The function and regulation of the U11-48K protein in U12-dependent splicing

Janne Turunen

Institute of Biotechnology

and

Division of Biochemistry and Biotechnology
Department of Biosciences
Faculty of Biological and Environmental Sciences

and

Helsinki Graduate Program in Biotechnology and Molecular Biology

University of Helsinki

ACADEMIC DISSERTATION

To be presented for public criticism, with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki, in auditorium 2 in Information Centre Korona (Viikinkaari 11, Helsinki) on January 20th 2012, at 12 noon.

HELSINKI 2012

Supervisor Docent Mikko Frilander
Institute of Biotechnology
University of Helsinki
Helsinki, Finland

Reviewers Docent Minna Poranen
Department of Biosciences and Institute of Biotechnology
University of Helsinki
Helsinki, Finland

Professor Anders Virtanen
Department of Medical Biochemistry and Microbiology
Uppsala University
Uppsala, Sweden

Opponent Professor Mark McNally
Department of Microbiology and Molecular Genetics
Medical College of Wisconsin
Milwaukee, Wisconsin

Custos Professor Kari Keinänen
Department of Biosciences
University of Helsinki
Helsinki, Finland

ISBN 978-952-10-7394-6 (paperback)
ISBN 978-952-10-7395-3 (PDF)
ISSN 1799-7372

Cover layout by Anita Tienhaara

Helsinki University Print, Helsinki 2012

"Every model we make tells us how our mind works,
as much as it tells us about the universe"

Robert Anton Wilson

Table of contents

List of original publications

The thesis is based on the following articles, which are referred to in the text by their Roman numerals:

I **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. *Molecular and Cellular Biology* 28: 3548-3560. (Reproduced here with permission from American Society for Microbiology)

II **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. *Molecular Cell* 37: 821-833. (Reproduced here with permission from Elsevier)

III **Turunen, J.J., Verma, B., Frilander, M.J.** HnRNP F/H, U1 snRNP and U11 snRNP co-operate to regulate the stability of the *U11-48K* pre-mRNA. *Unpublished manuscript.*

* Equal contribution

The author's contribution to each publication:

I JJT participated in planning the experiments and performed the *in vitro* characterization of the protein-RNA interactions as well as the analysis of the effect of U11-48K RNAi knockdown on *in vivo* splicing, and wrote the article together with the other authors.

II JJT planned and performed the *in vitro* characterization of the factors involved in recognizing the regulatory elements in both the *U11-48K* and *U11/U12-65K* transcripts, and wrote the article together with the other authors.

III JJT planned and performed the majority of the experiments, and wrote the manuscript.

Abbreviations

3'ss	3' splice site
3' UTR	3' untranslated region
5'ss	5' splice site
aa	amino acid(s)
AS-NMD	alternative splicing coupled to NMD
bp	base pair(s)
BP	branch point
BPS	branch point sequence
CERES	composite exonic regulatory elements of splicing
CTD	C-terminal domain (of RNAPII)
EJC	exon junction complex
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
hnRNP	heterogeneous nuclear RNP
ISE	intronic splicing enhancer
ISS	intronic splicing silencer
LECA	last eukaryotic common ancestor
mRNA	messenger RNA
mRNP	messenger RNP
NMD	nonsense-mediated decay
NMR	nuclear magnetic resonance
nt	nucleotide(s)
PPT	polypyrimidine tract
pre-mRNA	precursor mRNA
PTC	premature termination codon
qRRM	quasi-RRM
R	purine
RNAi	RNA interference
RNAPII	RNA polymerase II
RNP	ribonucleoprotein
RRM	RNA recognition motif
RS	arginine-serine-rich (domain)
RSV	Rous sarcoma virus
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
snRNP	small nuclear RNP
SR	serine-arginine-rich (protein)
SRE	splicing regulatory element
TEC	transcription elongation complex
TREX	transcription-export complex
USSE	U11 snRNP-binding splicing enhancer
Y	pyrimidine
ZnF	zinc finger

Summary

The removal of noncoding sequences, or introns, from the eukaryotic messenger RNA precursors is catalyzed by a ribonucleoprotein complex known as the spliceosome. In most eukaryotes, two distinct classes of introns exist, each removed by a specific type of spliceosome. The major, U2-type introns account for over 99 % of all introns, and are almost ubiquitous. The minor, U12-type introns are found in most but not all eukaryotes, and reside in conserved locations in a specific set of genes. Due to their slow excision rates, the U12-type introns are expected to be involved in the regulation of the genes containing them by inhibiting the maturation of the messenger RNAs. However, little information is currently available on how the activity of the U12-dependent spliceosome itself is regulated.

The levels of many known splicing factors are regulated through unproductive alternative splicing events, which lead to inclusion of premature STOP codons, targeting the transcripts for destruction by the nonsense-mediated decay pathway. These alternative splice sites are typically found in highly conserved sequence elements, which also contain binding sites for factors regulating the activation of the splice sites. Often, the activation is achieved by binding of products of the gene in question, resulting in negative feedback loops.

In this study, I show that U11-48K, a protein factor specific to the minor spliceosome, specifically recognizes the U12-type 5' splice site sequence, and is essential for proper function of the minor spliceosome. Furthermore, the expression of U11-48K is regulated through a feedback mechanism, which functions through conserved sequence elements that activate alternative splicing and nonsense-mediated decay. This mechanism is conserved from plants to animals, highlighting both the importance and early origin of this mechanism in regulating splicing factors. I also show that the feedback regulation of U11-48K is counteracted by a component of the major spliceosome, the U1 small nuclear ribonucleoprotein particle, as well as members of the hnRNP F/H protein family. These results thus suggest that the feedback mechanism is finely tuned by multiple factors to achieve precise control of the activity of the U12-dependent spliceosome.

1 Review of the literature

1.1 Introduction

Splicing of messenger RNA precursors (pre-mRNAs) is a ubiquitous eukaryotic process that affects virtually all protein-coding transcripts in humans, and thus has wide implications for normal gene expression, as well as for disease (reviewed by Wang and Cooper, 2007). However, the discovery that eukaryotic genes are discontinuous (Berget *et al.*, 1977; Chow *et al.*, 1977) was initially met with amazement and even incredulity. Understandably, the reason for maintaining such apparently useless DNA was initially difficult to fathom. Even more puzzling was the question of how the intervening regions, or introns, as they became known, were properly recognized and removed from the expressed regions, or exons. There seemed to be very little information in the introns themselves, and no pattern similar to the well-known protein-encoding genetic code was to be found. Yet, the precise removal of introns was crucial to maintaining the correct reading frame in the resulting messenger-RNA (mRNA), and thus for the expression of the correct proteins. The observation that some of the small nuclear RNAs (snRNAs) bore sequences compatible to the splice site sequences (Lerner *et al.*, 1980; Rogers and Wall, 1980), paved way for the realization that the splicing machinery, the spliceosome, is itself a ribonucleoprotein (RNP) complex. While highly degenerate, the splice sites are nonetheless recognized by snRNA-containing small nuclear RNPs (snRNPs) through base-pairing interactions with the snRNAs (Mount *et al.*, 1983; Black *et al.*, 1985).

The situation became even more complex with the discovery of alternative splicing. This process was originally found in immunoglobulin transcripts, the alternative isoforms of which produce the membrane-bound and soluble antibodies (Alt *et al.*, 1980; Early *et al.*, 1980; Rogers *et al.*, 1980). While this observation offered an insight into how splicing could be useful, i.e. by producing mRNAs for alternative protein isoforms from one gene (reviewed by Breitbart *et al.*, 1987), it underscored the inherent puzzle in splicing: How are the proper splice sites selected? It became clear that the splice site sequences themselves were not the only determinants of splicing, as the recognition of splice site sequences was found to be affected by exonic sequences (Somasekhar and Mertz, 1985; Reed and Maniatis, 1986). Over time, it was recognized that splicing regulatory elements play a crucial role in splice site activation in both constitutive and alternative splicing. Alternative splicing was also found to be a more common event than originally expected, particularly after completion of the human genome sequencing project (Lander *et al.*, 2001), when it became obvious that the complexity of the human proteome heavily depends on alternative splicing to produce the protein isoforms that are many-fold more numerous than the genes encoding them (reviewed by Nilsen and Graveley, 2010).

Almost as surprising as the discovery of splicing itself was the discovery of a rare class of introns (Jackson, 1991; Hall and Padgett, 1994) that were found to be excised by a separate spliceosome with its own set of snRNPs analogous to, but distinct from those of the canonical major spliceosome (reviewed by Patel and Steitz, 2003). Why should this be? What possible

use could an organism have for two machineries performing essentially identical tasks? The low frequency, inefficient excision and extremely conserved splice site sequences of these minor, or U12-type introns, as well as their absence from many organisms, initially suggested that they may be molecular fossils, slowly being purged from the genomes of eukaryotes. However, their retention in homologous locations over hundreds of millions of years of evolution suggests that they have important functions. The importance of U12-dependent splicing is likely linked to its regulatory role in controlling the expression of the specific set of genes containing U12-type introns and, ultimately, the downstream effects of those genes. The emerging view is that the various splicing processes are linked to each other as well as other steps of the gene expression pathway, with cross-regulation apparent at every level.

1.2 Overview of eukaryotic pre-mRNA processing

The initial step in the expression of proteins in all cellular organisms is transcription, during which the nucleotide sequence of DNA is copied into a complementary sequence in RNA. In eukaryotes, chromosomes are tightly packed within the nucleus of the cell, and transcriptionally active and inactive parts of the chromosomes are sequestered in different domains of the nucleus (reviewed by Lanctôt *et al.*, 2007). To become accessible to RNA polymerase, the DNA must be unpacked from the chromatin complexes. A number of enzymes regulate the affinity of histones for DNA within the chromatin by modifying a specific set of histone amino acid residues. Once the DNA has become accessible, a set of transcription initiation factors is assembled within the promoter region of a gene, and they in turn recruit the RNA polymerase (reviewed by Valen and Sandelin, 2011). This process is regulated by enhancer and silencer elements, which can be located thousands of nucleotides away from the promoter, and exert their effect by binding regulatory factors which either promote or suppress transcription initiation. As transcription proceeds, the nascent transcript is processed in various ways, including capping at its 5' end, splicing to remove the non-coding intronic sequences and finally polyadenylation at its 3' end (see the following chapter). All of these steps are generally required before the mature mRNA is transported into the cytoplasm (see chapter 1.7.2), where it can function as a template for protein synthesis. Among these steps, splicing has emerged as a highly versatile process, which not only removes introns with high fidelity, but can also give rise to alternative mRNA isoforms, and thus proteins, as well as regulate gene expression by producing non-viable mRNA variants.

1.2.1 Co-transcriptional nature of pre-mRNA processing

In eukaryotes, pre-mRNAs are transcribed by RNA polymerase II (RNAPII). Together with the chromatin template, the growing transcript and other associated factors, the polymerase forms a transcription elongation complex (TEC; reviewed by Perales and Bentley, 2009). The TEC is a multifunctional assembly, which co-ordinates a large number of interdependent RNA as well as chromatin processing events (reviewed by Phatnani and Greenleaf, 2006). One of the central elements in this complex is the C-terminal domain (CTD) of the large subunit of RNAPII, composed of a large number (52 in mammals) of YSPTSPS heptapeptide repeats. The phosphorylation state of these repeats controls the initiation and elongation phases of transcription, and also co-ordinates the assembly of pre-mRNA processing factors (reviewed

by Phatnani and Greenleaf, 2006). For example, as the transcript is initiated, the CTD is phosphorylated on the Ser5 residues of its heptapeptide repeats, allowing the loading and activation of the capping enzyme (Ho and Shuman, 1999). Towards the end of the transcript the CTD becomes phosphorylated predominantly at Ser2, and recruits and activates 3' end-processing factors (reviewed by Perales and Bentley, 2009). The interactions are also reciprocal, with pre-mRNA processing factors activating or inhibiting transcription elongation (Perales and Bentley, 2009).

Whether splicing factors associate physically with RNAPII is not equally well established, although a number of interactions between them have been observed. Splicing factors are found in complexes with RNAPII (Das *et al.*, 2007; Sapra *et al.*, 2009), and can also activate recruitment of transcription factors to the promoter and enhance elongation (Damgaard *et al.*, 2008; Lin *et al.*, 2008). On the other hand, splicing factors appear to be recruited only to intron-containing, but not intronless transcripts, suggesting that RNAPII does not directly promote their binding or at least is not the only factor doing so (Listerman *et al.*, 2006; Moore *et al.*, 2006). Despite the lack of information on the physical relationship between splicing factors and RNAPII, the co-transcriptional nature of splicing is well documented. Looped RNAs attached to chromatin have been observed by electron microscopy (Beyer and Osheim, 1991), spliced mRNAs are associated with isolated chromatin (Bauren and Wieslander, 1994; Pandya-Jones and Black, 2009) and spliced mRNAs can be detected on their chromosomal gene loci by RNA *in situ* hybridization (Zhang *et al.*, 1994). Also, synthetic pre-mRNAs are spliced less efficiently than co-transcriptionally spliced transcripts both *in vitro* and *in vivo*, and CTD phosphorylation is required for efficient co-transcriptional splicing (Bird *et al.*, 2004; Das *et al.*, 2006). Finally, with long genes, the removal of most introns is completed before the pre-mRNA is fully transcribed (Singh and Padgett, 2009), verifying that splicing takes place co-transcriptionally. Splicing is also linked to other steps of pre-mRNA processing, including capping (Görnemann *et al.*, 2005) and export, as well as quality control mechanisms in the nucleus, such that only correctly and rapidly processed mRNAs are packages into mature messenger RNPs (mRNPs) that are exported into the cytoplasm for translation (see chapter 1.7).

1.2.2 Variation and origins of spliceosomal introns

1.2.2.1 Prevalence of introns

Introns are one of the features typical of eukaryotic genomes, but different eukaryotes vary widely with respect to the length and density of their introns. Some unicellular eukaryotes contain only a few introns in total in their genomes (Nixon *et al.*, 2002; Vanacova *et al.*, 2005), and at least one cryptophyte has lost its introns altogether (Lane *et al.*, 2007). In contrast, multicellular eukaryotes with larger genomes typically have a larger number of introns, which are also longer and in some cases constitute a much larger fragment of their genomes than the protein-coding exonic sequences (Deutsch and Long, 1999). In humans, the lengths of the introns vary from less than 100 basepairs (bp) to several hundred kilobases (kb), with the mean length of 3300 bp (Lander *et al.*, 2001) and the average number of introns per

gene (intron density) of 6.9 (Csuros *et al.*, 2011). The most extreme case of intron density and coverage in the human genome is the gene which codes for the muscle-specific protein dystrophin. The total length of the gene is 2.5 million bp, and it contains 79 exons, which constitute only 1% of the total length of the gene (Pozzoli *et al.*, 2002). On the other hand, in baker's yeast *Saccharomyces cerevisae*, a commonly used model organism in biology, only a subset of genes harbor introns, and typically no more than one intron per gene, with an average intron density of 0.1 (Csuros *et al.*, 2011). Introns in the yeast genome are also considerably shorter than in mammals, with lengths varying between 50–1000 bp.

1.2.2.2 Introns and evolution

Such a wide range of varying intron sizes and densities suggest that the importance of introns and splicing has varied in different evolutionary lineages. How, then, did this variation arise during evolution? A number of theories accounting for the origins of introns have been proposed, and they can be roughly divided into two models known as "introns-late" and "introns-early" (Darnell, 1978; Doolittle, 1978). According to the introns-late model, introns were introduced (possibly through horizontal gene transfer) to a eukaryotic ancestral protein-coding DNA genome, or independently into several eukaryotic lineages. The introns-early model postulates that introns arose before or together with protein coding sequences or else were inserted into protein-coding RNA genomes before DNA was adopted as the hereditary material. In a more stringent version of the introns-early model, known as the "introns-first" model (Jeffares *et al.*, 1998; Penny *et al.*, 2009), introns are seen as remnants of the original RNA world (Gilbert, 1986), representing genes for ribozymes and other functional RNAs, which later became interspersed with spacers that evolved into protein coding sequences of today. Both the introns-early and introns-first models necessitate that spliceosomal introns were already extant before the last eukaryotic common ancestor (LECA), and may also have been present in the ancestors of modern archaea and/or bacteria.

Yeast is one of the most common model organisms used in splicing studies. As mentioned, the yeast genome is very intron-poor. Yeast introns typically have very precisely defined splice site sequences and are not alternatively spliced. Such a minimal, stringent splicing system has traditionally been considered to be the original primitive mode of splicing in LECA, thus lending support for the introns-late theory (Logsdon, 1998). However, recent comparisons across eukaryotes have clearly shown that the common eukaryotic ancestor instead had introns and spliceosomes very similar to most modern eukaryotes, i.e. high intron density, weakly conserved splice sites, both U2 and U12-type introns and complex spliceosomes (reviewed by Koonin, 2009; Roy and Irimia, 2009). It appears that intron loss is a common phenomenon in many eukaryotic lineages, while massive expansion in intron numbers has only taken place a few times during evolution, most notably in the common ancestor of Metazoa (Csuros *et al.*, 2011). In very intron-sparse lineages, such as yeast, intron loss is correlated with reduction in the genome size and with the emergence of highly regular splice sites, suggesting that all these characteristics are a product of high selection pressure on these fast-reproducing organisms (reviewed by Koonin, 2009; Penny *et al.*, 2009). Thus, introns appear to have arisen well before the LECA, but whether their origins are in the RNP or RNA world remains unresolved.

1.2.2.3 Relationship of spliceosomal introns to self-splicing introns

In addition to spliceosomal introns, many types of organisms carry self-splicing introns, which are divided into group I and group II introns based on their different catalytic mechanisms (reviewed by Bonen and Vogel, 2001; Haugen *et al.*, 2005; Fedorova and Zingler, 2007). They are rarely found in archaea, but are common in bacteria and in eukaryotic mitochondria and chloroplasts, and are also present in some nuclear genes encoding ribosomal RNAs. Self-splicing introns of the same type show a high degree of structural conservation, which is required for their autocatalytic excision. They also often carry open reading frames encoding proteins that enhance the splicing activity and also enable the insertion of the excised intron into DNA, allowing the introns to function as retrotransposons. Excision of group II introns and spliceosomal introns proceeds through a similar catalytic mechanism, which has prompted the hypothesis that group II introns may be evolutionarily related to both the spliceosome and spliceosomal introns (Hickey, 1992; Stoltzfus, 1999). A number of functional similarities support this hypothesis: Both intron types are excised through a similar two-step transesterification reaction, and the structures of spliceosomal snRNAs are similar to the catalytically active domains of group II introns and they even have interchangeable activities in certain cases (reviewed by Valadkhan, 2010). Furthermore, constructs based on spliceosomal snRNAs can perform splicing reactions *in vitro* in the absence of proteins, similar to group II introns (Valadkhan *et al.*, 2009). However, there currently is no clear evidence to show whether the spliceosome is derived from group II introns or shares a common ancestor with it, and it has been noted that group II introns and spliceosomal introns may merely have gained similar characteristic through parallel evolution due to performing similar chemical functions (Weiner, 1993).

1.3 Minor and major introns

1.3.1 Intron structure

The excision of self-splicing introns relies on specific structures within the introns themselves, and the introns are therefore highly conserved sequence elements. In contrast, spliceosomal introns are removed by a *trans*-acting spliceosome, and have therefore diverged significantly, with little sequence conservation between different introns. However, they do contain three regions of moderate conservation. The scissile phosphodiester bond at the 5' splice site (5'ss) is located in a conserved element, known as the 5'ss sequence, which generally includes the first six nucleotides of the intron, as well as up to three nucleotides in the exon (Figure 1). Similarly, the 3' splice site (3'ss) is preceded by a conserved intronic region. Finally, the catalytically active branch point (BP) adenosine resides within the branch point sequence (BPS) located at a variable distance upstream of the 3'ss. These are the sequences that are recognized by the core spliceosomal components during spliceosome assembly and are also involved in the catalytic steps required for intron removal and exon ligation.

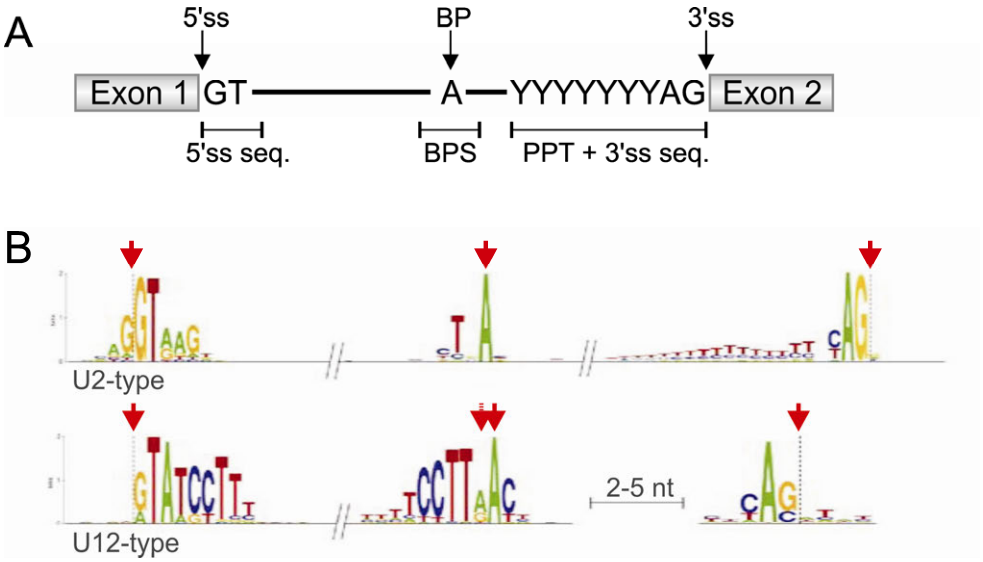


Figure 1. Intron structure and consensus sequences.
A) Schematic depiction of (U2-type) intron structure in the coding strand of a typical eukaryotic gene. Exons are depicted as boxes, and the intron as a line, with key sequence elements highlighted by letters. The locations of the scissile phosphodiester bonds at the splice sites (5'ss and 3'ss) are indicated by arrows, as well as the location of the branch point (BP) adenosine. The approximate locations of the conserved sequence elements are indicated below the intron, including the 5'ss sequence, the branch point sequence (BPS), and the polypyrimide tract (PPT) together with the 3'ss sequence.
B) Human U2 and U12-type intron consensus sequences The consensus sequences of the 5'ss, BPS, and 3'ss (with PPT included) of U2 and U12-type introns are depicted schematically, with the height of the letters proportional to the frequency of the corresponding base at that position. The 5'ss, BP and 3'ss are indicated by arrows. Adapted with modifications from Roy and Irimia (2009), with permission from Elsevier.

