes a major reorganization of splicing factors, which are first recruited from nuclear speckles to ring-like structured regions surrounding the replication
centers, the so-called peripheral replicative zones (25, 162). In the late phase of infection, when the viral mRNAs starts to accumulate in the cytoplasm, the snRNPs redistribute from the peripheral replicative zones to enlarged nuclear speckles that are distributed to the outer regions of the nucleus (9). The redistribution of splicing factors to enlarged speckles is shown to depend on proteins encoded by the early E4 unit (24).

The mechanism of spatial gene regulation ensures that proteins needed at a certain stage of virus infection are produced at the right time point. The large majority of the late genes are transcribed from the MLTU, which encodes some regulatory proteins but mostly packaging, assembly and capsid proteins. However, in the late phase, the host cell is reprogrammed to produce only viral proteins from the MLTU. The cellular protein production is turned off, because of a viral-induced inhibition of phosphorylation of the CAP-binding eIF4E translation factor (44, 84). Under these conditions, the late viral mRNAs are still recognized by the ribosome, because of their 5’ non-coding region, the so-called the tripartite leader, which initiates translation by an unusual ribosome shunting mechanism (210, 211). The produced viral proteins are transported to the nucleus and assembled into immature capsids. The L4-100K and L1-52,55K proteins are chaperones needed for correct capsid assembly. The packaging signal in the genome is bound by the IVa2 and L4-22K proteins, which are required for genome-specific packaging into the virion. As the final step in maturation, the virus-encoded and encapsidated protease performs numerous cleavages of multiple viral proteins to generate the mature, infectious virion (57). The adenoviral death protein, expressed from the E3 region in the late phase of infection, acts in promoting cell lysis and virus release (202).

The major late transcription unit
Activation of adenoviral gene expression is under a strict temporal control with an expression of early genes directly after infection, and late genes after the onset of viral genome replication. Studies of the MLTU has shown that the early-to-late switch in gene expression is controlled at the level of premature transcription termination, alternative splicing and poly(A) site choice.

The MLTU produces a primary transcript of approximately 28,000 nucleotides. This transcript can be polyadenylated at five different positions, producing five families of mRNAs with co-terminal 3’-ends, L1 to L5 (Fig. 3). Transcription of the MLTU is initiated at the MLP, which is activated within the first hour of infection and is expressing short transcripts gradually preterminating between the L1- and the L3 poly(A) sites (175, 209). Even though a low percentage of transcription proceeds beyond the L3 poly(A) site, only the L1 poly(A) site is efficiently used at the early phase of infection. At this stage, only the L1 52,55K mRNA accumulates in the cytoplasm.
After the onset of viral DNA replication transcription is extended and proceeds through to the L5 poly(A) site (3, 145). The use of the L2-L5 poly(A) sites are temporally divided into two phases. The intermediate phase, before the onset of late protein synthesis, gives a selective activation of the L4 poly(A) site (108). The late phase, when late protein synthesis has commenced, results in the activation of the L2, L3 and L5 poly(A) sites (107, 108). In addition, some of the L4 family members seem to be expressed from the recently proposed L4P, which might explain the selective intermediate expression of L4 mRNAs (135).

The boost in mRNA accumulation seen late in infection is not only due to a shift in poly(A) site usage or enhancement of the activity of the MLP, but is to a major part due to a temporal shift in alternative splicing. The MLTU transcripts use alternative splicing to produce approximately 20 different mRNAs. All spliced mRNAs contain a common 5’ non-coding sequence, called the tri-partite leader, which is required for the cap-independent translation (48). This leader sequence is made by joining the three first exons and some times also a fourth exon, the i-leader exon, located between exon two and three. The i-leader exon codes for a protein of unknown function that is expressed at intermediate times of infection (4). The mRNAs produced from the L1-L5 units are derived by alternative 3’ splice site usage. A common 5’ splice site at the third exon in the tripartite leader is spliced to multiple alternative 3’ splice sites within each unit. The L1 unit is the only unit in MLTU producing mRNAs both early and late after infection. All other units produce mRNAs only after the onset of viral DNA replication and all, except the L4 mRNAs, require late viral protein synthesis. This suggests that one or several viral factors inducing the final accumulation of late mRNAs might be encoded from the L4 region (107, 188).

The L4 region

The L4 family of mRNAs is expressed after the onset of genome replication, but prior to the initiation of late protein synthesis (108). L4 mRNAs expressed from the MLP share the common 5’ tripartite leader sequence, which is absent in the L4P expressed mRNAs (135). The L4 mRNAs differ in their 3’ terminus, because of alternative 3’ splice site selection. The L4-33K protein is encoded by a spliced version of the last exon while the L4-22K protein is encoded by an unspliced version (151, 152), Fig 4.

The only structural protein expressed from the L4 region is the pVIII protein, which is one of the major accessory proteins in the virus capsid, which connects the viral core to the hexons in the virus particle (165). The largest L4 protein, 100K, acts as an assembly and scaffolding protein, which is required for proper hexon folding and trimerization (31). It also has a major role in inhibiting cellular protein translation by displacing the cellular MAPKAP/p38 kinase (Mnk1) from the binding to eIF4G and in that way
blocking phosphorylation of eIF4E. This results in a destabilization of the interaction between a capped mRNA and the initiator complex (44). In addition, the L4-100K recruits the eIF4G to the 5’ tripartite leader sequence of the late viral mRNAs and forms a complex together with the poly(A) binding protein (PABP) and thus stimulate translation of the viral mRNAs by an alternative mechanism called ribosome shunting (205, 211). During the late 1970s it was thought that the L4-100K and L4-33K proteins share an extensive amino acid homology since the proteins have overlapping coding regions as well as both being phosphoproteins (11, 65). However, it was later discovered that the L4-33K protein is produced from a +1 frame shift compared to L4-100K in the coding region (151).

L4-33K is a nuclear phosphoprotein that previously was assigned a role in virus assembly (11, 65). Introduction of a stop codon at the amino-terminal (N-terminal) end of the protein resulted in a reduction in virus yield and introduction of stop codons in the C-terminal part of the protein resulted in a complete failure in the production of infectious particles (55, 56). In addition, L4-33K has also been suggested to bind to the packaging domain of the viral genome (5). The importance of the intermediate expression of L4 mRNAs was not acknowledged until L4-33K was suggested to have a role in the final boost of adenoviral late mRNA production (52). It was shown that L4-33K can increase the cytoplasmic accumulation of late mRNAs, thus being a key component of the early-to-late switch in adenovirus late gene expression (52).

