

Molecular mechanisms of alternative splicing regulation by antitumor drugs targeting U2 snRNP

Luisa Vigevani

Doctoral Thesis UPF / 2016

Thesis Director: Dr. Juan Valcárcel Juárez

Gene Regulation, Stem Cells and Cancer Program
Center for Genomic Regulation (CRG), Barcelona



Abstract

RNA splicing is the process of removal of introns from pre-mRNAs. It is carried out by the spliceosome, composed of five small nuclear ribonucleoproteins (snRNPs). Several small molecules with antitumor properties target SF3B1, a U2 snRNP component frequently mutated in cancer that helps U2 snRNP recruitment to Branch Point (BP) sequences required for splicing catalysis.

Together with collaborators from the University of Barcelona, we describe a new drug variant with improved splicing inhibitory and antiproliferative activities, as well as the novel concept of drug-antisense oligonucleotide conjugates.

We report that 3' splice site sensitivity to drugs is finely tuned by other BP and BP-like sequences 5' of the BP and by the strength of the polypyrimidine tract 3' of the BP.

We also show that different drug variants display both similarities and differences in alternative splicing modulation.

Collectively, our results reveal substantial plasticity in the effects of U2 snRNP-targeting drugs in alternative splicing.

Resumen

ARN splicing consiste en la eliminación de intrones de moléculas de pre-mARN. Es llevado a cabo por el spliceosoma, compuesto por cinco ribonucleoproteínas nucleares pequeñas (snRNPs). Varios compuestos con propiedades antitumorales tienen como diana SF3B1, un componente de U2 snRNP frecuentemente mutado en cáncer que ayuda al reclutamiento de U2 snRNP a secuencias branch point (BP) necesarias para la catálisis.

Junto a colaboradores de la Universidad de Barcelona, describimos una nueva variante de estos compuestos con mayor actividad de inhibición de splicing y anti-proliferativa, así como el nuevo concepto de conjugados fármaco-oligonucleótido.

Observamos que la sensibilidad de los sitios 3' de splicing a estos compuestos está moduladas por la presencia de otras secuencias similares a BP situadas 5' del BP, así como por la fortaleza del tramo de polipirimidina situado 3' del BP.

También mostramos que diferentes variantes de compuestos tienen actividades similares y diferentes en la modulación de distintos eventos de splicing alternativo.

En conjunto, nuestros resultados revelan un elevado grado de plasticidad en los efectos sobre splicing alternativo de compuestos que tienen como diana U2 snRNP.

Preface

The activity of the splicing machinery can confer therapeutic vulnerability to some types of cancer (Hsu et al., 2015) and splicing inhibitors displaying antitumor properties are emerging as a potential new class of therapeutics for cancer (Bonnal et al., 2012). In addition, new cancer-related mutations within splicing factors are increasingly found. Thus, the investigation of the spliceosome's role in cancer development and associated therapeutic opportunities is experiencing a boost in recent years. Not only clinical research in the field is greatly expanding, but also basic studies are needed to understand the mechanistic consequences of splicing modulation by splicing-targeting drugs and to identify the cause of their selective cytotoxicity for cancer cells and even for sub-classes of cancer cells.

The work of this thesis aimed at the generation of new drug variants to test their structural requirements and improve their activities and at the possibility of gaining improved specificity by conjugating them with antisense oligonucleotides. Our studies also investigated the reasons why different transcripts display higher or lower sensitivity to the splicing inhibitory effects of the drugs, as well as the differences in alternative splicing regulation induced by two related compounds.

Our results can contribute to the rational design of splicing regulatory molecules with improved activity and specificity.

Key Words

Alternative splicing
Antitumor drugs
SF3B1
Spliceosome
U2 snRNP

Abbreviations

5' – 5 prime
3' – 3 prime
5'ss – 5' splice site
3'ss – 3' splice site
A – adenine or adenosine
AdML – Adenovirus Major Late promoter
AG – AG dinucleotide forming the 3' splice site
AML – acute myeloid leukemia
AS – alternative splicing
ASO – antisense oligonucleotide
ATP – adenosine triphosphate
bp – base pair
BP – branch point
BPRS – branch point recognition sequence
BrU – 5-Bromouridine
C – cytidine or cytosine
cDNA – complementary DNA
CE – cassette exon
ChIP – chromatin immunoprecipitation
CLL – chronic lymphocytic leukemia
CMV – cytomegalovirus
cpm – counts per minute
cryo-EM – cryo-electron microscopy
CTD – C-terminal repeat domain
CTR – control
CTRL – control
DMEM – Dulbecco's modified Eagle's medium
DNA – deoxyribonucleic acid
dNTPs – deoxynucleotides
DTT – dithiothreitol
 Δ – delta or deletion
EDTA – ethylenediamine tetraacetic acid
ESE – exonic splicing enhancer
ESS – exonic splicing silencer
FPKM – fragments per kilobase per million mapped reads
G – guanine or guanosine
GE – gene expression
GO – gene ontology
h – hour
HC – high confidence
iCLIP – individual-nucleotide resolution cross-linking and immunoprecipitation
IP – immunoprecipitation
ISE – intronic splicing enhancer
ISS – intronic splicing silencer
kDa – kilodalton
k.d. – knockdown
ln – natural logarithm
MC – medium confidence
mRNA – messenger RNA
mut – mutant
n – number in a trial or sample
N – any nucleotide
NMD – nonsense mediated decay
nt – nucleotides
NTP – nucleoside triphosphate
PAGE – polyacrylamide gel electrophoresis
pol – polymerase
pre-mRNA – premature mRNA
PCR – polymerase chain reaction

PSI – Percent Splicing Index or
Percent Spliced In
Py-tract – polypyrimidine tract
PWM – position weight matrix
R – purine
R – side group in chemical structures
RI – retained intron
RBP – RNA binding protein
RNA – ribonucleic acid
RNA-Seq – RNA sequencing
RRM – RNA recognition motif
RS – ring sideroblasts
RT-PCR – reverse-transcription
polymerase chain reaction
SDS – Sodium dodecyl sulfate
siRNA – small interfering RNA

SMA – spinal muscular atrophy
snRNA – small nuclear RNA
snRNP – small nuclear
ribonucleoprotein
SR protein – serine/arginine-rich
protein
SSA – Spliceostatin A
Sud – Sudemycin
SVM – support vector machine
T – thymine or thymidine
TBE – Tris/Borate/EDTA
U – uracil or uridine
wt – wild type
Y – pyrimidine

Table of Contents

Abstract	III
Resumen	V
Preface	VII
Key Words	IX
Abbreviations	IX
Table of Contents	XI
Introduction	1
1. <i>RNA Splicing</i>	1
1.1. Central dogma of biology and RNA splicing	1
1.2. Alternative splicing (AS)	3
1.3. Evolution of splicing	6
1.4. The spliceosome	8
1.5. Regulation of alternative splicing	14
2. <i>The spliceosome in disease and cancer</i>	17
2.1. Splicing and genetic disease	17
2.2. Alternative splicing and cancer	19
2.3. Splicing factors and cancer	22
2.4. Splicing inhibitors: a novel class of drugs	24
3. <i>SF3B1: function and dysfunction</i>	28
3.1. U2 snRNP function and components	28
3.2. SF3B1 protein	30
3.3. Consequences of SF3B1 mutations	32
3.4. Mechanistic consequences of SF3B inhibition	36
3.5. Functional consequences of SF3B inhibition	38
Objectives	43
Results	45
<i>Part I.A. Analysis of structure-activity relationship for SF3B-targeting drug variants</i>	45
<i>Part I.B. Analysis of the activity of drug-antisense oligonucleotide conjugates</i>	73
<i>Part II. Identification of sequence elements determining alternative splicing sensitivity to pharmacological inhibition of SF3B1</i>	83
<i>Part III. Comparison of the effects of different drug variants on alternative splicing modulation</i>	127
Materials and Methods	143
Discussion	153

Conclusions	169
References	171
Annexes	189
Acknowledgements	231

Introduction

1. RNA Splicing

1.1. Central dogma of biology and RNA splicing

Gene regulation at a first glance

The information to build an organism is stored in the inheritable genetic material, typically DNA: linear sequences of four different nucleotides drive the synthesis of proteins, the cell's functional elements. Nevertheless, an intermediate molecule is necessary in the flow of information from DNA to proteins: the RNA. This principle is known as the central dogma of biology: DNA sequences are transcribed into RNA sequences that are finally translated into proteins, which exert key structural and enzymatic functions within cells and organisms (Crick, 1970). Although research of the last 60 years has added many more details that add complexity to it, the flow “DNA → RNA → protein” is still the fundament of molecular biology.

The process by which RNA is copied from one of the two DNA strands is called transcription (and for this reason transcribed RNA molecules are also called transcripts). RNA polymerases are the enzymes dedicated to transcribe the DNA. On the other hand, translation is the process of converting the information from “messenger” RNA molecules into proteins. Ribosomes are the

ribonucleoprotein complexes (i.e. complexes of nuclear proteins and RNA molecules) running this cellular function.

The genome is the complete set of genetic material of an organism. It consists of DNA comprising genes that carry the information to make different proteins and intergenic regions, originally thought to be devoid of functional elements but today recognized as essential for regulating gene expression as well as for generating transcripts that are not being translated into proteins.

The information is subject to several modifications at each step of the flow. Indeed, eukaryotic genes encompass coding and non-coding sequences (i.e. sequences that will or will not be translated into protein). Coding and non-coding sequences alternate in most of the genes and are known as exons and introns: while the first ones are present in mature messenger RNAs (mRNAs) that are translated into proteins, the second ones are only present in premature messenger RNAs (pre-mRNAs) and need to be removed to allow mRNA translation into proteins (Crick, 1979; Gilbert, 1978).

The process by which introns are excised and exons joined together is known as pre-mRNA splicing. Splicing is one of the major steps of pre-mRNA processing. It takes place within the nucleus of every eukaryotic cell and it is carried out by the spliceosome, a very dynamic and complex ribonucleoparticle (Wahl et al., 2009).

Decades of research have shown and keep expanding the fascinating functions that lie behind the need to remove parts of the primary transcripts. Although the origin of introns is still a matter of controversy (Irimia and Roy, 2014), several reasons can be put

forward for splicing to exist. Undoubtedly, the possibility of sophisticated regulation is one of the key advantages of this mechanism: for example, many noncoding RNAs with regulatory function, like snoRNAs or miRNAs (i.e. small nucleolar and micro RNAs) derive from removed introns (Hesselberth, 2013).

A new intriguing class of RNAs, circular RNAs, can also be processed from intronic sequences and be associated to biologically relevant functions (Hesselberth, 2013; Lasda and Parker, 2014). Moreover, it has also been proposed that one of the advantages of RNA splicing is the defense from parasitic nucleic acids without optimal splicing signals that would lead to spliceosomes stalling and degradation (Madhani, 2013).

1.2. Alternative splicing (AS)

A further step within gene regulation

In higher eukaryotes, entire exons or part of them can be included or not within the mature mRNAs, meaning that RNA splicing can occur in alternative ways: this mechanism is known as alternative splicing (Early et al., 1980) (Figure 1). As a consequence, different mature transcripts can be generated from the same gene. Eukaryotic genomes get a key advantage from this additional layer of gene regulation in the variety of proteins that can be produced from the same genome, greatly expanding its information content (Nilsen and Graveley, 2010). Functional differences between proteins encoded by alternatively spliced transcripts (isoforms) include the presence or

absence of protein domains as well as unstructured polypeptide regions that are important for protein-protein interactions (Buljan et al., 2013; Ellis et al., 2012a; Yang et al., 2016).

Nevertheless, in some cases alternative splicing does not cause a difference at the protein level, but rather different stability of the alternative transcripts. For instance, alternative splicing can be coupled to NMD (nonsense-mediated decay) such that transcripts with premature stop codons included in alternatively spliced RNAs are degraded after the first round of translation. In other cases, the functional difference between isoforms is dictated by differential binding of RNA-binding proteins (RBPs) that control other aspects of RNA metabolism, including mRNA stability or translation efficiency. Collectively, these examples illustrate that alternative splicing constitutes a complex and sophisticated step of gene regulation with a variety of implications for the cell's function (Braunschweig et al., 2013).

These include the different needs of cells belonging to distinct tissues or distinct stages of development: these needs not only require the activation or repression of different genes through transcription regulation, but also a different balance of RNA and protein isoforms, i.e. different transcripts or proteins generated from the same gene (Kalsotra and Cooper, 2011; Ye and Blelloch, 2014). Indeed, even if all the cells of a complex organism share the same genome, their transcriptome and proteome can be substantially different, and this variation is contributed by both transcriptional and post-transcriptional mechanisms.

Alternative splicing seems to have played a key role during evolution. As an example, even if human and *C. elegans* worms genomes have quite similar numbers of genes (around 20'000), the higher human complexity can be partially attributed to the increased incidence of alternative splicing (Nilsen and Graveley, 2010). Indeed, more than 95% of human genes can undergo alternative splicing (Pan et al., 2008; Wang et al., 2008) (Figure 1).

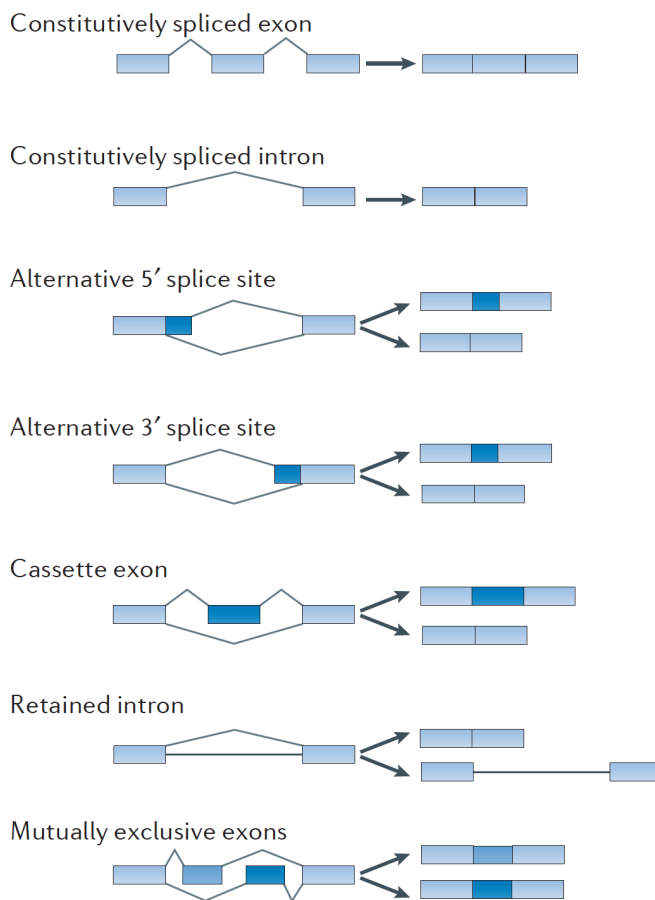


Figure 1. Types of alternative splicing. Splicing of a region is considered constitutive when it follows the same pattern in every cell: introns that are always spliced and exons that are always included in the mature transcripts are therefore known as constitutive. In contrast, alternative splicing

originates from the possibility of generating different splicing variants: alternative splice sites can be used (both 5' and 3'), exons can be skipped (and the events are named either cassette exon or exon skipping events), introns can be retained in the mature transcripts or exons can be mutually exclusive. Combinations of the different types of events are also possible, giving rise to complex events. Alternative splice sites that are only used in specific non-physiological conditions are often referred to as cryptic splice sites (adapted from Dvinge et al., 2016).

1.3. Evolution of splicing

From prokaryotic self-catalytic introns to a highly regulated mechanism conferring diversity among eukaryotes

Although splicing was first described for viral sequences (Berget et al., 1977; Chow et al., 1977), its complexity mainly developed during eukaryotic evolution. But where do introns come from? Spliceosomal introns (i.e. introns that are removed by the spliceosome machinery within the nucleus of eukaryotic cells) share high similarity with group II self-splicing introns (i.e. introns whose removal is catalyzed by the RNA molecule itself). While the first are exclusive for eukaryotic genomes, the second are found in bacteria, mitochondria and chloroplasts. Due to their extremely similar chemical steps, it is well accepted that spliceosomal introns evolved from self-splicing introns (Irimia and Roy, 2014). Indeed, comparisons of the detailed requirements for splicing catalysis revealed the highly similar arrangement of RNAs and magnesium ions at their catalytic cores (Papasaikas and Valcarcel, 2015; Robart et al., 2014; Yan et al., 2015).

It has been proposed that eukaryotic introns actually originated concomitantly with the integration of mitochondria in eukaryotic cells

and their presence imposed the need for cellular compartmentalization to avoid translation of pre-mRNAs before intron removal. Thus, the nucleus may have originated to separate the slow process of pre-mRNA splicing from the very efficient translation process, avoiding cotranscriptional translation of precursor mRNAs (Martin and Koonin, 2006).

Once "imported" into the eukaryotic system, splicing might have played a key role during evolution. Indeed, tissues diversity within an organism is largely due to different patterns of gene expression (in terms of the levels of transcripts coming from the same gene in each cell of the tissue, without considering different isoforms) and alternative splicing (i.e the level of expression of different splicing isoforms coming from the same gene). Interestingly, gene expression is much more conserved than alternative splicing when comparing transcriptomic profiles among the same organs from different organisms (Barbosa-Morais et al., 2012; Merkin et al., 2012). This observation highlights the fundamental role of alternative splicing for evolution and complexity.

The most widespread usage of alternative splicing occurs in the brain tissue, with alternative isoforms more conserved across organisms in this organ (Raj and Blencowe, 2015). Microexons form a recently discovered class of brain-specific exons: they are shorter than 21 nt (while the medium size of conventional exons is around 150 nt), highly conserved and regulated in a switch-like manner (Irimia et al., 2014). In addition, alternative isoforms of specific transcripts are finely regulated at the level of synapses, conferring plasticity to shape cell to cell connections within the brain (Raj and Blencowe, 2015).

Thus, the possibilities for regulation can be viewed as one of the keys to understand the reasons why evolution took advantage of the "intron invasion" to generate high complexity in gene expression and a fascinating variety of features among organisms.

1.4. The spliceosome

Components and structure of a dynamic regulatory tool

Splicing catalysis occurs in two steps (Figure 2). First, the nucleotide at the 5' end of the intron is covalently bound through a 2'-5' phosphodiester bond to the branch adenosine located within the branch point (BP) sequence, generating a free 5' exon and a lariat intron (Padgett et al., 1984; Ruskin et al., 1984). In the second step, the free 5' exon is ligated to the 3' exon and the lariat is excised. Each step consists in a trans-esterification reaction involving the breaking and forming of phosphodiester bonds: one electron donor oxygen carries out a nucleophilic attack on a phosphate within the RNA backbone (Fica et al., 2013; Papasaikas and Valcarcel, 2016; Scotti and Swanson, 2016).

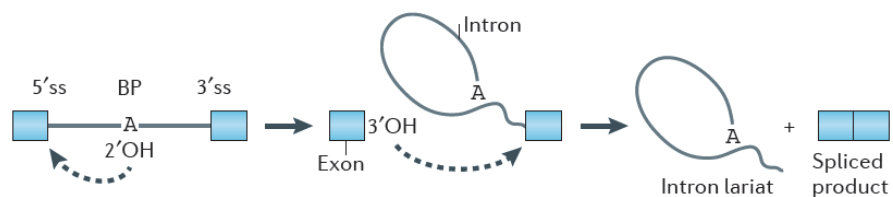


Figure 2. The two steps of splicing catalysis. During the first step, the 5'ss and branch adenosine are joined by a 2'-5' nucleophilic attack, leading to the formation of a lariat structure and a free upstream exon. The second

step consists in the 3'-5' nucleophilic bond between the free exon and the 3'ss (adapted from Scotti et al., 2016).

