Adenovirus Vector Systems Permitting Regulated Protein Expression and Their Use for \textit{in vivo} Splicing Studies

BY

MAGNUS MOLIN
We have constructed two adenovirus-based gene expression vector systems permitting regulated protein expression. They are based on the tetracycline-regulated Tet-ON- and the progesterone antagonist RU 486-regulated gene expression systems, which were rescued into E1-deficient adenovirus vectors. The vectors function in a number of cell types representing a broad species-variety and the regulation of protein expression was shown to be tightly controlled in cells not permissive for virus replication. Furthermore, the adenovirus-Tet-ON system was shown to perform in mice after intramuscular administration.

The novel adenovirus-vector systems were then used to study the effects of overexpression of selected proteins on adenovirus replication during a lytic infection, with focus on regulation of adenovirus alternative splicing. Expression of adenovirus transcription units is to a large extent temporally regulated at the level of alternative pre-mRNA splicing, where viral splice site usage shifts from proximal to distal splice site selection as infection proceeds. This makes adenovirus an appropriate model for mechanistic studies of regulated splicing. We show that overexpression of the essential host cell splicing factor ASF/SF2 inhibits this shift by promoting usage of proximal splice sites. As a consequence, the virus displayed a markedly inhibited growth. Interestingly, mRNA expression from the adenovirus major late promoter was almost completely lost as a consequence of ASF/SF2 overexpression. Collectively, the cellular splicing factor ASF/SF2 prevents adenovirus from entering the late phase of infection. This strongly argues for a need for the virus to block the splicing enhancer activity of ASF/SF2 for establishment of a lytic infection. Further, from analysis of the strict inhibition of late region 1 late pre-mRNA splicing we propose that the temporal regulation of alternative splicing is merely a consequence of fitness rather than profoundly deleterious effects of an unregulated expression. During our studies we noted that in 293 cells, which are used for growth of E1-deficient Ad vectors, an unwanted background reporter gene expression was evident in our vector systems. We therefore introduced an additional regulatory element, functioning as a transcriptional road-block, and showed that this methodological innovation represents a way to overcome the potentially deleterious effects of background reporter gene expression. This modified viral vector system should make it possible to reconstruct recombinant viruses expressing highly toxic proteins.

In conclusion, this work presents a new in vivo model system to study proteins involved in RNA splicing and other gene regulatory mechanisms.

Key words: Adenovirus, vector, inducible, gene expression, ASF/SF2, splicing.

Magnus Molin, Department of Medical Biochemistry and Microbiology, Biomedical Center, Box 582, Uppsala University, SE-751 23 Uppsala, Sweden

© Magnus Molin 2001

ISSN 0282-7476
ISBN 91-554-4926-3

Printed in Sweden by Fyris-Tryck, Uppsala 2001
To my family
A child of five would understand this. Send someone to fetch a child of five.
Groucho Marx (1895-1977)

The happiness of the bee and the dolphin is to exist.
For man it is to know that and wonder at it.
Jacques-Yves Cousteau (1910-1997)
This thesis is based on the following articles, which in the text are referred to by their roman numerals:


Reprints were made with the permission from the publishers; the American Society for Microbiology.
TABLE OF CONTENTS

ABBREVIATIONS ......................................................................................................................... 7

INTRODUCTION ............................................................................................................................. 8

INTRODUCTION TO ADENOVIRUSES ......................................................................................... 8
Taxonomy ........................................................................................................................................ 8
Pathogenesis ................................................................................................................................ 9
The structure of the virus particle ................................................................................................. 9
Attachment and entry .................................................................................................................. 10
Organization of the genome .......................................................................................................... 11
The goals of a virus ....................................................................................................................... 12
Cell cycle control .......................................................................................................................... 12
Transcription ................................................................................................................................ 13
Evasion of the host immune response .......................................................................................... 14
Alternative splicing and polyadenylation ..................................................................................... 14
mRNA transport ........................................................................................................................... 15
Translation and shut-off of host protein synthesis ......................................................................... 16
Replication .................................................................................................................................... 16
ADENOVIRUS VECTORS ................................................................................................................ 19
Different generations of Ad vectors ............................................................................................. 20
Inducible promoters ...................................................................................................................... 22
SPLICING ....................................................................................................................................... 24
The splicing machinery ................................................................................................................ 24
SR-proteins ................................................................................................................................... 26
Sequence elements required for splicing ....................................................................................... 30
Splicing enhancer/silencer sequences .......................................................................................... 30
Adenovirus as a model system ...................................................................................................... 31
Regulation of RNA splicing in other viral systems ....................................................................... 37

PRESENT INVESTIGATION AND DISCUSSION ............................................................................. 38

TWO NOVEL ADENOVIRUS VECTOR SYSTEMS PERMITTING REGULATED PROTEIN EXPRESSION IN GENE TRANSFER EXPERIMENTS (PAPER I) ........................................................................... 39
OVEREXPRESSION OF THE ESSENTIAL SPlicing FACTOR ASF/SF2 BLOCKS THE TEMPORAL SHIFT IN ADENOVIRUS pre-mRNA SPlicing AND REDUCES VIRUS PROGENY FORMATION (PAPER II) .............. 41
EXPRESSION OF THE STRUCTURAL IIIA PROTEIN DURING THE EARLY PHASE OF AN ADENOVIRUS INFECTION RESULTS IN DEFECTS OF MAJOR LATE mRNA ACCUMULATION (PAPER III) ...................... 44
AN ADENOVIRUS VECTOR DESIGNED FOR EXPRESSION OF TOXIC PROTEINS (PAPER IV) ........................................................................................................................................ 45

CONCLUSIONS ............................................................................................................................. 48

FUTURE PERSPECTIVES ................................................................................................................. 49

ACKNOWLEDGEMENTS ................................................................................................................ 50

REFERENCES ................................................................................................................................. 51
ABBREVIATIONS

Ad: adenovirus.
ASF/SF2: alternative splicing factor/splicing factor 2.
CAR: coxsackie-adenovirus receptor.
Δ: deletion mutant
DBP: adenovirus DNA binding protein.
DNA: deoxyribonucleic acid.
E. coli: Escherichia coli, a bacteria commonly used for plasmid cloning and recombinant protein expression.
eIF: eukaryotic (translational) initiation factor
hpi: hours post infection
kD: kilo Dalton, a molecular weight unit used for proteins.
MLTU: (adenovirus) major late transcription unit.
MOI: multiplicity of infection; number of infectious virus particles per cell used for infection.
MHC: major histocompatibility complex
mRNA: messenger ribonucleic acid.
ORF: open reading frame; potentially a coding part of a DNA molecule; begins with a start codon and ends with a stop codon.
p300: a 300 kD protein that interacts with the E1A-289R protein. Involved in cell differentiation
poly(A): A stretch of adenosine residues that is found attached to the 3’ end of most cellular and viral mRNAs.
pRb: retinoblastoma susceptibility protein; a cellular tumor suppressor protein
snRNP: small nuclear ribonucleoprotein particle.
3RE: IIIa repressor element. An intronic RNA sequence to which phosphorylated SR-proteins can bind and thereby prevent IIIa mRNA splicing.
3VDE: IIIa virus infection-dependent splicing enhancer.
TNF: tumor necrosis factor
Transgene: gene of interest inserted into a foreign context, such as a virus vector.
tRNA: transfer RNA
INTRODUCTION

The widespread expression of the receptors for adenovirus enables efficient gene transfer of a variety of cells using replication-defective adenovirus vectors. In contrast to most retroviruses, adenovirus infects cells independent of their mitotic status. The broad species- and tissue-specificity of adenovirus allows for quick scanning of several cell lines using the same vector construct. Consequently, replication-defective adenovirus vectors have become popular tools in gene transfer experiments.

Regardless of what construct (vector) the gene of interest (transgene) is inserted into, one has to carefully choose a promoter sequence that allows for a physiologically relevant production of the protein to be studied. Recent clinical gene therapy studies have confirmed this by showing that the therapeutic window for treatment in different disorders is narrow (reviewed in (29)). A constitutively active promoter rarely allows for physiologically relevant protein concentrations. In addition, expression of a potentially toxic gene is preferably kept silent during the vector cloning and amplification stage. In other words, the issue of transgene expression level is important. Thus, that was addressed in paper I and IV.

Moreover, adenovirus is interesting per se since studies on adenovirus replication and gene expression have yielded extensive knowledge about the general mechanisms regulating these events. Alternative splicing is one of the major mechanisms controlling gene expression in eukaryotic cells. Thus, 22% of all human genes have been estimated to undergo regulated alternative splicing (49). The mechanistic details of how alternative splice site (ss) choice is regulated are, however, to a large extent unknown. Expression from most adenovirus transcription units is subjected to a regulation at the level of alternative pre-mRNA splicing (reviewed in (3)). Hence, adenovirus represents a good research tool for studies of regulatory mechanisms controlling alternative splicing. This issue was explored in paper II and III.

Introduction to adenoviruses

Human adenovirus was first cultured and reported as a unique viral agent in 1953 (182) while researchers were trying to establish cell cultures from tonsils and adenoidal tissue, hence the name adenovirus.

Taxonomy

The adenovirus family, *adenoviridae*, is divided into two groups, aviadenovirus and mastadenovirus. Aviadenovirus are viruses with tropism exclusively for birds while mastadenovirus comprises viruses with human; simian; bovine; equine; porcine; ovine; canine; and opossum tropism. Human adenoviruses are divided into six subgroups (A-F) based on their ability to agglutinate red blood cells (98) and their oncogenic potential in rodent cells (reviewed in (188)). Briefly, subgroup A are highly oncogenic while subgroup B adenoviruses are weakly oncogenic and the non-oncogenic viruses were originally classified as subgroup C. By additional criteria (e.g. SDS-PAGE analysis of structural proteins), the non-oncogenic adenoviruses were subdivided into groups C, D, E, and F (205, 209). However, the evolutionary relationship between the various serotypes is probably best seen by DNA (or amino acid) sequence comparisons (10).
The degree of DNA homology between members of the same subgroup has been shown to be above 50%, whereas the homology between members of different subgroups is below 20%. As sequencing of the genome of human adenovirus serotype 5 (Ad5) was completed in 1991 a direct sequence comparison with the already known sequence of its closest relative within subgroup C, adenovirus serotype 2 (completed in 1984 (4, 5, 176)), was possible. DNA sequence comparisons show 94.7% homology between the two genomes (45). The percentage homology of amino acids (aa) differs for individual proteins, being in the range of 69 to 100%. While the fiber protein displays the largest aa difference (69 % homology or, if one prefers, 31 % difference), proteins like 52,55K and IIIa exhibit aa homologies of 99.3 and 99.8%, respectively. The high sequence homologies of e.g. 52,55K and IIIa most probably reflect highly optimized protein structures. The high variability of the fiber sequence has been suggested to result from recombination between viruses of different serotypes (45). Certain properties of the fiber structure, i.e. its ability to agglutinate red blood cells, constitutes the basis for classifying adenovirus and therefore it is not surprising to find differences between subgroups, within this protein sequence. In addition, the fibers and the hexon capsomeres of the virion (see below) are exposed to immune reactivity within the infected host, seen as host production of fiber and hexon specific antibodies, which can explain their high mutability.

Pathogenesis

The diagnosis of a human adenovirus infection is complicated by the existence of close to 50 viral serotypes (99), although certain subgroups are more likely to be involved in certain patient populations. In short, the human adenoviruses of subgroup A and F are all associated with infections of the gastrointestinal tract while the other subgroups largely infect the airways. Possible symptoms of adenovirus infection are fever with a variety of respiratory illnesses including pneumonia (subgroup C; serotypes 1, 2 and 5), gastroenteritis (subgroup F; serotypes 40 and 41), conjunctivitis, hepatitis, encephalitis and hemorrhagic cystitis. The clinical and histopathologic features of an adenovirus infection may resemble those of a cytomegalovirus infection, potentially complicating the diagnosis. Despite the oncogenic potential of some of the viral proteins, adenoviruses are not generally considered a causative agent of human cancer.

The structure of the virus particle

Adenoviruses (for a review see (188)) are non-enveloped viruses with an icosahedral capsid, characterized by 20 triangular facets and 12 vertices. The capsid is built up by 252 capsomeres and encapsulates one copy of a linear double stranded DNA genome, which varies in size between 30-36 kbp for mastadenoviruses.

The virus particle consists of a core particle surrounded by a protein shell, the capsid. The core particle comprises DNA bound to proteins V, VII, X, and the terminal protein (TP). Briefly, V interacts with VI in bridging the genome and the capsid. VII is a histone-like protein around which the viral DNA is wrapped (40). The terminal protein (TP) is covalently bound to the ends of the viral DNA where it serves as a primer for initiation of the first round of DNA synthesis following infection. The major component of the capsid, the hexon capsomer, is a trimer of polypeptide II and builds up the triangular facets where it is held together by the capsid-stabilizing (glue) proteins IX and IIIa. Pentamers of polypeptide III form the penton base, which construct the
vertices from where the fibers protrude. Fibers exist as trimers of polypeptide IV on the virion and are required for the initial attachment to the cell surface.

Figure 1. The adenovirus virion (taken from reference (188)). Structural proteins are by tradition denoted with roman numerals.

Maturation of the adenovirus virion includes trimming of five polypeptides that are made as precursors (pIIa, pV, pVI, pVII, and pTP) by the virus encoded L3-23K protease. IIIa for example, which is represented by 74 copies in the virus particle, exists in premature virions as a 67 kD monomer precursor that is later cleaved to 65.4 kD. Considering the role of phosphate groups in regulation of various biological processes, it interesting to note that IIIa has been demonstrated to be the only structural protein that is phosphorylated (184).

Generally, protein structures are more closely conserved than aa sequences. Despite a low DNA sequence homology, bacteriophage PRD1 and adenovirus share a similar capsid structure (13). Both viruses have similar capsid hexamers and the vertices are penton bases to which spikes (fibers) are attached. Moreover, both viruses have linear double-stranded DNA genomes with terminal repeats capped by terminal proteins that are required for DNA replication. This presumably represents an architecture that was perfected early in evolution. Accordingly, they have retained the same structure despite a different evolutionary origin; Gram-negative bacteria and cells of the human respiratory system, respectively.

