Promoter and 5'splice site interactions in retroviruses and retroviral vectors

Von der Naturwissenschaftlichen Fakultät der

Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktorin der Naturwissenschaften

Dr. rer. nat.

genehmigte Dissertation

von

Dipl. Biol. Daniela Zychlinski, geb. Müller

geboren am 28. Oktober 1978 in Hannover

2009

Referent: Prof. Dr. med. C. Baum

Korreferent: Prof. Dr. rer. nat. H. Holtmann

Tag der Promotion: 07. April 2009

meiner Oma Luise

II Abstract

Retrovirus-derived vectors are frequently used for the genetic modification of replicating cells and offer potential value for gene therapy. However, retroviral vector-mediated insertional mutagenesis has been shown to induce clonal imbalance and malignant transformation in animal models and humans, demonstrating the need for improved vector design to increase safety. The present study adressed the question of whether “self-inactivating” (SIN) retroviral vectors containing a weak cellular promoter/enhancer can reduce the risk of insertional mutagenesis. To evaluate the potential of retroviral vector sequences to activate neighbouring promoters, we compared cellular and retrovirus-derived promoter/enhancers in two different in-vitro assays. Furthermore, the impact of retroviral vector design on transforming potential was studied in cultured cells. Cellular promoters showed less enhancer activity than viral promoters, resulting in a strongly decreased potential to upregulate a cellular proto-oncogene *Evi1*, and thus a greatly decreased likelihood of immortalizing primary bone marrow cells. In summary, the cellular promoter/enhancers tested here are less likely to activate neighboring genes than are viral promoters, and are thus promising candidates to be incorporated in potential therapeutic vectors.

RNA processing signals may also play an important role in improving the safety of retroviral vectors. The presence of a 5’splice site (ss) in retroviral vectors enhances transgene expression and titer. However, a 5’ss potentially enables the generation of fusion transcripts that link retroviral and cellular exons. Gammaretroviral vectors used in clinical trials harbor the authentic splice sites of murine leukemia virus (MLV). So far alternative splicing of MLV is not understood in detail and retroviral vector development would benefit from insights into the underlying mechanisms. We were able to show that MLV attenuates its 5’ss, in contrast to other retroviruses that control alternative splicing via weak 3’ss. Attenuation is achieved by upstream sequences that fold into a complex secondary structure. Most of the structure could be functionally replaced by a heterologous stem loop. Mechanistically, the secondary structure reduces the accessibility of the 5’ss for U1 snRNP, which normally directs the RNA into the splicing pathway. To increase the complementarity of the 5’ss to U1 snRNA we constructed hyper-stable U1 snRNA mutants and could thereby overcome the attenuation exerted by the secondary structure. In conclusion, MLV hides a 5’ss by the stem loop structure to reduce binding of U1 snRNP and this inefficient recognition leads to the alternative splice regulation in MLV.

Taken together, this thesis offers improved insight into crucial determinants of retroviral gene regulation and gene vector biosafety.

III Kurzzusammenfassung

Retrovirale Vektoren werden häufig verwendet um replizierende Zellen genetisch zu modifizieren und sind für die Anwendung in der Gentherapie geeignet. Die Beobachtung, dass eine durch retrovirale Vektoren ausgelöste Insertionsmutagenese zu einem klonalen Ungleichgewicht und zu bösartigen Zellveränderungen im Tiermodell sowie beim Menschen führen kann, zeigte die Notwendigkeit, sicherere Vektoren zu entwickeln. Diese Arbeit beschäftigt sich mit der Frage, ob durch die Verwendung von selbstinaktivierenden retroviralen (SIN) Vektoren mit schwachen zellulären Promotor-/Enhancer-Sequenzen die Gefahr einer Insertionsmutagenese verringert werden kann. Um das Aktivierungspotential von retroviralen Vektor-Sequenzen auf benachbarte Promotoren zu validieren, verglichen wir zelluläre und retrovirale Promotor/Enhancer in zwei verschiedenen in-vitro Assays. Ferner wurde der Einfluss des retroviralen Vektordesigns auf das Transformationspotential von Zellkulturen untersucht. Zelluläre Promotoren zeigten eine geringere Enhancer-Aktivität als virale Promotoren, wodurch das Potenzial, ein Protoonkogen wie z.B. *Evi1* zu aktivieren, deutlich reduziert ist. Dies verringert in hohem Maße die Wahrscheinlichkeit, dass es zu einer Immortalisierung primärer Knochenmarkszellen kommt. Mit dieser Arbeit konnte gezeigt werden, dass eine Aktivierung benachbarter Gene durch die hier getesteten zellulären Promotor/Enhancer unwahrscheinlich ist, wodurch ein Einsatz in therapeutischen Vektoren vielversprechend ist.

RNA-Prozessierungssignale sind wahrscheinlich wichtig für die Verbesserung der Sicherheit retroviraler Vektoren. Das Vorhandensein einer 5'Spleißstelle (ss) in retroviralen Vektoren erhöht die Expression des Transgenes und des Titers. Hingegen können Fusionstranskripte aus retroviralen und zellulären Exons durch die Verwendung der retroviralen 5'ss entstehen. Die in klinischen Studien eingesetzten gammaretroviralen Vektoren besitzen die Spleißstellen des Murinen Leukämievirus (MLV). Bisher ist alternatives Spleißen von MLV nicht im Detail verstanden, jedoch könnte die weitere Entwicklung retroviraler Vektoren von der Aufklärung des Spleißmechanismus profitieren. Wir konnten zeigen, dass MLV im Gegensatz zu anderen Retroviren, die alternatives Spleißen über die Verwendung von schwachen 3'ss regulieren, die 5'ss attenuiert. Die Attenuierung wird erreicht durch stromaufwärts liegende Sequenzen, die eine komplexe Sekundärstruktur ausbilden. Der Hauptteil der Struktur konnte funktionell durch eine heterologe Sekundärstruktur ersetzt werden. Die Sekundärstruktur reduziert die Zugänglichkeit der 5'ss zu U1 snRNP, welches normalerweise das Spleißen der RNA dirigiert. Um die Komplementarität der 5'ss zu U1 snRNA zu erhöhen, konstruierten wir hyperstabile U1 snRNA-Mutanten und konnten damit die Attenuierung der 5'ss durch die Sekundärstruktur verringern. Im Ergebnis kann festgestellt werden, dass die Sekundärstruktur die 5'ss versteckt, wodurch die Bindung von U1 snRNP reduziert wird, was dann zu alternativem Spleißen beim MLV führt. Insgesamt erlauben die Ergebnisse dieser Doktorarbeit einen verbesserten Einblick in entscheidende Aspekte der retroviralen Genregulation und in die biologische Sicherheit von Genvektoren.

Schlagworte:

Retroviren
Alternatives Spleißen
Gentherapie

Keywords:

Retroviruses
alternative splicing
Gene therapy

I List of abbreviations

ADA	Adenosine deaminase deficiency
AIDS	Acquired immunodeficiency syndrome
att	Integrase attachment site
CA	Capsid
CGD	Chronic granulomatous disease
CTD	Carboxy-terminal domain
CTE	Constitutive transport element
DNA	Deoxyribonucleic acid
DSB	DNA double-stranded break
dsRed	Red fluorescent protein
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
FeLV	Feline leukemia virus
GaLV	Gibbon ape leukemia virus
GFP	Green fluorescent protein
G-SCF	Granulocyte-colony stimulation factor
HFV	Human foamy virus
HIV-1	Human immunodeficiency virus 1
hnRNP	Heterogeneous nuclear ribonucleoproteins
HR	Homologous recombination
HSC	Hematopoietic stem cell
HTLV-1	Human T-cell leukemia virus type 1
IP	Internal promoter
IRES	internal ribosomal entry site
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
KoRV	Koala retrovirus
LTR	Long terminal repeats
MA	Matrix
MESV	Murine embryonic stem cell virus
MLV	Murine leukemia virus
MoMLV	Moloney murine leukemia virus
MPMV	Mason-Pfizer monkey virus
MPSV	Myeloproliferative sarcoma virus
mRNA	Messenger RNA
NC	Nucleocapsid
NPC	Nuclear pore complex
NRS	Negative regulator of splicing
ORF	Open reading frame
pA	Poly(A) signal
PAP	Poly(A) polymerase
PBS	Primer binding site
PIC	Preintegration complex
PPT	Polypurine tract
PR	Protease
PRE	Posttranscriptional regulatory element
R region	Redundant region
RCR	Replication competent retrovirus
rER	Rough endoplasmic reticulum
RMT	Retroviral particle-mediated mRNA transfer
RNA	Ribonucleic acid
RNAP II	RNA polymerase II

RNP	Ribonucleoprotein
RRE	Rev responsive element
RSV	Rous sarcoma virus
RT	Reverse transcriptase
SFFV	Spleen focus forming virus
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein
SR protein	Serine/Arginine-rich protein
ss	Splice site
ss DNA	Single-stranded DNA
SU	Surface
TM	Transmembrane
tRNA	Transfer ribonucleic acid
USE	Upstream sequence enhancer
UTR	Untranslated region
VSV-G	Glycoprotein of vesicular stomatitis virus
WDSV	Walleye dermal sarcoma virus
Wt	Wild type
XMRV	Xenotropic MuLV-related virus
X-SCID	X-linked severe combined immune deficiencies
ZFN	Zinc finger nucleases
ψ	Retroviral packaging signal

Amino Acids

A	Ala	Alanine	M	Met	Methionine
C	Cys	Cysteine	N	Asn	Asparagine
D	Asp	Aspartic acid	P	Pro	Proline
E	Glu	Glutamic acid	Q	Gln	Glutamine
F	Phe	Phenylalanine	R	Arg	Arginine
G	Gly	Glycine	S	Ser	Serine
H	His	Histidine	T	Thr	Threonine
I	Ile	Isoleucine	V	Val	Valine
K	Lys	Lysine	W	Trp	Tryptophan
L	Leu	Leucine	Y	Tyr	Tyrosine

A Introduction

In the last 20 years the understanding and knowledge of the complex interplay of genetic information and cell fate has increased rapidly. One milestone was the sequencing of the human genome, finished at the beginning of the new millennium (Lander *et al.*, 2001). These sequence data enabled scientists to discover the genetic background of many diseases. Further advances in genetic, diagnostic and molecular biology resulted in new techniques to cure diseases by directly modifying the mutated gene. Diseases like cystic fibrosis, hemophilia, and β-thalassemia are predicated by a mutation of a single gene. In an autosomal recessive disorder like cystic fibrosis or adenosine deaminase deficiency (ADA) both alleles are mutated leading to the specific phenotype. Heterozygous individuals however, are often healthy due to compensation by the non-mutated allele. In the case of autosomal dominant disorders one allele is sufficient to show a phenotype. In addition, there are X chromosome-linked genetic diseases resulting in a dominant phenotype in males. Examples are the X-linked severe combined immune deficiencies (X-SCID) and the chronic granulomatous disease (CGD). Monogenetic diseases are potential candidates for gene therapy.

1 Gene therapy

Gene therapy is defined as the insertion of genes into the host cell genome in order to diagnose, prevent or to cure diseases (Mulligan, 1993). In general, somatic cells are used for genetic modifications, where the hematopoietic system, due to its regenerative capacity, is an ideal target for gene therapy (Fig.1). For a clinical approach hematopoietic stem cells (HSC) are isolated via a bone marrow biopsy or mobilization with G-CSF (granulocyte-colony stimulating factor). The genetic alteration is performed *ex vivo* and the modified cells are returned to the patient. Currently the most successful type of gene therapy to cure inherited disorders is gene addition (Aiuti *et al.*, 2002; Cavazzana-Calvo *et al.*, 2000; Gaspar *et al.*, 2004). Here, a correct copy of the disease-causing gene leads to correction of the phenotype at the cellular level.

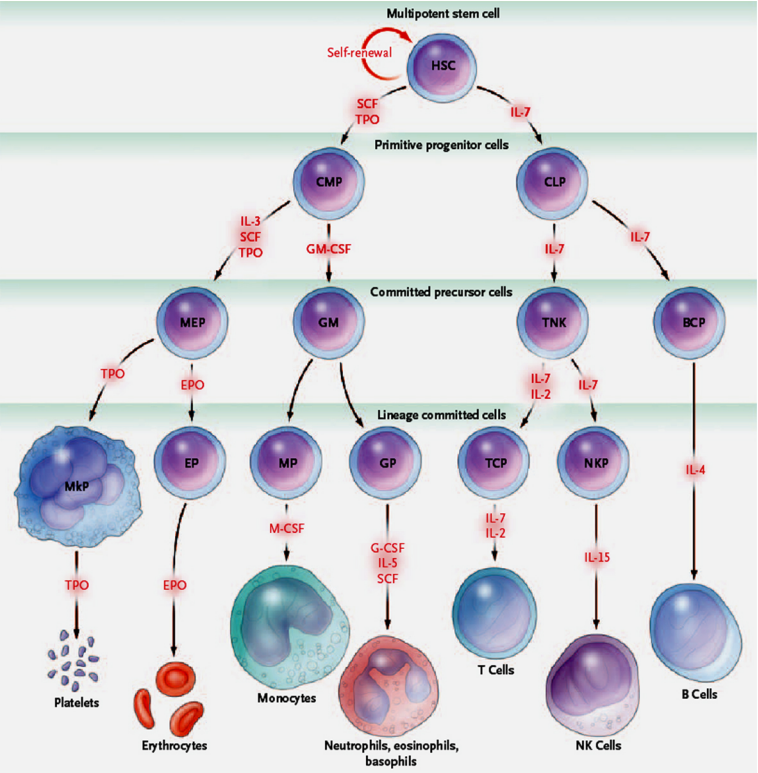


Figure 1: General model of hematopoiesis

HSC are the source of all blood cell development and are regulated by several factors. HSC have the potential to self-renew or to differentiate into blood progenitors. These processes are mediated by several cytokines. Abbreviations: HSC hematopoietic stem cell, CMP common myeloid progenitor, CLP common lymphoid progenitor, MEP megakaryocyte erythroid progenitor, GM granulocyte progenitor, TNK T-cells and natural killer cells, BCP B-cell progenitor, MkP megakaryocyte progenitor, EP erythroid progenitor, MP monocyte progenitor, GP granulocyte progenitor, TCP T-cell progenitor, NKP natural killer cell progenitor, IL interleukin, TPO thrombopoietin, SCF stem cell factor, M-CSF macrophage-colony stimulating factor, GM-CSF granulocytemacrophage CSF, EPO erythropoietin (Kaushansky, 2006).

Different techniques such as physicochemical gene transfer systems (Glover *et al.*, 2005) have been established to transport DNA into the nucleus of the target cell. Disadvantages of such approaches are limited efficacy, high cytotoxicity and a low frequency of stable integration. Viruses have been used as naturally occuring “gene-shuttle” systems. For clinical approaches so called viral based vector systems have been established on the basis of adenovirus, herpes virus, adeno-associated virus, poxvirus, vaccinia virus and retroviruses (Thomas *et al.*, 2003). Thereby, the therapeutic transgene is embedded into a full transcription unit, including promoter/enhancer and RNA processing elements.

Many vectors used in clinical gene therapy trials are based on retroviruses (~22%), which have the ability to catalyze the stable integration of their genome into the host cell chromosome (overview in http://www.wiley.co.uk/genetherapy/clinical/). The inserted DNA becomes part of the cellular genome and thus retroviral vectors offer the advantage of long-term gene expression. Furthermore, retroviral vectors have relatively simple genomes, are easy to use, show a broad cell tropism and a sufficient packaging capacity for the delivery of many therapeutic genes (Coffin *et al.*, 1997). Therefore retrovirus based vectors are a promising tool for gene therapy and have thus been optimized and investigated in the present study.

2 Retroviruses

The first retrovirus was identified as a tumour causing agent in chicken in 1911 by P. Rous (Rous sarcoma virus). D.Baltimore, H.M. Temin and S. Mituzami discovered in 1970 that retroviruses encode a specific enzyme, which causes a retrograde information flow from RNA to DNA (Coffin *et al.*, 1997). During their life cycle retroviruses reverse transcribe their RNA genome into double-stranded DNA and integrate as a "provirus" into the host genome. This feature gave retroviruses their name and the viral reverse transcriptase became an important tool for molecular biologists to study gene expression. Retroviruses belong to one of the best charaterized virus families, due to the discovery of the human immunodeficiency virus type 1 (HIV-1) in 1983 by L. Montagnier as the agent of the acquired immunodeficiency syndrome (AIDS) (Barre-Sinoussi *et al.*, 1983; Lever & Berkhout, 2008; Pincock, 2008). Retroviruses preferentially infect vertebrates, where they can induce multiple symptoms such as tumour diseases, immunodeficiencies or neurologic deficits, depending on the functional properties of the retroviral genome. In general, the family of the Retroviridae can be divided into exogenous retroviruses, where the infection occurs via viral particles, and endogenous retroviruses, which use vertical germline transfer. Based on their genome organization, structural morphology and their life cycle, retroviruses are subdivided into the following 7 genera (Modrow *et al.*, 2003).

- α-retrovirus (e.g. Rous sarcoma virus, RSV)
- β-retrovirus (e.g. Mason-Pfizer monkey virus, MPMV)
- γ-retrovirus (e.g. Murine leukemia virus, MLV)
- δ-retrovirus (e.g. Human T-cell leukemia virus type 1, HTLV-1)
- ε-retrovirus (Fish retroviruses, e.g. Walleye dermal sarcoma virus, WDSV)
- Lentivirus (e.g. Human immunodeficiency virus type 1, HIV-1)
- Spumavirus (e.g. Human foamy virus, HFV)

Depending on the presence of additional gene products encoded by the retroviral genome, retroviridae can be further classified into simple and complex retroviruses. Simple retroviruses (alpha, beta- and gammaretroviruses) only encode for three viral gene products (structural, enzymatic and envelope proteins) common to all

retroviruses. In contrast, complex retroviruses (delta-, lenti- and spumaviruses) harbor additional open reading frames (ORF) for accessory regulatory proteins (Coffin *et al.*, 1997). Since the aim of this PhD thesis was to improve gammaretrovirus derived gene transfer vectors, the following sections will focus on the viral life cycle, viral gene expression and the vector development of simple retroviruses.

2.1 The gammaretroviral genome

Retroviruses are enveloped, single-stranded and positive-sense RNA viruses (7-11kb) that carry two copies of their genome per virion (Coffin *et al.*, 1997). The viral RNA genome uses the cellular transcription machinery for replication and is similar to cellular messenger RNA (mRNA); exhibiting a 5'Cap and a 3'Poly(A) tail. All simple retroviruses are based on the same genomic organization as shown in Figure 2. A redundant region (R) is located at both termini of the viral RNA genome and includes the transcription start site and a weak polyadenylation signal. The 5'R region is followed by the U5 (U=unique) region, the primer binding site (PBS), the 5'splice site (5'ss) and the highly structured packaging signal (ψ) (D'Souza & Summers, 2005). Furthermore, the U5 region harbors one of the two attachment sites (att) whereas the PBS binds a cellular tRNA, which is used by the reverse transcriptase as a primer. The following coding region encodes three consecutive genes common for all retroviruses. The structural proteins (gag), the enzymes (pol; reverse transcriptase, integrase and protease) and the glycoprotein (env) that mediates viral entry. Whereas gag and pol are expressed from the same viral unspliced (genomic) RNA, the expression of env is directed from a spliced RNA using a 3'splice site (3'ss) located in pol. A polypurine tract (PPT) of at least 9 purines is located downstream of the open reading frames (ORFs) and is important for reverse transcription. The viral promoter/enhancer sequences and the second attachment site are located in the U3 region at the 3'end of the genomic RNA followed by the terminal 3'R region. During the viral life cycle the retroviral RNA genome is reverse transcribed into double-stranded DNA and integrated as a provirus into the host genome. In the process of reverse transcription the regulatory sequences are duplicated at both ends of the viral genome to create the long terminal repeats (LTR). The LTR include the U3, R and the U5 region in the same order at the 5' and at the 3'end of the DNA.

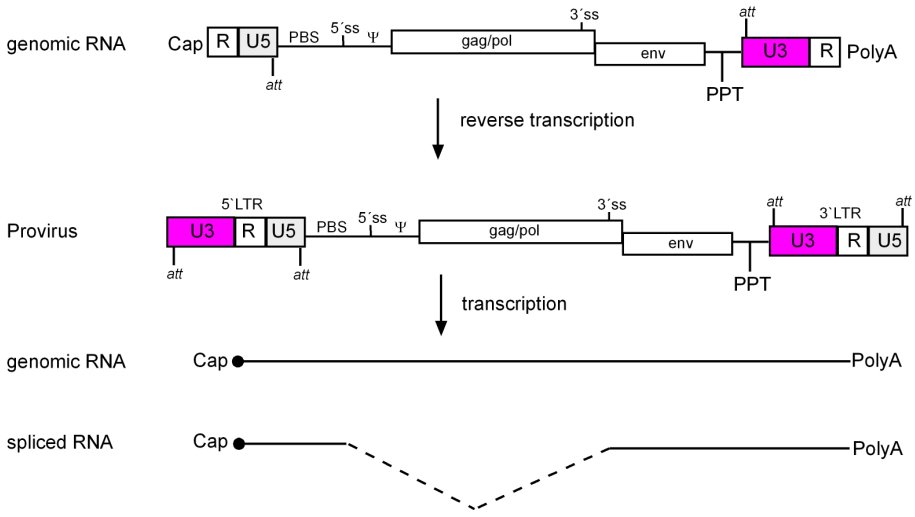


Figure 2: The gammretroviral RNA genome

The figure schematically illustrates the genomic RNA, the provirus and the transcribed RNA species. The retroviral genome harbors, like celluar mRNAs, a 5'Cap and a 3'Poly(A) tail. The 5'UTR contains the 5'R region, U5, PBS, the 5'ss and the packaging signal, followed by the open reading frames for gag/pol and env. The 3'UTR contains the viral enhancer promoter elements (U3), which are copied after reverse transcription to the 5'end of the integrated provirus, and the 3'R region. After reverse transcription, the attachment sites for the integrase are located in U5 and U3. The provirus is flanked by the long terminal repeats (LTR), each consisting of the U3, R and U5 regions. Transcription of the viral RNAs is carried out by cellular transcription machineries and is directed by the promoter located in U3. The 3'ss is located in pol and expression of env is dependent on a single splice event removing the "gag/pol intron".

2.2 The gammaretroviral life cycle

The whole retroviral RNA genome is packaged into a ribonucleoproteincomplex (RNP) consisting of the nucleocapsid and the reverse transcriptase and integrase proteins. Capsid proteins give the core its icosahedral structure. The outer shell of the virus is derived from the plasma membrane and contains viral envelope proteins like the surface (SU) and the transmembrane (TM) glycoprotein. Retroviruses use cell-surface molecules as specific receptors to enter their target cells by fusing with the host cell membrane, either directly or after endocytosis. The viral core is than released into the cytoplasm by membrane fusion (Coffin *et al.*, 1997). Subsequently, the viral core undergoes an uncoating process through which the capsid is removed and the viral reverse transcriptase retrogradely translate the RNA genome into DNA (Baltimore, 1992; Coffin *et al.*, 1997). Reverse transcription starts at the free 3'hydroxyl group of the cellular tRNA molecule that is linked via hydrogen bonds to the PBS of the 5'untranslated region (5'UTR). The reverse transcriptase (RT) uses

the genomic RNA as a template for the minus strand synthesis and the emerging DNA is used for the plus strand elongation (Fig.3). During the reverse transcription process RNA/DNA hybrids are formed and cleaved by the RNase H activity of the RT. So the RT exhibits two enzymatic properties, DNA polymerase and RNase H activity. Compared to cellular DNA polymerase, the RT offers no proofreading activity, which leads to the high mutation rate observed in retroviruses (Goodenow *et al.*, 1989; Preston *et al.*, 1988). During reverse transcription the regulatory sequences are duplicated at both ends of the viral genome to create the long terminal repeats (LTR).

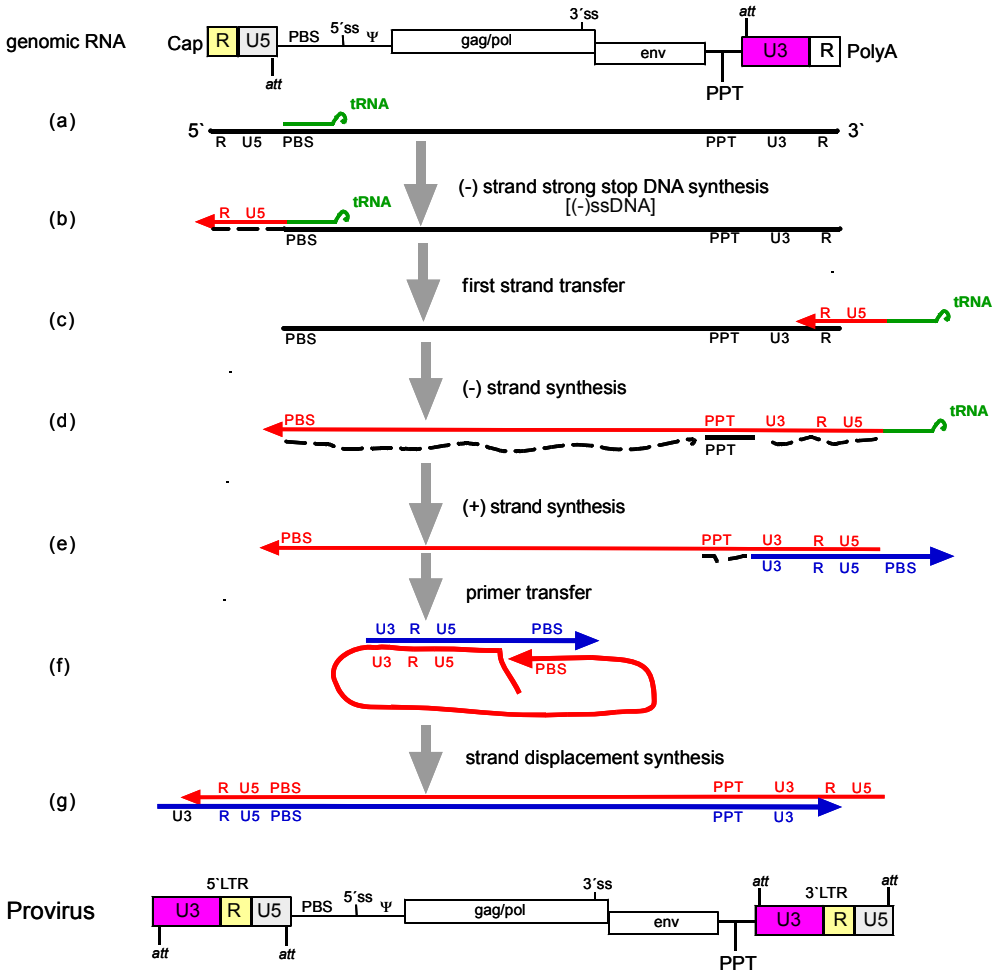


Figure 3: Reverse transcription

(a) Reverse transcription is initiated from the 3'end of a primer tRNA (green) partially annealed to the PBS region of the retroviral genomic RNA (black). (b) The reverse transcriptase extends through the 5'end and generates the minus-stranded strong stop DNA (red). The 5'end of the genomic RNA is degraded by RNase H. (c) strand transfer of the (-)ssDNA to the 3'R region of the retroviral genome. (d) Continued minus strand DNA synthesis occurs and RNase H degrades the RNA within the DNA/RNA hybrid except for the PPT sequences. (e) The PPT serves as a primer for plus strand synthesis (blue). Thereby the first 18 nucleotides of the tRNA primer are used for the generation of the PBS DNA sequence and degraded subsequently by RNase H. (f) The newly synthesized plus strand DNA is transferred to 5'end of the minus DNA strand. Base pairing of plus and minus DNA PBS sequences leads to a loop-like structure. (g) DNA synthesis resumes after second strand transfer with each strand using the other as template. Adapted and modified from (Basu *et al.*, 2008).

The stable inserted provirus is a functional transcription unit flanked by the viral LTRs that is transcribed by the cellular RNA polymerase II, partially spliced, polyadenylated and exported to the cytoplasm. The glycoprotein (env) is expressed as a precursor molecule from the spliced RNA (Fig.2) and synthesized at the rER. In the Golgi apparatus env is cleaved by the cellular furin protease to produce the SU and the TM glycoproteins during the transport via the secretory pathway to the plasma membrane of the cell. In contrast, gag and gag/pol are synthesized at free polyribosomes and expressed from the unspliced genomic RNA. The translation of the gag/pol polyprotein of Moloney murine leukemia virus (MoMLV) requires a process called “translational suppression of termination”. Here an amber stop codon (UAG) of the gag reading frame is occasionally misread by the tRNA for glutamine. More precisely, a purine rich sequence 3’ of the stop codon and an RNA pseudoknot structure further downstream are suggested to be responsible for the competition between the glutamine tRNA and translation termination/release factors by causing the ribosome to pause. Another strategy to express two proteins from the same mRNA template has evolved in the rous sarcoma virus (RSV) which uses ribosomal frameshifting. Here the ribosome moves into a different reading frame in the -1 direction before reaching the gag stop codon and than continues the translation of pol. Gag/pol precursor proteins are produced by misreading the stop codon in about 4 to 10% and by ribosomal frameshifting in about 5% of translation events (Flint, *et al* 2000). After synthesis of the viral proteins the assembly and budding of new virions occur at the plasma membrane of the cell. The gag and gag/pol precursor proteins are packaged into viral particles and cleaved into their single components by the viral protease. Gag consists of the matrix (MA, p15), p12, capsid (CA, p30), and nucleocapsid (NC, p10) proteins, whereas pol is processed into the protease (PR, p14), the reverse transcriptase (RT, p80) and the integrase (p46). The packaging of the genomic RNA is mediated by an interaction of the packaging signal ψ with the nucleocapsid domain of gag. Finally the maturation of infectious particles containing an activate env protein for fusion occurs by processing of TM proteins by the viral protease during virus budding to generate $p12^{env}$ and $p2^{env}$ molecules (Fig.4).

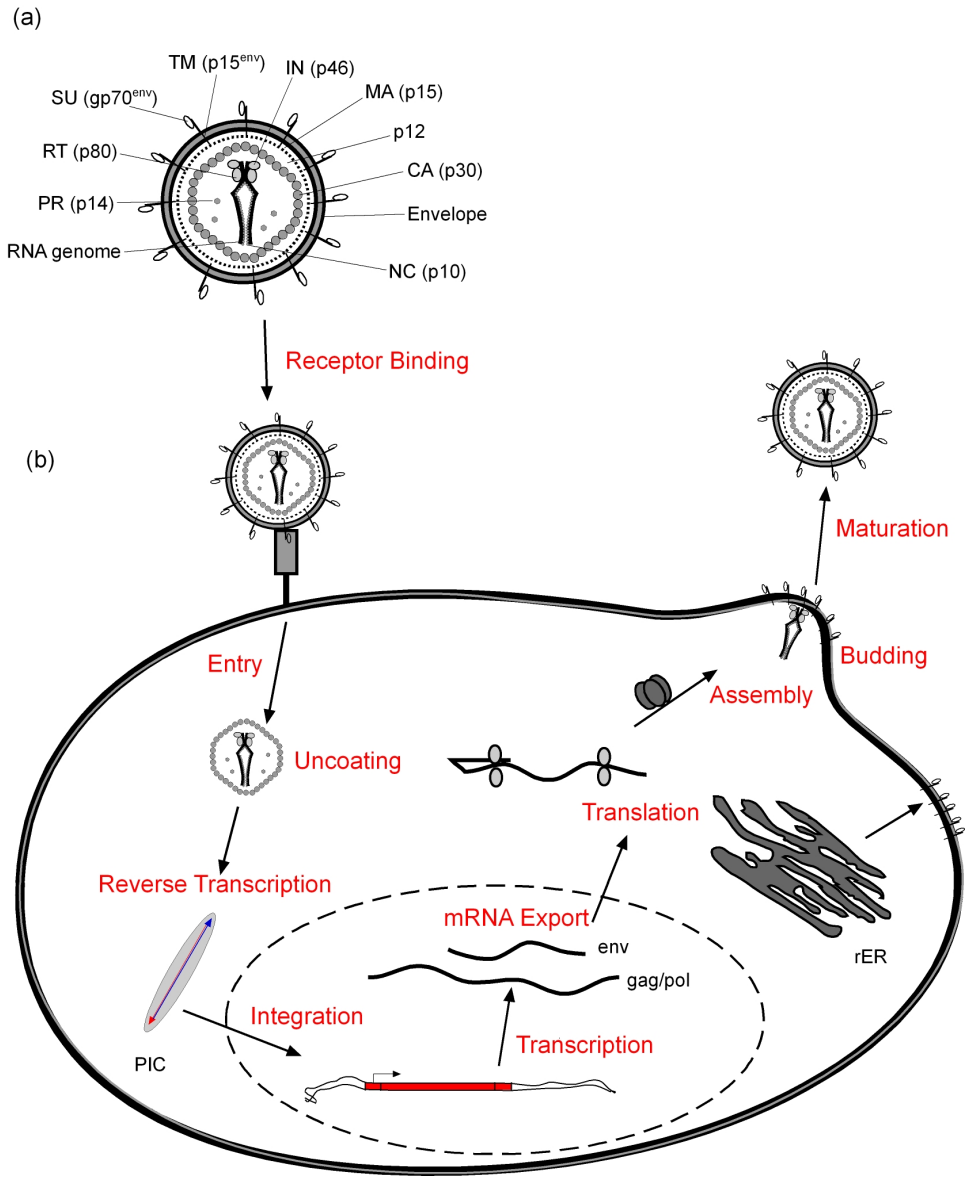


Figure 4: The gammaretroviral life cycle

(a) Highly schematic illustration of a mature MLV particle. All retroviral components and their relative locations within the particle are indicated. IN: integrase, MA: matrix, CA: capsid, NC: nucleocapsid, PR: protease, RT: reverse transcriptase, SU: surface subunit, TM: transmembrane subunit. Adapted from (Galla, 2008).
(b) The retroviral virion enters the cytoplasm in a receptor-mediated manner by fusing its envelope with the host cell membrane (entry). After uncoating, the reverse transcription of the viral genome is initiated and the resulting double-stranded DNA gains access to the host´s chromatin during mitosis. The viral integrase inserts the viral provirus into the host genome. Using the cellular transcription machinery the viral genome is transcribed, spliced in a defined ratio and the RNAs are exported to the cytoplasm. The gag and gag/pol precursor proteins are translated at free polyribosomes whereas the env proteins are synthesized at the rough endoplasmic reticulum (rER). The proteins are transported to the cell membrane where the assembly of the retroviral components and the budding process occurs. The maturation of infectious particles takes place during or shortly after budding (maturation).

2.3 Eukaryotic gene expression

Gene expression by simple retroviruses is carried out by cellular machineries and is regulated at many steps during transcription, splicing, polyadenylation, export and translation of the mRNA. In contrast to prokaryotes, transcription and translation in eukaryotes are separated into two distinct cellular compartments i.e. the nucleus and the cytoplasm. Transcription of all protein-coding genes and most of the small nuclear RNAs (snRNAs) is catalyzed by RNA polymerase II (RNAP II). The expression level of most genes is regulated by transcription factors that bind to DNA regulatory sequences (enhancer) up to 10 kb upstream of the TATA-box where transcription is initiated. The accessibility of the gene locus to the transcriptional apparatus plays a pivotal role in regulating gene expression. Untranscribed regions of the genome are packaged into highly condensed “heterochromatin”. In contrast, transcribed genes are present in the more accessible “euchromatin” (Richards & Elgin, 2002). To induce transcription, activator proteins decondense repressive chromatin structures. Therefore transcription factors need co-activators/mediators like chromatin-modifying enzymes which facilitate the access of DNA-binding proteins to DNA and the recruitment of the RNA polymerase enzymes to the transcriptional start site (McKenna & O'Malley, 2002; Narlikar *et al.*, 2002). For publication 1 it was important to understand the underlying mechanisms of promoter interactions and the ability of enhancers to activate neighboring genes over a long distance.

The carboxy-terminal domain (CTD) of the RNAP II is crucial for coupling transcription to pre-mRNA processing. The CTD consists of 27 heptad repeats in yeast and 52 in humans, and the phosphorylation status of these repeats during transcription coordinates the recruitment of processing factors crucial for capping, splicing and polyadenylation (Dahmus, 1996; Komarnitsky *et al.*, 2000). After initiation of transcription, capping of nascent transcripts is the first step of pre-mRNA proccessing and it occurs as soon as the 5’end emerges from the polymerase (Fig.5) (Hirose & Manley, 2000; Proudfoot *et al.*, 2002; Shatkin & Manley, 2000).

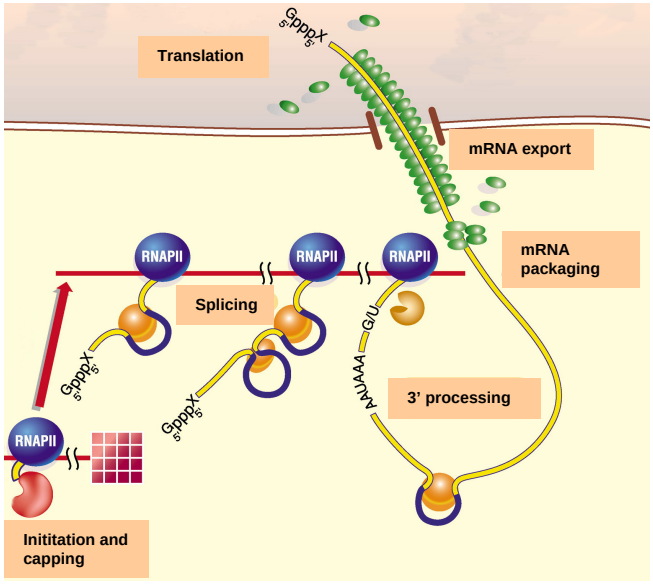


Figure 5: A contemporary view of gene expression

The figure shows co-transcriptional pre-mRNA processing. Capping of the pre-mRNA usually initiates after synthesizing 20-30 nucleotides of RNA. Splicing of introns occurs co- and post-transcriptionally, whereas the pre-mRNA must be cleaved before addition of around 200 adenosine residues to the 3'end takes place. During transcription the RNA is packaged as a ribonucleo-proteincomplex and exported through nuclear pores to the cytoplasm. (Orphanides & Reinberg, 2002)

The cap-binding complex including the capping enzymes is recruited via the CTD to the 5'end of the pre-mRNA. This 5'cap modification renders pre-mRNA and mRNA resistant to the activity of 5' to 3'exonucleases (Beelman & Parker, 1995). During the switch of initiation to elongation, the CTD becomes hyperphosphorylated on the first two serine residues in the heptad YSPTSPS. Normally mammalian genes consists of coding (exon) and non-coding (intron) sequences. In order to generate a functional message from the DNA template the introns are removed by mRNA splicing (Berget *et al.*, 1977). Thereby the exons are ligated together by a two-step transesterification reaction performed by the spliceosome. Since one important aim of the present study was to investigate how MLV regulates alternative splicing, the following paragraphs will focus on RNA splicing mechanisms and alternative splicing in more detail (see 2.4 et seq). Transcription termination and release of RNAP II depends on the recognition of a functional poly(A) signal consisting of a highly conserved AAUAAA hexamer and cleavage at the next CA dinucleotide downstream of the poly(A) signal. Cleavage is followed by the addition of around 200 adenosine residues by the poly(A) polymerase (PAP) (Wahle & Ruegsegger, 1999). This poly(A) tail stabilizes the RNA and facilitates downstream events like export and translation. After successful pre-mRNA proccessing the mRNA must be exported through the nuclear pore complexes (NPCs) (Allen *et al.*, 2000; Ryan & Wente, 2000) to the cytoplasm, where translation of the mRNA is carried out by the cellular ribosomes. In addition, the NPC has a quality control function in that intron-containing mRNAs are retained to prevent faulty translation due to the presence of premature stop codons. Export of bulk mRNA is thought to be facilitated by the general export factor TAP/NXF1:p15

and different adapter molecules like the REF/Aly and SR proteins (Dimaano & Ullman, 2004). All non-coding RNAs like tRNAs, rRNAs and U snRNAs are exported by members of the karyopherin family of Ran-dependent nucleoplasmic transport factors, also termed importins/exportins. For this to occur, nuclear export factors of the karyopherin class such as Crm1 or Exportin t bind their cargo in the presence of Ran-GTP in the nucleus and then release it via gradient Ran-GTP translocation to the cytoplasm (Allen *et al.*, 2000; Cullen, 2003; Gorlich & Kutay, 1999).

2.4 Splicing

Most eukaryotic genes are interrupted by non-coding sequences (introns) and are transcribed as pre-mRNAs that are then converted to mRNA by splicing. The introns are removed and the coding sequences (exons) are ligated together. This process is catalyzed by the spliceosome, a dynamic 60S ribonucleoprotein particle (Staley & Guthrie, 1998), assembled from five subcomplexes composed of five small nuclear ribonucleoproteins (U1, U2, U4, U5 and U6 snRNP) and as many as 150 other proteins (Jurica & Moore, 2003; Will & Luhrmann, 2001; Zhou *et al.*, 2002). Each exon/intron border exhibits consensus elements in *cis,* which are essential for the recognition by the spliceosome. Four short sequences define an intron. The 5'end of the intron, termed 5'splice site (5'ss), is marked by the consensus sequence AGI***GU***RAGU, whereas the 3'splice site (3'ss) defines the 3'end of the intron by Y_n-*Y****AG***IG (R, purine; Y, pyrimidine) (Gesteland, *et al* 2006). Another consensus sequence CUR**A**CUA (branchpoint) lies about 40 nt upstream of the 3'ss and harbors a highly conserved adenosine, which is important for the first transesterification reaction (Konarska & Query, 2005). Additionally, a pyrimidine-rich tract containing a run of about 13 pyrimidines is located between the branchpoint and the 3'ss. The branchpoint, the pyrimidine tract and the invariant AG dinucleotide define a 3'ss. The spliceosome assembles and rearranges itself in a highly ordered stepwise manner. Spliceosomal assembly proceeds through the E, A, B and C complex (Fig.6). Initial recognition starts with the binding of U1 snRNP to the 5'ss resulting in the E complex. The 5'ss is recognized through an RNA base pairing interaction with the consensus sequence of the 5'end of U1 snRNA. In parallel, a dimeric splicing factor U2AF binds the 3'ss and the adjacent pyrimidine tract which helps to recruit U2 snRNP to the branchpoint. The base pairing of the U2 snRNA to the branchpoint is the first ATP-dependent step, leading to the A complex. The addition of the

U4/U6·U5 triple snRNP to the spliceosome during B complex formation heralds a large number of RNA:RNA rearrangements (Nilsen, 1994), leading to the release of U4 snRNP and the replacement of U1 snRNP with U6 snRNP. As a result, the catalytically active C complex is formed by intramolecular base pairing between the U2, U6 and the pre-mRNA (Madhani & Guthrie, 1992; Sun & Manley, 1995). Base pairing between the U2 snRNA and the branch site sequence forms a duplex with a bulged adenosine, which engages in 2'OH linkage after attack of the guanosine at the 5'end of the intron. Crosslinking experiments showed that the U5 snRNP can interact with both 5' and 3'exon sequences (Newman, 1997) and is thought to stabilize the reactive complex and to align both exons for the second catalytic step. Here the 3'OH of the 5'exon attacks the intronic 3'guanosine and exon-ligation occurs. The final products are the intron-less mRNA and a lariat-like structure formed of U6 and the removed branchpoint oligo (Kramer, 1996).