The molecular machinery responsible for RNA splicing is known as the spliceosome (Brody and Abelson, 1985; Grabowski et al., 1985). This is a very dynamic and complex group of non-coding RNAs and proteins with the unique feature of assembling on a step-wise manner on each intron prior to splicing. The splicing machinery is composed of five snRNPs (small nuclear ribonucleoproteins: U1, U2, U4, U5 and U6 snRNPs), each one containing one specific RNA molecule (U1, U2, U4, U5 and U6 snRNAs because of their U-rich sequence) and several polypeptides, some of which (Sm core) are common to all spliceosomal snRNPs except for U6, and other proteins are specific of each complex. The machinery can assemble on each intron through recognition of specific sequence signals: a GURAGU consensus at the intronic 5' boundary (the 5' splice site, or 5'ss) and three elements at the 3' boundary (the 3' splice site, or 3'ss): a branch point sequence (or BP, with a degenerated YUNAY consensus in higher eukaryotes, where Y= pyrimidine, N= A,G,C or U), a stretch of pyrimidines (polypyrimidine tract, or Py-tract) and an AG dinucleotide at the 3' end of each intron (Wahl et al., 2009) (Figure 3).

In the first steps of spliceosome assembly, 5' splice site sequences are recognized by U1 snRNP through base-pairing interactions involving the 5' end of U1 snRNA and the 5' ss sequence, while the proteins SF1, U2AF65 (or U2AF2) and U2AF35 (or U2AF1) recognize the BP sequence, the polypyrimidine tract and the AG dinucleotide, respectively. The cooperative binding of these proteins leads to the formation of the early spliceosomal complex E (Figure 3) and helps to recruit U2 snRNP at the level of the BP sequence (Papasaikas and

Valcarcel, 2016; Wahl et al., 2009). The regulation of the recruitment of U1 and U2 snRNPs is important for alternative splicing (Braunschweig et al., 2013; Scotti and Swanson, 2016).

U2 snRNP assembly involves base-pairing interactions between the U2 snRNA component of the snRNP and nucleotides of the BP sequence (Parker et al., 1987). Once the U1 and U2 snRNPs have been recruited and SF1 is displaced, the complex A is formed (Figure 3). Subsequently, the tri-snRNP (composed by the U4, U5 and U6 snRNPs preassembled together) joins the complex, leading to B complex formation. When U1 and U4 snRNPs are destabilized and leave the complex, the activated B complex (B^{act}) is formed and gets catalytically activated (B^*). The spliceosome reaches the C complex conformation after the first step (Figure 3) and the post-splicing P complex at the end of the whole reaction, before being finally disassembled (Papasaikas and Valcarcel, 2016; Wahl et al., 2009).

With some possibilities of variation, these steps seem to operate for the removal of most eukaryotic introns. The complexity and dynamics of the spliceosome (among the most elaborate molecular machineries of the cell) open numerous possibilities for regulation at several levels during its assembly.

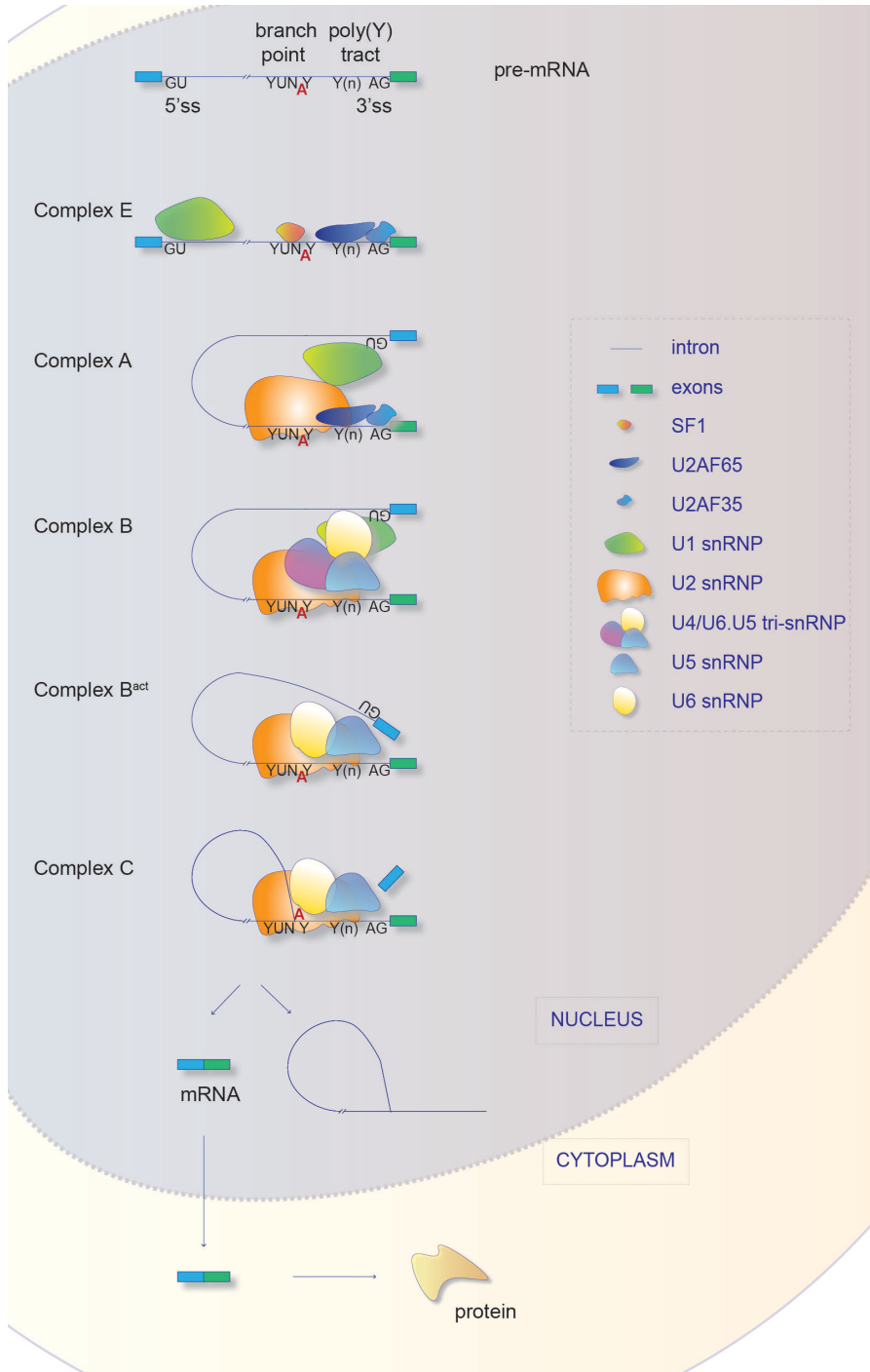


Figure 3 (previous page). Spliceosome assembly and catalysis of intron removal. Represented at the top is a typical precursor mRNA (pre-mRNA) with intronic sequences (thin line) and exonic sequences (coloured boxes), with the consensus sequences at the intron 5' and 3' ends (Y represents pyrimidines, N represents any nucleotide and (n) represents an undefined number of pyrimidines). The BP adenosine is represented in bold. The splicing factors represented in the box assemble on the pre-mRNA in a sequential manner, forming the indicated complexes (E, A and B). Conformational rearrangements within the assembled spliceosome (complexes Bact and C) lead to splicing catalysis, generating a mature mRNA and releasing the intron in a lariat configuration. U2 small nuclear ribonucleoprotein (snRNP) binding involves base-pairing interactions between U2 small nuclear RNA (snRNA) and nucleotides flanking the BP, as well as contacts between the U2 snRNP proteins splicing factor 3B subunit 1 (SF3B1) and p14 and the pre-mRNA. BPRS, BP Recognition Sequence; SF1, splicing factor 1; ss, splice site; U2AF, U2 snRNP auxiliary factor.

During spliceosome assembly, a large number of protein-protein, protein-RNA and RNA-RNA interactions are orchestrated to facilitate compositional and conformational rearrangements. Important changes occur in U2 snRNA (Figure 4). The latter ones (Figure 5) are coordinated by important helicases and are necessary to bring the 5'ss and the branch sequence close to each other for splicing catalysis (Matera and Wang, 2014; Patel and Steitz, 2003; Wahl et al., 2009). The structure of the yeast catalytically activated spliceosome has recently been solved, elucidating the details of these interactions at the core of catalysis (Rauhut et al., 2016; Yan et al., 2016).

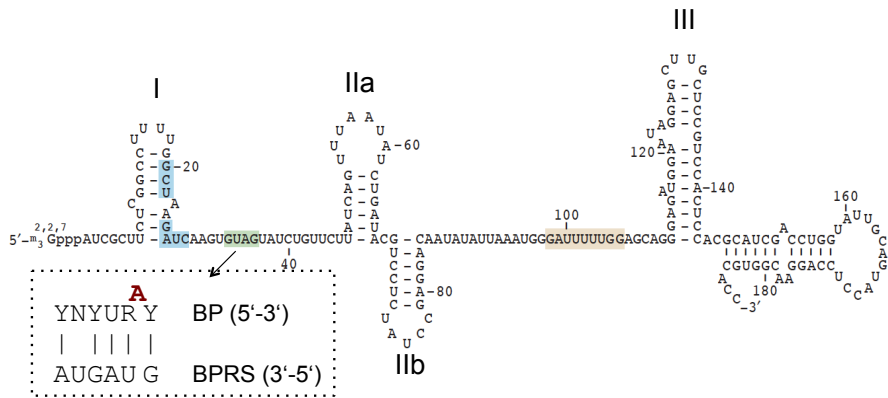


Figure 4. Secondary structure of U2 snRNA. The Sm-binding sites is shaded in light brown, the green box indicates interactions with the branch site and the blue box U2–U6 interactions. Stem-loop structures are indicated by roman numbers (adapted from Patel and Steitz, 2003). Stem-loops IIa and IIb and the BP recognition sequence undergo several remodelling steps during spliceosome assembly (Perriman and Ares, 2010; Perriman and Ares, 2007). Base-pairing between the pre-mRNA BP and the BP recognition sequence of the U2 snRNA (BPRS) leads to the branch adenosine protrusion from the pre-mRNA (as shown in the box).

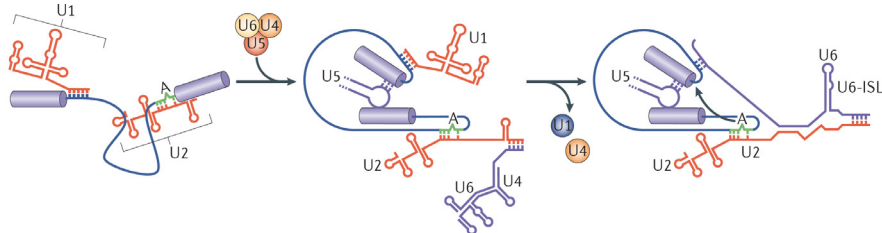


Figure 5. Remodelling of RNA-RNA interactions during spliceosome assembly. During the stepwise spliceosome assembly, RNA-RNA interactions are precisely coordinated to create the catalytic centre. In the complex A (left panel), U1 and U2 snRNAs base-pair with the 5'ss and the BP region, respectively. In the B complex (mid panel), U5 snRNA binds to exonic sequences close to the 5'ss and U6 to the U2 snRNA. U4 has some regions of base-pairing with the U6 snRNA. Thanks to precise helicase activities, U4 and U6 binding is unwound, U4 and U1 leave the spliceosome and U6 extends its base-pairing with the U2 snRNA and contacts the 5'ss as well. This conformation corresponds to the catalytically active B^{act} spliceosome (right panel), in which the branch adenosine and the 5'ss are brought close to each other, ready for the nucleophilic attack (adapted from Matera and Wang, 2014).

1.5. Regulation of alternative splicing

The role of splicing factors, chromatin and transcription

Alternative splicing can mainly manifest as exon skipping, intron retention, alternative 5' or 3' splice site usage and mutual exclusion of exons (Figure 1). These choices appear to be achieved through sophisticated regulation of the different steps of spliceosome assembly, especially in the early phases of intron recognition (Braunschweig et al., 2013).

Specific exonic and intronic sequence elements act as enhancers or silencers, favouring or disfavouring spliceosome assembly on specific intron boundaries. They are known as ESE (exonic splicing enhancers), ESS (exonic splicing silencers), ISE (intronic splicing enhancers) and ISS (intronic splicing silencers) (Figure 6). Generally, their effect in *cis* (i.e. within the same molecule, on the transcript they belong to) is mediated through the binding of trans-acting splicing factors (i.e. acting in *trans*, on molecules others than their own). These proteins work as accessory splicing factors, not strictly required for spliceosome assembly and splicing catalysis and therefore they do not belong to the *core* splicing machinery (that is the set of essential spliceosomal components), but they can influence its binding to specific introns by synergistic or antagonistic interactions (Braunschweig et al., 2013; Scotti and Swanson, 2016; Wahl et al., 2009). Given these combinatorial effects, the relative abundance of these regulators seems to be carefully controlled in different cell types; interestingly, alternative splicing is one of the ways in which splicing factors are regulated (Lareau and Brenner, 2015).

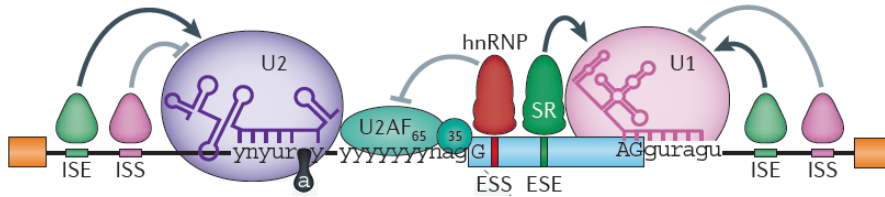


Figure 6. Representation of splicing silencers and enhancers. Thanks to the binding of mediator proteins, intronic splicing silencers and enhancers (ISE and ISS) can affect negatively or positively the binding of U2 or U1 snRNPs. A similar role is played by exonic silencers or enhancers (ESS and ESE). hnRNP and SR protein families are the best known RNA binding proteins recognizing these regulatory elements and leading to alternative splicing modulation. Different regulatory sequences within the same region display combinatorial effects. Thus, the abundance of the regulators is also a critical parameter to determine the splicing outcome (adapted from Scotti et al., 2016).

Current understanding of the role of primary sequences made it possible to predict, to a reasonable extent, broad tissue-specific or disease-related alternative splicing based on transcripts sequences and mutations affecting possible regulatory sequence motifs (Barash et al., 2010; Rosenberg et al., 2015; Xiong et al., 2015).

In addition to accessory splicing factors and regulatory sequences, the abundance of *core* spliceosomal proteins themselves can influence splicing and alternative splicing choices (Papasaikas et al., 2015; Saltzman et al., 2011; Wong et al., 2013). On the other hand, an important cross-talk with chromatin and transcription has been described (Iannone and Valcarcel, 2013; Kornblihtt et al., 2013). Indeed, although for a long time splicing has been considered -and it is still often classified as- a *post-transcriptional* mechanism of RNA processing, it is nowadays clear that most of splicing occurs cotranscriptionally, i.e. while RNA pol II is still transcribing the DNA

sequence into a pre-mRNA molecule (Beyer and Osheim, 1988; Nojima et al., 2015; Tilgner et al., 2012). The coupling of the two processes is so efficient that at least some introns can be spliced as soon as the 3' splice site is transcribed (Carrillo Oesterreich et al., 2016). Thus, it appears that there is a strong interplay between different layers of gene expression, with transcription kinetics, chromatin status and splicing choices influencing each other (Iannone and Valcarcel, 2013; Kornblihtt et al., 2013). In particular, it was shown in yeast that 3' splice site recognition causes RNA pol II stalling at intronic 3' ends, suggesting the existence of a checkpoint associated with cotranscriptional splicing (Alexander et al., 2010; Chathoth et al., 2014).

2. The spliceosome in disease and cancer

2.1. Splicing and genetic disease

Splicing defects are responsible for multiple genetic diseases

Being splicing a key cellular mechanism, it seems likely that its alteration has pathological consequences. Indeed, several mutations leading to disease reside within regulatory sequences important for splicing, or within genes coding for splicing factors. In both cases, the resulting misregulation of a subset of transcripts will determine a specific molecular defect and pathological manifestation (Daguenet et al., 2015).

One paradigmatic splicing-related disease is Spinal Muscular Atrophy (SMA), the leading genetic cause of infant mortality. This degenerative pathology is due to the loss-of-function of *SMN1* gene, coding for SMN protein, which is critical for snRNPs assembly. The homologous *SMN2* gene is not able to compensate for pathological mutations in *SMN1* because a nucleotide difference within *SMN2* exon 7 induces skipping of this exon and leads to the production of a truncated and unstable version of the SMN protein (Lefebvre et al., 1995; Lorson et al., 1999). It is still not clear why deficiency in this factor leads to a motor neuron-specific defect, but tissue-specificity is also associated to several other splicing-related diseases (Daguenet et al., 2015).

Interestingly, two main therapeutic approaches are being developed for SMA, and they are paradigmatic to summarize the current tools

for alternative splicing modulation. On one side, antisense oligonucleotides complementary to specific *SMN2* intronic silencer sequences are able to increase exon 7 inclusion and restore motor neuron function (Hua et al., 2010); on the other, small molecules have been found to have similar effects, with considerable specificity (Naryshkin et al., 2014; Palacino et al., 2015).

Among other splicing-related diseases, Duchenne Muscular Dystrophy (DMD) is induced by frame-shift mutations within the *dystrophin* gene, generating truncated proteins that are not able to maintain the integrity of muscular fibers and giving rise to the pathological condition. Antisense oligonucleotides approaches have been used to induce skipping of specific exons, in order to restore the coding frame. Because of the repetitive domains present in the large dystrophin protein, in-frame deletions caused by exon skipping generate proteins that can provide at least partial function and therefore therapeutic effects (Kole et al., 2012).

Finally, retinitis pigmentosa is a genetic disease consisting in progressive retina degeneration. It is caused by mutations within splicing factors genes, mainly *PRPF8* (Scotti and Swanson, 2016). This is the largest and most conserved spliceosomal protein, which plays a key function during splicing catalysis (Daguenet et al., 2015; Papasaikas and Valcarcel, 2016; Wahl et al., 2009; Yan et al., 2015).

The tissue-specific effects of splicing factor mutations are remarkable and still poorly understood. Even more striking is the fact the mutations within other spliceosomal components also cause pathological conditions restricted to specific but different organs. For example, *SF3B4* mutations are associated to Nager syndrome,

characterized by craniofacial and limb malformations (Bernier et al., 2012), while mutations in one of the mouse multicopy *U2 snRNA* genes lead to splicing alterations associated with ataxia and neurodegeneration (Jia et al., 2012). A deeper understanding of splicing-related diseases is one of the current challenges in the field.

2.2. Alternative splicing and cancer

Splicing isoforms are heavily altered in cancer

Tumor cells often display an altered balance of alternative isoforms that play relevant roles in preventing apoptosis or promoting proliferation and invasion (David and Manley, 2010).

FAS exon 6 skipping is among the most studied alternative splicing events. Full-length *FAS* (or CD95) is a membrane protein able to activate the apoptotic cascade when bound from the *FAS* ligand, *FASL*. Exon 6 skipping leads to the production of *FAS* short isoform, that is lacking the transmembrane domain and is secreted outside the cell, where it will be able to sequester *FASL* and therefore inhibit the apoptotic process (Figure 7). Several tumors switch *FAS* splicing towards the antiapoptotic isoform in order to escape apoptotic stimuli (Cheng et al., 1994). Recent high throughput screenings highlighted the dependence of this switch not only on splicing factors abundance but also on iron metabolism (Papasaikas et al., 2015; Tejedor et al., 2015).