**Attachment and entry**

Adenovirus was one of the first viruses who's entry was shown depend on two separate cellular receptors (reviewed in (158)). The sequential uptake process relies on an initial contact with a primary cellular receptor responsible for attachment (14, 203) followed by contact with a secondary cellular receptor responsible for internalization of the virus particle (219). Two primary receptors have been identified. One is shared between coxsackievirus B3 (a positive-stranded RNA virus) and adenovirus and was hence named coxsackie-adenovirus receptor (CAR) (14). The other receptor corresponds to a conserved region of the heavy chain of the major histocompatibility complex class I
molecules (MHC-I-α 2) (103). The initial contact between the virus particle and the cell surface is a result of an interaction between the fiber knob and either of CAR or MHC-I-α 2. The subsequent interaction between the Arg-Gly-Asp (RGD) motif (found in serotypes 2, 3, 5, and 12) of the penton base and a type of heterodimeric cell surface molecules called integrins (specifically integrin αvβ3 and αvβ5) (219) mediates virus internalization.

Attachment and internalization represent separate events since adenovirus can bind to cells lacking so called alpha-v-integrins but fail to become internalized. Internalization occurs by receptor-mediated endocytosis, where virus particles end up in coated pits (for a review see (159)). The endocytosed pits then fuse with endosomes within 10 minutes. Once in the cytosol, the virion is transported via microtubuli towards the nucleus. Meanwhile, the particle is dismantled by an ordered elimination of structural proteins so that when it reaches the nuclear membrane, only the core particle is left. Adenovirus uncoating culminates with the release of the viral DNA and DNA-associated protein VII into the nucleus via nuclear pore complexes (83).

The capsid contains a virus-encoded cysteine protease, L3/23K, (dormant in the extracellular virus), which is activated upon host cell entry. The L3/23K protease is activated by two separate signals, both which are required: interaction with the integrin receptor (inhibited with RGD peptides) and entry of the virus particle into the reducing environment in the endosome. Greber et al. used a temperature-sensitive mutant to grow viruses that lack a functional protease. Upon infection, this mutant failed to release the fibers and was thus not able to escape from the endosome into the cytosol (84). The results indicate that the viral protease is needed to disassemble the incoming virus, which is in agreement with previous studies (90).

Organization of the genome

All adenoviral genomes examined to date have so called inverted terminal repeats (ITRs) located at each end of the viral genome. The ITRs contain elements to which cellular and viral proteins are recruited, e.g. the adenovirus pre-terminal protein (pTP) which is bound to the 5' ends of the genome and functions as the primer for initiation of DNA replication. The overall genetic organization is conserved between the mammalian mastadenoviruses (for a review see (193)). Like most nuclear replicating DNA viruses, adenovirus gene expression can be divided into two sets of transcription units: early genes that are expressed before the onset of viral DNA replication and late genes, which are expressed preferentially after viral DNA replication. This classification also defines the early and late phase of the infectious cycle.

The adenovirus genome encodes for nine transcription units, whereof the following six are expressed at early times of infection: E1A, E1B, E2, E3, E4, and L1 (see Figure 2). E1A is the first gene to be expressed following infection as a result of the strong enhancer sequences found between the left ITR and the E1A TATA box (reviewed in (17)). E1A proteins then activate the rest of the viral early promoters. It is not known whether cellular histones replace the core protein VII of the viral chromatin. However, VII has been found associated with viral DNA during the early phase and possibly remain so throughout the entire infection (40). Interestingly, the onset of transcription of each region follows the location in the viral genome. In Figure 2 one can see that early region 1 and 4 are the ones closest to the ends of the DNA molecule and they are also the first regions to be transcribed. Transcription from internal units have been shown to be repressed by VII polypeptides (156).
Three transcription units become activated at intermediate (pIX and IVa2) and late (major late transcription unit, MLTU) times of infection. The MLTU gives rise to a primary transcript of about 28,000 nucleotides (nts) in length, which is processed by alternative splicing and the use of different polyadenylation sites into more than 20 cytoplasmic mRNAs. The MLTU mRNAs are grouped into five families (L1-L5) where each family consists of multiple alternatively spliced species with a common poly(A)-site.

The goals of a virus

During an infection, viruses can be said to be interested in one thing primarily and that is to produce virus progeny and hence secure its spreading. For adenovirus, this is achieved via three main steps. First, adenovirus optimizes the cellular conditions to govern its multiplication by activating the cellular DNA replication machinery (see "cell cycle control" section). Second, the host immune response has to be counteracted to avoid clearance of the infection (see "Evasion of the host immune response" section). Third, the virus must replicate its genome and for that to happen the viral replication proteins from early transcription units E2A and E2B must be efficiently expressed (see "replication" section). DNA replication in turn activates the MLTU and hence the production of capsid proteins.

Cell cycle control

During a normal cell cycle, so called cycline-dependent kinases (cdks) control the activity of the cellular transcription factor E2F via phosphorylation of its associated factor, the cellular tumor suppressor protein pRb. Phosphorylation of pRb disrupts the transcriptionally inactive Rb/E2F complex thereby releasing E2F. E2F then acts as a transcriptional activator of genes necessary for DNA synthesis, which allows the cell to enter the S phase (217).

E1A and E1B encode proteins that together have the potential to immortalize cells. During a lytic infection, E1A-289R and 243R proteins (also referred to as the 13S and 12S mRNA products, respectively) target pRb to force the cells into S phase. E1A proteins bind to and displace pRb whereby E2F is activated. Thus, the adenovirus E2 promoter, which gives rise to viral DNA replication proteins (see "Replication"), is in part controlled by E2F. In contrast, some cellular promoters, especially those that are dependent on the cellular co-activator proteins p300 and CBP, are repressed by the E1A proteins. p300 and CBP activate transcription of cellular genes involved in cell differentiation. Further, p300 and CBP are cellular histone acetylases (HATs) (11) that have been shown to acetylate the "tails" of histones, thereby activating transcription by relieving the repressive effects of chromatin (138). As a result, the p300/CBP complex, which interacts with RNA polymerase II, favor cell differentiation. Cell differentiation and S-phase are polar opposites and cell differentiation is believed to counteract efficient viral DNA replication. E1A proteins have been shown to inhibit associated (P/CAF) and intrinsic HAT activity of CBP and p300 (38, 88) thereby preventing cellular differentiation. Taken together, these events optimize the conditions for viral DNA replication.
However, E1A induced S-phase entry results in activation of the p53-dependent apoptosis pathway. To counteract this effect the proteins expressed from the E1B unit inhibit the E1A induced and p53-dependent apoptosis. E1B-55K possesses a transcriptional repression domain plus the ability to bind to p53. Hence, E1B-55K can repress p53-responsive genes by binding to p53 (136, 137, 227). Moreover, E1B-55K, together with the E4-ORF6 (also known as E4-34kD) protein, re-localize p53 to the cytoplasm and accelerate p53 degradation (180). In addition, E1B-19K, a functional homologue of the cellular apoptosis inhibitor Bcl-2, blocks apoptosis downstream of p53, by binding to and inactivating cellular apoptosis inducers such as Bax and Bak but not Bad (41).

**Transcription**

The E1A promoter is unique among adenovirus promoters since it can function during the earliest phase of infection. The E1A protein products are found in the nucleus and activate expression from all adenoviral early promoters (reviewed in (17)). The E1A proteins activates transcription via the same promoter elements to which basal transcription factors bind. E1A proteins do not bind sequence specifically to DNA (reviewed in (18)), instead they interact with the basal transcription machinery through a mediator complex (27). E1A 289R and 243R proteins are translated from alternatively spliced E1A mRNAs. Thus, E1A-243R differ from E1A-289R by lacking an internal aa domain designated conserved region 3 (CR3), which constitutes the main transcriptional activation activity of E1A. Activation of the E2A, E3 and E4 promoters is CR3-
dependent. The 289R protein can also activate other viral promoters such as the immediate early CMV promoter whilst 243R can not (147). Whereas E1A-243R, in contrast to E1A-289R, is a poor activator of viral promoters they both activate some cellular, e.g. the heat shock, promoters at roughly equal levels (189).

When adenovirus enters the late phase of infection transcription from the major late promoter (MLP) increases dramatically. Although the MLP is active also during the early phase of infection, seen by the production of the L1-52.55K protein, DNA replication per se dramatically increases production of MLTU mRNAs. Leong and Berk have shown that the increase in MLTU mRNA accumulation result from an increasing number of DNA templates, resulting from viral DNA replication, rather than activation of transcription from the MLP itself (127). However, some cellular and viral proteins have been shown to activate the MLP, e.g. the viral IVa2 protein (204), DBP (39) and the cellular transcription factors Sp1 and MAZ (166).

Evasion of the host immune response

The E3 region is dispensable for virus growth in tissue culture cells studies since E3 encodes for proteins that counteract the host immune response. Many virus infections stimulate TNF-α production by activated macrophages where after TNF-α binds to the cell-surface receptor Fas, expressed on most cell types, and triggers apoptosis of the infected cell. The E3-14.7 kDa and E3-10.4K/14.5K proteins have been shown to inhibit TNF-α activated cell death (reviewed in (220)). The E3-10.4K/14.5K heterodimer protein has recently been shown to reduce cell surface expression of the Fas receptor by internalization (59, 200), thereby protecting infected cells from Fas mediated apoptosis. In addition, the E3-19K glycoprotein has been reported to downregulate cell surface expression of MHC class I, thereby rescuing infected cells from cytotoxic T-lymphocyte mediated lysis (174).

Alternative splicing and polyadenylation

Due to the limited genome size, adenovirus has evolved an intricate arrangement of sites for transcriptional initiation, polyadenylation and splicing, which allow the virus to use its genetic information efficiently (reviewed in (107)). The expression of most adenovirus transcription units, early as well as late, is regulated by alternative splicing. As the adenovirus infection cycle proceeds the expression pattern changes from production of regulatory proteins to production of structural proteins needed for assembly of new virus particles. Following initiation of DNA replication, an increase in transcriptional elongation of MLTU produces more of the approximately 28,000 nts long full-length pre-mRNA. This transcript is processed by alternative splicing and differential polyadenylation to generate at least 20 cytoplasmic mRNAs (reviewed in (63) and (107)). The usage of five poly(A)-sites, yields mRNA families L1-L5 (mentioned above). During the early phase, the preferential use of the L1 poly(A) site is believed to be a consequence of its close location to the MLP, which gives a competitive advantage (56, 57, 87). In addition, a cis acting element upstream of the L1 poly(A) site has been shown to prevent export of mRNAs with unprocessed L1 poly(A) sites (170). Mini-constructs containing L1 and L3, but no normal ss has been inserted in place of the E1 region of adenovirus. Normal regulation of L1 and L3 poly(A) site usage could then be reproduced in vivo using such a virus (56, 63). Interestingly, superinfection of late-infected cells with the E1-recombinant virus gave an L1-specific polyadenylation from the recombinant. In other words, a virus-induced change in
processing factors is not likely to cause the shift in polyadenylation, rather a cis acting element probably prevails (63).

The multiple mRNAs expressed from each late poly(A) family result from alternative 3' ss usage (Figure 2). In contrast to E1A pre-mRNA splicing, splice sites within the MLTU do not interrupt the protein coding sequence. Adenovirus 5' and 3' ss usage is temporally regulated during adenovirus infection with a general shift towards more distal ss usage as infection proceeds (Figure 3). Moreover, splicing couples a 5' untranslated sequence (called the tripartite leader) to all mRNAs transcribed from the MLTU. The 5' tripartite leader sequence has been shown to increase the efficiency of transcription (6), mRNA transport, and mRNA stability during adenovirus infection (152). Finally, and perhaps most important, the tripartite leader enables initiation of translation, in the absence of a functional cap-complex, during the late phase of infection (19, 228, 229) (see "Translation and shut-off of host protein synthesis"). At early times, the tripartite leader of the L1 52.55K mRNA includes a protein coding sequence called the i-leader positioned between the second and third leader sequences (Figure 2). The protein product of the i-leader has been shown to act in cis to shorten the half-life of mRNAs containing its coding sequence (191). However, the significance of i-leader inclusion during early times of infection remains unknown.

**At early times of infection**

![Alternative 5' splice site usage](image1)

![Alternative 3' splice site usage](image2)

**At late times of infection**

*Figure 3.* Splicing of most adenovirus transcription units is temporally shifted. More of the primary transcript is removed, as infection proceeds, by choosing more distant splice sites.

**mRNA transport**

The adenovirus E1B-55K and E4-ORF6 proteins form a complex, which has been suggested to favor the export of viral mRNAs late after infection. E1B-55K localizes mainly in the cytoplasm when expressed alone, whereas E4-ORF6 is found in the nucleus. However, when co-expressed, both proteins exist predominantly in the nucleus, indicating that E4-ORF6 redirects E1B-55K to the nucleus, presumably by complex formation. Within the infected cell nucleus, the E4-ORF6 protein directs the E1B-55K protein to the periphery of the viral replication/transcription centers during the late phase of infection (164). It is not known how these proteins favor export of viral mRNAs. However, it has been suggested that the complex of E1B-55K and E4-ORF6 binds a cellular factor required for mRNA export and relocates it to the vicinity of viral replication/transcription centers, limiting the availability of this factor largely to viral
mRNA (164). In addition, the E1B-55K/E4-ORF6 protein complex has been shown to shuttle between the nucleus and cytoplasm, where E4-ORF6 was demonstrated to mediate both import and export (58).

Translation and shut-off of host protein synthesis

One main cellular defense mechanism against viral infections effectively shuts down translation at the level of initiation, by phosphorylation of the initiation factor eIF2α, whereby viral protein synthesis and hence multiplication and spread is prevented (114). This occurs when the cellular RNA-regulated protein kinase, PKR, is activated by double stranded RNA structures formed during a virus infection. eIF2α normally delivers GTP loaded initiator (Met)-tRNA to the 40S ribosomal subunit. However, as a consequence of PKR phosphorylation of eIF2α, eIF2α will remain bound in a non-functional complex with GDP after delivery of the Met-tRNA.

Some viruses, like adenovirus and vaccinia virus, have evolved strategies to prevent PKR activation (108, 162, 163). Efficient translation of adenovirus mRNA is secured by a small viral RNA, VA RNA I, to which PKR binds and is kept inactive (reviewed in (139)). The reason for why VA RNA bind but do not activate PKR has been attributed its relative smallness. That is, the stem loop structure that VA RNA forms is not thought to allow for dimer formation of PKR, which is a prerequisite for its activation (reviewed in (188)).

