The Multifunctional Nature of the Adenovirus L4-22K Protein

SUSAN LAN
Dissertation presented at Uppsala University to be publicly examined in A1:111a, BMC, Husargatan 3, Uppsala, Friday, 15 April 2016 at 13:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Professor Per Elias (University of Gothenburg, Department of Medical biochemistry and Cell Biology).

Abstract

The adenovirus major late transcription unit (MLTU) encodes for most of the mRNAs that are translated into the structural proteins of the virus capsid. Transcription from the MLTU is directed by the major late promoter (MLP), which is highly activated during the late phase of infection. This thesis discusses how the adenovirus-encoded L4-22K protein regulates the MLP at both the level of transcription and pre-mRNA splicing. The study shed new light on the complex regulation of the early to late shift of adenoviral gene expression.

Here we show that the L4-22K protein has opposing effects on MLP transcription, functioning both as an activator and a repressor protein. The stimulatory effect mainly depends on the direct interaction of the L4-22K protein with the downstream element (DE element) located approximately 100 nucleotides downstream of the transcription initiation site. In addition to the DE element we also show that the promoter-proximal upstream element (UPE) acts as an L4-22K responsive enhancer element in the MLP. Preliminary data suggests that the activation of MLP transcription via DE and UPE differs mechanistically. The transactivation domain of the L4-22K protein is localized to the conserved carboxy-terminus of the protein.

Our results also defined a novel low affinity L4-22K binding site, the R1 region, which functions as a repressor element in MLP transcription. At high concentrations L4-22K binds to R1 and recruits the cellular transcription factor Sp1 to a DNA segment covering the major late first leader 5′ splice site that is embedded in the R1 region. Sp1 binding to R1 results in a suppression of L4-22K-mediated activation of MLP transcription. This self-limiting effect on MLP transcription might have a function to fine-tune the MLTU gene expression.

Interestingly, the L4-22K protein binds with the same sequence specificity to both the R1 double-stranded DNA and R1 single-stranded RNA (ssRNA). L4-22K binds to the R1 ssRNA with the same polarity as the MLTU nascent RNA. This binding results in the recruitment of U1 snRNA to the major late first leader 5′ splice site. This enhanced U1 snRNA recruitment leads to a suppression of MLP transcription and simultaneously an increase of major late first intron splicing.

Keywords: L4-22K, MLP, RNA, transcription, splicing, adenovirus

Susan Lan, Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden.

© Susan Lan 2016

ISSN 1651-6206
urn:nbn:se:uu:diva-278135 (http://urn.kb.se/resolve?urn=nbn:se:uu:diva-278135)
To my family and beloved teachers

献给我的家人和爱师
List of Papers

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


III Lan, S., Akusjärvi, G. (2016) The adenovirus L4-22K protein regulates major late transcription and RNA splicing through a sequence-specific binding to single-stranded RNA. *Submitted*


Reprints of paper I and II were made with permission from Elsevier.
Members of the committee

Opponent

Per Elias, Professor  
Department of Medical biochemistry and Cell Biology  
University of Gothenburg

Members of the committee

Birgitta Tomkinsonm, Professor  
Department of Medical Biochemistry and Microbiology  
Uppsala University

Hongxing Zhao, Dr.  
Department of Immunology, Genetics and Pathology  
Uppsala University

Mikael Berg, Professor  
Department of Biomedical Sciences and Veterinary Public Health  
Swedish University of Agricultural Sciences
Contents

Introduction .................................................................................................................. 11

Transcription .................................................................................................................. 12
  The promoter .............................................................................................................. 12
  Initiation .................................................................................................................... 13
  Elongation .................................................................................................................. 14
  Transcription factors ............................................................................................... 14
  Transcription activation ......................................................................................... 14
  The coupling of transcription and RNA splicing ..................................................... 15

RNA splicing ................................................................................................................. 17
  Splicing signals and elements ................................................................................. 18
  The spliceosome ....................................................................................................... 18
  Splicing factors ........................................................................................................ 19
  Regulatory elements ............................................................................................... 20

Polyadenylation ............................................................................................................. 22

Adenovirus .................................................................................................................... 23
  Genome organization .............................................................................................. 24
  The Ad life cycle ..................................................................................................... 25
  The MLP and the MLTU ......................................................................................... 27
  The L4 region ........................................................................................................... 29

Present investigation .................................................................................................... 30
  Paper I ....................................................................................................................... 30
  Paper II ..................................................................................................................... 32
  Paper III ................................................................................................................... 34
  Paper IV ................................................................................................................... 36

Concluding remarks ..................................................................................................... 38

Acknowledgments ....................................................................................................... 42

References .................................................................................................................... 45
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad-NE</td>
<td>Nuclear extract prepared from adenovirus infected HeLa cells</td>
</tr>
<tr>
<td>Adpol</td>
<td>Ad DNA polymerase</td>
</tr>
<tr>
<td>ARD</td>
<td>Acute respiratory disease</td>
</tr>
<tr>
<td>BRE</td>
<td>TFIIB recognition element</td>
</tr>
<tr>
<td>CPSF</td>
<td>Cleavage/polyadenylation specificity factor</td>
</tr>
<tr>
<td>CstF</td>
<td>Cleavage-stimulating factor</td>
</tr>
<tr>
<td>CTD</td>
<td>Carboxy-terminal domain of the large subunit of RNA Pol II</td>
</tr>
<tr>
<td>C-terminal</td>
<td>Carboxy-terminal</td>
</tr>
<tr>
<td>DE element</td>
<td>Downstream element</td>
</tr>
<tr>
<td>DBD</td>
<td>DNA-binding domain</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dsDNA</td>
<td>double-stranded DNA</td>
</tr>
<tr>
<td>EC</td>
<td>Early elongation complex</td>
</tr>
<tr>
<td>EKC</td>
<td>Epidemic keratoconjunctivitis</td>
</tr>
<tr>
<td>ESE</td>
<td>Exonic splicing enhancer</td>
</tr>
<tr>
<td>ESS</td>
<td>Exonic splicing silencer</td>
</tr>
<tr>
<td>HAdV</td>
<td>Human adenovirus</td>
</tr>
<tr>
<td>hnRNP</td>
<td>heterogeneous nuclear ribonucleoprotein</td>
</tr>
<tr>
<td>Inr</td>
<td>Initiator</td>
</tr>
<tr>
<td>ISE</td>
<td>Intronic splicing enhancers</td>
</tr>
<tr>
<td>ISS</td>
<td>Intronic splicing silencers</td>
</tr>
<tr>
<td>ITC</td>
<td>Initial transcribing complex</td>
</tr>
<tr>
<td>ITRs</td>
<td>Inverted terminal repeats</td>
</tr>
<tr>
<td>L4</td>
<td>Late region 4 of MLTU of adenovirus</td>
</tr>
<tr>
<td>L4-22K</td>
<td>Protein of 22kDa expressed from the L4 region</td>
</tr>
<tr>
<td>L4-33K</td>
<td>Protein of 33kDa expressed from the L4 region</td>
</tr>
<tr>
<td>MLP</td>
<td>Major late promoter</td>
</tr>
<tr>
<td>MLTU</td>
<td>Major late transcription unit</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>NE</td>
<td>Nuclear extract prepared from HeLa cells</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>N-terminal</td>
<td>Amino-terminal</td>
</tr>
<tr>
<td>OC</td>
<td>Open promoter complex</td>
</tr>
<tr>
<td>ORFs</td>
<td>Open reading frames</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PABP</td>
<td>Poly (A) binding protein</td>
</tr>
<tr>
<td>PAP</td>
<td>Poly (A) polymerase</td>
</tr>
<tr>
<td>PIC</td>
<td>Pre-initiation complex</td>
</tr>
<tr>
<td>Pre-mRNA</td>
<td>Precursor mRNA</td>
</tr>
<tr>
<td>PPY</td>
<td>Polypyrimidine tract</td>
</tr>
<tr>
<td>pTP</td>
<td>precursor terminal protein</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNA pol</td>
<td>RNA polymerase</td>
</tr>
<tr>
<td>RNAP II</td>
<td>RNA Polymerase II</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal RNA</td>
</tr>
<tr>
<td>RRM</td>
<td>RNA recognition motifs</td>
</tr>
<tr>
<td>SF1</td>
<td>Splicing factor 1</td>
</tr>
<tr>
<td>SR protein</td>
<td>Serine and Arginine rich protein</td>
</tr>
<tr>
<td>snRNA</td>
<td>small nuclear RNA</td>
</tr>
<tr>
<td>ss</td>
<td>splice site</td>
</tr>
<tr>
<td>TAD</td>
<td>Trans-activation domain</td>
</tr>
<tr>
<td>TAF</td>
<td>TBP-associated factor</td>
</tr>
<tr>
<td>TAT-SF1</td>
<td>Tat specific factor 1</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>U2AF</td>
<td>U2 auxiliary factor U2AF</td>
</tr>
<tr>
<td>UPE</td>
<td>Upstream promoter element</td>
</tr>
<tr>
<td>USF</td>
<td>Upstream stimulatory factor</td>
</tr>
<tr>
<td>U snRNPs</td>
<td>Uridine-rich small nuclear ribonucleoprotein particles</td>
</tr>
<tr>
<td>VA RNAI</td>
<td>Virus associated RNA I</td>
</tr>
<tr>
<td>VA RNAII</td>
<td>Virus associated RNA II</td>
</tr>
</tbody>
</table>
Introduction

The genetic information is stored in the cell nucleus as a double stranded DNA molecule. The DNA is transcribed to a precursor messenger RNA (pre-mRNA) by RNA polymerase II. The pre-mRNA is subsequently subject to post-transcriptional modifications: 5′-end capping, RNA splicing and polyadenylation to form a mature mRNA. The mRNA is then exported from the nucleus to the cytoplasm where it is translated into an amino-acid sequence, which is finally folded into a native protein.