1.3.2 Characteristics of U2-type and U12-type introns

There are two parallel sets of spliceosomal introns in eukaryotes, the U2-type introns and U12-type introns (reviewed by Patel and Steitz, 2003). U2-type introns account for the majority of introns in any given eukaryote (over 99%), while U12-type introns are only found in a handful of genes, and the intron types are thus also known as the major and minor introns, respectively. The length of both intron types varies greatly, and they have similar average length (Levine and Durbin, 2001). However, there is a large subset of small U2-type introns of less than 100 bp, while such short U12-type introns are relatively rare (Levine and Durbin, 2001). This has led to the assumption that intron recognition may differ for the two intron types, with U12-type splicing relying more strictly on the so called exon definition interactions (see chapter 1.6.1).

The organization of both types of introns is similar, but they differ significantly in the precise sequence composition of their splice sites and BPS, such that they can be easily differentiated from one another based on these sequences (Figure 1B). The overall consensus for the U2-type 5'ss sequence is AG/GTAAGT, with the slash denoting the 5'ss (Sheth *et al.*, 2006). The level of sequence conservation varies greatly between different organisms, and organisms with low intron density typically also have very conserved 5'ss sequences (chapter 1.2.2.2). For

example, 84 % of the introns in *Cryptosporidium parvum* contain the precise consensus sequence, while in human introns the corresponding figure is only 14 % (Irimia *et al.*, 2007). The U2-type 3'ss is marked by a much shorter signal, CAG, at the very end of the intron. The U2-type BPS is generally located 20–40 bp upstream of the 3'ss, and shows considerable sequence variation, similar to the 5'ss sequence (Gao *et al.*, 2008; Corvelo *et al.*, 2010). Between the BPS and the 3'ss lies the polypyrimidine tract (PPT), the length of which varies significantly and is generally indicative of the strength of the 3'ss, and can also compensate for weak BPSs (Corvelo *et al.*, 2010).

In contrast to the U2-type splicing signals, the U12-type introns show a much higher degree of conservation in theirs. The U12-type 5'ss sequence almost invariably conforms to the consensus /RTATCCTTT (in which R is a purine) in all organisms, although occasional deviations occur (Figure 1B; Sheth *et al.*, 2006; Lin *et al.*, 2010). The BPS is similarly highly conserved, and is much more pyrimidine rich than that of U2-type introns. Conversely, no conserved PPT has been observed in U12-type introns. The 3'ss is marked only by the terminal nucleotides, which typically are YAG (where Y is a pyrimidine) in introns starting with GT, and YAC in introns with AT at the 5'ss, although many different dinucleotide combinations can be recognized as a 3'ss (Dietrich *et al.*, 2001; Levine and Durbin, 2001; Hastings *et al.*, 2005). However, the distance of a U12-type BPS from the 3'ss is typically much more restricted (10–20 bp) than for U2-introns, and it has been shown that the distance is a crucial factor for recognition of the 3'ss (Dietrich *et al.*, 2001; Levine and Durbin, 2001; Zhu and Brendel, 2003; Dietrich *et al.*, 2005).

1.3.3 Conversion of intron types, and evolution of U12-type introns

The existence of U12-type introns and/or U12-dependent splicing factors in almost all of the major eukaryotic lineages suggests that U12-type introns are of ancient origin, similarly to U2-type introns (Russell *et al.*, 2006; Bartschat and Samuelsson, 2010). However, they are entirely absent from many eukaryotes whose relatives nonetheless have them, suggesting repeated loss of U12-dependent splicing during evolution. The loss seems to have been most prevalent among diverse eukaryotic microbes, such as algae, unicellular fungi and amoebozoa, but some animals, including some but not all nematodes, have also lost their U12-type introns (Russell *et al.*, 2006; Bartschat and Samuelsson, 2010). Interestingly, the genome of the nematode *Caenorhabditis elegans* still contains vestigial U12-type introns that are now spliced by the U2-dependent spliceosome (Sheth *et al.*, 2006), supporting the idea that its U12-type introns have been converted to U2-type relatively recently.

As discussed in chapter 1.2.2.2, the loss of U12-type introns probably reflects the high selection pressures affecting intron density in general during the evolution of these organisms. How, then, do U12-type introns become lost? As the vestigial U12-type introns of *C. elegans* suggest, one likely pathway is the step-wise conversion of U12-type introns into U2-type introns. Natural examples as well as experimental approaches have also shown that simple point mutations in the 5'ss of a U12-type intron, especially in the GT-AG subtype, suffice to turn it into a U2-type 5'ss (Dietrich *et al.*, 1997; Burge *et al.*, 1998; Lin *et al.*, 2010), as U2-

type introns have much more degenerate splicing signals (Figure 1B). Furthermore, changes in the terminal dinucleotides of U12-type introns occur readily in nature (Sheth *et al.*, 2006; Lin *et al.*, 2010), thus enabling other U12-subtypes to become GT-AG introns. Sometimes the conversion can also occur by activation of a cryptic U2-type 5'ss near the original U12-type 5'ss (Bartschat and Samuelsson, 2010). It has been also noted that the pyrimidine-rich BPS of U12-type introns could efficiently function as a PPT for the newly created U2-type intron (Burge *et al.*, 1998). Given that the chance of creating the highly stringent U12-type 5'ss from a U2-type 5'ss is low, conversion in the reverse direction is a highly unlikely event. In fact, only one instance of a novel U12-type intron has been observed (Lin *et al.*, 2010). Thus, without natural selection acting in favour of maintaining U12-type introns, they are likely to be converted into U2-type introns during evolution.

Such a unidirectional conversion from U12-type introns to U2-type has raised the hypothesis that U12-type introns would have been much more prevalent during the early history of life, and that they might in fact be ancestral to U2-type introns. A further indication of this could be the fact that U12-type introns appear to resemble group II introns more closely (for discussion, see Roy and Irimia, 2009), assuming that spliceosomal introns are indeed derived from group II introns (see chapter 1.2.2.3). However, an analysis of amino acid distributions at intron insertion sites retained between human and *Arabidopsis thaliana* genomes suggested that most ancestral introns were already of the U2-type (Basu *et al.*, 2008b). Therefore, the precise origins of both U2 and U12-type introns remain unresolved, although it is clear that both have been present in the early eukaryotic ancestors.

1.3.4 Significance of U12-type introns

The apparent ease of loss of U12-type introns in some lineages and their low prevalence (0.35% of all introns in humans) might initially suggest that they are merely relics of the past that are fading away. However, the fact that they are retained in various highly divergent eukaryotes suggests they perform an essential function. It has been noted that U12-type introns are excised more slowly than U2-type introns in a wide variety of organisms from animals to plants (Patel *et al.*, 2002; Lewandowska *et al.*, 2004; Pessa *et al.*, 2006), giving rise to the notion that their removal may be a rate-limiting step in gene regulation. U12-type introns generally occur only once in a gene (Burge *et al.*, 1998; Levine and Durbin, 2001), suggesting that for such a rate-limiting function, only one U12-type introns is required. Furthermore, certain U12-type introns have retained their position in specific orthologous genes in distantly related organisms (Burge *et al.*, 1998; Basu *et al.*, 2008a). Together, these data suggest that the activity of U12-dependent splicing factors may have a conserved role in regulating a specific set of genes in response to external stimuli or in specific tissues or developmental stages. Indeed, U12-type intron-containing genes show a bias for being differentially expressed in bone marrow CD34 positive cells and B lymphoblast (Yeo *et al.*, 2007). More importantly, while U12-type introns account for a very small fraction of *Drosophila melanogaster* introns (Schneider *et al.*, 2004; Sheth *et al.*, 2006), U12-dependent splicing is nonetheless essential for *D. melanogaster* development (Otake *et al.*, 2002). U12-

type introns are not randomly scattered in genomes, but are most often found in genes that can be loosely categorized as information processing genes (Burge *et al.*, 1998; Yeo *et al.*, 2007). These genes have functions in DNA replication and repair, transcription, pre-mRNA processing, translation and signal transduction, suggesting that the essential function of U12-dependent spliceosome is linked to the regulation of these genes. However, alterations in U12-dependent splicing can also have wide effects on multiple downstream processes, as shown by the disruption metabolic functions in flies with defective U12-dependent splicing (Pessa *et al.*, 2010).

1.4 Spliceosome components

1.4.1 Components of the U2-dependent spliceosome

The core components of the major spliceosome include five small nuclear ribonucleoprotein particles (snRNPs), each of which contains, and is named after, one of the small nuclear RNAs (snRNAs): U1, U2, U4, U5 and U6 (Figure 2). Each snRNA, except U6, contains a binding site for Sm proteins, which together form a ring-like core structure for the snRNPs, while U6 has a similar structure composed of Sm-like (Lsm) proteins (reviewed by Patel and Bellini, 2008). The snRNPs also contain a varying number of specific protein factors (reviewed by Valadkhan and Jaladat, 2010): U1-specific proteins U1A and U1-70K bind directly to U1 snRNA, while the U1C protein associates through protein-protein interactions (Query *et al.*, 1989; Scherly *et al.*, 1989; Nelissen *et al.*, 1991; Pomeranz Krummel *et al.*, 2009). In addition to the Sm proteins, the core of the U2 snRNP contains the proteins U2A' and U2B", and is complemented by two multiprotein complexes, SF3a and SF3b (Boelens *et al.*, 1990; Brosi *et al.*, 1993; Krämer *et al.*, 1999). U4 and U6 snRNAs associate with one another through base-pairs formed between highly conserved sequence elements (Bringmann *et al.*, 1984; Hashimoto and Steitz, 1984), and together with specific proteins form the U4/U6 di-snRNP (Nottrott *et al.*, 2002). U5 snRNP associates with U4/U6 through protein-protein interactions (Cheng and Abelson, 1987; Konarska and Sharp, 1988; Black and Pinto, 1989), forming the U4/U6-U5 tri-snRNP together with additional tri-snRNP-specific protein factors (Behrens and Lührmann, 1991; Liu *et al.*, 2006 and references therein). U5 is the largest of the spliceosomal snRNPs (Black and Pinto, 1989), and contains several large proteins that make up a significant part of the catalytically active spliceosome, most notably the multifunctional protein Prp8 (reviewed by Valadkhan and Jaladat, 2010).

In addition to the snRNPs, a number of non-snRNP proteins are also found among the core components of the U2-dependent spliceosome, including splicing factor 1 (SF1) and U2 auxiliary factor (U2AF), which have important functions for recognizing the BPS, PPT and 3'ss (See chapter 1.5.2). To facilitate the conformational changes occurring during assembly, the spliceosome employs a number of peptidyl-prolyl *cis/trans* isomerases, helicases, kinases and other protein and RNA-modifying enzymes, some of which are snRNP components while others are non-snRNP proteins (reviewed by Staley and Guthrie, 1998; Smith *et al.*, 2008). The spliceosome also interacts with a large number of splicing factors that can suppress or

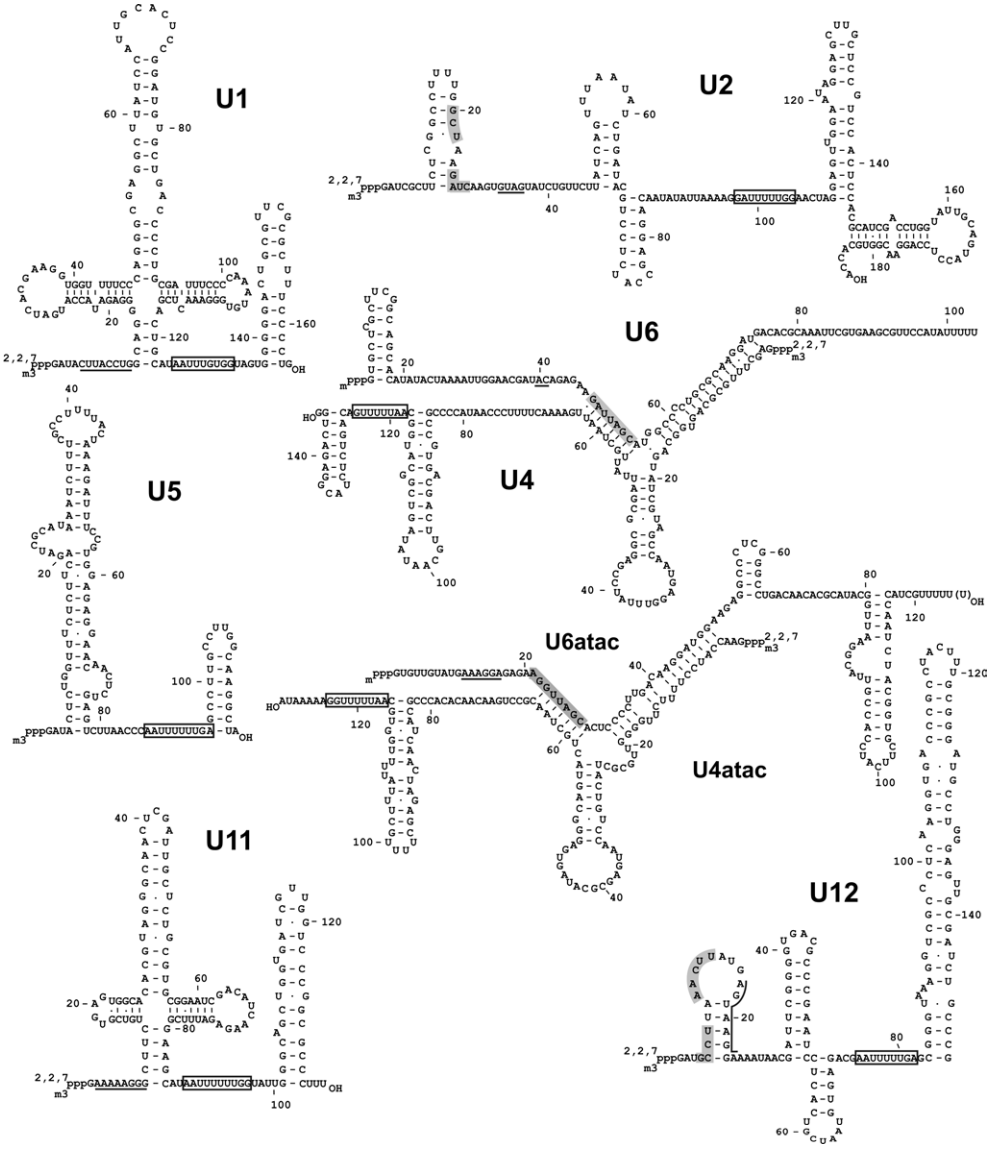


Figure 2. Secondary structures of spliceosomal snRNAs.
The predicted secondary structures of the human spliceosomal snRNAs are depicted here schematically. The binding sites for Sm proteins are boxed, the sequences interacting with the 5'ss or BPS are marked with a black line, and sequences involved in U2/U6 or U12/U6atac basepairing are highlighted in gray. Adapted with modifications from Pessa (2010). Structures are as published by Yu et al. (1999) for U1, U2, U4, U5 and U6, Tarn and Steitz (1997) for U11, Sikand and Shukla (2011) for U12, and Padgett and Shukla (2002) for U4atac and U6atac.

enhance splicing at a given site, such as heterogeneous nuclear ribonucleoproteins (hnRNPs) and serine-arginine-rich (SR) proteins (see chapter 1.6.4). Many of the protein factors associate with the spliceosome during specific stages of assembly, and can also be specific to certain introns or tissues only, making the spliceosome a very dynamic and variable molecular machine. It is therefore difficult to pinpoint the exact composition of the spliceosome, and estimates for the number of protein components vary between 150 and 300 (reviewed by Jurica and Moore, 2003; Valadkhan and Jaladat, 2010).

1.4.2 Components of the U12-dependent spliceosome

The U12-dependent spliceosome is similarly composed of five snRNPs, U11, U12, U4atac, U5 and U6atac. U5 is therefore a common component of both spliceosomes, and U11, U12, U4atac and U6atac are structural and functional counterparts of U1, U2, U4 and U6, respectively (Hall and Padgett, 1996; Tarn and Steitz, 1996a, b; Kolossova and Padgett, 1997; Yu and Steitz, 1997; Incorvaia and Padgett, 1998). The U12-type snRNAs do not resemble their U2-type counterparts in sequence: U11 and U12 are entirely different from U1 and U2 (Montzka and Steitz, 1988), and U4atac and U6atac share limited (ca. 40%) sequence similarity with U4 and U6, respectively (Tarn and Steitz, 1996a). However, the secondary structures of the analogous snRNAs are highly similar, highlighting the similarity of their functions (Figure 2).

The low abundance of minor introns is also reflected in the numbers of the minor snRNPs, which are approximately 100 times less abundant in human than the major snRNPs (Montzka and Steitz, 1988; Tarn and Steitz, 1996a). This has hindered the comprehensive analysis of their protein components, and consequently the composition of the minor snRNPs and spliceosomal complexes are known in less detail than their major counterparts (see e.g. Schneider *et al.*, 2002). However, most of the protein components of the two spliceosomes seem to be shared. The minor snRNAs are complexed with Sm or Lsm proteins similarly to the major snRNAs (Montzka and Steitz, 1988; Tarn and Steitz, 1996a; Will *et al.*, 1999; Schneider *et al.*, 2002). The minor U4atac/U6atac-U5 tri-snRNP also appears to contain most if not all of the protein components of the U4/U6-U5 tri-snRNP (Luo *et al.*, 1999; Schneider *et al.*, 2002). In contrast to U1 and U2, however, U11 and U12 snRNPs interact with one another already before spliceosome assembly, and are mainly present as preformed U11/U12 di-snRNPs (Wassarman and Steitz, 1992; see chapter 1.5.4). The protein composition of the U11/U12 di-snRNP is also different from that of the U1 and U2 snRNPs: U11/U12 entirely lacks U1-specific proteins and the U2-specific SF3a protein complex (Will *et al.*, 1999). However, it does contain the SF3b complex and three other proteins present in the major spliceosome (YB-1, hPrp43p and Urp), and additionally seven proteins (20K, 25K, 31K, 35K, 48K, 59K and 65K) specific to the U12-dependent spliceosome (Will *et al.*, 1999; 2001; 2004). Four of these (25K, 35K, 48K, 59K) are also found in free U11 snRNPs, while three appear to associate only with the di-snRNP (Will *et al.*, 2004). Although their functions remain undeciphered for the most part, their association with the U11/U12 di-snRNP suggests they are involved with 5'ss and BPS recognition and in interactions bridging these two sites (see chapter 1.5.5).

1.5 Spliceosome assembly and catalysis

1.5.1 Splicing catalysis

Both spliceosomes catalyze the same, two-step transesterification reaction that results in the formation of an excised intron lariat and ligation of the two exons (Figure 3; reviewed by Will and Lührmann, 2011). In the first step, the 2' hydroxyl group of the branch point adenosine performs a nucleophilic attack on the phosphate group at the 5'ss, resulting in formation of the lariat structure and a free 3' OH group at the end of the upstream exon. This 3' OH then attacks the phosphate at the 3'ss, resulting in the ligation of the two exons, and release of the intron lariat. Due to its nature as a transesterification reaction, splicing in itself is energetically neutral. However, ATP is consumed in many energy-requiring steps during the assembly of the spliceosome to ensure the specificity and unidirectionality of the reaction.

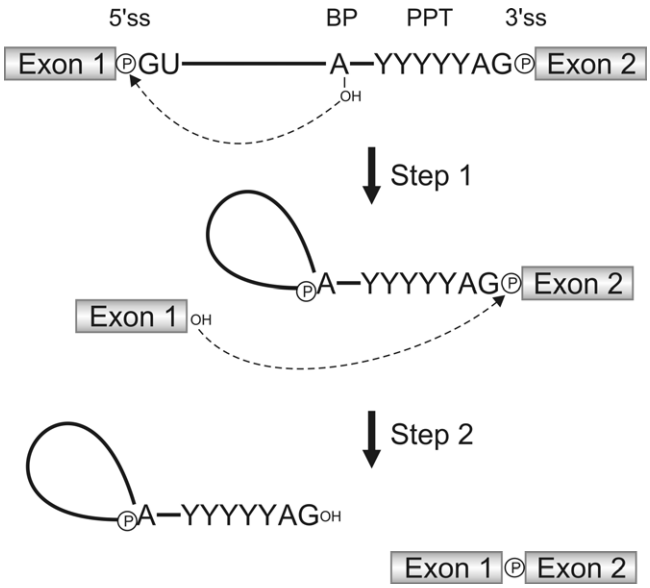


Figure 3. Catalytic mechanism of splicing.
Splicing occurs through a two-step transesterification reaction, depicted here schematically. Intron stucture is depicted as in Figure 1A, and additionally the reactive phosphates (P) and hydroxyl groups (OH) are indicated. In the first step of splicing, the 2' hydroxyl group of the BP adenosine attacks the 5'ss phosphate, resulting in the formation of a branched intron lariat structure, and a free 3' hydroxyl in exon 1. In the second step, this hydroxyl attacks the phosphate at the 3'ss, resulting in the ligation of the exons and release of the intron lariat.

1.5.2 Assembly of the U2-dependent spliceosome

Unlike another large ribonucleoprotein machinery, the ribosome, the components of the spliceosome must undergo a series of complicated rearrangements to become catalytically activated (Figure 4). While pre-formed higher order structures of snRNPs have been observed in some cases (Konarska and Sharp, 1988; Stevens *et al.*, 2002; Malca *et al.*, 2003), the interactions of snRNPs with each other and with the pre-mRNA are generally established in a

step-wise manner (Tardiff and Rosbash, 2006; Huranová *et al.*, 2010). The snRNPs are loosely associated with the pre-mRNA already at the earliest stages of spliceosome assembly, but their interactions with the pre-mRNA are tightly controlled in order to avoid premature and unspecific activation of the spliceosome (reviewed by Smith *et al.*, 2008; Wahl *et al.*, 2009). The reactive sites in the pre-mRNA are recognized multiple times by RNA or protein factors, a process often referred to as proofreading. The recognition events generally depend on interactions that are weak on their own, but are stabilized by the combined effects of multiple factors. Such intrinsically weak interactions also allow the exchange of binding partners, which is crucial for remodeling of the spliceosome at different stages of its assembly. Thus, the spliceosome is a highly dynamic molecular machine that undergoes dramatic changes both in its composition and its conformation during its maturation.

The association of snRNPs and other spliceosome components with the pre-mRNA during spliceosome assembly are often described in terms of complexes (E, A, B, B_{act}, B* and C; Figure 4) that can be separated by biochemical methods (reviewed by Will and Lührmann, 2011). The initial step in intron recognition is the formation of the commitment (or E) complex, during which U1 snRNP binds to the 5'ss sequence. U1 snRNA base-pairs to the 5'ss through a complementary sequence at its 5' end (Zhuang and Weiner, 1986), and this interaction is stabilized by protein-RNA interactions, most significantly by the U1C protein, which stabilizes the 5'ss/U1 helix (Heinrichs *et al.*, 1990; Pomeranz Krummel *et al.*, 2009). During this stage, the BPS is recognized by the protein factor SF1 (Berglund *et al.*, 1997), which defines the catalytically functional adenosine such that it bulges out of the structure (Liu *et al.*, 2001), making it available for the subsequent nucleophilic attack in the later stages of spliceosome maturation. At the same time, the PPT is recognized by the 65 kDa subunit of U2AF (Zamore and Green, 1989), while its 35 kDa subunit interacts with the 3'ss (Guth *et al.*, 1999; Wu *et al.*, 1999; Zorio and Blumenthal, 1999).