When adenovirus enters the late phase of infection, production of proteins from the MLTU increases. At the same time host protein synthesis is inhibited (Figure 4A) and by 36 hours post infection (hpi) cellular protein synthesis is completely abolished (51). Infection with many DNA and RNA viruses results in the selective shutdown of host cell protein synthesis (reviewed in (71)) and a common target for virus interference is the cellular cap-binding protein complex, eIF4F. eIF4F comprises eIF4E, eIF4E kinase Mnk1, eIF4A, and eIF4G. Infection with adenovirus and influenza virus causes dephosphorylation of the cap-binding subunit, eIF4E. The adenovirus L4 100K protein was recently shown to remove the Mnk1 kinase from the cap-binding complex (51), thereby blocking phosphorylation of eIF4E. eIF4E phosphorylation has been suggested to positively correlate with its ability to bind the cap-structure (reviewed in (173)). Initiation of translation of most cellular mRNAs is blocked during the late phase of infection because of their dependence on a stable cap-binding complex. The tripartite leader attached to all MLTU transcripts confers an ability to translate in the absence of a functional cap-binding complex. The tripartite leader as well as the 5' untranslated region of the late adenovirus IVa2 transcript contain conserved 18S rRNA complementary elements that enable eIF4E-phosphorylation independent translation (51). This occurs by a mechanism called "ribosome shunting" and possibly involves base-pairing between the tripartite leader and the 18S rRNA, reminiscent of prokaryotic translation initiation.

Replication

Viral DNA replication begins at about eight hpi and defines the beginning of the late phase of infection. Adenovirus DNA replication relies mainly on three viral proteins: the adenovirus encoded DNA polymerase (AdPol), the single stranded DNA binding protein (DBP), and the pre-terminal protein (pTP), which functions as primer for initiation of viral DNA synthesis (reviewed in (188)). During infection, adenovirus forms replication centers on the nuclear matrix. TP, which is bound to the incoming
viral genome, as well as pTP, is believed to anchor the viral DNA to these cellular structures (9, 68) via interaction with a trifunctional pyrimidine biosynthesis enzyme (CAD, carbamyl phosphate synthetase, aspartate transcarbamylase, and dihydroorotase) (8). TP is believed to prime the first round of DNA replication. Initiation of subsequent DNA synthesis requires de novo synthesized pTP and AdPol that form a tight heterodimer, which is recruited to the origin of replication. DNA synthesis initiates with the covalent attachment of a CMP residue to pTP, which provide AdPol with the necessary free 3' hydroxyl group. Initiation of replication is believed to proceed by a "jumping back" mechanism where pTP in a complex with a trinucleotide (CAT) sequence (118), complementary to the 3' GTAGTA sequence of the (Ad5) left end, base-pairs with the second GTA of the template strand. The pTP-CAT complex then jumps back three bases to permit the start of chain elongation. Notably, this mechanism allows for restoration of shortened genome ends. The polymerase leaves pTP as elongation starts (117). DBP is not necessary for the initiation of Ad DNA synthesis but instead for chain elongation where it binds to and protects newly synthesized single stranded DNA.

In addition, efficient viral DNA replication require three cellular proteins: nuclear factor I (NF I) and III (NF III), and Oct-1 (reviewed in (53)). Initiation is greatly enhanced by NF1 and Oct-1, which have been reported to stabilize the pTP-AdPol pre-initiation complex (206). Subsequently, NF III stimulates adenovirus DNA replication, by up to 50-fold, during the elongation phase (54).

Some of the E4 proteins have been shown to regulate adenovirus DNA replication (reviewed in (128)). The E4 products are not required for in vitro replication of the viral genome, rather they are believed to play an indirect role during infection. A number of E4 mutants have enabled the dissection of the roles of the different E4 products. Infections of cells with mutants lacking both ORF3 and ORF6 resulted in a dramatically reduced late protein synthesis (31). Two E4 mutants, lacking ORFs 2-6, were defective for viral late protein synthesis but showed no defect in viral DNA accumulation (31). In other words, DNA replication and late protein synthesis are not strictly correlated. A virus mutant that lacks expression of all E4 genes except E4-ORF4 is strongly attenuated in DNA replication and late viral protein synthesis. When this virus mutant was further mutated, abolishing E4-ORF4 expression, DNA replication went back to approximately normal (wild type) levels. Furthermore, the DNA replication inhibitory effect of E4-ORF4 on viral DNA replication was counteracted by expression of either of E4-ORF3 or E4-ORF6 (32). Due to the opposing effects on viral DNA replication E4-ORF3 and E4-ORF6 on one hand and ORF4 on the other, E4 proteins have been suggested to regulate the levels of DNA replication during infection. Moreover, infections with mutants lacking expression of all E4 proteins or defective in both E4-ORF3 and E4-ORF6 result in formation of concatemers of viral DNA, i.e. head-to-tail or tail-to-tail joining of multiple viral genomes (216). Possibly related to this, the protein products of E4 region has been implicated in inhibition of DNA double-strand-break repair in mammalian cells (26, 161).

Virus assembly begins with the formation of an empty capsid. The viral DNA then enters the capsid. Encapsidation of the DNA molecule requires the "packaging signal": a DNA element centered around 260 bp from the left end of the viral genome. The adenovirus IVa2 protein has previously been shown to interact with the 52,55K protein (86) and was recently shown to interact with the packaging signal (234). The 52 and 55 kD subunits of 52,55K are differentially phosphorylated forms of a single 48-kD polypeptide (93). Both phosphoforms are present within virus assembly intermediates (empty capsids contain 50 to 100 molecules and young virions 1 to 2 molecules) but not
within mature virions (93). After infection with a 52,55K deficient virus, Gustin and Imperiale observed an increase in the formation of empty capsids. This led them to conclude that the 52,55K protein plays a crucial role in the encapsidation of the viral genome (85). Taken together IVa2 and 52,55K probably bridges the viral genome and the capsid where IVa2 and 52,55K physically interacts with the DNA and the capsid, respectively.

A.

![Western blot diagram](image.png)

**Figure 4.** Shut-off of host cell protein synthesis. (A) $^{35}$S-labeled proteins from (1) MOCK infected and (2) dl309 infected 293 cells harvested at 30 hpi. (B) Transgene (e.g. IIIa) expression survives shut-off of host cell protein synthesis. $^{35}$S-labeled proteins from (1) AdTetTrip-IIIa, (2) AdTetTrip-IIIa plus doxycycline, and (3) dl309 infected 293 cells pulse-labeled for 3 hrs prior to harvest at 30 hpi.

Virus replication is completed approximately 40 hours post infection. Three viral proteins have been implicated in the release of newly synthesized virus particles, at the end of the infectious cycle. The L3-23K protease cleaves cellular cytokeratin (42, 235). In conjunction with the shut-off of host cell proteins these events disrupt the cytoskeleton and prevent renewal of filaments, respectively. Disruption of components of the cytoskeleton facilitates lysis of cells and spread of progeny virus. In addition, the E3 11.6 K protein ("Adenovirus Death Protein", ADP) is required for efficient lysis of infected cells (201, 202). Furthermore, the E4-ORF4 protein has been suggested to be
important for lysis of infected cells by activating p53 independent apoptosis late during infection (134, 135).

**Adenovirus vectors**

By virtue of the extensive knowledge of viral genome organization and control of viral gene expression, adenovirus was early on developed into a viral vector. Adenoviruses have a cloning capacity sufficient for most cDNAs of interest (see below). Furthermore, they are able to infect dividing as well as non-dividing cells; they are easy to purify to high titers; and they can transduce a broad variety of tissues both *in vitro* and *in vivo* (reviewed in (28)).

Construction of adenovirus vectors, and viral vectors in general, involves manipulation of the viral genome, whereby a foreign gene sequence replaces part of the original viral DNA sequence. This is followed by propagation of the resulting so-called virus vector (or "recombinant virus"). The E1 region was the first adenovirus locus to be replaced with foreign inserts. The rationale for replacing this region is the critical role of the E1A proteins (especially the 289R protein) for a lytic infection. Thus, by removing the E1A region the rest of the genome is silent. It has been shown for E1-deficient adenovirus vectors that the orientation of the construct replacing the E1 region determines transgene expression levels from the recombinant virus (1). A higher expression is achieved if the transgene is transcribed in a rightward direction.

The widespread use of E1-deleted adenovirus was promoted by the early existence of a complementing packaging cell line (293 cells (80)) that provide the E1 functions in *trans*. The 293 cell line was originally isolated in a screen where human embryonic kidney cells were transfected with restriction fragments of Ad5 to identify the location of the transforming genes of adenovirus. The potential to use 293 cells as complementary cells for construction and growth of E1 deficient adenovirus vectors was later recognized. Since the 293 cell line expresses both E1A and E1B proteins, the E1B region is also dispensable, which provides more space for larger inserts.

Due to the relatively large size of the adenovirus genome, there is a lack of unique restriction sites suitable for vector construction. Most adenovirus vectors are for that reason based on an Ad5 derived mutant (dl309), which has a unique *Xbal* restriction site at 3.7 m.u. that allows for easier rescuing of inserts or mutations into the E1 region. The dl309 mutant was obtained by restriction cleavage of Ad5 DNA and re-ligation followed by transfection into 293 cells (110). The resulting mutant (dl309) showed an altered *Xbal* restriction pattern having lost three out of four *Xbal* sites found in Ad5. Later, a foreign insert was found to substitute for part of the E3 region in dl309. The mutations in dl309 have been shown to affect the expression of E3 proteins (20). The E3 proteins 14.7K, 14.5K, and 10.4K were found to be absent during a dl309 infection. In addition, the expression of the 6.7K protein was strongly reduced. The remaining E3 proteins gp19K, 12.5K, and 11.6K are unaffected by the deletion/substitution event and are thus expressed at wild type levels. The foreign sequence in the E3 region of dl309 shows sequence similarity with salmon sperm DNA and has been proposed to result from unintentional recombination between Ad5 DNA and the carrier DNA used in the transfection (77). An additional (2 basepair, bp) deletion in dl309, accounting for the loss of the third *Xbal* site, locates to the transcriptional initiation site for VA RNA1 but does not affect virus growth or host cell shut off (198). All together, the mutations have reduced the genome of dl309 to 35823 bp, compared to the 35935 bp long genome of Ad5.
The unique XbaI restriction site at 3.7 m.u. in dl309 has enabled a frequently used approach, based on homologous recombination, for rescue of genes of interest in place of the E1 region. Thus, the gene of interest is cloned into a plasmid vector where it is flanked on one side by the left ITR and the packaging signal, and on the other side by an approximately 2 kb adenovirus sequence. The latter sequence allows for homologous recombination when the plasmid is transfected into 293 cells along with a replication deficient viral backbone (vector arm) (Figure 5).

Different generations of Ad vectors

Modifications of the adenovirus genome with successive deletion of more and more of the genome has lead to classifications termed first-, second- or third generation vectors (for a review see (101)). ΔE1 and ΔE1/ΔE3 constructs are usually referred to as first-generation vectors and can harbor inserts up to 4 kb (ΔE1) or 8 kb if both E1 and E3 regions are replaced. Recombinant genomes can not exceed the wild type genome size by more than 1.2 kb (i.e. 105% of the wild-type genome) without loosing genomic stability (21).

At high multiplicities of infection (MOI), first generation adenovirus vectors display a leaky viral gene expression which promotes viral DNA replication in certain cell lines and some level of toxicity in vivo (78, 129). To address this problem, E1-deleted recombinants have been further crippled by mutations in the E2A gene that encodes for DBP, which is required for viral DNA replication. In one study, the E2A gene was silenced by a temperature-sensitive mutation (60). At non-permissive temperature, this virus produced no late proteins. Furthermore, mutations of the E2B gene have given promising results (102). Double-mutant vector constructs aimed at further reducing late protein synthesis and viral DNA replication, e.g. ΔE1/ΔE2 and ΔE1/ΔE4 mutants are termed second-generation vectors. All E4 proteins, except the E4-ORF6 protein, can be removed with efficient growth still supported in 293 cells. A more efficient growth is however obtained if all E4 proteins are provided in trans in the 293 packaging cell line (213, 226). Furthermore, E4 expressing 293 cells also allows for total deletion of the E4 region. Notably, the presence of an intact E4 region in E1-deleted vectors prolongs the lives of certain tissue culture cells, by a presumed anti-apoptotic mechanism (172).

The E1 DNA sequence, that is found integrated in the chromosomes of 293 cells, overlaps with the sequences that flank the transgene of most E1-deleted vectors. As a consequence, replicating vector DNA is found to recombine with the cellular E1 sequence where wild type viruses typically appear after approximately five passages of E1-deleted vectors in 293 cells. Only one recombination event is required for this to happen since the integrated sequence is believed to go all the way to the left end of adenovirus. To prevent wild-type appearance, modified packaging cell lines have been established that contain only the coding sequences of E1A and E1B, and in which there are no overlapping sequences (64). Hence, no wild type virus can be detected using these cell lines.

The third generation comprises so-called "gut-less" vectors where most of the viral DNA sequence has been replaced by a foreign insert. These vectors allow for inserts up to 37 kb but are on the other hand dependent on a helper virus, and hence referred to as helper dependent virus vectors (HDV). The minimal size required for encapsidation has been reported to 27 kb (168) which often means that HDVs must be filled with stuffer-DNA. To date, the most elegant way of removing the helper virus is the approach described by Parks et al. where a helper virus in which the packaging signal flanked by loxP sites is used. Several passages in 293Cre cells (expressing the Cre recombinase)
yields decent amounts of HDVs while the continuously re-added helper virus genome is not packaged into new virus particles (167). Helper dependent adenovirus vectors get their capsid components from the helper virus. Thus, the cell type specificity of the HDV is decided by the serotype of the helper virus. This enables targeting of the HDV by use of different helper viruses.

Figure 5. A commonly used strategy for generation of E1-deleted adenovirus vectors. Homologous recombination between a transfer vector (plasmid) and a viral vector arm yields the recombinant virus. Adenovirus E1 proteins provided by the 293 cell line enable replication and packaging of recombinant virus particles. Virions were designed by Svend Petersen-Mahrt.
Although the use of first-generation vectors *in vivo* is associated with strong host immune response reactions, this is not necessarily a problem. In gene therapy protocols, adenovirus is probably best fitted for the treatment of cancer and local immune reactions, induced by the administered vector, can actually amplify their anti-tumor effect. Moreover, adenovirus vectors have a great potential for use as gene transfer vectors in cell culture studies.