Similarly, *MCL1* (Myeloid Cell Leukemia 1) transcript also encodes an apoptosis regulator whose function depends on alternative splicing

choices. MCL1 belongs to the Bcl-2 class of proteins, which regulate apoptosis by controlling mitochondrial outer membrane permeabilization. MCL1 displays antiapoptotic functions through its full-length isoform, usually predominant in abundance. Skipping of exon 2 gives rise to a protein isoform that misses trans-membrane domain and Bcl-like domains BH1 and BH2, only keeping the BH3 domain. This isoform exerts pro-apoptotic functions by losing interactions with MCL1 usual partners and sequestering the full-length protein (Bae et al., 2000) (Figure 7). Due to the involvement of MCL1 in the cancer and drug resistance processes, high throughput screenings have been performed to study its alternative splicing regulation (Moore et al., 2010; Papasaikas et al., 2015).

In other cases, alternative splicing can alter cellular metabolism of tumor cells: this is the case for example for *PKM*, which encodes the metabolic enzyme pyruvate kinase M. Two PKM isoforms with mutually exclusive exons can be generated: one is usually present in embryonic tissues, while the other is present in adult cells. Several tumor cells re-establish expression of the embryonic isoform, which favors aerobic glycolysis (Figure 7). This change is responsible for the so-called Warburg effect, namely the reactivation of aerobic glycolysis in tumor cells, which confers a growth advantage (Christofk et al., 2008).

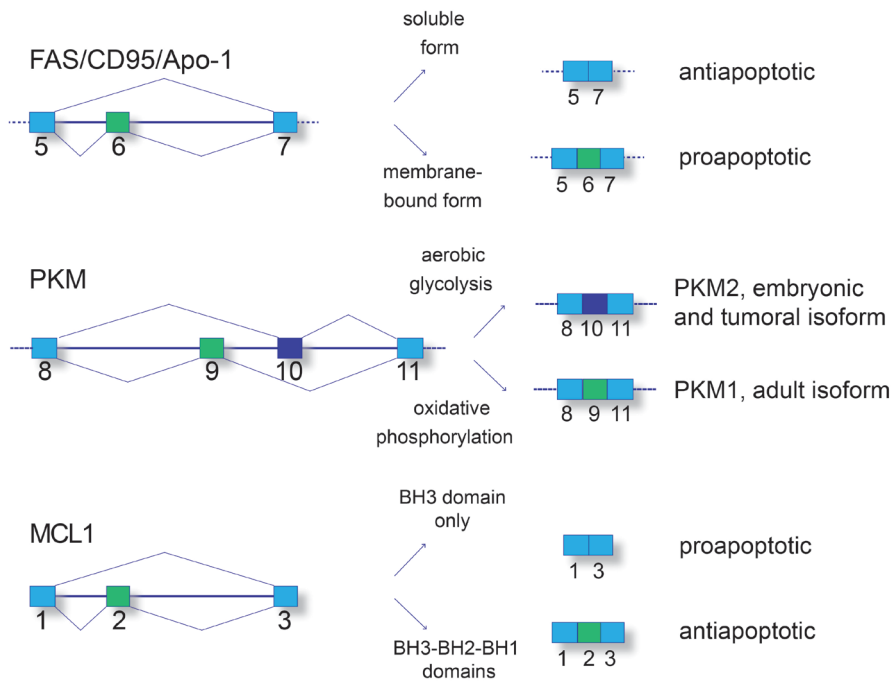


Figure 7. Examples of regulated alternative splicing events relevant for cancer progression. Alternative regions of the precursor mRNA are represented as green boxes and constitutive exons are shown in blue. The numbers correspond to the exons involved in the alternative splicing event for each of the indicated genes. The examples also illustrate the diverse functional outcomes of the encoded alternative protein products.

Interestingly, recent analyses documented that synonymous mutations are often acting as driver mutations in cancer. In several cases, this phenomenon could be related to a difference in regulatory sequences generating different splicing preferences in tumor cells. As a result, oncogenes and tumor suppressors can have their activity increased or inhibited through changes in splicing (Supek et al., 2014).

Many other examples have been reported so far, including the impact of signaling pathways activated in cancer cells on splicing aberrations, and how these splicing aberrations can control a variety of

mechanisms, including metastasis, angiogenesis and cell cycle (Bonnal et al., 2012; David and Manley, 2010).

2.3. Splicing factors and cancer

Several splicing factors are mutated in tumor cells

While cancer-related aberrant splicing has been known for long, mutations in splicing factors are only starting to be disclosed by large cancer sequencing efforts. Interestingly, several mutated proteins play a role in early steps of spliceosome assembly and more precisely in 3'ss recognition, suggesting that this level of regulation is somehow linked to cell proliferation control or other aspects of tumor progression. Also, most of the mutations are heterozygous and mutually exclusive, suggesting that different factors could have some common consequences and that complete loss of the wild-type versions of these factors is deleterious (Yoshida and Ogawa, 2014).

At first, *U2AF35*, *ZRSR2*, *SRSF2* and *SF3B1* mutations were found to be associated to myelodysplastic disorders (Yoshida et al., 2011). Strikingly, *SF3B1* mutations highly correlate with the presence of ring sideroblasts (RS), i.e. erythroblasts (erythrocyte precursors) with mitochondrial iron deposits forming perinuclear granules (Papaemmanuil et al., 2011). Mutations in this factor were then described for chronic lymphocytic leukemia (CLL) (Quesada et al., 2012) and solid tumors, including uveal melanoma (Furney et al., 2013). A flurry of clinical observations has followed these first

publications, associating the mutations with clear prognostic indicators (Bonnal et al., 2012; Yoshida and Ogawa, 2014).

Even if splicing is required for every gene to be expressed, relatively few transcripts have been found to be altered in mutant samples (Yoshida and Ogawa, 2014). Strikingly, sequence differences within affected transcripts are partially explaining the splicing alterations associated to *SF3B1*, *SRSF2* and *U2AF1* mutations.

In the case of *U2AF1*, mutations were described in hematological malignancies, lung cancer and other solid tumors (Dvinge et al., 2016). This splicing factor is involved in 3'ss recognition (Figure 3) and S34 mutation leads to a preference 3'ss where the AG is preceded by C or A rather than T, while Q157 mutation favors introns starting with G rather than A (Ilagan et al., 2015; Okeyo-Owuor et al., 2015; Przychodzen et al., 2013; Shirai et al., 2015).

On the other hand, *SRSF2* is a regulatory SR protein binding to exonic splicing enhancers (ESEs) (Figure 6). Pathological mutations induce a change in its RNA recognition motif (RRM), leading to a difference sequence preference: mutant *SRSF2* binds more efficiently to CCNG rather than GGNG motifs, while the wt protein can efficiently recognize both (Kim et al., 2015; Lee et al., 2016; Zhang et al., 2015).

Collectively, these observations show that the RNA affinity of mutated splicing factors is altered and that differential RNA binding can cause sequence-dependent alternative splicing changes, some of which can partially explain the pathologic phenotype (Dvinge et al., 2016). Nevertheless, also processes other than alternative splicing can

be affected and partially contribute to tumor progression, like described for the autophagy regulator *ATG7*, whose 3' end processing is altered in presence of *U2AF1* mutations (Park et al., 2016).

Although research in the field is progressing very fast, many mechanistic questions remain to be answered in order to be able to translate this knowledge into therapy.

2.4. Splicing inhibitors: a novel class of drugs

Small molecules targeting the spliceosome have antitumor properties

Several small molecules regulating splicing efficiency in a general or event-specific way have been identified (Berg et al., 2012; Naryshkin et al., 2014; O'Brien et al., 2008; Palacino et al., 2015; Zaharieva et al., 2012). Due to the increasing interest on such compounds, high throughput systems have been established to screen drugs libraries and find splicing modulators (Berg et al., 2012; Effenberger et al., 2015; Naryshkin et al., 2014; O'Brien et al., 2008; Palacino et al., 2015; Pawellek et al., 2014). Nevertheless, a direct interaction between the small molecule and *core* spliceosome components is known only for a few of them. Intriguingly, several of them have as common target the SF3B (splicing factor 3B) complex (Bonnal et al., 2012; Effenberger et al., 2016a; Zaharieva et al., 2012).

Screenings for antitumor candidates within bacterial *Pseudomonas* and *Streptomyces platenienseis* fermentation products lead to the discovery of FR901464, Pladienolides, GEX1A or Herboxidiene and Thailanstatines (Liu et al., 2013; Miller-Wideman et al., 1992;

Nakajima et al., 1996b; Sakai et al., 2004; Sakai et al., 2002b). These compounds display cytotoxic effects against multiple tumor cell lines, where they cause a cell cycle arrest in G1 and G2/M phases (Mizui et al., 2004; Nakajima et al., 1996a; Sakai et al., 2002b). Strong *in vivo* antitumor effects were also observed (Mizui et al., 2004; Nakajima et al., 1996b). The drugs' unique behavior compared to known antitumor drugs drew great interest and the involvement of a novel antitumor mechanism was hypothesized. Furthermore, the detection of *Cdc25a* and *Cdc2* short isoforms in GEX1-treated cells suggested that splicing could be critical for the effects of the drugs on cell proliferation control (Sakai et al., 2002a) .

The confirmation came a few years later, when a potent derivative of FR901464 was found to inhibit splicing by targeting SF3B complex and was therefore named as Spliceostatin A (Kaida et al., 2007). The same effect was observed for Pladienolide B (Kotake et al., 2007) and, more recently, for GEX1A (Hasegawa et al., 2011). These biochemical studies were consolidated by genetic results showing that a novel *SF3B1* mutation reverts both splicing and proliferation inhibition induced by Pladienolide B by suppressing the drug's binding to SF3B complex (Yokoi et al., 2011). Interestingly, drug-resistant mutation R1074H is located in very close proximity with the branch adenosine (Rauhut et al., 2016). Nevertheless, the detailed structural characterization of drug-target interactions has not yet been achieved for any of these compounds.

The high chirality and low solubility of the natural drugs restrict their use in *in vitro* and *in vivo* studies. To address these issues, FR901464 derivatives have been synthesized. Meayamycin and Meayamycin B

display more stability, potency and solubility (Albert et al., 2009; Osman et al., 2010), while Sudemycins are easier to synthesize chemically and recapitulate the hypothesized common pharmacophore shared by FR901464 and Pladienolides (Fan et al., 2011; Lagisetti et al., 2009; Lagisetti et al., 2008).

Indeed, the interest on the spliceosome as a therapeutic target and on the therapeutic potential of these molecules is expanding, with the double expectation that they could provide powerful tools for therapy as well as for the study of spliceosomal mechanisms (Disney, 2008; Jurica, 2008; Schneider-Poetsch et al., 2010). As antibiotics revealed to be extremely powerful tools for the study of the ribosome, small molecules targeting the spliceosome would definitely help to improve our knowledge of the most complex and dynamic ribonucleoproteic machine of the cell (Schneider-Poetsch et al., 2010; Wahl et al., 2009).

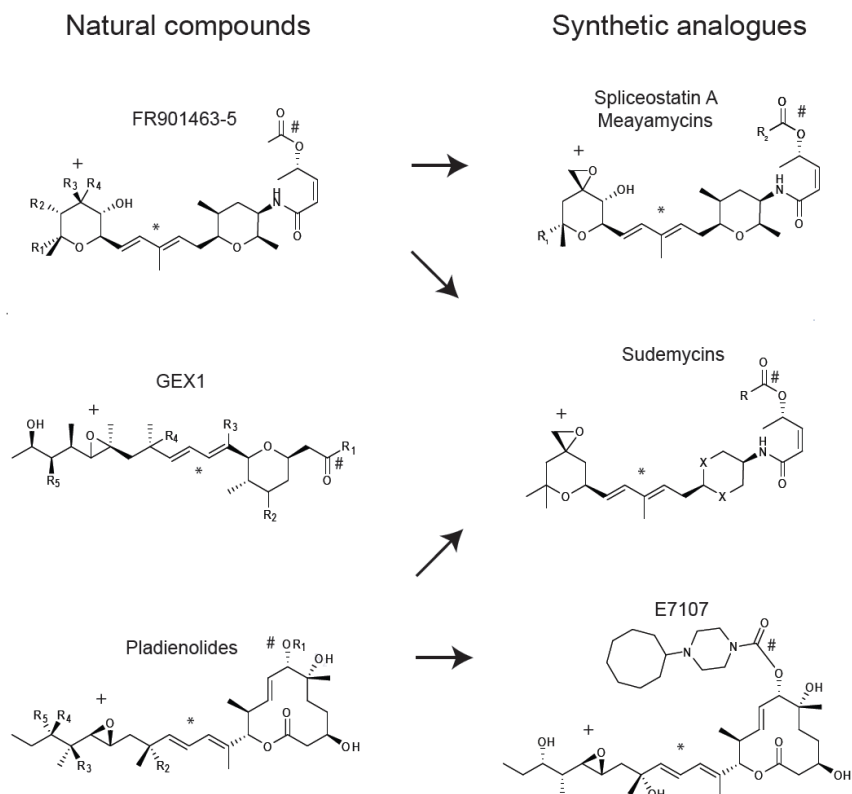



Figure 8. Antitumor compounds targeting SF3B1. General chemical structures of the main families of drugs are illustrated. Important groups for their function include the conjugated diene (*), the oxycarbonyl moiety (‡) and the epoxide (§). Synthetic analogues derived from natural products are represented in the right column, with arrows indicating the parental compound. GEX1 molecules are also known as Herboxidienes (Sakai et al., 2002a; Sakai et al., 2002b). FR901463-5-related variants of natural compounds were recently described as Thailanstatins (Liu et al., 2013). FD-895 is a Pladienolide-related compound, whose carbohydrate derivative was shown to have improved stability (Dhar et al., 2016). “R” indicates side groups that differ among variants of the same family of compounds (Bonnal et al., 2012; Effenberger et al., 2016a). The most active variants of each group and their side moieties correspond to the following: FR901464 (R₁: OH; R₂: H; R₃: epoxide; R₄: H); GEX1A (R₁: OH; R₂: H; R₃: CH₃; R₄: H; R₅: OCH₃); Pladienolide B (R₁: COCH₃; R₂: H; R₃: H; R₄: H; R₅: OH); Spliceostatin A (R₁: OCH₃; R₂: CH₃), Meayamycin (R₁: CH₃; R₂: CH₃); Meayamycin B (R₁: CH₃; R₂: ); Sudemycin C1 (R: CH(CH₃)₂; X: CH₂); Sudemycin D1 (R: N(CH₃)₂; X: CH₂); Sudemycin D6 (R: NHCH₃; X: CH₂).

3. SF3B1: function and dysfunction

3.1. U2 snRNP function and components

3' splice site binding and initial spliceosome assembly

The U2 snRNP complex is very dynamic and includes U2 snRNA and different protein complexes, including the complexes Splicing Factor 3A and 3B (SF3A and SF3B). SF3A is composed of three subunits (SF3A1, SF3A2 and SF3A3, also known as SF3a120, SF3a66 and SF3a60 based on their molecular weight), while SF3B encompasses at least eight subunits of 155, 145, 130, 49, 10, 14a (or p14), 14b and 125 kDa (also named on their molecular weight or SF3B1, SF3B2, SF3B3, SF3B4, SF3B5, SF3B6, SF3B7 and SF3B8). The precise function of most of these proteins is still unclear (Wahl et al., 2009; Will and Luhrmann, 2011).

The key role of U2 snRNP in the splicing process is the binding to the intronic 3'ss region after formation of an early spliceosome (complex E), leading to the pre-spliceosomal complex A (Figure 3). Regulation of the recruitment of U2 snRNP is believed to be one key way to regulate alternative splicing (Graveley et al., 2001; Graveley and Maniatis, 1998; Guth et al., 1999; Guth et al., 2001; Valcarcel et al., 1993).

Sequence-independent binding of conserved U2 snRNP proteins upstream of the branch site is required for assembly of complex A. This region was named “anchoring site” and it crosslinks with the SF3A and SF3B subunits in the apparent 5' to 3' linear order SF3B4,

SF3A1, SF3A3, SF3B2, SF3A2 and SF3B1, apparently in a sequence non-specific manner (Gozani et al., 1996). Interestingly, blocking the anchoring site with antisense oligonucleotides was found to switch BP selection in a model pre-mRNA containing duplicated BPs (Dominski and Kole, 1994).

Remarkably, some U2 snRNP components (the three members of SF3A complex and five of SF3B ones) are conserved also in *Cyanidioschyzon merolae*, a red algae with very few introns and with a very reduced spliceosome. Although missing many protein components, this spliceosomal machinery contains most of the proteins involved in BP recognition, suggesting that this is one of the most critical steps of the reaction (Hudson et al., 2015; Stark et al., 2015).

In addition to the conventional “major” spliceosome, several eukaryotic organisms contain also the “minor spliceosome”. This machinery processes a subset of introns with different canonical splice site sequences, named U12-introns because they are bound by U12 snRNP, in contrast with conventional U2-introns, bound by U2 snRNP. Interestingly, the minor spliceosome only shares few protein components with the major one: among these, members of SF3B (but not of SF3A) (Will and Luhrmann, 2005). This additional evidence highlights the importance of SF3B components across several classes of spliceosome machineries.

3.2. SF3B1 protein

A persistent player of the spliceosome and beyond

The correct recognition of the BP by base-pairing interactions with U2 snRNA is a key step for the splicing process. Indeed, base-pairing between the pre-mRNA BP region and the BP recognition sequence (BPRS) in U2 snRNA occurs in such a fashion that the BP adenosine is bulged out of the helix, a configuration that is important for the first catalytic reaction that leads to the formation of a phosphodiester bond between 5' splice site and branch adenosine (Figure 5) (Parker et al., 1987). p14 subunit of SF3B complex interacts with the branch adenosine and with the SF3B1 (also known as SF3b155) subunit (Schellenberg et al., 2011). The latter also interacts with U2AF65 and sequences upstream and downstream of the BP (Gozani et al., 1998).

SF3B1 is highly conserved across eukaryotes and its function is fundamental for stabilizing U2 snRNP recruitment (Wahl et al., 2009; Corriero et al., 2011). This protein is persistent within the spliceosome throughout its whole assembly and catalytic function (Rauhut et al., 2016; Wahl et al., 2009; Yan et al., 2016).

SF3B1 encompasses an unstructured N-terminal domain involved in protein-protein interactions and 22 carboxy-terminal HEAT (huntingtin, elongation factor 3, protein phosphatase 2A, target of rapamycin 1) repeats (Figure 10). HEAT repeats consist of 37–47 aminoacid residues forming two antiparallel α -helices and two turns, with conserved aspartic and arginine residues and short flexible linkers between repeats (Groves and Barford, 1999; Xing et al., 2006).

The architecture of SF3B1 HEAT-repeats results in a cavity surrounding the p14 protein: unknown conformational changes from a more open structure are required to allow the RNAs duplex and p14 to integrate within the shell, and HEAT-repeats curvature may play a role in it (Golas et al., 2003). These interactions play a key role to allow further spliceosome rearrangements.

SF3B1 undergoes several phosphorylation reactions on different residues of the N-terminal domain just before the catalytic step (Bessonov et al., 2010), but the function of these modifications has not been explained. Nevertheless, NIPP1 protein has been found to interact with SF3B1 depending on its phosphorylation status, which is highly increased during mitosis (Boudrez et al., 2002).