Although the U2 snRNP is already present in the E complex and in proximity to U1 (Dönmez *et al.*, 2007), it becomes stably associated with the BPS in the ATP-dependent A complex, also known as the prespliceosome (Bindereif and Green, 1987; Liao *et al.*, 1992). During the formation of the prespliceosome, U2AF recruits U2 snRNP to the BPS through interactions with the SF3b complex, displacing SF1 (Ruskin *et al.*, 1988; Valcárcel *et al.*, 1996; Gozani *et al.*, 1998; Rutz and Séraphin, 1999; Spadaccini *et al.*, 2006; see also Figure 4). Thus, in a proofreading event typical of the spliceosome, the task of defining the BP is transferred from SF1 to U2 snRNP. U2 snRNA associates with the BPS through base-pairing interactions (Wu and Manley, 1989; Zhuang and Weiner, 1989), which exclude the reactive adenosine (Query *et al.*, 1994), therefore retaining the bulged conformation. The specific recognition of the BP is additionally mediated by proteins of the SF3b complex, which together with the SF3a complex is also involved in stabilizing U2 binding to the BPS and to U2AF (Gozani *et al.*, 1996; 1998; Spadaccini *et al.*, 2006).

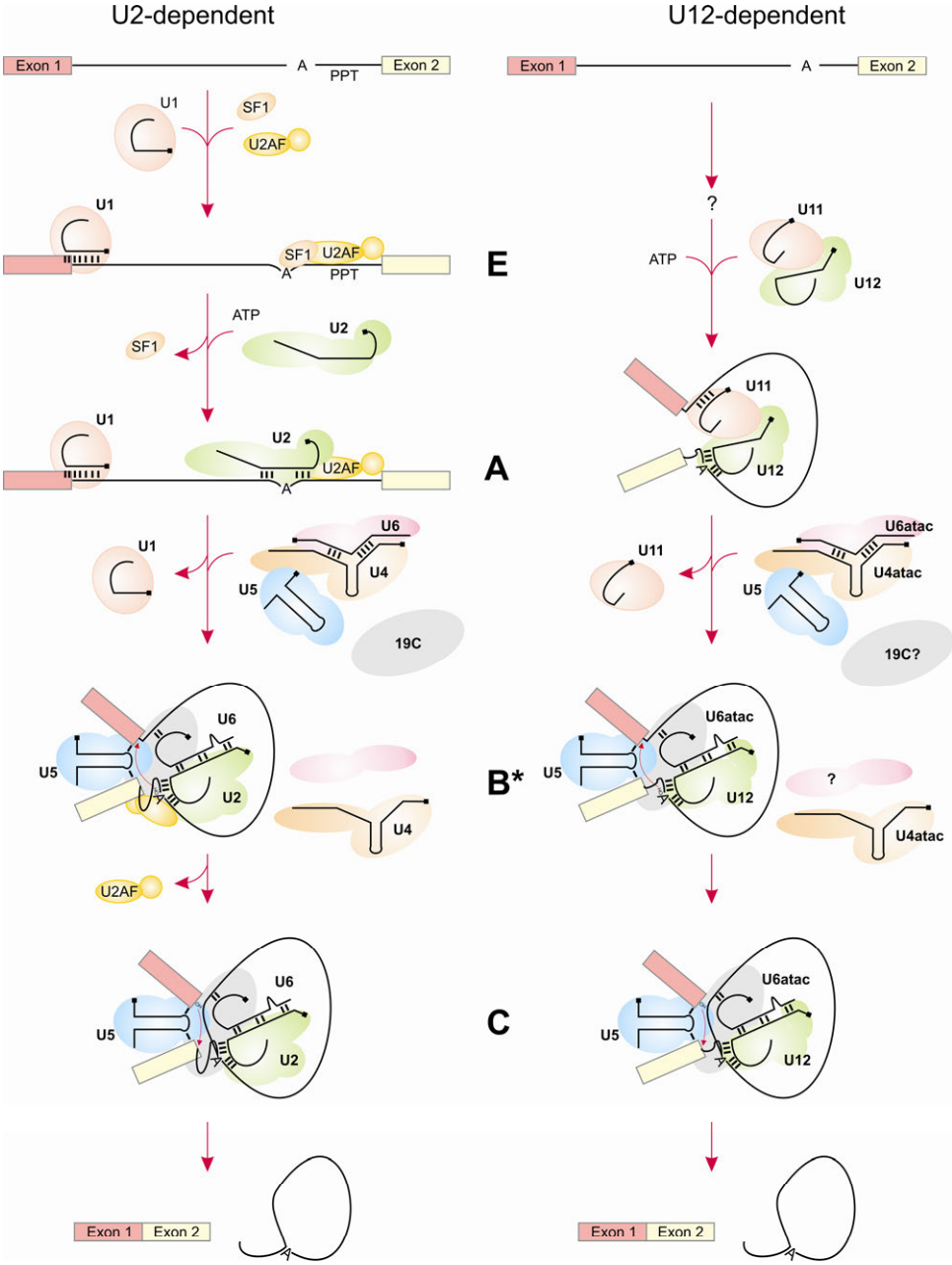


Figure 4. Spliceosome assembly.
The interactions of the spliceosomal snRNPs and some selected non-snRNP protein complexes at various stages of spliceosome assembly are depicted schematically for both the U2-dependent and U12-dependent spliceosomes, as indicated above the panel (see chapter 1.5 for details). Complexes E, A, B* and C are indicated in the middle.The Prp19/CDC5 complex is indicated by "19C". The association and dissociation of certain protein complexes is not known in detail for the U12-dependent spliceosome, and such events are marked with question marks.

The U4/U6-U5 tri-snRNP is also present already in the E complex, with the U5-specific protein Prp8 in direct contact with the 5'ss (Maroney *et al.*, 2000). However, compositional and conformational changes during B complex formation lead to the stable association of U4/U6-U5 tri-snRNP (Konarska and Sharp, 1987; Lamond *et al.*, 1988), followed by a number of significant rearrangements in RNA-RNA interactions to yield the pre-catalytic B_{act} complex: In another significant proofreading event, U1 is displaced from the 5'ss, which instead becomes base-paired to U6 (Kandels-Lewis and Séraphin, 1993). The helices between U4 and U6 snRNAs unwind and U4 snRNP is released from the spliceosome (Konarska and Sharp, 1987; Lamond *et al.*, 1988). U6 base-pairs to U2, forming RNA structures necessary for splicing catalysis (Wu and Manley, 1991; Madhani and Guthrie, 1992). U5 snRNA is not involved in sequence-specific base-pairing interactions, but seems to stabilize spliceosome structures by binding to both exons via its U-rich loop structure (Newman and Norman, 1992; Sontheimer and Steitz, 1993).

Interactions between the snRNAs are not the only ones to be remodeled at this stage. Many of the snRNP proteins dissociate, including all those of U1, U4 and U6, as well as a number of non-snRNP proteins (Bessonov *et al.*, 2008; Fabrizio *et al.*, 2009; Agafonov *et al.*, 2011). The spliceosome is also joined by a number of factors, including the Prp19/CDC5 complex, which is involved in rearranging many protein-protein and protein-RNA interactions, particularly those of U5 and U6 (Makarov *et al.*, 2002; Chan *et al.*, 2003; Makarova *et al.*, 2004). These changes eventually yield the activated spliceosome, or complex B* (Figure 4). The U5-specific Prp8 protein likely functions as a crucial platform for the catalytic core, as it forms a network of interactions with the catalytically active sites during the maturation of the spliceosome, eventually interacting with both splice sites and the BPS, as well as U5 and U6 snRNAs (reviewed by Grainger and Beggs, 2005). After the first catalytic step, the SF3a and SF3b complexes of U2 snRNP dissociate, and the spliceosome is joined by further factors, including specific helicases and peptidyl-prolyl *cis/trans* isomerases (Bessonov *et al.*, 2008; Fabrizio *et al.*, 2009; Agafonov *et al.*, 2011). These likely facilitate the conformational changes required for formation of complex C, which catalyzes the second catalytic step.

1.5.3 Is the spliceosome a ribozyme?

The similarities between the spliceosome and self-splicing group II introns suggest that it may also be an RNA enzyme (see chapter 1.2.2.3). Multiple lines of evidence suggest that U6 snRNA has a crucial role in the catalytic center. All U6-specific proteins, as well as the U4 snRNP, are released during the maturation of the catalytic spliceosome (Bessonov *et al.*, 2008; Agafonov *et al.*, 2011), suggesting that their function is merely to escort U6 snRNA into the catalytic core. Moreover, the binding of Mg^{2+} ions by U6 has been implicated in catalysis (Yean *et al.*, 2000; Lee *et al.*, 2010), and the first step of splicing takes place in the vicinity of an invariant sequence of U6 (reviewed by Valadkhan, 2010). Most strikingly, U2 and U6 snRNAs can catalyze a two-step splicing reaction between short oligonucleotides in the absence of proteins (Valadkhan *et al.*, 2009). However, it is clear that proteins play significant roles for the specificity, efficiency and fidelity of the spliceosome. Recent evidence also suggests that protein components of the spliceosome, particularly Prp8 and its RNase H-like

domain, might be involved in the catalysis itself (reviewed by Abelson, 2008). It is likely that both proteins and RNA contribute to the catalytic activity, but the question remains unresolved for the time being.

1.5.4 Assembly of the U12-dependent spliceosome

The overall assembly pathways in the two spliceosomes are similar (reviewed by Patel and Steitz, 2003). However, the initial recognition of U12-type introns differs from that of U2-type introns (Figure 4). The preformed U11-U12 di-snRNPs bind the intron as a unit, and the 5'ss and BPS are recognized in a co-operative manner, although U11/5'ss base-pairing can still be detected prior to stable base-pair formation between U12 and the BPS (Frilander and Steitz, 1999). Consequently, the first observed complex for U12-type introns in native gels is the ATP-dependent A complex (Tarn and Steitz, 1996b; Frilander and Steitz, 1999; Figure 4). U12-type introns do not have PPTs, and U2AF is not required for the recognition of U12-type introns (Shen and Green, 2007). However, a U2AF35-related protein, Urp, is required for A complex formation and 3'ss recognition (Shen *et al.*, 2010). Interestingly, Urp is also required for U2-dependent splicing, but only after the first catalytic step, when it apparently displaces U2AF from the PPT and 3'ss (Shen *et al.*, 2010), suggesting that the U2-type 3'ss is proofread several times by protein factors. In contrast, the lack of PPT, more conserved recognition sequences and more restricted BPS-3'ss distance (Figure 1) suggest that the recognition of U12-type introns relies more on RNA-RNA interactions, and protein-mediated proofreading is less important (Brock *et al.*, 2008).

As stated above (chapter 1.4.2), due to the low abundance of U12-type factors, our understanding of the specific interactions during minor spliceosome assembly is less complete than for the major spliceosome. However, despite the differences in intron recognition, it seems that the steps leading to catalytic core formation are similar to those of the major spliceosome (Figure 4). The U12/BPS duplex is highly analogous to the U2/BPS duplex, causing the bulging of the reactive adenosine (Tarn and Steitz, 1996b). The SF3b complex associates with U12 as well as U2 snRNPs, and helps to define the BPS in both spliceosomes (Gozani *et al.*, 1996; Will *et al.*, 2001; 2004). Upon formation of the B complex, the U4atac/U6atac/U5 tri-snRNP joins the forming spliceosome, leading to the displacement of U11 from the 5'ss by U6atac, which also base-pairs to U12, with the concomitant release of U4atac (Tarn and Steitz, 1996a; Yu and Steitz, 1997; Incorvaia and Padgett, 1998). The modifications taking place in the RNA-RNA interaction network are similar to those of the major spliceosome, but the order of events appears to be somewhat more flexible in the minor spliceosome (Frilander and Steitz, 2001). The active spliceosome then catalyses the two-step transesterification reaction, resulting in ligation of the exons and release of the intron lariat (Tarn and Steitz, 1996b; see Figures 3 and 4). Due to the identical chemistries of the splicing reactions, the catalytic cores of the two spliceosomes are likely to be similar. In support of this, the functional domains of U6 and U6atac snRNAs are highly similar, and are functional in splicing when replaced with one another (Shukla and Padgett, 2001).

1.5.5 Comparison of initial intron recognition in the two spliceosomes

The absence of U2AF and SF1, as well as the lack of an E complex during U12-dependent prespliceosome assembly, have led to the suggestion that recognition of the minor introns relies more on snRNA-RNA interactions, and requires less proofreading by protein factors than recognition of major introns (chapter 1.5.4; for discussion, see also Patel and Steitz, 2003). This may also be the consequence, at least partially, of the co-operative recognition of the minor 5'ss and BPS by the preformed U11/U12 di-snRNP (Frilander and Steitz, 1999), which imposes more stringent requirements on sequence recognition. However, snRNAs are not alone in recognizing the splice sites in the minor prespliceome. As mentioned above (chapter 1.5.4), proteins factors of the SF3b complex are involved in recognition of the U12-type BPS, although the SF3a complex present in the U2-type prespliceosome is absent (Will *et al.*, 1999; 2004). Interestingly, in contrast to major prespliceosome, the first three nucleotides of the U12-type introns (RUA; see Figure 1B) are not recognized by snRNAs at all, suggesting that specific protein factors are involved (see chapter 4.1). In the major spliceosome, the 5'ss/U1 helix spans the exon-intron junction, and is stabilized by interactions with the U1C protein. It has been noted that the U11/U12-20K protein has sequence similarity with U1C (Will *et al.*, 2004), but no interaction with the 5'ss and the 20K protein has been observed. U1/5'ss interactions are also stabilized by SR proteins, particularly SRSF1, which binds directly to the U1-70K protein (see chapter 1.6.4.1). The U11-35K protein shares similarity with U1-70K (Will *et al.*, 1999), and both have been observed to interact with the homologs of SRSF1 in *Arabidopsis thaliana* (Lorković *et al.*, 2004), suggesting that 5'ss recognition can be enhanced by similar mechanisms in both spliceosomes.

SR proteins have been shown to interact with both the 5'ss and BPS in both spliceosomes (Shen and Green, 2006, 2007), and may thus contribute to interactions that bring these two sites together. However, in the minor prespliceosome, the internal components of the U11/U12 di-snRNP are obviously involved in bridging the 5'ss and the BPS (Benecke *et al.*, 2005; **I**). Interestingly, the catalytically active 5' end of U12 is also brought close to the 5'ss during prespliceosome formation, suggesting that major rearrangements in the conformation of U12 are not necessary at later stages of spliceosome assembly (Frilander and Meng, 2005). Similarly, the U2-type 5'ss is in proximity to both the BPS and U2 already in the E complex, before U2 base-pairs to the BPS (Kent and MacMillan, 2002; Dönmez *et al.*, 2007). The 5' region of U2 snRNA has been shown to be close to the stem-loop SL3 of U1 snRNA, suggesting it might function as a binding platform for U2 snRNP (Dönmez *et al.*, 2007; Weber *et al.*, 2010). However, a number of non-snRNP proteins, including Prp5, have also been implicated in bridging U1 and U2 snRNPs (Xu *et al.*, 2004; Shao *et al.*, 2011), indicating that internal snRNP components are not sufficient for intron bridging in the major spliceosome. It is likely that these differences in intron recognition are used by spliceosome-specific regulatory mechanisms to control their activity.

1.6 Splice site definition and alternative splicing

1.6.1 Initial splice site definition over exons

In many multicellular eukaryotes, especially vertebrates, primary transcripts are mainly composed of intronic sequences, and some of the introns may be hundreds of kilobases long. This poses the splicing machinery with a difficult problem of correctly pairing splice sites that are separated by huge distances. On the other hand, vertebrate exons are generally rather short, suggesting that the initial splice site pairing in organisms with long introns in fact takes place over exons (Figure 5; reviewed by Berget, 1995). Indeed, 3'ss-recognizing factors of the upstream intron can interact with the 5'ss-recognizing factors of the downstream introns (Robberson *et al.*, 1990; Hoffman and Grabowski, 1992), and long exons flanked by long introns are recognized poorly (Sterner *et al.*, 1996; Fox-Walsh *et al.*, 2005). Such exon definition interactions are later replaced with interactions that pair the splicing factors within the same intron. The final splice site pairing generally takes place in the A complex (Lim and Hertel, 2004; see also Figure 4), but apparently the process is flexible, as direct conversion of exon-defined complexes into B complexes has also been observed (Schneider *et al.*, 2010).

Exon definition appears to be particularly important in mammals. The most common form of aberrant splicing in mammals is exon skipping (Nakai and Sakamoto, 1994), which is consistent with pairing of splicing factors located at the upstream and downstream exon, when defining the middle exon fails. The importance of exon definition in mammals is also apparent by the fact that mutations weakening one splice site flanking an exon are often compensated by mutations strengthening the other (Xiao *et al.*, 2007). In contrast, in organisms with shorter introns, such as plants, fungi and many invertebrates, compensatory mutations generally occur within the same intron (Xiao *et al.*, 2007), suggesting that shorter introns are defined directly, as also experimentally observed for mammalian introns by Sterner *et al.* (1996) and Fox-Walsh *et al.* (2005).

1.6.2 Splicing enhancers and silencers

Due to their short length and, especially in U2-type introns, degenerate nature, the core splicing signals within introns are not sufficient for unambiguous definition of splice sites (Burge *et al.*, 1999; Lim and Burge, 2001). Furthermore, true U2-type 5'ss signals are outnumbered many-fold by similar, non-functional splice sites (pseudo splice sites), and are often found in positions where they could define putative exons (pseudoexons) with upstream 3'ss-like sequences. Thus, the correct activation of splice sites requires additional information, which is provided by enhancer or silencer sequences in the vicinity of the splice sites (Figure 5). They are categorized based on their location into exonic splicing enhancers (ESEs) or silencers (ESSs) or intronic splicing enhancers (ISEs) or silencers (ISSs). Although each of these splicing regulatory element (SRE) subtypes has some typical characteristics, they are composed of short and variable sequences. As they can function in combinations, and either in a co-operative or antagonistic manner, their effects on splicing are highly dependent on context, and are difficult to predict based on sequence information alone (for review, see Wang and Burge, 2008). However, large-scale bioinformatic studies have provided

information about SRE distributions in true exons and pseudoexons, and the effect of short sequence motifs on splicing *in vitro* and *in vivo* has been tested experimentally by selective molecular evolution methods (reviewed by Chasin, 2007). Most promisingly, some progress has recently been made in this field by using computational methods to analyze and predict alternative splicing patterns in different tissues (Barash *et al.*, 2010). Among other things, this study revealed that alternative splicing is affected by intronic sequence elements located much further away (250–300 nt) from splice sites than previously thought.

While it has been established that the more conserved U12-type introns do contain enough information to be recognized correctly in a given transcript, their presence in only a handful of transcripts means that specifying their unique location within all transcripts requires additional information (Burge *et al.*, 1999). Accordingly, they also respond to regulation by SREs (Wu and Krainer, 1998; Hastings and Krainer, 2001; Lewandowska *et al.*, 2004), and U12-dependent splicing is enhanced by exon definition interactions with U2-type factors in neighbouring introns, and vice versa (Wu and Krainer, 1996; Lewandowska *et al.*, 2004; **II**).

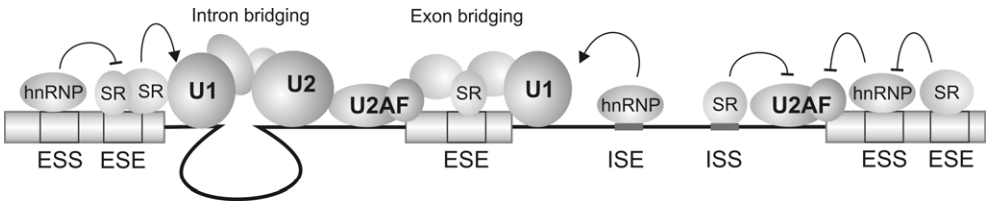
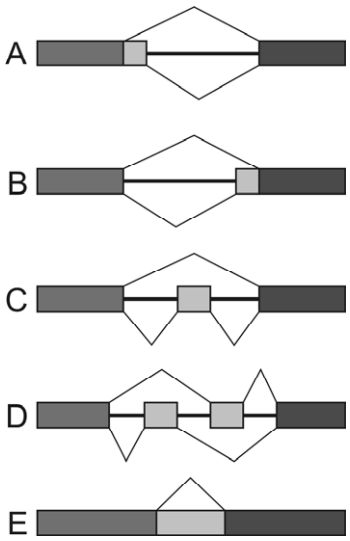


Figure 5. Splicing regulatory elements and splice site definition.
Splicing is regulated by trans-acting factors that bind to enhancer and silencer elements. These are also involved in bringing together spliceosomal components bound at the 5'ss and 3'ss, either through exon definition or intron definition interactions. Exons and the regulatory elements (ESEs, ESSs) within them are depicted as boxes, and introns and their regulatory elements (ISEs, ISSs) by lines. Interactions between splicing factors are depicted schematically, with arrows representing activating interactions, and blocked lines representing inhibitory interactions. See chapter 1.6 for details.

1.6.3 Alternative splicing increases proteome complexity

Regulation of splicing by SREs is important not only for distinguishing true exons from pseudoexons, but also for producing alternatively spliced mRNAs. Alternative splicing is the most significant mechanism for increasing proteome diversity, accounting for most of the estimated 5-fold excess of proteins over protein-coding genes in humans (reviewed by Nilsen and Graveley, 2010). In fact, it is utilized by almost all protein-coding genes, as recent studies indicate that 92–95 % of human multi-exon pre-mRNAs are spliced to produce at least two abundant isoforms (Pan *et al.*, 2008; Wang *et al.*, 2008). A typical case of alternative splicing in mammals involves the inclusion or exclusion of an alternative cassette exon (Figure 6C), reflecting the importance of exon definition interactions in these organisms (see chapter 1.6.1). Other, slightly less common modes of alternative splicing include use of alternative 5'ss or 3'ss, mutually exclusive exons or intron inclusion (Figure 6A, 6B, 6D and 6E, respectively). Thus, the majority of alternative splice site choices seem to take place between competing splice sites during formation of the commitment complex, either through exon definition or

intron definition. However, it has been shown that alternative splicing can also be affected during the transition from exon-defined complexes to intron-defined spliceosomes, so that exons already defined by splicing factors are nonetheless excluded from the mRNA. This can occur either through hyperstabilization of exon-defined complexes such that they cannot pair with other splicing complexes over introns, or by selectively inhibiting or enhancing the formation of intron definition interactions between specific exon-defined complexes (House and Lynch, 2006; Bonnal *et al.*, 2008; Sharma *et al.*, 2011). Other mechanisms during later steps are also occasionally involved, such as formation of stalled complexes (Lallena *et al.*, 2002; Giles and Beemon, 2005).