L4-33K is translated from a spliced mRNA where an intron interrupting the protein coding sequence has been excised, whereas the L4-22K protein is translated from the corresponding unspliced mRNA. Thus, the two proteins share a common N-terminus (the first 106 amino acids) but have unique C-terminal ends. L4-22K has been suggested to have a role in virus assembly and, just like L4-33K, it has been shown to bind the packaging sequence of the viral genome (51, 102, 152). The similarities in function might be due to the common N-terminus, but this remains to be investigated. Recently, L4-
33K was also suggested to function as a transcription factor stimulating the MLP (5). However, this role has been questioned since other groups presented evidence that L4-22K may function as the L4 protein stimulating MLP (12, 152). L4-22K is thought to stimulate MLP transcription in cooperation with the IVa2 protein by binding to sequence elements downstream of the promoter (134, 152). In addition, L4-22K-dependent accumulation of specific late mRNAs was also reported (134), suggesting a role for the protein in RNA processing. In conclusion, L4-22K and L4-33K seem to have complementary effects on late mRNA production, where both play key roles in the early-to-late switch in MLTU expression. However, their exact roles in these processes are still debated and needs further clarification.

\[
\text{3RE} = \text{IIIa virus infection-dependent splicing enhancer} \quad \text{3VDE} = \text{IIIa repressor element}
\]

*Figure 5.* The temporal splicing pattern of the L1 unit with the regulatory cis-elements in the upstream IIIa intron. Boxes represent exons, the thick line an intron, and thin lines with hooks the region, which is spliced out. The two adenosines at the branch point are indicated with asterisks (*).

**The L1 model system**

L1 is the only unit in MLTU producing messages both early and late during the virus life cycle. The 52,55K mRNA is produced both early and late after infection while the IIIa mRNA is only produced late. The 52,55K protein has a role in viral genome encapsidation by stabilizing the association of the genome with the empty capsid. The IIIa protein is a structural protein that is interspersed between the hexons in the capsid (4, 57). The functional reason for a temporal regulation of these proteins has been studied in tissue culture cells. Loss of 52,55K protein or gain of IIIa protein early in infection did not show any detrimental effects on virus growth, but it is not known whether the regulation is important when the virus infects its natural host (4).

The L1 unit has been extensively used as a model transcript for studies of temporally regulated alternative splicing. To date, a lot is known about this transcript; it uses a common 5' splice site and two alternative 3' splice sites to produce two mRNAs, the 52,55K mRNA (proximal 3’ splice site), and the IIIa mRNA (distal 3’ splice site), Fig. 5. Previous results show that the acti-
vation of the IIIa 3’ splice site is not dependent on cis-competition between the 52,55K and IIIa 3’ splice sites in vivo (76). Thus, by inactivating an ESE downstream of the 52,55K 3’ splice site, the temporal activation of IIIa in the late phase of infection was not destroyed.

IIIa splice site activation was shown to depend on two cis-elements in the transcript; the IIIa repressor element (3RE) and the IIIa virus infection-dependent splicing enhancer (3VDE) located in the intron just upstream of the IIIa 3’ splice site, Fig. 5. The 3RE binds highly phosphorylated SR proteins in uninfected and early-infected cells. This leads to an inhibition of IIIa splicing by blockage of spliceosome A complex formation and U2 snRNP recruitment to the branch point (91). In late-infected cells, the viral E4-ORF4 protein binds to the cellular PP2A and thus induces SR protein dephosphorylation. This dephosphorylation reduce the binding of SR proteins to the 3RE, which relieves the repressive function on IIIa splicing and makes the branch point accessible for U2 snRNP recruitment (92).

The mechanism of 3VDE is less clear. It has been shown to be the major cis-acting element responsible for IIIa 3’ splice site activation in adenovirus infected nuclear extracts (Ad-NE) (136, 137). The 3VDE contains the weak IIIa pyrimidine tract and is inefficiently bound by U2AF in uninfected HeLa cell nuclear extracts (HeLa-NE) (137, 213). This binding is even further reduced in Ad-NE (137). Splicing of the cellular IgM transcript, which like IIIa has a weak 3’ splice site, has been shown to proceed through the first catalytic step of splicing in U2AF depleted Ad-NE (119). These results indicate that there is a novel U2AF-independent pathway of splicing activation in adenovirus late-infected cells, which selectively supports splicing of weak 3’ splice sites. Our working hypothesis is that a viral factor or a virally induced factor takes over U2AFs function in recruiting U2 snRNP to the branch point and thereby activating IIIa splicing. We have chosen to call it 3VDE interacting factor (3VDF), Fig. 6.

![Figure 6](image_url)

**Figure 6.** The working model for L1 alternative splicing activation. The black box represents the IIIa exon. Thin boxes, white and gray, indicate the position of the 3RE and 3VDE respectively, and the thin line the rest of the intron. At late times of infection the hypophosphorylated SR proteins (SR) are released from the 3RE and a viral factor or a virally induced factor, the 3VDE interacting factor (3VDF), takes over U2AFs function in recruiting U2 snRNP to the branch point and thereby activating IIIa 3’ splice site usage.
Present investigation - results and discussion

Paper I
L4-33K, an adenovirus encoded alternative RNA splicing factor
The purpose of the intermediate expression of the L4 mRNAs was not known until it was proposed by Farley et al in 2004 that the early-to-late shift in MLTU expression require L4-33K protein expression (52). In this paper we extended the study by showing that L4-33K is an alternative RNA splicing factor, activating splicing of mRNAs expressed late during an infection. We further demonstrate that L4-33K favors splicing of transcripts with weak 3’ splice site sequence context, by enhancing an early step in spliceosome assembly. For the temporally regulated L1 transcript, we showed that the 3VDE sequence is the major element mediating L4-33K activation. This result implicates L4-33K, as being the 3VDE-interacting factor (3VDF) or being a part of the hypothetical 3VDF, in taking over the function of U2AF in the U2AF-independent pathway in adenovirus late-infected cells.

L4-33K is the key viral factor required for L1 alternative splicing
We were able to reproduce the cytoplasmic accumulation of late mRNAs from the L1 unit as Farley et al saw in their transient transfections. Hence, increasing amounts of a L4-33K expressing plasmid and an L1 reporter plasmid were transfected into HEK293 cells and resulted in a dose-dependent increase in IIIa mRNA expression. This increase in IIIa mRNA accumulation in vivo was due to a direct effect on RNA splicing, since we could mimic the enhancement of IIIa mRNA accumulation in in vitro splicing reactions using HeLa-NE supplemented with a recombinant L4-33K protein. Interestingly, the level of IIIa splice site activation with the recombinant L4-33K protein was essentially as efficient as the activation seen in Ad-NE. This result suggests, that L4-33K is the key viral factor required for the shift in L1 alternative splicing. The increase in IIIa mRNA splicing was not accompanied by a corresponding decrease in 52,55K mRNA splicing, suggesting that L4-33K acts as an alternative 3’ splice site activator.

