

# **Stony Brook University**



OFFICIAL COPY

**The official electronic file of this thesis or dissertation is maintained by the University Libraries on behalf of The Graduate School at Stony Brook University.**

**© All Rights Reserved by Author.**

**Physiological Role of Post-Translational Modifications of the  
Oncoprotein SF2/ASF**

A Dissertation Presented

by

**Rahul Sinha**

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

**Doctor of Philosophy**

in

**Molecular and Cellular Biology**

**(Biochemistry and Molecular Biology)**

Stony Brook University

**December 2009**

**Stony Brook University**

The Graduate School

**Rahul Sinha**

We, the dissertation committee for the above candidate for the  
Doctor of Philosophy degree, hereby recommend  
acceptance of this dissertation.

**Adrian R. Krainer**

**Professor, Cold Spring Harbor Laboratory**

**Rolf Sternglanz - Chairperson of Defense**

**Professor, Department of Biochemistry and Cell Biology**

**Linda Van Aelst**

**Professor, Cold Spring Harbor Laboratory**

**Rui-Ming Xu**

**Professor, Institute of Biophysics, Chinese Academy of Science**

This dissertation is accepted by the Graduate School

Lawrence Martin

Dean of the Graduate School

Abstract of the Dissertation

**Physiological Role of Post-Translational Modifications of the Oncoprotein  
SF2/ASF**

by

**Rahul Sinha**

**Doctor of Philosophy**

in

**Molecular and Cellular Biology  
(Biochemistry and Molecular Biology)**

Stony Brook University

**2009**

Alternative splicing and post-translational modifications (PTMs) are major sources of protein diversity in eukaryotic proteomes. The SR protein SF2/ASF is an oncoprotein that functions in pre-mRNA splicing, with additional roles in other post-transcriptional and translational events. Functional studies of SR-protein PTMs have exclusively focused on reversible phosphorylation of Ser residues in the C-terminal RS domain. Our proteomic analysis of SF2/ASF and other SR proteins obtained from mammalian cells revealed that they are extensively modified. We confirmed that human SF2/ASF is methylated at residues R93, R97, and R109, which were recently identified in a global proteomic analysis of Arg methylation, and further investigated whether these methylated residues regulate the functional properties of SF2/ASF. We show that the three arginines additively control the subcellular localization of SF2/ASF, and both the positive charge and the methylation state are important. Mutations that block methylation and remove the positive charge result in cytoplasmic accumulation of SF2/ASF. The consequent decrease in nuclear SF2/ASF levels prevents it from modulating alternative splicing of target genes, results in higher translation stimulation, and abrogates the enhancement of nonsense-mediated mRNA decay. This study addresses the mechanisms by which Arg methylation and the associated positive charge regulate the activities of SF2/ASF, and emphasizes the significance of localization control for an oncoprotein with multiple functions in different cellular compartments. Moreover, the various other types of modifications identified in this study of SF2/ASF, as well as other SR proteins, lay a foundation for future studies of the detailed mechanisms by which different PTMs may modulate the diverse functions of SR proteins.

I dedicate this work to my parents for their unconditional love and numerous sacrifices, without which, this feat would have been impossible to achieve.

# Table of Contents

List of Figures.....	ix
List of Tables.....	xii
Acknowledgments .....	xiii
Publications and Ongoing Work.....	xv
<b>Chapter 1. Background.....</b>	<b>1</b>
1.1. Introduction.....	2
1.1.1. The early beginnings.....	2
1.1.2. The basic mechanism and the splice-signals.....	2
1.1.3. The spliceosome and its assembly.....	3
a. The early spliceosome (E complex).....	4
b. The prespliceosome (A complex).....	5
c. The catalytically active spliceosome (the B* and the C complex).....	5
d. The spliceosome, a unique macromolecular machine.....	6
1.2. Splicing, a highly coordinated nuclear process.....	7
1.2.1. Splicing and transcription.....	7
1.2.2. Splicing and 5' cap formation.....	7
1.2.3. Splicing and 3'-end processing.....	8
1.3. Splicing history and its influence on downstream events.....	9
1.3.1. Splicing and mRNA export.....	9
a. The THO/TREX complex and mRNA export.....	9
b. The exon-junction complex and mRNA export.....	10
c. The SR proteins and mRNA export.....	10

1.4. Splicing and translation.....	11
1.5. Splicing and nonsense-mediated mRNA decay.....	11
1.6. Alternative splicing.....	12
1.6.1. The trans-acting regulators of alternative splicing.....	12
a. The SR proteins.....	12
i. The prototypical SR protein, SF2/ASF.....	13
b. The hnRNPs.....	14
1.6.2. Alternative splicing: the major source of protein diversity.....	14
1.6.3. Alternative splicing and disease.....	15
a. Alternative splicing and cancer.....	16
i. SF2/ASF and cancer.....	17
1.7. Concluding remarks.....	18
1.8. Figures and legends.....	20
<b>Chapter 2. Physiological Role of Post-Translational Modifications of the Oncoprotein SF2/ASF.....</b>	<b>24</b>
2.1. Introduction.....	25
2.2. Results.....	28
2.2.1. SF2/ASF is extensively modified.....	28
2.2.2. The methyl-arginines affect the nucleo-cytoplasmic distribution of SF2/ASF.....	30
2.2.3. The triple-Ala mutant, A <sub>1</sub> A <sub>2</sub> A <sub>3</sub> , fails to enhance NMD.....	30
2.2.4. Reduced activity of the A <sub>1</sub> A <sub>2</sub> A <sub>3</sub> mutant in general and alternative splicing in vitro.....	31
2.2.5. The A <sub>1</sub> A <sub>2</sub> A <sub>3</sub> mutant fails to modulate splicing of endogenous pre- mRNAs.....	32
2.2.6. The A <sub>1</sub> A <sub>2</sub> A <sub>3</sub> protein is more active than wt-SF2/ASF in enhancing translation in cells.....	32

2.2.7.	A nuclear-retained version of A <sub>1</sub> A <sub>2</sub> A <sub>3</sub> protein promotes NMD.....	33
2.2.8.	Tandem mass spectrometry reveals that the triple Lys mutant is dimethylated.....	33
2.2.9.	Mutation of the Gly residues in the RGG motifs restores SF2/ASF localization and function.....	34
2.2.10.	Interaction of SF2/ASF with import/export machineries is not disrupted by methylation or charge.....	35
a.	Interaction with TAP/NXF1.....	35
b.	Interaction with transportin-SR2.....	36
2.3.	Discussion.....	37
2.4.	Future Perspectives.....	42
2.5.	Materials and Methods.....	43
2.5.1.	Plasmids.....	43
2.5.2.	Expression and purification of recombinant proteins.....	44
2.5.3.	Cell culture, and transient and stable expression of proteins.....	45
2.5.4.	Indirect immunofluorescence (IF).....	46
2.5.5.	Shuttling assays.....	46
2.5.6.	Dual luciferase assay and cellular fractionation.....	47
2.5.7.	In vitro translation.....	47
2.5.8.	In vitro splicing.....	47
2.5.9.	RNase protection assay (RPA).....	48
2.5.10.	RT-PCR.....	48
2.5.11.	Antibodies and Western Blotting.....	48
2.5.12.	Mass spectrometry.....	48
2.5.13.	Immunoprecipitation and GST pull-down.....	49
2.6.	Acknowledgements.....	49



2.7. Tables.....	50
2.8. Figures and Figure legends.....	53
<b>Chapter 3. Antisense oligos: towards a cure for ‘life without pain or tears’...91</b>	
3.1. Introduction.....	92
3.2. Results.....	99
3.2.1. MOE-ASO walk reveals several enhancers and silencer regions in <i>IKBKAP</i> .....	99
3.2.2. Oligo-walk with an NMD-responsive minigene.....	100
3.2.3. High resolution micro-walk in the ISS-40 region.....	101
3.2.4. Effect of oligo 421992 in FD-derived fibroblasts.....	102
3.2.5. ISS-40 harbors at least two silencer elements.....	102
3.3. Discussion and future perspectives.....	103
3.4. Materials and methods.....	106
3.4.1. Oligonucleotide synthesis.....	106
3.4.2. Plasmids.....	106
3.4.3. Cell culture and transfections.....	106
3.4.4. RT-PCR.....	107
3.5. Acknowledgements.....	107
3.6. Tables.....	108
3.7. Figures and Legends.....	113
<b>References.....</b>	<b>143</b>
Chapter 1.....	143
Chapter 2.....	156
Chapter 3.....	164

## List of Figures

<b>Chapter 1. Background.....</b>	<b>1</b>
Figure 1.1. Consensus splice-site sequences in mammals.....	21
Figure 1.2. Diverse functions of SF2/ASF.....	23
<b>Chapter 2. Physiological Role of Post-Translational Modifications of the Oncoprotein SF2/ASF.....</b>	<b>24</b>
Figure 2.1. Mass-spec analysis reveals that SF2/ASF is extensively modified at multiple residues.....	54
Figure 2.2. Types of arginine methylation in mammals.....	56
Figure 2.3. Localization and shuttling activity of wild-type and mutant SF2/ASF.....	58
Figure 2.4. Localization of single or double Ala mutants of R93, R97, and R109 of SF2/ASF.....	60
Figure 2.5. Effect of overexpression of wild-type or mutant SF2/ASF on NMD.....	62
Figure 2.6. Effect of overexpression of wild-type or mutant SF2/ASF on alternative 5'ss selection and NMD.....	64
Figure 2.7. Effect of single- or double-Ala mutants of SF2/ASF on alternative 5'ss selection and sensitivity to a PTC upon overexpression...	66
Figure 2.8. In vitro splicing activity of wild-type and mutant SF2/ASF.....	68
Figure 2.9. In vitro splicing activity of unmodified wild-type or mutant SF2/ASF.....	70
Figure 2.10. Effects of wild-type or mutant SF2/ASF on alternative splicing of endogenous transcripts in IMR90 cells.....	72
Figure 2.11. Effect of wild-type or mutant SF2/ASF on ESE-dependent stimulation of translation.....	74
Figure 2.12. Enhancement of NMD by SF2/ASF or mutants fused to a nuclear-retention signal.....	76
Figure 2.13. Tandem mass spectrometry of the K <sub>1</sub> K <sub>2</sub> K <sub>3</sub> protein.....	78

Figure 2.14. Schematic representations of the motif mutations introduced to create SM and KM proteins.....	80
Figure 2.15. Subcellular localization of SM and KM mutants.....	82
Figure 2.16. Effect of overexpression of SM and KM mutants on NMD.....	84
Figure 2.17. Interactions between TAP/NXF1 and SF2/ASF.....	86
Figure 2.18. Unmodified SF2/ASF interacts with transportin-SR2.....	88
Figure 2.19. Dephosphorylated SF2/ASF interacts with transportin-SR2...	90
<b>Chapter 3. Antisense oligos: towards a cure for ‘life without pain or tears’.....</b>	<b>91</b>
Figure 3.1. Genomic structure of <i>IKBKAP</i> gene.....	114
Figure 3.2. The Major FD mutation causes predominant exon-20 skipping in <i>IKBKAP</i> .....	116
Figure 3.3. Schematic representation of the minigene constructs.....	118
Figure 3.4. Splicing patterns of expressed pre-mRNAs from the minigenes.....	120
Figure 3.5. The MOE-ASO structures.....	122
Figure 3.6. Coarse oligo-walk for upstream intron-19 using 19-21 minigene.....	124
Figure 3.7. Coarse oligo-walk for exon-20 using 19-21 minigenes.....	126
Figure 3.8. Coarse oligo-walk for downstream intron-20 using 19-21 minigenes.....	128
Figure 3.9. Coarse oligo-walk for upstream intron-19 using 19-22 minigenes.....	130
Figure 3.10. Coarse oligo-walk for exon-20 using 19-22 minigenes.....	132
Figure 3.11. Coarse oligo-walk for downstream intron-20 using 19-22 minigenes.....	134
Figure 3.12. High resolution microwalk for the ISS-40 region using 19-21 minigenes.....	136
Figure 3.13. Effect of best candidate oligo in patient fibroblasts and dose response.....	138

Figure 3.14. Mutation analysis of the first half of ISS-40.....	140
Figure 3.15. Mutation analysis of the second half of ISS-40.....	142

## List of Tables

Table 2.8.1	Modified residues identified in 9g8.....	50
Table 2.8.2	Modified residues identified in SRp20.....	50
Table 2.8.2	Modified residues identified in SC35.....	51
Table 2.8.2	Primers used for site-directed mutagenesis.....	51
Table 3.7.1	List of pO-MOE-ASOs.....	108
Table 3.7.2	List of PCR primers used to amplify <i>IKBKAP</i> genomic fragments.....	109
Table 3.7.3	List of primers used for RT-PCR.....	110
Table 3.7.3.a	Primers used to amplify minigene-specific mRNA.....	110
Table 3.7.1.b	Primers used to amplify endogenous <i>IKBKAP</i> -derived mRNA.....	110
Table 3.7.4	List of primers used for site directed mutagenesis.....	110

## Acknowledgements

First and foremost, I thank my advisor and mentor Adrian Krainer who has always been a tremendous source of encouragement, guidance, and support since the very beginning. He showed great confidence in me at every juncture during my career as a graduate student, especially the times when I was not sure about myself. In the long list of invaluable things that I had the opportunity to learn from him, his influence in the development of my whole outlook towards contemporary science is most noteworthy, as he taught me to be critical yet constructive at the same time. I am confident that the expertise that I have developed in the field of pre-mRNA splicing and the training that I have received under his guidance will have a major role to play in realizing my goal towards becoming an independent scientist.

I thank past and present members of the Krainer lab: Eric Allemand, Zuo Zhang, Rotem Karni, Phillip Smith, Michelle Hastings, Qing-Shuo Zhang, Stephanie Shaw, Hazeem Okunola, Lisa Manche, Xavier Roca, Yimin Hua, Shuying Sun, Mads Aaboe Jensen, Ying Hsiu Liu, Oliver Fregoso, Olga Anczukow-Camarda, Isabel Aznarez, Ruei-Ying Tzeng, Deblina Chatterjee, Shipra Das, Zhenxun Wang, Kentaro Sahashi and Martin Akerman for helpful discussions, as well as for creating an extremely comfortable and intellectually stimulating environment at the lab. I express special gratitude towards Eric Allemand, Zuo Zhang, Rotem Karni, Yimin Hua, Xavier Roca and Phillip Smith, who devoted their precious time and effort to teach and subsequently help me perfect my technical skills and knowhow. I feel fortunate to have their continued support. I also extend special thanks to Lisa Manche for critical reading of my dissertation drafts.

I am extremely grateful to the members of my present thesis committee, Rolf Sternglanz and Rui-Ming Xu, as well as the past members, Michael Myers and Senthil Muthuswamy, for their valuable insights and suggestions, which not only proved crucial as the project evolved, but also kept me focused on the right track of research during the entire course of the project. I thank Linda Van Aelst who has always extended her moral support over the years during my stay at Cold Spring Harbor Lab and generously agreed to serve on my thesis committee on a short notice, when the committee needed to be revised. Finally, I thank all the committee members for monitoring my progress as a graduate student and critical reading of my thesis proposal and dissertation.

I extend my gratitude towards the admissions committee of the MCB Graduate Program for accepting me as a student, and the Program support staff who was always there to help with the day to day affairs and requirements of the graduate school. I would also like to thank the support staff and advisors at the

International Services at Stony Brook for making my stay trouble-free in the United States as an international graduate student.

I am indebted by the love and care extended by my friends and classmates at Stony Brook, especially during my first year of stay, without which life on Long Island would have been a lot more difficult. I also thank the closely knit community of friends and colleagues at Cold Spring Harbor Laboratory, with whom I interacted scientifically and socially, for their generous help and support extended towards me over the years.

I thank my host family, Dan, Christine and Jonathan Rauch for always being there when I needed a boost in my spirits and for being my family away from family.

I express my special thanks to Justyna Janas, who has always stood by me and has benefitted me both scientifically as well as in other aspects of life with her numerous contributions.

Finally I thank my family, my mother and father and my two younger brothers Gautam and Kunal Sinha, whose love and inspiration have been instrumental in each and every one of my exploits.

## Publications and Ongoing Work

Sinha, R., Allemand, E., Zhang, Z., Karni, R., Myers, M.P. and Krainer, A.R. (2010). Arginine methylation controls the subcellular localization and functions of the oncoprotein SF2/ASF. *Mol Cell Biol*. *Provisionally accepted pending minor changes*.

The major portion of Chapter 2 is comprised of the above work to be published in *Molecular and Cellular Biology*. The mass spectrometry experiments were carried out by Michael P. Myers. The immunofluorescence and the heterokaryon assay were performed in part by Eric Allemand. Zuo Zhang and Rotem Karni provided valuable reagents. Adrian Krainer and I wrote the manuscript together.

Chapter 3 describes ongoing work and the advances made towards finding a therapy for Familial Dysautonomia. Yimin Hua provided generous help and advice during the initial design of the study. The future experiments mentioned in the text, involving transgenic FD mice, are currently being conducted primarily by Kentaro Sahashi. The work is part of a collaboration with ISIS Pharmaceuticals, where the antisense oligonucleotides are synthesized.



# **Chapter 1:**

## **Background**

## **1.1 Pre-mRNA splicing and its role in gene expression**

### **1.1.1 The early beginnings**

Studies of pre-mRNA splicing began with the discovery of split genes in adenovirus by two independent groups, headed by Richard Roberts and Phillip Sharp (Berget et al., 1977; Chow et al., 1977). Soon thereafter, increasing amounts of evidence made it clear that the majority of genes in higher eukaryotes are also split such that the coding sequences known as exons are interrupted by one, or most often by more than one, non-coding sequence called an intron. It was not long before it became evident that pre-mRNA splicing is a required step for the expression of the vast majority of human genes that involves transcription of a primary transcript (pre-mRNA) by RNA polymerase II. The transcribed pre-mRNA, which spans the whole region containing both exons and introns, is then processed such that the introns are removed and the exons are all joined together in the correct order. One other mechanism of expression of split genes, most notably the involvement of DNA-rearrangement during the expression of immunoglobulin heavy and light chains, was soon found to be an exception rather than the norm, and it was established that pre-mRNA splicing, in addition to 5'-capping and 3'-end cleavage and polyadenylation, is one of the crucial RNA processing events in eukaryotes leading to translation (Crick, 1979).

### **1.1.2 The basic mechanism and the splice-signals**

Today, through an analysis of all the annotated human genes, we know that the transcribed genome consists of approximately 230,000 exons and 210,000 introns, or roughly nine exons and eight introns per gene on average (Sakharkar et al., 2004). Splicing of the primary transcript, which involves precise excision of the introns and joining of the flanking exons as a result of two sequential transesterification reactions, is carried out in the nucleus (Burge et al., 1999; Wachtel and Manley, 2009).

In the first transesterification reaction, the 2'OH of a specific branchpoint adenosine nucleotide within the intron attacks the phosphoester linkage between the 3'OH of the last nucleotide of the first exon and the 5' phosphate of the first nucleotide of the intron. This creates the 2'-5' phosphodiester linkage between the branchpoint nucleotide and the first nucleotide of the intron, forming an intron lariat. Subsequently, the free 3'OH of the last nucleotide of the first exon attacks the phosphoester linkage between the 3'OH of last nucleotide of the intron and the 5' phosphate of the first nucleotide of the second exon, joining the exons by a regular 3'-5' phosphodiester linkage and releasing the branched intron lariat, thus completing the second transesterification.

The branchpoint nucleotide, as well as the first and last nucleotides of the introns, are part of degenerate consensus sequence motifs called the branchpoint sequence (BPS), the 5' splice site (5'ss), and the 3' splice site (3'ss), respectively (**Fig. 1.1**). The BPS is found usually 18-40 nucleotides upstream of the 3'ss. The stretch of sequence between the BPS and 3'ss is rich in pyrimidines and is known as the polypyrimidine tract (Py-tract). These cis-acting sequence elements show a low degree of conservation and are necessary, but not sufficient for splicing. Similar sequence elements are found within introns and exons, known as pseudo splice-sites or cryptic splice sites, which, taken together, almost always outnumber the real splice-signals, especially in long transcripts (Senapathy et al., 1990; Sun and Chasin, 2000). Pseudo splice sites are never used, whereas a cryptic splice site may become active when one of the real splice sites, e.g., in close vicinity, is inactivated by genetic mutation(s) (Roca et al., 2003). This clearly indicates that there are additional signals present within the pre-mRNA that help distinguish the real splice site from the pseudo and cryptic splice sites, and define the precise boundaries of exons and introns (see below).

### **1.1.3 The spliceosome and its assembly**

Recognition of these cis-acting sequence motifs (5'ss, BPS, Py-tract, and 3'ss) and catalysis of the transesterification reactions is carried out by the spliceosome, which is a complex macromolecular machine consisting of five small RNAs and more than 100 proteins (Black, 2003; Jurica and Moore, 2003). There are two different kinds of spliceosomes present in some plants and metazoans, including humans: the major and the minor spliceosome. The core components of the major spliceosome are the U1, U2, U4/U6, and U5 small nuclear ribonucleoproteins (snRNPs), named after the individual U-rich small nuclear RNAs (U1, U2, U4/U6, and U5 snRNAs) they carry, respectively. The minor spliceosome, which is responsible for splicing of a very small fraction of total introns in humans, is composed of distinct, but functionally analogous U11, U12, and U4atac/U6atac snRNPs, with U5 being common between both spliceosomes (Patel and Steitz, 2003). snRNPs, except U6 and U6atac, are associated with a heteroheptamer Sm ring at its core, which is comprised of SmB, SmD1, SmD2, SmD3, SmE, SmF, and SmG proteins. An alternative splice variant of SmB is the SmB' protein that can replace SmB. Additionally, a different SmN protein replaces SmB/B' in neuronal tissues. In contrast, a heteroheptameric ring composed of seven Lsm proteins (Sm like), Lsm2-8, forms the core of U6 and U6atac snRNPs (Beggs, 2005; Kiss, 2004). In addition to the core proteins, each snRNP harbors 3-15 factors that are specific to individual snRNPs (Matlin and Moore, 2007).

### a. The early spliceosome (E complex)

The spliceosome assembles on the pre-mRNA in a stepwise fashion, which involves sequential association and release of different snRNPs at specific times, finally leading to its activation and catalysis. Various assembly intermediates, which are defined according to their snRNP composition, exist transiently during the assembly process and help form the catalytically active site (Brow, 2002; Burge et al., 1999; Will and Lührmann, 2006). Assembly starts with the ATP-independent binding of U1 snRNP to the 5'ss, driven by base pairing between the 5' end of U1 snRNA and the 5'ss (**Fig.1**). This coincides with the binding of Splicing Factor 1/Branchpoint Binding Protein (SF1/BBP) to the BPS, and the binding of U2 Auxiliary Factor (U2AF) to both the Py-tract and the 3'ss, at the other end of the intron. U2AF consists of two subunits, U2AF65 and U2AF35, which are almost always found associated with each other. U2AF65 recognizes and binds to the Py-tract, whereas U2AF35 binds to the conserved AG dinucleotide within the 3'ss. The recognition of the BPS, the Py-tract, and the 3'ss by SF1/BBP and U2AF is highly cooperative, with the two subcomplexes helping recruit each other to the cis-elements. This is facilitated by the interaction of U2AF65 with SF1 through its carboxy-terminal (C-terminal) RNA recognition motif-3 (RRM3) domain. All the subsequent steps (rearrangements within the spliceosome) after the formation of the E complex are driven by ATP hydrolysis, although the transesterification chemistry per se does not require energy from ATP.

Formation of the E complex is arguably the most important step in spliceosome assembly, as it selectively recognizes the real 5'ss and the 3'ss among the sea of similar pseudo- and cryptic splice sites present on the pre-mRNA. One way this could happen is through a mechanism known as 'exon definition' (Berget, 1995). Given the fact that exons in mammals, on average, are 20 times shorter in length than introns, there is some evidence suggesting that exons, rather than introns, are detected by the spliceosome via cross-talk between the factors that occupy its two ends, the upstream 3'ss (SF1/BBP and U2AF) and the downstream 5'ss (U1 snRNP). However, by the same token, pseudo-exons can also be defined using pseudo 5'- and 3'ss, and just like pseudo splice sites, the pseudo-exons outnumber real exons by an order of magnitude (Senapathy et al., 1990; Sun and Chasin, 2000). The additional regulatory cis-acting markers on pre-mRNA that are used in the cell to distinguish real exons from pseudo exons are known as exonic or intronic splicing enhancers (ESE, ISE) and silencers (ESS, ISS), which together, with the real splice-signals (5'ss, BPS, Py-tract and 3'ss) constitute the so-called 'splicing code' (Fu, 2004). Both enhancers and silencers are 6-8 nucleotide long

degenerate motifs that are classified simply according to their effect and location. The trans-acting factors, which recognize these signals, are RNA-binding proteins, exemplified by two classes of proteins: Ser/Arg-rich (SR) proteins and heterogeneous ribonucleoproteins (hnRNP), which generally recognize ESSs and ISSs, respectively (see below).

Most of the crucial RNA-RNA interactions involved in the recognition of cis-elements by the spliceosome are weak in nature and believed to be aided by the SR proteins, most probably by direct interaction with various components of the spliceosome. Though precise mechanisms remain unclear, currently there are two models that could explain the role of SR proteins in exon inclusion, and there is evidence, which indicates that these mechanisms may not be mutually exclusive. One possibility is that the ESE/ISE-bound SR proteins may recruit and stabilize interactions between the U1 snRNP at the 5'ss and U2AF65 at the 3'ss, thereby leading to exon definition (Graveley et al., 2001). According to another model, the SR proteins bound to ESEs/ISEs may aid in exon inclusion by antagonizing the negative effects of hnRNPs, which recognize the silencers (ESS/ISS) (Zhu et al., 2001). Nevertheless, both the models are supported by the observations that SR proteins are found associated with the spliceosome throughout the entire assembly process and with each of the assembly intermediates, starting with the E complex (see below) (Behzadnia et al., 2007; Bessonov et al., 2008; Deckert et al., 2006).

#### **b. The prespliceosome (A complex)**

E complex is followed by the formation of the pre-spliceosome or A complex, which involves the ATP-dependent association of U2 snRNP at the branchpoint region through base pairing interactions between U2 snRNA and the BPS (**Fig. 1**). This base-pairing interaction is stabilized by the U2 snRNP associated factors, SF3a and SF3b (Gozani et al., 1996), as well as by the arginine-serine-rich (RS) domain of U2AF65 (Valcarcel et al., 1996). The engagement of U2 at the BPS leads to the removal of SF1/BBP and its replacement by SF3b14a/p14, which now interacts with the branchpoint adenosine (Will et al., 2001). The new complex is stabilized through interactions between SF3b155 and the C-terminal RRM3 of U2AF65 (Gozani et al., 1998).

#### **c. The catalytically active spliceosome (the B\* and the C complex)**

After the formation of the A complex, a preassembled tri-snRNP, U4/U6.U5, comprised of U4/U6 and U5 snRNPs is recruited to form the B complex. This catalytically inactive form of spliceosome undergoes major conformational changes that destabilize the U1 and U4 snRNPs, which are then

removed, causing profound rearrangements within the spliceosome. These rearrangements result in the activated spliceosome, also known as the B\* complex that catalyzes the first transesterification reaction. Subsequently, further rearrangements in the spliceosome give rise to the C complex (Konarska et al., 2006), which now catalyzes the second and final transesterification step. Additional sets of rearrangements are required at this stage to ensure the release of mature mRNA and recycling of the spliceosome.

#### **d. The spliceosome, a unique macromolecular machine**

The spliceosome is distinct from other macromolecular machines in the cell in two major ways. First, it does not have a preformed active site for catalysis, and second, it exhibits a high degree of compositional plasticity correlating with its function. This dynamic nature of the spliceosome not only allows for extensive regulation at various steps, but also ensures high fidelity and precision during catalysis. The latter is achieved through an elaborate sequential process in which the reactive groups of the pre-mRNA are recognized multiple times by different RNA or protein components of the spliceosome. Another hallmark of spliceosomal mechanics is the presence of various weak, but crucial binary RNA-RNA, RNA-protein, and protein-protein interactions, which are stabilized by a combination of other interactions, thus providing the flexibility to faithfully recognize a wide variety of highly degenerate and loosely related splice signals (Luhrmann and Stark, 2009). Finally, the similarity in the overall biochemical mechanism of transesterification between self-splicing group-II introns in eubacteria and nuclear spliceosomal introns in eukaryotes had led to earlier beliefs that the spliceosome is a ribozyme. In fact, the protein components of the spliceosome, in addition to providing structural integrity, directly cooperate with their RNA counterparts to form the active site and may even take part in catalysis, thus forming a complex ribonucleoprotein (RNP) enzyme (Abelson, 2008).

The dynamic plasticity observed in the case of the spliceosome is unprecedented among other macromolecular machines found in the cell. Multiple sets of spliceosome-associated protein factors provide the driving force necessary for the large scale remodeling of RNA-RNA, RNA-protein, and protein-protein interactions at key steps during spliceosome assembly. Most notable among these are enzymes such as DExD/H-type RNA-dependent ATPases/helicases (Staley and Guthrie, 1998) and peptidyl-prolyl cis/trans isomerases (PPIases) (Nagradova, 2007), which function in conjunction with various enzymes (kinases and phosphatases) responsible for posttranslational modifications on spliceosomal components (Wahl et al., 2009).

## **1.2 Splicing, a highly coordinated nuclear process**

Splicing is a coordinated nuclear process that takes place in conjunction with other RNA-processing events, specifically 5' cap formation and 3' cleavage/polyadenylation, all of which can occur cotranscriptionally (Fong and Bentley, 2001).

### **1.2.1 Splicing and transcription**

Though not obligatorily, splicing is usually coupled to transcription (Bentley, 2005; Kornblihtt et al., 2004). U1, U2, U5, and other components of the early spliceosome are recruited to the nascent transcripts of actively transcribed genes following a dynamic pattern that is consistent with the stepwise assembly of the spliceosome (Gornemann et al., 2005; Lacadie and Rosbash, 2005). First, U1 snRNP is recruited to the newly formed 5'ss, and as the rest of the intron comes into existence, the branchpoint binding factors and U2 snRNP associate at the other end of the intron. Subsequently, the recruitment of U5 takes place, leading to the splicing of the adjacent exons as transcription continues further. It has been demonstrated that RNA polymerase II (RNAP-II) associates and deposits U1 snRNP/SR proteins to the nascent transcripts. Subsequently, SR proteins play a key role in coupling splicing and transcription by directing the assembly of other components of the spliceosome, thus leading to efficient splicing of RNAP-II derived transcripts (Das et al., 2007).

A complex interplay exists between the transcription and the splicing machineries. The presence of promoter-proximal splice sites leads to increased transcription of RNAP-II derived genes, in part through enhancing the initiation step (Furger et al., 2002). Depletion of the SR proteins SF2/ASF and SC35 leads to a significant decrease in the production of nascent RNAP-II derived RNA in cells (Lin et al., 2008). It was shown that SC35 promotes transcriptional elongation by directly interacting with CDK9, a component of P-TEFb, which phosphorylates serine-2 of RNAP-II C-terminal repeats that are critical for elongation (Lin et al., 2008). Further, the U snRNPs enhance transcriptional elongation through the co-factor TAT-SF1 (Fong and Zhou, 2001). Finally, the rate of transcription, depending upon promoter usage, can influence alternative splicing (see below) (Kornblihtt, 2005).

### **1.2.2 Splicing and 5' cap formation**

The terminal exons in a pre-mRNA transcript are different from the internal exons. The boundaries of a 5'-terminal exon are marked by a 5' cap and the downstream 5'ss, whereas the boundaries of a 3'-terminal exon include an upstream 3'ss and a poly(A) signal at the 3' end, so that 'exon definition' of

terminal exon during splicing has to be distinct from that of the internal exons. There is some evidence suggesting that the 5' and 3' terminal exons are probably defined by cap-5'ss and 3'ss-poy(A) tail interactions, respectively (Robberson et al., 1990). Positive effects of the 5'-cap on the splicing of cap-proximal introns, independent of mRNA stability, were noticed twenty years ago (Inoue et al., 1989). The stimulatory effect on the splicing of cap-proximal introns is facilitated by the nuclear cap-binding complex (CBC), which recruits U1 snRNP to the cap-proximal 5'ss (Lewis et al., 1996). Subsequently, the cap also facilitates the replacement of U1 by U6 snRNP during remodeling of the spliceosome (O'Mullane and Eperon, 1998).

### **1.2.3 Splicing and 3'-end processing**

In a similar fashion, 3'-end processing can be either positively or negatively influenced by splicing, due to the presence of an extensive and coupled network of proteins involved in the two processes (Maniatis and Reed, 2002; Vagner et al., 2000a). During papillomavirus late gene expression in undifferentiated epithelial cells, poly(A) addition, but not cleavage, is inhibited by a cryptic 5'ss upstream of the polyadenylation site (Barksdale and Baker, 1995). U1-70K, a U1-snRNP-associated protein, interacts with and inhibits poly(A) polymerase (PAP) when U1 snRNP is bound to the cryptic 5'ss (Gunderson et al., 1998). In adeno-associated virus type-5 (AAV5), use of an alternative polyadenylation site within an intron is inhibited by U1 snRNP binding to the upstream 5'ss, and this inhibition is controlled by the distance of the promoter from the poly(A) site, such that an increase in distance leads to a decrease in inhibition (Qiu et al., 2004; Qiu and Pintel, 2004). A promoter in close proximity most probably leads to enhanced splicing of the intron containing the poly(A) signal, due to an increased positive interaction between the cap and the 5'ss and hence, suppression of 3'-end formation at the intronic poly(A) site. In contrast, the presence of a downstream 5'ss causes read-through of the proximal polyadenylation signal in the upstream long terminal repeat (LTR) of human immunodeficiency virus (HIV), probably mediated by U1 snRNP as above (Ashe et al., 2000; Ashe et al., 1995; Ashe et al., 1997). All of this taken together clearly shows an intricate interplay between all three RNA-processing events.

Conversely, another U1 snRNP associated protein, U1A, interacts with, and stabilizes the binding of cleavage and polyadenylation stimulating factor (CStF), which enhances simian virus 40 late RNA 3'-end formation (Lutz and Alwine, 1994; Lutz et al., 1996). In another example, it has been proposed that the U1A protein is involved in the efficient recognition of the polyadenylation site in exon 4 of the CT/CGRP gene, which was found to be dependent on U1 snRNP binding to a pseudo 5'ss downstream of this poly(A) site (Lou et al., 1996).



Further, it was demonstrated that PAP stimulates splicing by directly interacting with U2AF65 and recruiting it to the 3'ss just upstream of the 3' cleavage/polyadenylation signals (Vagner et al., 2000b). The SR-related protein SRm160, which, in general, acts as splicing activator, also promotes 3'-end cleavage. Strikingly, when expressed at high levels, SRm160 results in uncoupling of splicing and 3'-end formation, which leads to the accumulation of unspliced pre-mRNA in the cytoplasm (McCracken et al., 2002). Interestingly, the Py-tract binding protein (PTB), a well known splicing repressor, can both activate and repress 3' end formation (Castelo-Branco et al., 2004; Millevoi et al., 2009).

### **1.3 Splicing history and its influence on downstream events**

RNA-processing events in the nucleus, and splicing in particular, have far-reaching consequences for the subsequent steps of gene expression. The splicing history of the mRNA is arguably the most important, due to its regulatory influence on almost every downstream step in mRNA metabolism, including kinetics and efficiency of export, proper localization in the cytoplasm, efficiency of translation, and degradation via the nonsense mediated decay (NMD) pathway (Moore and Proudfoot, 2009). The molecular landscape of an mRNA, defined by a diverse set of associated messenger ribonucleoproteins (mRNPs), is highly complex and dynamic. A surprisingly large number of proteins are loaded on the mRNA in a splicing dependent manner, including the THO/TREX complex, the exon-junction complex (EJC) and the SR proteins, which can remain associated with mRNA after splicing (Merz et al., 2007).

#### **1.3.1 Splicing and mRNA export**

##### **a. The THO/TREX complex and mRNA export**

The THO/TREX complex is a pentameric complex in metazoans that associates with the mRNA in a strictly 5'-cap- and splicing-dependent manner. The THO/TREX complex was initially thought to be recruited by the EJC (see below). Later, however, it was shown that THO/TREX binds mRNA preferentially at a single location just downstream of the cap, held together by interactions between one of its components, REF/Aly, and the nuclear cap-binding protein CBP80 (Cheng et al., 2006; Custodio et al., 2004; Masuda et al., 2005). Though not absolutely essential, the splicing-dependent recruitment of the THO/TREX complex increases the efficiency and rate of mRNA export, and is believed to provide directionality to the export process, with the mRNA 5' end emerging first into the cytoplasm, which ensures efficient engagement of the translation machinery for protein synthesis (Valencia et al., 2008).

## **b. The exon-junction complex and mRNA export**

Unlike the THO/TREX complex, the core of the EJC is comprised of a tetramolecular complex, which is thought to be deposited approximately 20 nucleotides upstream of every exon-exon junction at a late stage of splicing (**Fig. 2**). Multiple EJCs loaded at exon-exon junctions presumably serve as tags for the downstream cellular machinery, marking the original positions of the introns in the mature mRNA (Le Hir and Andersen, 2008; Le Hir et al., 2000). EJCs bound to the mRNA are transported into the cytoplasm, where the EJCs that are present in the open reading frame (ORF) of the mRNA are removed by the translational machinery during the first ('pioneer') round of translation (Dostie and Dreyfuss, 2002; Lejeune et al., 2002). The core EJC components dynamically interact and recruit a large number of factors both in the nuclear and cytoplasmic compartments, thus exerting their widespread regulatory effects on various aspects of RNA metabolism, specifically, mRNA export, mRNA localization, translation, and NMD (see below) (Tange et al., 2004).

## **c. The SR proteins and mRNA export**

Another set of proteins that decorate the mRNA are the shuttling SR proteins. The shuttling SR proteins represent a subset of SR proteins, which shuttle rapidly and continuously between the nucleus and the cytoplasm (Cáceres et al., 1998). The SR proteins play a crucial role in splicing of pre-mRNA and thereafter some of them continue to escort the mRNA to the cytoplasm. Recent studies have shown that SR proteins are involved in recruitment of mRNPs in a splicing-dependent manner (Merz et al., 2007). Similar to the THO/TREX complex, the SR proteins interact with the general export receptor, TAP/P15, and act as adaptor proteins for mRNA export, which requires partial dephosphorylation of the C-terminal RS domain (see below) (Huang and Steitz, 2005). SR proteins engage in splicing when they are in a hyperphosphorylated state and as a result of partial dephosphorylation, assume a hypophosphorylated form by the end of the splicing process, which coincides with the hypophosphorylated form being required for export after splicing. This has led to the hypothesis that dephosphorylation of SR proteins might serve as a regulatory signal for mRNA export (Huang and Steitz, 2005; Kohler and Hurt, 2007). In the cytoplasm, a rephosphorylation event(s) destabilizes the interaction of mRNA with the SR proteins, which now interact with the nuclear-import apparatus and are recycled back to the nucleus (Lai et al., 2001).

## 1.4 Splicing and translation

It is a well established fact that splicing leads to enhanced translational efficiency of mRNA when compared to identical mRNA expressed from cDNA (Le Hir and Seraphin, 2008). At least two mechanisms have been described, which can explain this phenomenon. The first mechanism involves an EJC-bound factor, SKAR, which recruits the activated S6 kinase 1 (S6K1) to the newly formed mRNA and leads to an enhanced pioneer round of translation (Ma et al., 2008). S6K1 is a key kinase in the mTOR signaling pathway, which is associated with growth and protein expression. Upon direct activation by mTOR, S6K1 is known to enhance translation (Bhaskar and Hay, 2007).

In the second mechanism, the SR protein SF2/ASF promotes translation initiation by promoting phosphorylation of 4E-BP1, which is a competitive inhibitor of translation initiation (Michlewski et al., 2008). SF2/ASF directly activates mTORC1 (Karni et al., 2008) switching on a cascade that activates S6K1, which in turn phosphorylates 4E-BP1 rendering it inactive. Additionally, SF2/ASF also interacts with, and probably blocks, protein phosphatase 2A (PP2A), which is an antagonist of S6K1 phosphorylation, thus playing a dual role in promoting translation initiation: recruiting a kinase and blocking a phosphatase (Michlewski et al., 2008). It was previously known that SF2/ASF can promote translation of mRNA in an enhancer-dependent manner (Sanford et al., 2004) (**Fig. 2**).

## 1.5 Splicing and nonsense-mediated mRNA decay

Nonsense-mediated mRNA decay (NMD) is a translation-dependent decay pathway in which mRNAs that contain premature termination codons (PTC) are specifically degraded. NMD serves as a quality-control or surveillance mechanism that eliminates or reduces mutant or aberrantly spliced mRNA, which otherwise would lead to translation of truncated non-functional and perhaps toxic proteins. In mammals, the majority of the natural stop codons are found in the last exon, or within 50 base pairs upstream of the last exon-exon junction (Nagy and Maquat, 1998). The presence of one or more EJC 50 or more nucleotides downstream of a stop codon is presumed to target the mRNA for degradation (Chang et al., 2007; Stalder and Muhlemann, 2008). The EJC plays a central role in this pathway by recruiting the NMD-specific factors Upf2 and Upf3. During the pioneer round of translation, the helicase Upf1 binds to the stalled ribosome at the PTC as part of the SURF complex. The SURF complex comprises Upf1 and its kinase Smg1, and the translation release factors eRF1/eRF2. The interaction between the SURF complex and Upf2 at the EJC immediately downstream of the PTC triggers the phosphorylation of Upf1 by Smg1. Subsequently, through an unknown mechanism, phosphorylated Upf1 recruits the RNA-decay machinery,

which leads to rapid decapping and removal of the poly(A) tail, followed by exonucleolytic degradation of the mRNA from both 3' and 5' ends.

Though the detailed mechanisms remain unclear, the SR protein SF2/ASF also promotes NMD upon transient overexpression (Zhang and Krainer, 2004), which is consistent with its role in promoting the pioneer round of translation (see above) (**Fig. 2**).

## **1.6 Alternative splicing**

Alternative splicing is a process through which multiple mature mRNAs are generated as a result of differential inclusion of exons or exon segments from pre-mRNA derived from a single gene (Black, 2003; Matlin et al., 2005). The primary transcripts involved are either identical or may differ at one or both ends as a result of the use of alternative promoters or 3'-cleavage/polyadenylation sites. The basic mechanism leading to alternative splicing is the same as in constitutive splicing. It requires the recognition of cis-acting elements (5'ss, 3'ss, BPS, and Py-tract) by the basal spliceosome, and is influenced by additional regulatory cis-acting elements, specifically, enhancers (ESE, ISE) and silencers (ESS, ISS), and their corresponding trans-acting factors (**Fig. 1.1 and 1.2**). The main difference between alternative and constitutive splicing arises from the fact that alternative splice sites tend to be slightly weaker (Lear et al., 1990; Roca et al., 2005), and alternative exons have fewer ESEs than constitutive exons on average (Wang et al., 2005).

### **1.6.1 The trans-acting regulators of alternative splicing**

The trans-acting factors that regulate alternative splicing are typically RNA-binding proteins, exemplified by two classes of proteins: Ser/Arg-rich (SR) proteins and heterogeneous ribonucleoproteins (hnRNP), which generally recognize ESEs and ISSs/ESSs, respectively. Additionally, a wide variety of RNA-binding proteins, known as SR-like and/or SR-related proteins are also involved in modulating alternative splicing and may act as either activators or repressors or both (Long and Cáceres, 2009).

#### **a. The SR proteins**

SR proteins constitute a family of RNA-binding proteins with closely related sequences and a modular structure with one or two copies of the RNA-recognition motif (RRM) and a C-terminal Arg/Ser-rich (RS) domain that undergoes reversible phosphorylation. They are abundantly expressed across tissues and have multiple roles in constitutive and alternative splicing, including a general role in exon definition, which partly involves their ability to recognize and

bind cognate ESEs (Graveley, 2000). Recently, there has been an explosive burst in studies related to the role of SR proteins in various other aspects of gene expression, specifically, transcription, NMD, mRNA export, and translation, leading to the belief that SR proteins are in fact master regulators of gene regulation (see above) (Long and Cáceres, 2009; Zhong et al., 2009).

#### **a. (i) The prototypical SR protein, SF2/ASF**

SF2/ASF, the founding member of the SR protein family, has a modular structure with two RNA-recognition motifs (RRM), which recognize and bind a seven-nucleotide degenerate ESE motif on pre-mRNA, and an RS domain with numerous repeats of serine-arginine dipeptides (**Fig. 1**). The RS domain is phosphorylated at multiple serine residues and is believed to mediate protein-protein and RNA-protein interactions. SF2/ASF has a well characterized role in general splicing and was first discovered as a factor that was sufficient to promote spliceosome assembly and splicing of various pre-mRNAs in cytoplasmic S100 HeLa cell extracts, which are otherwise deficient in splicing (Krainer et al., 1990). SF2/ASF plays a crucial role in regulation of alternative splicing and usually promotes alternative exon inclusion and use of proximal alternative 5'ss or 3'ss in a concentration-dependent manner, in part through recognition of ESEs (Cartegni et al., 2002). Apart from its role as a broad-specificity splicing regulator, SF2/ASF strongly enhances NMD upon overexpression (Zhang and Krainer, 2004), enhances translation both in vivo and in vitro in an enhancer-dependent manner (Sanford et al., 2004), and functions as an adaptor protein for mRNA export (Huang and Steitz, 2005). Some of these recently identified functions of SF2/ASF are related to its ability to shuttle continuously between the nucleus and cytoplasm (Cáceres et al., 1998) (**Fig. 2**).

SF2/ASF is essential for cell viability and its knockout results in early embryonic lethality in mice (Lin et al., 2005; Xu et al., 2005). Loss of SF2/ASF results in G2 cell cycle arrest, apoptosis, as well as genomic instability in chicken DT40 cells (Li and Manley, 2005; Li et al., 2005). The effects of SF2/ASF in cell-cycle progression and maintenance of genomic stability are probably related to its ability to associate with interphase chromatin and its role in transcription (Lin et al., 2008; Loomis et al., 2009). However, DT40 cells are very rapidly dividing cells that lack p53 and hence are devoid of cell-cycle checkpoints, which makes them extremely susceptible to undergo apoptosis (Takao et al., 1999). Consistently, no apoptosis is observed in p53<sup>+/+</sup> mouse embryonic fibroblasts (MEFs) lacking another SR-protein, SC35, which is also essential for cell viability (Xiao et al., 2007).

In contrast to its depletion, a mere twofold overexpression of SF2/ASF transforms immortal rodent fibroblasts (see below) (Karni et al., 2007).

## **b. The hnRNPs**

Similar to SR proteins, hnRNPs are also abundant RNA-binding proteins that bind nascent pre-mRNA transcripts. They have multiple roles during various steps of pre-mRNA processing, with additional roles in mRNA export and surveillance, and consist of several protein subfamilies (Dreyfuss et al., 2002). One subset of hnRNPs, called the hnRNP A/B proteins, affect alternative splicing, most often by antagonizing SR proteins, which partly involves the recognition of ESSs (He and Smith, 2009; Zheng, 2004).