While different types of SREs are utilized for both constitutive and alternative splicing, certain differential patterns can be observed in genome-wide studies. Enhancing elements are expected to be important for efficient and precise splicing of constitutive exons, and ESEs are indeed enriched in constitutive exons, and are particularly underrepresented in pseudoexons (Wang *et al.*, 2005). ESSs also have an important role in distinguishing true exons from pseudoexons, but additionally function in regulating alternative splice site selection and intron retention (Wang *et al.*, 2004; 2006). Intronic SREs are likely to have particularly important functions for alternative splicing control, as intronic regions surrounding alternative exons are far more conserved than those surrounding constitutive exons (Sorek and Ast, 2003).

Figure 6. Types of alternative splicing.
Various types of alternative splicing are depicted schematically. Constitutive exonic sequences in the pre-mRNA are represented by the darker rectangles, and the intronic regions by lines. The light rectangles represent regions that can be either excluded or included in the mature mRNA, as indicated by the lines above and below the pre-mRNA. A) Splicing at alternative 5' splice sites. B) Splicing at alternative 3' splice sites. C) Alternative cassette exon (i.e. exon skipping or inclusion). D) Mutually exclusive exons. E) Intron retention.

1.6.4 *Trans*-acting factors affecting splice site choice

Although some SREs can affect splicing directly, e.g. through formation of specific RNA secondary structures (reviewed by Buratti and Baralle, 2004), most SREs function by recruiting *trans*-acting factors that either stimulate or suppress splicing (Figure 5). Thus, the effect of the SREs does not depend solely on their presence in the transcript, but also on the availability of different *trans*-acting factors, which varies in different cell types and developmental stages, giving rise to different splicing isoforms. There are various kinds of *trans*-acting factors, from riboswitch-binding small molecule ligands to small nucleolar RNAs (snoRNAs) and snRNPs (reviewed by Khanna and Stamm, 2010), but the most extensively characterized and possibly most common regulators of splicing are protein factors. They

generally bind to SREs through their RNA-binding domains, many different types of which have been described, including the RNA recognition motif (RRM), hnRNPK homology (KH) domain and zinc-binding domains (reviewed by Auweter *et al.*, 2006). In most cases, these domains recognize specific sequences in ssRNA, although certain domains, such as some RRMs, can show considerable flexibility in their target sequences. This is achieved through the recognition of a few specific bases by the RRM, which stabilizes the binding through non-specific interactions with RNA. Thus, while clear binding sites can be discerned for some splicing factors, many of them can bind to a wide variety of targets, increasing the complexity of alternative splicing. The final splice site selection is often decided through the co-operative and competitive actions between different splicing regulators. The best known regulators belong to one of two significant groups of proteins, the SR proteins and hnRNP proteins, which often, but not always, have antagonizing effects on splice site selection (Figure 5).

1.6.4.1 SR proteins

SR (serine-arginine-rich) proteins form a multi-functional group of proteins that operate in various steps of gene expression, from chromatin remodeling to mRNP transport and translation regulation (reviewed by Zhong *et al.*, 2009). SR proteins are structurally related, each of them having at least one N-terminal RRM and a downstream arginine-serine-rich (RS) domain (Manley and Krainer, 2010). Additionally, a large number of SR-like proteins contain RS domains, although they may lack the RRM domains and instead contain different RNA-binding domains or other functional domains (reviewed by Long and Caceres, 2009).

Many constitutive splicing factors are SR proteins or SR-like proteins, including U2AF65, U2AF35, Urp and U1-70K. However, most SR proteins are perhaps best known for their functions as splicing activators in both constitutive and alternative splicing, typically functioning through binding to ESEs (reviewed by Blencowe, 2000). SR proteins stimulate recruitment of splicing factors to nascent transcripts during transcription (Das *et al.*, 2007), and are required for committing pre-mRNA to the splicing pathway (Fu, 1993). Many SR proteins directly interact with core spliceosomal components and stabilize their binding, and they are essential for bridging the 5'ss and 3'ss in U2-type introns as well as mediating bridging interactions needed for exon definition (Boukis *et al.*, 2004). A well-characterized ESE-binding SR protein with multiple functions is SRSF1 (also known as ASF/SF2), which enhances binding of U1 snRNP to the 5'ss (Kohtz *et al.*, 1994; Cho *et al.*, 2011), and also interacts directly with the RNA at the BPS as well as with U2AF35 at the 3'ss (Wu and Maniatis, 1993; Shen and Green, 2004; Shen *et al.*, 2004). SR proteins are additionally involved in recruiting the U4/U6-U5 tri-snRNP (Roscigno and Garcia-Blanco, 1995; Makarova *et al.*, 2001). The recognition of the U12-type 5'ss and BPS by the U11/U12 di-snRNP is also promoted by SR proteins, and they directly contact U12-type introns during spliceosome assembly (Hastings and Krainer, 2001; Shen and Green, 2007).

Another way in which SR proteins stimulate splicing is to counteract the effect of inhibitory factors such as hnRNPs by either competing for binding sites or by blocking interactions required for silencing (see e.g. Zhu *et al.*, 2001; Crawford and Patton, 2006). Such

competition between splicing activators and suppressors is particularly important in alternative splicing. SR proteins can also change alternative splicing patterns by competing with each other, such that different SR proteins promote inclusion of different exons (Gallego *et al.*, 1997; Jumaa and Nielsen, 1997; Han *et al.*, 2011). Similar to other splicing regulators, the functions of SR proteins are context-dependent, and SR proteins with known functions in ESE-dependent splicing activation, including SRSF1, can also recognize ISSs and suppress inclusion of alternative exons (Kanopka *et al.*, 1996; Jiang *et al.*, 1998; Buratti *et al.*, 2007b).

An important mechanism for regulating the activities of SR proteins is phosphorylation at the serine residues of their RS domains. Such regulation is particularly important for the regulation of alternative splicing, as it enables changes in alternative splicing patterns in response to external stimuli, through kinase-mediated signaling cascades (reviewed by Lynch, 2007). Phosphorylation can also influence constitutive splicing. Phosphorylation of the RS domain of SRSF1 is required for a conformational switch, which enables the SRSF1 RRM domain bound to an ESE to also interact directly with the RRM domain of U1-70K, enhancing 5'ss recognition (Cho *et al.*, 2011). Conversely, dephosporylation of both SRSF1 and U1-70K is required for the catalytic step of splicing (Tazi *et al.*, 1993; Cao *et al.*, 1997).

1.6.4.2 HnRNP proteins

Heterogeneous nuclear RNP (hnRNP) proteins are another well-known group of splicing regulators. Like SR proteins, they are also involved in various other processes, including DNA repair, chromatin remodeling, transcription and mRNP export (reviewed by Han *et al.*, 2010). However, in contrast to SR proteins, they do not form a structurally unified group with similar domain structures. Some of them contain RRM domains, similar to SR proteins, but others contain RNA-binding domains that are structurally distinct from the RRM domain, such as the KH domain. Some hnRNPs also contain RS domains, and the distinction between hnRNPs and SR or SR-like proteins may in some cases be considered merely the result of a historical co-incidence, reflecting a presumed divide into activators and suppressors of splicing. Similar to SR proteins, hnRNPs are subject to modifications that regulate their functions. They undergo a variety of post-translational modifications, including phosphorylation and methylation. The production of different isoforms by alternative splicing is also widespread among hnRNPs, reflecting their varied functions. Many such isoforms are also specific to certain organisms or tissues or developmental phases.

Although hnRNPs have previously been regarded as splicing suppressors, it has become evident that their functional properties are as varied as their structures. For example, proteins of the hnRNPA/B family are well known for antagonizing the effects of SR proteins in splicing activation and promoting exon skipping (reviewed by Han *et al.*, 2010). A member of this family, hnRNPA1, has been shown to induce exon skipping by forming homodimers with the individual proteins bound on either side of the exon, thus making the exon unavailable to the splicing machinery (Chabot *et al.*, 1997; Han *et al.*, 2005). However, the same mechanism can also be employed to enhance exon inclusion by two hnRNPA/B molecules bound within an intron, which upon dimerizing bring together the two ends of the intron and promote

splicing (Martinez-Contreras *et al.*, 2006). HnRNPA/B proteins can also form similar bridging dimers with other hnRNPs, including members of the hnRNPF/H protein family (Martinez-Contreras *et al.*, 2006; Fisette *et al.*, 2010; see the following chapter). In addition to altering the RNA architecture, hnRNPs are also known to compete directly with spliceosome components. This is the case with hnRNPF/H proteins (see following chapter), and PTB (polypyrimidine tract binding protein, also known as hnRNPI), which binds to PPT, occluding binding of U2AF (Matlin *et al.*, 2007). Similar to hnRNPA/B proteins, several PTB molecules also often bind on both sides of exons (Xue *et al.*, 2009), and PTB interferes with the transition from exon-defined complexes to intron-defined complexes by binding next to the 5'ss and interacting directly with U1, thus blocking its interactions with splicing factors downstream (Sharma *et al.*, 2008; Sharma *et al.*, 2011).

1.6.4.3 G-run motifs and the hnRNPF/H protein family

Most hnRNPs have highly degenerate consensus binding sites. However, the members of the hnRNPF/H protein family have a fairly defined target sequence, centered around a G-run motif, DGGGD (where D is A or G or U) (Caputi and Zahler, 2001; Schaub *et al.*, 2007). This family is composed of five structurally related proteins known as hnRNPs F, H1 (or H), H2 (H'), H3 (2H9) and GRSF-1. Their RNA-binding domains resemble the canonical RRM domains, but have different folds, and have therefore been named quasi-RRMs (qRRMs; Honoré *et al.*, 1995). Like many hnRNPs, the members of the hnRNPF/H protein family are present in many alternatively spliced isoforms, some of which lack some or all of the qRRMs, suggesting that they may have variable functions in splicing as well as in other processes (Honoré, 2000). Each member of the hnRNPF/H group can bind to the G-run motif on its own *in vitro*, but they can also form aggregates with each other (Schaub *et al*., 2007). The members of the protein family also have variable functions, some of which appear to be redundant, while others are specific. For example, ESS-bound hnRNPF/H can in certain cases repress the inclusion of an exon in a redundant manner (Crawford and Patton, 2006; Mauger *et al.*, 2008; Coles *et al.*, 2009), while in other systems the repression takes place with hnRNPH1 and/or H2, but not with hnRNPF (Buratti *et al.*, 2004; McNally *et al.*, 2006). Such differences may be essential in defining alternative splicing patterns in different tissues, as hnRNPH1 levels vary greatly from one tissue to another (Honoré *et al.*, 2004). Also, while hnRNPs F, H1 and H2 are each found in both the nucleus and the cytoplasm, hnRNPs H1 and H2 are primarily nuclear, and hnRNPF is more abundant in the cytoplasm (Matunis *et al.*, 1994; Honoré *et al.*, 2004).

Similar to other SREs, the functions of the G-run motif are context-dependent. Exonic G-runs generally function as ESSs, while intronic G-runs are ISEs, and are particularly prevalent downstream of 5'ss sequences (Xiao *et al.*, 2007; Xiao *et al.*, 2009). Evolutionary studies have shown that strengthening such G-run ISEs compensates for weakened 5' splice sites (Xiao *et al.*, 2007). Activation of 5' splice sites by ISE-bound hnRNPH1 has been observed in both U2 and U12-type introns (Han *et al.*, 2005; McNally *et al.*, 2006; Wang and Cambi, 2009). A large-scale RNA interference (RNAi) knockdown experiment revealed that hnRNPH1 activates splicing of introns with G-run ISEs, particularly introns with intermediate strength

5'ss sequences, suggesting hnRNPH1 can act as a buffer for splice site mutations (Xiao *et al.*, 2009). In contrast, the effects of hnRNPH1 as an ESS-bound repressor were strongest on exons with strong or weak 5' splice sites.

While these results show that the G-runs and hnRNPH1 (and other hnRNPF/H proteins) have widespread functions in splicing regulation, less is known about the precise molecular mechanisms affecting splicing. At least in some cases, G-run ISEs and hnRNPF/H proteins enhance U1 binding to the 5'ss (Wang and Cambi, 2009), while in other cases they enhance spliceosome assembly during the formation of the A-complex, but not during the initial U1 binding (Schaub *et al.*, 2007). The enhancing or suppressing effect of G-runs has also been attributed to formation of cross-intron or cross-exon hnRNPH1 homodimers or hnRNPH1/hnRNPA1 heterodimers (Martinez-Contreras *et al.*, 2006; Fisette *et al.*, 2010). In many cases the G-run ESSs are located close to one of the splice sites, suggesting steric competition with splicing factors (Jacquenet *et al.*, 2001; Buratti *et al.*, 2004; LeFave *et al.*, 2011). HnRNPF/H proteins have also been shown to repress splicing by directly competing with other splicing factors, such as SR proteins (Crawford and Patton, 2006; Mauger *et al.*, 2008).

1.6.4.4 Core spliceosomal components as splicing regulators

Originally, alternative splicing was thought to be regulated by auxiliary splicing factors, such as SR proteins and hnRNPs. However, it has become clear that the levels and activity of core spliceosomal factors can also affect splicing patterns. In an RNAi screen of *Drosophila melanogaster* splicing factors, several core spliceosomal factors were identified as alternative splicing factors, including U2AF and components of U1, U2 and U4/U6 snRNPs (Park *et al.*, 2004). Also in human cells, depletion of U2AF65 and its homologue PUF60 resulted in altered splicing patterns (Hastings *et al.*, 2007), and depletion of the common snRNP-component SmB/B' resulted in changes in splicing of alternative exons, but not constitutive exons (Saltzman *et al.*, 2011). The depletion particularly affected U1 snRNPs, and the alterations in alternative splicing patterns were most likely caused by competition between 5' splice sites, with strong splice sites being favoured over weak ones. Specific changes in alternative splicing patterns were also observed after the depletion of the U1C protein in zebrafish (Rösel *et al.*, 2011).

Among core spliceosome components, U1 appears to be a particularly prevalent regulator of alternative splicing. U2-type 5'ss-like sequences were found as potent suppressors of pseudoexon splice sites in a screen for exonic splicing silencers (Wang *et al.*, 2004), and constituted as much as 17% of elements downstream of exons in a screen for intronic SREs (Yeo *et al.*, 2007). Pseudoexon repression has been found in the transcripts of the *ATM* gene, where U1 suppresses exon inclusion by binding to an ESS close to the 3'ss and promoting unstable binding of U2 upstream (Dhir *et al.*, 2010). Inhibition of true 5' splice sites by U1-binding upstream has also been observed, and is most likely caused by steric hindrances (Siebel *et al.*, 1992; Cloutier *et al.*, 2008). U1 binding to intronic SREs has been associated with activation of an upstream 5'ss (McCullough and Berget, 2000). In this case, U1

recognized the ISEs through base-pairing interactions that did not resemble the canonical U1/5'ss base-pairing. Interestingly, U1 also appears to affect processes other than splicing, as suggested by a recent report showing widespread suppression of premature polyadenylation sites by U1 (Kaida *et al.*, 2010).

Multiple snRNPs can be involved in splicing regulation, as exemplified by the *gag* gene of Rous sarcoma virus (RSV; reviewed by McNally, 2008). In this system, a pseudo-intron resides within an authentic intron, and a full complement of U2-type snRNPs assembles on it. However, the spliceosome assembly is defective, and splicing does not occur. Instead, the defective complex also suppresses the splicing of the authentic intron, possibly by preventing formation of cross-intron interactions. This inhibition can be reversed by binding of U11/U12 di-snRNP or U11 snRNP at a U12-type pseudo-5'ss overlapping the U2-type pseudo-5'ss. This also does not lead to U12-dependent splicing, but prevents binding of U1 at the pseudo-5'ss and permits splicing of the authentic intron. Multiple factors further regulate these events, with SR proteins promoting U1 binding, while hnRNPH1 promotes U11 binding.

1.6.4.5 Complex splicing regulatory elements

In the study of alternative splicing, SREs are in most cases viewed as single building blocks, which bind either repressing or activating factors, with the sum of all SRE-protein interactions defining the alternative splicing pattern. This view has been influenced by the fact that the effects of simple SREs, when present, are naturally much easier to detect than those of highly complex SREs, which may bind competing factors that only cause subtle effects on splicing. Complex regulatory elements have been observed in several systems, including the chicken *cTNT* exon 5 (Ladd *et al.*, 2005), and the human *c-src* exon N1 (Chou *et al.*, 2000; Sharma *et al.*, 2005) and *CFTR* exons 9 and 12 (Pagani *et al.*, 2003a, b). The *CFTR* exon 12 is almost entirely covered with splicing regulatory elements, including complex CERES elements (composite exonic regulatory elements of splicing) (Haque *et al.*, 2010). These elements are less than 10 nt long, but are recognized by a large number of competing factors, including SRSF1 and SRSF6 and hnRNPs A1, A2, C2 and U, and mutations within these elements can affect the binding of several factors at once, resulting in rather unpredictable splicing patterns (Haque *et al.*, 2010).

1.6.4.6 Other factors affecting alternative splicing

As mentioned in the previous chapters, the availability of splicing factors can affect splice site choice. The distribution of splicing factors within a cell is therefore an important factor regulating splicing (reviewed by Nilsen and Graveley, 2010). Due to its co-transcriptional nature, splicing is also affected by kinetic factors linked to transcription. Pausing of RNAPII at exons has been observed, and such changes in the speed of RNAPII may favour certain splicing events over others (reviewed by Nilsen and Graveley, 2010; Carrillo Oesterreich *et al.*, 2011), and affect RNA secondary structure, which in turn has been shown to affect splicing, e.g. by masking splice sites (Buratti *et al.*, 2007a; Yu *et al.*, 2008).

Most intriguingly, it has recently been shown that chromatin structure can also affect splicing. Although the causal relationship between chromatin modifications and splicing still remains unclear, several lines of evidence suggest that the two processes are linked: The average size of exons closely correlates with the average length of DNA bound by a mononucleosome, nucleosomes are subtly enriched in exons vs. introns, and methylation of the lysine 36 residue of histone 3 (H3K36me3) shows similar distribution (reviewed by Schwartz and Ast, 2010). SR proteins can interact directly with chromatin (Kress *et al.*, 2008; Loomis *et al.*, 2009), and PTB has been shown to affect alternative splicing in a way that depends on its binding to specifically methylated histones via a linker protein (Luco *et al.*, 2010). Conversely, inhibition of splicing has been shown to disrupt normal patterns of H3K36 methylation (Kim *et al.*, 2011). While it is not clear what initiates the pattern of histone methylations and how these are used to guide splicing, a tempting hypothesis is that splicing guides the establishment and/or maintenance of the pattern, which in turn affects splicing by recruiting splicing factors and/or by affecting transcription kinetics during the following rounds of transcription (Kim *et al.*, 2011).

1.6.5 Alternative splicing of U12-type introns

Alternative U12-type splicing events are rare, possibly due to the more rigid sequence constraints on the U12-type 5'ss and BPS and their co-operative recognition. Also, exon skipping (Figure 6C), which is the most common form of alternative splicing in mammals, is not observed due to the fact that U12-type introns rarely occur more than once in a gene and U2-type and U12-type splice sites are incompatible with each other. The distance constraints between a U12-type BPS and 3'ss also limit splicing at alternative 3'ss AG dinucleotides, although theoretically use of closely spaced sites would be possible. Such splicing events have been observed, but whether they are true regulated alternative splicing events or noise in the splicing process is not clear (Levine and Durbin, 2001; Zhu and Brendel, 2003; Chang *et al.*, 2007a). In some cases, retention of U12-type introns affects a significant portion of transcripts (Pessa *et al.*, 2006). In general, transcripts containing U12-type introns are more likely to give rise to alternative isoforms through retention of the U12-type intron than through alternative U12-type splice site usage, but the biological significance of these isoforms is also unclear.

Alternative usage of mutually exclusive U12-type and U2-type splice sites has been observed in at least two cases. The *Drosophila melanogaster* gene *prospero* contains a "twintron", a U12-type intron which in itself contains a U2-type intron (Hall and Padgett, 1994). The decision to splice by either the U12 or the U2-dependent pathway is affected by an intronic element which binds the *D. melanogaster* homologues of hnRNPA1 (Scamborova *et al.*, 2004; Borah *et al.*, 2009). In vertebrates, the members of the *JNK* gene family contain a hybrid intron with a U12-type 5'ss and U2-type BPS and 3'ss. These can be spliced in a mutually exclusive fashion to either the 5'ss of the upstream U2-type intron or to the 3'ss of the downstream U12-type intron (Chang *et al.*, 2007a), and at least in mice these isoforms show tissue-specificity, with the latter isoform preferentially expressed in neurons (Casanova *et al.*, 1996).

1.7 Quality control and nonsense-mediated decay

1.7.1 Nuclear degradation of aberrant transcripts

Despite the multiple mechanisms that ensure fidelity in splicing, low levels of mis-splicing are quite common (Melamud and Moult, 2009), producing "noise" transcripts, which might potentially code for deleterious protein isoforms (Faustino and Cooper, 2003). Cells contain a number of mechanisms to avoid and remove such aberrant transcripts. One way is to increase the fidelity of splicing. Indeed, long transcripts that contain many introns, and are therefore more vulnerable to splicing mistakes, contain more ESEs (Melamud and Moult, 2009). Transcripts that do not fully undergo splicing are generally retained in the nucleus and degraded by the exonuclease Rat1p/Xrn2, or by the nuclear exosome, which has both exo- and endonuclease activity, and associates with the elongating RNAPII (reviewed by Egecioglu and Chanfreau, 2011). The mechanism by which unspliced transcripts become specifically retained in the nucleus remains unclear. One possibility is that general retention factors may prevent the export of all transcripts, and those which are not processed fast enough become degraded, while efficiently processed transcripts gain markers that qualify them for export, including a poly-A tail and specific protein complexes.

1.7.2 Splicing is linked to mRNP export

One of the markers linking splicing to export are exon-junction complexes (EJCs), which are protein complexes deposited 20–24 nt upstream of the exon ligation sites, and whose recruitment is likely initiated by interactions with spliceosome components (reviewed by Bono and Gehring, 2011). The EJC core serves as a platform for various protein factors involved in splicing, mRNP export and mRNP quality control. The transcription-export complex (TREX) contacts the EJC, as well as 5' cap-binding proteins, components of the 3' end-processing machinery and the spliceosome (reviewed by Björk *et al.*, 2009). Among the proteins bound by TREX is the export receptor Nxf1, which also directly interacts with hypophosphorylated SR proteins, suggesting that the recruitment of export factors is linked to the spliceosomal rearrangements leading to splicing catalysis (reviewed by Björk *et al.*, 2009; Zhong *et al.*, 2009; see also chapter 1.6.4.1). Consistently, spliced RNA is exported more efficiently than identical RNA expressed from complementary DNA (Valencia *et al.*, 2008).