**Inducible promoters**

Several characteristics are expected by an ideal inducibly regulated gene expression system. It should not respond to endogenous activators; exhibit a low basal (uninduced) expression combined with a high induced gene expression upon addition of inducer; and be able to produce physiologically relevant concentrations of a given protein through a wide inducer dose-response interval. Only four existing expression systems fulfill these criteria, namely the tetracycline inducible system (Tet-ON and Tet-OFF systems); the RU 486 inducible system; the FK506/rapamycin inducible system; and finally the ecdysone inducible system (for a review see (179)).

The possibility to construct inducible promoters stems from the finding that certain transcription factors can be divided into discrete units that function independently of each other. The yeast Gal4 protein could be separated into a DNA binding domain and an effector domain, necessary for activation of transcription, without loss of function. The modular structure of eukaryotic transcription factors was shown by fusing the transcriptional activation domain of Gal4 with the DNA binding domain of the *E. coli* repressor protein LexA. In yeast, the hybrid protein was found to activate transcription from genes cloned behind *lex* operator sites (30). This then led to the innovative design of what would be later referred to as the "yeast-two hybrid system" (66).

Specific regulation of inducibly regulated gene cassettes in mammalian cells is achieved by using components, i.e. operator sequences and DNA binding domains, taken from non-mammalian cells since mammalian proteins are not expected to bind such operator (i.e. DNA) sequences. Conversely, no mammalian genes are known that are activated by e.g. the yeast Gal4 protein. These insights have led to the development of a number of inducible systems.

The RU 486 (progesterone-antagonist) inducible system (214) (hereafter referred to as the "prog-system") and the tetracycline inducible "Tet-ON" system (79) (hereafter referred to as the "Tet-ON system") were both used in the work presented here and they function in a similar way (Figure 6). Both systems rely on the expression of a chimeric activator protein that binds to its cognate operator sequence positioned upstream of the gene of interest. The activator proteins respond to an inducer by binding specifically to the operator sequence from where transcription is activated through the strong transcriptional activation domain of VP16 (taken from herpes simplex virus) (reviewed in (18)).

The DNA-binding domains of the activator proteins differ. The prog system uses the DNA binding domain of the yeast Gal4 protein while the Tet-ON system relies on the reverse tetracycline-repressor (rtetR) (79), which is a mutant of the original *E. coli* tet-repressor (tetR) (tetR) and binds the tet operator in the presence of antibiotic. While VP16 plus the DNA binding and inducer responsive rtetR moiety make up the "reverse tetracycline-controlled transcriptional activator" (rtTA), the activator protein of the prog system includes a third component. Namely, the ligand-binding domain (LBD) of a 42 aa C-terminal deletion mutant (hPRB891) of the human progesterone receptor which
has been shown become activated by the progesterone-antagonist RU 486 but not by progesterone itself or other endogenous hormones (207, 214). Thus, RU 486 is believed to bind at another position on the ligand-binding domain than progesterone.

The RU 486 progesterone-antagonist induced system

The Tet-ON induced system
**Figure 6.** Schematic illustration of the two inducible adenovirus mediated systems used in this work. For details see text ("adenovirus vectors").

The progesterone receptor, which is a nuclear transcription factor, is kept in an inactive complex with heat-shock proteins hsp70 and hsp90, plus a number of co-chaperones (reviewed in (190)). Binding to its ligand, progesterone, results in a dramatic structural compaction, including dimerization, and hence activation of the receptor by dissociation of the hsps. Notably, the (C-terminal) hormone binding of the progesterone receptor, which is the domain that attracts chaperones, can be fused to heterologous proteins and still preserve its function (reviewed in (190)).

A few inducible promoters (reviewed in (46)) have been cloned into recombinant adenovirus constructs (92, 151, 157, 183), where the use of the Tet-OFF or Tet-ON system predominates.

**Splicing**

The mosaic structure of eukaryotic genes was discovered by researchers working on adenovirus transcription. By hybridization of genomic adenovirus DNA to cytoplasmic viral mRNA from adenovirus infected cells, RNA/DNA duplexes with bulges were seen in the electron microscope (15, 44). Sequences remote from each other in the adenovirus genome were found to be adjacent at mRNA level. This result was the first demonstration that eukaryotic genes are encoded on discontinuous gene segments and implied the existence of a process by which the intervening sequences was removed. This process was named splicing and the intervening sequences, removed by this process, were termed introns. The remaining sequences of the pre-mRNA, i.e. those that are also found in the cytoplasmic mRNA, were given the name exons.

Most eukaryotic genes contain introns that, to generate a functional mRNA, need to be removed. Unspliced pre-mRNAs are rarely transported from the nucleus to the cytoplasm, suggesting that splicing and mRNA transport are tightly coupled events. In fact, recent studies suggests that splicing is required for efficient export of mRNAs whereby splicing generates a specific nucleoprotein composition on the mRNA that targets it for export (131). Although rare, intronless cellular pre-mRNAs do exist. For example, the mouse histone H2a (105) and S-adenosylmethionine decarboxylase gene (169); G-protein coupled receptors (76); type I interferons (177); and a heat shock protein 70 homologue (126) are all expressed and efficiently transported to the cytoplasm without splicing. These transcripts possibly share similar cis-acting elements that bypass the requirement for splicing, which has been shown for the H2a gene (105).

**The splicing machinery**

The events leading to intron removal and subsequent joining of the flanking exon sequences are fairly well characterized. Less well known is why and how the cellular splicing machinery makes its choice when alternative splice sites compete.

Pre-mRNA splicing requires small nuclear ribonucleoprotein particles (snRNPs) and non-snRNP proteins. Five small nuclear (sn) RNAs (U1, U2, U4, U5 and U6 snRNA) all play central roles in the assembly and rearrangement of the spliceosome. Base-pairing between the snRNAs and the pre-mRNA yields part of the sequence specificity to the splicing reaction. Aside from U snRNAs, the snRNPs consist of a number of heterogeneous nuclear proteins (hnRNPs) that have their primary role in coordinating the assembly of the spliceosome. In addition, it is generally believed, but
not proven, that U snRNAs in the spliceosome carry out the catalytic process of intron removal (reviewed in (192)).

Assembly of the spliceosome appears to be a sequential event where U1 snRNP first binds the 5' ss via base-pairing between the U1 snRNA and the 5' ss RNA sequence. Subsequently, the splicing factor 1/branch-point-binding protein (SF1/BBP) binds to the branch point (16), the 65kD subunit of the U2 snRNP auxiliary factor (U2AF65) binds the polypyrimidine tract (232), while the 35 kD subunit U2AF35 binds the almost invariant AG dinucleotide of the 3' ss. A minority of mammalian pre-mRNAs are spliced without an AG at the 3' ss (reviewed in (197)). Thus, they are also independent of U2AF35, which is not surprising since the 35 kD subunit of U2AF has been shown to bind directly to the AG dinucleotide (146, 222, 237). All together, these events form the early ("commitment") complex. SR-proteins promote assembly of commitment complexes by assisting binding of both U1 snRNP and U2AF, thereby presumably connecting the 5' and 3' ss. A bridging interaction might also occur between the SF1/BBP and U1 snRNP components binding to the branch site and 5' ss, respectively. Next, U2 snRNP binding to the branch point is required to form the pre-spliceosomal complex. Both complexes involve interactions that are believed to bridge the 5' and 3' ss, which might occur across the intron destined to be spliced. However, another model, "the exon definition model" suggests that a family of splicing factors (SR-proteins, see below) would form a bridge across the exon. Subsequently, binding of U5 and U4/6 snRNPs allows the assembly of a mature spliceosome across the intron. The spliced mRNA is then released from the spliceosome and the resulting changes in hnRNP composition on the mRNA facilitate export to the cytoplasm.

In addition to constitutive splicing, i.e. when splicing cuts and pastes the exons in an invariable manner, splicing can also proceed by alternate ways. Alternative splicing can occur when the splicing machinery is given more than one set of splice sites to choose from. Approximately 22% of all human genes are estimated to mature via alternative splicing (49). The obvious advantages for any organism to have spliced genes include the possibility to express a gene or not depending on if the pre-mRNA is spliced or not. This is utilized for temporal, tissue specific or developmental expression of certain genes (for reviews see for example (25, 225)). In addition, alternative pre-mRNA splicing allows a single gene to produce multiple mRNAs, which are translated into different protein species, and thus represents a tool to regulate gene expression.

The best-studied case of regulated alternative splicing concerns the sex-determination pathway of Drosophila melanogaster (reviewed in (75)) (Figure 7). The presence of two X chromosomes, i.e. the X chromosome to autosome ratio, in female flies stimulates expression of the female-specific sex-letal protein (Sxl). Certain so-called numerator genes, located on the X chromosome, encode for transcriptional regulators that activate a female specific promoter driving the sxl gene. Male flies, which have only one X chromosome, produce less of the transcriptional regulatory proteins and, as a consequence, activate an alternative, male specific promoter. The transcriptionally activated production of a female-specific Sxl protein induces a feedback loop on splicing that maintains a production of the female-specific protein (12), where the male and female specific sxl pre-mRNAs are alternatively spliced by the generic splicing machinery. The male transcript yields a short peptide with no known function whereas the female-specific Sxl protein is biologically active and, aside from controlling its own splicing, also regulates the alternative splicing of the transformer (tra) and male-specific lethal-2 (msl-2) transcripts that encodes for components of the sex-determination pathway of Drosophila. Expression of female-specific Tra is achieved by Sxl mediated alternative 3' ss usage, thereby excluding an exon that is non-
sex-specific and contains a termination codon (24). Tra, in turn, regulate splicing of the double-sex (*dsx*) (33) and fruitless (95) transcripts encoding for factors that stimulate female differentiation and inhibit male behavior, respectively. The Tra protein, in conjunction with SR-proteins, regulates *dsx* alternative splicing by binding to a purine rich element (PRE) of a splicing enhancer within the *dsx* transcript (199). The *dsx* PRE, which contains two consecutively repeated AGGACAA motifs, can actually operate as a binding site for ASF/SF2 and function as an ASF/SF2 specific exonic splicing enhancer. Sxl mediated regulation of *msl-2* is somewhat different in that Sxl here binds to the 5’UTR and the 3’ part of the *msl-2* transcript and inhibit splicing. The unspliced transcript is then exported to the cytoplasm along with Sxl bound to it. Sxl subsequently inhibits translation possibly by preventing mRNA circularization, which has been proposed to be required for efficient initiation of mRNA translation (reviewed in (185)). Importantly, the presence of circular mRNAs has been confirmed by atomic force microscopy (218).

Sxl is an RNA-binding protein with an amino (N)-terminal domain of unknown function; an RBD that has affinity for certain stretches of Us (poly(Y)-tracts); and a carboxy (C)-terminal domain believed to mediate protein-protein interactions. Sxl is believed to regulate alternative splice site usage by competing with U2AF for binding to the poly(Y)-tract preceding the male-specific 3’ ss in *sxl* and *tra*. As a consequence, U2AF is forced to bind to the lower affinity female-specific site. However, recent data complicates this model by showing that a fusion protein containing only the N-terminal domain of Sxl, plus the β-galactosidase protein, functions as a dominant negative that blocks the autoregulatory feed-back loop of Sxl and kills females (55). In addition, the N-terminal domain fusion protein promotes female specific splicing in males of the *tra* pre-mRNA. Albeit well studied, a full understanding of the regulation of alternative splicing in *Drosophila* underlying its sex determination pathway is lacking.

Alternative splicing can also be used for temporal regulation of expression of different proteins, as shown in the adenovirus system. Early during an adenovirus infection, splicing of most viral genes occurs to the closest (i.e. most proximal) splice site. As infection proceeds, there is a general tendency for the virus to remove a larger portion of the primary transcripts by choosing more distal splice sites. As a consequence, expression of proteins needed for e.g. capsid assembly is enabled.

**SR-proteins**

The SR-protein family (reviewed in (69, 81, 132)) was originally identified using a monoclonal ab (mAb104) recognizing phosphorylated epitopes that colocalize with snRNPs on *Xenopus laevis* lampbrush chromosomes (181, 231). Subsequently, Roth and colleagues found a family of related proteins including the previously identified member ASF/SF2 (alternative splicing factor/splicing factor 2) (73, 120), that was rich in serine (S) and arginine (R) aa residues, hence the name SR-proteins. The 33 kD cellular splicing factor ASF/SF2 was originally isolated as an activity that controls alternative splicing of SV40 early pre-mRNA *in vitro* (73) and in parallel as an activity essential for splicing of the β-globin pre-mRNA *in vitro* (120). Members of the SR-protein family are characterized by an N-terminal RNA-binding domain (RBD) that interacts with the pre-mRNA and a C-terminal arginine-serine (RS) domain believed to mediate protein-protein interactions (7, 119, 221, 223). To date, the family comprises 10 members. Half of the known SR-proteins, including ASF/SF2, contain an additional degenerate RBD. The RBD2 of ASF/SF2 has been reported to play a dominant role in substrate specificity (142). Like many other proteins involved in post-transcriptional
regulation SR-proteins are modular in function, which means that they can exert positive or negative effects. For the effects of SR-proteins on splicing, this depends upon the specific pre-mRNA and the phosphorylation status of SR-proteins (223). The SR-protein family is evolutionary very well conserved within metazoa (animal kingdom). Conservation includes protein sequences as well as protein size on SDS-PAGE.

Figure 7. Sex-determination in *Drosophila* through regulated alternative pre-mRNA splicing. X-chromosome encoded proteins stimulate transcription from the *sxl* promoter. The Sxl protein stimulates its own splicing by preventing U2AF binding to the poly(Y)-tract upstream of a strong male-specific 3' ss. In parallel, the Sxl protein prevents usage of a non-productive exon from an intronic position on the *tra* transcript. In contrast, the Tra protein stimulates inclusion of a female-specific exon.
from an exonic position, in the \textit{dsx} transcript. The Dsx protein subsequently promotes female differentiation. Figure adapted from reference (75).