SF3B1's N-terminal 450 amino acid unstructured domain interacts with U2AF65 and with p14, that induces a folding transition (Spadaccini et al., 2006), while the C-terminal domain encompasses 22 tandem HEAT-repeats (Golas et al., 2003). Recently, enormous advances in the structural description of the yeast spliceosome have been achieved. The structure of SF3B1's homolog Hsh155 within this complex highlights its key role for catalysis. Sequences immediately downstream of the branch adenosine enter the HEAT-repeats cavity, which has a positively charged surface. Prp2 interacts with Hsh155, possibly promoting the exposure of the branch adenosine for splicing catalysis (Yan et al., 2016) that is otherwise masked by Hsh155 HEAT repeats to prevent premature catalytic activation (Rauhut et al., 2016). In this context, Prp2 moves along the pre-mRNA and induces Hsh155 HEAT repeats remodeling to make the adenosine accessible for catalysis (Rauhut et al., 2016).

SF3B1 was found to crosslink upstream and downstream of the BP and to interact with U2AF65 and U2AF35 (Gozani et al., 1998). It was observed that the distance between the BP and the Py-tract is important for efficient complex A formation and that among multiple BPs, the downstream is used if it is not too close to the Py-tract (Gozani et al., 1998).

Apart from being a key but still enigmatic component of the spliceosome, SF3B1 is also involved in other functions. On one hand, it plays a role in mouse development as a component of the epigenetic repressor Polycomb complex (Isono et al., 2005). On the other, it was found to interact with chromatin and proposed to participate as such in alternative splicing regulation (Kfir et al., 2015). SF3B1 binds nucleosomes located on GC rich exons, flanked by long introns. It was proposed that it "jumps" from chromatin to the 3'ss of nascent RNAs of lowly expressed genes in order to facilitate splice site recognition (Kfir et al., 2015). The association of SF3B1 with several cellular processes keeps expanding and it sheds light on the importance of this factor for cell physiology.

3.3. Consequences of SF3B1 mutations

A role in disease progression and in branch point choice

Mutations in *SF3B1*, as well as in other 3' splice site-recognizing factors, are recurrent in cancer (Bonnal et al., 2012; Yoshida and Ogawa, 2014). *SF3B1* mutations are particularly frequent in Myelodysplastic Syndromes with Refractory Anemia and Ring

Sideroblasts (RARS) (Papaemmanuil et al., 2011; Yoshida et al., 2011) and in chronic lymphocytic leukemia (CLL) (Quesada et al., 2012; Rossi et al., 2011; Wang et al., 2011), but also in uveal melanoma (Furney et al., 2013; Harbour et al., 2013), mesothelioma (Bueno et al., 2016), breast, ovary and pancreas cancers (Ellis et al., 2012b; Pleasance et al., 2010; Wood et al., 2007). *SF3B1* mutations are prevalently missense, leading to single residues substitutions (Figure 9).

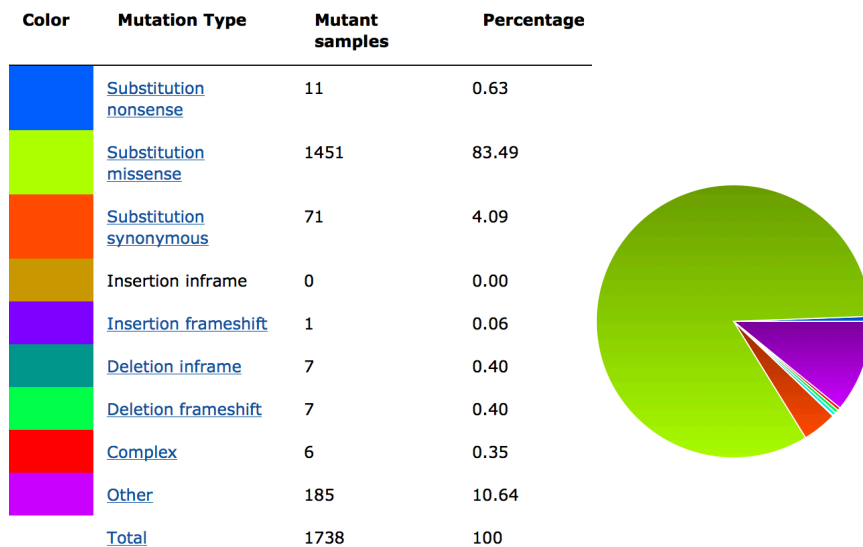


Figure 9. Distribution of *SF3B1* mutations from the COSMIC Database. The frequency of the different types of mutation for *SF3B1* in tumor samples is shown (from COSMIC database, <http://cancer.sanger.ac.uk/cosmic/>).

Intriguingly, most mutations are located in hotspots within specific *SF3B1* HEAT repeats (Bueno et al., 2016; Darman et al., 2015; Harbour et al., 2013; Quesada et al., 2012; Rossi et al., 2011) (Figure

10). Depending from the disease, these mutations can be related to either a better or poorer prognosis (Malcovati et al., 2011; Papaemmanuil et al., 2011; Visconte et al., 2012; Wang et al., 2011).

It was recently shown that most of the alternative splicing events related to the presence of *SF3B1* mutation are cryptic 3'ss, whose activation is sequence-dependent (Alsafadi et al., 2016; Bueno et al., 2016; Darman et al., 2015; DeBoever et al., 2015; Ferreira et al., 2014; Kesarwani et al., 2016).

A first hypotheses was that cryptic splice site activation was due to a shift in the steric protection of sequences downstream of the BP by SF3B1: in presence of *SF3B1* mutation, the protection would be less tight and AG dinucleotides proximal to the BP would be used as splice sites (DeBoever et al., 2015). In contrast, follow up functional studies revealed that the usage of alternative 3' ss is also associated to a change in BP usage in the presence of *SF3B1* hotspots mutations but not in the presence of rarer mutations (Alsafadi et al., 2016; Darman et al., 2015). These changes require the canonical Py-tract and BP to be intact and they have been related to a change of charge induced by the hotspot mutations within the HEAT repeats (Darman et al., 2015), which can be interpreted as alterations in electrostatic interactions with the RNA according to the recent cryo-EM structure of the yeast spliceosome (Rauhut et al., 2016; Yan et al., 2016). More than displaying a loss- or gain- of- function, *SF3B1* mutations would therefore have a change-of-function, where higher preference to strong BP sequences would be manifested (Alsafadi et al., 2016). Recent results also show that cryptic 3'ss are generally inaccessible because of RNA secondary structures and they get activated in a

mutation-dependent manner (Kesarwani et al., 2016).

An interesting observation came from the comparison of the splicing alterations induced by *SF3B1* mutations in human and mouse cells: sequence features associated with cryptic splice sites usage are similar in both organisms (i.e. presence of short and weak Py-tracts and enrichment of As), but affected junctions have minimum overlap, due to low intronic sequence conservation (Darman et al., 2015; Mupo et al., 2016; Obeng et al., 2016). This evidence suggests that the phenotype might depend on global effects linked to splicing disruption rather than on specific alternative splicing events (Mupo et al., 2016). For example, recent studies related SF3B1 function to epigenetic regulation (Kfir et al., 2015), or its mutation and inhibition to DNA damage (Te Raa et al., 2015; Wan et al., 2015) and these pathways may contribute to cancer phenotypes.

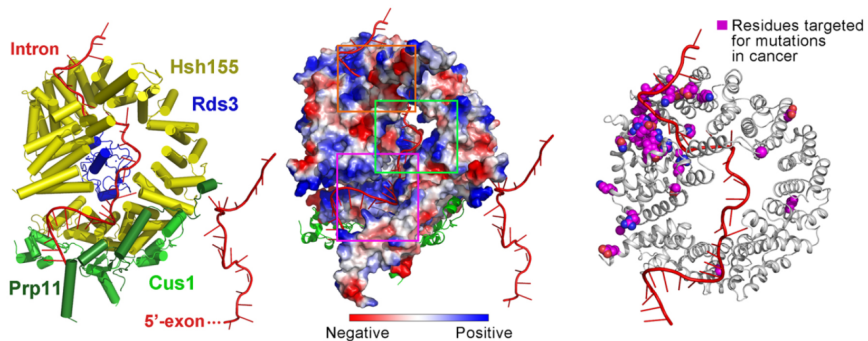


Figure 10. Structure of yeast SF3B1 and position of mutated residues.

Intronic sequences around the BP are recognized by SF3B1 (yeast Hsh155) and SF3B7 (yeast Rds3) (left panel), due to electrostatic interactions with their positively charged surface (mid panel). Most of the residues mutated in human cancers map to the 3rd and 4th HEAT-repeats, corresponding to a surface with positive charge that bind to pyrimidine-rich sequences downstream of the BP (adapted from Yan et al., 2016). Yeast Hsh155: human SF3B1 (or SF3B155); yeast Rds3: human SF3B7 (or SF3B14B or

PHF5A); yeast Cus1: human SF3B2 (or SF3B145); yeast Prp11: human SF3A2 (or SF3A66).

3.4. Mechanistic consequences of SF3B inhibition

Troubled branch point recognition and binding

In cultured cells Spliceostatin A induces a similar splicing inhibition to those induced by morpholinos that deplete U2 snRNA (Kaida et al., 2010) or by SF3B1 knockdown (Corrionero et al., 2011). Early studies on Spliceostatin A showed that it blocks early stages of spliceosome assembly (Roybal and Jurica, 2010). It was then described that the drug binds SF3B1 and prevents its interaction with the pre-mRNA, leading to the recruitment of U2 snRNA to “decoy” sequences upstream to its real binding site, the BP sequence (Corrionero et al., 2011). On the other hand, the Pladienolide-derivative E7107 alters the balance between alternative U2 snRNA conformations, implying a role for SF3B in ATP-dependent remodelling of the U2 snRNA structure (Folco et al., 2011). This structural switch depends on Prp5 helicase activity and it involves the region base-pairing with the BP: drug treatment disfavours the conformation required for stable spliceosome assembly (Folco et al., 2011; Perriman and Ares, 2010). Thus, SF3B1 seems to be involved in multiple interactions important for U2 snRNP binding at early steps of spliceosome assembly (Figure 11).

More recently it was shown that non-active and active drugs can compete with each other for binding to their target, suggesting that binding is not the only requisite for inhibition but rather that the

active ones induce a specific conformational change. Also, if added at late spliceosome steps, drugs can inhibit exon ligation, highlighting a role of SF3B complex and U2 snRNA also at late steps of spliceosome assembly (Effenberger et al., 2016b).

Apart from a direct effect on spliceosome dynamics, SF3B1 inhibition has been related also to cotranscriptional and epigenetic mechanisms. Indeed, it was shown that Sudemycin E binds SF3B1 and induces dissociation of U2 snRNPs and a decrease in their interaction with nucleosomes, resulting in a short-term reversible modulation of alternative splicing and in a long-term persistent regulation of gene expression as a consequence of epigenetic effects. It has been proposed that Sudemycin E interferes with the ability of U2 snRNP to maintain an H3K36me3 modification in actively transcribed genes (Convertini et al., 2014). Distribution of this epigenetic mark is indeed heavily altered by drug treatment (Kim et al., 2011). In addition, splicing inhibition results in a reduction of Ser2 phosphorylation in RNA pol II CTD (C-terminal repeat domain). This mark is important for transcription elongation and it induces a premature release of RNA pol II from some transcripts, in a gene-specific manner (Koga et al., 2015), although it was suggested that general RNA pol II elongation rate and snRNPs recruitment are not affected by drug treatment (Brody et al., 2011; Schmidt et al., 2011).

Spliceostatin A treatment affects the association of RNA pol II with the template after poly(A) addition and abrogates a quality control mechanisms that constrains unprocessed RNAs at the transcription site (Martins et al., 2011).

Overall, it is clear that SF3B inhibitors cause very diverse effect in treated cells. Assuming that they only target SF3B complex, this could be due to SF3B's multifaceted activities also outside the core spliceosome, or to the strict coupling of several gene expression processes (Moore and Proudfoot, 2009).

3.5. Functional consequences of SF3B inhibition

Targeting spliceosome-dependent cancer cells

SF3B-targeting drugs induce a cell cycle arrest in G1 and G2/M phases (Mizui et al., 2004; Nakajima et al., 1996a; Sakai et al., 2002a), displaying low nanomolar *in vitro* activity and significant antitumor effects in animal models (Mizui et al., 2004; Nakajima et al., 1996b; Sakai et al., 2002a; Sakai et al., 2002b). Furthermore, toxicity is generally higher in cancer cells (Albert et al., 2009; Fan et al., 2011; Kaida et al., 2007; Kotake et al., 2007), even in cases of multidrug-resistance (Albert et al., 2009). Thus, substantial interest was raised on these drugs as potential novel anti-cancer therapeutics, although clinical trials with E7107 (a Pladienolide variant) were stopped for visual side effects in a subset of patients (Dvinge et al., 2016). Clinical studies with H3B-8800 are being started to specifically target cancer cells with mutations in SF3B1 (Chakradhar, 2016).

What would be the physiological drug targets for these effects? Short isoforms of cell division-related proteins (including p27, Cdc25a and Cdc2) accumulate in cells treated with Spliceostatin A (SSA) (Kaida et al., 2007) or GEX-1 (Sakai et al., 2002a). At least in some cases,

production of truncated proteins affecting cell cycle derives from leakage of unspliced pre-mRNAs to the cytoplasm (Kaida et al., 2007; Kanada et al., 2007; Satoh and Kaida, 2016) (Figure 11).

SF3B1 inhibition induces abundant intron retention, with some genes being particularly sensitive. For example, intron retention and downregulation of vascular endothelial growth factor (*VEGF*) mRNA expression by Pladienolide (Mizui et al., 2004) and SSA (Furumai et al., 2010) could contribute to the drugs' anti-angiogenic and tumor regression effects. SSA was observed to affect alternative splicing as well, leading mostly to exon skipping. Some of the induced isoforms contain premature stop codons and are degraded by Nonsense Mediated Decay (NMD). Consequently, some cell cycle control genes might be downregulated (Corrionero et al., 2011) (Figure 11). In the case of MDM2, a negative regulator of p53, SSA and Sudemycin C1 induce short isoforms with reduced activity (Fan et al., 2011). Consistently, splicing disruption activates p53 pathway (Allende-Vega et al., 2013).

Regarding the specific effects of SF3B inhibitors in tumor cells, studies in chronic lymphocytic leukemia (CLL) indicated that drugs induce more apoptosis in CLL cells than in healthy lymphocytes (Kashyap et al., 2015; Xargay-Torrent et al., 2015). Selectivity for leukemia stem cell maintenance was found also in acute myeloid leukemia (AML) cells (Crews et al., 2016). *MCL1* splicing modulation, previously reported for this class of drugs (Gao and Koide, 2013; Papsaikas et al., 2015), was also found to be important in CLL and AML cells and possibly mediate the main cytotoxic effects of the drugs (Crews et al., 2016; Larrayoz et al., 2016; Xargay-Torrent et al.,

2015). In these contexts, cells with *SF3B1* mutations appear to be more sensitive to SF3B1 pharmacological inhibition than wild type cells (Obeng et al., 2016; Xargay-Torrent et al., 2015).

In mesothelioma models, in contrast, SSA treatment affects more the SF3B1 wild-type NCI-H2803 than the SF3B1-mutated NCI-H2595 cell line. Of potential relevance, gene expression but not alternative splicing changes are induced by the drug in both cell lines (Bueno et al., 2016), consistently with previous studies reporting drug-related gene expression alterations (Convertini et al., 2014; Lee et al., 2016).

Promising results were obtained also in melanoma cells: an intronic mutation induces exon skipping (with predicted redundant BPs playing a role in the mutant sequence), resulting in vemurafenib-resistance. SF3B1 inhibitors prevent the splicing change, reverting the vemurafenib-resistant phenotype (Salton et al., 2015).

Interestingly, overexpression of *cMYC* oncogene makes tumour cells more sensitive to SF3B inhibition (Hsu et al., 2015; Hubert et al., 2013), suggesting that cMYC-induced transcriptional activation generates the need of maintaining high levels of splicing activity and therefore higher sensitivity to splicing inhibitors. As a consequence, the spliceosome may confer a therapeutic vulnerability for a subset of cancers (Hsu et al., 2015). In conclusion, interfering with the spliceosome machinery in conditions where splicing becomes rate-limiting (because of increased transcription, mutations in pre-mRNA sequences or in splicing factors) has more dramatic consequences for the cell. This brings up the concept of synthetic lethality that was also suggested to explain the lack of evidence for coexistence of splicing factor mutations (Dvinge et al., 2016).

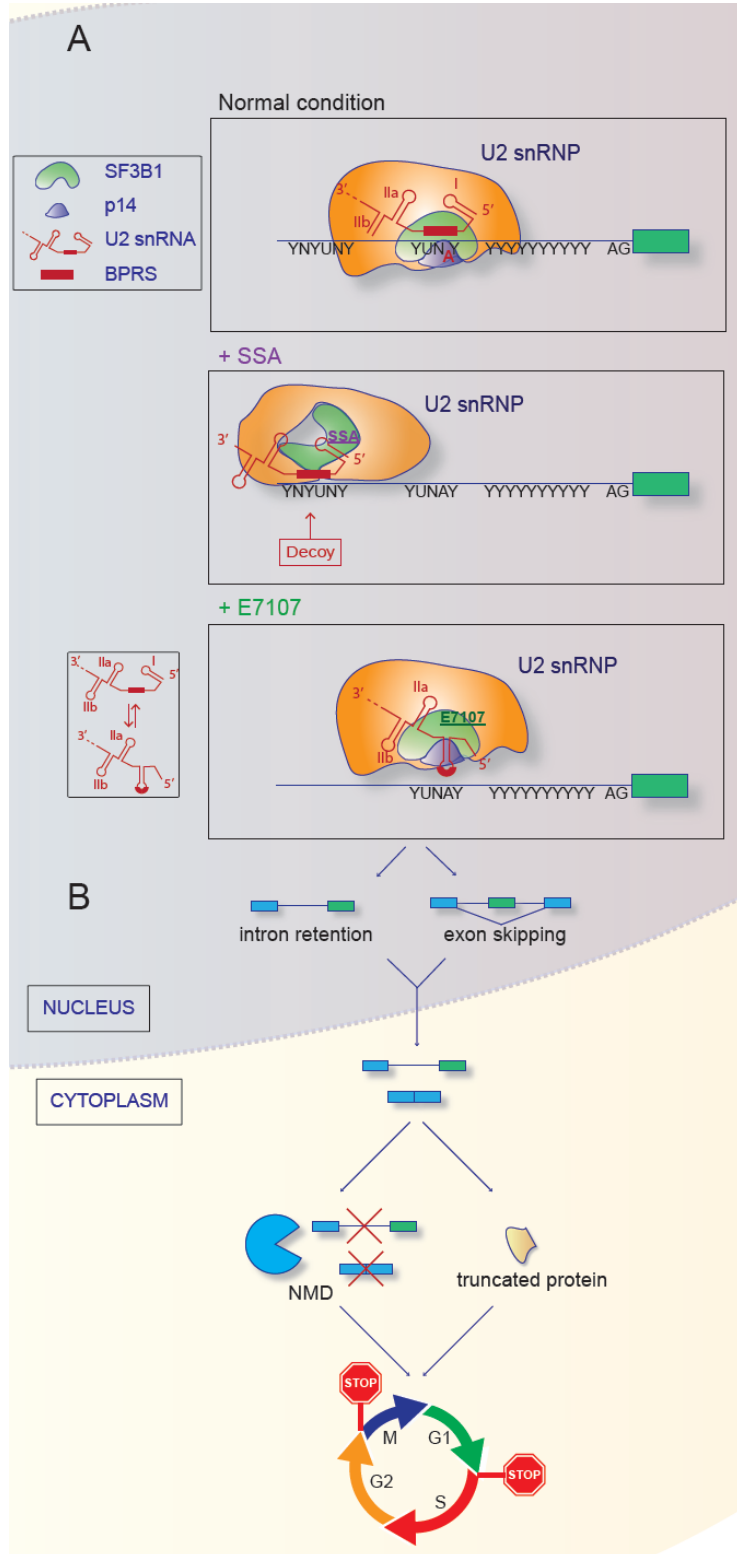


Figure 11 (previous page). Molecular effects of SF3B-targeting drugs on 3' splice site recognition and other steps of gene expression. (A) Effects of the drugs on 3' splice site recognition. U2 small nuclear ribonucleoprotein (snRNP) assembly involves base-pairing interactions between the BP recognition sequence (BPRS) in U2 small nuclear RNA (snRNA) and nucleotides flanking the BP adenosine, with the adenosine bulged out from the base-paired region. In addition, splicing factor 3B subunit 1 (SF3B1) crosslinks both 5' and 3' to the BP, and another SF3B protein, p14, contacts the BP adenosine. Spliceostatin A (SSA) prevents SF3B1–precursor RNA (pre-mRNA) interactions and induces base-pairing of U2 snRNA with decoy sequences upstream to the BP sequence. E7107 favours a conformation of U2 snRNA that establishes initial interactions with the BP while preventing an alternative conformation that allows stable U2 snRNP assembly. This stabilized initial conformation is known as BP-interacting stem-loop and its disruption is dictated by Prp5 helicase activity (Perriman and Ares, 2010). (B) Downstream effects of the drugs on gene expression. The splicing inhibitory effects of the drugs result in intron retention or exon skipping. Unprocessed or incorrectly processed mRNAs can leak to the cytoplasm and be translated or be subject to nonsense-mediated decay (NMD). One consequence of these effects is arrest in phases G1 and G2/M of the cell cycle. Exons are shown as blue or green boxes.