After their release from the transcription site, the mRNPs move through the nucleus apparently by diffusion, possibly directed by channels formed between volumes of chromatin (reviewed by Björk *et al.*, 2009). Upon reaching the nuclear pore complex, the mRNP structure and composition changes, and some of the associated protein factors are released. However, some of the protein complexes, including the EJCs, remain attached to the mRNP in the cytoplasm, where they play important roles in enhancing and regulating translation and in further quality control, such as in the nonsense-mediated decay (NMD) pathway.

1.7.3 Nonsense-mediated decay

Despite nuclear quality control mechanisms, some aberrantly processed mRNAs are transported into the cytoplasm, and further cytoplasmic control mechanisms are therefore needed. NMD is the most well-known of these mechanisms (reviewed by Chang *et al.*, 2007b). Although NMD can be activated by several mechanisms, the most common one is through EJCs and premature termination codons (PTCs). During the initial round of translation, the ribosome releases EJCs from the mRNP. As the ribosome reaches the STOP codon, the release factors eRF1 and eRF2 likely recruit the NMD factor Upf1. Although the specific details of NMD activation are not known, it is likely that NMD is initiated if Upf1 can interact with Upf2 and Upf3, which are components of the EJC. Such an interaction can only take place with an EJC that has not been displaced from the transcript and must therefore be downstream of the STOP codon, signaling that the STOP codon is upstream of a splicing event and therefore premature. Additionally, the EJC has to be sufficiently far downstream, ca. 50 nt, probably to allow for sufficient space for both the translation termination complex and the EJC. Activation of NMD is counteracted by the poly-A-binding protein, PABP, when located close to the STOP codon, thus ensuring that correctly processed mRNAs escape NMD (Eberle *et al.*, 2008; Singh *et al.*, 2008). Once activated, NMD leads to decapping of the transcript and degradation by the cytoplasmic exosome and other nucleases.

1.8 Conserved sequence elements and regulation of splicing factors through nonsense-mediated decay

1.8.1 Most NMD events are caused by splicing errors

NMD is not only utilized in the degradation of accidentally mis-spliced transcripts, but can also be functionally coupled to alternative splicing (AS-NMD). Alternative splicing can introduce PTCs into mRNAs either by inclusion of PTC-containing sequence fragments, or by inclusion of other fragments that cause a frameshift in downstream sequences, resulting in the formation of a PTC. Some of these events can be used to regulate final mRNA levels. Initial computational approaches predicted 35% of observed mRNA isoforms in the Refseq database to be subject to NMD (Green *et al.*, 2003; Lewis *et al.*, 2003). However, it was later discovered experimentally that the number of events subject to NMD was much lower, and that only a fraction of all AS-NMD events show any signs of being regulated and are greatly outnumbered by stochastic noise caused by the splicing machinery (Mendell *et al.*, 2004; Pan *et al.*, 2006; Wittmann *et al.*, 2006).

1.8.2 Conserved NMD-inducing elements in splicing factor genes

An evolutionary approach searching for conservation of AS-NMD events led to the discovery of a group of proteins that are commonly regulated through AS-NMD, namely those involved in RNA processing, and particularly in splicing (reviewed by McGlincy and Smith, 2008). The PTC-introducing alternative splicing events occur within conserved regions that are 100–200 nt long, and take place in transcripts coding for core spliceosomal proteins, SR proteins, hnRNPs, and other splicing factors, such as Nova and Fox proteins (Lareau *et al.*, 2007; Ni *et al.*, 2007; Yeo *et al.*, 2007; Saltzman *et al.*, 2008). One of the striking discoveries was that all

human SR proteins are regulated through AS-NMD, and the pattern of alternative splicing is also conserved in mouse (Lareau *et al.*, 2007). Similar AS-NMD events are also found in the SR protein transcripts in other animals and in fungi, but the location and even the number of PTC-inducing alternative splicing events varies, and transitions from one AS-NMD event to another are apparent between certain species (L. Lareau, personal communication). Thus, it appears that it is not the precise position of the alternative splicing event that is conserved during evolution, but the mode of AS-NMD regulation as a whole. It is therefore likely that AS-NMD is an ancient and wide-spread mechanism for regulating splicing factors. As a process that is directed by splicing-dependent EJCs, AS-NMD is indeed well suited to function in regulation of splicing factors. The fact that no changes are generally required in the protein coding region of the genes, especially when nonsense exons are involved, also helps to explain the apparent high preference for preserving AS-NMD regulation in these genes, as well as how it has apparently arisen independently several times during evolution.

1.8.3 Feedback regulation of splicing factors

Although the precise mechanism regulating the activation of AS-NMD is not known in all cases, certain common features suggest that feedback regulation by the products of the same gene is often involved (Ni *et al.*, 2007): In SR protein transcripts, PTCs are commonly introduced by inclusion of alternative nonsense exon, suggesting SR proteins activate their inclusion through ESEs. Conversely, in hnRNP transcripts (including hnRNPs H1, K and M and nPTB) PTCs occur due to frameshifts caused by exon skipping, suggesting splicing repression through ESSs. A computational search also identified a number of splicing factor genes with conserved elements, in which the NMD-inducing alternative splice sites were surrounded by known binding sites for the product of the same gene (Yeo *et al.*, 2007). Taken together, these results suggest that AS-NMD may be utilized for widespread homeostatic regulation of splicing factors, which are possibly regulated in a co-ordinated fashion (Ni *et al.*, 2007).

The feedback regulation of several SR proteins and hnRNPs has been studied in detail. The SR proteins SRSF2 (also known as SC35) and SRSF7 (9G8) activate alternative splice sites in their respective pre-mRNAs, resulting in the inclusion of PTCs (Lejeune *et al.*, 2001; Sureau *et al.*, 2001). In many cases, the feedback regulation of splicing factors is also linked to crossregulation of closely related factors. SRSF1 (ASF/SF2) and SRSF3 (Prp20) both promote inclusion of PTCs in their own transcripts, while SRSF1 inhibits the PTC-inducing splicing event in the *SRSF3* transcript (Jumaa and Nielsen, 1997; Sun *et al.*, 2010). The core spliceosomal factor SmB/B' regulates its own expression through AS-NMD, but this alternative splicing event can apparently also be affected by the levels of its homolog SmN (Saltzman *et al.*, 2008; Saltzman *et al.*, 2011). The alternative splicing isoforms of hnRNPA1 are regulated by itself and hnRNPA2 (Chabot *et al.*, 1997; Hutchison *et al.*, 2002), and the *hnRNPA2B1* transcript, the splicing isoforms of which code for hnRNPs A2 and B1, is also subject to autoregulation through AS-NMD (McGlincy *et al.*, 2010). HnRNPL regulates its own expression as well as that of the related hnRNPLL protein (Rossbach *et al.*, 2009). A very interesting cycle of tissue-specific crossregulation involves PTB and its paralogs, the neuron-

specific nPTB, and ROD1, which is mainly expressed in haematopoietic cells. PTB and nPTB both regulate their own levels through AS-NMD, and PTB activates AS-NMD of the *nPTB* transript in non-neuronal cells (Wollerton *et al.*, 2004; Spellman *et al.*, 2007). PTB and nPTB also activate formation of a PTC in the *ROD1* transcript through exon skipping, although this does not necessarily lead to NMD (Spellman *et al.*, 2007).

Like many other aspects of splicing, the mechanisms of feedback regulation are also linked to other processes of the gene expression pathway. For example, the U1 snRNP component U1A is downregulated by inhibition of polyadenylation caused by binding of two U1A molecules and U1 snRNP at the 3' untranslated region (UTR) of the transcript (Boelens *et al.*, 1993; Guan *et al.*, 2007). Same factors can also be regulated on multiple post-transcriptional levels: SRSF1 promotes the splicing of various different *SRSF1* isoforms, some of which code for proteins or are degraded by NMD, while others suppress gene expression on a translational level independently of NMD (Sun *et al.*, 2010). Conversely, some ribosomal proteins are regulated through AS-NMD, which they can even promote themselves (Cuccurese *et al.*, 2005; Russo *et al.*, 2011), thus highlighting the multiple roles of RNA-binding proteins.

The reason for the association of AS-NMD with highly conserved sequences remains an unresolved issue. In general, introns flanking alternative exons show more conservation than constitutive exons, but splicing regulation is also typically quite redundant, and most highly regulated splicing events do not require such a high degree of conservation (reviewed by McGlincy and Smith, 2008). The reason may in part lie in the fact that many of these AS-NMD events are likely to function in the homeostasis in splicing factors. It has been observed that transcriptional feedback loops with strongly repressive regulation tend to result in oscillatory effects on gene expression, while lowly repressive regulation results in more constant expression (McGlincy and Smith, 2008). It is possible that the homeostasis of splicing factors similarly requires a particularly controlled level of AS-NMD-induced repression, and, consequently, a large number of various factors fine-tuning that repression. These in turn may require extensive sequence elements as their platforms, as well as highly defined sequences within these platforms in order to maintain the (possibly overlapping) binding sites for all factors.

2 Aims of the study

This study was initiated in order to characterize in more detail factors involved in the recognition of U12-type introns. Unlike the 5'ss nucleotides of U2-type introns, the first three, highly conserved nucleotides (RUA) of U12-type introns do not base-pair to snRNAs, suggesting that unidentified protein components are responsible for recognizing this motif. This indicated mechanistic differences in intron recognition between the two spliceosomes. These differences were likely to be linked to the regulation of the U12-dependent spliceosome, and as such of great interest to deciphering its role in controlling the expression of genes with U12-type introns. In order to shed light on these matters, I aimed at finding and characterizing the protein component or components specifically interacting with this motif, and defining how they are regulated or could regulate the U12-dependent spliceosome. After the discovery of the U11-48K protein as the 5'ss-recognizing factor, it became obvious that the aberrantly spliced isoforms of its pre-mRNA, as well as those of the U11/U12-65K protein, could be diagnostic of feedback regulation of U11/U12 di-snRNP levels through a conserved sequence element present in both pre-mRNAs. The latter part of my project was therefore aimed at verifying the interaction of U11/U12 di-snRNP with the 5'ss-like motifs within this elements, as well as finding other splicing regulatory elements within the conserved sequence, and characterizing the effects of *trans*-acting factors binding to them.

3 Materials and methods

Methods used in this study are listed in Table 1. For detailed description of the methods, see the original publications.

Table 1. Methods used in this study.

Method	Article
Cell fractionation	I, II
Cell lines and culture	I, II, III
Far-Western overlays	II
Glycerol gradient centrifugation	I
Immunoprecipitation	I, II, III
In vitro splicing and spliceosome assembly assays	I, II
In vivo splicing block using oligonucleotides	II
Luciferase expression assays	II
NMD inhibition	II
Northern blotting	I, II, III
Protein-RNA crosslinking	I, III
Quantitative RT-PCR	I, II
Reporter plasmids	II, III
RNAi knockdown	I, II
RNA-RNA crosslinking	II, III
RNase H protection assays	I
RT-PCR	I, II, III
Sequence alignments	I, II, III
StrepII tag pulldown assays	I
Streptavidin pulldown assays	I, II
Translation inhibition	II
Western blotting	II
Yeast two-hybrid assays	I

4 Results and discussion

4.1 U11-48K protein recognizes the U12-type 5' splice site and is essential for U12-dependent splicing (I)

4.1.1 U11-48K specifically recognizes the U12-type 5' splice site during initial intron recognition

The main difference in the assembly of the two spliceosomes seems to be the initial intron recognition phase, which takes place in multiple steps in the major spliceosome, while minor introns are recognized by the preformed U11/U12 di-snRNP (chapter 1.5.5; Figure 4). The initial base-pairing interactions at the 5'ss are also different in the two spliceosomes. The consensus 5'ss sequence of major introns is complementary to U1 snRNA from position -2 to +6. In contrast, the first three nucleotides (RUA) of minor introns are not recognized by U11 snRNA (see Figure 1 in **I**). However, the almost universal conservation of the RUA motif (Sheth *et al.*, 2006), as well as the fact that the first two nucleotides are critical for splicing (Dietrich *et al.*, 2005), suggested that the motif is recognized through sequence-specific interactions with some other splicing factor. As the functions of most of the proteins specific to the U11/U12 di-snRNP (Will *et al.*, 2004) remained unknown, it appeared likely that one or several of them might be involved in 5'ss recognition.

To test this hypothesis, I performed *in vitro* splicing assays in HeLa nuclear extracts with the widely used *P120* splicing substrate containing a U12-type intron with AU-AC termini. As there was no prior experimental evidence for the significance of the third nucleotide of the U12-type 5'ss sequence (A+3), I initially tested spliceosome assembly and splicing catalysis of substrates mutated at the +3 position. I found that at least in the context of the *in vitro* assays, A+3 is essential for both spliceosome assembly and catalysis. An A+3G mutant was then used as a negative control in RNA-protein crosslinking assays. To detect proteins specifically binding to the 5' end of the intron, I placed a single photoactivatable 4-thio-uridyl residue followed by a radioactive phosphate group at the +2 position of the *P120* splicing substrate. Using the A+3G mutation and a mutation blocking U11 snRNA binding (CC+5+6GG) as negative controls, I discovered that a single protein crosslinks specifically to the wild type +2 position (see Figure 3 in **I** for substrate RNAs and crosslinked proteins). By performing immunoprecipitation with antibodies specific to the proteins of the U11/U12 di-snRNP, this protein was recognized as U11-48K (henceforth referred to as 48K). This interaction was also observed with a novel splicing substrate derived from the mouse *Vps16* gene, which contains a U12-type intron with GU-AG termini, verifying that the 48K interaction is not restricted to the AU-AC subtype of U12-type introns.

To address the generality of 48K functions, 48K was knocked down by RNAi in HeLa cells, which resulted in reduction of *48K* mRNA levels to ca. 10–15 % of the original level. This led to a significant cell growth arrest (70% reduction), comparable to that caused by knockdown of another essential splicing factor, Prp8 (90%). Thus, 48K has an essential function for cell

viability. I and my colleagues also observed reduced splicing of a number of U12-dependent introns, confirming that 48K has an essential and general function in U12-type intron recognition. In addition to reduced U12-dependent splicing, we observed activation of cryptic U2-type splice sites in the vicinity of U12-type splice sites. Importantly, this indicates that the splicing defect caused by 48K knockdown affects the intron recognition step, and not a later phase in the spliceosome assembly or catalysis, as such introns would already be committed to splicing by the U12-type spliceosome. This is also confirmed by my *in vitro* crosslinking analyses: The 48K-5'ss interaction precedes the formation of the stable, ATP-dependent A complex (see chapter 1.5.4), as it does not require ATP. It also precedes the formation of U12/BPS base-pairs, as it can form in the presence of an oligonucleotide blocking U12/BPS basepairing, and even in the absence of a BPS. However, 2–4-fold increase in 48K crosslinking was observed with substrates that were able to form U12/BPS basepairs. These results are consistent with previous results showing that while U11 basepairing to the 5'ss precedes U12/BPS interactions, U11 binding is enhanced by co-operative recognition of the BPS (Frilander and Steitz, 1999).

Such kinetic effects are also likely to be linked to differences in splicing efficiency of different U12-type introns, as reduced splicing after 48K knockdown was not observed for all U12-type introns. Similarly, depletion of major snRNPs has been observed to affect the splicing of specific introns, most likely caused by competition between strong and weak sites (see chapter 1.6.4.4). Kinetic competition may become particularly significant in the U12-dependent spliceosome due to co-operative recognition of the 5'ss and BPS, as under normal circumstances binding defects at one site could be compensated by enhanced binding at the other site. Indeed, the introns that were most severely affected after 48K knockdown generally had poor 5'ss and/or BP sequences (see Figure 5F in **I**), suggesting that their recognition has been compromised due to reduced affinity for the 5'ss. These results also support the notion that 48K functions by stabilizing 5'ss/U11 snRNA binding.

Due to our inability to express recombinant 48K in significant amounts, we were unable to study the interactions of 48K and the 5'ss in more detail *in vitro* and show how 48K stabilizes U11 binding, or whether the sequence specificity is conferred by 48K itself or by an additional splicing factor not detected by the crosslinking assay. Also, we were not able to show which domain of 48K is responsible for 5'ss recognition, although we did detect an evolutionarily conserved novel CHHC zinc finger (ZnF) motif in the N-terminal part of 48K, which we suggested as a possible RNA or protein recognition motif (see Supplementary Figure S1 in **I**). Fortunately, these questions were later addressed by nuclear magnetic resonance (NMR) studies of Tidow and co-workers. They showed that the 48K ZnF domain adopts a novel fold found in a few other RNA binding proteins (Andreeva and Tidow, 2008), and that this domain binds to short RNA oligonucleotides with a U12-type 5'ss sequence both in the presence and absence of an oligonucleotide mimicking the 5' end of U11 (Tidow *et al.*, 2009). 48K binding to the U11/5'ss helix was found to be independent of salt concentration, but binding to 5'ss RNA alone occurred only under low salt conditions. Combined with our results, it is thus

evident that 48K binds to the U11/5'ss helix during its formation or shortly thereafter, and also interacts with the unbase-paired nucleotides of the RUA motif, stabilizing the U11/5'ss helix.

As no RUA mutants were used in the study by Tidow *et al.* (2009), it still remains unclear whether 48K alone recognizes the RUA motif in a sequence-specific manner. SR proteins have been shown to interact directly with the U12-type 5'ss (Shen and Green, 2007), and might be involved in guiding or stabilizing the 48K-5'ss interaction. Similarly, other proteins in the U11/U12 di-snRNP might affect 5'ss recognition (see chapter 1.5.5). It was previously suggested that the U11/U12-20K protein (hereafter 20K) might function in 5'ss recognition analogously to the U1C protein. However, we were unable to detect 20K in our crosslinking assays (**I**), and knockdown of 20K had minimal effects on the feedback regulation of 48K and 65K proteins (Figure 5C in **II;** see chapter 4.2), suggesting it is not involved in 5'ss recognition. While 48K does not share sequence similarity with U1C, there are certain similarities between the two proteins that suggest that they may nonetheless function in a similar manner in 5'ss recognition in the two spliceosomes, as discussed in more detail in the following chapter.

4.1.2 Parallels between 5' splice site recognition and cross-intron interactions in the two spliceosomes

U1C has been implicated in stabilizing the U1/5'ss helix through a CCHH ZnF domain (Muto *et al.*, 2004; Pomeranz Krummel *et al.*, 2009). Our results, as well as those of Tidow *et al.* (2009) support a similar role for 48K. Somewhat controversially, it has been suggested that U1C can also recognize the 5'ss in the absence of U1/5'ss base-pairing both in humans and in baker's yeast (Rossi *et al.*, 1996; Du and Rosbash, 2002). These results need not be contradictory, as our results and those of Tidow *et al.* (2009) suggest that the 48K can also interact with ssRNA (the RUA motif) and the U11/5'ss helix. However, despite a similar overall fold, the ZnF motifs of 48K and U1C are different. The NMR structure of the free 48K ZnF domain has two β-strands followed by an α-helix, (interrupted by one π-turn; Tidow *et al.*, 2009). The U1C ZnF is composed of two β-strands followed by three α-helices, A, B and C, the latter two of which are reorganized in the crystal structure of U1 snRNP to yield one long α-helix B (Muto *et al.*, 2004; Pomeranz Krummel *et al.*, 2009), which is analogous to the α-helix in the 48K ZnF. In 48K, it is the N-terminal part of the α-helix and the preceding β-turn that have been implicated in RNA binding based on chemical shifts (Tidow *et al.*, 2009). In the structure of the U1 snRNP, the 5' ends of two U1 snRNAs base-pair to one another in a way that has been suggested to be similar to U1/5'ss base-pairing, and in this structure U1C contacts the RNA helix using the α-helix and β-turn analogous to those of 48K. Interestingly, the short helix A of U1C ZnF, which has no counterpart in 48K, seems to bind the minor groove of the RNA helix at the site corresponding to the first two nucleotides of the intron (Pomeranz Krummel *et al.*, 2009), which do not participate in base-pairing in the minor spliceosome. Thus, despite their dissimilar sequences, 48K and U1C may have evolved similar ZnF domains in parallel, and may use similar strategies in recognizing their targets, but are specifically equipped to deal with the differences in their RNA environment. It is important to

notice, however, that no structure of the 48K ZnF domain in complex with RNA has been published, and this might prove to be quite different from that of the solution structure or the U1C-RNA complex.

To study how 48K associates with the U11/U12 di-snRNP, we performed yeast two-hybrid assays with 48K and other U11/U12 proteins. The only observed interaction was with U11-59K protein (henceforth 59K). This result was further verified by far-Western overlays and pulldown assays, which also helped in localizing the interacting domains to the central part of 59K and the N-terminal domain of 48K. This N-terminal (amino acids 1–92) part of 48K also contains the ZnF domain interacting with the 5'ss. Interestingly, N-terminal fragment containing the ZnF domain of U1C (aa 1–45), can also associate with the U1 snRNP on its own (Nelissen *et al.*, 1991; Muto *et al.*, 2004). It is possible that this proximity of the 5'ss-binding and protein interaction interfaces serves a purpose in bringing the 5'ss close to other functional parts in the spliceosome. Indeed, both U12 and U2 are close to the 5'ss during the earliest phases of intron recognition in their respective prespliceosomes (Frilander and Meng, 2005; Dönmez *et al.*, 2007). In the major spliceosome, the 5'ss and BPS are brought together by non-snRNP proteins interacting with U1 and U2 (see chapter 1.5.5). Our observation of 48K interaction with 59K provides insights into how the 5'ss and BPS are bridged in the minor prespliceosome by internal components of the U11/U12 di-snRNP. Previous studies have indicated that 59K resides at the interface of U11 and U12 snRNPs and interacts with the U11/U12-65K protein, which directly binds to the 3' end of U12 snRNA (Benecke *et al.*, 2005). Thus, 48K is part of a molecular chain that links the two catalytically functional intron sites during prespliceosome assembly (see Figure 8 in **I** for a model).