### **1.6.2 Alternative splicing: the major source of protein diversity**

Alternative splicing is widespread and is a major source of protein isoform diversity in higher eukaryotes. A recent projection based on RNA deep sequencing of 15 different human tissues and cell lines estimates that ~100% of all multi-exon human genes undergo some alternative splicing and ~86% percent of these multi-exon genes produce significant levels ( $\geq 15\%$ ) of the minor isoform (Pan et al., 2008; Wang et al., 2008). The widespread role of alternative splicing in gene regulation, especially in those events that govern cellular differentiation during development, as well as allelic variation in human populations, is obvious, as 94% of all human genes comprise multiple exons. These estimates are significantly higher than earlier ones from microarray-based studies, which had limitations inherent to the microarray technology, especially those pertaining to the depth of coverage of the human transcriptome and inability to detect less abundant isoforms (Johnson et al., 2003). The estimates of alternative splicing events (~10-30%) that affect relative levels of spliced isoforms due to allelic variations among individuals were also higher than in previous microarray-based studies (Graveley, 2008; Kwan et al., 2008), but were similar to an integrated analyses of multiple data types, which had placed these estimates close to 21% (Nembaware et al., 2004). However, the most striking observation was that the majority of alternative splicing events, ranging from 52-74%, depending upon the type of alternative splicing pattern, are regulated in a tissue specific manner, thus supporting the notion that alternative splicing is the main contributing factor in generating the complexity and diversity of the human proteome.

Alternative splicing can directly influence the localization, stability, and translation of the resulting isoforms. Alternatively spliced mRNAs are either co-expressed or their expression is tightly regulated in a cell-type or tissue-specific manner, developmentally, or in response to signaling. Multiple protein isoforms

resulting from the alternatively spliced variants can have tissue-specific related, distinct, or antagonistic functions, or they can have different subcellular localization.

Very often, alternative splicing acts as an on/off switch to control gene expression at the post-transcriptional level, such that one mRNA isoform results in a functional protein, whereas the other(s) is subject to degradation by NMD, or is retained in the nucleus, or may result in a truncated/inactive or mislocalized protein. Such regulation in the case of cellular oncogenes and tumor suppressors has widespread implications. One important example in this regard is the *p73* tumor-suppressor gene, which gives rise to six different proteins with closely related sequences, *p73 $\alpha$* , *p73 $\beta$* , *p73 $\gamma$* , *p73 $\delta$* , and *p73 $\zeta$* , as a result of extensive alternative splicing in the C-terminal domain. These C-terminal splice variants exhibit different transcriptional and functional properties and show differential expression across normal human tissues and cell lines. Further, the use of an alternative promoter in exon 3 of the full-length *p73* gene results in the  $\Delta$ N variants of *p73*,  $\Delta$ N*p73 $\alpha$*  and  $\Delta$ N*p73 $\beta$* , which lack the N-terminal transactivation domain, and act as dominant-negative forms of both *p73* and *p53* with oncogenic activity (Ozaki and Nakagawara, 2005). Similar regulation through alternative splicing, combined with the use of alternative promoters, has been reported in the case of *p63* and *p53*, which belong to the same gene family as *p73* (Bourdon et al., 2005; Yang et al., 2002; Yang et al., 1998).

### **1.6.3 Alternative splicing and disease**

According to an earlier report, 15% of all genetic diseases result from mutations that disrupt splicing (Krawczak et al., 1992). More recent estimates have raised the proportion to as high as 60% (Lopez-Bigas et al., 2005). The difficulty to correctly estimate the precise frequency of splicing mutations reflects our incomplete understanding of the splicing code, which is primarily due to the degenerate nature of enhancers (ESE and ISE) and silencers (ESS and ISS), and the intricate interdependencies among themselves or in conjunction with their trans-acting factors (Roca et al., 2008). In general, most of the intronic mutations and silent exonic mutations, or genetic polymorphisms that do not lie within the splice sites (5'ss and 3'ss), tend to be ignored as neutral when identified, as they are not expected to cause any changes at the protein level. Moreover, a major fraction of isoforms generated by alternative splicing, even for extensively characterized genes, is unknown. In an analysis of 50 well characterized genes involved in cancer, nearly two thirds were found to express previously unknown isoforms and surprisingly, in 40% of the cases, the novel splice variants were found to be the major isoform in normal tissues (Roy et al., 2005). Functional studies, conducted in an attempt to identify disease-causing

mutations when the predominant isoform is not known in the affected tissue type, may be misleading if the isoform under analysis is irrelevant for the disease. Further, functional differences between the isoforms can conceal the true pathogenic effects of a mutation (Wang and Cooper, 2007). Nonetheless, the role played by misregulation of alternative splicing in the manifestation of a significant proportion of genetic disorders, including cancer, is certain and has been well documented (Cartegni et al., 2002; Cooper et al., 2009; Wang and Cooper, 2007) (also, see Chapter 3).

#### **a. Alternative splicing and cancer**

There is an ever-increasing number of instances in which misregulation of splicing has been associated with various types of cancers (Brinkman, 2004; Kalnina et al., 2005; Pio and Montuenga, 2009; Skotheim and Nees, 2007; Venables et al., 2009). These involve either mutations in the cis-acting sequence elements or alterations in the expression/function of trans-acting factors that regulate splicing. In the first case, mutations may involve key splice-site signals, including the 5'ss, 3'ss, and BPS, or regulatory sequences, including ESEs, ESSs, and ISSs. For example, a nonsense mutation in *BRCA1*, which is associated with increased risk of breast cancer, disrupts an ESE and leads to skipping of exon 18 (Liu et al., 2001; Mazoyer et al., 1998).

In the second case, changes in the level of expression of splicing regulators, which mainly include SR proteins and hnRNPs, may cause aberrant splicing of target genes. It has been shown that SR proteins display tumor-stage-specific changes in expression levels during mammary tumorigenesis, pointing towards possible concentration-dependent effects on splice-site selection (Stickeler et al., 1999). Frequently, changes in alternative splicing of oncogenes and tumor suppressors, such as *p53*, *BIN1*, *MDM2*, and *Bcl-x*, or the *CD44* and estrogen receptor genes, are associated with breast cancer (Bourdon et al., 2005; Ge et al., 2000a; Lukas et al., 2001; Rochaix et al., 1999; Stickeler et al., 1999; Zhang et al., 1996). Notably, in the case of *CD44*, enhanced inclusion of exons v4 and v5 is due to an increase in the levels of the SR-like protein, Tra2- $\beta$ 1, and this effect appears to be antagonized by an SR protein, SC35 (Watermann et al., 2006). Altered expression levels and/or subcellular localization of different hnRNP A/B proteins are associated with lung cancer (Peebles et al., 2007; Tockman et al., 1994; Tockman et al., 1997; Zerbe et al., 2004). These lines of evidence suggest that changes in the levels of various splicing factors can lead to cancer.



### a. (i) SF2/ASF and cancer

Recently we found that SF2/ASF is a potent proto-oncogene whose overexpression by as little as twofold is sufficient to transform immortal rodent fibroblasts (Karni et al., 2007). Loss of SF2/ASF induces G2 cell-cycle arrest, apoptosis and genomic instability (Li and Manley, 2005; Li et al., 2005). It has been shown that changes in the levels of SF2/ASF lead to altered splicing of the apoptotic protein caspase-9 (Massiello and Chalfant, 2006) and the *Ron* proto-oncogene (Ghigna et al., 2005). In the latter case, SF2/ASF promotes the expression of a constitutively active isoform of Ron,  $\Delta$ Ron, which results after skipping exon 11 of the *Ron* pre-mRNA. Increased levels of  $\Delta$ Ron are known to confer invasive growth properties to the cell through increased cell proliferation, decreased apoptosis, increased mobility, and promotion of cell dissociation, which are hallmarks of the epithelial to mesenchymal transition (EMT) as well as tumor progression and metastasis.

We have recently identified the tumor suppressor *BIN1*, which is an inhibitor of Myc (Sakamuro et al., 1996), as a splicing target of SF2/ASF. *BIN1* is frequently downregulated or inactivated in various cancers (DuHadaway et al., 2003; Ge et al., 2000b; Tajiri et al., 2003; Tvarusko et al., 1999), including breast cancer (Ge et al., 2000a). *SFRS1*, the gene encoding SF2/ASF, resides on Chromosome 17q23, a region that is commonly amplified in breast cancer and correlates with poor prognosis (Sinclair et al., 2003). The gene *RPS6KB1* also resides on the 17q23 amplicon, and we found that its splicing pattern was dramatically altered upon overexpression of SF2/ASF in primary cells, which led to identification of a novel splice variant of S6K1 that results in a 31-kDa protein (Karni et al., 2007). *RPS6KB1* encodes the protein S6K1, which phosphorylates the 40S ribosomal subunit protein S6 (also, see section 1.4) (Erikson and Maller, 1985). The protein has two known isoforms resulting from alternative translation start sites: a 70-kDa and an 85-kDa protein (Berven and Crouch, 2000). The 70-kDa isoform, known as p70S6K, is the best studied and has been implicated in the regulation of the cell cycle (Feng et al., 2000), cell growth, proliferation (Dufner and Thomas, 1999), and cell migration (Lambert et al., 2002). Our studies, using various cell lines derived from breast cancer tumors, MCF7, BT474, ZR751, HS578T, and T47D, confirm and extend previous findings. We observed high levels of SF2/ASF in the first three cell lines, which strongly correlated with the altered splicing of the target *RPS6KB1*, when compared to immortal MCF10A cells (Karni et al., 2007). In two instances, MCF7 and BT474, we observed a direct correlation between genomic amplification of 17q23 and increased expression of SF2/ASF.

Although now we know that SF2/ASF and probably other splicing regulators can act as oncogenes, we still have very little understanding of the detailed molecular pathways that lead to SF2/ASF-mediated cellular transformation. Various clues provided by our overexpression studies in different cell types suggest that there may be a stepwise mechanism for transformation by SF2/ASF: first, an early event, involving changes in the splicing patterns of key oncogenes and tumor suppressors. Most notable among these are S6K1, *MNK2*, the putative tumor suppressor *BIN1* and the Ron oncogene (see above). These changes may cause enough perturbations in cellular homeostasis to trigger a second step, in which some of the splicing-independent functions of SF2/ASF now assume a dominant role and help maintain the transformed state. Many of the recently identified roles of SF2/ASF, specifically, the dynamic association of SF2/ASF with chromatin in a cell-cycle dependent manner, its role in transcription stimulation as well as the recent finding that it can influence translation via the mTOR-pathway, are of special interest in this regard (see above). In at least one cancer cell line, this phenotype could be reversed by restoring the normal levels of SF2/ASF (Karni et al., 2007), which strongly suggests that SF2/ASF may be required at higher levels to maintain the transformed state. Further, we recently found that SF2/ASF regulates its own expression through an intricate process that involves multiple mechanisms acting at different levels (Sun et al., 2009). It is highly likely that such an autoregulatory loop has profound influence on the expression of various genes in the 17q23 region, such as *RPS6KB1*, as well as that of additional genes involved in oncogenesis. Moreover, post-translational modifications (PTMs) of SF2/ASF could be an important determinant of its activity and its role as an oncogene, a possibility that remains unexplored (see Chapter 2). Future studies designed to explore one or more scenarios presented above may provide us with a better understanding of the underlying molecular mechanisms that govern SF2/ASF-mediated cancer pathogenesis.

## 1.7 Concluding remarks

Pre-mRNA splicing is an extremely complex process. Though we have come a long way since splicing was discovered more than 30 years ago, we are only beginning to understand the details of how splicing is regulated. For the most part, we understand in great detail which factors are involved (e.g. SR proteins/hnRNPs) and what is their effect (inclusion/skipping), but we have very little or no knowledge of either the mechanism of action or the mode of regulation that triggers the effects. For example, we know that SR proteins are required for general splicing and can modulate alternative splicing, in part by binding to ESEs, but we still do not know the details of how this may happen, nor do we

understand fully how the activity of SR proteins itself may be regulated. A similar theme is apparent when other aspects of SR protein function are considered, specifically, their roles in transcription, NMD, or translation. Further, deciphering the splicing code still remains the biggest challenge in the field.

The prime importance of splicing as a cellular process is reflected by its ability to influence almost every subsequent step in gene expression (see above). For the most part, alternative splicing accounts for the organismal complexity in higher eukaryotes through its contribution to proteomic diversity. Given the central and crucial role played by splicing in maintaining the normal physiology of most eukaryotes, it is imperative that we invest considerable effort to achieve a better overall understanding of splicing regulation, for such understanding is clearly lacking. For example, splicing may respond to various intra- as well as extra-cellular stimuli, resulting in different splicing patterns of specific genes and at specific stages, which may be the key to development and tissue differentiation as well as 'dedifferentiation' leading to cancer metastasis. A better grasp of such events may better equip us to fight the numerous diseases related to splicing defects, as well as various types of cancer.

## 1.9 Figures and legends

**Figure 1.1 Consensus splice-site sequences in mammals.** The 5' splice site, branchpoint site, polypyrimidine-tract and the 3'splice site are depicted with their consensus motifs in the blow up. The height of the letters is directly proportional to the observed frequencies of the nucleotides at the corresponding positions (Burge et al., 1998). The U1 snRNP-5'ss and the U2 snRNP-BPS base-pairing interactions and the possible role of SR proteins in recruitment of snRNPs as well as in bridging the interactions at both ends of introns during spliceosome assembly are also shown. (N=any nucleotide, R=purine, Y=pyrimidine).

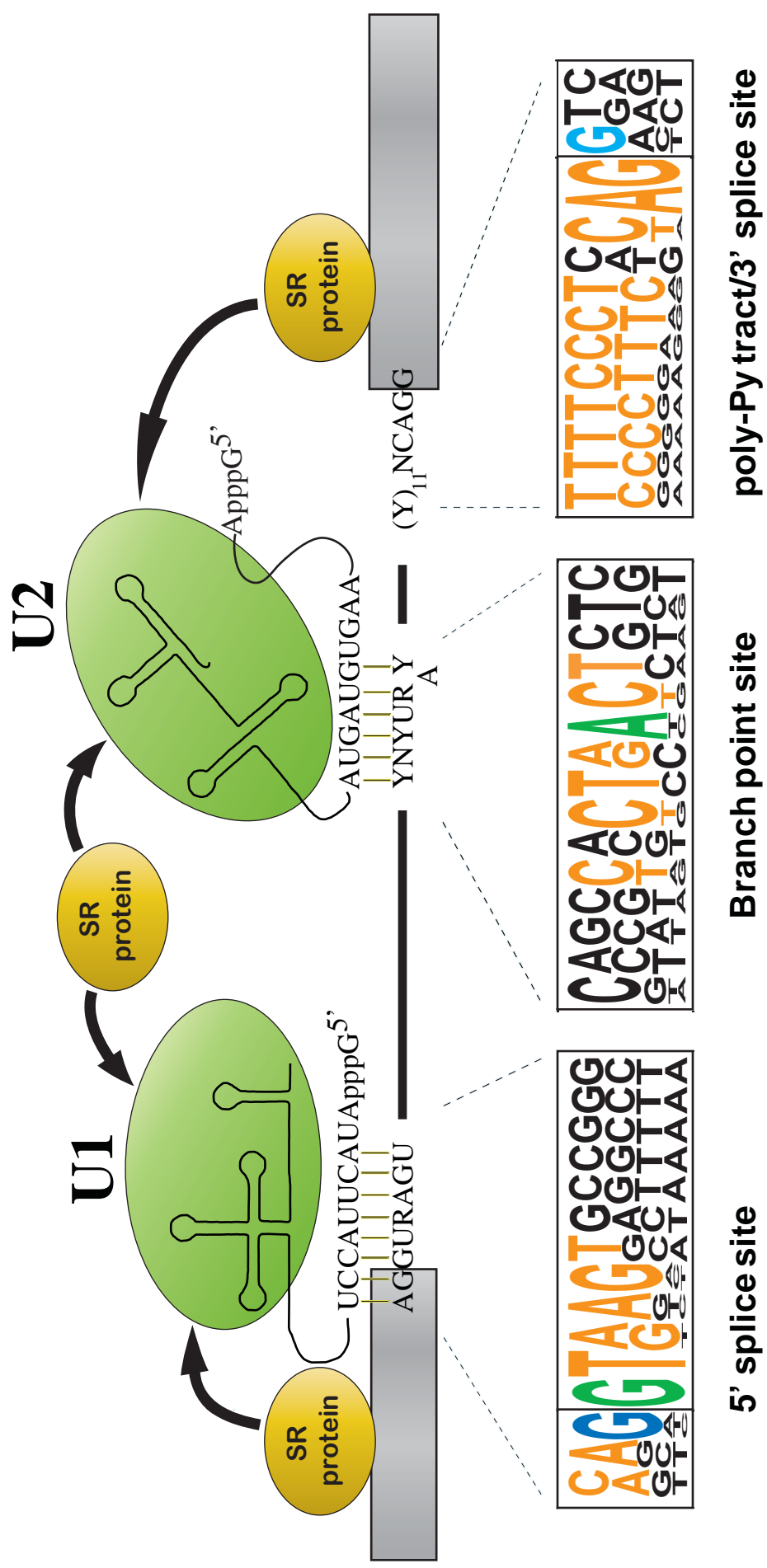


Figure 1

**Figure 1.2 Diverse functions of SF2/ASF.** Multifaceted roles of SF2/ASF in RNA metabolism across different subcellular compartments. In steady state SF2/ASF is localized in the nucleus where it accumulates in the interchromatin granule clusters (IGC; nuclear speckles). From speckles it is recruited to the sites of active transcription, where it is involved in general splicing, as well as regulates alternative splicing in part by binding to 6-8 nucleotide enhancer motifs (ESEs). By an unknown mechanism that requires a nuclear function of SF2/ASF, it also promotes nonsense-mediated mRNA decay (NMD) of a premature termination codon containing mRNA. Further, SF2/ASF acts as an adapter protein for mRNA export via its interaction with the TAP/p15 complex. In the cytoplasm SF2/ASF stimulates translation in an enhancer dependent manner. Finally, via its interaction with Transportin-SR proteins, SF2/ASF is shuttled back to the nucleus.

# Multiple functions of SF2/ASF

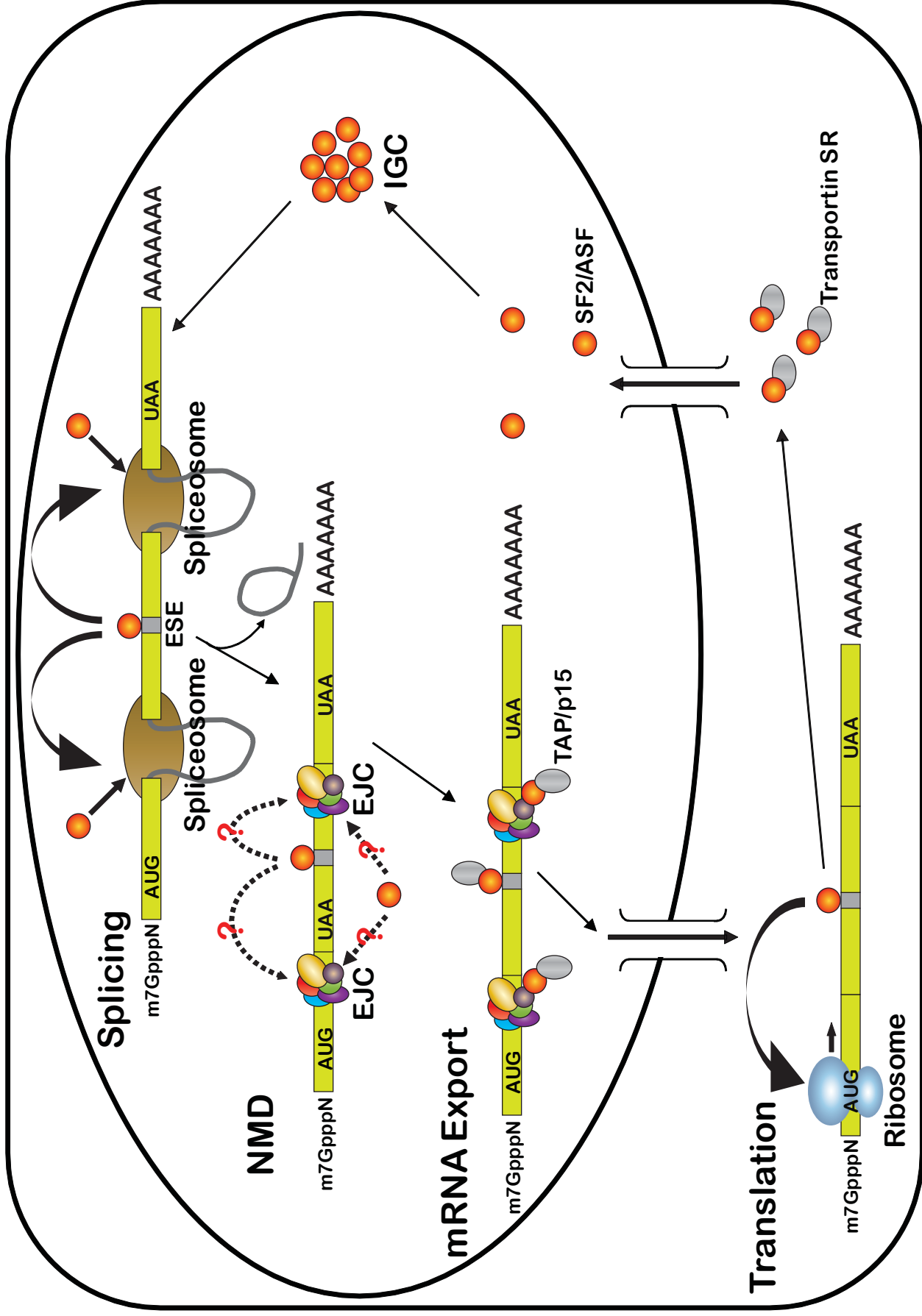


Figure 2

## **Chapter 2:**

# **Physiological Role of Post-Translational Modifications of the Oncoprotein SF2/ASF**

### **Abstract**

Alternative splicing and post-translational modifications (PTMs) are major sources of protein diversity in eukaryotic proteomes. The SR protein SF2/ASF is an oncoprotein that functions in pre-mRNA splicing, with additional roles in other post-transcriptional and translational events. Functional studies of SR-protein PTMs have exclusively focused on reversible phosphorylation of Ser residues in the C-terminal RS domain. Our proteomic analysis of SF2/ASF and other SR proteins obtained from mammalian cells revealed that they are extensively modified. We confirmed that human SF2/ASF is methylated at residues R93, R97, and R109, which were recently identified in a global proteomic analysis of Arg methylation, and further investigated whether these methylated residues regulate the functional properties of SF2/ASF. We show that the three arginines additively control the subcellular localization of SF2/ASF, and both the positive charge and the methylation state are important. Mutations that block methylation and remove the positive charge result in cytoplasmic accumulation of SF2/ASF. The consequent decrease in nuclear SF2/ASF levels prevents it from modulating alternative splicing of target genes, results in higher translation stimulation, and abrogates the enhancement of nonsense-mediated mRNA decay. This study addresses the mechanisms by which Arg methylation and the associated positive charge regulate the activities of SF2/ASF, and emphasizes the significance of localization control for an oncoprotein with multiple functions in different cellular compartments. Moreover, the various other types of modifications identified in this study of SF2/ASF, as well as other SR proteins, lay a foundation for future studies of the detailed mechanisms by which different PTMs may modulate the diverse functions of SR proteins.



## 2.1 Introduction

The SR protein SF2/ASF is a master regulator of gene expression, due to its widespread influence on almost every aspect of gene expression, ranging from transcription to translation. In addition to RNA metabolism, the scope of regulatory effects exerted by SF2/ASF extends further with its involvement in cell-cycle progression and the maintenance of genomic instability. Reflecting its multifaceted role as a regulator in various cellular processes, SF2/ASF is essential for cell viability and embryonic development, and can act as a potent oncoprotein with only a slight increase in its expression level (see Chapter 1). Surprisingly, even though the splicing of SF2/ASF pre-mRNA gives rise to six different splice variants, only one of them is translated, resulting in a protein product (Sun et al., 2009). It becomes difficult to explain how a single protein, which was originally identified as a splicing factor (Krainer et al., 1990), can influence such a wide spectrum of cellular processes and more strikingly, across different cellular compartments.

One of the ways through which the various activities of SF2/ASF could be regulated is by post-translational modifications (PTMs). Like alternative splicing, PTMs are an integral part of gene regulation and provide another way of creating protein isoform diversity, with widespread functional consequences. The chemical changes, which modify the side chains of specific amino acids in a polypeptide chain after translation (hence PTMs), greatly extend the functional diversity of individual proteins. In general, proteins are modified through one and/or any possible combination of the three basic mechanisms, which are: 1) by addition of new functional groups, including acetate, phosphate, lipids, carbohydrates, sulfate, etc., or by modifying the properties of existing functional groups through methylation or citrullination; 2) by contributing to the three-dimensional (3-D) structure of the proteins by forming disulfide bridges; and 3) by editing the primary polypeptide sequence, which includes either removal of the first amino acid, the amino-terminal (N-terminal) methionine, or a part of the polypeptide sequence through proteolytic cleavage. The modifications, if present, not only ensure that a single protein is able to perform many different or related functions, but also provide a powerful means for dynamic regulation of the protein activity by acting like on/off switches. Often, this is achieved simply by changing the localization of the protein to a different compartment in the cell as a result of modifications. Such type of regulation is most notable in epigenetic control of gene regulation and cancer. Different PTMs may arise in key proteins at different stages of cancer development and affect the degree of invasiveness. Many oncogenes are protein tyrosine kinases (PTK), which act as regulatory

on/off switches. In fully differentiated somatic cells, the expression of most genes is under epigenetic control, regulated by PTMs of histone tails, errors in which often lead to various diseases and cancer (Cho et al., 2004; Esteller, 2006).

Descriptions of PTMs, in the case of SR proteins, have mostly focused on reversible phosphorylation at multiple Ser residues within the C-terminal RS domain (**Fig. 2.3a**), which is regulated by a combination of phosphatases and kinases (Misteli, 1999). SF2/ASF and SR proteins, in general, are phosphorylated by a family of SR protein-specific kinases (SRPK) (Gui et al., 1994) and a family of dual specificity kinases Clk/Sty (Colwill et al., 1996). Recently, it was determined that Akt also promotes phosphorylation of SF2/ASF, directly or indirectly, and enhances its activities in splicing and translation (Blaustein et al., 2005). SF2/ASF localizes in nuclear speckles and its release to active sites of transcription is modulated by phosphorylation and various protein-protein interactions (Misteli et al., 1998; Misteli et al., 1997; Misteli and Spector, 1996). A requirement for phosphorylation for the activity of SF2/ASF in splicing has been reported (Xiao and Manley, 1997). However, the fact that the RS domain is dispensable for the activity of SF2/ASF in constitutive splicing for some and perhaps all substrates (Shaw et al., 2007; Zhu and Krainer, 2000), and for its concentration-dependent effects on alternative splicing (Caceres and Krainer, 1993; Zuo and Manley, 1993), suggests that phosphorylation of residues other than those in the RS domain could be essential for the activity of SF2/ASF in splicing of certain substrates.

Apart from its subnuclear localization and activity in splicing, phosphorylation of SF2/ASF also modulates its subcellular localization (Koizumi et al., 1999). The RS domain is required for shuttling of SF2/ASF (Cáceres et al., 1998) and contributes to its nuclear localization (Cáceres et al., 1997). It was shown that phosphorylation of the RS domain is essential for interaction of SF2/ASF with transportin-SR2 (TRN-SR2), which acts as a receptor for nuclear import of SR proteins (Lai et al., 2001). Further, the phosphorylation state of SF2/ASF has been implicated in its activity as an adaptor protein for TAP-mediated mRNA export (Huang et al., 2004; Lai and Tarn, 2004).

More recently, three methylated Arg residues (R93, R97, and R109) were reported in human SF2/ASF as part of a global analysis of protein methylation (Shao-En Ong, 2004) (**Fig. 2.3a**). Quantitatively, Arg methylation is one of the most extensive protein modifications that takes place in the cell, with a large repertoire of proteins that get modified. Three distinct classes of methyltransferases, Type I-III, exist in mammals. All three types of enzymes can recognize the unmodified side chain of Arg residues in a polypeptide and modify one of the terminal

guadinino nitrogen atoms ( $\omega$ -N or  $\omega$ -N') to  $\omega$ -N<sup>G</sup>-monomethylarginine ( $\omega$ MMA). It should be noted that both of the terminal  $\omega$ -N atoms are chemically equivalent, due to resonance, which leads to delocalization of the positive charge.  $\omega$ MMA is further modified to either  $\omega$ -N<sup>G</sup>, N<sup>G</sup>-dimethylarginine (asymmetric-dimethylarginine; ADMA) by type-I enzymes or  $\omega$ -N<sup>G</sup>, N<sup>G</sup>-dimethylarginine (symmetric-dimethylarginine; SDMA) by type-II enzymes (Bedford and Clarke, 2009) (**Fig. 2.2**).

In budding yeast, the protein Npl3p, which resembles both hnRNPs and SR proteins, based on sequence analysis (Birney et al., 1993), undergoes Arg methylation at RGG motifs in its C-terminal RGG/RS domain (Siebel and Guthrie, 1996). Npl3p is a shuttling protein with predominant nuclear localization at steady state, and is involved in mRNA export (Lee and Silver, 1997; Singleton et al., 1995). Npl3p is methylated by the type-I methyltransferase Hmt1p and fails to exit the nucleus in cells lacking Hmt1p (Shen et al., 1998). It was further shown that hypermethylation of RGG motifs blocks phosphorylation of the neighboring RS/SR dipeptides by Sky1p, and this in turn blocks the nuclear import of Npl3p (Yun and Fu, 2000). In mammals, methylation of RGG motifs in the C-terminal Gly-rich domain of the splicing factor hnRNP A2 promotes its nuclear localization (Nichols et al., 2000). The methylated Arg residues in SF2/ASF are part of Arg/Gly-rich motifs, which are very similar to those in Npl3p and hnRNP A2, though these motifs differ in their location in SF2/ASF. Indeed, it was shown that SF2/ASF is methylated by the human homolog of Hmt1p, PRMT1 (Goulet et al., 2007). PRMT1 is a type-I methyltransferase that catalyzes the formation of monomethyl and asymmetric dimethyl Arg on proteins (Weiss et al., 2000; Wooderchak et al., 2008), preferentially at RGG/GRG and related motifs (Hyun et al., 2000; Kim et al., 1997).

Besides Ser-phosphorylation and Arg-methylation described above, no other type of modifications are known for SF2/ASF or for SR proteins in general. We report an astonishing finding that may provide clues regarding largely unknown molecular mechanisms, by which SF2/ASF may regulate various important cellular processes. Our results show the presence of an extensive and diverse set of modifications on SF2/ASF at multiple residues. We further extend these findings to two other SR proteins that showed a similar high degree of diverse modifications at multiple positions. Finally, we confirmed the presence of previously reported methyl-Arg residues (Shao-En Ong, 2004), and after an extensive analysis of the various functions of SF2/ASF, we report that these three methylated Arg residues in the inter-RRM linker region control SF2/ASF's subcellular localization, which in turn determines the protein's activity in

enhancing NMD and regulating alternative splicing of endogenous pre-mRNA targets.

In each and every aspect of SF2/ASF-mediated regulation of different cellular processes, perturbations in SF2/ASF levels lead to dramatic consequences. These studies emphasize the importance of PTM-directed changes in subcellular localization of a regulatory protein, like SF2/ASF, as a means of controlling its local concentration in different compartments of the cell, which result in differential effects on individual cellular processes, depending upon the location where they take place. Additionally, these findings also lay the groundwork for future studies focusing on the physiological roles of other types of modifications, not only in the case of SF2/ASF, but also for other SR proteins.

## **2.2 Results**

### **2.2.1 SF2/ASF is extensively modified**

We hypothesized that PTMs might be of great importance in regulating the activities of SF2/ASF in different cellular processes, which in turn might be influenced by key signaling events in the cell. Also, there might be a complex interplay between different kinds of PTMs affecting such regulation. Most studies to date of SF2/ASF PTMs using proteomic approaches (Velazquez-Dones et al., 2005) or other techniques (Ngo et al., 2005; Sanford et al., 2005) have focused on phosphorylation of serines in the RS domain. Even though phosphorylation of the RS domain is well documented, with some insights into its effects, the studies to date are incomplete and have not fully revealed the exact sites of phosphorylation and their precise functional roles (see above).

Therefore, in order to thoroughly map all the possible modifications of SF2/ASF protein, we analyzed C-terminal 6xHis tagged SF2/ASF, expressed and purified from human 293-EBNA1 (293E) cells, by mass spectrometry (mass-spec). For the identification of specific modified residues, the purified protein was further separated by SDS-PAGE to remove minute impurities. After staining with mass-spec-compatible dyes, the bands corresponding to SF2/ASF were excised and the protein was subjected to trypsin digestion that resulted in SF2/ASF peptides containing individual modified residues. In general, two to three closely migrating bands of SF2/ASF could be resolved by electrophoresis, reflecting different modification states of the protein. Each of the individual bands was carefully excised and analyzed separately. The trypsin-derived peptides were then subjected to MALDI/TOF, followed by MS/MS analysis for some of the selected peptides. As a control, full-length untagged SF2/ASF purified from

bacteria, mimicking the unmodified mammalian protein, was included. A truncated version of SF2/ASF lacking the RS domain, expressed in and purified from 293E cells, the C-terminal 6xHis tagged SF2/ASF- $\Delta$ RS, was analyzed to specifically look for phosphorylation events that may occur elsewhere in the protein other than the RS domain. Further, N-terminal 6xHis tagged SF2/ASF coexpressed with Clk/Sty in bacteria was also included in the analysis in an attempt to identify phosphorylation events specifically catalyzed by Clk/Sty.

To our surprise, after a thorough comparative analysis of the four different versions of SF2/ASF, consisting of multiple repeat experiments and followed by mapping of individual PTMs, we found the presence of extensive modifications, including Lys methylation and acetylation, Ser and Thr phosphorylation, Tyr phosphorylation, and many novel sites of Arg methylation, in addition to those described previously (**Fig. 2.1**). We further verified the results by MS/MS analysis for some of the peptides, which also confirmed that SF2/ASF is methylated at Arg residues 93, 97, and 109, consistent with the previous observations from global proteomics (Shao-En Ong, 2004).

Similar analysis of 9g8, SRp20, and SC35 all expressed in and purified from 293E cells, revealed fewer modifications (**Table 1, 2, and 3**). The lower number of modifications found in these proteins may be a direct reflection of a far less comprehensive analysis done in these cases as most of our effort was focused on SF2/ASF. In general, this type of analysis is especially difficult in the case of SR and SR-related proteins. Trypsin, which is usually used in these experiments, cleaves after unmodified Arg or Lys residues in a polypeptide chain. The presence of continuous repeats of RS/SR dipeptides in the RS domain of these proteins makes the identification of PTMs in this region extremely difficult, as it is almost impossible to map the observed masses in the mass-spec to a specific peptide at a specific position, when many such short peptides exist elsewhere in the same RS domain. This problem is further compounded by the lack of other mass-spec-compatible proteolytic enzymes with well defined cleavage specificities that are different from trypsin's. For example, we used chymotrypsin in some of our experiments, which preferentially cleaves after aromatic residues (Phe, Trp, and Tyr) with high specificity, but in addition it also cleaves after Leu, Met, and His with lower specificity, thus making the analysis of the results difficult. Nonetheless, using optimized trypsin digestion conditions, which allowed for up to three miscleavages in our searches, we were able to obtain more than 70% peptide coverage in our analysis of SF2/ASF.

### 2.2.2 The methyl-arginines affect the nucleo-cytoplasmic distribution of SF2/ASF

Since the regulatory roles of SF2/ASF span both cytoplasmic and nuclear compartments, we hypothesized that proper localization, controlled by post-translational modification, is essential for its activities. Based on the studies of Npl3p in yeast and other RNA-binding proteins, we further suspected that the sites of Arg methylation in SF2/ASF (**Fig. 2.3A**) might influence its localization. We anticipated that changes in the distribution of SF2/ASF between the nucleus and cytoplasm may affect alternative splicing of target genes, as well as its other functions. Since a very large number of Arg residues were found to be methylated, we decided to focus on the three methyl-Arg residues that were already described (Shao-En Ong, 2004). To investigate the functional significance of the modifications, we mutated these sites singly or in combination to Ala, to prevent methylation, as well as abolish the positive charge. Similarly, as a control, we also mutated these residues to Lys, to retain the positive charge but presumably prevent methylation. We determined the subcellular localization of wild-type SF2/ASF (wt-SF2/ASF) and the missense mutants by indirect immunofluorescence (IF) via an N-terminal T7 epitope tag after transient transfection of the cDNAs into HeLa cells. We observed that the triple-Ala mutant, A<sub>1</sub>A<sub>2</sub>A<sub>3</sub>, was predominantly cytoplasmic, in contrast to the corresponding triple-Lys substitution mutant, K<sub>1</sub>K<sub>2</sub>K<sub>3</sub>, or wt-SF2/ASF, which localized to nuclear speckles (**Fig. 2.3B**). The single- or double-Ala mutants displayed intermediate phenotypes (**Fig. 2.4**).

To test if the A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> mutant could still shuttle, we treated transiently transfected HeLa cells with a nuclear export inhibitor, Leptomycin B. The accumulation of T7 signal in the nucleus, upon treatment with the inhibitor, indicates that A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> shuttles between the nuclear and cytoplasmic compartments (**Fig. 2.3C**). We also tested the shuttling of the nuclear protein, K<sub>1</sub>K<sub>2</sub>K<sub>3</sub>, using a heterokaryon assay, and found it to be the same as that of wt-SF2/ASF (**Fig. 2.3D**). The heterokaryon assay (Piñol-Roma and Dreyfuss, 1992), which is used to test the shuttling activity of a nuclear protein, could not be used in the case of the A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> mutant, as it showed predominant cytoplasmic localization.

### 2.2.3 The triple-Ala mutant, A<sub>1</sub>A<sub>2</sub>A<sub>3</sub>, fails to enhance NMD

Splicing is a nuclear process (Black, 2003) and a nuclear activity of SF2/ASF appears to be required for its stimulatory effect on NMD (Zhang and Krainer, 2004). Therefore, we next assayed the activity of the above SF2/ASF

mutants to enhance NMD, using an RNase-protection assay (RPA). We used a full-length  $\beta$ -globin gene and a mutant version with a nonsense mutation in codon 39 in exon 2 as reporters. The nonsense mutation introduces a premature termination codon (PTC), making the mRNA susceptible to NMD (**Fig. 2.5D**) (Thermann et al., 1998; Zhang et al., 1998). RPA showed that upon overexpression, the wt-SF2/ASF, as well as the  $K_1K_2K_3$  version, strongly enhanced NMD of the T39 (mut) mRNA, whereas the mislocalized  $A_1A_2A_3$  mutant was similar to the empty vector control (**Fig. 2.5A and C**). The single- and double-Ala mutants had intermediate effects (**Fig. 2.5B and C**).

To assay both alternative splicing and NMD using RPA, we used a derivative of  $\beta$ -globin with a 5' splice site (5'ss) duplication in the first exon (5'ssD-wt) (Zhang and Krainer, 2004). Because of alternate 5'ss usage, the pre-mRNA gives rise to two distinct mRNAs: a longer proximal mRNA and a shorter distal mRNA. The mutant version of this gene (5'ssD-mut) carries a point mutation that introduces a PTC in the proximal, but not the distal, spliced mRNA (**Fig. 2.6C**). As reported previously (Zhang and Krainer, 2004), SF2/ASF promoted the selection of the proximal 5'ss and enhanced NMD of the resulting mRNA. In contrast, the  $A_1A_2A_3$  mutant efficiently promoted proximal splice-site selection, but failed to enhance NMD; the  $K_1K_2K_3$  mutant, however, showed identical activity as wt-SF2/ASF (**Fig. 2.6A, B, and D**). Again, as observed above, the single- and double-Ala mutants had intermediate effects (**Fig. 2.7**).

#### **2.2.4 Reduced activity of the $A_1A_2A_3$ mutant in general and alternative splicing in vitro**

To further analyze the activity of these proteins in splicing, untagged recombinant SF2/ASF and the  $A_1A_2A_3$  and  $K_1K_2K_3$  versions were purified to homogeneity from *E. coli*. Post-translationally modified versions of these proteins were also purified after expression in 293E cells with a C-terminal 6xHis tag (**Fig. 2.8A**). Purified proteins were then tested for their ability to complement HeLa S100 extract to carry out general splicing of various pre-mRNAs. We observed that the  $K_1K_2K_3$  mutant was as active as wt-SF2/ASF, but the  $A_1A_2A_3$  mutant was approximately fourfold less active with  $\beta$ -globin pre-mRNA (**Fig. 2.8B, D and 2.9**). We obtained similar results with other substrates, including Ftz, AdML, IgM-M1M2, IgM-C1C2, and  $\delta$ -crystallin (data not shown).

To test alternative splicing activity in vitro, we used a model  $\beta$ -globin pre-mRNA with a duplicated 5'ss in intron 1 (Zhang and Krainer, 2004). Again, we observed that the  $K_1K_2K_3$  mutant was as efficient as wt-SF2/ASF in promoting proximal 5'ss use in HeLa nuclear extract, whereas the  $A_1A_2A_3$  mutant was

slightly less active (**Fig. 2.8C and E**). In all the assays, the trend was similar and did not vary when the proteins were obtained from bacteria or mammalian cells. However, the specific activity of SF2/ASF obtained from 293E cells was consistently higher than that from bacteria (data not shown), which may reflect differences in PTMs and/or in purification/renaturation protocols and solubility. The A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> protein, irrespective of its source, was slightly less soluble under splicing conditions, especially at higher concentrations. This may be one of the reasons why the triple-Ala mutant was less active in these cell-free assays, in contrast to the transient-transfection experiments (**Fig. 2.6**).

### **2.2.5 The A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> mutant fails to modulate splicing of endogenous pre-mRNAs**

We recently identified *BIN1* and *MNK2* pre-mRNAs as endogenous targets of SF2/ASF (Karni et al., 2007). To test the effect of the mutant proteins on alternative splicing of these transcripts, we stably expressed cDNAs encoding T7-tagged, wt-SF2/ASF, or the mutant A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> and K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> versions in primary human lung fibroblasts (IMR90) using retroviral transduction. We analyzed total RNA extracted from transductant pools by RT-PCR to detect the isoforms of the above target genes. We observed that overexpression of both wt-SF2/ASF and the K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> mutant promoted the inclusion of exon 12a in *BIN1* and led to an increase in the levels of the *MNK2* alternatively spliced isoform MNK2b; however, the K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> mutant was less active due to its lower level of expression. In contrast, the relative levels of alternatively spliced isoforms of the target genes upon overexpression of the A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> mutant were similar to the empty vector control (**Fig. 2.10**). In contrast to the transient co-transfection experiments, which result in substantial overexpression in the cells that take up the DNA, these experiments involving stable transduction result in only about twofold overexpression of SF2/ASF and the mutants (**Fig. 2.10A**). Because the A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> mutant retains some ability to shuttle, the high levels of overexpression presumably allow enough A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> protein to transit through the nucleus to alter the splicing of  $\beta$ -globin pre-mRNA with a 5'ss duplication (**Fig. 2.3C and 2.6**). In contrast, with more physiological levels of overexpression, the triple-Ala mutant fails to modulate the splicing of target genes (**Fig. 2.10**).

### **2.2.6 The A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> protein is more active than wt-SF2/ASF in enhancing translation in cells**

To test whether the triple-Ala mutant of SF2/ASF can enhance translation, we used previously described reporter constructs, in which either a wild-type or a mutant ESE recognized by SF2/ASF is fused at the 5' end of firefly luciferase



(Sanford et al., 2004). Co-transfection of the reporters, along with wild-type or mutant versions of SF2/ASF and subsequent analysis of luciferase activity, revealed that the A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> protein is twice as active as wt-SF2/ASF in promoting translation in an enhancer-dependent manner (**Fig. 2.11A**). The increase in luciferase activity in the cytoplasm upon overexpression of A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> could be due to three different possibilities: increased translational-stimulatory activity due to mutation, greater accumulation of A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> in the cytoplasm, or increased levels of luciferase mRNA in the cytoplasm due to enhanced export. RT-PCR analysis of luciferase mRNA levels in nuclear and cytoplasmic fractions from the transfected HeLa cells revealed that A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> does not enhance mRNA export (**Fig. 2.11B**). Furthermore, recombinant A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> was equally active as wt-SF2/ASF in enhancing translation of in-vitro-transcribed luciferase reporter in a cell-free assay with translationally competent HeLa extract (**Fig. 2.11C**). We conclude that the apparent increased activity of the A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> mutant in enhancing translation in cells is due to its increased residence time in the cytoplasm.

### **2.2.7 A nuclear-retained version of A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> protein promotes NMD**

Based on the above observations, it seemed plausible that mislocalization of the A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> mutant was responsible for its loss of function in NMD. To test this hypothesis, we fused a nuclear-retention signal (NRS) from the non-shuttling SR protein SC35 to the A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> and K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> versions of SF2/ASF at their respective C-termini. C-terminal fusion of the NRS sequence, in the case of wt-SF2/ASF, results in loss of shuttling activity (Cazalla et al., 2002). Furthermore, this nuclear-retained version of SF2/ASF is active in splicing and NMD (Zhang and Krainer, 2004), but fails to enhance translation in cells (Sanford et al., 2004). Immunofluorescence analysis confirmed that fusion of the NRS sequence to the A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> mutant restored nuclear localization (**Fig. 2.3B**). Furthermore, when assayed for NMD, the NRS version of the A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> mutant was as efficient as wt-SF2/ASF in enhancing NMD, clearly indicating that nuclear localization is essential for this activity (**Fig. 2.12**).

### **2.2.8 Tandem mass spectrometry reveals that the triple Lys mutant is dimethylated**

Surprisingly, our results showed that in all the assays, the K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> mutant had very similar effects as wt-SF2/ASF. To address the possibility that the substituted Lys residues at these positions undergo methylation, we expressed the K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> mutant in 293E cells and analyzed the purified protein by mass spectrometry. MS/MS analysis of one of the peptides corresponding to the region of substitution revealed that the substituted Lys at position 109 was indeed

dimethylated (**Fig. 2.13**). This unexpected finding points towards the possibility that both the methylation state, and the positive charge at these positions, could be responsible for the localization of SF2/ASF (see Discussion).

### **2.2.9 Mutation of the Gly residues in the RGG motifs restores SF2/ASF localization and function**

To address the question of whether charge or methylation or both are involved in controlling localization of SF2/ASF, we mutated the Gly residues surrounding the methyl-arginine residues to Ala as well as Arg at position 111 to Lys, resulting in the SM mutant (**Fig. 2.14**). As a control, we also made the corresponding mutations (Gly to Ala, and Arg<sup>111</sup> to Lys) in the K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> version, giving rise to the KM mutant (**Fig. 2.14**). All the Gly to Ala changes disrupt the RGG/GRG motifs whereas the Arg<sup>111</sup> to Lys substitution disrupts a RGG-related RGR motif (Weiss et al., 2000).

We hypothesized that if methylation is responsible for the mislocalization of SF2/ASF and methylation at these positions is dependent upon RGG motifs, then disrupting the glycines should result in a similar phenotype, as observed for the A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> mutant. The K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> mutant was functionally indistinguishable from SF2/ASF, which earlier suggested that charge, and not methylation was important at these positions, but the unexpected discovery that at least one of the substituted lysines could be methylated brought back the possibility that methylation, or more likely both charge and methylation, are involved in localization control (**Fig. 2.13**). We argued that if methylation of the substituted Lys residue also depended upon KGG motif, which was highly unlikely, the KM mutant would display activity similar to the A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> mutant.

Alternatively, if charge and not methylation is involved in the regulation of localization, then both SM and KM mutants should behave like the wt-SF2/ASF. However, this conclusion would be entirely based upon one very crucial assumption, that the presence of the RGG-type motif is essential for the methylation of the Arg residues, which does not seem to be the case, since methylation by PRMT1 is not restricted to RGG and related motifs (see discussion). Therefore, if SM and KM both show activities similar to wt-SF2/ASF, then we still cannot rule out the involvement of methylation in localization control as both the Arg residues in SM and Lys residues in KM could still be methylated.

When analyzed by IF, both the SM and the KM versions of SF2/ASF showed predominant nuclear localization with residual cytoplasmic staining (**Fig. 2.15**). As expected, when assayed for activity in enhancing NMD, both proteins

showed similar activity as the wt-SF2/ASF (**Fig. 2.16**). As discussed above, these results pointed to the same conclusion that charge and methylation together control localization of SF2/ASF (see discussion).