Objectives

Given the potential relevance of small molecules for understanding a) the complex process of spliceosome assembly and differential intron removal, and b) the properties of these compounds to control cancer cell proliferation, we set up the following goals for this thesis:

1) Exploring the key structural features of these compounds impacting on their effects on spliceosome assembly, alternative splicing modulation and cancer cell proliferation, as well as generating and functionally testing novel drug variants that could expand the chemical space of the backbone of Sudemycins: **analysis of structure-activity relationship for SF3B-targeting drug variants (part I.A).**

2) Exploring the possibility of achieving splice site-specific inhibition of splicing by conjugating SF3B-targeting drugs to antisense oligonucleotides (ASOs), with the goal of targeting the drugs and making them act at concentrations typical of ASOs, thus potentially improving the efficiency and selectivity of their function. Applications could include the specific modulation of cancer-related splicing events, limiting the collateral general inhibition potentially harmful for non cancer cells: **analysis of the activity of drugs-ASO conjugates (part I.B).**

3) Exploring the reasons why drugs targeting SF3B components, which would in principle affect general steps in 3' splice site recognition, can produce inhibitory effects specific of particular introns/splice sites and specifically affect cell growth of cancer cells. Our hypothesis is that

some sequence elements are critical for the transcripts' sensitivity to the drug and that they are present in genes that play key roles in cell proliferation regulation, thus potentially explaining the effects of these compounds as anti-proliferative reagents: **identification of sequence elements determining alternative splicing sensitivity to pharmacological inhibition of SF3B1 (part II).**

4) Exploring whether structurally similar -but not identical- drugs cause distinct effects on alternative splicing and, if so, what are the molecular basis for these differences: **comparison of the effects of different drug variants on alternative splicing modulation (part III).**

Results

Part I.A. Analysis of structure-activity relationship for SF3B-targeting drug variants

Submitted Manuscript

Sudemycin K: a synthetic anti-tumor splicing inhibitor with improved activity and versatile chemistry

Makowski K, Vigevani L, Albericio F, Valcárcel J, Álvarez M. [Sudemycin K: A Synthetic Antitumor Splicing Inhibitor Variant with Improved Activity and Versatile Chemistry](#). ACS Chem Biol. 2017 Jan 20;12(1):163–73. DOI: 10.1021/acscchembio.6b00562

Part I.B. Analysis of the activity of drug-antisense oligonucleotide conjugates

Collaborative project with Clément Paris, Kamil Makowski, Mercedes Álvarez and Enrique Pedroso (University of Barcelona)

Introduction

Antisense oligonucleotides (ASOs) complementary to sequences within the pre-mRNA are efficiently used to modulate alternative splicing events for either research or clinical purposes. Their therapeutic effects are promising for several pathologies, with SMA on top of the list (Daguenet et al., 2015; Havens and Hastings, 2016; Hua et al., 2010; Kole et al., 2012).

In contrast with pharmacological treatments, ASOs target specific transcripts, since their effect is based on the steric blockage of the access of proteins and complexes to the specific RNA region they are complementary to. On the other hand, drugs targeting the activity of splicing factors provide unique tools to modify the function of the spliceosome, but they may affect multiple transcripts within the cell and therefore their use in therapy may be limited by possible side effects. Intriguingly, recent studies identified pharmacological compounds targeting *SMN2* splicing and displaying a satisfying specificity, without extensive effects on other RNAs (Naryshkin et al., 2014; Palacino et al.,

2015), although trials were paused because of collateral effects on primates' eyes (Chakradhar, 2016).

While considering advantages and disadvantages of each of these approaches, we came up with the possibility to combine the power of splicing inhibiting drugs with the specificity provided by ASOs. The main concept behind this approach would be to direct the inhibitory effects of the drugs to specific transcripts, such that the drugs could be used at concentrations at which ASOs work, which are several order of magnitude lower.

Conjugation of ASOs with different chemical moieties has been extensively studied with the aim of increasing their cell specificity (Havens and Hastings, 2016). A conjugation strategy has been recently proposed for improving the potency for *SMN2* modulation by linking *SMN2*-specific oligonucleotides to a known PP1 phosphatase inhibitor. This approach was expected to increase splicing factor Tra2 β phosphorylation in the proximity of *SMN2* transcripts to favour exon inclusion. Unfortunately, conjugated molecules produced only limited gain in *SMN2* splicing modulation compared to unconjugated ASOs (Kwiatkowski et al., 2016).

Here we report the conjugation of Sudemycin drugs with ASOs and the evaluation of their effects to repress U2 snRNP assembly *in vitro*: a Sudemycin variant suitable for chemical conjugation was generated, coupled with an ASO complementary to the sequence 5' of the branch point (BP) region of Adenovirus Major Late promoter intron 1 and the activity of the conjugate tested in biochemical assays. The results indicate that this strategy can be used to enhance the splicing inhibitory

effects of ASOs, but that the distance range at which these conjugates can work is limited, constraining the applicability of this approach.

Results

The azide Sudemycin N retains splicing inhibitory activity

The replacement of two methyl groups associated to the ester side group of Sudemycin C1 by an azide ($-N_3^+$) led to the generation of a new Sudemycin variant, Sudemycin N, which is suitable for conjugation with ASOs. The activity of this compound was initially characterized by *in vitro* A3' complex formation, cytotoxicity and *MCL1* alternative splicing modulation assays in HeLa cells, as for previous compounds (Part I.A). The results showed that the compound retains activity in the three assays, although at concentrations over 100-fold higher than Sudemycin K (Figure I.B.1 and Table I.B.1). Nevertheless, we considered that the possibility of conjugation made the drug potentially useful for targeted splice site inhibition.

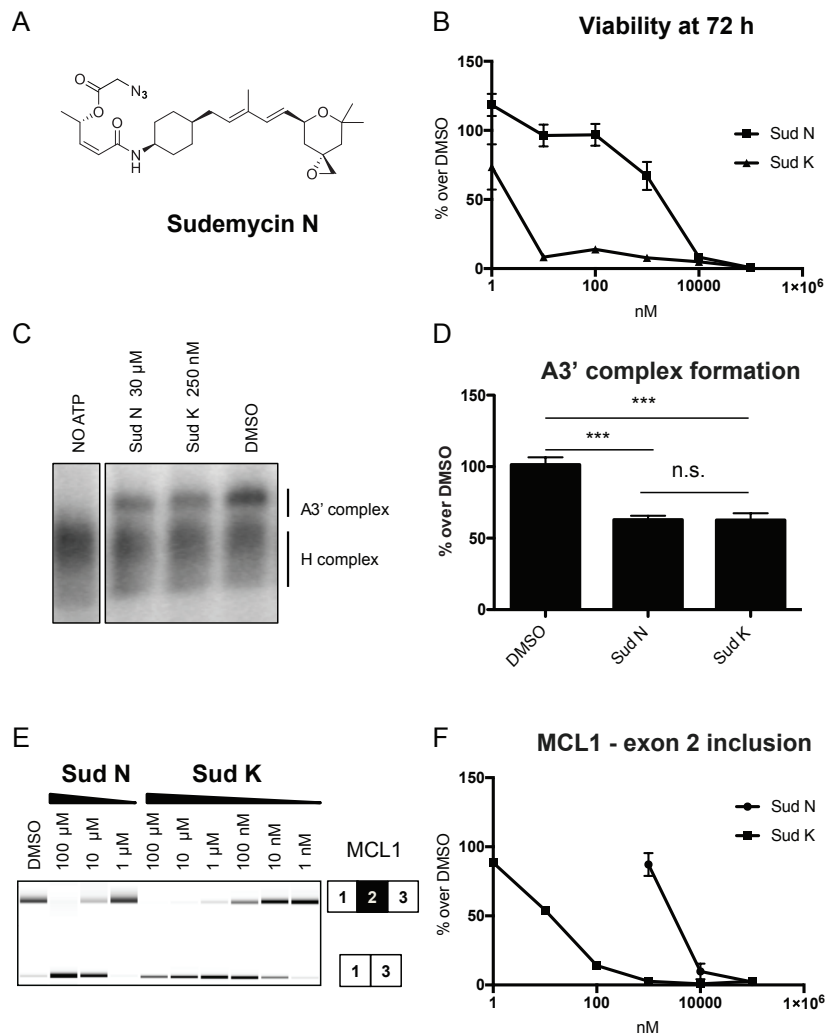


Figure I.B.1. Relative activity of novel Sudemycins N and K. (A) Chemical structure of Sudemycin N. (B) Cytotoxicity assays: cell viability was measured using Resazurin assays 72 h after drug exposure. Graphs indicate fraction of living cells compared to control DMSO treatment. All treatments were performed in triplicate and standard deviations are shown. (C) Representative Phosphorimager pictures of electrophoretic separation of H and A3' complexes assembled on Adenovirus Major Late promoter intron 1 3' splice site region in the presence of the indicated amounts of drug. All the reactions were treated with heparin. (D) Quantification of the results reported in C. The assays were repeated for a minimum of three times and statistically evaluated using a t-test (***: p-value < 0.0001; n.s.: p-value > 0.01). Different concentrations for the two drugs were adjusted to provide approximately 50% inhibition in complex formation. (E) Capillary electrophoresis profiles of RT-

PCR amplification of *MCL1* alternatively spliced products from RNA isolated 3 h after drug exposure. One representative example per condition is shown. (F) Quantification of data shown in E for duplicate experiments. Sudemycin N was synthesized by Kamil Makowski.

DRUG	<i>In vitro</i> A3' complex formation – IC ₅₀ (nM)	<i>MCL1</i> alternative splicing regulation – IC ₅₀ (nM)	Cytotoxicity in HeLa cells – IC ₅₀ (nM)
Sud K	≈250	≈15	2,3 ± 0,81
Sud N	≈30000	≈3700	1890 ± 574

Table I.B.1. Summary of activities of Sudemycins N and K. IC₅₀ values corresponding to *in vitro* inhibition of A3' complex formation, *MCL1* alternative splicing regulation and cytotoxicity in HeLa cells are indicated.

Conjugates of antisense oligonucleotides with Sudemycin N display increased inhibitory activity *in vitro*

Sudemycin N was conjugated or not to the 5' end of an ASO complementary to the region immediately 5' of Adenovirus Major Later promoter (AdML) branch point (BP), in order to bring the drug in close proximity of the BP sequence (Figure I.B.2). A control sequence not predicted to be complementary to any target was also synthesized and conjugated or not with Sudemycin N. All ASOs were 2'-O-methyl-phosphorothioate modified to improve their stability and RNA binding (Kole et al., 2012).

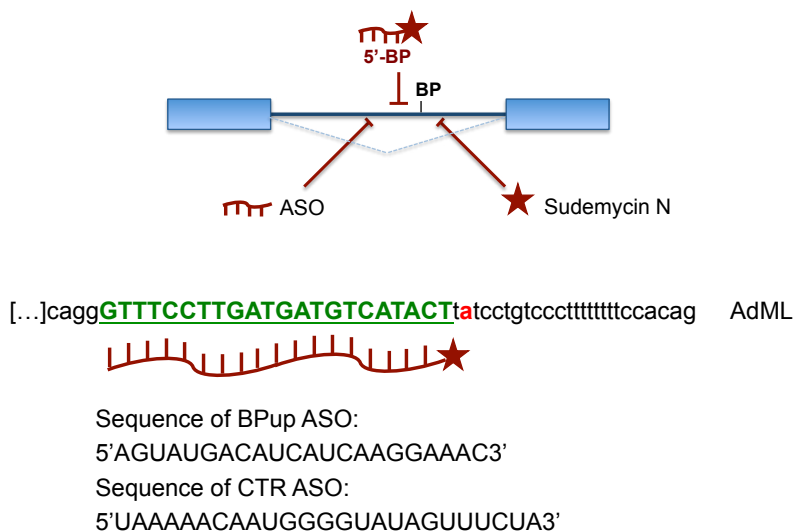


Figure I.B.2. Scheme of the conjugation strategy and the targeted region. The aim of the ASO-drug conjugation was to take advantage of the specificity conferred by base pairing interactions afforded by the ASO with the splicing inhibitory effects of the drug, thus obtaining a synergistic effect (upper panel). Lower panel: sequence of the 3' end of AdML intron 1 transcript and ASOs (BPup: specifically targeting the sequence 5' of the BP -in green-, CTR: control oligonucleotide). The BP adenosine is indicated in red.

In vitro A3' complex formation assays indicated that the inhibitory activity of the ASO complementary to the region 5' of the BP can be enhanced by conjugation with Sudemycin N (Figures I.B.3A and I.B.3B, compare lanes 6-8 with 12-14, corresponding to ASO alone and ASO-Sudemycin N conjugate respectively). The control ASO, conjugated or not with the drug, did not display any effect in these assays (Figures I.B.3A and I.B.3B, lanes 3-5 and 9-11). While the conjugated oligonucleotides displayed effects at 10-100 fold excess over the target RNAs (i.e. at 1 or 10 nM concentration), the non-conjugated Sudemycin N displayed inhibitory effects only at much higher (100 μ M) drug concentrations (Figures I.B.3A and I.B.3B, compare lanes 15 and 16-17).

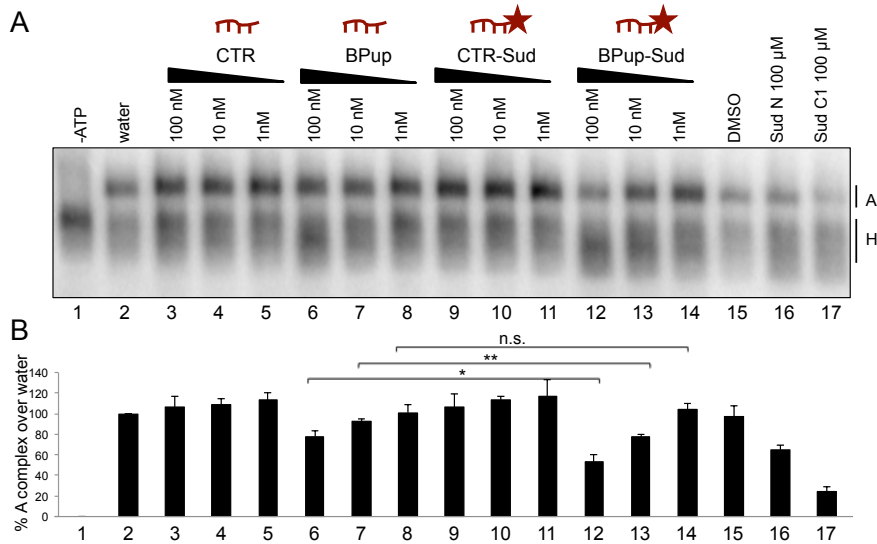


Figure I.B.3. *In vitro* effects of ASO-Sudemycin N conjugates. (A) Phosphorimager picture of electrophoretic separation of H and A complexes assembled on AdML full length construct in the presence of non-conjugated ASOs (BPup and CTR) or conjugated ASOs (BPup-Sud and CTR-Sud), or non-conjugated drugs Sud N and Sud C1 (used as internal control). ASOs were added at 100, 10 or 1 molar ratio relative to the target RNA, while drug treatments were 100 μ M (i.e. 100'000 molar ratio with the RNA). A control without ATP (-ATP) was included to document the migration of H complex and the lack of formation of A complex. All reactions were treated with heparin. (B) Quantification of the results for two independent experiments with full length or 3' intron 1 and exon 2 of AdML. t-test: (* and **: p-value < 0.01 and 0.001; n.s.: p-value > 0.01). ASOs and conjugates were synthesized by Clément Paris.

At a concentration that is highly effective for conjugates (100 nM), the drug alone does not induce detectable inhibition of A3' complex formation (Figure I.B.4A). As a confirmation that the synergistic effect is due to the conjugation and not to the co-presence of both ASO and drug, we cotreated the assembly reactions with the ASO (CTR or BPup) and DMSO or Sud N. Only BPup ASO is associated to the reduction of

A3' complex formation, without a detectable difference upon addition of DMSO or Sud N (Figure I.B.4B), confirming the hypothesis. Targeting specificity was studied by replacing the natural sequence complementary to the BPup ASO by the sequence complementary to the CTR ASO: under these conditions, inhibition was detected with CTR but not BPup ASO (Figure I.B.4C, lanes 3 and 5 versus lanes 4 and 6). Nevertheless, the conjugate did not display increased activity compared to the free ASO (Figure I.B.4C, lane 3 versus lane 5). In order to preserve the BP sequence intact, the distance between the 5' end of the ASO and the BP was increased by 4 nucleotides in this mutant (Figure I.B.4D: the mutation brings the ASO to a 6 nt distance from the BP in the CTR -6 mutant, while in wt AdML the ASO BPup's target is 2 nt apart from the BP). We checked the consequences of increasing the distance also between BPup complementary sequence and the BP (Figure I.B.4E, BPup -6 mutant) and we observed once again comparable activities between free ASOs and conjugates (Figure I.B.4D, lanes 10 and 12), meaning that the distance increase (and not the sequence difference) reduced the inhibitory activity of the conjugate. Hence, proper positioning of the drug relative to the BP is critical for the efficiency of this strategy.