Eukaryotic cells can be infected by different pathogens, such as viruses, bacteria, and parasites. Viruses are obligate parasites. Replication of their DNA or RNA genomes relies on the transcription, pre-mRNA processing and translation machineries of the host cell. In its simplest form, a virus consists of a DNA or RNA genome encapsidated by a protein shell. The size of the genome that can be introduced into a given virus capsid is limited. Therefore, viruses have to maximize their own coding potentials in order to express the proteins needed to facilitate virus multiplication.

Adenovirus infects various vertebrates, including humans, and typically causes upper respiratory tract and gastroenteritis infections as well as conjunctivitis. Its early to late switch of gene expression is strictly regulated. The adenoviral major late promoter (MLP) is a strong promoter, driving transcription from the major late transcription unit (MLTU) and producing maximal amount of structure proteins during the late phase of infection. The basal MLP contains upstream activating elements and downstream elements (DE elements) located within the first intron. The binding of transcription factors to these elements controls the activity of the MLP. This project focuses on the impact that the virus-encoded L4-22K protein has on MLP activity, both as a regulator of transcriptional and post-transcriptional events. The multifunctional L4-22K protein is important for the temporal regulation of the adenoviral life cycle.
Transcription

Transcription is the process where the DNA is transcribed to an RNA transcript by an enzyme called the RNA polymerase (RNA pol). All eukaryotes use (at least) three nuclear RNA polymerases: RNA pol I transcribes genes coding for the ribosomal RNA (rRNA) and RNA pol III catalyzes transcription of small RNAs like transfer RNA (tRNA), 5S ribosomal RNA (5S rRNA), a few primary microRNAs (pri-miRNA) and some other small RNAs. RNA pol II (RNAP II), which is the main interest of this thesis, transcribes protein-encoding pre-mRNAs, small nuclear RNAs (snRNAs), long non-coding RNAs (lincRNA) and the majority of pri-miRNAs (128). To regulate transcription, RNAP II relies on temporal changes of phosphorylation status of its C-terminal domain (CTD), which consists of 52-tandem repeats of the amino acid sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser (YSPTSPS) in human RNAP II.

The promoter

RNAP II transcription is initiated by the polymerase binding to a core promoter, whose activity is controlled by local regulatory elements typically positioned 100-500 base pairs (bp) upstream of the transcription start site. The whole region is referred to as the promoter. In addition, distant sequences called enhancers or silencers also contribute to regulation of the promoter activity (35). The core promoter extends approximately 35 bp upstream of the initiation site but in some cases includes around 35 bp downstream (118).

The core promoter can be divided into different regions: the TATA box, highly conserved among eukaryotes, is located 20-25 bp upstream of the start site and is bound by the TATA-binding protein (TBP), a subunit of the general transcription factor IID (TFIID); the initiator region (Inr), which is located at the transcription start site, is recognized by the TBP-associated factor 1 (TAF1) and TAF2 (56, 66, 118); the TFIIB recognition element (BRE), which can be positioned both upstream and downstream of the transcriptional start site (52). Dependent on the promoter additional elements might be present in the core promoter.
Initiation

The transcription initiation event is well characterized. It begins with the recognition of the TATA box by TFIID. This binding is stabilized by TFIIA that binds to TFIID and counteracts factors binding to repressive elements in the promoter. TFIIIB is recruited and binds to the BRE (60, 114). This is followed by TFIIIF, which binds to hypophosphorylated RNAP II complex and recruits it to the promoter. Finally the TFIIE and TFIIH factors are recruited. TFIIH, which has helicase activity, helps RNAP II to access to the template by unwinding the DNA strands ("promoter melting") (14, 42). After formation of the pre-initiation complex (PIC) the transcription starts with the production of short transcripts (abortive transcription) due to the promoter-proximal pausing (1).

Figure 1. The stepwise initiation of transcription by the RNAP II (modified from (100)).

During the assembly process the CTD is hypophosphorylated. However, after PIC formation, TFIIH phosphorylates serine 5 in the heptad repeats, a process that stimulates the transition from the initiation to the elongation phase of transcription (117, 135).
Elongation

As the initiation phase moves towards the elongation phase, a transient open promoter complex (OC) is converted to an initial transcribing complex (ITC). As TFII B and RNAP II are stabilized, short RNAs (2-6 nts) are produced (17). When 7-nucleotide-long RNAs are generated, the transcription complex is remodeled to form an early elongation complex (EC), along with the release of TFII B when the RNAP II leaves the promoter. Negative elongation factors that interact with the hypophosphorylated CTD in the Pol II/RNA complex make the RNAP II pause at the 5′ region of the transcription unit (promoter-proximal pausing).

When around 20-nucleotide-long RNA has exited the RNAP II, capping enzymes, which have been recruited to the Ser5 phosphorylated CTD tail, add a 7-methylguanosine cap to the nascent RNA (46). This cap is required for mRNA stability and translational efficiency. Following this step, the positive transcription elongation factor b (P-ETFb) phosphorylates the Ser2 residues in the CTD heptad repeats, relieving the impacts of negative elongation factors. Therefore, transcription elongation is resumed via the association of P-ETFb and other positive elongation factors. For example, TFIIF and elongin (SIII) stimulate the catalytic rate of elongation, and TFIIS suppresses the RNAP II arrest (93).

Transcription factors

Transcription factors are sequence-specific DNA-binding proteins that stimulate or repress transcription (activators or repressors, respectively). Most of transcription factors bind DNA as dimers and belong to the protein families of related proteins with the same type of domains for DNA binding and dimerization. The activators contain a DNA-binding domain (DBD) and one or more trans-activation domains (TAD). The TADs might share properties such as acidic amino acids (VP16, p53) or be enriched in certain amino acids like glutamine (Sp1). Some TADs transmit a signal by binding directly to the basal transcription factors, whereas others interact via cofactors (reviewed in (67)). Transcription factors can also contain regulatory domains for ligand binding (nuclear receptors) or interaction surfaces for binding of regulatory proteins (for example, NFkB-IkB).

Transcription activation

Transcription can be activated by diverse mechanisms: recruitment of basal transcription factors and RNAP II, conformational changes in the PIC, phosphorylation of the CTD, removal of repressors, or stimulation of promoter
clearance or elongation. The TADs can be divided into three classes: those that stimulate initiation (type I), elongation (Type IIA, predominantly) and both initiation and elongation (type IIB), respectively (11). These functional differences might result from the different general transcription factors that the activators interact with. Type I and IIA activators stimulate transcription synergistically (11). Mechanisms by which transcription factors synergistically activate promoters include chromatin remodeling, cooperative DNA binding and stabilization of interaction with general transcription factors.

Chromatin remodeling has received much attention as a regulatory mechanism in transcription since the access of basal transcription factors and the RNAP II to a promoter region can be controlled by modulating the structure of chromatin. In a simplistic model acetylation of the N-terminal tails of histones loosens up the chromatin structure and therefore enhances transcription whereas methylation of the histone tails results in chromatin compression and as a consequence transcription repression.

The coupling of transcription and RNA splicing

Transcription and RNA splicing are extensively co-regulated. Thus, RNAP II transcription directly stimulates pre-mRNA splicing (7, 39, 59, 63, 70) and the promoter strength also affects the outcome of alternative RNA splicing events (29, 53). These coupled events are believed to, at least in part, rely on the fact that the CTD binds RNA splicing factors that become tethered to the active transcription sites (26, 85). The coupling between these two processes can be illustrated by the facts that a hyperphosphorylated CTD tail stimulates RNA splicing, whereas the hypophosphorylated tail does not (9, 45). Deletion of the CTD tail inhibits splicing (9, 81), and tethering of the phosphorylated CTD tail to a heterologous RNA binding protein increases splicing efficiency (83). Furthermore, the rate of transcription, transcription factors and cofactors and chromatin structure are important regulators of splicing.

Interestingly, introns are required for efficient gene expression because RNA splicing enhances mRNA export from the nucleus to cytoplasm (110). Further, introns increase the efficiency of polyadenylation (127), which in turn also promotes higher mRNA levels. Removal of a promoter proximal 5´ splice site (ss) or increased spacing between the 5´ss and transcriptional start site reduces transcription efficiency (36). A 5´ss stimulates transcription even in the absence of splicing because U1 snRNP binding to a 5´ss stimulates the recruitment of basal transcription initiation factors during PIC assembly (19). Further, highly purified fractions of TFIIH contain the spliceosomal U1 snRNA, which also has a function as an enhancer of transcription initiation (62). U snRNPs association with the transcription-splicing factor Tat specific factor 1 (TAT-SF1) and the elongation factor P-ETFb also par-
ticipates in transcription activation by enhancing transcription elongation (33).

Further, there is a highly conserved 10-residue motif, IPSDDSDEEN, located at the C-terminus of the CTD, which in its phosphorylated form is required for efficient transcription, splicing and 3’ end processing (32).
RNA splicing

A pre-mRNA is composed of exons (sequences retained in the final mRNA, often protein-coding sequences) and introns (sequences removed by RNA splicing). RNA splicing is the process where exons are ligated together and introns are excised, producing mature mRNAs. There are two types of splicing: constitutive splicing generating only one type of mRNA from a pre-mRNA and alternative RNA splicing where exons are combined in alternative ways to produce multiple mature mRNAs with different coding capacities and therefore with potentially different biological functions.