L4-33K specifically activates weak 3’ splice sites
Next, splicing substrates with different 3’ splice sites were tested in vitro. The results showed, that the substrates could be divided into three major
categories based on their responsiveness to L4-33K; non-, weakly- and strongly responsive substrates. The substrates with classical, long pyrimidine tract like AdML, penton, 100K, globin, Ftz, and M-Tra ended up in the non-responsive group, while substrates with short and interrupted pyrimidine tract like IIIa, pV, and pVII, or hexon and IgM, were grouped into the strongly-responsive or weakly-responsive substrates, respectively. We found one exception to this rule and that was the 52,55K 3’ splice site that has a long pyrimidine tract but still is weakly activated by L4-33K. These data suggest, that L4-33K has a specificity towards activating 3’ splice sites with short and interrupted pyrimidine tracts that has a low basal activity in HeLa-NE. This is also what is observed during an infection since non-consensus-type introns like IIIa, and pV are highly activated in Ad-NE whereas consensus type introns are not stimulated (136). The transcripts activated by L4-33K are expected to bind U2AF65 inefficiently due to their poor pyrimidine tracts, while the non-responsive transcripts with long pyrimidine tracts bind U2AF65 efficiently. The pV and IIIa pyrimidine tracts have been shown to bind U2AF65 with a 10-50 times lower efficiency compared to penton and 52,55K (136). Splicing of the IgM transcript, with a weak pyrimidine tract, have been shown to require both U2AF65 and U2AF35 in uninfected nuclear extracts (77). However, the first catalytic step of IgM splicing has been shown to be independent of the U2AF dimer in Ad-NE (119), which implies that splicing of weak 3’ splice sites is enhanced by a switch from U2AF to L4-33K requirement during infection.

IIIa splice site activation was previously mapped to two cis-elements, the 3RE and 3VDE, located immediately upstream of the IIIa 3’ splice site. By constructing chimeric β-globin-IIIa mini-constructs we could show that exchanging the 3VDE with the corresponding sequence element from the rabbit β-globin gene (branch point, pyrimidine tract and AG dinucleotide) resulted in a decreased L4-33K responsiveness in vitro. Exchanging the 3RE with the corresponding sequence from the β-globin gene resulted in an increase in the basal level of IIIa splicing. This result was expected, since the 3RE previously has been shown to repress IIIa splicing by binding the SR family of splicing factors (91). In the presence of L4-33K, splicing of this transcript was enhanced to the same level as the normal IIIa transcript. In contrast, the double mutant lacking both the 3RE and the 3VDE was completely non-responsive to L4-33K. The 3VDE was also able to convert the L4-33K non-responsive β-globin transcript to a transcript that was strongly activated by L4-33K. Transfer of the 3RE to β-globin resulted in a transcript that was modestly activated by L4-33K. Taken together, these results suggest that the 3VDE is the major element mediating L4-33K splicing activation with a smaller contribution by 3RE.
L4-33K stimulates spliceosome assembly at an early step

Spliceosome assembly was studied to gain further support that L4-33K activated splicing directly. Spliceosome A complex formation is ATP dependent, so we used ATP-depleted HeLa-NE to show that the L4-33K induced complexes were related to spliceosome assembly and not some ATP-independent complexes formed in the extract. Also, by using U2 snRNA-depleted HeLa-NE we could show that L4-33K enhancement of spliceosome assembly was dependent on a functional U2 snRNP.

Potentially, L4-33K may work via a hit and run mechanism, modifying and activating a cellular factor to stimulate splicing of IIIa, or together with a cellular splicing factor that mediates the direct RNA binding function. However, we have no direct proof that L4-33K binds to the 3VDE. Previously, we hypothesized that a viral factor or a virally induced factor, the 3VDF, takes over U2AFs function in recruiting U2 snRNP to the branch point and thereby activating IIIa splicing (described in the section “The L1 model system”, Fig. 6). Thus, our results implicate, L4-33K in being the 3VDF or, more likely, being the factor that together with a cellular protein in the 3VDF takes over the function of U2AF as the factor recruiting U2 snRNP to the branch point. In agreement with such a hypothesis, we have shown that L4-33K stimulates spliceosome assembly at the level of U2 snRNP recruitment. Previously, we have observed A complex formation on the IgM transcript in U2AF-depleted adenovirus-infected nuclear extracts (119). These results indicate that there is a novel U2AF-independent pathway of splicing activation in adenovirus late-infected cells, which selectively supports splicing of weak 3’ splice sites. However, it remains to be tested whether the formation of A complex is dependent on L4-33K activity during an infection.

Figure 7. A schematic drawing of the L4-33K protein showing different motifs with an expansion of the 28 carboxyl-terminal amino acids corresponding to the so-called ds region. The tiny RS repeat is shown with bold amino acids in the sequence and the specific serines are indicated with their positional numbers.
A tiny RS repeat is required for L4-33K splicing enhancer activity

By testing deletion mutants of the L4-33K protein in the L1 in vitro splicing assay, we found that the first 157 N-terminal amino acids in L4-33K were non-essential for L4-33K activation of IIIa splicing. However, the N-terminus seems to have a slight stimulatory effect on L4-33K function, possibly by stabilizing the protein. Recent work from our group shows that the L4-22K protein, which shares the first 106 amino acids with L4-33K, is not activating IIIa splicing (12). This result further supports the conclusion that the L4-33K splicing enhancer function domain is not located in the N-terminus. It is possible that the acidic N-terminus serves other functions attributed to the L4-33K protein (5, 55, 56).

When a deletion mutant lacking a conserved C-terminal domain (ds) made L4-33K non-active as a splicing enhancer protein we found that this region contained a tiny RS-repeat, Fig. 7. RS repeats in SR proteins are known to function as mediators of protein-protein interactions and also important for the subcellular and subnuclear distribution of SR proteins (summarized in the section “Splicing factors”). Further, substituting the serines to glycines in the tiny RS repeat suggested that the serines might serve redundant functions, with the serine at position 192 (S192) being most important for the L4-33K splicing enhancer phenotype. The tiny RS repeat in L4-33K is not long enough to classify L4-33K as a classical SR protein. This was further supported when L4-33K was not able to complement splicing in S100 extracts, which is one of the requirements for being a classical SR protein. However, it should be noted that it appears to be important for L4-33K function and subcellular distribution, as discussed in paper II.