### **2.2.10 Interaction of SF2/ASF with import/export machineries is not disrupted by methylation or charge**

#### **a. Interaction with TAP/NXF1**

We next attempted to investigate the factors that may influence the localization of SF2/ASF. Since shuttling ability is not affected, the predominant cytoplasmic localization of the A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> mutant may result due to either enhanced export or decreased import. SF2/ASF acts as an adapter for mRNA export via its interaction with TAP/NXF1 and this interaction is dependent upon partial dephosphorylation of the RS domain of SF2/ASF (Huang et al., 2004). We sought to investigate whether the introduced Ala mutations in SF2/ASF disrupt or more likely enhance its interactions with TAP/NXF1, especially, since the A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> protein was found to be predominantly cytoplasmic (**Fig. 2.3C**). We transfected plasmids expressing, either T7-tagged wt-SF2/ASF, or its mutant versions, into HeLa cells and two days later cells were harvested and lysed in a buffer with or without phosphatase inhibitors. The cell lysate without phosphatase inhibitors were further subjected to treatment with calf intestinal phosphatase (CIP) to strip the phosphates en masse from the modified proteins. We then performed immunoprecipitation (IP) using T7 antibody coupled to agarose beads to pull down the T7-tagged proteins and their associated factors from both CIP-treated as well as CIP-untreated cell lysates. As expected, there was no interaction between SF2/ASF or its mutants with TAP/NXF1 when the proteins were phosphorylated (**Fig. 2.17A**). Quite unexpectedly, we did not observe any interaction with SF2/ASF or its mutants even when the proteins were not phosphorylated (**Fig. 2.17B**).

It has been reported previously that TAP/NXF1 interacts with SF2/ASF when it is hypophosphorylated (Huang et al., 2004). It may be possible that in the case where we tried to preserve the different phosphorylation states (hypo- and hyper-) of SF2/ASF using phosphatase inhibitors, the amount of hypophosphorylated protein was low enough in the cell lysates that we were unable to detect any interaction with TAP/NXF1 (**Fig. 2.17A**). In the second case the near complete dephosphorylation of SF2/ASF caused by CIP may have prevented any interaction with TAP/NXF1, as certain level of phosphorylation is required for interaction (**Fig. 2.17B**). We were unable to investigate the SF2/ASF-

TAP/NXF1 interaction in great detail using in vitro pull down assays as we could not purify GST-tagged TAP/NXF1 from bacteria due to technical difficulties.

## **b. Interaction with transportin-SR2**

Transportin-SR and transportin-SR2 proteins belong to the importin-beta family of proteins, which are involved in the nuclear import. SR proteins, including SF2/ASF bound to the newly exported mRNA, are phosphorylated in their RS domain by SRPK1, which causes their release from the mRNA. The phosphorylated SF2/ASF then interacts with the transportin-SR proteins and is subsequently imported back to the nucleus (Huang and Steitz, 2005; Lai et al., 2001).

To test if mutation of the three Arg residues changes the ability of SF2/ASF to interact with Transportin-SR2, we expressed and purified N-terminal GST-tagged transportin-SR2 from bacteria and performed a series of GST-pull down assays with wt-SF2/ASF and its mutant versions (**Fig. 2.18A**). We observed that GST-transportin-SR2, but not GST alone, strongly interacted with the untagged wt-SF2/ASF, as well as its mutants (A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> and K<sub>1</sub>K<sub>2</sub>K<sub>3</sub>), which were purified from bacteria. Further, this interaction was lost in an ATP dependent manner, when the unmodified proteins were incubated in S100 extract prior to the pull-down assay (**Fig. 2.18B**). Incubation of SF2/ASF in the cytoplasmic S100 extract, in the presence of ATP, results in the phosphorylation of the RS domain. We also performed similar experiments with the C-terminal His-tagged and modified versions of SF2/ASF and its mutants purified from 293E cells. We observed that the 293 proteins, with their phosphorylated RS domains, failed to interact with transportin-SR2 (**Fig. 2.19**). Consistent with the previous pull-down assay involving the proteins obtained from bacteria, the 293 proteins interacted strongly when they were treated with calf intestinal phosphatase (CIP) (**Fig. 2.19**), which dephosphorylates almost all the phosphorylated residues (**Fig. 2.13A**). Since all three versions of proteins interacted with transportin-SR2, the predominant cytoplasmic localization of the A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> protein was not due to inhibition of nuclear import mediated by transportin-SR2.

Intriguingly, we observed that SF2/ASF interacts with transportin-SR2 only when it is dephosphorylated, which is in direct opposition to the previous observations where phosphorylation of RS domain of SF2/ASF was required for its interaction with transportin-SR2 (Lai et al., 2001). This discrepancy may be caused due to the fact that GST-transportin-SR2, used in our experiments, was obtained from bacteria, thus lacking any modifications, especially, the phosphorylations of its RS domain. In contrast, Lai and coworkers added GST-

tagged purified SF2/ASF, either phosphorylated or unphosphorylated, to HeLa extracts and showed that only phosphorylated SF2/ASF was able to pull-down the endogenous transportin-SR2.

### **2.3 Discussion**

The coordinated and sequential events that lead to regulated gene expression in eukaryotes are extremely complex, spanning across cellular compartments and requiring numerous protein and RNA factors at various stages and at specific locations. When a single protein factor, such as SF2/ASF, is involved in multiple post-transcriptional events, it is essential that its movements between cellular compartments be tightly regulated. SF2/ASF is a shuttling protein that shows predominant nuclear localization in the steady state (Cáceres et al., 1998). Within the nucleus, SF2/ASF accumulates in nuclear speckles, and its recruitment to active sites of transcription is modulated by phosphorylation of Ser residues in the RS domain and various protein-protein interactions (Misteli, 1999). Phosphorylation of SF2/ASF also modulates its subcellular localization (Koizumi et al., 1999). The RS domain is required for shuttling of SF2/ASF (Cáceres et al., 1998) and contributes to its nuclear localization (Cáceres et al., 1997).

Here we have demonstrated that additional signals, which control the cellular localization of SF2/ASF, are present in the linker between RRM1 and RRM2. The Arg residues in this linker region, in particular R93, R97, and R109, are methylated (Shao-En Ong, 2004) and are important for correct localization, as we found that mutating these residues, simultaneously to Ala, resulted in cytoplasmic, rather than nuclear accumulation. These arginines are part of RGG-type motifs that are typically recognized by PRMT1, the predominant methyltransferase, which accounts for >85% of asymmetric dimethylated Arg residues in mammalian proteins (Bedford and Clarke, 2009; Dreyfuss et al., 1993), including SF2/ASF (Goulet et al., 2007).

The three guanidino nitrogen atoms in the Arg side chain can act as hydrogen bond (H-bond) donors and potentially form five H-bonds, with H-bond acceptors in both RNA and protein molecules, if present in the proper orientation (Borders et al., 1994; Calnan et al., 1991). In the case of RNA, the acceptors may include oxygen atoms in both phosphate and ribose 2' OH, or groups on the nitrogenous bases, potentially resulting in a network of H-bond interactions with the Arg side chain. Such interactions are not possible with Lys, which has a single terminal amino group. For example, a short Arg-rich basic peptide from

HIV-1 Tat protein binds specifically to TAR RNA, and Lys substitution results in loss of binding and transactivation (Calnan et al., 1991).

SF2/ASF has two RRM domains through which it binds RNA. Though the structures of individual RRM domains of SF2/ASF have been reported (He, 2005; Ngo et al., 2008; Tintaru et al., 2007a; Tintaru et al., 2007b), the complete structure of SF2/ASF or its two RRM domains in complex with RNA is lacking. In cases where the structure of two-RRM containing proteins is known, such as UP1, sex-lethal, and nucleolin, the inter-RRM linkers are disordered when not bound to RNA. In contrast, in complexes with both RRM domains bound to RNA or single-stranded DNA, the interdomain linker cooperates with the RRM domains in binding the nucleic acid, providing increased affinity and specificity, though the nature of the linker-mediated interactions varies in each case (Maris et al., 2005). In our study, the triple-Lys mutant was functionally indistinguishable from wt-SF2/ASF, and the triple-Ala mutant was as active as wt-SF2/ASF in promoting translation *in vitro*, suggesting that RNA binding was not affected by these substitutions.

Methylation of the two terminal amino groups in the Arg side chain does not alter the positive charge, but increases the hydrophobicity, makes the side chain bulkier, and most importantly, blocks any potential H-bond formation. This could provide a potential means of regulating protein-RNA as well as protein-protein interactions, such that methylation of Arg residues abolishes some interactions based on H-bonding, while leaving electrostatic interactions unaffected. For example, Pro-rich motifs in Sam68 interact with both SH3 and WW domains present in interacting partners; methylation of RG repeats that flank the Pro-rich motifs reduce the binding of Sam68 to the SH3 domains of p59<sup>lyn</sup> and phospholipase C $\gamma$ -1, without affecting the binding to the WW domain of FBP30 (Bedford et al., 2000).

Because the motif recognized by PRMT1 is not strictly limited to RGG and related sequences (Wooderchak et al., 2008), it is possible that other Arg residues of SF2/ASF are also modified by PRMT1. One precedent is methylation of R3 in histone H4 by PRMT1, in which the methylated Arg is not part of an Arg/Gly-rich region (Wang et al., 2001). As in the case of histones, the methylation state in the SF2/ASF linker region may control various protein-protein interactions, either directly or by influencing other modifications of SF2/ASF, such as phosphorylation of the RS domain. Such regulation via Arg methylation, in conjunction with phosphorylation-dephosphorylation cycles of the RS domain, could play a role in localization and trafficking of SF2/ASF between cellular compartments. A precedent for this type of cross-talk has been observed for Npl3p in budding yeast, in which hypermethylation of Arg residues in the C-

terminal RGG/RS domain interferes with phosphorylation of neighboring Ser residues by the kinase Sky1p, which in turn blocks the import of Npl3p to the nucleus (Yun and Fu, 2000). Phosphorylation of the RS domain is essential for the interaction of SF2/ASF with transportin-SR2 (TRN-SR2), which acts as a receptor for the nuclear import of SR proteins (Lai et al., 2001). Furthermore, the phosphorylation state of SF2/ASF influences its activity as an adaptor protein for TAP-mediated mRNA export (Huang et al., 2004; Lai and Tarn, 2004). However, when we analyzed the interactions of wt-SF2/ASF and its mutant versions (A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> and K<sub>1</sub>K<sub>2</sub>K<sub>3</sub>) with either TAP or TRN-SR2, we observed that all versions of SF2/ASF interacted similarly with these two proteins, arguing against the possibility that these three Arg residues, or their methylation states, affect these interactions.

In addition to changing the properties of binding sites and affecting other modifications, methylated arginines are also directly involved in protein-protein interactions. The Tudor domain of SMN protein directly interacts with symmetric dimethyl Arg residues in proteins with this modification (Côté and Richard, 2005). Tudor domains have highly conserved Trp, Tyr, and Phe residues, and it has been suggested that methyl groups on arginines may increase their affinity for aromatic rings by increasing stacking and hydrophobic interactions through the expanded surface area of the guanidino group, which ensures greater delocalization of positive charge (Hughes and Waters, 2006). Proteins and their respective domains that may bind to asymmetric dimethyl Arg residues in SF2/ASF, and other proteins with the same modification, are yet to be discovered.

The triple-Ala mutant of SF2/ASF was unable to enhance NMD and failed to modulate alternative splicing of endogenous target pre-mRNAs when modestly overexpressed. Furthermore, due to the accumulation of the protein in the cytoplasm, A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> was more efficient at enhancing translation of a luciferase reporter when compared to wt-SF2/ASF. When we restored the nuclear localization of A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> protein by C-terminal fusion of a nuclear-retention sequence from the non-shuttling SR protein SC35 (Cazalla et al., 2002), the resulting A<sub>1</sub>A<sub>2</sub>A<sub>3</sub>-NRS1 protein was as effective in promoting NMD as wt-SF2/ASF, demonstrating that the effects observed with A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> were due to mislocalization. However, we note that at high levels of overexpression, A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> was as effective as SF2/ASF in promoting selection of the proximal 5' splice sites of the  $\beta$ -globin model pre-mRNA with duplicated 5' splice sites. The A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> mutant retained the ability to shuttle, such that in the context of high overexpression, there was enough protein in the nucleus to modulate alternative splicing, whereas at lower

levels of expression, A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> failed to alter splicing of target genes. Yet it is interesting that even at high levels of overexpression and with the ability to shuttle, A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> was inactive in promoting NMD. The ability of SF2/ASF to regulate splicing in a concentration-dependent manner is well documented (Cartegni et al., 2002), whereas the precise mechanisms underlying the effect of SF2/ASF in NMD (Zhang and Krainer, 2004) remain largely unknown, although, it was shown recently, that transient overexpression of SF2/ASF promotes an increase in the efficiency of the pioneer round of translation (Sato et al., 2008). Nonetheless, we have conclusively confirmed the initial observation (Zhang and Krainer, 2004) that nuclear localization of SF2/ASF is essential for its activity in enhancing NMD.

Though the detailed mechanisms through which SF2/ASF promotes transformation are understood only in part (Karni et al., 2007), improper cellular localization of SF2/ASF may be one of the ways through which it exerts its oncogenic activity, especially due to its regulatory roles in multiple post-transcriptional events, both in the nucleus and cytoplasm. For example, an increase in the cytoplasmic levels of another splicing factor, hnRNP A2, in sputum samples serves as a powerful predictor of lung cancer almost a year prior to clinical detection (Tockman et al., 1994), suggesting that such a change may be a prerequisite for the transformation of lung epithelial cells.

Intriguingly, the K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> mutant was functionally indistinguishable from wt-SF2/ASF, pointing towards the importance of charge at these positions, as opposed to the methylation state, per se. Studies of the yeast SR-like protein, Npl3p (Shen et al., 1998; Yun and Fu, 2000), other shuttling RNA-binding proteins in mammals, such as some hnRNPs (Herrmann et al., 2004; Liu and Dreyfuss, 1995; Nichols et al., 2000) and the transcription factors TAF15 and EWS (Araya et al., 2005; Jobert et al., 2008), have shown that Arg methylation of RGG motifs is involved in controlling the nucleo-cytoplasmic distribution of these proteins. However, no obvious defect in the nuclear export or localization of Npl3p was observed upon substitution of Arg residues in the RGG motifs to Lys (McBride et al., 2005; Xu and Henry, 2004). A similar observation was made with the yeast protein Hrp1p, which resembles hnRNPs and also contains RGG motifs; but in this case, additional substitutions of the Arg residues to Glu or Gln resulted in cytoplasmic localization, most likely through impaired nuclear import (Xu and Henry, 2004). In mammals, the Ewing's Sarcoma oncoprotein (EWS) interacts with components of both the transcriptional machinery, via its N-terminal transactivation domain, as well as with the splicing machinery via its C-terminal domain, and is thought to couple transcription and splicing (Yang et al., 2000).



EWS is methylated at two RGG-boxes, RGG2 and RGG3, which are required for its nuclear localization (Araya et al., 2005; Belyanskaya et al., 2001). However, only the substitution of Arg residues in the RGG boxes to Ala, and not to Lys, altered the transcriptional activity of EWS (Alex and Lee, 2005).

Our observation, that at least one of the substituted lysines (R109K) in the K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> version of SF2/ASF was dimethylated, suggests that both the methylation state and the positive charge at these positions may contribute to the localization of SF2/ASF. Lys methylation in histone tails, and its role in epigenetic regulation, has been extensively studied. In mammals, two distinct set of Lys methyltransferases have been identified: the SET domain family and the DOT1 family (Smith and Denu, 2009). More recently, Lys methylation was also described in non-histone proteins, including p53, TAF7, and TAF10 (Lan and Shi, 2009). From our own analysis, we know that naturally occurring Lys residues in SF2/ASF are methylated, but it was totally unexpected for us to find methylation of one of the substituted Lys residues. Also, in the absence of quantitative mass-spectrometry data, we do not know what fraction of the total K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> protein expressed in 293E cells underwent this modification, so it is possible that dimethylation of the substituted Lys109 is present in only a small fraction of the protein. Unfortunately, further experiments involving the SM and the KM mutants, which showed activities similar to wt-SF2/ASF, did not help solve the conundrum of charge vs. methylation, due to a possibility that even in the absence of surrounding glycines, the Arg residues in SM protein could be methylated by PRMT1 (see above). Unless the presence of methyl-Arg residues in SM protein is ruled out by direct observation (mass-spec), the possibility that methylation is involved in regulation of localization still exists.

Here, we used site-specific mutagenesis to study the specific effect of Arg methylation on the various activities of SF2/ASF. Studies of other proteins with Arg methylation that used a similar approach, likewise found that substituting the Arg residues to Ala invariably resulted in mutant proteins with loss of activity (Alex and Lee, 2005). In cases in which substitution to Lys was included as a control for positive charge, the mutant protein was functionally indistinguishable from the wild-type protein (Alex and Lee, 2005; McBride et al., 2005; Xu and Henry, 2004), except when the substituted Arg is directly involved in RNA binding (Calnan et al., 1991). As both the methylation state and the positive charge of a particular Arg residue may be involved in regulating the activity of any given protein, it is conceivable that certain activities of the protein that depend on electrostatic intra- or intermolecular interactions through the positively charged Arg residues will not be affected by Lys substitution. In at least one instance in

which Lys was able to functionally substitute for Arg, it was inferred that charge, and not Arg methylation (of EWS), is critical for activity (Alex and Lee, 2005). However, in light of our observation that a substituted Lys can itself undergo methylation, it appears that both the methylation state and the positive charge of a given Arg residue can control the activities of a protein, in this case SF2/ASF.

In summary, our findings underscore the importance of proper localization of SF2/ASF for its activity in key nuclear and cytoplasmic processes. We have further identified the signals that control the distribution of SF2/ASF between nucleus and cytoplasm, and also generated mutants of SF2/ASF with partial loss of function. Such mutants will prove useful in future studies to dissect the mechanisms through which SF2/ASF affects various normal cellular processes, as well as oncogenic transformation.

## **2.4 Future Perspectives**

The proteomic analysis to thoroughly identify all the possible modifications in SF2/ASF as well as other SR proteins, and the subsequent functional analysis characterizing the effects of these modifications, remains incomplete. Our proteomic analysis was done with SF2/ASF that was transiently overexpressed in 293E cells. The PTMs identified may not entirely reflect the true physiological state of the protein in the cell, as overexpression might lead to over or under modification or might be a consequence of mis-localization. Purification of native SF2/ASF from primary, immortal, and cancer cells and subsequent comparative proteomic analysis should enable us to thoroughly analyze all the modifications present on endogenous SF2/ASF and map them in a more complete and physiologically relevant manner. At the same time, it might also provide us with clues regarding specific PTM signatures or mutations of SF2/ASF that are potentially associated with cancer.

Nonetheless, in our initial analysis we found a number of novel modifications on SF2/ASF, most notable among them Tyr phosphorylation, Lys acetylation, and Lys methylation, which were not known to occur on SR proteins. The list of novel modifications also included additional sites of Ser phosphorylation other than those found in the RS domain, as well as many new methylated Arg residues. The crucial role played by Tyr kinases in regulation of gene expression and cancer via signal transduction is well known. It is quite likely that some of these Tyr kinases regulate the activity of SF2/ASF by changing its properties in a dynamic way by modifying one or more Tyrosines found to be phosphorylated in our study. The functional characterization of these modified Tyrosines may provide a better understanding of the mechanism of action of

SF2/ASF in various cellular processes. Simultaneously, the characterization of the upstream events and the specific Tyr kinases that are involved might provide insights into how SF2/ASF may function in response to various cellular stimuli, with obvious implications in SF2/ASF-mediated transformation.

Another such modification that has potential for widespread implications in SF2/ASF's functions, as well as its role in cancer, is Lys acetylation/methylation. The central role played by Lys acetylation and methylation of histone tails in the epigenetic regulation of gene expression through chromatin remodeling, especially in cellular differentiation during development, has been studied extensively. Further, perturbations in these modification events are thought to be one of the major ways by which a normal differentiated cell undergoes 'dedifferentiation', leading to its transformation and subsequent metastasis. This is achieved again by chromatin remodeling events that result in an increase in expression of oncogenes and repression of tumor suppressors, due to newly formed euchromatin and heterochromatin regions, respectively. In a recent study, SF2/ASF was found to be associated with interphase chromatin, which is released from chromosomes during metaphase and reassociates after completion of telophase, in a His3 Ser10 phosphorylation-dependent manner (Loomis et al., 2009). Additionally, SF2/ASF was shown to be essential for the release of heterochromatin protein 1 (HP1) from the metaphase chromosomes via direct interaction. These activities of SF2/ASF appear to be related to its role in maintaining genomic stability by destabilizing the formation of R-loops during transcription, thus preventing double-strand breaks in DNA (Li and Manley, 2005; Lin et al., 2008). The interaction of SF2/ASF with chromatin places it in the direct vicinity of histone acetylases and deacetylases (HATs and HDACs), which are responsible for modifying the histone tails during chromatin remodeling.

Lastly, it would be extremely interesting to investigate if there is any cross-talk between various modifications present on SF2/ASF, and their roles in modulating the activities of SF2/ASF. Precedents for such cross-talk already exists in the SR-like yeast protein Npl3p (see above).

## **2.5 Materials and Methods**

### **2.5.1 Plasmids**

The plasmid pSP64-H $\beta$  $\Delta$ 6 (Krainer et al., 1984) and the duplicated 5' ss version pSP64-5'ssD-wt (Zhang and Krainer, 2004) have been described. The plasmid, pUC $\beta$ 128SV, has a full-length  $\beta$ -globin gene driven by its own promoter and the SV40 enhancer (Cáceres et al., 1994). The mutant version pUC $\beta$ 128SV-

T39 and the wt and mutant duplicated 5'ss versions, pUC-5'ssD-wt and pUC-5'SSD-mt have been described (Zhang and Krainer, 2004). The mammalian expression plasmid, pCGT-SF2/ASF, encodes a T7-tagged version of human SF2/ASF (Cáceres et al., 1997). We used site-directed mutagenesis to create the various T7-tagged pCGT versions of single, double, or triple missense mutants, in which the Arg residues at positions 93, 97, and 109 were changed to either Ala or Lys (**Table 4**). Further, using pCGT-A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> and pCGT-K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> as templates, the pCGT-SM and the pCGT-KM were also created by site directed mutagenesis, in which the Gly residues were changed to Ala. We replaced a ClaI-BstBI fragment in the pCGT-SF2/ASF-NRS1 plasmid (Cazalla et al., 2002) by the corresponding fragment from pCGT-A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> and pCGT-K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> to create pCGT-A<sub>1</sub>A<sub>2</sub>A<sub>3</sub>-NRS1 and pCGT-K<sub>1</sub>K<sub>2</sub>K<sub>3</sub>-NRS1, respectively. We replaced the SacII-BamHI fragment in pET9c-SF2/ASF(R/S) (Krainer et al., 1991) by the corresponding fragment from pCGT-A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> and pCGT-K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> to create pET9c-A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> and pET9c-K<sub>1</sub>K<sub>2</sub>K<sub>3</sub>, respectively, for bacterial expression and purification. We added a C-terminal 6xHis-tag to the cDNAs of wt-SF2/ASF, A<sub>1</sub>A<sub>2</sub>A<sub>3</sub>, and K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> by PCR using their pCGT versions as templates. We then individually subcloned these amplicons as XbaI-BamHI fragments in the mammalian expression vector pTT3 (Durocher et al., 2002). Similarly, we inserted the amplified cDNAs of A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> and K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> with an N-terminal T7-tag as BamHI-EcoRI fragments in the retroviral vectors pBABE-puro and pWZL-hygro. Both pBABE-T7SF2/ASF and pWZL-T7SF2/ASF have been described (Karni et al., 2007). The plasmids, pLCS-EDA and pLCS-EDA<sup>mt</sup>, used for translation assays involving transient transfection, as well as the constructs used for in vitro translation, pLuc-3xEDA and pLuc-3xEDA<sup>mt</sup> were generous gifts from Javier Cáceres, and have been described (Sanford et al., 2004). pGEX-transportin-SR2 was generously provided by Rui-Ming Xu.

## 2.5.2 Expression and purification of recombinant proteins

We purified recombinant untagged SF2/ASF as described (Krainer et al., 1991). Briefly, we expressed wild-type or mutant SF2/ASF cDNAs in *E. coli* using pET9c-SF2/ASF(R/S)-type plasmids, followed by a series of purification steps including cesium-chloride gradient centrifugation to remove nucleic acids, selective precipitation in low salt, urea denaturation, ion exchange chromatography, and subsequent refolding of the protein. We purified C-terminal 6xHis tagged SF2/ASF expressed in human 293-EBNA1 (293E) cells grown in suspension using a variation of a published method (Durocher et al., 2002). We transfected 293E cells with wild-type or mutant pTT3-SF2/ASF-His and harvested the cells 3-4 days after transfection. We resuspended the cells in lysis

buffer (50 mM Tris-HCl pH 8.0, 1 M NaCl, 10 mM imidazole, 20 mM  $\beta$ -mercaptoethanol, 0.1% [v/v] Triton-X100) and centrifuged them after sonication to remove insoluble material. We then precipitated most of the cellular proteins using ammonium sulfate at 40% saturation at 4 °C, and removed them by centrifugation. We then purified the His-tagged proteins from the supernatant by nickel-agarose affinity chromatography.

For the purification of GST-transportin-SR2, we grew bacterial cultures of *E. coli* strain, BL21, transformed with pGEX-transportin-SR2, at 37 °C till OD<sub>600</sub> reached 0.6, following which we induced the cultures with 1 mM IPTG and transferred to 21 °C incubator and grew for additional 8 h. subsequently, we harvested the cells and lysed them using sonication in GST-lysis buffer (50 mM Tris pH 8, 150 mM NaCl, 1mM EDTA, 10% glycerol, and 1% triton). We removed the insoluble cellular debris using centrifugation. The soluble GST-tagged transportin-SR2 in the clarified cell extract were then pulled-down using their affinity towards Glutathion agarose beads (GE Healthcare) using the batch method, and further, eluted the proteins bound to the beads using GST-lysis buffer supplemented with 10 mM reduced Glutathione. The purified protein was then dialyzed in Buffer D (50 mM Hepes pH 8, 100 mM KCl, 1 mM EDTA, 1 mM DTT, and 20% glycerol), aliquoted, snap frozen in liquid nitrogen and stored in -80 °C till further use.

### **2.5.3 Cell culture, and transient and stable expression of proteins**

We grew HeLa and IMR90 cells in Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% fetal bovine serum (Invitrogen) at 37 °C with 5% CO<sub>2</sub>. We grew 293E suspension cultures in Joklik's modified Eagle's medium (JMEM) (Invitrogen) supplemented with 5% calf serum (Invitrogen). We transfected HeLa cells with Fugene 6 reagent (Roche Diagnostics), as per the manufacturer's recommendations, unless otherwise stated. We transfected 293E cells using linear polyethylenimine (PEI 25000; Polysciences Inc.). We prepared a transfection mix consisting of 1  $\mu$ g plasmid DNA, 2  $\mu$ g PEI, and 50  $\mu$ l of JMEM for every 1 ml of cell culture to be transfected. We then incubated the mix at room temperature for 10-15 min and added it to the suspension culture, which we maintained at a density of  $2 \times 10^5$  cells/ml. For stable expression, we generated retroviruses by co-transfecting pBABE or pWZL constructs of wild-type or mutant SF2/ASF, along with a vesicular stomatitis G protein (VSV-G) expression construct, into Phoenix cells (a gift from Scott Lowe), an ecotropic packaging cell line expressing essential viral proteins (Hemann et al., 2004). We used the resulting viruses to infect primary human lung fibroblasts (IMR90) to establish

stable cell lines overexpressing wild-type or mutant SF2/ASF, as described (Karni et al., 2007).

#### **2.5.4 Indirect immunofluorescence (IF)**

We determined the subcellular localization of transiently expressed wild-type or mutant versions of SF2/ASF by indirect immunofluorescence using the N-terminal T7 epitope tag, as described (Cazalla et al., 2002). Briefly, we transfected HeLa cells grown on coverslips in 6-well plates with 1  $\mu$ g of wild-type or mutant pCGT-SF2/ASF. We fixed the cells with 4% para-formaldehyde in phosphate-buffered saline (PBS) 36 h after transfection for 30 min, followed by incubation for 5 min in 0.2% (v/v) Triton X-100 to permeabilize the cells. We then incubated the cells with anti-T7 monoclonal antibody (1:1000; Novagen) for 1 h, washed with PBS, and incubated with Alexa Fluor 488-conjugated goat anti-mouse IgG antibody (1:1000; Invitrogen) for 1 h. We mounted the coverslips on slides using ProLong Gold Antifade Reagent with DAPI (Invitrogen) and imaged the cells with a fluorescent microscope (Axioskop 2 Plus, Carl Zeiss) equipped with a camera (AxioCam, Carl Zeiss).

#### **2.5.5 Shuttling assays**

We transfected HeLa cells grown on coverslips with T7-tagged wild-type or A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> mutant SF2/ASF, as above. Twenty four hours after transfection, we incubated the cells in tissue-culture medium supplemented with 10 ng/mL Leptomycin B (Calbiochem) for an additional 12 h. We then fixed and permeabilized the cells, and analyzed them by IF, as above.

The heterokaryon assay to assess the shuttling activity of nuclear proteins has been described (Cáceres et al., 1998; Piñol-Roma and Dreyfuss, 1992). In short, to achieve high transfection efficiency, we electroporated 1 X 10<sup>6</sup> HeLa cells with 4  $\mu$ g of either pCGT7-SF2/ASF or pCGT-K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> and plated them on coverslips; we included pCGT-SC35 as a negative control. We then co-incubated the cells with an excess of untransfected mouse NIH 3T3 cells for 3 h in media supplemented with 50  $\mu$ g/ml cycloheximide; we then increased the concentration of cycloheximide to 100  $\mu$ g/ml, with additional incubation for 30 min. We fused the cells using polyethylene glycol (PEG), and further incubated the resulting heterokaryons for 2 h in medium containing 100  $\mu$ g/ml cycloheximide. We visualized the cells by IF using T7 antibody, as above. We distinguished mouse nuclei (NIH3T3) from human nuclei (HeLa) in the heterokaryons by the former's characteristic punctate DAPI-staining pattern.

### **2.5.6 Dual luciferase assay and cellular fractionation**

We transfected 40-50% confluent HeLa cells grown in 6-well plates with 0.4 µg pLCS reporter (Sanford et al., 2004), 0.2 µg TK-Renilla Luciferase (Promega), and 1.0 µg of wild-type or mutant pCGT-SF2/ASF. We lysed the cells using passive lysis buffer (Promega) 48 h post-transfection, and assayed the levels of Firefly and Renilla luciferase using a Dual Luciferase Assay Kit (Promega), measuring the luminescence with a Monolight 2010 (Analytical Luminescence Lab). For cellular fractionation, we harvested the cells 48 h after transfection by trypsinization, resuspended them in PBS, and used 1/10<sup>th</sup> of the cell suspension for the luciferase assay. We collected the remaining cells by centrifugation, resuspended them in a hypotonic buffer (10 mM Hepes pH 8.0, 10 mM KCl, 1 mM MgCl<sub>2</sub>, 1 mM dithiothreitol, 0.5% [v/v] Nonidet P-40), incubated them on ice for 5 min, and lysed them by gentle pipetting. We centrifuged the lysate with intact nuclei and collected the supernatant as the cytosolic fraction. We washed the nuclear pellet with PBS and extracted the nuclear proteins by resuspending the pellet in Laemmli buffer. We analyzed both nuclear and cytoplasmic fractions by Western blotting.

### **2.5.7 In vitro translation**

In vitro translation reactions consisted of 200 ng of in vitro-transcribed pLuc reporter mRNA, 200 ng of recombinant wild-type or mutant SF2/ASF, and translation-competent HeLa cell extracts (Bergamini et al., 2000), incubated for 30 min at 37 °C, as described (Sanford et al., 2004). We stopped the reactions by dilution with 100 µl of passive lysis buffer (Promega) and measured luminescence as above, using Luciferase-activating reagent (Promega).

### **2.5.8 In vitro splicing**

We prepared radiolabeled pre-mRNAs by in vitro transcription of pSP64-H $\beta$  $\Delta$ 6 and pSP64-5' ssD-wt plasmids linearized with BamHI, with SP6 RNA polymerase, in the presence of  $\alpha$ -<sup>32</sup>P-UTP. We incubated the pre-mRNAs under splicing conditions in HeLa nuclear or cytoplasmic S100 extracts, as described (Mayeda and Krainer, 1999a, b). We extracted the RNAs and analyzed them by denaturing PAGE and phosphorimage analysis on a FUJIFILM FLA-5100 instrument (Fuji Medical Systems USA, Inc.). We quantified the band intensities using Multi Gauge software Version 2.3 (FUJIFILM).

### **2.5.9 RNase protection assay (RPA)**

We co-transfected 40-50% confluent HeLa cells, grown in 10-cm plates, with 0.2 µg GFP plasmid, 0.5 µg wild-type or mutant pCGT-SF2/ASF plasmid, and 2.5 µg wt-β-globin or mutant (T-39) NMD reporter plasmids. To test both NMD and 5'ss selection, we co-transfected 0.2 µg GFP plasmid, 0.7 µg SF2/ASF plasmid, and 2.1 µg 5'ssD-wt or 5'ssD-mt plasmid. We isolated total RNA 36 h after transfection by ultracentrifugation of the cell lysates layered on a 5.7 M CsCl cushion at 20 °C, and carried out RPA as described (Huang and Carmichael, 1996; Zhang and Krainer, 2004). We analyzed the radioactive protected fragments by denaturing PAGE and phosphorimaging, as above.

### **2.5.10 RT-PCR**

We extracted total RNA from IMR90 cells stably expressing wild-type or mutant SF2/ASF using Trizol reagent (Invitrogen) and reverse transcribed 2 µg from each RNA sample using Superscript-II (Invitrogen). We amplified the cDNAs corresponding to transcripts of endogenous target genes of SF2/ASF by PCR using Taq-gold polymerase (Invitrogen) with specific primers and conditions described previously (Karni et al., 2007).

### **2.5.11 Antibodies and Western Blotting**

We lysed cells in Laemmli sample buffer and analyzed the proteins by Western blotting with various primary antibodies specific for SF2/ASF (mAb AK96), SRp55 (mAb 9-1-56), Myc (mAb 9E10), T7 (Novagen), β-Catenin (Transduction Labs), TAP/NXF1 (Abcam) and Caspase-3 (a gift from Yuri Lazebnik). The secondary antibody was goat anti-mouse IgG conjugated to Alexa Fluor 532 (Invitrogen). After washing, we scanned the nitrocellulose membranes (Whatman) on a phosphorimager, and quantified fluorescence band intensities as above.

### **2.5.12 Mass spectrometry**

Wild-type or mutant versions of SF2/ASF were purified from 293E cells as described above, and analyzed by mass spectrometry. Briefly, we separated the purified proteins by SDS-PAGE and stained them with Gelcode Blue (Pierce). We digested and processed the bands for analysis as described (Allemand et al., 2007). We analyzed the tryptic peptides either by LC-MS/MS on an LTQ mass spectrometer (Thermo) or by targeted fragmentation on a MALDI TOF/TOF spectrometer (Applied Biosystems) (Bish et al., 2008). We analyzed the LC-MS/MS runs using the X! Tandem search engine (Craig et al., 2004) using the



following parameters: variable modifications include mono-, di-, and trimethylation of Lys and mono- or dimethylation of Arg residues. For targeted fragmentation, we first analyzed the precursor mass list using either Profound or Peptidemap software packages (<http://prowl.rockefeller.edu/>). We then selected peaks with precursor masses consistent with methylated peptides for fragmentation and analyzed the resulting spectra using MS-Product (<http://prospector.ucsf.edu/>). We resolved ambiguities by comparison with theoretical spectra derived from all possible methylation patterns.

### **2.5.13 Immunoprecipitation and GST pull-down**

For immunoprecipitation (IP), we transfected 1 µg of pCGT-SF2/ASF or its mutants into 40-50% confluent HeLa cells grown in 10 cm petridishes and forty eight hours later we harvested the cells by trypsinization and lysed them by mild sonication in IP buffer (Tris-HCl pH 7.4, 150 mM NaCl, 1 mM DTT, 1 mM EDTA, Complete EDTA free protease inhibitors [Roche diagnostics]) with or without phosphatase inhibitors (2 mM activated Na<sub>3</sub>VO<sub>4</sub>, 20 mM NaF and 25 mM β-glycerophosphate) and quantified for protein content using Micro-BSA kit (Pierce). Protein complexes associated with T7-tagged SF2/ASF or its mutants were pulled-down using anti-T7-agarose beads (Novagen) from the lysates containing 500 µg of total protein in 1 ml volume. The pulled-down complexes were resolved using SDS-PAGE followed by Western blot analysis.

For GST pull-down experiments, we added 2 µg of both the bait (GST-transportin-SR2) and the test proteins (different versions of SF2/ASF and its mutants) to 1 ml of Buffer D (50 mM Hepes pH 8, 400 mM KCl, 1 mM EDTA, 1 mM DDT, and 20% glycerol) and mixed by rotating gently at 4 °C for an hour. We then pulled-down the interacting proteins using Glutathione agarose beads (GE Healthcare), separated the proteins by SDS-PAGE and subsequently, analyzed them by Western blotting. Wherever necessary the proteins were incubated with S100 extract with or without ATP in buffer D with low salt (100 mM KCl) to phosphorylate the bacterially obtained proteins before conducting the pull-downs.

CIP treatment in both cases, wherever necessary, was carried out prior to the pull-downs with 20 units of CIP (New England Biolabs).

## **2.6 Acknowledgements**

We thank Eric Allemand for actively participating in helpful discussions while designing experiments and for the help extended during heterokaryon assays. We thank Zuo Zhang for providing β-globin constructs for RPA, as well

as 9g8 and SF2/ASF- $\Delta$ RS proteins purified from 293 cells. We thank Rotem Karni for providing essential reagents necessary for creating stable IMR90 cells, and Michael Myers for performing the mass-spec experiments. We thank Rui-Ming Xu for helpful discussions and providing bacterially purified SF2/ASF protein that was coexpressed with Clk/STY. We thank Javier Cáceres for providing the pLCS- and pLuc- constructs, Yuri Lazebnik for Caspase-3 antibody, Scott Lowe for Phoenix cells, and Philip Smith for providing mammalian SRp30c. We thank Rolf Sternglanz for helpful discussions, and Lisa Manche for critical reading of the manuscript. This work was supported by grant CA13106 from the National Cancer Institute.

## 2.7 Tables

**Table 1. Modified residues identified in 9g8**

<b>Modification</b>	<b>Residues</b>	<b>Positions</b>
Monomethylation	Arg	156, 195, 214, 230
Monomethylation	Lys	185
Dimethylation	Arg	3, 6, 93, 172, 180, 222
Phosphorylation	Ser	32, 89, 128, 130 157, 159, 165, 167, 198, 200, 202, 204, 208, 209
Phosphorylation	Tyr	33, 34
Ubiquitination	Lys	K12, K212, k219

**Table 2. Modified residues identified in SRp20**

<b>Modification</b>	<b>Residues</b>	<b>Positions</b>
Monomethylation	Arg	43, 57, 163
Dimethylation	Lys	11, 85
Dimethylation	Arg	28, 86, 159
Phosphorylation	Ser	81, 87, 148
Phosphorylation	Tyr	104
Ubiquitination	Lys	146

**Table 3. Modified residues identified in SC35**

<b>Modification</b>	<b>Residues</b>	<b>Positions</b>
Monomethylation	Arg	5, 178, 190
Monomethylation	Lys	36, 52, 197
Dimethylation	Arg	146, 150, 188, 205
Phosphorylation	Ser	128, 145, 147, 191, 196
Ubiquitination	Lys	148, 170, 172, 207

**Table 4: Primers used for site-directed mutagenesis**

<b>Primer</b>	<b>Sequence</b>
MUTSF2AF1	GTGGAGTTTCCTCGAAGCGGCGCTGGAACAGGCCGAG GCGGCGGC
MUTSF2AF2	CGAAGCGGCCGTGGAACAGGCGCAGGCCGGCGGCGGG GGTGGAGGT
MUTSF2KF1	GTGGAGTTTCCTCGAAGCGGCAAAGGAACAGGCCGAG GCGGCGGC
MUTSF2KF2	CGAAGCGGCCGTGGAACAGGCAAAGGCCGGCGGCGGG GGTGGAGGT
MUTSF2KF3	GGTGGAGGTGGCGGAGCTCCCAAAGGTCGCTATGGCC CCCATCC
MUTSF2AR1	GCCGCCGCCTCGGCCTGTTCCAGCGCCGCTTCGAGGA AACTCCAC
MUTSF2AR2	ACCTCCACCCCGCCGCGCCTGCGCCTGTTCCACGG CCGCTTCG
MUTSF2KR1	GCCGCCGCCTCGGCCTGTTCTTTGCCGCTTCGAGGAA ACTCCAC
MUTSF2KR2	ACCTCCACCCCGCCGCGCCTTTGCCTGTTCCACGGC CGCTTCG
MUTSF2KR3	GGATGGGGGGCCATAGCGACCTTTGGGAGCTCCGCCA CCTCCACC

MUTSF2F1	GTGGAGTTTCCTCGAAGCGGCGCTGGAACAGGCGCAG GCGGCGGCGGGGGTGGAGGT
MUTSF2F2	GGTGGAGGTGGCGGAGCTCCC GCAGGTCGCTATGGCC CCCATCC
MUTSF2R1	ACCTCCACCCCGCCGCGCCTGCGCCTGTTCCAGCG CCGCTTCGAGGAAACTCCAC
MUTSF2R2	GGATGGGGGGCCATAGCGACCTGCGGGAGCTCCGCCA CCTCCACC

## 2.8 Figures and Figure legends

**Figure 2.1 Mass-spec analysis reveals that SF2/ASF is extensively modified at multiple residues.** We purified SF2/ASF expressed in 293E cells, separated them by SDS-PAGE after either with or without the treatment with CIP and stained them with GelCode blue. Next, we excised individual bands, digested them with trypsin, and analyzed them by MALDI-TOF. We also analyzed the unmodified SF2/ASF obtained from bacteria, purified SF2/ASF- $\Delta$ RS from 293 cells, and the purified SF2/ASF coexpressed with Clk/STY in *E. coli* the same way for comparison. The different modifications identified after the extensive comparative analyses were mapped and their individual positions are represented on the polypeptide sequence of SF2/ASF. The orange bars show the boundaries of conserved RNP1 and RNP2 motifs in both the RRM of SF2ASF (Birney et al., 1993). The secondary structure elements based on the available NMR structures are also depicted. Arrows:  $\beta$ -strands, cylinders:  $\alpha$ -helix (He, 2005; Tintaru et al., 2007a; Tintaru et al., 2007b).



**Figure 2.2 Types of arginine methylation in mammals.** The mechanism of action of the three different types of protein arginine methyl-transferases (PRMTs) is shown. All three types, PRMT-I, II, and III catalyze the formation of  $\omega$ -monomethylarginine ( $\omega$ -MMA) from arginine in an unmodified polypeptide chain. Further, PRMT type-I enzymes convert  $\omega$ -MMA to  $\omega$ -asymmetricdimethylarginine ( $\omega$ -ADMA), whereas PRMT type-II convert  $\omega$ -MMA to  $\omega$ -symmetricdimethylarginine ( $\omega$ -SDMA). Type-III PRMTs only form  $\omega$ -MMA (Bedford and Clarke, 2009). The added methyl groups are depicted in red. Another class of enzymes, protein arginine demethylases (PADs), can only use  $\omega$ -MMA as substrate and convert them to citrulline by releasing methylamine. SAM: s-adenosylmethionine, SAH: S-adenosylhomocysteine.