SR-proteins are believed to function at the earliest steps of spliceosome assembly by binding to the pre-mRNA and subsequently recruiting essential splicing factors to the pre-mRNA. This model is based on the observation that early spliceosomal complexes are not formed in extracts deficient of SR-proteins (i.e. cytoplasmic S-100 extracts) (reviewed in (69)). Addition of a single SR-protein to such extracts is sufficient to restore spliceosome assembly and splicing. The prototypical SR-protein ASF/SF2 has been shown to interact both with U1-70K (a component of the U1 snRNP) (109, 119) and the 3’ ss AG binding protein U2AF35 (221). Like many other protein components of the splicing machinery, U1-70K and U2AF35 share the property of having an RS(-like) domain (reviewed in (23)). Thus, recruitment probably occurs via RS-RS domain interactions (37). Since SR-proteins recruit splicing factors binding to the 5’ ss as well as the 3’ ss they have been proposed to function as bridging proteins, connecting the 5’ and 3’ ss. In addition, a number of \textit{in vitro} studies have provided data arguing that SR-proteins partake during virtually every step of the spliceosome assembly and splicing reaction (43, 178).

SR-proteins possess RBDs with different RNA-binding specificities (for a review see (81, 195)), which probably contribute to their individual roles as splicing regulators. However, SR-proteins are thought to be functionally redundant with respect to constitutive splicing since any individual family member can restore the splicing activity to inactive S-100 extracts (70, 74, 231). The contradiction between different RNA-binding specificities on one hand and functional redundancies on the other was somewhat explained by use of a substrate lacking a 3’ ss, the downstream exon and with only the 5’ most proximal nt left of the upstream exon (97). In the S-100 complementation assay, SR-proteins were shown to be required for lariat formation. Even though an intact RBD was required for lariat formation no significant binding of SR-proteins to the 5’ ss could be detected. Therefore the authors concluded that the first step of the splicing reaction only requires unspecific binding of SR-proteins to the intron (97).

The functions of individual SR-proteins appear to be more unique in the case of regulated alternative splicing (for a review see (195)). \textit{In vivo}, SR-proteins are believed to determine which alternative 5’ or 3’ ss will be used by recruitment of U1-70K and U2AF, respectively. This can presumably only occur when SR-proteins bind to the exons, at a position adjacent to the ss. Regulated alternative splicing generally involves combining splice sites that are weak, by deviating considerably from the ss consensus sequence, and hence are inefficiently spliced without splicing activators such as SR-proteins. Thus, individual SR-proteins likely plays a more decisive role in regulating alternative splicing by recognition of and binding to various splicing enhancer elements. In addition, the use of 5’ splice sites that some SR-proteins promote can be counteracted by certain non-essential splicing factors, like hnRNPs (141). For example, ASF/SF2 generally promote proximal splice site usage while hnRNP A1 antagonizes this and thus promote distal 5’ splice site usage (36, 140).

Only a few of the SR-proteins encoded by metazoans have been shown to be essential for cell survival and embryonic development. Conditional knock-out of ASF/SF2 protein expression in a chicken cell line resulted in cell death (211). In this experimental set-up the endogenous ASF/SF2 alleles were replaced with an ASF/AF2 allele expressed from a tetracycline-repressible promoter. Noteworthy, ASF/SF2 was
also shown to be regulated by an autoregulatory loop, thereby limiting its own expression (211). RNAi mediated gene silencing in Caenorhabditis elegans (C. elegans) demonstrated that the ASF/SF homologue rsp-3 was essential for early embryonic development. However, individual RNAi mediated silencing of six other SR-protein encoding genes had no effect on embryonic development, although silencing of all six was lethal (130). Further, a null-mutation of the Drosophila B52/SRp55 gene results in lethality during fly embryogenesis (175).

<table>
<thead>
<tr>
<th>ASF/SF2</th>
<th>RBD1</th>
<th>G</th>
<th>RBD2</th>
<th>RS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>98</td>
<td>106</td>
<td>182</td>
<td>198</td>
</tr>
</tbody>
</table>

**Figure 8.** Schematic representation of the SR-protein ASF/SF2. G denotes a glycine-rich region. The sequence between aa 182 and 198 has been described as a "hinge" region.

The RS domains of SR-proteins are to a large extent interchangeable as shown e.g. using the conditional knock-out strategy mentioned above. Wang et al. showed that replacing the RS domain of ASF/SF2 with one from other SR-proteins (e.g. SRp20, SC35, SRp40) rescued cell viability in chicken cells depleted of ASF/SF2 expression. Interestingly, the RS-like domains from so called SR-related proteins (reviewed in (23)) such as U2AF65 was not able to functionally replace the RS domain of ASF/SF2 with respect to cell survival. Previous studies have claimed that the RS domain of ASF/SF is dispensable for its regulatory role in alternative splicing, both in vitro (34, 238) and in vivo (210). The discrepancy between the results from the latter studies and the knock-out strategy mentioned above can be explained by the different methods used where the cell viability assay represent a functional test in which the importance of ASF/SF2 for total cellular splicing is tested. The splicing activation capacity of SR-proteins on weak splice sites, i.e. when recruited to splicing enhancer elements, has been reported to positively correlate with the length of the RS-domain (82). In addition, from recent in vitro data it has been proposed that the RS-domains have their main function in activation of weak, i.e. enhancer dependent, splice sites (236). It should be mentioned that SR-proteins activate splicing via enhancer sequences not only on alternatively spliced pre-mRNAs but also on constitutively spliced pre-mRNAs. Constitutively spliced pre-mRNAs with divergent and weak splice sites are believed to display the greatest need for the enhancer activity of SR-proteins. However, high affinity binding sites for SR-proteins have also been found in efficiently spliced transcripts such as β-globin (186).

Importantly, phosphorylated but not unphosphorylated SR proteins can restore the activity in splicing deficient extracts. Furthermore, phosphorylation of the RS-domain affects interactions between SR-proteins and various splicing factors. For example, phosphorylation of ASF/SF2 enhances its interaction with U1-70K (223) whereas interaction between ASF/SF2 and U2AF35 is unaffected by phosphorylation (224). Increased phosphorylation of the RS-domain correlates with reduced interactions of SR-proteins with themselves, as shown for ASF/SF2 and SRp40 (224). Moreover, unspecific RNA-binding by SR-proteins decreases with phosphorylation (194, 223).

SR-proteins are to a large extent found within the cell nucleus in so called intranuclear granules (speckles), which are believed to constitute storage sites for
inactive splicing factors. It has been shown that activation of transcription recruits SR-proteins from the speckles to transcriptionally active sites (149), for which phosphorylation of the RS-domain is required (148). Moreover, reports have shown that some SR-proteins shuttle between the nucleus and the cytoplasm. ASF/SF2 for example was shown by use of the heterokaryon approach to be exported to the cytoplasm in an RS-domain and pre-mRNA dependent way (35).

In summary, the functions of SR-proteins as splicing regulators have been extensively studied in vitro and in transiently transfected cells whereas the work done in vivo in animals is more limited.

**Sequence elements required for splicing.**

Four sequences within a pre-mRNA are necessary for recruitment of the splicing machinery. These are the 5' and 3' ss with the almost invariable GU and AG sequences, respectively, and furthermore the branch point with its catalytically active adenosine residue and finally, in the case of mammalian cells, a stretch of pyrimidines ("poly(Y)-tract") located between the branch site and the 3'ss. A comprehensive statistical comparison of sequence elements required for splicing (233), shows that the branch point of human introns is most often found 15 to 30 nts upstream of the 3' ss. Thus, the poly(Y)-tract is expected to be no longer than 30 nts. In reality the median size of the poly(Y)-tract is 4-8 nts (233). The length of the poly(Y)-tract, which correlates with its ability to recruit U2AF, is often weakened by the presence of intervening purines.

Interestingly, sequences surrounding the splice sites and the branch point are highly conserved in budding yeast whereas they are not in metazoan. In addition, in contrast to metazoan no cases of alternative splicing in budding yeast have been reported. This probably relates to the fact that budding yeast lack metazoan SR-protein homologues. None has been found in *Saccaromyces cerevisiae* while only two SR-proteins have been found in fission yeast, *Schizosaccaromyces pombe* (reviewed in (115)). A more complex splicing machinery is probably needed in metazoan due to the presence of divergent splice site sequences and the need to regulate alternative splicing.

Cellular transcripts that are prematurely terminated, not polyadenylated or incorrectly spliced rarely reach the cytoplasm. Many of these are probably degraded by a cellular mechanism referred to as “nonsense-mediated mRNA decay” (NMD) (reviewed (96, 133)). Further, proofreading of correct splicing potentially occurs at the level of pre-mRNA degradation since splice site consensus sequences in adenovirus RNA has been shown to be preferentially accessible to nucleases (155). That is, if there is no assembly of the splicing machinery on or around a splice site then the mRNA is targeted for degradation.

**Splicing enhancer/silencer sequences**

Beside their role in basal splicing, SR-proteins function as components of splicing enhancer complexes. SR-proteins most often function as splicing regulators by binding to enhancer sequences found in the downstream exon. However, they have also been shown to bind to silencer sequences within the intron (22, 112) or to sequences of the upstream exon (95). Whether the SR-protein binding site in the pre-mRNA represents an enhancer or a silencer sequence is position dependent. Interestingly, SR-proteins can in rare cases act as silencers from a position in the exon, for example in splicing of IgM exons (111).
The function of splicing enhancer sequences is distance dependent. That is, the further away from the splice site they are positioned, the weaker is the splicing activation potential. Consequently, most splicing enhancers are located within 100 nts from the 3' ss. Multiplying the number of SR-protein binding sites increases the strength of a splicing enhancer element (82). Similarly, increasing the number or RS repeats increases the enhancer strength of an SR-protein (82). The downstream 5' ss has also been shown to act as a splicing enhancer (122, 215). Introns with weak poly(Y)-tracts generally require SR-protein binding to a downstream exonic enhancer. From this position SR-proteins recruit U2AF to the 3' ss, thereby enabling assembly of the spliceosome.

Adenovirus as a model system

All adenovirus pre-mRNAs, except the transcript encoding for pIX, are processed by splicing. Further, most adenovirus transcription units are temporally regulated at the level of alternative splicing (Figure 2). This enables the virus to express several proteins from one primary transcript, thereby maximizing the coding capacity of a limited genome size as well as enabling viral proteins to be differentially expressed during the early and late phase of the virus life cycle. Several in vitro and a few in vivo studies have shown that SR-proteins regulate alternative splicing of adenovirus transcripts (100, 112, 210, 212). However, with exception of the work presented here (150), no studies have been performed to assay for the role of SR-proteins on viral pre-mRNA splicing throughout a lytic adenovirus infection.

Adenovirus encodes for a few transcripts that are regulated by alternative 5' ss selection (E1A and E1B) although most are regulated by alternative 3' ss usage (e.g. MLTU). All MLTU transcripts are generated through alternative 3’ ss usage where the noncoding tripartite leader sequence is coupled to the respective 5’ ends of the mRNA bodies. A few of the MLTU mRNAs also display removal of part of the exon sequence to yield additional subspecies. The MLTU region that exhibits the most strict regulation is region 1 (L1), which makes it an appropriate model for mechanistic studies of regulated splicing. L1 gives rise to two proteins: the non-structural 52,55K protein and the structural IIIa protein (Figure 9). The 52,55K mRNA is expressed during the early phase of infection while IIIa splicing is undetectable at this time.

Mühlemann et al. in one study used mini-constructs of several of the MLTU mRNAs. The temporal shift in splicing of the L2 region was nicely reconstituted in vitro where penton base mRNA splicing was inhibited in Ad nuclear extracts, compared to uninfected HeLa extract (153). Splicing of pVII and pV in particular, which represents more distal splice sites, was on the other hand activated in Ad nuclear extracts. Thus, the L2 region could also serve as a model substrate for studies of regulation of alternative splicing.

As infection proceeds to the late phase splicing of the L1-IIIa mRNA increases, which presumably results from a virus-induced modulation of cellular factors that redirects the splicing machinery. This occurs by stimulation of IIIa 3’ ss usage combined with a repression of 52,55K splicing (121). As shown on mini-constructs of the L1 gene, the IIIa splice site functions poorly in uninfected nuclear extracts. In addition, SR-proteins have been shown to bind to a purine-rich element, containing three repeated GAGGA motifs, upstream of the IIIa branch point. The IIIa repressor element (3RE) GAGGA sequence show similarity to the two ASF/SF2 high-affinity binding sites that were obtained by SELEX (systematic evolution of ligands by exponential enrichment), namely an RGAAGAAC octamer and an AGGACAGAGC.
decamer (196). Replacing the 3RE with consensus binding sites for ASF/SF2 still results in an ASF/SF2 mediated inhibition of IIIa splicing (112). Recent experiments, where the 3RE was replaced with the bacteriophage MS2 operator site, has shown that a fusion protein consisting of the MS2 coat protein and the RBD2 of ASF/SF2 is sufficient to reproduce the repressor activity of ASF/SF2 (52). SR-proteins binding to the 3RE were initially proposed to sterically outcompete U2snRNP binding to the branch point sequence. However, later experiments by Kanopka et al. demonstrated that the 3RE functions as a repressor of splicing at various positions in the intron preceding the IIIa 3' ss. Likewise, a MS2 operator site (above) mediate repression from different places within the intron (52). The 3RE has also been shown to act as a splicing-enhancer element when placed in the downstream exon, while it has essentially no effect on splicing when positioned just upstream of the 5' ss. The general conclusion that emerged from these studies was that SR proteins function as activators or repressors of splicing depending on where on the pre-mRNA they bind. SR-proteins to the 3RE are dependent on the phosphorylation status, with hyperphosphorylated SR-proteins binding most efficiently (113). SR-proteins purified from adenovirus late-infected cells are hypophosphorylated and the viral E4-ORF4 protein has been shown to mediate a virus

---

**Figure 9.** L1 pre-mRNA splicing is temporally regulated by alternative 3' ss usage. IIIa mRNA splicing is inhibited during the early phase of infection by hyperphosphorylated SR-proteins, which bind to the 3RE. As infection proceeds, the E4-ORF4 protein relieves this repression by dephosphorylation of SR-proteins. The poly(Y)-tract sequence (including the branch site) of the IIIa 3' splice site functions as a virus infection-dependent splicing enhancer (3VDE) that recruits an unknown factor, substituting for U2AF, late during infection.
induced dephosphorylation of SR-proteins (113)), possibly by recruitment of the cellular protein phosphatase 2 A (PP2A).