Paper II

Serine 192 in the tiny RS repeat of the adenoviral L4-33K protein is essential for nuclear localization and reorganization of the protein to the periphery of viral replication centers

Since L4-33K is a virus-encoded alternative splicing factor, we wanted to investigate how the protein is distributed within the cell. L4-33K is known to localize to the nucleus (65), but no nuclear compartments have been assigned. In this paper, we show that L4-33K does not localize to nuclear speckles as one might have expected from its activity as a splicing factor, but rather has a diffuse localization in the nucleus with enrichment in the nuclear membrane. Further, the conserved C-terminal domain containing the tiny RS repeat is essential for nuclear localization of the protein. In the context of a virus infection, L4-33K redistributes to discrete ring-like structures in the
surrounding of viral replication centers. In addition, we show that the functionally important S192 in the tiny RS repeat is necessary for the nuclear localization and redistribution of L4-33K during infection. Interestingly, we found a surprisingly good correlation between the functional domain (paper I) and the domain responsible for subcellular distribution.

L4-33K does not localize to nuclear speckles

We could follow the subcellular distribution of L4-33K by transfecting a FLAG-epitope tagged L4-33K expressing plasmid into human embryonic kidney (HEK) 293 cells and immunostain for the FLAG-tagged protein. In uninfected cells, L4-33K has a diffuse nuclear distribution with enrichment in the nuclear membrane. In addition, some unidentified spots outside the nucleus were typically observed. By co-staining for the SR protein SRSF2, which is a marker for splicing speckles, we concluded that the L4-33K protein does not specifically localize to nuclear speckles. This finding is a bit puzzling, since we know that L4-33K is active as a splicing factor in uninfected cells. Splicing factors, like SR proteins, are typically enriched in nuclear speckles, which serve as a storage and reassembly site. However, L4-33K is not a classical SR protein and storage in cellular compartments seem unnecessary for viral proteins, which are produced in infected cells to perform their actions directly.

Approximately 60% of the cells showed co-localization of L4-33K with Lamin B at the nuclear membrane. It is known that cellular regulators e.g. transcription factors are sequestered in the nuclear membrane in inactive complexes. The nuclear lamina is binding transcription factors and DNA to the nuclear membrane (8, 178). Implying that, L4-33K might also be coupled to the lamin network. This localization pattern was not seen for the related L4-22K protein, indicating that the domain responsible for membrane localization is located in the unique C-terminal part of L4-33K.

L4-33K localizes to ring-like structures in the nucleus during infection

Since both L4-22K and L4-33K are localized in the nucleus it seemed reasonable to suspect that a nuclear localization signal was located in the shared N-terminus of L4-33K and L4-22K. However, both biochemical fractionation and indirect immunofluorescence studies revealed that the common N-terminus was not necessary for nuclear localization of the L4-33K protein. Thus, N-terminal deletion mutants, removing part or all of the shared amino acids plus an additional 51 amino acids unique to L4-33K, also localized to the nucleus. Collectively, these results suggest that L4-33K and L4-22K have distinct nuclear localization signals.

In 1981, Gambke and Deppert showed by using indirect immunofluorescence with antiserum directed against purified SDS-denatured L4-33K, that the L4-33K protein localizes exclusively to the nucleus and more specifically in small dots allocated in ring-like structures during infection (65). We
could mimic the L4-33K localization pattern, by transfecting 911 cells with the FLAG-epitope tagged L4-33K expressing plasmid and subsequently 4 h later infect cells with wild type Ad5 for 20 h. The wild type L4-33K protein was efficiently reorganized from the nuclear membrane to the ring-like structures in the nucleus. By co-staining with an anti-DBP antibody (E2A-72K), we concluded that L4-33K is localized to the outer borders of replication centers during infection, where active transcription and RNA processing occurs. Interestingly, adenovirus causes a major reorganization of splicing factors from nuclear speckles to ring-like structures at replication centers called the peripheral replicative zone or transcription sites. This is followed by an accumulation in enlarged nuclear speckles localized to the periphery of the nucleus, in parallel with the E4-ORF4 induced dephosphorylation of SR proteins (25). When following the SRSF1 localization and L1 alternative splicing pattern during an infection it seems like IIIa 3’ splice site starts to be used more efficiently during the transition from “rings” i.e. the transcription sites to enlarged speckles (64). IIIa 3’ splice site repression is relieved upon dephosphorylation of SR proteins and account for some of the IIIa splicing activation seen during infection. However, in paper I we showed that L4-33K was responsible for the other part of IIIa 3’ splice site activation and here the localization to “rings” observed for L4-33K further strengthens this observation. In addition, the group of Blair showed in a spatial distribution that SRSF1 is observed in “rings” directly after onset of DNA replication and that this is followed by the L4-33K expression and an activation of alternative splicing (87).

The tiny RS repeat is important for subnuclear localization

We further showed that the tiny RS repeat was critical for nuclear localization and subnuclear distribution. By mutating serines in the RS repeat we found that S192 is not only crucial for the splicing enhancer function of the protein, but also for the localization of the protein. However, it appears that the other three serines in the region also contributes to the subcellular localization, since a triple mutant has an increased cytoplasmic distribution compared to the wild type L4-33K protein. We found a good correlation between the domain responsible for the L4-33K splicing enhancer phenotype and the domain responsible for nuclear distribution, i.e. the tiny RS repeat. In comparison, a good correlation between the domain responsible for splicing enhancer function and subcellular distribution is also observed for SR proteins and their RS repeats, as discussed in section “The SR proteins”.

When 911 cells transfected with the RS repeat mutants were infected with wild type Ad5 virus we observed that the splicing defective mutant proteins that lost their exclusive nuclear localization, did not redistribute to viral replication centers during infection. This further strengthens the correlation between the domain involved in splicing enhancer function and nuclear reorganization of L4-33K. However, one should keep in mind that the splicing
enhancer domain was established \textit{in vitro} (paper I), which means that the localization of the mutant proteins does not affect the previous conclusion that the tiny RS repeat is necessary for L4-33K splicing enhancer function.

The observation that serines are important for localization might be due to their ability to become phosphorylated. Reversible phosphorylation is a very common post-translational modification used by proteins for distribution between cellular compartments. As described in the section “Splicing factors”, phosphorylation of SR proteins by their specific kinases SRPK1 and CLK/Sty are important for their redistribution between cytoplasm to nucleus and from speckles to active sites of transcription. Since L4-33K is a phosphoprotein it would be interesting to test whether its redistribution during infection is dependent on phosphorylation. However, a recombinant L4-33K protein was not phosphorylated by SRPK1 nor by CLK/Sty \textit{in vitro} (data not shown, paper I), implying that other protein kinases must be involved.

\textbf{The ds region is not sufficient for nuclear localization of L4-33K}

In an attempt to further narrow down the L4-33K nuclear localization signal we isolated the ds region, containing the tiny RS repeat, and fused it to enhanced green fluorescent protein (eGFP). By transient transfection followed by Ad5 infection and fluorescence imaging we concluded that the ds region was necessary but not sufficient for nuclear localization of L4-33K and redistribution to the peripheral replicative zone. Sequences upstream or downstream of the conserved ds region are most likely needed for nuclear import of L4-33K. Preliminary data implies the presence of a bipartite signal, which needs to be further characterized.