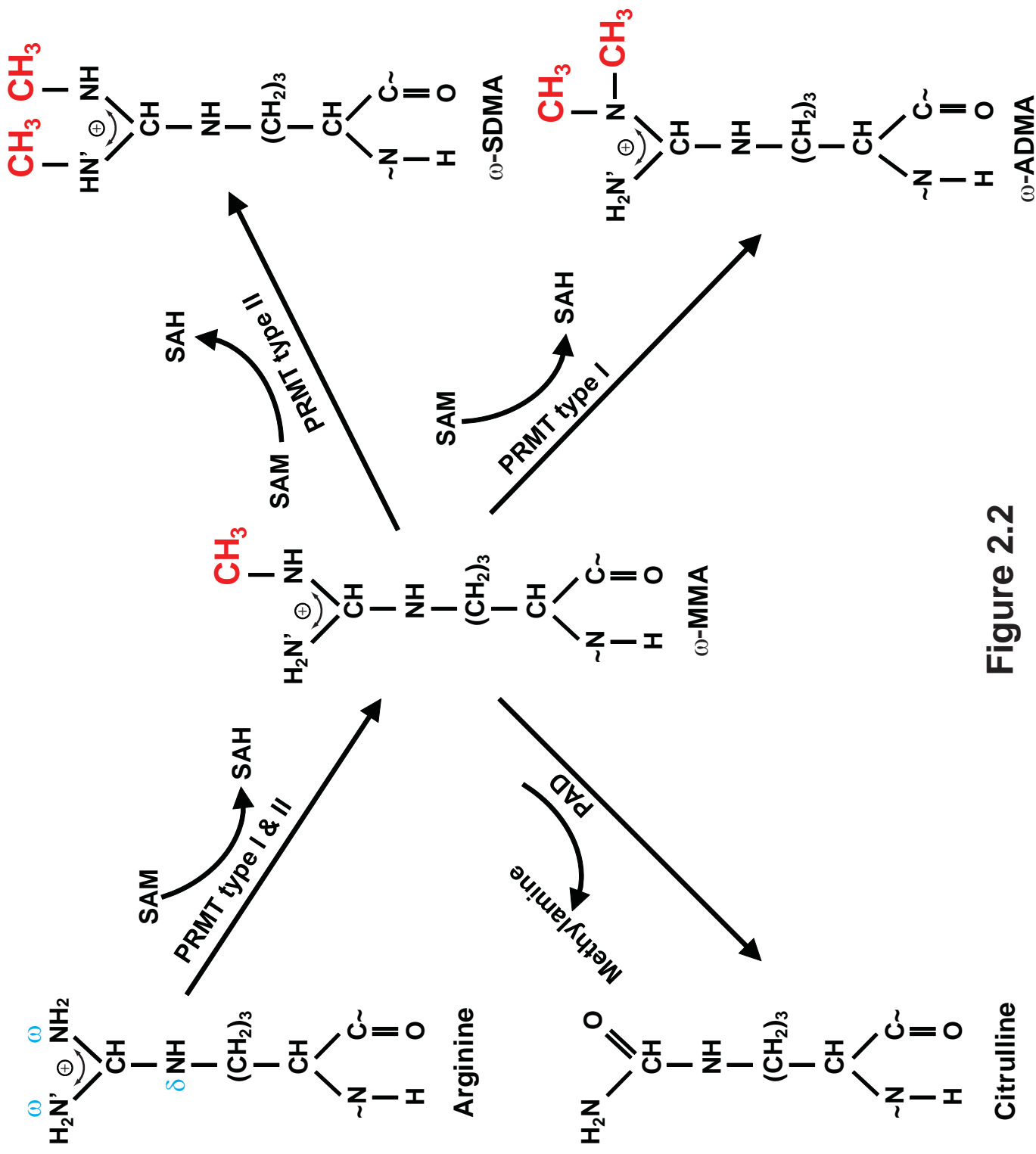
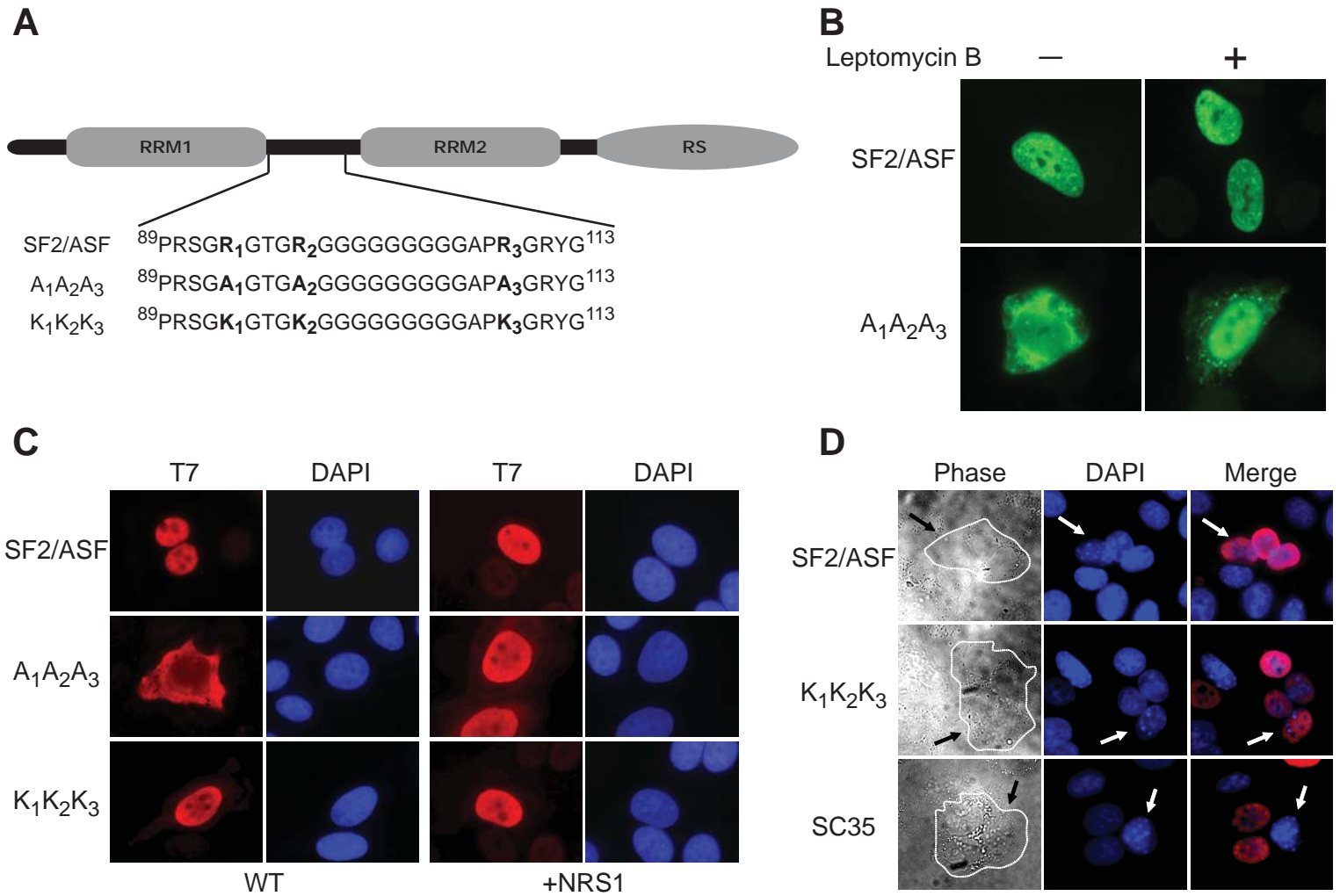


Figure 2.2

Citrulline

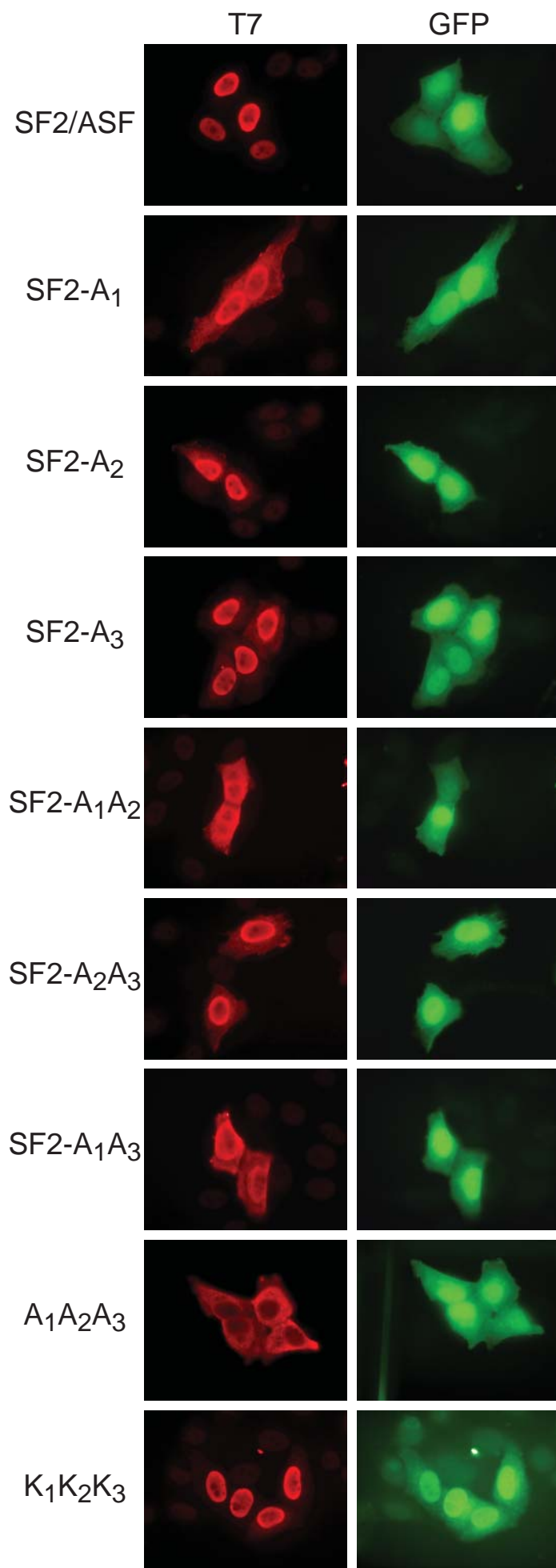


**Figure 2.3 Localization and shuttling activity of wild-type and mutant SF2/ASF.** (A) Modular structure of SF2/ASF and mutants in the inter-RRM linker region. The three Arg residues in bold are methylated, and we mutated each one to Ala or Lys. (B) Indirect immunofluorescence of HeLa cells transfected with wild-type or mutant T7-tagged SF2/ASF cDNA, with or without fusion to a nuclear-retention signal peptide from SC35 (NRS1) (Cazalla et al., 2002). (C) Indirect immunofluorescence of HeLa cells transfected with T7-tagged SF2/ASF or A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> mutant, with or without prior treatment with 10 ng/ml leptomycin B for 12 h. (D) Heterokaryon assay (Cáceres et al., 1998; Piñol-Roma and Dreyfuss, 1992). We transfected HeLa cells with T7-tagged SF2/ASF, K<sub>1</sub>K<sub>2</sub>K<sub>3</sub>, or SC35 as a control. After 24 h, we treated the cells with cycloheximide, fused them with mouse NIH 3T3 cells using polyethylene glycol, and incubated them for 2 h. After fixation, we determined the localization of the transiently expressed proteins in the heterokaryons by indirect immunofluorescence. The panels on the right column show the merged images of T7 and DAPI signals. The arrows indicate mouse nuclei within human-mouse heterokaryons, distinguished from human nuclei by their characteristic DAPI staining (middle column). The panels on the left column show phase-contrast images of the same heterokaryons.



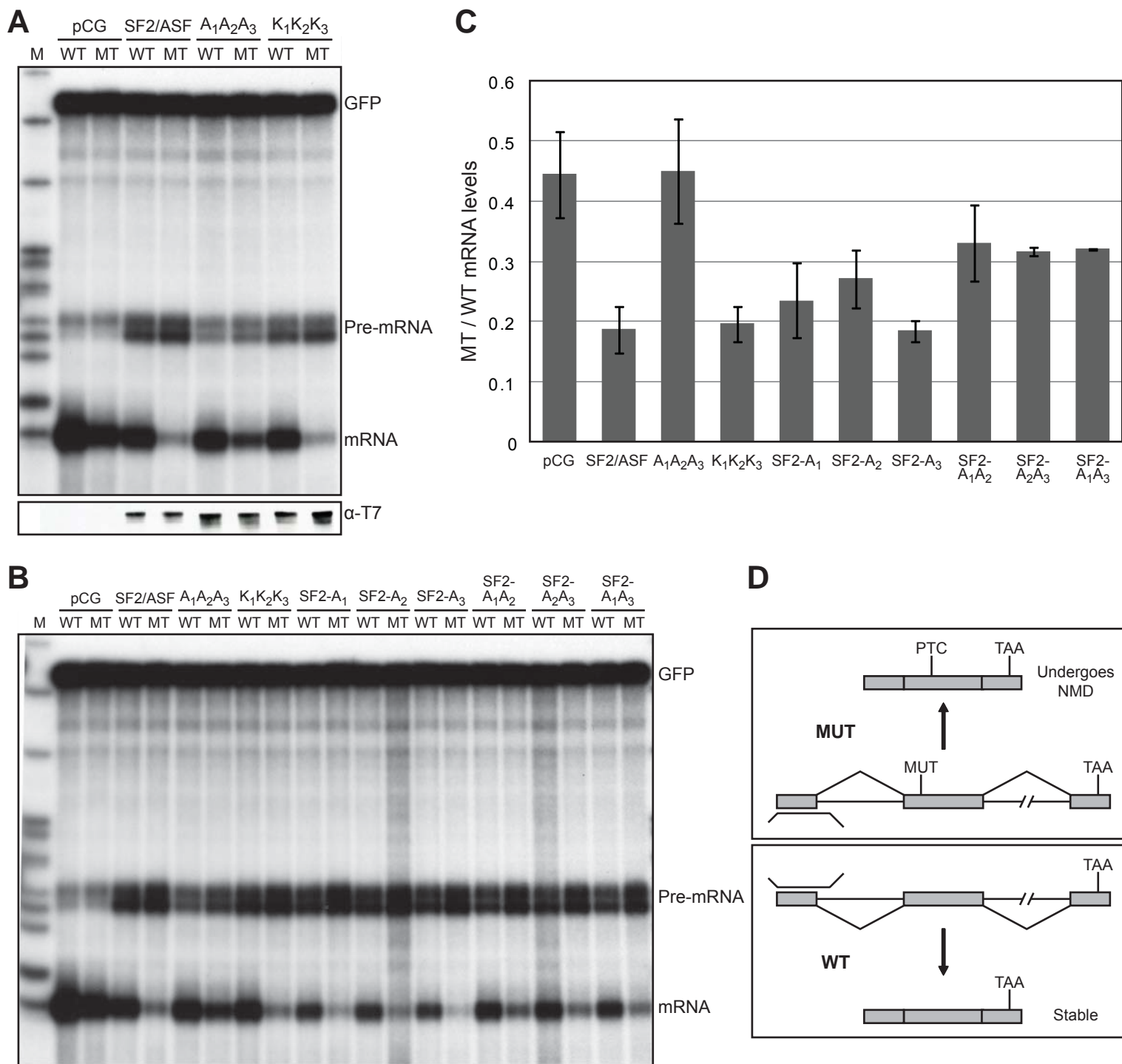
**Figure 2.3**

**Figure 2.4 Localization of single or double Ala mutants of R93, R97, and R109 of SF2/ASF.** The panels show indirect immunofluorescence of HeLa cells co-transfected with GFP and wild-type or mutant T7-tagged SF2/ASF.



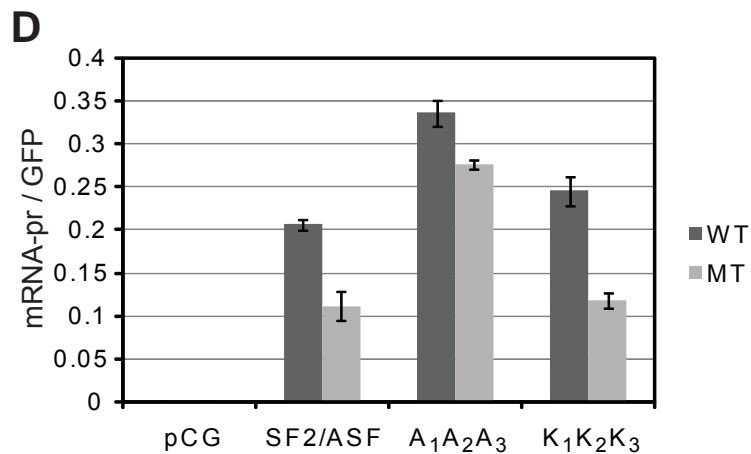
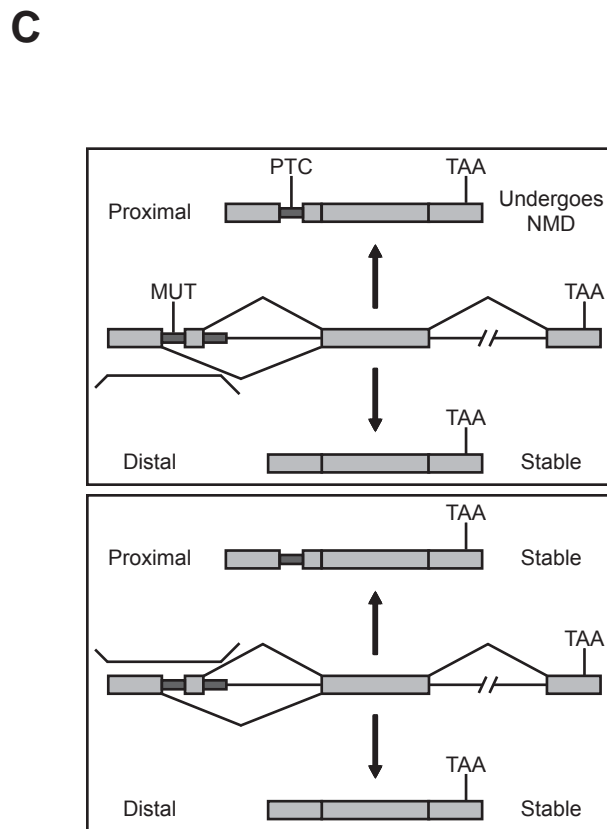
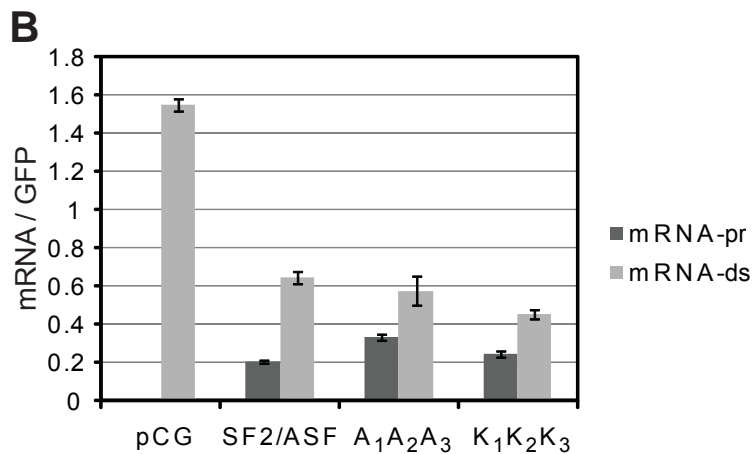
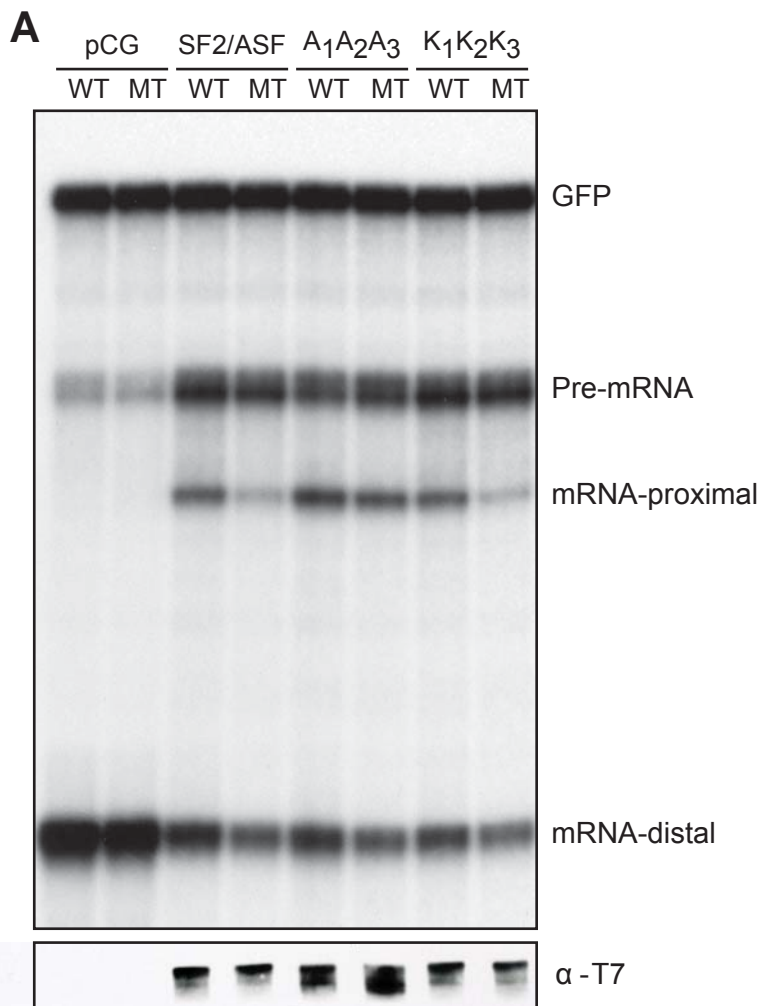
**Figure 2.4**

**Figure 2.5 Effect of overexpression of wild-type or mutant SF2/ASF on NMD.** (A) We co-transfected a human  $\beta$ -globin gene (WT) or a mutant (MT) version with an in-frame PTC, with wild-type or mutant versions of SF2/ASF into HeLa cells. We extracted total RNA 36 h after transfection and analyzed it by RPA. We also co-transfected a GFP-expressing plasmid as an internal reference, and used  $\beta$ -globin and GFP probes simultaneously for RPA. Western blot with T7 antibody in the bottom panel shows expression of each tagged protein. (B) Effect of single or double Ala mutants of SF2/ASF on sensitivity to a PTC upon overexpression. We performed transient co-transfection and RPA as in (A). (C) Quantification of phosphorimager signals from multiple experiments as in (A) and (B). We normalized the individual mRNA intensities to the corresponding GFP intensities in each lane. We plotted the ratios of normalized MT mRNA to WT mRNA intensities to show the relative NMD enhancement upon overexpression of each protein.  $n = 2-5$ ; error bars = standard deviation (SD). (D) Schematic representation of the WT and MT  $\beta$ -globin reporters and the RPA probe.



**Figure 2.5**

**Figure 2.6 Effect of overexpression of wild-type or mutant SF2/ASF on alternative 5'ss selection and NMD.** (A) A co-transfected  $\beta$ -globin minigene (WT) has duplicated 5'ss, and can be spliced via either the proximal or the distal 5'ss; the mutant (MT) version has an in-frame PTC, such that the proximal mRNA is susceptible to NMD (Zhang and Krainer, 2004). We analyzed total RNA recovered from HeLa cells after transient co-transfection with wild-type or mutant versions of SF2/ASF by RPA. We co-transfected GFP plasmid as a reference, and used  $\beta$ -globin and GFP probes simultaneously. The bottom panel shows the expression of each tagged protein analyzed by western blotting. (B) Quantification of phosphorimager signals from multiple experiments as in (A). The plot shows individual WT proximal and WT distal-mRNA intensities normalized to the corresponding GFP intensities in each lane.  $n = 2-5$ ; error bars = SD. All three versions of SF2/ASF promote use of the proximal 5'ss. (C) Schematic representation of the WT and MT  $\beta$ -globin reporters with duplicated 5'ss, and the RPA probe. (D) Same as (B). We plotted the normalized intensities of both WT and MT proximal-mRNA levels to show the NMD enhancement upon overexpression of the depicted proteins. The wild-type and  $K_1K_2K_3$  versions promoted NMD, whereas the  $A_1A_2A_3$  version was inactive.



**Figure 2.6**



**Figure 2.7 Effect of single- or double-Ala mutants of SF2/ASF on alternative 5'ss selection and sensitivity to a PTC upon overexpression.** We cotransfected  $\beta$ -globin and GFP probes simultaneously into HeLa cells as in Fig. 2.6 and performed RPA with GFP signal as a reference. All versions of SF2/ASF promote use of the proximal 5'ss. Individual Ala substitutions promote NMD as efficiently as the wild-type or K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> version; the double-Ala substitutions in every possible combination have intermediate effects when compared to the triple mutant.

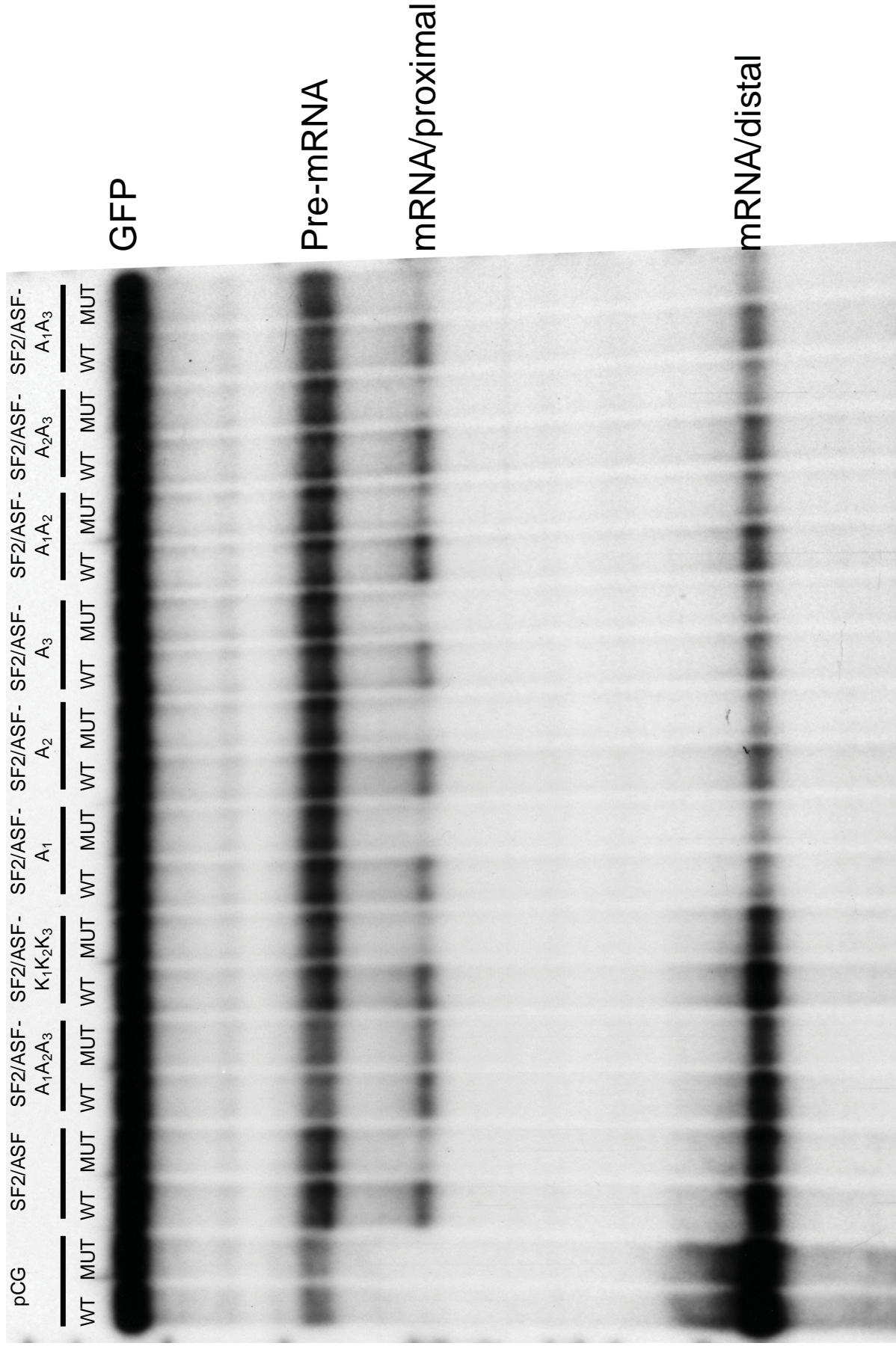
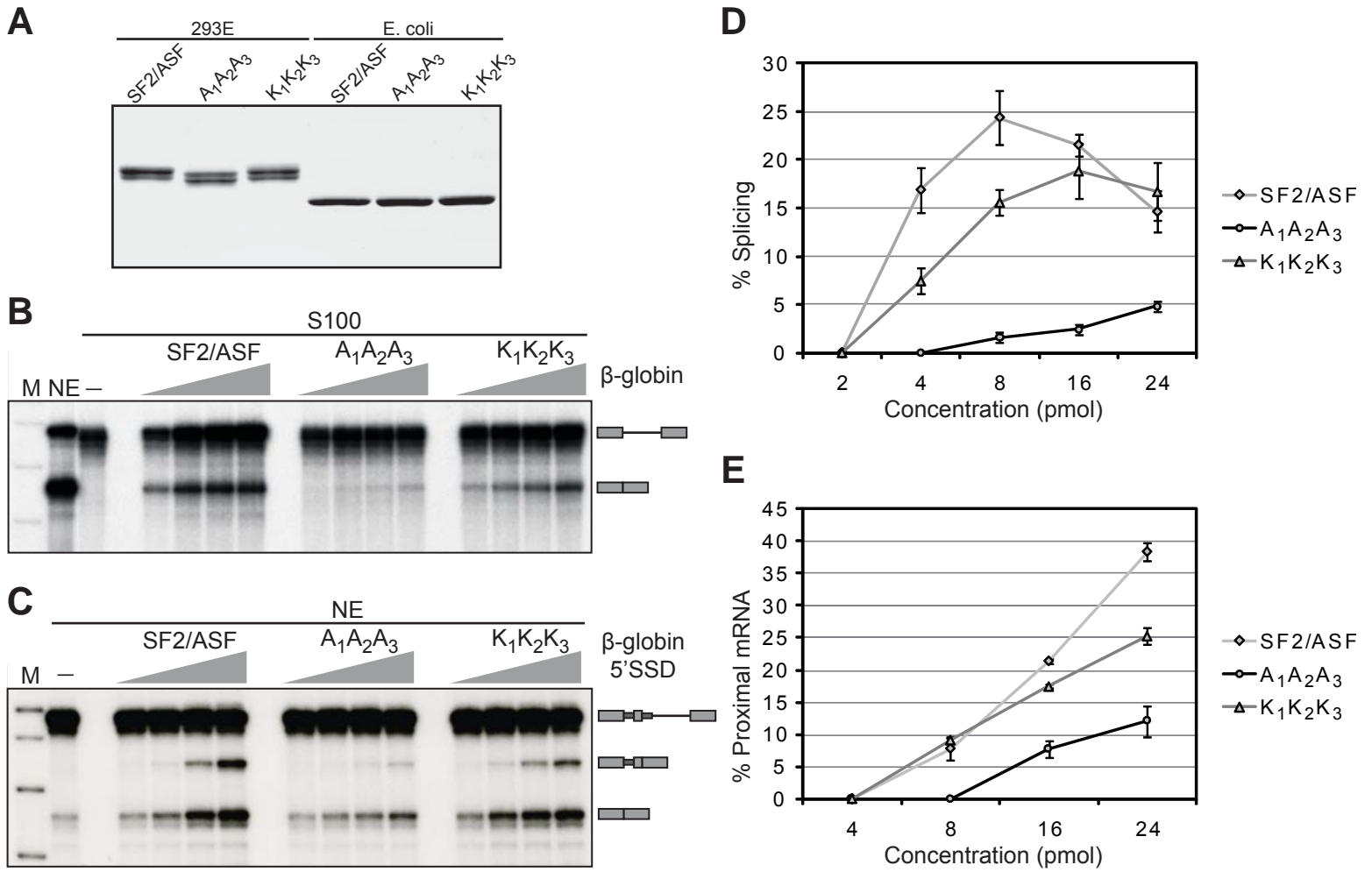


Figure 2.7

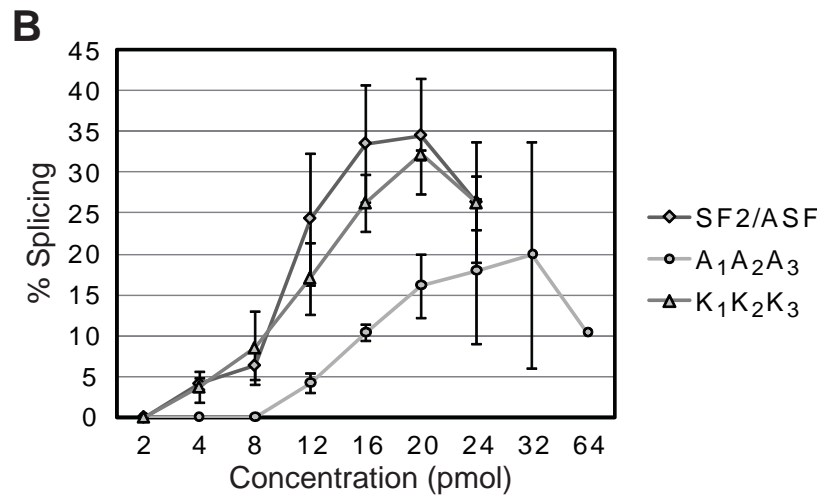
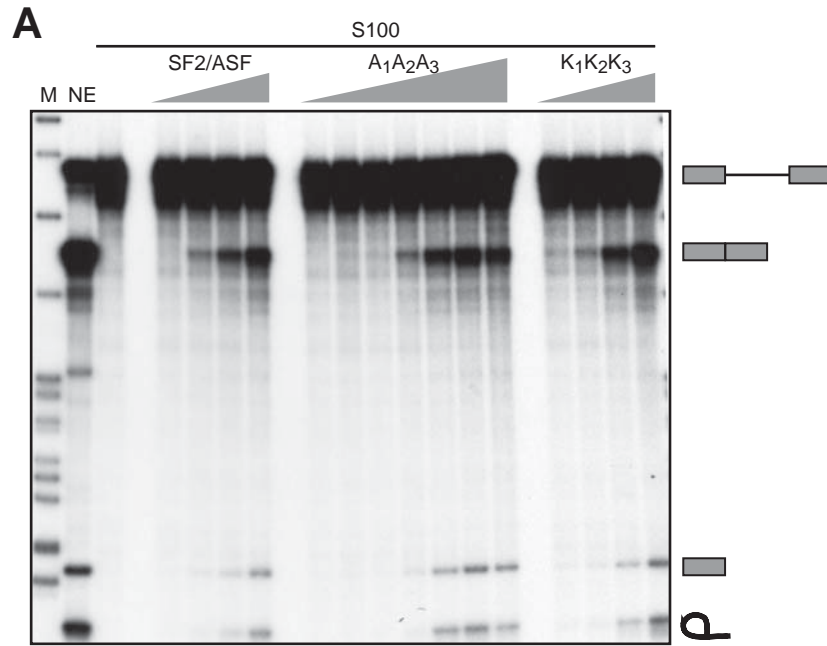
**Figure 2.8 In vitro splicing activity of wild-type and mutant SF2/ASF.**

(A) Coomassie stain of SDS-polyacrylamide gel showing the different versions of recombinant SF2/ASF purified from either 293E cells or *E. coli*. (B) S100-complementation assay to test the general splicing activity of wild-type or mutant SF2/ASF purified from 293E cells, using a radiolabeled  $\beta$ -globin minigene pre-mRNA. We added increasing amounts of each protein (4, 8, 16, 24 pmol) to otherwise identical reactions. The mobilities of pre-mRNA and mRNA are indicated on the right. M: molecular-weight marker; NE: splicing in nuclear extract. (C) In vitro splicing assay to test the alternative splicing activity of wild-type or mutant SF2/ASF purified from 293E cells, using a model  $\beta$ -globin pre-mRNA with a duplicated 5'ss derivative (Zhang and Krainer, 2004). The positions of the pre-mRNA and the two spliced mRNA isoforms are indicated on the right. We supplemented the splicing reactions in HeLa nuclear extract with increasing amounts of each protein (4, 8, 16, 24 pmol). (D) Graph showing quantification of phosphorimager signals from multiple experiments as in (B). We normalized the mRNA intensities to the corresponding pre-mRNA intensities in each lane, and plotted them as the percentage of splicing as a function of the amount of protein added.  $n = 2-3$ ; error bars = SD. (E) Graph showing quantification of phosphorimager signals from multiple experiments as in (C). We divided the proximal-mRNA intensities by the sum of proximal and distal mRNA intensities in each lane, and plotted them as the percentage of proximal mRNA as a function of the amount of protein added.  $n = 2$ ; error bars = SD.



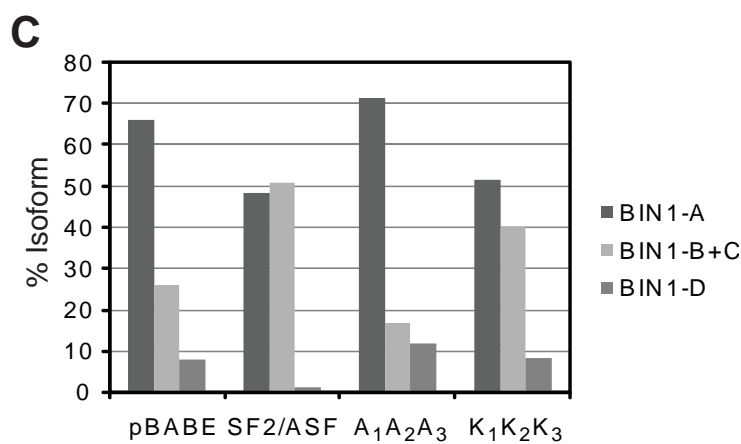
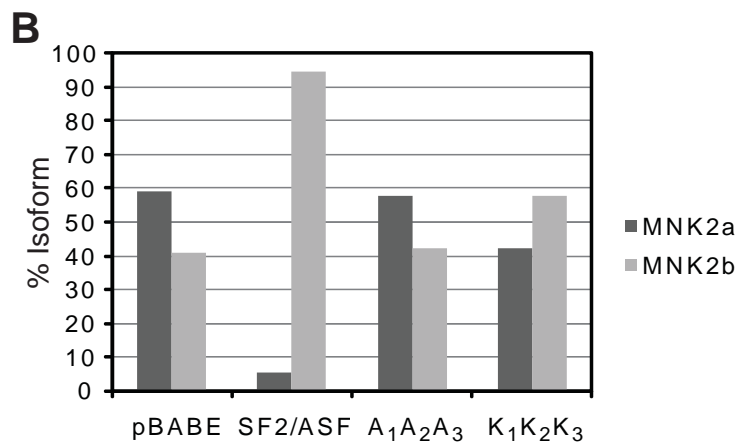
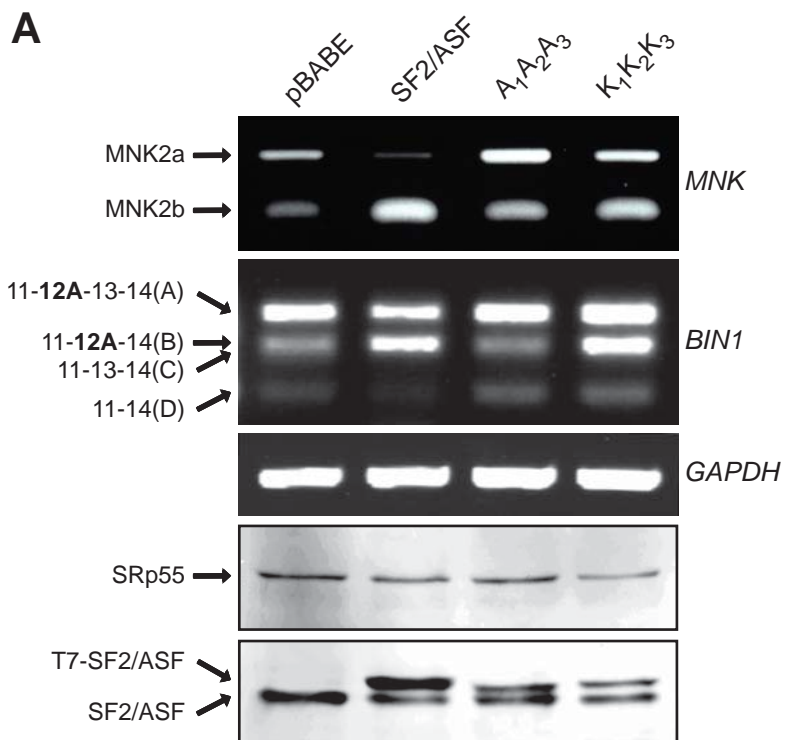
**Figure 2.8**

**Figure 2.9 In vitro splicing activity of unmodified wild-type or mutant SF2/ASF.** (A) S100-complementation assay to test the general splicing activity of recombinant wild-type or mutant SF2/ASF purified from *E. coli*, using radiolabeled  $\beta$ -globin pre-mRNA. Pre-mRNA, splicing intermediates, and products are depicted on the right. M: molecular weight markers; NE: splicing in nuclear extract. The A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> mutant was less active than the wild-type or K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> versions. (B) Quantification of phosphorimager signals from several experiments as in (A). We normalized the mRNA intensities to the corresponding pre-mRNA intensities in each lane, and plotted the percentage of splicing as a function of the amount of recombinant protein added (2, 4, 8, 12, 16, 20, 24, 32, 64 pmol). n = 1-5; error bars = SD.



**Figure 2.9**

**Figure 2.10 Effects of wild-type or mutant SF2/ASF on alternative splicing of endogenous transcripts in IMR90 cells.** (A) We extracted total RNA and protein from IMR90 cells stably overexpressing the indicated proteins after retroviral transduction (Karni et al., 2007). We analyzed total RNA by RT-PCR and detected the expressed T7-tagged proteins and endogenous SRp55 and SF2/ASF by Western blotting. (B) & (C) Graphs showing the levels of different mRNA isoforms of *MNK2* and *BIN1*, respectively, after quantification of the corresponding lanes in (A).



**Figure 2.10**



**Figure 2.11 Effect of wild-type or mutant SF2/ASF on ESE-dependent stimulation of translation.** (A) Diagram of the translation reporter with an in-frame fibronectin EDA wild-type or mutant ESE upstream of the Firefly luciferase ORF (Sanford et al., 2004) (top panel). We co-transfected HeLa cells with the reporter, wild-type or mutant T7-tagged SF2/ASF cDNA, and a Renilla luciferase reporter as a control for transfection efficiency. We harvested the cells after 48 h and analyzed them by the Dual Luciferase Reaction (DLR) assay. The bottom panel shows the ratios of Firefly to Renilla luciferase activity (bottom panel). n = 4; error bars = SD. (B) Subcellular distribution of the luciferase reporter mRNAs upon overexpression of wild-type or mutant SF2/ASF. We prepared nuclear and cytosolic fractions from aliquots of the transfected cells used for the DLR assay in (A), and measured the levels of the Firefly reporter mRNA by RT-PCR, as well as endogenous GAPDH mRNA as a reference (bottom two panels). Western blots show the levels of nuclear c-Myc protein and cytosolic Caspase-3 protein to verify proper fractionation (top two panels). (C) Effect of wild-type or mutant SF2/ASF on stimulation of translation in vitro. Diagram of the in vitro reporter (Sanford et al., 2004) (top panel). We incubated in vitro transcribed reporter mRNAs (200 ng) with three copies of wild-type or mutant EDA ESE in a HeLa cell translation extract with or without 200 ng of recombinant wild-type or mutant SF2/ASF at 37 °C for 30 min. The histogram shows the luciferase activity for each reaction.

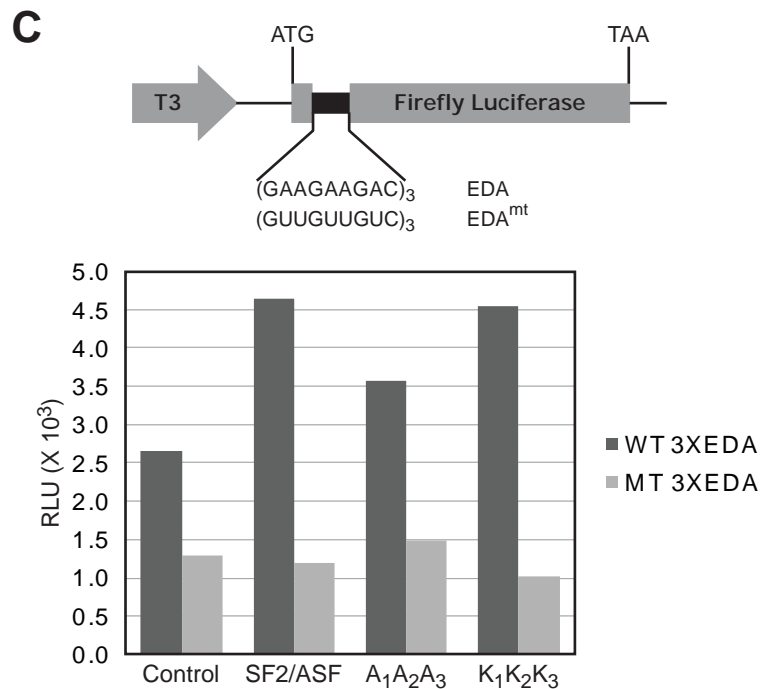
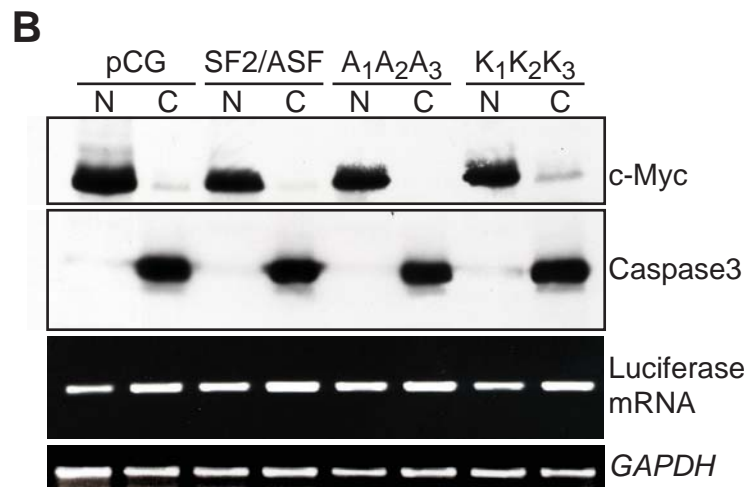
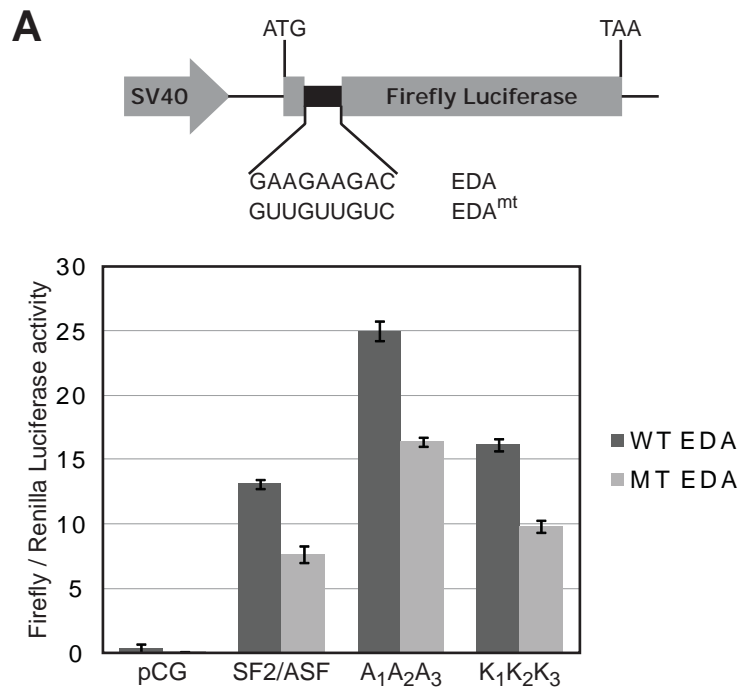
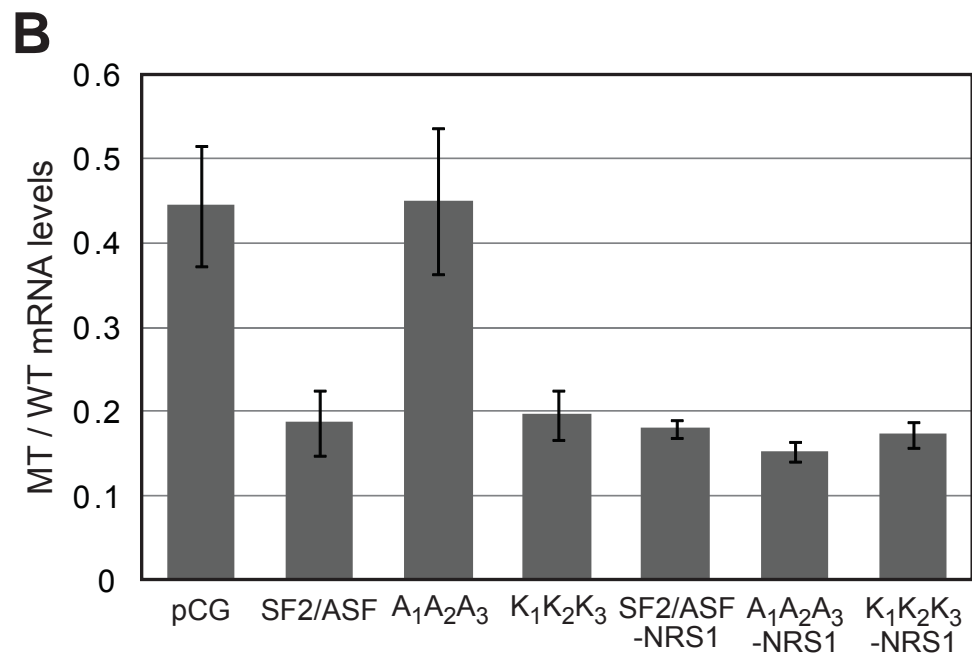
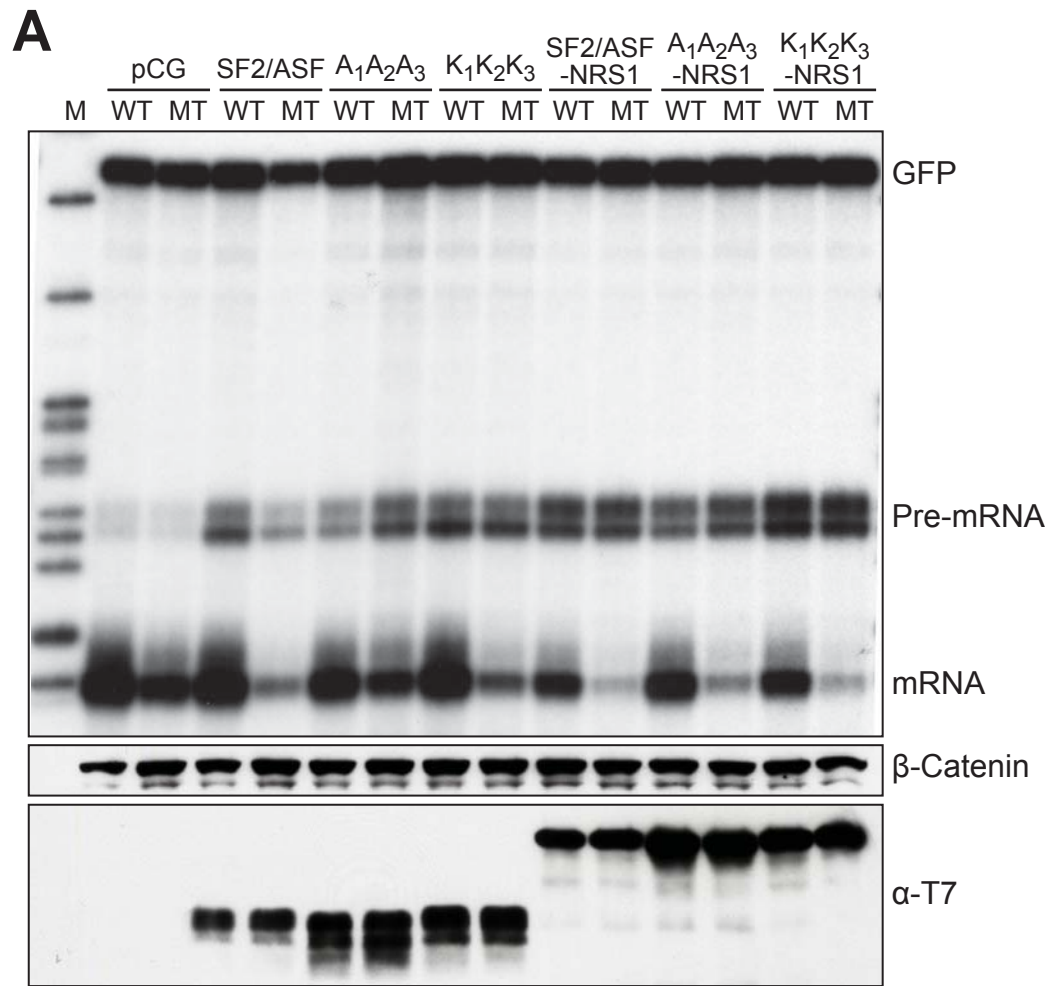