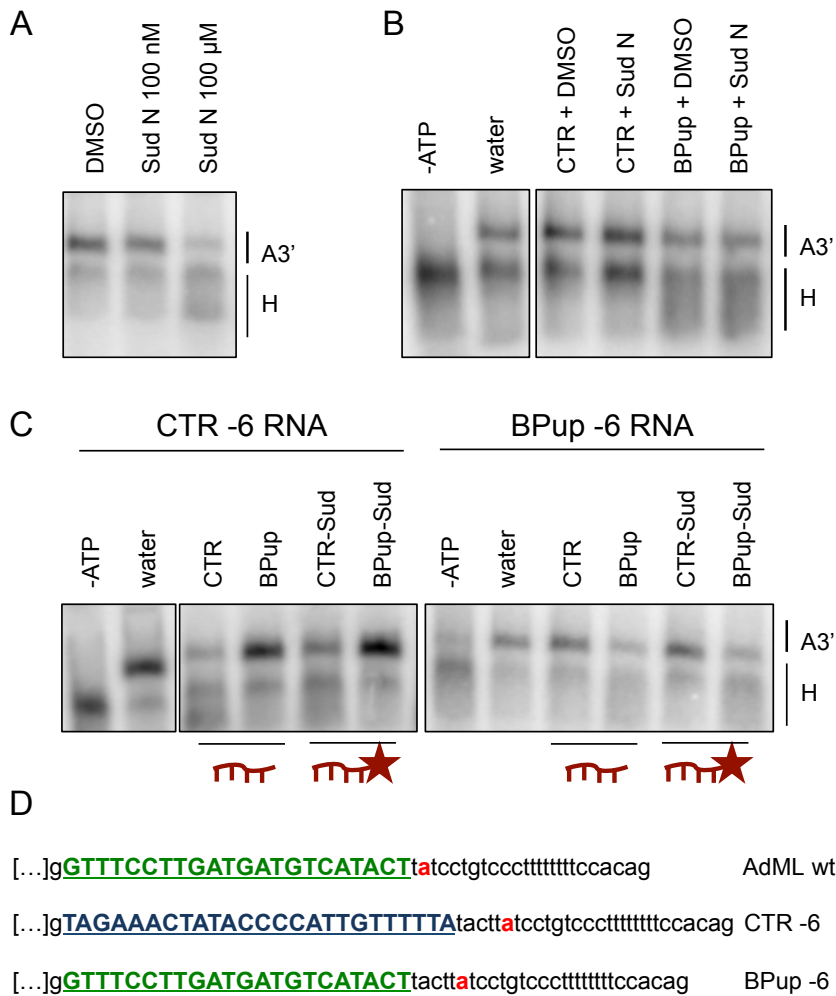


Figure I.B.4. *In vitro* effects of ASO-drug cotreatment and of ASO-BP increased distance. (A) Phosphorimager picture of A3' complexes assembled on AdML 3' construct in the presence of DMSO, 100 μ M or 100 nM Sud N. Effects are not detectable at 100 nM, which is the highest concentration tested for the conjugates in Figure 3. (B) A3' complexes assembled on AdML 3' construct upon cotreatment with 100 nM ASOs (CTR or BPup) and 100 nM Sud N (or same volume of DMSO control). Heparin was added to all the reactions. (C) A3' complexes assembled on CTR -6 or BPup -6 mutants upon addition of 100 nM ASOs or conjugates. Controls without ATP (-ATP) or with water were also performed. All the reactions were treated with heparin. (D) Detailed sequence of intronic 3' end of wt AdML or CTR -6 and BPup -6 mutants used for Figure 4C. The regions targeted by BPup and CTR oligos are shown in green and blue, respectively. The BP position is highlighted in red.

Inhibition of AdML intron 1 was also tested in cells transfected with a minigene containing AdML pre-mRNA under a CMV promoter. However, no effects on splicing of these transcripts were detectable, suggesting either poor cell penetration or limited stability of these reagents in viable cells.

While these issues may limit the effectiveness of drug-ASO conjugates, our results illustrate the principle that conjugation can enhance the splicing inhibitory activity of ASOs if properly positioned relative to key splicing signals. More work will be required to test the general applicability of the approach and improve its efficiency.

Appendix – Drug-ASO conjugation strategy

RNAs were synthesized in an automatic oligonucleotide synthesizer using phosphoramidite chemistry. The synthesized RNAs were fully modified with 2'-OMe groups and phosphorothioate internucleotide linkages. The RNAs were also derivatized with an alkyne functional group at their 5' terminal position to allow conjugation with the azide-derivatized Sudemycin through click Huisgen cycloaddition. The crudes of the conjugation reactions were immediately purified by reverse-phase HPLC, aliquoted and directly freeze-dried to prevent their degradation.

Synthetic procedures were optimized and performed by Clément Paris (University of Barcelona).

Results

Part II. Identification of sequence elements determining alternative splicing sensitivity to pharmacological inhibition of SF3B1

Key role for *MCL1* alternative splicing on cell viability

MCL1 (Myeloid Cell Leukemia 1) is a protein with anti-apoptotic function, often overexpressed in tumors and considered a good target for cancer therapy (Glaser et al., 2012; Opferman, 2016; Tiedemann et al., 2012; Wei et al., 2012). Due to its rapid turnover both at the protein and RNA levels, *MCL1* levels are highly affected by transcription and translation inhibitors, with toxic effects for tumor cells depending on Bclx levels and activity (Wei et al., 2012). *MCL1* exon 2 skipping generates a pro-apoptotic isoform and therefore promoting this alternative splicing event could be a valid strategy to enhance tumor cell death (Bae et al., 2000).

Interestingly, individual depletion of several splicing factors, including SF3B1, induces *MCL1* exon 2 skipping (Laetsch et al., 2014; Moore et al., 2010; Papasaikas et al., 2015). A similar regulation towards the pro-apoptotic isoform was also reported for SF3B-targeting drugs: indeed, it has been proposed that Spliceostatin A induces apoptosis in chronic lymphocytic leukemia (CLL) cells mainly through *MCL1* downregulation (Larrayoz et al., 2016), and cell lines resistant to the Bcl-XL inhibitor ABT-737 reacquire sensitivity when treated with

Meayamycin, due to *MCL1* splicing regulation and reduction of anti-apoptotic MCL1 levels (Gao and Koide, 2013).

Among a panel of alternative splicing events involved in proliferation and apoptosis control, several splicing inhibitors most prominently affect *MCL1* (Papasaikas et al., 2015) (Figure II.1A). While Meayamycin, Spliceostain A and Sudemycin C1 share SF3B1 as their cellular target (Bonnal et al., 2012), Isoginkgetin's a splicing inhibitor with antitumor effects thought to target a tri-snRNP component and inhibit A to B complex transition (O'Brien et al., 2008), also prominently affects *MCL1* splicing among the selected events (Figure II.1A). In contrast, other alternative splicing regulators, like TG003 (CLK kinases inhibitor) (Muraki et al., 2004) and Cyclosporin (cyclophillins inhibitor) (Horowitz et al., 2002) cause a different and more moderate spectrum of splicing changes (Figure II.1A).

Previous biochemical assays with Adenovirus Major Late promoter transcripts showed that Spliceostatin A (SSA) destabilizes A3' complex (i.e. A complex lacking U1 snRNP, since RNAs used for this assay lack the 5'ss). The destabilizing effects were detected only in the presence of heparin, which, due to the repetitive negative charge of the polymer, can mimic the phosphate backbone of RNAs and therefore displace weak complexes assembled on RNA (Corrionero et al, 2011).

In vitro A3' complex formation assays confirmed that while Sudemycin C1 inhibits A3' complex assembly on AdML intron 1 3' splice site region in the presence of heparin (as previously reported for Spliceostatin A), Isoginkgetin does not inhibit this step, even when high amounts of drug and heparin are used in the assays (Figure II.1B). This

result is concordant with its role at later stages of spliceosome assembly (O'Brien et al., 2008).

Consistent with the importance of *MCL1* levels for cell viability, and with the control of *MCL1* by alternative splicing, SF3B1 and *MCL1* knockdowns or treatment with Sudemycin C1 resulted in reduced cell viability, measured using Resazurin/Alamar Blue assays (O'Brien et al., 2000) (Figure II.2). Interestingly, even a pulse of as short as 3 h with Sudemycin C1 was sufficient to compromise cell's long-term viability (Figure II.2, compare upper and lower right panels). This could be explained indirectly by induction of irreversible changes in chromatin environment by the drug (Convertini et al., 2014) and other splicing regulators. It can also mean that, once internalized, Sudemycin C1 is persistent in the cells, although this is not concordant with reports relative to short cell stability of related drug variants (Convertini et al., 2014).

Collectively, these results highlight the sensitivity of *MCL1* alternative splicing to different types of splicing inhibition (Figure II.1) and the importance of this event, as well as SF3B1 activity, for normal proliferation of human cells (Figure II.2), but the reason of its strong regulation by splicing inhibition is not clear.

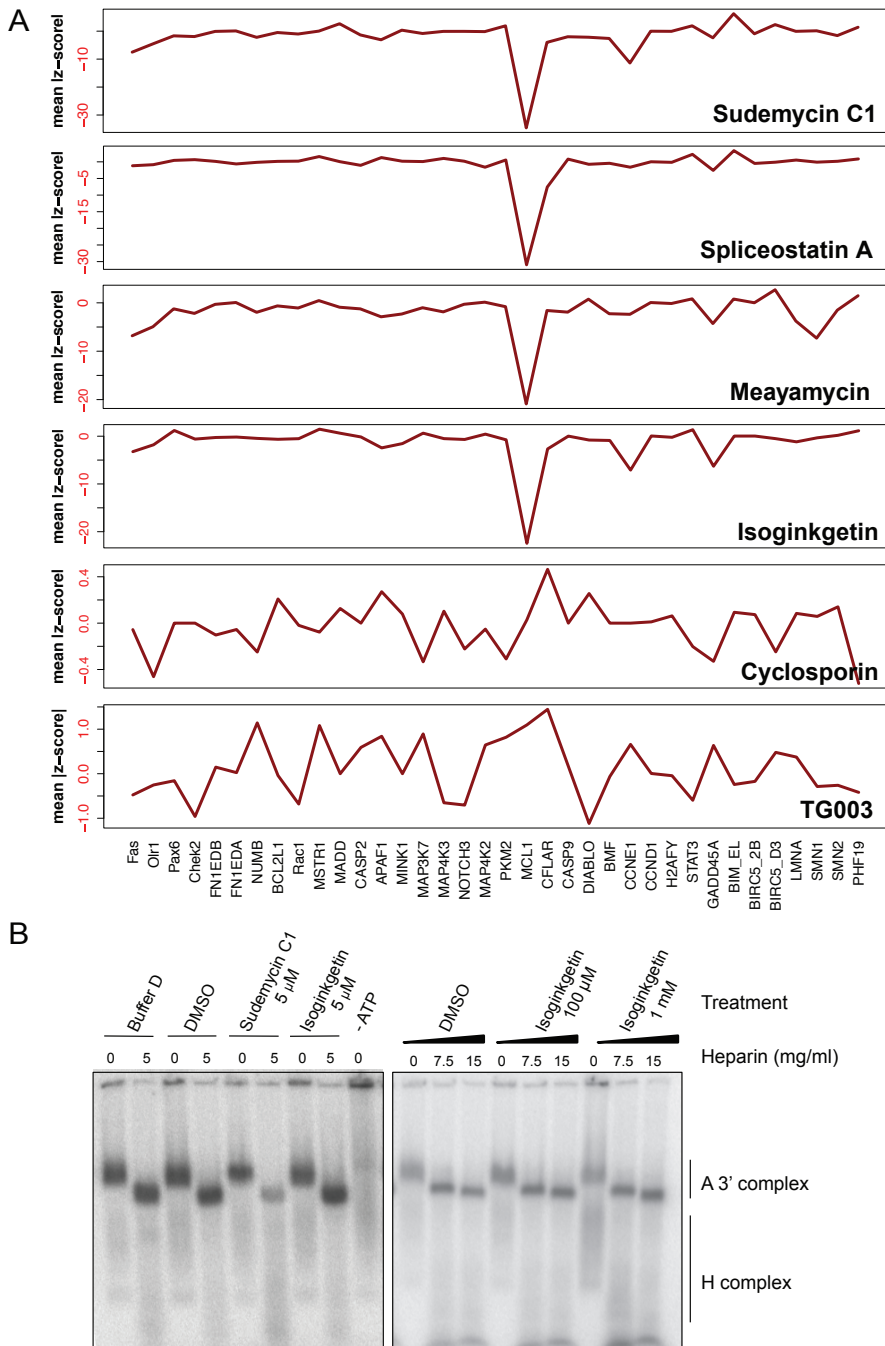


Figure II.1. Antitumor splicing inhibitors strongly affect *MCL1* alternative splicing. (A) Splicing perturbation profiles of different splicing inhibitors. Changes toward inclusion (>0) or skipping (<0) are represented, quantified as a robust Z score (Papasaikas et al., 2015), which is considered

significant if $\geq |2|$. Drug treatments were performed in HeLa cells (Sudemycin C1: 20 μM , 8 h; Spliceostatin A: 260 nM, 3 h; Meayamycin: 20 nM, 8 h; Isoginkgetin: 20 μM , 8 h; Cyclosporin A: 100 μM , 8 h; TG003: 10 μM , 8 h). (B) Distinct mechanisms of splicing inhibition by Sudemycin C1 and Isoginketin. *In vitro* A3' complex formation assay for AdML RNA with the indicated concentrations of Isoginketin or Sudemycin C1 and increasing concentrations of heparin. Meayamycin and Spliceostatin A were kindly provided by Drs Kazunori Koide (University of Pittsburgh) and Minoru Yoshida (RIKEN Institute), respectively.

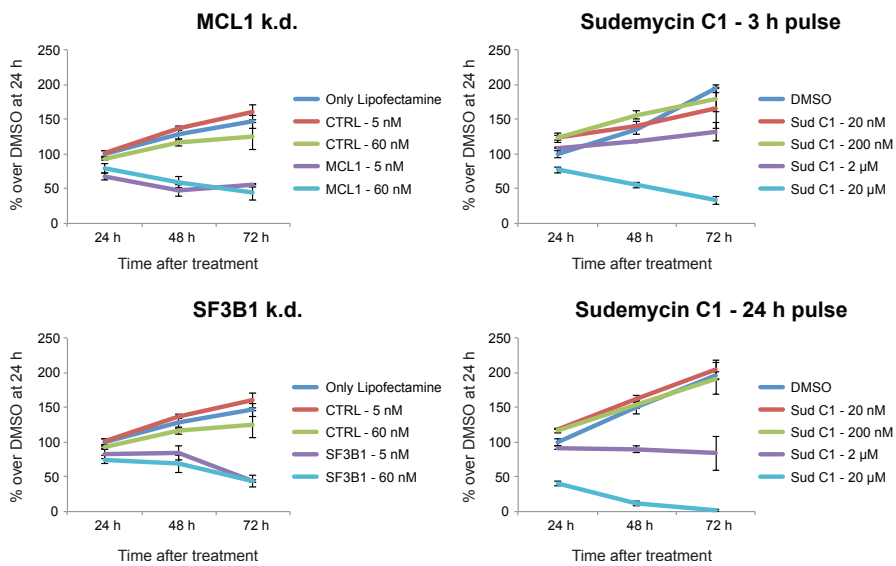


Figure II.2. MCL1 levels are important to maintain cell viability. Viability of HeLa cells was measured using Resazurin assays (O'Brien et al., 2000) upon MCL1 or SF3B1 knock down by RNAi, using the indicated concentrations of siRNAs (left panels) or upon treatment with the indicated concentrations of Scudemycin C1, 24, 48 and 72 hours after the different treatments. CTRL: scrambled siRNA.

Sequence elements 5' of the Branch Point repress drug inhibition

SF3B1 inhibitors Spliceostatin A and E7107 were described to alter the balance between alternative U2 snRNA conformations and/or to induce interactions of U2 snRNA with non-functional “decoy” sequences located 5' of the Branch Point (BP) (Corrionero et al., 2011; Folco et al., 2011).

To further investigate the hypothesis that alternative splicing regulation induced by SF3B1-targeting drugs is influenced by the sequence environment around the BP sequence, we selected two model alternative splicing events: *MCL1* exon 2 skipping, based on its strong regulation by splicing inhibitors and its important regulatory function for cell viability (Figures II.1 and II.2), and *PDCD10* exon 7 skipping, which is only weakly affected by the drug (Corrionero et al., 2011) (Figure II.3A).

To facilitate the study of sequences that modulate drug sensitivity, we used minigene assays in which alternatively spliced genomic sequences are expressed under the control of a Cytomegalovirus (CMV) promoter. Minigene-specific patterns of splicing are distinguished by RT-PCR using oligonucleotides complementary to vector-specific sequences present in the primary transcripts. RNAs expressed from these minigenes recapitulate the different sensitivity to the drug of the endogenous *MCL1* and *PDCD10* transcripts (Figures II.3C and 4A).

Recapitulation of *PDCD10*'s low drug sensitivity (Figure II.3C) occurs despite the fact that, in endogenous transcripts, exon 7 skipping causes a frameshift that could lead to RNA degradation by Nonsense-Mediated

Decay, while NMD cannot play a role for minigene-derived transcripts because of the absence of a natural ORF in these RNAs.

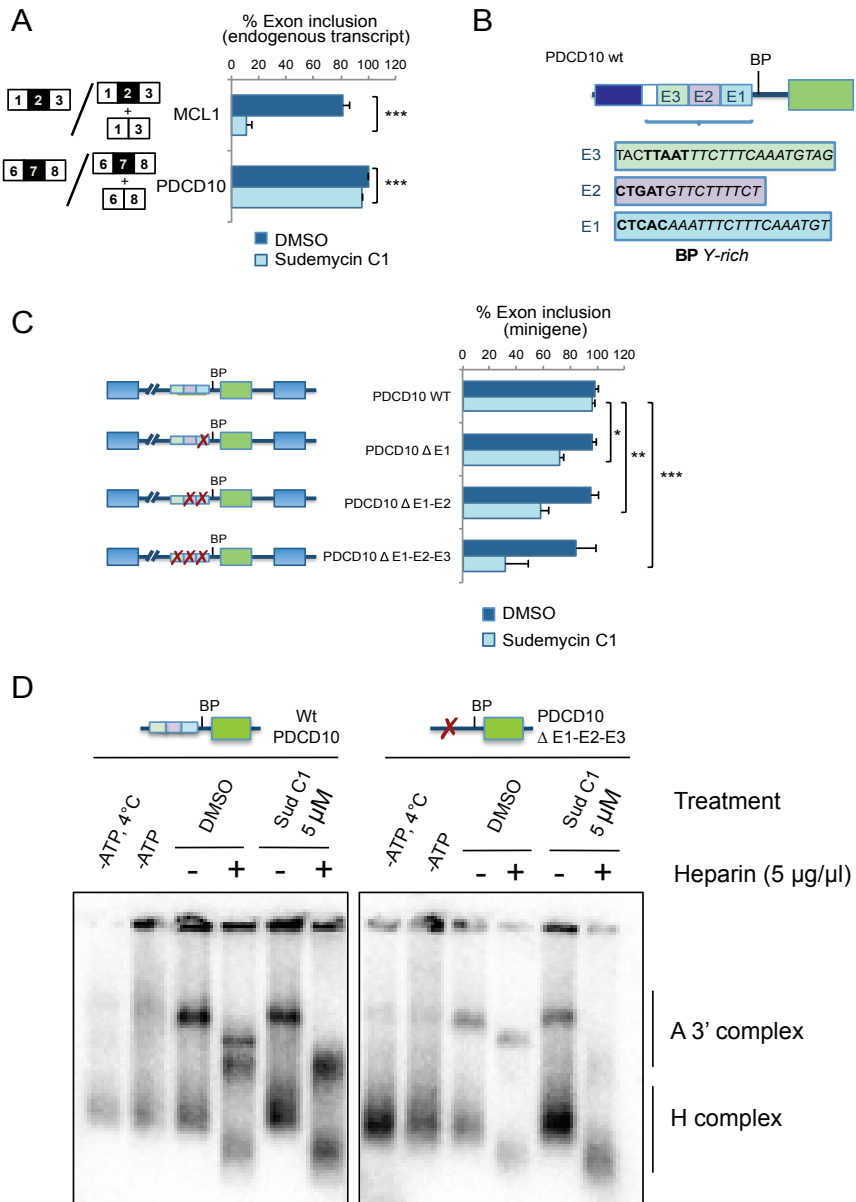
To explore the role of sequences 5' of the BP in Sudemycin C1 effects, we first generated three mutants that progressively delete sequences upstream of *PDCD10* putative BP (Figure II.3B). Each of these deletions increased the transcripts' sensitivity to Sudemycin C1, detected as a progressive increment in exon skipping in the presence of the drug (Figure II.3C). These results confirm that sequences 5' of the BP can strongly influence drug responses and display additive effects.