\textbf{Paper III}

Two cellular protein kinases, DNA-PK and PKA, phosphorylate the adenoviral L4-33K protein and have opposite effects on adenoviral alternative RNA splicing

As described in paper I, L4-33K is an alternative splicing factor activating the shift from L1-52,55K to L1-IIIa mRNA. To gain further knowledge about how L4-33K enhances alternative splicing during infection, we set out to find interacting proteins. The results show that L4-33K specifically associates with the catalytic subunit of DNA-PK (DNA-PKcs) in uninfected and adenovirus-infected cells. We also showed that L4-33K is highly phosphorylated by DNA-PK \textit{in vitro}, in an unusual dsDNA-independent manner. Further, DNA-PK appears to block the early-to-late switch in L1 alternative splicing, suggesting that DNA-PK has an activity as a regulator of adenovirus MLTU alternative splicing. Moreover, we showed that L4-33K also is phosphorylated by PKA \textit{in vitro}, and that PKA has an enhancer effect on L4-
33K-activated IIIa splicing. L4-33K is one of the major phosphoproteins expressed in adenovirus-infected cells and this is the first time that two cellular protein kinases are shown to phosphorylate L4-33K. Collectively, our data suggest novel regulatory roles for DNA-PK and PKA in adenovirus alternative splicing.

**L4-33K and DNA-PK**

By using two separate isolation methods, GST-pull-down and immunoprecipitation, we could verify that the catalytic subunit of DNA-PK is a stable interactor with L4-33K. DNA-PKcs associates with recombinant L4-33K both in uninfected nuclear extract and adenovirus-infected nuclear extracts. In addition, DNA-PKcs also associate with a FLAG-epitope tagged L4-33K expressed from a recombinant adenovirus, suggesting a formation of a L4-33K:DNA-PKcs complex during lytic infection.

DNA-PK is a heterotrimeric protein kinase consisting of DNA-PKcs and two regulatory subunits, Ku86 and Ku70. However, in our system we were not able to detect the regulatory subunits Ku86 or Ku70 as components in the L4-33K:DNA-PKcs complex. Interestingly, DNA-PKcs can phosphorylate proteins in a Ku-independent manner, which implies that Ku proteins might not be essential for stable L4-33K:DNA-PKcs complex formation (208).

The L4-33K interaction with DNA-PKcs was not expected since DNA-PK has not to our knowledge previously been shown to be involved in pre-mRNA splicing. However, since DNA-PK can modulate transcription, which in turn is known to regulate alternative splicing, there might be a connection between the two proteins, making the interaction more interesting. Also, the discovery of an interacting protein kinase was intriguing since L4-33K is one of the major phosphoproteins expressed during infection (11). When separating the L4-33K protein immunoprecipitated from transient transfected HEK293 cells, on two-dimensional SDS-PAGE, we observed that L4-33K exist in several different isoelectric charged forms, suggesting post-translational modifications. In addition, we could confirm that L4-33K is a phosphoprotein in vivo. Thus, we observed a size reduction corresponding to dephosphorylation of L4-33K by alkaline phosphatase treating immunopurified L4-33K protein.

We found that DNA-PK phosphorylates L4-33K in vitro. Also, when comparing the phosphorylation of L4-22K and a deletion mutant L4-33Kds, we could conclude that the majority of the DNA-PK phosphorylation sites reside within the highly conserved ds region. This region lacks the obvious consensus phosphorylation sites for DNA-PK, which are serines or threonines followed by glutamine (Ser/Thr-Gln). However, this observation correlate with the finding that DNA-PK can phosphorylate serines and threonines in non-consensus motifs as well (36).
DNA-PK is activated by dsDNA breaks in the genome, which induces the cellular DSBR system. In vitro, the activation can be mimicked by addition of linear dsDNA resulting in a 10 times higher phosphorylation rate of the substrate proteins (180). However, we could not detect any increase in phosphorylation of L4-33K by addition of dsDNA. This finding implies that, L4-33K is phosphorylated by DNA-PK by an atypical DNA-independent mechanism. Although unusual, other proteins have also been reported to show the same phenotype, like the nuclear receptor co-activator thyroid hormone receptor-binding protein which itself can enhance DNA-PK activity (98). Interestingly, the phosphorylation activation mechanism was changed from being DNA-independent into DNA-dependent when L4-33Kds was assayed. In summary, it appears that two L4-33K phosphorylation mechanisms co-exist:

1. the classical dsDNA-dependent phosphorylation mechanism, which appears to apply to phosphorylation sites in L4-33K located outside of the ds region.
2. the unusual DNA-independent mechanism, which appears to apply to phosphorylation sites within the L4-33K ds region.

An RNA-dependent DNA-PK phosphorylation mechanism has also been reported for the hnRNP C protein and the nuclear DNA helicase II protein (NDH II) (215). Thus, it would be interesting to test whether the DNA-independent phosphorylation pattern seen for L4-33K actually is dependent on the presence of RNA.

As discussed in paper I and II, the ds region in the C-terminus of L4-33K holds a tiny RS repeat, which is important for both the function in alternative splicing and the subcellular localization. This makes the observed DNA-PK phosphorylation pattern highly intriguing. Since DNA-PK phosphorylates L4-33K specifically in this region, we decided to investigate the role of DNA-PK in adenovirus pre-mRNA splicing. By infecting a DNA-PKcs defective cell line (MO59J) and a wild type counterpart (MO59K) and analyzing the late viral mRNA levels we observed that the level of IIIa mRNA accumulation clearly increased in the DNA-PKcs defective cell line as compared to wild type cells. This finding suggests a potential inhibitory role of DNA-PK on the switch from the early-to-late pattern of L1 alternative splicing.

DNA-PK has a known repressive function on adenovirus DNA replication by producing concatamers of the linear dsDNA viral genome (198). It was later found that the E4-ORF3 and E4-ORF6 proteins rescue the genome from concatemerization by inhibiting DNA-PK function in the DSBR system (21, 82). Our group previously showed that E4-ORF3 and E4-ORF6 behave like splicing factors facilitating MLTU i-leader exon inclusion or skipping, respectively, in transient transfection assays and during infection (149). One could suspect that a concentration-dependent increase in free unbound E4
proteins, because of the non-existing DNA-PKcs in infected MO59J cells would have an effect on L1-IIIa splicing. Nevertheless, this is unlikely because the E4-ORF3 and E4-ORF6 proteins have not been shown to have a stimulatory effect on L1-IIIa splicing (149). It is anyhow intriguing that adenovirus encodes for three different proteins that all target DNA-PK.