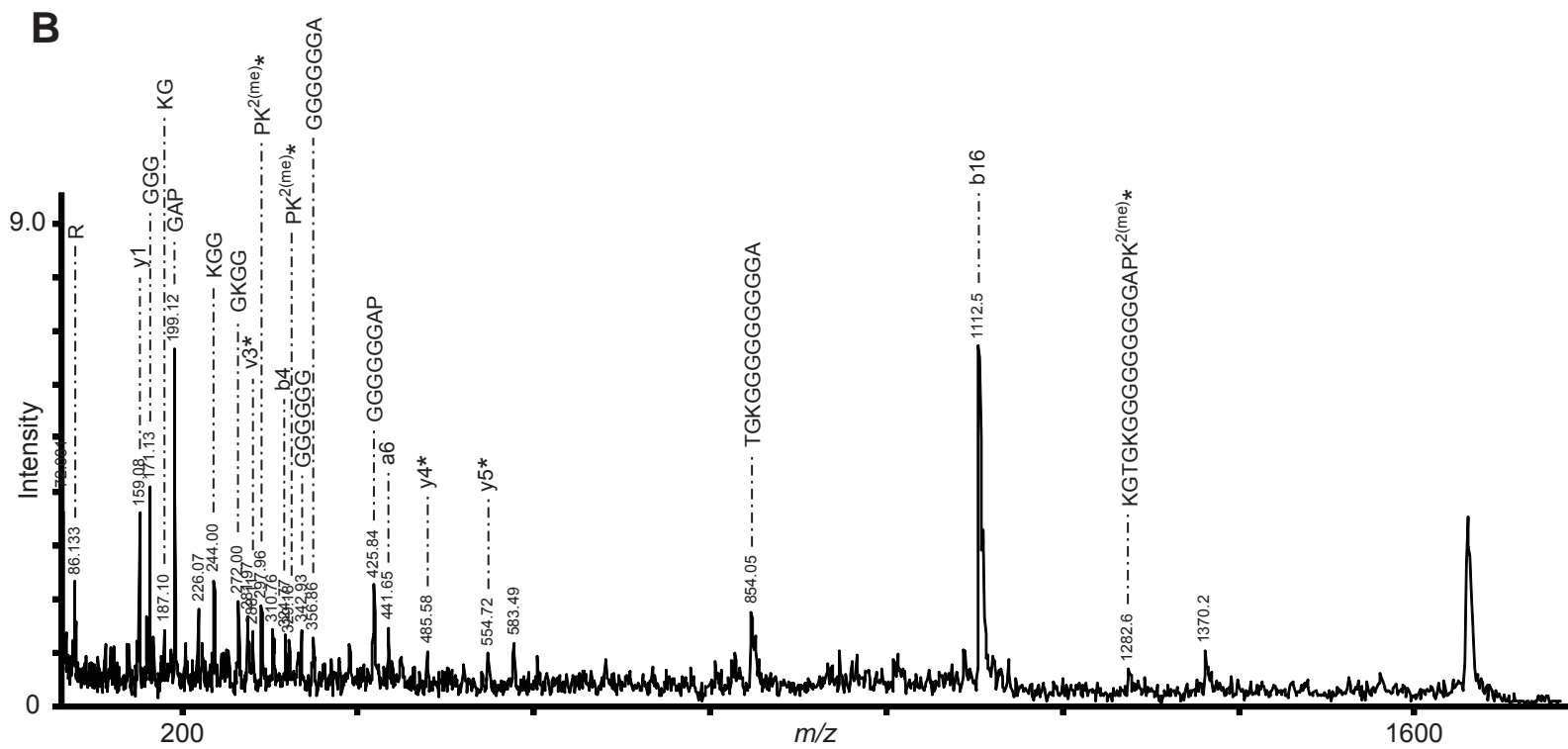
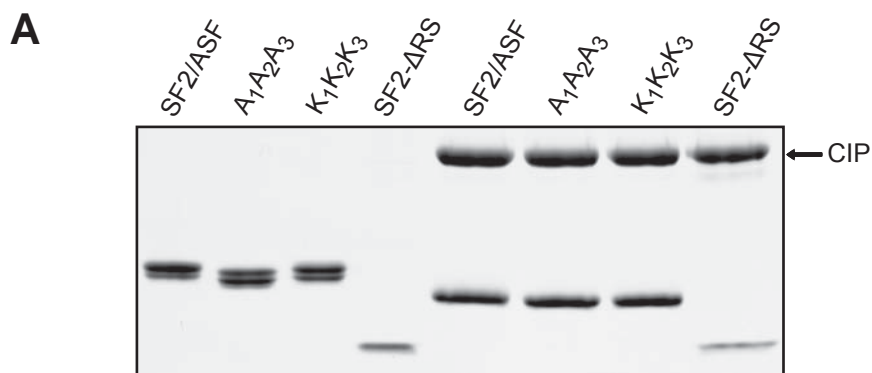
Figure 2.11

**Figure 2.12 Enhancement of NMD by SF2/ASF or mutants fused to a nuclear-retention signal.** We co-transfected T7-tagged, wt-SF2/ASF, or the A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> and K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> mutants, with or without the NRS1 peptide from SC35 fused at the C-terminus, with  $\beta$ -globin with or without a PTC. (A) RPA of total RNA after transient expression, as in Figure 4 (A). The bottom panels show western blots with anti-T7 antibody to detect the expressed tagged proteins, and endogenous  $\beta$ -catenin as a reference. (B) Quantification of phosphorimager signals from multiple experiments as in (A). We normalized the individual mRNA intensities to GFP mRNA, and plotted the ratios of normalized MT to WT mRNA intensities as in Figure 4 (C). n = 2-5; error bars = SD.



**Figure 2.12**

**Figure 2.13 Tandem mass spectrometry of the K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> protein.** (A) Coomassie stained SDS-polyacrylamide gel showing the different versions of SF2/ASF purified from 293E cells, before and after treatment with calf intestinal phosphatase (CIP). (B) We purified wild-type and mutant SF2/ASF expressed in 293E cells, treated them with CIP, separated them by SDS-PAGE, and stained them with GelCode blue. Next, we excised individual bands, digested them with trypsin, and analyzed them by MALDI-TOF. We selected peptides corresponding to the region of interest and analyzed them by MS/MS. The MS/MS spectrum shown is from one such peptide from the K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> protein with two methyl groups. The largest peak (b16 = 1112.5) has a match only when one of the three Lys residues is dimethylated, and not if two of them are monomethylated. The peptides marked with an asterisk correspond to peptide mass peaks that can occur only when Lys at position 109 is dimethylated. The ion-matches observed, when one of the three Lys is dimethylated, (K<sup>2(me)</sup>) is represented at the bottom, which indicates that Lys109 is indeed dimethylated in this peptide.



1684.8 =  $\left[ \begin{array}{l} \text{SGKGTGKGGGGGGGGGAPK}^{2(\text{me})}\text{GR} = 23/28 \text{ ions match} \\ \text{SGK}^{2(\text{me})}\text{GTGKGGGGGGGGGAPKGR} = 12/25 \text{ ions match} \\ \text{SGKGTGK}^{2(\text{me})}\text{GGGGGGGGGAPKGR} = 13/25 \text{ ions match} \end{array} \right.$

**Figure 2.13**

**Figure 2.14 Schematic representations of the motif mutations introduced to create SM and KM proteins.** Figure shows the different Gly→Ala as well as Arg<sup>111</sup>→Lys substitutions introduced in the inter-RRM linker region of SF2/ASF as well as the K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> mutant to alter the RGG and RGG-related motifs. The individual motifs are underlined. The Arg residues under investigation in wt-SF2/ASF are shown in orange and the corresponding initial triple mutations (A<sub>1</sub>A<sub>2</sub>A<sub>3</sub> and K<sub>1</sub>K<sub>2</sub>K<sub>3</sub>) are shown in red. The Ala substitutions of the surrounding Gly residues as well as the substituted Lys at position 111, for both SF2/ASF and K<sub>1</sub>K<sub>2</sub>K<sub>3</sub>, are depicted in green.

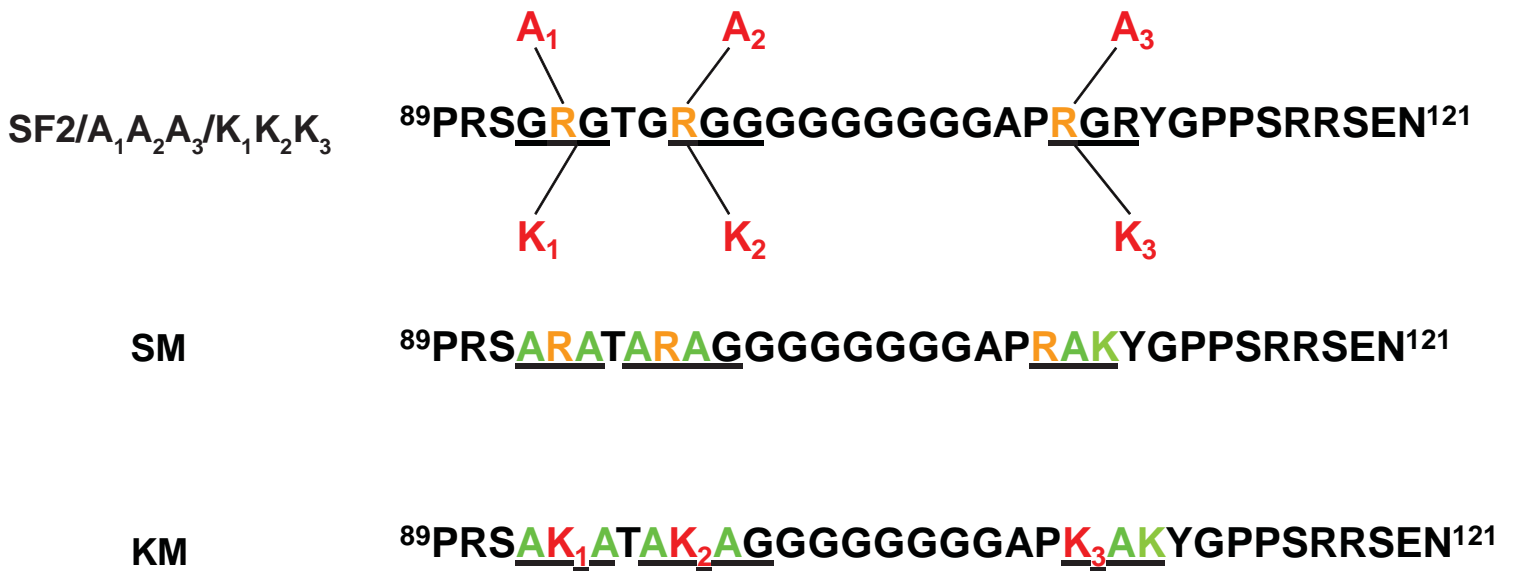
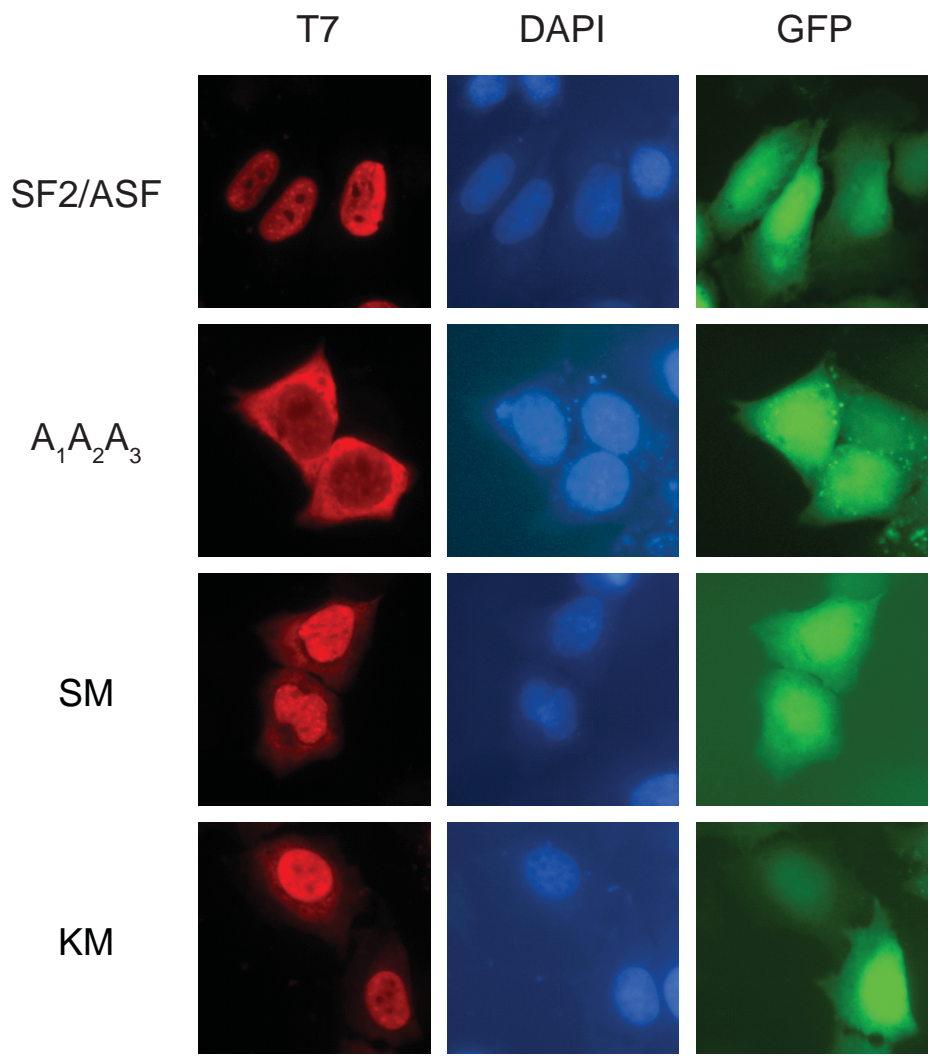


Figure 2.14



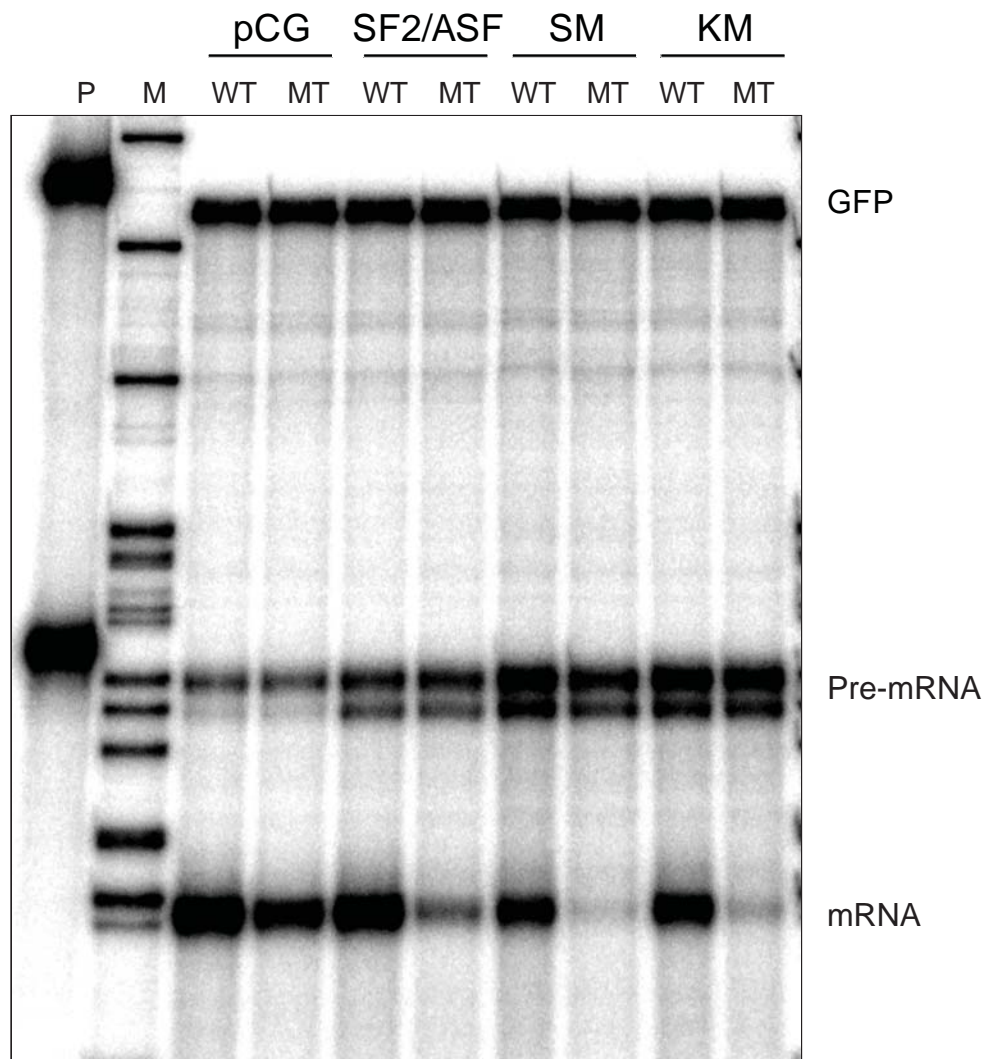
**Figure 2.15 Subcellular localization of SM and KM mutants.** Indirect immunofluorescence of HeLa cells co-transfected with GFP and wild-type or mutant T7-tagged SF2/ASF.



**Figure 2.15**

**Figure 2.16 Effect of overexpression of SM and KM mutants on NMD.**

We cotransfected  $\beta$ -globin and GFP probes simultaneously into HeLa cells as in Fig. 2.5 and performed RPA with GFP signal as a reference. Both SM and KM versions of SF2/ASF promote NMD as efficiently as the wild-type or K<sub>1</sub>K<sub>2</sub>K<sub>3</sub> version.

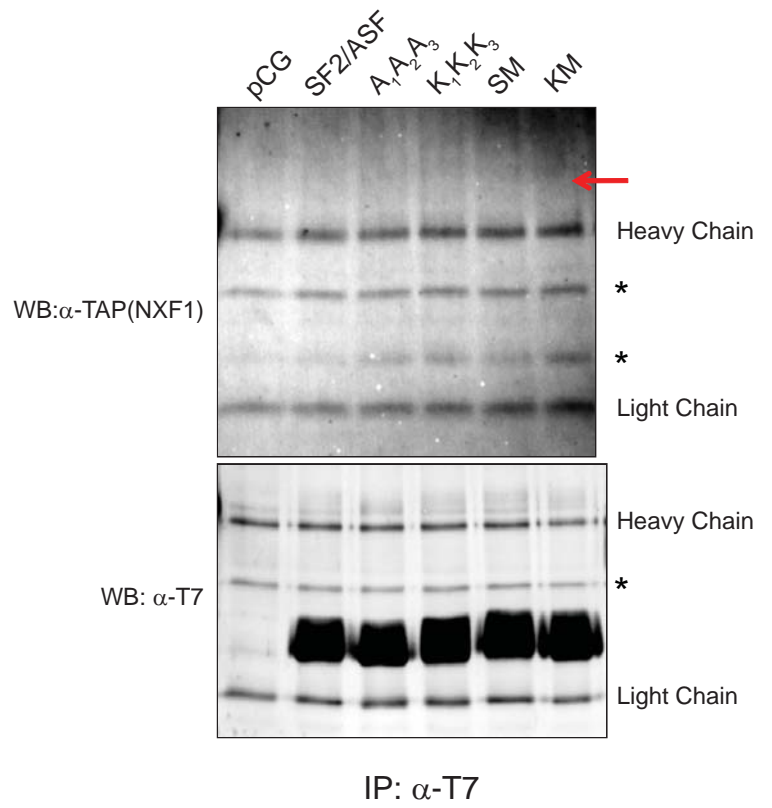
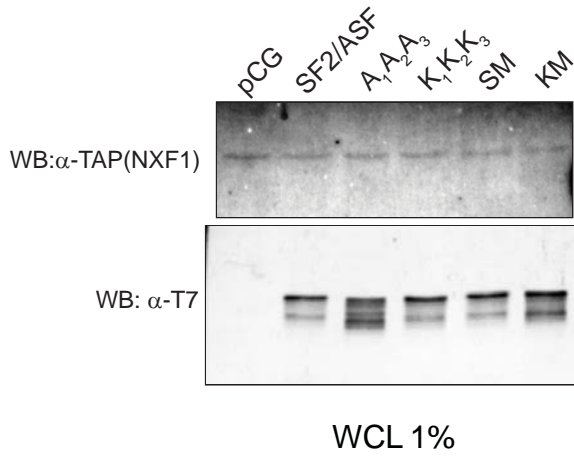


**Figure 2.16**

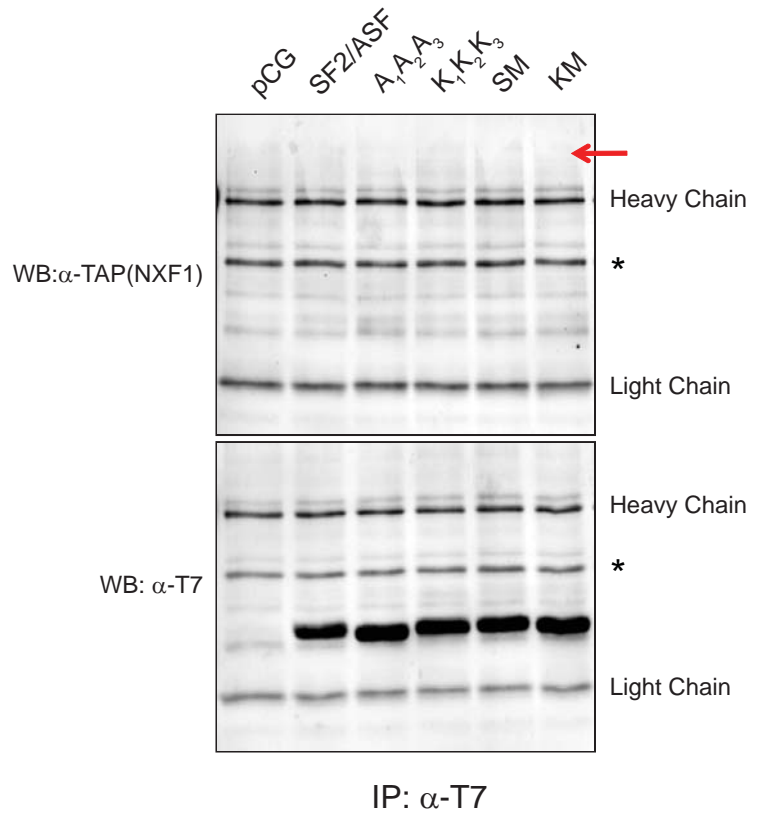
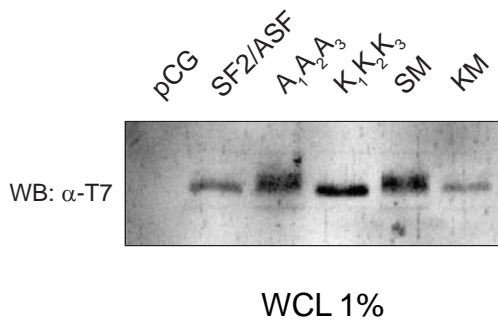
**Figure 2.17 Interactions between TAP/NXF1 and SF2/ASF.** (A) We transfected HeLa cells with T7-tagged SF2/ASF or its mutants and 48 h later, prepared the cell extracts with phosphatase inhibitors. We used 500 µg of the lysates for immunoprecipitation (IP) using anti-T7 agarose beads (Novagen), and eluted the final immunoprecipitated proteins by resuspending the beads in Laemmli buffer. We used 1% each of the eluted fraction as well as the whole cell lysate, resolved the constituent proteins by SDS-PAGE and analyzed by Western blotting (WB) using anti-T7 antibody. We analyzed the remaining eluted immune complexes by SDS-PAGE followed by WB with anti-TAP/NXF1. (B) Same as in (A) except cell lysates were prepared without phosphatase inhibitors and treated with calf intestinal phosphatase (CIP) before conducting the IP. The red arrows point to the expected position of band corresponding to TAP/NXF1.

**A**

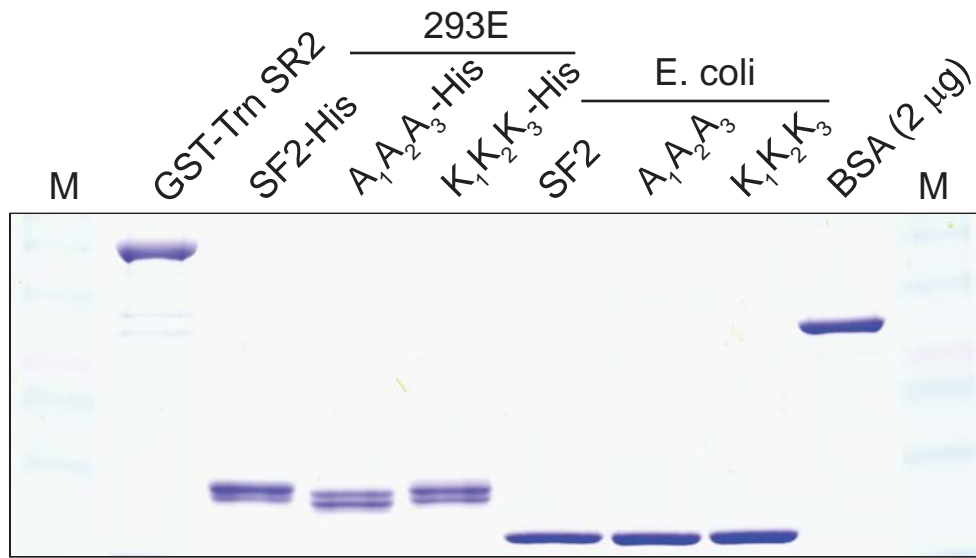
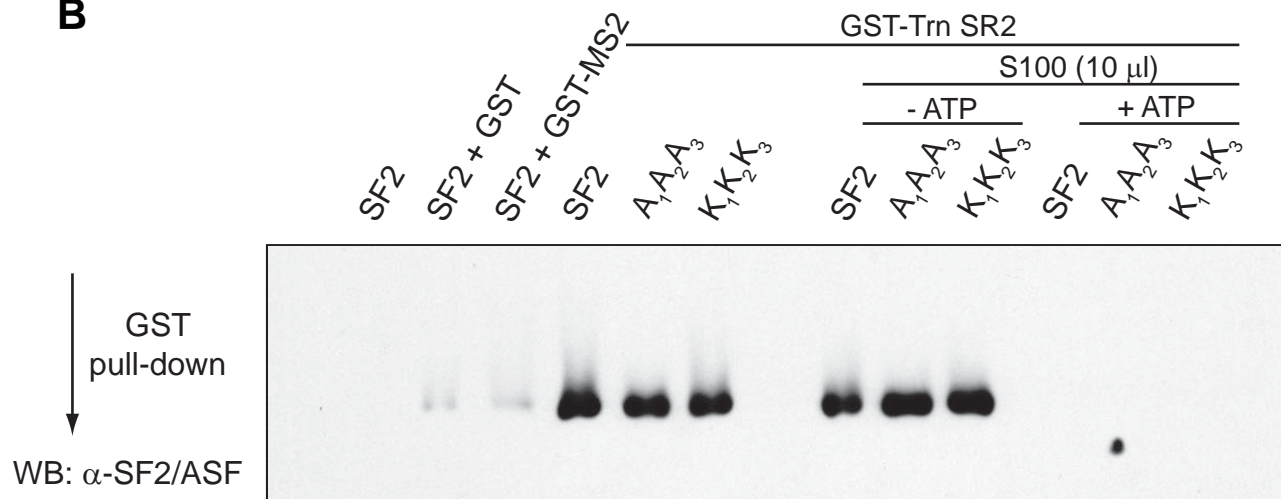
Phosphatase inhibitor

**B**

CIP treated

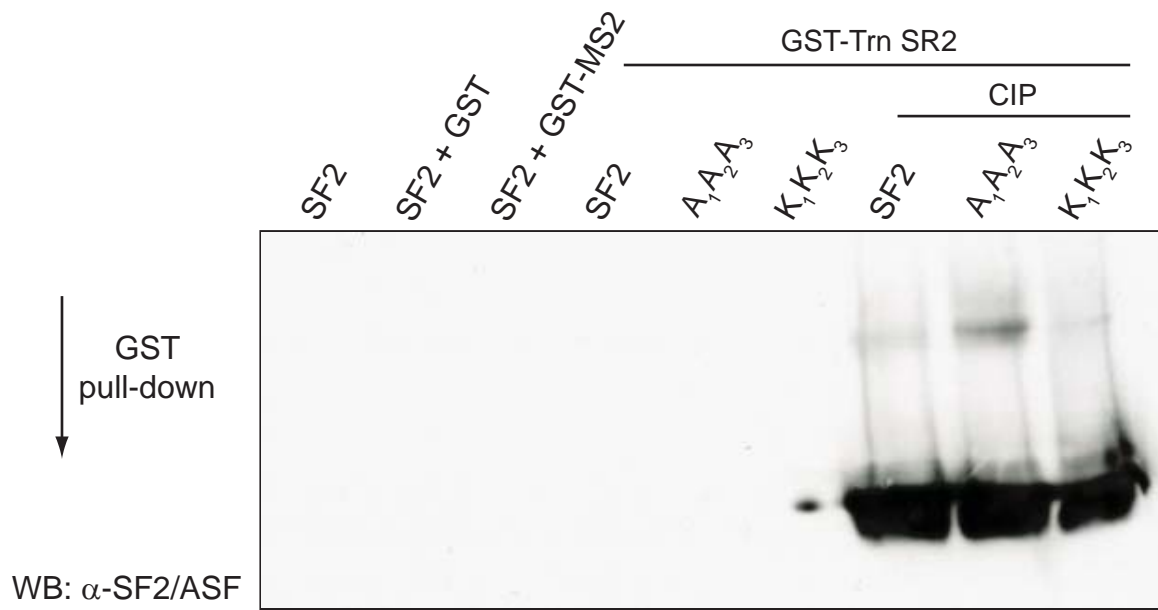
**Figure 2.17**

**Figure 2.18 Unmodified SF2/ASF interacts with transportin-SR2.** We purified GST-transportin-SR2 as well as SF2/ASF and its mutants from *E. coli* and used 2 µg of each protein for the GST pull-down. To visualize the interaction we resuspended the beads in Laemmli buffer, separated them by SDS-PAGE and analyzed by Western blot using anti-SF2/ASF (mAb96). In order to see the effect of phosphorylation on the interaction, we incubated SF2/ASF or its mutants in cytoplasmic S100 extract with ATP for 30 minutes prior to GST pull-down. The wt-SF2/ASF incubated in S100 extract without ATP was used as a control.

**A****B****Figure 2.18**



**Figure 2.19 Dephosphorylated SF2/ASF interacts with transportin-SR2.** We purified SF2/ASF or its mutants from 293E cells and used 2  $\mu$ g of each protein for GST pull-down using GST-transportin-SR2 as bait either with or without prior treatment with calf intestinal phosphatase (CIP).



**Figure 2.19**

## **Chapter 3:**

### **Antisense oligos: towards a cure for 'life without pain or tears'**

### 3.1 Introduction

It has been estimated that a substantial proportion of inherited genetic diseases are the results of splicing mutations (Lopez-Bigas et al., 2005). As described earlier, there are two ways splicing mutations can result in disease. The first mechanism involves mutations in the cis-acting elements present on the pre-mRNA, specifically, 5'ss, 3'ss, BPS, enhancers (ESE and ISE), and silencers (ESS and ISS) (see Chapter 1). The second mechanism involves changes in the physiological levels of trans-acting splicing regulatory proteins, which recognize enhancers and silencers. Perturbation of any kind, either mutations or changes in the levels of the core splicing machinery, rarely results in disease, probably because such events may be almost always incompatible with life and survival (Tazi et al., 2009).

In the first mechanism, a major fraction of disease-causing genetic mutations, which may either disrupt or create new cis-acting splicing elements, remain undetected due to our lack of understanding of the extremely degenerate enhancers and silencers, as well as our inability to distinguish between real and pseudo splice sites. As a direct result, even in the present era when the sequence of the entire human genome is available, complete information about all possible splice variants of a vast majority of important genes involved in diseases is lacking (see Chapter 1) (Wang and Cooper, 2007). However, our accumulated knowledge about splicing, as well as its involvement in disease, is ever increasing, with a growing list of diseases caused by splicing mutations (Cartegni et al., 2002; Wang and Cooper, 2007). In one study, about 10% of approximately 80,000 mutations reported in the human gene mutation database (HGMD) affect 'known' splice sites (Cooper et al., 2006).

Familial dysautonomia (FD) is one such autosomal condition that is caused by a recessive 5'ss mutation. FD or Riley-Day syndrome, as it was known earlier (Riley, 1952; Riley et al., 1949), is a hereditary sensory autonomic neuropathy type-III (HSAN-III), which is characterized by poor development and progressive degeneration of sensory and autonomic neurons (Axelrod and Simson, 2007). FD is a rare genetic disorder found almost exclusively in the Ashkenazi Jewish population, with a frequency of 1/3600 live births and a carrier frequency of 1/30, which increases up to 1/18 in Ashkenazi Jews of Polish descent (Maayan et al., 1987; Slaugenhaupt and Gusella, 2002). Almost all of the symptoms of FD result from widespread neuronal dysfunction caused by the continued degeneration of unmyelinated neurons involved in sensory and autonomic pathways. The most notable symptoms include: anhidrosis (inappropriate sweating) leading to hypothermia; absence of fungiform papillae on the tongue resulting in decreased taste; depressed deep tendon reflexes;

postural hypotension; mild to moderate loss of pain and temperature perception; lack of overflow tears with emotional crying (alacrima); gastroesophageal reflux; and scoliosis (Axelrod and Simson, 2007). The extent and severity of the symptoms vary among patients, but even with advanced management, the disease leads to premature death, with only half of the patients surviving to 40 years of age (Slaugenhaupt and Gusella, 2002).

Sixteen years ago, FD was found to be genetically linked to the long arm of Chromosome 9 (9q31) (Blumenfeld et al., 1993), but the gene responsible for FD, called *IKBKAP*, was identified and cloned only later (Anderson et al., 2001). The *IKBKAP* gene consists of 37 exons spanning a region of approximately 68 kb that encodes for the 150-kDa IKAP protein, comprised of 1332 amino acids (**Fig. 3.1**). A single recessive intronic point mutation in exon 20 (IVS20+6T→C) of the *IKBKAP* gene was found linked with the disease in over 99.5% of the cases. The mutation weakens the 5'ss of intron 20, as it replaces the conserved thymidine at position +6 with a non-consensus nucleotide, cytosine, leading to the skipping of exon 20 (Anderson et al., 2001). The skipping of exon 20 causes a frameshift, introducing a PTC in exon 21, which results in a truncated protein of only 699 amino acids (**Fig. 3.2**). The presence of a PTC in the skipped mRNA also makes it potentially susceptible to degradation via the NMD pathway (see Chapter 1). Additionally, two more missense mutations were identified, specifically, a single point mutation in exon 19 (G→C; R696P) and another mutation in exon 26 (P914L). Both the missense mutations represent minor mutations, as they are always accompanied by the major intronic non-coding mutation on the other allele (Leyne et al., 2003). Though it is clear that FD involves a splicing defect, it is not fully understood why pathogenesis is specific and exclusive to the peripheral nervous system, especially when disruption of *IKBKAP* splicing is ubiquitous across different tissues. Moreover, homozygous mutant cells derived from FD patients, as well as various patient tissues, express both included and skipped versions of the *IKBKAP* mRNA (Gold-von Simson and Axelrod, 2006; Slaugenhaupt et al., 2001). It has been observed that the relative levels of *IKBKAP* splice variants differ among different tissues in patients, with the levels of the full-length mRNA being the lowest in central and peripheral nervous systems, thus leading to the proposal that the restriction of phenotype exclusively to the neuronal tissue may result from tissue-specific differences in splicing and/or NMD (Cuajungco et al., 2003; Slaugenhaupt et al., 2001).

Very little is known about the mechanism by which depletion of the IKAP protein results in the pathogenesis of FD. IKAP was first described as an IκB kinase (IKK) complex-associated protein that acts as a scaffold and binds both NF-κB-inducing kinase and IKKs through separate domains, thus assembling

them into an active kinase complex (Cohen et al., 1998); this claim was subsequently repudiated (Krappmann et al., 2000). Today, it is established that IKAP is part of the human elongator complex and has a role in transcriptional elongation (Hawkes et al., 2002; Kim et al., 2002). Elongator complex was first identified as part of RNAP-II holoenzyme and isolated from yeast chromatin (Otero et al., 1999). Further characterization showed that the elongator is comprised of two sub-complexes that combine to form the holo-elongator, consisting of 6 subunits, elongator proteins 1-6 (Elp1, Elp2, Elp3, Elp4, Elp5, and Elp6) (Fellows et al., 2000; Krogan and Greenblatt, 2001; Li et al., 2001; Winkler et al., 2001). The core complex consists of Elp1-3, with Elp1 (IKAP) being the largest subunit. Another constituent of the core, Elp3, is similar to the GCN5 family of histone-acetyltransferases (HATs) based on homology (Wittschieben et al., 2000; Wittschieben et al., 1999). Though Elp3 is part of the core complex, it is not sufficient for HAT activity and requires the association of a smaller sub-complex comprised of Elp4-6. It was demonstrated that the holo-elongator (Elp1-6) preferentially acetylates histone H3, with a weaker affinity for histone H4 (Winkler et al., 2002), and promotes RNAP-II transcription from chromatin, but not naked DNA, in an acetyl-CoA-dependent manner (Kim et al., 2002). Subsequent evidence showed that the elongator complex associates with the transcription machinery after initiation of transcription and promotes elongation (Kouskouti and Talianidis, 2005; Metivier et al., 2003). In a more recent study, microarray analysis of IKAP/Elp1-depleted HeLa cells revealed that a subset of genes regulated by elongator are involved in cell migration. Consistently, both IKAP-depleted HeLa cells, as well as FD-patient-derived fibroblasts, show pronounced defects in cell migration and motility. These are processes critical for development and maintenance of the nervous system, suggesting the role of loss of elongator in FD pathogenesis (Close et al., 2006). In addition to its role in transcription, the elongator has been implicated in various other cellular events, including stress signaling, exocytosis, and tRNA modification (Svejstrup, 2007). However, how these additional functions may be related to FD phenotype is not known.

FD is undoubtedly the most studied among the HSANs and the only one for which a genetic screen is commercially available. Although there have been various attempts to find a therapy, to date prevention by prenatal screening remains the only weapon in the fight against FD. Almost all of the attempts to find a cure have focused on correcting the splicing defect by increasing the levels of exon 20 inclusion through treatment with various small molecules, with varying degrees of success. Most notable among these small molecules are tocotrienols, epigallocatechin gallate (EGCG), and kinetin. It was shown that treatment of FD-derived human fibroblasts, as well as neuronal-derived cells with  $\delta$ -tocotrienol, a

member of the vitamin A family, results in the increased transcription of the *IKBKAP* gene, which in turn leads to increased levels of full-length IKAP protein (Anderson et al., 2003b). EGCG, a polyphenol, was shown to increase the levels of full-length IKAP mRNA and protein by increasing the inclusion of exon 20 in FD-derived cells, probably by downregulation of the splicing repressors hnRNP A2/B1. Combined treatment with both  $\delta$ -tocotrienol and EGCG resulted in a synergistic increase in IKAP protein levels in FD patient fibroblasts (Anderson et al., 2003a). Similar to EGCG, a plant cytokine, kinetin, has also been shown to promote the inclusion of exon 20 in FD-derived lymphoblasts (Slaugenhaupt et al., 2004). In a recent clinical study, which involved 29 healthy individuals that were heterozygous for the major FD mutation, orally administered kinetin was shown to increase the levels of full-length IKAP mRNA after 8 days of treatment (Gold-von Simson et al., 2009).

Besides small molecules, antisense oligonucleotides (ASOs) provide another avenue towards therapy for various genetic disorders, including those caused by splicing mutations. ASOs have great therapeutic potential due to the ease with which they can be designed, synthesized, and then used to specifically micromanipulate a given aspect of gene expression without affecting other processes. For example, ASOs can be used to block translation through formation of a duplex with the mRNA, to direct RNaseH-mediated degradation of the target mRNA through the formation of RNA-DNA hybrids, or to modulate alternative splicing of target pre-mRNA by blocking splice sites and/or enhancers/silencers. More recently, a new class of ASOs that utilize the RNA interference machinery to shut off target genes have emerged as potential therapeutic agents. These new ASOs are: small interfering RNA (siRNA), short hairpin RNA (shRNA), and micro RNA (miRNA). Besides ASOs, there are other classes of oligonucleotides that are being used or have the potential to be used as therapies; most notable are aptamers, ribozymes, and RNA decoys. The concept that the cellular machinery can be redirected by ASOs (or by other oligonucleotides), towards a more favorable outcome, especially in the case of various genetic diseases, has existed for a long time. However, only in a couple of instances have the oligonucleotides made it to the clinic as drugs. The first such drug to be ever approved by the Food and Drug Administration (FDA) for treatment of a human disease was Fomivirsen, which is an ASO that blocks translation from mRNA of an important cytomegalovirus gene and is used to treat retinitis (Holmlund, 2003). The other FDA-approved drug is an aptamer, Pegaptanib, which acts as an anti-angiogenic agent by binding to VEGF and is used in the treatment of age-related macular degeneration (AMD) (Bonetta, 2009).

The various hurdles that have impeded the development of ASOs as drugs are multilayered, and include stability, tissue-specific delivery, and specificity. For a long time, the major hurdle has been the lack of desired levels of stability of the ASOs. Unmodified ASOs, if administered, are rapidly cleared, often in a matter of hours, by the action of various nucleases. Recent advances in chemical modifications that alter the backbone of ASOs, have largely overcome stability issues and have ensured that ASOs persist in the vascular system long enough (most often weeks to months), and in sufficient quantities, to fully exert their therapeutic potential. Some of these modifications include morpholino, peptide nucleic acid (PNA), locked nucleic acid (LNA), 2'-O-methyl, phosphothioate, and 2'-O-methoxyethyl (ISIS Pharmaceuticals) (Karkare and Bhatnagar, 2006). Due to these advances, there has been a renewed interest in ASOs as therapeutic agents, with an increasing number of potential ASOs reaching the stage of clinical trials (Bonetta, 2009).

Other criteria that are crucial, especially if ASOs are intended to modulate the splicing of target genes, are nuclear accumulation and affinity of binding. The ASOs must not only be able to reach the nucleus, but also should be able to overcome and outcompete various RNA-RNA and RNA-protein interactions between the spliceosome and the splice-signals, such that splicing is directed towards a more desirable alternative outcome. It is expected, however, that not all oligos can overcome the native ribonucleoprotein (RNP) interactions. Besides, information about cis-acting sequences other than splice-sites (enhancers and silencers), is usually lacking, which makes it difficult to predict the precise region where an oligo should be targeted. Therefore, multiple overlapping oligos should be tested by tiling stretches of pre-mRNA of varying lengths in order to identify the best functional oligo that can affect splicing.

In collaboration with ISIS Pharmaceuticals, our laboratory has successfully developed a few antisense-oligonucleotide (ASO) drug candidates that have the potential to treat or even cure spinal muscular atrophy (SMA) by correcting *SMN2* exon 7 splicing (Hua et al., 2007; Hua et al., 2008). These oligos carry either a 2'-O-methoxyethylribose-phosphorothioate or 2'-O-methoxyethylribose-phosphate backbone (referred to as pS-MOE-ASOs and pO-MOE-ASOs, respectively) instead of the ribose-phosphate backbone present in the natural RNA, which confers to them not only very high resistance to both exo- and endonucleases, but also much higher affinity for the target sequences (**Fig. 3.5**). Moreover, these oligos do not support cleavage of hybridized mRNA by RNase H and the modifications do not affect their nuclear penetration. The pO-MOE-ASOs are used in cell culture assays, whereas pS-MOE-ASOs are used in animal



model systems, as they are more stable and are more efficiently internalized in comparison.

SMA is a common motor neuron degenerative disease, which is caused by the loss or inactivation of survival of motor neuron (SMN) protein (Lefebvre et al., 1995; Talbot and Davies, 2001). SMN is essential for cell viability, as it plays a crucial role in snRNP biogenesis (Neuenkirchen et al., 2008). In humans, there are two genes that encode the SMN protein, survival of motor neuron 1 (*SMN1*) and 2 (*SMN2*), both of which have identical open reading frames. However, the majority of full-length functional protein in the cell results from expression of the *SMN1* gene. A single nucleotide difference between *SMN1* and *SMN2* at position 6 of exon 7 (C→T) leads to skipping of exon 7 in *SMN2*. Though the naturally occurring single-nucleotide substitution in exon 7 of *SMN2* does not alter the amino acid sequence, it inactivates an ESE and probably creates an ESS, leading to the skipping of exon 7 in a major fraction of *SMN2* transcripts (Cartegni et al., 2006). The skipped splice variant of *SMN2* results in a truncated protein, SMN $\Delta$ 7, which is highly unstable and is non-functional. Only a small fraction of the *SMN2* transcripts, ranging from 10-20% depending on tissue type, is spliced into full-length mRNA leading to functional SMN protein. All SMA patients show either loss or inactivation of *SMN1*, but preserve at least one copy of *SMN2*. As the severity of SMA is inversely proportional to SMN levels, the severity of the disease symptoms vary widely among patients, depending upon the number of copies of *SMN2* present in the genome (Wirth et al., 2006). One approach for SMA therapy is to correct the splicing defect in *SMN2* by promoting exon 7 inclusion, and thus alleviating the SMN deficiency. In our laboratory, Hua and coworkers employed the tiling approach (see above) to scan exon 7 and the surrounding intronic sequences with 15-20mer MOE-ASOs and thus identified several oligos that had a strong positive effect on exon 7 inclusion. Exon inclusion is promoted when ASOs bind to cis-acting silencers (ESS and ISS) and conversely, oligos that target enhancers (ESE and ISE) lead to exon skipping. The screen not only characterized oligos that were able to correct the *SMN2* splicing defect but also identified several new enhancers and silencers that control the splicing of exon 7, both within exon 7 and in the surrounding intronic regions (Hua et al., 2007; Hua et al., 2008). A few such MOE-ASOs that target an intronic region immediately downstream of the 5'ss of intron 7 could achieve up to ~95% *SMN2* exon 7 inclusion in both patient-derived cells and in various tissues in type III SMA mice, after systemic administration (Hua et al., 2008). The increase in mRNA was correlated with a significant increase in full-length SMN protein. Further analysis of the target intronic region revealed the precise silencer element (ISS) and its trans-acting factor, hnRNPA1, providing mechanistic insight into the regulation of exon 7 splicing (Hua et al., 2008).