4.1.3 U11-48K contributes to U11/U12 di-snRNP formation or stability

In order to see how these protein-protein interactions of 48K affect the U11/U12 di-snRNP structure, the fraction of intact di-snRNPs in nuclear extracts from 48K knockdown cells was assayed by glycerol gradient fractionation and by co-immunoprecipitation with an antibody specific to the U12-associated SF3b155 protein. The results indicated a ca. 50% decrease in the level of mature 18S U11/U12 di-snRNPs in the knockdown extracts vs. control extracts. This suggests that 48K does contribute to either the formation or stability of the di-snRNP. This reduction also likely contributes to the decrease observed in U12-dependent splicing observed in the knockdown cells. Due to the low levels of U12-type factors in the cells, and the relatively weak antibodies at our disposal, we were not able to ascertain whether the 50% reduction in di-snRNP levels directly reflects 48K protein levels, or whether 48K levels are even lower. However, it is unlikely that a 50% reduction in di-snRNP levels alone would cause the observed effects, as snRNPs are present in a large excess in the cell, and even residual levels (< 20%) of the most scarce U12-type snRNA, U4atac, are enough to support splicing at wild type levels (Pessa *et al.*, 2006). Therefore, 48K protein levels are likely to be reduced by more than 50%, as suggested by the reduction in the mRNA levels, and the absence of 48K likely affects the splicing of U12-type intron mainly by directly hindering proper 5'ss recognition.

4.2 The level of U11/U12 di-snRNPs is regulated through a conserved feedback mechanism (II)

4.2.1 Highly conserved sequence elements in the genes encoding U11-48K and U11/U12-65K

The discovery of 48K as a key factor in 5'ss recognition sparked an interest in its regulation, as regulating its function could directly affect the activity of the U12-dependent spliceosome. An investigation into the transcripts expressed from the gene encoding 48K, *SNRNP48*, revealed that it is spliced at alternative sites which are expected to give rise to PTCs, and these events take place in highly conserved intronic sequence elements (see Figure 1 in **II**), similar to the AS-NMD-mediated feedback regulation of SR proteins and hnRNPs (see chapter 1.8.2). The expressed mRNAs contain two PTC-containing isoforms. In one of them, an alternative U2-type 3'ss is activated in intron 4, leading to inclusion of a PTC residing in the intronic region between that 3'ss and exon 5. The second isoform is spliced at the same 3'ss, but also at a 5'ss only 8 nt downstream, causing inclusion of a small exon (exon 4i), resulting in a frameshift and PTC inclusion in exon 5. Most intriguingly, these alternative U2-type splice sites are situated upstream of an element containing a tandem repeat of two perfect U12-type 5'ss motifs. The U12-type sites were not found to be active splice sites in any of the expressed mRNAs, suggesting they perform a regulatory function. Thus, it appeared likely that the 5'ss-like motifs could act as sensors for functional U11 snRNP or U11/U12 di-snRNP levels, such that their binding could activate splicing at the upstream U2-type splice sites through exon definition interactions, eventually resulting in destruction of the transcripts by NMD. Thus, we refer to the element formed by the two U12-type 5'ss-like motifs as the U11 snRNP-binding splicing enhancer, or USSE.

The larger, ca. 110 bp element containing the USSE is almost completely conserved in mammals, more so than some of the exons in the *SNRNP48* gene. In more distantly related organisms, such as fish and insect species, the conservation is mainly limited to the USSE itself and the upstream U2-type splice sites, especially the 3'ss, suggesting that these sequences comprise the core elements of the feedback regulation. Strikingly, the USSE is also found in the plant genes coding for 48K. However, in plants it is located in the 3' UTR of the genes, and is associated with an upstream U2-type 3'ss. Our RT-PCR result show that in *Populus trichocarpa* and *Arabidopsis thaliana* the transcripts are indeed spliced alternatively at the 3'ss upstream of the USSE to produce mRNAs with 3' UTRs of differing length. The apparent similarity in the structures of the USSE and the associated elements suggest that this mode of regulation has had a significant function already in the common ancestor of plants and animals. Furthermore, the differences in the location of the element suggest that it is the mechanism of feedback regulation that is conserved, not the site of regulation, as also suggested for AS-NMD-mediated regulation of SR proteins (see chapter 1.8.2).

A bioinformatic search for conserved USSE elements in other genes revealed a similar element only in one gene, *RNPC3*, which codes for another U11/U12-specific protein, 65K. The *65K* USSE is located in the 3' UTR, similar to the plant *48K*, and alternative U2-type activation upstream of the USSE also leads to expression of transcripts with longer 3' UTRs. The *65K* USSE is well conserved in vertebrates, but we failed to find it in other eukaryotes, suggesting the 65K feedback regulation may have arisen independently in the ancestor of vertebrates, possibly by transfer of the USSE element from the *SNRNP48* gene. As 48K and 65K are both components of the U11/U12 di-snRNP, this may have improved the regulation of the activity of the U12-dependent spliceosome.

4.2.2 The USSE directs alternative splicing and mRNA destabilization

To verify the function of the USSE in the regulation of 48K and 65K, human embryonic kidney cells (HEK293) were treated with oligonucleotides blocking the binding of U11/U12 di-snRNPs to the USSE. As expected, blocking the USSE resulted in decreased inclusion of the nonsense exon or in the absence of the long 3' UTR isoform in the endogenous mRNAs coding for 48K or 65K, respectively. The same results were obtained with reporter plasmids containing mutations in one or both of the 5'ss-like motifs in the USSE, suggesting both of the motifs are required for splicing enhancement by USSE. Most importantly, the effect of USSE mutations was rescued by expression of a U11 snRNA with compensatory mutations, indicating that U11 snRNA base-pairing to the USSE is essential for its activity.

The inclusion of PTCs suggests that *48K* is regulated through AS-NMD, similar to SR proteins. This hypothesis was tested by inhibition of translation by cycloheximide, and by inhibiting NMD by expressing a dominant negative form of Upf1. Both approaches resulted in stabilization of nonsense exon-containing isoforms of *48K* mRNA, indicating that alternative splicing activated by the USSE results in destruction of the mRNA by NMD. In contrast, *65K* transcripts were unaffected by inhibition of NMD. It was found that the *65K* transcripts with long 3' UTRs were specifically destabilized in the cytoplasm, suggesting that a cytoplasmic process other than NMD degrades the *65K* transcripts.

Knockdown of 48K also resulted in a significant increase in the levels of *65K* mRNA, indicating that it is actively down-regulated by the USSE. Similarly, knockdown of 35K led to increase in both *48K* and *65K* mRNA levels. Interestingly, 35K interacts with SR proteins, similar to U1-70K (see chapter 1.5.5), which functions in exon-definition interactions in the major spliceosome. It is therefore possible that 35K mediates the activation of the upstream splice sites, potentially through SR proteins. It is also notable that while previous results have shown enhanced U12-dependent splicing through exon-definition interactions with U2-type factors (see chapter 1.6.2), our results show for the first time that components of the U12-dependent spliceosome can also activate U2-dependent splicing.

4.3 Activation of AS-NMD in the *U11-48K* transcript is regulated by multiple factors (II, III)

4.3.1 The USSE binds U11/U12 di-snRNPs (II)

To study the splicing activation by the USSE, I chose to characterize the components interacting with the *48K* USSE in more detail *in vitro*. Native gel analyses indicated that two spliceosome-like complexes assembled in HeLa nuclear extracts on short RNAs containing the USSE. I then performed pulldown experiments with short biotinylated RNAs to identify snRNAs binding to the USSE, as well as psoralen crosslinking experiments to detect RNA-RNA interactions. U11 snRNA binding to the USSE was observed with both methods. U12 snRNAs were also found to be present in the USSE-binding complex, but no RNA-RNA interactions were detected. U11 crosslinking and U11 and U12 binding was completely lost with substrates containing A+3G or CC+5+6GG mutations in both 5'ss-like motifs, indicating that recognition by both U11 snRNA and 48K is required, similar to recognition of a *bona fide* U12-type 5'ss. Intermediate levels of U11 and U12 snRNA binding in pulldown experiments was observed with substrates mutated only at one 5'ss-like motif. Similarly, U11 crosslinks formed to both motifs independently. This suggests that the USSE motifs bind U11/U12 di-snRNPs independently (at least *in vitro*), and that both motifs are occupied in wild type substrates. This is consistent with the results indicating that both motifs are required for splicing enhancement by USSE *in vivo*. Simultaneous binding may stabilize the complex, or the two di-snRNPs may form a platform that is required for interactions with factors mediating the activation of upstream U2-type splice sites.

However, it is not clear whether the activation requires complete U11/U12 di-snRNPs, or whether U11 snRNP alone is sufficient, similar to U11 function in regulating the splicing of the RSV *gag* transcript (McNally *et al.*, 2004). My pulldown experiments suggest that each RNA substrate binds similar amounts of U11 and U12 snRNAs, suggesting the USSE is recognized by U11/U12 di-snRNPs, although the results lack statistical significance. U12 does not appear to participate in RNA binding in the USSE: No U12 crosslinks were formed, and RNase H-mediated digestion of U12 snRNA had no effect on U11 binding. However, the fragments of digested U12 were pulled down with the complex, indicating that the U12 moiety of the di-snRNP did not disintegrate (data not shown). Thus, it may be that the U12 snRNA is not necessary for USSE recognition, but it or the protein factors associated with it may be required for interactions with other factors. As both the 48K and 65K are targets of the feedback regulation, it is likely that 65K, and thus the U12 snRNP, is required for the activation of the USSE. Indeed, knocking down 65K by RNAi has proven to be particularly difficult (**II**; Benecke *et al.*, 2005), likely due to the feedback regulation through the USSE, suggesting that the process requires 65K.

4.3.2 AS-NMD is suppressed by hnRNPF/H and U1 snRNP (II, III)

The USSE itself constitutes only a small part of the conserved sequence element, and other conserved features also likely affect the feedback regulation. One of the observations arising from our *in vivo* studies is that exon 4i is included very weakly in the absence of the USSE, despite the strong splice sites surrounding it, suggesting the presence of inhibitory elements. A notable feature in exon 4i is a run of four guanosines, which is among the most conserved elements in mammalian as well as fish species (see Figure 1 in **III**). The sequence and location of this element suggested that it might function as an ESS by binding hnRNPF/H proteins. To test this hypothesis, I performed protein-RNA crosslinking experiments with longer RNA substrates, containing the USSE as well as exon 4i and part of the preceding intron. By immunoprecipitation, I identified a strong crosslink to hnRNPH1, as well as a weaker one to hnRNPF, both of which occured with wild type substrates but not with substrates with G-run mutations. A relatively strong crosslink also formed to a 35 kDa protein, possibly hnRNPH3, suggesting that several hnRNPF/H proteins may recognize the G-run in exon 4i. However, I was unable to identify this protein by immunoprecipitation. Crosslinks to hnRNPA1 and phospho-SR proteins were also detected, but these were not specific to the wild type sequence.

Binding of U1 to the exon 4i 5'ss was verified by pulldown experiments and psoralen crosslinking using the same substrates (see Figures 2 and 3 in **II**, and Figure 2 in **III**). Due to the proximity of the 5'ss and the G-run, it seemed likely that U1 and hnRNPH1 (or F) could be competing for binding. Competition has previously been observed for similarly overlapping sites (Buratti *et al.*, 2004), and ESS-bound hnRNPH1 has been shown to be a potent suppressor of strong 5' splice sites (Xiao *et al.*, 2009), such as that of exon 4i. To test this, U1 binding was inhibited with a 2'-*O*-methyl oligonucleotide, with an increase in hnRNPH1 crosslinking by ca. 50%, suggesting that U1 and hnRNPH1 are competing for binding. However, other protein crosslinks were also intensified upon U1 blockage, and the reciprocal experiment of mutating the G-run had no significant effect on U1/5'ss crosslinks.

To better define the effect of hnRNPF/H on exon 4i inclusion and U1 binding, I transfected HEK293 cells with reporter plasmids containing various mutations in the G-run and the exon 4i 5'ss, as well as in the USSE. The 5'ss mutants cannot be spliced at the 5'ss, but they are still spliced at the exon 4i 3'ss, resulting in the inclusion of the downstream intron, similar to the longer endogenous NMD-inducing mRNA. Thus, for the purposes of this study, exon 4i inclusion was defined as any event resulting in inclusion of the exon, regardless of whether the downstream intron was included or not. The results indicate that the G-run does indeed function as an ESS, and that U1 inhibits hnRNPF/H binding. However, the results are not consistent with a simple model, in which hnRNPF/H represses and U1 activates exon inclusion. Instead, splicing at the exon 4i 3'ss occurs most efficiently when both the G-run and 5'ss are mutated, suggesting that both hnRNPF/H and U1 suppress the 3'ss. However, hnRNPF/H appears to be a more dominant repressor, as in the absence of U1 and USSE it can completely block splicing to exon 4i, while transcripts with mutations in the hnRNPF/H binding site and USSE still allow exon 4i inclusion at low levels.

It may appear contradictory that the U1 snRNP bound at the exon 4i 5'ss can function in repressing the 3'ss, while also activating the 5'ss itself, at least for a subset of transcripts. It is likely that these two functions take place at different stages: The 5'ss of exon 4i is too close to the 3'ss to allow U1 to engage in exon definition interactions, as previously observed for other short exons (Sterner and Berget, 1993; Hwang and Cohen, 1997; Lewandowska *et al.*, 2005), and U1 binding at the 5'ss could inhibit 3'ss recognition sterically. If the 3'ss is activated by the USSE, however, U1 snRNP can then function in defining the 5'ss of the downstream intron. Therefore, our results suggest that the main target of splicing activation by USSE is the 3'ss. Such a strict preferential order for upstream vs. downstream intron removal was also observed for an alternative exon in the *ATM* transcript (Lewandowska *et al.*, 2005). The primacy of the 3'ss for the feedback regulation is also supported by the fact that the 5'ss activation is not necessary for PTC inclusion in the naturally occurring isoforms.

Although my *in vivo* results clearly demonstrate the USSE-dependent activation of the 3'ss, I failed to detect recruitment of 3'ss-recognizing factors in my *in vitro* assays. No U2/BPS crosslinks were observed, and only a minimal increase in U2 binding was observed with the long vs. short substrates in the biotin pulldown assays (**II**). The precise location of the BPS in the intron upstream of exon 4i has not been mapped, and there are several putative U2-type BPSs, all of which are highly degenerate. Thus, the proper recognition of the BPS may require additional enhancing elements not present in our *in vitro* RNA substrate.

The conservation of the 110 bp element also suggests that there are a number of other elements and factors that may regulate the activity of the USSE and exon 4i inclusion, and a number of putative regulatory elements were discovered computationally (see Supplemental Figure S3 in **II**). The presence of enhancer elements is also suggested by the activation of cryptic splice sites in reporters with mutations blocking the binding of U1 and U11/U12 di-snRNPs to the 5'ss or the USSE, respectively. Such enhancers may function to recruit U1 and/or U11/U12 under normal circumstances, while activating binding of U1 to cryptic sites when the natural sites are inactivated. Also, the TAATT sequence preceding the G-run may also be important, as highly similar sequences at the 5' end of exons were found to be significant for hnRNPH1-mediated exon silencing (LeFave *et al.*, 2011). However, our data show that simultaneous mutation of the G-run, 5'ss and the USSE results in inclusion levels similar to the wild type reporter, indicating that the USSE is not strictly required under these circumstances. Thus, it appears that the most crucial interactions regulating exon 4i inclusion are those between the U11/U12 di-snRNPS bound at the USSE and hnRNPF/H proteins and U1 snRNP bound in and adjacent to exon 4i.

The interplay between inhibitory and activating functions suggests a requirement for finely controlled and limited feedback regulation, which has been proposed to be essential for maintaining constant protein levels (see chapter 1.8.3). Indeed, the USSE itself also appears to be a highly sensitive element for monitoring functional U11 or U11/U12 di-snRNP levels. It forms complexes with snRNPs more stably than an isolated 5'ss, but less stably than authentic

U12-type introns. Consistent with this, a knockdown of 48K completely inhibits activation of the 3'ss upstream of the 65K USSE (see Figure 5D in **II**), but has only minor effects on the splicing of U12-type introns (Figure 5 in **I**). Additional regulation may be provided by varying the levels or activity of hnRNPF/H in different cell lines. It is also possible that the sequences surrounding exon 4i form a CERES-like complex splicing regulatory element (see chapter 1.6.4.5), with multiple factors competing for binding, as suggested by our crosslinking results. Further studies are required to decipher the precise nature of these interactions.

4.4 Evolutionary implications of the USSE

The presence of a USSE in the genes coding for 48K in animals as well as plants suggests that this mode of regulation may have existed in the last common ancestor of eukaryotes (LECA), and thus probably represents the oldest known SRE. It also provides further support for a well-established U12-dependent spliceosome in LECA, as opposed to recent gain in independent lineages (see chapters 1.2.2.2 and 1.3.3). Furthermore, it is possible that the activity of the USSE is related to the loss of U12-type introns in many lineages. Inactivation of the USSE would likely result in disruption of minor spliceosome homeostasis, possibly leading to the loss of its rate-limiting function. This in turn would remove the selection pressure against mutations that result in conversion of U12-type introns to U2-type introns, ultimately resulting in loss of all U12-type introns and the U12-dependent spliceosome.

Although highly conserved elements regulating AS-NMD in splicing factor genes are common (see chapter 1.8), the extreme conservation of the *48K* regulatory element in mammals is puzzling, considering the fact that in other organisms the conservation is restricted to the USSE itself and the upstream 3'ss (see Figure 1 in **II**). Naturally, the fact that mammals are the most extensively studied taxon may cause a bias in our results. Also, mammals are a relatively recent taxon, and may not have had as much time to diverge to the same extent as some of the other taxa studied here. Interestingly, recent results indicate that, unlike most evolutionary lineages studied, the lineage leading from the LECA to mammals shows no signs of extensive intron loss (Csuros *et al.,* 2011). It is thus possible that regulation of gene expression through splicing has had a more significant role in the past and present members of this particular lineage, accounting for the wide-spread use of feedback mechanisms controlling the levels of splicing activators and inhibitors, as well as constitutive splicing factors in both the major and minor spliceosome.

5 Concluding remarks

The recognition of the U12 and U2-type introns differs with respect to the specific components involved and the order in which they interact with the intronic sequences. As the later steps of spliceosome assembly are highly similar in both cases, it is likely that the activity of a specific spliceosome is regulated through the components involved in intron recognition. In this study, I identified the U11-48K protein as an essential splicing factor that specifically recognizes U12-type 5' splice sites. The levels of U11-48K, as well as those of U11/U12-65K, are regulated through a feedback mechanism involving a highly conserved sequence element, USSE, which directs alternative splicing and nonsense-mediated decay. The components controlling this event include not only U11/U12 di-snRNPs, but also U1 snRNP and hnRNPF/H proteins. These results suggests that the activity of the minor spliceosome is controlled by the levels of U11/U12 di-snRNPs, which are fine-tuned by the combined actions of multiple splicing factors regulating the expression of U11-48K and U11/U12-65K.

The conservation of the regulatory mechanism in plants as well as in animals highlights the importance and early origin of feedback regulation of splicing factors through alternative splicing. Furthermore, the distribution of USSE in various eukaryotes suggests that the USSE may be essential for the regulation of the U12-dependent spliceosome and its function as a rate-limiting process in gene expression. Thus, the loss of the USSE may predispose the organisms in question to increased conversion of U12-type introns into U2-type introns, ultimately resulting in the loss of the U12-dependent spliceosome.

6 Acknowledgements

This study has been carried out at the Institute of Biotechnology of the University of Helsinki. I would like to thank the Institute for providing the excellent research infrastructure that has been essential for carrying out this work. Equally important, and highly appreciated, has been the funding provided by the Helsinki Graduate Program in Biotechnology and Molecular Biology.

Despite the excellent infrastructure, this study would never have succeeded without the guidance of my supervisor Mikko Frilander. He has given plenty of room for free thought, yet has also always been available for advice on any aspect of my work, for which I am most grateful. I am also indebted to the late Xiaojuan Meng, who tirelessly supervised me in the early stages of my work, and whose thoroughness and passion has been an inspiration to me. I have had the privilege of enjoying the companionship of the past and present members of the splicing lab: Heli, Lilli, Kaisa, Elina, Jouni, Annukka, Jens, James, Trang, Visa, Bhupendra and Ali, thanks to all of you as well as all the visitors who have contributed to the jolly atmosphere in our lab. In particular, I would like to thank Lilli and Annukka for all the help with the experiments, and Elina, Jens and Bhupendra for all the great ideas we have made reality together, as well as for all those times when you spent hours searching for the relevant samples for my experiments.

I am honoured and delighted to have professor Mark McNally as my opponent, and professor Kari Keinänen as the custos in my defense. I am grateful to Minna Poranen and professor Anders Virtanen for reviewing my thesis and providing me with insightful feedback. Anders has also been a member of my follow-up group along with professor Pekka Lappalainen, and I thank them for the much needed advice and encouragement I have received from them over the years.

My first two publications were the result of a very fruitful collaboration with Cindy Will, Michael Grote and professor Reinhard Lührmann from the Max Planck Institute of Biophysical Chemistry in Göttingen, as well as with Janne Ravantti from the Department of Biosciences here in Helsinki, and I am very grateful to them for all the hard work put into this joint effort. In particular, I would like to thank Cindy for her insightful and detailed comments on my results, as well as for her extensive advice on scientific writing. From her, I have truly learned a great deal about being both cautious and precise when interpreting my results.

I thank professor Tomi Mäkelä, as well as his predecessor, Mart Saarma, for running the smooth operation that is known as the Institute of Biotechnology, and helping to create such a pleasant working environment. Similarly essential to that have been interactions with the neighbouring research groups, headed by Harri Savilahti, Elina Ikonen, Tapio Heino, Oscar Puig, Osamu Shimmi and Tomi Mäkelä. Therefore, I wish thank all of our past and present neighbours, and particularly Lotta and Danielle, for their help and the refreshing interactions outside the realm of RNA splicing. The interdisciplinary interactions have also been fostered

by the Institute's research programmes, which have brought together researchers from various labs for meetings, retreats and scientific collaboration. The Developmental Biology Research Program and the Genome Biology Research Program have been particularly important for our group, and I thank their directors, professors Irma Thesleff and Ykä Helariutta for co-ordinating the programmes, respectively, as well as all the other people involved.

I'm very grateful to Lotta, Päivi and Sarah for being my peer support group, thanks for being there in moments of deepest despair as well as exuberant joy! Thanks to all of my friends for sharing all those special little every-day moments, especially to Tomi, Elina, Liisa and Jaakko for constantly reminding me that there indeed is life outside the university.

Olen kiitollinen vanhemmilleni, ja erityisesti äidilleni kannustuksesta ja niistä lapsuuden kokemuksista, jotka jo varhain herättivät minussa halun ilmaista itseäni kirjallisesti. Tämä kirja edustaa toivoakseni kunniakkaasti sen jatkumon viimeisintä, mutta ei viimeistä ilmentymää. Haluan lisäksi erityisesti kiittää Anja-mummoa, joka on aina kannustanut kaikkia lapsukaisiaan hankkimaan haluamansa koulutuksen ja tavoittelemaan unelmiaan.

Helsinki, December 2011

Janne

7 References

Abelson, J. (2008). Is the spliceosome a ribonucleoprotein enzyme? *Nat Struct Mol Biol* 15: 1235-1237.

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

Alt, F.W., Bothwell, A.L., Knapp, M., Siden, E., Mather, E., Koshland, M., and Baltimore, D. (1980). Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends. *Cell* 20: 293-301.

Andreeva, A., and Tidow, H. (2008). A novel CHHC Zn-finger domain found in spliceosomal proteins and tRNA modifying enzymes. *Bioinformatics* 24: 2277-2280.

Auweter, S.D., Oberstrass, F.C., and Allain, F.H. (2006). Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res* 34: 4943-4959.

Barash, Y., Calarco, J.A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B.J., and Frey, B.J. (2010). Deciphering the splicing code. *Nature* 465: 53-59.

Bartschat, S., and Samuelsson, T. (2010). U12 type introns were lost at multiple occasions during evolution. *BMC Genomics* 11: 106.

Basu, M.K., Makalowski, W., Rogozin, I.B., and Koonin, E.V. (2008a). U12 intron positions are more strongly conserved between animals and plants than U2 intron positions. *Biol Direct* 3: 19.

Basu, M.K., Rogozin, I.B., and Koonin, E.V. (2008b). Primordial spliceosomal introns were probably U2-type. *Trends Genet* 24: 525-528.

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

Behrens, S.E., and Lührmann, 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.

Berget, S.M. (1995). Exon recognition in vertebrate splicing. *J Biol Chem* 270: 2411-2414.

Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A* 74: 3171-3175.

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 Lührmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. *Nature* 452: 846-850.

Beyer, A.L., and Osheim, Y.N. (1991). Visualization of RNA transcription and processing. *Semin Cell Biol* 2: 131-140.

Bindereif, A., and Green, M.R. (1987). An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. *EMBO J* 6: 2415-2424.

Bird, G., Zorio, D.A., and Bentley, D.L. (2004). RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. *Mol Cell Biol* 24: 8963-8969.

Björk, P., Jin, S., Zhao, J., Singh, O.P., Persson, J.O., Hellman, U., and Wieslander, L. (2009). Specific combinations of SR proteins associate with single pre-messenger RNAs in vivo and contribute different functions. *J Cell Biol* 184: 555-568.

Black, D.L., Chabot, B., and Steitz, J.A. (1985). U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. *Cell* 42: 737-750.

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.

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

Boelens, W.C., Jansen, E.J., van Venrooij, W.J., Stripecke, R., Mattaj, I.W., and Gunderson, S.I. (1993). The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell* 72: 881-892.

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.

Bonnal, S., Martinez, C., Förch, P., Bachi, A., Wilm, M., and Valcárcel, J. (2008). RBM5/Luca-15/H37 regulates Fas alternative splice site pairing after exon definition. *Mol Cell* 32: 81-95.

Bono, F., and Gehring, N.H. (2011). Assembly, disassembly and recycling: the dynamics of exon junction complexes. *RNA Biology* 8: 24-29.

Borah, S., Wong, A.C., and Steitz, J.A. (2009). Drosophila hnRNP A1 homologs Hrp36/Hrp38 enhance U2-type versus U12-type splicing to regulate alternative splicing of the prospero twintron. *Proc Natl Acad Sci U S A* 106: 2577-2582.

Boukis, L.A., Liu, N., Furuyama, S., and Bruzik, J.P. (2004). Ser/Arg-rich protein-mediated communication between U1 and U2 small nuclear ribonucleoprotein particles. *J Biol Chem* 279: 29647-29653.

Breitbart, R.E., Andreadis, A., and Nadal-Ginard, B. (1987). Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Annu Rev Biochem* 56: 467-495.

Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, H., and Lührmann, 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.

Brock, J.E., Dietrich, R.C., and Padgett, R.A. (2008). Mutational analysis of the U12-dependent branch site consensus sequence. *RNA* 14: 2430-2439.

Brosi, R., Gröning, K., Behrens, S.E., Lührmann, R., and Krämer, A. (1993). Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. *Science* 262: 102-105.

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

Buratti, E., Baralle, M., De Conti, L., Baralle, D., Romano, M., Ayala, Y.M., and Baralle, F.E. (2004). hnRNP H binding at the 5' splice site correlates with the pathological effect of two intronic mutations in the NF-1 and TSHβ genes. *Nucleic Acids Res* 32: 4224-4236.

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., Padgett, R.A., and Sharp, P.A. (1998). Evolutionary fates and origins of U12-type introns. *Molecular Cell* 2: 773-785.

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

Cao, W., Jamison, S.F., and Garcia-Blanco, M.A. (1997). Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro. *RNA* 3: 1456-1467.

Caputi, M., and Zahler, A.M. (2001). Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. *J Biol Chem* 276: 43850-43859.

Carrillo Oesterreich, F., Bieberstein, N., and Neugebauer, K.M. (2011). Pause locally, splice globally. *Trends Cell Biol* 21: 328-335.

Casanova, E., Garate, C., Ovalle, S., Calvo, P., and Chinchetru, M.A. (1996). Identification of four splice variants of the mouse stress-activated protein kinase JNK/SAPK α-isoform. *Neuroreport* 7: 1320-1324.

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

Chan, S.P., Kao, D.I., Tsai, W.Y., and Cheng, S.C. (2003). The Prp19p-associated complex in spliceosome activation. *Science* 302: 279-282.
Chang, W.C., Chen, Y.C., Lee, K.M., and Tarn, W.Y. (2007a). Alternative splicing and bioinformatic analysis of human U12-type introns. *Nucleic Acids Res* 35: 1833-1841.
Chang, Y.F., Imam, J.S., and Wilkinson, M.F. (2007b). The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem* 76: 51-74.
Chasin, L.A. (2007). Searching for splicing motifs. *Adv Exp Med Biol* 623: 85-106.
Cheng, S.C., and Abelson, J. (1987). Spliceosome assembly in yeast. *Genes Dev* 1: 1014-1027.
Cho, S., Hoang, A., Sinha, R., Zhong, X.Y., Fu, X.D., Krainer, A.R., and Ghosh, G. (2011). Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proc Natl Acad Sci U S A* 108: 8233-8238.
Chou, M.Y., Underwood, J.G., Nikolic, J., Luu, M.H., and Black, D.L. (2000). Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing. *Mol Cell* 5: 949-957.
Chow, L.T., Gelinas, R.E., Broker, T.R., and Roberts, R.J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 12: 1-8.
Cloutier, P., Toutant, J., Shkreta, L., Goekjian, S., Revil, T., and Chabot, B. (2008). Antagonistic effects of the SRp30c protein and cryptic 5' splice sites on the alternative splicing of the apoptotic regulator Bcl-x. *J Biol Chem* 283: 21315-21324.
Coles, J.L., Hallegger, M., and Smith, C.W. (2009). A nonsense exon in the Tpm1 gene is silenced by hnRNP H and F. *RNA* 15: 33-43.
Corvelo, A., Hallegger, M., Smith, C.W., and Eyras, E. (2010). Genome-wide association between branch point properties and alternative splicing. *PLoS Comput Biol* 6: e1001016.
Crawford, J.B., and Patton, J.G. (2006). Activation of α-tropomyosin exon 2 is regulated by the SR protein 9G8 and heterogeneous nuclear ribonucleoproteins H and F. *Mol Cell Biol* 26: 8791-8802.
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.
Cuccurese, M., Russo, G., Russo, A., and Pietropaolo, C. (2005). Alternative splicing and nonsense-mediated mRNA decay regulate mammalian ribosomal gene expression. *Nucleic Acids Res* 33: 5965-5977.
Damgaard, C.K., Kahns, S., Lykke-Andersen, S., Nielsen, A.L., Jensen, T.H., and Kjems, J. (2008). A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Mol Cell* 29: 271-278.
Darnell, J.E., Jr. (1978). Implications of RNA-RNA splicing in evolution of eukaryotic cells. *Science* 202: 1257-1260.
Das, R., Dufu, K., Romney, B., Feldt, M., Elenko, M., and Reed, R. (2006). Functional coupling of RNAP II transcription to spliceosome assembly. *Genes Dev* 20: 1100-1109.
Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Gygi, S.P., and Reed, R. (2007). SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol Cell* 26: 867-881.
Deutsch, M., and Long, M. (1999). Intron-exon structures of eukaryotic model organisms. *Nucleic Acids Res* 27: 3219-3228.
Dhir, A., Buratti, E., van Santen, M.A., Lührmann, 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., Fuller, J.D., and Padgett, R.A. (2005). A mutational analysis of U12-dependent splice site dinucleotides. *RNA* 11: 1430-1440.
Dietrich, R.C., Incorvaia, R., and Padgett, R.A. (1997). Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns. *Mol Cell* 1: 151-160.
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.
Dönmez, G., Hartmuth, K., Kastner, B., Will, C.L., and Lührmann, R. (2007). The 5' end of U2 snRNA is in close proximity to U1 and functional sites of the pre-mRNA in early spliceosomal complexes. *Mol Cell* 25: 399-411.
Doolittle, W.F. (1978). Genes in pieces - were they ever together. *Nature* 272: 581-582.

Du, H., and Rosbash, M. (2002). The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature* 419: 86-90.
Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R., and Hood, L. (1980). Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. *Cell* 20: 313-319.
Eberle, A.B., Stalder, L., Mathys, H., Orozco, R.Z., and Muhlemann, O. (2008). Posttranscriptional gene regulation by spatial rearrangement of the 3' untranslated region. *PLoS Biol* 6: e92.
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.
Faustino, N.A., and Cooper, T.A. (2003). Pre-mRNA splicing and human disease. *Genes Dev* 17: 419-437.
Fedorova, O., and Zingler, N. (2007). Group II introns: structure, folding and splicing mechanism. *Biol Chem* 388: 665-678.
Fisette, J.F., Toutant, J., Dugré-Brisson, S., Desgroseillers, L., and Chabot, B. (2010). hnRNP A1 and hnRNP H can collaborate to modulate 5' splice site selection. *RNA* 16: 228-238.
Fox-Walsh, K.L., Dou, Y., Lam, B.J., Hung, S.P., Baldi, P.F., and Hertel, K.J. (2005). The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc Natl Acad Sci U S A* 102: 16176-16181.
Frilander, M.J., and Meng, X. (2005). Proximity of the U12 snRNA with both the 5' splice site and the branch point during early stages of spliceosome assembly. *Mol Cell Biol* 25: 4813-4825.
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. (1993). Specific commitment of different pre-mRNAs to splicing by single SR proteins. *Nature* 365: 82-85.
Gallego, M.E., Gattoni, R., Stevenin, J., Marie, J., and Expert-Bezancon, A. (1997). The SR splicing factors ASF/SF2 and SC35 have antagonistic effects on intronic enhancer-dependent splicing of the β-tropomyosin alternative exon 6A. *EMBO J* 16: 1772-1784.
Gao, K., Masuda, A., Matsuura, T., and Ohno, K. (2008). Human branch point consensus sequence is yUnAy. *Nucleic Acids Res* 36: 2257-2267.
Gilbert, W. (1986). Origin of life - the RNA world. *Nature* 319: 618-618.
Giles, K.E., and Beemon, K.L. (2005). Retroviral splicing suppressor sequesters a 3' splice site in a 50S aberrant splicing complex. *Mol Cell Biol* 25: 4397-4405.
Görnemann, J., Kotovic, K.M., Hujer, K., and Neugebauer, K.M. (2005). Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol Cell* 19: 53-63.
Gozani, O., Feld, R., and Reed, R. (1996). Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev* 10: 233-243.
Gozani, O., Potashkin, J., and Reed, R. (1998). A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol Cell Biol* 18: 4752-4760.
Grainger, R.J., and Beggs, J.D. (2005). Prp8 protein: at the heart of the spliceosome. *RNA* 11: 533-557.
Green, R.E., Lewis, B.P., Hillman, R.T., Blanchette, M., Lareau, L.F., Garnett, A.T., Rio, D.C., and Brenner, S.E. (2003). Widespread predicted nonsense-mediated mRNA decay of alternatively-spliced transcripts of human normal and disease genes. *Bioinformatics* 19 Suppl 1: i118-121.
Guan, F., Caratozzolo, R.M., Goraczniak, R., Ho, E.S., and Gunderson, S.I. (2007). A bipartite U1 site represses U1A expression by synergizing with PIE to inhibit nuclear polyadenylation. *RNA* 13: 2129-2140.

Guth, S., Martínez, C., Gaur, R.K., and Valcárcel, J. (1999). Evidence for substrate-specific requirement of the splicing factor U2AF(35) and for its function after polypyrimidine tract recognition by U2AF(65). *Mol Cell Biol* 19: 8263-8271.

Hall, S.L., and Padgett, R.A. (1994). Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J Mol Biol* 239: 357-365.

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, K., Yeo, G., An, P., Burge, C.B., and Grabowski, P.J. (2005). A combinatorial code for splicing silencing: UAGG and GGGG motifs. *PLoS Biol* 3: e158.

Han, S.P., Tang, Y.H., and Smith, R. (2010). Functional diversity of the hnRNPs: past, present and perspectives. *Biochem J* 430: 379-392.

Haque, A., Buratti, E., and Baralle, F.E. (2010). Functional properties and evolutionary splicing constraints on a composite exonic regulatory element of splicing in CFTR exon 12. *Nucleic Acids Res* 38: 647-659.

Hashimoto, C., and Steitz, J.A. (1984). U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. *Nucleic Acids Res* 12: 3283-3293.

Hastings, M.L., Allemand, E., Duelli, D.M., Myers, M.P., and Krainer, A.R. (2007). Control of pre-mRNA splicing by the general splicing factors PUF60 and U2AF(65). *PLoS One* 2: e538.

Hastings, M.L., and Krainer, A.R. (2001). Functions of SR proteins in the U12-dependent AT-AC pre-mRNA splicing pathway. *RNA* 7: 471-482.

Hastings, M.L., Resta, N., Traum, D., Stella, A., Guanti, G., and Krainer, A.R. (2005). An LKB1 AT-AC intron mutation causes Peutz-Jeghers syndrome via splicing at noncanonical cryptic splice sites. *Nat Struct Mol Biol* 12: 54-59.

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.

Hickey, D.A. (1992). Evolutionary dynamics of transposable elements in prokaryotes and eukaryotes. *Genetica* 86: 269-274.

Ho, C.K., and Shuman, S. (1999). Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol Cell* 3: 405-411.

Hoffman, B.E., and Grabowski, P.J. (1992). U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. *Genes Dev* 6: 2554-2568.

Honoré, B. (2000). The hnRNP 2H9 gene, which is involved in the splicing reaction, is a multiply spliced gene. *Biochim Biophys Acta* 1492: 108-119.

Honoré, B., Baandrup, U., and Vorum, H. (2004). Heterogeneous nuclear ribonucleoproteins F and H/H' show differential expression in normal and selected cancer tissues. *Exp Cell Res* 294: 199-209.

Honoré, B., Rasmussen, H.H., Vorum, H., Dejgaard, K., Liu, X., Gromov, P., Madsen, P., Gesser, B., Tommerup, N., and Celis, J.E. (1995). Heterogeneous nuclear ribonucleoproteins H, H', and F are members of a ubiquitously expressed subfamily of related but distinct proteins encoded by genes mapping to different chromosomes. *J Biol Chem* 270: 28780-28789.

House, A.E., and Lynch, K.W. (2006). An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. *Nat Struct Mol Biol* 13: 937-944.

Huranová, M., Ivani, I., Benda, A., Poser, I., Brody, Y., Hof, M., Shav-Tal, Y., Neugebauer, K.M., and Staněk, D. (2010). The differential interaction of snRNPs with pre-mRNA reveals splicing kinetics in living cells. *J Cell Biol* 191: 75-86.

Hutchison, S., LeBel, C., Blanchette, M., and Chabot, B. (2002). Distinct sets of adjacent heterogeneous nuclear ribonucleoprotein (hnRNP) A1/A2 binding sites control 5' splice site selection in the hnRNP A1 mRNA precursor. *J Biol Chem* 277: 29745-29752.

Hwang, D.Y., and Cohen, J.B. (1997). U1 small nuclear RNA-promoted exon selection requires a minimal distance between the position of U1 binding and the 3' splice site across the exon. *Mol Cell Biol* 17: 7099-7107.
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.
Irimia, M., Penny, D., and Roy, S.W. (2007). Coevolution of genomic intron number and splice sites. *Trends Genet* 23: 321-325.
Jackson, I.J. (1991). A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res* 19: 3795-3798.
Jacquenet, S., Méreau, A., Bilodeau, P.S., Damier, L., Stoltzfus, C.M., and Branlant, C. (2001). A second exon splicing silencer within human immunodeficiency virus type 1 *tat* exon 2 represses splicing of Tat mRNA and binds protein hnRNP H. *J Biol Chem* 276: 40464-40475.
Jeffares, D.C., Poole, A.M., and Penny, D. (1998). Relics from the RNA world. *J Mol Evol* 46: 18-36.
Jiang, Z.H., Zhang, W.J., Rao, Y., and Wu, J.Y. (1998). Regulation of Ich-1 pre-mRNA alternative splicing and apoptosis by mammalian splicing factors. *Proc Natl Acad Sci U S A* 95: 9155-9160.
Jumaa, H., and Nielsen, P.J. (1997). The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *EMBO J* 16: 5077-5085.
Jurica, M.S., and Moore, M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* 12: 5-14.
Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468: 664-668.
Kandels-Lewis, S., and Séraphin, B. (1993). Involvement of U6 snRNA in 5' splice site selection. *Science* 262: 2035-2039.
Kanopka, A., Mühlemann, O., and Akusjärvi, G. (1996). Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* 381: 535-538.
Kent, O.A., and MacMillan, A.M. (2002). Early organization of pre-mRNA during spliceosome assembly. *Nat Struct Biol* 9: 576-581.
Khanna, A., and Stamm, S. (2010). Regulation of alternative splicing by short non-coding nuclear RNAs. *RNA Biol* 7: 480-485.
Kim, S., Kim, H., Fong, N., Erickson, B., and Bentley, D.L. (2011). Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proc Natl Acad Sci U S A* 108: 13564-13569.
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.
Kolossova, I., and Padgett, R.A. (1997). U11 snRNA interacts in vivo with the 5' splice site of U12-dependent (AU-AC) pre-mRNA introns. *RNA* 3: 227-233.
Konarska, M.M., and Sharp, P.A. (1987). Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell* 49: 763-774.
Konarska, M.M., and Sharp, P.A. (1988). Association of U2, U4, U5, and U6 small nuclear ribonucleoproteins in a spliceosome-type complex in absence of precursor RNA. *Proc Natl Acad Sci U S A* 85: 5459-5462.
Koonin, E.V. (2009). Intron-dominated genomes of early ancestors of eukaryotes. *J Hered* 100: 618-623.
Krämer, A., Grüter, P., Gröning, 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.
Kress, T.L., Krogan, N.J., and Guthrie, C. (2008). A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast. *Mol Cell* 32: 727-734.
Ladd, A.N., Stenberg, M.G., Swanson, M.S., and Cooper, T.A. (2005). Dynamic balance between activation and repression regulates pre-mRNA alternative splicing during heart development. *Dev Dyn* 233: 783-793.
Lallena, M.J., Chalmers, K.J., Llamazares, S., Lamond, A.I., and Valcárcel, J. (2002). Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SPF45. *Cell* 109: 285-296.

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.

Lanctôt, 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.

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.

Lane, C.E., van den Heuvel, K., Kozera, C., Curtis, B.A., Parsons, B.J., Bowman, S., and Archibald, J.M. (2007). Nucleomorph genome of *Hemiselmis andersenii* reveals complete intron loss and compaction as a driver of protein structure and function. *Proc Natl Acad Sci U S A* 104: 19908-19913.

Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C., and Brenner, S.E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* 446: 926-929.

Lee, C., Jaladat, Y., Mohammadi, A., Sharifi, A., Geisler, S., and Valadkhan, S. (2010). Metal binding and substrate positioning by evolutionarily invariant U6 sequences in catalytically active protein-free snRNAs. *RNA* 16: 2226-2238.

LeFave, C.V., Squatrito, M., Vorlova, S., Rocco, G.L., Brennan, C.W., Holland, E.C., Pan, Y.-X., and Cartegni, L. (2011). Splicing factor hnRNPH drives an oncogenic splicing switch in gliomas. *EMBO J* 30: 4084-4097.

Lejeune, F., Cavaloc, Y., and Stevenin, J. (2001). Alternative splicing of intron 3 of the serine/arginine-rich protein 9G8 gene. Identification of flanking exonic splicing enhancers and involvement of 9G8 as a trans-acting factor. *J Biol Chem* 276: 7850-7858.

Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L., and Steitz, J.A. (1980). Are snRNPs involved in splicing? *Nature* 283: 220-224.

Lewandowska, D., Simpson, C.G., Clark, G.P., Jennings, N.S., Barciszewska-Pacak, M., Lin, C.F., Makalowski, W., Brown, J.W., and Jarmolowski, A. (2004). Determinants of plant U12-dependent intron splicing efficiency. *Plant Cell* 16: 1340-1352.

Lewandowska, M.A., Stuani, C., Parvizpur, A., Baralle, F.E., and Pagani, F. (2005). Functional studies on the ATM intronic splicing processing element. *Nucleic Acids Res* 33: 4007-4015.

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 U S A* 100: 189-192.

Liao, X.C., Colot, H.V., Wang, Y., and Rosbash, M. (1992). Requirements for U2 snRNP addition to yeast pre-mRNA. *Nucleic Acids Res* 20: 4237-4245.

Lim, L.P., and Burge, C.B. (2001). A computational analysis of sequence features involved in recognition of short introns. *Proc Natl Acad Sci U S A* 98: 11193-11198.

Lim, S.R., and Hertel, K.J. (2004). Commitment to splice site pairing coincides with A complex formation. *Mol Cell* 15: 477-483.

Lin, C.F., Mount, S.M., Jarmołowski, A., and Makałowski, W. (2010). Evolutionary dynamics of U12-type spliceosomal introns. *BMC Evol Biol* 10: 47.

Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., and Fu, X.D. (2008). The splicing factor SC35 has an active role in transcriptional elongation. *Nat Struct Mol Biol* 15: 819-826.

Listerman, I., Sapra, A.K., and Neugebauer, K.M. (2006). Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat Struct Mol Biol* 13: 815-822.

Liu, S., Rauhut, R., Vornlocher, H.P., and Lührmann, R. (2006). The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. *RNA* 12: 1418-1430.

Liu, Z., Luyten, I., Bottomley, M.J., Messias, A.C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Krämer, A., and Sattler, M. (2001). Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* 294: 1098-1102.

Logsdon, J.M., Jr. (1998). The recent origins of spliceosomal introns revisited. *Curr Opin Genet Dev* 8: 637-648.

Long, J.C., and Caceres, J.F. (2009). The SR protein family of splicing factors: master regulators of gene expression. *Biochem J* 417: 15-27.

Loomis, R.J., Naoe, Y., Parker, J.B., Savic, V., Bozovsky, M.R., Macfarlan, T., Manley, J.L., and Chakravarti, D. (2009). Chromatin binding of SRp20 and ASF/SF2 and dissociation from mitotic chromosomes is modulated by histone H3 serine 10 phosphorylation. *Mol Cell* 33: 450-461.

Lorković, Z.J., Lopato, S., Pexa, M., Lehner, R., and Barta, A. (2004). Interactions of *Arabidopsis* RS domain containing cyclophilins with SR proteins and U1 and U11 small nuclear ribonucleoprotein-specific proteins suggest their involvement in pre-mRNA splicing. *J Biol Chem* 279: 33890-33898.

Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. *Science* 327: 996-1000.

Luo, H.R., Moreau, G.A., Levin, N., and Moore, M.J. (1999). The human Prp8 protein is a component of both U2- and U12-dependent spliceosomes. *RNA* 5: 893-908.

Lynch, K.W. (2007). Regulation of alternative splicing by signal transduction pathways. *Adv Exp Med Biol* 623: 161-174.

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.

Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., and Lührmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* 298: 2205-2208.

Makarova, O.V., Makarov, E.M., and Lührmann, R. (2001). The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes. *EMBO J* 20: 2553-2563.

Makarova, O.V., Makarov, E.M., Urlaub, H., Will, C.L., Gentzel, M., Wilm, M., and Lührmann, R. (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *EMBO J* 23: 2381-2391.

Malca, H., Shomron, N., and Ast, G. (2003). The U1 snRNP base pairs with the 5' splice site within a penta-snRNP complex. *Mol Cell Biol* 23: 3442-3455.

Manley, J.L., and Krainer, A.R. (2010). A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev* 24: 1073-1074.

Maroney, P.A., Romfo, C.M., and Nilsen, T.W. (2000). Functional recognition of 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. *Mol Cell* 6: 317-328.

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

Matlin, A.J., Southby, J., Gooding, C., and Smith, C.W. (2007). Repression of α-actinin SM exon splicing by assisted binding of PTB to the polypyrimidine tract. *RNA* 13: 1214-1223.

Matunis, M.J., Xing, J., and Dreyfuss, G. (1994). The hnRNP F protein: unique primary structure, nucleic acid-binding properties, and subcellular localization. *Nucleic Acids Res* 22: 1059-1067.

Mauger, D.M., Lin, C., and Garcia-Blanco, M.A. (2008). hnRNP H and hnRNP F complex with Fox2 to silence fibroblast growth factor receptor 2 exon IIIc. *Mol Cell Biol* 28: 5403-5419.

McCullough, A.J., and Berget, S.M. (2000). An intronic splicing enhancer binds U1 snRNPs to enhance splicing and select 5' splice sites. *Mol Cell Biol* 20: 9225-9235.

McGlincy, N.J., and Smith, C.W. (2008). Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends Biochem Sci* 33: 385-393.

McGlincy, N.J., Tan, L.Y., Paul, N., Zavolan, M., Lilley, K.S., and Smith, C.W. (2010). Expression proteomics of UPF1 knockdown in HeLa cells reveals autoregulation of hnRNP A2/B1 mediated by alternative splicing resulting in nonsense-mediated mRNA decay. *BMC Genomics* 11: 565.

McNally, L.M., Yee, L., and McNally, M.T. (2004). Two regions promote U11 small nuclear ribonucleoprotein particle binding to a retroviral splicing inhibitor element (negative regulator of splicing). *J Biol Chem* 279: 38201-38208.

McNally, L.M., Yee, L., and McNally, M.T. (2006). Heterogeneous nuclear ribonucleoprotein H is required for optimal U11 small nuclear ribonucleoprotein binding to a retroviral RNA-processing control element: implications for U12-dependent RNA splicing. *J Biol Chem* 281: 2478-2488.

McNally, M.T. (2008). RNA processing control in avian retroviruses. *Front Biosci* 13: 3869-3883.

Melamud, E., and Moult, J. (2009). Stochastic noise in splicing machinery. *Nucleic Acids Res* 37: 4873-4886.

Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F., and Dietz, H.C. (2004). Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat Genet* 36: 1073-1078.

Montzka, K.A., and Steitz, J.A. (1988). Additional low-abundance human small nuclear ribonucleoproteins: U11, U12, etc. *Proc Natl Acad Sci U S A* 85: 8885-8889.

Moore, M.J., Schwartzfarb, E.M., Silver, P.A., and Yu, M.C. (2006). Differential recruitment of the splicing machinery during transcription predicts genome-wide patterns of mRNA splicing. *Mol Cell* 24: 903-915.

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

Muto, Y., Pomeranz Krummel, D., Oubridge, C., Hernandez, H., Robinson, C.V., Neuhaus, D., and Nagai, K. (2004). The structure and biochemical properties of the human spliceosomal protein U1C. *J Mol Biol* 341: 185-198.

Nakai, K., and Sakamoto, H. (1994). Construction of a novel database containing aberrant splicing mutations of mammalian genes. *Gene* 141: 171-177.

Nelissen, R.L., Heinrichs, V., Habets, W.J., Simons, F., Lührmann, R., and van Venrooij, W.J. (1991). Zinc finger-like structure in U1-specific protein C is essential for specific binding to U1 snRNP. *Nucleic Acids Res* 19: 449-454.

Newman, A.J., and Norman, C. (1992). U5 snRNA interacts with exon sequences at 5' and 3' splice sites. *Cell* 68: 743-754.

Ni, J.Z., Grate, L., Donohue, J.P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T.A., Blume, J.E., and Ares, M., Jr. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev* 21: 708-718.

Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463: 457-463.

Nixon, J.E., Wang, A., Morrison, H.G., McArthur, A.G., Sogin, M.L., Loftus, B.J., and Samuelson, J. (2002). A spliceosomal intron in Giardia lamblia. *Proc Natl Acad Sci U S A* 99: 3701-3705.

Nottrott, S., Urlaub, H., and Lührmann, R. (2002). Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J* 21: 5527-5538.

Otake, L.R., Scamborova, P., Hashimoto, C., and Steitz, J.A. (2002). The divergent U12-type spliceosome is required for pre-mRNA splicing and is essential for development in Drosophila. *Mol Cell* 9: 439-446.

Padgett, R.A., and Shukla, G.C. (2002). A revised model for U4atac/U6atac snRNA base pairing. *RNA* 8: 125-128.

Pagani, F., Buratti, E., Stuani, C., and Baralle, F.E. (2003a). Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9. *J Biol Chem* 278: 26580-26588.

Pagani, F., Stuani, C., Tzetis, M., Kanavakis, E., Efthymiadou, A., Doudounakis, S., Casals, T., and Baralle, F.E. (2003b). New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. *Hum Mol Genet* 12: 1111-1120.

Pan, Q., Saltzman, A.L., Kim, Y.K., Misquitta, C., Shai, O., Maquat, L.E., Frey, B.J., and Blencowe, B.J. (2006). Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. *Genes Dev* 20: 153-158.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40: 1413-1415.
Pandya-Jones, A., and Black, D.L. (2009). Co-transcriptional splicing of constitutive and alternative exons. *RNA* 15: 1896-1908.
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 U S A* 101: 15974-15979.
Patel, A.A., McCarthy, M., and Steitz, J.A. (2002). The splicing of U12-type introns can be a rate-limiting step in gene expression. *EMBO J* 21: 3804-3815.
Patel, A.A., and Steitz, J.A. (2003). Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol* 4: 960-970.
Patel, S.B., and Bellini, M. (2008). The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res* 36: 6482-6493.
Penny, D., Hoeppner, M.P., Poole, A.M., and Jeffares, D.C. (2009). An overview of the introns-first theory. *J Mol Evol* 69: 527-540.
Perales, R., and Bentley, D. (2009). "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Mol Cell* 36: 178-191.
Pessa, H.K. (2010). *U12-type spliceosome: Localization and effects of splicing efficiency on gene expression* (Helsinki, Helsinki University Print).
Pessa, H.K., Greco, D., Kvist, J., Wahlström, G., Heino, T.I., Auvinen, P., and Frilander, M.J. (2010). Gene expression profiling of U12-type spliceosome mutant Drosophila reveals widespread changes in metabolic pathways. *PLoS One* 5: e13215.
Pessa, H.K., Ruokolainen, A., and Frilander, M.J. (2006). The abundance of the spliceosomal snRNPs is not limiting the splicing of U12-type introns. *RNA* 12: 1883-1892.
Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* 20: 2922-2936.
Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 Å 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.
Query, C.C., Bentley, R.C., and Keene, J.D. (1989). A specific 31-nucleotide domain of U1 RNA directly interacts with the 70K small nuclear ribonucleoprotein component. *Mol Cell Biol* 9: 4872-4881.
Query, C.C., Moore, M.J., and Sharp, P.A. (1994). Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev* 8: 587-597.
Reed, R., and Maniatis, T. (1986). A role for exon sequences and splice-site proximity in splice-site selection. *Cell* 46: 681-690.
Robberson, B.L., Cote, G.J., and Berget, S.M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* 10: 84-94.
Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L., and Wall, R. (1980). Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin mu chain. *Cell* 20: 303-312.
Rogers, J., and Wall, R. (1980). A mechanism for RNA splicing. *Proc Natl Acad Sci U S A* 77: 1877-1879.
Roscigno, R.F., and Garcia-Blanco, M.A. (1995). SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA* 1: 692-706.
Rösel, T.D., Hung, L.H., Medenbach, J., Donde, K., Starke, S., Benes, V., Rätsch, G., and Bindereif, A. (2011). RNA-Seq analysis in mutant zebrafish reveals role of U1C protein in alternative splicing regulation. *EMBO J* 30: 1965-1976.
Rossbach, O., Hung, L.H., Schreiner, S., Grishina, I., Heiner, M., Hui, J., and Bindereif, A. (2009). Auto- and cross-regulation of the hnRNP L proteins by alternative splicing. *Mol Cell Biol* 29: 1442-1451.

Rossi, F., Forné, T., Antoine, E., Tazi, J., Brunel, C., and Cathala, G. (1996). Involvement of U1 small nuclear ribonucleoproteins (snRNP) in 5' splice site-U1 snRNP interaction. *J Biol Chem* 271: 23985-23991.

Roy, S.W., and Irimia, M. (2009). Splicing in the eukaryotic ancestor: form, function and dysfunction. *Trends Ecol Evol* 24: 447-455.

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.

Russell, A., Charette, J., Spencer, D., and Gray, M. (2006). An early evolutionary origin for the minor spliceosome. *Nature* 443: 863-866.

Russo, A., Catillo, M., Esposito, D., Briata, P., Pietropaolo, C., and Russo, G. (2011). Autoregulatory circuit of human rpL3 expression requires hnRNP H1, NPM and KHSRP. *Nucleic Acids Res* 39: 7576-7585.

Rutz, B., and Séraphin, 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., Kim, Y.K., Pan, Q., Fagnani, M.M., Maquat, L.E., and Blencowe, B.J. (2008). Regulation of multiple core spliceosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay. *Mol Cell Biol* 28: 4320-4330.

Saltzman, A.L., Pan, Q., and Blencowe, B.J. (2011). Regulation of alternative splicing by the core spliceosomal machinery. *Genes Dev* 25: 373-384.

Sapra, A.K., Änkö, M.L., Grishina, I., Lorenz, M., Pabis, M., Poser, I., Rollins, J., Weiland, E.M., and Neugebauer, K.M. (2009). SR protein family members display diverse activities in the formation of nascent and mature mRNPs in vivo. *Mol Cell* 34: 179-190.

Scamborova, P., Wong, A., and Steitz, J.A. (2004). An intronic enhancer regulates splicing of the twintron of Drosophila melanogaster prospero pre-mRNA by two different spliceosomes. *Mol Cell Biol* 24: 1855-1869.

Schaub, M.C., Lopez, S.R., and Caputi, M. (2007). Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes. *J Biol Chem* 282: 13617-13626.

Scherly, D., Boelens, W., van Venrooij, W.J., Dathan, N.A., Hamm, J., and Mattaj, I.W. (1989). Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA. *EMBO J* 8: 4163-4170.

Schneider, C., Will, C.L., Brosius, J., Frilander, M.J., and Lührmann, R. (2004). Identification of an evolutionarily divergent U11 small nuclear ribonucleoprotein particle in Drosophila. *Proc Natl Acad Sci U S A* 101: 9584-9589.

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.

Schneider, M., Will, C.L., Anokhina, M., Tazi, J., Urlaub, H., and Lührmann, R. (2010). Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. *Mol Cell* 38: 223-235.

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.

Shao, W., Kim, H.S., Cao, Y., Xu, Y.Z., and Query, C.C. (2011). A U1-U2 snRNP interaction network during intron definition. *Mol Cell Biol*.

Sharma, S., Falick, A.M., and Black, D.L. (2005). Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex. *Mol Cell* 19: 485-496.

Sharma, S., Kohlstaedt, L.A., Damianov, A., Rio, D.C., and Black, D.L. (2008). Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat Struct Mol Biol* 15: 183-191.

Sharma, S., Maris, C., Allain, F.H., and Black, D.L. (2011). U1 snRNA directly interacts with polypyrimidine tract-binding protein during splicing repression. *Mol Cell* 41: 579-588.

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.

Shen, H., and Green, M.R. (2006). RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. *Genes Dev* 20: 1755-1765.
Shen, H., and Green, M.R. (2007). RS domain-splicing signal interactions in splicing of U12-type and U2-type introns. *Nat Struct Mol Biol* 14: 597-603.
Shen, H., Kan, J.L., and Green, M.R. (2004). Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol Cell* 13: 367-376.
Shen, H., Zheng, X., Luecke, S., and Green, M.R. (2010). The U2AF35-related protein Urp contacts the 3' splice site to promote U12-type intron splicing and the second step of U2-type intron splicing. *Genes Dev* 24: 2389-2394.
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.
Shukla, G.C., and Padgett, R.A. (2001). The intramolecular stem-loop structure of U6 snRNA can functionally replace the U6atac snRNA stem-loop. *RNA* 7: 94-105.
Siebel, C.W., Fresco, L.D., and Rio, D.C. (1992). The mechanism of somatic inhibition of Drosophila P-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5' splice site control U1 snRNP binding. *Genes Dev* 6: 1386-1401.
Sikand, K., and Shukla, G.C. (2011). Functionally important structural elements of U12 snRNA. *Nucleic Acids Res* 39: 8531-8543.
Singh, G., Rebbapragada, I., and Lykke-Andersen, J. (2008). A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. *PLoS Biol* 6: e111.
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.
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.
Somasekhar, M.B., and Mertz, J.E. (1985). Exon mutations that affect the choice of splice sites used in processing the SV40 late transcripts. *Nucleic Acids Res* 13: 5591-5609.
Sontheimer, E.J., and Steitz, J.A. (1993). The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* 262: 1989-1996.
Sorek, R., and Ast, G. (2003). Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res* 13: 1631-1637.
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.
Spellman, R., Llorian, M., and Smith, C.W. (2007). Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. *Mol Cell* 27: 420-434.
Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92: 315-326.
Sterner, D.A., and Berget, S.M. (1993). In vivo recognition of a vertebrate mini-exon as an exon-intron-exon unit. *Mol Cell Biol* 13: 2677-2687.
Sterner, D.A., Carlo, T., and Berget, S.M. (1996). Architectural limits on split genes. *Proc Natl Acad Sci U S A* 93: 15081-15085.
Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. (2002). Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol Cell* 9: 31-44.
Stoltzfus, A. (1999). On the possibility of constructive neutral evolution. *J Mol Evol* 49: 169-181.
Sun, S., Zhang, Z., Sinha, R., Karni, R., and Krainer, A.R. (2010). SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control. *Nat Struct Mol Biol* 17: 306-312.
Sureau, A., Gattoni, R., Dooghe, Y., Stévenin, J., and Soret, J. (2001). SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *EMBO J* 20: 1785-1796.
Tardiff, D.F., and Rosbash, M. (2006). Arrested yeast splicing complexes indicate stepwise snRNP recruitment during in vivo spliceosome assembly. *RNA* 12: 968-979.

Tarn, W.Y., and Steitz, J.A. (1996a). Highly diverged 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.
Tarn, W.Y., and Steitz, J.A. (1997). Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem Sci* 22: 132-137.
Tazi, J., Kornstädt, U., Rossi, F., Jeanteur, P., Cathala, G., Brunel, C., and Lührmann, R. (1993). Thiophosphorylation of U1-70K protein inhibits pre-mRNA splicing. *Nature* 363: 283-286.
Tidow, H., Andreeva, A., Rutherford, T.J., and Fersht, A.R. (2009). Solution structure of the U11-48K CHHC zinc-finger domain that specifically binds the 5' splice site of U12-type introns. *Structure* 17: 294-302.
Valadkhan, S. (2010). Role of the snRNAs in spliceosomal active site. *RNA Biol* 7: 345-353.
Valadkhan, S., and Jaladat, Y. (2010). The spliceosomal proteome: at the heart of the largest cellular ribonucleoprotein machine. *Proteomics* 10: 4128-4141.
Valadkhan, S., Mohammadi, A., Jaladat, Y., and Geisler, S. (2009). Protein-free small nuclear RNAs catalyze a two-step splicing reaction. *Proc Natl Acad Sci U S A* 106: 11901-11906.
Valcárcel, J., Gaur, R.K., Singh, R., and Green, M.R. (1996). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of 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.
Valencia, P., Dias, A.P., and Reed, R. (2008). Splicing promotes rapid and efficient mRNA export in mammalian cells. *Proc Natl Acad Sci U S A* 105: 3386-3391.
Vanacova, S., Yan, W., Carlton, J.M., and Johnson, P.J. (2005). Spliceosomal introns in the deep-branching eukaryote Trichomonas vaginalis. *Proc Natl Acad Sci U S A* 102: 4430-4435.
Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136: 701-718.
Wang, E., and Cambi, F. (2009). Heterogeneous nuclear ribonucleoproteins H and F regulate the proteolipid protein/DM20 ratio by recruiting U1 small nuclear ribonucleoprotein through a complex array of G runs. *J Biol Chem* 284: 11194-11204.
Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, 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, J., Smith, P.J., Krainer, A.R., and Zhang, M.Q. (2005). Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes. *Nucleic Acids Res* 33: 5053-5062.
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.
Wang, Z., Rolish, M.E., Yeo, G., Tung, V., Mawson, M., and Burge, C.B. (2004). Systematic identification and analysis of exonic splicing silencers. *Cell* 119: 831-845.
Wang, Z., Xiao, X., Van Nostrand, E., and Burge, C.B. (2006). General and specific functions of exonic splicing silencers in splicing control. *Mol Cell* 23: 61-70.
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.
Weber, G., Trowitzsch, S., Kastner, B., Lührmann, R., and Wahl, M.C. (2010). Functional organization of the Sm core in the crystal structure of human U1 snRNP. *EMBO J* 29: 4172-4184.
Weiner, A.M. (1993). mRNA splicing and autocatalytic introns: distant cousins or the products of chemical determinism? *Cell* 72: 161-164.
Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb Perspect Biol* 3: a003707.
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., MacMillan, A.M., Katopodis, N.F., Neubauer, G., Wilm, M., Lührmann, R., and Query, C.C. (2001). A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO J* 20: 4536-4546.
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.
Wittmann, J., Hol, E.M., and Jäck, H.M. (2006). hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. *Mol Cell Biol* 26: 1272-1287.
Wollerton, M.C., Gooding, C., Wagner, E.J., Garcia-Blanco, M.A., and Smith, C.W. (2004). Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol Cell* 13: 91-100.
Wu, J., and Manley, J.L. (1989). Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. *Genes Dev* 3: 1553-1561.
Wu, J.A., and Manley, J.L. (1991). Base pairing between U2 and U6 snRNAs is necessary for splicing of a mammalian pre-mRNA. *Nature* 352: 818-821.
Wu, J.Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* 75: 1061-1070.
Wu, Q., and Krainer, A.R. (1996). U1-mediated exon definition interactions between AT-AC and GT-AG introns. *Science* 274: 1005-1008.
Wu, Q., and Krainer, A.R. (1998). Purine-rich enhancers function in the AT-AC pre-mRNA splicing pathway and do so independently of intact U1 snRNP. *RNA* 4: 1664-1673.
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.
Xiao, X., Wang, Z., Jang, M., and Burge, C.B. (2007). Coevolutionary networks of splicing cis-regulatory elements. *Proc Natl Acad Sci U S A* 104: 18583-18588.
Xiao, X., Wang, Z., Jang, M., Nutiu, R., Wang, E.T., and Burge, C.B. (2009). Splice site strength-dependent activity and genetic buffering by poly-G runs. *Nat Struct Mol Biol* 16: 1094-1100.
Xu, Y.Z., Newnham, C.M., Kameoka, S., Huang, T., Konarska, M.M., and Query, C.C. (2004). Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. *EMBO J* 23: 376-385.
Xue, Y., Zhou, Y., Wu, T., Zhu, T., Ji, X., Kwon, Y.S., Zhang, C., Yeo, G., Black, D.L., Sun, H., *et al.* (2009). Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol Cell* 36: 996-1006.
Yean, S.L., Wuenschell, G., Termini, J., and Lin, R.J. (2000). Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* 408: 881-884.
Yeo, G.W., Van Nostrand, E.L., and Liang, T.Y. (2007). Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements. *PLoS Genet* 3: e85.
Yu, Y.-T., Scharl, E., Smith, C.M., and Steitz, J.A. (1999). The growing world of small nuclear ribonucleoproteins. In *The RNA world*, Gesteland, R.F., Cech, T.R., and Atkins, J.F., eds. (Cold Spring Harbour, NY, Cold Spring Harbour Laboratory Press), pp. 487-524.
Yu, Y., Maroney, P.A., Denker, J.A., Zhang, X.H., Dybkov, O., Lührmann, R., Jankowsky, E., Chasin, L.A., and Nilsen, T.W. (2008). Dynamic regulation of alternative splicing by silencers that modulate 5' splice site competition. *Cell* 135: 1224-1236.
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 U S A* 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.
Zhang, G., Taneja, K.L., Singer, R.H., and Green, M.R. (1994). Localization of pre-mRNA splicing in mammalian nuclei. *Nature* 372: 809-812.
Zhong, X.Y., Wang, P., Han, J., Rosenfeld, M.G., and Fu, X.D. (2009). SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Mol Cell* 35: 1-10.
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.

Zhu, W., and Brendel, V. (2003). Identification, characterization and molecular phylogeny of U12-dependent introns in the *Arabidopsis thaliana* genome. *Nucleic Acids Res* 31: 4561-4572.
Zhuang, Y., and Weiner, A.M. (1986). A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* 46: 827-835.
Zhuang, Y., and Weiner, A.M. (1989). A compensatory base change in human U2 snRNA can suppress a branch site mutation. *Genes Dev* 3: 1545-1552.
Zorio, D.A., and Blumenthal, T. (1999). Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*. *Nature* 402: 835-838.