Stony Brook University

Characterization of the Mechanistic Differences in Splicing of SF2/ASF RS Domain-Independent and RS Domain-Dependent Pre-mRNAs

A Dissertation Presented

by

Stephanie Denise Shaw

to

The Graduate School

in Partial fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Molecular and Cellular Biology

Stony Brook University

August 2007

Stony Brook University

The Graduate School

Stephanie Denise Shaw

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Dr. Adrian R. Krainer--Dissertation Advisor
Professor, Cold Spring Harbor Laboratory
Department of Molecular and Cellular Biology, Stony Brook University

Dr. Gregory Hannon--Chairperson of Defense
Professor, HHMI, Cold Spring Harbor Laboratory
Department of Molecular and Cellular Biology, Stony Brook University

Dr. A. Wali Karzai, Associate Professor, Department of Biochemistry and Cell Biology, Stony Brook University
Department of Molecular and Cellular Biology, Stony Brook University

Dr. Rui-Ming Xu, Professor, Department of Pharmacology, New York University

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Characterization of the Mechanistic Differences in Splicing of SF2/ASF RS Domain-Independent and RS Domain-Dependent Pre-mRNAs

by

Stephanie Denise Shaw

Doctor of Philosophy

in

Molecular and Cellular Biology

Stony Brook University

2007

Removal of introns from pre-mRNAs is an essential step of eukaryotic gene expression. In yeast, pre-mRNAs are short and typically have highly conserved splicing signals, but mammalian pre-mRNAs are very large, with degenerate splicing signals and long introns, creating a situation where exon/intron boundaries are poorly defined, and necessitating additional mechanisms for splice-site definition.

SR proteins are essential pre-mRNA splicing factors that can aid in establishing the location of introns by recruiting components of the splicing machinery to splicing signals. SF2/ASF is a prototypical member of the SR protein family, and is composed of two N-terminal RRM domains and a C-terminal RS domain. SR proteins bind to pre-mRNA via their RRM domains and are thought to recruit other components of the splicing machinery through protein-protein interactions mediated by their RS domains. However, our lab previously demonstrated that splicing can be accomplished for some

but not all pre-mRNAs *in vitro* with a mutant SF2/ASF lacking its RS domain ("ΔRS"). Therefore pre-mRNAs could be classified as either RS domain-dependent or RS domain-independent based on their ability to be spliced with an SR protein lacking an RS domain.

To identify pre-mRNA sequence elements that confer an RS domain requirement, we tested a large number of RS domain-dependent pre-mRNAs in which specific sequences were mutated or replaced by sequences from RS domain-independent pre-mRNAs in the *in vitro* splicing assay with SF2/ASF and ΔRS. We have identified sequence elements spanning the 5' splice site that confer RS domain-dependence, and also show that improvement of the pyrimidine tract abrogates the RS domain requirement. Collectively our findings suggest that RS domain-dependence is a consequence of a deficiency in intron definition, the assembly of spliceosomal complexes across the intron to establish the locations of the 5' and 3' splice sites, which is a prerequisite to splicing catalysis.

To try to understand how ΔRS can support splicing *in vitro* for some pre-mRNAs, we tested the abilities of a series of mutant SF2/ASF proteins to splice various RS domain-dependent and RS domain-independent pre-mRNAs. Deletion of the short N-terminal segment preceding the first RRM domain of SF2/ASF increased the amount of splicing supported by both SF2/ASF and ΔRS. Surprisingly, deletion of this inhibitory segment in the context of ΔRS results in a protein ΔNΔRS that supports splicing for both RS domain-independent and RS domain-dependent pre-mRNAs, indicating that the RS domain is dispensable for splicing *in vitro.*

Characterization of splicing of an RS domain-dependent pre-mRNA with the ΔNΔRS protein reveals a kinetic dependence on the RS domain, such that the RS

domain-dependent pre-mRNA is spliced more slowly with ΔNΔRS than with SF2/ASF or ΔNSF2/ASF. Deletion of the N-terminal segment from ΔRS greatly improves binding of the protein to the RS domain-dependent pre-mRNA, suggesting that the N-terminal segment inhibits splicing by interfering with the ability of SF2/ASF to bind to RNA. However, deletion of the N-terminus from SF2/ASF does not significantly increase the ability of the protein to bind to the RS domain-dependent substrate, suggesting that the RS domain also influences RNA binding. These data suggest that RS domain-dependent substrates may be particularly sensitive to the inhibition conferred by this N-terminal segment, because binding of SF2/ASF to RNA is normally assisted by the RS domain in these pre-mRNA contexts.

List of Figures

List of Tables

Acknowledgements

The following figures have been reproduced with copyright permission:

Figure 3 Spliceosomal Complex Assembly

Permission for use of the figure entitled “The Spliceosomal Splicing Cycle”, Figure 11-19 in Molecular and Cellular Biology, 4th Edition, by James Darnell, Paul Matsudaira, Lawrence Zipursky, Harvey Lodish, Arnold Berk, and David Baltimore, was obtained from W. H. Freeman and Company/Worth Publishers.

Figure 7 Structure of the RRM domain of an SR protein

Permission for use of Figure 3C, “Overview of the solution structure of the SRp20 RRM in complex with CAUC: The most representative structure of the complex in ribbon (protein backbone) and stick (RNA) representation.”, in Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8, The EMBO Journal (2006) 25, 5126–5137, by Yann Hargous, Guillaume M Hautbergue, Aura M Tintaru, Lenka Skrisovska, Alexander P Golovanov, James Stevenin, Lu-Yun Lian, Stuart A Wilson, and Frédéric H-T Allain, was obtained from the Nature Publishing Group.

Chapter 1

Background

1.1 Eukaryotic Gene Expression and pre-mRNA splicing

1.1.1 Split genes and the discovery of pre-mRNA splicing

In the 1970's biologists knew that eukaryotic mRNAs, the products of gene transcription and the templates for production of protein by the ribosome, were capped at their 5' ends and polyadenylated at their 3' ends, but they were puzzled by the identification of nuclear RNAs with the same 5' and 3' modifications which were much greater in length than the cytoplasmic mRNAs. Hybridization of adenovirus mRNA to the DNA of the gene encoding for it revealed that some segments of the gene looped out and were therefore not included in its mRNA (Berget et al, 1977 and Chow et al, 1977), immediately suggesting that large segments were somehow removed from these nuclear RNAs during the production of mature mRNAs. The simultaneous discovery of the phenomenon of "split genes" at Cold Spring Harbor and MIT ignited a race to identify the enzyme responsible for "splicing" of RNA and to understand the mechanism by which it removes segments from the nuclear RNA and joins the remaining segments together.

At Walter Gilbert's suggestion, segments transcribed but not included in the mRNA, or intervening sequences, came to be known as "introns", while segments spliced together to be included in the mRNA, or expressed sequences, were termed "exons". An examination of the sequences of introns and exons revealed the presence of consensus sequences that overlapped their boundaries, the donor (5'-splice junction) and acceptor (3'-splice junction) sites (Breathnach and Chambon, 1981). Mutation of these 5' and 3' splice site consensus sequences was found to prevent splicing.

Development of efficient *in vitro* systems for the biochemical study of splicing (Hernandez and Keller, 1983, Krainer et al, 1984, Hardy et al 1984) permitted characterization of the intermediates and products of the splicing reaction, including the discovery that introns are spliced out of pre-mRNAs as branched RNAs with a lariat structure (Ruskin et al, 1984, Grabowski et al, 1984). The branch of the lariat intermediate of the splicing pathway was shown to be created by a covalent linkage of the 5' end of the intron to an adenosine residue upstream of the 3' splice site, revealing that intronic sequences in addition to those at the 5' and 3' splice sites were involved in splicing. A comparison of the highly conserved sequences surrounding the branch site in yeast introns (Langford at al, 1983) to the less conserved sequences surrounding the branch site in mammalian introns permitted the identification of analogous but weakly conserved consensus sequences (Keller and Noon, 1984). Subsequently, an intronic pyrimidine-rich tract upstream of the 3' splice site was shown to be required for the formation of the lariat in metazoans (Reed and Maniatis, 1985).

By this time, a group of small nuclear RNAs (snRNAs) existing in the form of protein-RNA complexes had been identified, but their function was unknown. The involvement of these small nuclear ribonucleoproteins (snRNPs) in splicing was first suggested by the observations that the 5' end of the U1 snRNA was complementary to the 5' splice site (Lerner et al, 1980, Rogers and Wall, 1980) and that a portion of the U2 snRNA was complementary to the consensus branch site sequences (Keller and Noon, 1984). Several early experiments supported the hypothesis that snRNPs might be involved in splicing. In sedimentation analyses of nuclei, snRNPs were identified in 30-60S complexes that contained high molecular weight RNAs (Deimel et al, 1977, Howard,

1978, Zieve and Penman, 1981). Preincubation of splicing extracts with antibodies against a protein common to all snRNPs inhibited splicing (Padgett et al, 1983, Kramer et al, 1984). U1 and U2 snRNPs were further implicated in splicing when RNAse digestion experiments demonstrated that during the *in vitro* splicing reaction U1snRNP binds to the 5' splice site and U2 snRNP binds to the region of an intron including the branchpoint (Black et al, 1985, Krainer and Maniatis, 1985). By similar methods, the U4/U6 di-snRNP was also demonstrated to be required for splicing (Berget and Robberson, 1986, Black and Steitz, 1986), and U5 snRNP was shown to bind to the 3' splice site (Chabot et al, 1985).

The term "spliceosome" was coined by Brody and Abelson to describe a 40S yeast splicing complex identified by a glycerol gradient centrifugation characterization of complexes associated with labeled splicing precursors, intermediates, and products of *in vitro* splicing reactions (Brody and Abelson, 1985). A mammalian 60S spliceosome complex was also identified and characterized (Grabowski et al, 1985, Frendewey and Keller, 1985). Amazingly, the splicing apparatus appeared to be as large as the ribosome, and was evidently also a ribonucleoprotein machine.

In vitro studies of complexes associated with splicing substrates revealed a stepwise assembly of different spliceosomal complexes prior to the execution of the catalytic steps of splicing (Pikielny et al, 1986, Konarska and Sharp, 1987, Bindereif and Green, 1987, Cheng and Abelson, 1987, Michaud and Reed, 1991). (See section 1.2.2 for a review of the spliceosome assembly pathway and associated protein factors, and section 1.2.3 for a review of splicing catalysis.)

As data accumulated from the many labs working to biochemically characterize the splicing reaction and the machinery that accomplishes it, the number of proteins known to play roles in pre-mRNA splicing continued to grow. By the early 1990's, over 30 different non-snRNP protein factors that participate in the assembly of spliceosomal complexes had been identified (Bennett et al, 1992, Gozani et al 1994). By the mid-1990's, the number of protein factors implicated in splicing had grown to over 70 (Will and Luhrmann, 1997), and rose to 100 by the end of the decade (Burge et al, 1999). More recently, when purified functional spliceosomes were subjected to mass spectroscopic analysis it was discovered that the mammalian splicing machinery may contain as many as 300 different protein components (Hartmuth et al, 2002, Makarov et al, 2002, Zhou et al, 2002). Thus, merely identifying functions for all the *bona fide* protein factors of the spliceosome is a huge challenge unto itself, a task that will continue to go hand in hand with trying to understand how such a large and complex macromolecular machine can accomplish the daunting task of pre-mRNA splicing with speed and precision, particularly in higher metazoans, given the degeneracy of the splicing signals it must recognize. (For a discussion of the degeneracy of splicing signals in metazoan pre-mRNAs, see section 1.2.1 regarding splice site consensus sequences.)

1.1.2 Alternative splicing and the complexity of the proteome

One of the most surprising observations made through a comparison of the many genomes now available is that the complexity of an organism is not correlated with the number of protein coding genes in its genome, suggesting that organismal complexity may be conferred by mechanisms that regulate or diversify gene functions (Blencowe,

2006). Alternative splicing, the process by which different combinations of exons are joined together by the spliceosome to produce different mRNA variants, is one of the major mechanisms employed by the cell to generate diversity within its proteome. Earlier estimates indicated that at least two-thirds of all human genes code for alternative exons (Johnson et al, 2003) and that there are between one and two alternative splicing events per multi-intron gene (Lander et al, 2001, Johnson et al, 2003), but these numbers are now thought to be an underestimation of the number of alternative splicing events and are expected to increase as the databases amass expressed sequence tag (EST) data from the more complex tissue types, which on average tend to produce a more varied repertoire of mRNA isoforms per gene (Modrek et al, 2001, Yeo et al, 2004).

The repertoire of protein isoforms produced from a gene through alternative splicing can be regulated developmentally, in a tissue specific manner, and/or in response to cellular signaling events. The protein product of a gene can be functionally altered by alternative splicing through the addition or subtraction of an entire protein domain (Resch et al, 2004) or through the insertion or deletion of a small peptide (Stetefeld and Ruegg, 2005). Even small alterations to a protein sequence introduced through alternative splicing can have large functional consequences, such as changing a protein's subcellular localization, altering its ability to be phosphorylated by a kinase, or affecting its ability to bind to a ligand. Numerous examples of changes in structure and function of a protein as a consequence of alternative splicing have been documented for transcription factors, signaling molecules, extracellular matrix and cell adhesion molecules, ion channels, receptors, RNA binding proteins, cytoskeletal proteins, factors that regulate apoptosis, etc. (Lopez, 1998). Alternative splicing can also drastically alter the gene product by

introduction of a stop codon, resulting in either expression of a truncated protein product or targeting of the mRNA for degradation by the nonsense-mediated decay (NMD) pathway (Baek and Green, 2005). The many different types of alternative splicing that can result in the generation of functionally distinct mRNAs are depicted in Figure 1.

Just as the essential splicing signals of the 5' splice site, 3' splice site, and branchpoint consensus sequences were identified through comparison of existing sequence data in the early years of the splicing field, considerable insight can now be gained in the understanding of the regulation of alternative pre-mRNA splicing and its consequent effects on the proteome merely through in-depth comparisons of the data already available from the sequencing of the human genome and genomes of other organisms and the information contained within EST databases. Such bioinformatic analyses have already produced some very interesting observations about alternative splicing and how it shapes the proteome. In a recent study, a database of human alternative splicing events was compiled from genomic and EST data and translated *in silico* to produce an "alternatively spliced protein isoforms" database, which was then curated to include only alternative isoforms that would result in a functional protein product different from the major protein isoform. When these *in silico* generated protein isoforms for each gene were compared to protein domain databases, striking effects of alternative splicing on the composition of the proteome and the control of biological pathways were revealed. In approximately half of these cases of alternative protein isoforms, entire domains were removed as a consequence of alternative splicing (Resch et al, 2004). The majority of these splicing-regulated domains are known protein-protein interaction domains, and their removal by alternative splicing in many cases completely

redirects the outcome of biological pathways by modulating protein-protein interaction networks. For example, transcription factors are converted from transcriptional activators to transcriptional repressors, and vesicle secretion can be turned off when GTPases that regulate vesicle plasma membrane targeting are prevented from interacting with exocyst complexes at the plasma membrane (Resch et al, 2004).

The recent availability of the sequences of entire genomes has made possible the development of new technologies for the large scale profiling of alternative splicing. Conventional microarray technology has been adapted by the splicing field to catalog splice variants using exon and/or splice site junction probes (Johnson et al, 2003). Alternative splicing can also be profiled using a new fiber optic array technology (Yeakley et al, 2002). Global scale studies of mRNAs using these new technologies show that at least in some cases regulated alternative splicing events that dictate protein isoform expression for multiple factors in a biological pathway are controlled in a coordinated manner (Ule et al, 2005), yet unexpectedly indicate that sets of genes whose expression is regulated by alternative splicing in specific cell and tissue types do not significantly overlap with sets of genes regulated by transcription in the same cells and tissues (Le et al, 2004, Pan et al 2004, 2006). Thus, transcription regulates the quantity of the proteome, whereas alternative splicing regulates the diversity of the proteome.

1.1.3 Disease caused by defects in pre-mRNA splicing

Defects in pre-mRNA splicing are associated with disease in several different ways. *Cis*-acting splicing disorders occur when mutations in a gene directly interfere with normal splicing mechanisms for a specific pre-mRNA by damaging splice sites or

splicing regulatory elements, leading to disease by preventing expression of a protein. In fact, 15% of disease-associated single nucleotide mutations affect splice donor and acceptor sites (Krawczak et al, 2007), while a great many more point mutations disrupt splicing by weakening or strengthening splicing regulatory elements including exonic splicing enhancers (ESEs) and exonic splicing silencers (ESSs) (Cartegni et al, 2002, Licatalosi and Darnell, 2006). Point mutations not affecting splice sites were formerly assumed to cause disease through altering amino acid coding sequences, but a great number of these have been shown to instead result in skipping of entire exons or activation of cryptic splice sites, leading to expression of truncated protein or the introduction of a premature termination codon that targets the mRNA to NMD so that no protein product is even produced. Human disease genes with mutations affecting ESEs include *BRCA1* (breast cancer 1, early onset), *CFTR* (cystic fibrosis transmembrane conductance regulator), *HPRT1* (hypoxanthine phosphoribosyltransferase 1), *MAPT* (microtubule-associated protein tau), *SMN1* (survival of motor neuron 1), *FBN1* (fibrillin-1), *NF1* (neurofibromatosis 1), and *CD40* (CD40 molecule, TNF receptor superfamily member 5), *GH-1* (growth hormone), *VWF* (Von Willebrand factor), *GJA9* (connexin 36), and *APC* (adenomatosis polyposis coli) (Cartegni et al, 2002, Faustino and Cooper, 2003, Caputi et al, 2002, Colapietro et al, 2003, Ferrari et al, 2001, Moseley et al, 2002, James at al, 2004, Mas et al, 2004, Aretz et al, 2004). Splice site mutations can also cause exon skipping, activation of cryptic splice sites, creation of a pseudo-exon within an intron, or intron retention (Nakai and Sakamoto, 1994), which can similarly result in expression of truncated protein or no protein at all.

Trans-acting splicing disorders can occur when repeat expansion mutations in genes create dominant-negative sinks that sequester alternative splicing factors and prevent them from reaching their normal targets; this type of splicing misregulation has been implicated in both Myotonic Dystrophy and Fragile-X-Associated Tremor/Ataxia Syndrome (Faustino and Cooper, 2003, Ranum and Cooper, 2006, Ranum and Day, 2004). Lack of expression and ectopic expression of splicing regulatory factors can also lead to disease through splicing misregulation *in trans*, as documented in Prader Willi syndrome (Kishore and Stamm, 2006) and the paraneoplastic neurologic disorders (Licatalosi and Darnell, 2006), respectively.

Finally, cancer is also associated with both changes in the alternative splicing of single and multiple transcripts and alterations in the levels of splicing factors, but the mechanisms underlying these changes and their consequences are not always well understood. Significantly, cancer-specific alternative splicing often occurs in the absence of genomic mutations (Venables, 2004). Affected proteins include tumor supressors, cell signaling molecules, transcription factors, and extracellular matrix components (Ge et al, 1999, Kumar et al, 2003, Tennenbaum et al, 1995, Steinman et al, 2004). For example, changes in the expression of specific isoforms of the CD44 cell surface adhesion protein have been associated with metastasis in a number of tissues (Cooper and Dougherty, 1995, Faustino and Cooper, 2003).

Alterations in the relative levels of expression of splicing factors with antagonistic functions are seen in different tumor types (Trockman et al, 1997, Ghigna et al, 1998), and this has been proposed as a mechanism underlying the global changes in alternative splicing observed in tumor versus normal tissues (Xu et al, 2003, Wang et al, 2003).

Specifically, changes in the levels of SR proteins have been documented in several types of cancers (He et al, 2004, Fischer et al, 2004, Ghigna et al, 2005), and mammary tumor progression has been associated with progressive alterations in expression of SR proteins (Stickeler et al, 1999). Overexpression of the SR protein SF2/ASF was shown to transform fibroblasts and induce tumor formation in nude mice, concomitant with changes in alternative splicing of tumor supressors (Karni et al, 2007).

1.2 Pre-mRNA and the spliceosome

1.2.1 Pre-mRNA splice sites

Pre-mRNA splicing is highly complex and involves interplay between multiple pre-mRNA sequence elements and trans-acting protein and snRNA factors. The spliceosome is a "multi-megadalton machine", composed of some 300 different polypeptides in addition to its complement of spliceosomal snRNAs (Jurica and Moore, 2003), that is responsible for the daunting task of recognizing what are often quite degenerate splicing signals with extraordinary fidelity. During the course of splicing, to remove an intron and join together the exons flanking it the spliceosome must correctly identify a set of splice site signals consisting of the 5' splice site (5' ss) that overlaps the boundary of the upstream exon and the intron, intronic sequences consisting of the the branchpoint sequence (BPS) and downstream pyrimidine tract (Py), and the 3' splice site (3' ss) that overlaps the boundary of the intron and the downstream exon (Reed, 1996).

There are at least two classes of spliceosomal introns, each recognized by a different spliceosome. U2-dependent introns account for 99.9% of all introns and are recognized by the major spliceosome, composed of the U1, U2, U4/U6, and U5 snRNPs. U2

dependent introns have degenerate branchpoint sequences and 5' and 3' splice site consensus sequences. U12-dependent introns account for only 0.1% of introns and are recognized by the minor spliceosome, composed of the U11, U12, U4atac/U6atac, and U5 snRNPs. Unlike U2-type introns, U12-dependent introns have highly conserved 5'ss signals and branch point consensus sequences, but lack a pyrimidine tract (Burge et al, 1999). Consensus motifs for 5'splice sites, branch sites, and 3' splice sites recognized by the major and minor spliceosomes are depicted in Figure 2.

Notably, in higher eukaryotes these essential splicing signals recognized by the major spliceosome do not contain enough information to specify the locations of the exons (Burge et al 1999, Sun and Chasin, 2000). Human pre-mRNAs are typically tens of thousands of nucleotides (nt) in length, and consist of small exons of 150-300nt flanked by introns of approximately 3000nt in length on average (Burge et al, 1999, Lander et al, 2001). The 5' splice site consensus is MAG/**GU**RAGU (where the GU in bold lettering is invariant, R equals A or G, and M equals A or C), the branch point has the extremely loose consensus of YNYURAY (where Y equals C or U, and R equals A or G), and the 3' splice site consensus is merely CAG/G (Sun and Chasin, 2000). In human genes donor sites should occur at random once every 290nt, and acceptor sites (consisting of the pyrimidine tract and 3' splice site) should occur every 450nt (Burge et al, 1999). The great size of human genes coupled with the degeneracy of these essential splicing signals creates a situation where authentic splice sites are camouflaged among many pseudo-splice sites, sequences that by virtue of their conformity to the consensus sequences look as much like a splice site as the authentic splice sites (Sun and Chasin, 2000).

1.2.2 Spliceosome assembly

During the process of splicing, different spliceosomal complexes--E, A, B, and C--assemble onto the pre-mRNA (Figure 3), directed in large part by the RNA sequences at the splice sites. Spliceosomal complex assembly is similar for both the major and minor spliceosomes; assembly of the major spliceosome is described below. (For more information about the minor spliceosome and its assembly, see Burge et al, 1999.) The first complex to assemble on the pre-mRNA is the E complex, which commits the pre-mRNA to the splicing pathway. For the assembly of E complex, U1 snRNP is recruited and the U1 snRNA base pairs to the 5'ss (Michaud and Reed, 1991), the protein factor SF1/mBBP binds to the branchpoint (Abovich and Rosbash, 1997), the large subunit of the U2 Auxiliary Factor, U2AF65, binds to the pyrimidine tract (Ruskin et al, 1988, Zamore and Green, 1989, Bennett et al, 1992), and the small subunit, U2AF35, to the 3' ss AG (Wu et al, 1999, Merendino et al, 1999, Zorio and Blumenthal, 1999). Conversion of the E complex to the A complex is achieved when the U2 snRNP binds and displaces SF1, permitting the U2 snRNA to base pair to the branchpoint (Bennett et al, 1992). B complex is formed upon entry of the tri-snRNP, consisting of the U4, U5, and U6 snRNPs, into the spliceosome (Reed and Palandjian, 1997). Transition from B complex to C complex involves significant intraspliceosomal rearrangements, resulting in the replacement of U1 snRNP by U6 snRNP at the 5'ss and bridging of the 5' and 3' exons by U5 snRNP (Staley and Guthrie, 1998). The C complex is the first catalytically active form of the spliceosome.

There are a large number of protein factors that aid in the assembly of and transition between the different spliceosomal complexes. The SR proteins are essential splicing factors that help commit pre-mRNAs to splicing in the formation of the E complex (Fu, 1993, Reed, 1996) by recruitment of the U1 snRNP to the 5' splice site through interactions with the U1 snRNP protein U1-70K (Kohtz et al, 1994) and indirect recruitment of U2 snRNP through interactions with the small subunit of U2AF (Zuo and Maniatis, 1996). (See section 1.3.1 for more information about how SR proteins help to promote assembly of the spliceosome.)

Helicases help with the many rearrangements of RNA-RNA and RNA-protein interactions required for the transitions from E to A, from A to B, and from B to C complexes, and various proteins help to stabilize these interactions. The DEAD box helicase Prp5 and the DECD box helicase UAP56 assist in the replacement of SF1 with U2 snRNA at the branchpoint (Staley and Guthrie, 1998) for the transition between the E and A complexes. The ternary splicing factor complex SF3a, composed of the protein subunits SF3a60, SF3a66, and SF3a120, is also required for the stable association of U2 snRNP with the branchpoint required for E to A transition (Weist et al, 1996).

In the B complex, the 5' splice site and the branchpoint are brought together through interactions mediated by the tri-snRNP. The Prp28 and Brr2 helicases mediate the switch from U1 snRNA to U6 snRNA base pairing at the 5' splice site and the destabilization of U4 snRNA base pairing to U6 snRNA, respectively, upon the entry of the tri-snRNP to form B complex (Staley and Guthrie, 1998). In the final steps of spliceosomal rearrangements to form C complex, a base pairing interaction between U6 snRNA and U2 snRNA juxtaposes the 5' splice site and branch site to produce the catalytic site for

the first transesterification reaction; this transition is assisted by the actions of the RNA helicase Prp2 (Teigelkamp et al, 1994, Roy et al, 1995).

1.2.3 Splicing catalysis

Pre-mRNA is spliced to mRNA in two sequential transesterification reactions that occur within C complex (Figure 4). In the first transesterification reaction, the 5' exon is cleaved from the 5' splice site, and the 5' splice site end of the intron is covalently linked to the branch point nucleotide in a 2'-5' phosphodiester bond. The products of the first transesterification reaction are the 5' exon with a free 3'-OH terminus and the lariat intermediate consisting of the branched intron connected to the 3' exon. In the second transesterification reaction, the 3' exon is cleaved from the lariat intermediate at the 3' splice site, the exons are ligated together via a 3'-5' phosphodiester bond, and the intron is released as a lariat with a free 3'-OH terminus (Moore et al, 1993).

The spliceosome shares several important similarities with Group II ribozymes, and employs identical reaction pathways and stereo-chemistry for the two transesterification reactions (Moore and Sharp, 1993). The spliceosome and Group II ribozymes both require a metal ion cofactor and use the same catalytic strategy, in which a magnesium ion coordinates with the 3′-oxyanion leaving groups to permit a build-up of negative charge in the transition state for both catalytic steps of splicing (Gordon et al, 2000, Sontheimer, 2001).

Recent work implicates the U2 and U6 snRNAs as the putative catalysts for the first transesterification reaction (Valadkhan and Manley, 2001 and 2003), demonstrating that the spliceosome is likely a ribozyme, as has long been suspected. In these experiments, a

protein-free human U6–U2 RNA complex can bind a small RNA consisting of the intron branch site sequence and juxtapose the catalytic AGC triad conserved in U6 snRNA and the branch site adenosine prior to catalyzing a magnesium-dependent covalent linkage between the branch site adenosine, the reactive group of the first step of splicing, and the AGC triad.

The protein factor Prp8 helps position active site elements and coordinates the progression of spliceosome catalysis (Grainger and Beggs, 2005). Helicases are also required during splicing catalysis for conformational rearrangements of the spliceosome between the first and second transesterification reactions, including Prp16, Prp17, and Slu7 (Schwer and Guthrie, 1992, Jones et al, 1995, Umen and Guthrie, 1995, and Zhou and Reed, 1998).

1.3 Regulators of splicing: cis elements and trans-acting factors

1.3.1 Exonic Splicing Enhancers and SR proteins

Exonic splicing enhancers (ESEs) are degenerate 6-8nt motifs within the exons of pre-mRNAs that promote exon inclusion through the action of SR proteins (Cartegni et al, 2002, Blencowe, 2000, Graveley, 2000, Manley and Tacke, 1996, Fu, 1995). SR proteins are a family of conserved splicing factors (Zahler et al, 1992) that possess either one or two N-terminal RNA recognition motifs (RRM) and a C-terminal arginine and serine-rich (RS) domain (Birney et al, 1993). SR proteins bind to pre-mRNA via their RRM domains, while their RS domains are thought to function as protein-protein interaction domains for the recruitment of components of the basal splicing machinery.

There are two models for how SR proteins function at early steps of splicing to promote inclusion of ESE-dependent exons. In the recruitment model, the SR protein binds the ESE through its RRM domain(s) and acts as an adapter to recruit components of the splicing machinery via protein-protein interactions mediated by its RS domain (Graveley, 2000). In the antagonism model, the SR protein bound to the ESE counteracts the negative effect of an exonic splicing silencer (ESS) and the inhibitory protein(s) bound to it (Kan and Green, 1999) (see section 1.3.2 and Figure 6). The recruitment and antagonism models are not mutually exclusive, and SR proteins may perform either or both functions depending upon the pre-mRNA context in which they act (Cartegni et al, 2002).

Exonic splicing enhancers were first identified as purine-rich elements in the mouse immunoglobulin μ heavy chain pre-mRNA (Watakabe et al 1993), and subsequently a large number of both purine-rich and non-purine-rich sequences have been shown to function as ESEs. Many studies have been undertaken to systematically identify the ESE motifs recognized by different SR protein family members. Some of these studies identified high affinity binding sites for specific SR proteins (Tacke and Manley, 1995, Cavaloc et al, 1999), while others identified motifs that act as functional ESEs in *in vitro* splicing (Liu et al, 1998, Liu et al, 2000). ESE consensus motifs for specific SR proteins are surprisingly degenerate, which may reflect the evolutionary constraints placed upon the exonic sequences in which they reside; in several cases ESE consensus sequences can be recognized by more than one SR protein (Tacke and Manley, 1995, Liu et al, 1998).

Significantly higher numbers of ESE motifs are located in exons than in the introns flanking them (Liu et al, 1998, Liu et al, 2000, Fairbrother et al, 2002, Zhang and Chasin,

2004), and more ESE motifs are located in authentic exons than in pseudoexons (Wang et al, 2005). The presence of multiple ESE elements within a single exon seems to aid in the identification of the exon by increasing the probability of an interaction between an SR protein and the splicing machinery (Hertel and Maniatis, 1998). Interestingly, several studies have found fewer ESEs in alternative exons than in constitutive exons (Zhang and Chasin, 2004, Fairbrother et al, 2002, Wang et al, 2005), suggesting that exon definition may be weaker for alternative splicing events.

1.3.2 Exonic Splicing Silencers and hnRNP proteins

Exonic Splicing Silencers (ESSs) are pre-mRNA regulatory motifs that promote exon skipping or intron retention (Ladd and Cooper, 2002). Although many exonic splicing silencers have been identified, unlike ESEs, ESS sequences share little obvious similarity (Zheng, 2004). In addition, the mechanisms by which ESSs act to inhibit splicing are not well understood. In many cases ESSs are known to act by recruiting specific inhibitory factors (Zahler et al, 2004, Domsic et al, 2003, Caputi et al, 1999, Chen et al, 1999, Del Gatto-Konczak et al, 1999, Baba-Aissa et al, 1998, Si et al, 1997, Staffa et al, 1997, Del Gatto-Konczak et al, 1996, Gallego et al, 1996, Del Gatto-Konczak et al, 1995, Amendt et al, 1995, Caputi et al, 1994, Amendt et al, 1994, Graham et al, 1992), but how these factors accomplish the silencing of splicing is not always clear. A few studies have shed light on the varying mechanisms by which ESSs may exert their negative effects on splicing. In one study, hnRNPA1 was shown to bind to a single ESS and through cooperative binding propagate along an exon to prevent an SR protein from binding to, or displace it from, its cognate ESE (Zhu et al, 2001). While in some cases

ESS elements can function in heterologous contexts (Wang et al, 2004), in the case of the fibronectin EDA exon the activity of the ESS is completely context-dependent; this ESS is thought to function by disrupting an RNA secondary structure that functions to display an adjacent ESE for recognition by its cognate SR protein(s) (Muro et al, 1999). Irrespective of their mechanism of action, an enrichment of exonic splicing silencers in pseudoexons has been postulated as one possible mechanism by which the identification of authentic exons may be achieved (Sironi et al, 2004, Zhang and Chasin, 2004).

Exonic splicing silencers are often recognized by members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family, including hnRNPA1, hnRNPA2/B1, hnRNP C1/C2, hnRNP F, hnRNP H, hnRNP I (also known as PTB), hnRNP L, and hnRNP K. hnRNP protein family members vary in structure, but all possess at least one RNA binding domain. Both hnRNPA1 and hnRNPA2/B1 have two N-terminal RRM domains and a C-terminal glycine-rich domain (Dreyfuss et al, 1993). hnRNP C1/C2 possesses only one RRM domain which is followed by a domain rich in aspartic acid and glutamic acid (Dreyfuss et al, 1993). hnRNP I/PTB is known to regulate a large number of different splicing events (Spellman et al, 2005); its four RRM domains bound to their target motifs have been extensively studied (Amir-Ahmady et al, 2005, Oberstrauss et al, 2005). hnRNP L is similar in structure to hnRNP I/PTB and also has four RRM domains, while hnRNP K has another type of RNA-binding motif, the KH domain (Dreyfuss et al, 1993).

1.3.3 Trans-acting factors and splice site recognition in constitutive splicing

In yeast, pre-mRNAs are much shorter than in mammals, possess splice site signals which are very well conserved, and typically consist of longer exons interspersed with one, or occasionally two, shorter introns. Evidence suggests that yeast use intron definition, or pairing of splice sites across the intron, as an initial step in splicing (Berget, 1995, Guthrie, 1991, Ruby and Abelson, 1991). In vertebrates, the essential splicing signals of the 5'ss, the branchpoint, the pyrimidine tract, and the 3'ss are clearly insufficient to specify the locations of authentic exons, as intronic sequences contain many pseudo splice sites. However, at least some of the additional information content required for accurate identification of exons by the spliceosome is contained within the ESE and ESS sequences. Introns are a great deal longer than the exons in the much longer vertebrate pre-mRNAs, and the initial steps of splicing are thought to occur through exon definition, in which complexes are initially assembled across the smaller exons to specify the locations of splice sites relative to the exonic sequences (Berget, 1995).

The first support for the exon definition model of splice site recognition came when Berget and coworkers observed that the presence of a 5'ss at the downstream end of an exon promotes splicing of the immediately upstream intron (Robberson et al, 1990). In the exon definition model, splice sites are recognized in exon defined pairings of upstream 3' splice sites and downstream 5' splice sites. SR proteins bound to ESEs are believed to facilitate exon definition by recruitment of U2AF to the upstream pyrimidine tract through interactions with its small subunit U2AF35 (Wu and Maniatis, 1993, Wang et al, 1995, Zuo and Maniatis, 1996) and via recruitment of U1 snRNP to the downstream

5'ss through interactions with the U1-specific protein U1-70K (Kohtz et al, 1994, Wu and Maniatis, 1993) (Figure 6).

Ultimately, bringing together of the splice sites across the intron to juxtapose the 5'ss and the branchpoint must occur for splicing to be accomplished. Insight about how splice site pairing might occur across long vertebrate introns has been gained from *in vitro* trans splicing studies, in which an RNA containing both an exon and a 5'ss can be trans-spliced efficiently to another RNA containing a 3'ss if it contains either a downstream 5'ss or an ESE (Bruzik and Maniatis, 1995, Chiara and Reed, 1995). This trans-splicing reaction can also occur if the 3'ss RNA is preassembled into the A complex (Chiara and Reed, 1995), suggesting that intron definition may occur subsequent to the establishment of A complex, ie after both U1 snRNP and U2 snRNP have been recruited with the assistance of SR proteins. Thus, in higher organisms recognition of the correct splice sites and thus definition of authentic exons and their flanking intronic sequences is accomplished by the accumulated recognition of multiple signals and the establishment of a network of interactions of trans-acting factors across both exons and introns.