The Ku86 subunit of DNA-PK has been functionally associated with RNAP II elongation sites independently of DNA-PKcs (131). Further, Ku86 resembles the hnRNP family of splicing factors because of its high affinity to G-rich sequences and it has also been shown to associate with the hnRNP family (215). These findings indicate a possible role for Ku86 in pre-mRNA processing, maybe as a splicing factor. It is intriguing to speculate that the L1-IIIa 3’ splice site stimulation seen during infection of the DNA-PKcs negative cell line might be due to the free unbound Ku86 subunit.

**L4-33K and PKA**

Based on previous reports that PKA phosphorylates a subset of splicing factors, we tested whether L4-33K is a PKA substrate. *In vitro* phosphorylation assays revealed that L4-33K, but not L4-22K, was highly phosphorylated by PKA Cα1. PKA preferentially phosphorylates (Arg-Arg-X-Ser/Thr) motifs but other motifs are also used. By analyzing the amino acid sequence of L4-33K we identified five potential PKA phosphorylation sites in the protein; T100, S127, S136, S189 and T194. Interestingly, two of them are located in the functionally important ds region.

The outcome of PKA phosphorylation of its substrate proteins is versatile. It can, for example, regulate the transport from the nucleus to the cytoplasm as for the poly-pyrimidine tract binding protein (PTB), which is an hnRNP involved in pre-mRNA processing. PTB translocates from the nucleus to the cytoplasm upon PKA phosphorylation, a function that has been coupled to mRNA localization (120, 206). Also the reverse, import to the nucleus can be regulated by PKA phosphorylation as described for the Drosophila transcription factor Dorsal (26). In addition, apart from regulating the subcellular localization of a protein, PKA can induce other functions as for the NF-κB transcription factor, which activates transcription upon PKA phosphorylation (217). Furthermore, PKA is targeted to splicing speckles probably by the splicing factor arginine/serine-rich 17A (SFRS17A) and has a role in regulating alternative splicing by phosphorylating SRSF1, which in turn enhances inclusion of tau alternative exon 10 (88, 179). In addition, PKA promotes distal 5’ splice site selection of the adenoviral E1A transcript probably by a cAMP-independent mechanism (104).

Since PKA phosphorylation has been connected to alternative splicing regulation we wanted to investigate whether it also has a role in L4-33K activated L1 alternative splicing regulation. By co-transfecting HEK293T cells with an L1-reporter plasmid and plasmids expressing L4-33K plus/minus PKA Cα1 we observed that L4-33K enhancement of IIIa pre-
mRNA splicing was further enhanced by the PKA Cα1 subunit. This result suggests that PKA stimulates the activity of L4-33K as a splicing enhancer factor. This result may indicate, that L4-33K has at least two different biochemical states; the uninduced state, with low splicing enhancer activity, and the PKA induced state, where PKA Cα1 phosphorylates L4-33K, which obtains even higher splicing enhancer activity. In addition, we found that PKA Cα1 also increases transcription from the MLP.

PKA is activated upon adenovirus entry and is required for the microtubule-mediated nuclear targeting of the virus capsid (187). PKA is also involved in regulating the early phase of infection together with one of the E1A proteins. The E1A12S functions as a viral AKAP that binds to the RIIα subunit and relocates it to the nucleus to enhance transcription from the E2 promoter (53). In addition, our results suggest that PKA also has a role in inducing the late phase of infection by targeting L4-33K. However, the exact mechanism is far from clear. It might be that L4-33K redistributes with the Cα1 subunit to the nucleus upon PKA activation, however we have no evidence of PKA and L4-33K interacting with each other. More likely, PKA phosphorylation of L4-33K changes its conformation into becoming an even stronger splicing enhancer protein. Also, PKA phosphorylation of L4-33K might change its affinity towards interaction with nuclear import proteins, facilitating the nuclear localization of the protein. Anyhow, we have yet not investigated the role of PKA on L4-33K activity during a lytic infection.

PKA is probably not the missing component of 3VDF, since we were not able to co-immunoprecipitate PKA with L4-33K (data not shown) and PKA is not known to bind RNA. However, it is not unlikely that PKA activates another protein with RNA binding properties, which together with L4-33K takes over U2AF function in adenovirus late-infected cells. Nevertheless, the previously described cellular contribution to 3DVF is still missing. It might be that PKA and L4-33K are the missing pieces in 3VDE-dependent activation of IIIa splicing and that the activation is not dependent on a 3VDE-interacting protein, with increased binding to the 3VDE. However, an unknown 46 kDa protein was found to cross-link to IIIa pyrimidine tract in Ad-NE (137), which further strengthens the statement that the cellular contribution to 3VDF still is missing.

**Two protein kinases with opposite effects on L1 alternative splicing**

It has been over 30 years since L4-33K was shown to be a nuclear phosphoprotein. The protein kinases discussed in this paper are the first two that have been established as L4-33K phosphorylating enzymes. Phosphorylation by DNA-PK could even be mapped to the functional important ds region of L4-33K. In addition, there are putative PKA phosphorylation sites also located in this region, where preliminary data (data not shown) implies that those sites are phosphorylated by PKA. Phosphorylation of proteins is an important regulatory mechanism, which can activate or inactivate proteins de-
pending on the cellular environment. Alternative splicing is sensitive to the phosphorylation status of SR proteins, as described earlier. It is not unlikely that the activity of L4-33K on alternative splicing, as a viral SR-related protein, is subject to phosphorylation regulation.

Our results show that DNA-PK and PKA have opposite effects on L1 alternative splicing; DNA-PK functions as an inhibitor and PKA as an activator of IIIa pre-mRNA splicing. Counteracting proteins contribute to the diversity of regulated mechanisms. By reaching a balance in the environment they can fine-tune the expression, activity, and distribution of factors in the surroundings. Interestingly, PKA has a role in regulating the subcellular distribution of DNA-PK by promoting nuclear entry upon cAMP activation. In addition, DNA-PK is a substrate of PKA (85). However, the functional relevance of this is not known.

In addition, another link between DNA-PK and PKA has been reported for the Epstein Barr virus protein, EBNA-LP, which acts as a co-activator of the EBNA-2-mediated transcription (80, 148). EBNA-LP has been shown to associate with both DNA-PKcs and PKA C subunits. The latter through a binding via the HA95 protein, which interestingly previously has been shown to be involved in alternative splicing regulation (78, 79, 104). The purpose of the EBNA-LP association to these proteins is not clear, but they are suggested to play a role in virus-mediated transcriptional activation.