We performed A3' complex formation assays by incubating radioactive wt and mutant *PDCD10* sequences with HeLa nuclear extracts, in the absence or presence of Sudemycin C1 and heparin. Upon heparin addition, two complexes formed on the wt transcript (Figure II.3D, compare lanes 3 and 4). However, upon drug treatment only the lower band is formed, suggesting that alternative complexes can be formed on the transcript and only one of them (the upper one) is sensitive to the drug (Figure II.3D, compare lanes 4 and 6). On the other hand, Sudemycin C1 induces a clear inhibition of the only complex formed on *PDCD10* transcripts lacking E sequence elements (*PDCD10* Δ E1-E2-E3) in the presence of heparin (Figure II.3D, compare lanes 10 and 12). This result is compatible with the concept that transcripts lacking E elements are more sensitive to Sudemycin C1, as documented in Figure II.3C. The results also suggest that the reduced drug sensitivity conferred by the E sequence elements is largely due to one of the two distinct complexes that can form on these RNAs.

Examining the sequence of these regions, we noticed that each of them includes a sequence motif that resembles the consensus BP “YUNAY”

(Gao et al., 2008) followed by a stretch of nucleotides rich in pyrimidines (Figure II.3B). These elements are similar to the configuration of 3' splice sites, except that they are not followed by the AG dinucleotide that marks the 3' end of introns. It was therefore possible that the additional BP consensus and/or the pyrimidine-rich sequences, repress the effects of the drug. To address these issues, we next tested whether the effects of these elements could be transferred to other transcripts.

Figure II.3 (next page). Mapping of sequence elements 5' of the BP that repress Sudemycin C1 effects on *PDCD10* alternative splicing. (A) Regulation of endogenous *MCL1* exon 3 skipping and *PDCD10* exon 7 skipping by Sudemycin C1 treatment (20 μ M for 8 h) in HeLa cells. *PDCD10*: Programmed Cell Death 10, associated to apoptosis regulation. (B) *PDCD10* sequence elements upstream of exon 7. Possible Branch Point (BP) consensus sequences are highlighted in bold. (C) *PDCD10* sequence elements E1, E2 and E3 repress the drug effects in an additive manner (HeLa cells treated with 20 μ M Sudemycin C1 for 8 h, 16 h after transfection). *, **, ***: t-test p-value < 0.01, 0.001 and 0.0001, respectively. A minimum of three replicates was used for each condition. “//” in the scheme indicates the deletion of part of *PDCD10* intron 6 (since the intron is 7902 nt long, only the first 290 and the last 250 were inserted in the minigene, for ease of cloning and efficiency of expression). The short bar after the third exon indicates that additional 20 nt of the downstream intron were included in the minigene. (D) A3' complex assembly assay using radioactive RNAs encompassing the 3' end of *PDCD10* intron 6, the whole exon 7 and the first 25 nt of the downstream intron. RNAs were incubated under standard splicing conditions with HeLa nuclear extracts and reactions were separated on a native gel to detect A3' and H complexes. Sudemycin C1 was added at 5 μ M concentration or an equivalent volume of DMSO was added as control. Sudemycin C1 used for these and the following experiments was kindly provided by Dr. Thomas Webb (St. Jude's Children Hospital).



Repressive elements are transferable and resemble BP sequences

In order to test the relevance of *PDCD10* repressive E1-E2-E3 sequence elements outside of their endogenous context, we inserted them 5' of the BP of intron 1 of the highly responsive *MCL1* minigene, replacing a sequence of similar length located at an equivalent position in the RNA. We first verified that deletion of this *MCL1* sequence did not significantly alter the response to the drug (Figure II.4A, compare constructs *MCL1* wt vs *MCL1* Δ , top two rows) and that insertion in this deletion mutant of another sequence further upstream from *PDCD10* intron 6 did not affect it either (Figure II.4A, compare constructs *MCL1* vs *MCL1* Δ + upstream *PDCD10*, 2nd and 3th row). In sharp contrast, insertion of the E1-E2-E3 element resulted in strong inhibition of the *MCL1* exon 2 skipping effects of Sudemycin C1 (Figure II.4A, compare *MCL1* vs *MCL1* Δ + E3-E2-E1, 1st and 4th row), confirming that these sequence elements are necessary and can be sufficient to inhibit the effects of the drug on alternative splicing.

To further dissect the contribution of sequence elements within “E” regions, we separately tested the potential effects of BP consensus and Pyrimidine-rich sequences discussed above. Remarkably, insertion of a BP consensus (CTCTCAC) led to inhibition of the effects of the drug almost as strong as the insertion of the E3-E2-E1 elements (Figure II.4A, compare *MCL1* vs *MCL1* Δ + BP vs *MCL1* Δ + E3-E2-E1, 1st, 4th and 5th rows). Importantly, deletion of the adenosine in this sequence element, which can potentially work as the BP residue for 2'-5' phosphodiester bond formation after the first catalytic step, completely abrogated this effect (Figure II.4A, compare *MCL1* Δ + BP vs *MCL1* Δ + decoy, 5th and 6th rows). This result argues that a fully functional BP is

required to prevent the effect of the drug, because a non-functional decoy BP sequence (able to base-pair with U2 snRNA but inactive for splicing) does not prevent (and possibly facilitates) the effects of the drug.

Interestingly, insertion of the sequence 5'-TTTCTTTCAAATG also resulted in reduced drug effects (Figure II.4A, compare *MCL1* Δ with *MCL1* Δ + Y rich seq, 2nd and 7th rows). The combination of the pyrimidine-rich element and the BP consensus (even in a non-natural order) resulted in full inhibition of the effects of the drug (Figure II.4A, *MCL1* Δ + Y rich seq + BP, 8th row), arguing that these elements can work in a cooperative fashion to prevent U2 snRNP inhibition by Sudemycin C1.

To further dissect the contribution of the pyrimidine-rich element, we noticed that E3 and E1 contain an identical stretch of pyrimidines (TTTCTTTC) followed by an identical AAATG sequence (while E2 contains only a related pyrimidine tract, TTCTTTTC). While insertion of a TTTCTTTCT sequence did not inhibit the effect of the drug, insertion of a AAATGT sequence did (Figure II.4B, compare *MCL1* Δ + TTTCTTTCT vs *MCL1* Δ + AAATGT, 2nd and 3th rows). Conversely, deletion of this sequence in the *PDCD10* minigene leads to a significant gain in drug sensitivity (compare *PDCD10* wt in Figure 1C with *PDCD10* Δ CAATGTAG in Figure II.4B), similar to the deletion of the complete E1 element (Figure II.3), confirming that this sequence does play an important role in limiting drug responses.

The AAATGT sequence potentially acts as a back-up BP and thus contributes to drug resistance. Therefore, we speculate that the role of

this sequence is to offer alternative BP sites for efficient splicing of the intron even in presence of SF3B-targeting drugs.

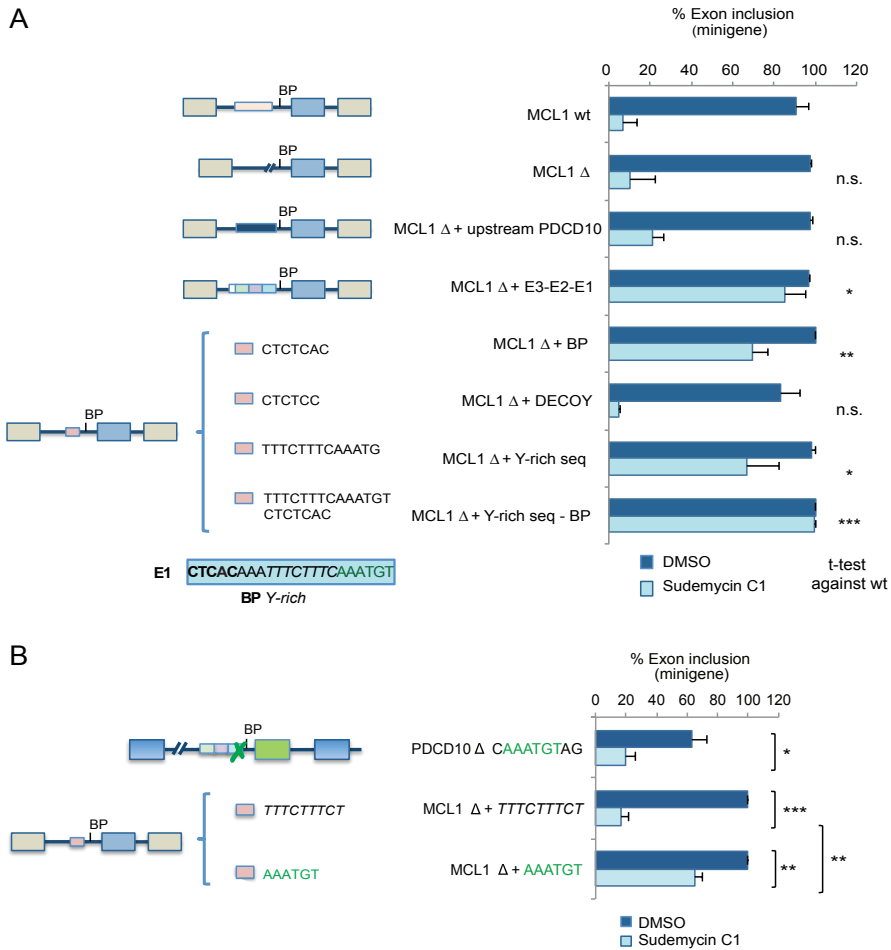


Figure II.4. Drug effects can be attenuated by sequences 5' of the BP identified in drug-resistant 3' splice sites. (A) Mapping of sequence elements from *PDCD10* that inhibit the effects of Sudemycin C1 on *MCL1* alternative splicing. The structure of the transcripts derived from the *MCL1* minigene variants is represented on the left, while the % of exon inclusion of minigene transcripts expressed in HeLa cells in presence of 20 μ M Sudemycin C1 or DMSO control (8 h of treatment, 16 h after minigene transfection) is quantified on the right for a minimum of three independent biological replicas. Statistical significance refers to results of t-test comparison between Sudemycin C1-treated mutated minigenes vs Sudemycin C1-treated wt minigenes *, **, ***, n.s.: p-value < 0.01, 0.001, 0.0001 or > 0.01, respectively.

MCL1 minigenes carry the complete sequence of the two introns and the alternative exon, but only part of exon 1 (i.e. the last 114 bp out of the total 896 bp) and exon 3 (i.e. the first 97 bp out of 2941 bp), for ease of cloning and efficient expression. (B) Minigenes experiments where the second part of E1 element was deleted (PDCD10 Δ CAATGTAG, with the putative regulatory element highlighted in green). *MCL1* minigenes where the sequence upstream of the BP was replaced either with the Y-rich sequence from E1 (highlighted in italics) or with the A-rich sequence element (highlighted in green) (*MCL1* Δ + TTTCTTTCT and *MCL1* Δ + AAATGT). *, **, ** and n.s.: t-test comparison between treated and non-treated minigenes or between mutant minigenes vs wt minigenes with the same treatment with p-value < 0.01, 0.001, 0.0001 or > 0.01, respectively; n = 3. HeLa cells treated with 20 μ M Sudemycin C1 for 8 h, 16 h after transfection.

BP mapping reveals degenerate BP usage and weak BP sequences as the basis of *MCL1* drug sensitivity

Given that SF3B1 (i.e. Sudemycin's target) binds to the pre-mRNA in the proximity of the BP (Corrionero et al., 2011; Gozani et al., 1996) and that sequences upstream of the BP (which resemble BPs) affect the response to Sudemycin C1 (Figures II.3 and II.4), it became important to unambiguously determine the sequences in *PDCD10* and *MCL1* that serve as BPs. With this aim we undertook two approaches: i) amplification of lariat RNAs derived from splicing of the endogenous transcripts in cells, and ii) analysis of the effects of mutation of the predicted positions in minigenes.

For the first approach we isolated total RNA from HeLa cells and carried out RT-PCR assays using oligonucleotides annealing 3' of the 5' splice site and 5' of the putative BP region, which exploit the capacity of reverse-transcriptase to retro-transcribe through a 2'-5' phosphodiester bond to generate PCR amplification products specific of lariat RNAs that join together the 5' splice site with the BP region (Figure II.5A)

(Vogel et al., 1997). The position of the BP can therefore be determined by sequencing of the amplification product (Figure II.5A). Because of imprecision in reverse transcription through the 2'-5' phosphodiester bond, ambiguity of 1-2 nucleotides and mis-incorporation of nucleotides in the location of the BP (usually T for the branch A) is common using this technique (Conklin et al., 2005; Vogel et al., 1997). Using this method, we could confirm Adenosine -25 as the most likely BP in *PDCD10* intron 6 in K562 cells (a cell line known to accumulate lariat RNAs), as predicted from its close resemblance to the BP consensus (http://regulatorygenomics.upf.edu/Software/SVM_BP/) (Corvelo et al., 2010) and its location relative to the polypyrimidine tract / 3' splice site AG (Figure 5B), while Adenosine -23 seemed to be used in HeLa cells, despite the lower base-pairing potential of its surrounding sequence with U2 snRNA (Figure II.3B). These PCR products could even correspond to a BP in -25 position, with a short insertion due to nucleotides mis-incorporation. The BP of this intron had not been reported previously.

In contrast, several potential BPs in the region -19 to -33 of *MCL1* intron 1 were identified in HeLa cells using this technology (Figure 5C), possibly including non-adenosine residues, as previously proposed by the results of a genome-wide BP mapping approach (Mercer et al., 2015) and annotated in circbase (a database of circular RNAs, some of which are intronic lariats) (Jeck et al., 2013).

Importantly, no difference in BP utilization was observed in the absence or presence of Sudemycin C1 treatment for either *PDCD10* or *MCL1*.

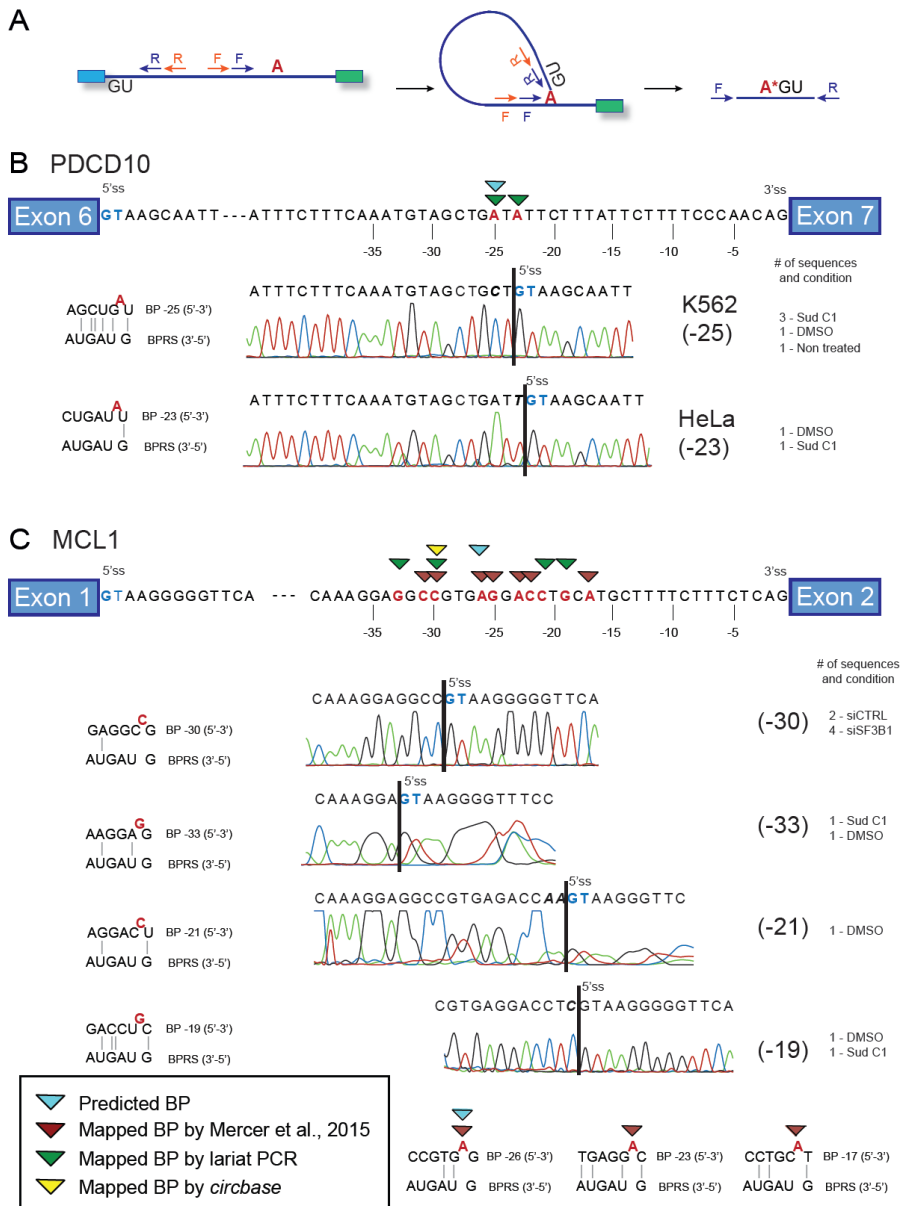


Figure II.5. BP mapping of *PDCD10* and *MCL1* transcripts by across the lariat RT-PCR. (A) Scheme of the position of oligonucleotides for RT-PCR assays and expected amplification products (Vogel et al., 1997): external primers are represented as blue arrow, nested primers as orange arrows. The branch A is often mutated in this assay (as indicated by an asterisk in the scheme). (B) BP mapping of *PDCD10* intron 6 lariats using two different cell lines. Representative electropherograms of different clones of the amplification products are shown. 5' terminal GT sequences corresponding to the intron 5' end are shown in blue. Vertical bars indicate split between

sequences at the 3' end of the intron and the 5' splice site; the split should be diagnostic of the position of the BP, with 1-2 nucleotide resolution (Gao et al., 2008; Mercer et al., 2015). The cell line, the numerical positions of the putative BPs (nearest adenosine to the split position), the number of sequenced products and the treatment condition are indicated on the right of each result. Schemes of base-pairing complementarity with U2 snRNA corresponding to the potential BPs are also indicated. siCTRL and siSF3B1: HeLa cells transfected with scrambled and SF3B1 siRNAs. (C) BP mapping of *MCL1* intron 1 lariats, carried out in HeLa cells as in (B). Various potential BPs were detected, suggesting degenerate BP usage, consistent with a previous report (Mercer et al., 2015). The most probable computationally inferred BPs within the last 60 nt of the intron are shown (Corvelo et al., 2010): *PDCD10* BP scoring value svm_scr is 1.34, *MCL1* BP scoring value svm_scr is -0.34, highlighting the weakness of *MCL1* intron 1 BPs. At the bottom of the figure, base-pairing with U2 snRNA are shown for positions previously mapped (Mercer et al., 2015) and predicted by SVM_BP (Corvelo et al., 2010). A mapped BP from a circular RNAs database (<http://www.circbase.org/>) (Jeck et al., 2013) is also shown for *MCL1* (hsa_circ_0002364 annotated circular RNA).