However, other hurdles mentioned above still remain. For example, ASOs are unable to cross the blood-brain barrier, and upon systemic administration, some organs, like liver and kidney, may be better targeted than others, like pancreas and muscle (Bonetta, 2009; Hua et al., 2008). Though specificity of an ASO can be easily predicted by BLAST search, off-target effects of a given oligo must be empirically determined through studies in animal model systems, and subsequent trials in primates, and finally, human subjects, like any other putative drug.

The widespread malfunction of the peripheral nervous system observed in FD patients, is a direct result of poor development and progressive loss of the sensory and autonomous neurons. Although consistent neuropathological condition related to central nervous system (CNS) has not been documented in case of FD, quite frequently, generalized atrophy of certain regions including cerebellum are observed in older patients resulting in their inability to maintain proper balance and ataxic gait. Further, changes in mental processing as well as progressive atrophy of optical nerve have also been observed in some adult FD patients (Axelrod and Simson, 2007). Since it is not entirely clear whether these anomalies are a direct manifestation of the disease, or are secondary in nature, resulting due to the labile blood pressure observed in FD patients, delivery of MOE-ASOs to the central nervous tissue may be critical for their effectiveness as a therapy. Recently, to circumvent the blood brain barrier, Hua and coworkers used surgically inserted osmotic pumps for direct intracerebroventricular (ICV) infusion to deliver the MOE-ASOs into the central nervous system (CNS) of SMA type III mice. Direct administration of MOE-ASOs to the nervous tissue resulted in complete rescue of splicing, leading to significant increase in human SMN protein in both brain and spinal cord (Hua et al., 2009). Remarkably, in another approach, a single ICV injection in prenatal embryos resulted in complete rescue of tail and ear loss, which are the overt phenotypic traits of SMA type III mice.

Here we show that the observations and the progress made towards a cure for SMA using MOE-ASOs, can be extended to other splicing-based genetic disorders like FD. In a similar approach employed previously for SMA, we used a tiling screen with 15mer overlapping MOE-ASOs to scan the entire sequence of exon 20 and the flanking upstream and downstream intronic regions in the *IKBKAP* pre-mRNA. We have identified extremely promising MOE-ASOs that are fully able to restore the splicing of mutant *IKBKAP* in patient-derived fibroblasts and are currently testing their effectiveness in correcting splicing defects in a human *IKBKAP* transgenic mouse model (Hims et al., 2007). Combined with mutational analysis of the *IKBKAP* minigene, we have identified novel silencer

elements in *IKBKAP* pre-mRNA, and are engaged in characterizing the trans-acting factors that regulate the splicing of exon 20.

## 3.2 Results

### 3.2.1 MOE-ASO walk reveals several enhancers and silencer regions in *IKBKAP*

The genomic region that spans the *IKBKAP* gene is very large (~68 kb) (**Fig. 3.1**). For ease of manipulation, we created minigenes by cloning the genomic fragment of *IKBKAP*, ranging from either exon 19 to exon 21 (wt19-21) or exon 19 to exon 22 (wt19-22), under the control of the cytomegalovirus (CMV) promoter. We introduced the major mutation found in FD (IVS20+6T→C) into both versions of the minigenes via site-directed mutagenesis to get the corresponding mutant minigenes, mt19-21 and mt19-22 (**Fig. 3.3**). All these different minigenes carried an artificial in-frame ATG as the first codon, within a Kozak sequence at the 5' end, downstream of the CMV promoter, as well as a stop codon at the 3' end upstream of the poly(A) signal. All the minigenes were transfected in 293 cells and the splicing patterns of the expressed RNA were analyzed by RT-PCR. We observed that the mutant versions of the minigenes (mt19-21 and mt19-22) consistently showed predominant skipping of exon 20, thus recapitulating the effects observed in FD patients (**Fig. 3.4A**).

We designed a total of 49 15mer pO-MOE-ASOs (**Fig. 3.5**), which were complementary to the entire *IKBKAP* exon 20, as well as the 100-nucleotide regions immediately upstream and downstream of exon 20. Each pair of consecutive oligos had a 10-nucleotide overlapping region, such that a different ASO was tiled every five nucleotides, moving from 5' to 3' in the genomic region specified above. An oligo with unrelated sequence was used as a negative control.

To test whether some of these oligos could promote the inclusion of exon 20 in the context of the major FD mutation in cells, we co-transfected each oligo individually with the mt19-21 minigene in 293 cells by electroporation, and two days later assayed the splicing behavior of expressed RNA by RT-PCR. The lanes with wt19-21 and mt19-21 minigenes alone were used as a point of reference for exon 20 inclusion levels, whereas the lane that contained mt19-21 with the control oligo, served as a control for non-specific effects of pO-MOE-ASO on splicing (**Fig. 3.6, 3.7, and 3.8**).

To our surprise, six consecutive oligos (oligo# 414161-66), which target a 40-nucleotide long intronic region immediately downstream of the 5'ss of exon 20, dramatically increased the inclusion levels of exon 20, suggesting the

presence of multiple silencer elements within this region, which we termed ISS-40 (**Fig. 3.8**). The positive effect of the ASOs decreased slightly as the distance of their target sequence increased from the 5'ss. Three more oligos (414135-37), which target a 20-nucleotide long region in the upstream intron 19 (ISS-20), also had a positive effect on exon 20 inclusion and resulted in about a twofold increase (**Fig. 3.6**). In contrast, exon 20 is apparently full of enhancers, as almost every oligo targeting the exonic sequences resulted in nearly complete skipping (**Fig. 3.7**). However, the skipping caused by the oligos targeting the extreme 5' end (414147) and extreme 3' end (414159) of the exon 20, was in part due to the fact that they target the 3'ss and 5'ss, respectively. Besides these exonic oligos, there were other intronic oligos that caused increased skipping, most probably because they targeted cis-splice signals. These are oligos 414143-45, which target the Py-tract (**Fig. 3.6**), and oligo 414160 that targets the 5'ss of exon 20 (**Fig. 3.8**). We also identified at least two oligos (414167-68) that apparently target an intronic enhancer element (ISE-20) and decreased the levels of included mRNA isoform by half when compared to the untreated control (**Fig. 3.8**).

### 3.2.2 Oligo-walk with an NMD-responsive minigene

We repeated the experiment, this time utilizing the wt- and mt19-22 minigenes, in order to check for reproducibility, and more importantly, to determine the role played by nonsense-mediated mRNA decay (NMD) in the stability of the skipped mRNA isoform (**Fig. 3.9, 3.10, and 3.11**).

As mentioned earlier, the skipping of exon 20 causes a frameshift and introduces a premature termination codon (PTC) in the skipped mRNA. The PTC lies in exon 21 and is located more than 50 nucleotides upstream of the last exon-exon junction between exon 21 and exon 22, therefore making the skipped mRNA potentially susceptible to degradation via NMD according to the characterized rules of the mRNA surveillance pathway (see Chapter 1) (Nagy and Maquat, 1998). However, unlike the natural context, the skipped mRNA resulting from the 19-21 minigene does not have an exon junction downstream of the PTC, and therefore, is incapable to respond to NMD. This was a crucial aspect of experimental design, as it allowed us to look at the effect of the oligos exclusively in altering splicing without interference from NMD in the previous experiment (**Fig. 3.6, 3.7, and 3.8**).

As far as the effect of individual oligos was concerned in the inclusion or skipping of exon 20, we observed exactly the same pattern as seen previously when we used the 19-21 minigenes, thus giving an indication of reproducibility (**Fig. 3.9, 3.10, and 3.11**). Further, we did not find any evidence that would

suggest that the skipped mRNA isoform is subjected to NMD. If the skipped mRNA (19-21-22) were unstable, we would expect to see an increase in inclusion percentages all across the oligo-walk; on the contrary, we observed that the inclusion levels of exon 20 were lower than observed with the 19-21 minigenes, in all cases. This suggested that even though the skipped mRNA isoform has all the characteristics of an NMD substrate, it somehow evades the surveillance machinery and is stable.

To confirm this finding, we corrected the frameshift that occurs due to skipping of exon 20, by a single nucleotide deletion in exon 21 of the mt19-22 minigene, giving rise to the mt19-22/FC minigene (**Fig. 3.3B**). The restored reading frame abolishes the PTC, and hence in theory, should make the skipped mRNA isoform more stable. We transfected the different minigenes into 293 cells and analyzed the expressed RNA by RT-PCR. Consistent with the observation made during the oligo walk, we found that the skipped mRNAs, either with or without the PTC, were equally stable, thus confirming that at least in 293 cells the skipped mRNA isoform is not subjected to degradation via NMD (**Fig. 3.4B**).

### **3.2.3 High resolution micro-walk in the ISS-40 region**

Since the ISS-40 region showed the most dominant repressive effect on inclusion of exon 20 in the context of the major FD mutation, we hypothesized that the optimal ASO that could restore the splicing defect with maximal effect must target this region. To search for such an oligo, we designed 10 new 20mer overlapping MOE-ASOs, which were complementary to and covered the first 30-nucleotide stretch of the ISS-40 with one-nucleotide increments, starting from the +6 position in exon 20. We cotransfected these oligos with the wt19-21 and mt19-21 minigenes in 293 cells, followed by RT-PCR, as earlier, to analyze the expressed mRNA levels. The microwalk led to the discovery of oligo 421992, which strikingly, almost completely restored exon 20 inclusion levels (up to 96%) in the case of the mutant minigene (**Fig. 3.12**). This clearly suggests that this oligo blocks a very strong silencer, or probably more than one silencer, in the region complementary to its sequence, which spans from +6 to +26 in intron 20. In general, the 20mer MOE-ASOs exhibited a stronger positive effect on exon 20 splicing than their 15mer counterparts, which targeted the same region, probably because the longer oligos were better able to interfere with the activities of multiple silencer elements in the region and/or because they hybridized more efficiently to the target RNA.

### 3.2.4 Effect of oligo 421992 in FD-derived fibroblasts

To investigate the effect of our best candidate oligo, 421992, in a more natural context, we acquired the patient fibroblast line GM04899 (Coriell Cell Repository), which was isolated from an individual homozygous for the major FD mutation. As the primary fibroblasts from patients are not amenable to electroporation (Hua et al., 2007), we used Lipofectamine 2000 to deliver increasing amounts of the oligo, ranging from 2-800 nM. Even though the basal level of exon 20 inclusion was higher in the FD fibroblasts (~64%) compared to those resulting from mutant minigenes described above, we observed that very low amounts of the 421992 oligo (10 nM) were sufficient to almost completely suppress the splicing defect (**Fig. 3.13**). Further increase in oligo dosage resulted in saturation of the included isoform, indicating that the effective oligo dose in these patient fibroblasts was somewhere between 2-10 nM. The experiments to determine whether the increase in levels of the included isoform observed after oligo treatment also results in the accumulation of full-length IKAP protein in these cells are underway.

### 3.2.5 ISS-40 harbors at least two silencer elements

Having identified ISS-40 as one of the major determinants that controls splicing of exon 20 in *IKBKAP*, as well as promising target for therapy of FD, we focused our efforts on demarcating the precise boundaries of the silencers present in the region. Mapping of the silencers is crucial, as it may provide clues regarding the corresponding trans-acting factors that regulate *IKBKAP* exon 20 splicing. The knowledge of the factors involved, may in turn, lead to a better understanding of the underlying molecular mechanisms, which are responsible for the strict tissue-specific manifestation of the disease, even when the splicing defect is ubiquitous and not restricted to the nervous system alone (see Introduction).

To delineate the silencers within the ISS-40 region, we individually mutated each of the first 31 constituent nucleotides of this region by changing them to their complementary bases (transversion) in the mt19-21 minigene through site-directed mutagenesis. The mutations resulted in a series of 31 new minigenes, each carrying two mutations: first, the major FD mutation at position +6 in intron 20 that was common to all, and a second mutation that consisted of one of the transversions in ISS-40, starting at position +7 and going up to position +37. Further, in addition to the already existing mt19-21 minigene that carried the major FD mutation (IVS20+6T →C), two additional minigenes were created where the thymine at position +6 was changed to either a guanosine (IVS20+6T→G) or an adenine (IVS20+6T→A). All of the 33 new minigenes, along

with the wt19-21 and mt19-21 minigenes (total 35), were individually transfected into 293 cells, and two days later, the levels of the expressed spliced isoforms were analysed by RT-PCR, as above. The analyses revealed eight different individual transversions that promoted exon 20 inclusion in the context of the major FD mutation, and clustered in two different 6- to 8-nucleotide regions within the ISS-40 region (**Fig. 3.14 and 3.15**). The first cluster consisted of 3 mutations, +10A→T, +13G→C, and +15A→T, which resulted in a three- to fourfold increase in exon 20 inclusion levels, suggesting that the hexamer, ATTGTA, is part of a putative silencer (**Fig. 3.14**). The second cluster comprised five mutations, +28A→T, +29G→C, +31T→A, +32A→T, and +33G→C, which span an eight-nucleotide region, **TAGTTAGC**, containing two signature hnRNPA1 core (UAG) motifs (Abdul-Manan and Williams, 1996; Burd and Dreyfuss, 1994; Si et al., 1998). The individual mutations in the hnRNP core sequence resulted in a three- to sevenfold increase in exon 20 inclusion levels (**Fig. 3.15**). Together with previous observations that showed that the AG dinucleotide in the core sequence is crucial for hnRNPA1 binding (Hua et al., 2008), our data strongly suggest that this region constitutes an hnRNPA1 binding site. The close proximity of the two core motifs raises interesting questions regarding the mechanisms through which hnRNPA1 might recognize and bind this sequence to carry out its silencing activity, as simultaneous recognition of both UAG cores seems sterically impossible.

### 3.3 Discussion and future perspectives

The recent advances in antisense technology have led to a renewed interest in ASOs as therapeutic agents, with an increasing number of ASOs and other therapeutic oligos, reaching clinical trials for a wide variety of ailments (Bonetta, 2009). Towards this end, our laboratory has also made significant contributions by developing a system and methodology to correct splicing defects caused by genetic mutations and further utilized such methodology to make considerable advances in the direction of treatment or even a cure for SMA (Hua et al., 2009; Hua et al., 2007; Hua et al., 2008). The potential of ASOs in treating genetic disorders that affect splicing is especially noteworthy and holds great promise, due to the ease with which different oligos can be designed and synthesized to target the regulatory cis-elements towards a favorable splicing outcome (see Introduction). The current study not only extends the previous findings, but also proves that the two-step methodology, pioneered in our laboratory to identify the most effective oligo for therapy (Hua et al., 2007; Hua et al., 2008), is an extremely effective approach, and can be applied to other genetic diseases. The two-step approach involves an initial coarse walk in a larger region to identify potential regulatory sequences as target regions,

followed by a high-resolution microwalk in the identified small target regions to search for the optimal oligo.

Using an identical approach, we were able to identify three specific regions that influence the splicing of *IKBKAP* exon 20 in the context of the major FD mutation. These regulatory cis-elements are: the ISS-20 in upstream intron 19, the strong ISS-40 silencing region immediately downstream of exon 20, the ISE-20 intronic enhancer downstream of ISS-40, and an enhancer region right in the middle of exon 20, probably spanning most of the exon (see results). As inclusion of exon 20 is the desired outcome for therapeutic purposes, based on our data it was clear that the ISS-40 region is the strongest candidate as a target sequence. This prompted us to concentrate our efforts towards the search for the most effective oligo that hybridizes within the ISS-40 region, which resulted in the discovery of the 20mer MOE-ASO, 421992. We show that the 421992 MOE-ASO could not only rescue the splicing defect in the mutant minigenes (mt19-21 and mt19-22), but could also almost completely restore exon 20 inclusion levels at very low doses in patient fibroblasts.

Encouraged by these promising results, we moved to the next logical step, and acquired the transgenic mice that have a few copies of the entire human *IKBKAP* gene with the major FD mutation, in addition to the mouse *Ikbkap* (Hims et al., 2007). Prior to the creation of the transgenic mice, attempts by others to humanize the mice by introducing the major FD mutation (IVS20+6T→G) in the mouse *Ikbkap* gene had failed, as the mutation did not result in the skipping of exon 20 in the resulting mouse pre-mRNA. We presume that the failure was due to the fact that the mouse genome lacks the ISS-40 region entirely, as we have seen in our experiments that the ISS-40 region has the most dominant silencing effect on the inclusion of exon 20. Though the transgenic mice are completely normal and do not show any overall phenotype, due to the presence of the mouse *Ikbkap* gene, the expressed mRNA from the introduced copies of the mutant human *IKBKAP* gene do show a pattern of skipping similar to that of FD patients. This allows us to deliver the best candidate oligo in this mouse model and observe the effects at the level of splicing by RT-PCR analysis of the expressed mRNA using human-specific primers. We are currently investigating whether oligo 421992 leads to an increase in exon 20 included mRNA levels in various tissues in the FD mice upon systemic delivery, as well as in the nervous tissue after direct delivery through either postnatal or prenatal ICV injections, as done previously in case of SMA mice (Hua et al., 2009; Hua et al., 2008). We are also using a commercial human-specific IKAP antibody to monitor if the increase in full-length mRNA levels of *IKBKAP* correlates with the increase in the levels of IKAP protein. In our study, we use osmotic pumps to deliver oligos directly into



the mouse brain. The pump carrying the oligo is surgically placed under the skin on the back of the mice. Once implanted, the pump releases a small and continuous dose of oligo by osmosis. Similar, but more advanced and programmable pumps are already in existence for use in humans, which are transplanted inside the patients and used to deliver pain killers directly to the pain receptors in the spinal cord (Medtronic). Such pumps can be adapted for the delivery of pS-MOE-ASOs to the cerebrospinal fluid, and may be instrumental in future clinical trials.

The antisense technology employed in our experiments not only aided us in finding a therapeutic oligo, but also led to the identification and the characterization of various cis-regulatory elements that are involved in controlling the splicing of *IKBKAP* exon 20 (see above). Further analysis of the ISS-40 region by a 'mutation walk', involving individual transversions of most of the constituent nucleotides, revealed at least two silencer elements, one of which is possibly recognized by hnRNP A1. hnRNPA1 is a well known splicing repressor (see Chapter 1) and in a previous study from our laboratory, it was shown that it is responsible for the silencing of exon 7 inclusion in *SMN2* by binding to a cis-element (ISS-N1) immediately downstream of exon 7 (Hua et al., 2008). However, among the two putative silencer motifs identified in *IKBKAP* in the ISS-40 region, the hnRNP A1 motif is more distally located from exon 20. Still, the presence of the whole ISS-40 region immediately downstream of exon 20 in *IKBKAP* has strong parallels to the regulatory environment of exon 7 in *SMN2*, and raises some intriguing questions. For example, are all or most alternatively spliced exons influenced by a strong silencer immediately downstream of the 5'ss? Or are exon 20 in *IKBKAP* and exon 7 in *SMN2* exceptions? Further, does the extent of silencing depend upon the distance from the 5'ss? Why does the silencer only affect the weaker mutant splice site and not the stronger wild-type splice site? Exploration of these and many related questions may provide us with a more generalized understanding of alternative splicing regulation, which is of prime importance in gene expression and the main contributor of proteomic diversity in higher eukaryotes.

My major focus at present is the identification and characterization of trans-acting factors that bind to the ISS-40 region and regulate the splicing of exon 20 in *IKBKAP*. Knowledge of these factors and their mode of action may provide important clues regarding the mechanics of disease manifestation in FD. For example, these factors may be differentially expressed or their activities may be regulated differently among various tissues, especially neurons, which may explain why the degenerative nature of the disease is specifically restricted to the peripheral nervous system.

In summary, this study is the first of its kind to utilize ASOs toward finding a cure for FD. Our findings have set the stage for work in animal models, which we are actively pursuing with our experiments in transgenic FD mice (Hims et al., 2007). *IKBKAP* knockout mice are embryonic lethal; however, recently, an *Ikbkap*-knockout mouse has been described in which the embryonic lethality was rescued by the human *IKBKAP* gene (Chen et al., 2009). This creates an interesting possibility in which the embryonic lethality in *IKBKAP* knockout mice may be rescued by providing the mutant human *IKBKAP* gene carrying the major FD mutation. This may result in a mouse model that mimics some visible FD phenotypes. Moreover, by varying the copy number of the mutant human *IKBKAP* gene, it may even be possible to modulate the severity of the phenotype. Use of our oligo in such a mouse model to rescue the FD-like phenotypes would be the next logical step and the success of such studies may one day lead us to clinical trials.

### **3.4 Materials and methods**

#### **3.4.1 Oligonucleotide synthesis**

All the MOE-ASOs used in our study were synthesized and purified at ISIS Pharmaceuticals using an Applied Biosystems 380B automated DNA synthesizer (Applied Biosystems) as described previously (Hua et al., 2007). We dissolved the oligos in water before use. The list of all the oligos and their sequences are provided in Table 1.

#### **3.4.2 Plasmids**

We amplified the *IKBKAP* genomic fragments spanning exons 19-21 and 19-22 using specific primers with linkers and human genomic DNA (Promega) as template. We cloned the genomic fragments into pCDNA3.1 (Invitrogen), using the sites BamHI and XbaI, giving rise to pCDNA3.1-wt19-21 and the pCDNA3.1-wt19-22 plasmids. We introduced the major FD mutation (IVS20+6T →C) using site-directed mutagenesis to create the minigenes pCDNA3.1-mt19-21 and pCDNA3.1-mt19-22. All the other transversions for the mutation walk were also introduced by site-directed mutagenesis using pCDNA3.1-mt19-21 as the template. The primers used for amplification of genomic fragments, as well as those used for site-directed mutagenesis, are listed in Tables 2 and 3, respectively.

#### **3.4.3 Cell culture and transfections**

We grew 293 cells in Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% fetal bovine serum (Invitrogen) at 37 °C with 5% CO<sub>2</sub>.

We grew the patient fibroblast cell line GM04899 (Coriell Cell Repository) in minimal essential medium (Invitrogen) supplemented with non-essential amino acids (Invitrogen) and 10% fetal bovine serum also at 37 °C with 5% CO<sub>2</sub>. We cotransfected 3 µg of the minigenes and 100 nM of the pO-MOE-ASOs into 293 cells for the macro- as well as microwalk, by electroporation, as described previously (Hua et al., 2007). We used 12 µl Lipofectamine 2000 transfection reagent (Invitrogen) to transfect different amounts of pO-MOE-ASOs, ranging from 2-800 nM, in 40-50% confluent patient fibroblasts grown in 10 cm dishes, according to the manufacturer's recommendations. For the mutation walk, 3 µg of each individual minigene was transfected into 40-50% confluent 293 cells grown in 10 cm plates using Fugene 6 transfection reagent (Invitrogen) as per the manufacturer's instructions.

#### **3.4.4 RT-PCR**

We extracted total RNA from 293 cells or patient fibroblasts using Trizol reagent (Invitrogen) and reverse transcribed 2 µg from each RNA sample using Superscript-II (Invitrogen) and oligo-dT primers. We amplified the cDNAs corresponding to transcripts of either minigenes or the endogenous *IKBKAP* gene by PCR using AmpliTaq DNA polymerase (Invitrogen) with vector-specific (pCDNA3.1) and human-specific primers, respectively, as per conditions described earlier (Hua et al., 2007). The list of primers used for RT-PCR analysis can be found in Table 4.

To deduce exon 20 inclusion levels, the  $\alpha$ -<sup>32</sup>P-dCTP labeled PCR-amplicons were analyzed by native PAGE followed by phosphorimage analysis on a FUJIFILM FLA-5100 instrument (Fuji Medical Systems USA, Inc.). We quantified the band intensities using Multi Gauge software Version 2.3 (FUJIFILM), which were normalized for G+C content before representation.

#### **3.5 Acknowledgements**

We thank Yimin Hua for helpful discussions and guidance during the course of the project, ISIS Pharmaceuticals for providing the ASOs and Lisa Manche for critical reading of the manuscript.

### 3.6 Tables

**Table 1 List of pO-MOE-ASOs**

<b>Oligo #</b>	<b>Sequence</b>	<b>Length</b>
4141-29	AGAGAATTACCACAA	15mer
4141-30	TTCACAGAGAATTAC	15mer
4141-31	AACTCTTCACAGAGA	15mer
4141-32	TACCTAACTCTTCAC	15mer
4141-33	CATTTTACCTAACTC	15mer
4141-34	TACACCATTTTACCT	15mer
4141-35	CAGGATACACCATTT	15mer
4141-36	ATAGCCAGGATACAC	15mer
4141-37	TTTAAATAGCCAGGA	15mer
4141-38	AAACATTTAAATAGC	15mer
4141-39	GTAGAAAACATTTAA	15mer
4141-40	ATTAAGTAGAAAACA	15mer
4141-41	TTTTAATTAAGTAGA	15mer
4141-42	AACATTTTTAATTAA	15mer
4141-43	GCAGTAACATTTTTTA	15mer
4141-44	TTAAAGCAGTAACAT	15mer
4141-45	ATAAATTAAGCAGT	15mer
4141-46	CTTAAATAAATTTAA	15mer
4141-47	GTTTCCCCTTGGCAT	15mer
4141-48	TCTAAGTTTCCCCTT	15mer
4141-49	CAACTTCTAAGTTTC	15mer
4141-50	ATGAACAACCTTCTAA	15mer
4141-51	CGATGATGAACAACCT	15mer
4141-52	GGGCTCGATGATGAA	15mer
4141-53	AACCAGGGCTCGATG	15mer
4141-54	GCTAAAACCAGGGCT	15mer
4141-55	TCTGAGCTAAAACCA	15mer
4141-56	CCGAATCTGAGCTAA	15mer
4141-57	CACTTCCGAATCTGA	15mer
4141-58	CCAACCACTTCCGAA	15mer
4141-59	TTGTCCAACCACTTC	15mer
4141-60	TACAATGGCGCTTAC	15mer
4141-61	AACAGTACAATGGCG	15mer
4141-62	TCGCAAACAGTACAA	15mer
4141-63	ACTAGTCGCAAACAG	15mer
4141-64	AGCTAACTAGTCGCA	15mer
4141-65	TCACAAGCTAACTAG	15mer

4141-66	ATAAATCACAAGCTA	15mer
4141-67	CACACATAAATCACA	15mer
4141-68	GTCTTCACACATAAA	15mer
4141-69	TTATTGTCTTCACAC	15mer
4141-70	AATACTTATTGTCTT	15mer
4141-71	AATAAAATACTTATT	15mer
4141-72	ATTGTAATAAAATAC	15mer
4141-73	TCGAAATTGTAATAA	15mer
4141-74	AGTTCTCGAAATTGT	15mer
4141-75	TTTTAAGTTCTCGAA	15mer
4141-76	CATAATTTTAAGTTC	15mer
4141-77	CTTTTCATAATTTTA	15mer
42-1991	TCGCAAACAGTACAATGGCG	20mer
42-1992	GTCGCAAACAGTACAATGGC	20mer
42-1993	AGTCGCAAACAGTACAATGG	20mer
42-1994	TAGTCGCAAACAGTACAATG	20mer
42-1995	CTAGTCGCAAACAGTACAAT	20mer
42-1996	ACTAGTCGCAAACAGTACAA	20mer
42-1997	AACTAGTCGCAAACAGTACA	20mer
42-1998	TAAGTTCGCAAACAGTAC	20mer
42-1999	CTAACTAGTCGCAAACAGTA	20mer
42-2000	GCTAACTAGTCGCAAACAGT	20mer

**Table 2 List of PCR primers used to amplify *IKBKAP* genomic fragments**

<b>Primer</b>	<b>Sequence</b>
IKAP19F6	GGGGAAGGATCCGCCATGGAGTTAATGGTGTGTTTAGCATTACAGG
IKAP21R3	GGGGAATCTAGACTTAGGGTTATGATCATAAATCAGATTGAG
IKAP22R	GGGGAATCTAGATTACTTCAATTCTGTAAAAACAAGTTAATATG

**Table 3 List of primers used for RT-PCR**

**a. Primers used to amplify minigene-specific mRNA**

<b>Primer</b>	<b>sequence</b>
pCDNAF	TAATACGACTCACTATAGGG
IKAP21R4	CTTAGGGTTATGATCATAAATCAG
IKAP22R2	TTCAATTCTGTAAAAACAAG

**b. Primers used to amplify endogenous *IKBKAP*-derived mRNA**

<b>Primer</b>	<b>Sequence</b>
IKAPRT19F	GTTCTGCGGAAAGTGGAGAG
IKAPRT21R	CATGCATTCAAATGCCTCTTT

**Table 4 List of primers used for site directed mutagenesis**

<b>Primer</b>	<b>Sequence</b>
IKAPIn20MtR 4	AACTAGTCGCAAACAGTACAATG
IKAPI20+40A TF	ACTGTTTGCGACTAGTTAGCTTGTGTTTTATGTGTGAAGAC AATAAGTATTTT
IKAPI20+39G CF	ACTGTTTGCGACTAGTTAGCTTGTCAATTTATGTGTGAAGAC AATAAGTA
IKAPI20+38T AF	ACTGTTTGCGACTAGTTAGCTTGAGATTTATGTGTGAAGAC AATAAG
IKAPI20+37G CF	GTACTGTTTGCGACTAGTTAGCTTCTGATTTATGTGTGAAG ACAATAAG
IKAPI20+36T AF	GTACTGTTTGCGACTAGTTAGCTAGTGATTTATGTGTGAAG ACAATAA
IKAPI20+35T AF	GTACTGTTTGCGACTAGTTAGCATGTGATTTATGTGTGAAG ACAATA
IKAPI20+34C GF	TGTACTGTTTGCGACTAGTTAGGTTGTGATTTATGTGTGAA GACAA
IKAPI20+33G CF	CATTGTACTGTTTGCGACTAGTTACCTTGTGATTTATGTGTG AAGACA
IKAPI20+32A TF	CATTGTACTGTTTGCGACTAGTTTGCTTGTGATTTATGTGTG AAGAC
MIKAPIn20Mt R3	AAACAGTACAATGGCGCTTAC

IKAPI20+31T AF	GCCATTG TACTG TTTGCGACTAGTAAGCTTGTGATTTATGT GTGAAG
IKAPI20+30T AF	GCCATTG TACTG TTTGCGACTAGATAGCTTGTGATTTATGT GTGAAG
IKAPI20+29G CF	GCCATTG TACTG TTTGCGACTACTTAGCTTGTGATTTATGT GTGAAG
MIKAPI20+28 ATF	GCGCCATTG TACTG TTTGCGACTTGTAGCTTGTGATTTAT GTGTGAAG
MIKAPI20+27 TAF	GCGCCATTG TACTG TTTGCGACAAGTTAGCTTGTGATTTAT GTGTGAA
MIKAP20+26 CGF	GCGCCATTG TACTG TTTGCGAGTAGTTAGCTTGTGATTTAT GTGTG
MIKAPI20+25 ATF	AGCGCCATTG TACTG TTTGCGTCTAGTTAGCTTGTGATTTA TGTGTG
MIKAPI20+24 GCF	AAGCGCCATTG TACTG TTTGCCACTAGTTAGCTTGTGATTT ATGTG
MIKAPI20+23 CGF	GTAAGCGCCATTG TACTG TTTGGGACTAGTTAGCTTGTGAT TTATGTG
MIKAPI20+22 GCF	GTAAGCGCCATTG TACTG TTTCCGACTAGTTAGCTTGTGAT TTATG
MIKAPI <sub>n</sub> 20Mt R2	ATGGCGCTTACTTGTCCAACC
MIKAPI20+21 TAF	GGACAAGTAAGCGCCATTG TACTG TTAGCGACTAGTTAGCT TGTGATT
MIKAPI20+20 TAF	GGACAAGTAAGCGCCATTG TACTGTATGCGACTAGTTAGCT TGTG
MIKAPI20+19 TAF	GGACAAGTAAGCGCCATTG TACTGATTGCGACTAGTTAGCT TGTG
MIKAPI20+18 GCF	GGACAAGTAAGCGCCATTG TACTCTTTGCGACTAGTTAGCT TGTG
MIKAPI20+17 TAF	GGACAAGTAAGCGCCATTG TACAGTTT GCGACTAGTTAGCT TGTG
MIKAPI20+16 CGF	GGACAAGTAAGCGCCATTG TAGTGT TTTGCGACTAGTTAGCT TG
MIKAPI20+15 ATF	GTTGGACAAGTAAGCGCCATTG TTCTG TTTGCGACTAGTTA GCTTG
MIKAPI20+14 TAF	GTTGGACAAGTAAGCGCCATTG AACTG TTTGCGACTAGTTA GC
MIKAPI20+13 GCF	GTTGGACAAGTAAGCGCCATTG TACTG TTTGCGACTAGTTA GC
MIKAPI20+12 TAF	GGTTGGACAAGTAAGCGCCATAGTACTG TTTGCGACTAGTT AG
IKAPI <sub>n</sub> 20MtR 1	CTTACTTGTCCAACCACTTCC
MIKAPI20+11 TAF	GTGGTTGGACAAGTAAGCGCCAATGTACTG TTTGCGACTA GTTAG

MIKAPI20+10 ATF	GTGGTTGGACAAGTAAGCGCCTTTGTA CTGTTTGCGACTAGTTAG
MIKAPI20+9 CGF	GTGGTTGGACAAGTAAGCGCGATTGT ACTGTTTGCGACTAG
MIKAPI20+8 CGF	GAAGTGGTTGGACAAGTAAGCGGCAT TGTACTGTTTGCGACTAG
MIKAPI20+7 GCF	GAAGTGGTTGGACAAGTAAGCCCCAT TGTACTGTTTGCGACTAG
IKAPI20+6TG F	GGAAGTGGTTGGACAAGTAAGGGCC ATTGTACTGTTTGCGACTAC
IKAPI20+6TA F	GGAAGTGGTTGGACAAGTAAGAGCC ATTGTACTGTTTGCGACTAC
MUTIKAP-F	GATTCGGAAGTGGTTGGACAAGTAAG CGCCATTGTACTGTTTGCGACTAGTTAG
MUTIKAP-R	CTAACTAGTCGCAAACAGTACAATGG CGCTTACTTGTCCAACTTCCGAATC
MUTIKAPF2	CTTTCTCTGTCTTCTCACAGACTATG TTTAAAGAGGCATTGAAATG
MUTIKAPR2	CATTCAAATGCCTCTTTAAACATAGT CTGTGAGAAGACAGAGAAAG



### 3.7 Figures and Legends

**Figure 3.1 Genomic structure of *IKBKAP* gene.** *IKBKAP* gene spans across a 68 kb region in the 9q31 region of the Chromosome 9. The gene has 37 exons and gives rise to two mRNA species of different lengths, a shorter 4.8 kb and a longer 5.9 kb. The two different mRNA species are a result of alternative polyadenylation sites and vary only in the lengths of their 3' untranslated region (UTR).

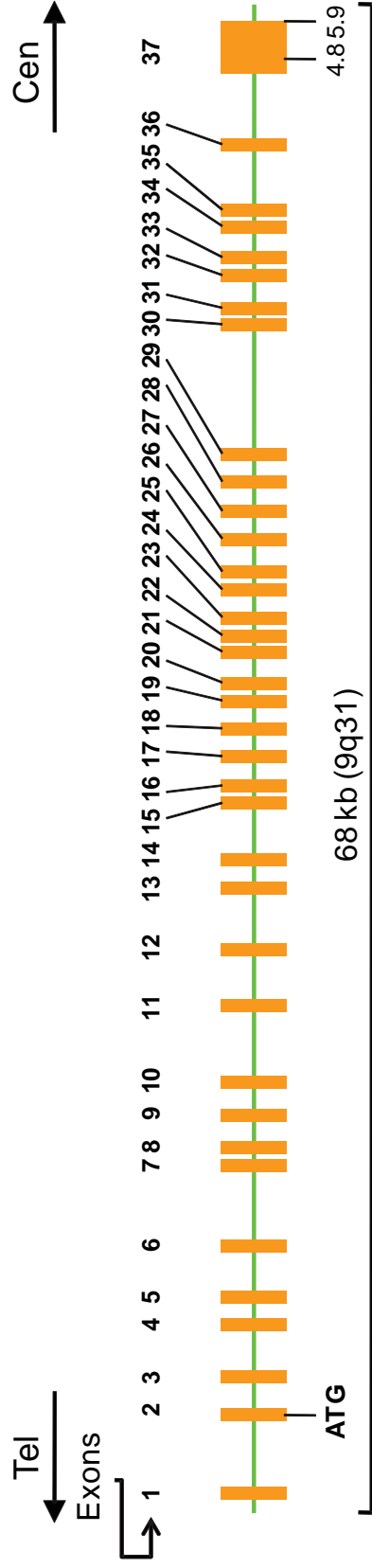


Figure 3.1

**Figure 3.2 The Major FD mutation causes predominant exon 20 skipping in *IKBKAP*.** The genomic region of *IKBKAP* spanning exon 19 to exon 21 is shown. Both wild-type (WT) and the mutant (MT) alleles are shown with the T→C mutation at position +6 in intron 20 shown in the blow up of the 5'ss region. The consensus 5'ss sequence found in humans is also represented for reference in which the height of the letters correspond to their frequency of occurrence (Burge et al., 1998). The T→C change, also known as the major FD mutation, results in predominant skipping of the exon 20. The skipping of 74-nucleotide long exon 20 causes a frame shift, which introduces a premature stop codon (PTC) in the exon 21 in case of the skipped mRNA. The PTC in skipped splice variant of *IKBKAP* not only results in a truncated (699 amino acids instead of 1332 amino acids) but also, in theory, makes the mRNA susceptible to degradation via nonsense-mediated mRNA decay (NMD).

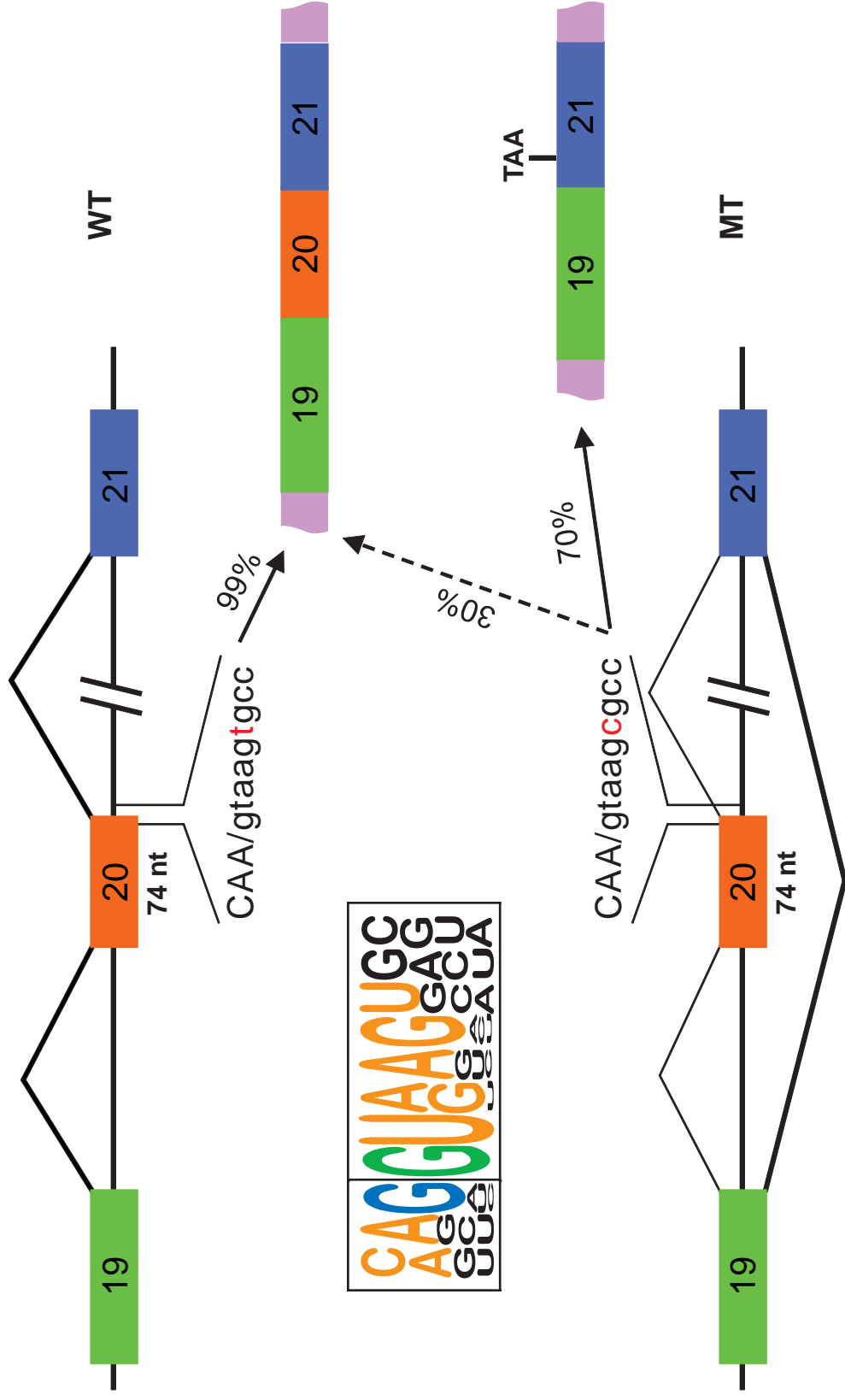
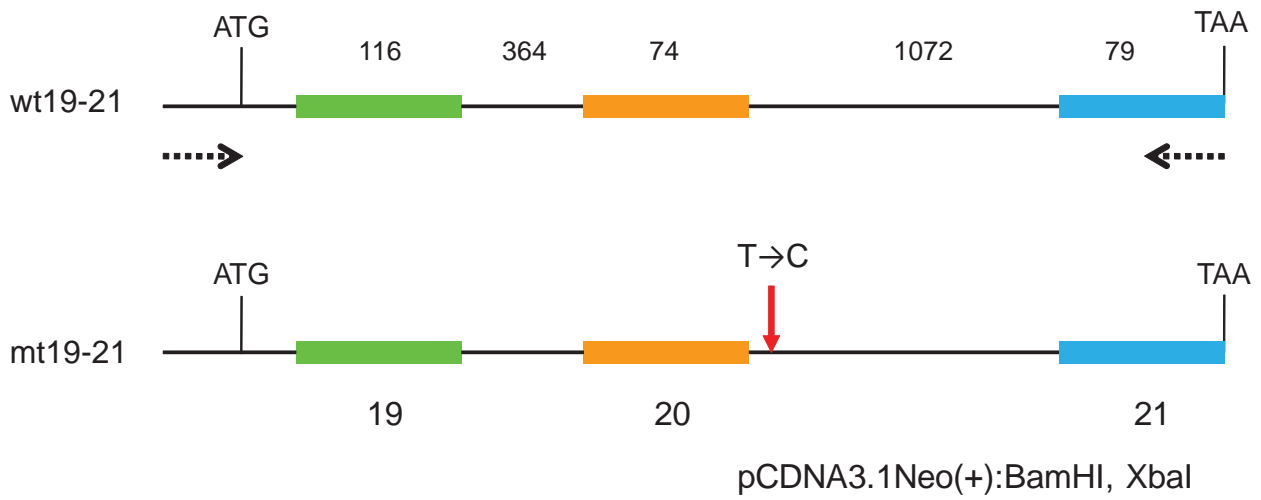


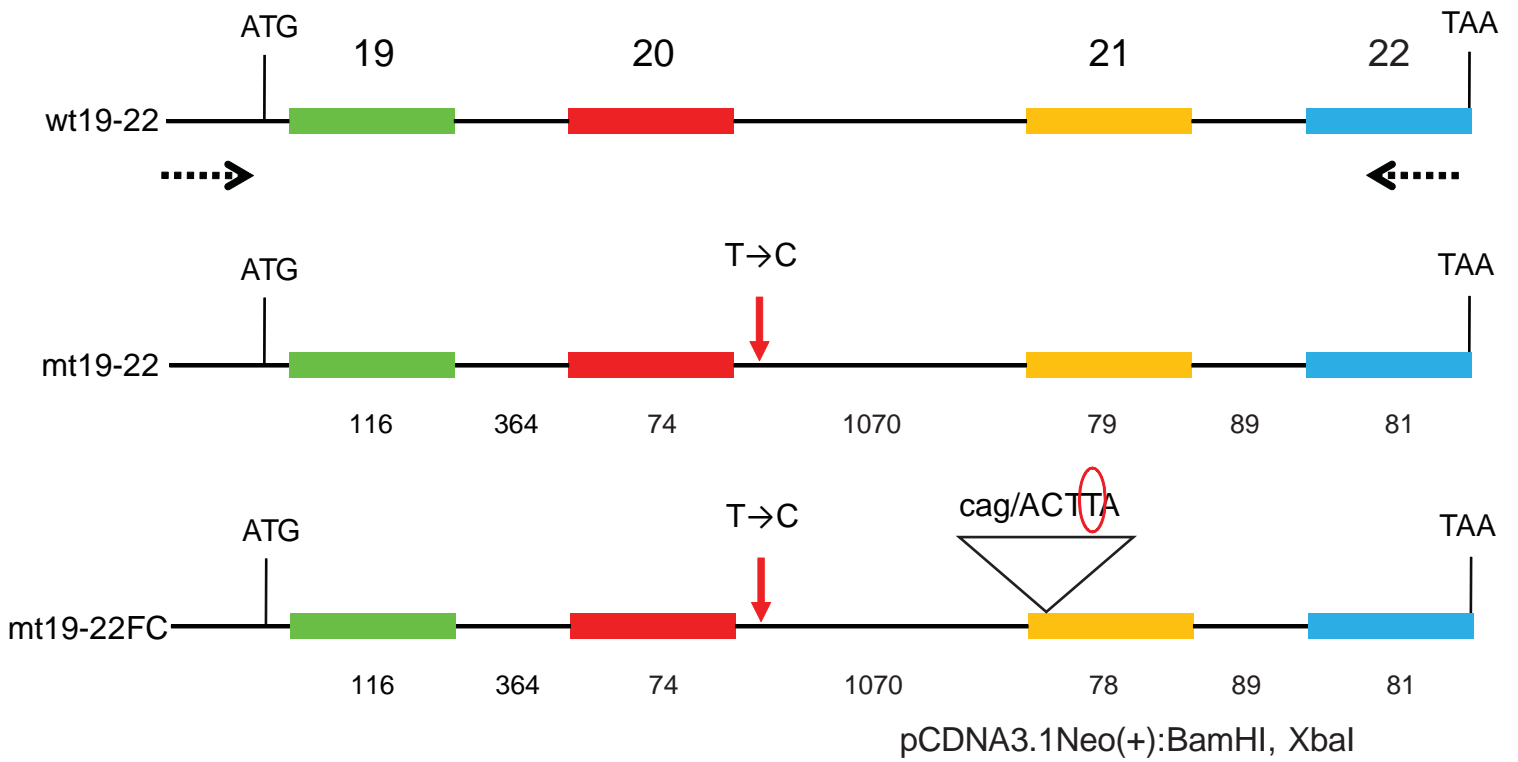
Figure 3.2

**Figure 3.3 Schematic representation of the minigene constructs.** (A) Both wild-type (WT) and the mutant (MT) versions of the 19-21 series of minigenes are shown. The major FD mutation (IVS+6T →C) was introduced in the case of MT minigene by site directed mutagenesis and is shown by a red arrow. The broken arrows represent the location of the complementary sequences of the minigene-specific primers, which were used for RT-PCR analysis (Table 3.3a). The lengths of each individual exon and intron are depicted at the top in smaller font. The restriction sites used to clone the the minigenes cassette into pCDNA3.1Neo(+) are mentioned at the bottom. (B) Same as in (A) except this panel shows the 19-22 series of minigene constructs. In the case of the mt19-22FC minigene, the red circle shows the deleted T nucleotide from exon 21, which occurs at position +4 in the natural context. The deletion of T nucleotide restores the frameshift that is caused by skipping of exon 20 in the skipped splice variant.