RNA splicing and alternative RNA splicing was discovered in the adenovirus system in 1977. Generally, alternative RNA splicing occurs when splicing signals are weak and the usage of splice sites relies on the accumulation of diverse splicing factors. By this type of splicing, multiple protein isoforms are produced, in specific tissues, or in response to different environmental stimuli, adding to the complexity of organisms and evolutionary diversification. Adenovirus, as many other small DNA viruses, expands its genetic information by producing a complex set of alternatively spliced mRNAs. Alternative splicing is considered as a hallmark of higher multicellular eukaryotes and it is most extensively used in vertebrates. In humans, more than 90% of all genes produce alternatively spliced mRNAs. In general, the splice sites in lower eukaryotes are more conserved than those in higher eukaryotes, thereby limiting the possibility for lower eukaryotes to use alternative RNA splicing as a mean to expand the coding capacity.

There are five basic models of alternative splicing events: exon skipping, mutually exclusive exons (cassette exon), alternative 5’ss selection, alternative 3’ss selection and intron retention (57). All of these different types of splicing events can be regulated by trans-acting factors, cell type specific factors and post-translational modification of both types of factors. Also, splicing occurs co-transcriptionally (35). The phosphorylation status of the CTD, available transcription factors and cofactors, the chromatin structure and the rate of transcription are important parameters regulating alternative RNA splicing (54, 90). In a simplistic model a slow rate of transcription favors exon inclusion whereas a rapidly moving RNA polymerase presents multiple splice sites for spliceosome assembly and as a consequence an increase in the accumulation of alternatively spliced mRNAs.
Splicing signals and elements

Pre-mRNA splicing signals are the sequences defining the borders of exons and introns. The upstream border between an exon and an intron is defined as the 5´ss and the downstream intron/exon boundary as the 3´ss. There are four classes of introns: classes I, II and III are ribozymes that catalyze their own splicing (30, 89, 124, 129) whereas class IV is the nuclear introns that are spliced by the spliceosome. The class IV introns are further subdivided into U2- and U12-dependent introns that are spliced by the major and the minor spliceosome, respectively. Most U2-dependent introns contain GU-AG dinucleotide pairs at the 5´ and 3´ splice sites, respectively, whereas U12-dependent introns contain AU-AC splice site consensus sequences (122, 129). In this thesis splicing only refers to the U2-dependent introns and the major spliceosome.

Figure 2. Splicing signals in an U2-dependent intron. Exons are illustrated as boxes and the intron as the horizontal line. ss=splice site

At the 5´ end of an intron the consensus sequence AG/GU acts as a splice donor site during splicing catalysis. The 3´ end of the intron contains the branch point element (18–40 nucleotides upstream of the 3´ss) harboring one or two conserved adenine residues that form the intermediate lariat structure during the first step of catalysis. The polypyrrimidine tract (PPY, rich in U residues) is positioned between the branch point sequence and the conserved AG dinucleotide at the 3´ss that functions as the splice acceptor site during the second step of catalysis.

The spliceosome

The spliceosome is assembled de novo on the pre-mRNA in the nucleus. It consists of five uridine-rich small nuclear ribonucleoprotein particles (U snRNPs): the U1, U2, U4, U5, and U6 snRNPs, splicing factors and many other proteins. In total the spliceosome contains more than 170 proteins with a size exceeding 4.8 MDa. The U snRNPs are composed of small nuclear RNA molecules (U1, U2, U4, U5 and U6 snRNA) and seven common Sm or Sm-like proteins. In addition each snRNP possess its own particle-specific proteins. In spliceosome assembly the snRNPs serve a key function as factors recognizing the 5´ and 3´ ss via a direct RNA-RNA base pairing (132).
The spliceosome assembly pathway has been well defined in vitro (Fig. 3). First, the pre-mRNA is committed to the spliceosome assembly pathway by U1 snRNP binding to the 5’ss, thereby forming the E complex. The 3’ss is defined by splicing factor 1 (SF1) binding to the branch point sequence and U2 snRNP auxiliary factor (U2AF) interacting with the pyrimidine tract at the 3’ss. The E complex is converted to the A complex by recruitment of U2 snRNP to the branch point. U2AF binding to the pyrimidine tract promotes U2 snRNP recruitment to the pre-mRNA (31). In the subsequent step the preformed tri-snRNP (U4/U5/U6) arrives to form the B complex. Through major rearrangements of RNA-RNA and RNA-protein interactions the U1 and U4 snRNPs dissociate from the B complex, forming a catalytically active spliceosome that processes the first step in splicing. This yields the C complex, which completes the catalytic activity of pre-mRNA splicing (reviewed in (109)). After catalysis the spliceosome disassembles and releases the mature mRNA. The released snRNPs are reused for additional rounds of splicing.

Figure 3. Spliceosome complexes. Only U snRNPs assembled on the pre-mRNA for different spliceosome complexes are illustrated.

Splicing factors

In addition to the snRNPs, the splicing process requires a large number of additional proteins, including spliceosomal core components, spliceosome assembly factors, regulatory factors (enhancer or repressor proteins) and enzymes modifying protein activity by post-translational modifications. Two protein families, the SR (serine and arginine rich proteins) and the hnRNP
(heterogeneous nuclear ribonucleoproteins) proteins have received much attention (116).

SR proteins are key regulators of spliceosome assembly. They possess one or two N-terminal RNA recognition motifs (RRMs) and a domain rich in arginine serine dipeptide repeats (the RS domain) at the C-terminus. The RRM s determine the specificity of RNA binding. In general the RRMs are not highly sequence-specific and splicing factors therefore interact with a range of elements with similarity in sequence composition (reviewed in (72)), providing multi-functions in splicing regulation. SR proteins can function both as enhancer and repressor proteins in spliceosome assembly. The RS domain mediates interactions between the pre-mRNA and other splicing factors to regulate spliceosome assembly and the selection of splice sites. For example, SRSF1 (previously known as ASF/SF2) by binding to an exonic splicing enhancer (ESE) recruits and stabilizes U1 snRNP binding to a downstream 5´ss while simultaneously assisting in the recruitment of U2AF to the upstream 3´ss. Also, the phosphorylation status of the RS domain in an SR protein controls its activity as a splicing factor and the nuclear localization and/or shuttling between the nucleus and the cytoplasm (88). The proteins in the SR family are structurally related and phylogenetically conserved.

The hnRNPs contain one or two RRMs plus an auxiliary domain. In contrast to SR proteins they typically act as negative regulators of splicing. Many alternative splicing events are co-regulated by an SR protein and an hnRNP protein. For example, hnRNP A1 binding to splicing silencer elements inhibits the usage of the closest 5´ss, therefore favoring the distal 5´ss selection, while SRSF1 binding to an ESE favors the proximal 5´ss usage. Similarly, PPY binding protein (PTB) competes with U2AF binding to a PPY. U2AF activates 3´ss usage whereas PTB binding inhibits splicing (120).

U2AF, which is a SR-related protein, recruits U2 snRNP to the branch point and is essential for 3´ss definition (18). U2AF is a heterodimer of the 65 kD protein (U2AF<sup>65</sup> subunit) that binds to the PPY (18) and 35 kD protein (U2AF<sup>35</sup> subunit) that interacts with the AG dinucleotide at the 3´ss, thereby stabilizing the interaction of U2AF<sup>65</sup> with the PPY (61).

The ratio of different splicing factors determines the efficiency of splice site selection since positive and negative factors affect alternative splice site selection simultaneously (reviewed in (119)).

---

**Regulatory elements**

There are several splicing regulatory elements that splicing factors interact with. They are classified according to their position and effect on splicing. Thus, exonic splicing enhancers (ESE) activate splicing whereas exonic...
splicing silencers (ESS) inhibit splicing. The same nomenclature can be used for elements positioned in the intron: intronic splicing enhancers (ISE) and intronic splicing silencers (ISS). Enhancers are more common in exons than in introns and vice versa for silencers. ESEs mostly interact with SR proteins and the binding recruits and stabilizes U1 snRNP binding to the downstream 5´ss. SR proteins also enhances the recruitment of U2AF to the upstream 3´ss, which results in the definition of an exon as a unit (the so-called exon definition model) (144).

Splicing silencers typically bind hnRNP proteins, a binding that most often blocks splice site usage (121). The function of SR proteins as activators or repressors of splicing depends on where in the pre-mRNA they bind: binding to exons activates splicing whereas binding to an intronic position leads to splicing repression (115). ESSs function best on suboptimal splice sites that are engaged in alternative splicing. By disturbing the recruitment of splicing factors to a 3´ss or binding of SR proteins to an ESE, factors binding to an ESSs can inhibit upstream 3´ss recognition and exon definition (12, 144).
Polyadenylation

The final step in pre-mRNA synthesis is the polyadenylation reaction (or so-called 3′ end processing). The core elements controlling polyadenylation are the AAUAAA sequence located around 10-30 nucleotides upstream of the actual cleavage site on the pre-mRNA and a G/U-rich element directly downstream of the cleavage site. When polyadenylation initiates, a multimeric protein complex is assembled at the poly (A) site. The first step in the assembly involves the binding of the cleavage/polyadenylation specificity factor (CPSF) to the AAUAAA sequence followed by a binding of the cleavage-stimulating factor (CstF) to the G/U-rich element. These two events orchestrate the recruitment of cleavage factor I (CFI), cleavage factor II (CFII) and the poly (A) polymerase (PAP) to the polyadenylation site (20). CFI and CFII direct an ATP-dependent cleavage of the pre-mRNA (76), followed by a PAP-dependent extension of a 200-300-nt-long poly (A) tail to the newly cleaved 3′ hydroxyl end (46). The poly (A) tail, which is covered by the poly(A) binding protein (PABP) plays a crucial role in the stability, nuclear export and translation of the mature mRNA (76).