The ability of ESE-bound SR proteins to facilitate exon definition can be antagonized by the effects of ESS-bound hnRNPs (Figure 6), though the mechanisms by which splicing silencers exert their negative effects on exon definition are not entirely understood. Therefore the efficiency of exon definition in constitutive splicing and the outcome of exon inclusion versus exon skipping in alternative splicing are thought to be specified through a combinatorial control exerted by SR proteins and hnRNPs. Indeed, the ratios between specific SR proteins and hnRNP proteins are under tight spatial and temporal control, and vary in different tissue types and during different developmental

stages (Kamma et al, 1995, Hanamura et al, 1998, Pollard et al, 2000). In addition, some SR proteins and hnRNPs shuttle between the nucleus and cytoplasm (Caceres et al, 1998, Pinol-Roma and Dreyfuss, 1992), and the concentrations of these factors in the nucleus can be altered as a consequence of signal transduction events (van der Houven van Oordt et al, 2000, Xie et al, 2003, Allemand et al, 2005).

1.4 Modular organization of SR protein domains

1.4.1 The RRM domain and its functions

SR protein family members have either one or two N-terminal RRM domains (Birney et al, 1993) (Figure 5). All SR proteins have in common the first and typical RRM1 domain, while SR proteins with two RRM domains possess a second "pseudo" RRM2 domain. The RRM domain is one of the most abundant protein domains in eukaryotes and is coded for by approximately 2% of human genes (Maris et al, 2005). Although first characterized as an RNA-binding domain (Dreyfuss et al, 1988), the RRM is a versatile protein domain that can engage in interactions with both nucleic acids and proteins (Maris et al, 2005). RNA binding is thought to be the principal function of the RRM domain(s) of SR proteins, bringing the proteins to ESEs so that their RS domains can engage in protein-protein interactions which promote spliceosomal assembly (see sections 1.3.1 and 1.4.2). The substrate specificity of SR proteins is determined by their RRMs (Chandler et al, 1997, Mayeda et al, 1999). At this time, protein-protein interactions through the RRM domains of SR proteins have not been documented.

The RRM domain is typically around 90 amino acids in length, and adopts a canonical secondary structure in a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology comprised of a four-stranded

antiparallel beta sheet surface, with beta strands in the order of $\beta_4\beta_1\beta_3\beta_2$, backed by two perpendicular alpha helices (Nagai et al, 1990, Maris et al, 2005, Figure 7). The primary amino acid sequences of RRMs are characterized by two conserved segments referred to as RNP1 (with consensus K/R-G-F/Y-G/A-F/Y-V/I/L-X-F/Y, and located in the β_3 strand) and RNP2 (with consensus V/I/L-F/Y-V/I/L-X-N/L, and located in the β_1 strand) composed primarily of aromatic and positively charged residues that are involved directly in RNA binding (Dreyfuss et al, 1988, Birney et al, 1993, Auweter et al, 2006).

The structures of many RRM domains have been solved by either NMR or crystallography, either in isolated form or in complex with nucleic acid and/or protein (Maris et al, 2005, Auweter et al, 2006), and much is already known about the modes by which RRMs can interact with RNA and how specific interactions are achieved. From ten published structures of RRMs in complex with nucleic acids, Maris et al have derived a common structural archetype for RRM-nucleic acid interactions consisting of three characteristic contacts that involve two nucleotides and conserved protein side chains of the RRM in the RNP2 and RNP1 motifs of the central beta strands β_1 and β_3, respectively. In the first two contacts, aromatic rings from β_1 (RNP2 position 2) and β_3 (RNP1 position 5) engage in stacking interactions with two nucleotides, orienting the nucleic acid strand with its 5' end located on the $\beta_4\beta_1$ side of the beta sheet surface and the 3' end on the $\beta_3\beta_2$ side. In the third contact, a third aromatic residue from β_3 (RNP1 position 3) interacts hydrophobically with the sugar rings of the two nucleotides (Maris et al, 2005, Auweter et al, 2006). These interactions typify the core dinucleotide binding pocket common to all RRM domains, but the RRM domain can accommodate binding of

between two and seven nucleotides, with the typical RRM containing four nucleotide binding sites (Auweter et al, 2006).

Binding of RRM domains to single stranded RNA is understood to initially occur due to the electrostatic affinity of the positively charged surface of the RRM domain to RNA and as a consequence of the hydrophobic nature of the beta sheet surface. Subsequent to this initial sequence-independent attraction, sequence-specificity of binding is achieved through shape recognition of the RNA molecule by binding pockets provided by the protein, and the RNA is locked into place through stacking and hydrogen bonds (Auweter et al, 2006). In the case of SR proteins, the RRM domains are thought to bind at least somewhat sequence specifically to single stranded RNA. The solution structures of the isolated RRM1 domain of the SR proteins 9G8 and SRp20 have recently been solved, in addition to the structure of SRp20 in complex with the RNA 5'CAUC3' (Hargous et al, 2006, Figure 7). The SRp20 RRM domain-RNA structure reveals that four nucleotides are contacted by the beta sheet surface, each interacting with one of four aromatic residues. The mode of interaction of SRp20 with RNA fits precisely with the common structural archetype for the core dinucleotide binding pocket derived by Maris et al; nucleotides C1 and A2 stack on RNP2 via Y13 in β_1 and RNP2 via F50 in β_3, and residue F48 in β_3 inserts between the sugar rings of C1 and A2 (Hargous et al, 2006, Figure 7). Interestingly, the structure reveals that of the four nucleotides bound by the SRp20 RRM, only the 5' cytidine is recognized in a sequence specific manner, perhaps explaining the degeneracy in sequences recognized by SRp20 (Cavaloc et al, 1999, Schaal and Maniatis, 1999).

1.4.2 The RS domain and its functions

SR proteins possess C-terminal "RS" domains, so named because they are predominantly composed of multiple segments of alternating arginine and serine dipeptides (Figure 5). The primary function of the RS domain in promoting splicing is thought to be to engage in protein-protein interactions to recruit essential components of the splicing machinery. In the mid-1990's, the SR protein SF2/ASF was shown to interact with itself and with the U1-specific protein U1-70K and the small subunit of U2AF, U2AF35, and these protein-protein interactions required the RS domains of each protein (Wu and Maniatis, 1993, Kohtz et al, 1994). Subsequently it was proposed that SR proteins can recruit the U1 snRNP to the 5'ss through SR protein RS domain-mediated interactions with the U1 snRNP-specific protein U1-70K (Staknis and Reed, 1994). However, the RS domain of SF2/ASF alone (in its unphosphorylated state) is unable to interact with U1-70K *in vitro* (Xiao and Manley, 1997). Enhancer-bound SR proteins are also thought to escort U2AF65 to the pyrimidine tract through an RS domain-mediated recruitment of U2AF35 (Zuo and Maniatis, 1996, Wang et al, 1995, Bouck et al, 1998). This role for SR proteins in bringing U2AF65 to the pyrimidine tract is supported by several experiments in which improving the pyrimidine tract can relieve the requirement for an ESE for pre-mRNAs with enhancer-dependent introns (Tian and Maniatis, 1994, Lorson and Androphy, 2000, Graveley et al, 2001). However, other experiments failed to observe changes in U2AF recruitment in the presence versus in the absence of an ESE (Li and Blencowe, 1999, Kan and Green, 1999).

SR protein RS domains vary in length (Figure 5), and the potency of the RS domain to activate splicing has a direct relationship to the number of RS dipeptides present

(Graveley et al, 1998, Cartegni and Krainer, 2003). Interestingly, SR protein RS domains are at least somewhat interchangeable (Chandler et al, 1997, Wang et al, 1998, Graveley et al, 1998, Mayeda et al, 1999), and can even function to promote ESE-dependent splicing when tethered to a heterologous RNA binding domain (Graveley and Maniatis, 1998, Cartegni and Krainer, 2003).

The serine residues within RS domains are targets of phosphorylation by multiple kinases, including the SR protein kinases SRPK1 (Gui et al, 1994a) and SRPK2 (Wang et al, 1998), Clk/Sty (Colwill et al, 1996), and DNA topoisomerase I (Rossi et al, 1996). Phosphorylation of RS domains influences both the subcellular localization of SR proteins (Gui et al, 1994b, Colwill et al, 1996, Misteli et al, 1998) and RS domain-mediated protein-protein interactions (Wang et al, 1998, Xiao and Manley, 1998). The phosphorylation state of the RS domain has a significant influence on SR protein function, as both hyper- and hypophosphorylated SR proteins are unable to support splicing (Kanopka et al, 1998, Prasad et al, 1999, Sanford and Bruzik, 1999), and SR protein kinases have been shown to influence the splicing of specific pre-mRNAs during development (Du et al, 1998, Sanford and Bruzik, 1999) and viral infection (Kanopka et al, 1998).

The RS domain may also function in direct pre-mRNA contacts during the course of pre-mRNA splicing. An ESE-bound RS domain has been shown to interact directly with the branchpoint in A complex; this RS domain-branchpoint interaction is transient, as it was excluded from E complex, identified in A complex, and subsequently excluded in B complex (Shen and Green, 2004). This ESE-bound RS domain-RNA interaction was proposed to occur with a hypophosphorylated RS domain through an interaction of the

positively charged arginine residues of the RS domain with the negatively charged phosphate backbone of the RNA (Shen and Green, 2004, Shen et al, 2004, Valcarcel et al, 1996), in which case phosphorylation of the serine residues of the RS domain could switch the ability of the RS domain to interact with RNA and proteins (Hertel and Graveley, 2005). This is an interesting hypothesis, as it has already been demonstrated that dynamic changes in SR protein phosphorylation state are required for the completion of spliceosome assembly (Cao et al, 1997, Xiao and Manley, 1998).

1.5 References

Abovich, N., Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell *89*, 403-412.

Allemand, E., Guil, S., Myers, M., Moscat, J., Caceres, J. F., Krainer, A. R. (2005). Regulation of heterogeneous nuclear ribonucleoprotein A1 transport by phosphorylation in cells stressed by osmotic shock. Proc Natl Acad Sci USA *102*, 3605-3610.

Amendt, B. A., Hesslein, D., Chang, L. J., Stoltzfus, C. M. (1994). Presence of negative and positive *cis*-acting RNA splicing elements within and flanking the first Tat-coding exon of human immunodeficiency virus type 1. Mol Cell Biol *14*, 3960–3970.

Amendt, B. A., Si, Z. H., Stoltzfus, C. M. (1995). Presence of exon splicing silencers within human immunodeficiency virus type 1 Tat exon 2 and Tat-Rev exon 3: evidence for inhibition mediated by cellular factors. Mol Cell Biol *15*, 4606–4615.

Amir-Ahmady, B., Boutz, P. L., Markovtsov, V., Phillips, M. L., Black, D. L. (2005). Exon repression by polypyrimidine tract binding protein. RNA *11*, 699-716.

Aretz, S., Uhlhaas, S., Sun, Y., Pagenstecher, C., Mangold, E., Caspari, R., Moslein, G., Schulmann, K., Propping, P., Friedl, W. (2004). Hum Mutat *24*, 370-380.

Auweter, S. D., Oberstrass, F. C., Allain, F. H. (2006). Sequence-specific binding of single-stranded RNA: is there a code for recognition? Nucleic Acids Res *34*, 4943-59.

Baba-Aissa, F., Van Den Bosch, L., Wuytack, F., Raeymaekers, L., Casteels, R. (1998). Regulation of the sarco/endoplasmic reticulum Ca(2+)-ATPase (SERCA) 2 gene transcript in neuronal cells. Brain Res Mol Brain Res *55*, 92–100.

Baek, D., Green, P. (2005). Sequence conservation, relative isoform frequencies, and nonsense-mediated decay in evolutionarily conserved alternative splicing. Proc Natl Acad Sci USA *102*, 12813-12818.

Bennett, M., Michaud, S., Kingston, J, Reed, R. (1992). Protein components specifically associated with prespliceosome and spliceosome complexes. Genes Dev *6*, 1986-2000.

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

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

Berget, S. M., Robberson, B. L. (1986). U1, U2, and U4/U6 small nuclear ribonucleoproteins are required for *in vitro* splicing but not polyadenylation. Cell *46*, 691-696.

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

Birney, E., Kumar, S., Krainer, A. R. (1993). Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. Nucleic Acids Res *21***,** 5803-5816.

Black, D. L., Chabot, B., Steitz, J.A. (1985). U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. Cell *42*, 737-50.

Black, D. L., Steitz, J. A. (1986). Pre-mRNA splicing in vitro requires intact U4/U6 small nuclear ribonucleoprotein. Cell *46*, 697-704.

Blencowe, B. J. (2000). Exonic splicing enhancers: mechanism of action, diversity, and role in human genetic diseases. Trends Biochem Sci *25*,106-10.

Blencowe, B. J. (2006). Alternative splicing: new insights from global analyses. Cell *126*, 37-47.

Bouck, J., Fu, X.-D., Skalka, A. M., Katz, R. A. (1998). Role of the constitutive splicing factors U2AF65 and SAP49 in suboptimal RNA splicing of novel retroviral mutants. J Biol Chem *273*, 15169-15176.

Breathnach, R., Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Annu Rev Biochem *50*, 349-383.

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

Brody, E., Abelson, J. (1985). The "spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. Science *228*, 963-967.

Bruzik, J., Maniatis, T. (1995). Enhancer dependent interaction between 5' and 3' splice sites in trans. Proc Natl Acad Sci USA *92*, 7056-7059.

Caceres, J. F., Screaton, G. R., Krainer, A. R. (1998). A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. Genes Dev *12*, 55-66.

Cao, W., Jamison, S. F., 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., Casari, G., Guenzi, S., Tagliabue, R., Sidoli, A., Melo, C. A., Baralle, F.E. (1994). A novel bipartite splicing enhancer modulates the differential processing of the human fibronectin EDA exon. Nucleic Acids Res *22*, 1018–1022.

Caputi, M., Kendzior, R. J., Jr., Beemon, K. L. (2002). A nonsense mutation in the fibrillin-1 gene of a Marfan syndrome patient induces NMD and disrupts an exonic splicing enhancer. Genes Dev *16*, 1754-1759.

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

Cartegni, L., Krainer, A. R. (2003). Correction of disease-associated exon skipping by synthetic exon-specific activators. Nat Struct Biol *10*, 120-5.

Cartegni, L., Chew, S L., Krainer, A. R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet *3*, 285-298.

Cavaloc, Y., Bourgeois, C. F., Kister, L., Stevenin, J. (1994). The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. RNA *5*, 468-483.

Chabot, B., Black, D. L., LeMaster, D. M., Steitz, J.A. (1985). The 3' splice site of pre-messenger RNA is recognized by a small nuclear ribonucleoprotein. Science *230*, 1344-1349.

Chandler, S. D., Mayeda, A., Yeakley, J. M., Krainer, A. R., and Fu, X.-D. (1997). RNA splicing specificity determined by the coordinated action of RNA recognition motifs in SR proteins. Proc Natl Acad Sci USA *94*, 3596-3601.

Chen, C. D., Kobayashi, R., Helfman, D. M. (1999). Binding of hnRNPH to an exonic splicing silencer is involved in the regulation of alternative splicing by the rat beta-tropomyosin gene. Genes Dev *13*, 593-606.

Cheng, S.-C., Abelson, J. (1987). Spliceosome assembly in yeast. Genes Dev *1*, 1014-1027.

Chiara, M. D. and Reed, R. (1995). A two-step mechanism for 5' and 3' splice site pairing. Nature *375*, 510-513.

Chow, L.T., Gelinas, R. E., Broker, T. R., Roberts, R. J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell *12*, 1-8.

Colapietro, P., Gervasini, C., Natacci, F., Rossi, L., Riva, P., Larizza, L. (2003). NF1 exon7 skipping and sequence alterations in exonic splice enhancers (ESEs) in a neurofibromatosis 1 patient. Hum Genet *113*, 551-554.

Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J. L., Bell, J. C., Duncan, P. I. (1996). The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. EMBO J *15*, 265-275.

Cooper, D. L., Dougherty, G. J. (1995). To metastasize or not? Selection of CD44 splice sites. Nat Med *1*, 635-637.

Deimel, B., Louis, Ch., Sekeris, C. E. (1977). The presence of small molecular weight RNAs in nuclear ribonucleoprotein particles carrying HnRNA. FEBS Lett *73*, 80-84.

Del Gatto, F., Breathnach, R. (1995). Exon and intron sequences, respectively, repress and activate splicing of a fibroblast growth factor receptor 2 alternative exon. Mol Cell Biol *15*, 4825–4834.

Del Gatto, F., Gesnel, M. C., Breathnach, R. (1996). The exon sequence TAGG can inhibit splicing. Nucleic Acids Res *24*, 2017–2021.

Del Gatto-Konczak, F., Olive, M., Gesnel, M. C., Breathnach, R. (1999). hnRNPA1 recruited to an exon in vivo can function as an exonic splicing silencer. Mol Cell Biol *19*, 251-260.

Domsic, J. K., Wang, Y., Mayeda, A., Krainer, A. R., Stoltzfus, C. M. (2003). Human immunodeficiency virus type 1 hnRNPA/B-dependent exonic splicing silencer ESSV antagonizes binding of U2AF65 to viral polypyrimidine tracts. Mol Cell Biol *23*, 8762-8772.

Dreyfuss, G., Matunus, M. J., Pinol-Roma, S., Burd, C. G. (1993). hnRNP proteins and the biogenesis of mRNA. Annu Rev Biochem *62*, 289-321.

Dreyfuss, G., Swanson, M. S., Pinol-Roma, S. (1988). Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. Trends Biochem Sci *13*, 86-91.

Du, C., McGuffin, M. E., Dauwalder, B., Rabinow, L., Mattox, W. (1998). Protein phosphorylation plays a critical role in the regulation of alternative splicing and sex determination in Drosophila. Mol Cell *2*, 741-750.

Fairbrother, W. G., Yeh, R. F., Sharp, P. A., Burge, C. B. (2002). Predictive identification of exonic splicing enhancers in human genes. Science *297*, 1007-1013.

Faustino, N. A., Cooper, T. A. (2003). Pre-mRNA splicing and human disease. Genes Dev *17*, 419-437.

Ferrari, S., Giliani, S., Ilsalaco, A., Al-Ghonaium, A., Soresina, A. R., Loubser, M., Avanzini, M. A., Marconi, M., Badolato, R., Ugazio, A. G., Levy, Y., Catalan, N.,

Durandy, A., Tbakhi, A., Notarangelo, L. D., Plebani, A. (2001). Proc Acad Natl Sci USA *98*, 12614-12619.

Fischer, D. C., Noack, K., Runnebaum, I. B., Watermann, D. O., Kieback, D. G., Stamm, S., Stickeler, E. (2004). Expression of splicing factors in human ovarian cancer. Oncol Rep *11*, 1085–1090.

Frendewey, D., Keller, W. (1985). Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. Cell *42*, 355-367.

Fu, X.-D. (1993). Specific commitment of different pre-mRNAs to splicing by single SR proteins. Nature *365*, 82 – 85.

Fu, X.-D. (1995). The superfamily of arginine/serine-rich splicing factors. RNA *1*, 663-680.

Gallego, M. E., Sirand-Pugnet, P., Durosay, P., Clouet d'Orval, B., d'Aubenton-Carofa, Y., Brody, E., Expert-Bezancon, A., Marie. J. (1996). Tissue-specific splicing of two mutually exclusive exons of the chicken beta-tropomyosin pre-mRNA: positive and negative regulations. Biochimie *78*, 457–465.

Ge, K., DuHadaway, J., Du, W., Herlyn, M., Rodeck, U., Prendergast, G. C. (1999). Mechanism for elimination of a tumor suppressor: aberrant splicing of a brain-specific exon causes loss of function of Bin1 in melanoma. Proc Natl Acad Sci USA *96*, 9689–9694.

Ghigna, C., Giordano, S., Shen, H., Benvenuto, F., Castiglioni, F., Comoglio, P. M., Green, M. R., Riva, S., Biamonti, G. (2005). Cell motility is controlled by SF2/ASF through alternative splicing of the *Ron* protooncogene. Mol Cell *20*, 881–890.

Ghigna, C., Moroni, M., Porta, C., Riva, S., Biamonti, G. (1998). Altered expression of heterogenous nuclear ribonucleoproteins and SR factors in human colon adenocarcinomas. Cancer Res *58*, 5818–5824.

Gordon, P. M., Sontheimer, E. J., Piccirilli, J. A. (2000). Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: extending the parallels between the spliceosome and group II introns. RNA *6*, 199-205.

Gozani, O., Patton, J.G., Reed, R. (1994). A novel set of spliceosome-associated proteins and the essential splicing factor PSF bind stably to pre-mRNA prior to catalytic step II of the splicing reaction. EMBO J *13*, 3356-3367.

Grabowski, P. A., Padgett, R. A., Sharp, P. A. (1984). Messenger RNA splicing in vitro: an excised intervening sequence and a potential intermediate. Cell *37*, 415-427.

Grabowski, P. A., Seiler, S. R., Sharp, P. A. (1985). A multicomponent complex is involved in the splicing of messenger RNA precursors. Cell *42*, 345-353.

Graham, I. R., Hamshere, M., Eperon, I. C. (1992). Alternative splicing of a human alpha-tropomyosin muscle-specific exon: identification of determining sequences. Mol Cell Biol *12*, 3872–3882.

Grainger, R. J., Beggs, J. D. (2005). Prp8 protein: at the heart of the spliceosome. RNA *11*, 533–557.

Graveley, B. R. (2000). Sorting out the complexity of SR protein functions. RNA *6*, 1997-1211.

Graveley, B. R., Hertel, K. J., Maniatis, T. (1998). A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. EMBO J *17*, 6747-6756.

Graveley, B. R., Hertel, K. J., Maniatis, T. (2001). The role of U2AF35 and U2AF65 in enhancer-dependent splicing. RNA *7*, 806-18.

Graveley, B. R., Maniatis, T. (1998). Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. Mol Cell *1*, 765-771.

Gui, J. F., Lane, W. S., Fu, X.-D. (1994b). A serine kinase regulates intracellular localization of splicing factors in the cell cycle. Nature *369*, 678-682.

Gui, J. F., Tronchere, H., Chandler, S. D., Fu, X.-D. (1994a). Purification and characterization of a kinase specific for the serine- and arginine-rich pre-mRNA splicing factors. Proc Natl Acad Sci USA *91*, 10824-10828.

Guthrie, C. (1991). Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. Science *253*, 157-163.

Hanamura, A., Caceres, J. F., Mayeda, A., Franza, B. R., Jr., Krainer, A. R. (1998). Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. RNA *4*, 430-444.

Hardy, S. F., Grabowski, P. J., Padgett, R. A., and Sharp, P. A. (1984). Cofactor requirements of splicing of purified messenger RNA precursors. Nature *308*, 375-377.

Hargous, Y., Hautbergue, G. M., Tintaru, A. M., Skrisovska, L., Golovanov, A. P., Stevenin, J., Lian, L. Y., Wilson, S. A., Allain, F. H. (2006). Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. EMBO J *25*, 5126-37.

Hartmuth, K., Urlaub, H., Vornlocher, H. P., Will, C. L., Gentzel., M., Wilm, M., Luhrmann, R. (2002). Protein composition of human spliceosomes isolated by a tobramycin affinity-selection method. Proc Natl Acad Sci USA *99*, 16719-16724.

He, X., Ee, P.L., Coon, J.S., Beck, W.T. (2004). Alternative splicing of the multidrug resistance protein 1/ATP binding cassette transporter subfamily gene in ovarian cancer creates functional splice variants and is associated with increased expression of the splicing factors PTB and SRp20. Clin Cancer Res *10*, 4652–4660.

Hernandez, N., and Keller, W. (1983). Splicing of in vitro synthesized messenger RNA precursors in HeLa cell extracts. Cell *35*, 89-99.

Hertel, K. J., Graveley, B. R. (2005). RS domains contact the pre-mRNA throughout spliceosome assembly. Trends Biochem Sci *30*, 115-8.

Hertel, K. J., Maniatis, T. (1998). The function of multisite enhancers. Mol Cell *1*, 449-455.

Howard, E. F. (1978). Small nuclear RNA molecules in nuclear ribonucleoprotein complexes from mouse erythroleukemia cells. Biochemistry, *17*, 3228-3236.

James, P. D., O'Brien, L. A., Hegadorn, C. A., Notley, C. R., Sinclair, G. D., Hough, C., Poon, M. C., Lillicrap, D. (2004). A novel type 2A von Willebrand factor mutation located at the last nucleotide of exon 26 (3538G>A) causes skipping of 2 nonadjacent exons. Blood *104*, 2739-2745.

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

Jones, M. H., Frank, D. N., Guthrie, C. (1995). Characterization and functional ordering of Slu7p and Prp17p during the second step of pre-mRNA splicing in yeast. Proc Natl Acad Sci USA *92*, 9687-9691.

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

Kamma, H., Portman, D. S., Dreyfuss, G. (1995). Cell-type specific expression of hnRNP proteins. Exp Cell Res *221*, 187-196.

Kan, J. L. C., Green, M. R. (1999). Pre-mRNA splicing of IgM exons M1 and M2 is directed by a juxtaposed splicing enhancer and inhibitor. Genes Dev *13*, 462-471.

Kanopka, A., Muhlemann, O., Petersen-Mahrt, S., Estmer, C., Ohrmalm. C., Akusjarvi, G. (1998). Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. Nature *393*, 185-187.

Karni, R., de Stanchina, E., Lowe, S. W., Sinha, R., Mu, D., Krainer, A. R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. Nat Struct Mol Bio *14*, 185-193.

Keller, E. B., and Noon, W. A. (1984). Intron splicing: a conserved internal signal in introns of animal pre-mRNAs. Proc Natl Acad Sci USA *81*, 7147-7420.

Kishore, S., Stamm, S. (2006). The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. Science *311*, 230–232.

Konarska, M. M., Sharp, P. A. (1987). Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. Cell *49*, 763-764.

Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Luhrmann, R., Garcia-Blanco, M. A., Manley, J. L. (1994). Protein-protein interactions and 5' splice-site recognition in mammalian mRNA precursors. Nature *368*, 119-124.

Krainer, A. R., Maniatis, T. (1985). Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. Cell *42*, 725-736.

Krainer, A. R., Maniatis, T., Ruskin, B., Green, M. R. (1984). Normal and mutant human β-globin pre-mRNAs are faithfully and efficiently spliced *in vitro*. Cell *36*, 993-1005.

Kramer, A., Keller, W., Appel, B., Luhrmann, R. (1984). The 5′ terminus of the RNA moiety of U1 small nuclear ribonucleoprotein particles is required for the splicing of messenger RNA precursors. Cell *38*, 299-307.

Krawzak, M., Thomas, N. S., Hundrieser, B., Mort, M., Wittig, M., Hampe, J., Cooper D. N. (2007). Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. Hum Mutat *28*, 150-158.

Kumar, A., Baker, S. J., Lee, C. M., Reddy, E. P. (2003). Molecular mechanisms associated with the regulation of apoptosis by the two alternatively spliced products of c-Myb. Mol Cell Biol *23*, 6631–6645.

Ladd, A. N., Cooper, T. A. (2002). Finding signals that regulate alternative splicing in the post-genomic era. Genome Biol *3*, reviews0008.1-0008.16.

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.

Langford, C. J., Klinz, F.-J., Donath, C., Gallwitz, D. (1984). Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal sequence in yeast. Cell *36*, 645-653.

Le, K., Mitsouras, K., Roy, M., Wang, Q., Xu, Q., Nelson, S. F., Lee, C. (2004). Detecting tissue-specific regulation of alternative splicing as a qualitative change in microarray data. Nucleic Acids Res *32*, e180.

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.

Li, Y., Blencowe, B. J. (1999). Distinct factor requirements for exonic splicing enhancer function and binding of U2AF to the polypyrimidine tract. J Biol Chem *274*, 35074-35079.

Licatalosi, D. D., Darnell, R. B. (2006). Splicing regulation in neurologic disease. Neuron *52*, 93-101.

Lopez, A. J. (1998). Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. Ann Rev Genet *32*, 279-305.

Lorson, C. L., Androphy, E. J. (2000). An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene SMN. Hum Molec Genet *9*, 259-265.

Manley, J. L., Tacke R. (1996). SR proteins and splicing control. Genes Dev *10*, 1569-79.

Maris, C., Dominguez, C., and Allain, F. H.-T. (2005). The RNA recognition motif, a plastic RNA binding platform to regulate post-transcriptional gene expression. FEBS Journal *272*, 2118-2131.

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

Mas, C., Taske, N., Deutsch, S., Guipponi, M., Thomas, P., Covanis, A., Friis, M., Kjeldsen, M. J., Pizzolato, G. P., Villemure, J. G., Buresi, C., Rees, M., Malafosse, A., Gardiner, M., Antonarakis, S. E., Meda, P. (2004). Association of the connexin36 gene with juvenile myoclonic epilepsy. J Med Genet *41*, e93.

Mayeda. A., Screaton, G. R., Chandler, S. D., Fu, X.-D., Krainer, A. R. (1999). Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. Mol Cell Biol *19*, 1853-1863.

Merendino, L., Guth, S., Bilbao, D., Martinez, C., Valcarcel, J. (1999). Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. Nature *402*, 838-841.

Michaud, S., Reed R. (1991). An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. Genes Dev *5*, 2534-2546.

Misteli, T., Caceres, J. F., Clement, J. Q., Krainer, A. R., Wilkinson, M. F., Spector, D. L. (1998). Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. J Cell Biol *143*, 297-307.

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

Moore, M. J., Query, C. C., Sharp, P. A. (1993). Splicing of precursors to mRNA by the spliceosome. In The RNA World, Gesteland, R., Atkins, J., eds., (Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press), 303-357.

Moore, M. J., Sharp, P. A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature *365*, 364–368.

Moseley, C. T., Mullis, P. E., Prince, M. A., Phillips, J. A., 3rd. (2002). An exon splice enhancer mutation causes autosomal dominant GH deficiency. J Clin Endocrinol Metab *87*, 847-852.

Muro, A. F., Caputi, M., Pariyarath, R., Pagani, F., Buratti, E., Baralle, F. E. (1999). Regulation of fibronectin EDA exon alternative splicing: possible role of RNA secondary structure for enhancer display. Mol Cell Biol *19*, 2657-2671.

Nagai, K., Outbridge, C., Jessen, T. H., Li, J., Evans, P. R. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. Nature *348*, 515-520.

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

Oberstrauss, F. C., Auweter, S. D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D. L., Allain, F. H. (2005). Structure of PTB bound to RNA: specific binding and implications for splicing regulation. Science *309*, 2054-2057.

Padgett, R. A., Mount, S. M., Steitz, J. A., Sharp, P. A. (1983). Splicing of messenger RNA precursors is inhibited by antisera to small nuclear ribonucleoprotein. Cell *35*, 101-107.

Pan, Q., Shai, O., Misquitta, C., Zhang, W., Saltzman, A. L., Mohammad, N., Babak, T., Siu, H., Hughes, T. R., Morris, Q. D., et al (2004). Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. Mol Cell *16*, 929-941.

Pan, Q., Bakowski, M. A., Morris, Q., Zhang, W., Frey, B. J., Hughes, T. R., Blencowe, B. J. (2005). Alternative splicing of conserved exons is frequently species-specific in human and mouse. Trends Genet *21*, 73-77.

Pikielny, C. W., Rymond, B. C., Roshbash, M. (1986). Electrophoresis of ribonucleoproteins reveals an ordered assembly pathway of yeast splicing complexes. Nature *324*, 341-345.

Pinol-Roma, S., Dreyfuss, G. (1992). Shuttling of pre-mRNA binding proteins between the nucleus and cytoplasm. Nature *355*, 730-732.

Pollard, A. J., Sparey, C., Ronson, S. C., Krainer, A. R., Eurpose-Finner, G. N. (2000). Spatio-temporal expression of the trans-acting splicing factors SF2/ASF and heterogeneous ribonuclear proteins A1/A1B in the myometrium of the pregnant human uterus: a molecular mechanism for regulating regional protein isoform expression in vivo. J Clin Endocrinol Metab *85*, 1928-1936.

Prasad, J., Colwill, K., Pawson, T., Manley, J. L. (1999). The protein kinase Clk/Sty directly modulates SR protein activity: Both hyper- and hypophosphorylation inhibit splicing. Mol Cell Biol *19*, 6991-7000.

Ranum, L. P., Cooper, T. A. (2006). RNA-mediated neuromuscular disorders. Annu Rev Neurosci *29*, 259–277.

Ranum, L. P., Day, J. W. (2004). Pathogenic RNA repeats: an expanding role in genetic disease. Trends Genet *20*, 506–512.

Reed, R. (1996). Initial splice-site recognition and pairing during pre-mRNA splicing. Curr Opin Genet Dev *6*, 215-220.

Reed, R., Maniatis, T. (1985). Intron sequences involved in lariat formation during pre-mRNA splicing. Cell *41*, 95-105.

Reed, R., Palandjian, L. (1997). Spliceosome assembly. In Krainer, A. R., ed., Eukaryotic mRNA processing. Oxford, IRL Press, 103-129.

Resch, A., Xing, Y., Modrek, B. Gorlick, M., Riley, R., Lee, C. (2004). Assessing the impact of alternative splicing on domain interactions in the human proteome. J. Proteome Res *3*, 76-83.

Robberson, B. L., Cote, G. J., Berget, S. M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol Cell Biol *10*, 84-94.

Rogers, J., Wall, R. (1980). A mechanism for RNA splicing. Proc Natl Acad Sci USA *77*, 1877-1879.

Rossi, F., Labourier, E., Forne, T., Divita, G., Derancourt, J., Riou, J. F., Antoine, E., Cathala, G., Brunel, C., Tazi, J. (1996). Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. Nature *381*, 80-82.

Roy, J., Kim, K., Maddock, J. R., Anthony, J. G., Woolford, Jr., J. L. (1995). The final stages of spliceosome maturation require Spp2p that can interact with the DEAH box protein Prp2p and promote step 1 of splicing. RNA *1*, 375-390.

Ruby. S. W., Abelson, J. (1991). Pre-mRNA splicing in yeast. Trends Genet *7*, 79-85.

Ruskin, B., Krainer, A. R., Maniatis, T., Green M. R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing *in vitro*. Cell *38*, 317-331.

Ruskin, B., Zamore, P. D., Green, M. R. (1988). A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. Cell *52*, 207-219.

Sanford, J. R., Bruzik, J. P. (1999). Developmental regulation of SR protein phosphorylation and activity. Genes Dev *13*, 1513-1518.

Schaal, T. D., Maniatis, T. (1999). Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. Mol Cell Biol *19*, 1705-1710.

Schwer, B., Guthrie, C. (1992). A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. EMBO J *17*, 2086-2094.

Shen, H., 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., Kan, J. L., Green, M. R. (2004). Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. Mol Cell 13, 367-76.

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

Sironi, M., Menozzi, G., Riva, L., Cagliani, R., Comi, G. P., Bresolin, N., Giorda, R., Pozzoli, U. (2004). Silencer elements as possible inhibitors of pseudoexon splicing. Nucleic Acids Res *32*, 1783-1791.

Sontheimer, E. J. (2001). The spliceosome shows its metal. Nat Struct Biol *1*, 11-13.

Spellman, R., Rideau, A., Matlin, A., Gooding, C., Robinson, F., McGlincy, N., Grellscheid, S. N., Southby, J., Wollerton, M., Smith, C. W. J. (2005). Regulation of alternative splicing by PTB and its associated factors. Biochem Soc Trans *33*, 457-460.

Staffa, A., Acheson, N. H., Cochrane, A. (1997). Novel exonic elements that modulate splicing of the human fibronectin EDA exon. J Biol Chem *272*, 33394–33401.

Staknis, D., Reed, R., (1994). SR proteins promote the first specific recognition of pre-mRNA and are present together with U1 snRNP in a general splicing enhancer complex. Mol Cell Biol *14*, 7670-7682.