**A** Minigene constructs 19-21

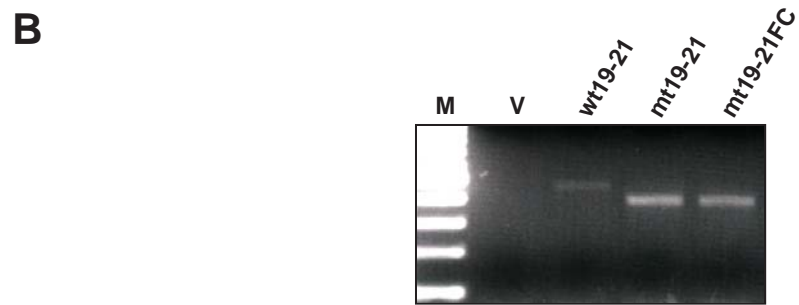
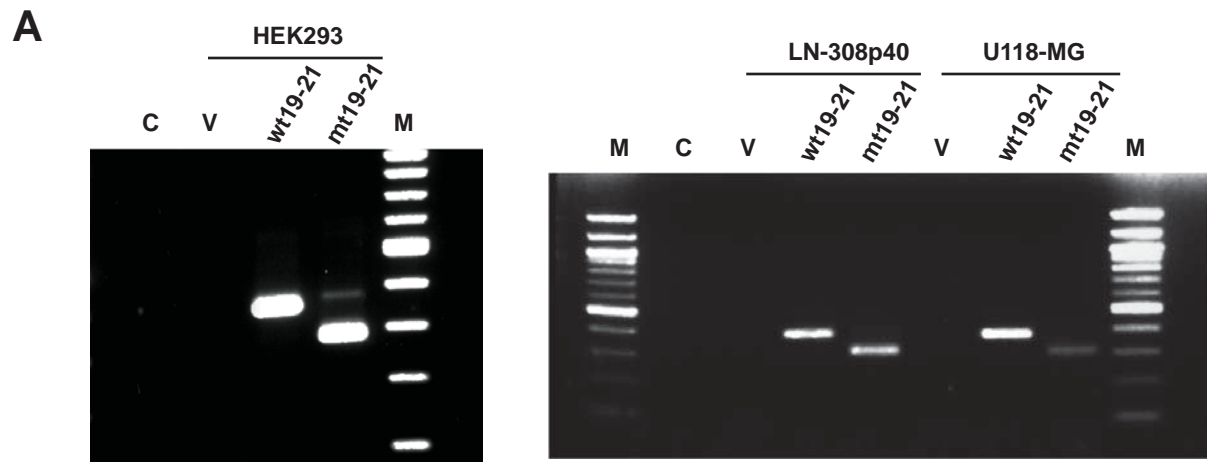


**B** Minigene constructs 19-22



**Figure 3.3**

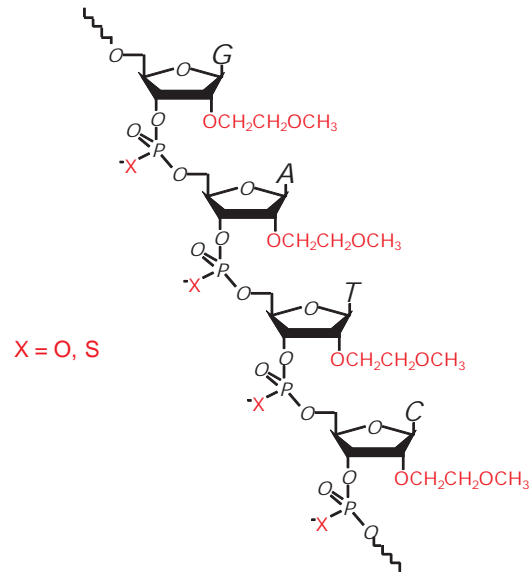
**Figure 3.4 Splicing patterns of expressed pre-mRNAs from the minigenes.** (A) We transfected the empty vector (V), the wt19-21, and the mt19-21 minigenes in three different cell lines, which are: human epithelial kidney 293 (HEK293) cell line and the two neuroblastoma lines, LN308p40 and U118-MG. The panels show the ethidium bromide stained agarose gels showing the pattern of splicing of the expressed RNA as analyzed by RT-PCR using minigene specific primers shown in Fig. 3.3a and Table 3.3a. M: marker, C: PCR water control, V: pCDNA3.1Neo(+). (B) Same as in (A) except this panel shows pattern of splicing of the expressed 19-22 minigenes in HEK293 cells as analyzed by RT-PCR using primers described in Fig. 3.3b and Table 3.3a.



**Figure 3.4**



**Figure 3.5 The MOE-ASO structures.** The figure shows the structure and modifications of the anti-sense oligonucleotides used in the study. In all oligos the 2' OH in the ribose is replaced by a 2' methoxyethyl group. When used in animal model systems, the phosphate in the backbone is replaced by phosphorothioate moiety.



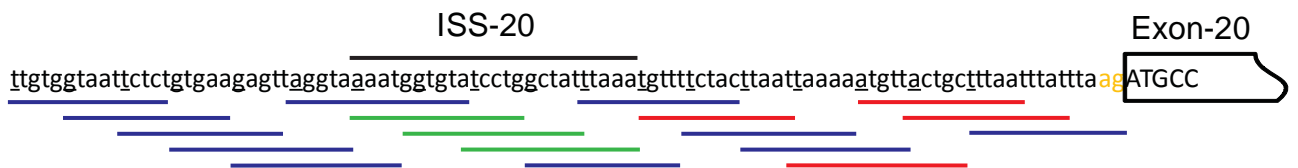
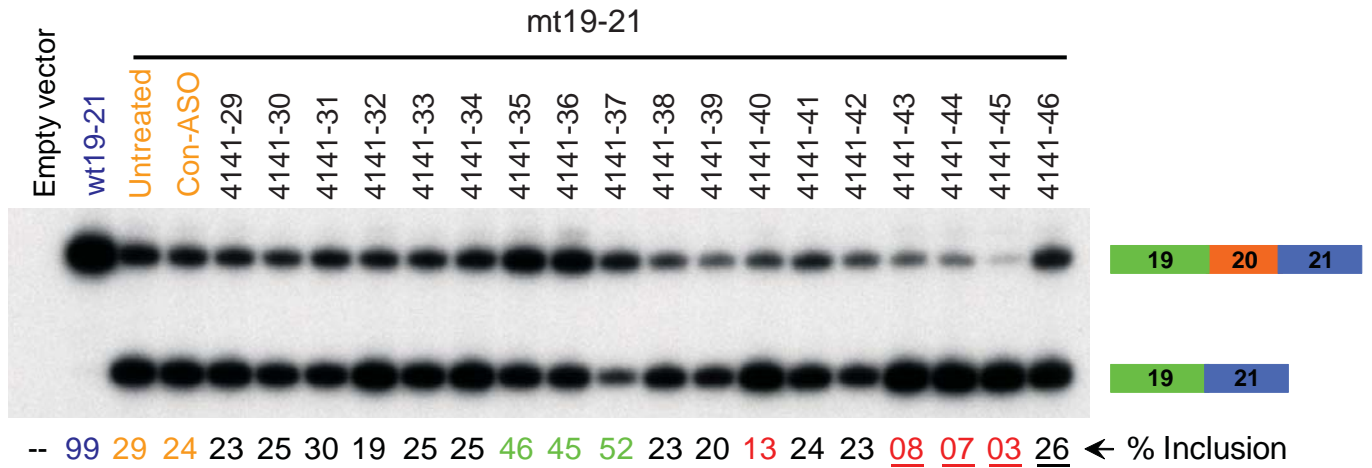
P=O (phosphodiester)  
P=S (2'-deoxy-phosphorothioate-RNA)

2'-O-methoxyethyl ribose backbone

**Figure 3.5**

**Figure 3.6 Coarse oligo-walk for upstream intron 19 using 19-21 minigenes.** We cotransfected 293 cells with the mt19-21 minigene along with the individual 15mer pO-MOE-ASOs by electroporation; two days later we isolated total RNA and performed reverse transcription (RT) to obtain the expressed cDNAs, which were then subjected to PCR analysis in the presence of radioactive  $\alpha$ -<sup>32</sup>P-dCTP using primers described in Fig. 3.3a and Table 3.3a. The panel shows the scanned phosphorimage of the labeled PCR products after they were resolved by native PAGE. We quantified the band intensities, normalized for G+C content, and then calculated the ratio of included to skipped band, which are depicted as inclusion percentages at the bottom of each individual lane. As a reference, the first and the second lanes show exon 20 inclusion levels in the expressed mRNAs from the wt19-21 and mt19-21 minigenes, respectively, in the absence of any ASO. The third lane shows mRNA levels when the mt19-21 minigene was cotransfected with a control oligo of unrelated sequence and serves as a control for non-specific effects of pO-MOE-ASO on splicing. The inclusion ratios where an ASO is suspected to target the cis splice-signals are underlined. Tiling of the 100-nucleotide region immediately upstream of exon 20 by overlapping ASOs at 5-nucleotide intervals is schematically represented at the bottom. Each underlined nucleotide in the intronic sequence (lower case letters) marks the start of the target sequence by an ASO. Each line below the intronic sequence represents an ASO and is coded by color based on its effect on exon 20 inclusion; blue: neutral, green: positive and red: negative. The ISS-20 region is demarcated by a line above the intronic sequence.

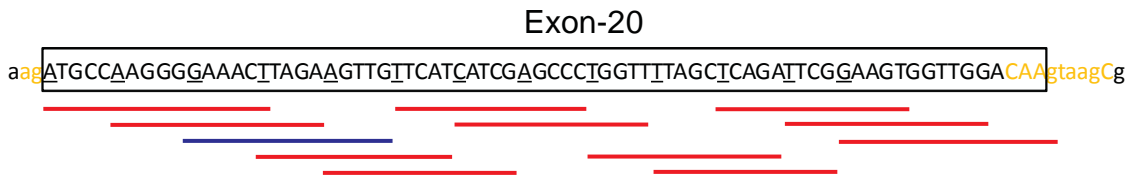
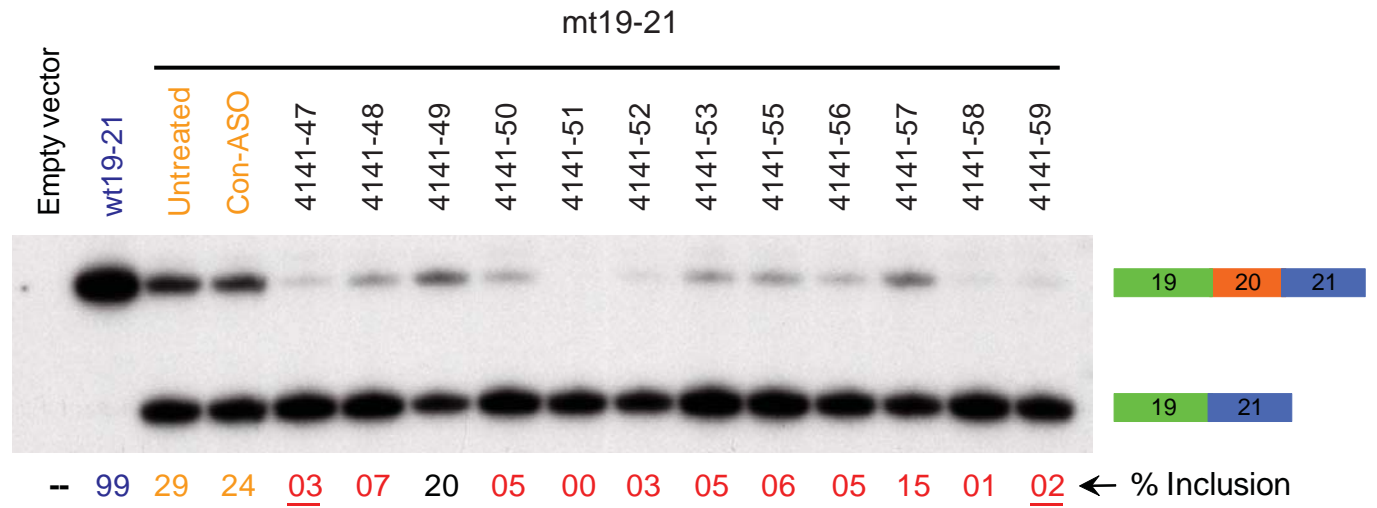
## Antisense oligo-walk for usptream intron-19



**Figure 3.6**

**Figure 3.7 Coarse oligo-walk for exon 20 using 19-21 minigenes.** Same as in Fig. 3.6. The exonic sequence is shown in upper case letters. The conserved AG dinucleotide of the 3'ss and the 5'ss sequence at the ends of the exon are shown in orange.

## Antisense oligo-walk for exon-20



**Figure 3.7**

**Figure 3.8 Coarse oligo-walk for downstream intron 20 using 19-21 minigenes.** Same as Fig. 3.6. The ISS-40 as well as the ISE-20 are marked by lines and shown above the intronic sequence.

## Antisense oligo-walk for downstream exon-20

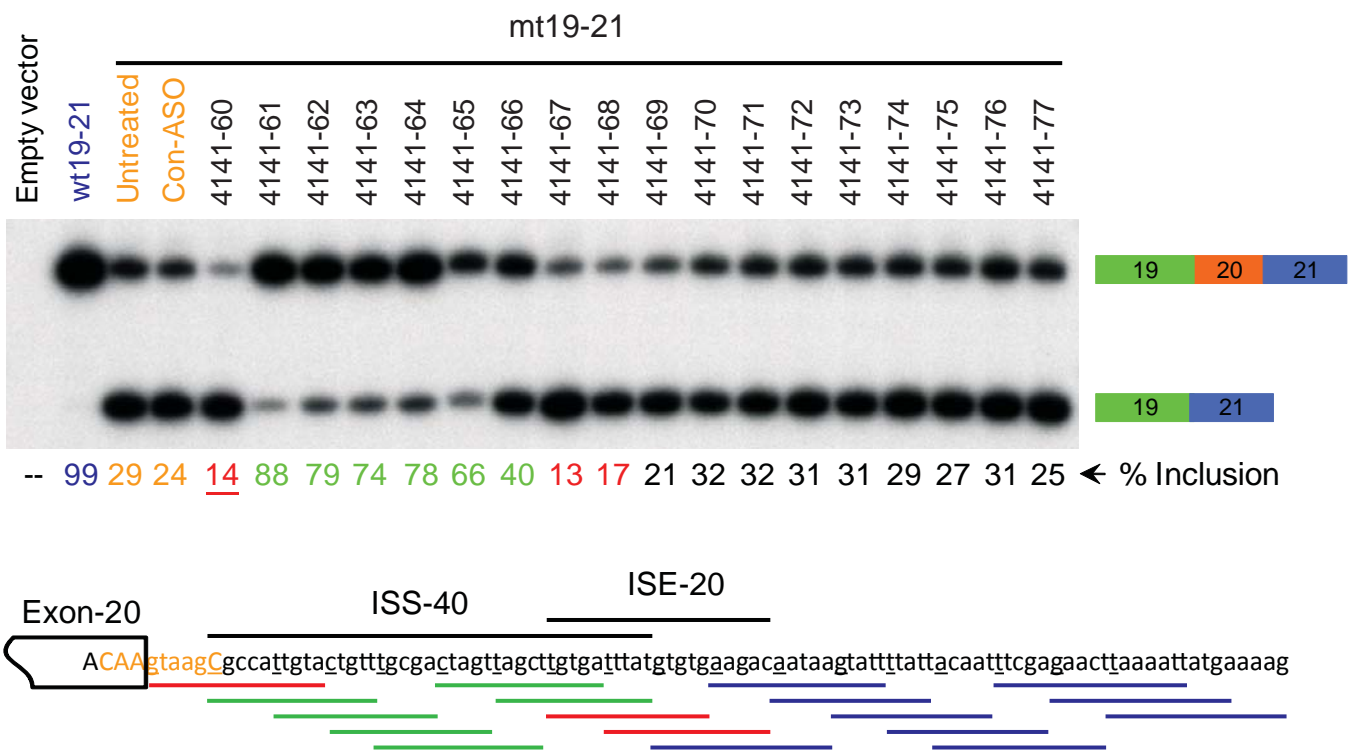
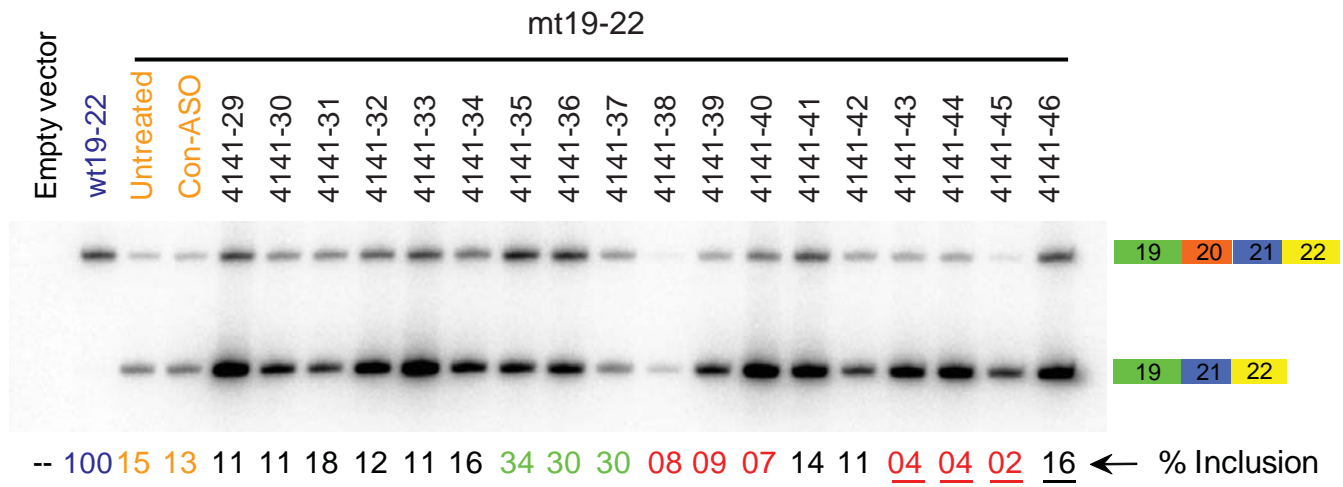


Figure 3.8



**Figure 3.9 Coarse oligo-walk for upstream intron 19 using 19-22 minigenes.** We cotransfected the mt19-22 minigenes with the individual 15mer pO-MOE-ASO into 293 cells and performed RT-PCR analysis as shown in Fig. 3.6. First two lanes with the wt19-22 and the mt19-22 transfected alone serve as references. The cotransfection of mt19-22 with an unrelated oligo in the third lane serves as control. The band intensities were normalized and the inclusion ratios were calculated as in Fig. 3.6 and are presented as inclusion percentages at the bottom of each lane. The tiling as well as the effects of individual oligos is schematically represented at the bottom as in Fig. 3.6.

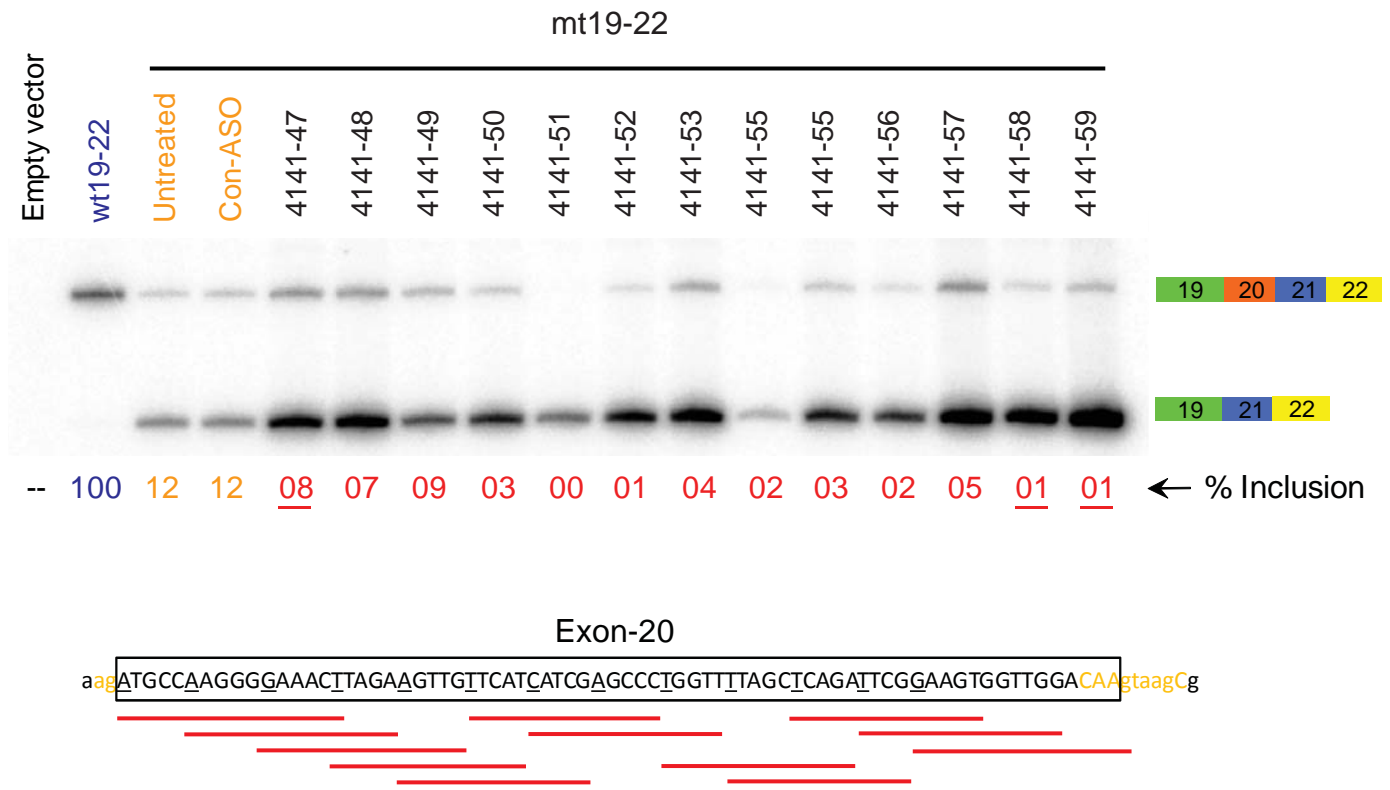
## Antisense oligo-walk for usstream intron-19



**Figure 3.9**

**Figure 3.10 Coarse oligo-walk for exon 20 using 19-22 minigenes.**  
Same as Fig. 3.9.

## Antisense oligo-walk for exon-20



**Figure 3.10**

**Figure 3.11 Coarse oligo-walk for downstream intron 20 using 19-22 minigenes. Same as Fig. 3.9.**

## Antisense oligo-walk for downstream exon-20

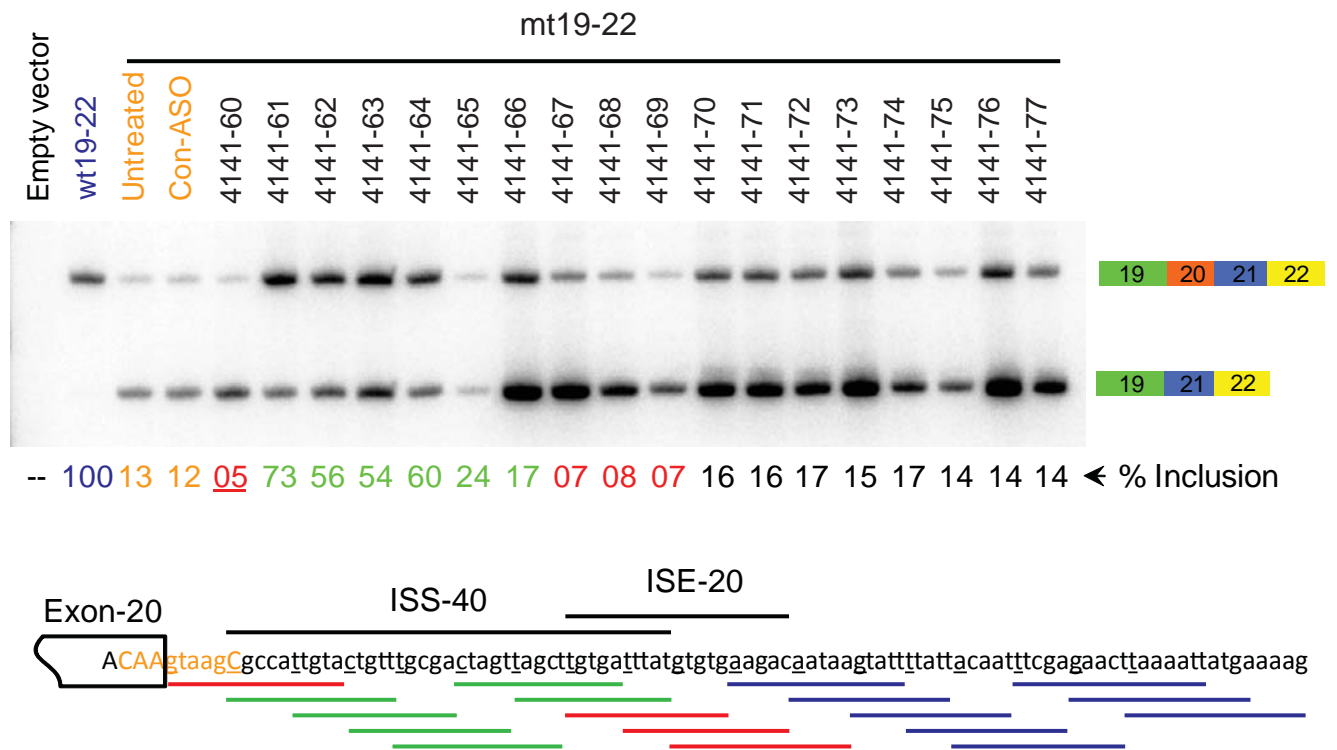
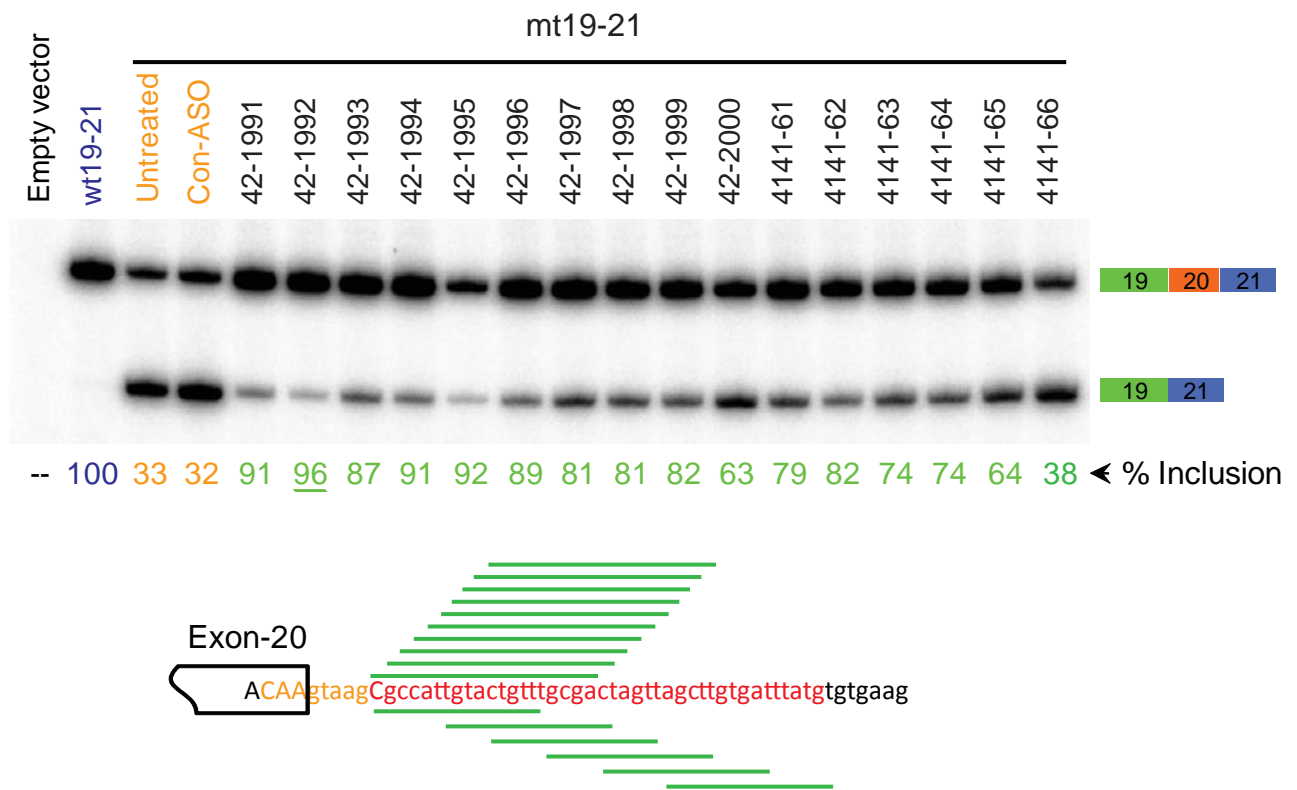


Figure 3.11

**Figure 3.12 High resolution microwalk for the ISS-40 region using 19-21 minigenes.** We cotransfected the mt19-21 minigenes with the individual 20mer pO-MOE-ASO into 293 cells and performed RT-PCR analysis as shown in Fig. 3.6. First two lanes with the wt19-21 and the mt19-21 transfected alone serve as references. The cotransfection of mt19-21 with an unrelated oligo in the third lane serves as control. Cotransfections with the 15mer oligos (414161-66), which had positive effects in the earlier experiment (Fig. 3.8) were included for comparison. The inclusion ratios were calculated as in Fig. 3.6 and are presented as percent-inclusion at the bottom of each lane. The first 30 nucleotides of the ISS-40 region (shown in red) were tiled with the overlapping 20mer ASOs (421991-2000) at one-nucleotide intervals and are shown schematically at the bottom following the same color scheme as in Fig. 3.6. The tiling of 15mer oligos used for comparison is also shown.

## Antisense oligo microwalk in ISS-40 region

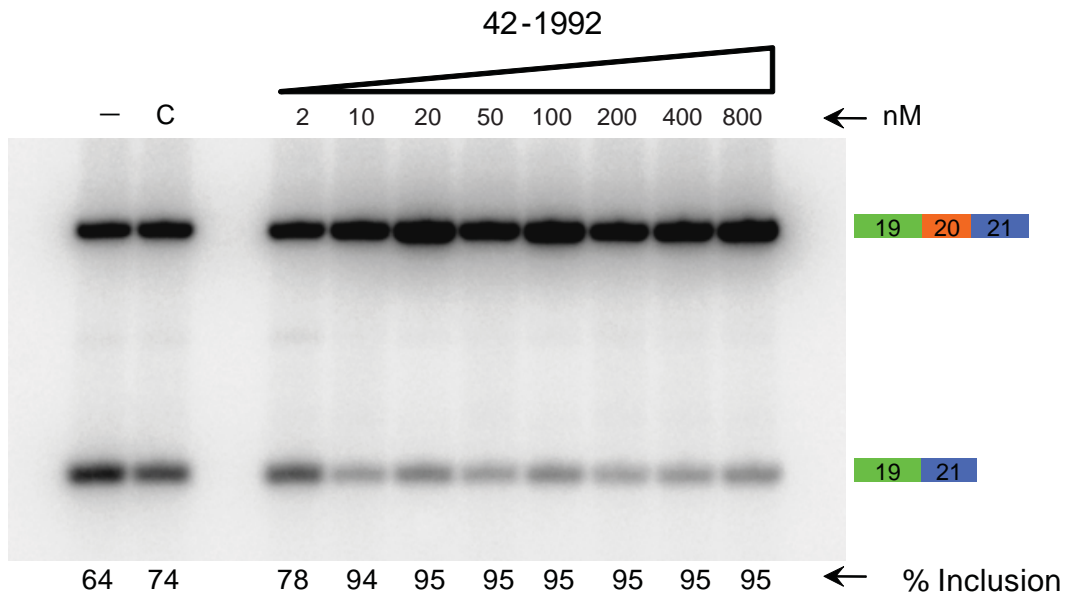


**Figure 3.12**

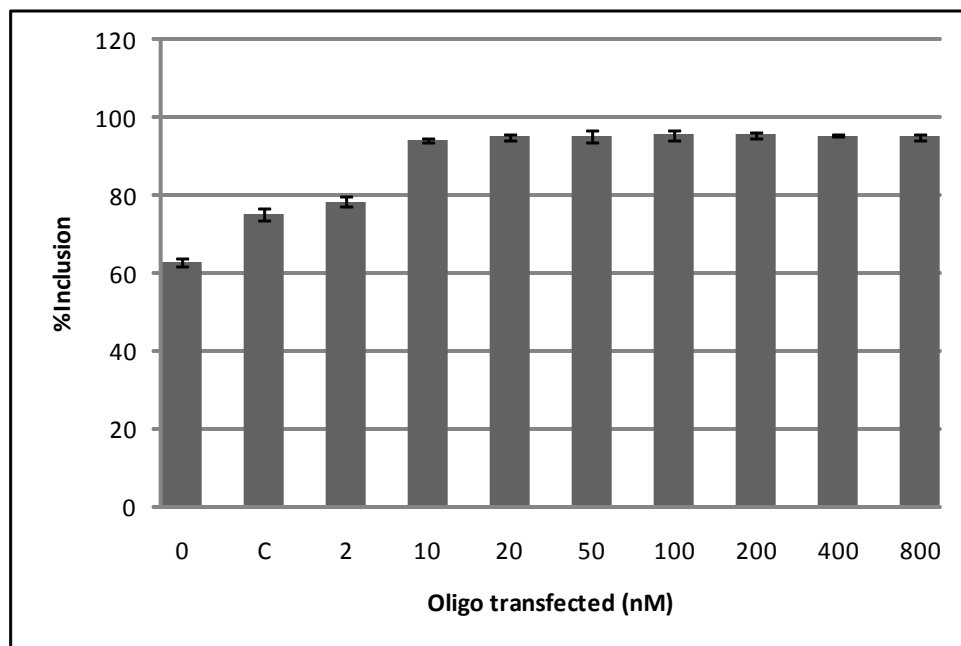


**Figure 3.13 Effect of best candidate oligo in patient fibroblasts and dose response.** (A) We transfected the patient-derived GM04899 cell line with increasing doses (2, 10, 20, 50, 100, 200, 400, and 800 nM) of the 421992 oligo using Lipofectamin 2000 and analyzed the levels of endogenous *IKBKAP* mRNA levels by radioactive RT-PCR using specific primers shown in Table 3.3b. The radioactive RT-PCR analysis and the subsequent calculation of inclusion percentages were performed as in Fig. 3.6. (B) The inclusion-percentages from three independent experiments as in (A) are plotted as a function of oligo dosage. Error bars = standard deviation.

**A**



**B**



**Figure 3.13**

GM04899; Lipofectamine 2000

**Figure 3.14 Mutation analysis of the first half of ISS-40.** We mutated each nucleotide of the ISS-40 region, one at a time, starting at position +7 in intron 20, by site directed mutagenesis using the mt19-21 minigene as template giving rise to a series of minigenes, which carried a single transversion in the ISS-40 region in context of the major FD mutation. We also changed the thymine at position +6 in intron 20 in the wt19-21minigene to either a guanine or adenine giving rise to the +6T →G and +6T→A mutants, respectively, which are analogous to the mt19-21 carrying the major FD mutation (+6T →C). We transfected wt19-21, mt19-21, +6T→G, +6T→A, and all the mt19-21 minigenes with individual transversions from intron 20 position +7 to position +20 into 293 cells using Fugene 6 transfection reagent. We isolated the total RNA days after transfection, analyzed the expressed mRNA isoforms by RT-PCR, and calculated the exon 20 inclusion-percentages as done in Fig. 3.6. The nucleotide transversions, which showed a positive effect on exon 20 inclusion in context of the major FD mutation, are labeled as green. The putative silencer motif (or part of a silencer motif) is underlined.

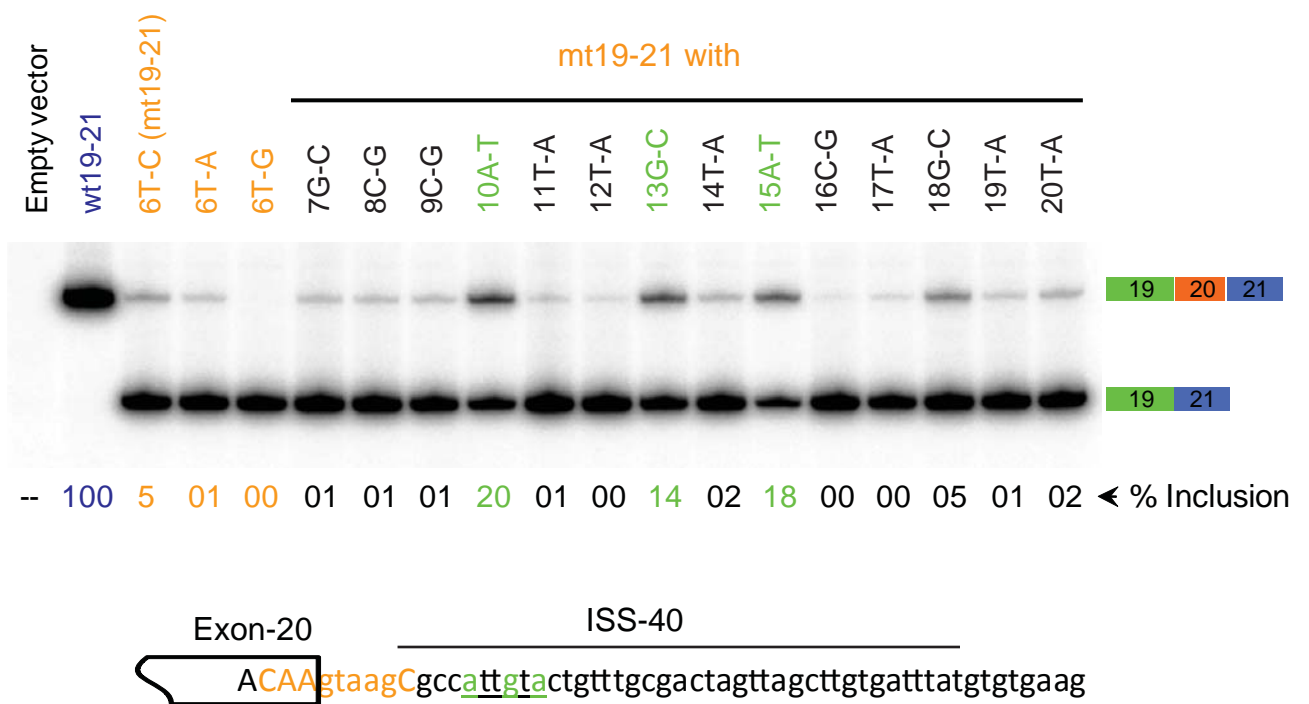


Figure 3.14

**Figure 3.15 Mutation analysis of the second half of ISS-40.** Same as Fig. 3.14. The wt19-21, mt19-21, and the different mt19-21 minigenes with individual transversions from position +21 to position +37 in intron 20 were transfected and analyzed. The nucleotide transversions, which have the positive effect on the inclusion of exon 20 are shown in green and the putative hnRNP core motifs are underlined.

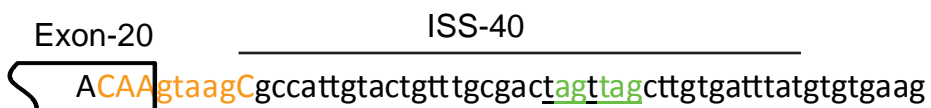
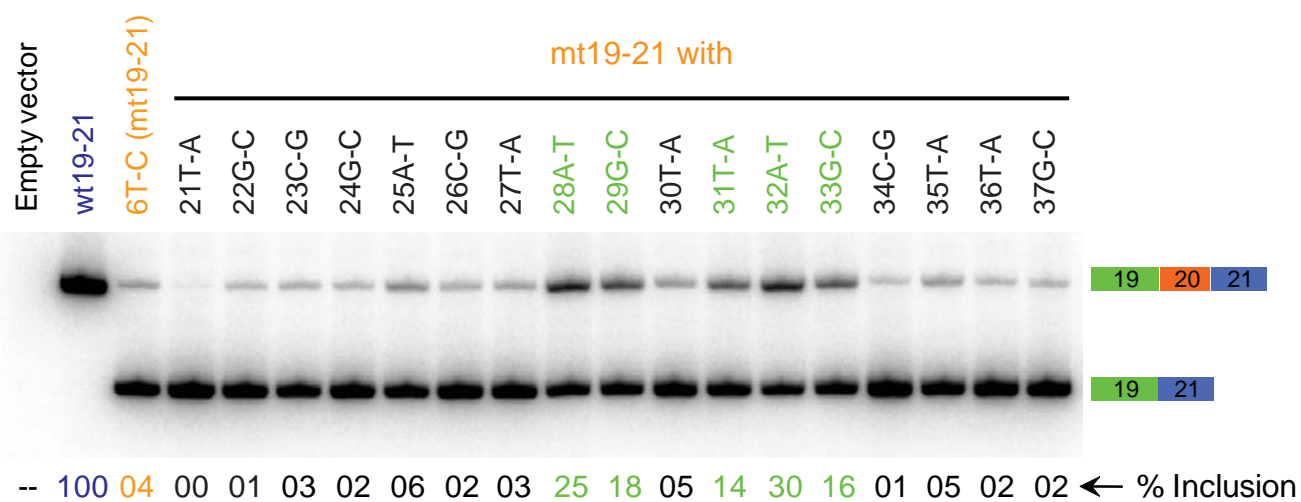


Figure 3.15

## References

### Chapter 1

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

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

Ashe, M.P., Griffin, P., James, W., and Proudfoot, N.J. (1995). Poly(A) site selection in the HIV-1 provirus: inhibition of promoter-proximal polyadenylation by the downstream major splice donor site. *Genes Dev* 9, 3008-3025.

Ashe, M.P., Pearson, L.H., and Proudfoot, N.J. (1997). The HIV-1 5' LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. *Embo J* 16, 5752-5763.

Barksdale, S., and Baker, C.C. (1995). Differentiation-specific alternative splicing of bovine papillomavirus late mRNAs. *J Virol* 69, 6553-6556.

Beggs, J.D. (2005). Lsm proteins and RNA processing. *Biochem Soc Trans* 33, 433-438.

Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H., and Luhrmann, R. (2007). Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *Embo J* 26, 1737-1748.

Bentley, D.L. (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr Opin Cell Biol* 17, 251-256.

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

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

Berven, L.A., and Crouch, M.F. (2000). Cellular function of p70S6K: a role in regulating cell motility. *Immunol Cell Biol* 78, 447-451.

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

- Bhaskar, P.T., and Hay, N. (2007). The two TORCs and Akt. *Dev Cell* 12, 487-502.
- Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 72, 291-336.
- Bourdon, J.C., Fernandes, K., Murray-Zmijewski, F., Liu, G., Diot, A., Xirodimas, D.P., Saville, M.K., and Lane, D.P. (2005). p53 isoforms can regulate p53 transcriptional activity. *Genes Dev* 19, 2122-2137.
- Brinkman, B.M. (2004). Splice variants as cancer biomarkers. *Clin Biochem* 37, 584-594.
- Brow, D.A. (2002). Allosteric cascade of spliceosome activation. *Annu Rev Genet* 36, 333-360.
- Burge, C.B., Padgett, R.A., and Sharp, P.A. (1998). Evolutionary fates and origins of U12-type introns. *Mol Cell* 2, 773-785.
- Burge, C.B., Tuschl, T., and Sharp, P.A. (1999). 20 Splicing of Precursors to mRNAs by the Spliceosomes.
- Cáceres, J.F., Sreaton, G.R., and Krainer, A.R. (1998). A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes Dev* 12, 55-66.
- Cartegni, L., Chew, S.L., and Krainer, A.R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 3, 285-298.
- Castelo-Branco, P., Furger, A., Wollerton, M., Smith, C., Moreira, A., and Proudfoot, N. (2004). Polypyrimidine tract binding protein modulates efficiency of polyadenylation. *Mol Cell Biol* 24, 4174-4183.
- Chang, Y.F., Imam, J.S., and Wilkinson, M.F. (2007). The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem* 76, 51-74.
- Cheng, H., Dufu, K., Lee, C.S., Hsu, J.L., Dias, A., and Reed, R. (2006). Human mRNA export machinery recruited to the 5' end of mRNA. *Cell* 127, 1389-1400.
- Chow, L.T., Gelinas, R.E., Broker, T.R., and Roberts, R.J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 12, 1-8.



Cooper, T.A., Wan, L., and Dreyfuss, G. (2009). RNA and disease. *Cell* 136, 777-793.

Crick, F. (1979). Split genes and RNA splicing. *Science* 204, 264-271.

Custodio, N., Carvalho, C., Condado, I., Antoniou, M., Blencowe, B.J., and Carmo-Fonseca, M. (2004). In vivo recruitment of exon junction complex proteins to transcription sites in mammalian cell nuclei. *Rna* 10, 622-633.

Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Gygi, S.P., and Reed, R. (2007). SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol Cell* 26, 867-881.

Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H., Urlaub, H., and Luhrmann, R. (2006). Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol Cell Biol* 26, 5528-5543.

Dostie, J., and Dreyfuss, G. (2002). Translation is required to remove Y14 from mRNAs in the cytoplasm. *Curr Biol* 12, 1060-1067.

Dreyfuss, G., Kim, V.N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol* 3, 195-205.

Dufner, A., and Thomas, G. (1999). Ribosomal S6 kinase signaling and the control of translation. *Exp Cell Res* 253, 100-109.

DuHadaway, J.B., Lynch, F.J., Brisbay, S., Bueso-Ramos, C., Troncoso, P., McDonnell, T., and Prendergast, G.C. (2003). Immunohistochemical analysis of Bin1/Amphiphysin II in human tissues: diverse sites of nuclear expression and losses in prostate cancer. *J Cell Biochem* 88, 635-642.

Erikson, E., and Maller, J.L. (1985). A protein kinase from *Xenopus* eggs specific for ribosomal protein S6. *Proc Natl Acad Sci U S A* 82, 742-746.

Feng, L.X., Ravindranath, N., and Dym, M. (2000). Stem cell factor/c-kit up-regulates cyclin D3 and promotes cell cycle progression via the phosphoinositide 3-kinase/p70 S6 kinase pathway in spermatogonia. *J Biol Chem* 275, 25572-25576.