How L4-33K is activating L1 alternative splicing still remains unclear, nevertheless our data suggest a novel interplay between the viral L4-33K protein and two cellular protein kinases, DNA-PK and PKA, in regulating the adenovirus late alternative splicing.
Concluding remarks and future aims

In this thesis, I have discussed today’s knowledge about the adenoviral protein L4-33K. At the beginning of the work presented in this thesis, L4-33K was suggested to be an important regulator of the early-to-late shift in adenovirus gene expression (52). In addition, previous research from our group had pointed out the existence of a virus infection-dependent splicing factor in the regulation of the L1 mRNA expression (137). When we investigated the possible connection between these two observations we found that the L4-33K protein is an alternative RNA splicing factor, activating splicing of a subset the late mRNAs (paper I). We further demonstrated that L4-33K favors splicing of cellular and viral transcripts with weak 3’ splice sites by enhancing an early step in spliceosome assembly. Thus, some of the proposed hypothetical 3VDF-criteria were filled except one, we were not able to demonstrate that the L4-33K protein binds to RNA. Since we knew from our in vitro assays that L4-33K is the only viral factor needed to convert an uninfected HeLa-NE to an extract with essentially the same splicing properties as an Ad-NE we began to search for cellular interacting proteins.

As described in paper III, we found that L4-33K was interacting with the protein kinase DNA-PK, which we later observed to phosphorylate L4-33K, in what seems to be the functionally important ds region. In addition, we found that the protein kinase PKA also phosphorylates L4-33K. These protein kinases seem to have opposite effects on L1 alternative splicing; DNA-PK functions as an inhibitor and PKA as an activator of IIIa pre-mRNA splicing.

The present model of L1 alternative splicing activation is described in Fig. 8. Early in infection the L1 unit only expresses the 52,55K mRNA, whereas the IIIa 3’ splice site is repressed due to a blockage of U2 snRNP recruitment to the branch point by phosphorylated SR proteins binding to 3RE. Further, the weak pyrimidine tract in 3VDE is weakly bound by U2AF, preventing stable A complex formation. In late-infected cells, the viral E4-ORF4 protein binds to the cellular PP2A and induces SR protein dephosphorylation. In addition, L4-33K is produced in a transient intermediate phase possibly from the recently proposed L4P, which seem to be activated by accumulating linear dsDNA genomes (135). L4-33K is recruited to the nucleus and to the peripheral replicative zone, where late viral transcription occurs. It seems possible that L4-33K replaces the SR proteins in that location, since the binding of dephosphorylated SR proteins to the 3RE is
reduced. They are redistributed from the peripheral replicative zone to enlarged nuclear speckles in the periphery of the nucleus at the same time (87). The release of SR proteins relieves their repressive function on IIIa splicing and makes the branch point accessible for U2 snRNP recruitment.

Figure 8. A hypothetical model for the temporal control of L1 alternative splicing. The black box represents the IIIa exon. Thin boxes, white and gray, indicate the position of the 3RE and 3VDE respectively, and the thin line the rest of the intron. Cellular factors are labelled accordingly SR proteins (SR), U2 snRNP (U2), U2 auxiliary factor (U2AF), Protein phosphatase 2A (PP2A), Protein kinase A (PKA), DNA-dependent protein kinase (DNA-PK), unknown protein of viral or host origin (?) and viral factors by their names. Phosphorylation is indicated by addition of phosphate group (P) to the substrate factor and dephosphorylation by loss of P. The model is described in the text in section “Concluding remarks and future aims”.

46
L4-33K is phosphorylated by DNA-PK and PKA in vitro. However the phosphorylation pattern of L4-33K during infection needs to be established. Hypothetically, L4-33K becomes phosphorylated and forms a complex with an unknown cellular RNA-binding protein with specificity towards weak pyrimidine tracts. This complex, 3VDF, then replaces the function of U2AF during infection. This enables the recruitment of U2 snRNP to the branch point and thus activation of the weak IIIa 3’ splice site. Whether this is a general 3’ splice site activation mechanism for late mRNAs also needs to be established.

The model raises many questions concerning the precise function of the L4-33K protein in splice site activation. Firstly, is there a functional relevance of phosphorylation of L4-33K by PKA and/or DNA-PK? And if so, what is the consequence? To be able to answer these questions the specific phosphorylation sites of L4-33K needs to be established. Further a mutational analysis, where phosphorylation sites are mutated and the effect of these changes are analyzed for splicing enhancer function, would answer some of these questions. This applies specially for PKA, which was shown to have a stimulatory effect on L4-33K enhanced L1 alternative splicing. To further investigate the role of PKA in L1 alternative splicing, it would be interesting to inhibit PKA in cells or deplete nuclear extracts from PKA and test whether L4-33K can activate IIIa splicing in the absence of PKA.

Secondly, what is the composition of 3VDF? In an early attempt to find interactors, we used a FLAG-epitope tagged L4-33K expression plasmid to transfect HEK293 cells for preparation of cytoplasmic extracts and found proteins involved in transcription, RNA processing and chromatin remodelling (data not shown). However, we were not able to verify these interactions in either uninfected or infected nuclear extracts in the assays used in paper III. It would of course be interesting to further investigate these potential L4-33K interactors and their involvement in L1 alternative splicing regulation. We have failed to detect an RNA binding activity for L4-33K in UV cross-linking experiments. However, it might still be a possibility that L4-33K acts by its own in recruiting U2 snRNP to the branch point, for example if the RNA binding affinity of L4-33K increases upon phosphorylation.

Thirdly, is the Ku86 subunit of DNA-PK important for splicing activation of the late genes? DNA-PK has been shown to phosphorylate the CTD of RNAP II, which implies an indirect role in of DNA-PK in splicing. Ku86 has been functionally associated with RNAP II elongation sites independently of DNA-PKcs (131). This could mean that Ku86 has a role in pre-mRNA splicing. Further, Ku86 resembles the hnRNP family of splicing factors because of its high affinity to G-rich sequences (215). The 3RE and 3VDE in the IIIa intron has a high G content, implying a high binding affinity for Ku86. It is an intriguing possibility that the 3VDF might contain the Ku86 subunit and thus regulate L1 alternative splicing during infection. This however, needs to be experimentally verified.
Fourthly, what decides the subcellular localization and redistribution of L4-33K during an infection? It seems like a viral factor or virus-induced factor is responsible for L4-33K redistribution. L4-33K has been shown to bind two viral proteins, pV and L4-100K (103), but the functional relevance for these interactions is not known. However, it seems unlikely that these viral proteins help in redistributing L4-33K to replication centers, since they are expressed as a consequence of L4-33K splicing activation function. The motif responsible for localization and redistribution of L4-33K i.e. the RS-repeat lies within the domain phosphorylated by DNA-PK. Thus it would be interesting to investigate whether the subcellular distribution of L4-33K is dependent on phosphorylation either by DNA-PK or PKA.
Acknowledgements

First and foremost, I would like to thank my supervisor professor Göran Akusjärvi for giving me this opportunity to explore the world of science. I am really grateful for you believing in me and for your help and support during the years. I really appreciate that you have allowed me to explore things independently but also that you have pushed me in the right direction when needed. Your sometimes tricky questions have taught me to think at many different levels when planning and proceeding with the projects.