Importantly, recent in vitro experiments show that the IIIa 3′ ss usage is activated by a mechanism other than the relief of SR-protein inhibition (154). Interestingly, activation of IIIa 3′ ss usage in adenovirus infected extracts does not correlate with increased binding by U2AF65 to the poly(Y)-tract preceding the IIIa 3′ ss (154), suggesting a role for an additional (viral?) protein substituting for U2AF65. In fact, in extracts from adenovirus late-infected cells, the poly(Y)-tract plus the branch site of the IIIa 3′ splice site acts as a splicing enhancer and was hence named IIIa virus-infection dependent enhancer (3VDE) (Figure 9). Further, 3VDE can confer enhanced splicing to unrelated pre-mRNAs, shown by replacing the poly(Y)-tract of the B-globin pre-mRNA with 3VDE (154).

Since SR-proteins are essential factors needed for e.g. spliceosome assembly, they are also needed during the later stages of infection. At this time, SR-proteins show a reduced RNA-binding capacity as a consequence of the virus induced dephosphorylation (113). Hence, one can assume that the specific RNA-binding properties of SR-proteins are of less importance during the late phase of infection. Thus, they might primarily function in protein-protein interactions, i.e. as bridging factors, during the late phase of infection. If splicing of late viral mRNAs is independent of SR-proteins that have the ability to specifically bind RNA, one can assume that these pre-mRNAs are spliced efficiently without aid from SR-protein dependent enhancer complexes, since formation of the latter ones require sequence specific RNA binding. Collectively, this indicates that the late ss of adenovirus have evolved to include enhancer sequences, like 3VDE, that function exclusively in a late-infected environment.

All Ad5 splice sites except those encoded from the E3 region, which is partially deleted in most adenovirus vectors, are listed in table 1. Splice site sequences for Ad2 has been added in the few cases where the closely related serotypes differ. The degree of DNA sequence homology between Ad5 and Ad2 is, as previously mentioned (see “Taxonomy”), 94.7%. Out of the 19 5′ and 29 3′ splice sites listed in Table 1 only four are found to be different in Ad2. Thus, the DNA sequence homology of the listed splice sites are 99.2% identical, meaning that splice sites are more conserved than the overall DNA sequence of Ad. More than 60% of the splice sites are acceptor splice sites, which simply reflects the more prevalent use of alternative 3′ than alternative 5′ ss usage during an adenovirus infection. Further, almost 50% of the listed splice sites include a nonsense codon that is removed by splicing. This could represent a check-point for proofreading of correct splicing and a potential target for NMD.

Like mammalian splice sites, the 5′ ss derived from adenovirus pre-mRNAs shows a high complementarity to U1 snRNA. The mammalian ss consensus is AG/GU/RAGU and the adenovirus 5′ ss consensus sequence is AG/GU/AAG (/ denotes the exon/intron boundary, R=A or G) (Figure 10, upper panel). The mammalian 3′ ss consensus sequence is short, representing YAG/N, whereas the 3′ ss consensus sequence from adenovirus pre-mRNAs exhibits a higher degree of conservation and thus represents YYURCAG/R (intron/exon, Y=C or U and N=any nucleotide) (compare with the adenovirus 3′ ss in Figure 10, lower panel). Notably, all E4 3′ ss use a U at position -5. No individual adenovirus 5′ ss displays a full eight nt sequence complementarity to U1 snRNA. However, it is evident that in some of the late genes, such as the MLTU first leader, L4 33K and IVa2 5′ ss, there is a striking similarity with only one nt difference from the mammalian consensus sequence. Hence, it is interesting to note that these late genes represent normal and presumably optimal splice sites in a late-infected milieu that
has been proposed to represent a severely altered splicing machinery with regards to SR-protein status.

<table>
<thead>
<tr>
<th>mRNA</th>
<th>5' splice sites</th>
<th>3' splice sites</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Right strand transcripts</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9S</td>
<td>GAAGAGGUACUGG</td>
<td></td>
</tr>
<tr>
<td>12S</td>
<td>AAGAGGGUGAGGA$^1$</td>
<td></td>
</tr>
<tr>
<td>13S</td>
<td>UCUACAGUAAGUG$^1$</td>
<td></td>
</tr>
<tr>
<td>9, 12 and 13S</td>
<td>UUAAAAGGUCCUG$^3$</td>
<td></td>
</tr>
<tr>
<td>E1B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13S and 22S</td>
<td>GUACAGGUGGCUG</td>
<td></td>
</tr>
<tr>
<td>13S</td>
<td>GUAACGGAGGGG</td>
<td></td>
</tr>
<tr>
<td>Second 5' ss for E1B</td>
<td>AUUGAGGUACUGA</td>
<td></td>
</tr>
<tr>
<td>Second 3' ss for E1B</td>
<td>UUUGCAGCAGCCG</td>
<td></td>
</tr>
<tr>
<td>MLTU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1$^\text{st}$ leader</td>
<td>GUUAGGGUGAGUA$^4$</td>
<td></td>
</tr>
<tr>
<td>2$^\text{nd}$ leader</td>
<td>CGAAGGGUAAGAG$^1$</td>
<td>UCCACAGUCGC$^1$</td>
</tr>
<tr>
<td>i-leader</td>
<td>CGUACAGGCAGGC</td>
<td>AUGCAGGUGACA$^1$</td>
</tr>
<tr>
<td>3$^\text{rd}$ leader</td>
<td>CGCAAGGGUGCC$^1$</td>
<td>UGUAGGUGACUC$^1$</td>
</tr>
<tr>
<td>L1</td>
<td>52,55K</td>
<td>UUCCAGAUGCA$^2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UGAUCAGAUGA$^2$</td>
</tr>
<tr>
<td>L2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penton (III)</td>
<td>CCCUUGAUAGGA$^{1,2}$</td>
<td></td>
</tr>
<tr>
<td>pVII</td>
<td>CUCACAGAUCACG</td>
<td></td>
</tr>
<tr>
<td>pV</td>
<td>UAUCCAGCCGCGG</td>
<td></td>
</tr>
<tr>
<td>L3</td>
<td>pVI</td>
<td>UUUGAGAAGUGGA$^{1,2}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L4</td>
<td>Hexon (II)</td>
<td>CCGCAGAGGAGC</td>
</tr>
<tr>
<td></td>
<td>23K</td>
<td>UUUGAGUCUUUG$^3$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L5</td>
<td>100K</td>
<td>CCAUAGGCAGAA$^1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L5</td>
<td>33K</td>
<td>UGUAGGUGACUC$^1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF-2 to 6/7 (D1)</td>
<td>UUACUGGUAGGC$^1$</td>
<td></td>
</tr>
<tr>
<td>ORF-2 (A1a)</td>
<td></td>
<td>UAUGCAAGAACCAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AAUGCAAGAACC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Ad2)</td>
</tr>
<tr>
<td>ORF-3 (A1b)</td>
<td></td>
<td>UUAUCAGGUCUGU</td>
</tr>
<tr>
<td>ORF-1 and 3/4 (D2a)</td>
<td>GAGAAGGGUGGCGA</td>
<td>GUAAGGGUGGCGA$^1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Ad2)</td>
</tr>
<tr>
<td>ORF-4 (A1c)</td>
<td></td>
<td>UGUAGGUUUGAC</td>
</tr>
<tr>
<td>ORF-6 (A1d)</td>
<td></td>
<td>CUUCAAGCUCUUC</td>
</tr>
<tr>
<td>ORF-6 (D2b)</td>
<td></td>
<td>GAAUGUGUAGGUU$^1$</td>
</tr>
<tr>
<td>ORF-1, 3/4 and 6/7</td>
<td></td>
<td>GCUUCAAGAAAUA</td>
</tr>
</tbody>
</table>
Table 1. 5' and 3' ss sequences of most adenovirus type 5 transcription units. The Ad2 sequences have been included in the cases where they differ from Ad5. ¹ The splice site contains a nonsense codon (UAA, UAG or UGA). ² The splice site is succeeded by a start codon (AUG). ³ An A at position -3 in the 3' ss has been reported to reduce splicing efficiency by 70% (2). Designations for E4 region splice sites given in parenthesis refer to nomenclature used by Virtanen et al. (208).
Figure 10. Compilation of nucleotide frequencies around 5’ splice sites (A) and 3’ splice sites (B) derived from Table 1. A single nucleotide was regarded as consensus when showing a frequency above 50%. Purine (R) or pyrimidine (Y) consensus was assumed when A and G or C and U together reached above 60%. Please note that other criteria can of course be applied.
Regulation of RNA splicing in other viral systems

The reason for why many viruses extensively use alternative splicing could simply be that it represents a mechanistic way to maximize the use of a limited genome size. As mentioned above, alternative splicing provides the ability to regulate the level of the expressed protein as well as a temporal regulation of protein expression. Furthermore, when several mRNAs are derived from the same primary transcript fewer promoter and poly(A) signal sequences need to be preserved. In addition to adenovirus, most other viruses that replicate in the nucleus make use of splicing.

Many retroviruses make use of the host cell splicing machinery to produce unspliced and partially spliced messages that are transported to the cytoplasm. Rous sarcoma virus (RSV), is a relatively simple retrovirus and, as for other retroviruses, there is balance between production of unspliced and spliced transcripts. Transcription from the integrated retroviral genome yields a full length transcript that is used both for packaging of genomes into new virus particles as well as for splicing to produce shorter mRNAs that are translated to many viral proteins needed for virus multiplication. The balance between production of unspliced and spliced transcripts in RSV is partially controlled by an RNA sequence functioning as a negative regulator of splicing (NRS) (144). In vitro analysis of cellular splicing regulatory proteins with potential to interact with the NRS revealed that ASF/SF2, but not SC35 or SRp40, could bind to and form a complex on the NRS (47). ASF/SF2 has been proposed to bind to a purine rich stretch within the NRS (143). Just like the adenovirus 3RE sequence, the RSV NRS element can function as a splicing enhancer when placed in the downstream exon.

Nuclear export of unspliced and partially spliced HIV-1 mRNA is promoted by binding of the HIV Rev protein to the Rev-responsive element within the viral mRNA (reviewed in (104)). DNA viruses that to a large extent encode for intron less genes, such as the large Herpes Simplex virus and the very small Hepatitis B virus, use mechanisms similar to the HIV Rev-responsive element that circumvent the need for splicing for nuclear export to occur as shown for the HSV thymidine kinase gene (165) and the HBV surface antigen gene (106)).

The influenza viruses (segmented RNA viruses) (reviewed in (124)) and Borna disease virus shows that also negative single stranded RNA viruses that replicate in the nucleus use the host cell splicing machinery (50). Influenza viruses have exploited a variety of strategies to increase their genome coding capacities. These include unspliced, spliced, alternatively spliced and bicistronic mRNAs (reviewed in (124)). Just like retroviruses, Borna disease virus expresses certain viral proteins by suppressing splicing via a cis-acting RNA element (187). As a consequence, unspliced and partially spliced RNA accumulate in the cytoplasm.
PRESENT INVESTIGATION AND DISCUSSION

The specific aims with each of the studies were to:

I. Construct two vector systems based on inducible promoters that should be of general interest for use in gene transfer experiments.

II. Reconstruct a recombinant adenovirus expressing the cellular SR-protein ASF/SF2, known from *in vitro* studies to regulate viral splicing. To use this virus to analyze the effects of ASF/SF2 overexpression on viral pre-mRNA processing and viral replication during lytic viral growth.

III. Test whether the strict temporal control of L1 IIIa mRNA splicing is due to deleterious effects of unregulated IIIa protein expression on virus growth.

IV. Silence the E1A mediated activation of uninduced transgene expression that was noted during the work of paper I and II and that hampers the usefulness of the vector systems.
Two novel adenovirus vector systems permitting regulated protein expression in gene transfer experiments (paper I)

A vast majority of cell culture experiments relies on transfection of plasmids bearing the construct of interest. Commonly used reagents for introduction of plasmids into the cell line of choice display with few (if any) exceptions a limited efficiency, typically below 40-50%. This means that, even under optimal conditions, at least half of the transfected cells will actually never express the protein of interest. Another approach for cell culture studies is to establish stably transfected cell lines. This procedure includes clonal selection of successfully transfected cells. Therefore the resulting cell line is homogenous with regards to the transfected construct. In addition, this gives the possibility of selecting a clone that harbors a desired number of inserted constructs. However, integration of plasmid sequences into cellular chromosomes frequently represses transcription from the plasmid-contained construct (67). This silencing phenomenon has been suggested to result from integration of plasmid concatemers. Thus, several clones have to be screened to obtain one that gives the desired expression of the protein of interest. This approach is tedious and furthermore limited to the cell line chosen for transfection.

Conventional transfection methods share one major disadvantage. They do not efficiently transfer genes in vivo (i.e. to animals). Consequently, viral vectors have come to dominate gene transfer to animals and the field of gene therapy.

In the first work, summarized in paper I, we wanted to create expression vectors that could function in vitro as well as in vivo based on adenovirus where the transgene expression could be regulated by addition of an exogenous substance. We chose adenovirus since we were planning to later use the vector systems for in vivo studies, where the effects of selected proteins on regulation of adenovirus alternative splicing was to be analyzed. Furthermore, adenoviruses have many of the characteristics required by an optimal gene transfer vector (see “Adenovirus vectors” section). We included the adenovirus tripartite leader in our constructs as a 5’ untranslated region to ensure an efficient export and translation of the reporter mRNA during adenovirus infection (see “Translation and shut-off of protein synthesis”).

Both adenovirus systems are based on a double-infection strategy (Figure 6) where the activator protein is expressed from one viral chromosome. The activator protein then binds to and activates transcription from a gene-cassette contained on a co-infected viral genome. Binding is inducer dependent and thus relies on addition of RU 486 or doxycycline for the prog- and Tet-ON system, respectively. The use of a double-infection strategy enables calibration of the desired operator:activator protein ratio, which has been reported as advantageous in injections of animals where a lowered diffusion of virus particles leads to very high local viral titers at the site of injection (91). As a result, the excess of activator virus causes an elevated background expression. This can be avoided by usage of less activator virus (91). The background we observed from injections of the Balb/c mice probably relates to this. Thus, adjustment of the operator:activator protein ratio constitutes an additional and valuable level of regulation of transgene expression (Figure 11).
Figure 11. Transgene expression (CAT) resulting from different ratios between reporter (AdTetTrip-CAT) and activator virus (AdCMVrtTA). (1) 1:1; (2) 10:1; and (3) 10:10. NIH3T3 (mouse) cells infected with various ratios between AdTetTrip-CAT and AdCMVrtTA. Cell extracts from 16 hpi were analyzed by a CAT-assay.

The prog- and Tet-ON vector systems were shown to transduce a number of cell types: HeLa cells representing cervical epithelia; normal human foreskin fibroblasts (1523 cells); African green monkey epithelial kidney cells (Cos-7); and mouse fibroblasts (NIH 3T3). Thus, the vectors exhibited a broad species-specificity useful for screening of effects of a protein of interest in various hosts. Surprisingly, the prog-system induced poorly in NIH3T3 cells, which perhaps results from a tighter complex between mouse chaperones and the progesterone receptor LBD of the activator protein.

The regulation of protein expression was tightly controlled as shown by analysis of CAT reporter gene expression. Moreover, reporter gene expression paralleled the amount of inducer added to the cell culture medium. This allows for a fine-tuning of the level of protein expression. The amounts of expressed reporter protein were determined by CAT ELISA and were shown to be 43 and 15 ng per million cells, respectively. To obtain measurable readings of the uninduced background expression of CAT it was necessary to use 100-fold more concentrated extracts. Quantitation of the data revealed ratios for fold induction of 1800 (Tet-ON) and 600 (RU 486-regulated system) times the background expression in HeLa cells. Importantly, Tet-ON and RU 486-regulated CAT protein production exceeded by far CAT expression produced from a recombinant virus where CAT expression was driven by a CMV promoter. This shows that the inducible vector systems not only allow for stringent control of the expression levels but also give a high maximum yield of reporter protein. This is important e.g. when the vector systems are used for recombinant protein production. As expected, almost all cells were transduced, already at 10 MOI, as seen by staining for CAT reporter gene expression. Furthermore, the Tet-ON system was demonstrated to permit an efficient control of gene expression in mice after intramuscular administration of Balb/c mice where the inducer, doxycycline, was fed via the drinking water. Collectively these experiments demonstrate that the Tet-ON and the prog-adenovirus expression vectors systems can be used for a variety of cell culture or animal gene transfer experiments. In subsequent experiments, we could show that our inducibly expressed genes indeed survive the virus induced shut-off of host protein synthesis (Figure 4B).

Although the double-infection strategy can be advantageous in cell culture experiments (mentioned above) it is perhaps more appealing to have each system expressed from one viral chromosome. In vivo administration to experimental animals, e.g. by intramuscular injection, brings about limitations that are not encountered in cell cultures. The most striking difference is the existences of the extracellular matrix as well as components of the immune system that often counteract viral gene transfer.
Consequently, virus vectors usually transduce at lower efficiencies in animals. Hence, it is important to ensure that all components needed for transgene expression reach the target cells. Thus, current work in our lab is focused on reconstitution of the activator and reporter gene cassettes into one vector. The activator protein encoding sequence is inserted in place of part of the E3 region, which is dispensable for virus growth in cell cultures. Hence, this double-expressing virus can be grown in normal 293 cells. This new variant is based on the so called "AdEasy system" (94) where the recombinant virus is constructed in bacteria using standard cloning techniques. The resulting recombinant genome can be excised from the plasmid and transfected to 293 cells to yield a virus preparation. In addition, there should be no need for plaque purification of recombinant viruses constructed using this approach since clonal selection is performed in bacteria, prior to transfection into 293 cells.

**Overexpression of the essential splicing factor ASF/SF2 blocks the temporal shift in adenovirus pre-mRNA splicing and reduces virus progeny formation (paper II)**

ASF/SF2 is one of the few SR-proteins that has been shown to be essential for cell survival (211). Therefore, one can assume that expression of this cellular splicing factor is prevalent and that the protein performs essential cellular activities. Consequently, viruses such as adenovirus will become exposed to this splicing factor upon infection of host cells.

Several *in vitro* studies have shown that SR-proteins, and in particular ASF/SF2, control splicing of adenovirus pre-mRNAs (62, 100, 112, 210). Kanopka et al. previously showed that SR-proteins purified from uninfected HeLa cells inhibit splicing of the IIIa mRNA. This was shown using mini-constructs of the model substrate, the L1 pre-mRNA (112). Inhibition of IIIa splicing correlates with binding of SR-proteins the 3RE. Adenovirus was later shown to inactivate SR-proteins as splicing repressor proteins by a virus-induced dephosphorylation that diminished their RNA-binding capacity (113), thereby relieving their inhibitory effect on IIIa mRNA splicing. Recent data from our group shows that the E4-ORF4 protein selectively targets the ASF/SF2 and SRp30c (62).

Based on these results, we predicted that overexpression of ASF/SF2 would present an obstacle to regulation of adenovirus alternative pre-mRNA splicing. A prediction from *in vitro* studies, where ASF/SF2 has been shown to mostly promote proximal splice site usage, was that ASF/SF2 overexpression would potentially prevent the temporal shift in adenovirus alternative splicing that occurs when the virus enters the late phase of infection. To test this hypothesis we constructed a recombinant virus expressing ASF/SF2 under the inducible control of RU 486-regulated gene cassette (see "Inducible promoters").

ASF/SF2 overexpression was shown to prevent the early to late shift in E1A, E1B and L1 alternative splicing. That is, these regions were withheld in a proximal splice site pattern. E1A 13S splicing was slightly stimulated on the expense of the 12S and the minor 9S mRNAs. The E1A 13S and 12S protein products have been shown to activate the E2 promoter (in a CR1- and 2-dependent way) (reviewed in (160)). The adenovirus E2A product DBP, in turn, has been reported to activate its own promoter, the E1A promoter as well as the MLP (39). We could show that the DBP protein levels were enhanced late during infection as a consequence of ASF/SF2 overexpression.
Furthermore, high ASF/SF2 expression drastically reduced the accumulation of viral MLTU mRNAs, DNA replication, late protein synthesis and virus production. Notably, even though background expression of ASF/SF2 did not significantly affect the L4 100K mRNA levels, the shut-off of host protein synthesis was less obvious.

In all transcription units examined, ASF/SF2 promoted proximal splice site selection. For the MLTU, this was evident by looking at the effects of uninduced background expression of ASF/SF2 seen in 293 cells that in itself was sufficient to severely affect the temporal shift in MLTU pre-mRNA splicing, delay DNA replication, and inhibit late protein synthesis. Collectively, these in vivo results provide support for the hypothesis that viral inactivation of SR-protein function is important for the correct temporal expression of early and late viral mRNAs during a lytic infection, and hence efficient virus multiplication.

Since phosphorylation of SR-proteins has been shown to positively correlate with their ability to regulate adenovirus pre-mRNA splicing (62, 113) it was also important to show that the overexpressed ASF7SF2 was indeed phosphorylated. Interestingly, detection of the overexpressed ASF/SF2 (via its His-tag) showed that it becomes increasingly hypophosphorylated as infection proceeds (Figure 12), possibly as a result of virus induced dephosphorylation. This is in line with predictions from previous studies (113).

![CIP](image)

**Figure 12.** Hypophosphorylated forms of ASF/SF2 accumulate as infection proceeds. Western blot detection of His-tagged ASF/SF2 (using an anti His antibody) from AdG5(His)-ASF infected 293 cells, harvested at 12 (lanes 1 and 3) and 18 hpi (lanes 2 and 4). A temporal difference in phosphorylation remains after treatment with "calf intestinal alkaline phosphatase" (CIP).

The observation that the reduction of L1 mRNA accumulation was relieved when ASF/SF2 expression was turned on at the start of viral DNA suggests that ASF/SF2 exerts its effects on MLTU mRNA levels via alterations of some early product(s). This is, however, a bit complicated since late regions 2 and 4 do not display the same relief upon a later ASF/SF2 induction. Previous results have indicated that viral DNA replication, per se, is enough to induce accumulation of normal levels of L1 mRNAs (125). However, under such conditions the temporal shift in alternative splicing was shown to be incomplete with a preferential accumulation of i-leader containing 52,55K and IIIa mRNA species (125). The latter is in agreement with our observations that ASF/SF2 also produces the IIIa+i mRNA species. Interestingly, the background of ASF/SF2, which was sufficient to alter L1 and L4 splicing did not significantly affect the relative levels of the L2 mRNA species. This contrasts the in vitro results described by Mühlemann et al. (153).

The connection between inhibition of viral DNA replication and the observed increase in DBP protein levels (as a consequence of ASF/SF2 overexpression) could simply be explained by earlier reports that prevention of viral DNA replication by drugs or mutations results in accumulation of E2 proteins (72), although extended incubation
(24-48 hpi) was required in this case. This, however, does not give any detailed clues to the specific regulatory mechanisms.

Since ASF/SF2 stimulates E1A 13S splicing with an accompanying increase of synthesis of the E1A-289R transcriptional activating protein, the other early promoters should be activated at least to wild type levels. This suggests that the observed reduction of MLTU mRNA accumulation is not a consequence of lowered transcription of any early region. Rather, it seems more plausible that the altered splicing patterns of early regions are the primary event causing the phenotypical alterations of the ASF/SF2 virus. For example, ASF/SF2 overexpression could potentially result in a disturbed balance of E4 protein expression. Since ASF/SF2 favors proximal splice site usage, splicing of the ORF3 mRNA should be governed at the expense of ORF4 and ORF6 mRNA production. The E4-ORF4 protein inhibits viral DNA replication while the products of E4-ORF3 and 6 antagonize the effect of ORF4. Consequently, one has to assume that ASF/SF2 favors ORF4 splicing to explain the inhibition of viral DNA replication and late protein synthesis. However, a virus expressing ORF4 in an ORF3 and ORF6 background shows decreased steady-state levels of the viral E2 replication proteins, e.g. DBP (145). Further, the inhibitory effect of ORF4 on DNA replication is counteracted by the presence of either of ORF3 or ORF6 (32). Collectively, we do not believe that ASF/SF2 induced changes in E4 pre-mRNA splicing account for the observed effects on viral DNA replication and late protein synthesis.

Work done with E4-ORF3 and ORF6 mutant viruses show that some mutants have a defective late protein synthesis while DNA replication is normal (31). This suggests that viral DNA replication and late viral protein synthesis are not strictly correlated, which is particularly interesting since DNA replication is believed to be responsible for activation of the MLP. Other virus mutants with severe deficiencies in transcription from the MLP display a marked increase in expression of the E1A, E1B, and E2 early genes (65). In addition, expression from the late gene pIX, which is not contained within the MLTU, is also increased when the MLP is not active (65). These effects on early gene expression are similar to our observations from the ASF/SF2 virus. However, ASF/SF2 decreased the accumulation of the (non-spliced) pIX mRNA. Nonetheless, we believe that the dramatic reduction of major late mRNAs could reflect a decrease in MLP transcription.

The C-terminal domain (CTD) of the largest subunit of RNA polymerase II (pol II), which have been shown play a role in coupling between transcription and mRNA processing (reviewed in (171)), is heavily enriched in serines and also found to be hyperphosphorylated and negative in charge. This part of the polymerase has been shown to represent a loading site for mRNA processing factors (reviewed in (48)). Interestingly, immunoprecipitations with specific antibodies show that SR-proteins co-precipitate with pol II, and vice versa (116). This is apparently not dependent on the pre-mRNA, which points to direct via protein-protein interactions. Further, antibodies against the CTD inhibit in vitro splicing (230). Whether SR-proteins interact with the CTD directly or via pol II associated factors is currently not known. In any case, overexpression of a potential pol II "load", such as ASF/SF2, could change the composition of pol II and its associated factors, as a consequence of ASF/SF2 "overload". The fact that ASF/SF2 overexpression specifically affected MLTU mRNA accumulation is intriguing since the processing fate of a pre-mRNA has been proposed and partly shown to depend on the composition of the transcriptional pre-initiation complex (pol II plus associated factors). Perhaps the MLP specific pol II complex is particularly sensitive for ASF/SF2 "overload".
Concerning other cellular RNA binding proteins beside ASF/SF2, one model suggests that at limiting concentrations of SR-proteins, hnRNPs bind to and cover most of the pre-mRNA, thereby hindering unspecific SR-protein binding. If an SR-protein like ASF/SF2 is expressed at multifold higher levels than normal then it is likely to bind more non-specifically to pre-mRNAs. This could interfere with the splicing machinery, possibly to such a degree that the pre-mRNA is spliced in a cryptic way. Hence, the pre-mRNA could be denied access to the transport machinery and hence be sentenced to NMD. However, the high degree of phosphorylation of ASF/SF2, when expressed from our recombinant virus, would be predicted to increase the RNA-binding specificity of ASF/SF2 (see "SR-proteins").

Regarding the general proximal splice site promoting potential of ASF/SF2, hnRNP A1 induces distal splice site selection, which is the opposite activity compared to ASF/SF2. The balance between splicing regulatory protein like SR-proteins and some hnRNPs have been proposed to contribute to e.g. tissue specific splicing (89). Thus, varying ratios between e.g. ASF/SF2 and hnRNP A1 could equally well contribute to a temporal regulation of adenovirus pre-mRNA splicing. hnRNP A1 binds pre-mRNAs and has been shown to compete with U1 snRNP for binding to the 5' ss (61). A model based on the results by Eperon et al. would suggest that U1 snRNP is deprived of some of its affinity for the 5' ss, as a consequence of hnRNP A1 competition. Thus, U1 snRNP might only bind when encountering highly complementary 5’ ss, which in turn will direct 5’ ss usage. Further, elevated concentrations of ASF/SF2 enhances U1 snRNP binding to all 5’ ss (61), which makes them equally competent for usage. A conclusion would be that, hnRNP A1 favors usage of high affinity ss while ASF/SF2 counteracts this and lets the generic splicing machinery decide, which by default (?) use the proximal ss. This model would get support if one could show that adenovirus early 5’ ss bind U1 snRNP poorly and thus are dependent on enhancer activity from SR-proteins.

Interestingly, 13 nts downstream of the i-leader/leader 3 splice junction is a potential ASF/SF2 binding site, GAGGGA, resembling the purine-rich elements of 3RE to which ASF/SF2 bind. If one assumes that this represents an exonic splicing enhancer, it could explain our observations that ASF/SF2 overexpression stimulated the (abnormal) production of IIIa+i mRNA species, i.e. i-leader inclusion.

Since SR-proteins can stimulate usage of 5’ ss from a close position in the upstream exon, it is surprising to find that the 5’ ss of the E1A 12S and 9S mRNAs, which were both repressed by ASF/SF2 overexpression, are both flanked by a GAAGA sequence (see Table 1 (the first G of the 12S GAAGA sequence is not included)). However, 21 nts upstream of the E1A 13S 5’ ss, a GAGGA sequence is found. If ASF/SF2 does bind to this element this would indicate that there is a minimal distance required for ASF/SF2 to act as an enhancer. When attracted to close to the 5’ ss it might for some reason interfere with the splicing machinery.

Expression of the structural IIIa protein during the early phase of an adenovirus infection results in defects of major late mRNA accumulation (paper III)

The significance of the strict inhibition of IIIa mRNA splicing during the early phase of infection has previously not been investigated. Both the 52,55K and the IIIa protein are required during the late stage of virus particle formation. Thus, it is not obvious why adenovirus confines IIIa mRNA production exclusively to the late stage of infection.
However, if the IIIa protein, for some reason, has negative effects on virus growth if expressed too early during infection then the tight control of L1 alternative splicing appears reasonable. Temporal regulation L1 pre-mRNA splicing could also simply reflect an adoption to a general regulatory mechanism used by the virus. Although it may not be important for L1 protein expression per se, the use of a common regulatory mechanism, by which the virus uses a minimum of effort, could represent optimal fitness.

Previous in vitro studies have shown that the 52,55K 3’ ss is a strong splice site in extracts from uninfected HeLa cells while the IIIa 3’ ss is weak. Conversely, the IIIa ss is the most active L1 3’ ss when assayed in vitro in extracts from adenovirus late-infected cells. Thus, uninfected cells contain factors that inhibit IIIa splicing (SR proteins) while adenovirus infected cells do not and accommodate for a factor that activate IIIa splicing (154). Interestingly, reversing the order of 3’ ss on a tandem pre-mRNA resulted in an almost exclusive IIIa splicing, which shows that the order of 3’ ss presentation is important for the outcome of alternative L1 splicing (123). Importantly, recent experiments show that the IIIa 3’ ss usage is activated in adenovirus infected extracts by a mechanism other than the relief of SR-protein inhibition (154) where the poly(Y)-tract preceding the IIIa 3’ ss stimulates IIIa mRNA splicing. In addition, it has been shown that viral DNA replication and expression of late proteins are needed for the splice site shift in L1 to occur (125). Thus, a late viral protein could mediate some of this effect. However, one should keep in mind that some early genes, such as E4, might actually be expressed at their maximum during the late phase of infection.

To test the hypothesis that IIIa mRNA splicing is strictly inhibited during the early phase of a lytic infection because inappropriate IIIa protein expression has negative effects on virus growth, we constructed a recombinant adenovirus expressing IIIa from a Tet-ON-regulated gene-cassette (see "Inducible promoters").

Unregulated expression of the IIIa protein throughout the infection resulted in a reduction in the steady-state levels of MLTU encoded late structural proteins. Moreover, viral DNA replication was reduced by overexpression of an unregulated IIIa gene. The observation that also a CAT expressing virus shows a decrease in DNA replication could partly result from a bias caused by the experimental setup. Since we first infected HeLa cells with the recombinant viruses, the subsequent uptake of the superinfecting wild type virus could be negatively affected, for example by the fact that cellular receptors are “consumed” upon the primary infection, by receptor mediated endocytosis. Furthermore, 52,55K mRNA splicing was selectively inhibited as compared to IIIa mRNA splicing.

An unregulated overexpression of the IIIa protein from the start of virus infection has only minor effects on DNA replication and late protein synthesis. We therefore propose that the reason that IIIa splicing is strictly controlled during the early phase of infection merely reflects an adoption to an overall mechanism used by adenovirus to turn on late gene expression. The finding that overexpression of the IIIa protein appears to enhance its own production is intriguing. It invites to further investigation regarding the potential role of an autoregulatory mechanism controlling the temporal shift in L1 pre-mRNA splicing.

An adenovirus vector designed for expression of toxic proteins (paper IV)
In paper I we presented two versatile and efficient research tools where E1-deficient adenovirus vectors were combined with regulatory promoters. These were shown to allow for a stringent control of gene expression, with no detectable background, when used for infection of cells not permissive for virus replication.

However, during the work described in paper I and II, we became aware of a problem exclusively associated with the growth of the recombinant viruses in the 293 producer cell line. When the vector systems are used for infection of the 293 cells, a troublesome background expression is evident, where basal transgene expression is active in the absence of inducer.

293 cells express both E1A and E1B proteins and are therefore used as complementing cells for amplification of E1-deficient adenovirus vectors. E1-deficient adenovirus vectors must inevitably pass through 293 cells for rescue of recombinant virus particles. Cloning of a potentially toxic gene may be difficult, if not impossible, in a situation where the gene to be cloned has inhibitory effects on virus multiplication. When reconstructing recombinant adenovirus vector it is hence highly desirable to silence the expression of the transgene during the cloning and amplification phase. The logical conclusion from our observations would be that either of the E1 proteins is the cause of the unwanted transcriptional activation of transgene expression. The main candidate would then be the E1A-289R transcriptional activator protein.

We predicted that the unwanted background expression resulted from E1A activation of transcription from the TATA sequence of the minimal promoters used to drive the reporter gene rather than via interaction with the E1A enhancer elements located around the packaging signal. This can be concluded from a study in which the Tet-responsive element was cloned in a leftward direction near the right ITR so as to avoid potential interaction with the E1A enhancer elements. Still, a troublesome (uninduced) background transcription that reduced virus growth was observed from this Tet-OFF regulated gene-cassette (183).

To test whether the E1A 289R protein contributes to the unwanted transgene expression in our vector constructs, we infected a U2OS cell line expressing the E1A-289R protein under the inducible control of a Tet-ON-regulated promoter. When E1A protein expression was induced eight hours prior to infection, background expression levels of ASF/SF2 became detectable (Figure 13).

![Cells treated with dox 8 h prior to infection](image)

**Figure 13.** The E1A-289R protein mediates activation of uninduced basal reporter gene expression. A U2OS cell line, expressing the E1A-289R protein under the inducible control of a Tet-ON regulated promoter, was MOCK infected (lanes 1 and 5), infected with AdG2(His)-ASF plus AdCMVProg in the absence (lanes 2 and 6) or presence of RU 486 (lanes 3 and 7) or wild-type dl309 virus (lanes 4 and 8). Cells pre-treated with doxycycline (thus expressing the E1A-289R protein) (lanes 5-8) gave rise to an increased expression of (uninduced) reporter gene expression (lane 6).
To prevent E1A-289R mediated activation of basal transcription, we introduced an additional regulatory element comprising the E. coli lac operator to which the Lac repressor protein bind with high affinity. The approach relied on construction of a stable 293 cell line expressing the Lac repressor protein (293-LacI) and reconstruction of a new ASF/SF2 reporter virus where three lac operator sequences was inserted in the intron located downstream of the ASF/SF2 gene. This vector was named AdG5(His)-ASF(Lac). The idea was to use the Lac system to restrict background expression, in the absence of inducer, by preventing production of full-length mRNA from promiscuously initiated transcripts.

In conjunction with a LacI expressing 293-cell line, which we established, high affinity binding of the LacI repressor protein to the lac operator sequence was shown to prevent transcriptional elongation. Although markedly reduced, the background expression was still evident on long exposures of autoradiograms. Consequently, the ASF/SF2(Lac) virus did not behave exactly like a wild type virus with respect to accumulation of L1 mRNAs.

In 293 cells, the normally strict regulation of the prog system is insufficient to prevent an E1A mediated activation of basal expression. The promiscuous E1A transcriptional activator protein activates transcription from the same promoter elements used by the basal transcription machinery (reviewed in (17)). Hence, it should be difficult to silence E1A mediated transcription without simultaneously affecting the ability to transcribe at all from e.g. minimal promoters used in inducible regulatory sequences. Inducible regulation is dependent on at least some minimal element for basal transcription factors to function. Further, E1A transcriptional enhancer elements located in the left ITR are not thought to play an important role here due to reasons mentioned above. Even though our vector system was not perfect through the double-regulated system we believe that the substantially lowered background expression will more readily allow for reconstruction of recombinant viruses expressing proteins with significant toxicity to adenovirus replication. Future studies will tell whether this is true.
CONCLUSIONS

I. Inducible regulated promoters can be successfully introduced into the E1 region of adenovirus. There is no problem associated with having the activator protein expressed from a separate viral chromosome. In fact, this two-vector infection strategy allows for the calibration of the desired ratio between activator and reporter virus (Figure 11). Transgene expression levels can be fine-tuned over a wide inducer dosage range. In addition, inclusion of the tripartite leader is important for efficient transgene expression during the late stages of infection, when studies such as those described in paper II and III are undertaken. Having showed that the vector systems function in vitro as well as in vivo, we believe that they should be of general interest for many gene transfer experiments where for example transduction is hampered by low transfection efficiencies and expression of the transgene is only desired during a certain period of the experimental protocol.

II. The prototypical SR-protein, ASF/SF2, stimulates both proximal 5’ and 3’ ss usage during an adenovirus infection. This general activity can be explained by the affinity of ASF/SF2 for purine-rich elements in the vicinity of the splice sites. However, I favor an alternative mechanism where overexpression of ASF/SF2 counteracts the functional antagonism of cellular hnRNPs, such as hnRNP A1. Furthermore, ASF/SF2 may directly inhibit transcription from the MLP, a finding which may be significant since transcription and splicing has recently been proposed to be coupled.

III. L1 pre-mRNA splicing is regulated by alternative splicing, which restricts IIIa mRNA and protein expression exclusively to the late phase of infection. This seems not to be due to deleterious effects of an early IIIa protein expression. Rather, the marginal effect on virus growth argues for a selective advantage of this temporal regulation, through increased fitness of the virus. However, an interesting effect of IIIa protein expression on L1 pre-mRNA splicing was noted possibly implying the existence of an auto-regulatory feed-back loop.

IV. We show that combining an inducible promoter and a LacI based roadblock to transcriptional elongation serves as an efficient strategy to silence unwanted background transgene expression during recombinant virus growth in the 293 producer cell line. In turn, this should make it more likely to successfully rescue genes encoding for toxic proteins into adenovirus vectors.
FUTURE PERSPECTIVES

The inducible viral vectors presented here should have great potential for future use in different gene transfer experiments. With the high transgene expression levels (1-2% of total protein, paper I and II) obtained in 293 cells they challenge e.g. baculovirus for use as eukaryotic expression vectors. In contrast to baculovirus, adenovirus vectors enable expression of proteins in mammalian cells, which confers more relevant post-translational modifications if human proteins are studied. In addition, the broad species-specificity allows for identification of interacting proteins in various hosts, i.e. isolation of species homologues.

The ASF/SF2 virus is currently used for continued studies of the role of individual SR-proteins in regulation of adenovirus splicing. The strategy is to purify His-ASF/SF2 from early and late-infected 293 cells and use the purified protein in *in vitro* splicing reactions to see if it can reproduce the activity of the SR-protein fractions from uninfected and adenovirus infected cells (compare early and late-isolated His-ASF/SF2, respectively). Previous work with SR-proteins purified from infected cells can not exclude the existence of contaminating factors co-precipitated in the SR-protein containing fraction.

Our current model for temporal regulation of L1 pre-mRNA splicing proposes that the viral encoded E4-ORF4 protein interacts specifically with ASF/SF2 (62). In addition, E4-ORF4 interacts with the cellular phosphatase PP2A and induces phosphorylation. As a consequence, SR-protein mediated inhibition of IIIa mRNA splicing is relieved in late adenovirus infected cells. This model will be further investigated using the ASF/SF2 virus where immunoprecipitations of ASF/SF2 from early and late phase of infection, respectively, will reveal its interacting partners where among others E4-ORF4 is expected to show up, possibly in different disguises at different time points.

A thorough examination of the mechanism(s) by which ASF/SF2 specifically reduces accumulation of MLTU mRNAs could possibly reveal an interaction between the CTD of RNA polymerase II and ASF/SF2. The resulting complex possibly determines the rate of MLP transcription. In addition promoter sequence specific loading of pre-mRNA processing factors, onto RNA Pol II, could affect splicing of the primary transcript.

Finally, I think it would be worthwhile to statistically further investigate the differences in splicing decisive sequences to see in what way early and late genes differ. That is, to see what decides whether the splicing machinery regards a splice site as early or late.
ACKNOWLEDGEMENTS
This work was carried out at the department of Medical Biochemistry and Microbiology, Section of Microbiology. The project was supported by grants from the Swedish Cancer Society and the Göran Gustafsson Foundation for Natural Science.
I wish to thank the following people for support and encouragement:

Göran Akusjärvi, my supervisor, for giving me the opportunity to join his lab where I was given stimulating projects and where I could develop a self-reliant and scientific way of working, and last but not least for careful reading and valuable comments on my thesis.

Göran Magnusson for giving me the opportunity to explore the academic career.

Mats Nilsson for introducing me into the “hands-on” world of molecular virology.

All the staff at IMBIM for support during the years and especially Erika and Olav for assistance with the preparations needed for the dissertation.

All the people at the lab.

My present group: Anders, Anette, Baigong, Camilla, Catharina, Cecilia, Christina, Dan, Edyta, Josef, Martin, Peter, Regine, Saideh, Shaoan, Tanel and Vita.

Members of my former group and others: Alexis (for being an uplifting person to talk to), HongYun, Shouting, Carin, Hamid, Brian and all the rest.

Mimmi Shoshan and Stig Linder for nice and, as it turned out, fruitful co-operation.

My dear friends Patrik, Micke, Mange, Patrik O, Niklas, Ola, Malin, Lisen, Ullis, Lene, Jenny, Charlotte, Oskar and the rest for innumerous scientific as well as non-scientific discussions, stimulating friendship, sports events, sauna sessions, pub evenings et c.

My mother, my father and my brother for everything.

My late brother Fredrik for all we experienced.

Laura for her love and support.
REFERENCES