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

- Fu, X.D. (2004). Towards a splicing code. *Cell* 119, 736-738.
- Furger, A., O'Sullivan, J.M., Binnie, A., Lee, B.A., and Proudfoot, N.J. (2002). Promoter proximal splice sites enhance transcription. *Genes Dev* 16, 2792-2799.
- Ge, K., Duhadaway, J., Sakamuro, D., Wechsler-Reya, R., Reynolds, C., and Prendergast, G.C. (2000a). Losses of the tumor suppressor BIN1 in breast carcinoma are frequent and reflect deficits in programmed cell death capacity. *Int J Cancer* 85, 376-383.
- Ge, K., Minhas, F., Duhadaway, J., Mao, N.C., Wilson, D., Buccafusca, R., Sakamuro, D., Nelson, P., Malkowicz, S.B., Tomaszewski, J., and Prendergast, G.C. (2000b). Loss of heterozygosity and tumor suppressor activity of Bin1 in prostate carcinoma. *Int J Cancer* 86, 155-161.
- Ghigna, C., Giordano, S., Shen, H., Benvenuto, F., Castiglioni, F., Comoglio, P.M., Green, M.R., Riva, S., and Biamonti, G. (2005). Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. *Mol Cell* 20, 881-890.
- Gornemann, J., Kotovic, K.M., Hujer, K., and Neugebauer, K.M. (2005). Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol Cell* 19, 53-63.
- Gozani, O., Feld, R., and Reed, R. (1996). Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev* 10, 233-243.
- Gozani, O., Potashkin, J., and Reed, R. (1998). A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol Cell Biol* 18, 4752-4760.
- Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. *Rna* 6, 1197-1211.
- Graveley, B.R. (2008). The haplo-spliceo-transcriptome: common variations in alternative splicing in the human population. *Trends Genet* 24, 5-7.
- Graveley, B.R., Hertel, K.J., and Maniatis, T. (2001). The role of U2AF35 and U2AF65 in enhancer-dependent splicing. *Rna* 7, 806-818.

- Gunderson, S.I., Polycarpou-Schwarz, M., and Mattaj, I.W. (1998). U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol Cell* 1, 255-264.
- He, Y., and Smith, R. (2009). Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. *Cell Mol Life Sci* 66, 1239-1256.
- Huang, Y., and Steitz, J.A. (2005). SRprises along a messenger's journey. *Mol Cell* 17, 613-615.
- Inoue, K., Ohno, M., Sakamoto, H., and Shimura, Y. (1989). Effect of the cap structure on pre-mRNA splicing in *Xenopus* oocyte nuclei. *Genes Dev* 3, 1472-1479.
- Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R., and Shoemaker, D.D. (2003). Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302, 2141-2144.
- Jurica, M.S., and Moore, M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* 12, 5-14.
- Kalnina, Z., Zayakin, P., Silina, K., and Line, A. (2005). Alterations of pre-mRNA splicing in cancer. *Genes Chromosomes Cancer* 42, 342-357.
- Karni, R., de Stanchina, E., Lowe, S.W., Sinha, R., Mu, D., and Krainer, A.R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol* 14, 185-193.
- Karni, R., Hippo, Y., Lowe, S.W., and Krainer, A.R. (2008). The splicing-factor oncoprotein SF2/ASF activates mTORC1. *Proc Natl Acad Sci U S A* 105, 15323-15327.
- Kiss, T. (2004). Biogenesis of small nuclear RNPs. *J Cell Sci* 117, 5949-5951.
- Kohler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. *Nat Rev Mol Cell Biol* 8, 761-773.
- Konarska, M.M., Vilardell, J., and Query, C.C. (2006). Repositioning of the reaction intermediate within the catalytic center of the spliceosome. *Mol Cell* 21, 543-553.

- Kornblihtt, A.R. (2005). Promoter usage and alternative splicing. *Curr Opin Cell Biol* 17, 262-268.
- Kornblihtt, A.R., de la Mata, M., Fededa, J.P., Munoz, M.J., and Nogues, G. (2004). Multiple links between transcription and splicing. *Rna* 10, 1489-1498.
- Krainer, A.R., Conway, G.C., and Kozak, D. (1990). Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes Dev* 4, 1158-1171.
- Krawczak, M., Reiss, J., and Cooper, D.N. (1992). The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet* 90, 41-54.
- Kwan, T., Benovoy, D., Dias, C., Gurd, S., Provencher, C., Beaulieu, P., Hudson, T.J., Sladek, R., and Majewski, J. (2008). Genome-wide analysis of transcript isoform variation in humans. *Nat Genet* 40, 225-231.
- Lacadie, S.A., and Rosbash, M. (2005). Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast. *Mol Cell* 19, 65-75.
- Lai, M.C., Lin, R.I., and Tarn, W.Y. (2001). Transportin-SR2 mediates nuclear import of phosphorylated SR proteins. *Proc Natl Acad Sci U S A* 98, 10154-10159.
- Lambert, J.M., Karnoub, A.E., Graves, L.M., Campbell, S.L., and Der, C.J. (2002). Role of MLK3-mediated activation of p70 S6 kinase in Rac1 transformation. *J Biol Chem* 277, 4770-4777.
- Le Hir, H., and Andersen, G.R. (2008). Structural insights into the exon junction complex. *Curr Opin Struct Biol* 18, 112-119.
- Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J. (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *Embo J* 19, 6860-6869.
- Le Hir, H., and Seraphin, B. (2008). EJC's at the heart of translational control. *Cell* 133, 213-216.
- Lear, A.L., Eperon, L.P., Wheatley, I.M., and Eperon, I.C. (1990). Hierarchy for 5' splice site preference determined in vivo. *J Mol Biol* 211, 103-115.

- Lejeune, F., Ishigaki, Y., Li, X., and Maquat, L.E. (2002). The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. *Embo J* 21, 3536-3545.
- Lewis, J.D., Izaurralde, E., Jarmolowski, A., McGuigan, C., and Mattaj, I.W. (1996). A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev* 10, 1683-1698.
- Li, X., and Manley, J.L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* 122, 365-378.
- Li, X., Wang, J., and Manley, J.L. (2005). Loss of splicing factor ASF/SF2 induces G2 cell cycle arrest and apoptosis, but inhibits internucleosomal DNA fragmentation. *Genes Dev* 19, 2705-2714.
- Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., and Fu, X.D. (2008). The splicing factor SC35 has an active role in transcriptional elongation. *Nat Struct Mol Biol* 15, 819-826.
- Lin, S., Xiao, R., Sun, P., Xu, X., and Fu, X.D. (2005). Dephosphorylation-dependent sorting of SR splicing factors during mRNP maturation. *Mol Cell* 20, 413-425.
- Liu, H.X., Cartegni, L., Zhang, M.Q., and Krainer, A.R. (2001). A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. *Nat Genet* 27, 55-58.
- Long, J.C., and Cáceres, J.F. (2009). The SR protein family of splicing factors: master regulators of gene expression. *Biochem J* 417, 15-27.
- Loomis, R.J., Naoe, Y., Parker, J.B., Savic, V., Bozovsky, M.R., Macfarlan, T., Manley, J.L., and Chakravarti, D. (2009). Chromatin binding of SRp20 and ASF/SF2 and dissociation from mitotic chromosomes is modulated by histone H3 serine 10 phosphorylation. *Mol Cell* 33, 450-461.
- Lopez-Bigas, N., Audit, B., Ouzounis, C., Parra, G., and Guigo, R. (2005). Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett* 579, 1900-1903.
- Lou, H., Gagel, R.F., and Berget, S.M. (1996). An intron enhancer recognized by splicing factors activates polyadenylation. *Genes Dev* 10, 208-219.

- Luhrmann, R., and Stark, H. (2009). Structural mapping of spliceosomes by electron microscopy. *Curr Opin Struct Biol* 19, 96-102.
- Lukas, J., Gao, D.Q., Keshmeshian, M., Wen, W.H., Tsao-Wei, D., Rosenberg, S., and Press, M.F. (2001). Alternative and aberrant messenger RNA splicing of the mdm2 oncogene in invasive breast cancer. *Cancer Res* 61, 3212-3219.
- Lutz, C.S., and Alwine, J.C. (1994). Direct interaction of the U1 snRNP-A protein with the upstream efficiency element of the SV40 late polyadenylation signal. *Genes Dev* 8, 576-586.
- Lutz, C.S., Murthy, K.G., Schek, N., O'Connor, J.P., Manley, J.L., and Alwine, J.C. (1996). Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro. *Genes Dev* 10, 325-337.
- Ma, X.M., Yoon, S.O., Richardson, C.J., Julich, K., and Blenis, J. (2008). SKAR links pre-mRNA splicing to mTOR/S6K1-mediated enhanced translation efficiency of spliced mRNAs. *Cell* 133, 303-313.
- Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. *Nature* 416, 499-506.
- Massiello, A., and Chalfant, C.E. (2006). SRp30a (ASF/SF2) regulates the alternative splicing of caspase-9 pre-mRNA and is required for ceramide-responsiveness. *J Lipid Res* 47, 892-897.
- Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., and Reed, R. (2005). Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev* 19, 1512-1517.
- Matlin, A.J., Clark, F., and Smith, C.W. (2005). Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* 6, 386-398.
- Matlin, A.J., and Moore, M.J. (2007). Spliceosome assembly and composition. *Adv Exp Med Biol* 623, 14-35.
- Mazoyer, S., Puget, N., Perrin-Vidoz, L., Lynch, H.T., Serova-Sinilnikova, O.M., and Lenoir, G.M. (1998). A BRCA1 nonsense mutation causes exon skipping. *Am J Hum Genet* 62, 713-715.
- McCracken, S., Lambermon, M., and Blencowe, B.J. (2002). SRm160 splicing coactivator promotes transcript 3'-end cleavage. *Mol Cell Biol* 22, 148-160.

- Merz, C., Urlaub, H., Will, C.L., and Luhrmann, R. (2007). Protein composition of human mRNPs spliced in vitro and differential requirements for mRNP protein recruitment. *Rna* 13, 116-128.
- Michlewski, G., Sanford, J.R., and Cáceres, J.F. (2008). The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1. *Mol Cell* 30, 179-189.
- Millevoi, S., Decorsiere, A., Loulergue, C., Iacovoni, J., Bernat, S., Antoniou, M., and Vagner, S. (2009). A physical and functional link between splicing factors promotes pre-mRNA 3' end processing. *Nucleic Acids Res* 37, 4672-4683.
- Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* 136, 688-700.
- Nagradova, N. (2007). Enzymes catalyzing protein folding and their cellular functions. *Curr Protein Pept Sci* 8, 273-282.
- Nagy, E., and Maquat, L.E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci* 23, 198-199.
- Nembaware, V., Wolfe, K.H., Bettoni, F., Kelso, J., and Seoighe, C. (2004). Allele-specific transcript isoforms in human. *FEBS Lett* 577, 233-238.
- O'Mullane, L., and Eperon, I.C. (1998). The pre-mRNA 5' cap determines whether U6 small nuclear RNA succeeds U1 small nuclear ribonucleoprotein particle at 5' splice sites. *Mol Cell Biol* 18, 7510-7520.
- Ozaki, T., and Nakagawara, A. (2005). p73, a sophisticated p53 family member in the cancer world. *Cancer Science* 96, 729-737.
- Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40, 1413-1415.
- Patel, A.A., and Steitz, J.A. (2003). Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol* 4, 960-970.
- Peebles, K.A., Dwyer-Nield, L.D., and Malkinson, A.M. (2007). Altered expression of splicing factor, heterogeneous nuclear ribonucleoprotein A2/B1, in mouse lung neoplasia. *Mol Carcinog* 46, 887-900.

Pio, R., and Montuenga, L.M. (2009). Alternative splicing in lung cancer. *J Thorac Oncol* 4, 674-678.

Qiu, J., Nayak, R., and Pintel, D.J. (2004). Alternative polyadenylation of adeno-associated virus type 5 RNA within an internal intron is governed by both a downstream element within the intron 3' splice acceptor and an element upstream of the P41 initiation site. *J Virol* 78, 83-93.

Qiu, J., and Pintel, D.J. (2004). Alternative polyadenylation of adeno-associated virus type 5 RNA within an internal intron is governed by the distance between the promoter and the intron and is inhibited by U1 small nuclear RNP binding to the intervening donor. *J Biol Chem* 279, 14889-14898.

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

Roca, X., Olson, A.J., Rao, A.R., Enerly, E., Kristensen, V.N., Borresen-Dale, A.L., Andresen, B.S., Krainer, A.R., and Sachidanandam, R. (2008). Features of 5'-splice-site efficiency derived from disease-causing mutations and comparative genomics. *Genome Res* 18, 77-87.

Roca, X., Sachidanandam, R., and Krainer, A.R. (2003). Intrinsic differences between authentic and cryptic 5' splice sites. *Nucleic Acids Res* 31, 6321-6333.

Roca, X., Sachidanandam, R., and Krainer, A.R. (2005). Determinants of the inherent strength of human 5' splice sites. *Rna* 11, 683-698.

Rochaix, P., Krajewski, S., Reed, J.C., Bonnet, F., Voigt, J.J., and Brousset, P. (1999). In vivo patterns of Bcl-2 family protein expression in breast carcinomas in relation to apoptosis. *J Pathol* 187, 410-415.

Roy, M., Xu, Q., and Lee, C. (2005). Evidence that public database records for many cancer-associated genes reflect a splice form found in tumors and lack normal splice forms. *Nucleic Acids Res* 33, 5026-5033.

Sakamuro, D., Elliott, K.J., Wechsler-Reya, R., and Prendergast, G.C. (1996). BIN1 is a novel MYC-interacting protein with features of a tumour suppressor. *Nat Genet* 14, 69-77.

Sakharkar, M.K., Chow, V.T., and Kanguane, P. (2004). Distributions of exons and introns in the human genome. *In Silico Biol* 4, 387-393.



- Sanford, J.R., Gray, N.K., Beckmann, K., and Cáceres, J.F. (2004). A novel role for shuttling SR proteins in mRNA translation. *Genes Dev* 18, 755-768.
- Senapathy, P., Shapiro, M.B., and Harris, N.L. (1990). Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol* 183, 252-278.
- Sinclair, C.S., Rowley, M., Naderi, A., and Couch, F.J. (2003). The 17q23 amplicon and breast cancer. *Breast Cancer Res Treat* 78, 313-322.
- Skotheim, R.I., and Nees, M. (2007). Alternative splicing in cancer: noise, functional, or systematic? *Int J Biochem Cell Biol* 39, 1432-1449.
- Stalder, L., and Muhlemann, O. (2008). The meaning of nonsense. *Trends Cell Biol* 18, 315-321.
- Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92, 315-326.
- Stickeler, E., Kittrell, F., Medina, D., and 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.
- Sun, S., Zhang, Z., Sinha, R., Karni, R., and Krainer, A.R. (2009). SF2/ASF autoregulation involves multiple mechanisms. *Nat Struct Mol Biol* *Accepted*.
- Tajiri, T., Liu, X., Thompson, P.M., Tanaka, S., Suita, S., Zhao, H., Maris, J.M., Prendergast, G.C., and Hogarty, M.D. (2003). Expression of a MYCN-interacting isoform of the tumor suppressor BIN1 is reduced in neuroblastomas with unfavorable biological features. *Clin Cancer Res* 9, 3345-3355.
- Takao, N., Kato, H., Mori, R., Morrison, C., Sonada, E., Sun, X., Shimizu, H., Yoshioka, K., Takeda, S., and Yamamoto, K. (1999). Disruption of ATM in p53-null cells causes multiple functional abnormalities in cellular response to ionizing radiation. *Oncogene* 18, 7002-7009.
- Tange, T.O., Nott, A., and Moore, M.J. (2004). The ever-increasing complexities of the exon junction complex. *Curr Opin Cell Biol* 16, 279-284.

Tockman, M.S., Erozan, Y.S., Gupta, P., Piantadosi, S., Mulshine, J.L., and Ruckdeschel, J.C. (1994). The early detection of second primary lung cancers by sputum immunostaining. LCEWDG Investigators. Lung Cancer Early Detection Group. *Chest* 106, 385S-390S.

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., and 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.

Tvarusko, W., Bentele, M., Misteli, T., Rudolf, R., Kaether, C., Spector, D.L., Gerdes, H.H., and Eils, R. (1999). Time-resolved analysis and visualization of dynamic processes in living cells. *Proc Natl Acad Sci U S A* 96, 7950-7955.

Vagner, S., Ruegsegger, U., Gunderson, S.I., Keller, W., and Mattaj, I.W. (2000a). Position-dependent inhibition of the cleavage step of pre-mRNA 3'-end processing by U1 snRNP. *Rna* 6, 178-188.

Vagner, S., Vagner, C., and Mattaj, I.W. (2000b). The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. *Genes Dev* 14, 403-413.

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

Valencia, P., Dias, A.P., and Reed, R. (2008). Splicing promotes rapid and efficient mRNA export in mammalian cells. *Proc Natl Acad Sci U S A* 105, 3386-3391.

Venables, J.P., Klinck, R., Koh, C., Gervais-Bird, J., Bramard, A., Inkel, L., Durand, M., Couture, S., Froehlich, U., Lapointe, E., *et al.* (2009). Cancer-associated regulation of alternative splicing. *Nat Struct Mol Biol* 16, 670-676.

Wachtel, C., and Manley, J.L. (2009). Splicing of mRNA precursors: the role of RNAs and proteins in catalysis. *Mol Biosyst* 5, 311-316.

Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701-718.

- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470-476.
- Wang, G.S., and Cooper, T.A. (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 8, 749-761.
- Wang, J., Smith, P.J., Krainer, A.R., and Zhang, M.Q. (2005). Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes. *Nucleic Acids Res* 33, 5053-5062.
- Watermann, D.O., Tang, Y., Zur Hausen, A., Jager, M., Stamm, S., and Stickeler, E. (2006). Splicing factor Tra2-beta1 is specifically induced in breast cancer and regulates alternative splicing of the CD44 gene. *Cancer Res* 66, 4774-4780.
- Will, C.L., and Lührmann, R. (2006). 13 Spliceosome Structure and Function.
- Will, C.L., Schneider, C., MacMillan, A.M., Katopodis, N.F., Neubauer, G., Wilm, M., Luhrmann, R., and Query, C.C. (2001). A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *Embo J* 20, 4536-4546.
- Xiao, R., Sun, Y., Ding, J.H., Lin, S., Rose, D.W., Rosenfeld, M.G., Fu, X.D., and Li, X. (2007). Splicing regulator SC35 is essential for genomic stability and cell proliferation during mammalian organogenesis. *Mol Cell Biol* 27, 5393-5402.
- Xu, X., Yang, D., Ding, J.H., Wang, W., Chu, P.H., Dalton, N.D., Wang, H.Y., Bermingham, J.R., Jr., Ye, Z., Liu, F., *et al.* (2005). ASF/SF2-regulated CaMKII $\delta$  alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle. *Cell* 120, 59-72.
- Yang, A., Kaghad, M., Caput, D., and McKeon, F. (2002). On the shoulders of giants: p63, p73 and the rise of p53. *Trends Genet* 18, 90-95.
- Yang, A., Kaghad, M., Wang, Y., Gillett, E., Fleming, M.D., Dotsch, V., Andrews, N.C., Caput, D., and McKeon, F. (1998). p63, a p53 homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. *Mol Cell* 2, 305-316.
- Zerbe, L.K., Pino, I., Pio, R., Cospers, P.F., Dwyer-Nield, L.D., Meyer, A.M., Port, J.D., Montuenga, L.M., and Malkinson, A.M. (2004). Relative amounts of antagonistic splicing factors, hnRNP A1 and ASF/SF2, change during neoplastic lung growth: implications for pre-mRNA processing. *Mol Carcinog* 41, 187-196.

Zhang, Q.X., Hilsenbeck, S.G., Fuqua, S.A., and Borg, A. (1996). Multiple splicing variants of the estrogen receptor are present in individual human breast tumors. *J Steroid Biochem Mol Biol* 59, 251-260.

Zhang, Z., and Krainer, A.R. (2004). Involvement of SR proteins in mRNA surveillance. *Mol Cell* 16, 597-607.

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.

Zhong, X.Y., Wang, P., Han, J., Rosenfeld, M.G., and Fu, X.D. (2009). SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Mol Cell* 35, 1-10.

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

## **Chapter 2**

Alex, D., and Lee, K.A. (2005). RGG-boxes of the EWS oncoprotein repress a range of transcriptional activation domains. *Nucleic Acids Res* 33, 1323-1331.

Allemand, E., Hastings, M.L., Murray, M.V., Myers, M.P., and Krainer, A.R. (2007). Alternative splicing regulation by interaction of phosphatase PP2C $\gamma$  with nucleic acid-binding protein YB-1. *Nat Struct Mol Biol* 14, 630-638.

Araya, N., Hiraga, H., Kako, K., Arao, Y., Kato, S., and Fukamizu, A. (2005). Transcriptional down-regulation through nuclear exclusion of EWS methylated by PRMT1. *Biochem Biophys Res Commun* 329, 653-660.

Bedford, M.T., and Clarke, S.G. (2009). Protein arginine methylation in mammals: who, what, and why. *Mol Cell* 33, 1-13.

Bedford, M.T., Frankel, A., Yaffe, M.B., Clarke, S., Leder, P., and Richard, S. (2000). Arginine methylation inhibits the binding of proline-rich ligands to Src homology 3, but not WW, domains. *J Biol Chem* 275, 16030-16036.

Belyanskaya, L.L., Gehrig, P.M., and Gehring, H. (2001). Exposure on cell surface and extensive arginine methylation of ewing sarcoma (EWS) protein. *J Biol Chem* 276, 18681-18687.

Bergamini, G., Preiss, T., and Hentze, M.W. (2000). Picornavirus IRESes and the poly(A) tail jointly promote cap-independent translation in a mammalian cell-free system. *RNA* 6, 1781-1790.

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.

Bish, R.A., Fregoso, O.I., Piccini, A., and Myers, M.P. (2008). Conjugation of complex polyubiquitin chains to WRNIP1. *J Proteome Res* 7, 3481-3489.

Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 72, 291-336.

Blaustein, M., Pelisch, F., Tanos, T., Munoz, M.J., Wengier, D., Quadrana, L., Sanford, J.R., Muschietti, J.P., Kornblihtt, A.R., Cáceres, J.F., *et al.* (2005). Concerted regulation of nuclear and cytoplasmic activities of SR proteins by AKT. *Nat Struct Mol Biol* 12, 1037-1044.

Borders, C.L., Jr., Broadwater, J.A., Bekeny, P.A., Salmon, J.E., Lee, A.S., Eldridge, A.M., and Pett, V.B. (1994). A structural role for arginine in proteins: multiple hydrogen bonds to backbone carbonyl oxygens. *Protein Sci* 3, 541-548.

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

Cáceres, J.F., Misteli, T., Sreaton, G.R., Spector, D.L., and Krainer, A.R. (1997). Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J Cell Biol* 138, 225-238.

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

Cáceres, J.F., Stamm, S., Helfman, D.M., and Krainer, A.R. (1994). Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science* 265, 1706-1709.

- Calnan, B.J., Tidor, B., Biancalana, S., Hudson, D., and Frankel, A.D. (1991). Arginine-mediated RNA recognition: the arginine fork. *Science* 252, 1167-1171.
- Cartegni, L., Chew, S.L., and Krainer, A.R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 3, 285-298.
- Cazalla, D., Zhu, J., Manche, L., Huber, E., Krainer, A.R., and Cáceres, J.F. (2002). Nuclear export and retention signals in the RS domain of SR proteins. *Mol Cell Biol* 22, 6871-6882.
- Cho, K.S., Elizondo, L.I., and Boerkoel, C.F. (2004). Advances in chromatin remodeling and human disease. *Curr Opin Genet Dev* 14, 308-315.
- Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J.L., Bell, J.C., and Duncan, P.I. (1996). The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *Embo J* 15, 265-275.
- Côté, J., and Richard, S. (2005). Tudor domains bind symmetrical dimethylated arginines. *J Biol Chem* 280, 28476-28483.
- Craig, R., Cortens, J.P., and Beavis, R.C. (2004). Open source system for analyzing, validating, and storing protein identification data. *J Proteome Res* 3, 1234-1242.
- Dreyfuss, G., Matunis, M.J., Piñol-Roma, S., and Burd, C.G. (1993). hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* 62, 289-321.
- Durocher, Y., Perret, S., and 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.
- Esteller, M. (2006). Epigenetics provides a new generation of oncogenes and tumour-suppressor genes. *Br J Cancer* 94, 179-183.
- Goulet, I., Gauvin, G., Boisvenue, S., and Côté, J. (2007). Alternative splicing yields protein arginine methyltransferase 1 isoforms with distinct activity, substrate specificity, and subcellular localization. *J Biol Chem* 282, 33009-33021.
- Gui, J.F., Tronchere, H., Chandler, S.D., and Fu, X.D. (1994). Purification and characterization of a kinase specific for the serine- and arginine-rich pre-mRNA splicing factors. *Proc Natl Acad Sci U S A* 91, 10824-10828.

- He, F., Muto, Y., Inoue, M., Kigawa, T., Shirouzu, M., Terada, T., Yokoyama, S. (2005). Solution structure of RRM domain in splicing factor SF2. Unpublished structure, DOI:10.2210/pdb2211x2214a/pdb.
- Hemann, M.T., Zilfou, J.T., Zhao, Z., Burgess, D.J., Hannon, G.J., and Lowe, S.W. (2004). Suppression of tumorigenesis by the p53 target PUMA. *Proc Natl Acad Sci U S A* *101*, 9333-9338.
- Herrmann, F., Bossert, M., Schwander, A., Akgun, E., and Fackelmayer, F.O. (2004). Arginine methylation of scaffold attachment factor A by heterogeneous nuclear ribonucleoprotein particle-associated PRMT1. *J Biol Chem* *279*, 48774-48779.
- Huang, Y., and Carmichael, G.G. (1996). A suboptimal 5' splice site is a cis-acting determinant of nuclear export of polyomavirus late mRNAs. *Mol Cell Biol* *16*, 6046-6054.
- Huang, Y., and Steitz, J.A. (2005). SRprises along a messenger's journey. *Mol Cell* *17*, 613-615.
- Huang, Y., Yario, T.A., and Steitz, J.A. (2004). A molecular link between SR protein dephosphorylation and mRNA export. *Proc Natl Acad Sci U S A* *101*, 9666-9670.
- Hughes, R.M., and Waters, M.L. (2006). Arginine methylation in a beta-hairpin peptide: implications for Arg-pi interactions, DeltaCp(o), and the cold denatured state. *J Am Chem Soc* *128*, 12735-12742.
- Hyun, Y.L., Lew, D.B., Park, S.H., Kim, C.W., Paik, W.K., and Kim, S. (2000). Enzymic methylation of arginyl residues in -gly-arg-gly- peptides. *Biochem J* *348 Pt 3*, 573-578.
- Jobert, L., Argentini, M., and Tora, L. (2008). PRMT1 mediated methylation of TAF15 is required for its positive gene regulatory function. *Exp Cell Res*.
- Karni, R., de Stanchina, E., Lowe, S.W., Sinha, R., Mu, D., and Krainer, A.R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol* *14*, 185-193.
- Kim, S., Merrill, B.M., Rajpurohit, R., Kumar, A., Stone, K.L., Papov, V.V., Schneiders, J.M., Szer, W., Wilson, S.H., Paik, W.K., and Williams, K.R. (1997). Identification of N(G)-methylarginine residues in human heterogeneous RNP

protein A1: Phe/Gly-Gly-Gly-Arg-Gly-Gly-Gly/Phe is a preferred recognition motif. *Biochemistry* 36, 5185-5192.

Koizumi, J., Okamoto, Y., Onogi, H., Mayeda, A., Krainer, A.R., and Hagiwara, M. (1999). The subcellular localization of SF2/ASF is regulated by direct interaction with SR protein kinases (SRPKs). *J Biol Chem* 274, 11125-11131.

Krainer, A.R., Conway, G.C., and Kozak, D. (1990). Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes Dev* 4, 1158-1171.

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

Krainer, A.R., Mayeda, A., Kozak, D., and Binns, G. (1991). Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and *Drosophila* splicing regulators. *Cell* 66, 383-394.

Lai, M.C., Lin, R.I., and Tarn, W.Y. (2001). Transportin-SR2 mediates nuclear import of phosphorylated SR proteins. *Proc Natl Acad Sci U S A* 98, 10154-10159.

Lai, M.C., and Tarn, W.Y. (2004). Hypophosphorylated ASF/SF2 binds TAP and is present in messenger ribonucleoproteins. *J Biol Chem* 279, 31745-31749.

Lan, F., and Shi, Y. (2009). Epigenetic regulation: methylation of histone and non-histone proteins. *Sci China C Life Sci* 52, 311-322.

Lee, M.S., and Silver, P.A. (1997). RNA movement between the nucleus and the cytoplasm. *Curr Opin Genet Dev* 7, 212-219.

Li, X., and Manley, J.L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* 122, 365-378.

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

Liu, Q., and Dreyfuss, G. (1995). In vivo and in vitro arginine methylation of RNA-binding proteins. *Mol Cell Biol* 15, 2800-2808.



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

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

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

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

McBride, A.E., Cook, J.T., Stemmler, E.A., Rutledge, K.L., McGrath, K.A., and Rubens, J.A. (2005). Arginine methylation of yeast mRNA-binding protein Npl3 directly affects its function, nuclear export, and intranuclear protein interactions. *J Biol Chem* 280, 30888-30898.

Misteli, T. (1999). RNA splicing: What has phosphorylation got to do with it? *Curr Biol* 9, R198-200.

Misteli, T., Caceres, J.F., Clement, J.Q., Krainer, A.R., Wilkinson, M.F., and 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.

Misteli, T., Caceres, J.F., and Spector, D.L. (1997). The dynamics of a pre-mRNA splicing factor in living cells. *Nature* 387, 523-527.

Misteli, T., and Spector, D.L. (1996). Serine/threonine phosphatase 1 modulates the subnuclear distribution of pre-mRNA splicing factors. *Mol Biol Cell* 7, 1559-1572.

Ngo, J.C., Chakrabarti, S., Ding, J.H., Velazquez-Dones, A., Nolen, B., Aubol, B.E., Adams, J.A., Fu, X.D., and Ghosh, G. (2005). Interplay between SRPK and Clk/Sty kinases in phosphorylation of the splicing factor ASF/SF2 is regulated by a docking motif in ASF/SF2. *Mol Cell* 20, 77-89.

Ngo, J.C., Giang, K., Chakrabarti, S., Ma, C.T., Huynh, N., Hagopian, J.C., Dorrestein, P.C., Fu, X.D., Adams, J.A., and Ghosh, G. (2008). A sliding docking interaction is essential for sequential and processive phosphorylation of an SR protein by SRPK1. *Mol Cell* 29, 563-576.

- Nichols, R.C., Wang, X.W., Tang, J., Hamilton, B.J., High, F.A., Herschman, H.R., and Rigby, W.F. (2000). The RGG domain in hnRNP A2 affects subcellular localization. *Exp Cell Res* 256, 522-532.
- Piñol-Roma, S., and Dreyfuss, G. (1992). Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature* 355, 730-732.
- Sanford, J.R., Ellis, J.D., Cazalla, D., and Cáceres, J.F. (2005). Reversible phosphorylation differentially affects nuclear and cytoplasmic functions of splicing factor 2/alternative splicing factor. *Proc Natl Acad Sci U S A* 102, 15042-15047.
- Sanford, J.R., Gray, N.K., Beckmann, K., and Cáceres, J.F. (2004). A novel role for shuttling SR proteins in mRNA translation. *Genes Dev* 18, 755-768.
- Sato, H., Hosoda, N., and Maquat, L.E. (2008). Efficiency of the pioneer round of translation affects the cellular site of nonsense-mediated mRNA decay. *Mol Cell* 29, 255-262.
- Shao-En Ong, G.M., Matthias Mann (2004). Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nature Methods* 1, 119-126.
- Shaw, S.D., Chakrabarti, S., Ghosh, G., and Krainer, A.R. (2007). Deletion of the N-terminus of SF2/ASF permits RS-domain-independent pre-mRNA splicing. *PLoS One* 2, e854.
- Shen, E.C., Henry, M.F., Weiss, V.H., Valentini, S.R., Silver, P.A., and Lee, M.S. (1998). Arginine methylation facilitates the nuclear export of hnRNP proteins. *Genes Dev* 12, 679-691.
- Siebel, C.W., and Guthrie, C. (1996). The essential yeast RNA binding protein Np13p is methylated. *Proc Natl Acad Sci U S A* 93, 13641-13646.
- Singleton, D.R., Chen, S., Hitomi, M., Kumagai, C., and Tartakoff, A.M. (1995). A yeast protein that bidirectionally affects nucleocytoplasmic transport. *J Cell Sci* 108 ( Pt 1), 265-272.
- Smith, B.C., and Denu, J.M. (2009). Chemical mechanisms of histone lysine and arginine modifications. *Biochim Biophys Acta* 1789, 45-57.
- Sun, S., Zhang, Z., Sinha, R., Karni, R., and Krainer, A.R. (2009). SF2/ASF autoregulation involves multiple mechanisms. *Nat Struct Mol Biol* *Accepted*.

- Thermann, R., Neu-Yilik, G., Deters, A., Frede, U., Wehr, K., Hagemeyer, C., Hentze, M.W., and Kulozik, A.E. (1998). Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J* 17, 3484-3494.
- Tintaru, A.M., Hautbergue, G.M., Hounslow, A.M., Hung, M.L., Lian, L.Y., Craven, C.J., and Wilson, S.A. (2007a). Structural and functional analysis of RNA and TAP binding to SF2/ASF. *EMBO Rep* 8, 756-762.
- Tintaru, A.M., Hautbergue, G.M., Hounslow, A.M., Lian, L.Y., Wilson, S.A., and Craven, C.J. (2007b). Assignment of <sup>1</sup>H, <sup>13</sup>C, and <sup>15</sup>N resonances for SF2 RNA recognition motif 2. *J Biomol NMR* 38, 193.
- Tockman, M.S., Erozan, Y.S., Gupta, P., Piantadosi, S., Mulshine, J.L., and Ruckdeschel, J.C. (1994). The early detection of second primary lung cancers by sputum immunostaining. LCEWDG Investigators. Lung Cancer Early Detection Group. *Chest* 106, 385S-390S.
- Velazquez-Dones, A., Hagopian, J.C., Ma, C.T., Zhong, X.Y., Zhou, H., Ghosh, G., Fu, X.D., and Adams, J.A. (2005). Mass spectrometric and kinetic analysis of ASF/SF2 phosphorylation by SRPK1 and Clk/Sty. *J Biol Chem* 280, 41761-41768.
- Wang, H., Huang, Z.Q., Xia, L., Feng, Q., Erdjument-Bromage, H., Strahl, B.D., Briggs, S.D., Allis, C.D., Wong, J., Tempst, P., and Zhang, Y. (2001). Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. *Science* 293, 853-857.
- Weiss, V.H., McBride, A.E., Soriano, M.A., Filman, D.J., Silver, P.A., and Hogle, J.M. (2000). The structure and oligomerization of the yeast arginine methyltransferase, Hmt1. *Nat Struct Biol* 7, 1165-1171.
- Wooderchak, W.L., Zang, T., Zhou, Z.S., Acuna, M., Tahara, S.M., and Hevel, J.M. (2008). Substrate profiling of PRMT1 reveals amino acid sequences that extend beyond the "RGG" paradigm. *Biochemistry* 47, 9456-9466.
- Xiao, S.H., and Manley, J.L. (1997). Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev* 11, 334-344.
- Xu, C., and Henry, M.F. (2004). Nuclear export of hnRNP Hrp1p and nuclear export of hnRNP Npl3p are linked and influenced by the methylation state of Npl3p. *Mol Cell Biol* 24, 10742-10756.

Yang, L., Chansky, H.A., and Hickstein, D.D. (2000). EWS.Fli-1 fusion protein interacts with hyperphosphorylated RNA polymerase II and interferes with serine-arginine protein-mediated RNA splicing. *J Biol Chem* 275, 37612-37618.

Yun, C.Y., and Fu, X.D. (2000). Conserved SR protein kinase functions in nuclear import and its action is counteracted by arginine methylation in *Saccharomyces cerevisiae*. *J Cell Biol* 150, 707-718.

Zhang, J., Sun, X., Qian, Y., and Maquat, L.E. (1998). Intron function in the nonsense-mediated decay of beta-globin mRNA: indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm. *Rna* 4, 801-815.

Zhang, Z., and Krainer, A.R. (2004). Involvement of SR proteins in mRNA surveillance. *Mol Cell* 16, 597-607.

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

### **Chapter 3**

Abdul-Manan, N., and Williams, K.R. (1996). hnRNP A1 binds promiscuously to oligoribonucleotides: utilization of random and homo-oligonucleotides to discriminate sequence from base-specific binding. *Nucleic Acids Res* 24, 4063-4070.

Anderson, S.L., Coli, R., Daly, I.W., Kichula, E.A., Rork, M.J., Volpi, S.A., Ekstein, J., and Rubin, B.Y. (2001). Familial dysautonomia is caused by mutations of the IKAP gene. *Am J Hum Genet* 68, 753-758.

Anderson, S.L., Qiu, J., and Rubin, B.Y. (2003a). EGCG corrects aberrant splicing of IKAP mRNA in cells from patients with familial dysautonomia. *Biochem Biophys Res Commun* 310, 627-633.

Anderson, S.L., Qiu, J., and Rubin, B.Y. (2003b). Tocotrienols induce IKBKAP expression: a possible therapy for familial dysautonomia. *Biochem Biophys Res Commun* 306, 303-309.

- Axelrod, F.B., and Simson, G.G.V. (2007). Hereditary sensory and autonomic neuropathies: types II, III, and IV. *Orphanet J Rare Dis* 2, -.
- Blumenfeld, A., Slaugenhaupt, S.A., Axelrod, F.B., Lucente, D.E., Maayan, C., Liebert, C.B., Ozelius, L.J., Trofatter, J.A., Haines, J.L., Breakefield, X.O., and et al. (1993). Localization of the gene for familial dysautonomia on chromosome 9 and definition of DNA markers for genetic diagnosis. *Nat Genet* 4, 160-164.
- Bonetta, L. (2009). RNA-based therapeutics: ready for delivery? *Cell* 136, 581-584.
- Burd, C.G., and Dreyfuss, G. (1994). RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *Embo J* 13, 1197-1204.
- Burge, C.B., Padgett, R.A., and Sharp, P.A. (1998). Evolutionary fates and origins of U12-type introns. *Mol Cell* 2, 773-785.
- Cartegni, L., Chew, S.L., and Krainer, A.R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 3, 285-298.
- Cartegni, L., Hastings, M.L., Calarco, J.A., de Stanchina, E., and Krainer, A.R. (2006). Determinants of exon 7 splicing in the spinal muscular atrophy genes, SMN1 and SMN2. *Am J Hum Genet* 78, 63-77.
- Chen, Y.T., Hims, M.M., Shetty, R.S., Mull, J., Liu, L., Leyne, M., and Slaugenhaupt, S.A. (2009). Loss of mouse Ikbkap, a subunit of elongator, leads to transcriptional deficits and embryonic lethality that can be rescued by human IKBKAP. *Mol Cell Biol* 29, 736-744.
- Close, P., Hawkes, N., Cornez, I., Creppe, C., Lambert, C.A., Rogister, B., Siebenlist, U., Merville, M.P., Slaugenhaupt, S.A., Bours, V., et al. (2006). Transcription impairment and cell migration defects in elongator-depleted cells: implication for familial dysautonomia. *Mol Cell* 22, 521-531.
- Cohen, L., Henzel, W.J., and Baeuerle, P.A. (1998). IKAP is a scaffold protein of the IkkappaB kinase complex. *Nature* 395, 292-296.
- Cooper, D.N., Stenson, P.D., and Chuzhanova, N.A. (2006). The Human Gene Mutation Database (HGMD) and its exploitation in the study of mutational mechanisms. *Curr Protoc Bioinformatics Chapter 1, Unit 1* 13.

Cuajungco, M.P., Leyne, M., Mull, J., Gill, S.P., Lu, W., Zagzag, D., Axelrod, F.B., Maayan, C., Gusella, J.F., and Slaugenhaupt, S.A. (2003). Tissue-specific reduction in splicing efficiency of IKBKAP due to the major mutation associated with familial dysautonomia. *Am J Hum Genet* 72, 749-758.

Fellows, J., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (2000). The Elp2 subunit of elongator and elongating RNA polymerase II holoenzyme is a WD40 repeat protein. *J Biol Chem* 275, 12896-12899.

Gold-von Simson, G., and Axelrod, F.B. (2006). Familial dysautonomia: update and recent advances. *Curr Probl Pediatr Adolesc Health Care* 36, 218-237.

Gold-von Simson, G., Goldberg, J.D., Rolnitzky, L.M., Mull, J., Leyne, M., Voustianiouk, A., Slaugenhaupt, S.A., and Axelrod, F.B. (2009). Kinetin in familial dysautonomia carriers: implications for a new therapeutic strategy targeting mRNA splicing. *Pediatr Res* 65, 341-346.

Hawkes, N.A., Otero, G., Winkler, G.S., Marshall, N., Dahmus, M.E., Krappmann, D., Scheidereit, C., Thomas, C.L., Schiavo, G., Erdjument-Bromage, H., *et al.* (2002). Purification and characterization of the human elongator complex. *J Biol Chem* 277, 3047-3052.

Hims, M.M., Shetty, R.S., Pickel, J., Mull, J., Leyne, M., Liu, L., Gusella, J.F., and Slaugenhaupt, S.A. (2007). A humanized IKBKAP transgenic mouse models a tissue-specific human splicing defect. *Genomics* 90, 389-396.

Holmlund, J.T. (2003). Applying antisense technology: Affinitak and other antisense oligonucleotides in clinical development. *Ann N Y Acad Sci* 1002, 244-251.

Hua, Y., Sahashi, K., Hung, G., Rigo, F., Passini, M.A., Bennett, C.F., and Krainer, A.R. (2009). Antisense correction of SMN2 splicing rescues type III SMA mice. *Nat. Genet.*, submitted.

Hua, Y., Vickers, T.A., Baker, B.F., Bennett, C.F., and Krainer, A.R. (2007). Enhancement of SMN2 exon 7 inclusion by antisense oligonucleotides targeting the exon. *PLoS Biol* 5, e73.

Hua, Y., Vickers, T.A., Okunola, H.L., Bennett, C.F., and Krainer, A.R. (2008). Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. *Am J Hum Genet* 82, 834-848.

- Karkare, S., and Bhatnagar, D. (2006). Promising nucleic acid analogs and mimics: characteristic features and applications of PNA, LNA, and morpholino. *Appl Microbiol Biotechnol* 71, 575-586.
- Kim, J.H., Lane, W.S., and Reinberg, D. (2002). Human Elongator facilitates RNA polymerase II transcription through chromatin. *Proc Natl Acad Sci U S A* 99, 1241-1246.
- Kouskouti, A., and Talianidis, I. (2005). Histone modifications defining active genes persist after transcriptional and mitotic inactivation. *Embo J* 24, 347-357.
- Krappmann, D., Hatada, E.N., Tegethoff, S., Li, J., Klippel, A., Giese, K., Baeuerle, P.A., and Scheidereit, C. (2000). The I kappa B kinase (IKK) complex is tripartite and contains IKK gamma but not IKAP as a regular component. *J Biol Chem* 275, 29779-29787.
- Krogan, N.J., and Greenblatt, J.F. (2001). Characterization of a six-subunit holo-elongator complex required for the regulated expression of a group of genes in *Saccharomyces cerevisiae*. *Mol Cell Biol* 21, 8203-8212.
- Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., and et al. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* 80, 155-165.
- Leyne, M., Mull, J., Gill, S.P., Cuajungco, M.P., Oddoux, C., Blumenfeld, A., Maayan, C., Gusella, J.F., Axelrod, F.B., and Slaugenhaupt, S.A. (2003). Identification of the first non-Jewish mutation in familial Dysautonomia. *Am J Med Genet A* 118A, 305-308.
- Li, Y., Takagi, Y., Jiang, Y., Tokunaga, M., Erdjument-Bromage, H., Tempst, P., and Kornberg, R.D. (2001). A multiprotein complex that interacts with RNA polymerase II elongator. *J Biol Chem* 276, 29628-29631.
- Lopez-Bigas, N., Audit, B., Ouzounis, C., Parra, G., and Guigo, R. (2005). Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett* 579, 1900-1903.
- Maayan, C., Kaplan, E., Shachar, S., Peleg, O., and Godfrey, S. (1987). Incidence of familial dysautonomia in Israel 1977-1981. *Clin Genet* 32, 106-108.

- Metivier, R., Penot, G., Hubner, M.R., Reid, G., Brand, H., Kos, M., and Gannon, F. (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* 115, 751-763.
- Nagy, E., and Maquat, L.E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci* 23, 198-199.
- Neuenkirchen, N., Chari, A., and Fischer, U. (2008). Deciphering the assembly pathway of Sm-class U snRNPs. *FEBS Lett* 582, 1997-2003.
- Otero, G., Fellows, J., Li, Y., de Bizemont, T., Dirac, A.M., Gustafsson, C.M., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (1999). Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. *Mol Cell* 3, 109-118.
- Riley, C.M. (1952). Familial autonomic dysfunction with defective lacrimation. *AMA Am J Dis Child* 84, 503-504.
- Riley, C.M., Day, R.L., and et al. (1949). Central autonomic dysfunction with defective lacrimation; report of five cases. *Pediatrics* 3, 468-478.
- Si, Z.H., Rauch, D., and Stoltzfus, C.M. (1998). The exon splicing silencer in human immunodeficiency virus type 1 Tat exon 3 is bipartite and acts early in spliceosome assembly. *Mol Cell Biol* 18, 5404-5413.
- Slaugenhaupt, S.A., Blumenfeld, A., Gill, S.P., Leyne, M., Mull, J., Cuajungco, M.P., Liebert, C.B., Chadwick, B., Idelson, M., Reznik, L., et al. (2001). Tissue-specific expression of a splicing mutation in the IKBKAP gene causes familial dysautonomia. *Am J Hum Genet* 68, 598-605.
- Slaugenhaupt, S.A., and Gusella, J.F. (2002). Familial dysautonomia. *Curr Opin Genet Dev* 12, 307-311.
- Slaugenhaupt, S.A., Mull, J., Leyne, M., Cuajungco, M.P., Gill, S.P., Hims, M.M., Quintero, F., Axelrod, F.B., and Gusella, J.F. (2004). Rescue of a human mRNA splicing defect by the plant cytokinin kinetin. *Hum Mol Genet* 13, 429-436.
- Svejstrup, J.Q. (2007). Elongator complex: how many roles does it play? *Curr Opin Cell Biol* 19, 331-336.
- Talbot, K., and Davies, K.E. (2001). Spinal muscular atrophy. *Semin Neurol* 21, 189-197.



Tazi, J., Bakkour, N., and Stamm, S. (2009). Alternative splicing and disease. *Biochim Biophys Acta* 1792, 14-26.

Wang, G.S., and Cooper, T.A. (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 8, 749-761.

Winkler, G.S., Kristjuhan, A., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (2002). Elongator is a histone H3 and H4 acetyltransferase important for normal histone acetylation levels in vivo. *Proc Natl Acad Sci U S A* 99, 3517-3522.

Winkler, G.S., Petrakis, T.G., Ethelberg, S., Tokunaga, M., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (2001). RNA polymerase II elongator holoenzyme is composed of two discrete subcomplexes. *J Biol Chem* 276, 32743-32749.

Wirth, B., Brichta, L., and Hahnen, E. (2006). Spinal muscular atrophy: from gene to therapy. *Semin Pediatr Neurol* 13, 121-131.

Wittschieben, B.O., Fellows, J., Du, W., Stillman, D.J., and Svejstrup, J.Q. (2000). Overlapping roles for the histone acetyltransferase activities of SAGA and elongator in vivo. *Embo J* 19, 3060-3068.

Wittschieben, B.O., Otero, G., de Bizemont, T., Fellows, J., Erdjument-Bromage, H., Ohba, R., Li, Y., Allis, C.D., Tempst, P., and Svejstrup, J.Q. (1999). A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme. *Mol Cell* 4, 123-128.