There is clear evidence that polyadenylation and RNA splicing are coregulated. SR proteins have been shown to regulate the efficiency of polyadenylation. Further, both CPSF and a subunit of CstF have been shown to interact with the hyperphosphorylated form of the CTD (106), indicating that splicing and polyadenylation are coupled events.
Adenovirus

Viruses rely on the host cell to replicate and express its genes. Viruses are considered as useful model systems to study various processes within a eukaryotic cell, like transcription regulation and pre-mRNA splicing, etc. since they are easily modified and express potent regulatory factors that interfere with the biosynthetic machineries in the cell.

Table 1. Classification of human adenoviruses (modified from (100)).

<table>
<thead>
<tr>
<th>Species</th>
<th>Types</th>
<th>Tumorigenic in animals</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAdV-A</td>
<td>12, 18, 31, 61</td>
<td>High</td>
<td>Enteric infection</td>
</tr>
<tr>
<td>HAdV-B</td>
<td>3, 7, 11, 14, 16, 21, 34-35, 50, 55, 66</td>
<td>Moderate</td>
<td>Conjunctivitis Acute respiratory disease Hemorrhagic cystitis (CNS)</td>
</tr>
<tr>
<td>HAdV-C</td>
<td>1-2, 5-6, 57</td>
<td>Low or none</td>
<td>Endemic infection Respiratory symptoms</td>
</tr>
<tr>
<td>HAdV-D</td>
<td>8-10, 13, 15, 17, 19-20, 22-30, 32-33, 36-39, 42-49, 51, 53-54, 56, 58-60, 62-65, 67</td>
<td>Low or none</td>
<td>Keratoconjunctivitis in immunocompromised patients</td>
</tr>
<tr>
<td>HAdV-E</td>
<td>4</td>
<td>Low or none</td>
<td>Conjunctivitis Acute respiratory disease</td>
</tr>
<tr>
<td>HAdV-F</td>
<td>40-41</td>
<td>Unknown</td>
<td>Infantile diarrhea</td>
</tr>
<tr>
<td>HAdV-G</td>
<td>52</td>
<td>Unknown</td>
<td>Gastroenteritis</td>
</tr>
</tbody>
</table>

Adenoviruses (Ads) are small non-enveloped icosahedral viruses infecting a wide range of hosts (21). Of the approximately 100 Ad types characterized so far 67 have been shown to infect humans. The human types have been classified into seven subgroups: A to G (Table 1), based on the capability to agglutinate red blood cells, oncogenic potential in rodents, genome homology and immunological characteristics (51, 130).
Human adenoviruses (HAdVs) typically cause acute respiratory disease (ARD), pharyngitis, mild infections in conjunctiva or epidemic keratoconjunctivitis (EKC), gastrointestinal infections and urinary tract infections (Table 1). More severe infections might occur in immunocompromised patients (for instance, AIDS, immunosuppressive treatment for organ or tissue transplantation and cancer) and patients with simultaneous infections or individual susceptibility (44, 48). Ads can stay in a dormant state, causing long-term persistent infection in specific cells/tissues of the host, if they are not cleared after the initial amplification. It is not uncommon that the virus reactivates in immunocompromised individuals (25, 38). Even though Ads have not been associated with human cancer, members of species A efficiently induce cancer in rodents (92). The rate of Ad infection relies on the patient age, type of immunosuppression, background disease and so on. Today, there are no specific antiviral therapies against adenovirus infections.

In this thesis we worked with Ad serotype 2 and 5 (Ad2, Ad5), which belong to the non-oncogenic species C and are commonly used for studies of gene regulation. Ad5 also functions as the principal adenoviral vector in cancer/gene therapy. The residual activity of the late genes from the early generation of Ad vectors results in a low persistence of transgene expression (82, 139, 142). Therefore, an understanding of the regulatory events controlling late gene expression during a lytic infection needs to be further clarified in order to develop improved adenovirus delivery vehicles for cancer/gene therapy.

Genome organization

Members of the Ad family have linear double-stranded DNA genomes that range in size from 26-45 kbp and encode for some 30 to 40 proteins, depending on the Ad type. Both strands of the Ad DNA contain open reading frames (ORFs) and are transcribed into cytoplasmic mRNAs. The genome contains inverted terminal repeats (ITRs) at both ends, which functions as the origins for viral DNA replication (47). The viral genes, produced by RNAP II, are divided into early (E1A, E1B, E2A, E2B, E3, and E4), intermediate (pIX and IVa2) and late transcription units (L1-L5) (Fig. 4) based on the time of expression during the infectious cycle (Figure 4).
Figure 4. Schematic map of the human adenovirus 5 genome. Early genes are indicated by light blue arrows, the intermediate genes in green, the late genes in yellow and VA RNAs in pink.

The expression of early genes initiates when the Ad genome has been delivered to the cell nucleus and continues at different levels during the entire lytic life cycle. In a simplistic view the early genes mainly function to force the host cell into the S-phase (E1A), block apoptosis (E1B), stimulate viral transcription (E1A), activate viral DNA replication (E2) and stimulate production of viral proteins required for evasion of the host cell immune response (E3), (78, 130). Following initiation of viral genome replication the pattern of viral gene expression switches to a production of the late structural proteins required for viral capsid formation. In addition, some viral proteins with regulatory functions are also expressed. The intermediate protein pIX is a transcriptional activator and also serves as a glue protein that stabilizes the virus capsid (104). The IVa2 protein is involved in viral genome packaging (95) and has also been implicated as a factor stimulating transcription from the major late promoter (MLP) (102), which directs transcription from the so-called major late transcription unit (MLTU).

In addition, the Ad genome encodes for two RNA polymerase III products, the virus associated RNA I and II (VA RNAI and VA RNAII) that accumulate primarily during the late phase of infection (79). These small, non-coding RNAs form hairpin structures and accumulate to large amounts at late times of infection (109, 110). They serve an important function as viral factors protecting the infected cell from the cellular immune responses triggered by interferon (80) and the RNAi/miRNA systems (4, 138).

The Ad life cycle

The Ad particles enter into cells via the endosomal pathway. The fibers protruding from the corners of the icosahedral capsid interact with the cellular coxsackievirus and Ad receptor (CAR), which is the primary Ad receptor, whereas the penton at the base of the fiber interacts with cellular integrins, which function as secondary receptors (reviewed in (47)). In the acidic endo-
somal environment the virus particle is partially dismantled (94), followed by a degradation of the endosomal membrane triggered by the viral polypeptide VI, which leaks out from the interior of the capsid. The partly dismantled capsid is released into the cytoplasm (131) and transported on microtubuli to the nuclear membrane where the viral genome is imported to the nucleus in a process assisted by nuclear import receptors and the viral pVII protein (133).

The early phase of an Ad infection reprograms the cell for efficient virus replication. E1A is the first viral gene to be expressed. E1A encodes for two major mRNAs, the 12S and the 13S mRNAs, which encodes for proteins that function as activator of the other viral transcription units and repressor of multiple cellular genes. Further, E1A is required to drive the infected cell into the S phase (34). The proteins expressed from the E1B region have a major function as suppressors of apoptosis. The E1B-55K protein is multifunctional. Two of its activities directly relate to a control of apoptosis. First, E1B-55K binds to p53 and blocks its function as a transcriptional activator. Second, E1B-55K together with E4-ORF6 forms a complex that induces proteasomal degradation of p53 via the ubiquitination pathway (105). The E1B-55K/E4-ORF6 complex also facilitates cytoplasmic transport of late viral mRNAs (10). The E4 unit encodes for several proteins regulating transcription, RNA processing and additional events. For example, the E4-ORF6 and E4-ORF3 proteins block the cellular DNA repair system to prevent formation of Ad DNA concatemers during infection, and further have been shown to enhance the accumulation of spliced late mRNAs (91, 123). The E4-ORF4 promotes the early to late shift in gene expression by inducing dephosphorylation of SR proteins (e.g. SRSF1 and SRSF9) and transcription factors (e.g. E1A, AP1) (13, 55, 123). The E3 unit encodes for the viral proteins that suppress the host cell immune response. The best studied example is the E3-gp19K protein that sequesters MHC class I antigens in the endoplasmic reticulum thereby rescuing Ad infected cells from being recognized by cytotoxic T lymphocytes (69).

Ad DNA replication occurs in discrete foci in the nucleus, called nuclear replication centers. Three viral proteins encoded by the E2 region are required for viral DNA replication: the Ad DNA polymerase (Adpol), the precursor terminal protein (pTP) and the single-stranded DNA binding protein (E2A-72K) (71). The replication initiates when cellular factors NFI and NFIII bind to Adpol and pTP, respectively, forming the pre-initiation complex at the end of the genome (23). The terminal nucleotide of viral DNA is covalently attached by pTP, followed by synthesis of the nascent DNA by Adpol in a 5’ to 3’ direction. Finally, Adpol elongates the newly generated strand to full-length DNA with the help of the E2A-72K that coats the second displaced strand of viral DNA (22). Simultaneously with viral DNA replication occurring in the middle of the replication centers, Ad late tran-
scription and RNA splicing take place in ring-like structures surrounding the replication centers (5, 15, 37).

After initiation of Ad DNA replication, the MLTU becomes fully activated, producing all the structural proteins required for capsid assembly, except the pIX protein, which is expressed from a separate transcription unit embedded within the E1B region.