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

Steinman, H.A., Burstein, E. A., Lengner, C., Gosselin, J., Pihan, G., Duckett, C. S., Jones, S. N. (2004). An alternative splice form of Mdm2 induces p53-independent cell growth and tumorigenesis. J Biol Chem *279*, 4877–4886.

Stetefeld, J., Ruegg, M. A. (2005). Structural and functional diversity generated by alternative splicing. Trends Biochem Sci *30*, 515-521.

Stickeler, E. Kittrell, F., Medina, D., Berget, S.M. (1999). Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis. Oncogene *18*, 3574-3582.

Sun, H., and Chasin, L. A. (2000). Multiple splicing defects in an intronic false exon. Mol Cell Biol *20*, 6414-6425.

Tacke, R., Manley, J. L. (1995). The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. EMBO J *14*, 3540-3551.

Tennenbaum, T., Belanger, A. J., Glick, A. B., Tamura, R., Quaranta, V. (1995). A splice variant of alpha 6 integrin is associated with malignant conversion in mouse skin tumorigenesis. Proc Natl Acad Sci USA *92*, 7041–7045.

Teigelkamp, S., McGarvey, M., Plumpton. M., Beggs, J. D. (1994). The splicing factor PRP2, a putative RNA helicase, interacts directly with pre-mRNA. EMBO J *13*, 888-897.

Tian, M., Maniatis, T. (1994). A splicing enhancer exhibits both constitutive and regulated activities. Genes Dev *8*, 1703-1712.

Tockman, M. S., Mulshine, J.L., Piantadosi, S., Erozan, Y.S., Gupta, P.K., Ruckdeschel, J.C., Taylor, P.R., Zhukov, T., Zhou, W.H., Qiao, Y.L., Yao, S.X. (1997). Prospective detection of preclinical lung cancer: results from two studies of heterogeneous nuclear ribonucleoprotein A2/B1 overexpression. Clin Cancer Res *3*, 2237–2246.

Ule, J., Ule, A., Spencer, J., Williams, A., Hu, J. S., Cline, M., Wang, H., Clark, T., Fraser, C., Ruggiu, M., et al (2005). Nova regulates brain-specific splicing to shape the synapse. Nat Genet *37*, 844-852.

Umen, J. G., Guthrie, C. (1995). A novel role for a U5 snRNP protein in 3' splice site selection. Genes Dev *9*, 855-868.

Valadkhan, S., Manley, J. L. (2001). Splicing-related catalysis by protein-free snRNAs. Nature *413*, 701–707.

Valadkhan, S., Manley, J. L. (2003). Characterization of the catalytic activity of U2 and U6 snRNAs. RNA *9*, 892–904.

Valcarcel, J., Gaur, R. K., Singh, R., Green, M. R. (1996). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. Science *273*, 1706-9.

van Der Houven van Oordt, W., Diaz-Meco, M. T., Lozano, J., Krainer, A. R., Moscat, J., Caceres, J. F. (2000). The MKK(3/6)-p38-signaling cascade alters the subcellular distribution of hnRNPA1 and modulates alternative splicing regulation. J Cell Biol *143*, 297-307.

Venables, J. P. (2004). Aberrant and alternative splicing in cancer. Cancer Res *64*, 7647-7654.

Wang, J., Smith, P. S., Krainer, A. R., Zhang, M. R. (2005). Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes. Nucleic Acids Res *33*, 5053-5062.

Wang, Z., Hoffman, H. M., Grabowski, P. J. (1995). Intrinsic U2AF binding is modulated by exon enhancer signals in parallel with changes in splicing activity. RNA *1*, 21-35.

Wang, Z., Lo, H.S., Yang, H., Gere, S., Hu, Y., Buetow, K.H., Lee, M.P.,Wang, Z. (2003). Computational analysis and experimental validation of tumor-associated alternative RNA splicing in human cancer. Cancer Res *63*, 655–657.

Wang, Z., Rolish, M. E., Yeo, G., Tung, V., Mawson, M., Burge, C. B. (2004). Systematic identification and analysis of exonic splicing silencers. Cell *119*, 831-845.

Wang, J., Xiao, S. H., Manley, J. L. (1998). Genetic analysis of the SR protein ASF/SF2: interchangeability of RS domains and negative control of splicing. Genes Dev *12*, 2222-2233.

Watakabe, A., Tanaka, K., Shimura, Y. (1993). The role of exon sequences in splice site selection. Genes Dev *7*, 407-418.

Weist, D. K., O'Day, C. L., Abelson, J. (1996). In vitro studies of the Prp9.Prp11.Prp21 complex indicate a pathway for U2 small nuclear ribonucleoprotein activation. J Biol Chem *271*, 33268-33276.

Will, C. L., Luhrmann, R. (1997). Protein functions in pre-mRNA splicing. Curr Opin Cell Biol *9*, 320-328.

Wu, J. Y., Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell *75*, 1061-1070.

Wu, S., Romfo, C. M, Nilsen, T. W., Green, M. R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. Nature *402*, 832-835.

Xiao, S.-H., Manley, J. L. (1997). Phosphorylation of the ASF/AF2 RS domain affects both protein-protein interactions and is necessary for splicing. Genes Dev *11*, 334-344.

Xiao, S.-H., Manley, J. L. (1998). Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. EMBO J *17*, 6359-67.

Xie, J., Lee, J. A., Kress, T.L., Mowry, K. L., Black, D.L. (2003). Protein kinase A phosphorylation modulates transport of the polypyrimidine tract-binding protein. Proc Natl Acad Sci USA *100*, 8776-81.

Xu, Q., Lee, C. (2003). Discovery of novel splice forms and functional analysis of cancer-specific alternative splicing in human expressed sequences. Nucleic Acids Res *31*, 5635–5643.

Yeakley, J.M., Fan, J.-B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M. S., Fu, X.-D. (2002). Profiling alternative splicing on fiber-optic arrays. Nat Biotechnol *20*, 353-358.

Yeo, G., Holste, D., Kreiman, G., Burge, C. B. (2004). Variation in alternative splicing across human tissues. Genome Biol *5*, R74.

Zahler, A. M., Damgaard, C. K., Kjems, J., Caputi, M. (2004). SC35 and heterogeneous ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/silencer element to regulate HIV-1 tat exon 2 splicing. J Biol Chem *279*, 10077-10084.

Zahler, A. M., Lane, W. S., Stolk, J. A., Roth, M. B. (1992). SR proteins: a conserved family of pre-mRNA splicing factors. Genes Dev *6*, 837-847.

Zamore, P. D., Green, M. R. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. Proc Natl Acad Sci U S A *86*, 9243-7.

Zhang, X. H., Chasin, L. A. (2004). Computational definition of sequence motifs governing constitutive exon splicing. Genes Dev *18*, 1241-1250.

Zheng, Z.-M. (2004). Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression. J Biomed Sci *11*, 278-294.

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

Zhou, Z., Reed, R. (1998). Human homologs of yeast prp16 and prp17 reveal conservation of the mechanism for catalytic step II of pre-mRNA splicing. EMBO J *17*, 2095-2106.

Zhu, J., Mayeda, A., Krainer, A. R. (2001). Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNPA1 and enhancer-bound SR proteins. Mol Cell *8*, 1351-1361.

Zieve, G., Penman, S. (1981). Subnuclear particles containing a small nuclear RNA and a heterogeneous nuclear RNA. J. Mol. Biol *145*, 501-523.

Zorio, D. A., Blumenthal, T. (1999). Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*. Nature *402*, 835-838.

Zuo, P., Maniatis, T. (1996). The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. Genes Dev *10*, 1356-1368.

1.6 Figures

Figure 1. Types of alternative splicing. Many different types of alternative splicing contribute to mRNA and thus protein diversity. Pre-mRNAs consist of exons (depicted as boxes) interrupted by introns (depicted as lines). Cassette alternative exons are either included or skipped in an mRNA. Alternative 5' and 3' splice sites permit the introduction of changes as small as a single codon, or less, into mRNAs. A composite exon results from intron retention. Mutually exclusive exons allow a choice between two coding sequences. Alternative transcriptional promoters are linked to their own first exons, or alternative first exons. Likewise, alternative polyadenylation sites can be coded through alternative terminal exons.

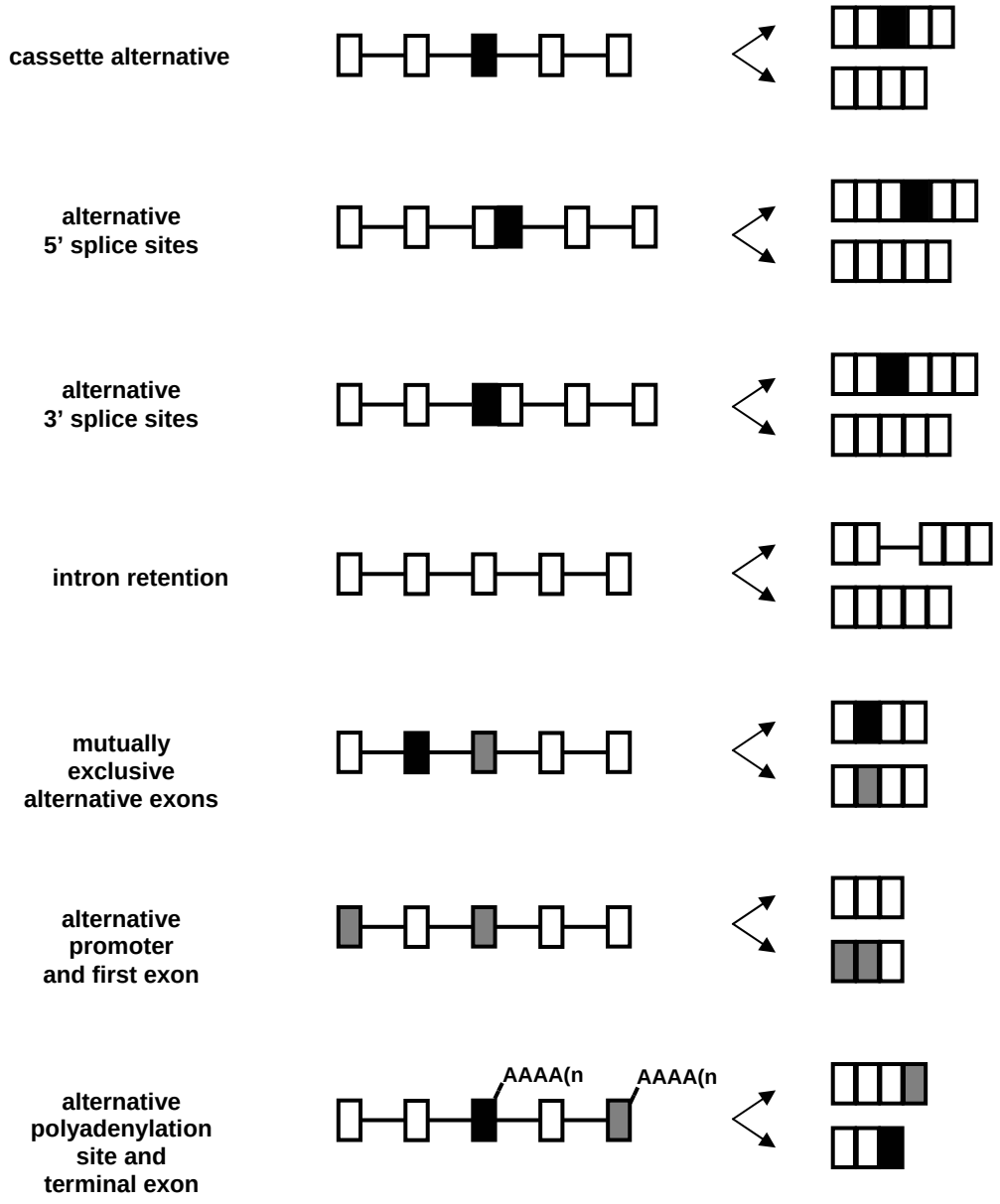


Figure 2. Essential splicing signals and their consensus sequences. Pre-mRNA sequence elements recognized by the splicing machinery include the 5' splice site, the branchpoint, the polypyrimidine tract, and the 3' splice site. Introns recognized by the major spliceosome begin with 5'-GU and end with 3'-AG. Sequences overlapping the exon-intron boundaries, the 5' and 3' splice sites, exhibit conservation. The vertebrate 5' splice site consensus sequence is ag / **GU**RAGU (where R = G or A). The vertebrate 3' splice site consensus sequence is Y>11 NY**AG** / g (where N = any nucleotide, and Y = C or T), and incorporates the polypyrimidine tract, a C/T rich stretch between the branchpoint and the 3' splice site proper. The branchpoint is typically located 18-40 nucleotides upstream from the 3' splice site, and exhibits a consensus sequence of YNYUR**A**C or YNYUR**A**Y in vertebrates. The adenosine branchpoint residue indicated in bold lettering is covalently linked to the first nucleotide of the intron during the first transesterification reaction.

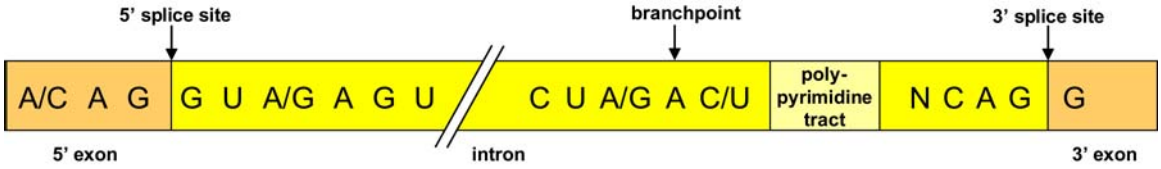
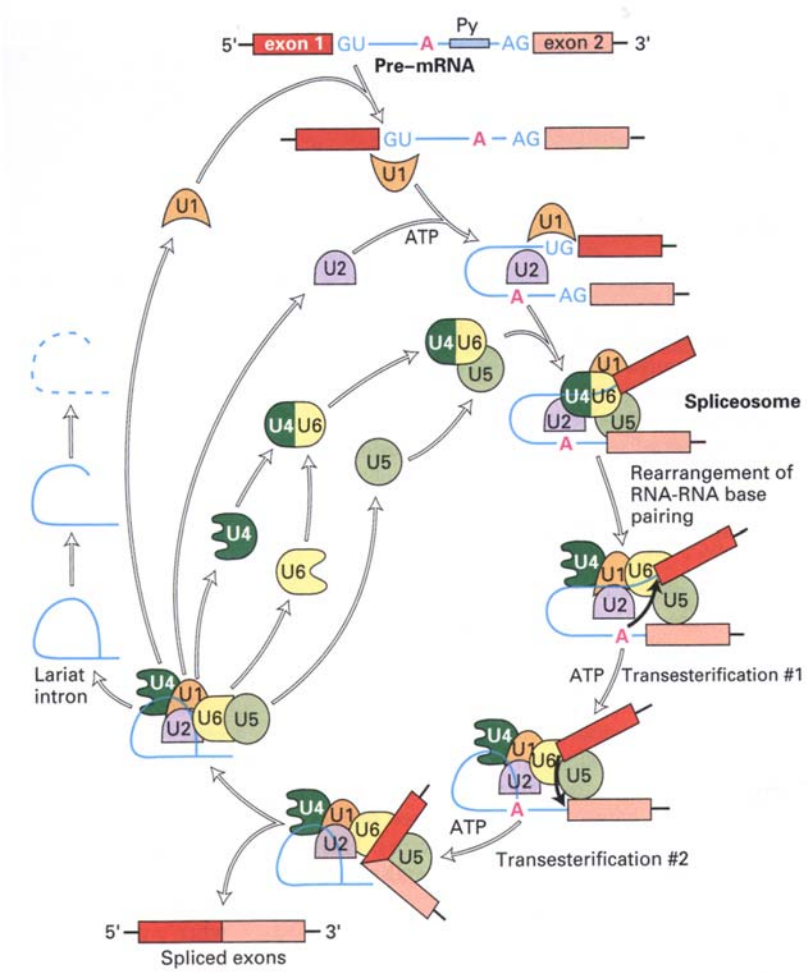


Figure 3. Spliceosomal complex assembly. In the first step of splicing, U1 snRNP is recruited to the pre-mRNA and U1 snRNA base pairs with the 5'ss, leading to the formation of the E complex. The recruitment of U2 snRNP and the base pairing of U2 snRNA with the branchpoint results in the formation of the A complex. The U4/U6 di-snRNP and U5 snRNP, together in a complex called the tri-snRNP, join with U1 and U2 to produce the C complex, also called the spliceosome. Next, U4 snRNA is displaced so that U6 snRNA can bind to U2 snRNA, and U6 snRNA displaces U1 snRNA and binds to the 5'ss to create the catalytic spliceosome. In the first transesterification reaction, the 5' exon is cleaved from the intron and a lariat intermediate is produced. In the second transesterification reaction the exons are joined together to produce the mRNA and release the lariat intron. After each round of splicing, the snRNPs are recycled and the lariat is degraded.



Lodish, H. F., Berk, A., Zipursky, L. S., Matsudaira, P., Baltimore, D., Darnell, J. (2000). Molecular Cell Biology, 4th Edition, New York, W.H. Freeman and Company, section 11.2. Reproduced with permission from W. H. Freeman and Company/Worth Publishers.

Figure 4. Splicing is accomplished by two sequential transesterification reactions. On the pre-mRNA depicted at the top of the diagram, the intron is indicated by a line, and exons are indicated by boxes; the upstream exon is labeled L1 and the downstream exon is labeled L2. Positions of the 5' splice site (5'ss), branch site (BP), and 3' splice site (3'ss) on the pre-mRNA are indicated. In the first transesterification reaction, the pre-mRNA is cleaved at the 5' splice site and the 5' end of the intron is joined by a 2'-5' phosphodiester bond to the adenosine at the branch site to produce the lariat intermediate composed of the intron and the L2 exon. The L1 exon 5' splice site with its free 3' hydroxyl group is held in close proximity by the spliceosome for the subsequent reaction. In the second transesterification reaction, using the phosphate at the 3' splice site to form a 3'-5' phosphodiester bond, the upstream exon L1 is joined to the downstream exon L2, releasing the lariat product, which ends in a 3' hydroxyl group.

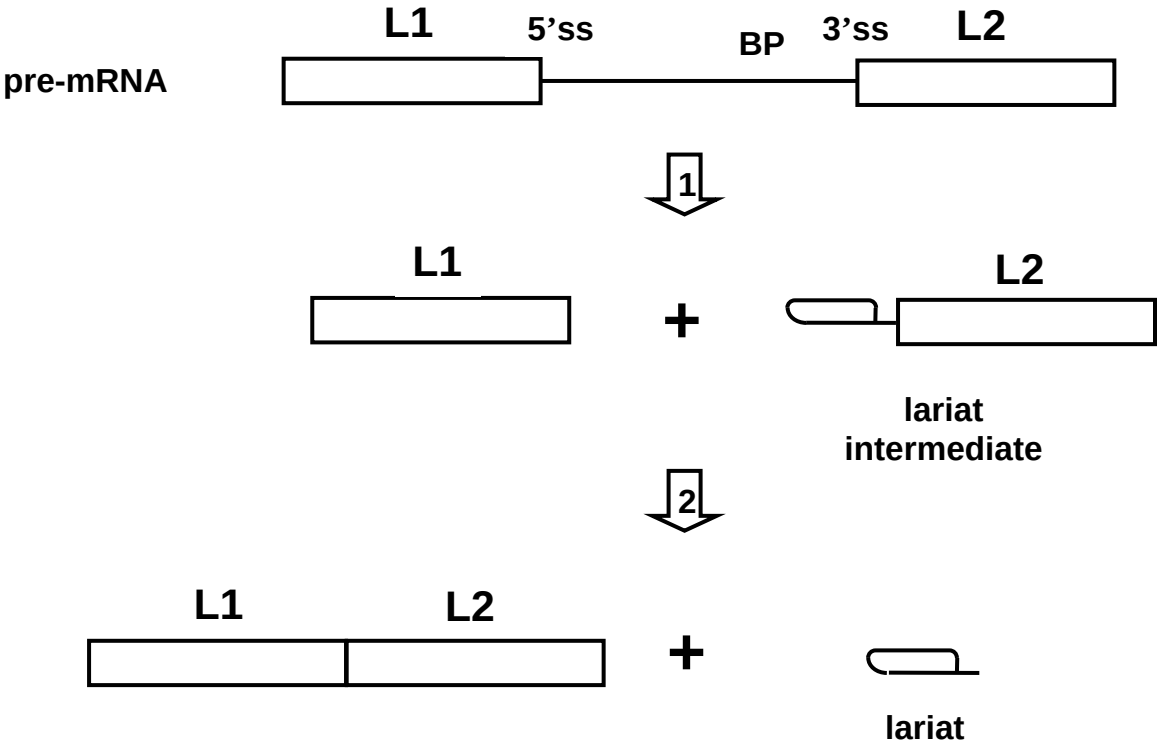


Figure 5. Human SR protein family members and their protein domain organization. SR proteins have either one or two RRM domains and an RS domain of variable length. RRM: RNA recognition motif, RRMH: RNA recognition motif homology (also known as ψRRM: pseudo RNA recognition motif), RS: arginine/serine-rich domain, Zn: zinc-knuckle domain.

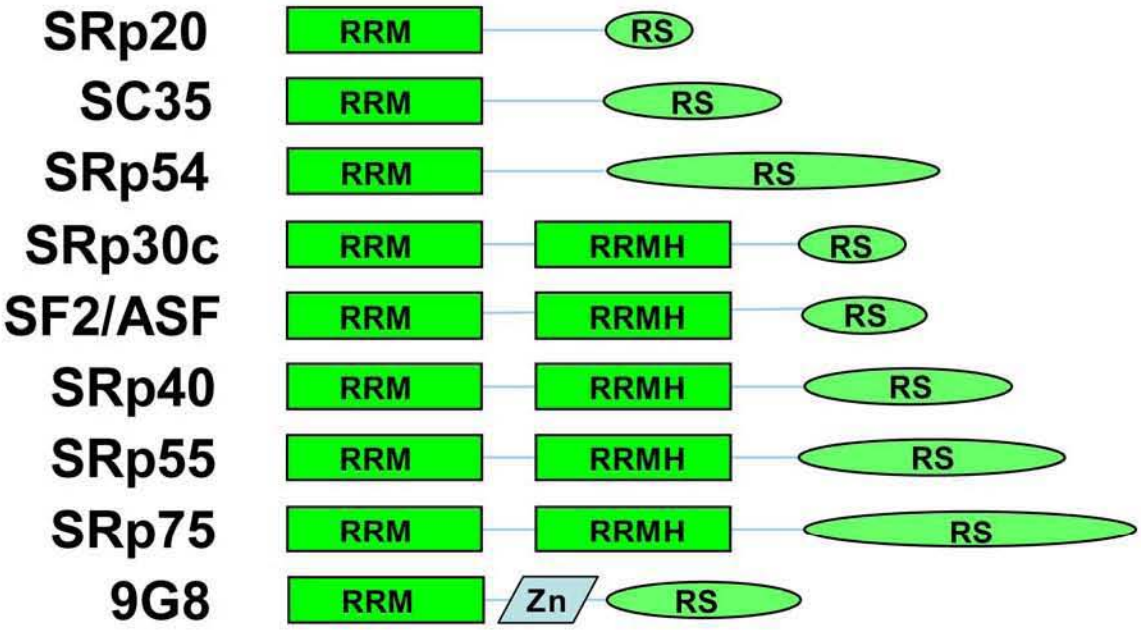


Figure 6. SR proteins promote splicing by recruiting positive acting factors and inhibiting negative acting factors. ESE-bound SR proteins engage in protein-protein interactions with U1-70K, U2AF35, and the splicing coactivator SRm160 to recruit spliceosomal components U1 snRNP and U2 snRNP to the 5'ss and the branchpoint, respectively. ESE-bound SR proteins also counteract the effects of inhibitory proteins bound to an ESS.

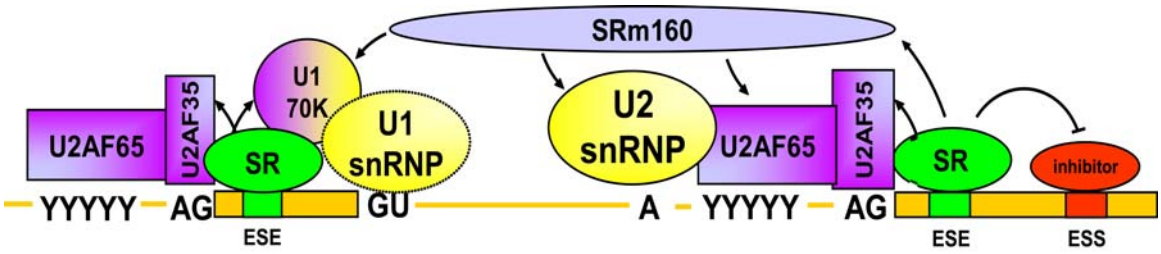


Figure 7. Structure of the RRM domain of an SR protein. Shown is the solution structure of residues 9-83 of the SRp20 RRM domain in complex with the RNA 5'CAUC3', in a ribbon and stick representation. The protein backbone is shown in cyan. RNA heavy atoms are shown in yellow (carbon), red (oxygen), and blue (nitrogen and phosphorous). Important side chains involved in RNA interactions are shown in green.

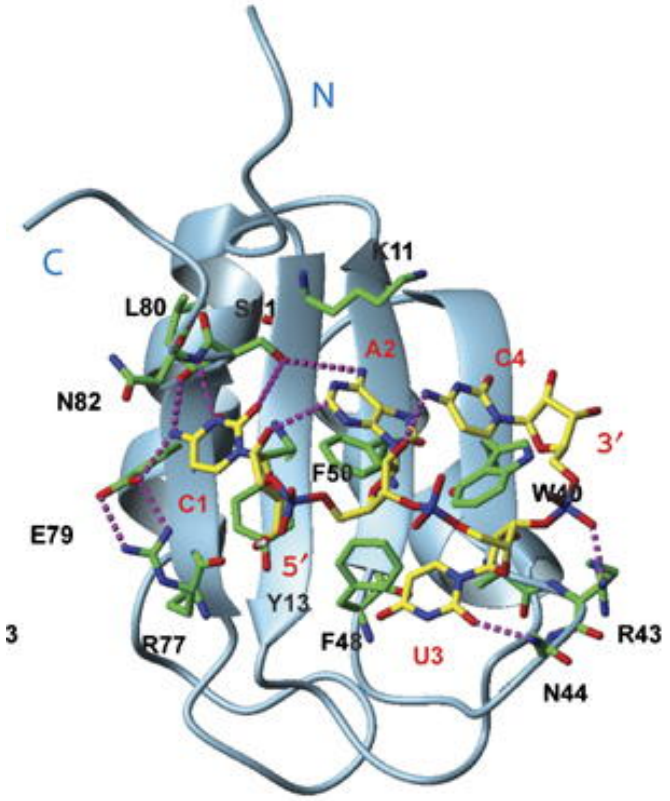


Figure 3C from: Hargous, Y., Hautbergue, G. M., Tintaru, A. M., Skrisovska, L., Golovanov, A. P., Stevenin, J., Lian, L. Y., Wilson, S. A., Allain, F. H. (2006). Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. EMBO J *25*, 5126-37. Reproduced with permission from Nature Publishing Group.

Chapter 2

Pre-mRNA sequence elements and SF2/ASF RS domain-dependence

2.1 Abstract

SR proteins are essential pre-mRNA splicing factors that bind to pre-mRNA via their RRM domains and are thought to recruit components of the splicing machinery through protein-protein interactions mediated by their RS domains. Our lab previously discovered that constitutive splicing can occur for some but not all ESE-dependent pre-mRNAs *in vitro*, with a mutant SR protein lacking its RS domain ("ΔRS") as the sole source of SR protein. Therefore, pre-mRNAs can be classified as either RS domain-dependent or RS domain-independent based on their ability to be spliced *in vitro* with an SR protein lacking its RS domain. We sought to identify sequence-specific requirements for the RS domain within different pre-mRNA contexts as a means to understand the mechanisms by which the domains of SR proteins promote splicing. We have demonstrated that improvement of the pyrimidine tract, an essential intronic splicing signal upstream of the 3' splice site, abrogates the RS domain requirement. We have also identified sequence elements spanning the 5' splice site that confer RS domain-dependence. Collectively these findings suggest that RS domain-dependence is caused by a defect in intron definition, the assembly of spliceosomal complexes across the intron to establish the locations of the 5' and 3' splice sites prior to splicing catalysis.

2.2 Introduction

Pre-mRNA splicing is a critical yet highly complex step in gene expression which is accomplished by the extremely large ribonucleoprotein machine called the spliceosome (Jurica and Moore, 2003). To properly specify the locations of introns to be removed prior to splicing catalysis, the mammalian spliceosome must somehow correctly identify a set of highly degenerate splice site signals consisting of the 5' splice site (5' ss), branchpoint sequence (BPS), pyrimidine tract (Py), and 3' splice site (3' ss) among a myriad of "decoy" splicing signals within what are typically very long pre-mRNAs (Burge et al, 1999, Sun and Chasin, 2000).

SR proteins are essential metazoan splicing factors that assist the spliceosome in identifying the locations of exons and introns, and are understood to promote constitutive and alternative splicing through multiple modes (Graveley, 2000). Exonic splicing enhancers (ESEs) are degenerate 6-8 nucleotide motifs within the exons of pre-mRNAs that promote exon inclusion through the action of SR proteins (Liu et al, 1998, Liu et al, 2000, Blencowe, 2000, Cartegni et al, 2002). SR protein family members possess N-terminal RNA recognition motifs (RRM) and C-terminal arginine and serine-rich (RS) domains (Zahler et al, 1992, Birney et al, 1993). SR proteins bind to ESEs via their RRMs (Hargous et al, 2006), and their RS domains are thought to function as protein-protein interaction modules to facilitate exon inclusion by recruiting components of the basal splicing machinery to the flanking splice sites (Reed, 1996). Exonic splicing silencers (ESSs) are pre-mRNA regulatory motifs that inhibit exon inclusion in both constitutive and alternative splicing (Pozzoli and Sironi, 2005). Most of the well-characterized ESS elements exert their silencing functions through the binding of various

repressor proteins, which are typically members of the hnRNP protein family (see section 1.3.2). SR proteins can antagonize the negative regulation conferred by ESSs, but less is known about how this is achieved mechanistically (Cartegni et al, 2002).

Roles for SR proteins have been identified at virtually every step of the spliceosomal complex assembly pathway. SR proteins are thought to function early in splice site recognition during the formation of the E, or commitment complex, through recruitment of splicing factors to the 5' and 3' splice sites, or by antagonizing the function of splicing silencer sequences and their cognate repressors, or by both of these mechanisms (Graveley, 2000, Figure 6). For the assembly of E complex, U1 snRNP must be recruited to the 5'ss. SR proteins are thought to recruit the U1 snRNP through interactions with the U1 snRNP-specific protein U1-70K (Reed, 1996). Also required for the assembly of E complex is the binding of U2AF65 to the pyrimidine tract. SR proteins escort U2AF65 to the pyrimidine tract through recruitment of the small subunit of U2AF, U2AF35 (Zuo and Maniatis, 1996). The mechanisms by which SR proteins overcome the effects of splicing silencers are not well understood.

Within the A complex, SR proteins are thought to promote splicing in part by interacting with the coactivator SRm160/300, which in turn interacts with both the U1 snRNP and the U2 snRNP (Eldridge et al, 1999). SR proteins reportedly aid in recruitment of the tri-snRNP for the formation of B complex (Roscigno and Garcia-Blanco, 1995, Tarn and Steitz, 1995). Our lab has observed that SR proteins promote the second catalytic step of the splicing reaction in an ESE-dependent role (Chew et al, 1999), but the protein-protein and protein-RNA interactions important for this function remain to be discovered.

Although many of the aforementioned functions of SR proteins are thought to occur via SR protein RS domain-mediated protein-protein interactions, it has not yet been demonstrated that such interactions occur in the context of a functional spliceosome (Hertel and Graveley, 2005). The RS domain of SR proteins was at one time thought to be indispensable for constitutive splicing *in vitro,* yet dispensable for concentration-dependent effects on alternative splice site selection (Caceres and Krainer, 1993, Zuo and Manley, 1993). However, our lab subsequently reported that the RS domain of SF2/ASF is dispensable for constitutive splicing of several pre-mRNAs *in vitro*, including tat23, an ESE-dependent pre-mRNA known to be regulated by an ESS (Zhu and Krainer, 2000). An SR protein lacking its RS domain was also found to promote splicing of a related HIV tat pre-mRNA (Tange and Kjems, 2001). Thus pre-mRNAs can be classified as either RS domain-dependent or RS domain-independent based on their ability to be spliced with an SR protein lacking its RS domain.

Splicing in the absence of the RS domain does not seem to agree with the recruitment model for SR protein function, as the RS domain should be required for protein-protein interactions to bring in the basal splicing machinery, including U1-70K and U2AF65/35 (Hertel and Graveley, 2005). U1-70K and U2AF35 interactions with the SR protein RS domain were observed by yeast two-hybrid assays and Far-Western experiments (Wu and Maniatis 1993, Kohtz et al, 1994), and formed the basis of the original recruitment model. Our lab's finding that the tat23 ESE-dependent pre-mRNA containing an ESS can undergo splicing when provided with an SR protein lacking the RS domain predicts two distinct mechanistic possibilities for the splicing of RS domain-independent pre-mRNAs. First, if the recruitment function of the SR protein is required

in these contexts, then SR proteins may not always need their RS domains to recruit the splicing machinery. The possibility that the RRM domains of SR proteins can act as protein-protein interaction modules to promote splicing has not been explored, but RRMs in other essential splicing proteins function in protein-protein interactions (Price et al, 1998, Kielkopf et al, 2001, Selenko et al, 2003, Shi and Xu, 2003). Second, if the RS domain is absolutely required for SR protein recruitment functions, then for those pre-mRNAs in which splicing can occur without the RS domain, the ESE's predominant function should be to antagonize the ESS.

Based on the observations that the RS domain can be dispensable for splicing *in vitro* of some substrates but not others, we attempted to investigate how the requirements for the domains of SR proteins vary within different pre-mRNA contexts as a means to refine the paradigms about the functions of the RS domains of SR proteins in pre-mRNA splicing. To identify pre-mRNA sequence elements that confer the RS domain requirement for splicing of RS domain-dependent pre-mRNAs, we tested in the *in vitro* splicing assay with SF2/ASF and ΔRS (a mutant SF2/ASF lacking its RS domain) a large number of pre-mRNAs in which specific sequences were mutated or replaced by sequences from RS domain-independent pre-mRNAs. We have demonstrated that improvement of the pyrimidine tract, an essential intronic splicing signal upstream of the 3' splice site, abrogates the RS domain requirement. We have also identified a sequence element spanning the 5' splice site that confers RS domain-dependence.

2.3 Results

To identify pre-mRNA sequences that confer an RS domain requirement, we have generated and tested derivative IgM M1-M2 (RS domain-dependent) and tat23 (RS domain-independent) pre-mRNAs in the *in vitro* splicing assay with the SF2/ASF and ΔRS proteins. We employed two different strategies in the design of these pre-mRNAs: creation of chimeric pre-mRNAs in which tat23 and IgM M1-M2 sequences are swapped (section 2.3.1), and mutation of specific IgM M1-M2 sequence elements based on prevailing models for SR protein function (sections 2.3.2 and 2.3.3).

2.3.1 Chimeric substrates reveal IgM M1-M2 sequences that confer RS domain-dependence

By means of *in vitro* splicing assays employing ΔRS, our lab discovered that splicing can occur for some but not all ESE-dependent pre-mRNAs (Zhu and Krainer, 2000). Two of the model pre-mRNAs tested with the ΔRS protein were found to have opposite RS domain requirements: HIV tat23 spliced nearly as well with the ΔRS SR protein mutant as with its full-length counterpart SF2/ASF, whereas IgM M1-M2 could not be spliced with the ΔRS protein (Zhu and Krainer, 2000).

Our first approach to try to identify sequences that confer SR protein RS domain dependence exploited the opposite RS domain requirements of these well-studied pre-mRNAs, ie IgM M1-M2 and HIV tat23. Chimeric pre-mRNAs generated by swapping sequences between the RS domain-dependent IgM M1-M2 and the RS domain-independent HIV tat23 pre-mRNAs were tested by *in vitro* splicing with SF2/ASF and ΔRS to try to identify the minimal context changes that abolish the RS domain

requirement(s) for the IgM M1-M2 pre-mRNA, and conversely to generate an RS domain-independent IgM M1-M2 by replacing sequences in the IgM M1-M2 pre-mRNA with sequences from the HIV-tat23 pre-mRNA. This chimeric substrate approach had previously been employed successfully in our lab to identify pre-mRNA sequence elements that confer specificity for recognition by different SR proteins (Mayeda et al, 1999).

Chimeric substrates were generated according to the following strategy, with the nomenclature system borrowed from Mayeda et al, 1999. Both the IgM M1-M2 and tat23 constitutive splicing substrates have a two-exon and one-intron structure; the upstream exons are called M1 and T2 and the downstream exons are called M2 and T3 for IgM M1-M2 and tat23, respectively (Figure 8). By conceptually dividing these two pre-mRNAs into four portions—the 5' exon (M1 or T2), the 5' half of the intron (Mi or Ti), the 3' half of the intron (Mi or Ti), and the 3' exon (M2 or T3)—14 possible chimeric pre-mRNA substrates could be generated by swaps between the tat23 (ie, T2TiTiT3) and IgM M1-M2 (ie, M1MiMiM2) pre-mRNAs (Figure 8), for a total of 16 substrates including the parental IgM M1-M2 and tat23 pre-mRNAs. The swap points were designed not to disrupt the essential splicing signals of the 5' splice site, the branchpoint, the pyrimidine tract, and the 3' splice site; the upstream intronic halves Mi and Ti included three nucleotides of the preceding exons M1 and T2, and the downstream intronic halves Mi and Ti included two nucleotides of the following exons M2 and T3.

To study the functions of the RS domain, the *in vitro* splicing assay was performed using HeLa cell S100 extract, which is deficient in SR proteins but provides the other necessary spliceosomal components (Mayeda and Krainer, 1999a), to which

either the full-length SR protein SF2/ASF or the mutant ΔRS protein was added. Eleven of the 14 possible chimeric pre-mRNAs were generated as described above and shown in Figure 8 (see Table 1), and a titration was done for each pre-mRNA to determine the concentration of magnesium chloride at which the most splicing was achieved in S100 complementation with SF2/ASF. Magnesium chloride optima are listed in Table 1. The *in vitro* splicing assay was then performed for each of the substrates at its optimum magnesium chloride concentration in S100 complementation with SF2/ASF and ΔRS. A representative *in vitro* splicing assay for some of the chimeric substrates is shown in Figure 9.

Although the chimeric pre-mRNAs vary in their RS-domain dependence, there seems to be no simple correlation between any one single portion of the IgM M1-M2 pre-mRNA included in the chimeric splicing substrates and RS domain-dependence (see Table 1). However, in the context of the chimeric substrates, the combination of two sequences—the IgM M1-M2 upstream exon M1 and the upstream half of the IgM M1-M2 intron—was associated with RS domain-dependence (see Table 1, substrates IgM M1-M2, M1MiMiT3, M1MiTiT3, and see also Figure 11, subchimeric substrate M1TiMiM2:A), and disruption of this combination is associated with conferring RS domain-independence (Table 1, substrates M1TiMiM2, M1TiTiM2, T2MiMiM2, and Figure 11, subchimeric substrates M1TiMiM2:B, M1TiMiM2:C, and M1TiMiM2:D). With the exception of the original context of the parental substrate IgM M1-M2, the M1Mi combination is also associated with reduced splicing efficiency (M1MiMiT3, M1MiTiT3, M1MiTiM2, and Figure 11, subchimeric substrate M1TiMiM2:A).

Some of the chimeric substrates show striking changes in RS domain-dependence relative to their parental IgM M1-M2 and tat23 pre-mRNAs. For example, in the M1TiMiM2 chimeric substrate, replacement of the upstream half of the IgM M1-M2 intron with the corresponding tat23 intronic sequences renders the RS domain-dependent IgM M1-M2 RS domain-independent, suggesting that either the Ti sequence activates splicing in the absence of the RS domain, or the upstream Mi sequence represses splicing, or is insufficient to permit splicing, in the absence of the RS domain. To try to identify sequences within the upstream half of the intron of the IgM M1-M2 pre-mRNA that confer RS domain-dependence, we attempted to restore the RS domain-dependence to M1TiMiM2 by adding back smaller portions of the IgM M1-M2 upstream intronic sequence. M1TiMiM2 "subchimeric" pre-mRNAs were generated in which four consecutive segments of the Mi sequences were added back in place of the corresponding Ti sequences, as diagrammed in Figure 10.

Splicing of the M1TiMiM2 subchimeric pre-mRNAs is shown in Figure 11. Add-back of the Ma segment spanning the 5' splice site is sufficient to completely restore RS domain-dependence to M1TiMiM2. The data also suggest that each of the three added back Mi segments Mb, Mc, and Md may contribute combinatorially towards an RS domain requirement, with the Mc segment making the smallest contribution. SR proteins can promote splicing by aiding in the recruitment of splicing factors but also by interfering with the binding to pre-mRNA of factors that prevent splicing. It is possible that replacement of the Mi sequences in M1MiMiM2 with Ti sequences to produce M1TiMiM2 may permit splicing to occur in the absence of the RS domain by relieving a requirement of the RS domain for preventing the binding of inhibitory factors to intronic

splicing silencers (ISSs) within some portion or portions of the Mi sequences. To test the hypothesis that there may be inhibitory factor(s) binding to ISSs within these IgM M1-M2 intronic sequences and conferring an RS domain requirement, we attempted to titrate any such factors away from the IgM M1-M2 intronic portions of the M1TiMiM2 subchimerics; splicing reactions were assembled without pre-mRNA and in the presence of an excess of RNA competitors corresponding to each of the Ma, Mb, Mc, and Md segments and preincubated under splicing conditions prior to adding the M1TiMiM2 subchimeric substrates containing the sequence of each RNA competitor. Preincubation of the splicing reactions without substrate rendered RS domain-dependent substrates RS domain-independent, including the control IgM M1-M2 substrate and the M1TiMiM2:A substrate, and unexpectedly even in the absence of RNA competitors. Unfortunately we were unable to follow up on this finding due to difficulties with the reproducibility of ΔRS activity in the *in vitro* splicing assay (see section 2.4.2).

2.3.2 Mutation of IgM M1-M2 PTB site I does not significantly improve splicing

In our second approach to identify IgM M1-M2 sequences that confer an RS domain-requirement, we created derivative IgM M1-M2 substrates in which specific sequence elements already demonstrated to have a relationship to SR protein function were mutated or deleted. Based on prevailing models for SR protein function in the IgM M1-M2 pre-mRNA context, three specific IgM M1-M2 sequence elements were selected: the ESE, the pyrimidine tract, and the ESS.

The IgM M1-M2 exonic splicing silencer was originally identified functionally by progressive deletion of the M2 exon from the 3' end, and therefore was mapped only to

the last half of the M2 exon (Watakabe et al, 1993); deletion of this 70 nt SpeI to XbaI fragment resulted in an increase in the amount of splicing for IgM M1-M2 in the *in vitro* splicing assay with nuclear extract (NE) (Watakabe et al, 1993). We had originally attempted to investigate whether there is a role of the ESS in the requirement for the RS domain for splicing of IgM M1-M2 using this Watakabe ESS deletion pre-mRNA, in which the ESS consisting of the last 70 nt of exon M2 was deleted. However, we were unable to adequately test the contribution of the 70 nt ESS to the RS domain-dependence of IgM M1-M2 due to degradation of this pre-mRNA in our S100 extracts.

The function of this 70 nt IgM M1-M2 ESS was subsequently confirmed in *in vitro* splicing assays done with NE, and the ESS was more precisely mapped to a single 11 nt motif called PTB site I and its cognate repressor protein identified as pyrimidine tract binding protein (PTB) (Shen et al, 2004). Using an IgM M1-M2 derivative substrate in which the ESE has been replaced by a consensus binding site for the MS2 bacteriophage coat protein (Schneider et al, 1992) for targeting of an SR protein RS domain to the ESE position via an MS2-RS fusion protein, Shen et al showed in S100 complementation assays with SF2/ASF that immunodepletion of the PTB protein from the S100 permitted splicing of this derivative substrate in the absence of an MS2-RS protein targeted to the ESE position (Shen et al, 2004), suggesting that the primary function of the ESE is to counteract the ESS and that an ESE-bound SR protein is not needed for recruitment functions.

We wanted to test whether there is a role for the ESS in the RS domain-dependence of IgM M1-M2, and first attempted to verify the function of this 11 nt PTB site I ESS in *in vitro* splicing assays with NE and S100 complementation with SF2/ASF.

IgM M1-M2 derivative substrates were generated in which the PTB site I ESS was mutated to no longer bind PTB and function as an ESS, as shown by Shen et al. The IgMPTB substrate, in which the PTB site I was mutated from UCUUACGUCUU to ACAUACGACAU, was tested in the *in vitro* splicing assay with NE or in S100 complementation with SF2/ASF (Figure 12). Contrary to our expectations, we did not see an increase in splicing for the IgMPTB substrate relative to the IgM M1-M2 substrate, either in NE or S100 complementation with SF2/ASF. Moreover, mutation of the PTB site I ESS did not relieve the SR protein requirement at the ESE position for splicing of IgM M1-M2 (Figure 14, substrate IgMΔKGESEPTB, an IgM M1-M2 derivative substrate with the same PTB site I mutation and the deletion of the ESE as defined by Kan and Green, 1999). Therefore we have not yet been able to test whether the RS domain of SF2/ASF is required to antagonize the function of the ESS.

2.3.3 Improvement of the pyrimidine tract relieves RS domain-dependence

We next investigated whether the RS domain was required for U2AF recruitment functions of SF2/ASF when it is bound to the ESE position by testing a series of IgM M1-M2 derivative substrates with mutations or deletions in the pyrimidine tract, ESE, and ESS (see Figure 13). Previous work in our lab demonstrated a link between an RS domain requirement and the strength of the pyrimidine tract (Zhu and Krainer, 2000). Weakening the pyrimidine tract of an RS domain-independent β-globin pre-mRNA renders the pre-mRNA RS domain-dependent (Zhu and Krainer, 2000). We were interested to learn whether the requirement for an SR protein RS domain to overcome a

weak pyrimidine tract is specific to the β-globin pre-mRNA context or is a more general requirement for pre-mRNAs with weak pyrimidine tracts.

To determine whether improvement of the pyrimidine tract could relieve the RS domain-requirement for splicing of IgM M1-M2, we generated the IgM M1-M2 derivative substrate IgMPy↑ in which the wild-type pyrimidine tract 5'ACACUGUCUCUGUCACCUG3' was replaced by the improved pyrimidine tract 5'UUUUUUCCCUUUUUUUUC3' (as in Graveley et al, 2001) and tested it in S100 complementation with SF2/ASF and ΔRS (Figure 14). In the context of IgM M1-M2, improvement of the pyrimidine tract partly relieves the RS domain requirement (Figure 14). We also made the same pyrimidine tract improvement in all of the chimeric substrates that contained the downstream Mi segment. Figure 15 shows the *in vitro* splicing assays for the RS domain-dependent chimeric substrates and their Py↑ derivatives; in all cases, improvement of the pyrimidine tract completely relieved the RS domain requirement.

2.4 Discussion

2.4.1 RS domain-dependence and recruitment functions of SF2/ASF

The function of the ESE-bound SR protein in the context of IgM M1-M2 has been controversial, and there are two competing models for the mechanism by which SF2/ASF promotes splicing at the ESE position. In the recruitment model, SF2/ASF binds via its RS domain to U2AF35 to indirectly recruit U2AF65 to the pyrimidine tract (Graveley et al, 2001), whereas in the antagonism model, SF2/ASF prevents the PTB protein from binding to a downstream ESS (Shen et al, 2004). As mentioned in section 2.3.3, we were not able to determine whether the RS domain of SF2/ASF is required to counteract the

IgM M1-M2 ESS, and therefore turned our attention to investigating whether the RS domain is required for recruitment functions.

For all of the substrates we tested, improvement of the pyrimidine tract at least partially relieves the requirement for the RS domain, suggesting that the RS domain is required for recruitment of U2AF. In the T2TiMiT3 chimeric substrate, substitution of the tat23 downstream half of the intron with the corresponding intronic sequences from IgM M1-M2 renders the RS domain-independent tat23 pre-mRNA RS domain-dependent. Although the Mi segment includes sequences in addition to the pyrimidine tract, the Py↑ mutant of T2TiMiT3 is RS domain-independent, suggesting that the IgM M1-M2 pyrimidine tract confers the RS domain requirement and that this requirement is for the recruitment of U2AF. Other experiments with the IgM M1-M2 substrate strongly support the model for SR protein function in which an ESE-bound RS domain recruits U2AF to the pyrimidine tract. Using a derivative of the IgM M1-M2 pre-mRNA in which the ESE had been replaced with an MS2 binding site and the ESS had been deleted, Graveley et al showed that an ESE-bound RS domain was necessary for splicing in the context of the wild-type IgM M1-M2 pyrimidine tract, but not with an improved pyrimidine tract, and also demonstrated that in the presence of the wild-type pyrimidine tract, the ESE-bound RS domain was needed for recruitment of both U2AF35 and U2AF65 (Graveley et al, 2001).

From the chimeric splicing substrate in vitro splicing data, we noticed that the combination of the IgM M1-M2 upstream exon M1 and the upstream half of the IgM M1-M2 intron confers RS domain-dependence. Consistent with this, replacement of the first quarter of Ti in the RS domain-independent substrate M1TiMiM2 with the Ma

segment of the IgM M1-M2 upstream intronic sequences renders it RS domain-dependent.

Two possible mechanisms can be proposed by which this segment spanning the 5' splice site could function to confer RS domain-dependency. First, we note that the 5' splice site of IgM M1-M2 is predicted by several different algorithms to be intrinsically somewhat weaker than the tat23 5' splice site (Shapiro and Senapathy: IgM 69.63, tat 77.13; Neural Network: IgM 0.64, tat 0.93; Maximum Entropy: IgM 5.69, tat 9.07; Multiple Dependence Decomposition: IgM 10.18, tat 13.18; First Order Markov Model: IgM 4.39, tat 7.27; see Roca et al, 2005 for a description of these algorithms). Therefore, the RS domain of SF2/ASF might be more essential for binding of the U1 snRNP to the IgM M1-M2 5' splice site via recruitment interactions with U1-70K.

Second, the Ma segment could contain an intronic splicing silencer (ISS) whose negative effects can only be overcome by an SR protein with its RS domain. In the case of the regulation of inclusion of the c-src N1 alternative exon, an ISS immediately downstream of the N1 exon 5' splice site has been characterized as preventing inclusion of the N1 exon by interfering with 5' splice site-dependent assembly of U2AF into the E complex across the downstream intron. Interestingly, this c-src ISS does not function by preventing U1 snRNP from being recruited to the 5'ss, but instead blocks the interaction of the U1 snRNP with U2AF, preventing assembly of U2AF at the downstream 3′ splice site and formation of the E complex on this intron (Sharma et al, 2005).

In the case of IgM M1-M2, we find that either replacement of the Mi region spanning the 5'ss with tat23 sequences, or improvement of the pyrimidine tract, relieves RS domain-dependence, suggesting that the RS domain may be needed for intron

definition by recruitment of U2AF, which could be accomplished either indirectly from the 5'ss or more directly to the 3'ss via an ESE-bound SR protein. One possibility for SF2/ASF RS domain's function in intron definition for IgM M1-M2 is bridging the U1 and U2 snRNPs through the splicing coactivator SRm160/300. SRm160/300 was previously demonstrated to be required for the ESE-dependent recognition of a weak 3' splice site by a mechanism in which the U1 snRNP and ESE function together to recruit SRm160/300 and U2 snRNP, and intron definition assisted by the SRm160/300 coactivator was proposed to occur through interactions mediated by one or more SR proteins between SRm160/300 and U1 snRNP, and between SRm160/300 and the ESE (Eldridge et al, 1999).

2.4.2 SF2/ASF lacking its RS domain does not consistently support splicing *in vitro* in S100 complementation

Although we were at times able to achieve equivalent amounts of splicing for the tat23 substrate and other RS domain-independent substrates in S100 complementation with the ΔRS protein as with the SF2/ASF protein, the ΔRS protein behaved inconsistently in our *in vitro* splicing assays. We know that our preparations of ΔRS are fully active, as they are capable of demonstrating activity equivalent to that of SF2/ASF, as is evident from the data presented in this chapter. However, we also found that the same preparations of ΔRS protein sometimes supported less splicing than SF2/ASF, and in some assays demonstrated no activity whatsoever.

Through extensive troubleshooting, we believe we have ruled out several possibilities as explanations for the inconsistent behavior of ΔRS. We know that a failure

of activity of ΔRS in one assay is not a consequence of damage to the protein from freezing and thawing, as fully active ΔRS that subsequently shows no activity after refreezing and thawing can be again refrozen and thawed and demonstrates activity comparable to SF2/ASF. Likewise, we cannot attribute the poor behavior of the protein to the methods used for expression and purification. We have expressed and purified active ΔRS protein in two different ways, and ΔRS protein from multiple preparations from both methods behaves in the same way. Originally the ΔRS protein was expressed in *E. coli* as an untagged protein and purified away from nucleic acid by CsCl differential density centrifugation followed by denaturation in urea and purification by anion- and cation-exchange chromatography prior to refolding. We have subsequently expressed ΔRS as a C-terminally His-tagged fusion protein in 293-EBNA1 suspension cells and purified it by an ammonium sulfate cut followed by Ni-NTA agarose affinity chromatography. We have experienced the same inconsistent behavior from ΔRS protein we and others prepared by either method. Because proteins purified after expression in *E. coli* and in 293E cells behave in the same manner, we can also rule out the absence or presence of post-translational modifications as a contributing factor.

As there are many components in the *in vitro* splicing assay, we have also eliminated the order of addition of reagents during the assembly of the splicing reaction as a contributor to the inconsistent behavior. In an attempt to identify splicing assay conditions in which ΔRS exhibits consistent behavior, we have titrated polyvinyl alcohol, magnesium chloride, ATP and creatine phosphate, S100, and splicing substrate. We compared the behavior of ΔRS with substrate that was subjected to gel purification after *in vitro* transcription to substrate that was not. The behavior of the protein was checked

after assembly of the splicing reactions at both room temperature and on ice. We have also looked at the consistency of the behavior of the protein at different time points in the splicing assay. To try to help the ΔRS protein bind to the pre-mRNA, we preincubated the protein and pre-mRNA on ice prior to the addition of the other reaction components, but this also did not ameliorate the inconsistent behavior we see with ΔRS.

We can suggest several possible reasons for the variability in the amount of splicing that can be supported by ΔRS, but we believe that the most likely explanation is that the protein sometimes precipitates when added to the splicing reaction. The estimated isoelectric point of ΔRS is 8.36, as compared to the estimated isoelectric point of unmodified SF2/ASF of 10.36, suggesting that ΔRS would be less soluble than SF2/ASF, and probably not soluble at all under in vitro splicing assay conditions of pH 8.0 and 60mM salt. Whether purified after expression in *E. coli* or in mammalian cells, unlike SF2/ASF the purified ΔRS protein must be dialyzed against a high salt buffer for it to remain soluble, even if the majority of the RNA bound to it is removed. In fact, we have exploited the insolubility of ΔRS at lower salt concentrations as a final purification step to remove any contaminating proteins; lowering of the salt from 0.4M to 0.1M leads to fairly rapid precipitation of ΔRS and leaves contaminating proteins in the supernatant, after which the protein can be resuspended in 0.4M salt buffer with no apparent loss of activity, since protein purified in this manner can support the same amount of splicing as an equal quantity of the wild-type protein.

Several factors could contribute to precipitation of the ΔRS protein during assembly into the splicing reaction. We already know that the protein precipitates at 4°C when moved from its 0.4M salt storage buffer to a 0.1M salt buffer, but during assembly

of the splicing reaction on ice, the protein is diluted into a reaction with a final salt concentration of 60mM. In addition, the polyvinyl alcohol routinely added to the reaction to reduce the effective volume and promote assembly of splicing complexes would likely exacerbate any propensity for precipitation. Finally, SR proteins lacking their RS domains are known to come out of solution when they are at high concentrations and in a low salt buffer in the presence of nucleic acids (Ying Huang, personal communication).

If the ΔRS protein is indeed precipitating during the assembly of the splicing reactions, it may sometimes be able to assemble into splicing complexes in sufficient amounts to support splicing prior to its precipitation. In our troubleshooting, we have titrated the ΔRS protein to make sure we are assembling assays within the working concentration. When the amount of ΔRS in the reaction is doubled from 16 to 32 pmol, we more often see at least some splicing, though the amount of splicing supported is still inconsistent from assay to assay. This observation would be consistent with a failure of the ΔRS protein to be effectively incorporated into productive splicing interactions, rather than inactivity in splicing through a dominant-negative mechanism.

2.5 Materials and Methods

2.5.1 Cloning procedures

Chimeric splicing substrates were generated from the templates pSP64-tat23 (Krainer et al, 1990) and pSP65-μM1-M2 (Watakabe et al, 1993) by overlapping PCR (Pont-Kingdon, 2003) using Pfu Turbo polymerase (Stratagene) and DMSO at varying concentrations from 0-10%, with outside primers upstream of the SP6 transcription start

site and downstream of the restriction enzyme sites used for the transcription runoffs. PCR products were cloned into the pSP65 (Promega), pSP64 (Promega), or pCR-Blunt (Invitrogen) vectors.

Subchimeric splicing substrates were generated from pCR-Blunt-M1TiMiM2 by either overlapping PCR (pCR-Blunt-M1TiMiM2:B) as described above, or site-directed mutagenesis (pCR-Blunt-M1TiMiM2:A, pCR-Blunt-M1TiMiM2:C, and pCR-Blunt-M1TiMiM2:D). Site directed mutagenesis was performed according to the Stratagene Quikchange Site-Directed Mutagenesis Kit protocol, and DpnI-treated PCR reactions were transformed into either Top10 (Invitrogen) or DH5αF' *E. coli*.

The Watakabe ESS deletion substrate described in section 2.3.2 was generated by *in vitro* transcription of the pSP65-μM1-M2 (Watakabe et al, 1993) plasmid linearized at the SpeI site.

IgM M1-M2 substrates with mutations or deletions in the pyrimidine tract, ESE, and/or ESS (as shown in Figures 13 and 14) were generated as follows. The IgMPy↑ template was generated by overlapping PCR with pSP65-μM1-M2 as a template, using inside primers to introduce the mutant pyrimidine tract 5'UUUUUUCCCUUUUUUUUC3' in place of the wild-type pyrimidine tract 5'ACACUGUCUCUGUCACCUG3' as shown in Figure 1, Graveley et al, 2001, and cloned into the pCR-Blunt vector. The IgMΔKGESE template was generated by overlapping PCR with pSP65-μM1-M2 as a template, using inside primers to delete the 23 nt enhancer 5'GAAGGACAGCAGAGACCAAGA3' shown in Figure 9, Kan and Green, 1999, and cloned into the pCR-Blunt vector. The IgMPTB template was generated by site-directed mutagenesis of the pSP65-μM1-M2 plasmid to introduce the

mutant PTB site I 5'ACAUACGACAU 3' in place of the wild-type PTB site I 5'UCUUACGUCUU3' as shown in Figure 3A, Shen et al, 2004. The IgMPy↑ΔKGESE template was generated by overlapping PCR with pCR-Blunt- IgMΔKGESE as a template, using inside primers as described above to introduce the mutant pyrimidine tract in place of the wild-type pyrimidine tract. The IgMΔKGESEPTB template was generated by site-directed mutagenesis of the pCR-Blunt-IgMΔKGESE plasmid to introduce the mutant PTB site I as described above. Likewise, the IgMPy↑PTB and IgMPy↑ΔKGESEPTB templates were generated by site-directed mutagenesis of the pCR-Blunt- IgMPy↑ and pCR-Blunt-IgMPy↑ΔKGESE plasmids, respectively. Outside primers used for overlapping PCR were the same as described above for the construction of the chimeric substrates. After cloning, all transcription template sequences were verified by sequencing.

2.5.2 Protein expression and purification

SF2/ASF and ΔRS proteins were expressed as C-terminally His-tagged fusion proteins from the pTT3-SF2His and pTT3-SF2ΔRSCHis plasmids, respectively, after transfection with polyethylenimine (PEI) into 293-EBNA1 cells (Durocher et al, 2002). 293-EBNA1 cells (Invitrogen) were maintained in suspension culture at a density of 2.5 x 10^5 cells/mL in MEM Joklik suspension modification medium with L-glutamine (US Biological) supplemented with 5% calf serum (Gibco) and Penicillin/Streptomycin. For transfection, 1L of cells at 2.5 x 10^5 cells/mL were allowed to grow for 24 hours and then transfected by the addition of 1mg of plasmid (either pTT3-SF2His or pTT3-SF2ΔRSCHis), 50 mL media, and 2 mg PEI linear MW=25,000 (Polysciences) to the suspension cell culture.

After transfection, cultures were grown for three days to allow for protein expression. Pelleted cells were washed with PBS and resuspended in a lysis buffer consisting of 1 M NaCl, 0.1% Triton-X100, 20 mM β-mercaptoethanol, 50 mM Tris-HCl pH 8.0, 10 mM imidazole, supplemented with the EDTA-free Complete Mini protease inhibitor tablet (Roche) prior to lysis by sonication. The sonicate was subjected to centrifugation at 15,000 rpm, and a 0-30% ammonium sulfate cut was done on the supernatant, followed by another 15,000 rpm centrifugation. The second supernatant was diluted with an equal volume of lysis buffer without salt and then incubated at 4° C with Ni-NTA Agarose beads (Qiagen) in batch. Beads were washed on column with 50 volumes of lysis buffer without TX-100, and proteins were eluted in lysis buffer with 1 M NaCl and 300 mM imidazole and without TX-100. Fractions containing SR protein were combined and dialyzed twice against Buffer D, consisting of 20 mM Hepes-KOH pH 8.0, 0.2 mM EDTA, 20% glycerol v/v, 0.4 M KCl, 1 mM DTT, and 0.5 mM PMSF.

2.5.3 In vitro splicing assays

Splicing substrates were transcribed from linearized plasmid templates using SP6 RNA polymerase (Promega), essentially as described in Mayeda and Oshima, 1988, except G(5')ppp(5')G cap analog (NEB) was used instead of ^{7m}GpppG cap analog. Plasmids carrying templates for substrates with the IgM M1-M2 second exon M2 were linearized with XbaI, whereas plasmids carrying templates for substrates with the tat23 second exon T3 were linearized with BamHI. The tat23 and IgM M1-M2 transcripts were transcribed from the linearized plasmids pSP64-tat23 (Krainer et al, 1990) and pSP65-μM1-M2 (Watakabe et al, 1993), respectively. All transcripts were gel-purified.

Hela cell cytoplasmic (S100) and nuclear extracts were prepared as described in Mayeda and Krainer, 1999a. *In vitro* splicing assays were carried out essentially as described in Mayeda and Krainer, 1999b. Briefly, 10 μL reactions containing 0.5 mM ATP, 20 mM creatine phosphate, 20 mM HEPES-KOH pH 7.3, 2.6 % polyvinyl alcohol, 1.6-3.2 mM magnesium chloride, 20 fmol α^{32}P-UTP labeled splicing substrate, and either 40% S100 and 16 pmol SR protein or 30-40% nuclear extract were set up on ice. For S100 complementation reactions with SR proteins in 0.4 M KCl Buffer D, the final salt concentration was adjusted to 60mM using Buffer D without salt. Splicing reactions were incubated at 30 °C for four hours, and then subjected to phenol extraction and ethanol precipiation. RNA pellets were resuspended in formamide/bromophenol blue/xylene cyanol FF loading dye, and run at 700 V for 1 hour and 40 minutes in a 5.5% acrylamide/8.3M urea gel (National Diagnostics Sequagel). Bands were visualized by autoradiography using X-OMAT film (Kodak), or by exposure to a FUJI PhosphorImager screen, and were quantified using the Image Gauge software (Fujifilm Sciencelab 2003 Version 4.2).

2.5.3 RNA competitor experiments

Competitor RNAs corresponding to the Ma, Mb, Mc, and Md IgM M1-M2 intronic sequences were transcribed using the Ambion Megashortscript kit as per the manufacturer's instructions. Briefly, DNA oligos containing the Ma, Mb, Mc, and Md sequences were annealed to a common T7 promoter oligo to produce transcription templates for each competitor RNA. Transcription reactions were assembled according to the Megashortscript protocol, and were incubated at 37 °C overnight, followed by treatment with 3 μL of RQ1 RNAse-free DNAse (Promega) and removal of

unincorporated nucleotides by purification in a Microspin G-25 column (Amersham). RNAs were then phenol extracted, precipitated with sodium acetate, glycogen, and ethanol, resuspended in milliQ water, and quantified by UV spectrometry.

For the RNA competition experiments, splicing reactions were assembled on ice as described above, except without the 20 fmol α^{32}P-UTP-labeled splicing substrate, and RNA competitors were added at a 25-, 50-, or 100-fold excess. Splicing reactions were incubated at 30 °C for 10 minutes, returned to ice for addition of splicing substrate, and then incubated at 30 °C for four hours.

2.6 Acknowledgements

We would like to thank Jeanne Wiggins, Carmelita Bautista, and Margaret Falkowski for assistance with HeLa suspension cell culture and cell culture media, Phil Smith for the pTT3-SF2His plasmid and the 293E transfection protocol, and Zuo Zhang for the pTT3-SF2ΔRSCHis plasmid and the 293E ΔRS purification protocol. We would also like to thank Khalid Siddiqui for suggesting the use of DMSO in overlapping PCR reactions, Michelle Hastings for suggesting the use of the pCR-Blunt vector, and Gordon Zhang for an introduction to site directed mutagenesis.

2.7 References

Birney, E., Kumar, S., and Krainer, A. R. (1993). Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. Nucleic Acids Res *21*, 5803-5816.

Blencowe, B. J. (2000). Exonic splicing enhancers: mechanism of action, diversity, and role in human genetic diseases. Trends Biochem Sci *25*,106-10.

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

Caceres, J. F., Krainer, A. R. (1993). Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. EMBO J *12*, 4715-4726.

Cartegni, L., Chew, S L., Krainer, A. R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet *3*, 285-298.

Chew, S.L., Liu, H.-X., Mayeda, A., Krainer, A.R. (1999). Evidence for the function of an exonic splicing enhancer after the first catalytic step of pre-mRNA splicing. Proc Natl Acad Sci USA 96, 10655-10660.

Durocher, Y., Perret, S., Kamen, A. (2002). High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. Nucleic Acids Res *30*, E9.

Eldridge, A. G., Li, Y., Sharp, P. A., Blencowe, B. J. (1999). The SRm160/300 splicing coactivator is required for exon-enhancer function. Proc Natl Acad Sci USA *96*, 6125-6130.

Graveley, B. R. (2000). Sorting out the complexity of SR protein functions. RNA *6*, 1997-1211.

Graveley, B. R., Hertel, K. J., Maniatis, T. (2001). The role of U2AF35 and U2AF65 in enhancer-dependent splicing. RNA *7*, 806-18.

Hargous, Y., Hautbergue, G. M., Tintaru, A. M., Skrisovska, L., Golovanov, A. P., Stevenin, J., Lian, L. Y., Wilson, S. A., Allain, F. H. (2006). Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. EMBO J *25*, 5126-37.

Hertel, K. J., Graveley, B. R. (2005). RS domains contact the pre-mRNA throughout spliceosome assembly. Trends Biochem Sci *30*, 115-8.

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

Kan, J. L. C., Green, M. R. (1999). Pre-mRNA splicing of IgM exons M1 and M2 is directed by a juxtaposed splicing enhancer and inhibitor. Genes Dev *13*, 462-471.

Kielkopf, C. L., Rodionova, N. A., Green, M. R., Burley, S. K. (2001). A novel peptide recognition mode revealed by the x-ray structure of a core $U2AF^{35}/U2AF^{65}$ heterodimer. Cell 106, 595-605.

Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Luhrmann, R., Garcia-Blanco, M., Manley, J. L. (1994). Protein-protein interactions and 5'splice site recognition in mammalian mRNA precursors. Nature 368, 119-124.

Krainer, A. R., Conway, G. C., Kozak, D. (1990). Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. Genes Dev *4*, 1158-71.

Liu, H.-X., Zhang, M., and Krainer, A.R. (1998). Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. Genes Dev *12*, 1998-2012.

Liu, H.-X., Chew, S. L., Cartegni, L., Zhang, M. Q., and Krainer, A. R. (2000). Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. Mol Cell Biol *20*, 1063-1071.

Mayeda, A., Krainer, A.R. (1999a). Preparation of Hela cell nuclear and cytosolic S100 extracts for in vitro splicing. Methods Mol. Biol *118*, 309-314.

Mayeda, A., Krainer, A.R. (1999b). Mammalian in vitro splicing assays. Methods Mol Biol *118*, 315-321.

Mayeda, A., Oshima, Y. (1988). Short donor site sequences inserted within the intron of beta-globin pre-mRNA serve for splicing in vitro. Mol Cell Biol *8*, 4484-4491.

Mayeda, A., Screaton, G. R., Chandler, S. D., Fu, X. D., Krainer, A. R. (1999). Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. Mol Cell Biol *19*, 1853-1863.

Nagai, K., Oubridge, C., Jessen, T. H., Li, J., Evans, P. R. (1990). Crystal structure of the RNA-binding domain of the U1 small ribonucleoprotein A. Nature 345, 502-506.

Pont-Kingdon, G. (2003). Creation of chimeric junctions, deletions, and insertions by PCR.Methods Mol Biol *226*, 511-6.

Pozzoli, U., Sironi, M. (2005). Silencers regulate both constitutive and alternative splicing events in mammals. Cell Mol Life Sci *62*, 1579-1604.

Price, S. R., Evans, P. R., Nagai, K. (1998). Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA. Nature *394*, 645-650.

Reed, R. (1996). Initial splice-site recognition and pairing during pre-mRNA splicing. Curr Opin Genet Dev *6*, 215-220.

Roca, X., Sachidanandam, R., Krainer, A. R. (2005). Determinants of the inherent strength of human 5' splice sites. RNA *11*, 683-98.

Roscigno, R. F., Garcia-Blanco, M. A. (1995). SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. RNA *1*, 692-706.

Schneider, D., Tuerk, C., and Gold, L. (1992). Selection of high affinity RNA ligands to the bacteriophage R17 coat protein. J Mol Biol *228*, 862-869.

Selenko, P., Gregorovic G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A., Sattler, M. (2003). Structural basis for the molecular recognition between human splicing factors $U2AF^{65}$ and SF1/mBBP. Mol Cell 11, 965-976.

Sharma, S., Falick, A. M., 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*, 485-96.

Shen, H., Kan, J. L., Ghigna, C., Biamonti, G., Green, M. R. (2004). A single polypyrimidine tract binding protein (PTB) binding site mediates splicing inhibition at mouse IgM exons M1 and M2. RNA *10*, 787-794.

Shi, H., Xu, R.-M. (2003). Crystal structure of the *Drosophila* Mago nashi-Y14 complex. Genes Dev 17, 971-976.

Sun, H., and Chasin, L. A. (2000). Multiple splicing defects in an intronic false exon. Mol Cell Biol *20*, 6414-6425.

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

Tarn, W. Y, Steitz, J. A. (1995). Modulation of 5' splice site choice in pre-messenger RNAs by two distinct steps. Proc Acad Natl Sci USA *92*, 2504-2508.

Watakabe, A., Tanaka, K., Shimura, Y. (1993). The role of exon sequences in splice site selection. Genes Dev *7*, 407-418.

Wu, J. Y., Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell *75*, 1061-1070.

Zahler, A. M., Lane, W. S., Stolk, J. A., Roth, M. B. (1992). SR proteins: a conserved family of pre-mRNA splicing factors. Genes Dev *6*, 837-847.

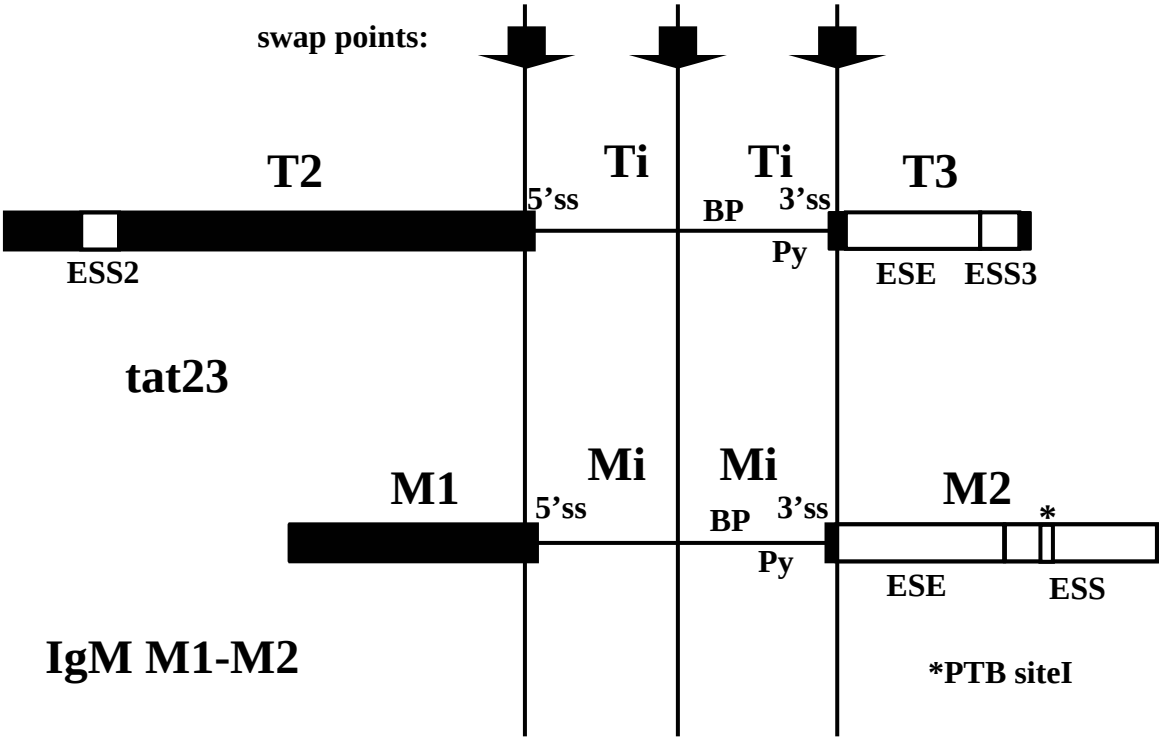
Zhu, J., Krainer, A.R. (2000). Pre-mRNA splicing in the absence of an SR protein RS domain. Genes Dev *14*, 3166-3178.

Zuo, P., Maniatis, T. (1996). The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. Genes Dev *10*, 1356-1368.

Zuo, P., Manley, J. L. (1993). Functional domains of the human splicing factor ASF/SF2. EMBO J. *12*, 4727-4737.

2.8 Figures

Figure 8. Schematic of chimeric substrates. A diagram of tat23 is shown in the upper half of the schematic. The tat23 pre-mRNA consists of the upstream exon T2 (black box) with its exonic splicing silencer ESS2 (white box), the intronic sequence (black line), and the downstream T3 exon (black box) with its exonic splicing enhancer ESE and exonic splicing silencer ESS3 (white boxes). A diagram of IgM M1-M2 is shown in the lower half of the schematic. The IgM M1-M2 pre-mRNA consists of the upstream exon M1 (black box), the intronic sequence (black line), and the downstream exon M2 (black box) with its exonic splicing enhancer ESE and its exonic splicing silencer ESS (white boxes). The position of PTB siteI identified within the IgM M1-M2 M2 exonic splicing silencer is marked with an asterisk. For both pre-mRNAs, the positions of the 5' splice site (5'ss), the branchpoint (BP), the pyrimidine tract (Py), and the 3' splice site (3'ss) are indicated. The three swap points used in the generation of chimeric substrates are indicated by arrows connected to lines that intersect with the pre-mRNAs at the swap positions.



Legend for Figure 9. In vitro splicing of chimeric substrates in S100 complementation with SF2/ASF and SF2/ASF lacking its RS domain.

Chimeric substrates with the M1 first exon and the T3 second exon are to the left of the first marker lane. A bracket on the left side of the gel indicates the positions of these substrates. The positions of the M1T3 mRNAs are indicated on the left side of the gel, and the bands corresponding to these splicing products are marked on the gel with a diamond. The positions of the M1 first exon splicing products for these chimeric substrates are indicated to the left side of the gel. The positions of the lariat products for M1MiMiT3 and M1TiMiT3 are indicated to the left side of the gel as "lariat", while the positions of the M1MiTiT3 and M1TiTiT3 lariats are indicated to the right side of the gel as "Ti" and the bands corresponding to these splicing products marked on the gel with arrowheads.

The parental tat23 and IgM M1-M2 substrates are shown between the marker lanes. The position of the T2T3 mRNA is indicated on the right side of the gel, and the bands corresponding to these splicing products are marked on the gel with a star. The positions of the tat23 lariats are indicated to the right side of the gel and the bands corresponding to these splicing products marked on the gel with arrowheads. The position of the M1M2 mRNA is indicated on the right side of the gel, and the bands corresponding to these splicing products are marked on the gel with a filled circle. The positions of the IgM M1-M2 lariats are indicated to the right side of the gel as "Mi".

Chimeric substrates with the M1 first exon and the M2 second exon are to the right of the second marker lane. The positions of the M1M2 mRNAs are indicated on the right side of the gel, and the bands corresponding to these splicing products are marked on the gel with a filled circle. The positions of the lariats for these chimeric substrates are indicated to the right side of the gel as "Mi". The positions of the M1 first exon splicing products for these chimeric substrates are indicated to the left side of the gel.

Figure 9. In vitro splicing of chimeric substrates in S100 complementation with SF2/ASF and SF2/ASF lacking its RS domain. (legend on previous page)

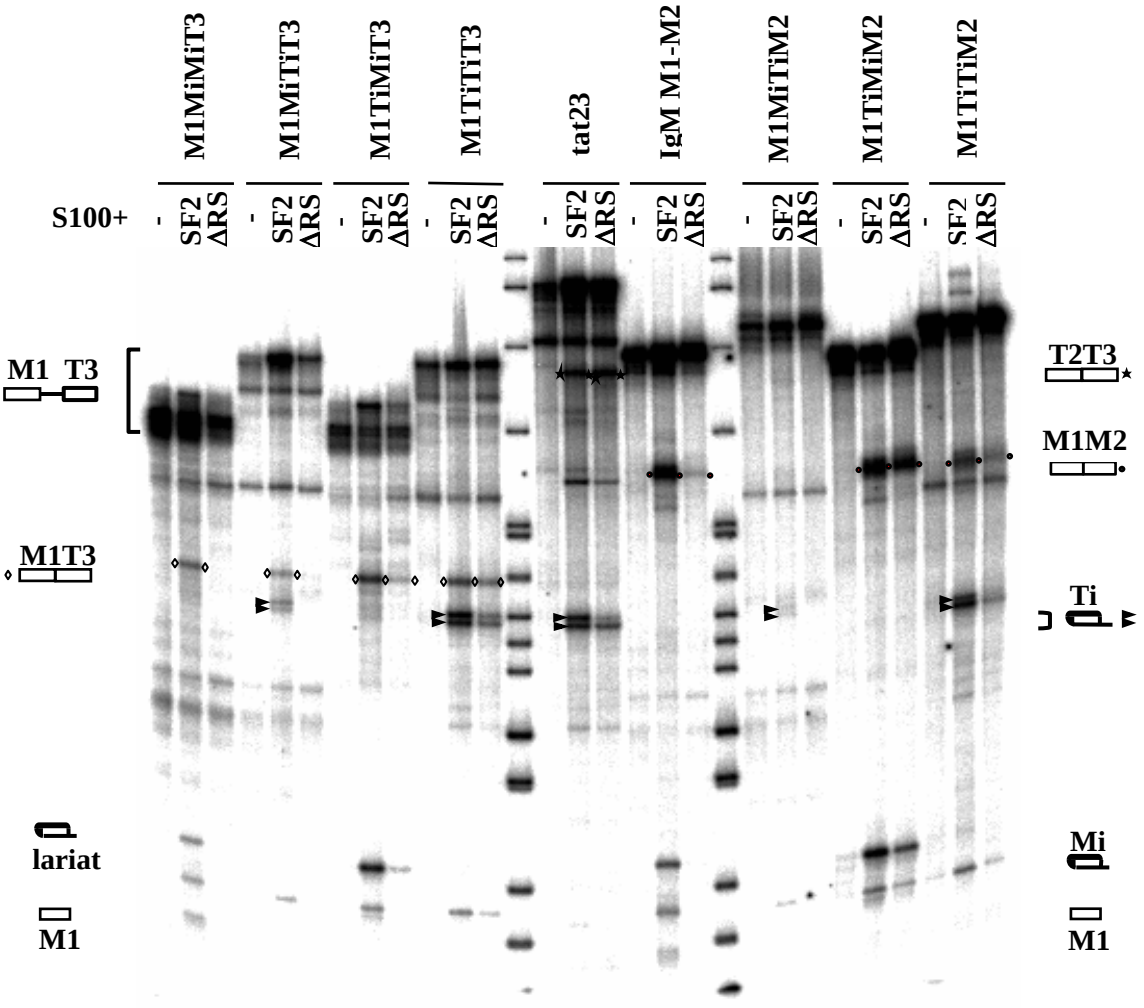


Table 1. Chimeric substrates and their RS domain-dependence. Each one of the 16 possible substrates in the set of chimerics is listed in the first column; substrates which were not tested are shown in italics, and parental substrates are in bold type. The magnesium chloride optima for splicing of each of the substrates in S100 complementation with SF2/ASF are indicated in the second column. The third and fourth columns show the percentage splicing for each substrate with SF2/ASF and ΔRS, respectively. Percentage splicing is calculated as mRNA divided by pre-mRNA x 100 as measured by densitometry from phosphoimaging at the end of the splicing reaction. As a measure of RS domain-dependence, the fifth column shows the percentage of splicing with ΔRS divided by the percentage of splicing with SF2/ASF for each substrate.

chimeric substrate	$MgCl_2$ optimum (mM)	% splicing with SF2	% splicing with ΔRS	% splicing with ΔRS / % splicing with SF2	mRNA with ΔRS / mRNA with SF2
T2TiTiM2	2.8	0	0	-	-
T2TiMiM2	3.2	17	31	1.82	0.90
T2MiMiM2	2.4	19	18	0.95	0.68
T2MiTiM2	-	-	-	-	-
T2MiTiT3	-	-	-	-	-
T2TiMiT3	2.4	50	0	0.00	0.00
T2MiMiT3	-	-	-	-	-
T2TiTiT3	3.2	11	14	1.28	1.09
M1MiMiT3	2.0	4	0	0.00	0.00
M1MiTiT3	2.8	4	0	0.00	0.00
M1TiTiT3	3.2	20	13	0.65	0.61
M1TiMiT3	2.8	35	8	0.23	0.18
M1TiMiM2	2.0	31	18	0.58	0.77
M1MiTiM2	2.4	0	0	-	-
M1TiTiM2	2.8	18	4	0.22	0.33
M1MiMiM2	1.6	22	3	0.14	0.09

Figure 10. Schematic of M1TiMiM2 subchimeric substrates. Four sequential portions--denoted as Ma, Mb, Mc, and Md--of the Mi intronic sequence were added back into chimeric substrate M1TiMiM2 in place of the corresponding Ti intronic sequence, to create the subchimeric substrates M1TiMiM2:A, M1TiMiM2:B, M1TiMiM2:C, and M1TiMiM2:D, respectively.

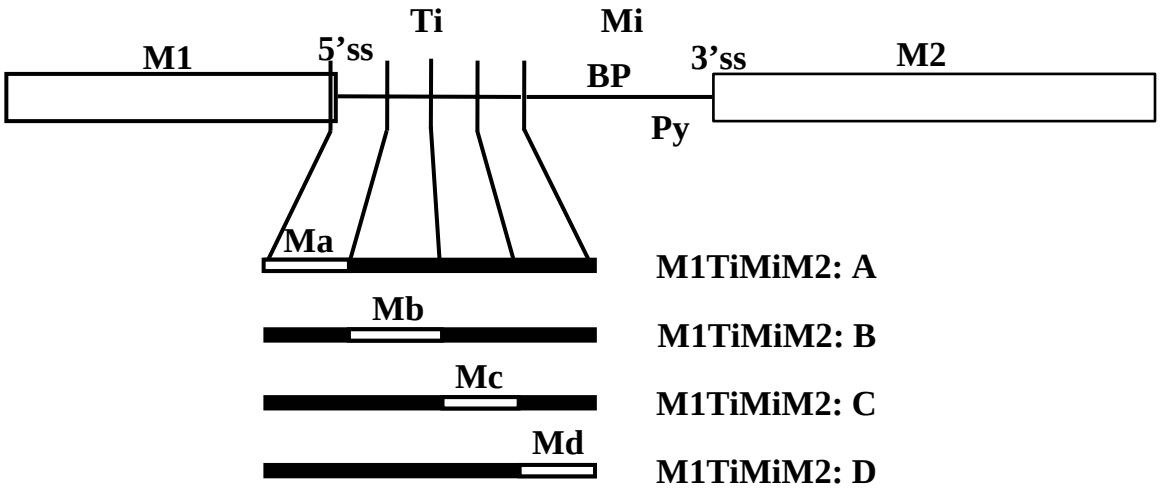


Figure 11. In vitro splicing of M1TiMiM2 subchimeric substrates in S100 complementation with SF2/ASF and SF2/ASF lacking its RS domain. The parental tat23 and IgM M1-M2 substrates are on the left hand portion of the gel. The position of the T2T3 mRNA is indicated on the left side of the gel, and the bands corresponding to these splicing products are marked on the gel with a star. The positions of the tat23 lariats are indicated to the left side of the gel and the bands corresponding to these splicing products are marked on the gel with arrowheads. The position of the M1M2 mRNAs are indicated on the right side of the gel, and the bands corresponding to these splicing products are marked on the gel with a filled circle. The positions of the IgM M1-M2, M1TiMiM2, and M1TiMiM2 subchimerics M1 first exon splicing products are indicated to the right side of the gel as "M1" and the positions of the lariats are indicated to the right side of the gel as "Mi".

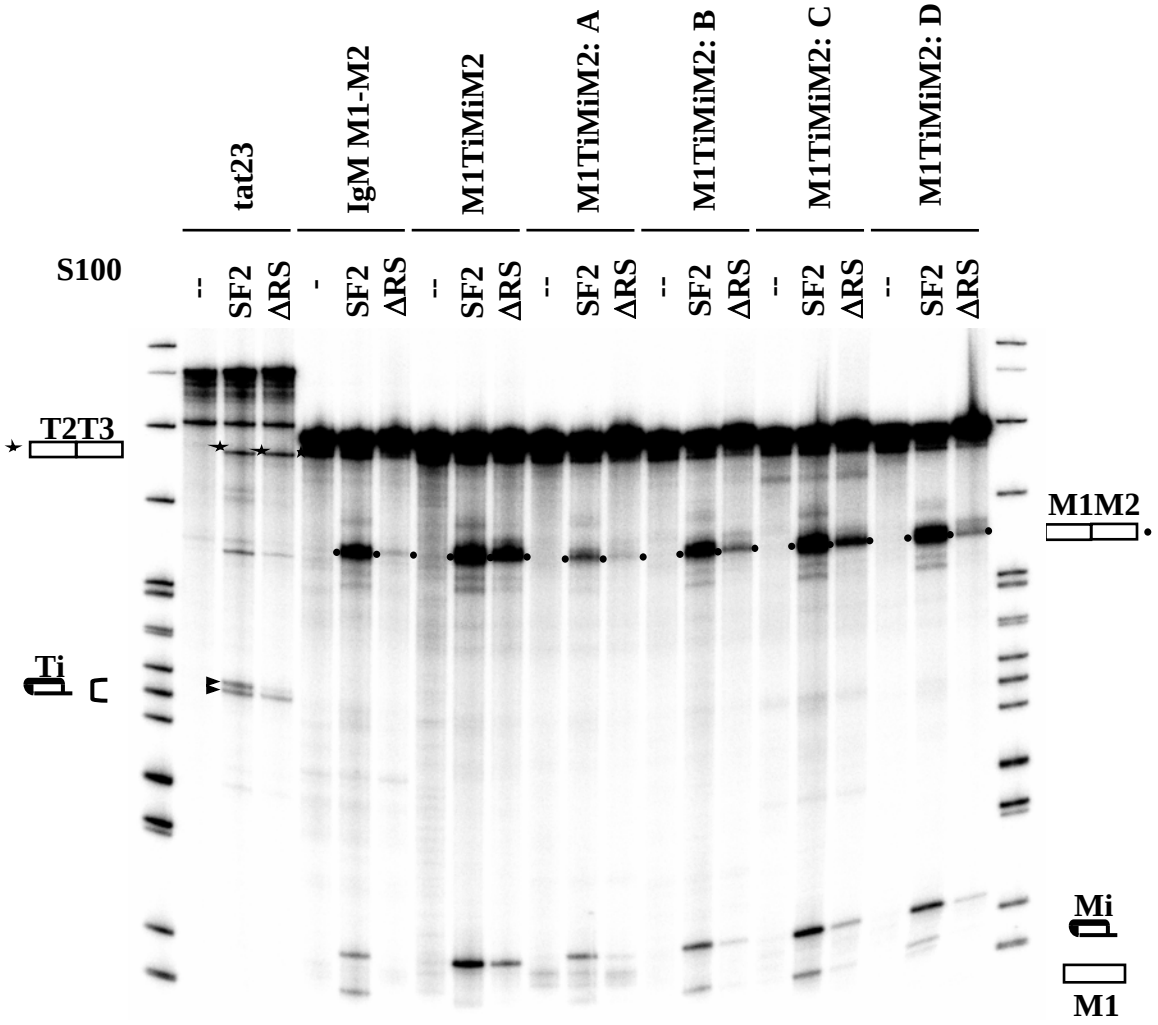


Figure 12. In vitro splicing of IgM M1-M2 PTB site I mutant substrate in NE and S100 complementation with SF2/ASF. The IgMPTB substrate is identical to the IgM M1-M2 pre-mRNA except for the PTB site I ESS mutation of UCUUACGUCUU to ACAUACGACAU. The positions of the M1M2 mRNAs are indicated on the left side of the gel, and the bands corresponding to these splicing products are marked on the gel with a filled circle. The positions of the M1 first exon splicing products are indicated to the left side of the gel as "M1" and the position of the lariats are indicated to the left side of the gel as "Mi".

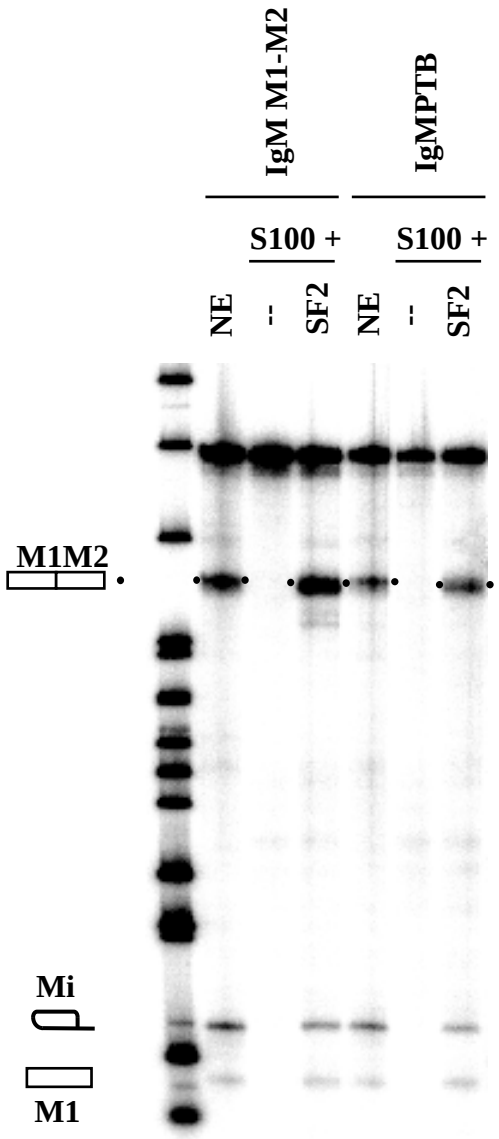


Figure 13. Schematic of IgM M1-M2 pyrimidine tract, ESE, and PTB site I mutant substrates. The pyrimidine tract located in the IgM M1-M2 intron and indicated by "Py" was changed from the wild-type pyrimidine tract 5'ACACUGUCUCUGUCACCUG3' to the mutant pyrimidine tract UUUUUUCCCUUUUUUUUC to produce Py substrates. The 23nt GAAGGACAGCAGAGACCAAGA exonic splicing enhancer, indicated by the box inside the M2 exon with an "x" marked through it, was deleted to produce ΔESE substrates. The PTB site I exonic splicing silencer located within the last half of the M2 exon was changed from the wild-type PTB site I of UCUUACGUCUU to the mutant site ACAUACGACAU to produce PTB substrates.

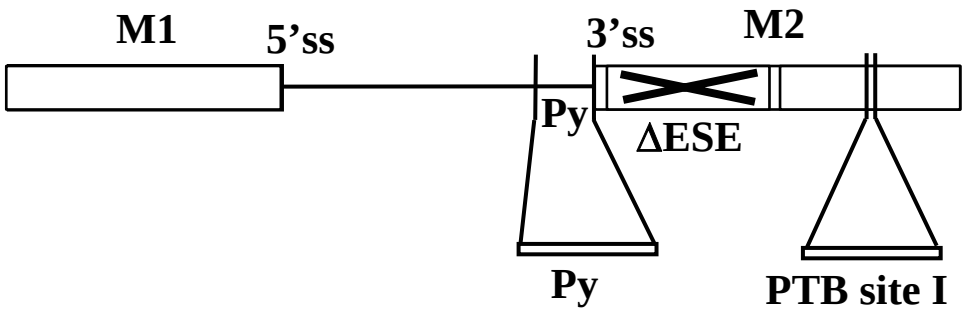


Figure 14. In vitro splicing of IgM M1-M2 pyrimidine tract, ESE, and PTB site I mutant substrates in S100 complementation with SF2/ASF and SF2/ASF lacking its RS domain. The position of the T2T3 mRNA is indicated on the left side of the gel, and the bands corresponding to these splicing products are marked on the gel with a star. The positions of the tat23 lariats are indicated to the left side of the gel and the bands corresponding to these splicing products are marked on the gel with arrowheads. The positions of the M1M2 mRNAs are indicated on the right side of the gel, and the bands corresponding to these splicing products are marked on the gel with a filled circle. The position of the IgM M1-M2 M1 first exon splicing product is indicated to the right side of the gel as "M1" and the position of the lariat is indicated to the right side of the gel as "Mi".

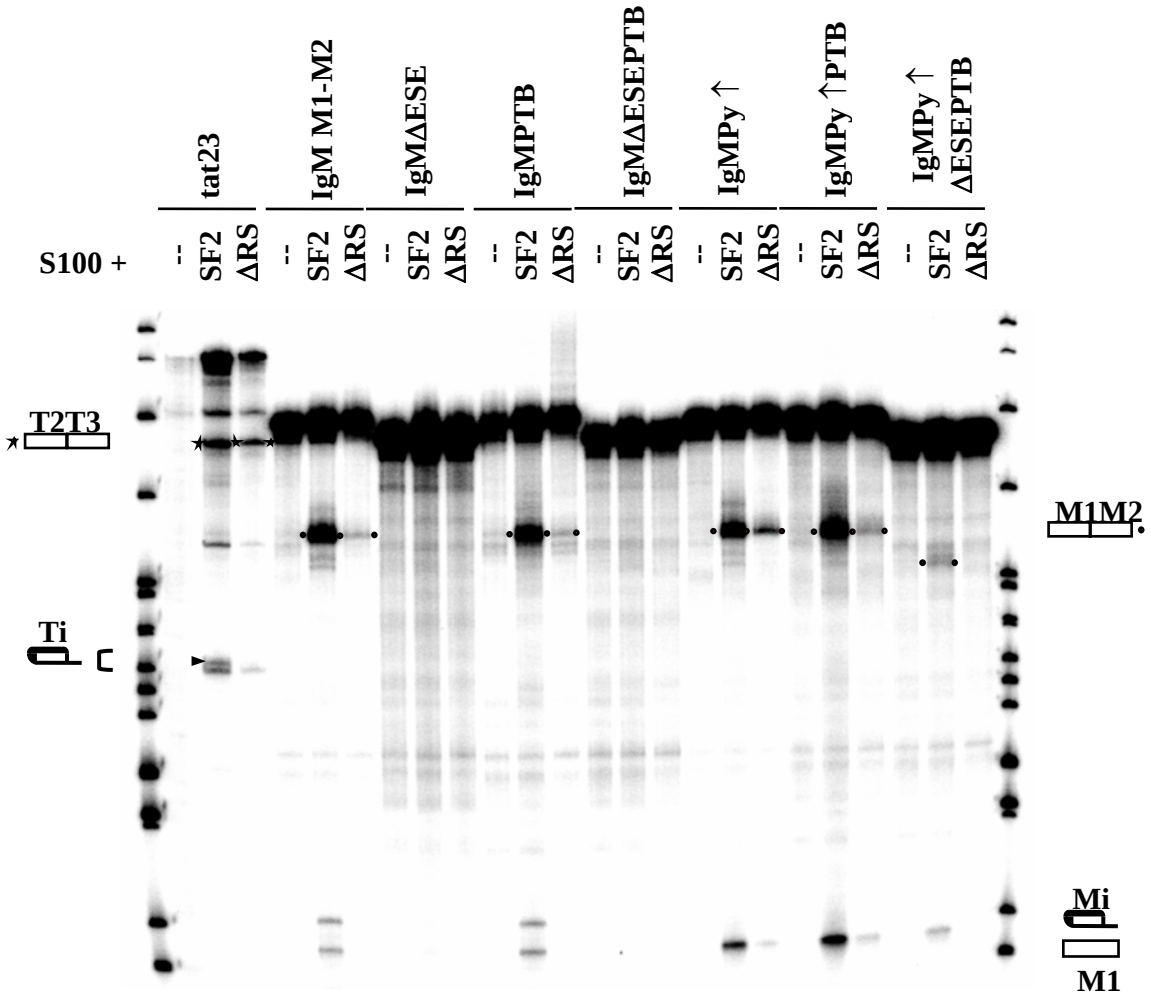
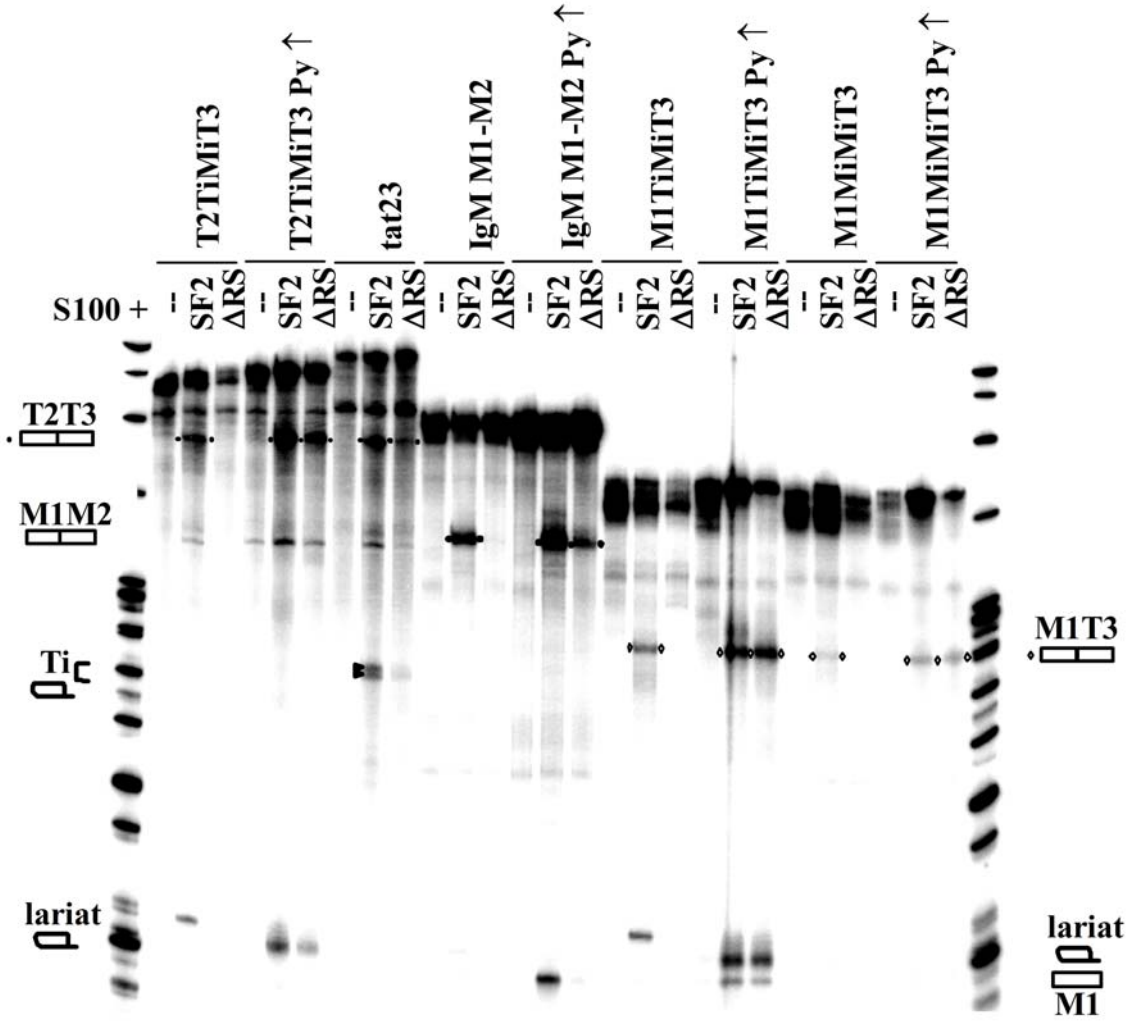


Figure 15. In vitro splicing of RS domain-dependent chimeric substrates with improved pyrimidine tracts in S100 complementation with SF2/ASF and SF2/ASF lacking its RS domain. The position of the T2T3 mRNA is indicated on the left side of the gel, and the bands corresponding to these splicing products are marked on the gel with a star. The positions of the tat23 lariats are indicated to the left side of the gel and the bands corresponding to these splicing products marked on the gel with arrowheads. The position of the M1M2 mRNA is indicated on the left side of the gel, and the bands corresponding to these splicing products are marked on the gel with a filled circle. The positions of the M1M2 mRNAs are indicated on the left side of the gel, and the bands corresponding to these splicing products are marked on the gel with a filled circle. The position of the M1T3 mRNAs are indicated on the right side of the gel, and the bands corresponding to these splicing products are marked on the gel with a diamond. The positions of the lariats for IgM M1-M2 and all of the chimeric substrates are indicated to the left and right sides of the gel as "Mi". The positions of the M1 first exon splicing products are indicated to the right side of the gel.



Chapter 3
The N-terminus of SF2/ASF is inhibitory for pre-mRNA splicing in vitro

3.1 Abstract

Using a mutant form of SF2/ASF lacking its RS domain, we previously confirmed that the RS domain was not required for *in vitro* splicing of some substrates, and therefore that pre-mRNAs could be classified as RS domain-independent or RS domain dependent. However, we have now identified a short inhibitory domain at the N-terminus of SF2/ASF that when deleted permits splicing in the absence of the RS domain of pre-mRNAs previously classified as RS domain-dependent. Therefore, we report that the RS domain of SF2/ASF is not required for constitutive splicing *in vitro*.

3.2 Introduction

SR proteins are composed of either one or two N-terminal RRM domains and a C-terminal RS domain. SF2/ASF is a prototypical member of the SR protein family, and consists of two N-terminal RRM domains separated by a glycine-rich linker region, followed by a C-terminal arginine- and serine-rich (RS) domain. SR proteins bind to exonic splicing enhancers (ESEs) via their RRM domains, and are thought to promote splicing by recruiting other components of the splicing machinery through interactions of their RS domains with the RS domains of other splicing factors (Graveley, 2000, Cartegni et al, 2002, see also section 1.4.2 for review). However, we have previously identified some pre-mRNAs that do not require the RS domain of SF2/ASF for splicing *in vitro* (Zhu and Krainer, 2000, Chapter 2), suggesting that SR proteins may not always need their RS domains to recruit other splicing factors.

3.3 Results

3.3.1 SF2/ASF lacking both its RS domain and the N-terminal segment preceding RRM1 consistently supports splicing in vitro in S100 complementation

In order to identify pre-mRNA sequence elements that confer an RS domain requirement for splicing of RS domain-dependent pre-mRNAs, we had previously tested a large number of different splicing substrates in the *in vitro* splicing assay with SF2/ASF and ΔRS, a mutant SF2/ASF lacking its RS domain. However, we were not able to achieve consistent behavior in the *in vitro* splicing assay with our fully active preparations of the ΔRS protein. In order to try to circumvent this problem, we obtained and tested a slightly different version of SF2/ASF lacking its RS domain from Dr. Gourisankar Ghosh's laboratory at UCSD. We were pleased to find that the Ghosh protein behaved reproducibly in our S100 complementation assays to splice the RS domain-independent pre-mRNA tat23.

The Ghosh protein was purified from *E. coli*, is untagged, and consists of amino acids 12-196 of SF2/ASF, deleting the N-terminal 11 amino acids of the protein and removing the C-terminal RS domain, which is generally defined as beginning with amino acid 198. We have made the ΔRS protein by two different methods. Our *E. coli* version of ΔRS is untagged and consists of amino acids 1-203, and we have also expressed amino acids 1-196 in 293-EBNA1 cells with a C-terminal His tag. As the activity of the ΔRS protein does not seem to be correlated with the exact C-terminal truncation or whether the protein is tagged at the C-terminus, we hypothesized that the difference between the consistency in the activity of the Ghosh protein and our ΔRS proteins may be due to the

N-terminal truncation of amino acids 1-11. Indeed, we saw the same inconsistent activity we had experienced with our ΔRS proteins with an untagged version of SF2/ASF consisting of amino acids 1-196 that was expressed in *E. coli* and purified for us by the Ghosh lab. For the sake of simplicity, we refer to versions of ΔRS lacking the N-terminal portion preceding the first RRM as "ΔNΔRS", and versions of ΔRS that contain this N-terminal segment as "ΔRS".

3.3.2 RS domain-dependent substrates can be spliced with SF2/ASF lacking both its N terminus and its RS domain

After establishing that the Ghosh ΔNΔRS protein consistently supported splicing of the RS domain-independent pre-mRNA tat23, we tested it with the RS domain-dependent pre-mRNA IgM M1-M2 in the hope that we had identified a protein that could be used to study functions of the RS domain of SF2/ASF. Surprisingly, the Ghosh ΔNΔRS also consistently supported splicing of IgM M1-M2, suggesting that the RS domain of SF2/ASF may not be required for splicing of constitutive substrates *in vitro*. To confirm this, we tested a number of our other RS domain-dependent substrates with ΔNΔRS, and in all cases the ΔNΔRS protein was able to complement S100 for splicing of substrates we had previously classified as RS domain-dependent. However, we did observe that substrates that we had previously classified as RS domain-dependent were spliced less well with ΔNΔRS than with SF2/ASF, and substrates that we had previously classified as RS domain-independent were spliced as well as or better with ΔNΔRS than with SF2/ASF. To illustrate this point, the splicing of the set of M1TiMiM2 and its subchimerics with ΔNΔRS is shown in Figure 17, as these substrates demonstrated a

range of RS domain-dependencies when previously tested with ΔRS (see Figure 11 and section 2.3.1).

We also tested the IgM M1-M2 Py/ESE/PTB mutant substrates shown in Figure 14 with ΔNΔRS, and found that although deletion of the N-terminal segment of SF2/ASF permits splicing of RS domain-dependent substrates in the absence of the RS domain, deletion of the N-terminus does not bypass the requirement of the ESE for splicing of IgM M1-M2.

3.3.3 Deletion of amino acids 2-11 increases the activity of both SF2/ASF and SF2/ASF lacking its RS domain

To verify that deleting the N-terminal segment confers upon ΔRS the ability to splice RS domain-dependent substrates, we purified our own version of the ΔNΔRS protein, ΔRS:Δ2-11, as well as the control protein SF2:Δ2-11, from 293-EBNA1 cells. (Our ΔNΔRS consisted of amino acids 12-196 with a C-terminal His tag, and our ΔNSF2 consisted of amino acids 12-248 with a C-terminal His tag.) Deletion of the N-terminal 11 amino acids increased the amount of splicing that could be supported by both ΔRS and SF2/ASF, with either the RS domain-independent tat23 or the RS domain-dependent IgM M1-M2 pre-mRNA (Figure 16, Figure 19, Figure 21). The increase in splicing for tat23 with ΔRS:Δ2-11 and SF2:Δ2-11 is apparently attributable to a stabilization of the pre-mRNA in the splicing extract, as for reactions containing these proteins the amounts of both pre-mRNA and mRNA are greater at the end of the splicing reaction (Figure 16, Figure 19), whereas the increase in splicing for IgM M1-M2 with ΔRS:Δ2-11 and SF2:Δ2-11 may be occurring through a different mechanism.

3.3.4 RS domain-dependent substrates are spliced at a slower rate with SF2/ASF lacking both its N terminus and its RS domain than with SF2/ASF

Although both RS domain-independent and RS domain-dependent substrates could be spliced with ΔRS:Δ2-11--demonstrating that the RS domain of SF2/ASF is not required for constitutive splicing *in vitro*--we saw that less splicing could be supported for the RS domain-dependent pre-mRNAs with ΔRS:Δ2-11 than with SF2/ASF or SF2:Δ2-11 (Figure 16, Figure 17, Figure 20). Although the RS domain is not strictly required for splicing of the RS domain-dependent substrates, it is possible that it plays a role in the efficiency or kinetics of splicing of these pre-mRNAs.

To test this hypothesis, we looked at the *in vitro* splicing of IgM M1-M2 in S100 complementation with the SF2/ASF, SF2:Δ2-11, ΔRS, and ΔRS:Δ2-11 proteins at different time points (Figure 20). The maximum amount of splicing for the proteins lacking the RS domain was observed at a later time point in the splicing reaction than for the proteins with the RS domain. For example, in this particular experiment, for SF2/ASF the maximum amount of splicing was seen at 2.0 hours, whereas for ΔRS:Δ2-11 the maximum amount of splicing was seen at 2.5 hours.

Although in this experiment, at the 2.0 hour time point the amount of mRNA produced with SF2/ASF is greater than the amount produced with ΔRS:Δ2-11, at the 2.5 hour time point the amount of mRNA produced with SF2/ASF is decreased and the amount produced with ΔRS:Δ2-11 is increased, such that they appear to be more or less equivalent. We observed that the ratios between the amount of mRNAs produced with each of the four proteins are different at the various time points, and we suggest that the

absolute amount of mRNA observed at each time point is a consequence of the combination of three different factors: the rate at which the pre-mRNA can be spliced with a given protein, the amount of splicing that can be supported by that protein, and the degradation (or stability) of the mRNA (and pre-mRNA) over time in the S100 extract during the incubation of the splicing reaction. The combined influences of these three different parameters explains why each different time point provides a different picture of the relative abilities of these four proteins to splice the IgM M1-M2 pre-mRNA.

We have repeated the time-course experiment several times and always obtain the same trend that the ΔRS:Δ2-11 protein splices IgM M1-M2 more slowly than SF2/ASF does, but we also observe variability in the kinetics of the splicing reactions from experiment to experiment. For example, in the experiment shown in Figure 20, the maximum amount of splicing with the ΔRS:Δ2-11 is achieved at the 2.5 hour time point, whereas in another experiments the splicing maximum occurs as much as a half hour earlier or later. We previously incubated all of our splicing reactions for 4.0 hours, and thus the variability of the kinetics of splicing in S100 and the degradation of the mRNAs and pre-mRNAs probably explains why for a given substrate we sometimes have more splicing with ΔNSF2 than with SF2/ASF and other times equivalent amounts of splicing with ΔNSF2 and SF2/ASF, and sometimes more splicing with SF2/ASF than with ΔRS:Δ2-11, but other times almost equivalent amounts of splicing with these two proteins.

3.3.5 Mutational analysis of the N terminus of SF2/ASF reveals amino acids that contribute to its inhibitory effect on splicing in vitro

In the time course experiment (Figure 20), we again observed that deletion of the 11 N-terminal amino acids from SF2/ASF and ΔRS significantly increased the amount of splicing that could be supported by these proteins, suggesting that the N-terminus has an inhibitory function. In an attempt to identify amino acids within this N-terminal region that may contribute to inhibition of splicing, we made and tested SF2/ASF and ΔRS proteins with mutations in the N-terminus. Amino acids 1-11 of SF2/ASF are MSGGGVIRGPA, and amino acids 5-10 are predicted to have beta sheet propensity (GOR4, Biology Workbench, San Diego Supercomputer Center, University of California at San Diego; Subramaniam, 1998). In support of the beta strand prediction for amino acids 5-10, several other proteins identified through a Basic Local Alignment Search Tool search (BLAST, National Center for Biotechnology Information) as having similar motifs to GVIRGP are known to adopt a beta sheet structure with their homologous residues, including DNA gyrase subunit A and Ton B-dependent ligand-gated channel. We made both SF2/ASF and ΔRS proteins with the following mutations at the N-terminus: deletion of amino acids 5-10 (the predicted beta strand), a triple mutant for amino acids 6-8 consisting of V6A/I7A/R8A and called VIR>AAA, V6A, I7A, R8A, P10A, and R8E.

The SF2/ASF and ΔRS N-terminus mutant proteins were tested in the *in vitro* splicing assay with IgM M1-M2 (Figure 19) to determine whether mutation of any of these amino acids relieves the inhibitory effect of the N-terminus. Most of the N-terminal mutations had little or no effect on the amount of splicing of IgM M1-M2 when made in the context of SF2/ASF (SF2:Δ5-10, SF2:VIR>AAA, SF2:V6A, SF2:I7A, and SF2:P10A) although one mutant SF2/ASF protein supported somewhat less splicing of

IgM M1-M2 than SF2/ASF (SF2:R8A). However, the SF2:R8E protein showed a drastic reduction in the amount of splicing with IgM M1-M2 relative to SF2/ASF. Of the ΔRS N-terminus mutant proteins, ΔRS:Δ5-10 and ΔRS:R8E showed a significant increase in splicing relative to their parental protein ΔRS, and for the ΔRS:Δ5-10 protein the level of splicing was similar to that seen with ΔRS:Δ2-11. We cannot at this time explain why the R8E mutation increases the ability of ΔRS to splice IgM M1-M2, but decreases the ability of SF2/ASF to splice the same substrate. However, the increase in splicing of IgM M1-M2 relative to ΔRS with the ΔRS:Δ5-10 and ΔRS:R8E proteins suggests that residues within amino acids 5-10 contribute to the inhibitory effect of the N-terminus, and that R8 may play a role in this inhibition. We note that this arginine residue is conserved in SR proteins with N-terminal extensions (Figure 18).

3.3.6 The N-terminal extension of RRM1 influences the ability of SF2/ASF to bind pre-mRNA

The N-terminal and C-terminal extensions of RRM domains have been demonstrated to play roles in the regulation of RNA binding and/or the accessibility of the antiparallel beta-sheet surface to nucleic acid (see section 3.4.3), so we wanted to know whether the SF2/ASF N-terminal extension has any influence on the affinity of the protein for splicing substrates. The ability of SF2/ASF, SF2:Δ2-11, ΔRS, ΔRS:Δ2-11, ΔRS:Δ5-10, and ΔRS:R8E to bind to IgM M1-M2 pre-mRNA was tested by UV crosslinking of the purified recombinant proteins to denatured RNA under splicing buffer conditions (Figure 21). Although there was little difference between the amount of RNA crosslinking observed for the SF2/ASF and SF2:Δ2-11 proteins, deletion of the amino

terminal extension greatly increased the ability of ΔRS to bind to the IgM M1-M2 substrate. These data strongly suggests that the inhibitory effect of the N-terminus on splicing of IgM M1-M2 with the ΔRS protein is due to the negative influence of this segment on the protein's ability to bind RNA.

3.4 Discussion

3.4.1 The RS domain of SF2/ASF is not required for constitutive splicing in vitro

We discovered through deletion of an inhibitory N-terminal segment of SF2/ASF that the RS domain of SF2/ASF is not strictly required for splicing *in vitro*. Although our finding at first seems to contradict the previous report from our laboratory that the RS domain is required for splicing of some pre-mRNAs (Zhu and Krainer, 2000) as well as our own observations about the RS domain-dependency of some splicing substrates, it is actually both another step forward in our understanding of the importance of the N-terminus of SF2/ASF in the regulation of its function, and an even stronger indicator that the existing models for how the domains of SR proteins function to promote splicing should be completely reevaluated. Initial experiments to test whether the RS domain of SF2/ASF is required for splicing were carried out with versions of ΔRS that were tagged at the N-terminus, and these experiments consistently showed that the RS domain of SF2/ASF was required for constitutive splicing (Caceres and Krainer, 1993, Zuo and Manley, 1993, Mayeda et al, 1999). Subsequent experiments using a version of ΔRS with no tag at the N-terminus unexpectedly yet clearly showed that the RS domain was completely dispensable for splicing of some substrates (Zhu and Krainer, 2000), and we have independently confirmed this finding with the same substrates employed by Zhu

(Chapter 2 and data not shown), as well as additional substrates (Chapter 2). We now find that deletion of the majority of the amino acids preceding RRM1 in the context of the ΔRS protein permits splicing of all of the substrates we have tested, whether previously characterized as RS domain-independent or as RS domain-dependent, demonstrating that the RS domain is not required for constitutive splicing *in vitro*.

However, we did observe that for pre-mRNAs we had previously characterized as RS domain-dependent, splicing with ΔRS:Δ2-11 was less efficient than with SF2/ASF. Because there is such a large increase in the amount of splicing supported for RS domain-dependent substrates with ΔRS by deletion of the N-terminal segment, our *in vitro* splicing data indicates that "RS domain-dependent" substrates are somehow more sensitive to the presence of the inhibitory N-terminal segment of SF2/ASF. We interpret these data as indicating that for these substrates, the RS domain normally performs an important function to overcome the inhibition conferred by the N-terminal segment of SF2/ASF preceding the first RRM domain.

One of the more interesting and puzzling aspects about splicing is that the spliceosome must be assembled onto a dizzying variety of pre-mRNA sequences that exist as folded RNA molecules coated with various proteins from 5' to 3' end. We propose that one possible explanation for "RS domain-dependence" could be that within the context of RS domain-dependent substrates, the binding of RRM1 of SF2/ASF to its target sequence(s) within the pre-mRNA is somewhat inhibited due to unfavorable secondary structure of the pre-mRNA or steric block by proteins bound adjacent to the RRM1 target, or both, and that the RS domain can assist SF2/ASF in initial recruitment

to its target through charge-mediated contacts with adjacent RNA sequences, bringing in the N-terminal RRM domain so that it can directly contact the RNA. In this model, the N-terminal domain may be inhibitory by functioning as a damper upon the RRM1 domain to interfere with its binding to the RNA, and the assistance provided by the RS domain normally overcomes the inhibition conferred by the N-terminus. There is precedent for interactions between SR protein RS domains and pre-mRNA, as an ESE-bound RS domain has been reported to make transient but direct contacts with specific RNA sequences during the course of splicing (Shen and Green, 2004); these contacts are postulated to occur with an unphosphorylated RS domain through an interaction of the positively charged arginine residues of the RS domain with the negatively charged phosphate backbone of the RNA (Shen and Green, 2004, Shen et al, 2004, Valcarcel et al, 1996).

3.4.2 Mechanisms for recruitment of spliceosomal components without an SR protein RS domain

Our discovery that the RS domain is not required for constitutive splicing directly contradicts prevailing models about how SR proteins function to promote splicing by recruiting other spliceosomal components. The previous finding that the RS domain was dispensable for splicing of some substrates but not for others (Zhu and Krainer, 2000, Chapter 2) hinted that SR proteins may be functioning to recruit spliceosomal components in a manner that does not involve SR protein RS domain-mediated protein-protein interactions, but because some substrates were found to require the RS domain, RS domain-mediated recruitment functions of SF2/ASF could not be formally eliminated

and were still hypothesized to occur for RS domain-dependent substrates. However, as we now know that the RS domain of SF2/ASF is not strictly required for *in vitro* splicing of constitutive substrates, it is unlikely that protein-protein interactions occurring through the RS domain of SR proteins are essential for recruitment of components of the splicing pathway.

In fact, when one considers what is already known about the requirements for the composition of RS domains in the context of SR proteins, it is quite difficult to envision how specific protein-protein interactions could occur through this unstructured domain. RS domains consist predominantly of alternating arginine and serine residues, with other amino acids interspersed between them; for example, the SF2/ASF RS domain sequence is: RSPSYGRSRSRSRSRSRSRSRSNSRSRSYSPRRSRGSPRYSPRHSRSRSRT. Several studies have shown that the RS domains of SR proteins are functionally interchangeable (Chandler et al, 1997, Wang et al, 1998, Graveley et al, 1998), and that the ability of an RS domain to activate splicing is directly related simply to the number of arginine and serine residues it possesses (Graveley et al, 1998, Cartegni and Krainer, 2003). The RS domain of SF2/ASF can even be replaced with a synthetic RS domain consisting merely of ten RS dipeptides (Zhu and Krainer, 2000, Cartegni and Krainer, 2003). Furthermore, a number of other splicing factors contain RS domains, including U2AF35 and U2AF65; snRNP components U1-70K, U5-100K, U4/U6.U5-27K, U4/U6.U5-65K, hLuc7p; splicing coactivators SRm160 and SRm300; RNA helicases Prp16 and HRH1; and the Clk/Sty protein kinases (Graveley, 2000, Zhou et al, 2002). It is difficult to understand how such a degenerate domain that is present in so many different splicing factors could engage in specific protein-protein interactions.

It is clear that an RS domain at the position of the ESE, whether synthetic (Zhu and Krainer, 2000, Cartegni and Krainer, 2003) or authentic (Graveley et al, 1998), and whether targeted there via an SR protein RRM domain (Zhu and Krainer, 2000), a heterologous RNA-binding domain such as the MS2 coat protein (Graveley et al, 1998, Shen et al, 2004), or an antisense oligonucleotide (Cartegni and Krainer, 2003), can function to promote splicing, and likely does so at least in part through influencing the recruitment of other splicing factors. From this we may conclude that one of the primary functions of ESEs is to recruit an RS domain. However, we also note that an ESE is still required for splicing of the "RS domain-dependent" substrate IgM M1-M2 with our ΔRS:Δ2-11 protein, which lacks an RS domain.

The traditional recruitment models for SR protein function assume that the RS domain of an SR protein interacts with the RS domain of another splicing component such as U2AF35 or U1-70K. However, if our ΔRS:Δ2-11 protein is indeed functioning to recruit other splicing components, it cannot be doing so through RS-RS domain interactions. Although there is abundant evidence that SR proteins promote the recruitment of specific essential splicing factors such as U2AF35 and U1-70K, which themselves have RS domains, it is not clear that this recruitment requires the RS domains of both of the involved proteins.

For example, the RS domain of SF2/ASF is not required for enhancement of U1 snRNP binding to alternative 5' splice sites (Eperon et al, 2000), which is presumed to occur through an interaction with the RS domain-containing protein U1-70K. Indeed, it is unlikely that the SF2/ASF RS domain can function to recruit U1 snRNP through

interactions with U1-70K, as the isolated RS domain of SF2/ASF is not sufficient for interaction with U1-70K (Xiao and Manley, 1997), and it is more probable that some portion of SF2/ASF other than the RS domain engages in protein-protein interactions with U1-70K to recruit U1 snRNP, as a GST-ΔRS fusion protein can engage in RNA independent protein-protein interactions with U1-70K (Xiao and Manley, 1997). Some labs have reported that at least one of either of the RS domains of SF2/ASF and U1-70K is required for interactions between these two proteins (Kohtz et al, 1994, Cao and Garcio-Blanco, 1998), but their experiments did not demonstrate that the RS domains of both proteins were required for their interaction with each other. In addition, these studies employed Far Western assays in which the target proteins were denatured, which would presumably eliminate any protein-protein interactions that can occur only through discontinuous epitopes created by the three-dimensional folding of protein domains, such as the RRM domains of both of these proteins.

The proposed recruitment function of SR proteins for which the most experimental evidence has been assembled is the model in which an ESE-bound SR protein engages via its RS domain in protein-protein interactions with the RS domain of U2AF35 to aid in the recruitment of U2AF65 to the pyrimidine tract (Zuo and Maniatis 1996, Graveley, 2000). Using a derivative of the IgM M1-M2 pre-mRNA in which the ESE had been replaced with an MS2-binding site and the ESS had been deleted, an MS2-RS domain fusion protein was found to be required for recruitment of both U2AF35 and U2AF65 in the presence of the wild-type pyrimidine tract of IgM M1-M2 (Graveley et al, 2001). However, an ESE is still required for splicing of the IgM M1-M2 substrate with our ΔRS:Δ2-11 protein that lacks an RS domain, suggesting that an SR protein may be

able to recruit U2AF without using its RS domain. Both the large and small subunits of U2AF have RS domains, and several lines of evidence suggest that only one of either of these two RS domains is required for efficient splicing (Zamore et al, 1992, Rudner et al, 1998a, Rudner et al, 1998b), leaving open the possibility that the RS domain of U2AF35 is likewise not needed for interaction with an ESE-bound SR protein.

Much as the U2AF heterodimer requires only one of its two RS domains for normal function, we suggest that dimers involving other RS domain-containing factors such as SF2/ASF may only require one RS domain to promote splicing. Thus the role of SR proteins in recruitment may not necessarily have to occur through their own RS domains, but could instead be through accomplishing the recruitment of at least one RS domain, whether an SR protein RS domain or an RS domain from another protein to which they bind, such as U2AF35 or U1-70K. Interestingly, unphosphorylated RS domains have been reported to directly contact the branchpoint and the 5' splice site during the course of pre-mRNA splicing (Shen and Green, 2004), and phosphorylated RS domains selectively contact double-stranded RNA regions, thus possibly promoting splicing by stabilizing base-pairing between U snRNAs and pre-mRNA splicing signals, particularly when the splicing signals are less than optimal for such base pairing (Shen and Green, 2006). In *Saccharomyces cerevisiae,* in which introns are short and have splicing signals that conform to strict consensus sequences (Lim and Burge, 2001), SR proteins and U2AF35 homologs are absent (Sanford et al, 2003), and U1-70K lacks an RS domain (Smith and Barrell, 1991). In *S. cerevisiae*, mutations in the 5'ss or branchpoint that reduce the ability of U1 and U2 snRNAs to base pair with these splicing signals interfere with splicing, but in the presence of such mutations, splicing can be

rescued by tethering an RS domain to the downstream exon in a position analogous to an ESE (Shen and Green, 2006).

We hypothesize that SF2/ASF lacking its RS domain can promote splicing by engaging in protein-protein interactions to recruit other spliceosomal components which have RS domains, and that these interactions could be occurring through surfaces of the SR protein RRM domains not involved in RNA binding. In fact, the RRM domains of the SR proteins SF2/ASF and SRp20 have already been demonstrated to interact with proteins not known to be involved in splicing (Ge et al, 1998, Elliott et al, 2000). There is precedent for the RRM domains of splicing factors functioning as protein-protein interaction domains. For example, the U2B" and U2A' proteins, both components of the U2 snRNP, must interact with each other for U2B" to bind hairpin IV of U2 snRNA; in this interaction, an alpha helix of the U2B" RRM domain on the opposite side from its RNA-binding surface is bound by the concave surface of a parallel β sheet within the leucine-rich repeat region of the U2A' protein (Price et al. 1998). The mode of dimerization of U2AF65 and U2AF35 has been determined to involve interactions between the RRM domain of U2AF35 and the proline-rich region of U2AF65 located between its N-terminal RS domain and its first RRM domain (Kielkopf et al, 2001). The splicing factor SF1 recognizes the branchpoint during the formation of E complex (Berglund et al, 1997, Kramer, 1992) and this recognition is stabilized through protein-protein interactions with U2AF65 (Berglund et al, 1998); the interaction between these two factors occurs through an N-terminal peptide of SF1 interacting with the face of RRM3 of U2AF65 opposite to its RNA-binding surface (Selenko et al, 2003). If the RRM domains of SR proteins can engage in protein-protein interactions with other

spliceosomal components, these interactions may also promote the RRM domain's recognition of RNA through changes in the conformation of the RNA-binding surface that occur as a consequence of protein binding.

3.4.3 Possible mechanisms of inhibition of splicing by the N terminus of SF2/ASF

The RRM domain exhibits a canonical secondary structure in a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology (see section 1.4.1), but RRM domains can have N-terminal and C-terminal extensions augmenting this core structure that are usually poorly ordered but in some cases adopt a secondary structure (Maris et al, 2005). Many SR proteins have N-terminal extensions preceding their first RRM domain (see Figure 18), including SF2/ASF. The N-terminal extension preceding the first RRM domain of SF2/ASF is approximately 15 amino acids in length, consisting of the primary amino acid sequence of (M)SGGGVIRGPAGNNDC.

Where their functions have been documented, N-terminal and C-terminal extensions of RRM domains regulate the ability of the RRM domain to bind nucleic acids. Protein segments N-terminal or C-terminal to the core RRM domain have been demonstrated to be critical for nucleic acid recognition for several splicing factors, including U1-70K (Query et al, 1989), U1A (Nagai et al, 1990, Oubridge et al, 1994, Jessen et al, 1991), PTB (Simpson et al, 2004), hnRNP C (Gorlach et al, 1994), and hnRNPA1 (Ding et al, 1999). In the U1A protein, the C-terminal helix extension has two modes of interaction with the RRM domain core (Allain et al, 1996). In the first mode, interaction of the protein with RNA is prevented when the helix is folded over onto the β-sheet surface, burying conserved hydrophobic residues that engage in RNA interactions

into a hydrophobic core composed of residues from the α-helix and the β-sheet. To allow the protein to interact with RNA, the helix rotates 136° away from the β sheet to expose the conserved hydrophobic RNA-binding residues, and instead interacts with an alternative hydrophobic patch on the protein. Likewise, the PTB RRM1 domain C-terminal extension folds over onto the β-sheet surface of the domain in an extended conformation (Simpson et al, 2004). In the case of hnRNPC, deletion of only five residues of the N-terminal extension abolishes RNA binding (Gorlach et al, 1992), whereas the C-terminal extension determines RNA-binding specificity (Gorlach et al, 1994). The N-terminal extension of the first RRM domain of hnRNPA1 forms a short 3_{10} helix that contacts bound nucleic acid (Shamoo et al, 1997, Xu et al, 1997, Ding et al, 1999).

In other RNA-binding proteins, N-terminal and C-terminal extensions also influence the availability of the RNA-binding surface of RRM domains. The La protein, which binds to the 3′ poly(U)-rich elements of nascent RNA polymerase III transcripts, has an atypical RRM domain composed of a five-stranded, antiparallel β-sheet terminated by a long C-terminal α-helix. Although this RRM domain of La has not yet been demonstrated to bind RNA, the C-terminal α-helix extends for five turns across what would be the RNA-binding surface of this RRM domain, and aliphatic residues from the helix engage in several hydrophobic interactions with an apolar region on the β-sheet (Jacks et al, 2003). The CstF-64 polyadenylation factor's N-terminal RRM is preceded by a short N-terminal helix with two turns, reminiscent of the N-terminal helix of hnRNPA1, and followed by a C-terminal helix that folds over onto the RNA-binding surface of the RRM and makes hydrophobic contacts with the conserved aromatic

residues of the RNA-binding motif; the C-terminal helix unfolds upon RNA binding to extend into the protein's hinge domain, where interactions with factors responsible for assembly of the polyadenylation complex occur (Perez Canadillas et al, 2003).

Based on our data and the documented functions of other RRM domain extensions, the most likely explanation for the inhibitory effect of amino acids 2-11 of SF2/ASF on splicing is that the N-terminal segment negatively regulates RNA binding by RRM1. We have observed in a UV crosslinking assay that deletion of the N-terminal extension greatly enhances binding of the ΔRS protein to IgM M1-M2 (Figure 21), and deletion of these amino acids increases the activity of both SF2/ASF and ΔRS (see section 3.3.3), and also increases the stability of the pre-mRNA and mRNA in some reactions (Figure 16). Progressive additions to the N-terminus of the ΔRS:Δ2-11 protein are correlated with reduced activity, as ΔRS supports splicing inconsistently (Chapter 2) and with only some substrates (Chapter 2, Zhu and Krainer, 2000), whereas N-terminal His-tagged ΔRS fails to support constitutive splicing in S100 complementation (Caceres and Krainer, 1993). We note that adding an N-terminal polyhistidine to the N-terminus of hnRNPA1 decreases its activity in splice-site switching *in vitro* relative to the untagged protein (L. Manche and A. R. Krainer, unpublished results), presumably by interfering with the regulation of RNA binding by the short 3_{10} helix (Ding et al, 1999). We propose that in the context of the splicing reaction the SF2/ASF RS domain normally assists in overcoming the inhibitory effect of the N-terminus, most probably by helping to recruit the RRM domains to the RNA. In agreement with this, we see little difference between UV crosslinking of SF2/ASF and SF2:Δ2-11 to IgM M1-M2 pre-mRNA, whereas the ΔRS protein binds less efficiently to IgM M1-M2 than does either SF2/ASF

or SF2:Δ2-11, but when the N-terminal extension is deleted from ΔRS its binding to the pre-mRNA is greatly enhanced (Figure 21).

3.5 Materials and Methods

3.5.1 Cloning procedures

Site-directed mutagenesis to delete sequences coding for amino acids 2-11 from the pTT3-SF2His and pTT3-SF2ΔRSCHis plasmids was performed according to the Stratagene Quikchange Site-Directed Mutagenesis Kit manufacturer's protocol, and Dpn I-treated PCR reactions were transformed into either Top10 (Invitrogen) or DH5αF' *E. coli*. Plasmids for expression of N-terminus mutant proteins were created by deletion or mutation of sequences coding for the specific amino acids from the pTT3-SF2His and pTT3-SF2ΔRSCHis plasmids by a site-directed mutagenesis strategy using a common reverse primer and mutagenic forward primers that overlap at their 5' ends with the reverse primer. The protein-coding regions for all protein expression plasmids were verified by sequencing.

Please see Chapter 2 Materials and Methods section 2.5.1 for detailed information about cloning of splicing substrates.

3.5.2 Protein expression and purification

The various SF2/ASF and ΔRS proteins were expressed as C-terminally His-tagged fusion proteins from the pTT3-SF2His and pTT3-SF2ΔRSCHis plasmids or derivatives of these plasmids as described in 3.5.1, respectively, after transfection with polyethylenimine (PEI) into 293-EBNA1 cells (Durocher et al, 2002). Proteins were expressed and purified as described in Chapter 2 Materials and Methods, section 2.5.2,

with the following modifications. After elution from the Ni-NTA Agarose column, fractions containing SR protein were combined, and if protein concentration was at least 3mg/mL after combining fractions, the proteins were dialyzed directly twice against Buffer D, consisting of 20mM Hepes-KOH pH 8.0, 0.2mM EDTA, 20% glycerol, 0.4M KCl, 1 mM DTT, and 0.5mM PMSF. If protein concentration was less than 3mg/mL after combining fractions, proteins were denatured by dialyzing into Buffer D with 6M urea at 0.1M KCl, and then concentrated at 4°C to approximately 3mg/mL using a Centricon-10 concentration device (Millipore Corporation) according to the manufacturer's instructions. Concentrated proteins were refolded by sequential dialyses in Buffer D containing 3M urea and 0.4M KCl, 1.5M urea and 0.4M KCl, 0.75M urea and 0.4M KCl, and into a final dialysis buffer of Buffer D consisting of 20mM Hepes-KOH pH 8.0, 0.2mM EDTA, 20% glycerol, 0.4M KCl, 1 mM DTT, and 0.5mM PMSF.

3.5.3 In vitro splicing assays

In vitro transcription and *in vitro* splicing assays were carried out as described in Chapter 2 Materials and Methods, section 2.5.3, except splicing reactions were incubated for 4 hours unless otherwise indicated.

3.5.4 UV crosslinking assays

Protein binding was performed in 20 μL reactions containing 32 pmol of protein and 80 fmol of uncapped IgM M1-M2 RNA labeled with all four rNTPs to a specific activity of 1.3×10^8 cpm/μg, under splicing reaction buffer conditions (excluding polyvinyl alcohol) containing final concentrations of 1.6 mM magnesium chloride, 20 mM HEPES-KOH pH 7.3, 0.5 mM ATP, 20 mM creatine phosphate, and 60 mM potassium chloride. RNA was denatured prior to incubation with protein; RNA mixes

containing 80 fmol of labeled IgM M1-M2, magnesium chloride, and water were assembled at 4 °C, heated at 95 °C for five minutes, and then returned to ice. 32 pmol of BSA or SR protein in buffer D was then added to each reaction along with the other splicing buffer components, and reactions were incubated at 30 °C for 30 minutes. Binding reactions were spotted onto parafilm and placed on ice prior to UV crosslinking at 0.864 J/cm^2 in a Spectrolinker XL-1000 UV Crosslinker (Spectronics Corporation). Reactions were returned to eppendorf tubes, and 3 μL of 27mg/mL RNAse A and 2 μL of 1000U/μL RNAse T1 (Roche) were added prior to incubation at 37°C for 15 minutes. Proteins and RNA were separated by 12% SDS-PAGE prior to visualization by autoradiography and phosphorimaging.

3.6 Acknowledgements

We thank Sutapa Chakrabarti for the *E. coli* ΔNΔRS (amino acids 12-196) and for the *E. coli* ΔRS (amino acids 1-196). We would also like to thank Jeanne Wiggins, Carmelita Bautista, and Margaret Falkowski for assistance with HeLa suspension cell culture and cell culture media, Phil Smith for the pTT3-SF2His plasmid and the 293E transfection protocol, and Zuo Zhang for the pTT3-SF2ΔRSCHis plasmid and the 293E ΔRS purification protocol. We are grateful to Hazeem Okonula for sharing his UV crosslinking protocol.

3.7 References

Allain, F. H., Gubser, C. C., Howe, P. W., Nagai, K., Neuhaus, D., Varani, G. (1996). Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. Nature *380*, 646-650.

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

Berglund, J. A., Abovich, N., Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. Genes Dev *12*, 858-867.

Blencowe, B. J. (2000). Exonic splicing enhancers: mechanism of action, diversity, and role in human genetic diseases. Trends Biochem Sci *25*,106-10.

Cao, W., Garcia-Blanco, M. A. (1998). A serine/arginine-rich domain in the human U1 70k protein is necessary and sufficient for ASF/SF2 binding. J Biol Chem *273*, 20629-20635.

Cáceres, J. F., Krainer, A. R. (1993). Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. EMBO J *12*, 4715-4726.

Cartegni, L., Krainer, A. R. (2003). Correction of disease-associated exon skipping by synthetic exon-specific activators. Nat Struct Biol *10*, 120-125.

Cartegni, L., Chew, S L., Krainer, A. R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet *3*, 285-298.

Chandler, S. D., Mayeda, A., Yeakley, J. M., Krainer, A. R., and Fu, X.-D. (1997). RNA splicing specificity determined by the coordinated action of RNA recognition motifs in SR proteins. Proc Natl Acad Sci USA *94*, 3596-3601.

Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R., Xu, R.M. (1999). Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. Genes Dev *13*, 1102-1115.

Durocher, Y., Perret, S., Kamen, A. (2002). High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. Nucleic Acids Res *30*, E9.

Elliott, D. J., Bourgeois, C. F., Klink, A., Stévenin, J., Cooke, H. J. (2000). A mammalian germ cell-specific RNA-binding protein interacts with ubiquitously expressed proteins involved in splice site selection. Proc Natl Acad Sci USA *97*, 5717-5722.

Eperon, I. C., Makarova, O. V., Mayeda, A., Munroe, S. H., Caceres, J. F., Hayward, D. G., Krainer AR. (2000). Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. Mol Cell Biol *20*, 8303-8318.

Ge, H., Si, Y., Wolffe, A. P. (1998). A novel transcriptional coactivator, p52, functionally interacts with the essential splicing factor ASF/SF2. Mol Cell 2, 751-759.

Gorlach, M., Wittekind, M., Beckman, R. A., Mueller, L., Dreyfuss, G. (1992). Interaction of the RNA-binding domain of the hnRNP C proteins with RNA. EMBO J *11*, 3289-3295.

Gorlach, M., Burd, C. G., Dreyfuss, G. (1994). The determinants of RNA-binding specificity of the heterogeneous nuclear ribonucleoprotein C proteins. J Biol Chem *269*, 23074-23078.

Graveley, B. R. (2000). Sorting out the complexity of SR protein functions. RNA *6*, 1997-1211.

Graveley, B. R., Hertel, K. J., Maniatis, T. (1998). A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. EMBO J *17*, 6747-6756.

Graveley, B. R., Hertel, K. J., Maniatis, T. (2001). The role of U2AF35 and U2AF65 in enhancer-dependent splicing. RNA *7*, 806-18.

Jacks, A., Babon, J., Kelly, G., Manolaridis, I., Cary. P. D., Curry, S., Conte, M. R. (2003). Structure of the C-Terminal Domain of Human La Protein Reveals a Novel RNA Recognition Motif Coupled to a Helical Nuclear Retention Element. Structure *7*, 833-843.

Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C., Nagai, K. (1991). Identification of molecular contacts between the U1 A small nuclear ribonucleoprotein and U1 RNA. EMBO J *10*, 3447-3456.

Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Luhrmann, R., Garcia-Blanco, M. A., Manley, J. L. (1994). Protein-protein interactions and 5’ splice-site recognition in mammalian mRNA precursors. Nature *368*, 119-124.

Kramer, A. (1992). Purification of splicing factor SF1, a heat-stable protein that functions in the assembly of a presplicing complex. Mol Cell Biol *12*, 4545-4552.

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

Maris, C., Dominguez, C., and Allain, F. H.-T. (2005). The RNA recognition motif, a plastic RNA binding platform to regulate post-transcriptional gene expression. FEBS Journal *272*, 2118-2131.

Mayeda, A., Oshima, Y. (1988). Short donor site sequences inserted within the intron of beta-globin pre-mRNA serve for splicing in vitro. Mol Cell Biol *8*, 4484-4491.

Mayeda, A., Screaton, G. R., Chandler, S. D., Fu, X.-D., Krainer, A. R. (1999). Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. Mol Cell Biol *19*, 1853-1863.

Nagai, K., Oubridge, C., Jessen, T. H., Li, J., Evans, P. R. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. Nature *348*, 515–520.

Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., Nagai, K. (1994). Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. Nature *372*, 432-438.

Perez Canadillas, J. M., Varani, G. (2003). Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. EMBO J *22*, 2821-2830.

Price, S. R., Evans, P. R., Nagai, K. (1998). Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA. Nature *394*, 645-650.

Query, C. C., Bentley, R. C., Keene, J. D. (1989). A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. Cell *57*, 89-101.

Rudner, D. Z., Breger, K. S., Rio, D. C. (1998a). Molecular genetic analysis of the heterodimeric splicing factor U2AF: the RS domain on either the large or small *Drosophila* subunit is dispensable in vivo. Genes Dev *12*, 1010-1021.

Rudner, D. Z., Breger, K. S., Kanaar, R., Adams, M. D., Rio, D. C. (1998b). RNA binding activity of heterodimeric splicing factor U2AF: at least one RS domain is required for high affinity binding. Molec Cell Biol *18*, 4004-4011.

Sanford, J. R., Longman, D., Caceres, J. F. (2003). Multiple roles of the SR protein family in splicing regulation. Prog Mol Subcell Biol *31*, 33-58.

Selenko, P., Gregorovic G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A., Sattler, M. (2003). Structural basis for the molecular recognition between human splicing factors $U2AF^{65}$ and SF1/mBBP. Mol Cell *11*, 965-976.

Simpson, P. J., Monie, T. P., Szendroi, A., Davydova, N., Tyzack, J. K., Conte, M. R., Read, C. M., Cary, P. D., Svergun, D. I., Konarev, P. V., Curry, S., Matthews, S. (2004). Structure and RNA interactions of the N-terminal RRM domains of PTB. Structure. *12*, 1631-1643.

Shamoo, Y., Krueger, U., Rice, L. M., Williams, K. R, Steitz, T. A. (1997). Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. Nat Struct Biol, *4*, 215–222.

Shen, H., 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., 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., Kan, J. L., Green, M. R. (2004). Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. Mol Cell 13, 367-76.

Smith, V., Barrell, B. G. (1991). Cloning of a yeast U1 snRNP 70K protein homologue: functional conservation of an RNA binding domain between humans and yeast. EMBO J *10*, 2627-2634.

Subramaniam, S. (1998). The biology workbench - A seamless database and analysis environment for the biologist. Proteins *32*, 1-2.

Valcarcel, J., Gaur, R. K., Singh, R., Green, M. R. (1996). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. Science *273*, 1706-9.

Wang, J., Xiao, S. H., Manley, J. L. (1998). Genetic analysis of the SR protein ASF/SF2: interchangeability of RS domains and negative control of splicing. Genes Dev *12*, 2222-2233.

Watakabe, A., Tanaka, K., Shimura, Y. (1993). The role of exon sequences in splice site selection. Genes Dev *7*, 407-418.

Wu, J. Y., Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell *75*, 1061-1070.

Xiao, S.-H., Manley, J. L. (1997). Phosphorylation of the ASF/AF2 RS domain affects both protein-protein interactions and is necessary for splicing. Genes Dev *11*, 334-344.

Xu, R.-M., Jokhan, L., Cheng, X., Mayeda, A., Krainer, A. R. (1997). Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. Structure, *5*, 559–570.

Zamore, P. D., Patton, J. G., Green, M. R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. Nature *355*, 609-614.

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

Zhu, J., and Krainer, A.R. (2000). Pre-mRNA splicing in the absence of an SR protein RS domain. Genes Dev *14*, 3166-3178.

Zuo, P., Maniatis, T. (1996). The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. Genes Dev *10*, 1356-1368.

Zuo, P., Manley, J. L. (1993). Functional domains of the human splicing factor ASF/SF2. EMBO J *12*, 4727-4737.

3.8 Figures

Figure 16. In vitro splicing of tat23 and IgM M1-M2 substrates in S100 complementation with SF2/ASF, SF2/ASF lacking its N-terminus, and SF2/ASF lacking both its N-terminus and its RS domain. The activity of SF2/ASF lacking its N-terminal 11 amino acids ("SF2:Δ2-11") and SF2/ASF lacking both its N-terminal 11 amino acids and its RS domain ("ΔRS:Δ2-11") was assessed by titration of increasing amounts (7 and 14 pmol) of protein as compared to 16 pmol of SF2/ASF.

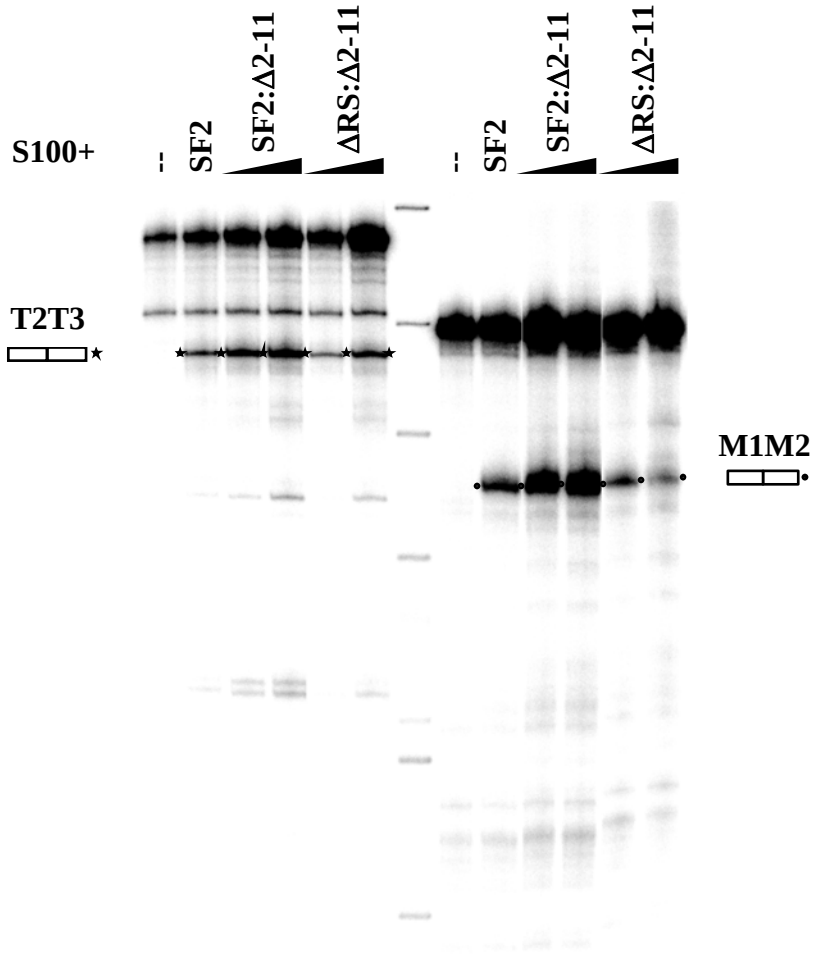


Figure 17. In vitro splicing of M1TiMiM2 subchimeric substrates in S100 complementation with SF2/ASF and SF2/ASF lacking both its N-terminus and RS domain. M1TiMiM2 subchimerics were spliced with either SF2/ASF ("SF2") or SF2/ASF lacking both its N-terminal 11 amino acids and its RS domain ("ΔNΔRS"). The parental tat23 and IgM M1-M2 substrates are on the left hand portion of the gel. The position of the T2T3 mRNA is indicated on the left side of the gel, and the bands corresponding to these splicing products are marked on the gel with a star. The positions of the tat23 lariats are indicated to the left side of the gel and the bands corresponding to these splicing products marked on the gel with arrowheads. The position of the M1M2 mRNA is indicated on the right side of the gel, and the bands corresponding to these splicing products are marked on the gel with a filled circle. The positions of the lariats for IgM M1-M2, M1TiMiM2, and M1TiMiM2 subchimerics are indicated to the right side of the gel as "MiMi/TiMi".

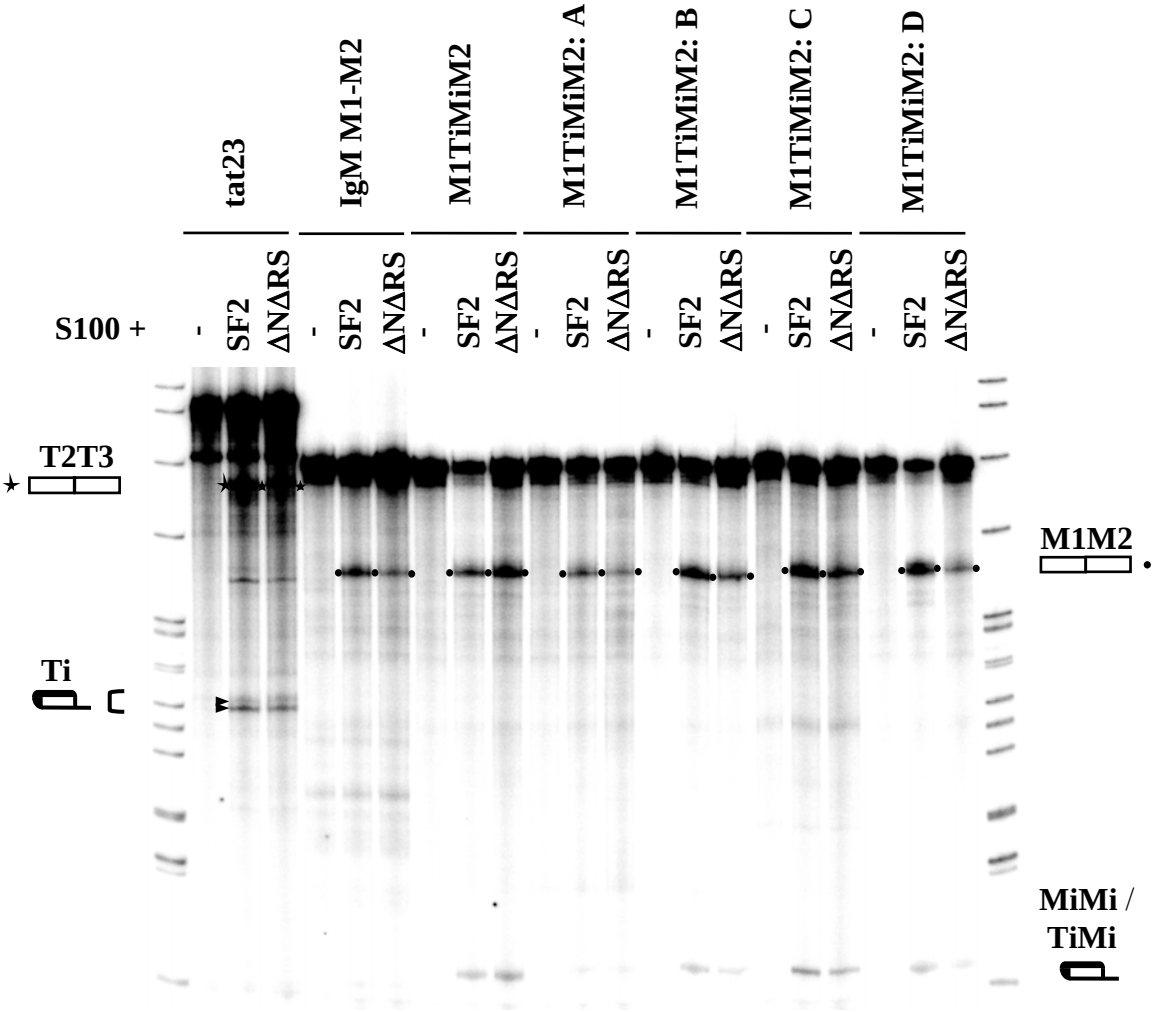


Figure 18. Phylogenetic Alignment of the N-terminus of SF2/ASF paralogs and orthologs. SR protein N-terminal RRM extensions were aligned using the ClustalW program. Accession numbers are provided for each sequence in the alignment. Sequences in the β1 strand are indicated by bold lettering, with the RNP2 motif underlined.

```
NP_008855 SF2/ASF H. sapiens          ------MSGGGVIRG-PAGNNDCRIYVGNL
NP_001071635 SF2/ASF M. musculus      ------MSGGGVIRG-PAGNNDCRIYVGNL
AAH95586 SF2/ASF D. rerio             ------MSGG-VIRG-PAGNNDCRIYVGNL
NM_001006918 SF2/ASF X. tropicalis    ------MSGGGVIRG-PAGNNDCRIYVGNL
AAF43413 SF2/ASF D. melanogaster      ----------------MGSRNECRIYVGNL
NP_499649 rsp-3 C. elegans            ------MP--------RGGSEDQKVYVGNL
Q13242 SRp30c H. sapiens              ------MS-GWADE--RGGEGDGRIYVGNL
NP_079849 SRp30c M. musculus          ------MSSGWADE--RGGEGDGRIYVGNL
NP_998064 SRp30c D. rerio             ------------------MSDGRIYVGNL
AAH84289 SRp30c X. laevis             ------MS-GWDREASRSGSGDGRIYVGNL
NP_001034554 SRp40 H. sapiens         ------------------MSGCRVFIGRL
NP_001073162 SRp40 M. musculus        ------------------MSGCRVFIGRL
NP_001002610 SRp40 D. rerio           ------------------MSGCRIFIGRL
AAH44085 SRp40 X. laevis              ------------------MSGCRVFIGRL
CAA91394 SRp40 C. elegans             ------------------MV--RVYIGRL
NP_006266 SRp55 H. sapiens            ------------------MP--RVYIGRL
NP_080775 SRp55 M. musculus           ------------------MP--RVYIGRL
NP_001008732 SRp55 D. rerio           ------------------MP--RVYIGKL
NP_788668 B52 D. melanogaster         ----------------MVGSRVYVGGL
NP_005617 SRp75 H. sapiens            ------------------MP--RVYIGRL
CAM14963 SRp75 M. musculus            ------------------MP--RVYIGRL
Q802Y1 D. rerio                       ------------------MS--RVYVGKL
CAD59160 SRp75 C. elegans             ----------------MAA--RIYIGRL
NP_003007 SC35 H. sapiens           MSYG---------RPPPDVEGMTSLKVDNL
NP_035488 SC35 M. musculus          MSYG---------RPPPDVEGMTSLKVDNL
AAH45229 SC35 X. laevis             MSYG---------RPPPDVEGMTSLKVDNL
AAF53192 SC35 D. melanogaster       MSNGGGAGGLGAARPPPRIDGMVSLKVDNL
NP_495014 SC35/rsp-4 C. elegans     MSRGGG----GDRRAAPDINGLTSLKIDNL
AAK72066 SC35/rsp-5 C. elegans        ------------------MP--RLYLGKI
AAK54350 SRp46 H. sapiens           MSCG---------RPPPDVDGMITLKVDNL
NP_003008 SRp20 H. sapiens            -----------MHR-DSCPLDCKVYVGNL
P84104 SRp20 M. musculus              -----------MHR-DSCPLDCKVYVGNL
AAH46661 SRp20 X. laevis              -----------MHR-DSCPLDCKVYVGNL
AAF54555 SRp20 D. melanogaster        -----------MPRYREWDLACKVYVGNL
AAF48264 SRp20 D. melanogaster        -----------MPRYREWDLACKVYVGNL
AAA82270 rsp-6 C. elegans             ------------------MDAKVYVGGL
NP_0010268549G8 H. sapiens            -----------MSRYGRYGGETKVYVGNL
NP_666195 9G8 M. musculus             -----------MSRYGRYGGETKVYVGNL
AAF52454 9G8 D. melanogaster          -----------MSR---HPSDRKVYVGDL
NP_004759 SRp54 H. sapiens            --------------MS--NTTVVPSTAGP
NP_081265 SRp54 M. musculus           --------------MINNHRDQTLYFSEH
NP_955870 SRp54 D. rerio              --------------MTSSSTS-VIQVTNV
AAF52825  SRp54 D. melanogaster       --------------MAGGNTPRVIQVTNI
O01159 SRp54/rsp-7 C. elegans         ------------MSGEEKEKVKILHVANI
CAA20690  SRp1p S. pombe              ---------------MSRRSLRTLYVTGF
CAB57400  SRp2p S. pombe              ------------------MSETRLFVGRI
```

Figure 19. In vitro splicing of IgM M1-M2 in S100 complementation with N-terminal mutants of SF2/ASF and SF2/ASF lacking its RS domain. The IgM M1-M2 substrate was spliced with SF2/ASF ("SF2"), SF2/ASF lacking its N-terminal 11 amino acids ("SF2:Δ2-11"), SF2/ASF lacking its RS domain ("ΔRS"), SF2/ASF lacking both its N-terminal 11 amino acids and its RS domain ("ΔRS:Δ2-11"), and N-terminal mutants of SF2/ASF and SF2/ASF lacking its RS domain, as indicated. The position of the M1M2 mRNA is indicated on the left side of the gel.

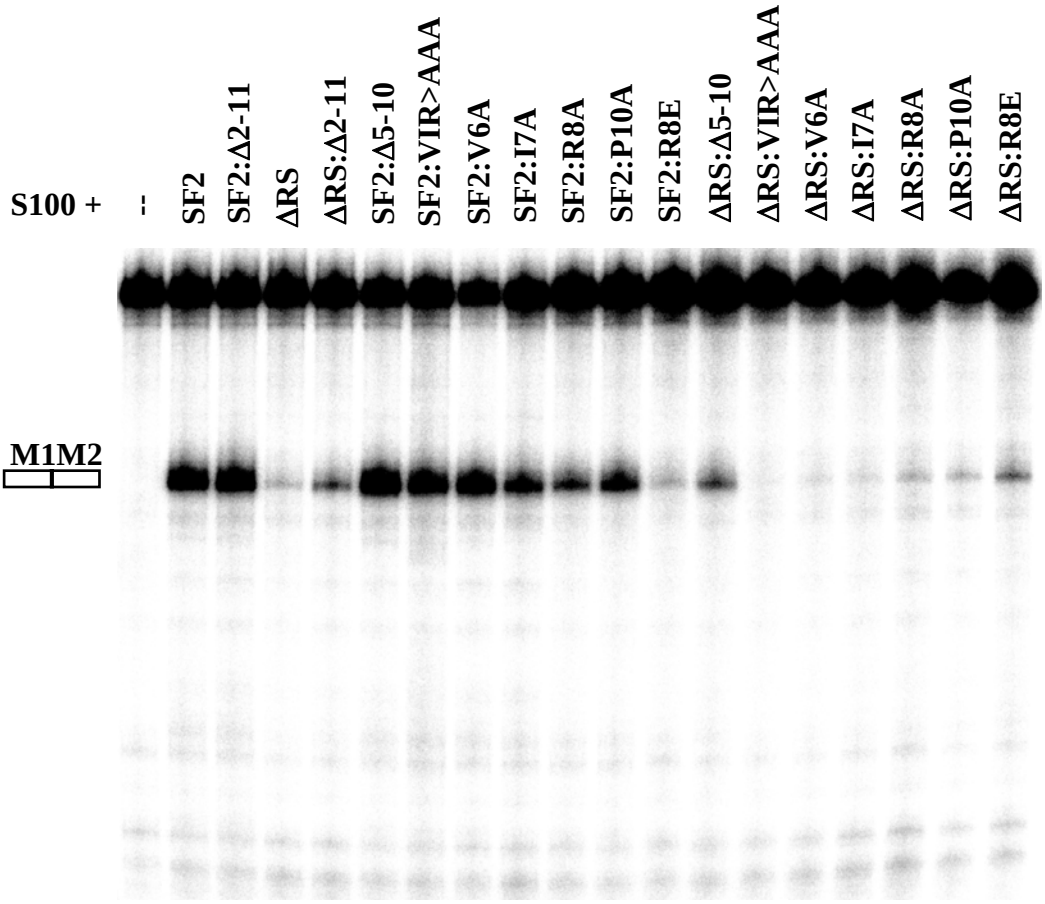


Figure 20. Time course of in vitro splicing of IgM M1-M2 in S100 complementation with SF2/ASF, SF2/ASF lacking its N-terminus, SF2/ASF lacking its RS domain, and SF2/ASF lacking both its N-terminus and RS domain. The IgM M1-M2 substrate was spliced with SF2/ASF ("SF2"), SF2/ASF lacking its N-terminal 11 amino acids ("SF2:Δ2-11"), SF2/ASF lacking its RS domain ("ΔRS"), and SF2/ASF lacking both its N-terminal 11 amino acids and its RS domain ("ΔRS:Δ2-11") for 2, 2.5, 3, 3.5, and 4 hours, as indicated. The position of the M1M2 mRNA is indicated on the left side of the gel.

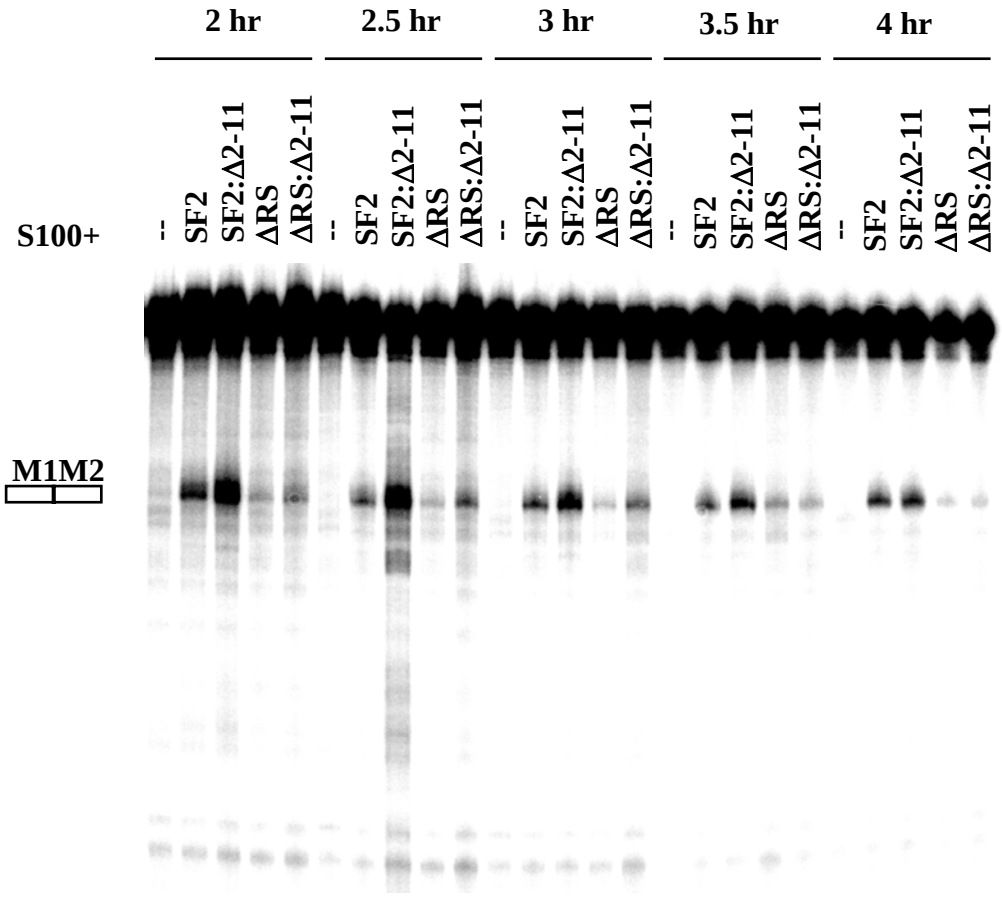
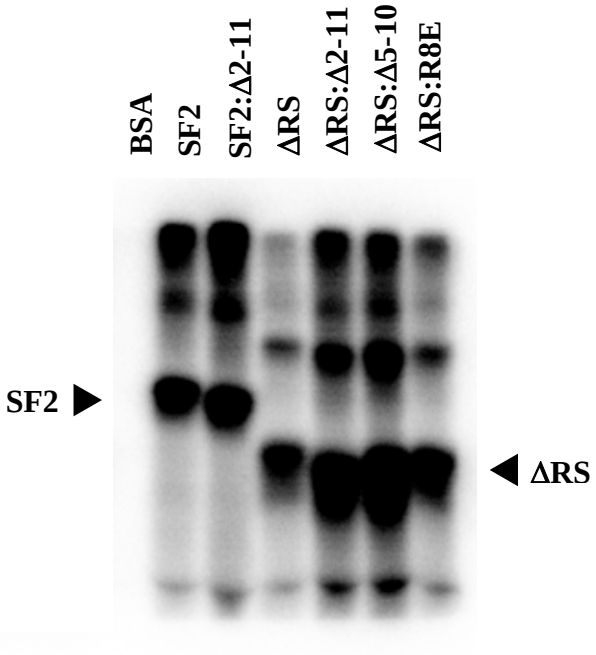


Figure 21. UV crosslinking of SF2/ASF and SF2/ASF N-terminal and RS domain mutant proteins to IgM M1-M2. 32 pmol of BSA or SR protein was incubated in splicing buffer with 80fmol of radiolabeled IgM M1-M2 pre-mRNA prior to UV crosslinking, RNAse digestion, and separation by SDS-PAGE.



Chapter 4
RRM domain mutations of SF2/ASF alter its splicing activity

4.1 Abstract

Using a mutant SF2/ASF protein lacking both its RS domain and its N-terminal 11 amino acids, we have demonstrated that the RS domain of SF2/ASF is not required for constitutive splicing *in vitro*. This discovery directly contradicts the traditional recruitment model for SR protein function, in which the RS domain of the SR protein is understood to promote splicing by engaging in protein-protein interactions with the RS domains of other essential splicing factors. Based on these findings we have undertaken a mutational analysis of the RRM domains of SF2/ASF to facilitate identification of amino acids outside of the RS domain that can affect SF2/ASF function, and have identified several residues that when mutated alter the activity of SF2/ASF in *in vitro* splicing.

4.2 Introduction

SR proteins are thought to promote pre-mRNA splicing by aiding in the recruitment of other essential splicing factors and by antagonizing the function of splicing silencers (Graveley, 2000, Cartegni et al, 2002). The recruitment functions of SR proteins are thought to require the arginine and serine-rich (RS) domain of the SR protein, and strong support for the model involving an RS domain-mediated mechanism of recruitment comes from experiments in which tethering an RS domain to the exonic splicing enhancer (ESE) position leads to the activation of splicing (Graveley et al, 1998, Cartegni and Krainer, 2003, Shen et al, 2004) and promotes the recruitment of U2AF35/65 to the branchpoint and pyrimidine tract (Graveley et al, 2001). However, through deletion of the N-terminal 11 amino acids of SF2/ASF, we have learned that the

RS domain of SF2/ASF is not required for constitutive splicing *in vitro* (see Chapter 3), calling into question this traditional model for the mechanism of recruitment of splicing factors by the RS domains of SR proteins.

Two interpretations about SR protein functions can be made based on our discovery that the RS domain of SF2/ASF is dispensable for splicing *in vitro*. First, as the RS domain of the SR protein is thought to be needed for recruitment of the splicing machinery, then the recruitment functions of SR proteins may not be required for splicing of some pre-mRNAs, and in such contexts the SR protein may be required solely for antagonism functions. Second, if the recruitment functions of the SR protein are required in any pre-mRNA context, then SR proteins may not need their RS domains to recruit the splicing machinery. We hypothesize that SR proteins can accomplish recruitment of other splicing factors through protein-protein interactions involving surfaces other than their RS domains, including their RRM domains, N-terminal and C-terminal extensions of the RRM domains, or linker regions that connect their RRM domains. Indeed, there is ample precedent for RRM domains functioning as protein-protein interaction domains in other essential splicing factors (see section 3.4.2), and SR proteins lacking their RS domains have already been documented to engage in protein-protein interactions with proteins not involved in splicing (Ge et al, 1998, Elliott et al, 2000).

SF2/ASF is composed of two N-terminal RRM domains separated by a glycine-rich linker and followed by a C-terminal RS domain. The RRM domain secondary structure typically exhibits a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology, resulting in a four-stranded antiparallel beta-sheet surface used to bind RNA (Hargous et al, 2006, Figure 7). RRM domains have two signature conserved motifs, RNP2 in the β_1 strand and RNP1 in the β_3

strand of the antiparallel beta-sheet, which are composed primarily of aromatic and positively charged residues involved directly in RNA binding (Dreyfuss et al, 1988, Birney et al, 1993, Auweter et al, 2006). The beta-sheets are backed by two perpendicular alpha helices, which we hypothesize could be protein-protein interaction surfaces for recruitment of other splicing factors when an SR protein is bound to RNA.

The effects of some different mutations in regions of SF2/ASF other than the RS domain on *in vitro* splicing have been assessed. Some of these mutations have established functions for amino acids in the conserved RNP motifs of the RRM domains. For example, within the β_3 strand of the RNP-1 motif of the SF2/ASF RRM1 are two highly conserved phenylalanine residues that participate in essential ring-stacking interactions with the RNA ligand (Birney et al, 1993), and mutation of these phenylalanines to aspartic acid abolishes RNA binding for SF2/ASF (Cáceres and Krainer, 1993). The roles of RRM1, RRM2, and the RS domain of SF2/ASF in localization and in constitutive and alternative splicing have been addressed in experiments with domain-deletion mutants of the protein (Cáceres and Krainer, 1993, Zuo and Manley, 1993, Caceres et al, 1997, Mayeda et al, 1999). Some studies have looked at the effect of mutating specific amino acids in RRM1 of SF2/ASF (Cáceres and Krainer, 1993), but with the exception of mutations made in the conserved heptapeptide SWQDLKD (Birney et al, 1993 Lutzelberger et al, 1999, Dauksite and Akusjarvi, 2002, Chiodi et al, 2004) the contributions of specific amino acids within RRM2 of SF2/ASF to splicing have been largely ignored.

In light of our finding that the RS domain is not required for constitutive splicing and therefore is unlikely to be engaging in protein-protein interactions essential for

splicing, we wanted to investigate whether amino acids in regions of SF2/ASF outside of its RS domain and not on the antiparallel beta-sheet RNA binding surface are important for SF2/ASF function. As any mutation can potentially disrupt the secondary structure of a protein, rather than mutate or delete portions of domains or multiple consecutive amino acids, we chose to address this question by introducing single amino acid substitutions so that the overall structure of the protein would be more likely preserved. To select the mutations to be made we employed CHAIN (Contrast Hierarchical Alignment and Interaction Network) analysis, a statistically based bioinformatics approach for identification of amino acids involved in critical structural features within a protein family (Neuwald et al, 2003, Neuwald 2006). CHAIN analysis utilizes Bayesian procedures to align all of the available amino acid sequences for a specific protein domain, and to then subdivide them based on shared sequence features into subfamily groups in a series of telescoping aligned sequence sets called a hierarchical alignment. Comparison of these aligned sequence sets against each other permits the identification of conserved sequence features unique to each subfamily group, which should reflect mechanisms that are functionally conserved after family divergence.

CHAIN analysis was performed by Dr. Andrew Neuwald using over 6000 RRM domain sequences, resulting in the identification of six highly constrained positions within the RRM domains of SF2/ASF: D46, R65, and D69 within RRM1, and W134, R142, and C148 within RRM2 (Figure 21). To try to infer where these amino acids may be located on SF2/ASF, the residues were modeled onto analogous positions on the crystal structure of UP1 (Xu et al, 1997, Ding et al, 1999), another splicing factor with two N-terminal RRM domains. Notably, five of the six amino acids identified by

CHAIN analysis were predicted to be located on the opposite face of the RRM domains from the antiparallel beta-sheet surface: R65, D69, W134, R142, and C148.

4.3 Results

Alanine mutations were generated for those amino acids identified by CHAIN analysis that were predicted to be on the face of SF2/ASF opposite the antiparallel beta-sheet of each RRM, in an attempt to disrupt any protein-protein interactions that could potentially occur through these residues. For amino acids located on the same RRM domain, we generated single alanine substitutions and also changed the amino acids to alanine in combination, to create the following mutations: R65A, D69A, and R65AD69A in RRM1; and W134A, R142A, C148A, R142AC148A, and W134AR142AC148A in RRM2. D46 was predicted to be in β1 on the antiparallel beta-sheet RNA-binding surface of SF2/ASF, so we chose to mutate the aspartic acid to an amino acid with the opposite charge (D46R) to disrupt any potential RNA interactions it might engage in.

4.3.1 The D69A mutation in RRM1 and the C148A mutation in RRM2 increase the activity of SF2/ASF in constitutive splicing in vitro

To determine whether any of the amino acids identified by CHAIN analysis are important for SF2/ASF function in constitutive splicing, CHAIN mutant proteins were tested for their ability to splice the tat23 substrate in S100 complementation (Figures 22 and 23). Of the RRM1 domain mutants, SF2:D46R and SF2:R65A proteins showed little difference with respect to the wild-type protein in their ability to splice tat23 (Figure 22). However, the D69A mutation increases the activity of SF2/ASF, as both SF2:D69A and

SF2:R65AD69A proteins demonstrated the ability to complement S100 at lower concentrations relative to the wild-type protein (Figure 22).

Of the RRM2 domain mutants, R142A and R142AC148A showed little difference with respect to the wild type protein in their ability to splice tat23 (Figure 23). S100 complementation with the SF2:R142A protein resulted in the production of slightly more tat23 mRNA relative to complementation with SF2/ASF, but both the R142A and R142AC148A proteins increased the stability of the pre-mRNA. However, the C148A mutation greatly increased the activity of SF2/ASF; SF2:C148A complemented S100 at much lower concentrations relative to the wild-type protein, and was also associated with a significant increase in RNA stability. Curiously, the significant increase in activity of the C148A mutant protein was abolished by the addition of the R142A mutation in the double mutant protein SF2:R142AC148A, suggesting that either the effect of the R142A mutation is dominant to C148A, or possibly that R142 and C148 work together.

4.3.2 The W134A mutation in RRM2 renders SF2/ASF inactive for constitutive splicing in vitro

The most striking effect created by changing a single amino acid identified by CHAIN analysis was seen in the case of the W143A mutation in RRM2 (Figures 23 and 24). SF2/ASF has only a single tryptophan in its entire sequence, and changing this amino acid to alanine completely inactivates it for constitutive splicing, whether in the context of the full-length protein (SF2:W134A and SF2:R142AC148AW134A, Figure 23; SF2:Δ2-11:W134A, Figure 24), or in the absence of the RS domain (ΔRS:Δ2-11:W134A, data not shown). Significantly, the inactivating W134A mutation is

dominant to the activating C148A mutation, as seen in the *in vitro* splicing assay with SF2:W134AR142AC148A. Inactivation of SF2/ASF by replacement of the tryptophan with alanine suggested that an aromatic residue was essential at this position, but our SF2:W134F protein was also completely inactive for constitutive splicing, demonstrating a specific requirement for the tryptophan at position 134 (Figure 24).

W134 is located within the SWQDLKD motif in the first alpha helix of RRM2 of SF2/ASF, a signature heptapeptide that is 100% conserved in all SR proteins with two RRM domains from *H. sapiens* to *S. pombe* (Birney et al, 1993). This motif has been mutated by other groups, also resulting in protein inactivation. Mutation of the SWQDLKD motif to ALQNVRN abolishes hSF2/ASF's activity as a splicing repressor (Dauksite and Akusjarvi, 2002). Changing the SWQDLKD motif to ALQNVRN in the *S. pombe* SR protein Srp2 has a negative effect on growth, presumably due to an accumulation of unspliced pre-mRNA (Lutzelberger et al, 1999). Within this conserved heptapeptide, a WDK to AAA mutation strongly decreases the activity of SF2/ASF in promoting mRNA translation, without decreasing its ability to bind to cytoplasmic mRNA (Sanford et al, 2005). A WDK to AAA mutation also abolishes recruitment of SF2/ASF to nuclear stress bodies induced by heat shock, and cotransfection of a plasmid expressing SF2/ASF with the WDK to AAA mutation alters the alternative splicing pattern of adenoviral E1A transcripts produced from a transfected E1A minigene (Chiodi et al, 2004).

4.4 Discussion

4.4.1 The RRM domains of SF2/ASF as potential protein-protein interaction domains

Our discovery that the RS domain of SF2/ASF is not required for constitutive splicing *in vitro* challenges the traditional model that SR proteins require their RS domains for protein-protein interactions to recruit other splicing factors. We therefore suggest that portions of SF2/ASF other than the RS domain may be involved in recruitment functions. Our data implicates the amino acid W134 in RRM2 as being essential for activity of SF2/ASF in constitutive splicing, and due to its location we propose that it may be involved in protein-protein interactions essential for SF2/ASF's activity in splicing. NMR structures of the isolated RRM1 and RRM2 domains of SF2/ASF have recently been determined by the Yokoyama group (He et al, 2005), revealing that W134 is displayed on the surface of SF2/ASF in a position that would permit it to be involved in protein-protein interactions.

Tryptophan is already known to play central roles in protein-protein interactions between other essential splicing factors. The U2AF35/65 heterodimer is locked together via reciprocal tryptophan interactions between the two subunits; U2AF65 Trp92 anchors the protein between the two alpha helices of the atypical RRM domain of U2AF35, whereas Trp134 of U2AF35 located on an exposed loop between $\alpha 2$ and $\beta 4$ interacts with a series of conserved proline residues in the U2AF65 linker region (Kielkopf et al, 2001). In an interaction that resembles the molecular interface between U2AF and U2AF65, SF1/mBBP inserts its conserved tryptophan into a hydrophobic pocket between the two alpha helices of U2AF65 RRM3 (Selenko et al, 2003).

Other groups have already suggested that residues within the SWQDLKD heptapeptide of SF2/ASF could be involved in protein-protein interactions (Dauksite and Akusjarvi, 2002, Chiodi et al, 2004, Sanford et al, 2005). An alternative explanation for the inactivation of SF2/ASF in splicing by the W134A and W134F mutations is that W134 regulates the binding of SF2/ASF to its RNA target; however, in the NMR structure W134 is located on the opposite side of the protein from the canonical RNA-binding surface. In addition, the WQD to AAA mutation does not affect SF2/ASF's ability to bind to cytoplasmic RNA in its translational regulation function (Sanford et al, 2005), and therefore mutating the single amino acid tryptophan to alanine or phenylalanine would not be expected to interfere with its ability to bind to pre-mRNA.

4.4.2 SF2/ASF RNA binding and splicing substrate stability

We have observed that several SF2/ASF mutant proteins tend to be associated with increased stability of splicing substrates and products during the splicing reaction, particularly SF2:C148A (this Chapter), and SF2:Δ2-11 and ΔRS:Δ2-11 (Chapter 3). This increase in RNA stability usually seems to go hand in hand with the ability of the mutant protein to support the production of a larger amount of mRNA product. However, we have also many times observed the reverse phenomenon, in which substrates that are not spliced are subject to increased degradation in the splicing reaction, such as when substrates are incubated in S100 without the addition of SR protein or in the cases in which the splicing of a specific substrate cannot be supported by the SR protein provided for complementation. Thus, we cannot automatically assume that the increase in splicing activity of our mutant proteins that also promote RNA stability is a direct consequence of

improving their ability to bind to the pre-mRNA substrates. The increase in RNA stability we have observed with these mutant proteins may be merely a consequence of their ability to recruit pre-mRNA species into splicing complexes, thereby blocking the accessibility of RNA to enzymes in the crude extract that are responsible for their degradation. In this way the increase in stability of RNA conferred by the mutant proteins could be a direct consequence of improved pre-mRNA binding, and/or an indirect consequence of an improved ability to recruit spliceosomal components subsequent to pre-mRNA binding.

4.5 Materials and Methods

4.5.1 CHAIN analysis

Contrast Hierarchical Alignment and Interaction Network analysis was performed as described in Neuwald et al, 2003.

4.5.2 Cloning procedures

Site-directed mutagenesis to delete sequences coding for amino acids identified by CHAIN analysis from the pTT3-SF2His, pTT3-SF2His:Δ2-11, and pTT3-SF2ΔRSCHis:Δ2-11 plasmids was performed according to the Stratagene Quikchange Site-Directed Mutagenesis Kit manufacturer's protocol, and Dpn I-treated PCR reactions were transformed into either Top10 (Invitrogen) or DH5αF' *E. coli*. The protein coding regions for all protein expression plasmids were verified by sequencing.

4.5.3 Protein expression and purification

SF2/ASF and ΔRS proteins were expressed as C-terminally His-tagged fusion proteins from the pTT3-SF2His and pTT3-SF2ΔRSCHis plasmids or derivatives of these

plasmids as described in 4.5.2, after transfection with polyethylenimine (PEI) into 293-EBNA1 cells (Durocher et al, 2002). Protein expression and purification was carried out as described in Chapter 3 Materials and Methods, section 3.5.2

4.5.4 In vitro splicing assays

In vitro transcription and *in vitro* splicing assays were carried out as described in Chapter 2 Materials and Methods, section 2.5.3.

4.6 Acknowledgements

We are grateful to Dr. Andrew Neuwald for his CHAIN analysis to identify residues for mutation in SF2/ASF, and to Dr. Rui-Ming Xu for modeling the amino acids identified by CHAIN analysis onto the UP1 crystal structure. We would also like to thank Jeanne Wiggins, Carmelita Bautista, and Margaret Falkowski for assistance with HeLa suspension cell culture and cell culture media, Phil Smith for the pTT3-SF2His plasmid and the 293E transfection protocol, and Zuo Zhang for the pTT3-SF2ΔRSCHis plasmid and the 293E ΔRS purification protocol.

4.7 References

Auweter, S. D., Oberstrass, F. C., Allain, F. H. (2006). Sequence-specific binding of single-stranded RNA: is there a code for recognition? Nucleic Acids Res *34*, 4943-59.

Birney, E., Kumar, S., Krainer, A. R. (1993). Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. Nucleic Acids Res *21*, 5803-5816.

Cáceres, J., Krainer, A.R. (1993). Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. EMBO J *12*, 4715-4726.

Caceres, J. F., Misteli, T., Screaton, G. R., Spector, D. L., Krainer, A. R. (1997). Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. J Cell Biol *138*, 225-38.

Cartegni, L., Chew, S L., Krainer, A. R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet *3*, 285-298.

Cartegni, L., Krainer, A. R. (2003). Correction of disease-associated exon skipping by synthetic exon-specific activators. Nat Struct Biol *10*, 120-125.

Chiodi, I., Corioni, M., Giordano, M., Valgardsdottir, R., Ghigna, C., Cobianchi, F., Xu, R. M., Riva, S. & Biamonti, G. (2004). RNA recognition motif 2 directs the recruitment of SF2/ASF to nuclear stress bodies. Nucleic Acids Res *32*, 4127–4136.

Dauksite, V., Akusjarvi, G. (2002). Human splicing factor ASF/SF2 encodes for a repressor domain required for its inhibitory activity on pre-mRNA splicing. J Biol Chem *277*, 12579-12586.

Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R., Xu, R.M. (1999). Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. Genes Dev *13*, 1102-1115.

Dreyfuss, G., Swanson, M. S., Pinol-Roma, S. (1988). Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. Trends Biochem Sci *13*, 86-91.

Durocher, Y., Perret, S., Kamen, A. (2002). High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. Nucleic Acids Res *30*, E9.

Elliott, D. J., Bourgeois, C. F., Klink, A., Stévenin, J., Cooke, H. J. (2000). A mammalian germ cell-specific RNA-binding protein interacts with ubiquitously expressed proteins involved in splice site selection. Proc Natl Acad Sci USA *97*, 5717-5722.

Ge, H., Si, Y., Wolffe, A. P. (1998). A novel transcriptional coactivator, p52, functionally interacts with the essential splicing factor ASF/SF2. Mol Cell 2, 751-759.

Graveley, B. R. (2000). Sorting out the complexity of SR protein functions. RNA *6*, 1997-1211.

Graveley, B. R., Hertel, K. J., Maniatis, T. (1998). A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. EMBO J *17*, 6747-6756.

Graveley, B. R., Hertel, K. J., Maniatis, T. (2001). The role of U2AF35 and U2AF65 in enhancer-dependent splicing. RNA *7*, 806-18.

Hargous, Y., Hautbergue, G. M., Tintaru, A. M., Skrisovska, L., Golovanov, A. P., Stevenin, J., Lian, L. Y., Wilson, S. A., Allain, F. H. (2006). Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. EMBO J *25*, 5126-37.

He, F., Muto, Y., Inoue, M., Kigawa, T., Shirouzu, M., Terada, T., Yokoyama, S.; Riken Structural Genomics/Proteomics Initiative (RSGI) (2005). Solution structure of RRM domain in splicing factor SF2. Brookhaven Protein Data Bank (PDB), accessible via Molecules To Go interface through the National Institutes of Health (NIH), at: http://molbio.info.nih.gov/doc/mtg/mol_r_us.html.

Kielkopf, C. L., Rodionova, N. A., Green, M. R., Burley, S. K. (2001). A novel peptide recognition mode revealed by the x-ray structure of a core $U2AF^{35}$/$U2AF^{65}$ heterodimer. Cell *106*, 595-605.

Lutzelberger, M., Gros, T., Kaufer, N F. (1999). Srp2, an SR protein family member of fission yeast: *in vivo* characterization of its modular domains. Nucleic Acids Res *27*, 2618-2626.

Mayeda, A., Oshima, Y. (1988). Short donor site sequences inserted within the intron of beta-globin pre-mRNA serve for splicing in vitro. Mol Cell Biol *8*, 4484-4491.

Mayeda, A., Screaton, G. R., Chandler, S. D., Fu, X.-D., Krainer, A. R. (1999). Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. Mol Cell Biol *19*, 1853-1863.

Neuwald, A. F. (2006). Bayesian shadows of molecular mechanisms cast in the light of evolution. Trends Biochem Sci *31*, 374-382.

Neuwald, A. F., Kannan, N., Poleksic, A., Hata, N., Liu, J. S. (2003). Ran's c-terminal, basic patch, and nucleotide exchange mechanisms in light of a canonical structure for Rab, Rho, Ras, and Ran GTPases. Genome Res *13*, 673-692.

Sanford, J. R., Ellis, J. D., Cazalla, D., Caceres, J. F. (2005). Reversible phosphorylation differentially affects nuclear and cytoplasmic functions of splicing factor 2/alternative splicing factor. Proc Natl Acad Sci USA *18,* 15042-15047.

Selenko, P., Gregorovic G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A., Sattler, M. (2003). Structural basis for the molecular recognition between human splicing factors $U2AF^{65}$ and SF1/mBBP. Mol Cell *11*, 965-976.

Shen, H., Green, M. R. (2004). A pathway of sequential arginine-serine rich domain-splicing signal interactions during mammalian spliceosome assembly. Mol Cell *16*, 363-373.

Xu, R.-M., Jokhan, L., Cheng, X., Mayeda, A., Krainer, A. R. (1997). Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. Structure, *5*, 559–570.

Zuo, P., Manley, J. L. (1993). Functional domains of the human splicing factor ASF/SF2. EMBO J *12*, 4727-4737.

4.8 Figures

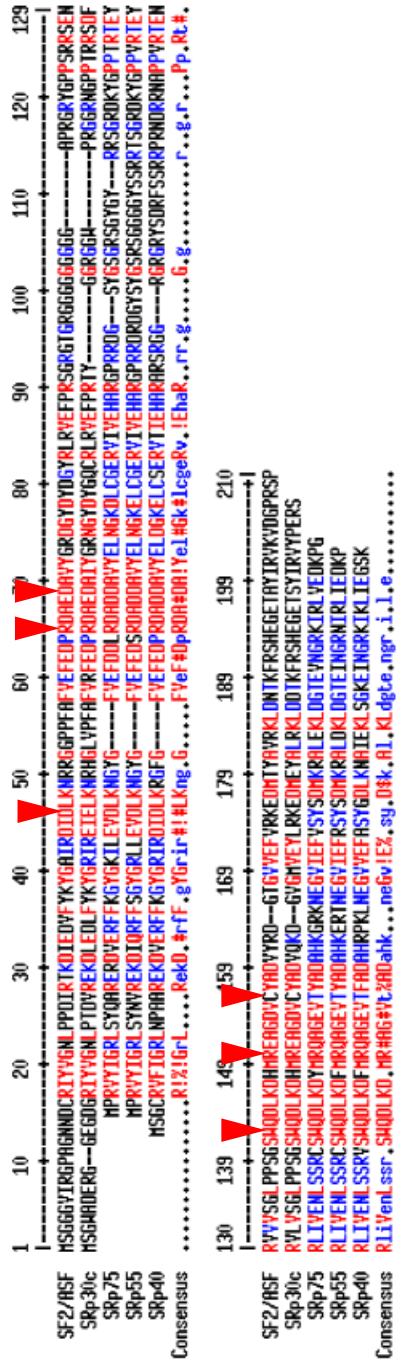


Figure 22. Residues identified by CHAIN analysis of SF2/ASF. An alignment of amino acid sequences of human SR proteins with two RRM domains is shown, truncated prior to the RS domain. Residues identified by CHAIN analysis for SF2/ASF--D46, R65, and D69 within RRM1, and W134, R142, and C148 within RRM2--are marked with arrowheads.

Figure 23. In vitro splicing of the tat23 substrate in S100 complementation with RRM1 domain mutants of SF2/ASF. The activity of SF2/ASF RRM1 domain CHAIN mutant proteins for splicing of tat23 was assessed by titration of increasing amounts (4, 8, 12, and 16 pmol) of SR protein, as compared to the same quantities of SF2/ASF. The positions of the T2T3 mRNA and tat23 lariat products ("Ti") are indicated on the left side of the gel.

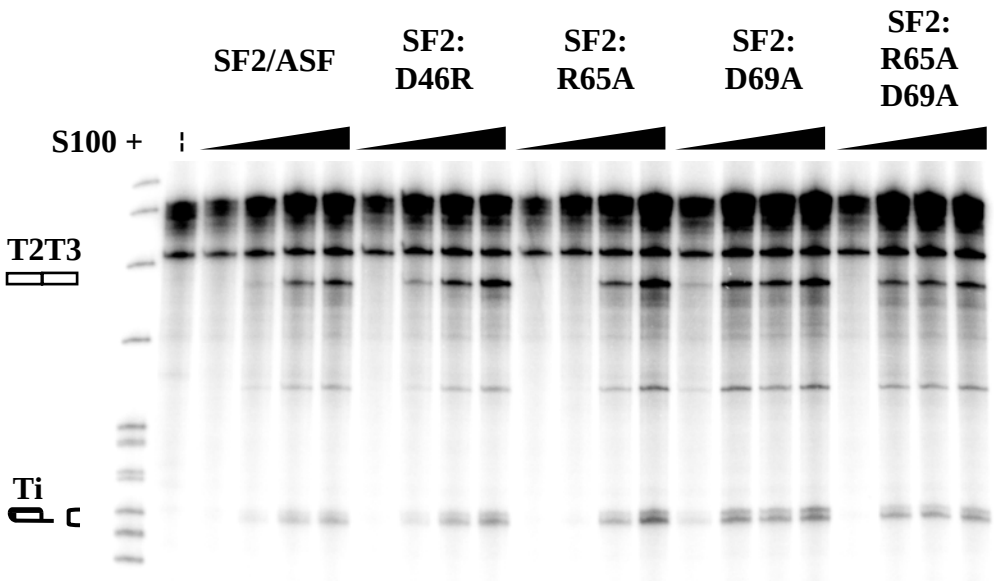


Figure 24. In vitro splicing of the tat23 substrate in S100 complementation with RRM2 domain mutants of SF2/ASF. The activity of SF2/ASF RRM2 domain CHAIN mutant proteins for splicing of tat23 was assessed by titration of increasing amounts (4, 8, 12, and 16 pmol) of SR protein, as compared to the same quantities of SF2/ASF. The positions of the T2T3 mRNA and tat23 lariat products ("Ti") are indicated on the left side of the gel.

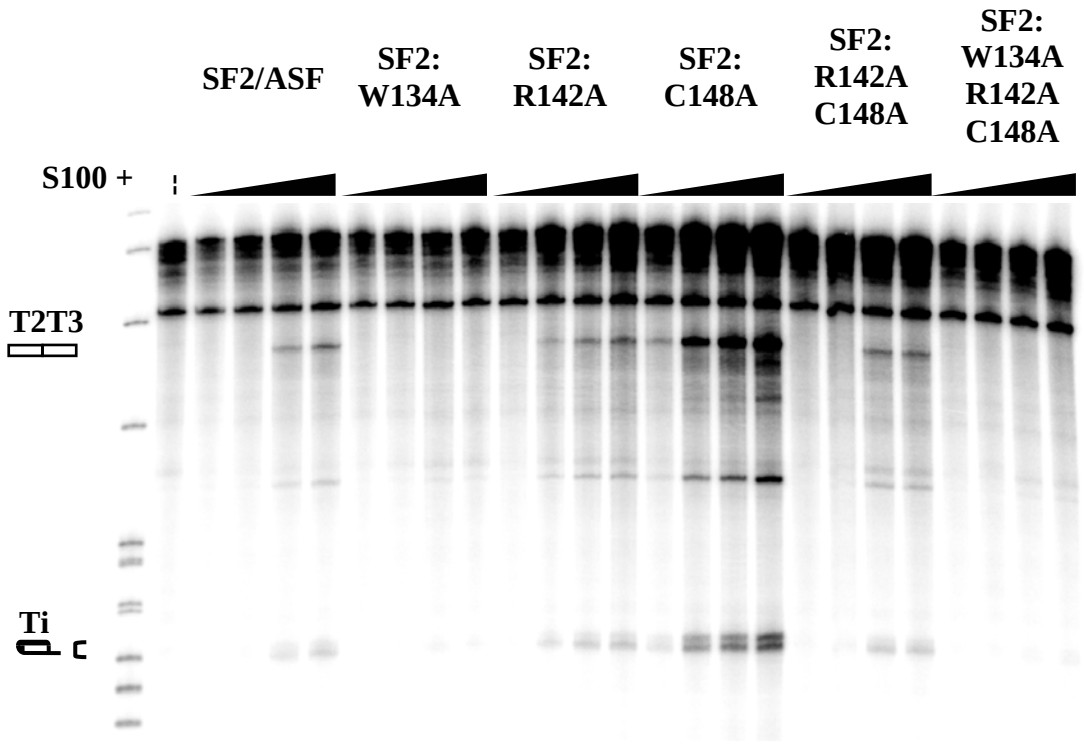
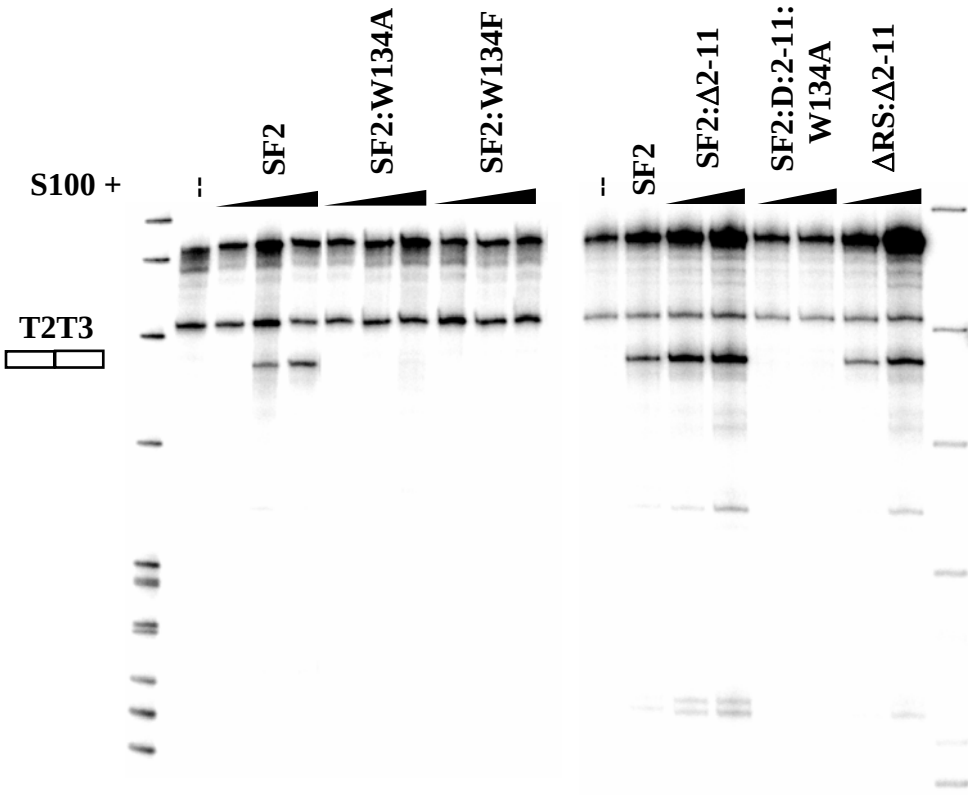


Figure 25. In vitro splicing of the tat23 substrate in S100 complementation with W134 mutants of SF2/ASF and SF2/ASF lacking its N-terminus. The activity of SF2/ASF W134 mutant proteins for splicing of tat23 was assessed by titration of increasing amounts of SR protein. The position of the T2T3 mRNA products is indicated on the left side of the gels.
Left: 8, 12, and 16pmol of the SF2:W134A and SF2:W134F mutant proteins were tested for their ability to complement S100 for splicing of tat23, as compared to the same quantities of SF2/ASF.
Right: 12 and 16pmol of the SF2:Δ2-11, SF2:Δ2-11:W134A, and ΔRS:Δ2-11 were tested for their ability to complement S100 for splicing of tat23, as compared to 16 pmol of SF2/ASF.



Chapter 5
Summary and Perspective

5.1 Reevaluation of the function of SR protein domains in pre-mRNA splicing

5.1.1 The RRM domains and recruitment and antagonism functions of SF2/ASF

The mechanisms by which ESE-bound SR proteins promote splicing have been controversial, with some data indicating that they are needed for recruitment of U2AF, and other data supporting the hypothesis that antagonism of silencers is their primary function (Graveley, 2000). Splicing in the absence of an SR protein RS domain suggests two possibilities for how the domains of SR proteins function to promote pre-mRNA splicing from the ESE position. First, the RS domain may indeed be required for recruitment functions of SR proteins, and in the absence of the RS domain the ΔRS portion of the protein promotes splicing merely by antagonizing the function of splicing silencers. Second, the ΔRS portion of the protein may be accomplishing recruitment of splicing factors, presumably via protein-protein interactions.

The RS domain is not always needed by SR proteins to counteract splicing silencers, as an SR protein lacking its RS domain can function from the position of the tat23 ESE to antagonize the exon 3 ESS (Zhu et al, 2001). However, the same ΔRS protein can also promote the recruitment of U2AF65 to the pyrimidine tract of a closely related tat substrate in an ESE-dependent manner, and this recruitment is antagonized by ESS3 (Tange and Kjems, 2001), suggesting that for the tat substrate recruitment and antagonism are occurring simultaneously, and that neither function requires an RS domain at the ESE position.

In light of our finding that the RS domain is not required for splicing *in vitro*, we suggest that the ΔRS portion of SF2/ASF, which consists of two RRM domains separated

by a glycine-rich linker, is likely functioning the same way in other pre-mRNA contexts such as IgM M1-M2 to promote splicing through both recruitment and antagonism capacities. Using the ΔRS:Δ2-11 protein, the roles of portions of SR proteins other than their RS domains in promoting splicing can easily be explored in the near future. Other splicing factors have been demonstrated to engage in protein-protein interactions via their RRM domains (reviewed in section 3.4.2), and it is likely that SR proteins are doing so as well. The spliceosome is an extremely large ribonucleoprotein machine that assembles onto the pre-mRNA to accomplish splicing. When the antiparallel beta-sheet side of the RRM domain of an SR protein is bound to pre-mRNA, the perpendicular alpha helices on the opposite face of the protein must be exposed towards the spliceosome. Given the gargantuan size of the spliceosome relative to the SR protein, and the SR protein's proximity to the splice site by virtue of its position at the ESE, it is difficult to imagine that the alpha helices or other surfaces of this side of the RRM domains of SR proteins are not in contact with some portion of the spliceosome during the splicing reaction.

5.1.2 SR protein RS domains in RNA-protein and protein-protein interactions

We have confirmed the previous finding that the RS domain of SF2/ASF is dispensable for splicing of some pre-mRNAs *in vitro* (Zhu and Krainer, 2000, Chapter 2), and have extended this finding with our discovery using the ΔRS:Δ2-11 protein that the RS domain is also not required even for splicing of substrates previously characterized as RS domain-dependent (Chapter 3). Thus, we have demonstrated that the RS domain is not essential for splicing *in vitro*, a finding which is apparently inconsistent with the prevailing models for recruitment functions of SR proteins.

In the traditional models, SR proteins are thought to recruit other essential splicing factors using their RS domains as protein-protein interaction modules to bind to the RS domains of the proteins to be recruited. Much support has been garnered for the SR protein RS domain-mediated recruitment model, particularly in the past few years through experiments that employ MS2-RS domain fusion proteins as splicing activators (Graveley and Maniatis, 1998, Graveley et al, 1998, Graveley et al, 2001, Shen et al, 2004b, etc.). Although it is quite clear that recruitment of an RS domain by any means, whether through an authentic SR protein, an MS2 fusion protein, or even an antisense oligonucleotide (Cartegni and Krainer, 2003), to the position of the ESE can function to promote splicing and likely does so at least in part through influencing the recruitment of other splicing factors (Graveley et al, 2001), such experiments do not demonstrate that the RS domain of an SR protein is either sufficient or essential for splicing. Indeed, in addition to showing that an RS domain can function as a splicing activator when tethered to the ESE position, these types of experiments have also revealed that an MS2-RS fusion protein is insufficient to activate splicing in S100 complementation, and that an SR protein must be added with the MS2-RS fusion protein for splicing to occur (Graveley and Maniatis, 1998).

It is obvious that the RS domains of SR proteins can activate splicing, yet our data strongly demonstrates that the SR protein RS domain is not required for an SR protein to activate splicing. This apparent paradox can be resolved if we consider that the function of an SR protein may not be to recruit other splicing factors through its RS domain, but rather simply to recruit an RS domain, whether its own or the RS domain of another splicing factor. In our recruitment model, the SR protein would interact with another

essential splicing protein, and this interaction must recruit at least one RS domain for splicing to be activated. An SR protein lacking its RS domain could activate splicing by interacting for example with U2AF35, which itself has an RS domain, or U1-70K, which also has its own RS domain. We note that for the essential splicing factors U2AF65 and U2AF35, which interact with each other and both have RS domains, only one of the two RS domains is required for splicing to occur (Zamore et al, 1992, Rudner et al, 1998a, Rudner et al, 1998b).

If only one RS domain may be required for a heterodimer interaction that promotes splicing, why then do many splicing factors that interact with each other all have RS domains? It is possible that the presence of two RS domains in a heterodimer activates splicing more strongly, or provides redundancy in case of steric interference with one of the RS domains. However, another reason could be simply that in addition to promoting splicing, the RS domain could regulate splicing factor binding to and/or release from RNA. RS domains are extensively phosphorylated and this phosphorylation has been shown to influence SR protein-RNA interactions. We also know that SR proteins are subjected to sequential dephosphorylation and phosphorylation during the course of splicing (Cao et al, 1997, Xiao and Manley, 1997, Xiao and Manley, 1998, Prasad et al, 1999), and that both hyper- and hypo-phosphorylated SR proteins are inactive for splicing (Kanopka et al, 1998, Prasad et al, 1999, Sanford and Bruzik, 1999). If the phosphorylation and dephosphorylation of RS domains regulates their ability to bind to RNA, it must therefore influence the ability of the SR protein to be recruited to pre-mRNA and to be released from it. While the RRM domain allows the SR protein to bind specifically to RNA, an unphosphorylated RS domain also assists the SR protein in

being nonspecifically recruited to RNA (Xiao and Manley, 1997), so that the protein can interact with other splicing factors. RS domain serine phosphorylation prevents strong sequence non-specific interactions with RNA *in vitro* (Tacke et al, 1997), so phosphorylation of the RS domain may assist the SR protein from being released from the pre-mRNA.

Through deletion of the inhibitory N-terminal segment of SF2/ASF preceding its first RRM domain, we have learned that the RS domain is not required for splicing *in vitro.* There is much precedent for N-terminal and C-terminal extensions of RRM domains regulating RNA binding (see section 3.4.3), and we suggest that the inhibitory N-terminal RRM domain extension of SF2/ASF normally functions to interfere with or downregulate RNA binding by RRM1. We propose that the RS domain normally helps to overcome this inhibition by aiding the SR protein in its binding to the pre-mRNA. Indeed, UV crosslinking with purified SR proteins shows little difference between binding of SF2/ASF and SF2:Δ2-11 to the IgM M1-M2 pre-mRNA, whereas the ΔRS protein binds less efficiently than either SF2/ASF or SF2:Δ2-11, but when the N-terminal extension is deleted from ΔRS its binding to the pre-mRNA is greatly enhanced, to a level greater than that of SF2/ASF (Chapter 3). For substrates that can be spliced with ΔRS:Δ2-11 but cannot be spliced with ΔRS ("RS domain-dependent substrates), a function of the RS domain in helping the SR protein to load onto the pre-mRNA in the context of the splicing reaction may be more critical due to inhibitory secondary structure of the RRM1 RNA target or because of proteins blocking access of the RRM domain to its target.

Our data strongly suggest that SR protein RS domain-mediated protein-protein interactions are not required for recruitment of essential splicing factors. If this is the case, then how might the RS domain activate splicing? We cannot rule out the possibility that RS domains engage in protein-protein interactions, and this would still be consistent with our hypothesis that only a single RS domain between two interacting proteins is required to activate splicing, and that this RS domain does not have to be supplied directly by the SR protein. However, an intriguing new model for how RS domains may activate splicing has recently been introduced, which would also be consistent with our data. In two studies, RS domains have been demonstrated to directly contact the pre-mRNA at the branchpoint and the 5' splice site within functional spliceosomes (Shen and Green, 2004, Shen et al, 2004a). In just the past year, these findings have been extended to show that phosphorylated RS domains can selectively contact double-stranded RNA regions and therefore possibly promote or stabilize base pairing between U snRNAs and pre-mRNA (Shen and Green, 2006).

Based on these findings and previous work about how the phosphorylation status of the RS domain regulates its interaction with RNA and the interaction of SR proteins with RNA, we propose the following model for how RS domains may activate splicing. Splicing factors with unphosphorylated RS domains are recruited non-specifically to pre-mRNA due to the phosphorylation status of their RS domains, helping the proteins assemble onto pre-mRNA so that their RNA-binding domains can engage in sequence-specific interactions with their nucleic acid targets. Interaction of the RNA-binding domains with the pre-mRNA may facilitate subsequent protein-protein interactions by altering the conformation of the proteins. RNA-bound splicing factors engage in specific

protein-protein interactions with each other, and these interactions at most require only one of the two RS domains of both factors, but may not require either RS domain, such as in the case of the U2AF heterodimer interaction. To promote the formation of E complex and to establish the positions of the splice sites, these specific protein-protein interactions either involve a protein specific to a U snRNP, or a protein that directly interacts with a U snRNP component. For example, SR proteins may interact with U1-70K, a U1-snRNP-specific protein, or with U2AF35, which indirectly recruits U2 snRNP. These specific protein-protein interactions, which may be enhanced by the presence of one or more RS domains, bring RS domains into close proximity to splicing signals where they can facilitate U snRNA binding to either the 5' splice site or the branchpoint by encouraging the formation or stabilization of double-stranded RNA as proposed by Shen and coworkers (Shen and Green, 2006). In this scenario, the phosphorylation state of the RS domain may function as a switch, where an unphosphorylated RS domain interacts with single-stranded RNA and a phosphorylated RS domain interacts with double-stranded RNA to promote RNA base-pairing.

5.2 References

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

Cartegni, L., Krainer, A. R. (2003). Correction of disease-associated exon skipping by synthetic exon-specific activators. Nat Struct Biol *10*, 120-125.

Graveley, B. R. (2000). Sorting out the complexity of SR protein functions. RNA *6*, 1997-1211.

Graveley, B. R., Maniatis, T. (1998). Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. Mol Cell *1*, 765-771.

Graveley, B. R., Hertel, K. J., Maniatis, T. (1998). A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. EMBO J *17*, 6747-6756.

Graveley, B. R., Hertel, K. J., Maniatis, T. (2001). The role of U2AF35 and U2AF65 in enhancer-dependent splicing. RNA *7*, 806-18.

Kanopka, A. Muhlemann, O., Petersen-Mahrt, S., Estmer, C., Ohrmalm, C., Akusjarvi, G. (1998). Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. Nature *393*, 185-187.

Prasad, J., Colwill, K., Pawson, T., Manley, J. L. (1999). The protein kinase Clk/Sty directly modulates SR protein activity: both hyper- and hypophosphorylation inhibit splicing. Mol Cell Biol 19, 6991-7000.

Rudner, D. Z., Breger, K. S., Rio, D. C. (1998a). Molecular genetic analysis of the heterodimeric splicing factor U2AF: the RS domain on either the large or small *Drosophila* subunit is dispensable in vivo. Genes Dev *12*, 1010-1021.

Rudner, D. Z., Breger, K. S., Kanaar, R., Adams, M. D., Rio, D. C. (1998b). RNA binding activity of heterodimeric splicing factor U2AF: at least one RS domain is required for high affinity binding. Molec Cell Biol *18*, 4004-4011.

Sanford, J. R., Bruzik, J. P. (1999). Developmental regulation of SR protein phosphorylation and activity. Genes Dev *13*, 1513-1518.

Shen, H., 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., 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., Kan, J. L., Green, M. R. (2004a). Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. Mol Cell *13*, 367-76.

Shen, H., Kan, J. L., Ghigna, C., Biamonti, G., Green, M. R. (2004b). A single polypyrimidine tract binding protein (PTB) binding site mediates splicing inhibition at mouse IgM exons M1 and M2. RNA *10*, 787-794.

Tacke, R., Chen, Y., Manley, J.L. (1997). Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. Proc Natl Acad Sci USA *94*, 1148-1153.

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

Xiao, S.-H., Manley, J. L. (1997). Phosphorylation of the ASF/AF2 RS domain affects both protein-protein interactions and is necessary for splicing. Genes Dev *11*, 334-344.

Xiao, S.-H., Manley, J. L. (1998). Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. EMBO J *17*, 6359-6367.

Zamore, P. D., Patton, J. G., Green, M. R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. Nature *355*, 609-614.

Zhu, J., Krainer, A.R. (2000). Pre-mRNA splicing in the absence of an SR protein RS domain. Genes Dev *14*, 3166-3178.

Zhu, J., Mayeda, A., Krainer, A. R. (2001). Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNPA1 and enhancer-bound SR proteins. Mol Cell *8*, 1351-1361.

List of Abbreviations

3'ss	3' splice site
5'ss	5' splice site
APC	adenomatosis polyposis coli
BLAST	basic local alignment search tool
bp	base pairs
BPS	branch point sequence
BRCA1	breast cancer 1, early onset
Brr2	Brr2p, an RNA-dependent ATPase RNA helicase
CD40	CD40 molecule, TNF receptor superfamily member 5
cDNA	complementary DNA
CTFR	cystic fibrosis transmembrane conductance regulator
DNA	deoxyribonucleic acid
ΔRS	SF2/ASF lacking its RS domain
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
EST	expressed sequence tag
FBN1	fibrillin-1
GH-1	growth hormone
GJA9	connexin 36
hnRNP	heterogeneous nuclear ribonucleoprotein
hnRNP A1	heterogeneous nuclear ribonucleoprotein A1
hnRNP A2/B1	heterogeneous nuclear ribonucleoprotein A2/B1
hnRNP C1/C2	heterogeneous nuclear ribonucleoprotein C1/C2
hnRNP H	heterogeneous nuclear ribonucleoprotein H
hnRNP I	heterogeneous nuclear ribonucleoprotein I
hnRNP L	heterogeneous nuclear ribonucleoprotein L
hnRNP K	heterogeneous nuclear ribonucleoprotein K
HPRT1	hypoxanthine phosphoribosyltransferase 1
IgM	immunoglobulin μ
ISS	intronic splicing silencer
MAPT	microtubule-associated protein tau
NE	nuclear extract
NF1	neurofibromatosis 1
NMD	nonsense-mediated decay
nt	nucleotide
pre-mRNA	precursor messenger RNA
Prp2	Prp2p, an RNA-dependent ATPase in the DEAH-box family
Prp5	Prp5p, RNA helicase in the DEAD-box family
Prp16	Prp16p, RNA helicase in the DEAH-box family
Prp17	Prp17p, also known as Cdc40p
Prp18	Prp18p
Prp28	Prp28p, RNA helicase in the DEAD-box family
PTB	polypyrimidine tract binding protein
Py	pyrimidine tract

RNA	ribonucleic acid
RNP1	ribonucleoprotein 1 consensus sequence
RNP2	ribonucleoprotein 2 consensus sequence
RRM	RNA recognition motif
RS domain	arginine serine rich domain
S100	HeLa cell cytoplasmic extract
SF1/mBBP	splicing factor 1/mammalian branchpoint bridging protein
SF2/ASF	splicing factor 2/alternative splicing factor
SF3a	splicing factor 3a
Slu7	Slu7p,
SMN1	survival of motor neuron 1
snRNP	small nuclear ribonucleoprotein
snRNA	small nuclear RNA
SRPK1	SR protein kinase 1
SRPK2	SR protein kinase 2
U1A	U1 small nuclear ribonucleoprotein protein A
U2A’	U2 small nuclear ribonucleoprotein protein A’
U2B”	U2 small nuclear ribonucleoprotein protein B”
U1-70K	U1 70 kilodalton protein
U2AF35	U2 auxiliary factor 35 kilodalton subunit
U2AF65	U2 auxiliary factor 65 kilodalton subunit
UAP56	UAP56p, RNA helicase in the DEAD box family
VWF	Von Willebrand factor