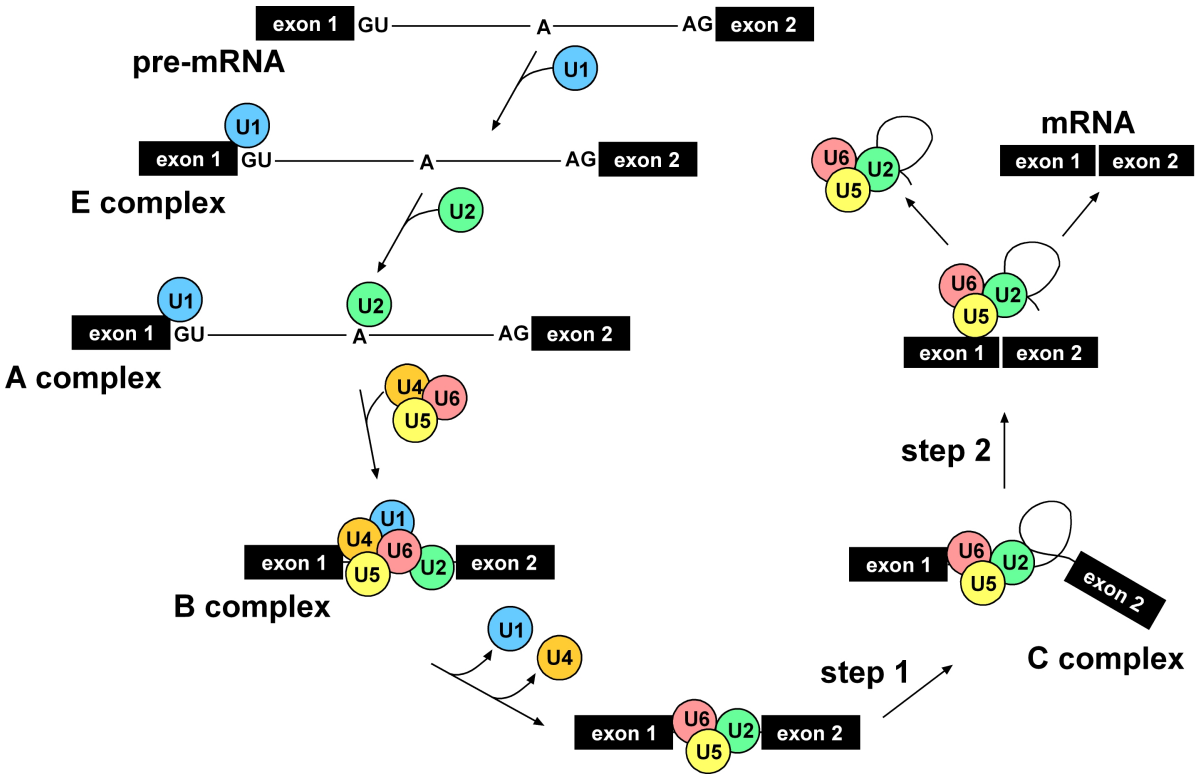


Figure 6: Assembly of the spliceosome and the splicing reaction

Stepwise assembly of the spliceosome with the complexes E, A, B, C. Step 1 and step 2 are both transesterification reactions. For more details see section 2.3.
Adapted from (Gesteland, *et al* 2006).

Pre-mRNA splicing begins co-transcriptionally and continues post-transcriptionally as observed for example in *Drosophila* (Beyer & Osheim, 1988), and humans (Tennyson *et al.*, 1995). Specifically, it has been shown for the Balbiani ring genes of *Chironomous tentans* that nascent mRNAs already lack introns at their 5'ends (Bauren & Wieslander, 1994; Wetterberg *et al.*, 1996). The time taken for RNAP II to reach the end of a transcription unit defines the maximal time in which splicing can occur co-transcriptionally. *Drosophila* RNAP II has a transcription rate of 1-1.5 kb/minute, leading to a ~3 minute transcription time for introns (with an average of 3.300 bp) and only ~30 seconds for exons (with an average of ~300 bp). Intron removal in *Drosophila* has been observed ~4 minutes after 3'ss recognition. In mammalian systems splicing occurs both co- and post-transcriptionally in general with a 5' to 3'direction. Consequently, terminal intron splicing often occurs post-transcriptionally and is linked to transcriptional termination and polyadenylation of the mRNA (Neugebauer, 2002; Rigo & Martinson, 2008).

2.5 Alternative splicing

Sequencing of the human genome yielded an unexpectedly low number of genes (~26,000; Lander *et al.*, 2001; Maniatis & Tasic, 2002) from which ca. 90,000 functional proteins (Woodley & Valcarcel, 2002) are expressed. Splicing can either be constitutive or alternative. Constitutive splicing describes a situation in which all exons are always included in a given mRNA, whereas in alternative splicing, different exon combinations can be used, resulting in several possible products. Alternative splicing is important in cellular differentiation (and gene expression) and aberrant splicing is involved in many human diseases (Hastings & Krainer, 2001; Stoilov *et al.*, 2002). In fact, alternative splicing is the rule and not the exception. Depending on the genome database and microarray data used, estimates range from 50 to 75% of human genes that are alternatively spliced (Johnson *et al.*, 2003; Modrek *et al.*, 2001). Thus, alternative splicing expands the proteome, resulting in a high number of protein isoforms encoded by a limited number of genes (Goldstrohm *et al.*, 2001). In alternative splicing, splice site choice is thought to be regulated by the presence of additional non splice site regulatory elements within the exon or intron. Exonic or intronic RNA sequences that act positively to stimulate spliceosome recruitment are termed exonic/intronic splicing enhancers (ESE/ISE). In contrast, other RNA

elements that block spliceosome assembly are called exonic/intronic splicing silencers (ESS/ISS) (Black, 2003). Many of these elements are bound by RNA binding proteins like SR proteins or hnRNP proteins. SR proteins exhibit one or two N-terminal RNA recognition motifs and a variable-length C-terminal arginine/serine-rich (RS) domain required for protein-protein interaction with other RS domains. Representative members of the SR protein family are ASF/SF2, SC35 and SRp55, which bind for example to exon sequences, where they enhance splicing of the adjacent intron (Graveley, 2000). Spliceosome recruitment can also be negatively influenced; for example by hnRNP binding to splicing silencer motifs. All hnRNPs share a common structure containing a RNA binding domain and a glycine-rich auxiliary domain that might mediate protein-protein interaction or facilitate protein localization (Krecic & Swanson, 1999). HnRNPs could interfere directly with the assembly of spliceosomal components by blocking recognition sites of the pre-mRNA, or they could block splicing activation by SR proteins binding to adjacent ESEs. In addition to this complexity, it has also been shown that pre-mRNA secondary structures can influence splicing (Buratti & Baralle, 2004). The most common mechanism is represented by secondary structures that hinder the accessibility of splicing factors to the pre-mRNA.

In general there are five major forms of alternative splicing (Fig. 7). Exon skipping is the most frequent event, which accounts for 38% of all alternative splicing events, whereas intron retention is responsible for less than 3% and is mainly observed in retroviruses. Furthermore, alternative usage of 5'ss or 3'ss (18% and 8%, respectively) can lead to either extended or shortened exons. Additionally, more complex events, such as mutually exclusive events, alternative transcription start sites and multiple polyadenylation sites account for the remaining 33% of the alternative splicing or processing events (Ast, 2004).

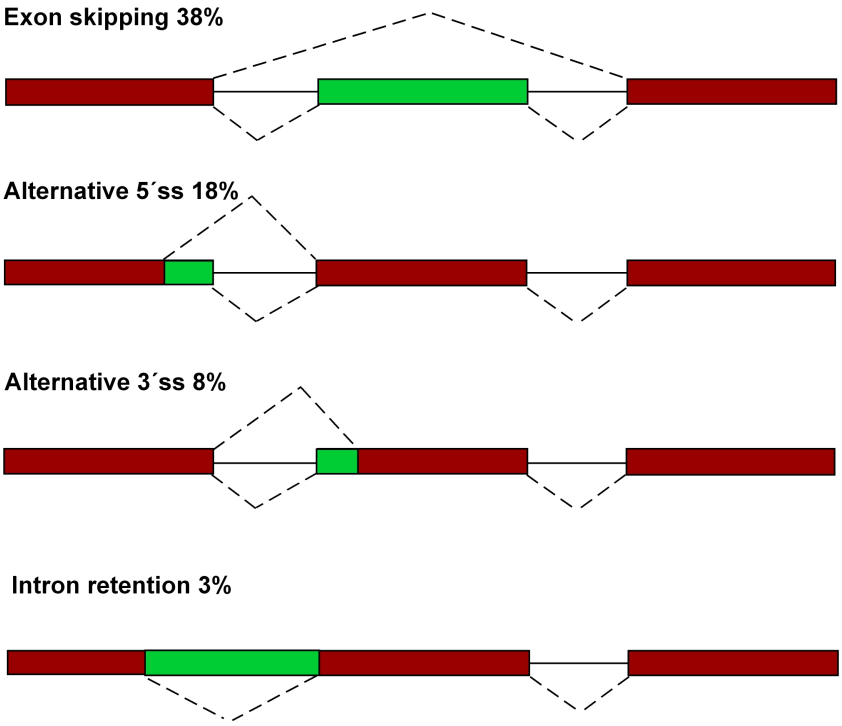


Figure 7: Types of alternative splicing

Constitutive exons are shown in red and alternatively spliced regions in green, whereas dashed lines indicate the splicing events. Relative amounts of alternative splicing events are shown as percentage. Adapted from (Ast, 2004).

2.6 Problems of retroviral gene expression

Since viral gene expression is dependent on the cellular spliceosome, elucidation of the regulatory mechanisms of viral splicing has provided deep insights into the general regulation of mammalian RNA processing. Retroviruses have evolved strategies to express proteins from unspliced, incompletely spliced and fully spliced transcripts (Cochrane *et al.*, 2006). As mentioned above, the best characterized system displaying intron retention are retroviruses, which harbor one or more introns in their genomic mRNA. For simple retroviruses like MLV, a single splice event is required to express the viral envelope protein (Boris-Lawrie *et al.*, 2001; Cullen, 1998). The unspliced RNA is needed for the expression of gag and gag/pol proteins and for packaging of genomic RNA into viral particles. So far the splicing regulation of the MoMLV is poorly understood, but several *cis*-acting elements have been proposed to be involved in viral splicing control (Hwang *et al.*, 1984; Logg *et al.*, 2007). In contrast, more complex retroviruses like HIV-1 encode for 6 additional regulatory proteins by processing the 9 kb transcript into over 30 different

mRNAs through alternative splicing (Purcell & Martin, 1993; Schwartz *et al.*, 1990). Splicing of the different transcripts is achieved through the combinatorial use of five 5'splice sites and nine 3'splice sites. It was shown that the 5'ss are highly active but the 3'ss are weak due to degenerated consensus sequences of the branchpoint and the pyrimidine tract in conjunction with ESEs and ESSs (Amendt *et al.*, 1995; Caputi *et al.*, 1999; Kammler *et al.*, 2006; O'Reilly *et al.*, 1995; Si *et al.*, 1997; Staffa & Cochrane, 1994; Tange *et al.*, 2001). RSV evolved another strategy, whereby the 5'ss competes with a decoy 5'ss for efficient recognition by the spliceosome. RSV harbors a negative regulator of splicing (NRS) near the actual 5'ss, which functions as an elaborate pseudo 5'ss that binds U1 snRNP and initiates the assembly of a non-productive splicing complex with the viral 3'ss.

During the viral life cycle, incompletely spliced RNAs must be exported to the cytoplasm. This is normally restricted through quality control mechanisms at the nuclear pore by preventing the export of intron containing mRNAs. To avoid this, retroviruses have evolved distinct strategies to export their incompletely spliced mRNAs. For example, HIV-1 encodes the viral accessory protein Rev (Fischer *et al.*, 1995; Malim *et al.*, 1988; Pollard & Malim, 1998) that binds to a highly conserved RNA stem loop structure located in the env gene, termed Rev responsive element (RRE). Rev functions as an adapter molecule for Crm1 (Fornerod *et al.*,1997; Ossareh-Nazari *et al.*, 1997; Stade *et al.*, 1997), shuttling genomic and incompletely spliced mRNAs through the nucleus by using the Crm1 protein export pathway. Simple retroviruses like the mason-pfizer monkey virus (MPMV) contain a *cis*-acting element at the 3'end of the genome, termed the constitutive transport element (CTE), which was shown to support Rev-independent HIV-1 structural protein expression. The CTE is a complex RNA secondary structure, which recruits the cellular mRNA export factor Tap directly to the genomic RNA (Bray *et al.*, 1994; Ernst *et al.*, 1997; Hammarskjold *et al.*, 1994). The export pathway of the unspliced genomic RNA of MLV is poorly defined (Hoshi *et al.*, 2002; Oshima *et al.*, 1998). It may be, similarly to MPMV, that a *cis*-acting mechanism operates to export intron-containing genomic RNA into the cytoplasm. So far a stem-loop structure in the R region of MLV genomes has been described to be important for the export of unspliced RNAs (Trubetskoy *et al.*, 1999).

Furthermore, retroviruses like HIV-1 and MLV exhibit polyadenylation signals in the R region at the 5' and 3'ends of the integrated provirus. Therefore they have evolved mechanisms to suppress the first polyadenylation signal for successful replication. In the case of HIV-1 the recruitment of U1 snRNP to the 5'ss suppresses the 5'polyadenylation signal (Ashe *et al.*, 2000; Gunderson *et al.*, 1998), whereas MLV adopted the use of inefficient signals, which due to failure RNA termination and readthrough transcripts including adjacent cellular DNA sequences (Furger *et al.*, 2001).

3 Viral vectors for gene therapy

Retroviruses are a promising tool for gene therapy based on their properties and their life cycle. Since retrovirus-based vectors enter the cytoplasm of the host cell in a receptor mediated manner and stably integrate their genomes into the host chromatin, they have been used for efficient gene transfer into mammalian cells for more than 20 years. Generally the development of such vectors comprised the construction of a transfer vector containing all *cis*-active sequences required for packaging, reverse transcription and integration, such as the packaging signal, the primer binding site and the long terminal repeats. Additionally, the coding sequences of the viral genes are removed to avoid the generation of replication competent retroviruses (RCR) and are replaced by the gene of interest (transgene). Production of infectious viral vector particles requires a packaging cell line, containing the genetic information of the viral proteins gag, pol and env (Fig.8) either stably integrated or transiently expressed through co-transfection of helper plasmids (Miller, 1990; Yu *et al.*, 2003).

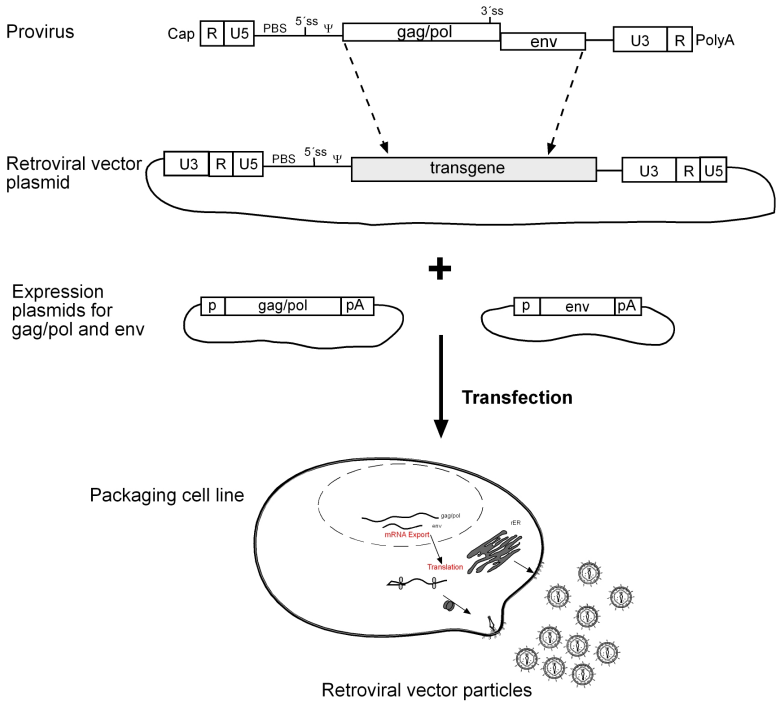


Figure 8: Turning a virus into a gene transfer vector

For the development of replication-deficient gammaretroviral vectors, the viral coding regions are replaced by the transgene. The gag/pol and env reading frames are separated on helper plasmids. Co-transfection into a packaging cell line together with the retroviral vector yields newly assembled retroviral vector particles in the supernatant (P: promoter; pA: Poly(A) signal).

Retroviral particles can be pseudo typed with env proteins from other viruses giving a broad cell tropism. For example the G-protein of the vesicular stomatitis virus (VSV-G), confering high stability of the particle, is used for the generation of gammaretroviral and lentiviral particles (Emi *et al.*, 1991; Yang *et al.*, 1995). VSV-G containing particles can easily be concentrated by ultracentrifugation to reach 10^9 infectious particles/ml (Burns *et al.*, 1993).

3.1 Gammaretroviral vector development

Retroviral vectors derived from gammaretroviruses (Miller & Rosman, 1989) or lentiviruses (Naldini *et al.*, 1996) are frequently used in gene therapy trials. The sequences controlling transcription are located in the U3 promoter/enhancer region. It was shown that MLV and other retrovirally derived enhancer sequences are blocked or essential transcription factors are not expressed in early embryonic stem cells (Baum *et al.*, 1999). This block was overcome by the development of a chimeric virus, containing all previously described positive features, leading to a new generation of vectors (Baum *et al.*, 1995; Grez *et al.*, 1990; Hildinger *et al.*, 1999; Schambach *et al.*, 2000). Recently, Trim28 has been identified as the factor responsible for the repression of the leader region of MoMLV in embryonic stem cells (Wolf & Goff, 2007). In detail the 5'LTR of the chimeric vectors contain sequences from MPSV (myeloproliferative sarcoma virus), the leader region is derived from MESV (murine embryonic stem cell virus) and the 3'LTR derives from SFFV (spleen focus forming virus). The so called FMEV (Friend-MCF-MESV hybrid) vector showed high transgene expression in hematopoietic cells (Baum *et al.*, 1995; Eckert *et al.*, 1996). Next, all viral (gag) sequences and AUGs were deleted, which leads to the SF11, SF110 and SF91 vectors, which vary by the presence or absence of functional splice sites (Fig.9). In the SF91 vector the packaging signal is embedded into an intron using the authentic MLV splice sites. This configuration ensures packaging of intron-containing genomes during particle production. Here, the splicing ratio of genomic to spliced RNA is similar to the balanced splicing pattern observed in MLV. The SF11 vector only exhibits a functional 5'ss, whereas both splice sites of the SF110 are mutated. Compared to the SF110 vector, SF91 and SF11 generate higher titers of virus (Hildinger *et al.*, 1999). It has been described that the presence of a 5'ss can positively influence RNA stability, RNA export, RNA transcription and elongation (Damgaard *et al.*, 2008; Fong & Zhou, 2001; Kornblihtt *et al.*, 2004; Reed

& Hurt, 2002). In the case of the SF91 and SF11 vectors, the 5'ss increased the amount of genomic RNA leading to more infectious particles. Inclusion of the woodchuck post-transcriptional regulatory element (wPRE, derived from woodchuck hepatitis virus) enhanced transgene expression and viral titer (Schambach *et al.*, 2000; Zufferey *et al.*, 1999) by mRNA stabilization and improvement of polyadenylation and export (Donello *et al.*, 1998; Loeb *et al.*, 1999; Popa *et al.*, 2002).

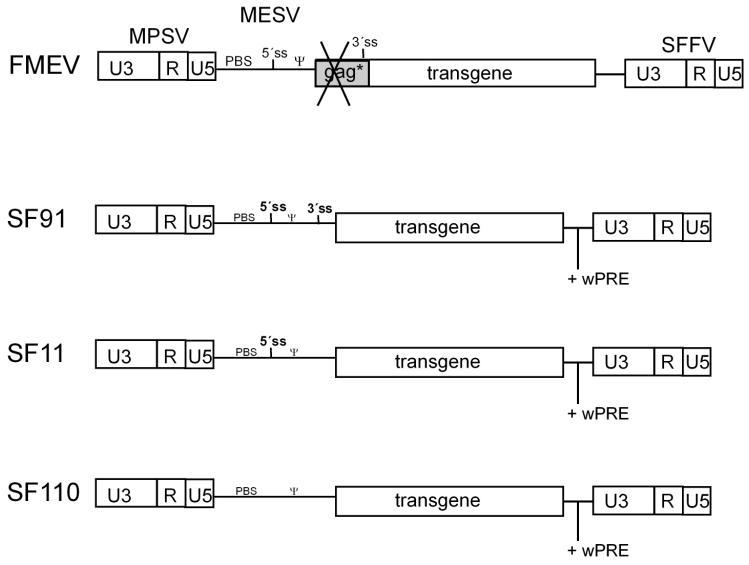


Figure 9: Gammaretroviral-based vector design

The SF91 vector is based on the FMEV (Friend-MCF-MESV) hybrid vector. All virus coding sequences (gag*) and AUG were deleted. The vector contains the whole leader region including the packaging signal flanked by the authentic viral splice sites. SF11 contain only the 5´ss, whereas the SF110 lacks both viral splice sites. Further modification leads to the introduction of the wPRE downstream of the transgene.

3.2 Self inactivating vectors (SIN vectors)

SIN vectors are termed as such because they generate, following reverse transcription, a defective, inactive promoter in the 5'LTR (Fig.10). In the packaging cell genomic RNA expression is driven by the 5'LTR, whereas after reverse transcription the expression of genomic RNA including the packaging signal and mobilization of endogenous viruses to generate RCRs is less likely. Due to the deletion of the U3 region expression of the transgene is driven, after reverse transcription and insertion into the host genome, by an internal promoter (Yu *et al.*, 1986). This configuration allows the usage of cellular promoters and reduces the promoter/enhancer elements from two to one. SIN vectors of the first generation were derived from MLV and are not applicable for gene therapy approaches due to

both low gene expression and low titers. Advanced SIN vector configurations were designed, based on the SF11 LTR vector, containing an internal SFFV promoter (Fig.10a). This configuration allowed a direct comparison of LTR- versus SIN-driven vectors with respect to safety aspects. Additionally, the introduction of the wPRE element and the presence of the viral 5'ss increased virus titer by stabilization of the RNA, enhanced RNA export and transcriptional elongation (Damgaard *et al.*, 2008; Kraunus *et al.*, 2004; Schambach *et al.*, 2000). The problem of low titers was solved by the substitution of the 5'MPSV promoter with the RSV (rous sarcoma virus) promoter. This substitution increased the ratio of packable full-length RNA to the RNA initiated at the internal promoter by overcoming promoter competition in the packaging cell (Schambach *et al.*, 2006).

The search for other positions for internal promoter insertion led to the generation of SINSF91 (Kraunus *et al.*, 2004). This vector showed a further drop in titer and was not suitable for clinical applications, but displayed an interesting splicing phenotype (Fig.10b). Compared with the balanced splicing ratio of the SF91, where the intron is located between the U5 sequences and the transgene, the same intron showed complete splicing when located between an upstream promoter and the transgene cassette (Kraunus *et al.*, 2004). This indicates that the splice sites of MLV per se are efficient for complete splicing. In addition, alternative splicing of MLV-derived vectors has still been observed after replacing the viral 3'ss with a strong cellular 3'ss (Lee *et al.*, 2004). This suggests alternative splicing in MLV is not regulated via weak 3'ss.

Retroviruses copy the R region, containing the weak polyadenylation signal, to the 5'end of the inserted provirus during reverse transcription. Therefore, retroviral SIN vectors suffer from inefficient 3'end processing in the 3'R region, increasing the probability of read-through and fusion transcripts. These effects are compensated by incorporation of the wPRE. To improve 3'end processing further, various polyadenylation enhancer elements (or upstream sequence enhancers, USEs) derived from viral or cellular genes were inserted into the deleted 3'U3 region of SIN vectors. Two copies of the USE derived from simian virus 40 (2x SV USE) showed the best results i.e. enhanced polyadenylation, 3'end processing, and suppression of transcriptional read-through. Furthermore, this configuration increases mRNA stability, titer and potentially the biosafety of SIN vectors (Schambach *et al.*, 2007).

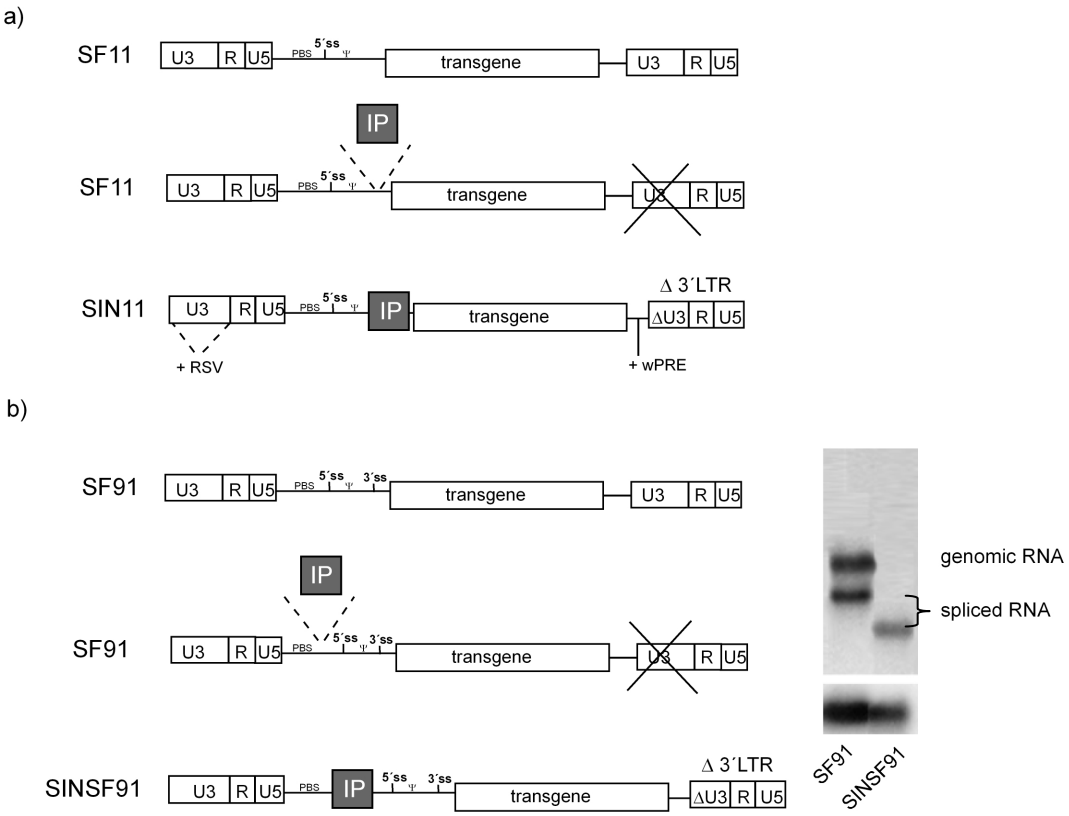


Figure 10: SIN vector development

a) Configuration of the SIN11 vector on the basis of the SF11 LTR vector. The internal promoter (IP) is located in front of the transgene and promoter/enhancer sequences of the 3´U3 are deleted. The introduction of a wPRE and the substitution of the 5'MPSV promoter by that of the RSV (rous sarcoma virus) further improved SIN vector design. b) SINSF91 contains an IP between the PBS and the 5'ss. As shown by northern blot analysis, this leads to complete splicing of the intron compared to the alternative splicing pattern of SF91 (Kraunus *et al.*, 2004).

3.3 Side effects and further modifications of viral vectors

The insertion of retroviruses into the cellular genome is a known risk factor for tumor development (Baum *et al.*, 2006; Baum *et al.*, 2004) by insertional mutagenesis. Different in vivo and *in vitro* assays have been established in order to characterize the mutagenic potential of the vector backbone in target cells (Du *et al.*, 2005; Modlich *et al.*, 2006; Montini *et al.*, 2006). Leukemia induction by MLV based vectors was reported in a murine model (Kustikova *et al.*, 2005; Li *et al.*, 2002; Modlich *et al.*, 2005), in macaques (Seggewiss *et al.*, 2006) and in five patients in two different gene therapy trials for X-linked SCID (Baum, 2007; Hacein-Bey-Abina *et al.*, 2008; Hacein-Bey-Abina *et al.*, 2003a; Howe *et al.*, 2008). All leukemic clones are characterized by insertion of the vector into loci known as proto-oncogenes (Calmels *et al.*, 2005;

Hacein-Bey-Abina *et al.*, 2003b; Li *et al.*, 2002; Schwarzwaelder *et al.*, 2007). The vector insertion leads to mutagenic effects including deregulation of neighboring genes (Fung *et al.*, 1983), the generation of read-through (fusion) transcripts (Benz *et al.*, 1980) or gene disruption. Gene disruption usually produces a monoallelic, recessive defect that needs further mutation in the second allele to show a phenotype (Suzuki *et al.*, 2006). In addition, fusion transcripts can be generated by the existence of a functional 5'ss in the vector backbone, thus leading to gene activation and insertional mutagenesis (Bokhoven *et al.*, 2008). Because of such side-effects due to retroviral vector integration, a major focus of ongoing work is to improve the safety of vector architecture. Moreover, it is important to understand the RNA processing machinery of retroviruses in more detail and in particular RNA splicing regulation to adopt possible new insights into further vector development.

B Aim of the study

Retroviral based vectors are efficient tools for the gene delivery of therapeutic transgenes. Successful gene therapy trials for the correction of inherited diseases e.g. immunodeficiencies like adenosine deaminase deficiency (ADA) (Aiuti *et al.*, 2002), chronic granulomatous disease (CGD) (Ott *et al.*, 2006) and X-linked severe combined immunodeficiency (Cavazzana-Calvo *et al.*, 2000) using retroviral vectors have been performed. Observations of adverse effects after transgene insertion from animal models and humans illustrated the need for improved vector design to increase safety. One simple approach is to reduce the transactivating potential of promoter enhancer elements and thereby the risk of insertional mutagensis by using SIN vectors for gene therapy. Nevertheless, it was shown using an *in vitro* cell culture assay that although SIN vectors carrying strong viral internal promoter/enhancers derived from SFFV are weaker insertional mutagens than LTR driven vectors, they are still able to transform cells by insertional mutagensis (Modlich *et al.*, 2006; Modlich *et al.*, 2008). One approach to enhance the safety of SIN vectors is the replacement of the strong viral internal promoter/enhancer to weak cellular promoter/enhancers. Furthermore, the introduction of additional chromatin elements into the deleted U3 region of SIN vectors may avoid interactions of the vector with surrounding cellular promoters. Examples of such shielding sequences are so called insulator elements such as the chicken HS4 (Ellis, 2005; Emery *et al.*, 2000; Rivella *et al.*, 2000; Yusufzai & Felsenfeld, 2004) or scaffold attachment regions (Bode & Maass, 1988; Lutzko *et al.*, 2003; Ramezani *et al.*, 2003). The aim of publication 1 was to analyze the impact of retroviral vector design on insertional transformation by using weaker cellular promoter/enhancer and insulator elements.

In some diseases correction of the phenotype needs a high expression level of the transgene (Lagresle-Peyrou *et al.*, 2006), whereas many transgenes applied for human gene therapy require no high expression rates (Thornhill *et al.*, 2008). However, further options to enhance transgene expression above the therapeutic threshold include improving the RNA processing of the vector by enhancing the 3'end processing or by using the positive effects of a 5'ss on transgene expression and titer (see above). Both splice sites of gammaretroviral vectors, which are used in clinical trials, are derived from MLV, but the mechanism of alternative splicing regulation of MLV is still not completely understood. The 5'ss matches the consensus sequence to

almost 100% and it is has been shown that the replacement of the viral 3'ss by a very strong cellular 3'ss (Lee *et al.*, 2004) still leads to alternative splicing. Additionally, RNA analysis of the SINSF91 vector showed that both authentic splice sites can splice very efficiently. These observations suggest that alternative splicing of MLV is regulated by an unknown *cis*-acting mechanism. The purpose of publication 2 and the manuscript was to investigate how MLV regulates alternative splicing to express viral genes in a balanced ratio.

C Publication 1

Physiological Promoters Reduce the Genotoxic Risk of Integrating Gene Vectors

Daniela Zychlinski, Axel Schambach, Ute Modlich, Tobias Maetzig, Johann Meyer, Elke Grassman, Anjali Mishra and Christopher Baum*

Published in Molecular Therapy, 2008 Apr; Vol. 16 p. 718-725

doi:10.1038/mt.2008.5

http://www.nature.com/mt/journal/v16/n4/abs/mt20085a.html

* corresponding author:

Christopher Baum,
Department of Experimental Hematology, Hannover Medical School,
Carl-Neuberg-Straße 1, 30625 Hannover,Germany.
E-mail: baum.christopher@mh-hannover.de

Author's contribution statement:

Daniela Zychlinski constructed the "minigene" reporter constructs, introduced the insulators into the used retroviral vector generation and performed the experiments and analyzed the data for Figs.2-5 and Suppl. Fig.2. Furthermore, she participated in designing the figures and in writing the manuscript. Axel Schambach designed and constructed the plasmids for Fig.1 and the basic retroviral SIN vectors (without insulator), helped in designing the "minigene" constructs, performed the experiments for Fig.1 and Suppl. Fig.1, and participated in designing the figures and in writing the manuscript. Ute Modlich introduced DZ into the *in vitro* immortalization assay (IVIM), participated in isolation of lineage bone marrow cells and in writing the manuscript. Tobias Mätzig participated in performing one IVIM assay. Johann Meyer provided the truncated CD34 receptor and supervised the "minigene" construction. Elke Grassman provided the chicken HS4 insulator constructs. Anjali Mishra performed backup *in vivo* experiments. Christopher Baum is the corresponding author, supervised all experiments and designed the study. In addition, CB critically evaluated and analyzed the data and wrote the paper.

Abstract

The possible activation of cellular proto-oncogenes as a result of clonal transformation is a potential limitation in a therapeutic approach involving random integration of gene vectors. Given that enhancer promiscuity represents an important mechanism of insertional transformation, we assessed the enhancer activities of various cellular and retroviral promoters in transient transfection assays, and also in a novel experimental system designed to measure the activation of a minigene cassette contained in stably integrating retroviral vectors. Retroviral enhancer-promoters showed a significantly greater potential to activate neighboring promoters than did cellular promoters derived from human genes, elongation factor-1*α* (*EF1α*) and phosphoglycerate kinase *(PGK)*. Self-inactivating (SIN) vector design reduced but did not abolish enhancer interactions. Using a recently established cell culture assay that detects insertional transformation by serial replating of primary hematopoietic cells, we found that SIN vectors containing the *EF1α* promoter greatly decrease the risk of insertional transformation. Despite integration of multiple copies per cell, activation of the crucial proto-oncogene *Evi1* was not detectable when using SIN-EF1*α* vectors. On the basis of several quantitative indicators, the decrease in transforming activity was highly significant (more than tenfold, $P < 0.01$) when compared with similarly designed vectors containing a retroviral enhancer–promoter with or without a well-characterized genetic insulator core element. In this manner, the insertional biosafety of therapeutic gene vectors can be greatly enhanced and proactively evaluated in sensitive cell-based assays.

Introduction

It has been observed in recent clinical trials and animal models that induction of clonal imbalance and malignant transformation can occur following the mutagenic insertion of gene vectors.[1–10] Clonal imbalance and transformation are particularly overt in, but not necessarily restricted to, hematopoietic cells that have a high capacity for clonal expansion. Murine models of bone marrow (BM) transplantation have demonstrated the leukemogenic potential of insertional mutagenesis in the contexts of various gammaretroviral vectors designed for human gene therapy.[4,11] A recent study showed that self-inactivating (SIN) lentiviral vectors with a relatively weak internal promoter derived from the human phosphoglycerate kinase gene (*PGK*) did not accelerate tumor induction in lymphoma-prone mice, in contrast to gammaretroviral vectors with strong enhancer-promoter sequences located in the long-terminal repeats (LTRs).[12] However, the *in vivo* assays that are generally used for quantifying the impact of vector design on the genotoxic risk are hampered by long observation times, the need for large numbers of animals, the requirement to set induction of severe disease as the experimental endpoint, and the high costs involved. It therefore remained unclear whether it was the type of enhancer-promoter used, other features of the vector's architecture, or the particular insertion pattern of lentiviral vectors that was responsible for the reported lack of tumorigenic side effects.[12]

As has been recently pointed out,[3] cell-based assays may provide quick and reliable results and mechanistic insights into crucial variables of gene vector biosafety. On the basis of a study by Du *et al.*,[7] we have developed an assay to detect the clonal dominance of insertional mutants which expand in initially polyclonal cultures of primary murine hematopoietic cells within 2 weeks after gene transfer and acquire serial replating ability, the extent of which provides a measure of clonal fitness.[6] Using this assay, we showed that gammaretroviral vectors with a SIN design, which are devoid of enhancer-promoter elements in the U3 region, are weaker insertional mutagens than their LTR driven counterparts. Nevertheless, SIN vectors with internal enhancer-promoters derived from spleen focus-forming virus (SFFV), an oncogenic murine leukemia virus (MLV), were able to transform cells by insertional activation of proto-oncogenes such as *Evi1*. It remained unclear whether relatively potent cellular promoters such as elongation factor-1*α* (*EF1α*)[13] which, in hematopoietic cells, is only two- to threefold less active than MLV enhancer-promoters,[14,15] may also induce clonal transformation. In order to address this important question, we used complementary approaches to evaluate enhancer promiscuity and transforming efficiency.

Results

Enhancer activity of frequently used promoters assessed in transient transfection assays

First, we used a plasmid-based reporter assay to analyze the enhancer interactions of various enhancer-promoter combinations on a minimal promoter driving a luciferase expression cassette. The elements of interest were cloned 3′ of a codon-optimized luciferase cassette in order to detect enhancer interactions with the upstream minimal promoter. Transcriptional readthrough was prevented by a strong termination site (**Figure 1a**). We focused our analysis on two widely used retroviral enhancer-promoters derived from the U3 region of myeloproliferative sarcoma virus (MPSV) or SFFV, and two well-known cellular promoters derived from human genes, namely, PGK and EFS, the latter representing a short, intron-less form derived from human *EF1α*. EFS is only two- to threefold less active in hematopoietic cells than SFFV-U3 is, and therefore potent enough to express clinically relevant genes.[15] Computer-assisted searches predict that EFS contains significantly fewer and less tightly clustered binding sites for transcription factors than the SFFV or MPSV sequences do (**Figure 1b**, for MPSV refer to ref. 16). Detailed functional studies of EFS[17] in hematopoietic cells are not available, in contrast to SFFV-U3.[16,18,19] The only motifs they both share are binding sites for the ubiquitous transcription factor SP1[17,18] (**Figure 1b**).

Transient transfection in 32D cells (murine nonleukemogenic myeloid progenitor cells) revealed that both cellular promoters (PGK, EFS) exhibit significantly weaker enhancer activity on the thymidine kinase (TK) promoter than do the retroviral promoters derived from MPSV or SFFV (**Figure 1c**; $P < 0.05$, unpaired *t*-test), which is a finding consistent with the transcriptional activity observed in retroviral and lentiviral SIN vectors.[15] Unexpectedly, the difference between SFFV and EFS did not reach statistical significance when testing *cis*-activation of the minimal TK promoter. In 32D cells, the cellular promoters did not significantly enhance the activity of either of the cryptic promoters tested. In contrast, the EFS promoter showed residual enhancer activity in 293T cells (**Supplementary Figure S1**).

We next used this assay to study the enhancer activity of a complete vector genome. The duplication of the MLV enhancer-promoter in the typical retroviral LTR configuration further increased the degree of reporter activation, whereas a SIN vector with an internal SFFV promoter showed a greatly reduced *cis*-activation potential. Interestingly, SIN vectors that

lacked an internal enhancer-promoter, nevertheless enhanced the minimal promoter (data not shown). The nature and localization of residual enhancer sequences in the SIN backbone remain to be identified. Importantly, when EFS or PGK were added as internal promoters, the enhancer activity of SIN vectors was not further increased (**Figure 1d**). Inserting the vector genome in the antisense orientation yielded similar results (**Figure 1d**).

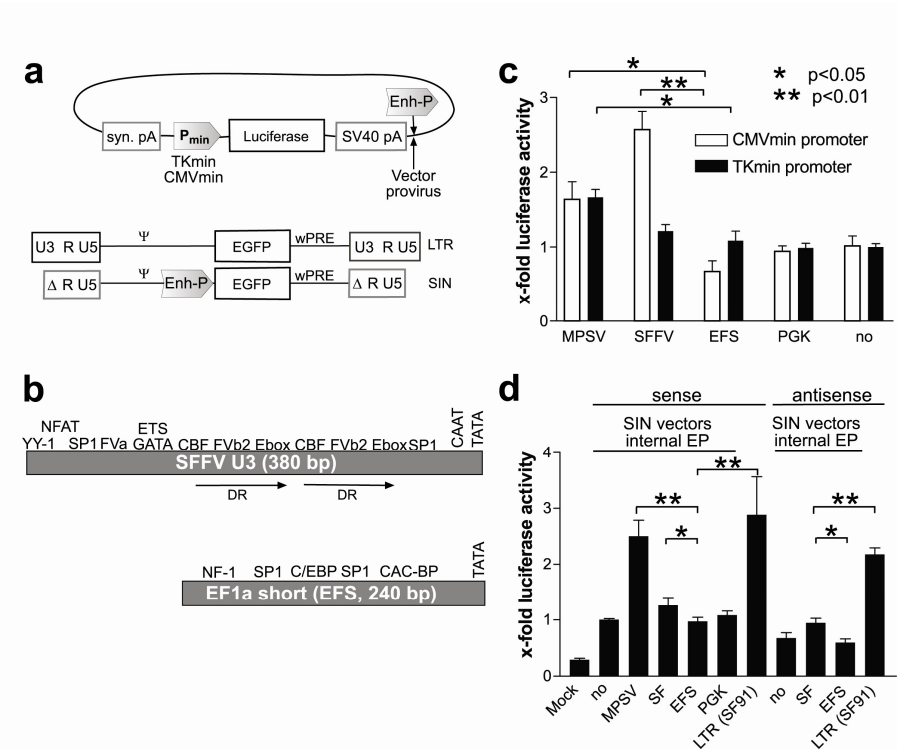


Figure 1: Transient transfection assays reveal low enhancer activity of cellular promoters and reduced enhancer activity of self-inactivating (SIN) vector in comparison with long-terminal repeat (LTR) vectors. **(**a) The reporter construct encodes a codon-optimized luciferase that is devoid of at least 300 known transcription factor binding sites. A synthetic polyadenylation (syn. pA) prevents potential readthrough from the enhancer-promoter cassette (Enh-P) cloned downstream of the termination site of the luciferase cassette. Alternatively, entire provirus sequences of LTR or SIN vectors were introduced in sense or antisense orientation. (**b**) Schema of transcription factor binding sites predicted by computer algorithms and functional data, revealing a dense clustering of binding sites in the SFFV-U3 region, in contrast to the cellular promoter elongation factor-1*α* (EF1*α*). (**c**) The retroviral Enh-P sequences derived from myeloproliferative sarcoma virus (MPSV) and spleen focus-forming virus (SFFV) show a significant induction of two different minimal promoters, in contrast to the cellular promoters [EFS and phosphoglycerate kinase (PGK)]. No, lack of Enh-P 3′ of the luciferase cassette. Error bars indicate standard deviations of four experiments. (**d**) When compared with LTR vectors, SIN vectors show a greatly decreased enhancer activity, even when they contain the same Enh-P. Enhancer activity is also more dependent on orientation (sense or antisense to the luciferase cassette) in the SIN configuration. CMV, cytomegalovirus; EGFP, enhanced green fluorescent protein; no, SIV vector lacking an interval Enh-P TK, thymidine kinase; WPRE, woodchuck hepatitis virus posttranscriptional regulatory element.

Enhancer interactions evaluated after stable gene insertion

Having established that the MLV enhancers are significantly more promiscuous than the cellular ones tested here, we evaluated enhancer interactions in the context of stable gene insertion. For this purpose, we constructed a gammaretroviral SIN vector harbouring a minigene in the 3′U3 deletion activatable by the enhancer of the internal enhanced green fluorescent protein (EGFP) cassette (**Figure 2a**). The minigene consisted of a minimal cytomegalovirus (CMV) promoter in front of a truncated version of the CD34 surface marker gene (tCD34). Target cells were transduced at a low multiplicity of infection (MOI) so as to achieve a low average transgene copy number. Concordant with the transient transfection studies, the internally located retroviral enhancer-promoter activated the minigene cassette to the greatest extent. In SC1 fibroblasts that are highly permissive in respect of the two cellular enhancer-promoters (EFS and PGK), we found a very low induction of the minigene cassette (white bars in **Figure 2b**), in contrast to the results achieved with vectors containing the internal SFFV enhancer-promoter. In additional experiments, we ruled out the possibility that the tCD34 activity mediated by these constructs could be explained by reinitiation of translation in the transcript originating from the internal promoter (**Supplementary Figure S2**).

Using these retroviral vectors, we could also test enhancer interactions in primary murine hematopoietic cells. When compared with fibroblasts, hematopoietic cells displayed a generally lower enhancer interaction. This finding was consistent with the data from the transfection studies referred to earlier, showing that only the internal SFFV enhancer-promoter significantly induced the minigene (gray bars in **Figure 2b**). In order to rule out the possibility that cellular enhancers located adjacent to the vector had contributed to the observed results, we added a well-characterized, 250 base pair (bp) insulator core element (1×HS4) with potential enhancer-blocking activity[20] into the LTR, upstream of the minigene cassette. This had no effect except that it revealed the high reproducibility of the assay conditions (**Figure 2b**). In view of the fact that the addition of an insulator did not result in a significant decrease in minigene activation, these data suggest that the minigene activity is mainly caused by the cryptic promoter in the 5′LTR. We assume that readthrough transcription from the internal promoter cassette suppresses the activity of the minigene located in the 3′LTR. These findings support the conclusion that the retroviral enhancer-promoter is significantly more likely to activate neighboring promoters through enhancer-mediated interactions than are the promoters derived from the human genes, *EF1α* and *PGK*.

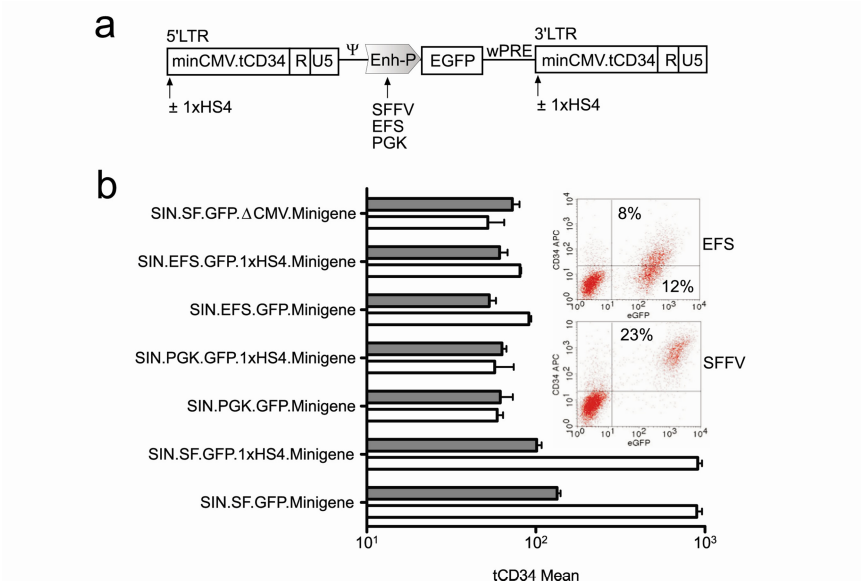


Figure 2: Stably integrated "minigene" reporter constructs reveal the low enhancer activity of the cellular promoters tested.
(**a**) The retroviral self inactivating (SIN) vector contains a minigene cassette in the U3 region of the long-terminal repeat (LTR), consisting of a minimal promoter followed by a truncated CD34 surface marker gene (*tCD34*). Termination is mediated by signals in the R region of the LTR. The same vectors were also tested with 1×HS4 insulator element inserted upstream of the minigene cassette. Enh-P sequences were introduced as indicated. A control vector contained no minimal promoter (ΔCMV). (**b**) In SC1 fibroblasts and lineage-negative primary murine hematopoietic cells, only the internal spleen focus–forming virus (SFFV) promoter led to a significant activation of the minigene cassette. Background levels are reflected in the tCD34 expression of the construct that lacks a minimal promoter. Data obtained from SC1 cells are shown in white columns and lineage-negative cells in grey ones. The insets show representative flow cytometry data (SC1 cells) for illustration of the assay, CD34 expression shown on the y-axis. Potential activation of the minigene by transcriptional readthrough could be ruled out (**Supplementary Figure S2**). CMV, cytomegalovirus; EFS, short form of EF1*α* EGFP, enhanced green fluorescent protein; PGK, phosphoglycerate kinase.

Reduced insertional risk of SIN vectors with internal cellular promoters

Our third approach was to test the same retroviral SIN configurations in the clonal dominance assay.[6] Previous studies had shown that a SIN vector with an internal SFFV-U3 has significantly less transforming efficacy than a vector with the same sequences contained in the LTRs,[6] a finding in line with the transfection studies described earlier (**Figure 1d**). We conducted experiments in which we made side-by-side comparisons of the transforming capability of gammaretroviral SIN vectors harboring an internal MLV enhancer-promoter (SIN.SF.GFP) and the EFS enhancer-Promoter (SIN.EFS.GFP). In order to exclude the possibility that a potential failure to detect a residual transforming capacity of the SIN.EFS.GFP vector was because of an insufficient transgene copy number, we modified the originally described protocol, and transduced hematopoietic cells on four consecutive days, each time with a high MOI. As a result, well over 95% of the cultured cells expressed the transgene. This led to a high average copy number (typically >20 copies/cell, **Table 1**). Even under these extreme conditions, in the course of seven independent assays, we did not recover a single clone with continued replating capacity following transduction with

SIN.EFS.GFP, whereas clones arose in all the cultures that had been transduced with SIN.SF.GFP (**Table 1**; **Figure 3**). The average copy number of the SIN.SF.GFP-transduced bulk culture, determined at 4 days after the last exposure to the particles, never exceeded 27 (19.9 ± 4.3), and similar copy numbers were achieved when using a tenfold lower MOI of this vector (**Table 1**). In contrast, under identical conditions, the vector SIN.EFS.GFP led to an average copy number of 40 ± 30 and showed a greater dependence on the MOI, thereby indicating a higher maximal tolerated dose. These data suggest an overdose inhibition effect encountered when using the more mutagenic vector (SIN.SF.GFP), perhaps caused by insertional disregulation of too many alleles in the same cell or by toxicity related to abundant expression of EGFP. Such overdose inhibition effects are known from studies with other genotoxic carcinogens,[21] and may have contributed to the relatively low replating efficiency observed in two of the six experiments conducted with SIN.SF.GFP (**Figure 3**; see Discussion). Summarizing our studies in the clonal dominance cultures (details in **Table 1**), the minimal copy number to transform hematopoietic cells was ~2 for SIN.SF.GFP (containing the SFFV enhancer-promoter),[6] and >40 for SIN.EFS. As the incidence of independent clones induced by SIN.SF.GFP was on average 2 per 100,000 exposed cells,[6] the assays conducted with this vector reflected at least 10 different mutants. In contrast, after transducing a total of 700,000 cells with SIN.EFS, no mutants with serial replating capacity were observed. The frequency of mutants was therefore ~2 $\times 10^{-5}$ for the vector with the internal SFFV promoter, and $<1.43 \times 10^{-6}$ for the EFS vector (more than tenfold lower).

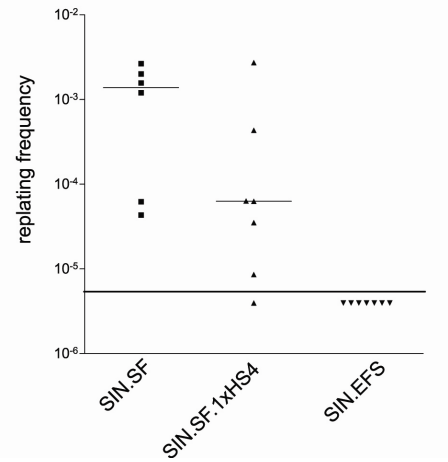


Figure 3: The clonal dominance assay shows a significant reduction in transformed cells following the use of a self-inactivating (SIN) vector with a physiological cellular promoter (EFS).
The ratio of replating frequency (determined by limiting dilution cloning) per vector copy number (detected by real-time PCR) 4 days after transduction, is a measure of the degree of transformation. A SIN vector with an internal EFS promoter shows no transforming capacity. The efficiency of transformation is significantly reduced in comparison to an unmodified or insulated SIN vector with an internal retroviral promoter. All seven experiments conducted with SIN.EFS vectors yielded results that were below the detection limit, indicated as a horizontal line. For negative values, calculations are based on the assumption that a replating clone would have been detected when plating 97 instead of 96 wells.

Table 1 Replating assay results

Vector	Exp	MOI	Positive wells (100 cells/well)	Positive wells (10 cells/well)	Replating frequency	Mean copy number (SD)	Replating frequency/ Mean copy number (SD)
SIN.SF.GFP	1	40	96	29	0.04340	16.4	0.0026
	2	40	84	16	0.02	16.7	0.0012
	3	40	13	4	0.0013	20.9	6.2×10^{-5}
	4	12	96	30	0.044	22	0.002
	5	40	9	2	0.0011	25.5	4.3×10^{-5}
	6	4	92	41	0.042	26.6	0.0016
	Mean (SD)	NA	NA	NA	0.01645 (0.002)[a]	19.9 (4.3)[a]	0.00119 (0.00115)
SIN.EFS.GFP	1	40	0	0	0	17.2	0
	2	40	0	0	0	37.8	0
	3	40	0	0	0	23.9	0
	4	40	0	0	0	27.6	0
	5	4	0	0	0	11	0
	6	12	0	0	0	33.3	0
	7	40	0	0	0	94.6	0
	Mean (SD)	NA	NA	NA	0	40.2 (31.3)[a]	0[b]
SIN.SF.GFP.1×HS4	1	40	62	3	0.0094	21.8	4.32×10^{-4}
	2	40	11	2	0.0013	20.5	6.33×10^{-5}
	3	40	12	1	0.0012	19.1	6.2×10^{-5}
	4	40	5	1	5.8×10^{-4}	16.5	3.5×10^{-5}
	5	4	93	23	0.0313	11.5	0.0027
	6	12	0	0	0	17.3	0
	7	40	2	0	1.91×10^{-4}	22.3	8.5×10^{-6}
	Mean (SD)	NA	NA	NA	0.0025 (0.0038)[a]	20.0 (2.3)[a]	0.00047 (0.001)

Abbrevations: Exp., experiment; NA, not applicable.
[a]Calculation of only the values obtained with a multiplicity of infection (MOI) of 40. [b]Corresponding values shown in the logarithmic scale of **Figure 3** assume that the 97th well would have been positive if 97 wells had been plated instead of 96.

A single copy of the HS4 insulator core element (250 bp) is unable to block the transforming potential of the internal retroviral enhancer-promoter

Further, using the clonal dominance assay, we investigated whether the well-characterized 250-bp insulator core element derived from the chicken HS4 locus, and containing the boundary functions as well as the enhancer-blocking activity of the cHS4 insulator,[20] could protect against being transformed by SIN.SF.GFP.

The insulator was inserted into the SIN-U3 region of the vector. Although a duplication of this element (dimer) may be more effective in blocking enhancer interactions,[20] it was found to lower vector titers, indicating genetic instability, and we therefore used the insulator as a monomer (**Figure 4**). In contrast, the vector with the single copy of the insulator located in the deleted U3 region of the LTR (SIN.SF.1×HS4.GFP) gave equal titers and expression levels of the transgene as the unmodified SIN vector (**Figure 4**). Using this vector, we conducted a total of seven clonal dominance assays, each time side-by-side with the corresponding uninsulated vector (SIN.SF.GFP). SIN.SF.GFP.1×HS4 induced a lower replating frequency

than the uninsulated counterpart in five of these assays, and a higher replating frequency in the two other assays (**Table 1**). The average copy number achieved was almost identical for both vectors (19.9±4.3 for SIN.SF.GFP; 20.0±2.3 for SIN.SF.GFP.1×HS4), as were the EGFP expression levels, underlining the reproducibility of our assay conditions. These experiments showed that incorporation of the insulator core element, although stably transmitted by the retroviral SIN vector, was not sufficient to protect against insertional transformation. However, the insulator clearly tended to reduce genotoxicity (P = 0.09, n = 7).

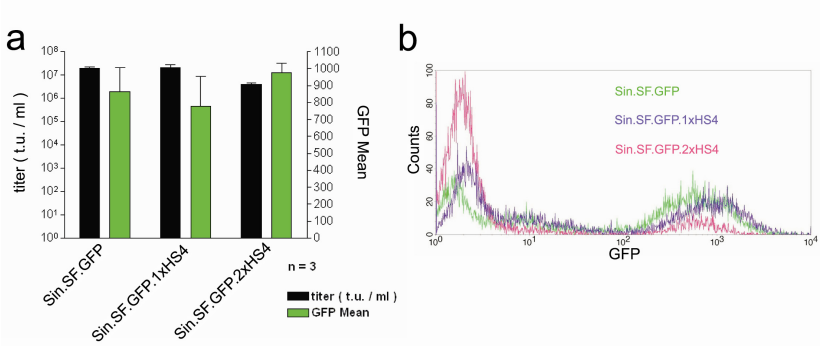


Figure 4: Self-inactivating (SIN) vectors containing an internal retrovirus-derived enhancer–promoter (SF) and insulator sequences (HS4 250 base pair core element) in the otherwise deleted U3 region of the long-terminal repeat.
(**a**) The vector with a single copy of the HS4 element (SIN.SF.GFP.1×HS4) has identical titers and expression levels (in SC1 fibroblasts) as the unmodified counterpart (SIN.SF.GFP) does. In contrast, a vector with a tandem repeat of the HS4 element had significantly reduced titers, indicative of genetic instability (SIN.SF.GFP.2xHS4). (**b**) Histogram of enhanced green fluorescent protein (EGFP) expression achieved with the different vectors under similar multiplicity of infection (SC1 fibroblasts). Differences in the replating frequency cannot be explained by variations in EGFP expression levels. t.u., transducing units.

Lack of molecular evidence for upregulation of *Evi1* when using the SIN vector with the internal cellular promoter

Finally we tested, at the molecular level, whether the EFS vector was able to induce the upregulation of a cellular proto-oncogene by random insertion. Southern blot experiments performed with genomic DNA harvested from replating clones revealed multiple insertions when using vectors SIN.SF.GFP or SIN.SF.GFP.1xHS4 (**Figure 5a**). This can be explained by the high MOI used for transduction. Because many of these insertions may represent innocent bystander mutations,[4] and previous studies had already confirmed that the replating assay correctly reflects insertional mutants,[6,7] we did not attempt to clone the insertion sites. We rather chose *Evi1* as an indicator allele for transcriptional disregulation because previous studies had shown that clones with serial replating capacity often have insertional events in this proto-oncogene, among others.[6,7] Northern blot experiments performed with messenger RNA harvested from replating cells revealed upregulation of *Evi1* transcripts, including low-molecular-weight forms that are likely to represent alternative splice products of this complex gene (**Figure 5b**). Because Northern blot analyses showed the expression of relatively short

transcripts, we chose primers that span exons 4 and 5 for real-time reverse transcriptase PCR (the entire *Evi1* gene contains 18 exons). Using this sensitive approach, we detected upregulation of *Evi1* transcripts as early as day 6 after transduction-but only after transduction with vectors SIN.SF.GFP or SIN.SF.GFP.1×HS4 (**Figure 5c**). The vector with the cellular promoter (SIN.EFS) was unable to activate this allele beyond the levels detected in mock-treated cultures (**Figure 5c**).

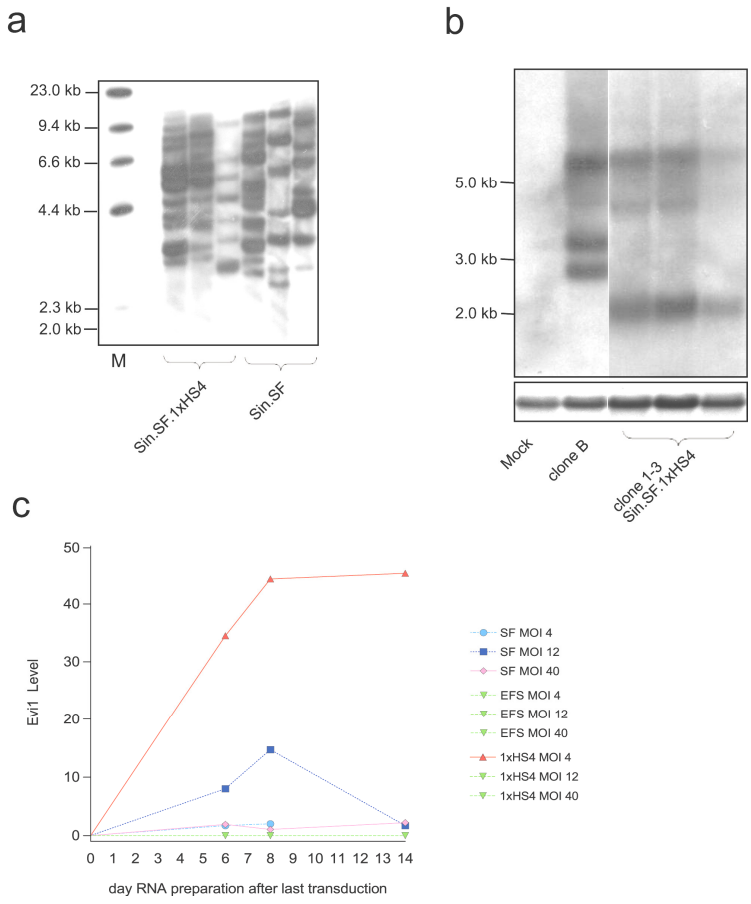


Figure 5: Molecular characterization of cultures and replating clones.
(**a**) Southern blot analysis of replating clones shows multiple vector insertion sites when using vectors SIN.SF.GFP or SIN.SF.GFP.1×HS4. (**b**) Northern blot revealing transcriptional upregulation of *Evi1* in replating clones. Mock-transduced lin−bone marrow cells cultured for >14 days served as negative control, and 18S RNA served as loading control. (**c**) Induction of *Evi1* transcripts can be detected as early as 6 days after gene transfer, thereby suggesting a quick outgrowth of mutants. In contrast, cultures transduced with a vector containing the EFS internal promoter showed no detectable induction of *Evi1* at any of the time points. MOI, multiplicity of infection; SIN, self-inactivating.

Cytological studies performed with cells harvested before replating corroborated this finding, by revealing an increase in the frequency of occurrence of blasts only with the use of vectors SIN.SF.GFP or SIN.SF.GFP.1×HS4 (**Figure 6**). This suggested that even if upregulation of *Evi1* could be induced by integration of SIN.EFS, the amount may be too small to trigger transient or permanent expansion under our experimental conditions.

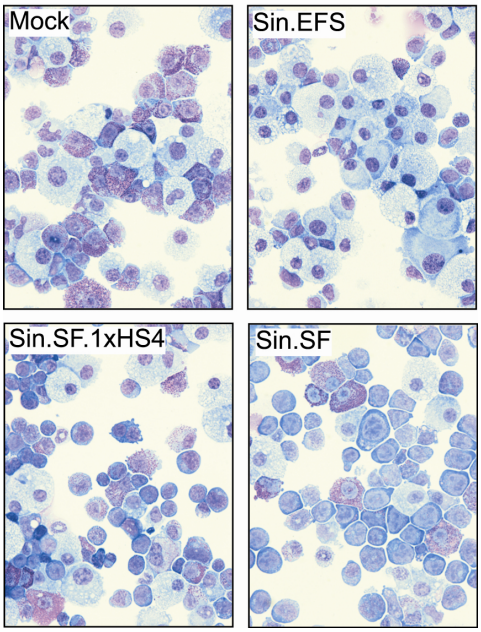


Figure 6: Cytology of cultures before replating (May-Grünwald–Giemsa stain of cytospin preparations). Mock-treated and SIN.EFS transduced cultures both show a predominance of mature myeloid cells, whereas numerous blasts are present in cultures transduced with vectors SIN.SF.GFP or SIN.SF.GFP.1×HS4; SIN, self-inactivating.

Discussion

The findings from our study show that, in hematopoietic cells, the cellular enhancer-promoters tested here (EFS and probably also PGK) are much less likely to activate neighboring genes than are viral promoters; retroviral enhancer-promoters were particularly “promiscuous”, activating various types of promoters in different configurations, probably as a result of the dense clustering of potent binding sites available for a variety of cellular transcription factors. The introduction of the cHS4 insulator 250-bp core element into both flanks of the internal cassette was not sufficient to prevent immortalization of hematopoietic cells triggered by random insertion of a retroviral enhancer–promoter, although it tended to reduce the replating activity of mutants (P = 0.09, n = 7). In contrast, SIN vectors using the EFS enhancer-promoter were unable to immortalize primary hematopoietic cells even in the absence of additional insulators, despite rather drastic dose escalation. Accordingly, these vectors did not lead to detectable levels of *Evi1* upregulation in cultured cells. Quantitative indicators of the reduced transforming efficacy of the vectors containing physiological promoters are the number of immortalized clones induced per treated cell (more than tenfold lower) and the average copy number required for inducing clonal dominance (>20-fold higher). The weaker enhancer promiscuity of the EFS sequence (**Figures 1** and **2**) appears to be the most likely explanation for its inability to induce clonal dominance in cultured cells, thereby suggesting that cellular proto-oncogene activation needs to overcome a threshold in

order to induce transformation. An alternative mechanism of enhanced biosafety of SIN vectors containing cellular promoters is an altered integration pattern. The preintegration complex, which is responsible for tethering of proviral DNA to certain chromatin loci, might contain transcription factors binding to enhancer sequences in the LTR or elsewhere in the proviral DNA. As *Evi1* does not seem to represent a hot spot of retroviral insertion events in unselected hematopoietic cells, thousands of insertion sites might have to be sequenced in order to address this question.[22] We conclude that it is possible to design integrating vectors with increased biosafety even for those gene therapy settings that require relatively high expression levels of therapeutic genes.

Our data are validated by previously published BM transplantation studies, for which a mechanistic explanation was provided. Database analyses of insertion sites associated with clonal dominance *in vivo* suggest that enhancer interactions account for the majority of transforming lesions induced by retroviral vectors.[23] We suggest that the lack of tumor-promoting potential observed with a lentiviral SIN vector is largely the result of the low enhancer promiscuity of the cellular promoter used (PGK).[12] Also in line with our findings, the *EF1α* promoter poorly activated a neighboring tissue-specific promoter of a lentiviral vector *in vivo*.[24] Although we observed a correlation between gene activation efficacy in reporter assays and transformation efficacy in primary hematopoietic cells (as determined by replating assays), the current study reveals that reporter assays alone would not be sufficient to assess the transforming potential of a given vector: the relatively weak enhancer activity of the complete SIN vector harboring an internal SFFV promoter (**Figure 1d**) stands in sharp contrast to the transforming potential detected in the replating assay (**Figure 3**). Major uncertainties that remain after analysis of the reporter assays are the predictive value of the cell background chosen and the threshold of gene activation required for cell transformation by upregulation of proto-oncogenes.[25]

Our results also encourage further investigation of the use of insulator elements to increase vector biosafety. In our experimental setting, the 250-bp core element, which is reported to have enhancer-blocking activity, tended to reduce the incidence of transformation of primary murine hematopoietic cells by a vector containing a strong retroviral enhancer–promoter. This result is in line with observations made in polyclonal cultures of lentivirally transduced Jurkat cells.[26] Modifications of this element, which is also of interest in relation to decreasing the position-dependence of gene expression, may further enhance vector biosafety,[27,28] thereby reaching the level of statistical significance in our assay. The insulated vector control

showed that the level of reporter gene expression had no transformation-promoting role in our experimental setting, as evidenced by a clear trend toward lower replating frequencies with the use of insulated vectors that mediated identical levels of EGFP expression. While EGFP may overload the cellular protein degradation machinery, there is no evidence hat its overexpression in hematopoietic cells is oncogenic.[29] In agreement with this, we earlier observed a comparable induction of replating activity when using gammaretroviral vectors that express a strongly attenuated EGFP or the dsRed protein under control of the MLV enhancer-promoter.[6] However, we cannot rule out the possibility that the extreme levels of EGFP expression mediated by multiple copies of SIN vectors with a strong internal promoter may have reduced cell fitness in our assay conditions. The gain in safety achieved with more physiological promoters may therefore be even higher than indicated in this report. Overdose inhibition effects may explain the finding that the maximal copy number achieved with vectors containing a strong enhancer-promoter (insulated or not) was ~20, whereas at least twice as high a maximal tolerated dose was observed with a vector containing a more physiological promoter. Fitness-reducing insertional lesions could also have contributed to the apparent overdose inhibition effect observed with vectors containing retroviral enhancer-promoters, in line with observations made with other genotoxic agents.[21] The vector containing the physiological promoter derived from EF1*α* showed no evidence of overdose inhibition, with the cell fitness remaining indistinguishable from that in mock cultures, and identical results were achieved over a wide range of gene dosage (11-94 copies/cell, **Table 1**).

In summary, this study not only demonstrates the robustness of the cell-culture-based clonal dominance assay for revealing the insertional genotoxicity of dose-escalated integrating vectors, but also reveals, for the first time, that a significant protection against insertional transformation can be achieved by modifying the vector's enhancer-promoter elements. Redesigned randomly integrating vectors will probably lead to an increased therapeutic index in future studies of gene therapy targeting hematopoietic and other cells, and may also reduce the "insertional bias" in the use of retroviral or lentiviral vectors in basic research.

Materials and Methods

Plasmid-based enhancer assay.

pRG-TK (Promega, Madison, WI) is a basic reporter plasmid harboring the codon-optimized Renilla luciferase gene (hRluc) cleared of more than 300 known transcription factor binding sites. In this construct the herpes simplex virus–TK promoter was substituted with a minimal herpes simplex virus–TK (TKmin) or CMV (CMVmin) promoter, which can be activated by neighboring enhancer–promoter combinations. Promoters (EF1*α* short promoter, EFS; MPSV U3, MPSV; PGK; Rous sarcoma virus U3, SFFV) or complete LTR-driven or SIN vector cassettes (integrated proviral forms) were introduced into *Bam*HI and *Sal*I sites located downstream of the SV40 pA signal (Figure 1a). In order to reduce unspecific luciferase background activity, a synthetic pA/transcriptional pause site was cloned upstream of the minimal promoter. Luciferase values were corrected for GFP expression mediated by a co-transfected plasmid so as to normalize for transfection efficiency.

Retroviral vectors and vector production.

The retroviral SIN vector backbones SIN.SF.GFP and SIN.PGK.GFP have been described earlier.[15] They harbor the respective internal promoters (SF, SFFV-U3; PGK), the eGFP transgene (Clontech, Mountain View, CA), and an X-protein ORF deleted version of the woodchuck hepatitis virus posttranscriptional regulatory element.[30] In order to construct SIN.EFS.GFP, the 240 bp intron-less promoter of EF1*α* short (EFS)[15] was introduced as a *Bam*HI/*Cla*I fragment. The cHS4 core insulator (250 bp; ref. 20) was cloned into the 3′ dU3 region using *Nhe*I in single or duplicated version, 1×HS4 or 2×HS4, respectively. The minigene consists of a minimal CMV promoter followed by a truncated 400 bp version of the CD34 receptor (tCD34), which can be detected by a monoclonal mouse-anti CD34 antibody (clone qbend/10; AbD Serotec, Dusseldorf, Germany). The minigene cassette was introduced into the *Eco*RI site of the residual 3′ U3 region (Figure 2) of SIN.SF.GFP. In order to construct SIN.SF.GFP.ΔCMV minigene, the minimal CMV promoter was excised using *Nhe*I and *Mlu*I restriction sites. The details of construction are available on request. In this vector configuration, the majority of the minigene activity is expected to result from the 5′LTR, because the cassette located in the 3′LTR is likely subject to transcriptional suppression by readthrough from the upstream-located EGFP cassette. Cell-free viral supernatants were generated by transient transfection of 293T cells with vector-plus-packaging constructs encoding gag–pol proteins (M57-DAW)15 and ecotropic envelope (K73eco).[31] Viral titers as determined on SC1 fibroblasts were in the range of 10^7 infectious transducing units/ml of

unconcentrated supernatant. All experiments were performed with thawed vector stocks of known titers.

Isolation of lineage-negative BM cells and retroviral transduction.

Lineage-negative (Lin−) BM cells of untreated C57BL6/J mice (Charles River Laboratories, Wilmington, MA) were transduced as described earlier.[32] Briefly, Lin− cells were isolated from complete BM by magnetic sorting, using lineage-specific antibodies (Lineage Cell depletion kit; Miltenyi, Bergisch Gladbach, Germany) and were cryopreserved in aliquots. Before retroviral transduction, Lin-BM cells were prestimulated for 2 days in Iscove's Modified Dulbecco's Medium (Biochrom, Berlin, Germany) containing 50 ng/ml murine stem cell factor, 100 ng/ml hFlt-3 ligand, 100 ng/ml hIL-11, 10 ng/ml mIL-3 (PeproTech, Heidelberg, Germany), 10% fetal calf serum, 1% penicillin/streptomycin, and 2 mmol/l glutamine at a density of 1–5 × 105 cells/ml. The cells were transduced on days 3, 4, 5, and 6 using 10^5 cells and an MOI of 1–10 per transduction. Virus preloading was carried out on RetroNectin-coated (10 µg/cm2; TaKaRa, Otsu, Japan) suspension culture dishes by spinoculation for 30 minutes at 4 °C. The transduction starts with 1 × 10^5 cells seeded into 500 µl medium and was increased by 250 µl medium on the following days, so that the culture volume was 1.25 ml on day 6. Real-Time PCR analysis (copy number) and flow cytometry (FACSCalibur; Becton-Dickinson, Heidelberg, Germany) were evaluated at the time points indicated in the text.

Clonal dominance assay.

After retroviral transduction, the BM cells were expanded as mass cultures for 2 weeks in Iscove's Modified Dulbecco's Medium containing 50 ng/ml murine stem cell factor, 100 ng/ml hFlt-3 ligand, 100 ng/ml hIL-11, 10 ng/ml mIL-3, 10% fetal calf serum, 1% penicillin/ streptomycin, and 2 mmol/l glutamine. During this time, the cell density was adjusted to 5 × 10^5 cells/ml every 3 days. After mass culture expansion for 14 days the BM cells were plated into 96-well plates at densities of 100 cells/well and 10 cells/well.[6] Two weeks later the positive wells were counted, and the frequency of replating cells was calculated based on Poisson statistics using L-Calc software (Stem Cell Technologies, Vancouver, Canada). Selected clones were expanded for further characterization.

Taqman real-time PCR analysis and Northern blot.

Quantitative PCR was performed on an Applied Biosystems 7300 Real-Time PCR System (Foster City, CA) using the Quantitect SYBR Green Kit (Qiagen, Hilden, Germany) as

described earlier.[6] The woodchuck hepatitis virus posttranscriptional regulatory element specific primers (forward 5′-GAGGAGTTGTGGCCCGTTGT-3′ and reverse 5′-TGACAGGTGGTGGCAATGCC-3′) amplified a 94 bp fragment. The woodchuck hepatitis virus posttranscriptional regulatory element specific signal was normalized by the signal of a housekeeping gene (flk-1 intron enhancer, gene ID AF061804, bases 352-459, forward 5′-gtgaattgcagagctgtgtgttg-3′ and reverse 5′-attcattgtataaaggtgggattg-3′). Results were quantified using the comparative *C*T method. A standard curve for the PCR was derived using a 32D cell clone with a known number of insertions as determined by Southern blot. Day 4 was chosen for the real-time PCR studies, because potential plasmid contaminations resulting from the use of supernatants produced by transient transfection of packaging cells were no longer detectable at this time point under our assay conditions (data not shown). In order to evaluate messenger RNA expression, RNA was extracted from expanded clones using RNAzol (WAK chemicals, Steinbach, Germany) and the RNAeasy micro kit (Qiagen, Hilden, Germany). Reverse transcription was performed with 0.5–2 µg RNA using PowerScript MLV reverse transcriptase (Becton-Dickinson), and real-time PCR for Evi1 expression using Taqman Gene Expression Assays (for Evi1 Mm00514810_m1, directed against exon boundary 4–5; Applied Biosystems, Foster City, CA) and a control assay against beta actin. Northern blots were performed using an Nco1 fragment of the Evi1 complementary DNA, the GFP complementary DNA or 18S RNA as described in refs. 33,34.

Statistical analysis.

Data from experiments are expressed as mean values ±SD. Student's unpaired *t*-test was used for comparing the variables in the transfection assays. For the immortalization assays, an analysis of variance test revealed that the differences in transforming efficacy observed in the replating assay were related to vector design. Student's paired *t*-test was used for comparing differences between indicated groups. $P < 0.05$ was considered significant.

Acknowledgments

We thank Sabine Knoess and Maimona Id (Hannover Medical School) for technical assistance with replating and transfection assays and Meghan Williams and Shawnagay Clarke (Cincinnati Children's Hospital) for assistance with cloning the insulator elements in self-inactivating vectors. The study was supported by grants from the the Else-Kröner foundation, the German Ministry for Research and Education (network grant TreatID), the Deutsche Forschungsgemeinschaft (network grant SPP1230 and Excellence Cluster

REBIRTH), the Deutsche Studienstiftung, the European Union (integrated project CONSERT, LSHB-CT-2004-005242; network of excellence CliniGene, LHSB-CT-2006-018933), and the National Cancer Institute (R01-CA107492-01A2).

Supplementary materials

Figure S1. Similar to 32D cells, 293T cells show a significantly greater enhancer activity of the retroviral promoters in comparison with the cellular promoters. Figure S2. Expression of the CD34 minigene contained in the retroviral reporter vector cannot be explained by transcriptional readthrough.

References

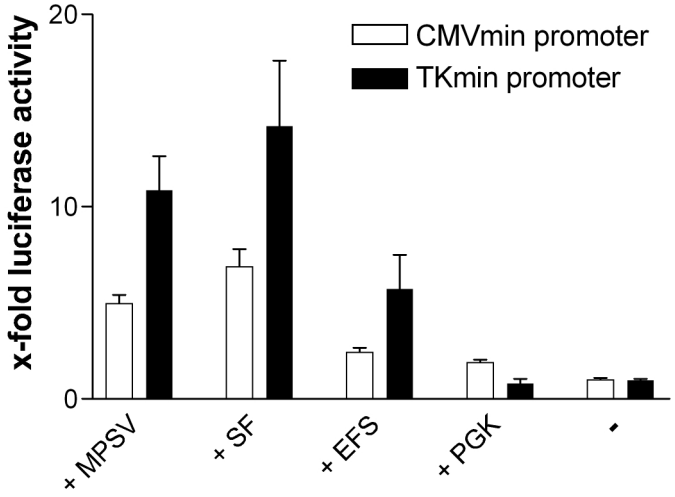
1. Li, Z, Dullmann, J, Schiedlmeier, B, Schmidt, M, von Kalle, C, Meyer, J *et al.* (2002). Murine leukemia induced by retroviral gene marking. *Science* **296**: 497.
2. Hacein-Bey-Abina, S, Von Kalle, C, Schmidt, M, McCormack, MP, Wulffraat, N, Leboulch, P *et al.* (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**: 415–419.
3. Nienhuis, AW, Dunbar, CE and Sorrentino, BP (2006). Genotoxicity of retroviral integration in hematopoietic cells. *Mol Ther* **13**: 1031–1049.
4. Modlich, U, Kustikova, O, Schmidt, M, Rudolph, C, Meyer, J, Li, Z *et al.* (2005). Leukemias following retroviral transfer of multidrug resistance 1 are driven by combinatorial insertional mutagenesis. *Blood* **105**: 4235–4246.
5. Kustikova, OS, Fehse, B, Modlich, U, Yang, M, Düllmann, J, Kamino, K *et al.* (2005). Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. *Science* **308**: 1171–1174.
6. Modlich, U, Bohne, J, Schmidt, M, von Kalle, C, Knoss, S, Schambach, A *et al.* (2006).
Cell-culture assays reveal the importance of retroviral vector design for insertional genotoxicity. *Blood* **108**: 2545–2553.
7. Du, Y, Jenkins, NA and Copeland, NG (2005). Insertional mutagenesis identifies genes
that promote the immortalization of primary bone marrow progenitor cells. *Blood* **106**: 3932–3939.
8. Du, Y, Spence, SE, Jenkins, NA and Copeland, NG (2005). Cooperating cancer-gene identification through oncogenic-retrovirus-induced insertional mutagenesis. *Blood* **106**: 2498–2505.
9. Seggewiss, R, Pittaluga, S, Adler, RL, Guenaga, FJ, Ferguson, C, Pilz, I *et al.* (2006). Acute myeloid leukemia associated with retroviral gene transfer to hematopoietic progenitor cells of a rhesus macaque. *Blood* **107**: 3865–3867.
10. Ott, MG, Schmidt, M, Schwarzwaelder, K, Stein, S, Siler, U, Koehl, U *et al.* (2006). Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat Med* **12**: 401–409.
11. Shou, Y, Ma, Z, Lu, T and Sorrentino, BP (2006). Unique risk factors for insertional mutagenesis in a mouse model of XSCID gene therapy. *Proc Natl Acad Sci USA* **103**: 11730–11735.
12. Montini, E, Cesana, D, Schmidt, M, Sanvito, F, Ponzoni, M, Bartholomae, C *et al.*(2006). Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. *Nat Biotechnol* **24**: 687–696.

13. Kim, DW, Uetsuki, T, Kaziro, Y, Yamaguchi, N and Sugano, S (1990). Use of the human elongation factor 1 alpha promoter as a versatile and efficient expression system. *Gene* **91**: 217–223.
14. Salmon, P, Kindler, V, Ducrey, O, Chapuis, B, Zubler, RH and Trono, D (2000). High-level transgene expression in human hematopoietic progenitors and differentiated blood lineages after transduction with improved lentiviral vectors. *Blood* **96**: 3392–3398.
15. Schambach, A, Bohne, J, Chandra, S, Will, E, Margison, GP, Williams, DA *et al.* (2006). Equal potency of gammaretroviral and lentiviral SIN vectors for expression of O6-methylguanine-DNA methyltransferase in hematopoietic cells. *Mol Ther***13**: 391–400.
16. Wahlers, A, Zipfel, PF, Schwieger, M, Ostertag, W and Baum, C (2002). *In vivo* analysis of retroviral enhancer mutations in hematopoietic cells: SP1/EGR1 and ETS/GATA motifs contribute to long terminal repeat specificity. *J Virol* **76**: 303–312.
17. Wakabayashi-Ito, N and Nagata, S (1994). Characterization of the regulatory elements in the promoter of the human elongation factor-1*α* gene. *J Biol Chem* **269**: 29831–29837.
18. Baum, C, Itoh, K, Meyer, J, Laker, C, Ito, Y and Ostertag, W (1997). The potent enhancer activity of SFFVp in hematopoietic cells is governed by a binding site for Sp1 in the upstream control region and by a unique enhancer core creating an exclusive target for PEBP/CBF. *J Virol* **71**: 6323–6331.
19. Wahlers, A, Kustikova, O, Zipfel, PF, Itoh, K, Koester, M, Heberlein, C *et al.* (2002). Upstream conserved sequences of mouse leukemia viruses are important for high transgene expression in lymphoid and hematopoietic cells. *Mol Ther* **6**: 313–320.
20. Bell, AC, West, AG and Felsenfeld, G (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**: 387–396.
21. Okey, AB, Harper, PA, Grant, DM and Hill, RP (2005). Chemical and radiation carcinogenesis. In: Tannock, IF, Hill, RP, Bristow, RG and Harrington, L (eds). *The Basic Science of Oncology*. McGraw-Hill: New York. pp. 25–48.
22. Cattoglio, C, Facchini, G, Sartori, D, Antonelli, A, Miccio, A, Cassani, B *et al.* (2007). Hot spots of retroviral integration in human CD34+ hematopoietic cells. *Blood* **110**: 1770–1778.
23. Kustikova, O, Geiger, H, Li, Z, Brugman, MH, Chambers, SM, Shaw, CA *et al.* (2007). Retroviral vector insertion sites associated with dominant hematopoietic clones mark “stemness” pathways. *Blood* **109**: 1897–1907.
24. Robert-Richard, E, Richard, E, Malik, P, Ged, C, de Verneuil, H and Moreau-Gaudry, F (2007). Murine retroviral but not human cellular promoters induce *in vivo* erythroid-specific deregulation that can be partially prevented by insulators. *Mol Ther* **15**: 173–182.
25. Weber, EL and Cannon, PM (2007). Promoter choice for retroviral vectors: transcriptional strength versus trans-activation potential. *Hum Gene Ther* **18**: 849–860.
26. Evans-Galea, MV, Wielgosz, MM, Hanawa, H, Srivastava, DK and Nienhuis, AW (2007). Suppression of clonal dominance in cultured human lymphoid cells by addition of the chs4 insulator to a lentiviral vector. *Mol Ther* **15**: 801–809.
27. Aker, M, Tubb, J, Groth, AC, Bukovsky, AA, Bell, AC, Felsenfeld, G *et al.* (2007). Extended core sequences from the cHS4 insulator are necessary for protecting retroviral vectors from silencing position effects. *Hum Gene Ther* **18**: 333–343.
28. Emery, DW, Yannaki, E, Tubb, J and Stamatoyannopoulos, G (2000). A chromatin insulator protects retrovirus vectors from chromosomal position effects. *Proc Natl Acad Sci USA* **97**: 9150–9155.
29. Baens, M, Noels, H, Broeckx, V, Hagens, S, Fevery, S, Billiau, AD *et al.* (2006). The dark side of EGFP: defective polyubiquitination. *PLoS ONE* **1**: e54.
30. Schambach, A, Bohne, J, Baum, C, Hermann, FG, Egerer, L, von Laer, D *et al.* (2006). Woodchuck hepatitis virus post-transcriptional regulatory element deleted from X protein and promoter sequences enhances retroviral vector titer and expression. *Gene Ther* **13**: 641-645.

31. Morita, S, Kojima, T and Kitamura, T (2000). Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Ther* **7**: 1063–1070.
32. Li, Z, Schwieger, M, Lange, C, Kraunus, J, Sun, H, van den Akker, E *et al.* (2003). Predictable and efficient retroviral gene transfer into murine bone marrow repopulating cells using a defined vector dose. *Exp Hematol* **31**: 1206–1214.
33. Schambach, A, Mueller, D, Galla, M, Verstegen, MM, Wagemaker, G, Loew, R *et al.* (2006). Overcoming promoter competition in packaging cells improves production of self-inactivating retroviral vectors. *Gene Ther* **13**: 1524–1533.
34. Schambach, A, Galla, M, Maetzig, T, Loew, R and Baum, C (2007). Improving transcriptional termination of self-inactivating gamma-retroviral and lentiviral vectors. Mol Ther 15: 1167–1173

Supplementary figure 1

Similar to 32D cells, 293T cells show a significantly greater enhancer activity of the retroviral promoters in comparison with the cellular promoters.



Supplementary figure 2

Similar to 32D cells (cf. to figure 1), 293T cells show a significantly greater enhancer activity of the retroviral promoters in comparison with the cellular promoters.

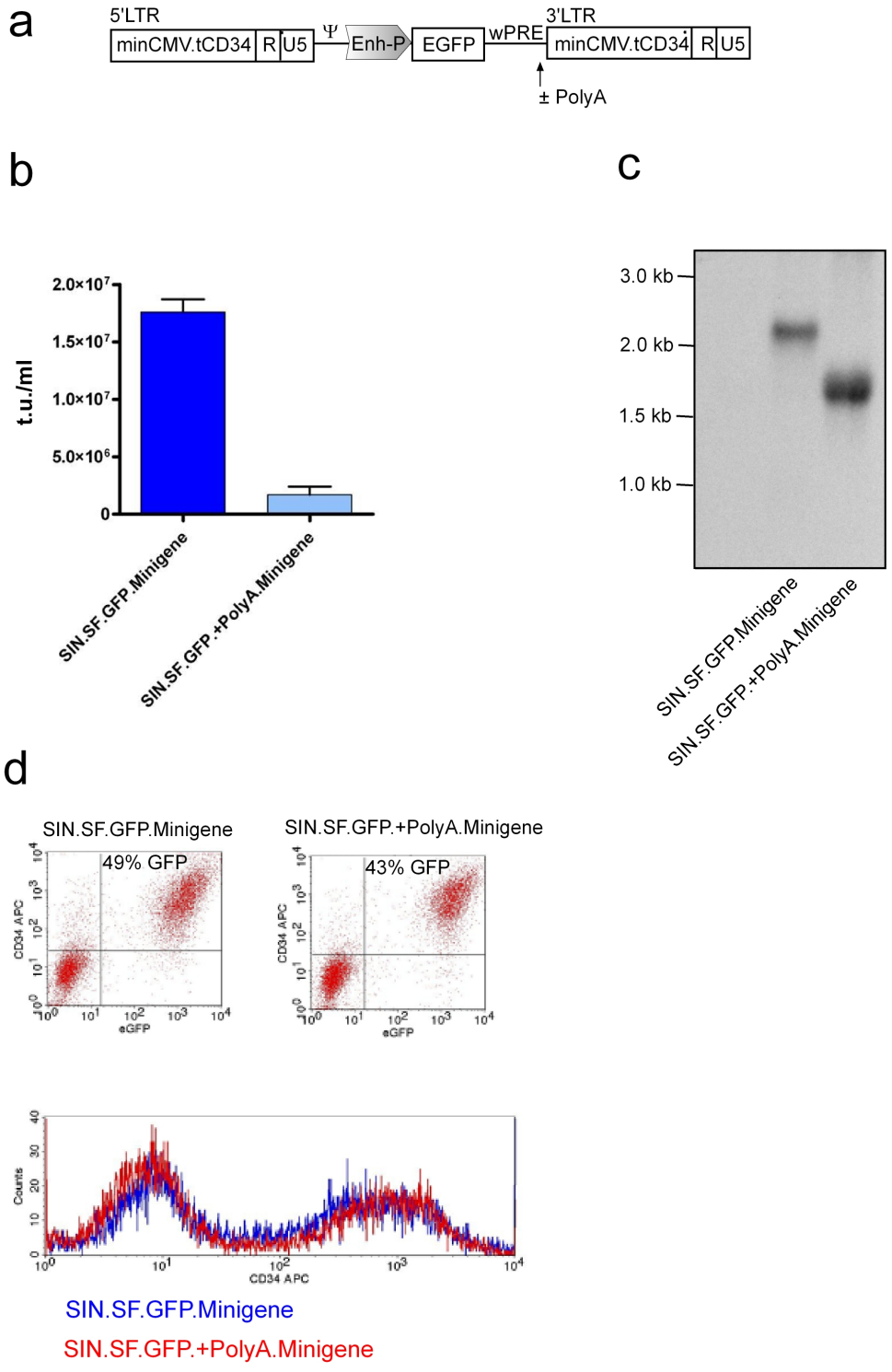


Figure legend:
a) Depiction of the integrated minigene vector where an additional poly(A) site is introduced upstream of the 3'LTR. b) Compared to the minigene construct lacking the additional poly(A) site the titer is about one log decreased. C) Northern blot analysis of total RNA from transduced cells displaying the two constructs. The transcriptional polyadenylation occurs nearly to 100% at the additional poly(A) site.
d) FACS analysis of the enhanced minigene cassette in transduced SC-1 cells 5 days after transduction.

D Publication 2

Murine Leukemia Virus Regulates Alternative Splicing through Sequences Upstream of the 5'Splice Site

Janine Kraunus, Daniela Zychlinski, Tilman Heise, Melanie Galla, Jens Bohne*, and Christopher Baum*

Published in Journal of biological chemistry, 2006 Oct; Vol. 28 p. 37381–37390

DOI 10.1074/jbc.M601537200

http://www.jbc.org/cgi/content/full/281/49/37381?view=long&pmid=17038324

* corresponding author:

Christopher Baum,
Department of Experimental Hematology, Hannover Medical School,
Carl-Neuberg-Straße 1, 30625 Hannover,Germany.
E-mail: baum.christopher@mh-hannover.de

Jens Bohne,
Department of Experimental Hematology, Hannover Medical School,
Carl-Neuberg-Straße 1, 30625 Hannover,Germany.
E-mail: bohne.jens@mh-hannover.de

Author's contribution statement:

Janine Kraunus constructed the SF91 mutant vectors shown in Figs.2-5, performed the experiments in Figs.2-5 and participated in writing the paper. Daniela Zychlinski constructed the SF91 sm3 and sm3comp mutants, performed the experiment in Fig.7 and participated in performing the experiments for Fig.3C. Tilman Heise contributed and labelled the GAPDH RNA probe used for all fractionation experiments. Melanie Galla constructed the SF91 sm1 and sm2 mutants and participated in performing the experiments in Fig.3C. Jens Bohne cloned the proviral mutants and performed the experiment in Fig.6. Both corresponding authors Jens Bohne and Christopher Baum designed the whole study. Both critically evaluated and analyzed the data and wrote the paper.

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 281, NO. 49, pp. 37381–37390, December 8, 2006

Murine Leukemia Virus Regulates Alternative Splicing through Sequences Upstream of the 5′ Splice Site*[S]

Received for publication, February 17, 2006, and in revised form, October 10, 2006 Published, JBC Papers in Press, October 11, 2006, DOI 10.1074/jbc.M601537200

Janine Kraunus[‡§¶], Daniela Zychlinski[‡], Tilman Heise[§], Melanie Galla[‡], Jens Bohne[‡1], and Christopher Baum[‡‖2]

From the [‡]Department of Experimental Hematology, Hannover Medical School, D-30625 Hannover, Germany, the [§]Department of General Virology, Heinrich-Pette-Institute D-20251 Hamburg, Germany, the [¶]Bone Marrow Transplantation, University Hospital Eppendorf, D-20246 Hamburg, Germany, and the [‖]Division of Experimental Hematology, Cincinnati Children's Hospital, Cincinnati, Ohio 45229

Alternative splicing of the primary transcript plays a key role in retroviral gene expression. In contrast to all known mechanisms that mediate alternative splicing in retroviruses, we found that in murine leukemia virus, distinct elements located upstream of the 5′ splice site either inhibited or activated splicing of the genomic RNA. Detailed analysis of the first untranslated exon showed that the primer binding site (PBS) activates splicing, whereas flanking sequences either downstream or upstream of the PBS are inhibitory. This new function of the PBS was independent of its orientation and primer binding but associated with a particular destabilizing role in a proposed secondary structure. On the contrary, all sequences surrounding the PBS that are involved in stem formation of the first exon were found to suppress splicing. Targeted mutations that destabilized the central stem and compensatory mutations of the counter strand clearly validated the concept that murine leukemia virus attenuates its 5′ splice site by forming an inhibitory stem-loop in its first exon. Importantly, this mode of splice regulation was conserved in a complete proviral clone. Some of the mutants that increase splicing revealed an opposite effect on translation, implying that the first exon also regulates this process. Together, these findings suggest that sequences upstream of the 5′ splice site play an important role in splice regulation of simple retroviruses, directly or indirectly attenuating the efficiency of splicing.

A characteristic feature of all retroviruses is the process of reverse transcription of the RNA genome into double-stranded DNA. Following integration into the host genome, the proviral DNA functions as one expression unit, which is transcribed by cellular RNA polymerase II, yielding a single polycistronic primary transcript that serves as genomic RNA for progeny virus. Productive infection and formation of new retroviral particles require the well balanced expression of all viral genes. This is accomplished by a combination of alternative splicing (intron retention) and regulated nuclear export of the primary transcript on the RNA processing level and proteolytic cleavage and translational read-through on the post-translational level (reviewed in Refs. 1–4).

The genomic organization of all retroviruses is similar (Fig. 1*A*). The *gag-pol* open reading frame (ORF)[3] encoding the inner structural proteins (Gag), and the replication enzymes (Pol) is located in the 5′ half of the transcript and expressed from the unspliced genomic RNA after nuclear export. The *gag-pol* ORF in all primary retroviral transcripts is defined as an intron through the presence of a preceding 5′ splice site (ss) in the 5′-untranslated region and a functional 3′ss located toward the end of the polymerase ORF (Fig. 1*A*). To express the glycoproteins (Env), which are encoded in the 3′ half of the genomic RNA, the *gag-pol* ORF is removed by a single splice event for subsequent export of the fully spliced RNA (1, 3, 4). This one-splice event strategy creates the challenge to export intron-containing RNAs, which is typically not supported by the cell and rather results in nuclear retention and degradation of the respective RNA (5, 6). For export of their unspliced RNA, retroviruses make use of constitutive transport elements as exemplified by Mason-Pfizer monkey virus or trans-acting factors as illustrated by HIV (7–9). Murine leukemia virus (MLV), a paradigmatic gammaretrovirus, supports the export of unspliced mRNA by a yet unknown mechanism involving the so-called R region stem-loop (RSL) formed by the cap-proximal 28 bases (10).

As simple retroviruses encode no *trans*-acting regulators of gene expression, alternative or inefficient splicing must be regulated entirely through *cis*-acting RNA motifs and cellular cofactors. Such motifs may include non-consensus 3′ss, decoy 5′ss, and splice modulatory sequences such as splicing enhancers and silencers (11, 12, 16, 25, 34, 35, 41).

In MLV, the 5′ss matches to almost 100% the cellular consensus sequence (Fig. 1*B*) and splices to a 3′ss within the *pol* reading frame (13) (Fig. 1*B*). The 3′ss misses two relatively

* The work was supported by Deutsche Forschungsgemeinschaft Grant 1837/Ba4 and the European Union (INHERINET and CONSERT grants). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains a supplemental table.

[1] To whom correspondence may be addressed: Dept. of Experimental Hematology, Hannover Medical School, Carl-Neuberg-Str. 1, D-30625 Hannover, Germany. Tel.: 49-511-532-6067; Fax: 49-511-532-6068; E-mail: bohne.jens@mh-hannover.de.

[2] To whom correspondence may be addressed: Dept. of Experimental Hematology, Hannover Medical School, Carl-Neuberg-Str. 1, D-30625 Hannover, Germany. Tel.: 49-511-532-6067; Fax: 49-511-532-6068; E-mail: baum.christopher@mh-hannover.de.

[3] The abbreviations used are: ORF, open reading frame; PBS, primer binding site; MLV, murine leukemia virus; ss, splice site(s); HIV, human immunodeficiency virus; RSL, R region stem-loop; PPT, polypyrimidine tract; GFP, green fluorescent protein; eGFP, enhanced GFP; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; CytC, cytochrome *c* oxidase II; pol, polymerase.

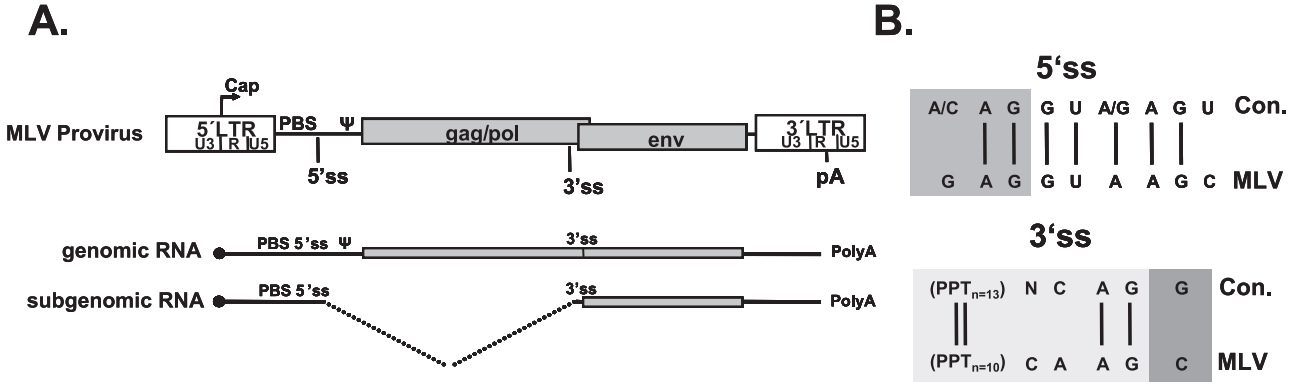


FIGURE 1. **Genomic organization, RNA species, and splice sites of MLV.** *A*, MLV proviral genome and mRNA species. The genome is flanked by the long terminal repeats (*LTRs*, consisting of U3, R, and U5). The ORFs are shown as *gray boxes*. The transcription start site (*cap*) and the polyadenylation site (*pA*) mark the length of the genomic RNA. In addition, the PBS, the packaging signal (Ψ), and the splice sites (5′ss/3′ss) are shown. Below the mRNA species are depicted: the genomic RNA corresponding to the full-length unspliced primary transcript and the spliced, subgenomic RNA. *B*, comparison of cellular and MLV consensus (*Con.*) 5′ and 3′ss. *Dark to light gray* marks the exon/intron junction, and *vertical lines* show matched nucleotides. The length of the PPT is indicated.

important nucleotides surrounding the AG. The polypyrimidine tract (PPT), a second key feature of all 3′ss, is of suboptimal length (10 nucleotides long when compared with 13 residues in average) but not interrupted by weakening purines as most PPTs of HIV (14). Interestingly, a recent study showed that in the context of MLV-based retroviral vectors, the 3′ss can be replaced by very efficient counterparts derived from the human EF1α gene (15). Although the infectious titer was reduced by about 1 order of magnitude, unspliced genomic RNA was still formed, arguing for the existence of additional splice inhibitory sequences.

Recently, we have observed that self-inactivating MLV vectors display complete splicing of the retroviral intron when the promoter was placed in between the primer binding site (PBS) and the 5′ss (16). Using a number of mutants, we here demonstrate that distinct sequence elements upstream of the 5′ss either promote or inhibit splicing. Interestingly, this regulation correlates with the stability of the RNA secondary structure proposed by Mougel *et al.* (17). These data suggest a novel mechanism of splice regulation in retroviruses, directly or indirectly regulating the activity of the 5′ss through upstream sequences.

EXPERIMENTAL PROCEDURES

Plasmids—The retroviral vector plasmids were derived from pSF91 (18). Mutant vectors lacking the U5 region of the 5′ long terminal repeat located 70–145 bp downstream of the cap site (pSF91delU5) or lacking the PBS located 146–163 bp downstream of the cap site (pSF91delPBS) were generated in a three-fragment ligation using PCR-based deletion strategies (primers are provided in the supplemental table). Using pSF91 as a template, two separate PCR products were generated. For the deletion of U5, the forward primer (forward, 5′-CAG ATG GTC CCC AGA TGC-3′, position −150 regarding cap) was used in combination with a reverse primer immediately upstream of U5. The 3′ primer (reverse, 5′-ACG CTG AAC TTG TGG CCG-3′) of the 3′ PCR product is located downstream of an NcoI site (+706 regarding cap). This primer was used in combination with a forward primer annealing just downstream of the U5 region. The two products were then digested with XbaI (5′ fragment) and NcoI (3′ fragment), leaving a blunt end in the middle. The two products were ligated into XbaI/NcoI digested pSF91. An analogous strategy was used for the deletion of the PBS.

All other vector plasmids including deletion, antisense, and other mutants were derived from pSF91 by overlap PCR. Again, pSF91 was used as template. The outer 5′ and 3′ primers are mentioned above. The inner primers bridge the deletion (in the case of pSF91del33–182: sense primer, 5′-AGT CGC CCG G-GA CCA CCG ACC CCC CCG CCG-3′; antisense primer, 5′-GTC GGT GGT C-CC GGG CGA CTC AGT CAA TCG-3′, overlap is underlined, hyphen marks the deletion) or carry the suitable point mutations (in the case of pSF91mlvPBS: sense primer, 5′-CAT TTG G**G**G G**C**T C**GT** CCG **G**GA TTT GGA GAC CCC TG-3′; antisense 5′-CCA AAT C**C**C GG**A** **C**GA **G**CC **C**CC AAA TGA AAG ACC CC-3′, overlap is underlined, mutations are bold). The overlapping PCR product was digested with XbaI and NcoI and ligated into the pSF91 backbone. For the deletions or mutations named sm1, sm2, sm3, and sm3comp, a different 3′ primer (5′-AAT GGG CCA CAA AAC GGG CCC CCG A-3′) was used, including an ApaI site. The final overlap PCR product was then cloned via XbaI and ApaI. All used primers (FW, forward; RV, reverse) are listed in the supplemental table, indicating name, sequence, and final vector construct. To transfer the PBS and the deletion 3′PBS to a complete proviral clone, an EcoRI/PstI fragment of pMOVGFP (kind gift from B. Schnierle, Paul-Ehrlich-Institute, Langen, Germany) was subcloned into pBluescript (Stratagene). Within this subclone, the AflII/PstI fragment (−464 to +563, regarding the cap site as +1) was exchanged to the corresponding pSF91 sequence and to that of SF91delPBS and SF91del3′PBS via AflII and PstI, yielding pMOVSFGFP, pMOVSFdelPBSGFP, and pMOVSFdel3′PBSGFP, respectively. Correct deletions or nucleotide replacements were confirmed by sequencing.

Cells, Transfections, and Reporter Assays—293T cells were grown in Dulbecco's modified Eagle's medium (Biochrom, Berlin, Germany) supplemented with 10% fetal calf serum, 2 mM glutamine, and 1 mM sodium pyruvate including antibiotics. 7×10^5 293T cells/well were seeded in a 6-well plate. For trans-

fection, the medium was exchanged, and 25 μM chloroquine (Sigma, Taufkirchen, Germany) was added. Retroviral vector DNA (0.9 μg) was transfected using the calcium phosphate precipitation method (19). Medium was exchanged 6 h after transfection, and the cells were harvested after 48 h. Transfection efficiencies ranging between 60 and 80% and protein expression were assessed by flow cytometry in a FACScalibur (BD Biosciences, Heidelberg, Germany) using CellQuest software (BD Biosciences).

Retroviral vector particles were produced by cotransfection of 0.9 μg of retroviral plasmid pSF91 with expression plasmids for MLV *gag-pol* (1.5 μg) and ecotropic envelope (0.3 μg) into 7×10^5 Phoenix GP (G. Nolan, Stanford University, Palo Alto, CA) cells. In the case of pSF91artPBS, plasmids coding for artificial tRNAs (tRNA-x2-Lys, tRNA-x2-Pro (20)) were co-transfected. Supernatants containing the viral particles were collected 48 h after transfection, filtered through a 0.22-μm filter, and used to transduce 1×10^5 target cells in serial dilutions for titer determination. Transduction was assisted by adding 4 μg/ml protamine sulfate and centrifugation for 60 min at 400 $\times$ *g* and 25–32 °C. Cells were grown for another 2 days before the percentage of enhanced green fluorescent protein (eGFP)-positive cells was determined by flow cytometry. Further analysis was limited to those experiments where less than 30% of target cells were productively transduced.

RNA Preparation and Northern Blot—For preparation of nuclear and cytoplasmic RNA, 8×10^6 cells were collected 48 h after transfection and treated according to the protocol of Weil *et al.* (21). Briefly, the cells were resuspended in 500 μl of Nonidet P-40 lysis buffer (0.5% Nonidet P-40, 0.14 M NaCl, 10 mM Tris, pH 8.4, 1.5 mM $MgCl_2$, 10 mM EDTA, pH 8.0) for 5 min at 0 °C. After centrifugation at 470 $\times$ *g* for 5 min at 4 °C, the supernatant containing the cytoplasmic fraction was harvested. The nuclear pellet was washed twice with lysis buffer. RNA was extracted from total cells or nuclear and cytoplasmic fractions using the RNA Instapure reagent according to the manufacturer's protocol (Eurogentec, Brussels, Belgium).

For Northern blot, 5–10 μg of RNA were separated at 2 V/cm in 1% agarose gels after denaturing RNA samples with glyoxal (6%) and Me_2SO (50%). Subsequently, RNAs were transferred to Biodyne B membrane (0.45 μm, Pall) by capillary transfer and UV cross-linked (Stratalinker, Stratagene). Specific probes used for hybridization corresponded to the cDNAs of eGFP, GAPDH, cytochrome *c* oxidase II, and the env fragment of MLV. Probes (25 ng) were radiolabeled using the Prime-It II kit (Stratagene, Amsterdam, The Netherlands) to an activity of at least 5×10^8 cpm/μg and separated from unincorporated nucleotides on spin columns (Molecular probes, Göttingen, Germany). DNA template used for the *in vitro* transcription to generate antisense RNA probes specific for the GAPDH intron B was raised by PCR on genomic DNA using reverse primer 5′-GGA CTA GT**T AAT ACG ACT CAC TAT A**GG GTG CGG TGG AGA TCT G-3′ containing the T7 RNA polymerase promoter sequence (shown in bold) and forward primer 5′-CAA GGA GAG CTC AAG GTC-3′. Transcription reactions were carried out with 0.5 μl of PCR product in a final volume of 20 μl in transcription buffer (Promega, Mannheim, Germany) containing 0.31 mM ATP, CTP, and GTP, 0.25 μM UTP, 5.0 μM [α-^{32}P]UTP (800 Ci/mmol; Hartmann Analytic, Braunschweig, Germany), 5 mM dithiothreitol, 20 units of RNasin (Promega), and 20 units of T7 RNA polymerase (Promega). The reaction was terminated by adding 10 μg of yeast tRNA and 1 unit of DNaseI (Promega) and incubated for 15 min at 37 °C. Unincorporated nucleotides were removed as above. Hybridization solutions had a final activity of 10^6 cpm/ml. Membranes were washed, sealed, and exposed to x-ray films (Kodak X-omat-AR, Kodak, Stuttgart, Germany) or quantified by PhosphorImager analysis (Fuji, Düsseldorf, Germany; Amersham Biosciences, Freiburg, Germany) analysis.

RESULTS

The MLV-derived Vector SF91 Shows Balanced Splicing—The MLV-based vector SF91 (18) (Fig. 2*A*) was used to investigate the potential splice regulatory role of sequences located upstream of the 5′ss, namely R, U5, PBS, and a short region downstream of PBS (here referred to as 3′PBS). The SF91 vector is derived from MLV by deleting the *gag-pol* ORF and introducing the *env* 3′ss including its PPT and branch point sequences 3′ of the packaging signal (Ψ) followed by the original *gag* ATG, which allows translation of a reporter gene such as eGFP. Importantly, the vector SF91 contains a 460-bp spliceable intron in the 5′-untranslated region, which uses the same 5′ss and 3′ss as the proviral gag-pol intron. For a detailed analysis of splicing, vector plasmids were transiently transfected into human 293T cells. These were harvested 48 h after transfection to prepare nuclear and cytoplasmic RNA. 5 or 10 μg of each RNA fraction were analyzed by Northern blotting.

The vector SF91 showed balanced splicing with a predominance of the unspliced RNA in the nucleus (Fig. 2*B*, *lane 2*, *left panel*), whereas the cytoplasmic compartment revealed an accumulation of spliced message, most likely caused by more efficient export of processed RNA (Fig. 2*B*, *lane 2*, *right panel*). Due to this imbalance in the export kinetics of spliced and unspliced RNAs, we investigated both RNA fractions, assuming that mutants with inhibited splicing would be underrepresented in total RNA and in the cytoplasmic fraction. To control the RNA fractionation procedure, two probes were devised. The first recognized intron B of GAPDH (Fig. 2*B*, *Intron*), whose signal should be confined to the nuclear fraction. Using the intron probe, we detected two specific bands at 6.8 kb and about 13 kb in the nucleus, corresponding to the pre-mRNA (NCBI accession number NC_000012.10) and a longer transcript, which was observed previously (22); these transcripts were much weaker or absent in the cytoplasm (Fig. 2*B*, *right panel*). Another band at 5 kb was detectable in both nuclear and cytoplasmic fractions (Fig. 2*B*). The size matched to the large ribosomal RNA (28 S), and since this was also seen with other probes, we regarded it as unspecific. The second probe was a fragment of cytochrome *c* oxidase II (CytC), a gene exclusively transcribed in the mitochondria, thus serving as a cytoplasmic marker. Due to incomplete removal of mitochondria from the nucleus, the nuclear fraction was never fully devoid of CytC RNA (Fig. 2*B*, *left panel*). In summary, we established a simplified system to study retroviral splice regulation and nuclear export of MLV RNA.

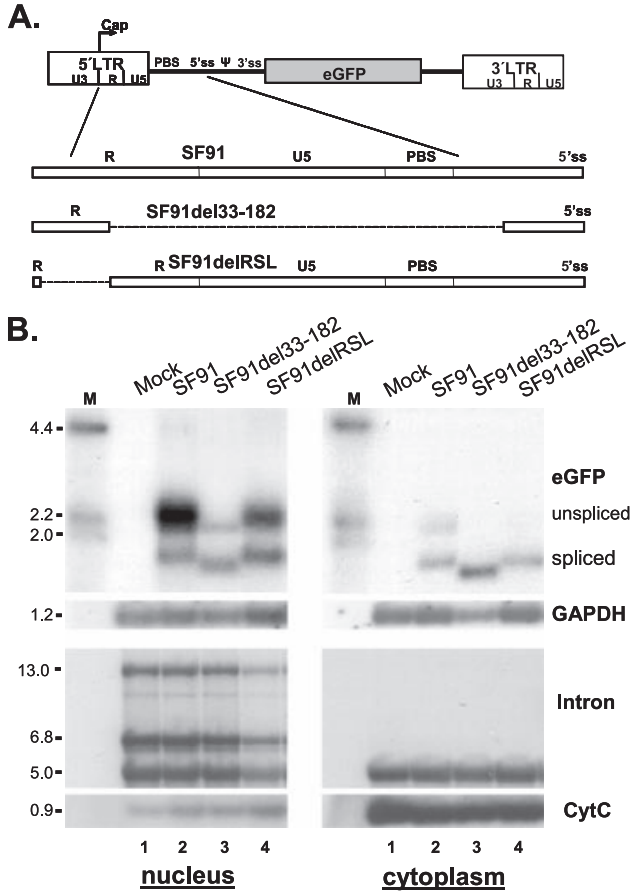


FIGURE 2. **Sequences of the untranslated first exon change the splicing pattern of the MLV-derived vector SF91.** *A*, in SF91 (*top*), coding sequences from MLV are replaced with eGFP using the original gag ATG site as the translational start site. The viral 3′ss, including branch point and PPT, is cloned upstream of eGFP. Thereby, a shortened intron comprising 460 nucleotides harboring identical splices sites as the MLV provirus is reconstituted. The region of the untranslated first exon (cap site to 5′ss) is enlarged below. The mutant SF91del33–182 (*middle*) lacks the bases +33 to +182 regarding cap as +1. The mutant SF91delRSL lacks the bases of the RSL that has been shown to be involved in accumulation of unspliced RNA in the cytoplasm (10). *LTR*, long terminal repeat. *B*, Northern blots from fractionated lysates of transfected 293T cells, prepared as described under "Experimental Procedures." The blots were hybridized with a radiolabeled eGFP-specific probe (*upper panel*). Rehybridization with a GAPDH-specific probe served as loading control (*second panel*). The efficiency of fractionation was checked with a probe specific for an intron of GAPDH (*Intron*) as a nuclear marker (*third panel*) and with a cytochrome *c* oxidase II-specific probe (*CytC*) as cytoplasmic marker (*lower panel*). The unspliced and spliced RNAs are identified on the *right*. Molecular mass standards (*M*) in kb are shown on the *left*. *Mock*, mock-transfected.

Distinct Elements in the Untranslated First Exon Regulate Splicing and RNA Export—Next we focused on the region upstream of the 5′ss and its role in splice regulation. Following earlier experiments (16), we deleted bases +33 to +182 (with the cap site being +1) upstream of the 5′ss, resulting in the vector SF91del33–182 (Fig. 2*A*). When analyzed in parallel with SF91, SF91del33–182 showed an increased splicing efficiency in the nuclear fraction (Fig. 2*B*, *lane 3*, *left panel*). Accordingly, the cytoplasm contained almost exclusively spliced RNA, whereas unspliced RNA could only be detected after prolonged exposure (data not shown). This experiment revealed that the deleted sequences, located upstream of the 5′ss, play a major role in retroviral RNA processing. This rather large deletion also seems to lower the RNA level (Fig. 2*B*, *lanes 2* and *3*, *left panel*). In contrast, smaller deletions that also increased the efficiency of splicing did not affect RNA levels (see below).

To examine the role of the RSL, previously shown to promote the appearance of unspliced RNA in the cytoplasm (10), we constructed another vector lacking the bases 4–31 of the R region (SF91delRSL, Fig. 2*A*) and thus the entire RSL. When compared with SF91, deletion of the RSL enhanced nuclear splicing by 2.5-fold (as determined by PhosphorImager analysis). In addition, deleting the RSL inhibited the accumulation of unspliced RNA in the cytoplasm (Fig. 2*B*, *lanes 2* and *4*). The ratio of spliced/unspliced RNA in the cytoplasm was 2.3 for SF91 and changed to 8.6 for SF91delRSL (3.8-fold enhancement). These data confirm that the RSL is involved in RNA export (10); moreover, our data suggest that the RSL participates in balancing the retroviral splice reaction.

Splicing Is Blocked in Vectors Lacking the PBS—To study the role of the remaining sequences in the untranslated first exon in more detail, we deleted parts of this region separately. The vectors as shown in Fig. 3*A* were SF91delR_2 (R deletion 3′ of the RSL), SF91delU5 (deletion of U5), SF91delPBS (deletion of the PBS), and SF91del3′PBS (deletion of bases +164 to +182, immediately downstream of the PBS). Following transient transfection of 293T cells and fractionation of nuclear and cytoplasmic RNA, Northern blots were performed as described above. Deletions of R_2, U5, or 3′PBS produced similar amounts of total RNA but strongly enhanced the accumulation of spliced message already in the nucleus (Fig. 3*B*, *lanes 3*, *4*, and *7*, *left panel*). The overloaded *lanes 4* and *7* revealed that unspliced RNA was still detectable with these deletion mutants. In contrast, splicing was nearly blocked when deleting the PBS (Fig. 3*B*, *lane 6*, *left panel*). Another deletion mutant with the combined deletion of PBS sequences and downstream sequences +164 to +182 (SF91delPBS-3′PBS, Fig. 3*A*) demonstrated a similar phenotype as SF91del3′PBS, in which only the bases +164 to +182 were deleted (Fig. 3*B*, compare *lanes 5* and *7*). Therefore, the latter mutation was dominant over the deletion of the PBS.

For all these mutants except the one with the deleted PBS, we detected only spliced RNA in the cytoplasmic fraction (Fig. 3*B*, *right panel*, compare *lane 6* with *lanes 3*, *4*, *5*, and *7*). This reflected the strength of the nuclear splicing reaction. As overexposure of the blot showed unspliced RNA in these cases (data not shown), a strong effect of the deleted sequences on RNA export was unlikely.

The correctness of the transcripts was proven by RT-PCR and subsequent sequence analysis (data not shown). The additional band observed in the nucleus was refractory to sequence analysis due to PCR-intramolecular hybridization. However, scoring the first untranslated exon in a splice site prediction program (available from the Berkeley Drosophila Genome Project) revealed a cryptic 5′ss at 262 nucleotides (3′ to the major 5′ss, consistent with the observed band). Interestingly, splicing at this site was not inhibited by deletion of the PBS (Fig. 3*B*, *lane 6*, *left panel*). In summary, these data revealed an exceptional

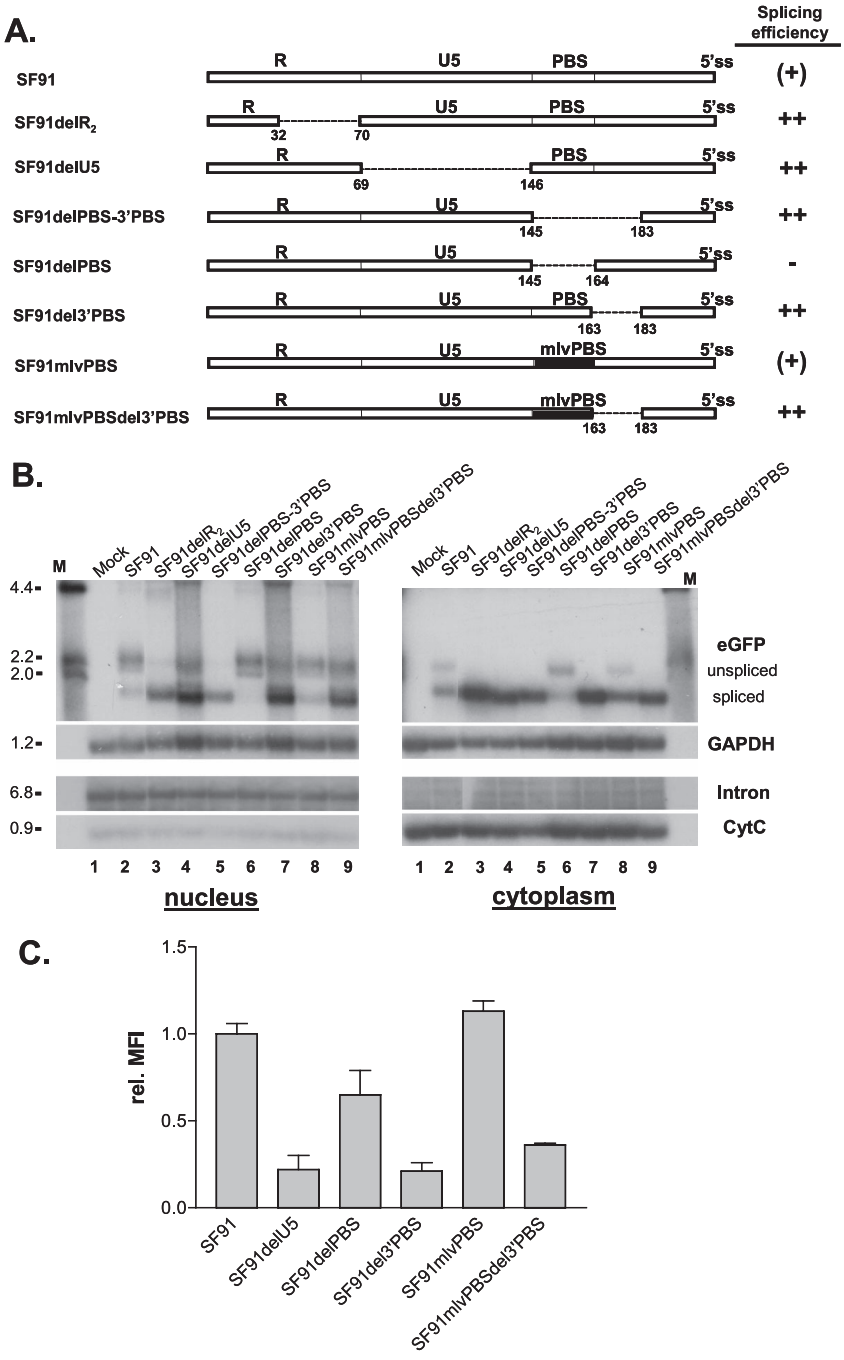


FIGURE 3. **The splice enhancing PBS is surrounded by splice inhibitory elements.** *A*, a schematic illustration of the first untranslated exon (cap to 5′ss) of SF91 (*top*) and the deletion mutants. Sequences are deleted as indicated: SF91delR$_2$ (+33-+69 deleted); SF91delU5 (+70-+145 deleted); SF91delPBS-3′PBS (+146-+182 deleted); SF91delPBS (+146-+163 deleted); SF91del3′PBS (+164-+182 deleted). In SF91mlvPBS and SF91mlvPBSdel3′PBS, the PBS sequence (annealing to tRNA^{Gln} in SF91) is mutated in the PBS sequence original used in MLV for tRNA^{Pro}. On the *right*, splicing efficiency as determined in *B* is shown. *B*, Northern blot analysis of RNA from transfected 293T cells. Using the eGFP-specific probe (*upper panel*), the blots represent the ratio of unspliced and spliced RNA in the nuclear (*left*) and cytoplasmic (*right*) fraction. GAPDH-, intron-, and CytC-specific probes served as controls for loading and fractionation. Molecular mass standards (*M*) in kb are shown on the *left*. *Mock*, mock-transfected. *C*, translation efficiency of the untranslated region mutants as indicated by mean fluorescence intensity (*MFI*) of eGFP determined in unsorted cell pools comprising 10,000 single events. Median and standard deviation of 3–9 independent experiments are given. *rel.*, relative.

role of the PBS being the only element within the first untranslated exon that promotes retroviral splicing from the 5′ss, whereas the surrounding sequences inhibit this process.

Deletion of U5 and 3′PBS Impairs Translational Utilization—Taking advantage of eGFP encoded by the vectors, we also determined the influence of the 5′-untranslated region deletions on the protein level in transfected 293T cells. Although the deletion of U5 and 3′PBS produced almost exclusively spliced RNA in the cytoplasm, the mean fluorescence intensity was markedly decreased (Fig. 3*C*). Both U5 and 3′PBS deletions reduced the mean fluorescence intensity by 80% when compared with the parental vector SF91. Thus, in these deletion mutants, the amount of spliced RNA in the cytoplasm did not correlate with translational utilization (compare Fig. 3*B*, *right panel* with Fig. 3*C*). However, deleting the PBS reduced GFP expression by 40%, which correlated with the reduced amount of spliced RNA. We extended these observations to 293T-based packaging cells, which stably produce MLV Gag-Pol. The presence of Gag did not alter the translational regulation (data not shown). Moreover, identical data were obtained in transfected HeLa cells (data not shown). These data suggest that the sequences involved in splice regulation also contribute to translational control of the retroviral transcript.

The Splice-promoting Effect of the PBS Is Independent of Its Genotype and Primer Binding—To find out whether the role of the PBS in splice regulation is dependent on the type of the corresponding tRNA and is therefore dependent on the PBS sequence or whether primer binding itself regulates splicing, we developed another set of PBS mutants. MLV typically contains a PBS with a specificity for the proline tRNA. Our vectors contain a PBS with a specificity for the glutamine tRNA as this PBS does not inhibit transcription in primitive embry-

onic and hematopoietic stem cells (23, 24). Exchanging the PBS for $tRNA^{Gln}$ used in SF91 to the PBS for $tRNA^{Pro}$ usually found in MLV (five point mutations within the 18 bases of the PBS) yielded SF91mlvPBS (Fig. 3*A*). In the transient transfection assay, SF91 and SF91mlvPBS showed similar splicing patterns, as shown in Fig. 3*B* (*lanes 2* and *8*, *left* and *right panel*). The independence of the kind of PBS and corresponding tRNA in splice regulation was also confirmed by a mutant containing the MLV-PBS followed by a deletion of +164 to +182 (SF91mlvPBSdel3′PBS, Fig. 3*A*). When compared with its SF91 counterpart (SF91del3′PBS), a similar splice alteration was observed in both cases (Fig. 3*B*, compare *lanes 7* and *9*, *left* and *right panel*). These data showed that the dominant splice inhibitory effect of the region 3′ of the PBS is independent of the type of the neighboring PBS and that the experimental system provided highly reproducible results.

Annealing of the tRNA to the PBS takes place during virion assembly (25). However, it is still possible that tRNA precursors, which are present in the nucleus (26), might anneal to the PBS of the pre-mRNA. To rule out any effects of primer tRNA binding in splice regulation, we designed vectors containing an artificial PBS (SF91artPBS), which does not bind any cellular tRNA, adapting the approach developed by Pedersen and colleagues (20), as described in Ref. 27. With this method, it is possible to examine splicing in the absence or presence of a bound tRNA primer. We found that the potential annealing of a matching tRNA did not alter splicing efficiency, strongly suggesting that binding of a tRNA primer is not important for splice regulation (data not shown).

Evidence for a Structural Role of the PBS in Splice Enhancement—The previous experiments imply that the role of the PBS in balanced splicing is not dependent on tRNA specificity or primer binding. To analyze whether the role of the PBS in splice regulation is largely sequence-independent, we developed a mutant that carries the PBS in antisense orientation (SF91asPBS, Fig. 4*A*). Indeed, this construct showed an identical splice phenotype as the wild type SF91 (Fig. 4*B*, compare *lanes 3* and *5*). This was confirmed by quantification of the RNA ratios (Fig. 5). Thus, the phenotype of the PBS deletion, namely a block in splicing, was reverted when introducing the PBS in antisense (Fig. 4*B*, *lanes 4* and *5*, and Fig. 5).

In great contrast to the deletion of the PBS, deleting the 19 bases downstream of the PBS (mutant SF91del3′PBS) strongly enhanced splicing (Fig. 3*B*, compare *lanes 6* and *7*; confirmed in Fig. 4*B*, *lanes 4* and *6*, and quantified in Fig. 5). To investigate whether the role of the sequence downstream of the PBS in splice regulation is dependent on its sequence, we cloned a mutant that contained this region in antisense orientation (SF91as3′PBS; Fig. 4*A*). In this case, the deletion and antisense orientation produced similar phenotypes, *i.e.* increased splicing (Fig. 4*B*, *lanes 6* and *7*; quantified in Fig. 5). Identical results were obtained when transfecting murine SC-1 fibroblasts, revealing that the observed effects are not species-specific (data not shown).

We conclude that the role of the PBS in splice regulation is largely independent of its sequence. Instead, our data suggest that the PBS modulates splicing by forming a spacer within a larger structural framework.

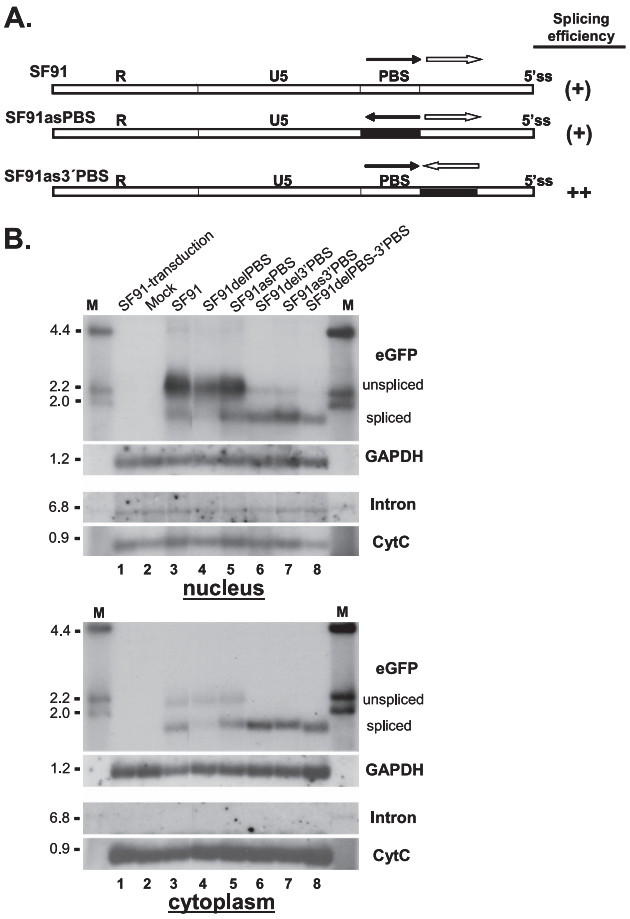


FIGURE 4. **Role of the PBS and downstream regions in RNA processing.** *A*, a drawing of the first untranslated exon of SF91. *Black arrows* (PBS) and *white arrows* (3′PBS) indicate the orientation of the two elements. SF91asPBS (*middle*) carries the PBS sequence in antisense orientation. The other mutant (SF91as3′PBS, *bottom*) carries the 19 bases downstream of the PBS in antisense orientation. The *right column* shows the splicing efficiency as determined in *B*. *B*, Northern blot data were obtained as described before. eGFP-transcript ratios (*upper panel*) of the wild type vector SF91 were directly compared with the transcripts of the PBS deletion (SF91delPBS) and the antisense (SF91asPBS) mutant. Additionally, the deletion mutant of the sequence 3′ of the PBS (SF91del3′PBS) was compared with its antisense counterpart SF91as3′PBS. The *last lane* shows the combined deletion of both regions PBS and 3′PBS. Control probes (*lower panels*) were used as indicated. Molecular mass standards (*M*) in kb are shown on the *left*. *Mock*, mock-transfected.

The Splice Regulation Mediated by the PBS and Adjacent Sequences Also Applies to a Complete MLV Proviral Clone—To test whether the above results apply to an entire proviral genome, we transferred the two deletion mutants with the most prominent phenotypes, namely the PBS deletion (inhibits splicing) and the 3′PBS deletion (increases splicing), into a replication-competent MLV construct called MOVGFP (28). MOVGFP contains the entire MLV proviral genome with the eGFP gene inserted into the proline-rich region of *env*. We cloned the different leader regions also containing a part of the SF91 U3 (SF) into the proviral plasmid MOVGFP yielding a MOVSFGFP and the respective deletion mutants, MOVSFdelPBSGFP and MOVSFdel3′PBSGFP. Based on analysis of total RNA, introducing the SF91 leader into MOVGFP slightly increased splicing of the *env* RNA (Fig. 6, compare *lanes 4* and

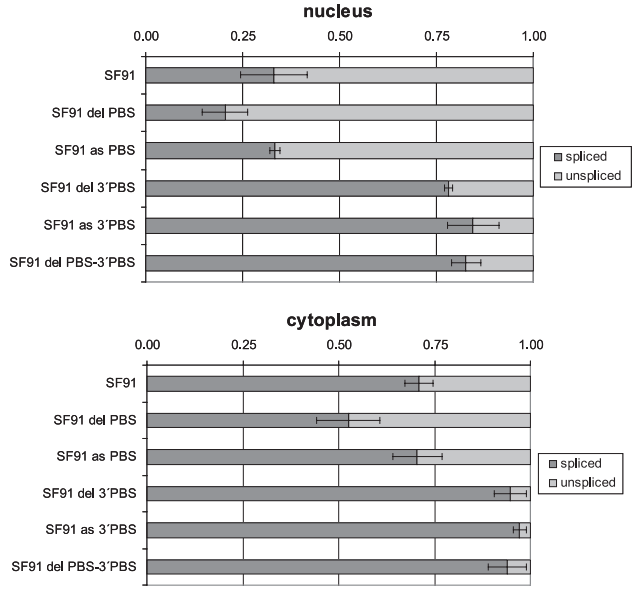


FIGURE 5. **Quantification of the splicing ratios.** Northern blots containing constructs from Figs. 2–4 were quantified using PhosphorImager technology. The relative amounts of spliced and unspliced RNA are shown. Values represent the median of 3–5 independent experiments.

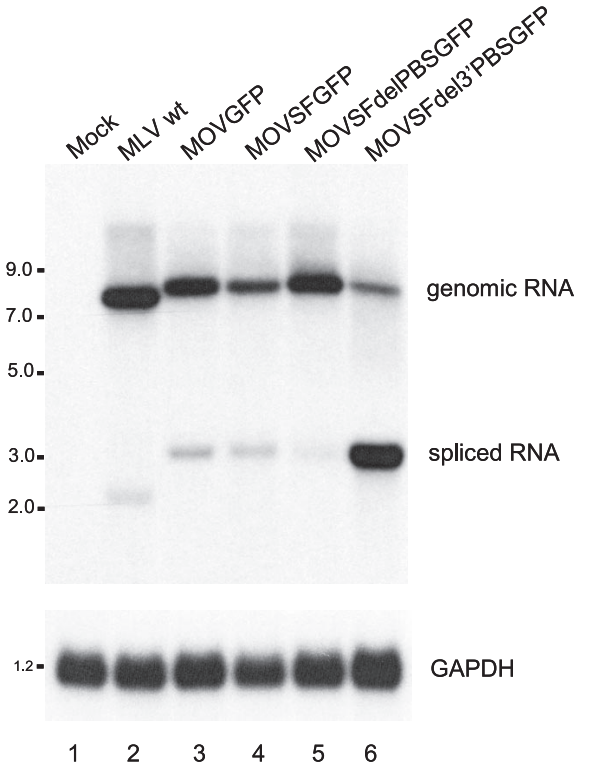


FIGURE 6. **Sequences upstream of the 5′ss regulate splicing in a complete MLV proviral clone.** Northern blot analysis of 293 T cells transfected with an MLV wild type plasmid (*MLV wt*) and a variant carrying the eGFP gene in the proline-rich region (MOVGFP; 28). The SF91 leader (SF) and two deletion mutants, mainly delPBS and del3′PBS, were cloned into MOVGFP, resulting in MOVSFGFP, MOVSFdelPBSGFP, and MOVSFdel3′PBSGFP. 10 μg of total RNA were separated on a denaturing agarose gel. MLVs specific RNAs were detected with a probe corresponding to the env ORF. Rehybridization with a GAPDH-specific probe served as loading control (*bottom panel*). Molecular mass standards in kb are given on the *left*, and the RNA species are identified on the *right*. *Mock*, mock-transfected.

3). Deleting the PBS led to accumulation of unspliced RNA, as predicted by the above studies (Fig. 6, *lane 5*). In contrast, deleting the region 3′ of the PBS severely reduced the amount of genomic RNA and generated much higher levels of spliced *env* RNA (Fig. 6, *lane 6*). As an additional control, an unmodified wild type MLV was used (Fig. 6, *lane 2*). Thus, the data obtained in the context of the subgenomic vector system (Figs. 2–5) were predictive for RNA processing of a complete provirus. The data also show that the distance to the splice acceptor is not a critical parameter of the splice inhibitory effects of the sequences located upstream of the 5′ss.

To test whether the *env* splice acceptor is required in this context, we designed vectors lacking this sequence, the preceding PPT, and the branch point; these constructs still allow cryptic splicing to a minor splice acceptor located in the packaging signal. In this context (data not shown) as well as in the vectors containing the authentic *env* splice acceptor (Fig. 4) or the proviral context shown above (Fig. 6), deleting sequences 3′ of the PBS enhanced splicing. Together, these data reveal that the splice regulation mediated by sequences upstream of the 5′ss is independent of the size of the intron and the type of the splice acceptor. This leaves the 5′ss as the most likely interaction partner of the splice regulatory sequences located in the first untranslated exon.

Detailed Mapping of the Sequence Downstream of the PBS Validates a Structural Model for Splice Regulation—To analyze whether splicing in MLV depends on sequence or structure of the first untranslated exon, we focused on the validated secondary structure (17) and introduced further mutations into the region 3′ to the PBS (Fig. 7*A*). First, the deletion 3′PBS was divided into smaller deletions named sm1 and sm2 (stem mutants 1 and 2). We then analyzed total RNA by Northern blot. The parental construct SF91 was compared with SF91delPBS and SF91del3′PBS (Fig. 7*B*, *lanes 2*, *3*, and *4*, *upper panel*), reproducing the effects on splicing from previous analyses performed with fractionated RNA. The deletion sm1, which compromises the first half of del3′PBS, shifted the balance toward the spliced RNA (Fig. 7*B*, compare *lanes 4* and *5*, *upper panel*). The deletion of the second half of del3′PBS (sm2) displayed the same phenotype as the complete del3′PBS (Fig. 7*B*, compare *lanes 4* and *6*, *upper panel*). Since a potential binding motif for hnRNPA1, a splice repressor, coincides with sm2 (AGGGA), we destroyed it by mutation to ACCGA (sm3; Fig. 7*A*, *boxed motif*). This mutant again showed a similar phenotype as sm2 and del3′PBS (Fig. 7*B*, *lanes 7*, *6*, and *4*), suggesting that potential hnRNPA1 binding at this site is not involved in MLV splice regulation. To ultimately test whether the sequence motif or the secondary structure is important, we introduced compensatory mutations into the corresponding bases on the ascending side of the stem (sm3comp, Fig. 7*A*, *boxed motif*). Importantly, these mutations restored the wild type splicing phenotype (Fig. 7*B*, compare *lanes 8* and *2*). A lower RNA amount for sm3 was only observed in this particular experiment and not in others (Fig. 7*B*, *two extra lanes*). PhosphorImager analysis confirmed that the splice inhibitory effect mainly

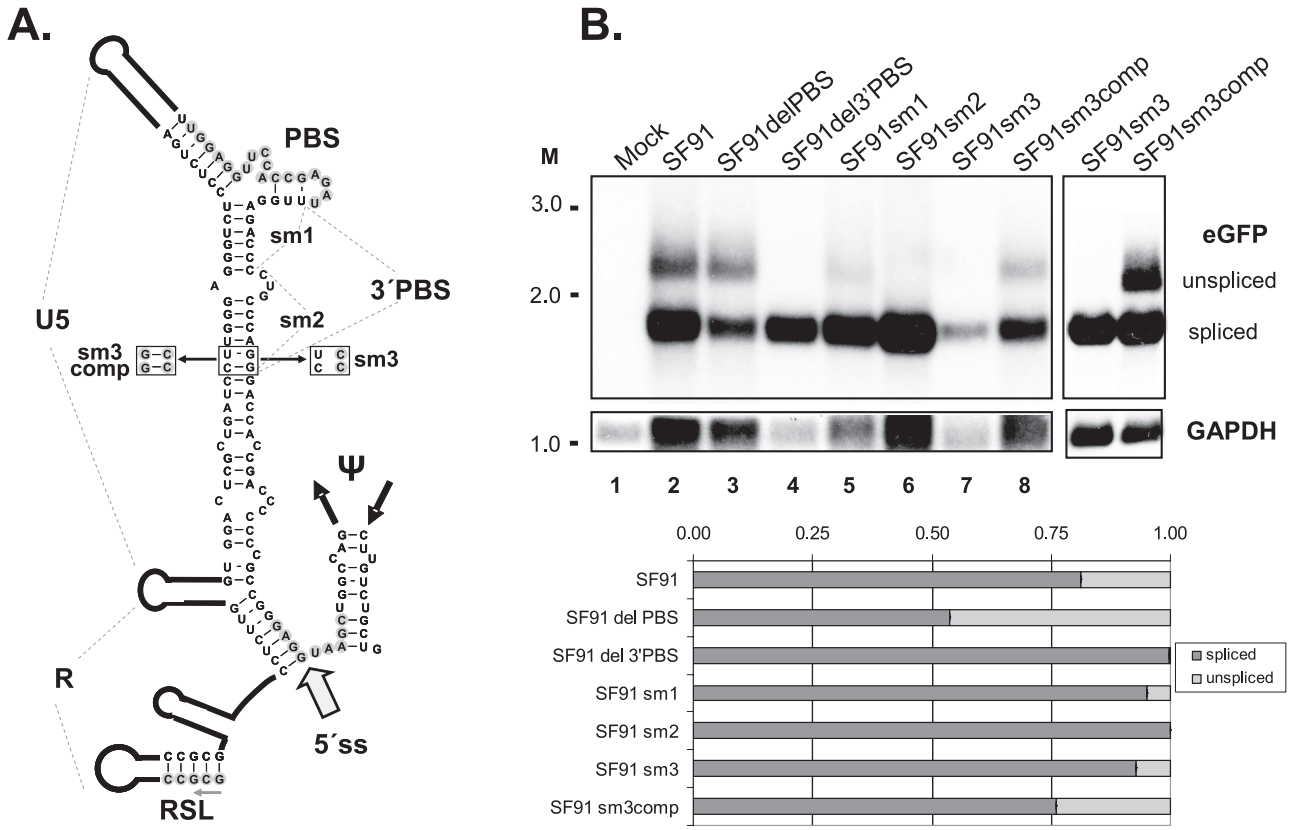


FIGURE 7. **The secondary structure upstream of the 5′ss regulates alternative splicing in MLV.** *A*, a schematic illustration of the secondary structure of the MLV first untranslated exon according to the structure by Mougel *et al.* (17). The RSL (first 32 bases), the U5 region, the PBS, the region downstream of the PBS (3′PBS), and the 5′ss are indicated. The *large gray arrow* points to the invariant GU dinucleotide at the 5′ss also being the first two nucleotides of the intron. Downstream of the 5′ss follows the packaging signal (Ψ). Key sequences are shown in detail; other stems are represented as *thick black lines*, and loops are represented as *thick black half-open circles*. For the RSL, the PBS, the consensus 5′ss, and the mutated nucleotides in sm3/sm3comp, the important nucleotides are emphasized by *gray circles*. *B*, *upper panel*, Northern blot of total RNA from the indicated constructs. RNA species are indicated on the right, and molecular size markers (*M*) are noted in kb on the *left*. A GAPDH-specific probe was used as a loading control. The *two extra lanes* represent a different experiment for the two most important mutations. The *lower panel* shows the quantification of the experiment depicted in the *upper panel* by PhosphorImager analysis. Spliced and unspliced RNA are indicated as in Fig. 5. *Mock*, mock-transfected.

resides in the lower part of the stem (del3′PBS, sm2 and sm3) and that the compensation of sm3 restores the wild type balance (Fig. 7*B*, *lower panel*). By RT-PCR and sequencing, all mutants were shown to splice to the correct 3′ss (data not shown). These findings strongly suggest that alternative splicing in MLV is regulated via a secondary structure upstream of the 5′ss.

DISCUSSION

The present study investigated the impact of sequences located upstream of the 5′ss of MLV, a paradigmatic gammaretrovirus, on retroviral splice regulation. Our data reveal that sequences surrounding and including the PBS are crucial determinants of alternative splicing. These sequences show a higher degree of conservation with other gammaretroviruses (such as gibbon ape leukemia virus, 84% identity) when compared with other regions (such as the active site of reverse transcriptase, 74% identity), whereas the packaging signal shows no homology. We have shown that the PBS plays an exceptional role in MLV splice regulation by counteracting adjacent strong splice inhibitory structures, without the need for primer binding. This points to a previously unknown mechanism of splice regulation in simple retroviruses, possibly also relevant for eukaryotic cellular genes. This mechanism involves sequences located upstream of the 5′ss to regulate alternative splicing, without the need for a virus-encoded *trans*-acting factor.

Interestingly, deleting the putative export element of MLV (RSL) slightly enhanced splicing. However, in the case of SF91, the export efficiency of the unspliced transcript is rather low so that most of the unspliced RNA stays in the nucleus. Therefore, it is not very likely that complete splicing of other mutants can be attributed to a prolonged nuclear retention. In addition, all of these mutants still contain the putative export element.

Regulation of alternative splicing is an essential step in the life cycle of all retroviruses and thus subject to tight evolutionary control. HIV, as a complex retrovirus, and RSV, as an example for a simple retrovirus, are the best studied viruses in this respect. In HIV, all 5′ss match the cellular consensus sequence (29), in contrast to the 3′ss, which are weakened by several means, including short and interrupted PPTs (14). Besides, the exons and introns of HIV contain a well balanced assembly of splicing silencers and enhancers (30–32). RSV explores a different mechanism in addition to weak 3′ss for *env* and *src* (32).

The *gag* gene contains a sequence known as negative regulator of splicing, which acts as a decoy 5′ss to generate a non-productive spliceosome, thereby reducing the efficiency of the actual upstream 5′ss (33).

The mechanisms of splice regulation in the otherwise well investigated MLV have not been subject to detailed studies (13, 34). Several lines of evidence led us to hypothesize that MLV negatively regulates the 5′ss instead of the 3′ss, in contrast to many other retroviruses. Although the MLV 3′ss does not fully match the consensus (Fig. 1*B*), the preceding PPT is close to the cellular average in length and, more importantly, not interrupted by attenuating purines. Such a relatively strong PPT can substitute for either a weak branch point or poor sequence conservation of the actual 3′ss (35). Moreover, introducing a very efficient cellular 3′ss into MLV vectors did not fully prevent the formation of genomic RNA (15). Finally, when constructing a new generation of retroviral self-inactivating vectors, we placed the internal promoter 18 bp downstream of the PBS and thus 24 bp upstream of the 5′ss. In this configuration, we observed complete splicing of the retroviral intron, strongly suggesting that sequences located upstream of the promoter insertion site (comprising the second half of R, U5, and 18 bp 3′ of the PBS) negatively regulate gammaretroviral splicing (16).

These observations and the opposite, splice-promoting effect of the PBS in MLV splice regulation can be explained by the secondary structure of the first untranslated exon. Mougel *et al.* (17) determined the structure of the leader region of MLV by chemical probing. Their stem-loop model is schematically shown in Fig. 7*A*. The 5′ss is located at the bottom of the stem, where the first G residue of the intron is still paired (Fig. 7*A*). Within the stem structure, the PBS is the only sequence element that loops out. This is likely a prerequisite for its main function, binding of the tRNA in the producer cell (17), and thus a consequence of a strong evolutionary pressure. Interestingly, deleting the PBS leads to an even more stable structure as determined by the MFOLD program (36), whereas deletion of U5 and sequences 3′ of the PBS destabilize the structure. Thus, the stability of this region correlates with the degree of 5′ss attenuation. In addition, fine mapping of the 3′PBS region revealed that in particular, the lower part of the stem is important to form a splice inhibitory structure (Fig. 7). Compensatory mutations strongly argue for the correctness of the structural model.

The stability may well affect the frequency of stem-loop formation. A similar phenomenon has also been reported for alternative splicing of the fibroblast growth factor receptor mRNA, where two regulating sequences are juxtaposed by formation of a stem (37, 38). If the stability of the upstream region is important for negative regulation of the 5′ss, one could envision that recognition of the 5′ss by U1snRNP would be impaired, thus leading to balanced splicing. Examples for this kind of splice regulation have been found in two cellular genes: adenosine deaminase (39) and tau (Ref. 40 and reviewed in Ref. 41). In these two cases, the complete 5′ss is part of the stem, whereas we favor a model for MLV of an inhibitory secondary structure formed by sequences located upstream of the 5′ss. This structure may determine the accessibility for cellular splice regulators, whose identity and recognition motifs remain to be determined. Interestingly, a secondary structure is also implicated in the function of the negative regulator of splicing of RSV (42). In this case, a single nucleotide deletion leads to an increase in stability of the stem-loop structure in analogy to the PBS deletion. As a result, the mutated negative regulator of splicing displays a reduced binding of U1snRNP, and the authors conclude that for balanced splicing, a moderately destabilized structure is necessary (42).

In summary, the complex structure of the first untranslated exon of MLV is important for several steps in the retroviral life cycle and represents a compromise of several evolutionary needs. The RSL is required for the export of the unspliced RNA, in line with previous results (10). The U5, in particular the stem structure upstream of the PBS and the pairing with the region 3′ of the PBS, is important for the initiation of reverse transcription (43). In addition, both deletions affect the translational utilization of the respective RNAs (Fig. 3*C*), reminiscent of the role of the R/U5 region of spleen necrosis virus in translational regulation (44, 45). The main finding of the current study is a novel function for the sequences downstream of the RSL, and especially for the PBS, in retroviral splice regulation.

Acknowledgments—We thank B. Schnierle for the plasmid MOVGFP and C. Stocking, K. Harbers, H. Will, and W. Ostertag for support of J. K.'s thesis at the Heinrich-Pette-Institute.

REFERENCES

1. Coffin, J. M., Hughes, S. H., and Varmus, H. E. (1997) *Retroviruses* (Coffin, J. M., ed) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
2. Frankel, A. D., and Young, J. A. (1998) *Annu. Rev. Biochem.* **67,** 1–25
3. Pollard, V. W., and Malim, M. H. (1998) *Annu. Rev. Microbiol.* **52,** 491–532
4. Wodrich, H., and Krausslich, H. G. (2001) *Results Probl. Cell Differ.* **34,** 197–217
5. Chang, D. D., and Sharp, P. A. (1989) *Cell* **59,** 789–795
6. Legrain, P., and Rosbash, M. (1989) *Cell* **57,** 573–583
7. Hammarskjold, M. L. (2001) *Curr. Top. Microbiol. Immunol.* **259,** 77–93
8. Bray, M., Prasad, S., Dubay, J. W., Hunter, E., Jeang, K. T., Rekosh, D., and Hammarskjold, M. L. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 1256–1260
9. Cullen, B. R. (2003) *Trends Biochem. Sci.* **28,** 419–424
10. Trubetskoy, A. M., Okenquist, S. A., and Lenz, J. (1999) *J. Virol.* **73,** 3477–3483
11. Bouck, J., Fu, X. D., Skalka, A. M., and Katz, R. A. (1995) *Mol. Cell. Biol.* **15,** 2663–2671
12. Zhang, L., and Stoltzfus, C. M. (1995) *Virology* **206,** 1099–1107
13. Lazo, P. A., Prasad, V., and Tsichlis, P. N. (1987) *J. Virol.* **61,** 2038–2041
14. Purcell, D. F., and Martin, M. A. (1993) *J. Virol.* **67,** 6365–6378
15. Lee, J. T., Yu, S. S., Han, E., and Kim, S. (2004) *Gene. Ther.* **11,** 94–99
16. Kraunus, J., Schaumann, D. H., Meyer, J., Modlich, U., Fehse, B., Brandenburg, G., Von Laer, D., Klump, H., Schambach, A., Bohne, J., and Baum, C. (2004) *Gene. Ther.* **11,** 1568–1578
17. Mougel, M., Tounekti, N., Darlix, J. L., Paoletti, J., Ehresmann, B., and Ehresmann, C. (1993) *Nucleic Acids Res.* **21,** 4677–4684
18. Hildinger, M., Abel, K. L., Ostertag, W., and Baum, C. (1999) *J. Virol.* **73,** 4083–4089
19. Chen, C., and Okayama, H. (1987) *Mol. Cell. Biol.* **7,** 2745–2752
20. Lund, A. H., Duch, M., Lovmand, J., Jorgensen, P., and Pedersen, F. S. (1997) *J. Virol.* **71,** 1191–1195
21. Weil, D., Boutain, S., Audibert, A., and Dautry, F. (2000) *RNA* (*Cold Spring Harbor*) **6,** 962–975
22. Dani, C., Piechaczyk, M., Audigier, Y., El Sabouty, S., Cathala, G., Marty,

L., Fort, P., Blanchard, J. M., and Jeanteur, P. (1984) *Eur. J. Biochem.* **145,** 299–304
23. Grez, M., Zornig, M., Nowock, J., and Ziegler, M. (1991) *J. Virol.* **65,** 4691–4698
24. Logan, A. C., Haas, D. L., Kafri, T., and Kohn, D. B. (2004) *J. Virol.* **78,** 8421–8436
25. Marquet, R., Isel, C., Ehresmann, C., and Ehresmann, B. (1995) *Biochimie* (*Paris*) **77,** 113–124
26. Iborra, F. J., Jackson, D. A., and Cook, P. R. (2001) *Science* **293,** 1139–1142
27. Galla, M., Will, E., Kraunus, J., Chen, L., and Baum, C. (2004) *Mol. Cell* **16,** 309–315
28. Erlwein, O., Buchholz, C. J., and Schnierle, B. S. (2003) *J. Gen. Virol.* **84,** 369–373
29. O'Reilly, M. M., McNally, M. T., and Beemon, K. L. (1995) *Virology* **213,** 373–385
30. Amendt, B. A., Si, Z. H., and Stoltzfus, C. M. (1995) *Mol. Cell. Biol.* **15,** 4606–4615
31. Bilodeau, P. S., Domsic, J. K., Mayeda, A., Krainer, A. R., and Stoltzfus, C. M. (2001) *J. Virol.* **75,** 8487–8497
32. Katz, R. A., Skalka, A. M., and Kotler, M. (1990) *Mol. Cell. Biol.* **10,** 696–704
33. Paca, R. E., Hibbert, C. S., O'Sullivan, C. T., and Beemon, K. L. (2001) *J. Virol.* **75,** 7763–7768
34. Hoshi, S., Odawara, T., Oshima, M., Kitamura, Y., Takizawa, H., and Yoshikura, H. (2002) *Biochem. Biophys. Res. Commun.* **290,** 1139–1144
35. Roscigno, R. F., Weiner, M., and Garcia-Blanco, M. A. (1993) *J. Biol. Chem.* **268,** 11222–11229
36. Zuker, M. (1989) *Methods Enzymol.* **180,** 262–288
37. Muh, S. J., Hovhannisyan, R. H., and Carstens, R. P. (2002) *J. Biol. Chem.* **277,** 50143–50154
38. Baraniak, A. P., Lasda, E. L., Wagner, E. J., and Garcia-Blanco, M. A. (2003) *Mol. Cell. Biol.* **23,** 9327–9337
39. Bratt, E., and Ohman, M. (2003) *RNA* (*Cold Spring Harbor*) **9,** 309–318
40. Grover, A., Houlden, H., Baker, M., Adamson, J., Lewis, J., Prihar, G., Pickering-Brown, S., Duff, K., and Hutton, M. (1999) *J. Biol. Chem.* **274,** 15134–15143
41. Buratti, E., and Baralle, F. E. (2004) *Mol. Cell. Biol.* **24,** 10505–10514
42. Cabello-Villegas, J., Giles, K. E., Soto, A. M., Yu, P., Mougin, A., Beemon, K. L., and Wang, Y. X. (2004) *RNA* (*Cold Spring Harbor*) **10,** 1388–1398
43. Miller, J. T., Ge, Z., Morris, S., Das, K., and Leis, J. (1997) *J. Virol.* **71,** 7648–7656
44. Roberts, T. M., and Boris-Lawrie, K. (2000) *J. Virol.* **74,** 8111–8118
45. Roberts, T. M., and Boris-Lawrie, K. (2003) *J. Virol.* **77,** 11973–11984

E Manuscript

Interplay of secondary structure and limited pairing to U1 snRNA results in dual attenuation of a retroviral 5'splice site

Daniela Zychlinski, Steffen Erkelenz, Vanessa Melhorn, Heiner Schaal, Christopher Baum and Jens Bohne

Author's contribution statement:

Daniela Zychlinski cloned the SF91 vector mutants, the SCSenv hybrid mutants, the heterologous stem loop mutants and designed and constructed the U1 snRNA mutants. Furthermore DZ performed the experiments and analyzed the data for Figs.1, 2D-E, 3-5 and participated in performing Suppl. Fig.1. In addition, she designed the figures and participated in writing the manuscript. Steffen Erkelenz constructed the MS2 plasmids. Vanessa Melhorn analyzed the titer of the proviral mutants. Heiner Schaal designed the MS2 constructs and participated in writing the manuscript. Christopher Baum supervised the study, critically evaluated the data and participated in writing the manuscript. Jens Bohne is the corresponding author, supervised and designed all experiments and mutants, perfomed the experiments for Fig. 2A-B, participated in performing Suppl. Fig.1 and wrote the manuscript

Abstract

Multiple types of regulation are used by cells and viruses to control alternative splicing. In murine leukemia virus, accessibility of the 5'splice site (ss) is regulated by an upstream region, which can fold into a complex RNA stem-loop structure. The underlying sequence of the structure itself is negligible, since most of it could be functionally replaced by a simple heterologous RNA stem loop preserving the wild-type splicing pattern. Increasing the RNA duplex formation between U1 snRNA and the 5'ss by a compensatory mutation in position +6 led to enhanced splicing. Interestingly, this mutation affects splicing only in the context of the secondary structure, arguing for a dynamic interplay between structure and primary 5'ss sequence. The reduced 5'ss accessibility could also be counteracted by recruiting an RS domain via a modified MS2 coat protein to a single binding site at the tip of the simple RNA stem loop. The mechanism of 5'ss attenuation was revealed using hyperstable U1 snRNA mutants, showing that restricted U1 snRNP access is the cause of alternative splicing.

Introduction

Cells and retroviruses heavily depend on alternative splicing for differential gene expression. In human cells alternative splicing expands the proteome, giving rise to roughly 100,000 protein isoforms from only 25,000 genes (Maniatis & Tasic, 2002). In retroviruses a single pre-mRNA corresponding to the complete genome undergoes alternative splicing in order to express all viral genes (Coffin et al., 1997).

The splicing reaction is executed by the spliceosome (Staley & Guthrie, 1998). The core of the spliceosome consists of the small nuclear RNPs (U snRNPs) (Will & Lührmann, 2006). Some of the U snRNPs participate in splice site recognition via RNA:RNA interactions (Valadkhan, 2007; Konarska & Query, 2005). The first steps towards mRNA splicing is the recognition of the 5'ss by the free, complementary 5'end of U1 snRNA (Mount et al., 1983; Rinke et al., 1984; Zhuang & Weiner, 1986). Therefore, the hydrogen bonding pattern between the 5'ss and U1 snRNA determines the intrinsic strength of a 5'ss and thus contributes to its recognition and frequency of usage, creating a first layer of regulation (Burge et al., 1999; Freund et al., 2003).

In contrast to yeast, where almost all splice sites strictly follow the consensus sequence (Burge et al., 1999), splice sites in retroviral and mammalian genomes are much more degenerated and recognition is assisted by a number of proteins (Hastings & Krainer, 2001; Stoltzfus & Madsen, 2006). In addition, regions surrounding the splice sites often represent exonic or intronic splicing enhancers (Dirksen et al., 1994; Blencowe, 2000) or silencers (Amendt et al., 1994; Tange et al., 2001). These elements modulate the intrinsic strength of splice sites via recruitment of splicing factors like SR proteins or hnRNPs (Hertel, 2008). The efficiency of splice sites is also modulated by RNA structure (Solnick, 1985; Buratti & Baralle, 2004). Alternatively, binding of splicing regulatory proteins may also depend on the local RNA structure (Marchand et al., 2002). On top of this, transcriptional elongation also regulates alternative splicing, illustrating the close connection between splicing and transcription (Kornblihtt et al., 2004).

Retroviruses represent very valuable model systems for studying alternative splicing (Cochrane et al., 2006). They synthesize only one polycistronic primary transcript, which undergoes alternative splicing for full viral gene expression

(Fig.1A; Coffin et al., 1997). Thus, retroviruses need to tightly control the use of their splice sites to ensure optimal levels of unspliced vs. spliced RNAs (Pollard & Malim, 1998). The unspliced or genomic RNA is both packaged into progeny virus and serves as a translation template for the structural and enzymatic proteins Gag and Pol, whereas the spliced RNA encodes the envelope protein (Env) in simple retroviruses (Fig.1A). Therefore the correct ratio between Gag and Env partly determines viral infectivity. Retroviral splicing is mainly regulated via weak 3'ss in conjunction with ESEs or ESSs (Amendt et al., 1995; Caputi et al., 1999; Tange & Kjems, 2001; Kammler et al., 2006). For HIV, it has been shown that most PPTs are interrupted by weakening purines (O'Reilly et al., 1995) and, in one case, a PPT is additionally attenuated by a secondary structure (Jacquenet et al., 2001). For simple retroviruses like Rous sarcoma virus (RSV) or murine leukemia virus (MLV), alternative splicing has been attributed to weak 3'ss too (McNally & Beemon, 1992; Lee et al., 2004; Logg et al., 2007). In addition, RSV harbors a decoy 5'ss, which redirects splicing activity from the actual 5'ss to the non-productive one (McNally et al., 1991; McNally & McNally, 1999).

We could previously show that in MLV the 5'ss instead of the 3'ss is negatively regulated via upstream sequences, which can form a secondary structure (Kraunus et al., 2006). Moreover, the stability and integrity of this structure correlates with 5'ss attenuation (Kraunus et al., 2006). We now demonstrate that this structure restricts access to U1 snRNP. A heterologous simple RNA stem loop was able to replace the structure while preserving the wild-type splicing pattern. Furthermore, we tethered SR proteins to the RNA structure via modified MS2 coat proteins. The RS domain was able to overcome the restricted access of U1 snRNP to the 5'ss presumably by stabilizing the RNA duplex formation. In addition, we discovered that the RNA secondary structure exerts its effect in conjunction with limited base pairing of the 5'ss to U1 snRNA at position +6. Interestingly, low complementarity affects splicing only in the context of the secondary structure, arguing for a novel dynamic interplay between structure and primary 5'ss sequence.

Results

Improvement of the 5'splice site at position +6 enhances splicing in conjunction with the secondary structure

In order to clarify the molecular mechanism of splicing regulation in murine leukemia virus (MLV), we used our previously described splicing reporter SF91 (Kraunus et al., 2006). SF91 is derived from an MLV provirus and contains the packaging signal (ψ) embedded in an intron flanked by the authentic MLV 5' and 3'splice sites (Fig.1A; Hildinger et al., 1999). Immediately downstream of the 3'ss the eGFP ORF was inserted (Fig.1A). The intron displays alternative splicing as the provirus (Fig.2A; Kraunus et al., 2006). We could previously show that a complex secondary structure attenuates the 5'ss based on the RNA structure of the leader region (Fig.1B; Mougel et al., 1993) and the functional evaluation of deletions upstream of the 5'ss (Kraunus et al., 2006).

In addition to the presence of an inhibitory secondary structure, we recognized that the primary sequence of the 5'ss deviates from the cellular consensus at position +6 (Fig.1B, inset). Genome wide, this position displays a low level of conservation (Hartmann et al., 2008), but it has been implicated in the recognition of weak 5'ss due to mismatches in other parts of the consensus sequence or the presence of splicing silencers (Ibrahim el et al., 2007; Hartmann et al., 2008). We therefore converted position +6 from C to U (Fig.1B), thereby increasing complementarity to U1 snRNA. In order to compare the splicing efficiency of the C6U mutant, we took advantage of a structural mutant (sm3; Fig.1B), which affects stem formation and leads to an eight-fold enhancement of splicing (Fig.1C, lane 3; Kraunus et al., 2006). After transient transfection of 293T cells, the C6U mutant displayed enhanced splicing, but not to the same extent as the sm3 mutation (Fig.1C, D, compare lanes 4,3). A combination of the two mutations resulted in a slightly stronger enhancement of splicing (Fig.1C, lane 5). The effect of the +6 mutation is also detectable in the context of a deletion of the primer binding site (dPBS; Fig.1B and C, lanes 6,7), which leads to a more stable stem structure and a stronger attenuation of the 5'ss (Fig.1B; (Kraunus et al., 2006). In summary, splicing at the 5'ss is regulated on two layers: secondary structure and primary splice site sequence.

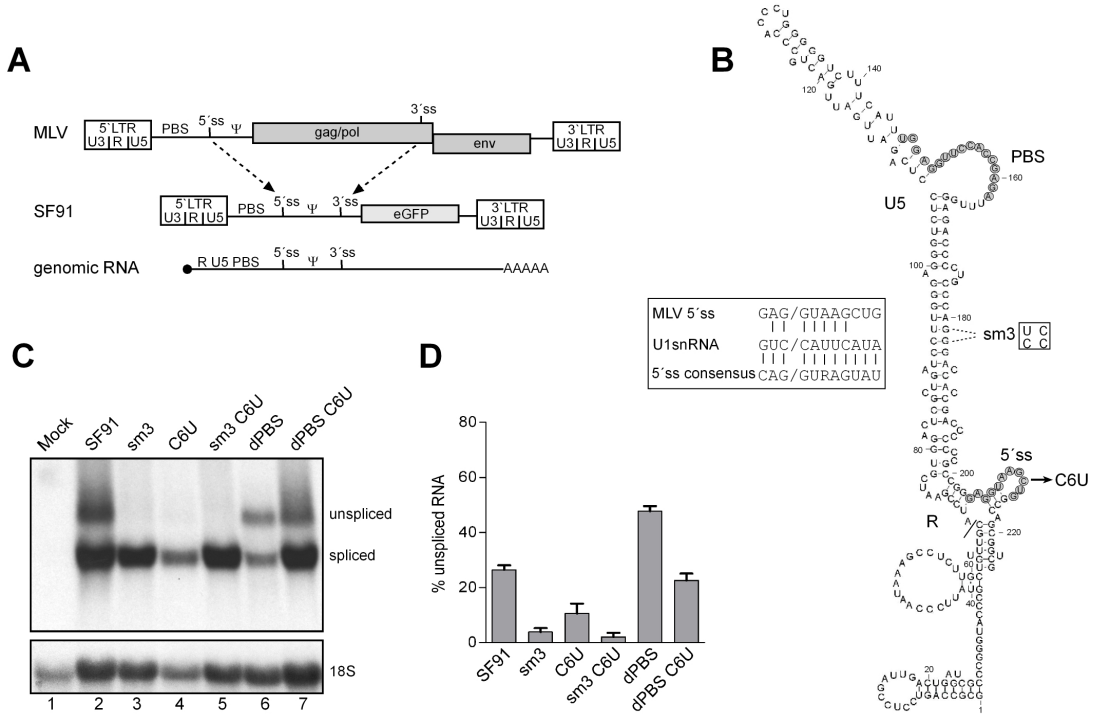


Figure 1: Dual attenuation of a 5'ss via the structure and a mismatch at position +6

(A) Schematic drawing of the murine leukemia provirus. The retrovirus is characterized by the terminal repeats (LTRs, open boxes), the presence of the genes *gag*, *pol* and *env* (gray boxes) and the primer binding site (PBS). The packaging signal (ψ) and the *gag*/*pol* reading frame is flanked by 5' and 3'ss. Below this, the MLV-derived splicing reporter SF91 is depicted. The proviral 3'ss was cloned downstream of the packaging signal and upstream of the eGFP start codon (gray box). Therefore, SF91 encodes a genomic RNA (black line), which contains an intron constituted by the authentic MLV splice sites. **(B)** RNA secondary structure of the region upstream of the packaging signal. The structure is adapted from the chemically validated structure (Mougel et al., 1993). The regions R and U5 are part of the LTR. The nucleotides representing the primer binding site and the 5'ss are highlighted by gray circles. At the 5'ss, the +6 C to U mutation is marked on the right. In the central region of the stem, the structural mutant sm3 including the nucleotide exchanges is depicted. The inset shows the MLV 5'ss pairing to U1 snRNA. Below this, the consensus 5'ss sequence is paired to U1 snRNA. Base pairs are depicted as solid vertical lines. **(C)** Northern blot using 10 µg total RNA from 293T cells transiently transfected with 5 µg of the indicated constructs and harvested 48h post-transfection. The blot was probed with an eGFP-specific probe. The identity of the RNA species is stated on the right. As a loading control, the blot was re-hybridized with a probe corresponding to 18S rRNA. **(D)** Phosphoimager analysis of Northern blot as shown in C. The extent of alternative splicing is given as a percentage of unspliced RNA. The mean values and standard deviations of three independent experiments are shown.

The regulation of splicing is conserved in a replication-competent provirus

We were previously able to show that deletion mutants of sequences upstream of the 5'ss affect splicing in the context of a replication-competent MLV provirus (Kraunus et al., 2006). We now wanted to confirm that splicing in the provirus is also regulated on two layers: structure and primary 5'ss sequence. Cloning of two key splicing mutations, namely sm3 (structural mutant; Fig.1C, lane 3) and C6U (splice site mutant; Fig.1C, lane 4), into the provirus (Fig.1A) yielded an enhancement of the

level of splicing identical to that seen in the splicing reporter (Fig.2A, lanes 1-3). Even though the provirus splices less efficiently than the reporter, the magnitude of splicing enhancement of the particular mutants is maintained (Fig.2B).

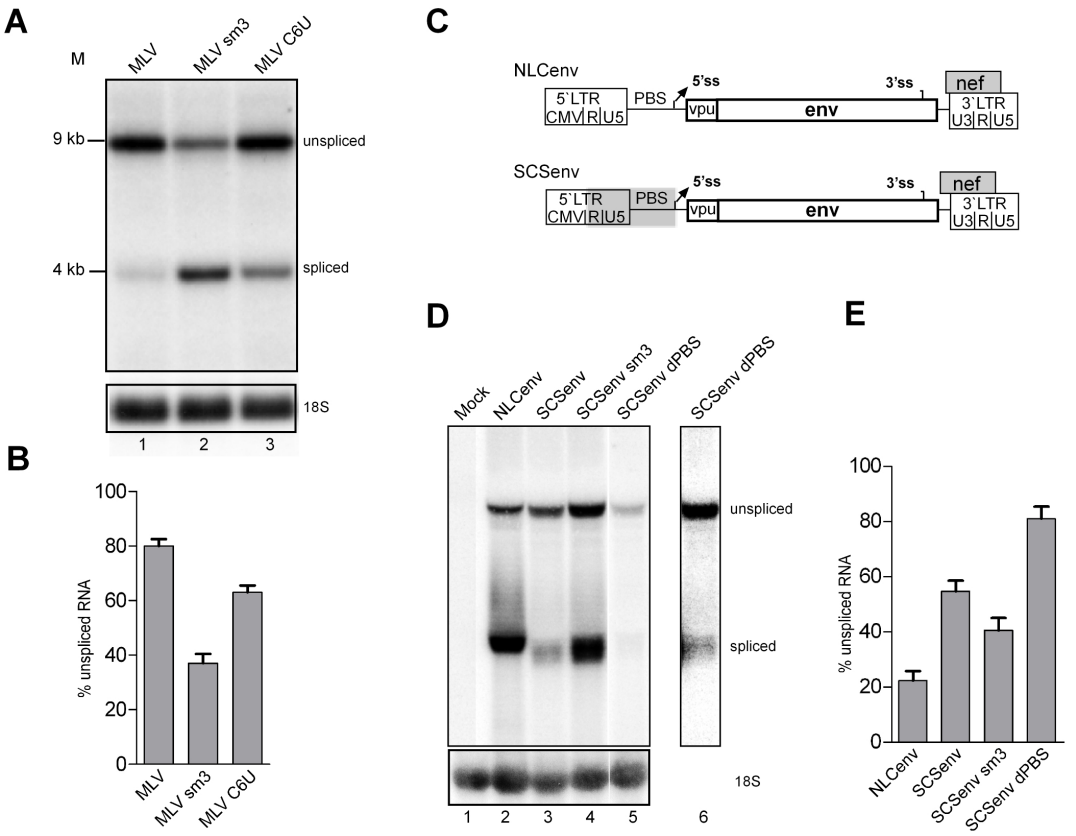


Figure 2: Splicing regulation is transferable to a complete provirus and an HIV 5'ss

(A) Northern blot of 10 µg of total RNA from 293T cells transfected with 10µg of the proviral constructs (Fig.1A) harboring two splicing mutants (sm3 and C6U; Fig.1C). Molecular weight standard is given on the left and the RNA species are named on the right. **(B)** Phospho-imager analysis. The extent of splicing is given as the percentage of unspliced RNA. The mean values and standard deviations of three independent experiments are shown. **(C)** Schematic drawing of the NLCenv plasmid. NLCenv is derived from the HIV molecular clone NL4-3. Expression is directed by the authentic LTR sequences except that the HIV promoter was replaced by CMV. ORFs are shown as boxes. The splice sites are indicated as well as the primer binding site. Below this, the hybrid construct SCSenv is shown. The gray region highlights the SF91-derived sequence in the HIV backbone. **(D)** Northern blot of 10 µg of total RNA from HeLaP4 cells transiently transfected with 2.75 µg of the indicated constructs. Two splicing mutants were transferred to SCSenv (sm3 and dPBS; Fig.1B). The RNA species are named on the right. The lane containing SCSenv dPBS was overexposed to display the ratio between unspliced and spliced RNA. **(E)** Phosphoimager analysis as described in Figure 2C. The mean values and standard deviations of five independent experiments are shown.

We also tested whether the mutants hamper viral replication. A 4-fold decrease in viral titer for the sm3 mutant after replication in murine cells compared to the wild-type was observed (data not shown). Thus, there is a correlation between the extent of splicing and viral fitness. Oversplicing of the genomic RNA leads to reduced levels

of full-length RNA to be packaged into viral progeny and to lower amounts of the structural proteins since they are translated only from the unspliced RNA (see Fig.1A).

To sum up, the regulation of alternative splicing is conserved in the context of the complete provirus and mutations that severely affect splicing ratios impede viral replication.

The attenuating effect of the secondary structure is transferable to a heterologous 5'splice site

Next we asked: does the inhibitory secondary structure represent a general mean of splicing regulation? And did it evolve and function only in murine leukemia virus? In addition, we wanted to rule out the possibility that sequences downstream of the 5'ss are involved in its attenuation. For this purpose, we used the previously described HIV-NLenv system (Fig.2C; (Bohne et al., 2005). Briefly, this subviral envelope expression system is based on the HIV-1 proviral clone NL4-3. Using the Tat-independent CMV promoter instead of the viral LTR (NLCenv), the construct expresses a sequence-identical HIV *env* mRNA, which can alternatively splice to yield *nef* mRNA (Fig.2C; Bohne et al., 2007). RNA analysis showed that NLCenv displays alternative splicing due to the weak 3'ss upstream of *nef* (Fig.2D, lane 2) (Zhu et al., 2001; Marchand et al., 2002; Bohne et al., 2005).

In order to transfer the splicing regulation from MLV to a heterologous 5'ss, we inserted the 5'leader sequence of MLV (bases 1-203; Fig.1B) immediately upstream of the HIV 5'ss, thereby creating the SCSenv construct (light gray box, Fig.2C). SCSenv expresses an MLV/HIV fusion transcript containing the MLV-derived secondary structure (Fig.1B) followed by the complete HIV 5'ss and downstream *env* mRNA sequences. In addition, we transferred two structural mutants (sm3; dPBS; Fig.1C) to SCSenv. Transfection into HeLaP4 cells revealed that the MLV-derived sequences can strongly attenuate the HIV 5'ss (Fig.2D, compare lanes 2,3). The level of unspliced RNA is enhanced 2.5-fold (Fig.2E) and exceeds the level of 5'ss attenuation observed for the basic construct SF91 containing the identical leader sequence (Fig.1D). Comparison of the two 5'ss using the HBond score algorithm (Freund et al., 2003) revealed that the HIV 5'ss is intrinsically weaker than MLV (HIV HBond score 15.7; MLV HBond score 17.1). This is in agreement with our hypothesis

that a 5'ss with lower complementarity should be even more prone to attenuation via the RNA secondary structure. Moreover, the mode of splicing regulation could be transferred onto a heterologous 5'ss since the sm3 mutant is spliced more efficiently and a deletion of the PBS leads to more unspliced RNA (Fig.2E, compare columns 4,5 and Fig.1). Furthermore, no downstream MLV-derived sequences were necessary for 5'ss attenuation. We noted that the sm3 mutant did not splice to the same extent as NLCenv, although the 5'ss should be accessible to the spliceosome (Fig.2E). However, this can firstly be explained by the presence of a purine-rich splicing enhancer only present in the HIV leader sequence upstream of the 5'ss (Kammler et al., 2001; Caputi et al., 2004). Secondly, the degree of splice site attenuation seems to influence the total amount of RNA (Fig.2D). The dPBS mutant in particular displayed the highest level of unspliced RNA, but the lowest amount of RNA in total (Fig.2D, lanes 5,6). In order to visualize the ratio of unspliced vs. spliced RNA, this part of the Northern blot had to be overexposed (Fig.2D, lane 6).

The experiments using the HIV *env* expression system demonstrate that the splicing regulation observed is not restricted to MLV, but can be transferred to a heterologous 5'ss. Moreover, this implies some sort of generality in the underlying mechanism of 5'ss attenuation.

Replacement of the RNA structure with random sequences capable of stem formation results in proper splicing regulation

In order to differentiate between whether mere structural requirements of the leader region or specific sequences harboring splicing regulatory protein binding sites cause 5'ss attenuation, we replaced the upper part of the RNA structure with a heterologous stem loop harboring the ability to form a stem with free energy similar to the wild-type structure (Fig.1B and 3A; SF91: ΔG = -49.6 kcal/mol; heterologous stem-loop: ΔG = -7.2 kcal/mol). Transfection of the stem-loop construct almost perfectly resembles the splicing pattern of the wild-type reporter (Fig.3B, lanes 1,2). Thus, 5'ss attenuation can be attributed to RNA stem-loop formation and not to the primary sequence, which forms the structure. As a control, we reversed the descending part of the stem, resulting in a stem-loop antisense (as) construct, which is unable to form the secondary structure. This vector displayed complete splicing (Fig.3B, lane 3). In contrast, strengthening the stem by a 20 bp extension on either side

(ΔG = - 68.8 kcal/mol) led to more unspliced RNA (Fig.3B,C; lane 4), reminiscent of the deletion of the PBS (Fig.1C, lane 6). Also, the C6U mutation enhances splicing as in the wild-type context, suggesting that this interplay does not require specific cellular proteins binding to the MLV stem loop (Fig.3B, lane 5). In addition, we noted that, reminiscent of the SCSenv plasmid (Fig.2D), the splicing efficiency correlated with the overall transcript levels (Fig.3B, compare lanes 3,4). Therefore, the lane containing the stem-loop antisense (as) was underloaded for proper visualization (Fig.3B, lane 3). In addition, the antisense mutant showed that the 5'ss could function highly efficiently in the absence of the stem loop despite the mismatch at position +6. This position became only critical in conjunction with the secondary structure, pointing to a novel dynamic interplay between structure and primary splice site sequence.

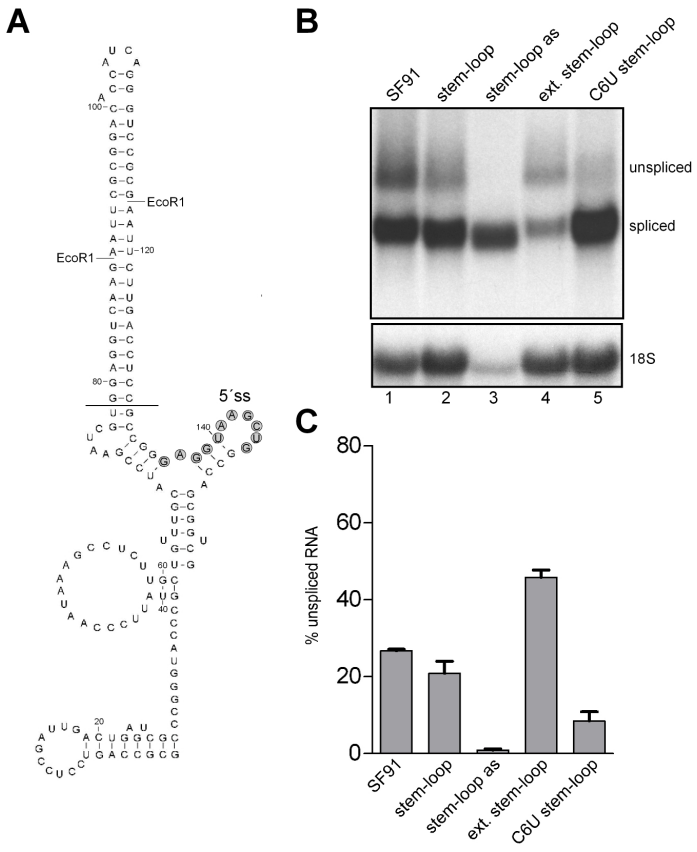


Figure 3: A heterologous stem loop perfectly resembles MLV splicing regulation

(A) The secondary structure is depicted as in Figure 1B. The vertical black line marks the insertion of heterologous sequences forming the synthetic stem-loop. On each side of the stem an EcoRI restriction site is indicated. The 5'ss is highlighted by gray circles. **(B)** Northern blot analysis of 10 µg total RNA from transfected 293T cells. The RNA species are marked on the right. Note that only 5 µg RNA were loaded in the stem-loop antisense (as) lane due to the high expression of this construct. **(C)** Phosphoimager analysis. Mean values and standard deviation represent three independent experiments.

SR-protein domains targeted to the heterologous stem-loop partially overcome 5'ss attenuation

It has been shown in yeast that 5' and 3'ss mutants can be rescued by SR proteins even though S. cerevisiae does not code for such proteins (Shen & Green, 2006). This hints at a mechanism where SR proteins might stabilize RNA:RNA interactions at weak splice sites. Since the secondary structure may restrict access to U1 snRNP, we anticipated that targeting an RS domain into the vicinity of the 5'ss would possibly overcome this attenuation. We made use of the MS2-fusion tethering system (Selby & Peterlin, 1990; Peabody, 1993). However, due to structural constraints, the heterologous RNA stem loop accommodates only one MS2 binding site. We therefore used the ΔFG variant of the MS2 coat protein leading to an increase in RNA binding affinity at the expense of dimer:dimer formation (Chao et al., 2008). This modification should theoretically result in enhanced binding to a single site *in vivo* (Fig.4A). In a co-transfection assay, we tested various RS domains fused to MS2ΔFG along with the SF91 construct and two RNA stem-loop variants harboring a single MS2 binding site (Fig.4B). Transfection of SF91 and a plasmid encoding the RS domain of SRp55 fused to MS2ΔFG is not neutral, as observed for SF91, and leads to a slight statistically significant enhancement of splicing (Fig.4B, lanes 1,2 and Fig.4C, *P* value = 0.03). However, in the context of the RNA stem loop and the extended stem loop construct, co-transfection of the MS2-SRp55 plasmid led to a highly significant enhancement of splicing (Fig.4B,C, compare lanes 3,4 and 5,6; *P* value = 0.009 and 0.002).

In general, these experiments proved for the first time *in vivo* that MS2ΔFG mutants can be targeted to a single binding site with reasonable efficiency. Moreover, tethering of an RS domain to the heterologous RNA stem loop partially overcomes 5'ss attenuation, in agreement with previous results that the RNA structure restricts access to U1 snRNP.

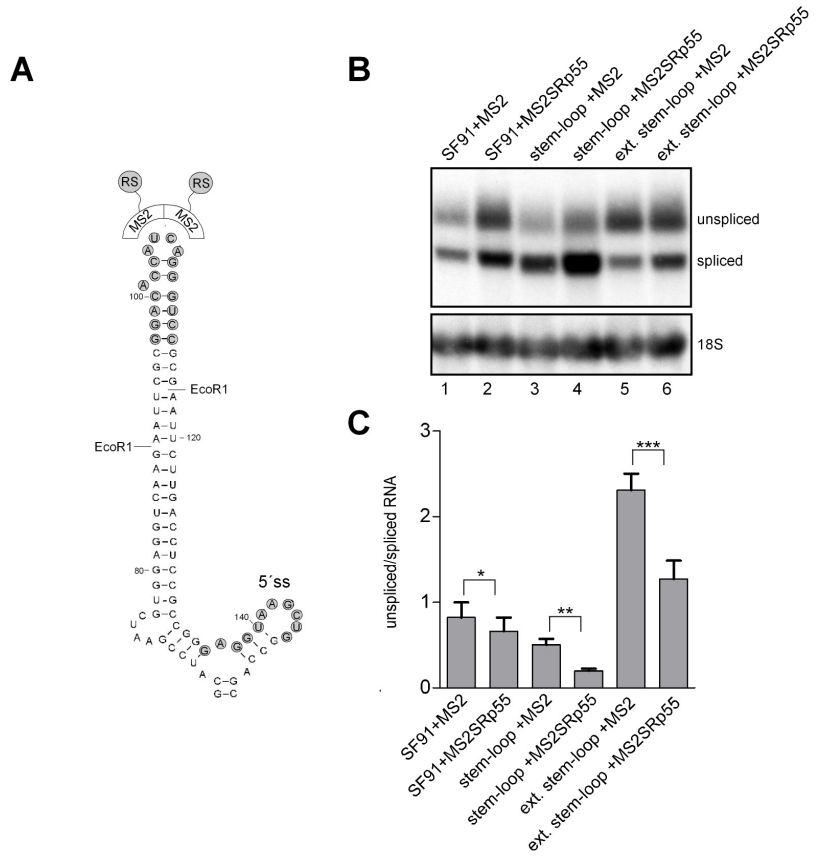


Figure 4: An RS domain can partially relieve 5'ss attenuation

(A) Depiction of the heterologous stem-loop structure as in Figure 4A. In addition to the 5'ss, the MS2 binding site is also highlighted by gray circles. MS2-RS fusion proteins recognize the MS2 binding site as dimers. **(B)** Northern blot performed as in Figure 1C. As a control, a plasmid encoding just the MS2 coat protein was co-transfected as indicated. The RS domain of SRp55 was fused to the modified MS2 protein. The RNA species are marked on the right. **(C)** Phosphoimager analysis. Splicing efficiency is displayed as unspliced/spliced RNA ratio. Student's t-test was performed using mean values from four independent experiments. *P* values are * = 0.03; ** = 0.009; *** = 0.002.

5'splice site attenuation can be rescued by hyperstable U1 snRNA suppressor mutants

In addition to protein:protein contacts of U1 snRNP with exonic or intronic sequences (Du & Rosbash, 2002; Lund & Kjems, 2002), recognition of 5'ss is initiated by RNA:RNA interactions (Zhuang & Weiner, 1986; Seraphin et al., 1988). Therefore, the hydrogen bonding pattern between U1 snRNA and the 5'ss is critical and can either be enhanced by mutations within the 5'ss (C6U mutant, Fig.1, lane 4) or by overexpression of U1 snRNA suppressor mutants that increase the complementarity to a given 5'ss. In yeast, hyperstable U1 snRNA mutants cannot be displaced by U6 snRNA and therefore splicing is inhibited (Staley & Guthrie, 1999). In contrast, in mammalian cells, an extended U1 snRNA/5'ss interaction does not decrease splicing

efficiency, but rather increases 5'ss recognition (Freund et al., 2005; Singh et al., 2007). We constructed two U1 snRNA mutants containing one substitution (U1 G11C) leading to 8 complementary base pairs with the MLV 5'ss (Fig.5A; HBond score 18.8) or a perfect match of U1 snRNA leading to 11 continuous base pairs (U1 perfect; Fig.5A; HBond score 23.8). Overexpression of these suppressor mutants along with wild-type U1 snRNA led to an increase in splicing depending on the intrinsic strength of the RNA duplex (Fig.5B,C). Even an increase in complementarity of one base pair, in the case of the U1 G11C mutant (Fig.6A, lower panel) led to a strong enhancement of splicing (Fig.5C). This argues for a dynamic balance between secondary structure and accessibility of the 5'ss to U1 snRNA. In addition, we looked at direct interaction between U1 snRNP and the 5'ss *in vitro* by EMSA. We incubated purified U1 snRNP with *in vitro* transcribed RNA containing either the heterologous stem loop or the antisense mutant (Fig.3A). We could observe an enhanced binding of U1 snRNP in the absence of the RNA secondary structure (i.e. the antisense mutant) and reduced binding upon formation of the structure (suppl. Fig.1).

These experiments clearly demonstrate that splicing regulation in MLV uses restricted access of U1 snRNP to the 5'ss exerted by the upstream secondary structure and limited complementarity of the primary 5'ss sequence to U1 snRNA.

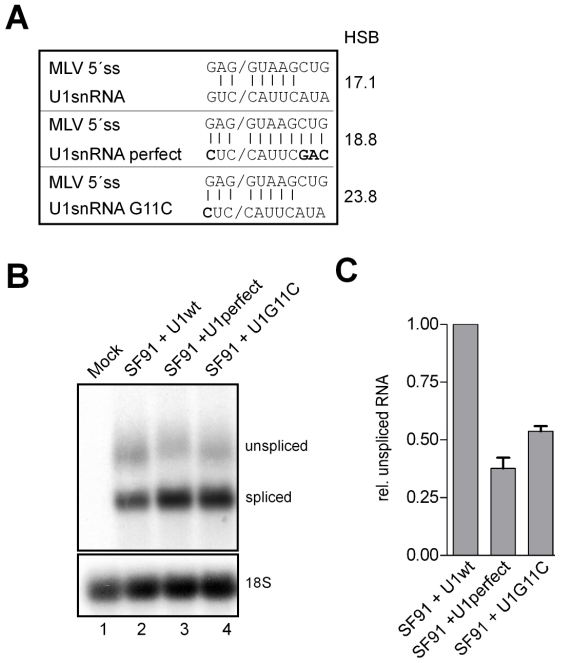


Figure 5:
Hyperstable U1 snRNA suppressor mutants enhance splicing

(A) The top line shows the MLV 5'ss. Possible base pairs to U1 snRNA are shown as solid vertical lines. In the middle and lower panel, pairing of the MLV 5'ss with the two U1 snRNA mutants is depicted. Mutated nucleotides in U1 snRNA are in bold. On the right-hand side, the HBond (HBS) score is indicated. **(B)** Northern blot using 10 µg of total RNA from transiently transfected 293T cells. Transfections were performed using 5 µg of SF91 plasmid and 10 µg of the respective U1 snRNA plasmid. The RNA species are named on the right. **(C)** Phosphoimager analysis. The efficiency of splicing is shown as the relative level of unspliced RNA. SF91 in the presence of co-transfected U1 wt plasmid was set to 1. Mean values and standard deviations represent five independent experiments.

Discussion

As presented here, murine leukemia virus uses a dynamic interplay between RNA secondary structure and primary 5'ss sequence to restrict access of U1 snRNP to its 5'ss, which ultimately results in alternative splicing and full viral gene expression.

Secondary structure has been implicated in alternative splicing early on (Solnick, 1985; Fu & Manley, 1987; Eperon et al., 1988). Cellular examples of attenuated 5'ss, which are part of a secondary structure, were discovered in association with different genetic diseases (Grover et al., 1999: Singh, 2007 #616; Bratt & Ohman, 2003). Modulation of splicing efficiency by RNA secondary structures has recently also been described for the HIV-1 leader RNA structure, where the major 5'ss is embedded in a semistable hairpin (Abbink & Berkhout, 2008). However, contrary to HIV-1, splicing regulation at the MLV 5'ss seems to be even more complex, since stem mutations 25 nucleotides upstream of the 5'ss already provoke an eight-fold enhancement of splicing (Fig.1). Certainly, secondary structures cannot only sequester 5'ss, but also modulate the binding efficiency of hnRNPs or SR proteins. On a global scale, it was shown that splicing enhancers and silencers are present mostly in single-stranded regions (Hiller et al., 2007). There is also a particular example from HIV, where a change in secondary structure allows hnRNP H to bind and influence splicing (Jablonski et al., 2008). RNA secondary structures are also statistically associated with alternative 5'ss (Shepard & Hertel, 2008), allowing splice site selection via conformational variability.

An exchange of the upper part of the structure with a heterologous stem-loop proved that the main function of the stem is to force the 5'ss into an inhibitory conformation and the extent of splicing inhibition correlates closely with the free energy of the structure (Fig.3). Not surprisingly, the complex RNA stem loop possesses additional functions in the viral life cycle. It is the scaffold to loop out the primer binding site, which binds a cellular tRNA as a primer to initiate reverse transcription (Coffin et al., 1997). Yet, splicing regulation could be transferred to a complete provirus and to a heterologous HIV 5'ss (Fig.2). Using the MLV/HIV hybrid plasmids, it seemed that the degree of 5'ss inhibition correlates inversely with the overall RNA amount (Fig.2D, compare lanes 5, 4 and 2). Similar effects have been observed in other studies, where splice sites are able to enhance transcriptional elongation (Fong & Zhou, 2001) and gene expression in general (Furger et al., 2002). It appears that the CMV

promoter is highly dependent on this positive feedback exerted by the interaction of U1 snRNP with a proximal 5'ss (Bohne & Krausslich, 2004; Damgaard et al., 2008). Thus, this was the first direct evidence that the secondary structure restricts access of U1 snRNP. Additional evidence was obtained by tethering RS domains to the heterologous stem loop (Fig.4). RS domains may play a dual role in splice site selection. They can enhance RNA:RNA interactions (Valcarcel et al., 1996; Shen & Green, 2006) at degenerated splice sites or engage in protein:protein interactions, since an excess of SR proteins can select 5'ss in the absence of U1 snRNP (Crispino et al., 1994). In our case, the RS domain may directly facilitate U1 snRNA binding to the attenuated 5'ss, since the strongest enhancement of splicing was observed with the long and continuous RS domain from SRp55 and not with four copies of the SF2/ASF RS domain (data not shown).

In addition, the retroviral 5'ss is characterized by a mismatch at position +6. Reversion of this position into a matching nucleotide with U1 snRNA enhances splicing 2.5-fold. Although this position is less conserved on a genome scale, it turns into a preserved nucleotide if the 5'ss is degenerated or weakened by surrounding elements (Hartmann et al., 2008). For example, a C6U mutation in the 5'ss flanking exon 20 of the IKBKAP gene causes exon skipping, resulting in familial dysautonomia (Slaugenhaupt et al., 2001). Here, attenuation is due to a weak upstream 3'ss and possible splicing silencers (Ibrahim el et al., 2007). In this line, position +6 of the MLV 5'ss regulates efficiency only in the context of the secondary structure. So there is a dynamic interplay between structure and the 5'ss sequence. The same result was obtained when we fused the complete second intron from β-globin to the MLV secondary structure. Alternative splicing occurred only after lowering the complementarity at positions +6 or -3 of the globin 5'ss (D. Z. and J. B., unpublished results). We also mutated positions +7 and +8 of the MLV 5'ss back to the consensus and observed an enhancement of splicing (data not shown). However, the enhancement was not as strong as observed for position +6. One may speculate that not solely complementarity to U1 snRNA, but also the length and neighborhood of the 5'ss:U1 snRNA duplex determines the strength of a splice site.

Succinctly, splicing regulation in MLV functions via a dynamic interplay of RNA secondary structure and primary 5'ss sequence adding another layer of complexity to the fine tuned regulation of alternative splicing and to the deciphering of the splicing code.

Materials and Methods

Plasmids

All retroviral vector plasmids were derived from pSF91 (Hildinger et al., 1999). The mutants sm3 (stem-loop mutant 3) and the compensation thereof (sm3comp) and the deletion of the primer binding site (dPBS) were described previously (Kraunus et al., 2006). All overlap PCRs used the outer primers XbaI and ApaI (Kraunus et al., 2006) to generate the final PCR fragment, which was cloned into the XbaI/ApaI sites of SF91. Vectors carrying the C6U mutation of the 5'ss were cloned by overlap PCR using SF91, SF91 sm3, SF91dPBS and SF91 stem-loop as a template (fw: 5`-TAA GTT GGC CAG CGG TCG TTT CG-3`; rv: 5`-GGC CAA CTT ACC TCC CGG C-3'). For the introduction of the heterologous stem loop (SF91stem-loop) into the leader region of SF91, the Kpn1/Msc1 fragment was replaced by a PCR product including the heterologous stem-loop sequences (fw: 5´-GTA CGG TAC CGT ATT CCC AAT AAA GCC TCT TGC TGT TTG CAT CCG AAT CGT GGA GGT CAA GAA TTC GCG GAC ACC ATC-3`; rv: 5`-GCA TCC TGG CCA GCT TAC CTC CCG GCG GAG GTC AAG AAT TCG CGG ACC CTG ATG GTG TCC GCG AAT TCT TGA CCT CCA CGA TTC-3`). The extended stem-loop was generated by introducing 20 additional base pairs into the EcoR1 sites of SF91stem-loop (fw:5`-[Phos] AAT TCG ATA TCC CGT GCG GAC ACC ATC AGG GTC CGC ACG GGA TAT CG -3`; rv: 5`[Phos] CGA TAT CCC GTG CGG ACC CTG ATG GTG TCC GCA CGG GAT ATC GAA TT-3` ) by ligation. By using the SF91stem-loop as a template, the SF91stem-loop antisense mutant (fw: 5`-CGC CTC CAG TTC TTA AGC GCC TCG GGA GGT AAG CTG GCC AGC GGT CGT TTC G-3`; rv: 5'-AGG CGC TTA AGA ACT GGA GGC GCC CTG ATG GTG TCC GCG AAT TCT TGA C-3`) was generated by overlap PCR and the final PCR product was cloned into the XbaI/ApaI sites of SF91. The SCSenv HIV/MLV hybrid constructs were also cloned by overlap PCR. The 5´PCR product was generated by using SCS11 (Schambach et al., 2006) as template (fw: 5`-GCG GTA TAC GCT AGC TTA AGT AG-3`; rv: 5`-TAC TTA CTG CCC GGC GGG GGG GTC GGT G-3`). To generate the 3`PCR product containing the HIV 5´ss NLenv was used as template (fw: 5`-CCC CCC CGC CGG GCA GTA AGT AGT ACA TGT AAT GC-3´; rv: 5`-GGT TGC TTC CTT CCA CAC AGG TAC-3`). Using the outer primers the final PCR product was cloned into the Bst1107I/BstEII sites of the NLCenv backbone (Bohne et al., 2007). In order to transfer the sm3 and the C6U mutant in an

MLV provirus an AflII/PstI fragment of the respective SF91 plasmids was generated and ligated into MOVSFGFP (Erlwein et al., 2003; Kraunus et al., 2006).

U1 snRNA mutants were cloned by PCR using a reverse primer (rv: 5`- CGC GGA TCC TCC ACT GTA GGA TTA AC-3') including a BamHI restriction site and different forward primers containing the U1 mutations flanked by a BglII restriction site (fw U1G11C: 5´-GCC CGA AGA TCT CAT ACT TAC CTC GCA G-3'; fw U1 perfect: 5`-GCC CGA AGA TCT CCA GCT TAC CTC GCA G-3`). The PCR products were cloned as a BamHI/BglII fragment into the pUC19 U1wt plasmid (kind gift of A. Weiner, Seattle, USA).

For construction of plasmid SV SD4/SA7 NLS-MS2 SRp55pA, the *Bam*HI/*Xho*I fragment of SV NLS-MS2 9G8 (Caputi et al., 2004) was replaced by a PCR product using primers (fw: 5'-GGT GGA TCC CGC ACA AGC CAT AGG CGA TC-3'; rv: 5'-AGA CTC GAG TTA ATC TCT GGA ACT CGA CCT GG-3') using CMV myc SRp55 (kindly provided by A. Cochrane, Toronto, Canada) as a template. Plasmid SV NLS-MS2-ΔSR has been described (Caputi et al., 2004). SV SD4/SA7 NLS-MS2ΔFG-ΔSR and SVSD4/SA7 NLS-MS2 ΔFG-SRp55 were generated by substitution of the *Xma*I/*Eco*RI fragment by a PCR-amplified fragment using primers (fw: 5'-GAC CCC GGG ATG GGG CCG CAA AAA ACG CCG C-3' and rv: 5'- TCG GAA TTC GTA GCG AAA ATT GGA ATG GTT AGT TCC ATA TTT AAG TAC GAA CGC CAG GCG CCT-3'), containing a deletion of the sequence encoding the flexible α-helical loop connecting the two β-strands F and G (FG loop) within the MS2 coat protein, using SV SD4/SA7 NLS-MS2-ΔSR as a template.

Cells and transfections

293T cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, 1 mM sodium pyruvate and 1% ampicillin. The day before transfection, 4 x 10^6 293T cells were seeded in a 10 cm plate. Transfections were performed using the calcium phosphate precipitation method with 5 µg of the SF91 plasmid (10 µg for proviral constructs). Medium was exchanged 8h after transfection, and the cells were harvested after 48h. Transfection efficiency was measured by FACS analysis and ranged between 40-60% (FACSCalibur; Becton-Dickinson, Heidelberg, Germany). The co-transfection assays (MS2 and U1) were performed using 5 µg SF91 plasmid and 10 µg U1 plasmid or 5 µg MS2-plasmid.

HeLaP4 cells (Kimpton & Emerman, 1992) were cultured in the same conditions used for 293T cells. The day before transfection, 6 x 10^5 HeLaP4 cells were seeded in a 6 cm plate. Cells were transfected using the ICAFectin™ 441 DNA transfection reagent (Eurogentec, Brussels, Belgium) and 2.75 µg plasmid DNA. RNA was harvested 36h later.

RNA preparation and analysis

Preparation of total RNA, gel electrophoresis, blotting and detection with a radiolabeled probe were performed as described previously (Wodrich et al., 2001). The eGFP-specific probe was generated by digestion of SF91. To detect 18S rRNA, a genomic fragment was PCR-amplified and subcloned into pCR2.1 (Invitrogen, Karlsruhe, Germany). The probes were radiolabeled using the DecaLabel Kit (Fermentas, St. Leon-Rot, Germany). RNA was quantified photometrically and 10µg were used for Northern blot analysis, if not stated otherwise.

Phosphoimager analysis

The different RNA species were quantified by phosphoimager analysis (Storm 820; GE Healthcare). The percentage of unspliced RNA was calculated using the following formula: (unspliced RNA / (spliced RNA + unspliced RNA)) x 100 = % unspliced RNA

Computer software tools

RNA folding analysis of the leader region and mutants thereof was performed using the mFOLD software (Zuker, 1989); web access: http://mfold.bioinfo.rpi.edu). RNA structures were assembled using the XRNA software (http://rna.ucsc.edu/rnacenter/xrna /xrna. html). Strength of 5'ss was determined using the HBond-score algorithm (http://www.uni-duesseldorf. de/ rna/ html /hbond_ score.php). Statistical analysis was performed using Graph Pad Prism 4 software.

Acknowledgements

We would like to thank M. Garcia-Blanco for critically reading the manuscript and for his helpful suggestions during the course of this work, M. Galla, J. Kraunus, A. Schambach and J. Meyer for their discussion and experimental advice, and T. Schulz for his generous support. This work was supported by DFG grant BO 2512/2-1 to J. Bohne and SCHA 909/2-2 to H. Schaal. Stiftung für AIDS-Forschung, Düsseldorf to H. Schaal. DFG cluster of excellence REBIRTH and SFB 738 project C4 to C. Baum.

References

Abbink TE, Berkhout B. 2008. RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor. *J Virol 82*:3090-3098.

Amendt BA, Hesslein D, Chang LJ, Stoltzfus CM. 1994. Presence of negative and positive cis-acting RNA splicing elements within and flanking the first tat coding exon of human immunodeficiency virus type 1. *Mol Cell Biol 14*:3960-3970.

Amendt BA, Si ZH, Stoltzfus CM. 1995. Presence of exon splicing silencers within human immunodeficiency virus type 1 tat exon 2 and tat-rev exon 3: evidence for inhibition mediated by cellular factors. *Mol Cell Biol 15*:4606-4615.

Blencowe BJ. 2000. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci 25*:106-110.

Bohne J, Krausslich HG. 2004. Mutation of the major 5' splice site renders a CMV-driven HIV-1 proviral clone Tat-dependent: connections between transcription and splicing. *FEBS Lett 563*:113-118.

Bohne J, Schambach A, Zychlinski D. 2007. New way of regulating alternative splicing in retroviruses: the promoter makes a difference. *J Virol 81*:3652-3656.

Bohne J, Wodrich H, Krausslich HG. 2005. Splicing of human immunodeficiency virus RNA is position-dependent suggesting sequential removal of introns from the 5' end. *Nucleic Acids Res 33*:825-837.

Bratt E, Ohman M. 2003. Coordination of editing and splicing of glutamate receptor pre-mRNA. *Rna 9*:309-318.

Buratti E, Baralle FE. 2004. Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol Cell Biol 24*:10505-10514.

Burge CB, Tuschl T, Sharp PA. 1999. Splicing of precursors to mRNAs by the spliceosomes. In: Cech TR, ed. *The RNA world*. Cold Spring Habor, NY. pp 525-560.

Caputi M, Freund M, Kammler S, Asang C, Schaal H. 2004. A bidirectional SF2/ASF- and SRp40-dependent splicing enhancer regulates human immunodeficiency virus type 1 rev, env, vpu, and nef gene expression. *J Virol 78*:6517-6526.

Caputi M, Mayeda A, Krainer AR, Zahler AM. 1999. hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. *Embo J 18*:4060-4067.

Chao JA, Patskovsky Y, Almo SC, Singer RH. 2008. Structural basis for the coevolution of a viral RNA-protein complex. *Nat Struct Mol Biol 15*:103-105.

Cochrane AW, McNally MT, Mouland AJ. 2006. The retrovirus RNA trafficking granule: from birth to maturity. *Retrovirology 3*:18.

Coffin JM, Hughes SH, Varmus HE. 1997. *Retroviruses*. Cold Spring Harbor, NY.

Crispino JD, Blencowe BJ, Sharp PA. 1994. Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science 265*:1866-1869.

Damgaard CK, Kahns S, Lykke-Andersen S, Nielsen AL, Jensen TH, Kjems J. 2008. A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Mol Cell 29*:271-278.

Dirksen WP, Hampson RK, Sun Q, Rottman FM. 1994. A purine-rich exon sequence enhances alternative splicing of bovine growth hormone pre-mRNA. *J Biol Chem 269*:6431-6436.

Du H, Rosbash M. 2002. The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature 419*:86-90.

Eperon LP, Graham IR, Griffiths AD, Eperon IC. 1988. Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? *Cell 54*:393-401.

Erlwein O, Buchholz CJ, Schnierle BS. 2003. The proline-rich region of the ecotropic Moloney murine leukaemia virus envelope protein tolerates the insertion of the green fluorescent protein and allows the generation of replication-competent virus. *J Gen Virol 84*:369-373.

Fong YW, Zhou Q. 2001. Stimulatory effect of splicing factors on transcriptional elongation. *Nature 414*:929-933.

Freund M, Asang C, Kammler S, Konermann C, Krummheuer J, Hipp M, Meyer I, Gierling W, Theiss S, Preuss T, Schindler D, Kjems J, Schaal H. 2003. A novel approach to describe a U1 snRNA binding site. *Nucleic Acids Res 31*:6963-6975.

Freund M, Hicks MJ, Konermann C, Otte M, Hertel KJ, Schaal H. 2005. Extended base pair complementarity between U1 snRNA and the 5' splice site does not inhibit splicing in higher eukaryotes, but rather increases 5' splice site recognition. *Nucleic Acids Res 33*:5112-5119.

Fu XY, Manley JL. 1987. Factors influencing alternative splice site utilization in vivo. *Mol Cell Biol 7*:738-748.

Furger A, O_Sullivan JM, Binnie A, Lee BA, Proudfoot NJ. 2002. Promoter proximal splice sites enhance transcription. *Genes & Development 16*:2792-2799.

Grover A, Houlden H, Baker M, Adamson J, Lewis J, Prihar G, Pickering-Brown S, Duff K, Hutton M. 1999. 5' splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10. *J Biol Chem 274*:15134-15143.

Hartmann L, Theiss S, Niederacher D, Schaal H. 2008. Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? *Front Biosci 13*:3252-3272.

Hastings ML, Krainer AR. 2001. Pre-mRNA splicing in the new millennium. *Curr Opin Cell Biol 13*:302-309.

Hertel KJ. 2008. Combinatorial control of exon recognition. *J Biol Chem 283*:1211-1215.

Hildinger M, Abel KL, Ostertag W, Baum C. 1999. Design of 5' untranslated sequences in retroviral vectors developed for medical use. *Journal of Virology 73*:4083-4089.

Hiller M, Zhang Z, Backofen R, Stamm S. 2007. Pre-mRNA secondary structures influence exon recognition. *PLoS genetics 3*:e204.

Ibrahim el C, Hims MM, Shomron N, Burge CB, Slaugenhaupt SA, Reed R. 2007. Weak definition of IKBKAP exon 20 leads to aberrant splicing in familial dysautonomia. *Human mutation 28*:41-53.

Jablonski JA, Buratti E, Stuani C, Caputi M. 2008. The secondary structure of the human immunodeficiency virus type 1 transcript modulates viral splicing and infectivity. *J Virol 82*:8038-8050.

Jacquenet S, Ropers D, Bilodeau PS, Damier L, Mougin A, Stoltzfus CM, Branlant C. 2001. Conserved stem-loop structures in the HIV-1 RNA region containing the A3 3' splice site and its cis-regulatory element: possible involvement in RNA splicing. *Nucleic Acids Res 29*:464-478.

Kammler S, Leurs C, Freund M, Krummheuer J, Seidel K, Tange TO, Lund MK, Kjems J, Scheid A, Schaal H. 2001. The sequence complementarity between HIV-1 5' splice site SD4 and U1 snRNA determines the steady-state level of an unstable env pre-mRNA. *Rna 7*:421-434.

Kammler S, Otte M, Hauber I, Kjems J, Hauber J, Schaal H. 2006. The strength of the HIV-1 3' splice sites affects Rev function. *Retrovirology 3*:89.

Kimpton J, Emerman M. 1992. Detection of replication-competent and pseudotyped human immunodeficiency virus with a sensitive cell line on the basis of activation of an integrated beta-galactosidase gene. *J Virol 66*:2232-2239.

Konarska MM, Query CC. 2005. Insights into the mechanisms of splicing: more lessons from the ribosome. *Genes Dev 19*:2255-2260.

Kornblihtt AR, de la Mata M, Fededa JP, Munoz MJ, Nogues G. 2004. Multiple links between transcription and splicing. *Rna 10*:1489-1498.

Kraunus J, Zychlinski D, Heise T, Galla M, Bohne J, Baum C. 2006. Murine leukemia virus regulates alternative splicing through sequences upstream of the 5' splice site. *J Biol Chem 281*:37381-37390.

Lee JT, Yu SS, Han E, Kim S. 2004. Engineering the splice acceptor for improved gene expression and viral titer in an MLV-based retroviral vector. *Gene Ther 11*:94-99.

Logg CR, Baranick BT, Lemp NA, Kasahara N. 2007. Adaptive evolution of a tagged chimeric gammaretrovirus: identification of novel cis-acting elements that modulate splicing. *J Mol Biol 369*:1214-1229.

Lund M, Kjems J. 2002. Defining a 5' splice site by functional selection in the presence and absence of U1 snRNA 5' end. *Rna 8*:166-179.

Maniatis T, Tasic B. 2002. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature 418*:236-243.

Marchand V, Mereau A, Jacquenet S, Thomas D, Mougin A, Gattoni R, Stevenin J, Branlant C. 2002. A Janus splicing regulatory element modulates HIV-1 tat and rev mRNA production by coordination of hnRNP A1 cooperative binding. *J Mol Biol 323*:629-652.

McNally LM, McNally MT. 1999. U1 small nuclear ribonucleoprotein and splicing inhibition by the rous sarcoma virus negative regulator of splicing element. *J Virol 73*:2385-2393.

McNally MT, Beemon K. 1992. Intronic sequences and 3' splice sites control Rous sarcoma virus RNA splicing. *J Virol 66*:6-11.

McNally MT, Gontarek RR, Beemon K. 1991. Characterization of Rous sarcoma virus intronic sequences that negatively regulate splicing. *Virology 185*:99-108.

Mougel M, Tounekti N, Darlix JL, Paoletti J, Ehresmann B, Ehresmann C. 1993. Conformational analysis of the 5' leader and the gag initiation site of Mo-MuLV RNA and allosteric transitions induced by dimerization. *Nucleic Acids Res 21*:4677-4684.

Mount SM, Pettersson I, Hinterberger M, Karmas A, Steitz JA. 1983. The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. *Cell 33*:509-518.

O'Reilly MM, McNally MT, Beemon KL. 1995. Two strong 5' splice sites and competing, suboptimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA. *Virology 213*:373-385.

Peabody DS. 1993. The RNA binding site of bacteriophage MS2 coat protein. *Embo J 12*:595-600.

Pollard VW, Malim MH. 1998. The HIV-1 Rev protein. *Annu Rev Microbiol 52*:491-532.

Rinke J, Appel B, Blocker H, Frank R, Luhrmann R. 1984. The 5'-terminal sequence of U1 RNA complementary to the consensus 5' splice site of hnRNA is single-stranded in intact U1 snRNP particles. *Nucleic Acids Res 12*:4111-4126.

Schambach A, Mueller D, Galla M, Verstegen MM, Wagemaker G, Loew R, Baum C, Bohne J. 2006. Overcoming promoter competition in packaging cells improves production of self-inactivating retroviral vectors. *Gene Ther 13*:1524-1533.

Selby MJ, Peterlin BM. 1990. Trans-activation by HIV-1 Tat via a heterologous RNA binding protein. *Cell 62*:769-776.

Seraphin B, Kretzner L, Rosbash M. 1988. A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. *Embo J 7*:2533-2538.

Shen H, Green MR. 2006. RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. *Genes Dev 20*:1755-1765.

Shepard PJ, Hertel KJ. 2008. Conserved RNA secondary structures promote alternative splicing. *Rna 14*:1463-1469.

Singh NN, Singh RN, Androphy EJ. 2007. Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. *Nucleic Acids Res 35*:371-389.

Slaugenhaupt SA, Blumenfeld A, Gill SP, Leyne M, Mull J, Cuajungco MP, Liebert CB, Chadwick B, Idelson M, Reznik L, Robbins C, Makalowska I, Brownstein M, Krappmann D, Scheidereit C, Maayan C, Axelrod FB, Gusella JF. 2001. Tissue-specific expression of a splicing mutation in the IKBKAP gene causes familial dysautonomia. *American journal of human genetics 68*:598-605.

Solnick D. 1985. Alternative splicing caused by RNA secondary structure. *Cell 43*:667-676.

Staley JP, Guthrie C. 1998. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell 92*:315-326.

Staley JP, Guthrie C. 1999. An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol Cell 3*:55-64.

Stoltzfus CM, Madsen JM. 2006. Role of viral splicing elements and cellular RNA binding proteins in regulation of HIV-1 alternative RNA splicing. *Curr HIV Res 4*:43-55.

Tange TO, Damgaard CK, Guth S, Valcarcel J, Kjems J. 2001. The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *Embo J 20*:5748-5758.

Tange TO, Kjems J. 2001. SF2/ASF binds to a splicing enhancer in the third HIV-1 tat exon and stimulates U2AF binding independently of the RS domain. *J Mol Biol 312*:649-662.

Valadkhan S. 2007. The spliceosome: a ribozyme at heart? *Biol Chem 388*:693-697.

Valcarcel J, Gaur RK, Singh R, Green MR. 1996. Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science 273*:1706-1709.

Will CL, Lührmann R. 2006. Spliceosome Structure and Function. *In The RNA World*: CSHL Press.

Wodrich H, Bohne J, Gumz E, Welker R, Krausslich HG. 2001. A new RNA element located in the coding region of a murine endogenous retrovirus can functionally replace the Rev/Rev-responsive element system in human immunodeficiency virus type 1 Gag expression. *J Virol 75*:10670-10682.

Zhu J, Mayeda A, Krainer AR. 2001. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell 8*:1351-1361.

Zhuang Y, Weiner AM. 1986. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell 46*:827-835.

Zuker M. 1989. Computer prediction of RNA structure. *Methods Enzymol 180*:262-288.

Supplementary figure 1

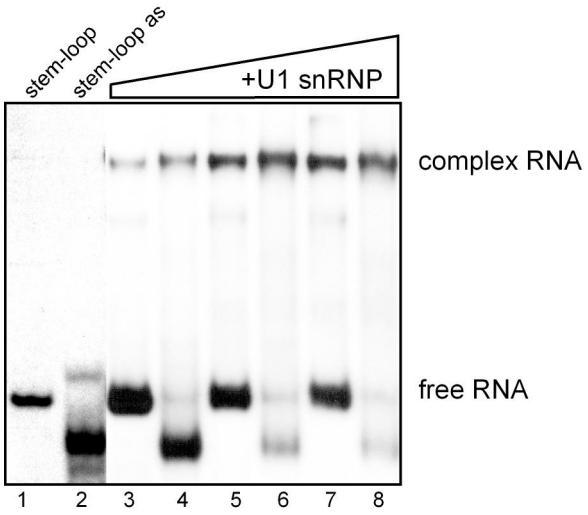


Figure legend

Folded free RNA was incubated with increasing amounts of purified U1snRNP (50 ng, 125 ng and 250 ng). The native gel readily reveals the folding of stem-loop containing RNA in contrast to the control where one part of the structure is present in antisense orientation. In the presence of U1snRNP complex formation takes place. At higher concentrations almost all of the free antisense RNA is bound to U1snRNP (lane 8). The stem-loop construct also bind U1snRNP, but with a lower efficiency.

Supplementary Methods

Using SF91stem-loop and stem-loop antisense templates PCR products were generated carrying a T7 promoter at the 5'end and the 5'ss at the 3'end. The products were ligated into pCR2.1. The resulting plasmids were linearized and approximately 500 ng were used for in-vitro transcription using Ambion's Maxiscript kit in the presence of 50 µCi of [32P]UTP (400 Ci/mmol, Hartmann Analytic) and unlabeled UTP to final concentration of 40 µM. Free nucleotides were removed by MoBiTec S300 columns and probes were purified using denaturing PAGE. The radiolabeled RNA Probe (25000 Counts) was incubated in 25 mM Tris-HCl, pH 7.9, 5mM $MgCl_2$, 10% (v/v) glycerol, 0.4 mM dithiothreitol, 0.5 mM EDTA, 10U RNAse Inhibitor, 4µg BSA and 1 µg of tRNA in a total volume of 15 µl for 30 min at 20 °C after denaturating for 2 min at 90°C to allow folding of the RNA structure. Purified U1snRNP (Phadia, Freiburg, Germany) were added 10 min prior loading on a 6% (60:1) polyacrylamid gel run in Tris-borate-EDTA buffer.

F Discussion & Outlook

The following section combines the results of publication 2 and the manuscript into an integrated model, explaining the function of the MLV leader region in the viral life cycle.

Alternative splicing is an essential step in the gene expression of all retroviruses and is carried out by the cellular spliceosome. As mentioned before the splicing regulation of MLV is poorly understood. Several *cis*-acting elements have been suggested to control alternative splicing in simple gammaretroviruses. A deletion of sequences downstream of the 3'ss resulted in a total loss of spliced RNAs (Oshima *et al.*, 1998), whereas the deletion of an element located 140 nt upstream of the branchpoint showed more spliced RNA (Hwang *et al.*, 1984). All these possible *cis*-acting sequences are not present in the SF91 vector, which still displays alternative splicing. Furthermore, splicing analysis of the SINSF91 vector (see above) showed that both splice sites per se are efficient, arguing that additional *cis*-acting sequences are required for the splice inhibitory effect.

Hide and seek: secondary structure and splicing control

The present study demonstrated that MLV uses a dynamic interplay between RNA secondary structure and primary 5'ss sequence to restrict the access of the U1 snRNP to its 5'ss, which leads to alternative splicing of the viral RNA. In general, RNA forms highly stable secondary and tertiary structures *in vitro* and *in vivo* (Brion & Westhof, 1997; Conn & Draper, 1998; Fontana *et al.*, 1993). It has been observed that altering of these structures represents a well-known regulatory mechanism for many cellular processes (Klaff *et al.*, 1996). Various RNA binding proteins have been found to bind single-stranded RNA in a sequence-specific manner (Antson, 2000; Chabot *et al.*, 1997; Perez-Canadillas & Varani, 2001). Others recognize double-stranded RNA independently of its sequence (Carlson *et al.*, 2003). The binding of several positive (SRp55, B52 and NOVA-1) and negative (hnRNP A1) splicing regulators depends on the primary sequence and in some cases on RNA secondary structure (Buckanovich & Darnell, 1997; Damgaard *et al.*, 2002; Jablonski *et al.*, 2008; Nagel *et al.*, 1998; Shi *et al.*, 1997). In the manuscript we have shown, by

replacing the wild type secondary structure with a heterologous stem loop, that MLV requires no further splicing regulatory proteins for 5'ss attenuation.

The most common mechanism of splicing regulation by secondary structures is a reduced accessibility of the conserved splice sites for splicing factors, as shown in humans for the *tau* gene (Grover *et al.*, 1999), the *Hprt* gene (Hennig *et al.*, 1995; Tu *et al.*, 2000), and the *hnRNP A1* gene (Blanchette & Chabot, 1997). Here the attenuated splice sites are part of the secondary structure and are not accessible to the spliceosome. However, in the case of MLV the 5'ss is located at the bottom of a stem and not completely paired, leading to the question: How is binding of U1 snRNA to the 5'ss restricted in MLV? In a first model, it is suggested that MLV genomic RNA may exist in two forms. A certain percentage of RNA molecules may form the described stem loop structure (Mougel *et al.*, 1993) in which the 5'ss cannot be recognized by the spliceosome. Another portion of the RNA molecules may exist with a relatively unfolded structure in which splicing can occur. Thus, the ratio of folded to unfolded RNA, i.e. the proportion of stem loop structure within the population may determine the accessibility of the 5'ss to U1 snRNA resulting in the production of a defined amount of spliced RNAs. In a second model the RNA is always folded into a stem loop structure. Here, recognition of the 5'ss is a very rare event, because of steric hindrance and the fact that only 4 positions are available for U1 snRNA binding. Binding to the other complementary positions, including the invariant positions +1 and +2 (GU), requires of the secondary structure to be dissolved. Here, a dynamic competition between base pairing of the 5'ss with U1 snRNA and sequestering of the 5'ss into the secondary structure could take place.

Both models fit to the results of the delPBS mutant and the extended stem loop mutants, where we observed more unspliced RNA by strengthening the stem structure. For the second model the extended or continuous stem increases the free energy of the stem, leading to reduced recognition and binding of U1 to the 5'ss. Moreover, it is likely that by strengthening the stem structure the ratio of folded to unfolded RNA species is shifted to folded RNA, also leading to more unspliced RNA in the second model. The observation that the 5'ss alone is strong enough to drive complete splicing, as shown for the stem loop antisense mutant, suggested that an unfolded leader region with an accessible MLV 5'ss would also result in complete splicing. This indicates that by using the second model we could only enhance

splicing by increasing the ratio of folded to unfolded RNA and not by increasing the complementarity of the 5'ss to U1. In contrast to this we have shown, using hyperstable U1 snRNA mutants, that we can enhance splicing by increasing the complementarity of U1 snRNA to the 5'ss. The same was observed for the C6U mutant, where position +6 of the primary splice site sequence is converted to the consensus sequence increasing the complementarity. These results apply to the second model, whereas additional base pairing of U1 snRNA to the 5'ss could facilitate the dissolving of the intramolecular paired splice site positions leading to enhanced splicing (Fig.11). Furthermore, the results of the primary splice site mutants +7 and +8, in which the splicing activity was again increased, support the hypothesis that splice sites lacking complementarity to U1 snRNA on their 5'end can be rescued by increasing the complementarity at their 3'end (Hartmann *et al.*, 2008).

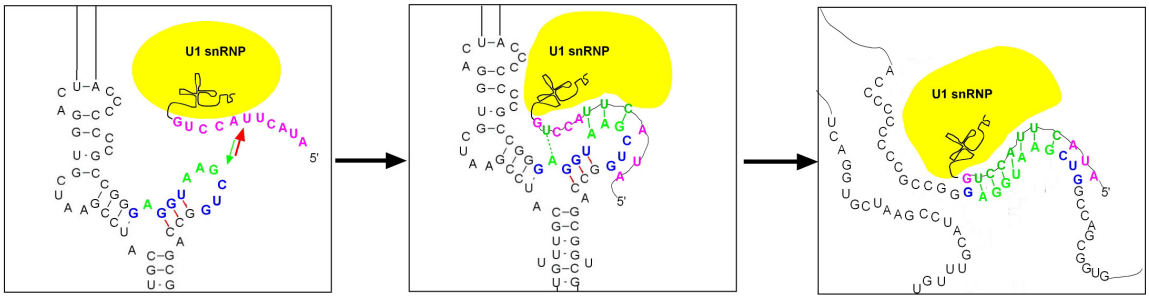


Figure 11: Reduced accessibility to U1 snRNA results in alternative splicing

The figure shows model 2, in which the MLV 5'ss (blue/green) is located at the bottom of the stem loop structure. Four possible binding sites (green) of the 5'ss are accessible for U1 snRNA (pink) binding, leading to reduced recognition of the 5'ss by the spliceosome. Base pairing to the four binding sites dissolves the stem loop structure. This allows base pairing at three additional positions of the 5'ss to U1 snRNA, including the invariant positions +1 and +2 (GU).

Moreover the results of the MS2-fusion tethering system strengthen the second model, showing that the RS domain of SRp55 increase splicing efficiency about 2-fold. SR proteins are not present in *Saccharomyces cerevisiae*, and probably not needed due to highly conserved consensus sequences of the splice sites. Recently it has been demonstrated that SR proteins can rescue mutated 5' and 3'ss in *S. cerevisiae* possible by stabilizing RNA:RNA interactions between U1 snRNA and 5'ss (Shen & Green, 2006). The authors suggest a common mechanism for one of the RS domain functions from yeast to humans. Since SR proteins contain an alternating arrangement of positively and negatively charged argenine and serine

residues, we hypothesize that the basic amino acid side chains may neutralize the negatively charged phosphates of the nucleic acids, thereby facilitating pairing of two RNA strands; in this case U1 snRNA and the 5'ss. By using the MS2-fusion tethering system we tested different RS domains and observed the strongest enhancement of splicing with the long and continuous RS domain from SRp55 and not with four copies of the SF2/ASF RS domain. With this experiment we did not increase complementarity of U1 snRNA with the 5'ss, but we facilitated annealing of U1 snRNA to a partially sequestered 5'ss. These results indicate for the first time that RS domains probably stabilize RNA:RNA interactions in mammals.

In conclusion we hypothesize that MLV hides its 5'ss by the stem loop structure to prevent binding of U1 snRNA and that this inefficient recognition leads to the alternative splicing in MLV.

The PBS loop is crucial for infection

In addition to the stem loop mutants we have designed a SF91PBScomp mutant, introducing additional complementary nucleotides on the opposite side of the PBS. Compared to the delPBS, the PBScomp strengthened the stem by adding 16 bp, leading to more unspliced RNA. Interestingly, viral particles produced with the PBScomp mutant were almost non-infectious. It has been shown by our lab that reverse transcription-deficient retroviral vectors, like the SF91 delPBS, mediate ectopic expression of proteins in target cells in a process called retroviral particle-mediated mRNA transfer (RMT) (Galla *et al.*, 2004). Here, infection delivers retroviral RNA, which serves directly as a template for translation. The delPBS cannot bind the tRNA and initiate reverse transcription, allowing RMT to occur. After infection with the PBScomp we observed GFP expression by RMT, showing that the mutated genomic RNA is packaged into viral particles during virus production. In contrast to the delPBS, the PBScomp still contains the 18 bp necessary for tRNA primer binding. To control for any negative effects of the additional 16 bp, a mutant was designed containing 16 non-complementary base pairs. In contrast to the PBScomp the control mutant is able to generate infectious particles, showing that the tRNA binds to the PBS and RNA is reverse transcribed into DNA. Consequently we suggest that the PBS has to be looped out for efficient tRNA primer binding, and that reverse transcription is inhibited when the PBS becomes a part of the stem (Fig.14). Further

experiments have to be performed to confirm this hypothesis. A simple, but relatively indirect approach to show that no tRNA primer can bind to the paired PBS, is the analysis of strong stop DNA (Fig.3).

Does MLV also regulate the second step of spliceosome assembly?

We have shown that U1 snRNA binding is disturbed by the secondary structure of the MLV leader region and that we can enhance splicing by increasing the number of possible binding sites of U1 snRNA to the 5'ss. In addition to the U1 snRNA suppressor mutants shown in Figure 5 of the manuscript, we mutated the corresponding position to the C6U mutant in the U1 snRNA (U1 A3G) in order to increase the number of possible binding sites from 7 to 8 (Fig.12).

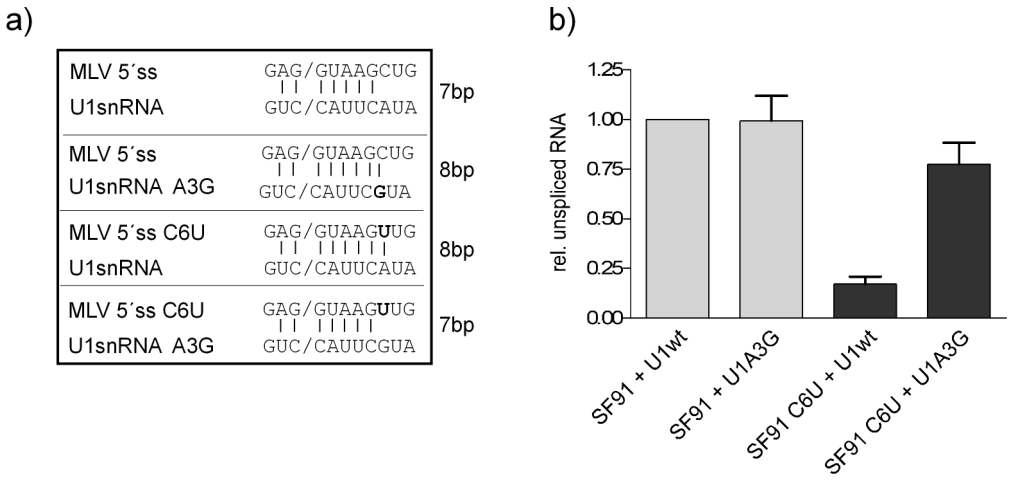


Figure 12: Mutation of +3 in U1 snRNA corresponding to +6 shows no enhanced splicing

a) The top line shows the 5'ss pairing to U1 snRNA in wild type MLV. Possible base pairs with U1 snRNA are shown as solid vertical lines. The next panel shows the new U1 snRNA mutant, and the two lower panels show pairing between the U1 snRNA wt and the U1 snRNA mutant with the SF91 5'ss mutant on position +6 (SF91 C6U). Mutated nucleotides in U1 snRNA and in SF91 are in bold. On the right-hand side, the number of possible binding sites is indicated. b) Phosphoimager analysis. The efficiency of splicing is shown as the relative level of unspliced RNA. SF91 in the presence of co-transfected U1 wt plasmid was set as 1. Mean values and standard deviations represent five independent experiments.

Surprisingly, we could not observe an enhancement of splicing for this mutant in comparison to the parental C6U mutant. The mutant in U1 snRNA shows the same splicing ratio as the wild type (Fig.12). There are two possible explanations for these results. The first and very simple explanation is that this U1 snRNA mutant has an

unknown defect. In a control experiment we co-transfected the new U1 snRNA mutant with the SF91 C6U splice site mutant. It is theoretically possible that a G·U non Watson-Crick-pair could form on position +6, but it has been described that non Watson-Crick-pairs only form when they are adjacent to Watson-Crick-pairs (Freund *et al.*, 2003). This is not the case in our mutant at position +7. Additionally by using this mutant pair we reduced the complementarity of U1 snRNA with the 5'ss from 8 to 7, i.e. to the same level as in the wild type. As a result we restored the level of unspliced RNA nearly to the wild type level (Fig.12). This indicates that the mutant is generally functional.

The 5'ss consensus sequence in higher eukaryotes corresponds to perfect Watson–Crick base pairing to the U1 snRNA 5'terminus (Horowitz & Krainer, 1994) and plays a critical role in 5'ss selection (Seraphin *et al.*, 1988; Siliciano & Guthrie, 1988; Zhuang & Weiner, 1986). After U1 recognizes the 5'ss, a conformational rearrangement during spliceosome assembly results in the replacement of U1 by U6 snRNA, which has been proposed to contribute to the high fidelity of the reaction (Staley & Guthrie, 1999). It has been reported that U6 snRNA and high concentrations of SR proteins can substitute for the absence, or limiting amount, of U1 snRNA and restore splicing in U1-depleted extracts (Crispino *et al.*, 1994; Crispino & Sharp, 1995; Tarn & Steitz, 1994). Thereby the SR proteins promote the association of U2 snRNP with the branchpoint. The distal nucleotides in the 5'ss (+2 to +6) may solely be recognized by U6 snRNA with the assistance of SR proteins. However, a second explanation for the unexpected results of the U1 snRNA mutant could be that the SF91 C6U splice site mutant displays an increased complementarity to both U1 and U6 snRNA (Fig.13). In the wild type position +6 can neither be recognized by U1 nor by U6 snRNA leading to 7 bp complementarity for U1 and 2 bp for U6 snRNA, respectively. If we mutate the corresponding position in U1 snRNA we increase the binding sites from 7 bp to 8 bp, whereas U6 binding is not affected and still exhibits only 2 possible binding sites. In contrast, the replacement of position +6 with the primary 5'ss consensus sequence leads to an increase of one Watson-Crick-basepair for both U1 snRNA and U6 snRNA. It is possible that the differences between the two situations on position +6 are not due to the increase in U1 snRNA binding to the 5'ss, but rather that the replacement of U1 snRNA by U6 snRNA in the second step of splicing is facilitated.

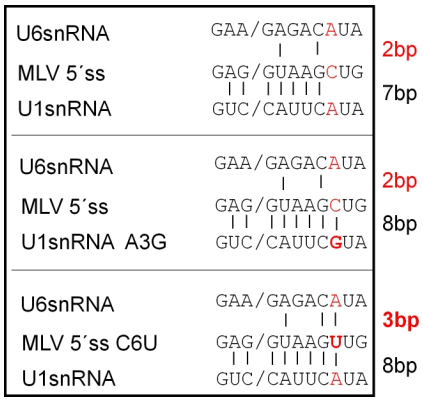


Figure 13: Watson-Crick-base pairing of U6 snRNA to the 5'ss of MLV

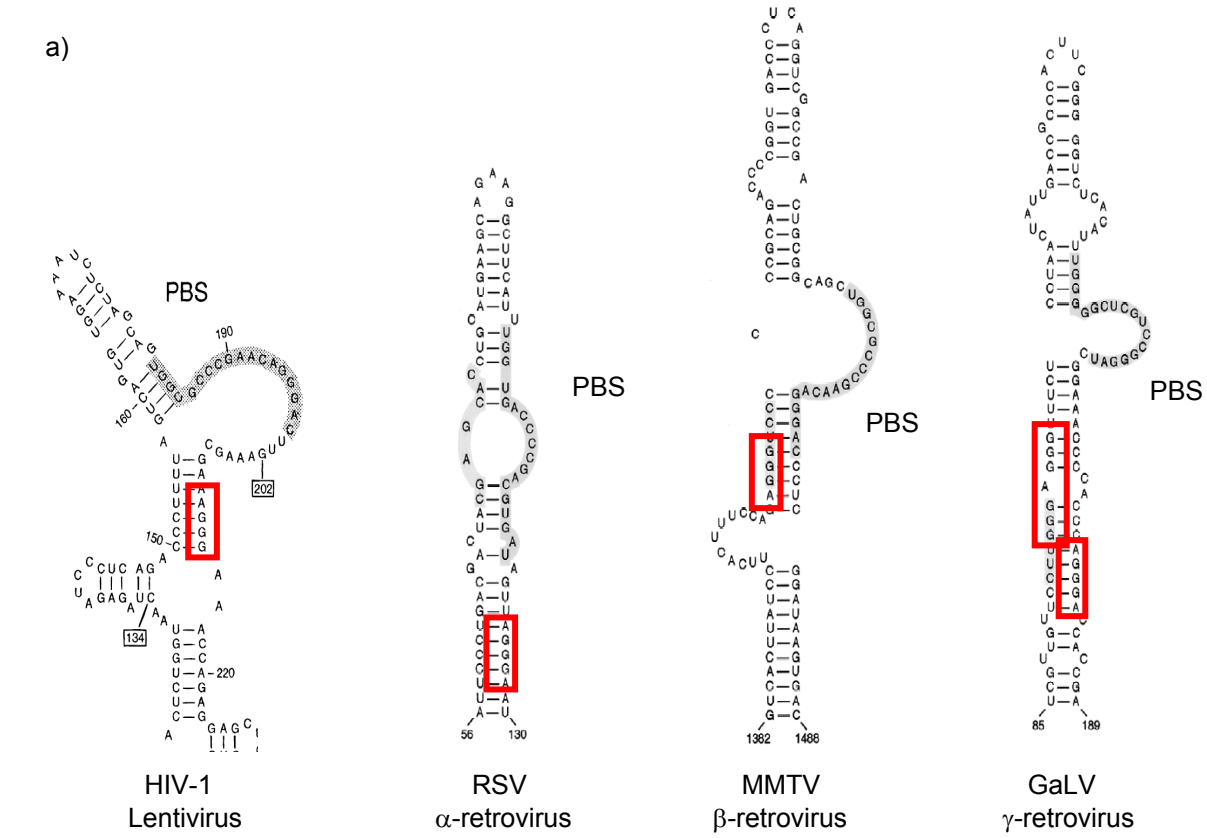
Possible base pairs to U1 snRNA and U6 snRNA are shown as solid vertical lines. The next panel shows SF91 wt combined with the U1 snRNA A3G mutant. The lower panel shows pairing of U1 snRNA wt with the SF91 5'ss mutant on position +6 (SF91 C6U). Mutated nucleotides in U1 snRNA and in SF91 are in bold and position +6 and corresponding positions in U1 and U6 snRNA are in red. On the right-hand side, the possible binding sites to U6 snRNA are indicated in red and to U1 snRNA in black.

To test our hypothesis, we replaced the corresponding nucleotides in U6 snRNA in the same manner as we had done for the U1 mutant. However, preliminary data from co-transfection of SF91 wt with U6 wt showed that over-expressing U6 snRNA is not neutral. We observed enhanced splicing compared to our control experiment using SF91 wt + pUC19. Moreover, the critical position of U6 snRNA binds intermolecularly to U2 snRNA and it is likely that changing the primary sequence leads to difficulties during spliceosomal assembly (Sun & Manley, 1995). For future work new experiments have to be designed. It is possible that we could establish an *in vitro* cross-linking assay for U6 snRNA (Valcarcel *et al.*, 1996).

Is that all? Or is there a third layer of splicing regulation?

All retroviruses harbor a highly structured RNA leader region, in which only the PBS is looped out to bind a cellular tRNA. In order to find conserved sequence motifs we analyzed various leader regions of different retroviruses and observed GGGA motifs in several genera (Fig.14). Such sequences are usually recognition motifs for RNA binding proteins belonging to the hnRNP family. By using two different prediction tools for RNA binding proteins (http://www.ebi.ac.uk/asd-srv/wb.cgi?method=8 and http://www.introni.it/splicing.html), we recognized several putative hnRNP recognition motifs in the MLV leader region adjacent to the PBS and 5'ss. Interestingly, the ascending side of the stem exhibits only one potential binding site. In addition to the stem mutants of publication 2 (Fig.7) we destroyed this putative recognition motif in the same manner as we did for the sm3/sm3comp mutants. As expected, the sm4 mutant alone showed the same splicing enhancement as the sm3 mutant. Surprisingly, the sm4comp could not restore the splicing pattern compared to sm3comp and still showed complete splicing. This result led to the hypothesis that binding of an hnRNP protein is prevented, resulting in enhanced splicing. This surprising observation did not match to the results of the artificial stem loop, which suggested that no additional splicing regulating proteins, such as hnRNPs or SR proteins are necessary for regulating alternative splicing in MLV. It is likely that the wild type stem loop of MLV needs additional proteins for stabilization, due to non-paired regions like the PBS. In contrast, the heterologous artificial stem probably needs no additional proteins because the stem is continuous.

Moreover, it has been described that proteins of the hnRNP family can self-interact and form intermolecular dimers (hnRNP duplex) if they are bound on opposite sides to an RNA stem (Martinez-Contreras *et al.*, 2006; Nasim *et al.*, 2002). It is likely that MLV uses hnRNP duplex formation to stabilize the stem structure adjacent to the looped out PBS. Hence, it is likely that by destroying the motif on the left and right side (sm4comp) we inhibited potential duplex formation of two hnRNP proteins (Fig.15). This would then disrupt the stem upstream of the 5'ss, finally leading to enhanced splicing.



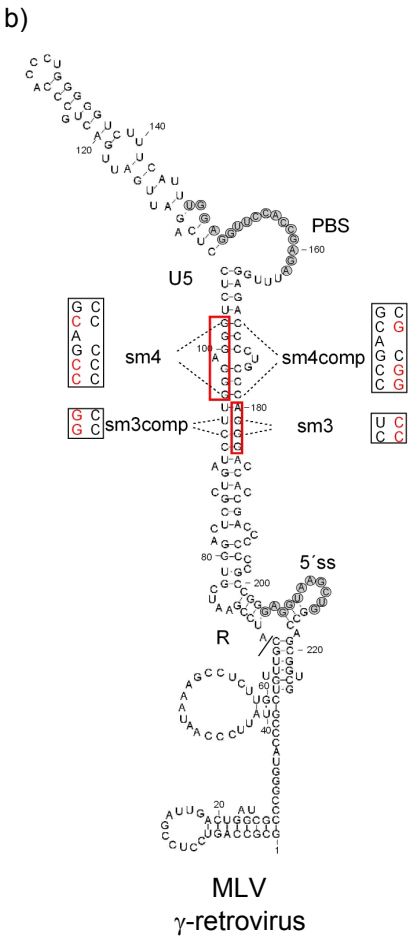


Figure 14: Conserved GGGA motifs

a) Depiction of various leader sequences of different retroviruses including GGGA sequence elements, boxed in red. A relatively conserved motif is located near the PBS loop either on the left or on the right side of the stem, adapted from (Berkhout & van Wamel, 2000; Beerens & Berkhout, 2002). b) RNA secondary structure of MLV. The structure is adapted from the chemically validated structure (Mougel *et al.*, 1993). GGGA sequences are boxed in red. The mutations destroying the GGGA sequence on the right side of the stem; sm3 and sm3comp are analyzed in publication 2 Figure 7. The mutant pair sm4/sm4comp is shown upstream of the stem, destroying the GGGAGGG motif on the left side of the stem in the same manner as for the sm3/sm3comp pair.

Further experiments are needed, including disrupting other potential hnRNP motifs that probably allow hnRNP duplex formation (upstream of the 5'ss and upstream the PBS) and siRNA approaches to elucidate a potential involvement of hnRNPs in the stabilization of the stem structure. Additionally, a combination of the sm4comp and the deletion of the PBS would be interesting, since such a mutant would probably need no additional hnRNPs, due to its continuous stem structure.

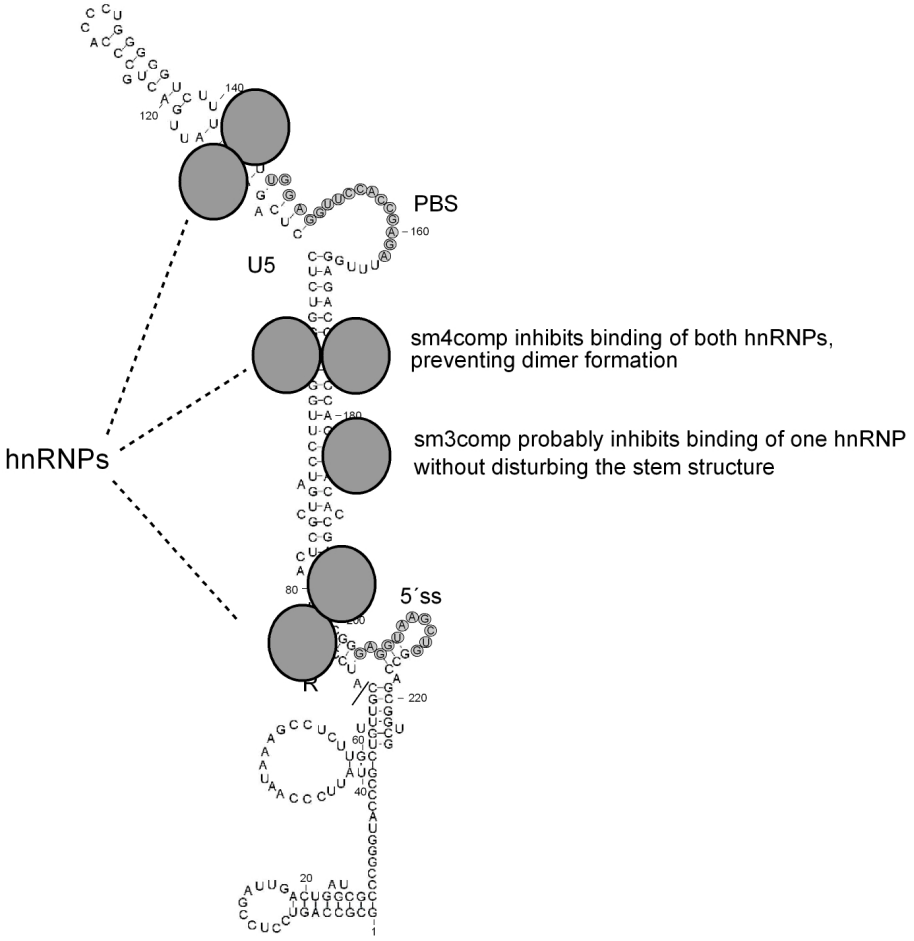


Figure 15: Putative hnRNP dimers stabilize the retroviral stem loop structure

Depiction of the stem-loop structure of MLV. Bioinformatic analysis of the primary sequence indicates putative hnRNP binding sides. HnRNPs are shown as grey circles. Three possible hnRNP dimer formations are indicated, whereas the putative hnRNP recognition motif, destroyed by the sm3comp mutation, has no counterpart on the opposite side of the stem.

Why so complicated?

During the viral life cycle unspliced or incompletely spliced viral RNA species must be exported to the cytoplasm. Usually, incompletely spliced RNAs are retained in the nucleus and degraded (Fasken & Corbett, 2005; Jensen *et al.*, 2003). Importantly, all retroviruses have to overcome this quality control and have evolved several strategies to circumvent the nuclear retention/degradation of their unspliced RNA (see introduction). As mentioned before the export mechanism of unspliced RNA from MLV and other gammaretroviruses has still not been elucidated; additionally, new data from our lab indicate that the stem loop structure of the R region (RSL) is not involved in the export of unspliced RNA as was previously suggested (Trubetskoy *et al.*, 1999). Since MLV expresses no accessory proteins such as Rev, it might have evolved a different strategy to escape the quality control of unused splice sites by the nuclear pore complex. The present study showed that in contrast to other retroviruses that use suboptimal 3'ss for differential gene expression MLV attenuates its 5'ss. What might be the reason for MLV to develop such a strategy? Our hypothesis is that by preventing the recognition of the 5'ss by U1 snRNP, MLV renders a defined portion of its genomic RNA intron-less to circumvent the cellular quality control. Alternative splicing control via a weak 3'ss without any export factors would not result in export of the unspliced RNA, as 5'ss recognition by U1 snRNP occurs co-transcriptionally and this would lead to mRNA retention and degradation.

For future approaches we want to use a mutant that results in complete splicing of the genomic RNA, like the sm3 mutant. This mutation will be combined with mutations that inactivate the actual 3'ss and two additional cryptic sites. All these modifications will result in an RNA in which the 5'ss can efficiently bind U1 snRNP, but which is not able to splice due to the lack of 3'ss. If our hypothesis is true we would generate an RNA that should be blocked in nuclear export, due to the recognition of the 5'ss by U1 snRNA.

Conclusions of the secondary structure model

We have shown that MLV attenuates its 5'ss and not its 3'ss to ensure alternative splicing. During evolution retroviruses had to solve two problems related to their genome organization. Since retroviruses synthesize a polycistronic pre-mRNA they have to export intron-containing RNA. MLV may solve the export problem by masking unspliced RNA or, in other words, the unspliced genomic RNA is not recognized as an intron-containing RNA and slowly exported. In addition to the export of the genomic RNA, a co-expression strategy is needed to express all viral proteins. Our hypothesis is that MLV evolved this complex splice regulation to ensure both export of its genomic RNA and alternative splicing. Furthermore, the stem loop structure is important for reverse transcription. We suggest that for efficient tRNA primer binding the PBS has to be looped out. These results fit to our hypothesis that MLV probably needs additional RNA binding proteins to stabilize the stem loop structure surrounding the PBS. In all retrovirus genera secondary structure analysis indicates that the PBS is looped out (Fig.14). We hypothesize that in general the stem loop structure functions as a scaffold to loop out the PBS and that this mechanism is conserved throughout all retroviruses.

Furthermore, we compared the primary sequences of MLV, the gibbon ape leukemia virus (GaLV), feline leukemia virus (FeLV), spleen focus forming virus (SFFV), koala retrovirus (KoRV) and the newly identified xenotropic MuLV-related virus (XMRV), which can produce authenticated human infections (Urisman *et al.*, 2006). The alignment showed that the leader region is highly conserved compared to other regions like the packaging signal which is very heterogeneous. In addition, the primary sequence of the 5'ss is highly conserved (Fig.16). All viruses exhibit a guanosine at position +6 instead of the conserved uracil and only have 7 possible binding sites of the 5'ss to U1 snRNA, including position –2 to +5.

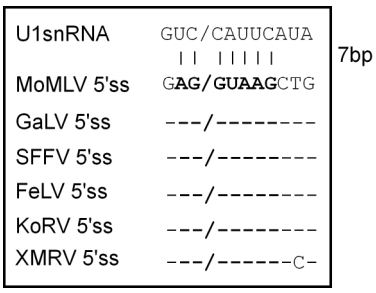


Figure 16: Comparison of 5'ss sequences of gammaretroviruses

The comparison of the 5'ss of various gammaretroviruses show high conservation and all viruses exhibit seven possible base pairs to U1 snRNA.

Using the Mfold software (http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi) we could show that the stem loop structure of the leader region is conserved, including the looped out PBS and the 5'ss at the bottom of the stem (data not shown). These results indicate that gammaretroviruses in general attenuate their 5'ss to ensure alternative splicing in the same manner as we have shown with the present study for MLV. Furthermore, it is likely that all gammaretroviruses use the same export strategy by masking their intron-containing genomic RNA as intron-less.

In conclusion, the stem loop structure of the leader region retroviruses is crucial for the viral life cycle. Gammretroviruses have evolved a complex alternative splice mechanism, maybe allowing additional export of their unspliced genomic RNA.

Safer vectors for gene therapy

The integration of MLV-based vectors into the host genome increases the risk of adverse events in gene therapy due to transcriptional deregulation of cellular genes which could potentially stimulate tumorigenesis. The occurrence of leukemia after vector integration in five patients in a gene therapy trial for SCID-X1 illustrated the requirement for improved vector design. We have shown that in hematopoietic cells a weak cellular promoter/enhancer reduced the genotoxic risk of retroviral-based vectors. In contrast to cellular promoters, viral promoter/enhancer elements consist of dense clusters of binding sites for a variety of cellular transcription factors (Baum *et al.*, 1997). Therefore, they are able to activate neighboring genes more potently than cellular promoters. New results from our lab showed (Modlich, 2008) that the strong viral enhancer sequences are responsible for their transforming capacity in primary hematopoietic cells. Two SF91 mutants containing either a deletion of the promoter or the enhancer sequences were compared. Using the clonal dominance assay no mutants with serial replating capacity were observed with the SF91 vectors lacking enhancer sequences. In contrast, the deletion of the promoter region showed nearly the same level of transforming capacity compared to the wild type SF91. This indicates that the choice of the enhancer sequences is crucial for the protection against insertional mutagenesis.

Cellular enhancers have also the potential to activate a number of neighboring genes over a large chromosomal region (Nobrega *et al.*, 2003). However, several cellular restriction mechanisms exist to shield certain gene loci from activation. This is achieved by specialized DNA sequences, termed insulators (Gaszner & Felsenfeld, 2006). We have also tested an insulator element to increase vector safety. We have shown that the insulator core element (250 bp) was not sufficient to protect against insertional transformation, but that it clearly tended to reduce genotoxicity. The distribution of the replating frequency is very heterogeneous, indicating that the efficiency of the shielding effect of this insulator depends on the distinct integration site. Recently, another group showed by using the clonal dominance assay (Modlich *et al.*, 2006) that the transforming potential of SIN retroviral and lentiviral vectors, containing a novel insulator element, was reduced to background levels compared to their unshielded counterparts (Ramezani *et al.*, 2008). These results suggest that

insulator may additionally improve the safety of retroviral vectors (Cathomen & Joung, 2008).

Furthermore, homologous recombination can be used instead of retroviral vectors for stable gene transfer into the host genome. Cells use homologous recombination (HR) to repair DNA damage. Therapeutic gene targeting is the process of replacing a gene by HR, using an exogenous donor fragment and the cellular repair mechanism (Yanez & Porter, 1998). HR frequency is low and therefore gene targeting is not very efficient in mammalian cells (Cathomen & Joung, 2008). However, it has been shown that the insertion of a DNA double-stranded break (DSB) using endonucleases can increase HR by several 100-fold in various cells (Rouet *et al.*, 1994; Smih *et al.*, 1995). Zinc finger nucleases (ZFNs) are used to introduce site-specific DSBs into a genomic locus. ZFNs are artificial nucleases exhibiting an engineered zinc finger domain linked to the cleavage domain of the restriction endonuclease *Fok1* (Kim *et al.*, 1996). The zinc finger domain binds sequence-specifically to the DNA and can be designed to recognize a target site consisting of 18 or 24 bp, which is statistically large enough to define a unique site in a human genome (Bohne & Cathomen, 2008). Besides the advantage of gene targeting over gene addition regarding the decreased risk of insertional mutagenesis several studies observed ZFN-induced cytotoxicity. Cell death and apoptosis is most likely caused by cleavage at unspecific (off-target) sites, which can be repaired by non-homologous end-joining leading to insertions or deletions and chromosomal translocations (Cornu *et al.*, 2008; Lieber *et al.*, 2003).

However, therapeutic gene targeting will be the gold standard for future gene therapy approaches. But today the technology is still in its infancy and has to be improved concerning delivery, efficacy and safety aspects before it could be used as a gene therapy tool (Bohne & Cathomen, 2008). At the moment there is no safer and more efficient gene transfer system for long term expression available than retrovirus-based vectors. But they have to be optimized to be as safe as possible for future gene therapy trials.

Positive and negative effects of the 5'ss in retroviral vector design

The existence of a 5'splice site in retroviral-based gene transfer vectors has been described to enhance RNA stability, RNA export, RNA transcription and elongation (Damgaard *et al.*, 2008; Fong & Zhou, 2001; Kornblihtt *et al.*, 2004; Reed & Hurt, 2002), leading to a higher titer. In contrast to these positive features the risk of insertional mutagenesis increases due to the generation of fusion transcripts (Dupuy *et al.*, 2005; Ott *et al.*, 2006). To further improve the vector design for clinical applications, the 5'ss has to be deleted. A splice site deficient vector (SF110) has already been described and characterized (Fig.9). Compared to the SF11, containing a 5'ss, the titer of the SF110 is about one log decreased (Hildinger, 1999). Therefore, an ideal vector should have functional 5'ss in the packaging cell line and not in the target cell. We have shown in the manuscript, that we can manipulate splicing activity, by using U1 snRNA suppressor mutants. Here, replacing the corresponding nucleotides in U1 snRNA to the 5'ss increased the complementarity of the 5'ss to U1 snRNA. For future vector design we could alter the primary sequence of the 5'ss in such a way that wild type cellular U1 snRNA could not bind to the mutated 5'ss (Fig.17). We could further design a U1 snRNA suppressor mutant which is complementary to the mutated 5'ss. This U1 snRNA mutant could be transiently or stably expressed during vector production and recognize the mutated 5'ss. This would be likely to enhance the virus titer, due to the positive effects of U1 snRNP binding to the pre-mRNA. In the target cell the mutated 5'ss would not be able to be recognized by the spliceosome, due to the altered consensus sequence, and so would prevent the production of fusion transcripts. This vector design would improve safety without the loss of titer.

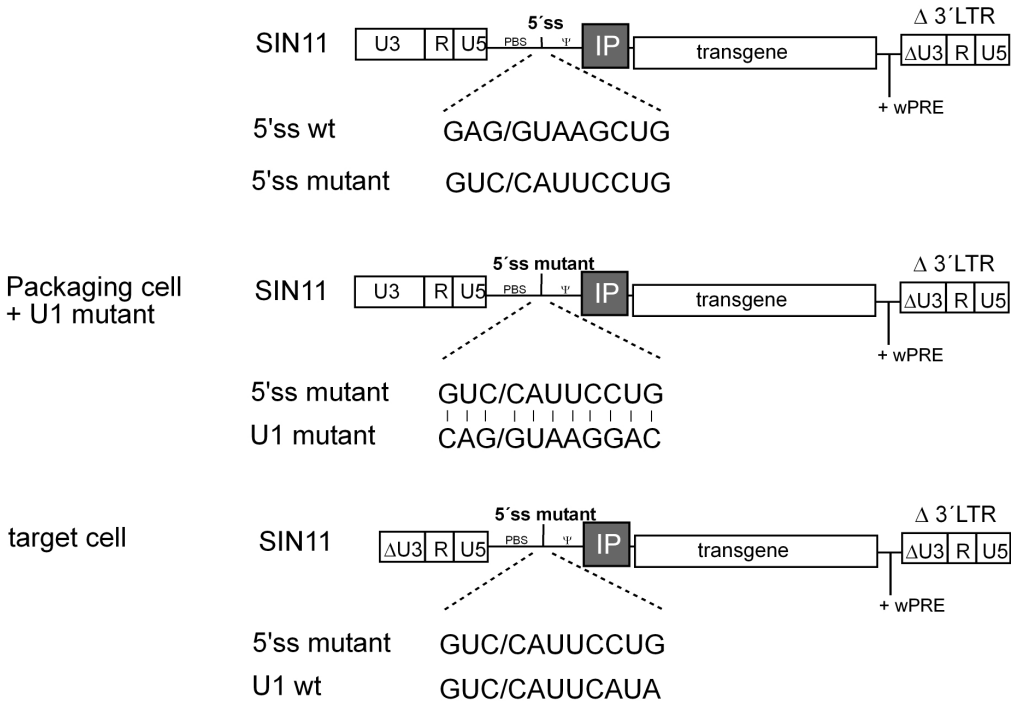


Figure 17: Mutated 5'ss for a safer vector

The top line shows the SIN11 vector backbone containing a 5'ss upstream of the internal promoter (IP). Possible mutation of the 5'ss would lead to inhibited recognition by U1 snRNA. The panel shows the situation in the packaging cell line where the mutated 5'ss could be recognized through a co-expressed U1 snRNA mutant by complementary base pairing. The lower panel shows the integrated vector backbone as a provirus, from which the mutated 5'ss would not be recognized by wild type U1 snRNA expressed by the cell.

Design of new co-expression vectors

In the present study we have designed "minigene" reporter constructs to analyze and compare the enhancer activity of the cellular and viral promoter/enhancer. Here the internal promoter/enhancer of the SIN vector activated the minigene cassette located in the deleted U3 region. The enhancer activity could be measured by flow cytometry. This configuration can also be used for the co-expression of two transgenes e.g. a fluorescence marker gene, a drug resistence gene or a surface antigen. Co-expression is usually achieved by using IRES-based dual gene transfer vectors (Martinez-Salas, 1999). The IRES configuration displayed some limitations because of the lower efficiency compared to Cap-dependent expression, and to the variablity in IRES performance according to the type of transgene and target cell (Hennecke *et al.*, 2001; Mizuguchi *et al.*, 2000). With our minigene design we could express two transgenes from two different transcripts by using the enhancer activity of the internal

promoter. So far we have tested this approach successfully for the co-expression of a transgene, driven by the internal promoter and a fluorescent marker gene (eGFP and dsRed) driven by the minimal CMV promoter, in retroviral and lentiviral based SIN vectors. This vector is not applicable for gene therapy,due to the observation that the minigene cassette, located after reverse transcription in the 5'LTR, is able to transcribe packable genomic RNA transcripts. But nevertheless it is a new option compared to the IRES configuration and an interesting tool for usage in co-expression studies.

Furthermore, alternative splicing as a co-expression strategy has been described (De Felipe & Izquierdo, 2000). For the development of a new co-expression vector it would be possible to use the SF91 stem loop mutant (manuscript; Fig.3), lacking viral sequences upstream of the 5'ss. In this vector the stem loop structure including the intron flanked by the 5'ss and 3'ss could be introduced downstream of the internal promoter. The intron could be replaced by a marker gene and if removed another transgene could be expressed. To manipulate the transgene expression we could use the same strategy discussed above. The primary 5'ss sequence can be mutated and is then only recognized by using U1 suppressor mutants complementary to the altered 5'ss. Transgene 2 would then only be expressed when U1 suppressor mutants are present in the target cell. Furthermore, we could use integrase-defective lentiviral episomal vectors (Vargas *et al.*, 2004) to express the mutated U1 snRNAs transiently in the target cell. This configuration would be not a potential vector for gene therapy but it would allow controlled expression of the transgene 2 and is so maybe an interesting tool to distinguish between effects due to integration of the vector backbone into the host genome and effects of a transgene. Additionally, we have shown that the ratio of spliced/unspliced RNA species is determined by the strength of the artificial stem loop. So it is possible that we could control the splicing ratio of transgene 1 to transgene 2 by strengthening the stem loop. Finally we could combine the minigene approach with the stem loop intron to an "all in one vector" and would thereby be able to express three different genes from one vector.

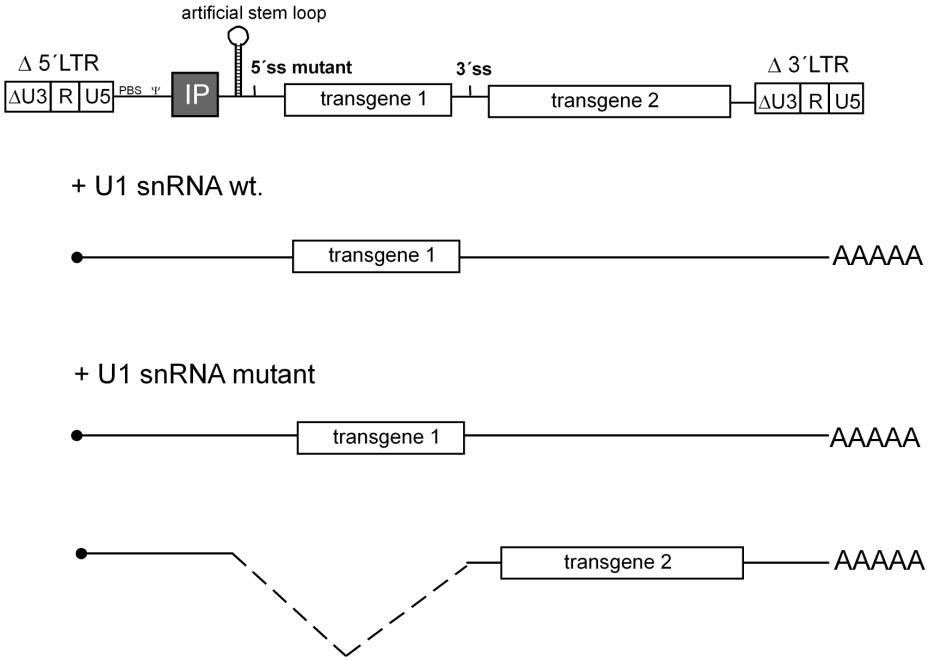


Figure 18: Splicing as co-expression strategy

A new possible SIN vector backbone is depicted, in which an additional intron is cloned downstream of the internal promoter (IP). The intron is derived from the stem loop mutants (manuscript; Fig. 3), where the mutated 5'ss is downstream of an artificial stem loop. Transgene 1 is located in the intron and expressed from unspliced RNA. Splicing and expression of transgene 2 could only occur if U1 snRNA mutant is expressed in the target cell due to the mutated 5'ss (see Fig.17).

Retrovirally based vectors are so far the best available vehicles for efficient gene transfer into hematopoietic stem cells. Non-viral gene delivery is potentially safer, but at present limited by the low efficiency. We have shown that SIN vectors containing weak enhancer elements decrease the risk of insertional transformation. In some diseases high expression of the transgene is required to correct the phenotype. Further improvement of the RNA processing of SIN vectors and of the transgene itself may lead to a general usage of weak promoter/enhancer for gene therapy applications.

G Appendix

1 References (Introduction & Discussion)

Aiuti, A., Slavin, S., Aker, M., Ficara, F., Deola, S., Mortellaro, A., Morecki, S., Andolfi, G., Tabucchi, A., Carlucci, F., et al. (2002). Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. *Science* **296**, 2410-2413.

Allen, T. D., Cronshaw, J. M., Bagley, S., Kiseleva, E. & Goldberg, M. W. (2000). The nuclear pore complex: mediator of translocation between nucleus and cytoplasm. *J Cell Sci* **113 (Pt 10)**, 1651-1659.

Amendt, B. A., Si, Z. H. & Stoltzfus, C. M. (1995). Presence of exon splicing silencers within human immunodeficiency virus type 1 tat exon 2 and tat-rev exon 3: evidence for inhibition mediated by cellular factors. *Molecular and cellular biology* **15**, 4606-4615.

Antson, A. A. (2000). Single-stranded-RNA binding proteins. *Current opinion in structural biology* **10**, 87-94.

Ashe, M. P., Furger, A. & Proudfoot, N. J. (2000). Stem-loop 1 of the U1 snRNP plays a critical role in the suppression of HIV-1 polyadenylation. *RNA (New York, NY* **6**, 170-177.

Ast, G. (2004). How did alternative splicing evolve? *Nature reviews* **5**, 773-782.

Baltimore, D. (1992). Viral RNA-dependent DNA polymerase. 1970. *Biotechnology* **24**, 3-5.

Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., et al. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science (New York, NY* **220**, 868-871.

Basu, V. P., Song, M., Gao, L., Rigby, S. T., Hanson, M. N. & Bambara, R. A. (2008). Strand transfer events during HIV-1 reverse transcription. *Virus Res*.

Baum, C. (2007). What are the consequences of the fourth case? *Mol Ther* **15**, 1401-1402.

Baum, C., Hegewisch-Becker, S., Eckert, H. G., Stocking, C. & Ostertag, W. (1995). Novel retroviral vectors for efficient expression of the multidrug resistance (mdr-1) gene in early hematopoietic cells. *J Virol* **69**, 7541-7547.

Baum, C., Itoh, K., Meyer, J., Laker, C., Ito, Y. & Ostertag, W. (1997). The potent enhancer activity of the polycythemic strain of spleen focus-forming virus in hematopoietic cells is governed by a binding site for Sp1 in the upstream control region and by a unique enhancer core motif, creating an exclusive target for PEBP/CBF. *J Virol* **71**, 6323-6331.

Baum, C., Richters, A. & Ostertag, W. (1999). Retroviral vector-mediated gene expression in hematopoietic cells. *Current opinion in molecular therapeutics* **1**, 605-612.

Baum, C., Schambach, A., Bohne, J. & Galla, M. (2006). Retrovirus vectors: toward the plentivirus? *Mol Ther* **13**, 1050-1063.

Baum, C., von Kalle, C., Staal, F. J., Li, Z., Fehse, B., Schmidt, M., Weerkamp, F., Karlsson, S., Wagemaker, G. & Williams, D. A. (2004). Chance or necessity? Insertional mutagenesis in gene therapy and its consequences. *Mol Ther* **9**, 5-13.

Bauren, G. & Wieslander, L. (1994). Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell* **76**, 183-192.

Beelman, C. A. & Parker, R. (1995). Degradation of mRNA in eukaryotes. *Cell* **81**, 179-183.

Beerens, N. & Berkhout, B. (2002). Switching the in vitro tRNA usage of HIV-1 by simultaneous adaptation of the PBS and PAS. *RNA (New York, NY* **8**, 357-369.

Benz, E. W., Jr., Wydro, R. M., Nadal-Ginard, B. & Dina, D. (1980). Moloney murine sarcoma proviral DNA is a transcriptional unit. *Nature* **288**, 665-669.

Berget, S. M., Moore, C. & Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 3171-3175.

Berkhout, B. & van Wamel, J. L. (2000). The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. *RNA (New York, NY* **6**, 282-295.

Beyer, A. L. & Osheim, Y. N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes & development* **2**, 754-765.

Black, D. L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annual review of biochemistry* **72**, 291-336.

Blanchette, M. & Chabot, B. (1997). A highly stable duplex structure sequesters the 5' splice site region of hnRNP A1 alternative exon 7B. *RNA (New York, NY* **3**, 405-419.

Bode, J. & Maass, K. (1988). Chromatin domain surrounding the human interferon-beta gene as defined by scaffold-attached regions. *Biochemistry* **27**, 4706-4711.

Bohne, J. & Cathomen, T. (2008). Genotoxicity in gene therapy: an account of vector integration and designer nucleases. *Current opinion in molecular therapeutics* **10**, 214-223.

Bokhoven, M., Stephen, S. L., Knight, S., Gevers, E. F., Robinson, I. C., Takeuchi, Y. & Collins, M. K. (2008). Insertional gene activation by lentiviral and gammaretroviral vectors. *J Virol*.

Boris-Lawrie, K., Roberts, T. M. & Hull, S. (2001). Retroviral RNA elements integrate components of post-transcriptional gene expression. *Life sciences* **69**, 2697-2709.

Bray, M., Prasad, S., Dubay, J. W., Hunter, E., Jeang, K. T., Rekosh, D. & Hammarskjold, M. L. (1994). A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 1256-1260.

Brion, P. & Westhof, E. (1997). Hierarchy and dynamics of RNA folding. *Annual review of biophysics and biomolecular structure* **26**, 113-137.

Buckanovich, R. J. & Darnell, R. B. (1997). The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. *Molecular and cellular biology* **17**, 3194-3201.

Buratti, E. & Baralle, F. E. (2004). Influence of RNA secondary structure on the pre-mRNA splicing process. *Molecular and cellular biology* **24**, 10505-10514.

Burns, J. C., Friedmann, T., Driever, W., Burrascano, M. & Yee, J. K. (1993). Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 8033-8037.

Calmels, B., Ferguson, C., Laukkanen, M. O., Adler, R., Faulhaber, M., Kim, H. J., Sellers, S., Hematti, P., Schmidt, M., von Kalle, C., et al. (2005). Recurrent retroviral vector integration at the Mds1/Evi1 locus in nonhuman primate hematopoietic cells. *Blood* **106**, 2530-2533.

Caputi, M., Mayeda, A., Krainer, A. R. & Zahler, A. M. (1999). hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. *The EMBO journal* **18**, 4060-4067.

Carlson, C. B., Stephens, O. M. & Beal, P. A. (2003). Recognition of double-stranded RNA by proteins and small molecules. *Biopolymers* **70**, 86-102.

Cathomen, T. & Joung, J. K. (2008). Zinc-finger nucleases: the next generation emerges. *Mol Ther* **16**, 1200-1207.

Cavazzana-Calvo, M., Hacein-Bey, S., de Saint Basile, G., Gross, F., Yvon, E., Nusbaum, P., Selz, F., Hue, C., Certain, S., Casanova, J. L., et al. (2000). Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science* **288**, 669-672.

Chabot, B., Blanchette, M., Lapierre, I. & La Branche, H. (1997). An intron element modulating 5' splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNP A1. *Molecular and cellular biology* **17**, 1776-1786.

Cochrane, A. W., McNally, M. T. & Mouland, A. J. (2006). The retrovirus RNA trafficking granule: from birth to maturity. *Retrovirology* **3**, 18.

Coffin, J. M., Hughes, S. H. & Varmus, H. E. (1997). *Retroviruses*. Cold Spring Harbor, NY.

Conn, G. L. & Draper, D. E. (1998). RNA structure. *Current opinion in structural biology* **8**, 278-285.

Cornu, T. I., Thibodeau-Beganny, S., Guhl, E., Alwin, S., Eichtinger, M., Joung, J. K. & Cathomen, T. (2008). DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Mol Ther* **16**, 352-358.

Crispino, J. D., Blencowe, B. J. & Sharp, P. A. (1994). Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science (New York, NY* **265**, 1866-1869.

Crispino, J. D. & Sharp, P. A. (1995). A U6 snRNA:pre-mRNA interaction can be rate-limiting for U1-independent splicing. *Genes & development* **9**, 2314-2323.

Cullen, B. R. (1998). Retroviruses as model systems for the study of nuclear RNA export pathways. *Virology* **249**, 203-210.

Cullen, B. R. (2003). Nuclear RNA export. *J Cell Sci* **116**, 587-597.

D'Souza, V. & Summers, M. F. (2005). How retroviruses select their genomes. *Nat Rev Microbiol* **3**, 643-655.

Dahmus, M. E. (1996). Reversible phosphorylation of the C-terminal domain of RNA polymerase II. *The Journal of biological chemistry* **271**, 19009-19012.

Damgaard, C. K., Kahns, S., Lykke-Andersen, S., Nielsen, A. L., Jensen, T. H. & Kjems, J. (2008). A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Molecular cell* **29**, 271-278.

Damgaard, C. K., Tange, T. O. & Kjems, J. (2002). hnRNP A1 controls HIV-1 mRNA splicing through cooperative binding to intron and exon splicing silencers in the context of a conserved secondary structure. *RNA (New York, NY* **8**, 1401-1415.

De Felipe, P. & Izquierdo, M. (2000). Tricistronic and tetracistronic retroviral vectors for gene transfer. *Human gene therapy* **11**, 1921-1931.

Dimaano, C. & Ullman, K. S. (2004). Nucleocytoplasmic transport: integrating mRNA production and turnover with export through the nuclear pore. *Molecular and cellular biology* **24**, 3069-3076.

Donello, J. E., Loeb, J. E. & Hope, T. J. (1998). Woodchuck hepatitis virus contains a tripartite posttranscriptional regulatory element. *J Virol* **72**, 5085-5092.

Du, Y., Jenkins, N. A. & Copeland, N. G. (2005). Insertional mutagenesis identifies genes that promote the immortalization of primary bone marrow progenitor cells. *Blood* **106**, 3932-3939.

Dupuy, A. J., Akagi, K., Largaespada, D. A., Copeland, N. G. & Jenkins, N. A. (2005). Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature* **436**, 221-226.

Eckert, H. G., Stockschlader, M., Just, U., Hegewisch-Becker, S., Grez, M., Uhde, A., Zander, A., Ostertag, W. & Baum, C. (1996). High-dose multidrug resistance in primary human hematopoietic progenitor cells transduced with optimized retroviral vectors. *Blood* **88**, 3407-3415.

Ellis, J. (2005). Silencing and variegation of gammaretrovirus and lentivirus vectors. *Human gene therapy* **16**, 1241-1246.

Emery, D. W., Yannaki, E., Tubb, J. & Stamatoyannopoulos, G. (2000). A chromatin insulator protects retrovirus vectors from chromosomal position effects. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 9150-9155.

Emi, N., Friedmann, T. & Yee, J. K. (1991). Pseudotype formation of murine leukemia virus with the G protein of vesicular stomatitis virus. *J Virol* **65**, 1202-1207.

Ernst, R. K., Bray, M., Rekosh, D. & Hammarskjold, M. L. (1997). A structured retroviral RNA element that mediates nucleocytoplasmic export of intron-containing RNA. *Molecular and cellular biology* **17**, 135-144.

Fasken, M. B. & Corbett, A. H. (2005). Process or perish: quality control in mRNA biogenesis. *Nat Struct Mol Biol* **12**, 482-488.

Fischer, U., Huber, J., Boelens, W. C., Mattaj, I. W. & Luhrmann, R. (1995). The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell* **82**, 475-483.

Flint, S. J., Enquist, L.W., Racaniello, V.R., Skalka, A.M. (2000). Principles of virology.

Fong, Y. W. & Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. *Nature* **414**, 929-933.

Fontana, W., Konings, D. A., Stadler, P. F. & Schuster, P. (1993). Statistics of RNA secondary structures. *Biopolymers* **33**, 1389-1404.

Fornerod, M., Ohno, M., Yoshida, M. & Mattaj, I. W. (1997). CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* **90**, 1051-1060.

Freund, M., Asang, C., Kammler, S., Konermann, C., Krummheuer, J., Hipp, M., Meyer, I., Gierling, W., Theiss, S., Preuss, T., et al. (2003). A novel approach to describe a U1 snRNA binding site. *Nucleic acids research* **31**, 6963-6975.

Fung, Y. K., Lewis, W. G., Crittenden, L. B. & Kung, H. J. (1983). Activation of the cellular oncogene c-erbB by LTR insertion: molecular basis for induction of erythroblastosis by avian leukosis virus. *Cell* **33**, 357-368.

Furger, A., Monks, J. & Proudfoot, N. J. (2001). The retroviruses human immunodeficiency virus type 1 and Moloney murine leukemia virus adopt radically different strategies to regulate promoter-proximal polyadenylation. *J Virol* **75**, 11735-11746.

Galla, M. (2008). Retroviral particle-mediated mRNA transfer. *University Hannover*.

Galla, M., Will, E., Kraunus, J., Chen, L. & Baum, C. (2004). Retroviral pseudotransduction for targeted cell manipulation. *Molecular cell* **16**, 309-315.

Gaspar, H. B., Parsley, K. L., Howe, S., King, D., Gilmour, K. C., Sinclair, J., Brouns, G., Schmidt, M., Von Kalle, C., Barington, T., et al. (2004). Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. *Lancet* **364**, 2181-2187.

Gaszner, M. & Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nature reviews* **7**, 703-713.

Gesteland, R. F., Cech, T.R, Atkins, J.F. (2006). *The RNA World*.

Glover, D. J., Lipps, H. J. & Jans, D. A. (2005). Towards safe, non-viral therapeutic gene expression in humans. *Nature reviews* **6**, 299-310.

Goldstrohm, A. C., Greenleaf, A. L. & Garcia-Blanco, M. A. (2001). Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. *Gene* **277**, 31-47.

Goodenow, M., Huet, T., Saurin, W., Kwok, S., Sninsky, J. & Wain-Hobson, S. (1989). HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. *J Acquir Immune Defic Syndr* **2**, 344-352.

Gorlich, D. & Kutay, U. (1999). Transport between the cell nucleus and the cytoplasm. *Annu Rev Cell Dev Biol* **15**, 607-660.

Graveley, B. R. (2000). Sorting out the complexity of SR protein functions. *RNA (New York, NY* **6**, 1197-1211.

Grez, M., Akgun, E., Hilberg, F. & Ostertag, W. (1990). Embryonic stem cell virus, a recombinant murine retrovirus with expression in embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 9202-9206.

Grover, A., Houlden, H., Baker, M., Adamson, J., Lewis, J., Prihar, G., Pickering-Brown, S., Duff, K. & Hutton, M. (1999). 5' splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10. *The Journal of biological chemistry* **274**, 15134-15143.

Gunderson, S. I., Polycarpou-Schwarz, M. & Mattaj, I. W. (1998). U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Molecular cell* **1**, 255-264.

Hacein-Bey-Abina, S., Garrigue, A., Wang, G. P., Soulier, J., Lim, A., Morillon, E., Clappier, E., Caccavelli, L., Delabesse, E., Beldjord, K., et al. (2008). Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *The Journal of clinical investigation* **118**, 3132-3142.

Hacein-Bey-Abina, S., von Kalle, C., Schmidt, M., Le Deist, F., Wulffraat, N., McIntyre, E., Radford, I., Villeval, J. L., Fraser, C. C., Cavazzana-Calvo, M. & Fischer, A. (2003a). A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med* **348**, 255-256.

Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M. P., Wulffraat, N., Leboulch, P., Lim, A., Osborne, C. S., Pawliuk, R., Morillon, E., et al. (2003b). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science (New York, NY* **302**, 415-419.

Hammarskjold, M. L., Li, H., Rekosh, D. & Prasad, S. (1994). Human immunodeficiency virus env expression becomes Rev-independent if the env region is not defined as an intron. *J Virol* **68**, 951-958.

Hartmann, L., Theiss, S., Niederacher, D. & Schaal, H. (2008). Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? *Front Biosci* **13**, 3252-3272.

Hastings, M. L. & Krainer, A. R. (2001). Pre-mRNA splicing in the new millennium. *Current opinion in cell biology* **13**, 302-309.

Hennecke, M., Kwissa, M., Metzger, K., Oumard, A., Kroger, A., Schirmbeck, R., Reimann, J. & Hauser, H. (2001). Composition and arrangement of genes define the strength of IRES-driven translation in bicistronic mRNAs. *Nucleic acids research* **29**, 3327-3334.

Hennig, E. E., Conney, A. H. & Wei, S. J. (1995). Characterization of hprt splicing mutations induced by the ultimate carcinogenic metabolite of benzo[a]pyrene in Chinese hamster V-79 cells. *Cancer research* **55**, 1550-1558.

Hildinger, M. (1999). Entwicklung retroviraler Vektoren für den Einsatz in der Gentherapie. *University Hamburg*.

Hildinger, M., Abel, K. L., Ostertag, W. & Baum, C. (1999). Design of 5' untranslated sequences in retroviral vectors developed for medical use. *J Virol* **73**, 4083-4089.

Hirose, Y. & Manley, J. L. (2000). RNA polymerase II and the integration of nuclear events. *Genes & development* **14**, 1415-1429.

Horowitz, D. S. & Krainer, A. R. (1994). Mechanisms for selecting 5' splice sites in mammalian pre-mRNA splicing. *Trends Genet* **10**, 100-106.

Hoshi, S., Odawara, T., Oshima, M., Kitamura, Y., Takizawa, H. & Yoshikura, H. (2002). cis-Elements involved in expression of unspliced RNA in Moloney murine leukemia virus. *Biochem Biophys Res Commun* **290**, 1139-1144.

Howe, S. J., Mansour, M. R., Schwarzwaelder, K., Bartholomae, C., Hubank, M., Kempski, H., Brugman, M. H., Pike-Overzet, K., Chatters, S. J., de Ridder, D., et al. (2008). Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *The Journal of clinical investigation* **118**, 3143-3150.

Hwang, L. S., Park, J. & Gilboa, E. (1984). Role of intron-contained sequences in formation of moloney murine leukemia virus env mRNA. *Molecular and cellular biology* **4**, 2289-2297.

Jablonski, J. A., Buratti, E., Stuani, C. & Caputi, M. (2008). The secondary structure of the human immunodeficiency virus type 1 transcript modulates viral splicing and infectivity. *J Virol* **82**, 8038-8050.

Jensen, T. H., Dower, K., Libri, D. & Rosbash, M. (2003). Early formation of mRNP: license for export or quality control? *Molecular cell* **11**, 1129-1138.

Johnson, J. M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P. M., Armour, C. D., Santos, R., Schadt, E. E., Stoughton, R. & Shoemaker, D. D. (2003). Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science (New York, NY* **302**, 2141-2144.

Jurica, M. S. & Moore, M. J. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Molecular cell* **12**, 5-14.

Kammler, S., Otte, M., Hauber, I., Kjems, J., Hauber, J. & Schaal, H. (2006). The strength of the HIV-1 3' splice sites affects Rev function. *Retrovirology* **3**, 89.

Kaushansky, K. (2006). Lineage-specific hematopoietic growth factors. *The New England journal of medicine* **354**, 2034-2045.

Kim, Y. G., Cha, J. & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 1156-1160.

Klaff, P., Riesner, D. & Steger, G. (1996). RNA structure and the regulation of gene expression. *Plant molecular biology* **32**, 89-106.

Komarnitsky, P., Cho, E. J. & Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes & development* **14**, 2452-2460.

Konarska, M. M. & Query, C. C. (2005). Insights into the mechanisms of splicing: more lessons from the ribosome. *Genes & development* **19**, 2255-2260.

Kornblihtt, A. R., de la Mata, M., Fededa, J. P., Munoz, M. J. & Nogues, G. (2004). Multiple links between transcription and splicing. *RNA (New York, NY* **10**, 1489-1498.

Kramer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annual review of biochemistry* **65**, 367-409.

Kraunus, J., Schaumann, D. H., Meyer, J., Modlich, U., Fehse, B., Brandenburg, G., Von Laer, D., Klump, H., Schambach, A., Bohne, J. & Baum, C. (2004). Self-inactivating retroviral vectors with improved RNA processing. *Gene therapy* **11**, 1568-1578.

Krecic, A. M. & Swanson, M. S. (1999). hnRNP complexes: composition, structure, and function. *Current opinion in cell biology* **11**, 363-371.

Kustikova, O., Fehse, B., Modlich, U., Yang, M., Dullmann, J., Kamino, K., von Neuhoff, N., Schlegelberger, B., Li, Z. & Baum, C. (2005). Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. *Science (New York, NY* **308**, 1171-1174.

Lagresle-Peyrou, C., Yates, F., Malassis-Seris, M., Hue, C., Morillon, E., Garrigue, A., Liu, A., Hajdari, P., Stockholm, D., Danos, O., et al. (2006). Long-term immune reconstitution in RAG-1-deficient mice treated by retroviral gene therapy: a balance between efficiency and toxicity. *Blood* **107**, 63-72.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.

Lee, J. T., Yu, S. S., Han, E. & Kim, S. (2004). Engineering the splice acceptor for improved gene expression and viral titer in an MLV-based retroviral vector. *Gene therapy* **11**, 94-99.

Lever, A. M. & Berkhout, B. (2008). 2008 Nobel prize in medicine for discoverers of HIV. *Retrovirology* **5**, 91.

Li, Z., Dullmann, J., Schiedlmeier, B., Schmidt, M., von Kalle, C., Meyer, J., Forster, M., Stocking, C., Wahlers, A., Frank, O., et al. (2002). Murine leukemia induced by retroviral gene marking. *Science (New York, NY* **296**, 497.

Lieber, M. R., Ma, Y., Pannicke, U. & Schwarz, K. (2003). Mechanism and regulation of human non-homologous DNA end-joining. *Nat Rev Mol Cell Biol* **4**, 712-720.

Loeb, J. E., Cordier, W. S., Harris, M. E., Weitzman, M. D. & Hope, T. J. (1999). Enhanced expression of transgenes from adeno-associated virus vectors with the woodchuck hepatitis virus posttranscriptional regulatory element: implications for gene therapy. *Human gene therapy* **10**, 2295-2305.

Logg, C. R., Baranick, B. T., Lemp, N. A. & Kasahara, N. (2007). Adaptive evolution of a tagged chimeric gammaretrovirus: identification of novel cis-acting elements that modulate splicing. *J Mol Biol* **369**, 1214-1229.

Lutzko, C., Senadheera, D., Skelton, D., Petersen, D. & Kohn, D. B. (2003). Lentivirus vectors incorporating the immunoglobulin heavy chain enhancer and matrix attachment regions provide position-independent expression in B lymphocytes. *J Virol* **77**, 7341-7351.

Madhani, H. D. & Guthrie, C. (1992). A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* **71**, 803-817.

Malim, M. H., Hauber, J., Fenrick, R. & Cullen, B. R. (1988). Immunodeficiency virus rev trans-activator modulates the expression of the viral regulatory genes. *Nature* **335**, 181-183.

Maniatis, T. & Tasic, B. (2002). Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418**, 236-243.

Martinez-Contreras, R., Fisette, J. F., Nasim, F. U., Madden, R., Cordeau, M. & Chabot, B. (2006). Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PLoS biology* **4**, e21.

Martinez-Salas, E. (1999). Internal ribosome entry site biology and its use in expression vectors. *Current opinion in biotechnology* **10**, 458-464.

McKenna, N. J. & O'Malley, B. W. (2002). Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell* **108**, 465-474.

Miller, A. D. (1990). Retrovirus packaging cells. *Human gene therapy* **1**, 5-14.

Miller, A. D. & Rosman, G. J. (1989). Improved retroviral vectors for gene transfer and expression. *Biotechniques* **7**, 980-982, 984-986, 989-990.

Mizuguchi, H., Xu, Z., Ishii-Watabe, A., Uchida, E. & Hayakawa, T. (2000). IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector. *Mol Ther* **1**, 376-382.

Modlich, U. (2008). Insertional genotoxicity is dependent on the enhancer and less on the vector context. *in preparation*.

Modlich, U., Bohne, J., Schmidt, M., von Kalle, C., Knob, S., Schambach, A. & Baum, C. (2006). Cell culture assays reveal the importance of retroviral vector design for insertional genotoxicity. *Blood*.

Modlich, U., Kustikova, O. S., Schmidt, M., Rudolph, C., Meyer, J., Li, Z., Kamino, K., von Neuhoff, N., Schlegelberger, B., Kuehlcke, K., et al. (2005). Leukemias following retroviral transfer of multidrug resistance 1 (MDR1) are driven by combinatorial insertional mutagenesis. *Blood* **105**, 4235-4246.

Modlich, U., Schambach, A., Brugman, M. H., Wicke, D. C., Knoess, S., Li, Z., Maetzig, T., Rudolph, C., Schlegelberger, B. & Baum, C. (2008). Leukemia induction after a single retroviral vector insertion in Evi1 or Prdm16. *Leukemia* **22**, 1519-1528.

Modrek, B., Resch, A., Grasso, C. & Lee, C. (2001). Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic acids research* **29**, 2850-2859.

Modrow, S., Falke, D. & Truyen, U. (2003). *Molekulare Virologie*. Berlin: Spektrum Akademischer Verlag.

Montini, E., Cesana, D., Schmidt, M., Sanvito, F., Ponzoni, M., Bartholomae, C., Sergi Sergi, L., Benedicenti, F., Ambrosi, A., Di Serio, C., et al. (2006). Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. *Nat Biotechnol* **24**, 687-696.

Mougel, M., Tounekti, N., Darlix, J. L., Paoletti, J., Ehresmann, B. & Ehresmann, C. (1993). Conformational analysis of the 5' leader and the gag initiation site of Mo-MuLV RNA and allosteric transitions induced by dimerization. *Nucleic acids research* **21**, 4677-4684.

Mulligan, R. C. (1993). The basic science of gene therapy. *Science (New York, NY* **260**, 926-932.

Nagel, R. J., Lancaster, A. M. & Zahler, A. M. (1998). Specific binding of an exonic splicing enhancer by the pre-mRNA splicing factor SRp55. *RNA (New York, NY* **4**, 11-23.

Naldini, L., Blomer, U., Gage, F. H., Trono, D. & Verma, I. M. (1996). Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 11382-11388.

Narlikar, G. J., Fan, H. Y. & Kingston, R. E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. *Cell* **108**, 475-487.

Nasim, F. U., Hutchison, S., Cordeau, M. & Chabot, B. (2002). High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5' splice site selection in support of a common looping out and repression mechanism. *RNA (New York, NY* **8**, 1078-1089.

Neugebauer, K. M. (2002). On the importance of being co-transcriptional. *J Cell Sci* **115**, 3865-3871.

Newman, A. J. (1997). The role of U5 snRNP in pre-mRNA splicing. *The EMBO journal* **16**, 5797-5800.

Nilsen, T. W. (1994). RNA-RNA interactions in the spliceosome: unraveling the ties that bind. *Cell* **78**, 1-4.

Nobrega, M. A., Ovcharenko, I., Afzal, V. & Rubin, E. M. (2003). Scanning human gene deserts for long-range enhancers. *Science (New York, NY* **302**, 413.

O'Reilly, M. M., McNally, M. T. & Beemon, K. L. (1995). Two strong 5' splice sites and competing, suboptimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA. *Virology* **213**, 373-385.

Orphanides, G. & Reinberg, D. (2002). A unified theory of gene expression. *Cell* **108**, 439-451.

Oshima, M., Odawara, T., Hanaki, K., Igarashi, H. & Yoshikura, H. (1998). cis Elements required for high-level expression of unspliced Gag-containing message in Moloney murine leukemia virus. *J Virol* **72**, 6414-6420.

Ossareh-Nazari, B., Bachelerie, F. & Dargemont, C. (1997). Evidence for a role of CRM1 in signal-mediated nuclear protein export. *Science (New York, NY* **278**, 141-144.

Ott, M. G., Schmidt, M., Schwarzwaelder, K., Stein, S., Siler, U., Koehl, U., Glimm, H., Kuhlcke, K., Schilz, A., Kunkel, H., et al. (2006). Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat Med* **12**, 401-409.

Perez-Canadillas, J. M. & Varani, G. (2001). Recent advances in RNA-protein recognition. *Current opinion in structural biology* **11**, 53-58.

Pincock, S. (2008). HIV discoverers awarded Nobel Prize for medicine. *Lancet* **372**, 1373.

Pollard, V. W. & Malim, M. H. (1998). The HIV-1 Rev protein. *Annu Rev Microbiol* **52**, 491-532.

Popa, I., Harris, M. E., Donello, J. E. & Hope, T. J. (2002). CRM1-dependent function of a cis-acting RNA export element. *Molecular and cellular biology* **22**, 2057-2067.

Preston, B. D., Poiesz, B. J. & Loeb, L. A. (1988). Fidelity of HIV-1 reverse transcriptase. *Science (New York, NY* **242**, 1168-1171.

Proudfoot, N. J., Furger, A. & Dye, M. J. (2002). Integrating mRNA processing with transcription. *Cell* **108**, 501-512.

Purcell, D. F. & Martin, M. A. (1993). Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J Virol* **67**, 6365-6378.

Ramezani, A., Hawley, T. S. & Hawley, R. G. (2003). Performance- and safety-enhanced lentiviral vectors containing the human interferon-beta scaffold attachment region and the chicken beta-globin insulator. Blood. 2003 Jun 15;101(12):4717-24. *Blood* **101**, 4717-4724.

Ramezani, A., Hawley, T. S. & Hawley, R. G. (2008). Combinatorial Incorporation of Enhancer Blocking Components of the Chicken {beta}-Globin 5'HS4 and Human T-Cell Receptor {alpha}/{delta} BEAD-1 Insulators in Self-Inactivating Retroviral Vectors Reduces their Genotoxic Potential. *Stem cells (Dayton, Ohio)*.

Reed, R. & Hurt, E. (2002). A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell* **108**, 523-531.

Richards, E. J. & Elgin, S. C. (2002). Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* **108**, 489-500.

Rigo, F. & Martinson, H. G. (2008). Functional coupling of last-intron splicing and 3'-end processing to transcription in vitro: the poly(A) signal couples to splicing before committing to cleavage. *Molecular and cellular biology* **28**, 849-862.

Rivella, S., Callegari, J. A., May, C., Tan, C. W. & Sadelain, M. (2000). The cHS4 insulator increases the probability of retroviral expression at random chromosomal integration sites. *J Virol* **74**, 4679-4687.

Rouet, P., Smih, F. & Jasin, M. (1994). Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 6064-6068.

Ryan, K. J. & Wente, S. R. (2000). The nuclear pore complex: a protein machine bridging the nucleus and cytoplasm. *Current opinion in cell biology* **12**, 361-371.

Schambach, A., Galla, M., Maetzig, T., Loew, R. & Baum, C. (2007). Improving Transcriptional Termination of Self-inactivating Gamma-retroviral and Lentiviral Vectors. *Mol Ther*.

Schambach, A., Mueller, D., Galla, M., Verstegen, M. M., Wagemaker, G., Loew, R., Baum, C. & Bohne, J. (2006). Overcoming promoter competition in packaging cells improves production of self-inactivating retroviral vectors. *Gene Ther* **13**, 1524-1533.

Schambach, A., Wodrich, H., Hildinger, M., Bohne, J., Krausslich, H. G. & Baum, C. (2000). Context dependence of different modules for posttranscriptional enhancement of gene expression from retroviral vectors. *Mol Ther* **2**, 435-445.

Schwartz, S., Felber, B. K., Benko, D. M., Fenyo, E. M. & Pavlakis, G. N. (1990). Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. *J Virol* **64**, 2519-2529.

Schwarzwaelder, K., Howe, S. J., Schmidt, M., Brugman, M. H., Deichmann, A., Glimm, H., Schmidt, S., Prinz, C., Wissler, M., King, D. J., et al. (2007). Gammaretrovirus-mediated correction of SCID-X1 is associated with skewed vector integration site distribution in vivo. *The Journal of clinical investigation* **117**, 2241-2249.

Seggewiss, R., Pittaluga, S., Adler, R. L., Guenaga, F. J., Ferguson, C., Pilz, I. H., Ryu, B., Sorrentino, B. P., Young, W. S., 3rd, Donahue, R. E., et al. (2006). Acute myeloid leukemia is associated with retroviral gene transfer to hematopoietic progenitor cells in a rhesus macaque. *Blood* **107**, 3865-3867.

Seraphin, B., Kretzner, L. & Rosbash, M. (1988). A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. *The EMBO journal* **7**, 2533-2538.

Shatkin, A. J. & Manley, J. L. (2000). The ends of the affair: capping and polyadenylation. *Nat Struct Biol* **7**, 838-842.

Shen, H. & Green, M. R. (2006). RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. *Genes & development* **20**, 1755-1765.

Shi, H., Hoffman, B. E. & Lis, J. T. (1997). A specific RNA hairpin loop structure binds the RNA recognition motifs of the Drosophila SR protein B52. *Molecular and cellular biology* **17**, 2649-2657.

Si, Z., Amendt, B. A. & Stoltzfus, C. M. (1997). Splicing efficiency of human immunodeficiency virus type 1 tat RNA is determined by both a suboptimal 3' splice site and a 10 nucleotide exon splicing silencer element located within tat exon 2. *Nucleic acids research* **25**, 861-867.

Siliciano, P. G. & Guthrie, C. (1988). 5' splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. *Genes & development* **2**, 1258-1267.

Smih, F., Rouet, P., Romanienko, P. J. & Jasin, M. (1995). Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic acids research* **23**, 5012-5019.

Stade, K., Ford, C. S., Guthrie, C. & Weis, K. (1997). Exportin 1 (Crm1p) is an essential nuclear export factor. *Cell* **90**, 1041-1050.

Staffa, A. & Cochrane, A. (1994). The tat/rev intron of human immunodeficiency virus type 1 is inefficiently spliced because of suboptimal signals in the 3' splice site. *J Virol* **68**, 3071-3079.

Staley, J. P. & Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315-326.

Staley, J. P. & Guthrie, C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Molecular cell* **3**, 55-64.

Stoilov, P., Meshorer, E., Gencheva, M., Glick, D., Soreq, H. & Stamm, S. (2002). Defects in pre-mRNA processing as causes of and predisposition to diseases. *DNA and cell biology* **21**, 803-818.

Sun, J. S. & Manley, J. L. (1995). A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes & development* **9**, 843-854.

Suzuki, T., Minehata, K., Akagi, K., Jenkins, N. A. & Copeland, N. G. (2006). Tumor suppressor gene identification using retroviral insertional mutagenesis in Blm-deficient mice. *The EMBO journal* **25**, 3422-3431.

Tange, T. O., Damgaard, C. K., Guth, S., Valcarcel, J. & Kjems, J. (2001). The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *The EMBO journal* **20**, 5748-5758.

Tarn, W. Y. & Steitz, J. A. (1994). SR proteins can compensate for the loss of U1 snRNP functions in vitro. *Genes & development* **8**, 2704-2717.

Tennyson, C. N., Klamut, H. J. & Worton, R. G. (1995). The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nat Genet* **9**, 184-190.

Thomas, C. E., Ehrhardt, A. & Kay, M. A. (2003). Progress and problems with the use of viral vectors for gene therapy. *Nature Rev Genet* **4**, 346-358.

Thornhill, S. I., Schambach, A., Howe, S. J., Ulaganathan, M., Grassman, E., Williams, D., Schiedlmeier, B., Sebire, N. J., Gaspar, H. B., Kinnon, C., et al. (2008). Self-inactivating gammaretroviral vectors for gene therapy of X-linked severe combined immunodeficiency. *Mol Ther* **16**, 590-598.

Trubetskoy, A. M., Okenquist, S. A. & Lenz, J. (1999). R region sequences in the long terminal repeat of a murine retrovirus specifically increase expression of unspliced RNAs. *J Virol* **73**, 3477-3483.

Tu, M., Tong, W., Perkins, R. & Valentine, C. R. (2000). Predicted changes in pre-mRNA secondary structure vary in their association with exon skipping for mutations in exons 2, 4, and 8 of the Hprt gene and exon 51 of the fibrillin gene. *Mutation research* **432**, 15-32.

Urisman, A., Molinaro, R. J., Fischer, N., Plummer, S. J., Casey, G., Klein, E. A., Malathi, K., Magi-Galluzzi, C., Tubbs, R. R., Ganem, D., et al. (2006). Identification of a novel Gammaretrovirus in prostate tumors of patients homozygous for R462Q RNASEL variant. *PLoS Pathog* **2**, e25.

Valcarcel, J., Gaur, R. K., Singh, R. & Green, M. R. (1996). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science (New York, NY* **273**, 1706-1709.

Vargas, J., Jr., Gusella, G. L., Najfeld, V., Klotman, M. E. & Cara, A. (2004). Novel integrase-defective lentiviral episomal vectors for gene transfer. *Human gene therapy* **15**, 361-372.

Wahle, E. & Ruegsegger, U. (1999). 3'-End processing of pre-mRNA in eukaryotes. *FEMS Microbiol Rev* **23**, 277-295.

Wetterberg, I., Bauren, G. & Wieslander, L. (1996). The intranuclear site of excision of each intron in Balbiani ring 3 pre-mRNA is influenced by the time remaining to transcription termination and different excision efficiencies for the various introns. *RNA (New York, NY* **2**, 641-651.

Will, C. L. & Luhrmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Current opinion in cell biology* **13**, 290-301.

Wolf, D. & Goff, S. P. (2007). TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* **131**, 46-57.

Woodley, L. & Valcarcel, J. (2002). Regulation of alternative pre-mRNA splicing. *Briefings in functional genomics & proteomics* **1**, 266-277.

Yanez, R. J. & Porter, A. C. (1998). Therapeutic gene targeting. *Gene therapy* **5**, 149-159.

Yang, Y., Vanin, E. F., Whitt, M. A., Fornerod, M., Zwart, R., Schneiderman, R. D., Grosveld, G. & Nienhuis, A. W. (1995). Inducible, high-level production of infectious murine leukemia retroviral vector particles pseudotyped with vesicular stomatitis virus G envelope protein. *Human gene therapy* **6**, 1203-1213.

Yu, S. F., von Ruden, T., Kantoff, P. W., Garber, C., Seiberg, M., Ruther, U., Anderson, W. F., Wagner, E. F. & Gilboa, E. (1986). Self-inactivating retroviral vectors designed for transfer of whole genes into mammalian cells. *Proc Natl Acad Sci U S A* **83**, 3194-3198.

Yu, S. S., Han, E., Hong, Y., Lee, J. T., Kim, S. & Kim, S. (2003). Construction of a retroviral vector production system with the minimum possibility of a homologous recombination. *Gene therapy* **10**, 706-711.

Yusufzai, T. M. & Felsenfeld, G. (2004). The 5'-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 8620-8624.

Zhou, Z., Licklider, L. J., Gygi, S. P. & Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**, 182-185.

Zhuang, Y. & Weiner, A. M. (1986). A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* **46**, 827-835.

Zufferey, R., Donello, J. E., Trono, D. & Hope, T. J. (1999). Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. *J Virol* **73**, 2886-2892.

2 List of own publications

Schambach A., **Mueller D.**, Galla M., Verstegen M.M.A., Wagemaker G., Baum C., Bohne J.
„Overcoming promoter competition in packaging cells improves production of self-inactivating retroviral vectors", *Gene Therapy 2006 Vol.13, p.1524-1533*

Kraunus J., **Zychlinski D.**, Heise T., Galla M., Bohne J., Baum C.
„Murine leukemia virus regulates alternative splicing through sequences upstream of the 5' splice site", *Journal of Biological Chemistry 2006 Vol.281, p.37381-37390*

Bohne J., Schambach A., **Zychlinski D.**
„New way of regulating alternative splicing in retroviruses: the promoter makes a difference", *Journal of Virology 2007 Vol.81, p.3652-3656*

Zychlinski D., Schambach A., Modlich U., Maetzig T., Meyer J., Grassman E., Mishra A., Baum C.
„Physiological promoters reduce the genotoxic risk of integrating gene vectors", *Molecular Therapy 2008, Vol.16, p.718-725*

Zychlinski D., Erkelenz S., Melhorn V., Schaal H., Baum C., Bohne J.
„Interplay of a secondary structure and limited pairing to U1 snRNP results in dual attenuation of a retroviral 5´splice site", *Manuscript*

3 Curriculum vitae

Persönliches

Daniela Zychlinski, geb. Müller
Mittelstraße 15a
D-30900 Wedemark

Telefon: 05130/928439
E-Mail: daniela@zychlinski.de

Geboren am 28.10.1978 in Hannover

Verheiratet

Ausbildung

1998 Allgemeine Hochschulreife an der Leibniz Schule Hannover

1998 – 1999 Studium der Chemie an der Universität Hannover

1999 – 2005 Studium der Biologie an der Universität Hannover

2005 Diplom mit der Note 1 (sehr gut)

Prüfungsfächer: Zellbiologie, Virologie und Biochemie

Titel der Diplomarbeit: „Optimierung von retroviralen Vektoren für die Gentherapie von hämatopoietischen Stammzellen“, betreut durch Prof. Dr. Christopher Baum Medizinische Hochschule Hannover, Labor für Experimentelle Zelltherapie

seit Juni 2005 Promotion betreut durch Prof. Dr. Christopher Baum, Medizinischen Hochschule Hannover, Abteilung Experimentelle Hämatologie

Praktika und berufliche Tätigkeiten

1999 – 2001 pflegerische Hilfskraft im Alten- und Pflegeheim Roderbruch des Deutschen Roten Kreuzes (studentischer Nebenjob)

2001 – 2003 Mitarbeiterin im Kino am Raschplatz (studentischer Nebenjob)

2002 Praktikum im Blutspendedienst des DRK in Springe

2003 Praktikum an der Medizinische Hochschule Hannover, Labor für Experimentelle Zelltherapie

2003 - 2004 studentische Hilfskraft an der Medizinischen Hochschule Hannover, Abteilung Hämatologie, Hämostaseologie und Onkologie

seit 2003 wissenschaftliche Hilfskraft im Büro der Bundestagsabgeordneten Caren Marks (Nebentätigkeit)

seit 2005 wissenschaftliche Mitarbeiterin an der Medizinischen Hochschule Hannover

Juni 2007 wissenschaftlicher Austausch mit den National Institutes of Health, Washington D.C.

Zusatzqualifikationen

Sprachen: Englisch; Französisch (Grundlagen)

Computer: Office-; Grafik-; Statistik-Programme

4 Danksagung

Diese Arbeit wurde von 06/2005-12/2008 an der Medizinischen Hochschule Hannover in der Abteilung Experimentelle Hämatologie und im Institut für Virologie erstellt.

Herrn Prof. Dr. Christopher Baum möchte ich für die Möglichkeit in diesem faszinierenden Forschungsfeld meine Doktorarbeit anfertigen zu können und für die hervorragende Betreuung ganz besonders danken.

Herrn Prof. Dr. Helmut Holtmann danke ich ganz herzlich für die Übernahme des Gutachtens.

Mein besonderer Dank gilt meinem Betreuer Dr. Jens Bohne, der mir stets zur Seite stand und meine Begeisterung für die Virologie und Zellbiologie förderte. Die vielen kritischen Diskussionen haben mein Projekt und mich zielführend weiter gebracht. Lieber Jens, vielen Dank für die hervorragende Zusammenarbeit in den letzten Jahren!

Herzlichen Dank an Dr. Axel Schambach für seine immer vorhandene Bereitschaft, mich zu unterstützen und für die große Hilfe beim vektorologischen Part dieser Arbeit. Ich werde den Klonkriegern auch 2009 wieder die Daumen drücken.

Weiterhin möchte ich mich bei Dr. Johann Meyer besonders bedanken für seine ständige Hilfsbereitschaft und dafür, dass er seine Erfahrungen stets mit mir teilte.

Danken möchte ich auch Dr. Ute Modlich für ihre vielfältige, freundliche Unterstützung und besonders für ihre Hilfe bei dem „IVIM assay“.

Sehr bedanken möchte ich mich bei Sabine Knöß für die ausgezeichnete Einarbeitung in zell- und molekularbiologische Techniken. Vielen Dank für die freundschaftliche Unterstützung während der gesamten Zeit meiner Promotion.

Herzlichen Dank an Dr. Penelope Kay-Jackson, die dieses Manuskript auf die englische Sprache geprüft hat.

Darüber hinaus möchte ich mich bei allen Mitgliedern der Arbeitsgruppe Baum bedanken - insbesondere bei Dr. Melanie Galla, Tobias Mätzig und Elke Barczak für ihre Unterstützung während der gesamten Doktorarbeit.

Für viele schöne und vor allem lustige Abende möchte ich mich bei allen Mitgliedern des „Doktoranden Clubs“ bedanken. Ich hoffe wir werden uns auch weiterhin regelmäßig treffen.

Ein weiterer Dank gilt allen Mitgliedern des Instituts für Virologie und seinem Leiter Prof. Dr. Thomas Schulz. Vielen Dank für die freundliche Aufnahme und Unterstützung zum Ende meiner Doktorarbeit. Besonders Vanessa Melhorn, Jörg Langemeier und Lars Koste danke ich für viel Spaß an der Arbeit.

Den größten Dank schulde ich meinen Eltern, meiner gesamten Familie, meinen Freunden und im Besonderen meinem Mann Helge für seine Liebe. Ihr habt mich immer unterstützt und an mich geglaubt. Ohne euch wäre ich niemals soweit gekommen.

5 Erklärung zur Dissertation

Hierdurch erkläre ich, dass die Dissertation „Promoter and 5'splice site interactions in retroviruses and retroviral vectors“ selbständig verfasst und alle benutzten Hilfsmittel sowie eventuell zur Hilfeleistung herangezogene Institutionen vollständig angegeben wurden.

Die Dissertation wurde nicht schon als Diplom- oder ähnliche Prufungsarbeit verwendet.

Wedemark, den 15.Dezember 2008

_____

Daniela Zychlinski