The structural proteins are imported into the nucleus, followed by oligomerization of the capsid proteins and then assembly of the empty procapsids. The Ad genome is packaged into the empty capsid through one of the open vertexes. The viral proteins involved in the assembly process are IVa2, L4-22K, L4-33K and L1-52,55K (98, 99). The capsid is maturated to an infectious particle by the L3 protease, which proteolytically cleavages six precursor proteins in the virus (pIIIa, pTp, pVI, pVII and p\u2013). Finally, the infected cell lyses releasing the newly formed infectious viruses. The lysis process is promoted by the adenoviral death protein (E3-11.6K also called ADP (98)).

The MLP and the MLTU

The adenoviral MLP is active both early and late after infection but shows a dramatic increase in activity following initiation of viral DNA replication. MLP drives transcription from the MLTU (Fig. 5), which generates more than 20 cytoplasmic mRNAs through an extensive use of alternative 3′ss and poly(A) site selection. The mRNAs from the MLTU are grouped into five families, L1-L5, where each family members share the same 3′ polyadenylation site. The basal MLP contains two upstream activating elements, the CAAT box and the upstream promoter element (UPE), a canonical TATA box, an initiator (Inr) covering the transcription start site as well as highly GC-rich regions surrounding the TATA box (113). The GC-rich regions are needed for high levels of transcription (16, 103).

The CAAT box binds the cellular transcription factor CP1 (also named CBF/NF-Y) whereas the UPE interacts with the cellular upstream stimulatory factor (USF). The TATA box, which interacts with TBP, is the most critical element controlling MLP transcription (40, 58). Point mutations in the CAAT box, UPE, TATA box or Inr only slightly affect transcription from the MLP in vivo (73, 107, 108, 143). However, transcription is diminished if several MLP elements simultaneously are mutated. Both the TATA box and Inr interact with TFII, promoting PIC formation. TFII-I binds to both the UPE and Inr sequences (111). In addition the Inr can also interact with USF, stimulating transcription (24). Together with the protein-protein interactions of TFII-I/USF (112), the functional robustness of the MLP can be explained.

The MLP also contains downstream elements (DE elements, +86 to +116) located within the major late first intron. The DE elements, which stimulate
MLP transcription, can be subdivided into DE1 and DE2a/b. The DE elements interact with two infection-dependent viral trans-acting factor complexes: DEF-A binds to both the DE1 and DE2a and DEF-B binds to the DE2b element (49, 68, 77). DEF-B is suggested to be a homodimer of the viral IVa2 protein, whereas DEF-A has been proposed to consist of a heterodimer of IVa2 and an unknown viral protein (74). Two adenoviral proteins L4-22K (27, 97, 136, 140, 141) and L4-33K (3) are potential candidates to be the missing viral component of DEF-A. Binding of IVa2 to the DE elements is required for efficient late-specific activation of the MLP in the context of a virus infection (126).

The MLTU produces an approximately 28 000 nt long pre-mRNA that is processed to produce essentially all mRNAs encoding for the structural proteins of the viral capsid. Activation of the MLTU takes place following the onset of viral DNA replication, except for the L1-52,55K mRNA, which is produced already during the early phase of infection. All mRNAs processed from the MLTU pre-mRNA share a tripartite leader sequence at their 5¢ end, which is needed for efficient translation during the late phase of infection. The tripartite leader consists of three small exons, where the 5¢ss of the third exon individually combines with the alternative 3¢ss forming different late viral mRNAs. The exclusive expression of the L1-52,55K mRNA at early times of infection is subjected to a complex regulation at multiple levels: transcription elongation, alternative splicing and polyadenylation. At a transient stage following the initiation of viral genome replication, mRNAs from both the L1 and L4 transcription units are selectively expressed (65). This expression pattern is then followed by a general increase in the production of mRNAs from the whole MLTU (L1-L5) (2).

Figure 5. Schematic drawing showing the organization of the MLTU and MLP (modified from (64)).
The L4 region

The L4 region encodes for five proteins, pVIII, two isoforms of L4-100K, L4-33K and L4-22K. The pVIII protein is a structural protein of the viral capsid whereas the other proteins are regulatory factors controlling translation and the temporal early to late shift in adenoviral gene expression. The L4-100K aids in the assembly of viral hexon trimers and functions as an enhancer protein stimulating translation of viral mRNAs containing the MLTU tripartite leader (43, 137). The L4-33K is a nuclear phosphoprotein, which localizes to the periphery of viral replication centers (101). It functions as a virus-encoded alternative RNA splicing enhancer protein that activates the production of the full set of MLTU mRNAs (28, 86, 125). Additionally, L4-33K has been reported to stimulate MLP transcription by binding to the DE element (3).

The L4-33K and 22K proteins share the 105 N-terminal amino acids but have distinct C-terminal domains (Fig. 6), resulting from an internal splicing event within the L4-33K coding region (86, 125). It has been suggested that both L4-33K and L4-22K are expressed from a novel L4 specific promoter (L4P) at intermediate times of infection (87), thereby explaining the transient activation of L4 mRNA expression immediately following viral DNA replication. Both proteins have been reported to bind to the conserved A-repeats in the viral DNA packaging domain (3, 27, 97-99), indicating that their role in viral assembly might correlate with their shared N-termini.

The L4-22K protein, together with IVa2, is also reported to bind to the DE element in the MLP (27, 97). The L4-22K protein functions as a transcription factor, activating MLP transcription (6, 86). Taken together, available experiments suggest that L4-22K is a component of the trans-acting factor complex DEF-A, alone or potentially together with the IVa2 protein. L4-22K has also been implicated as a factor stimulating L4P transcription (87) and activating L4-33K splicing (41). A viral mutant defective in L4-22K protein expression dramatically reduces expression of both structural and nonstructural late protein (e.g. penton, hexon, fiber, L1-52, 55K, L4-100K) as well as intermediate proteins IVa2 and pIX suggesting that L4-22K also regulates MLTU gene expression by post-transcriptional mechanisms (86).

Figure 6. Schematic organisation of the L4-22K and L4-33K proteins. The tripartite leader is indicated in red.
Present investigation

Paper I

**Adenovirus L4-22K stimulates major late transcription by a mechanism requiring the intragenic late-specific transcription factor-binding site.**

Since both L4-22K and L4-33K have been reported to activate MLP transcription (3, 86) and bind to the DE element we decided to investigate the effect of L4-22K and L4-33K on MLP transcription using a reductionistic approach in order to clarify whether the two proteins have unique or overlapping functions in transcriptional activation.

To identify the L4 products that stimulate transcription from the MLP, a reporter plasmid containing all the characterized upstream and downstream transcription elements of the MLP was co-transfected in 293 cells with an activator plasmid expressing both L4-22K and L4-33K or plasmid mutants with a premature stop codon in either coding sequence. The results showed that activator plasmids capable of expressing the full-length L4-22K protein activated MLP transcription.

Since the DE element is required for DEF-A dependent MLP activation (50, 74, 84), we tested whether L4-22K stimulation of transcription requires the DE element. Indeed, deletion of the DE element resulted in a reduced L4-22K activation of MLP transcription suggesting that L4-22K exerts most its effect via the DE element. However, the activation capacity was not eliminated. Thus, a low level of residual L4-22K activation of MLP transcription was observed in the transient transfection assays, indicating that additional L4-22K responsive elements might exist (see paper IV). Further, gel shift assays demonstrated that the recombinant L4-22K protein binds efficiently and sequence-specifically to the DE probe.

Previous studies have suggested the IVa2 protein is required for MLP activation at late times of infection. We therefore tested whether the L4-22K and the IVa2 proteins cooperated as activators of MLP transcription in transient transfection assays. The results showed that IVa2 alone was a weak activator of MLP transcription. Further, IVa2 in combination with L4-22K only weakly increased L4-22K activation of MLP transcription. We therefore concluded that the IVa2 protein does not contribute significantly to MLP transcription in transient transfection experiments. However, it should be noted that this does not exclude that the IVa2 protein functions as an enhancer protein of MLP transcription in the context of a virus infection.
To test whether the recombinant L4-22K protein has an impact on transcription from the MLP and other viral promoters, *in vitro* run-off transcription assays in HeLa-NE were performed. The results suggested that the effect of L4-22K was not restricted to the MLP, but the pIX promoter was also activated. Interestingly, the pIX protein itself has previously been shown to activate MLP transcription (75), suggesting a possible feed-forward mechanism in MLP transcription. In line with our observation previous studies have shown that an L4-22K-deficient viral genome expresses dramatically reduced levels of the pIX protein (86). The E1A promoter and the VA RNAI gene were unaffected whereas the E4 and E1B promoters were slightly repressed by the L4-22K protein. The absence of a stimulatory effect on the E1A promoter was slightly unexpected since the seven A-repeats, involved in viral DNA packaging, are located upstream of the E1A transcriptional start site. These A-repeats have previously been shown to bind to the L4-22K protein. Potentially efficient binding of L4-22K to the A-repeats might require the IVa2 protein, as has previously been suggested in the context of DNA packaging (27, 97). Collectively our results suggest that L4-22K might stimulate MLP transcription and regulate adenoviral late gene expression by a feed-forward mechanism.