Many thanks to my co-supervisor professor Stefan Schwartz, as well as professor Göran Magnusson for constructive criticism during the years. And a special thanks to professor Catharina Svensson for sharing your enormous knowledge in molecular biology and virology and for taking part in discussions during our Friday meetings. Thanks also to my examinator professor Dan Andersson.

To all my co-authors; Thank you Ellenor Backström Winquist (my old “lab-twin” and splice girl, for everything, but especially for teaching me how to splice, what would I have done without you? also a huge thank you! for proof-reading my thesis), Anette Carlsson (for answering all my questions about lab stuff, How? Where? When? and for spreading your calmness in the lab), Sara Östberg (Piff or Puff?, I never learn… for your contagious laughter, many fika times with important and not so important discussions), Anne-Katrine Kvissel (the Norwegian splice girl, for being such an optimistic soul, and for your engagement in my projects. I really hope we can have that glass of red wine soon!), Tanel Punga (my trx expert!, thanks for returning to the lab with such enthusiasm and for sharing your great scientific knowledge), Anne-Kristin Aksaas (for your help in combining two projects into one super-paper!), Åke Engström (for your proficiency in mass spectrometry), Björn Skällhegg (for your expertise in the protein kinase field and revision of paper III). I am very thankful for your engagement in our projects, it has really been a wonderful time and the knowledge you possess is invaluable.

To all the past and present members of the Adeno lab; what would I have done without you? You have all contributed to this thesis and I could never have finished it if it was not for you guys! Special thanks to Cecilia Johans-
son (for always having time to discuss everything from lab-life to “real” life, for your constructive criticism and also for proof-reading my thesis), Wael Kamel (the smartest student ever! for our nice chats and for sharing your good mood. Also, for finishing my old nuts and candy!), Xin Lan (the astrologist in the lab, for teaching me the importance of Chinese astrology), Alexis Fuentes (Lunch? Små barn, små problem. Stora barn, stora problem. Thanks for putting things in a different perspective), Daniel Öberg (guapo hombre, energizer x10, for spreading your optimism, wherever you are!), Raviteja Inturi (for never saying no to fika and for teaching me about the amazing India), Wu Cheng Jun “Troy” (for your kindness), Christina Öhrlalm (for being “the one to ask” when I was a newcomer in the lab), Josef Seibt (for all kinds of lab introductions e.g. lab safety), Sofia Gkountela (for sharing your knowledge in cloning and for your wonderful kindness), Ning Xu (for sharing your knowledge in different virology-related methods), Gunnar Alkemar (for keeping the project going when I was on maternity leave), Saideh, Xiaofu, Hongyan, Bosse, Edyta, Bai Lufeng “Frank”, Angelica, David, Kerstin, Anton, Endrit, Emilio…

To the SS-lab members; I had initially grouped you with the Adeno-group because I really feel so close to you guys! Anyway THANK YOU all HPV, HIV and influenza experts! You have all taught me a lot about splicing in these “not so important” ;) viruses. Monika Somberg (Ms Northern Blot, dancing queen and efficiency embodied, for making me realise the importance of taking care of yourself, please take care!), Anna Tranell (for always being optimistic and having time to discuss both lab- and life-related problems, also for cheering me up whenever needed), Carlos Cardoso (guap a ;) hmm I will never learn, however I will try to take it mañana from now on), Xiaoze Li (the beautiful Ms Li, for your inspiring style), Sofia (Fia) Lindström (happy, happy! for your amazing party mood and for the afterwork chats about everything and nothing), Susanne Tingsborg (for uncountable lunches and for always listening and caring), Roger, Samir, Margaret, Xiaomin, Görel, Hilary…

I would like to give my warmest thanks to Lena Möller, my lab mum, I still come up with “Lena- questions”…

To the administration staff; Thanks to you all for always smiling and caring. Barbro (for listening and double-checking all the paper-work, especially during these last months), Marianne, Erika, Rehné, Kerstin, Olav (Mr. Fix-It, for all the times you have saved my day by rescuing samples from a centrifuge, repairing incubators etc), Tony, Nalle, and Ylva

To present and former IMBIMers for sharing IMBIM-days, IphA-pubs, lunches, equipment, laughter, discussions and much, much more. A special
thanks to the former B11:3 people; Christina Tobin Kårström (for being you, a very inspiring Irish cailín, and for proof-reading the most read parts of the thesis...), Sanna Koskiniemi (the inspiring FIN, for trying to help me keep my Finnish spirit alive, for many parties, spexes and chats about almost everything!), Anna Zorzet (for always being happy and spreading your enthusiasm), Maria Pränting (for your cheerful charisma and theatrical skills), Linus Sandegren (for many, many nice lunches and chats about sick kids and other parental issues) and to the former and present B9:3 people especially Alejandro, Thomas, Ananya, Kathrine, Else, Julia and finally my new office mate Carolina (for encouraging words throughout the whole thesis writing process).

I want to thank some amazing people outside the lab that have helped me through the years and are particularly essential to me; Jenny, Signe, Sofia (Fia) Persson, Kalle, Carro, Micke, Kattis, Peter, Lisa, Agron, Anna, Johan, Emely, Henrik. You have really helped me with this thesis, by distracting my attention towards other things every now and then.

Ett stor tack till mina svärföäldrar Lena och Göte som alltid ställer upp när det behövs! Och även till Cissi och Danne som är ett stort stöd i vardagen och underbara vänner!

Oj, hur ska jag kunna tacka er? Min underbara Mamma och Pappa! Tack för allt ert stöd i allt jag åtagit mig i mitt liv. Ni har alltid trott på mig och det är till stor del därför jag har klarat av att utföra detta, att få ihop en avhandling. Jag älskar er!

Kim, my little brother, lilla-stora du... I am so incredibly grateful that you are who you are, always so helpful, interested and sooo exceptionally smart! Thanks for all your help over the years! And last but not least, for making the cover of my thesis the best cover ever!

To my lovely family; Stefan, words are not enough to describe my gratitude. Your support, love, patience and humour during all my years as a PhD student really made things easier. And seriously there would not have been a thesis if it were not for your cooking and care taking of our home and our son. Mika, min underbara lilla solstråle! Vad skulle jag göra utan dig?

Jag älskar er!
References


Foulkes, N. S., B. M. Laoide, F. Schlotter, and P. Sassone-Corsi. 1991. Transcriptional antagonist cAMP-responsive element modu-


dependent protein kinase type II from the Golgi complex to the nucleus. EMBO J 4:2801-6.


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 709

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine.