

Mutational analysis of splicing activation by the USSE motif in the
U11/U12-specific 65K gene

Prosanta Kumar Singha

Master's Thesis

University of Helsinki

MBIOT (Cellular Biotechnology)

February 2013

This work was carried-out at Institute of Biotechnology, University of Helsinki, Finland

Supervisor:

Docent Mikko Frilander

Institute of Biotechnology

University of Helsinki, Finland

Table of contents

Abbreviations	V
Abstract	VI
1 Review of the literature:	7
1.1 Introduction	7
1.2 Processing of eukaryotic pre-mRNA	8
1.3 Two parallel intron types in eukaryotes	8
1.3.1 Structure	9
1.3.2 Characteristics	9
1.4 Splicing machinery: Components of two Spliceosomes	10
1.5 Spliceosome assembly and catalysis	13
1.5.1 Splicing catalysis mechanism	13
1.5.2 Assembly of U2 and U12 dependent spliceosome	14
1.6 Alternative splicing	17
1.6.1 Splice site selection and alternative splicing	18
1.6.2 Factors affecting alternative splicing	19
1.6.2.1 Splicing enhancer and silencer	19
1.6.2.2 SR proteins	20
1.6.2.3 HnRNP proteins	20
1.6.2.4 Core spliceosomal components	21
1.6.2.5 Other factors	21
1.6.3 Significance of alternative splicing	22
1.7 Quality control of mRNA	22
1.7.1 Nuclear degradation	22
1.7.2 Cytoplasmic quality control	23
1.8 Alternative splicing in U11/U12-65K gene	24
1.8.1 Evolutionary conservation of U11/U12-65K gene	24
1.8.2 USSE mediated alternative splicing	25
2 Objectives of the study	26

3	Materials and Methods	27
3.1	Construction of Plasmids	27
3.1.1	Molecular Cloning	27
3.1.1.1	PCR and Gel extraction	31
3.1.1.2	Phosphorylation and Ligation of the DNA	32
3.1.1.3	Phenol-chloroform extraction and DNA precipitation	33
3.1.1.4	Electroporation	33
3.1.1.5	Screening by Minipreps and Maxipreps.	34
3.2	Cell culture and transfection of reporter plasmids	34
3.3	RNA isolation	35
3.4	RT-PCR	35
4	Results	38
4.1	Construction of reporter plasmids	38
4.2	USSE activation highly dependent on the distance between 3'ss and USSE	40
4.3	Interaction between two U11/12 snRNP is essential for splicing activation	42
4.4	Manipulation of element distances with RNA hairpin constructs	43
5	Discussions	45
5.1	Deletion might affect Enhancer/Repressor sites to 3'ss	45
5.2	Interaction between two U11/12 di-snRNP is essential for splicing activation	
5.3	Rescue construct didn't fully rescue the functionality of USSE	47
5.4	Construction of mutated plasmids not always succeeds	48
6	Conclusions	49
7	Acknowledgements	49
8	Lists of References	50

Abbreviations

3'ss	3' splice site
3' UTR	3' untranslated region
5'ss	5' splice site
AS	alternative splicing
bp	base pair(s)
BPS	branch point sequence
DMEM	Dulbecco's modified Eagle's medium
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
FBS	Fetal bovine serum
hnRNP	heterogeneous nuclear ribonucleoprotein
ISE	intronic splicing enhancer
ISS	intronic splicing silencer
KH Domain	K Homology Domain
Lsm	Sm-like proteins
LECA	last eukaryotic common ancestor
mRNA	messenger RNA
MQ	Milli-Q
NGD	'No-Go' decay
NMD	nonsense-mediated decay
NSD	nonstop decay
nt	nucleotide(s)
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PPT	polypyrimidine tract
pre-mRNA	precursor mRNA
PTC	premature termination codon
R	purine
REMD	ribosome extension-mediated decay
RNAPII	RNA polymerase II
RRM	RNA recognition motif
RNP	ribonucleoprotein
RS	arginine-serine-rich (domain)
RSV	Rous sarcoma virus
RT-PCR	Reverse transcription polymerase chain reaction
SF1	splicing factor 1
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
snRNP	small nuclear RNP
SR	serine-arginine-rich (protein)
SRSF1	serine/arginine-rich splicing factor 1
SRE	splicing regulatory element
U2AF	U2 auxiliary factor
USSE	U11 snRNP-binding splicing enhancer
Y	pyrimidine
wt	wild type

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty: Agriculture and Forestry		Laitos — Institution — Department: Master's Degree Programme in Biotechnology (MBIOT)	
Tekijä — Författare — Author : Prosanta Kumar Singha			
Työn nimi — Arbetets titel — Title: Mutational analysis of splicing activation by the USSE motif in the U11/U12-specific 65K gene			
Oppiaine — Läroämne — Subject : Biotechnology			
Työn laji — Arbetets art — Level Master's Thesis		Aika — Datum — Month and year February 2013	Sivumäärä — Sidoantal — Number of pages 52
Tiivistelmä — Referat — Abstract <p>Splicing of introns and joining of exons to yield mature and functional mRNA molecules is carried out through two sequential trans-esterification reactions by two distinct spliceosomes; the major U2-dependent and the minor U12-dependent spliceosomes. Alternative splicing is common in eukaryotes and occurs mostly with U2 type introns. The level of minor spliceosomes specific proteins, U11-48K and U11/U12-65K, are regulated via alternative splicing where U11 snRNP binds to the USSE (U11 snRNP binding splicing enhancer), a tandem repeat of 5'ss of U12-type introns, and activates an upstream 3'ss of the U2 dependent spliceosome. Here, I carried forward the study done in Verbeeren et al, 2010, through manipulation of the distance between the two 5'ss within the USSE element, and the distance between the upstream 3'ssplice site and USSE. The mechanisms through which USSE recognition by the U11 snRNP is achieved, as well as the distance constraints the USSE element imposes on alternative splicing, were analyzed. RT-PCR analysis shows that (a) occurrence of the alternative splicing event drops with reduced distance between 3'ss-USSE, and is finally lost when the distance is reduced to zero, (b) both increased and decreased distance between the two 5'ss within the USSE resulted in reduced long isoform formation. The data set suggests that the simultaneous binding of two 5'ss within USSE by U11/U12 snRNPs and their interaction is necessary for USSE mediated alternative splicing activation. So, the evolutionary conserved sequence within USSE, and the distance to 3'ss are critical for USSE mediated alternative splicing.</p>			
Avainsanat — Nyckelord — Keywords: pre-mRNA splicing, Alternative splicing, USSE, U11/U12-65K, 3' long UTR, RT-PCR, Evolutionary conservation.			
Säilytyspaikka — Förvaringsställe — Where deposited:			
Muita tietoja — Övriga uppgifter — Further information: Thesis Supervisor: Docent Mikko Frilander, Institute of Biotechnology, University of Helsinki, Finland			

REVIEW OF THE LITERATURE

1.1 Introduction:

Most of the protein coding genes in eukaryotes are interrupted by intervening non-coding sequences (introns) with the coding parts arranged as exons. The precise removal of introns by splicing is a ubiquitous process to yield mature, functional messenger RNA (mRNA) (Reviewed by Patel and Steitz, 2003). Splicing has a vital role in normal gene expression as well as disease (Wang and Cooper, 2007). Expression of proteins by maintaining correct reading frame in the resulting mRNA via proper recognition and removal of the introns and joining of the exons is a big challenge for splicing machinery.

Introns have a wide size distribution and density in different organisms. “Introns-late” and “Introns-early” are the two proposed models for origin of introns in different evolutionary lineages (Darnell, 1978; Doolittle, 1978). Introns-late model postulates that the introduction of introns in ancestral eukaryotic protein coding genome happened via horizontal gene transfer and independently in different evolutionary lineages whereas the introns-early model proposed that introns arose either before or at the same time as the protein coding sequences. Yeast, which is a commonly used model organism that has only few hundred constitutively spliced introns in its genome, supports the intron-late theory (Logsdon Jr, 1998). In contrast, most current models arising from genome sequences support intron-early model. Comparative genomics across eukaryotes suggests that common eukaryotic ancestors had high intron density and weakly conserved splice sites.

1.2 Processing of eukaryotic pre-mRNA

In eukaryotes, the initial step of gene expression is transcription where pre-mRNAs are copied by RNA polymerase II (RNAPII) from DNA to complementary sequence of RNA. Normally chromosomes are tightly packed by DNA-histone interaction and sequestered in different domains of the nucleus of the cell (Lanctot et al., 2007). To start

transcription, DNA must be unpacked and be accessible for RNA polymerase. RNA polymerase is recruited by a set of transcription initiation factors assembled within the promoter region of the gene (Valen and Sandelin, 2011). During transcription, the nascent transcript undergoes capping at its 5' end, splicing to remove intronic sequences and finally addition of a long poly-A tail to its 3' end. All of these modifications are required to form a mature mRNA that can be transported into cytoplasm for protein synthesis (Singh and Padgett, 2009).

1.3 Two parallel intron types in eukaryotes

Introns are the typical feature of eukaryotes and vary in length and density; some unicellular eukaryotes have only a few introns while multicellular eukaryotes typically have a large number of introns, which are also longer and constitute a larger fragment of their genome (Michael and Manyuan, 1999). In humans, intron length varies from less than 100 base pair to several hundred kilobases and average intron density is 6.9 introns/gene (Csuros et al., 2011). An extreme example is the gene that codes for dystrophin 2.5 million bp long and contain 79 exons, which nevertheless cover only 1% of the total length of the gene (Pozzoli et al., 2002).

1.3.1 Structure

Apart from spliceosomal introns, many organisms e.g. Archaea, bacteria and eukaryotic mitochondria and chloroplast carry self-splicing introns- group I and group II (Bonen and Vogel, 2001; Haugen et al., 2005). These fold on specific 3D-structures to carry out self-splicing reaction. In comparison, spliceosomal introns are excised by a *trans*-acting spliceosome and have three regions with moderate conservation viz. 1) 5'ss: Junction between 3'-end of exon and 5'-end of intron which comprises first six nucleotides of the intron and up to three nucleotides of exon, 2) 3'ss: junction between 3'-end of intron and 5'-end of exon which is preceded

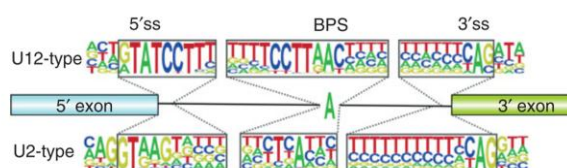


Fig. 1: Structure and consensus sequences of human U12- and U2- type introns. The height of the letters correlates frequency of corresponding nucleotides in that position (Turunen et al., 2013).

by a conserved intronic region and finally 3) Branch point sequence (BPS) - where a catalytically active adenosine (A) residue reside within the BPS, located upstream of 3'ss. These sequences are recognized by specific RNA-protein complexes termed as snRNP (small nuclear ribonucleoprotein) and form the catalytically active spliceosomal machinery to precise excision of introns and exon ligation (Reviewed by Schwartz and Ast, 2010).

1.3.2 Characteristics

The genome of most of the eukaryotes harbor two sets of spliceosomal introns viz. the U2- type intron and the U12- type introns. U12-type introns are present in most eukaryotes, account for less than 0.5% introns in any given genome (Reviewed by Turunen et al., 2013) while the U2- type intron constitute the majority of the introns (over 99.5 %). Thus the types of the introns are known as the minor and major introns, respectively (Reviewed by Patel and Steitz, 2003). The length of both introns varies greatly having a mean length of 4,130 base pairs (bp) for U2-type and 3,600 base pairs for U12 type introns.

Both intron types can be differentiated easily from each other by their precise sequence composition in splice sites and BPS. U2-type introns have consensus sequence AG/GTAAGT at 5' end where slash denotes intron-exon boundary (Sheth et al., 2006) (See Fig. 1). A shorter sequence signal CAG at the very 3' end of intron denotes 3'ss. BPS is located 20-40 bp upstream of 3'ss and shows specific consensus sequence. For example, human BPS is simply $\gamma\text{Un}\underline{\text{A}}\gamma$, where the underlined A is the branch point Adenosine (Gao et al., 2008) (See Fig. 1). The polypyrimidine tract (PPT) located between BPS and 3'ss has a variable length and contributes to the strength of 3'ss together with BPSs (Corvelo et al., 2010).

In contrast to U2-type splicing signals, the defining features of U12-types introns are highly conserved sequences at 5'ss and BPS position but they lack a distinct PPT (Reviewed by Turunen et al., 2013) (See Fig. 1). The 5'ss of U12-type introns have invariably conserved sequence /RTATCCTTT in all organisms (where slash denotes

intron-exon boundary and R denotes purine) (Sheth et al., 2006). The 3'ss is generally featured by terminal nucleotides as YAG and YAC (Y stands for a pyrimidine) in introns starting with GT and AT, respectively (Dietrich et al., 2001; Levine and Durbin, 2001). The distance between BPS and 3'ss of U12-type introns is a crucial factor and optimal at 11-13 nucleotides (nt) for intron recognition (Reviewed by Turunen et al., 2013).

1.4 Splicing machinery: Components of two Spliceosomes

Intron recognition and removal from pre-mRNA is carried out by two distinct splicing machineries- viz. spliceosomes that are specific to either the major U2-type or the minor U12-type introns. Both machineries consist of five small nuclear ribonucleoprotein particles (snRNPs) and the numerous non-snRNP protein components (Will and Luhrmann, 2011).

Each of the five snRNPs of the major U2-type spliceosome consists of one small nuclear RNA (snRNA) (U1, U2, U4, U5 and U6) and a number of protein components. Each of the U1, U2, U4 and U5 snRNAs are associated with a 7-protein ring-like structure. These sm-proteins bind to a specific sequence, so-called sm-site in the snRNA (See Fig. 2). U6 snRNA forms a similar structure with Sm-like (Lsm) proteins (Reviewed by Snehal Bhikhu and Michel, 2008). The snRNPs also contain some specific protein factors: U1 snRNP has 3 specific protein factors e.g. U1A, U1C and U1-70K. U2 snRNP has more than 11 protein factors: U2A', U2B' and two multiprotein complexes, SF3a and SF3b (Boelens et al., 1990; Kramer et al., 1999). U4 and U6 snRNAs form U4/U6 di-snRNP by base pairing with each other via highly conserved complementary sequences along with specific protein factors (Bringmann et al., 1984; Nottrott et al., 2002) (See Fig. 2). U5 snRNP connects with U4/U6 di-snRNP through protein-protein interaction and form U4/U6-U5 tri-snRNP together with specific protein factors (Behrens and Luhrmann, 1991; Black and Pinto, 1989). In addition to snRNPs, many non-snRNP proteins are also needed for intron recognition and splicing. For example: splicing factor 1 (SF1) and U2 auxiliary factor (U2AF) are important for recognizing the BPS, PPT and 3'ss (Berglund et al., 1997;

Ruskin et al., 1988; Wu et al., 1999; Zamore and Green, 1989; Zorio and Blumenthal, 1999).

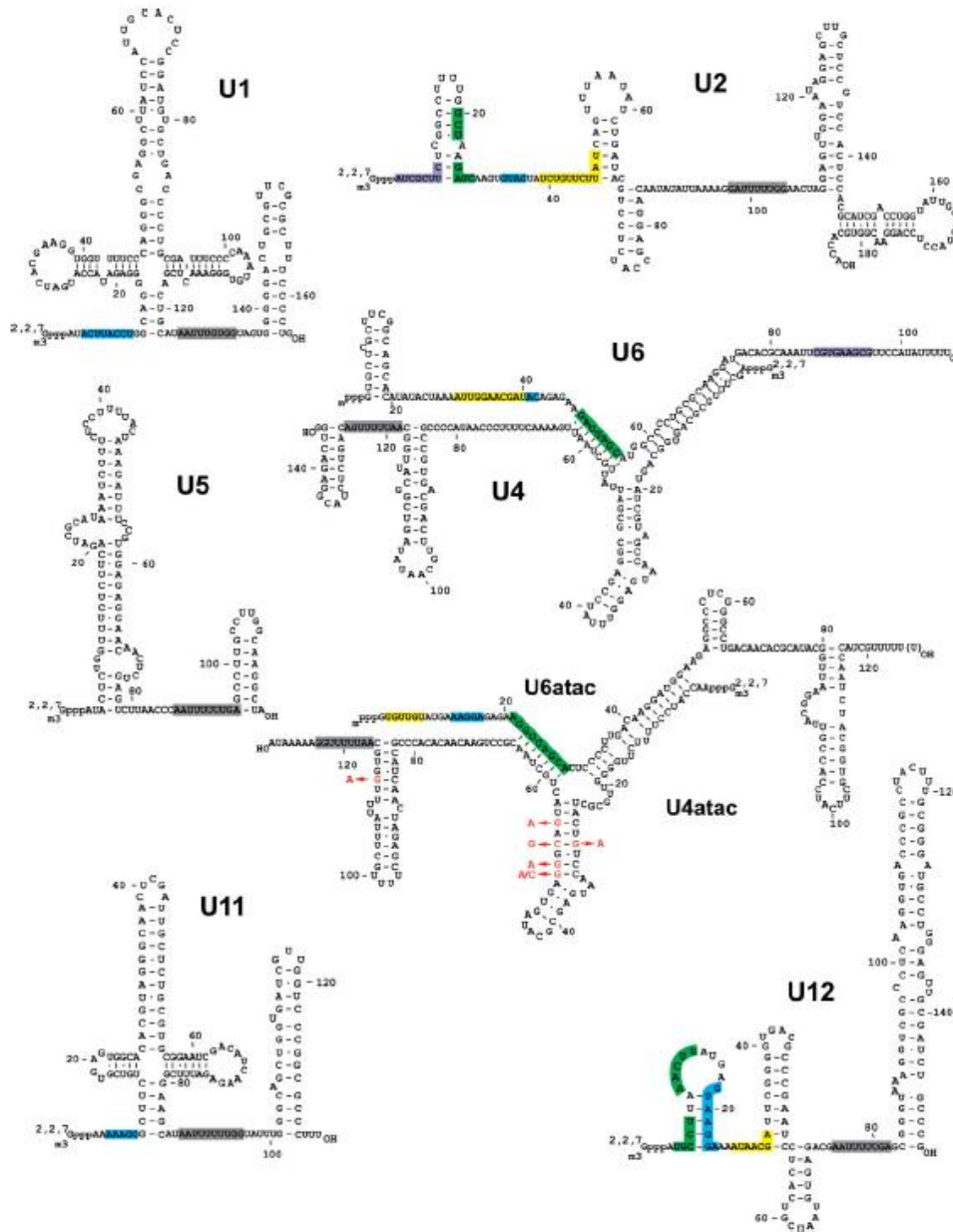


Fig.2: Sequences and predicted secondary structures of the human spliceosomal snRNAs. Functionally corresponding snRNAs from the two spliceosomes have similar secondary structures: compare U1 and U11, U2 and U12, and U4-U6 and Uatac-U6atac. Gray shade denotes binding site for Sm proteins, Cyan shade for sequences interacting with 5'ss or BPS. Green (helix I), purple (helix II) and yellow (helix III) colored sequences involved in various U2/U6 or U12/U6atac interactions (Reviewed by Turunen et al., 2013)

The U12 dependent spliceosome also comprises five snRNPs which are U11, U12, U4atac, U5 and U6atac. U5 is the common component of both spliceosomes where U11 and U1, U12 and U2, U4atac and U4, and U6atac and U6 are the structural and functional counterparts of each other (Hall and Padgett, 1996; Tarn and Steitz, 1996a, b). U11 and U12 snRNPs interact with each other and form a preformed di-snRNPs before spliceosome assembly (Wassarman and Steitz, 1992). Most of the protein components are shared between two spliceosomes. Minor snRNAs are complexed with Sm and Sm-like proteins in the similar manner as the major ones (Tarn and Steitz, 1996a). Only seven proteins are specific to minor spliceosome and are located within U11/U12 di-snRNP (Will et al., 2004; Will et al., 1999). U4atac/Uatac-U5 tri-snRNP share all of the protein component of the U4/U6-U5 tri-snRNPs (Schneider et al., 2002).

1.5 Spliceosome assembly and catalysis

1.5.1 Splicing catalysis mechanism:

Both spliceosomes have a similar mechanism of splicing which involves a two-step transesterification reaction resulting in an excised intron lariat and ligated exons (Figure 3; reviewed by Will and Luhrmann, 2011). The first step of the transesterification reaction generates a free 3'-OH group at the 5' exon and a lariat intron structure attached with the 3' exon through a 2'-5' phosphodiester bond between the 2' OH group of the BPS adenosine and intron 5' phosphate group. In the second step of reaction, the 3' OH of the 5' exon attacks the phosphodiester bond at 3'ss resulting in the release of the lariat intron structure and ligated exons. Splicing is itself an energetically neutral process but ATP consumption in different energy-requiring steps is necessary to ensure the specificity and unidirectionality of the reaction.

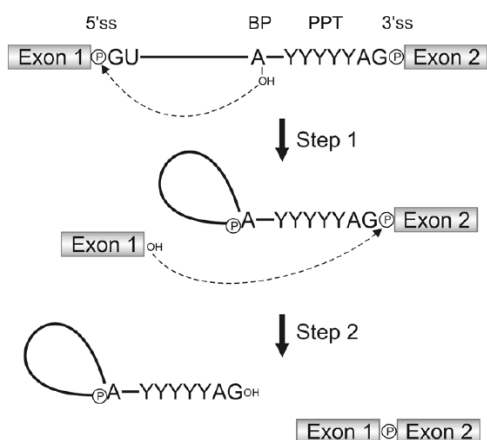


Figure 3: Catalysis mechanism of splicing.

A schematic pre-mRNA is shown here as single intron flanked by two exons. The first and second step of splicing involve nucleophilic attacks on the phosphate groups of 5'ss and 3'ss by 2'OH of BPS adenosine and 3'OH of 5'exon, respectively, producing ligated exons and lariat intron (Turunen, 2012).

1.5.2 Assembly of U2 and U12 dependent spliceosome

The interaction between snRNPs and pre-mRNA is tightly controlled and established in a stepwise manner to avoid premature and unspecific activation of spliceosome. During spliceosome assembly reactive sites of pre-mRNAs are recognized multiple times by snRNA or protein factors allowing binding partner exchange to remodel the spliceosome at different stage of its assembly, thereby making the spliceosome dynamic in composition and conformation during its maturation (Smith et al., 2008; Wahl et al., 2009).

The step-wise assembly of U2 dependent spliceosome can be described in terms of complexes (E, A, B* and C; Fig. 4) that are formed following the association of snRNP and other components with pre-mRNA. Formation of commitment complex (E complex) is the initial step of intron recognition where U1 snRNA base pairs with the 5'ss via its complementary sequence at the 5' end (Zhuang and Weiner, 1986). The 5'ss/U1 helix interaction is stabilized by RNA-protein interaction especially association with U1C protein (Heinrichs et al., 1990; Pomeranz Krummel et al., 2009). During this stage, the BPS is recognized and bound by splicing factor 1 (SF1) which bulges out the branch site adenosine to allow nucleophilic attack in the later stages of spliceosomal maturation. The PPT and 3'ss are recognized by 65kDa and 35kDa subunits of U2 auxiliary factor

(U2AF), respectively (Berglund et al., 1997; Wu et al., 1999; Zamore and Green, 1989; Zorio and Blumenthal, 1999).

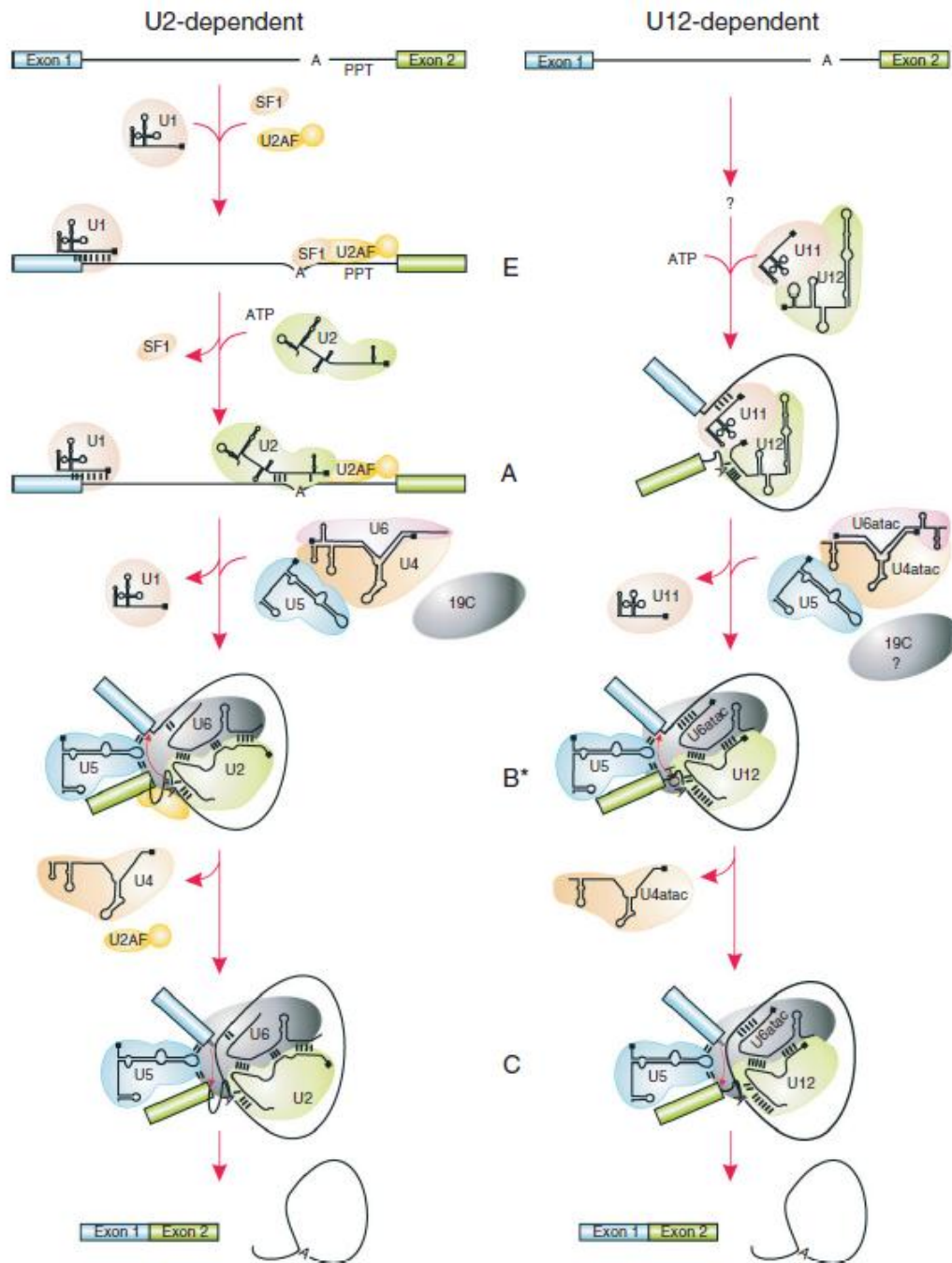


Figure 4: Pathways of assembly and catalysis of the major U2-dependent and minor U12-dependent spliceosomes. The interaction between spliceosomal snRNPs and some selected non-snRNP proteins are depicted schematically for both spliceosome at various stages of spliceosomal assembly. The pathways are mechanistically very similar except the early steps of spliceosome formation (Reviewed by Turunen et al., 2013).

During the formation of pre-spliceosome (ATP dependent A complex), SF1 is displaced followed by recruitment of U2 snRNP to the BPS through interaction with SF3b complex (Ruskin et al., 1988; Rutz and Seraphin, 1999; Spadaccini et al., 2006; Valcárcel et al., 1996) (See also Fig. 4). The U4/U6.U5 tri-snRNP associates with 5'ss after conformational changes in the pre-spliceosome, resulting in B complex formation where U1 is displaced from the 5'ss and replaced by U6. U4/U6 interaction is unwound and U4 is released from spliceosome (Lamond et al., 1988). U6 base pairs with U2 to form a catalytically active RNA structure (Kandels-Lewis and Séraphin, 1993; Madhani and Guthrie, 1992). Interactions between snRNAs and protein factors are remodeled, many snRNP proteins dissociate to yield the activated spliceosome or complex B* that catalyses the first transesterification reaction (Fabrizio et al., 2009)(reviewed by (Turunen, 2012). After the first catalytic step, U2 snRNP specific proteins dissociates and spliceosome is joined by further factors which facilitate formation of complex C to catalyze the second catalytic step (Agafonov et al., 2011; Bessonov et al., 2008; Fabrizio et al., 2009).

The overall assembly pathways of both spliceosomes are similar except for the absence of a separate commitment complex in the U12 dependent spliceosome where a pre-formed U11/U12 di-snRNP act as a unit to recognize the 5'ss and BPS in a cooperative manner, forming the A complex (Frilander and Steitz, 1999)(See Fig. 4). The initial recognition of 5'ss in both spliceosomes are different. U11 snRNA does not base pair with 5'ss boundary or even the first three nucleotides of the intron. Rather these nucleotides are recognized by the U11 snRNA-specific 48K protein which stabilizes the U11/5'ss helix (Turunen et al., 2008). Despite of the differences in intron recognition, steps leading to catalytic core formation are similar in both spliceosomes: U12/BPS duplex is highly analogous to U2/BPS duplex (Tarn and Steitz, 1996b), U4atac/U6atac.U5 tri-snRNP joins the pre-spliceosome to form B complex, leading to the displacement of U11 snRNP from 5'ss by U6atac, which also base pairs to U12, with parallel release of U4atac (Frilander and Steitz, 2001; Incorvaia and Padgett, 1998; Tarn and Steitz, 1996a; Yu and Steitz, 1997). Remodeling of RNA-RNA interaction network is similar to major

type spliceosomes, which forms catalytically active spliceosome to catalyze two steps transesterification reaction, resulting in ligation of exons and release of intron lariat (Tarn and Steitz, 1996b) (See Fig. 4).

1.6 Alternative splicing

Alternative splicing (AS) of pre-mRNA is a central process in eukaryotes that allows individual genes to produce two or more mRNAs isoforms which may encode functionally distinct proteins, thereby increasing protein diversity and giving an additional layer of control over posttranscriptional gene expression. Recent high-throughput sequencing technology revealed that 92-94% of multi-exon containing human genes are alternatively spliced (Wang et al., 2008). AS is dominant in U2-type introns but almost entirely absent in U12-type introns possibly because of their highly conserved sequence and distance constraints on the 5'ss, BPS, and 3'ss (Levine and Durbin, 2001).

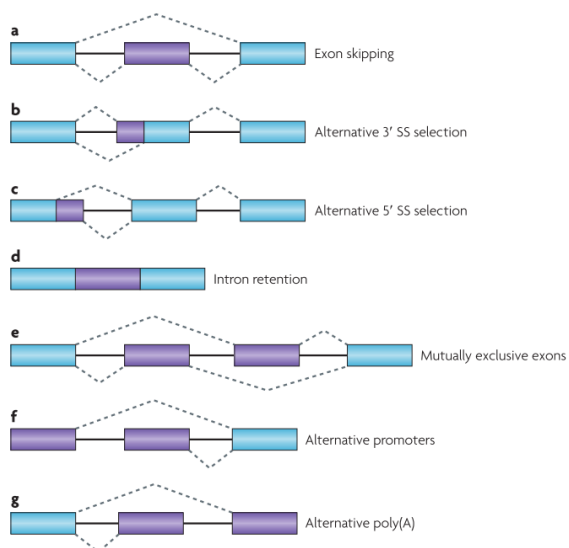


Figure 5: Types of alternative splicing. Various types of alternative splicing events are depicted schematically. Blue and purple boxes represent constitutive exons and alternatively spliced regions, respectively. Solid line represents introns, and dashed lines indicate splicing options (Reviewed by Keren et al., 2010).

The most common type of AS observed is the exon skipping (Fig. 5a), where a type of exon known as cassette exon is spliced out together with its flanking introns. Exon skipping is the most frequent AS event in higher eukaryotes (40% of AS events). Alternative 3'ss and 5'ss selection are (Fig. 5b and c) slightly less common and take place when more than one splice site is recognized at 5' or 3' end of exon, respectively. Intron retention (d) in the mature mRNA transcript is the rarest AS event in vertebrates and invertebrates (<5%) but most prevalent in lower eukaryotes. Some other less frequent AS events include mutually exclusive exons (e), usage of alternative promoter (f), and alternative polyadenylation (Reviewed by Keren et al., 2010).

1.6.1 Splice site selection and alternative splicing

Exon definition and intron definition are the two models to describe the mechanism of exon and intron selection. The ancient mechanism by which an intronic sequence unit is recognized and basal splicing machinery is recruited across introns termed as intron definition model. In the exon definition model, splice sites are recognized across exons which constraints the length of exons (Reviewed by Keren et al., 2010; See also Fig. 6).

Due to the co-transcriptional nature of splicing, the processing of nascent transcript of extremely long introns (especially vertebrates) becomes difficult because it can take hours for them to be transcribed entirely. In that case, initial splice site definition takes place via interaction of 3'ss-recognizing factors in the upstream intron with 5'ss-recognition factors in the downstream intron over considerably shorter exons (See Fig. 6). Then these exon definition interactions are replaced by the interaction between the 5'ss and 3'ss-recognition factors across the same intron. In contrast, plants, fungi and many invertebrates have shorter introns that are defined directly (Reviewed by Turunen et al., 2013). Exon definition is particularly important in mammals where exon skipping is the most common form of AS which results in from a failure to define the middle exon which, in turn, leads to paring of splicing factors located upstream and downstream exon (Nakai and Sakamoto, 1994).

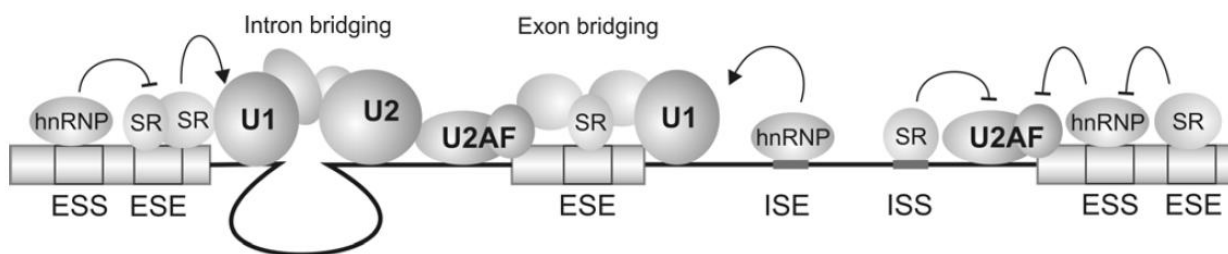


Figure 6: Splicing regulatory elements and splice site selection.

Splicing is regulated by trans-acting factors that bind to enhancer and silencer elements and interacting with 5'ss and 3'ss-recognition components, either through exon definition or intron definition interaction. Exons and exonic regulatory elements (ESEs, ESSs) are depicted as boxes, and intron and their regulatory elements (ISEs, ISSs) by lines. Arrows indicates activation of interaction between splicing factors, and blocked lines indicate suppression of interactions (Reviewed by Turunen et al., 2013).

1.6.2 Factors affecting alternative splicing

AS is usually regulated during recognition of splice sites by the spliceosome. It depends on the presence of short splicing regulatory elements (SREs) within introns and/or exons of pre-mRNAs and the availability of the *trans*-acting factors that vary from cell to cell and in different developmental stages. SREs can be recognized by *trans*-acting regulatory proteins that either enhance or suppress the stepwise assembly of spliceosomes. Canonical and well characterized *trans*-acting factors are most common protein factors, in particular SR proteins and hnRNP proteins, small nucleolar RNAs (snoRNA) and snRNPs (Reviewed by Khanna and Stamm, 2010). The combinatorial actions of activatory and inhibitory factors determine the choice of splice site by affecting the assembly rate of spliceosome at the competing splice site (Fu, 2004; Matlin et al., 2005; Wang and Burge, 2008).

1.6.2.1 Splicing enhancer and silencer

The core splicing signals (5'ss, 3'ss and BPS) within introns, especially in U2-type introns, are very short in length and degenerate in nature, thus are not enough to define the splice sites distinctly. Additionally, the 5'ss can be present within the same intron many times in positions which can define putative exons (pseudo exons) with upstream 3'ss like sequence (Burge et al., 1999; Lim and Burge, 2001). Thus, additional information is required to define splice sites. This information comes from enhancer or silencer sequences located close to the splice site (Fig. 6). They are named by their position either in exons or introns i.e., exonic splicing enhancers/silencers (ESEs/ESSs) and intronic splicing enhancers/silencers (ISEs/ISSs). Each of these splicing regulatory element (SRE) subtypes is composed of short and variable sequences, and works in combination. They can either suppress or activate splice sites and multiple sites can work either in co-operative or antagonistic manner. As SREs are typically rather degenerate, their effect on splicing are difficult to predict by sequence information only (Wang and Burge, 2008).

1.6.2.2 SR proteins

The best characterized trans-acting protein factors that positively regulate AS are called as serine-arginine-rich proteins (SR proteins). They are structurally related to each other, having a downstream RS domain (arginine-serine-rich domain) and at least one N-terminal RNA recognition motif (RRM). In addition to splicing regulation, they also operate diverse pathways in gene expression including chromatin remodeling, nuclear mRNA export, nonsense mediated decay (NMD) and regulation of translation (Reviewed by Shepard and Hertel, 2009). Since binding sites of the SR proteins are present within both constitutively and alternatively spliced exons, they are best characterized as a splicing activator that recruit spliceosomal components such as U2AF to the 3'ss and U1 snRNP to the 5'ss through binding to ESEs (Reviewed by (Busch and Hertel, 2012). A well-characterized SR protein with multiple function is SRSF1 which stimulates binding of U1 snRNP to the 5'ss and also stabilize U2/BPS and U2AF35/3'ss interactions (Kohtz et al., 1994; Shen and Green, 2004).

SR proteins also stimulate splicing by counteracting the inhibitory effect of splicing silencer hnRNPs (See below). Apart from ESE-dependent splicing activation, SR proteins also suppress the inclusion of alternative exon by recognizing intronic splicing silencer (ISSs) elements (Buratti et al., 2007b).

1.6.2.3 HnRNP proteins

HnRNP proteins are the best characterized family that typically negatively regulates splicing. Unlike SR proteins, hnRNP proteins do not form a unified group of clearly defined domains. All have RNA-binding domains, of which the RRM domain is most common, but others have structurally distinct RNA binding domains such as KH domains. HnRNPs, showing varied functions, are historically considered as a splicing silencer but splicing activation functions are also common among hnRNPs (Reviewed by Han et al., 2010).

HnRNP proteins influence pre-mRNA splicing via site-specific binding with the target RNA. This binding is stabilized by the RRM or the KH domains that mediate splicing decision by promoting protein-protein interactions. Known cases include repression of spliceosomal assembly and looping out entire exons (Reviewed by Busch and Hertel, 2012).

The competition for binding sites and blocking of silencing interaction of hnRNPs by SR proteins is important for AS that promotes inclusion of alternative exons (Han et al., 2011; Zhu et al., 2001).

1.6.2.4 Core spliceosomal components

SR proteins and hnRNP proteins are not the only regulators of alternative splicing, as core splicing components also affect splicing regulation. Several core spliceosomal factors like U2AF and protein components of U1, U2 and U4/U6 snRNPs were identified as factors of regulating alternative splicing during RNA interference screen using *Drosophila melanogaster* S2 cells (Park et al., 2004). In human cells, depletion of the common snRNP component SmB/B' reduced the level of inclusion of alternatively

spliced exons while the inclusion of constitutive exons remain unchanged (Saltzman et al., 2011). Core spliceosome component U1 snRNP appears to be a prevalent AS regulator. In *ATM* gene, U1 binds to an ESS close to 3'ss that leads to unstable binding of U2 at upstream sequence, resulting in suppression of exon inclusion (Dhir et al., 2010). Splicing can also be regulated by multiple snRNPs as illustrated by the *gag* gene of RSV (Rous sarcoma virus) where a pseudo-intron is located within a true intron and prevents splicing of true intron after being bound with full component of U2-type snRNP (Reviewed by McNally, 2008).

1.6.2.5 Other factors

Since splicing is a co-transcriptional event, factors affecting kinetics of RNA polymerase II (RNAPII) elongation also have affect on splicing. Pausing on certain exons and overall speed of RNAPII may favor some splicing event over other (Reviewed by Nilsen and Graveley, 2010). Similarly, masking of splice sites by RNA secondary structure has been shown to affect splicing (Buratti et al., 2007a). RNA structure can inhibit splicing by blocking the binding sites of splicing factors within the target pre-mRNAs. On the other hand, in some introns 3'ss and BPS are located hundreds of nucleotides away and RNA secondary structure brings them in close proximity to promote splicing (Reviewed by Warf and Berglund, 2010).

1.6.3 Significance of alternative splicing

AS is important not only for increasing proteome diversity but also has an important role in regulation of gene expression. A well-documented example of the regulation of gene expression is the coupling between nonsense-mediated decay (NMD) and alternative splicing, where “poison cassette exon” inclusion leads introduction of premature termination codon (PTC) that results in mRNA decay (Lewis et al., 2003).

1.7 Quality control of mRNA

In eukaryotic cells, normal and functional mRNA biogenesis is ensured by numerous quality control systems which preferentially degrade nonfunctional and aberrant RNAs.

RNA quality control processes exist both in nucleus and cytoplasm. Cytoplasmic quality control is carried out by the exosome which is composed of a ten-subunit core complex having 3' to 5' exonuclease activity, and Xrn1p which is a 5' to 3' exonuclease that degrades decapped mRNAs. The major quality control machinery in the nucleus is nuclear exosome which has both exo- and endonuclease activity. Besides, Xrn2/Rat1p, a paralog of Xrn1p, may also function in nuclear RNA degradation (Reviewed by Doma and Parker, 2007).

1.7.1 Nuclear degradation

Transcripts that do not complete splicing process are generally retained in the nucleus and degraded by nuclear exosome or by exonuclease Xrn2/Rat1p. Though the mechanism of unspliced mRNA retention is unclear, it is possible that all transcripts are retained in the nucleus by general retention factors (Reviewed by (Turunen, 2012)). Only fully processed transcripts are exported to cytoplasm after gaining a marker such as poly-A tail and/or other specific protein complexes, while unmarked transcripts undergo the degradation pathway by exosome or Xrn2/Rat1p (Reviewed by (Egecioglu and Chanfreau, 2011)).

1.7.2 Cytoplasmic quality control

There are several cytoplasmic quality control pathways that degrade eukaryotic mRNAs with abnormalities in translation. Adapter proteins are the key factor that interact with translation machinery and distinguish between normal mRNA and aberrant mRNA. Aberrant mRNAs are then directed to the degradation pathways depending on their defects. Aberrant mRNAs, containing premature translation termination codons (PTC), are distinguished by the conserved Upf protein and their interaction with translation termination complex, and go through nonsense-mediated decay (NMD) pathway (Reviewed by Doma and Parker, 2007). In NMD pathway, aberrant mRNAs are subjected to decapping and 5'>3' endonuclease degradation by Xrn1p; or accelerated deadenylation and 3'>5' degradation by exosome (Reviewed by Isken and Maquat,

2007). On the other hand, when mRNAs do not have any stop codon, ribosomes reach to the end of mRNA and recruit the exosome that cause rapid 3'→5' degradation, termed as nonstop decay (NSD). Similarly, mRNAs undergo another endonucleolytic cleavage pathway termed as 'No-Go' decay (NGD), when their translation elongation has been paused strongly. Finally, when translation continues beyond the normal stop codon into 3' UTR, it activates ribosome extension-mediated decay (REMD) pathway (Reviewed by Doma and Parker, 2007).

1.8 Alternative splicing in U11/U12-65K gene

1.8.1 Evolutionary conservation of U11/U12-65K gene.

An evolutionarily conserved 110bp sequence element in U11/U12-specific 65K gene was identified in a bioinformatic analysis of 29 vertebrate genome alignments (Verbeeren et al., 2010). This intronic element is highly conserved in mammals, birds and lizards. In phylogenetically distant organisms like mammal and fish, the conservation is limited into two regions- one contains a U2-type 3'ss and an upstream PPT, and the other contains a tandem repeat of two U12-type 5'ss motifs (Fig. 7). Importantly, these 5'ss sequences are not used for splicing as shown by RT-PCR analysis but rather regulate alternative splicing of the upstream U2-type 3'ss. This regulates the length of 3' UTR. The regulatory sequence having two U12-type 5'ss is termed as U11 snRNP-binding splicing enhancer (USSE). Similar sequence conservation for USSE was also found in another gene U11-48K in mammals, fishes, insects and plants. The presence of a USSE element in distantly related organisms and both in plant and animal genomes suggests that USSE mediated gene regulation has been evolved and existed in the last common ancestor of eukaryotes (LECA) and thus considered as the oldest known splicing regulatory element (SRE) (Verbeeren et al., 2010).

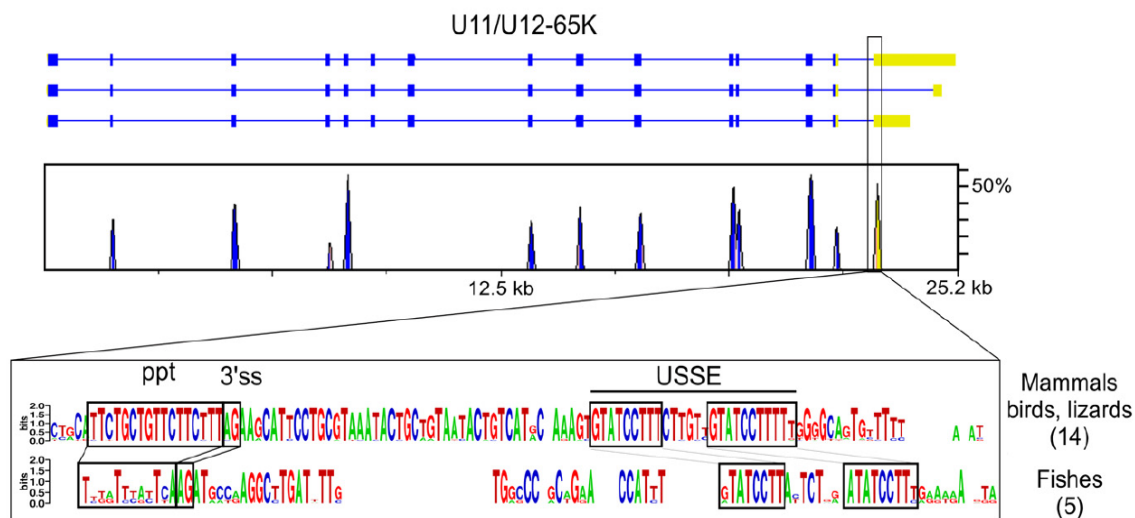


Figure 7: Conserved sequence element in animal U11/U12-65K genes. Genomic organization and splicing variation of each transcript is shown in upper part and phylogenetic conservation plot in lower part. A blow-up shows the residue-level conservation of the 3'UTR element. Protein coding sequences are represented in blue and UTRs in yellow. Exons and introns are indicated by solid boxes and horizontal lines, respectively (Modified from (Verbeeren et al., 2010)).

1.8.2 USSE mediated alternative splicing

The U11 snRNPs of the U12-dependent spliceosome recognizes the 5' splice site of U12-type introns and also acts as an activator of the U2 dependent alternative splicing. The recognition of U12-dependent 5'ss by U11 is found to be stimulated and established by U12/BPS interaction (Frilander and Steitz, 1999). So, the lack of BPS in neighboring sequence of USSE is believed to be compensated by the presence of duplicated 5'ss. The U11/5'ss interaction within USSE might be established by the communication between two U12 snRNPs, and thus activating the usage of upstream U2-type alternative 3'ss and regulating the level of minor spliceosomes specific proteins, U11-48K and U11/U12-65K (Verbeeren et al., 2010).

2 OBJECTIVES OF THE STUDY

An earlier study from Frilander laboratory revealed that USSE is recognized by U11 snRNP which promotes the use of an upstream U2-type 3'ss (Verbeeren et al., 2010). This alternative splicing event regulates the length of U11/12-65K 3' UTR. In this study, I investigated the distance requirement for upstream 3'ss activation by systemically modifying the distance of two 5'ss elements within USSE and the distance between USSE and 3'ss. Two hypotheses are taken into account for recognition and binding of the USSE by U11 snRNP and these are "simultaneous binding" and "affinity based binding" (Fig. 8). One of the aims was to distinguish between these two competing hypotheses for the duplication of 5'ss within USSE. The other aim was to investigate the distance constraints between 3'ss and USSE.

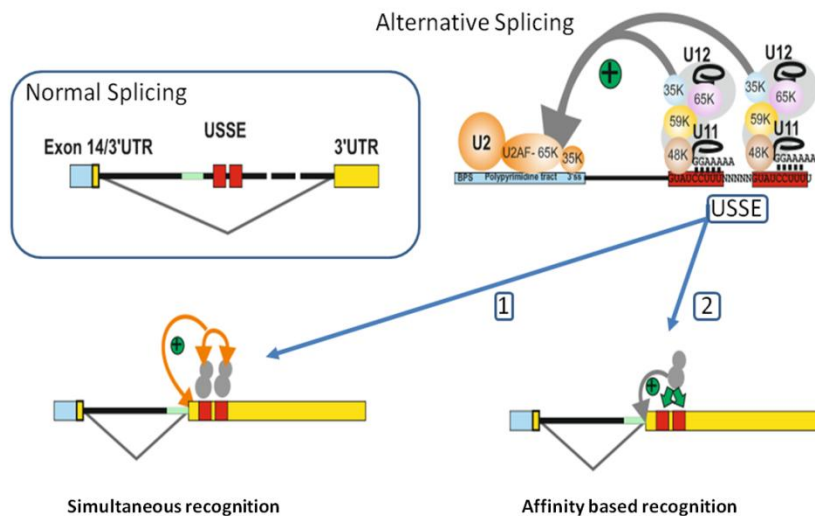


Figure 8: **USSE binding with U11 sn-RNP.** There are two hypotheses of USSE recognition and binding with U11 sn-RNPs: 1) Simultaneous recognition and b) increased affinity due to 5'ss duplication. In simultaneous recognition hypothesis, two 5'ss of USSE are recognized by two U11/U12 snRNPs and the stabilizing interaction between these two snRNPs play a vital role to promote the usage of upstream 3'ss, resulting in alternative splicing. In affinity based recognition hypothesis, U11/U12 snRNP recognizes either of the two 5'ss of USSE. The function of 5'ss duplication is to increase the U11/5'ss binding affinity. Normal splicing pattern e.g. the one not induced by USSE, is shown in a box.

3 MATERIALS AND METHODS

3.1 Construction of Plasmids

3.1.1 Molecular Cloning

The pGL4.13-65KUTR plasmid was obtained from common plasmid stock of *RNA splicing Lab*. It contained the unspliced 3' UTR with the USSE of the human *RNPC3* gene fused to pGL4.13 luciferase expression vector (Promega). Several plasmids described below were constructed using molecular cloning strategies using pGL4.13-65KUTR plasmid as a starting plasmid (wt). In wt plasmid, the distance between 3'ss and USSE motif is 44 nt, and the distance between two U12-type 5'ss of USSE is 6nt. PCR based mutagenesis was used to modify distances between upstream 3'ss and USSE, and between the two U12 type 5'ss motifs within the USSE (Fig. 9).

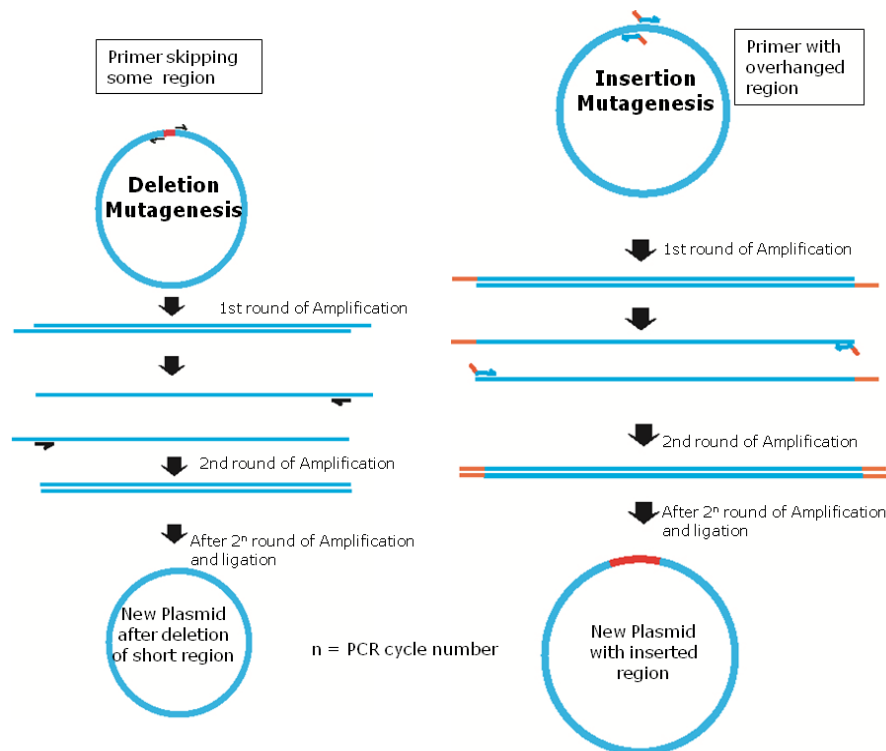


Figure 9: Site directed insertion and deletion mutagenesis by using specific set of primers. Primers are used to skip or add nucleotides during deletion or insertion mutagenesis. After every PCR cycle, the number of template DNA become double which results 2^n number of DNA fragments after completion of n cycles. Blue color denotes sequence derived from wild type plasmid while red color for deleted or inserted regions on specific plasmids.

Modification of the distance between 3'ss and USSE: pGL4.13-65KUTR-du10nt, -du20nt, -du30nt, and -d3ss44nt plasmids were constructed by deletion of 10, 20, 30, and 44 bp, respectively, in the sequence between 3'ss and USSE motifs (Fig. 10). A set of primers were used to amplify the wt plasmid skipping the required number of nucleotides resulting desired plasmid as denoted by the construct name. The pGL4.13-65KUTR denotes for wild type plasmid, "du" denotes for USSE side deletions, "3ss" denotes for 3'ss side deletions, and "**nt" is the number of nts deleted (Table 1-a).

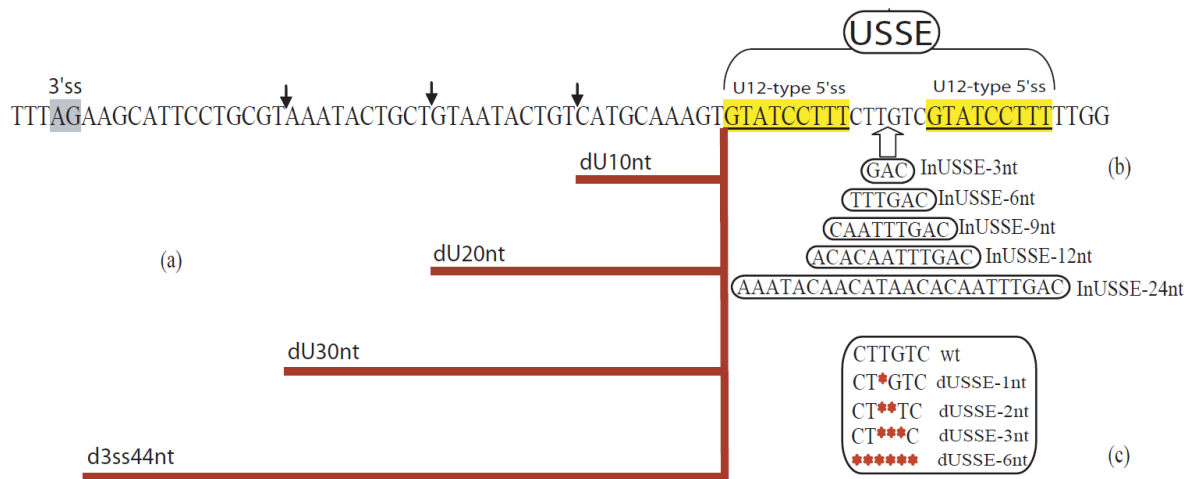


Figure 10: Constructs after deletion and insertion mutagenesis. (a) Deletion of 10, 20, 30 and 44 nucleotides were performed from USSE side towards 3'ss and denoted by dU10nt, dU20nt, dU30nt, and d3ss44nt, respectively. Arrows indicates the sites of deletions from USSE ends. (b) Insertion mutagenesis were performed in between USSE motifs where 3, 6, 9, 12, and 24 nt were inserted and denoted by InUSSE-3nt, -6nt, -9nt, -12nt, and -24nt, respectively. (c) Deletion of 1, 2, 3, and 6 nt in-between USSE motifs made dUSSE-1nt, -2nt, 3nt, and -6nt plasmids, respectively.

Modification of the distance between two U12-type 5'ss of USSE: pGL4.13-65KUTR-InUSSE-3nt, -6nt, -9nt, -12nt, -24nt and pGL4.13-65KUTR-dUSSE-1nt, -2nt, -3nt, and -6nt plasmids were constructed by inserting 3, 6, 9, 12, 24 and deleting 1, 2, 3, 6 nt, respectively (Fig. 10), in the sequence between two U12-type 5'ss of USSE. After deletion and insertion, the final distance between two U12-type 5'ss of USSE is 0, 3, 4, 5, 6 (wt), 9, 12, 15, 18 and 30nt. The primer sets used for construction of each plasmid are listed in table 1-a.

Rescue plasmid for USSE region: A third type of plasmid was constructed to regain the structural distances as above plasmids in different way, e.g. by making a hairpin structure. Plasmid pGL4.13-65KUTR-InUSSE-hp-24nt brings two USSE motifs close together in a distance of 6nt, e.g. same as wt. Primer sets used to make these plasmids are listed in table 1-a.

3.1.1.1 PCR and Gel extraction

For PCR reaction, the following reagents were used to make 50µl final reaction volume:

5×Phusion HF buffer	10 µl
10 mM dNTP	1 µl
100 µM Forward primer	0,5 µl
100 µM Reverse primer	0,5 µl
10 ng/µl plasmid pGL4.13-65KUTR	1,0 µl
2 U/µl Phusion polymerase	0,5 µl
MQ*	36, 5 µl
<hr/>	
Final volume =	50µl per sample

*Milli-Q water

A master mixture was made by multiplying the sample numbers with the amount of reagents listed above. Finally 49µl of master mix was distributed per PCR tube and 1µl of primer (0.5µl of each primer) was added before PCR starts. To perform final PCR reactions, a gradient PCR reaction was performed for each set of primers to check the optimum annealing temperature to get the maximum output from same input. Temperature was ranged from 61⁰C to 71⁰C during gradient PCR reaction. After the annealing temperature has been optimized, the final PCR reaction was performed according to the following program:

- 1: +98 °C 30 sec
- 2: +98 °C 15 s
- 3: +63* °C 20 s

4: +72 °C	4 min 15 sec**
5: Go to 2,	33 times
6: +72 °C	7 min
7: +15 °C	Forever

Where, * Annealing temperature determined from gradient PCR reaction.

** Elongation time depends on template size. Usually 30 sec/kb was used for template amplification.

PCR reaction mixture contains both methylated template plasmid DNA and newly synthesized DNA fragments. To avoid contamination from the template plasmid, the PCR reaction was treated with *DpnI* (Thermo-scientific) restriction enzyme (10U, +37°C, overnight). This step removes methylated template. Correct sizes were gel purified (0.8% topvision agarose) by using QIAquick gel extraction kit (*QIAGEN*).

3.1.1.2 Phosphorylation and Ligation of the DNA

Plasmid DNA fragments amplified by Phusion DNA polymerase have blunt ends with 5'-OH termini. To proceed into next stage of cloning, 5'-OH was phosphorylated by T4 polynucleotide kinase (T4 PNK) (Thermo-scientific). A modified protocol was used to manipulate the reaction volume as follows:

Approximately 100 ng of gel purified PCR products was used for phosphorylation. So, after gel purification, concentration was measured by Nano-drop spectrophotometer.

PCR product (100ng) and MiliQ	7.5µl
10×PNK buffer A	1.0µl
10 mM ATP	0.5µl
T4 PNK (polynucleotide kinase)	1.0µl
<hr/>	
Total reaction volume =	10.0µl

A master mixture was made by using required amount of reagents and divided accordingly with gel-purified DNA. Subsequently the reaction mixtures were incubated at +37°C for 30 min followed by +65°C for 20 min to inactivate the T4 PNK enzyme.

Ligation step was performed right after the completion of phosphorylation:

Kinase reaction (as above)	10µl
10×ligase buffer	5µl
MQ	34ml
5 U/µl T4 ligase	1µl
<hr/>	
Total reaction volume =	50µl

The ligation volume was 50µl to avoid intermolecular ligations. The reaction mixture was then incubated at either +16°C for over-night or +22°C for 3 hours.

3.1.1.3 Phenol-chloroform extraction and DNA precipitation

After successfully completion of phosphorylation and ligation reactions, plasmid DNAs were phenol-chloroform extracted followed by ethanol precipitation. Specifically the total volume was adjusted with MiliQ water to 200µl and 1µl glycogen was added. Extraction was done once with 50:48:2 phenol-chloroform-isoamylalcohol and once with 48:2 chloroform-isoamylalcohol. For ethanol precipitation 20µl 3M NaCl and 500µl 94% ethanol was added. After 30 min of centrifugation (13,000 rpm, +9°C), pellet was washed twice by 70% ethanol, dried, and dissolve in 5µl MQ.

3.1.1.4 Electroporation

For high transformation efficiency and maintenance of large plasmids, *Escherichia coli* DH10B (F⁻ *mcrA* Δ (*mrr-hsdRMS-mcrBC*) Φ 80*lacZ* Δ M15 Δ *lacX74* *recA1* *endA1* *araD139* Δ (*ara leu*) 7697 *galU* *galK* *rpsL* *nupG* λ -) was used to propagate the reporter plasmids.

25µl of electro-competent DH10B cells (lab stock) was taken with 2µl of precipitated plasmid DNAs above into an electro-cuvette (pre-chilled). The gene pulser (Bio-Rad) was set as resistance = 200 Ω, capacitance = 25 µF and voltage to 1.8 kV. Following the electric pulse 1ml SOC was added and then cells were incubated at +37°C shaker for 45 min. After incubation, cells were plated with different ratios into several plates of LB-ampicillin (100µg/ml) and incubated at +37°C over-night (16-18h).

3.1.1.5 Screening by Minipreps and Maxipreps

4-6 colonies were selected from over-night grown plate. Individual colonies were grown in 4 ml L-broth with ampicillin (100µg/ml) over-night at +37°C with shaking. Plasmid DNAs were purified by “Plasmid DNA Purification Kit-NucleoSpin® Plasmid (MACHEREY-NAGEL)” and screened by enzymatic digestion or by colony PCR. Positive samples were sent to “DNA sequencing and Genomics laboratory-Institute of Biotechnology” for sequencing. Correct clones were processed as maxi preps using “NucleoBond® Xtra Maxi kit (MACHEREY-NAGEL)”.

3.1.2 Cell culture and transfection of reporter plasmids

CHO (Chinese hamster ovary; ATCC) cells were grown overnight on a 10 cm cell culture plate with complete medium (88% DMEM, 10% FBS, 1% Penicillin-Streptomycin and 1% L-glutamine) at +37°C in a CO₂ incubator. The following day, cells were washed with 2-3 ml PBS and 2 ml of 0.25% Trypsin-EDTA (1X) was added into the plate followed by incubation for 2 min at +37°C. 10ml complete medium was added, mixed and divided onto several 10 cm plates with different dilutions e.g. 1:3, 1:4, 1:5, etc., and let them grow over-night to achieve an approximately 95 % confluency. The cells were then washed, trypsinized and divided onto 12-well plates and grown over-night. Total volume per well was 1ml containing medium without antibiotics (89% DMEM, 10% FBS and 1% L-glutamine). When the wells were 95% confluent, transfection was performed with

reporter plasmid (100 ng) as instructed in standard protocol of Lipofectamine 2000 (*Invitrogen*). After 24h of transfection, cells of each well were mixed with 450µl of Trizol reagent (Sigma) and collected into an eppendorf tube.

3.1.3 RNA isolation

A modified Trizol protocol was used to isolate RNA from CHO cell. Cells with 450µl Trizol were mixed with 80µl chloroform and centrifuged (13,000 rpm, +9°C) for 15 min followed by extraction of supernatant with 1 volume 50:48:2 phenol-chloroform-isoamylalcohol pH 5. Then 200 µl of 0.4 M NaCl was added and extraction was repeated. Finally ethanol precipitation step was performed. RNA pellets were dissolved in 20µl MQ.

3.1.4 RT-PCR

RNA was then treated with RQ DNase (Promega) (x µl of RNA (1µg), 1µl of 10xRQ1 buffer, 1µl of 1U/µl RQ1 DNase and 8-x µl MQ) and incubated at +37°C for 30 min. 1 µl RQ1 stop solution was added and incubated at +65°C for 20 min to inactivate DNase enzyme before the cDNA synthesis with RevertAid reverse transcriptase.

Two reaction mixtures were made for the +RT and –RT reactions. Initially primers were annealed by combining 2.5µl DNase treated RNA, 0.5 µl of 100ng/µl random primer and 0.5 µl of 10mM dNTP, and incubated at +65°C for 5 min and then placed on ice. Subsequently the following reagents were added for the +RT and –RT reactions:

	<u>For +RT</u>	<u>For -RT</u>
5x RT buffer	2 μ l	2 μ l
40 U/ μ l Ribolock (Fermentas)	0.25 μ l	0.25 μ l
RevertAid Premium reverse transcriptase (Fermentas)	0.5 μ l	---
MQ	3.75 μ l	4.25 μ l
<hr/>		
Amount to be added per reaction =	6.5 μ l	6.5 μ l

After mixing all reagents, samples were incubated at 25 °C for 10 min, 55 °C for 30 min and 85 °C for 10 min and finally placed on ice.

A multiplex PCR reaction was performed with three primers: h65K-80, h65K-79 and h65K-145. Primers were specific to long and short isoforms.

The PCR reaction mixture per 20 μ l volume was:

5x GC buffer (Thermo-scientific)	4 μ l
10 mM dNTPs	0.4 μ l
h65K-80	0.2 μ l
h65K-79	0.1 μ l
h65K-145	0.1 μ l
Phusion DNA polymerase (Thermo-scientific)	0.2 μ l
MQ	14 μ l
<hr/>	
Total =	19 μ l per reaction.

1 μ l cDNA was then added with 19 μ l reaction mixture and run by the following PCR program:

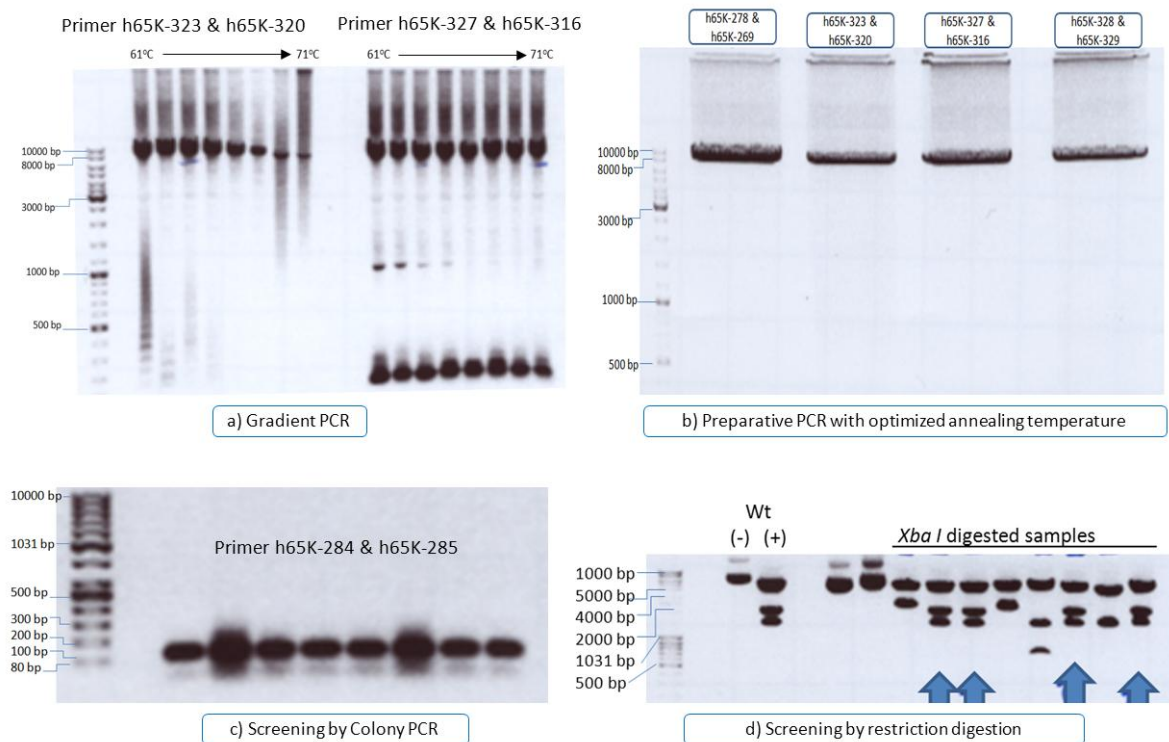
- 1: +98 °C 30 sec
- 2: +98 °C 10 s
- 3: +61 °C 20 s
- 4: +72 °C 5 sec
- 5: Go to 2, 31 times
- 6: +72 °C 7 min
- 7: +10 °C Forever
- 8: End

After completing PCR, reactions were loaded into a 3% Metaphor agarose gel containing ethidium bromide (0.5ug/ml) and run at 80 V for 2.5 h (14cm gel) or 120V for 2.5h (24cm gel). The image of the gel was taken by the LAS3000 imager (Fuji) and analyzed by Aida image analyzing software (Raytest) to compare the relative ratios of short and long isoforms.

4 RESULTS

4.1 Construction of reporter plasmids

Each of the experimental constructs used in this study are derived from wild type plasmid pGL4.13-65KUTR that was modified by site-directed mutagenesis. The new constructs were made by whole plasmid amplification which allowed a simple method for insertion/deletion mutagenesis. At the beginning of the cloning, a gradient PCR was performed to check the optimum temperature of a primer set to anneal with the template DNA. The temperature range of gradient PCR was 61°C to 71°C. The identified temperature was then used in a subsequent round to provide preparative amounts of PCR product which was gel-isolated. The techniques used to make a correct clone are schematically presented at Figure 11.



InUSSE-3nt

```

...TTTGCTTTAAACAGGTAATTTTAGTCATACTTAATTTGATGATTAATATGGTAGAATTAAGCATTTTAAATGTTAATGTTTGTACA
TTGTTCAAGAAATAAGTAGAAATATATTCTTTGTTTTTATTTAAATTTTGTCTCTGTAACTAAAAGAACACGAAATTTGGTCACAATTACTGGTGTAA
CTGCCAAATATGGGTAATAAGGGAAAATTTGTTAATTTAGTCTTCTGAGATGGCTTGAATATTTGAATTTTGTGTACGTCTATACTGGGTAGTCACAAGT
CTTATAAACACTTTAGAGGAAAGATGGATTCAGTCTGTATTTTAAACATCATTTTAAATCTGGGCTGAAAAATAAGAAAAAATTAACCTGCATTCTGCT
GTTCTTTTAGAAGCATTCTGCGTAAATACTGCTGAATACTGTATGCAAAAGTGTATCCTTCTTGTGTGTATCCTTGTGGGGcCAGgGtTTATG...

```

```

WT: ATTCTGCTGTTCTTCTTAGAAGCATTCTGCGTAAATACTGCTGAATACTGTATGCAAAAGTGTATCCTTCTTGTGTGTATCCTTGTGGGGCAGTGT

```

e) Confirmation of correct clones by sequencing

Figure 11: **Steps used in the construction of different mutated plasmids.** Representative gels of each set are shown. (a) Gradient PCR: Annealing temperature was optimized for each set of primers. The optimum annealing temperature of the PCR primers were checked by gradient PCR where temperature ranged from 61°C to 71°C.

(b) Preparative PCR was done using optimized annealing temperature, gel purified (1% agarose gels) and followed by downstream cloning steps. The primer sets used in the representative reactions PCR are shown on the top of each lane. The product size after successful PCR was 8102 bp ± mutation (bp).

(c) Following ligation and electroporation, the success of cloning was checked by colony PCR by using a set of primer: forward h65K-284 and reverse h65K-285 surrounding the site of insertions/deletions. The success rate was very high. The product size after PCR was 100 bp and interestingly in this gel, all the samples represented right sized product.

(d) The success of cloning and the intactness of plasmids were checked by restriction digestion. Digestion with restriction enzyme *Xba*I produced three fragment sizes: 4641 bp, 2006 bp ± mutation (bp) and 1455 bp for correct clones. Both digested and undigested wt plasmid (indicated) were used as controls. Blue arrows indicate the intact and successful clones.

(e) Finally correct clones were confirmed by sequencing at the DNA sequencing and Genomic laboratory, Institute of Biotechnology, University of Helsinki. A 3nt insertion within USSE region is shown as an example. Red colored highlighted sequences indicate 5'ss, the elements within USSE region of both mutated (Upper sequence) and wt (Lower sequence) constructs. The three nucleotides colored in green indicate insertion between two 5'ss within USSE of the mutated construct.

4.2 USSE activation highly dependent on the distance between 3'ss and USSE

To study the effects of mutations, all the newly constructed plasmids were transfected into CHO cells. 24h after the transfection, total RNA was isolated and equal amounts of RNA were used for cDNA production prior to RT-PCR. A set of three primers were used to investigate the mRNA isoform distribution. The analysis strategy is illustrated in Figure 12, which shows the primer location in the two isoforms. Primer set comprises one forward primer h65K-80 and two reverse primers h65K-79 and h65K-145. The short isoform was detected using an upstream h65K-80 primer that was specific to the short mRNA isoform together with the h65K-145 downstream primer that is common to both isoforms. Long isoform was amplified using isoform-specific primer h65K-79 that was used together with h65K-80.

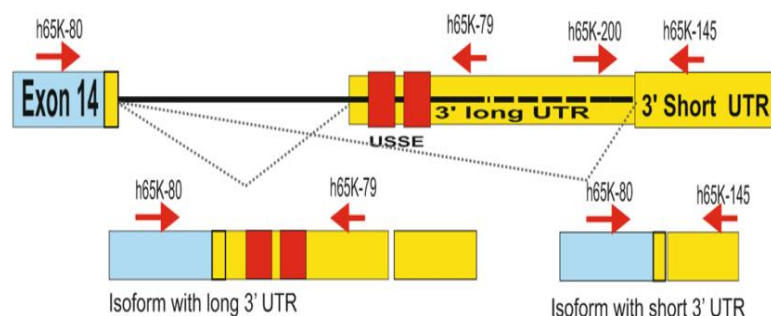


Figure 12: Primer binding sites for RT-PCR analysis. Red arrows represent primer binding sites on target sequence and direction of polymerization. Solid blue, red and yellow boxes indicate the last coding exon of 65K gene, USSE, and 3' UTR, respectively. Alternative splicing patterns are shown by dotted lines.

RT-PCR analysis allows amplification of the *in vivo* splicing pattern. I found that production of the two isoforms in each mutant varied based on the sizes of deletion (Figure 13). In wild type construct, the ratio of long to short isoform is approximately 45:55. However, there is transfection to transfection variation at the absolute ratio. Therefore a relative value, in which ratio from the wt construct has been set as 1.0, is used in the subsequent analyses.

The result in Figure 13 indicates a general trend where the long isoform level drops as the USSE is moved closer to the 3'ss, with no alternative splicing observed with the largest deletion. This results provides functional explanation for evolutionary conservation of distance between the USSE and 3'ss (Verbeeren et al., 2010); it is critical for the USSE mediated alternative splicing. Here the distance constraints on exon definition potential of the USSE.

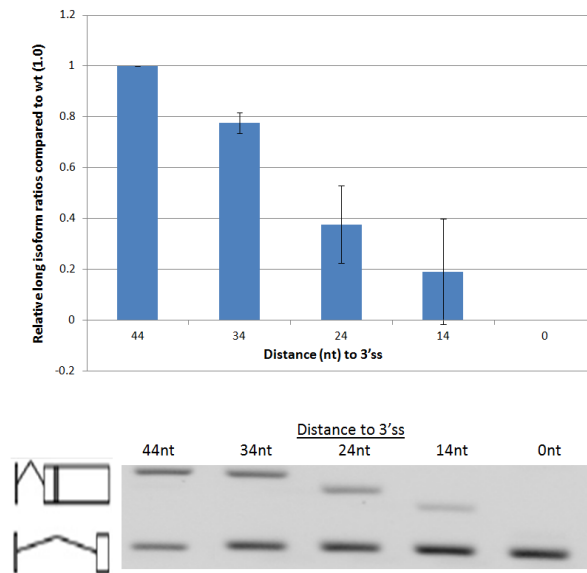


Figure 13: USSE mediated alternative long isoform formation depends on distance to 3'ss. Upper panel: Bar chart showing relative isoform levels from the RT-PCR analysis. Band intensities were determined from digital images and were then normalized using wild type construct that was set as 1.0. Results represent averages taken from at least three independent replications. Error bars indicates standard deviations from the mean. Lower Panel: A gel image of a respective RT-PCR analysis. The identities (Long Vs short isoform) are indicated on left.

4.3 Interaction between two U11/12 di-snRNP is essential for splicing activation

My next question was to ask if the distance between two 5'ss sequence within USSE affects alternative splicing. The distance between two U12-type 5'ss within USSE which in wt situation is 6nt, and was reduced to 0, 3, 4, or 5 nt or alternatively the distance was increased to 9, 12, 15, 18, or 30 nt. RT-PCR analysis showed that alternative splicing event was almost absent in the case of 0nt distance between two 5'ss (Fig. 14). In contrast, alternatively spliced long isoforms were observed for the 3, 4, and 5 nt distances, albeit with reduced levels compared to wt. The optimum result was obtained for the 9 nt distance between two 5'ss, which was 3nt larger than the wt distance. Constructs having larger distances between these two motifs resulted lower long isoform formation compared to wt. The overall scenario of this mutational analysis revealed that optimal distance is approximately 6-9 nt, but surprisingly long distances, up to 30nt, support alternative splicing, albeit weakly.

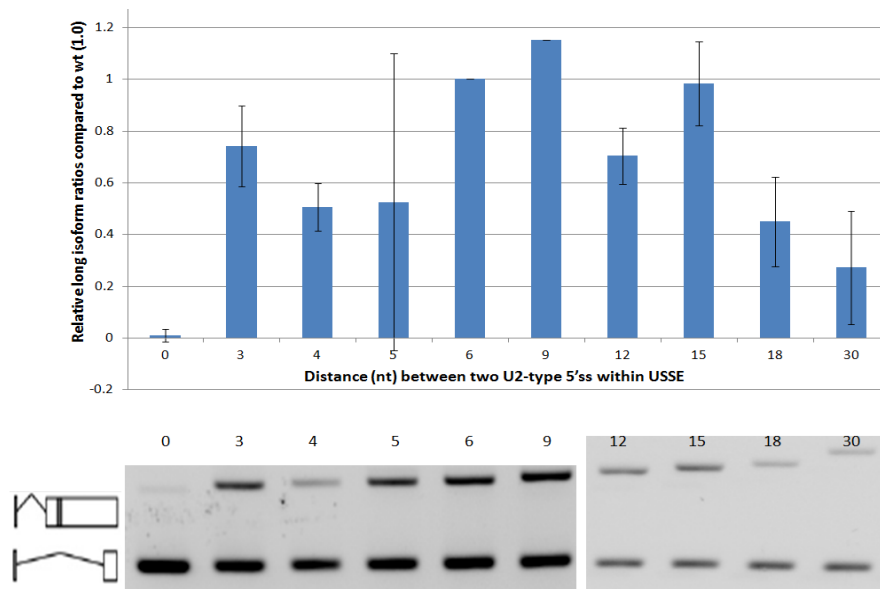


Figure 14: Alternative splicing events after manipulation of distance of 5'ss within the USSE. Upper panel shows a bar chart of relative band intensities (Normalized to 6 nt value (1.0)) and the lower panel shows a representative gel analysis of the RT-PCR reactions. The identities (Long Vs short isoforms) are indicated on left.

4.4 Manipulation of element distances with RNA hairpin constructs

The above results with distance manipulation have a complication that sequence context (e.g. specific nucleotides in the insertions and deletions) can also at least potentially affect the activity. To alleviate this issue I introduced a hairpin structure between the two U12 5'ss of the USSE element. The hairpin structure, though the actual distance between the two U12 5'ss would be 30 nucleotides, would bring the physical distance between the two U12 5'ss to about 6 nucleotides, which is similar as in the wt case. Therefore, I expected that the introduction of the hairpin would restore the splicing activity back to wt levels. However, this was not what I observed, as the hairpin showed similar splicing activity as the construct with a 30 nt distance between the two U12 5'ss.

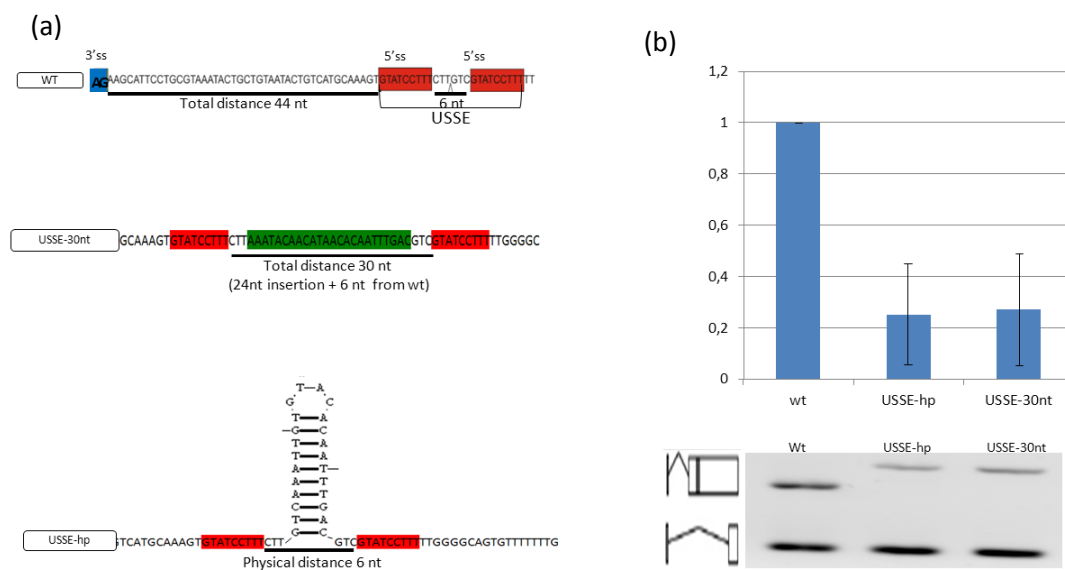


Figure 15: Rescue constructs to regain the lost functionality of USSE. (a) Schematic representation of different constructs used as control and rescue constructs. 3'ss, USSE and inserted sequences were indicated by solid boxes colored in blue, red and green, respectively. (b) Upper panel shows a bar chart of relative band intensities (Normalized to wt value (1.0)) and the lower panel shows a representative gel analysis of the RT-PCR reactions. The identities (Long Vs short isoforms) are indicated on left. USSE-hp construct had hairpin structure to rescue functionality for the USSE region.

5 DISCUSSIONS

In this thesis work I have investigated the distance constraints affecting the alternative splicing of U11/12-65K 3'UTR. The results described here extended the current understanding of the USSE mediated activation of alternative splicing. USSE is present in genes encoding the U11-48K and U11/12-65K proteins, and both are required for U12-type intron recognition and U11/12 di-snRNP stability (Benecke et al., 2005; Turunen et al., 2008; Verbeeren et al., 2010). As USSE elements are conserved from humans to plants (e.g. almost 1 billion years of evolution), this element is among the most conserved regulatory systems described in the literature (Verbeeren et al., 2010). Thus these findings will help to better understand the role of USSE and in particular the distance requirements between the different elements, as a part of the regulatory circuits controlling the level of these proteins in the cell.

5.1 Deletion might affect Enhancer/Repressor sites to 3'ss

In my study of USSE mediated alternative splicing, I was investigating whether the distance between the USSE element and the 3'ss activated by USSE is optimal or whether it can be shortened. Earlier evolutionary data from 21 mammalian, bird and lizard species showed very little variation, e.g. 44 ± 1 nt, in the 3'ss-USSE distance. In my experiments, I progressively shortened this distance to zero nucleotides, which resulted in concomitant decrease, and eventual loss in the use of upstream 3'ss (Fig. 13).

This result suggests that the observed evolutionary conservation in 3'ss-USSE distance in mammals (Verbeeren et al., 2010) displays functional significance in regulation of alternative splicing of U11/U12-65K mRNA isoform distribution. More importantly, the same distance constraints are also observed in more distantly related organism, such as in fish, which also displays a similar 3'ss-USSE spacing (Verbeeren et al., 2010). The region reside between the 3'ss and USSE possess several putative binding sites for splicing regulators such as SR proteins and this region is also conserved at sequence level (Verbeeren et al., 2010). Thus an alternative hypothesis is that instead of distance,

the trans-acting protein factors that could participate in the regulation of USSE mediated alternative splicing as observed with U11-48K gene (Turunen et al., 2013). However, comparison of sequence conservation between mammals and fishes suggests that this is not the case. There is high conservation between mammals and fishes at the sequence level, even though the distance is relatively well-conserved (Verbeeren et al., 2010). Thus my results together with earlier data indicate that the distance to 3'ss are critical for optimal level of USSE mediated alternative splicing.

5.2 Interaction between two U11/12 di-snRNP is essential for splicing activation

In addition to 3'ss-USSE distance, also the distance between the individual 5'ss elements within USSE is conserved. This suggests that this distance is also evolutionary optimized. My results are consistent with this hypothesis as the changes of distance between these two motifs, insertions or deletions, results in reduction of alternatively spliced long isoform formation (Fig. 14).

The other central question in this research was whether USSE element was recognized simultaneously by two U11 snRNPs, or whether it rather serves as high-affinity platform for a single U11 snRNP to bind more efficiently (See section 2 above). My results support the simultaneous binding hypothesis. The most important result to support this conclusion is the lack of long isoform with constructs that have the entire spacer between the individual 5'ss deleted (See Fig. 14). The molecular explanation is that if there are two U11 snRNPs particles binding on the two 5'ss, they will sterically clash with each other when the spacer have been removed, and are thus unable to support AS event.

What is then the purpose of duplicating 5'ss sequence within USSE? A further support for this model comes from earlier work by Frilander and Steitz (2001), who reported that binding of 5'ss by U11 snRNP is established by U12/BPS interaction. In essence, the

recognition of 5'ss and BPS is cooperative and this cooperativity is necessary for intron recognition (Frilander and Steitz, 1999). Both the U11-48K gene and the U11/U12-65K 3'UTR lack U12-type BPS sequence (Verbeeren et al., 2010) and therefore the duplication of 5'ss within USSE can be seen as a compensatory feature that allows U11 snRNP to bind in the absence of U12/BPS interaction. In this model the U11/U12 di-snRNPs bound to individual 5'ss sequence are expected to interact and stabilize each other's binding. Such interactions have not yet been observed and remain to be proven experimentally.

On the other hand, in this model the expansion of the distance between the individual 5'ss is expected to work because RNA molecules are flexible and can form loops that bring two motifs close enough for the two di-snRNPs to interact and to stabilize each other's bindings, thus leading to activation of AS. A surprising result is that alternative splicing is activated even when the distance between the two 5'ss is increased to 30 nt, indicating that RNA is able to loop out relatively large spacer elements. This result has significant impact on studies searching for USSE-like elements from various genomes. Earlier bioinformatic studies (Verbeeren et al., 2010) had used up to 16 nt spacer elements, but it is now obvious that even larger spacer sequences must be used in computational models when predicting functional binding sites.

5.3 Hairpin construct didn't fully rescue the functionality of USSE

I tried to manipulate the distance between the two 5'ss within USSE by RNA hairpin structures. These were compared to a corresponding 24 nt insertion construct. We expected the hairpin structure to restore the splicing activity back to wt levels, however this was not observed. The actual cause can be due to presence of the hairpin structure which interferes with either communication of di-snRNPs within USSE, or the communication between 3'ss and USSE. Also, the hairpin structure might interfere with cDNA and PCR methods.

5.4 Construction of mutated plasmids did not always succeed

Theoretically, it was easy to design a plasmid construct with desired mutation using the whole plasmid amplification, but in practice I faced a number of difficulties. First, although there were intense bands during PCR and abundant colony formation in electroporation, no correct clones were identified, in particularly those containing hairpin structure. Typical outcome in failed experiments was partial or total deletions of insertion sites. I had planned much more extensive hairpin construct study that was shown in the results, but that was abandoned due to cloning difficulties. It took a lot of efforts but remain unsuccessful after trying about 70 minipreps per sample out of approximately 700 colonies screened by PCR. The primers used in both cases contained hairpin structure, which might hinder the efficiency of DNA polymerase during PCR (Singh et al., 2000) or later in ligation reaction by making inaccessible of ligation sites for DNA ligase. The other explanation is that bacterial recombination mechanism may rearrange or delete such regions as they contain inverted repeats.

6 CONCLUDING REMARKS

As the components involved in intron recognition and their interaction with the intronic sequences are different in two distinct spliceosome, and later steps of the two spliceosomes formation are similar, this strongly indicates that the splice sites recognition and components involved in it play a crucial role in the regulation of splicing. The U2 and U12-dependent spliceosomes form a control circuit which uses a splicing enhancer element (USSE) to regulate the level of key proteins of U12- dependent spliceosomes. U11 snRNP or U11/U12 di-snRNP recognizes USSE and activates 3'ss of U2- type intron promoting alternative splicing. In this study, I showed that manipulation of distance within USSE and in 3'ss-USSE elements affect alternative splicing of upstream U2-type 3'ss. These results suggest that conservation distance constraints of these elements are crucial for the functionality of USSE mediated alternative splicing and thus important for the gene expression that houses them.

7 ACKNOWLEDGEMENTS

Words would not suffice to express my deepest gratitude to my supervisor Docent Mikko Frilander for his guidance and overall support. I could not be more thankful upon Janne Turunen and Jens Verbeeren for their unprecedented endeavor behind this work. I am highly indebted to Bhupendra Verma and Jouni Kvist for their help every now and then. I also have each and every one from *team Frilander* to thank for their valuable suggestions and help. Special thanks to Marja-Leena Peltonen (Lilli) to make the work environment full of entertainment! Inspiration from my family has been invaluable, which I remember with gratefulness.

8 LISTS OF REFERENCES

Agafonov, D.E., Deckert, J., Wolf, E., Odenwalder, P., Bessonov, S., Will, C.L., Urlaub, H., and Luhrmann, R. (2011). Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol Cell Biol* 31, 2667-2682.

Behrens, S.E., and Luhrmann, R. (1991). Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. *Genes Dev* 5, 1439-1452.

Benecke, H., Lührmann, R., and Will, C.L. (2005). The U11/U12 snRNP 65K protein acts as a molecular bridge, binding the U12 snRNA and U11-59K protein. *EMBO J* 24, 3057-3069.

Berglund, J.A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* 89, 781-787.

Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Luhrmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. *Nature* 452, 846-850.

Black, D.L., and Pinto, A.L. (1989). U5 small nuclear ribonucleoprotein: RNA structure analysis and ATP-dependent interaction with U4/U6. *Mol Cell Biol* 9, 3350-3359.

Boelens, W.C., Scherly, D., Mattaj, I.W., and van Venrooij, W.J. (1990). Analysis of the in vitro binding of A and B' snRNP proteins to U1 and U2 snRNAs using mutated A and B' proteins. *Mol Biol Rep* 14, 159.

Bonen, L., and Vogel, J. (2001). The ins and outs of group II introns. *Trends Genet* 17, 322-331.

Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, H., and Luhrmann, R. (1984). Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. *EMBO J* 3, 1357-1363.

Buratti, E., Dhir, A., Lewandowska, M.A., and Baralle, F.E. (2007a). RNA structure is a key regulatory element in pathological ATM and CFTR pseudoexon inclusion events. *Nucleic Acids Res* 35, 4369-4383.

Buratti, E., Stuani, C., De Prato, G., and Baralle, F.E. (2007b). SR protein-mediated inhibition of CFTR exon 9 inclusion: molecular characterization of the intronic splicing silencer. *Nucleic Acids Res* 35, 4359-4368.

Burge, C.B., Tuschl, T., and Sharp, P.A. (1999). Splicing of precursors to mRNAs by the spliceosome. In *The RNA world*, R.F. Gesteland, T.R. Cech, and J.F. Atkins, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 525-560.

Busch, A., and Hertel, K.J. (2012). Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdisciplinary Reviews: RNA* 3, 1-12.

Corvelo, A., Hallegger, M., Smith, C.W.J., and Eyras, E. (2010). Genome-Wide Association between Branch Point Properties and Alternative Splicing. *PLoS Comput Biol* 6, e1001016.

- Csuros, M., Rogozin, I.B., and Koonin, E.V. (2011). A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. *PLoS Comput Biol* 7, e1002150.
- Darnell, J.E. (1978). Implications of RNA-RNA splicing in evolution of eukaryotic cells, Vol 202.
- Dhir, A., Buratti, E., van Santen, M.A., Luhrmann, R., and Baralle, F.E. (2010). The intronic splicing code: multiple factors involved in ATM pseudoexon definition. *EMBO J* 29, 749-760.
- Dietrich, R.C., Peris, M.J., Seyboldt, A.S., and Padgett, R.A. (2001). Role of the 3' Splice Site in U12-Dependent Intron Splicing. *Mol Cell Biol* 21, 1942-1952.
- Doma, M.K., and Parker, R. (2007). RNA Quality Control in Eukaryotes. *Cell* 131, 660-668.
- Doolittle, W.F. (1978). Genes in pieces: were they ever together? *Nature* 272, 581-582.
- Egecioglu, D.E., and Chanfreau, G. (2011). Proofreading and spellchecking: a two-tier strategy for pre-mRNA splicing quality control. *RNA* 17, 383-389.
- Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2009). The Evolutionarily Conserved Core Design of the Catalytic Activation Step of the Yeast Spliceosome. *Mol Cell* 36, 593-608.
- Frilander, M.J., and Steitz, J.A. (1999). Initial recognition of U12-dependent introns requires both U11/5' splice-site and U12/branchpoint interactions. *Genes Dev* 13, 851-863.
- Frilander, M.J., and Steitz, J.A. (2001). Dynamic Exchanges of RNA Interactions Leading to Catalytic Core Formation in the U12-Dependent Spliceosome. *Mol Cell* 7, 217-226.
- Fu, X.-D. (2004). Towards a Splicing Code. *Cell* 119, 736-738.
- Gao, K., Masuda, A., Matsuura, T., and Ohno, K. (2008). Human branch point consensus sequence is γ UnAy. *Nucleic Acids Res* 36, 2257-2267.
- Hall, S.L., and Padgett, R.A. (1996). Requirement of U12 snRNA for in vivo splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science* 271, 1716-1718.
- Han, J., Ding, J.H., Byeon, C.W., Kim, J.H., Hertel, K.J., Jeong, S., and Fu, X.D. (2011). SR proteins induce alternative exon skipping through their activities on the flanking constitutive exons. *Mol Cell Biol* 31, 793-802.
- Han, S.P., Tang, Y.H., and Smith, R. (2010). Functional diversity of the hnRNPs: past, present and perspectives. *Biochem J* 430, 379-392.
- Haugen, P., Simon, D.M., and Bhattacharya, D. (2005). The natural history of group I introns. *Trends Genet* 21, 111-119.
- Heinrichs, V., Bach, M., Winkelmann, G., and Lührmann, R. (1990). U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site. *Science* 247, 69-72.
- Incorvaia, R., and Padgett, R.A. (1998). Base pairing with U6atac snRNA is required for 5' splice site activation of U12-dependent introns in vivo. *RNA* 4, 709-718.

- Isken, O., and Maquat, L.E. (2007). Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev* 21, 1833-3856.
- Kandels-Lewis, S., and Séraphin, B. (1993). Involvement of U6 snRNA in 5' splice site selection. *Science* 262, 2035-2039.
- Keren, H., Lev-Maor, G., and Ast, G. (2010). Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet* 11, 345-355.
- Khanna, A., and Stamm, S. (2010). Regulation of alternative splicing by short non-coding nuclear RNAs. *RNA Biol* 7, 480-485.
- Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Lührmann, R., Garcia-Blanco, M.A., and Manley, J.L. (1994). Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* 368, 119-124.
- Kramer, A., Gruter, P., Groning, K., and Kastner, B. (1999). Combined biochemical and electron microscopic analyses reveal the architecture of the mammalian U2 snRNP. *J Cell Biol* 145, 1355-1368.
- Lamond, A.I., Konarska, M.M., Grabowski, P.J., and Sharp, P.A. (1988). Spliceosome assembly involves the binding and release of U4 small nuclear ribonucleoprotein. *Proc Natl Acad Sci U S A* 85, 411-415.
- Lanctot, C., Cheutin, T., Cremer, M., Cavalli, G., and Cremer, T. (2007). Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat Rev Genet* 8, 104-115.
- Levine, A., and Durbin, R. (2001). A computational scan for U12-dependent introns in the human genome sequence. *Nucleic Acids Res* 29, 4006-4013.
- Lewis, B.P., Green, R.E., and Brenner, S.E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci USA* 100, 189-192.
- Lim, L.P., and Burge, C.B. (2001). A computational analysis of sequence features involved in recognition of short introns. *Proc Natl Acad Sci USA* 98, 11193-11198.
- Logsdon Jr, J.M. (1998). The recent origins of spliceosomal introns revisited. *Current Opinion in Genetics & Development* 8, 637-648.
- Madhani, H.D., and 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.
- Matlin, A.J., Clark, F., and Smith, C.W.J. (2005). Understanding alternative splicing: Towards a cellular code. *Nat Rev Mol Cell Biol* 6, 386-398.
- McNally, M.T. (2008). RNA processing control in avian retroviruses. *Frontiers in bioscience : a journal and virtual library* 13, 3869-3883.
- Michael, D., and Manyuan, L. (1999). Intron—exon structures of eukaryotic model organisms. *Nucleic Acids Res* 27, 3219-3228.

- Nakai, K., and Sakamoto, H. (1994). Construction of a novel database containing aberrant splicing mutations of mammalian genes. *Gene* 141, 171-177.
- Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457-463.
- Nottrott, S., Urlaub, H., and Luhrmann, R. (2002). Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J* 21, 5527-5538.
- Park, J.W., Parisky, K., Celotto, A.M., Reenan, R.A., and Graveley, B.R. (2004). Identification of alternative splicing regulators by RNA interference in *Drosophila*. *Proc Natl Acad Sci USA* 101, 15974-15979.
- Patel, A.A., and Steitz, J.A. (2003). Splicing double: Insights from the second spliceosome. *Nat Rev Mol Cell Biol* 4, 960-970.
- Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K.W., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5Å resolution. *Nature* 458, 475-480.
- Pozzoli, U., Sironi, M., Cagliani, R., Comi, G.P., Bardoni, A., and Bresolin, N. (2002). Comparative Analysis of the Human Dystrophin and Utrophin Gene Structures. *Genetics* 160, 793-798.
- Ruskin, B., Zamore, P.D., and Green, M.R. (1988). A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* 52, 207-219.
- Rutz, B., and Seraphin, B. (1999). Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. *RNA* 5, 819-831.
- Saltzman, A.L., Pan, Q., and Blencowe, B.J. (2011). Regulation of alternative splicing by the core spliceosomal machinery. *Genes Dev* 25, 373-384.
- Schneider, C., Will, C.L., Makarova, O.V., Makarov, E.M., and Lührmann, R. (2002). Human U4/U6.U5 and U4atac/U6atac.U5 Tri-snRNPs Exhibit Similar Protein Compositions. *Mol Cell Biol* 22, 3219-3229.
- Schwartz, S., and Ast, G. (2010). Chromatin density and splicing destiny: on the cross-talk between chromatin structure and splicing. *EMBO J* 29, 1629-1636.
- Shen, H., and Green, M.R. (2004). A Pathway of Sequential Arginine-Serine-Rich Domain-Splicing Signal Interactions during Mammalian Spliceosome Assembly. *Mol Cell* 16, 363-373.
- Shepard, P., and Hertel, K. (2009). The SR protein family. *Genome Biol* 10, 242.
- Sheth, N., Roca, X., Hastings, M.L., Roeder, T., Krainer, A.R., and Sachidanandam, R. (2006). Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res* 34, 3955-3967.
- Singh, J., and Padgett, R.A. (2009). Rates of in situ transcription and splicing in large human genes. *Nat Struct Mol Biol* 16, 1128-1133.
- Singh, V.K., Govindarajan, R., Naik, S., and Kumar, A. (2000). The effect of hairpin structure on PCR amplification efficiency. (*Molecular Biology Today*), pp. 67-69.

- Smith, D.J., Query, C.C., and Konarska, M.M. (2008). "Nought may endure but mutability": spliceosome dynamics and the regulation of splicing. *Mol Cell* 30, 657-666.
- Snehal Bhikhu, P., and Michel, B. (2008). The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res* 36, 6482-6493.
- Spadaccini, R., Reidt, U., Dybkov, O., Will, C., Frank, R., Stier, G., Corsini, L., Wahl, M.C., Lührmann, R., and Sattler, M. (2006). Biochemical and NMR analyses of an SF3b155-p14-U2AF-RNA interaction network involved in branch point definition during pre-mRNA splicing. *RNA* 12, 410-425.
- Tarn, W.-Y., and Steitz, J.A. (1996a). Highly divergent U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science* 273, 1824-1832.
- Tarn, W.-Y., and Steitz, J.A. (1996b). A novel spliceosome containing U11, U12 and U5 snRNPs excises a minor class (AT-AC) intron *in vitro*. *Cell* 84, 801-811.
- Turunen, J. (2012). The function and regulation of the U11-48K protein in U12-dependent splicing. University of Helsinki.
- Turunen, J.J., Niemelä, E.H., Verma, B., and Frilander, M.J. (2013). The significant other: splicing by the minor spliceosome. *Wiley Interdisciplinary Reviews: RNA* 4, 61-76.
- Turunen, J.J., Will, C.L., Grote, M., Lührmann, R., and Frilander, M.J. (2008). The U11-48K Protein Contacts the 5' Splice Site of U12-Type Introns and the U11-59K Protein. *Mol Cell Biol* 28, 3548-3560.
- Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* 136, 701-718.
- Valcárcel, J., Gaur, R.K., Singh, R., and Green, M.R. (1996). Interaction of U2AF65 RS region with pre-mRNA of branch point and promotion base pairing with U2 snRNA. *Science* 273, 1706-1709.
- Valen, E., and Sandelin, A. (2011). Genomic and chromatin signals underlying transcription start-site selection. *Trends Genet* 27, 475-485.
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470-476.
- Wang, G.-S., and Cooper, T.A. (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 8, 749-761.
- Wang, Z., and Burge, C.B. (2008). Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA* 14, 802-813.
- Warf, M.B., and Berglund, J.A. (2010). Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem Sci* 35, 169-178.
- Wassarman, K.M., and Steitz, J.A. (1992). The low-abundance U11 and U12 small nuclear ribonucleoproteins (snRNPs) interact to form a two-snRNP complex. *Mol Cell Biol* 12, 1276-1285.

- Verbeeren, J., Niemelä, E.H., Turunen, J.J., Will, C.L., Ravantti, J.J., Lührmann, R., and Frilander, M.J. (2010). An ancient mechanism for splicing control: U11 snRNP as an activator of alternative splicing. *Mol Cell* 37, 821-833
- Will, C.L., and Luhrmann, R. (2011). Spliceosome structure and function. *Cold Spring Harbor perspectives in biology* 3.
- Will, C.L., Schneider, C., Hossbach, M., Urlaub, H., Rauhut, R., Elbashir, S., Tuschl, T., and Lührmann, R. (2004). The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. *RNA* 10, 929-941.
- Will, C.L., Schneider, C., Reed, R., and Lührmann, R. (1999). Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science* 284, 2003-2005.
- Wu, S., Romfo, C.M., Nilsen, T.W., and Green, M.R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* 402, 832-835.
- Yu, Y.-T., and Steitz, J.A. (1997). Site-specific crosslinking of mammalian U11 and U6atac to the 5' splice site of an AT-AC intron. *Proc Natl Acad Sci USA* 94, 6030-6035.
- Zamore, P.D., and Green, M.R. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci U S A* 86, 9243-9247.
- Zhu, J., Mayeda, A., and Krainer, A.R. (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., and Weiner, A.M. (1986). A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* 46, 827-835.
- Zorio, D.A.R., and Blumenthal, T. (1999). Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*. *Nature* 402, 835-838.