To clarify whether L4-22K DNA binding was necessary for transcription activation an L4-22K/E2 fusion protein was constructed. This chimeric protein contains the DNA binding domain from the bovine papilloma virus type 1 E2 protein (BPV1 E2) fused at the carboxy terminus of L4-22K reading frame. The L4-22K/E2 fusion protein activated MLP transcription as efficiently as the wild type L4-22K protein suggesting that the fusion of the BPV E2 domain did not have negative effects on the activity of L4-22K as a transcription factor. This finding allowed us to construct reporter plasmids where the DE element was deleted and BPV E2 DNA binding sites were inserted at various locations in the MLP. Using a co-transfection approach we could show that the L4-22K/E2 protein activated transcription from the reporter plasmid where a BPV1 E2 binding site was in place of the DE element (MLP-E2) or E2 binding site was positioned 300 bp upstream of the MLP transcriptional initiation site. The activation of L4-22K/E2 on MLP-E2 was increased compared to that of L4-22K protein on the wild type MLP containing the DE element, indicating that the fusion protein might form more stable complexes on the E2 binding site compared to L4-22K on the DE sequence. To test whether L4-22K activation of transcription was restricted to the MLP two additional promoter constructs were tested: E2-E1B-CAT, consisting of a minimal adenoviral E1B promoter with an upstream E2 binding site and BPV1URR-Luc, which is the natural BPV1 promoter driving a luciferase gene. The results showed that the E1B promoter was activated by the L4-22K/E2 fusion protein. In contrast, the BPV1 promoter, which contains multiple E2 DNA binding sites, was non-responsive to L4-22K/E2 cotransfection. Collectively, our results suggest that tethering of the L4-22K
protein to the DNA was required for transcription stimulation. However, the results also suggest that tethering of the L4-22K protein to a promoter is not sufficient for transcriptional activation. Thus, additional transcription factors and/or promoter elements might be required to specify an L4-22K responsive promoter.

The L4-33K protein has previously been shown to function as an alternative RNA splicing factor involved in the temporal shift from early to late gene expression (28, 125). In this study we directly compared the activities of L4-22K and L4-33K as transcription and RNA splicing factors. The results clearly showed that L4-33K activated IIIa pre-mRNA splicing in a dose-dependent manner whereas L4-22K could not. Conversely, L4-22K stimulated MLP transcription in a does-dependent manner whereas L4-33K did not. Taken together the results suggest that the two proteins have distinct and non-overlapping functions in RNA synthesis and RNA processing, at least at the two reporter systems tested.

Paper II

A suppressive effect of the first leader 5′ splice site on L4-22K-mediated activation of major late transcription

The Ad MLP is active at a low level during the early phase of infection. The activity is dramatically increased after the onset of viral DNA replication. Maximal late-specific activation of MLP transcription requires the DE elements, which bind to the DEF-A and DEF-B factors (49, 68, 77). L4-22K, which appears to be the core component of DEF-A, binds to the DE sequence and activates MLP transcription (paper I, (6)). Here we identified a novel binding site for L4-22K in the major late first intron, the so-called R1 region.

DNase protection studies have previously shown that three regions, designated R1, R2 and R3, located in the major late first intron are protected in extracts from late-virus infected cells (49, 68). The R2 and R3 regions correspond to the DE1 and DE2 subdomains of the DE element whereas the R1 region is a novel site covering the major late first leader 5′ss that weakly interacts with an unknown viral late-specific protein. To begin the testing whether L4-22K was the mystery protein binding to the R1 region, 293 cells were cotransfected with the L4-22K expressing plasmid and an MLP reporter with or without the R1 region. The effect on MLP transcription was measured by the primer extension technique and the S1 nuclease protection assays. The results suggested that L4-22K-mediated activation of MLP transcription increased with the reporter lacking the R1 region. This result indicated the R1 might have a suppressive effect on MLP transcription.

To investigate whether the major late 5′ss, which is embedded in the R1 region, was responsible for the suppressive effect of R1 on L4-22K-mediated
MLP transcription, we mutated the 5’ss in the MLP reporter plasmid (four point mutations disrupting the conserved 5’ss consensus sequence). Using a cotransfection approach we studied the effect of the 5′ss mutation on L4-22K-mediated MLP transcription. Similar to the effect of the deletion of the whole R1 region, a mutation in the major late first leader 5′ss resulted in an increase in L4-22K-mediated activation of MLP transcription.

To further characterize the significance of the first leader 5′ss in L4-22K-activated MLP transcription we used a HeLa-NE in vitro run-off transcription approach. For this experiment the linearized MLP wild type or 5′ss mutant templates was incubated in HeLa-NE with the recombinant L4-22K protein. The results showed that, compared to the MLP wild type, the first leader 5′ss mutant enhanced the L4-22K-dependent MLP activation also under in vitro conditions.

To test whether L4-22K directly binds to the R1 region, the recombinant L4-22K protein was incubated with the radiolabeled R1 probe in the gel mobility shift assay. The data indicated that L4-22K directly binds to a 5’-CAAA-3′ motif located at the distal end of the R1 region. This sequence represents the reverse complement to the consensus L4-22K binding motif (5′-TTTG-3’) located in the DE element and the DNA packaging domain (96). Since L4-22K binding to the DE element activates MLP transcription (paper I, (6)), and L4-22K binding to R1 suppresses transcription the relative binding efficiency of L4-22K to DE and R1 was worth testing. Using competitive binding of radiolabeled and unlabeled probes we could show that L4-22K binding to the DE element was much stronger than binding to the R1 region. These observations suggested the model that L4-22K binds with a high affinity to the DE sequence and activates transcription. As the concentration of the protein increases it starts to bind to the R1 region resulting in a suppression of transcription. These results are consistent with the observation that protection of the R1 region was much weaker compared to the protection of the R2 and R3 sites in extracts from late-virus infected cells (68).

Since L4-22K did not bind directly to the major late first leader 5′splice site, which we showed was important for the repressive effect of L4-22K on MLP transcription, we tested whether L4-22K binding to the distal part of R1 resulted in a recruitment of cellular factors to the 5′ss. For this experiment the recombinant L4-22K protein was incubated with the radiolabeled R1 probe in HeLa-NE. The results demonstrated that L4-22K stimulated the formation of four specific complexes, whereof the two largest were formed at a region overlapping the major late 5′ss. Interestingly, using a super-shift strategy we could show that the major complex formed at the 5′ss contained the cellular Sp1 transcription factor. Collectively, the results suggested a model where L4-22K binding to the 5′-CAAA-3′ motif located at the distal part of the R1 stimulates the recruitment of Sp1 to a position in the DNA overlapping the major late 5′ss.
Sp1 can function both as an activator or repressor of transcription. To test the effect of Sp1 recruitment to the major late 5′ss we used a HeLa-NE *in vitro* transcription assay. Incubation of L4-22K and Sp1 with the wild type or 5′ss mutant template showed that Sp1 enhancement of L4-22K-mediated activation of the MLP was greater on the 5′ss mutant template DNA. Taken together these results therefore support the conclusion that Sp1 binding to the major late first leader 5′ss region has a negative effect on L4-22K-mediated activation of MLP transcription. However, it should be noted that Sp1 also functions as an enhancer protein of MLP transcription via interactions with GC-rich sequences located upstream of the transcriptional start site (103, 134).

Paper III

**The adenovirus L4-22K protein regulates major late transcription and RNA splicing through a sequence-specific binding to single-stranded RNA**

Besides functioning as a transcription factor, the L4-22K has also been implicated as a post-transcriptional regulator of the early to late switch in viral gene expression (reviewed in (8)). For example, L4-22K has been shown to control accumulation of both structural and nonstructural late proteins and further, to directly stimulate L4-33K pre-mRNA splicing (41). As shown in paper II the negative effects of L4-22K on MLP transcription require a sequence element overlapping the major late first intron 5′ss. With these findings in mind I analyzed whether L4-22K had a capacity to bind sequence-specifically to the single-stranded RNA (ssRNA) of the same polarity as the nascent MLTU transcript (sense-stranded RNA). If this was the case L4-22K could potentially regulate both splicing of the major late first intron and transcription from the MLP via its binding to the sense-stranded R1 RNA.

The L4-22K protein binds to the double-stranded DNA (dsDNA) form of the DE element and the R1 region *in vitro*. To test whether L4-22K binds to single-stranded nucleic acids the recombinant L4-22K protein was incubated with the radiolabeled R1 dsDNA or R1 dsRNA probe with or without increasing amounts of the unlabeled sense- or antisense-stranded R1 ssDNA or ssRNA as competitors. The results from the gel shift assays showed that L4-22K indeed binds to the R1 ssDNA and R1 ssRNA with a preference for the sense-strand in both cases.

L4-22K binds to the 5′-CAAA-3′ motif at the distal part of the R1 dsDNA (paper II, (64)). To study whether L4-22K binds to this consensus binding site also in the sense-stranded R1 ssRNA we used R1 ssRNA mutant probes in competition assays to investigate the specificity of L4-22K ssRNA binding. The results from these competition assays showed that a mutant containing a single point mutation in the 5′-CAAA-3′ motif failed to out-
compete L4-22K binding to the sense-stranded R1 ssRNA, suggesting that the L4-22K protein binds to the same consensus binding site in both dsDNA and R1 ssRNA.

Since L4-22K binds to the R1 ssRNA with a polarity identical to the nascent MLTU transcript we investigated whether L4-22K could stimulate recruitment of U1 snRNP to the major late first intron 5’ss. Thus, L4-22K might act as an RNA splicing factor enhancing late gene expression via a stimulation of first intron splicing. To test this possibility we used RNase H depletion of U1 snRNA from HeLa-NE to probe the effect of L4-22K binding to the R1 ssRNA and U1 snRNA recruitment to the major late first intron 5’ss. The recruitment of U1 snRNA was monitored by a gel shift assay where the radiolabeled major late leader1-R1 transcript was incubated in HeLa-NE or U1 snRNP depleted HeLa-NE together with the recombinant L4-22K protein. The results showed that addition of L4-22K resulted in a significant enhancement of U1 snRNP recruitment in the control HeLa-NE but not in the U1 snRNP depleted HeLa-NE, suggesting that L4-22K indeed recruits U1 snRNP to the major late first leader 5’ss. Since U1 snRNP recruitment to a 5’ss is a critical step in commitment of a pre-mRNA to splicing (see above) one would expect that the L4-22K-mediated increase in U1 snRNP recruitment to the major late 5’ss should enhance first intron splicing. To test this we incubated a major late first intron pre-mRNA with or without L4-22K under in vitro splicing conditions in HeLa-NE. The results showed that the L4-22K could function as an RNA splicing enhancer protein, activating MLTU first intron splicing.

The major late first leader 5’ss region functions as a repressor element of L4-22K-mediated activation of MLP transcription (paper II, (64)). To test whether the recruitment of U1 snRNP to the first leader 5’ss affected MLP transcription we used a reporter plasmid with a 5’ss mutation disrupting the functional U1 snRNA interaction. As expected from our previous result this mutant showed an increase in reporter gene expression. However, cotransfection of a plasmid expressing a modified U1 snRNA gene containing a compensatory mutation restoring the U1 snRNA/5’ss interaction resulted in a decrease in L4-22K-mediated activation of MLP transcription. To more directly demonstrate that U1 snRNP has an inhibitory effect on MLP transcription we showed that L4-22K enhancement of MLP transcription was more effective in in vitro transcription assay in HeLa-NE depleted of a functional U1 snRNP compared to wild type HeLa-NE. Taken together, the results suggest that an interaction between the U1 snRNA and the major late first leader 5’ss at the pre-mRNA level has a suppressive effect on L4-22K-mediated activation of MLP transcription.
Paper IV

Adenovirus L4-22K protein responsive elements in the adenovirus major late promoter

L4-22K stimulates MLP transcription in a DE-dependent manner (6). However, deletion of the DE element reduced but did not completely abolish L4-22K activation of the MLP. Therefore, additional L4-22K responsive enhancer elements might exist in the MLP (6). Here we analyzed the impact of upstream transcription factor binding sites on L4-22K-mediated activation of MLP transcription and further made a preliminary identification of the L4-22K transactivation domain.

To study the possible contribution of upstream transcription factor binding sites we used MLP reporter plasmids with mutations disrupting the CAAT, UPE or Inr elements. To avoid the complication of L4-22K regulating transcription via the characterized DE element and the R1 region these constructs were generated using a MLP promoter fused to MLTU tripartite leader. Using a cotransfection approach we could show that the wild type MLP construct was significantly stimulated by L4-22K. Deletion of the UPE site essentially abolished the capacity of L4-22K to stimulate transcription whereas deletions affecting the CAAT box or the Inr did not reduce the capacity of L4-22K to activate MLP transcription. These results strongly suggest that the UPE is a positive element mediating L4-22K stimulation of MLP transcription. To determine whether the UPE and DE were the sole L4-22K responsive enhancer elements in the MLP, a reporter with a double deletion was co-transfected with the L4-22K activator plasmid. The results showed that deletion of both the UPE and the DE elements essentially eliminated L4-22K-mediated activation of the MLP.

The DE element has previously been shown to bind to L4-22K and activate MLP transcription both in vivo and in vitro. Since the UPE is an L4-22K responsive element in vivo it became of interest to test whether the UPE element also activated MLP under in vitro transcription conditions. For this experiment the recombinant L4-22K protein was incubated with the wild type MLP and MLPΔUPE templates in HeLa-NE. Surprisingly, L4-22K failed to activate either the wild type MLP or the MLPΔUPE in the HeLa-NE, whereas L4-22K activated efficiently MLP transcription in a template containing the DE element. The basal level of MLP transcription was reduced with the MLPΔUPE compared to the wild type, in agreement with the conclusion that the UPE is a key regulatory element in the basal promoter. Collectively, these results indicate that UPE and DE elements might stimulate MLP transcription by different mechanisms: UPE is required for basal MLP activity and it acts as an L4-22K responsive enhancer element in vivo but not under in vitro conditions whereas the DE element functions both in vitro and in vivo. A reason for the difference in function might be that, in contrast to the DE element, the UPE does not contain the 5’-CAAA-3’ motif.
required for L4-22K DNA or RNA binding. Most likely L4-22K activates MLP transcription through a more indirect mechanism not requiring a direct association to the upstream MLP.

To make a first characterization of the L4-22K transcriptional activation domain we created amino-terminal deletion mutants of the L4-22K protein in the context of the previously described L4-22K/E2 plasmid. This plasmid expresses an L4-22K protein with a C-terminal fusion of the BPV E2 DNA binding domain. Using the reporter plasmid MLP-E2, which has the BPV DNA binding element replacing the DE sequence, makes it possible to study the location of the L4-22K transcriptional activation domain in the absence of a possible interference of the L4-22K DNA binding domain. The results from the cotransfection experiments suggested that the L4-22K transactivation domain was localized between amino acids 120 and 176 in the highly conserved carboxy-terminal domain of the L4-22K reading frame. However, further work will be needed to establish whether the natural L4-22K DNA binding domain overlaps with the transactivation domain or if they can be physically separated.
The closely related proteins L4-22K and L4-33K are required for the temporal early to late shift in viral mRNA expression. There are conflicting reports concerning the properties and functions of these two proteins, especially their roles in stimulation of MLP transcription. The aim of my study was to further characterize how the L4-22K protein regulates MLTU gene expression to better understand the significance of the selective expression of L4 proteins at the onset of viral genome replication.

Here we show that L4-22K but not L4-33K acts as a transcriptional enhancer protein stimulating transcription from the MLP. The L4-22K associated activation of the MLP transcription was independent of the adenoviral IVa2 protein both in transfection assays and in vitro. L4-22K activates MLP transcription by binding to the consensus 5′-CAAA-3′ motif within the DE element downstream of the MLP. The sequence is identical to the L4-22K binding sites in the DNA packaging domain, which is vital for the production of new viral capsids. Taken together, the results from paper I suggest that the L4-22K protein is the core factor of DEF-A.

My studies show that the DE element is the main L4-22K responsive enhancer element in the MLP with the UPE functioning as a second minor L4-22K responsive element. Based on our results, the DE and UPE function independently as L4-22K enhancer elements. Whether the DE and UPE can function redundantly, in an additive or cooperative manner in L4-22K-associated MLP transcription remains to be investigated. The UPE does not contain the consensus 5′-CAAA-3′ motif that L4-22K specifically binds to. In contrast to the DE element the UPE is not an L4-22K responsive element in HeLa-NE in vitro, indicating that L4-22K might not interact with the UPE sequence directly, and therefore might require another factor(s) as a bridging factor.

L4-22K stimulates expression of the pIX and the IVa2 proteins (86). In turn, pIX is a positive regulator of MLP transcription (75) and IVa2 has been suggested to activate MLP transcription in late infected cells (126). Our observation that L4-22K stimulates pIX transcription in vitro supports the suggestion that the pIX protein might be important for MLP transcription in late infection. Similarly, the IVa2 protein might also be required for the high activation of MLP transcription during a lytic infection although we do not find evidence that IVa2 is required for L4-22K-mediated activation of MLP transcription in transient transfection assays or in vitro transcription.
The elements required for promoter specificity are of interest, especially since we noticed that the E1A promoter was unresponsive to L4-22K transactivation. This was unexpected since the A repeats in the viral DNA packaging sequence, which is located upstream of the E1A promoter, contain multiple L4-22K DNA binding sites. The results suggest that L4-22K stimulation of transcription might not be general and require another transcription factor(s) for function. However, since L4-22K binding to the packaging domain has been suggested to require the IVa2 protein (27, 97), L4-22K-associated transcription from the E1A promoter needs to be further investigated to determine whether it requires IVa2 or some other DNA element(s) or transcription factor(s) to function. In addition, it might be worthwhile to test if L4-22K activates transcription from the IVa2 promoter, which is located close to the MLP driving transcription in the opposite direction. Thus, the DE element that activates MLP transcription might also function as an enhancer element stimulating IVa2 transcription.

The results from this thesis suggest that the L4-22K transactivation domain is localized at the highly conserved carboxy-terminal end of the protein. More amino-terminal and carboxy-terminal deletions in the L4-22K coding sequence are needed to further define the L4-22K transactivation domain. This is of importance since the DNA binding domain most likely is also located within this highly conserved region. Whether the transactivation domain and DNA binding domain overlaps awaits experimental verification.

The results suggest a novel model for a self-limiting regulation of MLP transcription. L4-22K binds to the R1 region, which spans the major late first leader 5´ss. L4-22K binds specifically to the 5´-CAAA-3´ motif located to the distal part of the R1. Deletion of the R1 region or mutating the first leader 5´ss leads to a higher L4-22K-mediated activation of MLP transcription, suggesting that R1 or the 5´ss specifically functions as a suppressor element of MLP transcription. Future studies will aim at determining whether the R1 region or the first leader 5´ss shows the same negative effect in the context of a virus-infected cell. The inhibitory effect of the first leader 5´ss might result from an L4-22K enhanced recruitment of cellular factors to the region overlapping the major late first leader 5´ss. We demonstrated that one such factor was the cellular transcription factor Sp1. Sp1 binding to the first leader 5´ss region results in a suppression of L4-22K-mediated activation of MLP transcription. It might be worthwhile to use chromatin immunoprecipitation assays to identify additional cellular factors that interact with the R1 region and the MLP. The significance of such factors in controlling MLP transcription needs to be unraveled.

The affinity of L4-22K binding to the DE element is greater than to the R1 region. This observation is consistent with previous results showing that DNase I footprint protection of the R1 region is much weaker compared to the DE element (68). The data shows that at low concentrations L4-22K binds preferentially to the DE element, which leads to a stimulation of MLP
transcription. This increase in transcription leads to a higher production of the L4-22K protein. At high concentrations L4-22K starts to bind to its low affinity site in the R1 region, resulting in a suppression of MLP transcription. This self-limiting mechanism might be important for the regulated production of late viral proteins.

The data from previous studies suggest that L4-22K might be involved in pre-mRNA splicing (41, 86). Here we show that L4-22K binds preferentially to the sense strand ssDNA and ssRNA of the R1 region. L4-22K binds specifically to the 5′-CAAA-3′ motif in the sense-strand R1ssRNA. This binding site is identical to the L4-22K binding site in dsDNA. L4-22K binding recruits U1 snRNA to the major late first leader 5′ss, resulting in a suppression of the L4-22K-mediated activation of MLP transcription. These results suggest that the negative effects of L4-22K on MLP transcription might be coupled to a malfunction of U1 snRNP as an enhancer of transcription initiation.

We show that the L4-22K induced recruitment of U1 snRNA to the first leader 5′ss stimulates first intron splicing. Other studies have shown that L4-22K activates L4-33K pre-mRNA splicing (41, 86) but not the L1-IIIa transcript (6). Thus, whether L4-22K acts as a sequence specific splicing enhancer protein remains to be studied. For example, it might require the consensus 5′-CAAA-3′ L4-22K binding site in close proximity to a 5′ss in order to activate splicing.

Collectively, the results suggest a model where the L4-22K protein regulates MLTU expression by a feedback regulatory mechanism (Figure 7). At low concentrations L4-22K binds to the DE element, stimulating MLP transcription. As the concentration increases, probably due to enhanced transcription of the MLP that leads to higher expression of MLTU genes, L4-22K starts to bind to the R1 region, both at the DNA and ssRNA levels. The binding to the R1 DNA recruits Sp1 to the major late first leader 5′ss region, resulting in a suppression of MLP transcription. Simultaneously, L4-22K binding to the nascent MLTU transcript recruits U1 snRNP to the first leader 5′ss, resulting in a further suppression of the L4-22K associated MLP transcription. However, the increase in U1 snRNP recruitment to the RNA also leads to an enhancement of first intron splicing. An interaction between L4-22K and U1 snRNP most likely alters the interaction of U1 snRNP with the basal transcriptional machinery thereby reducing MLP transcription. This model suggests a possible scenario of how the L4-22K protein might regulate MLTU gene expression at both the level of transcription and RNA splicing. In the future, work needs to be done to demonstrate how L4-22K interacts with U1 snRNP, Sp1 and other yet to be discovered cellular factors that participate in the regulation of MLP transcription.
During my thesis work I also observed that high amounts of L4-22K enhanced the accumulation of prematurely terminated MLP transcripts around 40 nucleotides long, suggesting a regulation at the level of transcription elongation (data not shown). This indicates that L4-22K can effectively regulate MLP transcription as well as the temporal expression from the MLTU by an additional self-limiting mechanism. However, more experiments will be required to characterize this mechanism.

Figure 7. Hypothetical model summarizing how the L4-22K protein regulates MLTU gene expression at the level of transcription and RNA splicing. Kindly provided by Göran Akusjärvi.
Acknowledgments

All the work was performed in Department of Medical Biochemistry and Microbiology, Faculty of Medicine, Uppsala University. The Swedish Cancer Society, EURASNET and the Swedish Research Council through a grant to the Uppsala RNA Research Centre supported the work. Here I would like to express my gratitude to my supervisor(s) and co-authors involved in the work presented in this thesis, and everyone who has supported and helped me during my PhD studies.

Göran Akusjärvi- It is my great honour to be your 24th PhD student in your academic career. I chose the group due to my interests in RNA splicing, the field in which you are one of the most talented scientists. Even though my projects are more related to transcription than splicing and they seem to be complicated, I was still obsessed with studying the mechanisms. I had difficulties of my projects, but luckily I managed to finish and decide to continue being a scientist after graduation. Thank you for your support, supervision and questioning, which were so helpful to lay a good foundation for my knowledge in molecular biology. Several important events happened in my life during this stage, and therefore thank you a million for all your support.

Catharina Svensson- for interesting discussions during our Friday meetings. Probably, we will have collaborations for international education in the future ☺; Göran Magnusson for nice discussion; my examiner Lena Kjellén for your kindness and encouragement; my co-author Tanel Punga for helpful discussions (especially Sp1 experiments), support and encouragement; my previous secondary supervisor Stefan Schwartz for nice discussion and suggestions; Staffan Johansson for your kindness, advice and encouragement; Daniel Öberg for your funny jokes and nice arrangement of the Virology course; my master thesis supervisor Francoise Raffalli-Mathieu for kind recommendation and teaching me molecular biology technologies; Karin Carlsson for your recommendation and warmly support; Hongxing Zhao for joyful and encouraging communication☺; Anders Bergqvist for happy time in Friday meeting and Xmas dinner; my secondary supervisor Pernilla Bjerling for fruitful conversations; Jin-ping Li for advices on preparing the thesis.

Many thanks to present and past members of the virology groups and Friday meetings! Sara- my co-author and good friend in the lab. Thanks for your smile, friendship and support. Ravi- you are always helpful. I am im-
pressed by your positive energy. Wael- I agree with Sara: you are so smart and talented. Thanks for your help and nice discussions between us. Anette- for your help and support. Troy (Chengjun)- I cherish all the happy times together. Gunnar- for your support, teaching me ‘gel shift assays’ and happy talking. Heidi- for teaching me how to take care of kids, friendship and advice. Ellenor- for being my co-author and your kindness. Alexis- helping me to prepare the defense. Roberta, Sibel, Helen and Priya- nice talking and good luck in the future! Monika- you are so energetic. Anna- supervising me how to work in the isotope lab. Susanne- your smile and advice. Lufeng, Mun, Songil, Maria, Farzaneh, Cecilia, Songil, Xiaoze, Edyta and many more.

The people at IMBIM for adding nice memory to my time! Rehne, Veronica and other administrative staff for your help, Olav for fixing things that need to be fixed, my master program classmate Ramesh for your smiling and jokes, all colleagues in the B9:3 corridor, Xiao-qun, Yani, Dirui, Benjamin, Jessika, Lisa, Alejandro and Barbro for helping me to be registered as a PhD student and many more.

张迎梅教授 (Professor Yingmei Zhang), 感谢知遇之恩 (You raise me up)! Very grateful for all your support during my late stage of PhD studies and the happy reunite in UK and Sweden. 感谢您及李叔叔对 Helen 和我们全家的爱。

竺一莘 (干妈) 和曹德骏 (干爹), 感谢多年的悉心陪伴和支持！竺妈妈的慈爱，鼓励，平静而宽阔的内心以及曹叔叔的榜样力量给了我勇敢前进的动力。

张晓华老师 (师父), 感谢将我领入奇妙的生物学世界和锻炼我的处事能力。

梅官霞老师, 感谢多年如姐姐般的陪伴和支持，我爱你！感谢干老师 的呵护，以及你们对我父母的关心和照顾。

赵尔宓爷爷 (Professor Ermi Zhao), 真正忘年交喔，感谢带我野外考察，支持我学生物和出国以及分享人生感悟！

All wonderful friends who have supported me and enriched my life during my PhD studies!

In Sweden: Anqi Xiong (妹)- for friendship and all warmly support; Ying Huang (姐)- for all your help, kindness and support for me and Helen; our lunch group Yanyu, Sha, Yanhong, Xiaofang, Ganlin, Xueliang, Gang, Chungang and Tao for sharing happiness, friendship and support; my only undergraduate classmate in Sweden Yuan for friendship; Ruixiang & Si for
friendship and all happy times together; **Haisha & Bo** and **Baofeng** for your help and friendship; **Shuainan & Peng, Lingguang & Alexis, Hua & Lei, Yanling & Di, Han, Chuan & Di, Suzhen & Shangfeng, Jun & Hua, Chaoying & Dirui, Tianyi, Shiyong, Zhirong, Lu, Juan, Minpeng, Wei, Yani, Yang & Hanna** and so on.

**In USA:** **Ailing Li** for friendship and encouragement. Let’s committed to YES®! **Zhuihui** for friendship. **In UK:** **Zhengming** and **Leifu** for happy time in England.

**In Canada:** **Weikai Qi,** 感谢支持，感谢曾经的所有。

**In China:** 张翼飞，你的建议和推荐让我选择了瑞典，感谢两次来瑞看望，在 WWF 期间的支持以及多年的友谊；**锦**, 本科最爱好的朋友，感谢所有的关爱；**才女颖昕**, 感谢一路陪伴，愿你健康 ☞, 璐峰, can I call you my student? ☞嘉, 学术路上勇敢向前吧; **如伟**, 感谢支持和友谊; **欣轩**, 我们都看好你喔! 佳, 为梦想努力吧! **谭英霞** 和 **张淑萍**, 感谢在瑞典的时光, 虽然你们贵为 PI; **建平, 万童**, 感谢鼓励和建议; **琳**, 感谢建议; **瑞, 冼** 等等。

感谢: 爸爸妈妈的养育之恩和对 Helen 的照顾，深沉的父爱；外公的深情厚爱；女儿 Helen 的爱和依赖；**Richard** 的陪伴；三姑于我选择生物学的大力支持；张叔叔和姨妈来瑞典给予的帮助；其他家人及至亲对我出国以及读博的支持和关爱。
References


100. Östberg, S. 2014. Functional Characterization of the Evolutionarily Conserved Adenoviral Proteins L4-22K and L4-33K.


134. 


135. 


136. 


137. 


138. 


139. 


140. 


141. 


142. 


143. 


144. 

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1185

Editor: The Dean of the Faculty of Medicine

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