Next we carried out mutational analysis of the predicted BP sequences. Adenosine at position -26 of *MCL1* intron 1 was found as one of the possible BPs (Mercer et al., 2015), but mutation to cytosine affected exon inclusion only slightly, consistent with the idea that multiple BPs can function at this 3' splice site region (Figure II.5C); the mutation did not affect drug sensitivity (Figure II.6, compare *MCL1* wt vs *MCL1* -26 mut). In contrast, replacement of the nucleotides around adenosine -26 by a consensus BP sequence (TACTAAC, known to be the optimal BP functional sequence from yeast to mammals) (Zhuang et al., 1989) resulted in efficient exon 2 inclusion and, importantly, complete resistance to the effects of the drug (Figure II.6, compare *MCL1* wt vs *MCL1* -26 TACTAAC, where the wt CCGTGAG sequence was replaced with TACTAAC consensus -BP A is underscored-). These results argue that the strength of the BP sequence is a determinant of drug response and, consequently, that the presence of multiple, rather

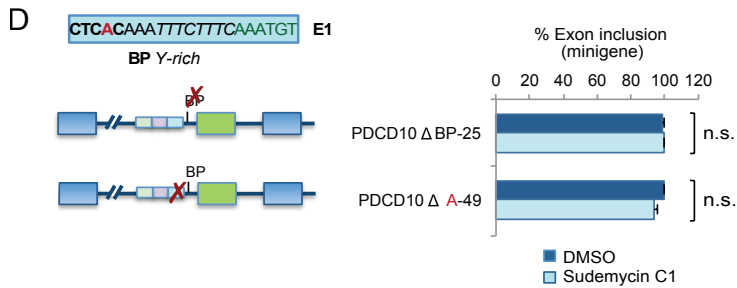
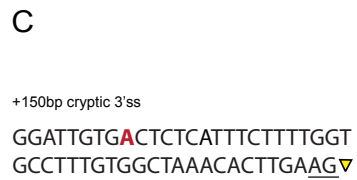
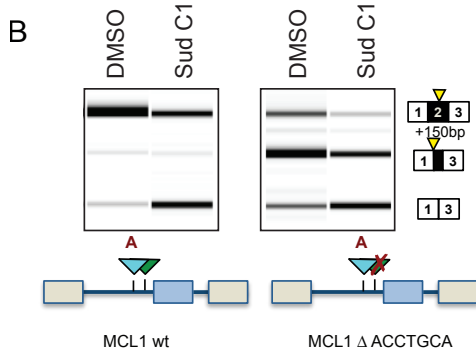
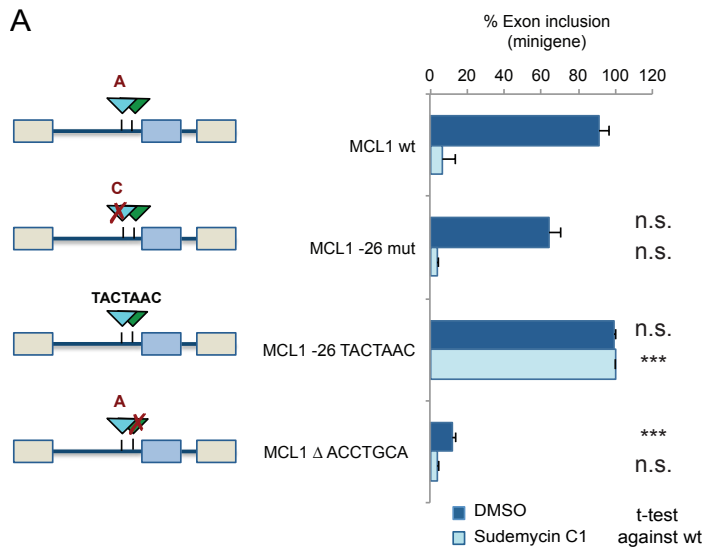
weak BPs sequences in *MCL1* intron 1 confers sensitivity to Sudemycin C1 treatment.

Several potential BPs were mapped by our RT-PCR approach and by previous work (Mercer et al., 2015) to a ACCTGCA sequence element located between positions -23 and -17 from the 3'ss (Figure II.5C). Interestingly, deletion of this entire sequence element significantly reduced the levels of exon inclusion even in control conditions (Figure II.6A), concomitantly with activation of a cryptic 3' splice site at position +150 within exon 2 (Figure II.6B). This result argues that these sequences play an important role in the efficient splicing of *MCL1* intron 1, possibly acting as the main BPs for the formation of lariats during its excision. Sudemycin C1 treatment resulted in further reduction of exon 2 inclusion, arguing that the drug efficiently inhibits both the mutated 3' splice site (using alternative BPs) as well as the cryptic 3' splice site activated by the mutation (Figure II.6B).

MCL1 resistance to drug-induced skipping was mainly induced by an optimal BP consensus sequence (TACTAAC, Figure II.6A), but also by a BP consensus from *PDCD10* E1 element (CTCTCAC, Figure II.4A). Deletion of the possible branch Adenosine in this element rescued the effects (CTCTCC, Figure II.4A), consistently with a role of this Adenosine in splicing catalysis. Moreover, the repression of the effect was even stronger if this last consensus was preceded by the second half of E1 sequence, switching the two sequence elements (TTTCTTTCAAATGT, Figure II.4A). This result indicates that the latter might display a synergistic repressive role together with the BP consensus for the drug effects. To investigate more in details this effect and validate the usage of upstream functional BP sequences in presence

of the drug, we focused first on *PDCD10* transcript again. Surprisingly, the deletion of the mapped BP in position -25 from the 3'ss (which would affect also the sequence for a BP in position -23) did not affect pre-mRNA splicing, as observed by almost complete exon inclusion, independently from drug-treatment (Δ BP-25, Figure II.6D). The same was observed by deleting the putative branch adenosine in the first half of E1 element (Δ BP-49, with the adenosine highlighted in red in E1 sequence), consistent with a more prominent role for the downstream A-rich sequence AAATGT (Figure II.6D, Figure II.4B). Therefore, both *MCL1* and *PDCD10* cases reveal flexibility for BP selection.

Figure II.6 (next page). Mutation of *MCL1* and *PDCD10* BP sequences reveals flexibility in BP selection. (A) Percentage of exon inclusion of the indicated *MCL1* minigenes in the absence or presence of Sudemycin C1, as in previous Figures. *MCL1* -26 mut indicates A at position -26 mutated to C; *MCL1* -26 TACTAAC indicates substitution of residues around A -26 by the BP consensus sequence; *MCL1* Δ ACCTGCA indicates deletion of residues -23/-17. *** and n.s.: t-test comparison of mutant minigenes vs wt minigenes with the same treatment (DMSO or Sud C1) with p-value < 0.01, 0.001, 0.0001 or > 0.01, respectively; n = 3. (HeLa cells treated with 20 μ M Sudemycin C1 for 8 h, 16 h after transfection). (B) Representative capillary electrophoresis profile of *MCL1* transcript isoforms from a minigene lacking the -23/-17 ACCTGCA sequence. Activation of an exonic cryptic 3'ss (at position +150 from the conventional one) was observed and confirmed by Sanger sequencing (data not shown). (C) 3'ss sequence of the cryptic 3'ss reported in B), with the AG dinucleotide highlighted and the computationally predicted BP in red (BP scoring value svm_scr: 1.54) (Corvelo et al., 2010), suggesting that the BP is stronger than the ones of the conventional 3'ss, which scores below 0 (Figure II.5). (D) Minigenes experiments where *PDCD10* BP A in position -25 or the A within E1 element were deleted (*PDCD10* Δ BP-25 and Δ -49, with the deleted A highlighted in red in E1 sequence).



Comparison between mouse and human *MCL1* regulation reveals a role for polypyrimidine tracts in modulating drug response

A relevant insight came from the realization that *MCL1* exon 2 is not known to be alternatively spliced in mouse cells. Although mouse *MCL1* transcripts use alternative splice sites within exon 1 (Kojima et al., 2010), to date, no exon 2 skipping was annotated in Genome Browsers and alternative isoform databases, including UCSC Genome Browser: <https://genome.ucsc.edu/> (Kent et al., 2002); Ensembl Genome Browser: <http://www.ensembl.org/> (Stalker et al., 2004); AS-ALPS: <http://as-alps.nagahama-i-bio.ac.jp/> (Shionyu et al., 2009); ExonMine: <http://www.imm.fm.ul.pt/exonmine/> (Mollet et al., 2010). Exon 2 from mouse *MCL1* is also absent in lists of exons that are alternative across a variety of species and of species-classifying events (Barbosa-Morais et al., 2012; Merkin et al., 2012). To explore this further, patterns of *MCL1* splicing in the absence or presence of Sudemycin C1 were investigated using minigenes carrying the relevant human or mouse genomic sequences (part of exons 1 and 3, introns 1 and 2 and exon 2). The analyses were carried out both in human HeLa cells as well as in mouse 3T3 cells, to explore the possibility that species-specific differences in *MCL1* alternative splicing regulation were due to differences in the splicing machinery of mouse and human cells, rather than to differences in *cis*-acting regulatory sequences.

The results of these experiments indicate that mouse *MCL1* transcripts are refractory to the exon skipping effects induced by Sudemycin C1 in human *MCL1* (Figures II.7A and II.7B). The results were similar in human and in murine cells, although the effects of the drug on human transcripts were slightly weaker in murine cells (Figures II.7A and

II.7B), possibly due to different drug sensitivity among cell lines, or among different organisms.

Next we generated chimeric minigenes where either the region 5' of the putative BP (from -104 to -29 bp from the 3'ss, being the putative BP at -23 bp) or the entire 3' splice site region (from -104 to the 3'ss, including putative BP and polypyrimidine tract) of the mouse transcript were replaced by equivalent sequences of the human *MCL1* gene. While replacement of the region 5' of the BP did not confer drug sensitivity to the mouse minigene-derived transcripts, replacement of the 3' splice site region did (Figures II.7A and II.7B, compare m-hMCL1 chimeras 1 and 2).

To get further insights into the determinants of the drug sensitivity of human *MCL1* transcripts, we compared the 3' splice site regions of the mouse and human genes (Figure II.7C). Three main noticeable differences include: a) a longer polypyrimidine tract in the mouse intron, b) absence of the ACCTGCA sequence that includes several potential BP sequences (as mapped by our and previous work) (Mercer et al., 2015), and c) a single nucleotide difference at position -29 of the human sequence, where a guanosine is replaced by adenosine in the mouse.

Mutation of G -29 to A in the human minigene resulted in some decrease in the effect of Sudemycin C1 treatment (Figure II.7D, compare *hMCL1* G-29A vs *hMCL1* wt in 7A), but the reverse mutation (A to G) at the equivalent position of the mouse 3' splice site region did not affect drug resistance of the mouse minigene (Figure II.7D, *mMCL1* A-26G).

Strikingly, replacement of the polypyrimidine tract of the human sequence (Figure II.7C, underlined residues in *hMCL1*) by the longer polypyrimidine tract sequence of the mouse sequence (Figure II.7C, underlined residues in *mMCL1*) conferred strong resistance to the drug effects (Figure II.7D, compare *hMCL1* mPY with *hMCL1* wt in Figure 7A). Resistance to the drug was maintained even after deletion of the -23/-17 ACCTGCA sequence (Figure II.7D, *hMCL1* mPY Δ ACCTGCA), suggesting that the presence of the longer polypyrimidine tract suppresses the effects of the drugs in the absence of a cluster of BPs found important for the function of the human 3' splice site, further arguing for a role of polypyrimidine tract length and/or sequence in conferring drug resistance.

As an additional confirmation of this hypothesis, we took advantage of the observation that inclusion of the alternative exon EDB of Fibronectin 1 (*FN1EDB*) was not repressed, but actually enhanced, by SF3B1 knock down (Papasaikas et al., 2015). Consistently, Sudemycin C1 treatment induces very limited skipping of this exon compared to the DMSO control (Figure II.7E). When the 3' splice site region of *MCL1* intron 1 was replaced by the equivalent region of the intron preceding the *FN1EDB* exon, the chimeric RNA was unresponsive to Sudemycin 1 treatment (Figure II.7F). As shown in Figure II.7G, *FN1EDB* is preceded by a very long polypyrimidine tract that positions the BP relatively far from the 3'ss, further suggesting that a strong polypyrimidine tract confers resistance to SF3B1-inhibitors.

Collectively, our results indicate that the strength and degeneracy of BPs, as well as the length/strength of the polypyrimidine tract are determinants of the sensitivity of 3' splice sites to treatment with

SF3B1-targeting drugs. Furthermore, the presence of decoy or *bona fide* BP mimics, located 5' of the functional BP, can also modulate drug response. Taken together, our results demonstrate that a variety of sequence features determine the differential sensitivity of 3' splice sites to anti-cancer drugs targeting the spliceosome, to the point that a single nucleotide difference can significantly alter their response to these compounds. These observations provide the basis to explain why splicing inhibitory drugs can have effects on cell proliferation without totally blocking the splicing process.

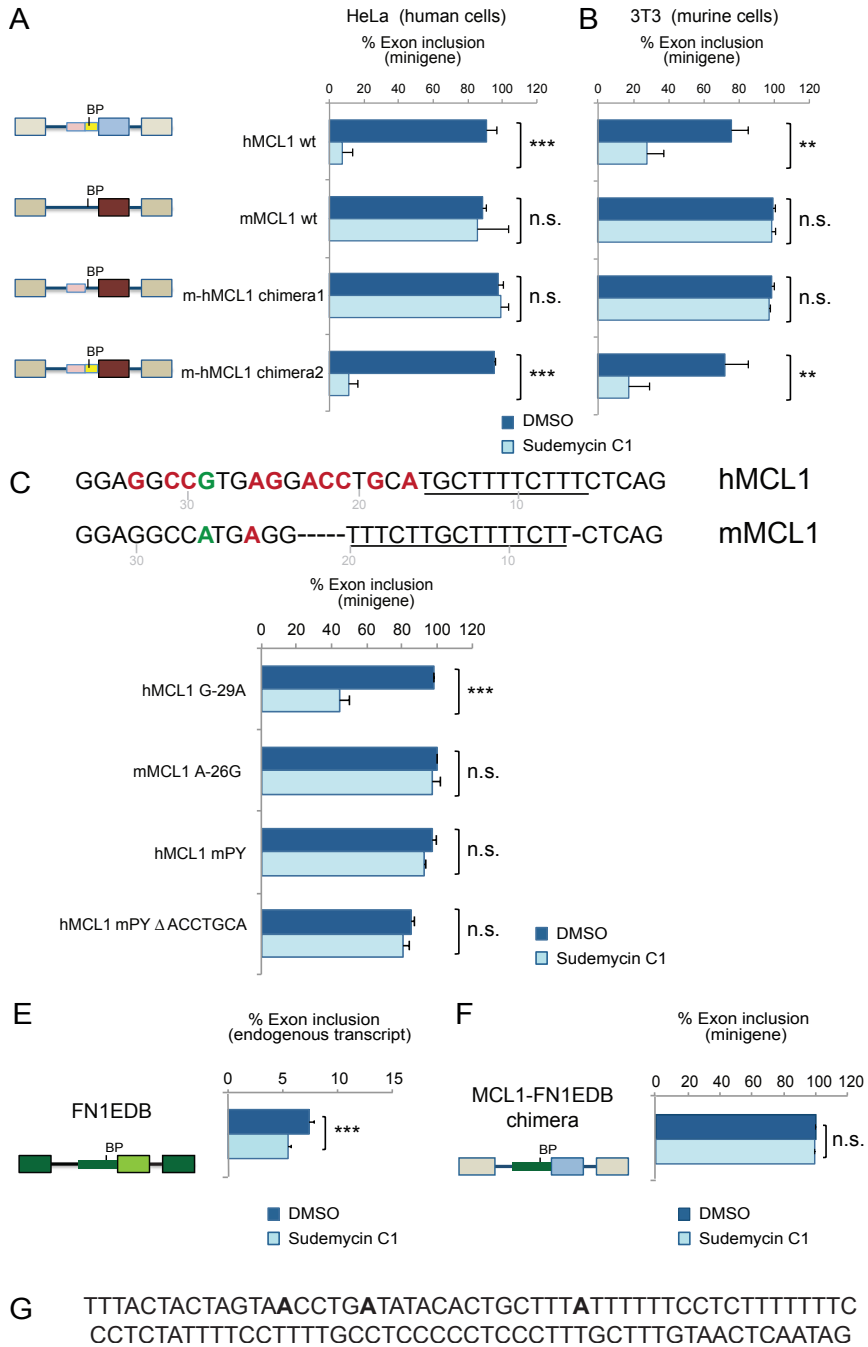


Figure II.7. Comparison between mouse and human transcripts regulation reveals a role for polypyrimidine tracts in modulating drug response. (A and B) Analysis of *MCL1* exon 2 inclusion in transcripts derived from human (hMCL1) or mouse (mMCL1) minigenes, tested in human (HeLa, A) or mouse (3T3, B) cell lines. In chimeric constructs, mouse sequences were replaced by the equivalent human ones, either 5' of the BP (chimera 1, positions -104 to -29 from the AG of the mouse minigene were replaced with sequences -103 to -28 of the human one), or the 3' splice site region (chimera 2, where all the last 103 nt of the intron were replaced with the human one, adding also the human BP and polypyrimidine tract apart from the whole sequence added in chimera 1). Cells were treated with DMSO or with 20 μ M Sudemycin C1 for 8 h, 16 h after transfection. **, ***, n.s.: t-test comparison of Sudemycin C1-treated vs DMSO-treated minigenes with p-value < 0.001, 0.0001 or > 0.01, respectively; n = 3. (C) Comparison of 3' end sequences of human and mouse *MCL1* intron 1. Underlined residues indicate the polypyrimidine tracts, G/A difference in the 5' part of the region is indicated in green and residues indicated in red correspond to the predicted BP (for mouse *MCL1*) and mapped BPs (for human *MLC1*). (D) Polypyrimidine tract length/sequence contributes to drug sensitivity. Assays as in (A) were carried out for the indicated mutant constructs: hMCL1 G-29A: human construct replacing G -29 by A; mMCL1 A-26G: mouse construct replacing the A at the equivalent position -XX by G; hMCL1-mPy: human construct replacing its polypyrimidine tract (underscored in C) by the corresponding polypyrimidine tract in the mouse 3' splice site (also underscored in C); hMCL1; hMCL1 mPy Δ ACCTGCA: same construct but with deletion of residues -23/-17 of the wild type human sequence. E) Effects of Sudemycin C1 treatment (20 μ M for 8 h) on alternative splicing of endogenous Fibronectin 1 *FN1EDB* exon. F) Lack of effects of Sudemycin C1 treatment on *MCL1-FN1EDB* chimera minigene in which the 3' 119 bp of *MCL1* intron 1 were substituted by the 3' 119 bp of the intron preceding *FN1EDB* exon. Experimental conditions and analysis of results were carried out as in (A).

RNA-Seq of Sudemycin C1-treated cells reveals extensive splicing and gene expression alterations

In order to expand the study of sequence features influencing Sudemycin C1-induced regulation to a genome-wide level, we performed both splicing sensitive microarrays (presented in the following section) and RNA-Seq analyses on Sudemycin C1-treated samples. Cultured HeLa cells were co-treated with 10 μ M Sudemycin C1 and 2 mM 5-Bromouridine (BrU) for 3 hours. This modified nucleotide gets incorporated during transcription and a specific antibody allows the isolation of BrU-containing (i.e. recently transcribed) transcripts by immunoprecipitation (IP) (Paulsen et al., 2014). Both total RNA and BrU-RNA were isolated and deeply sequenced, revealing widespread alternative splicing regulation and intron retention (Figure II.8A), as well as extensive gene expression changes (Figure II.8C). The number of detected changes was lower in BrU-IP RNA, most likely because of the lower yield and coverage of RNA-Seq. The majority of transcripts with splicing changes displayed changes in a single splicing event (Figure II.8B), supporting the idea that specific sequence features confer differential sensitivity to the drugs, even for introns from the same transcriptional unit.

Gene ontology analysis (GO) revealed that genes displaying splicing changes (Figure II.9) and gene expression changes (Figure II.10) are involved in various processes, including cytoskeleton organization, macromolecule metabolism, cell cycle and proliferation, transcription and RNA processing, pointing towards a relevant role of SF3B1 in the control of vital functions.

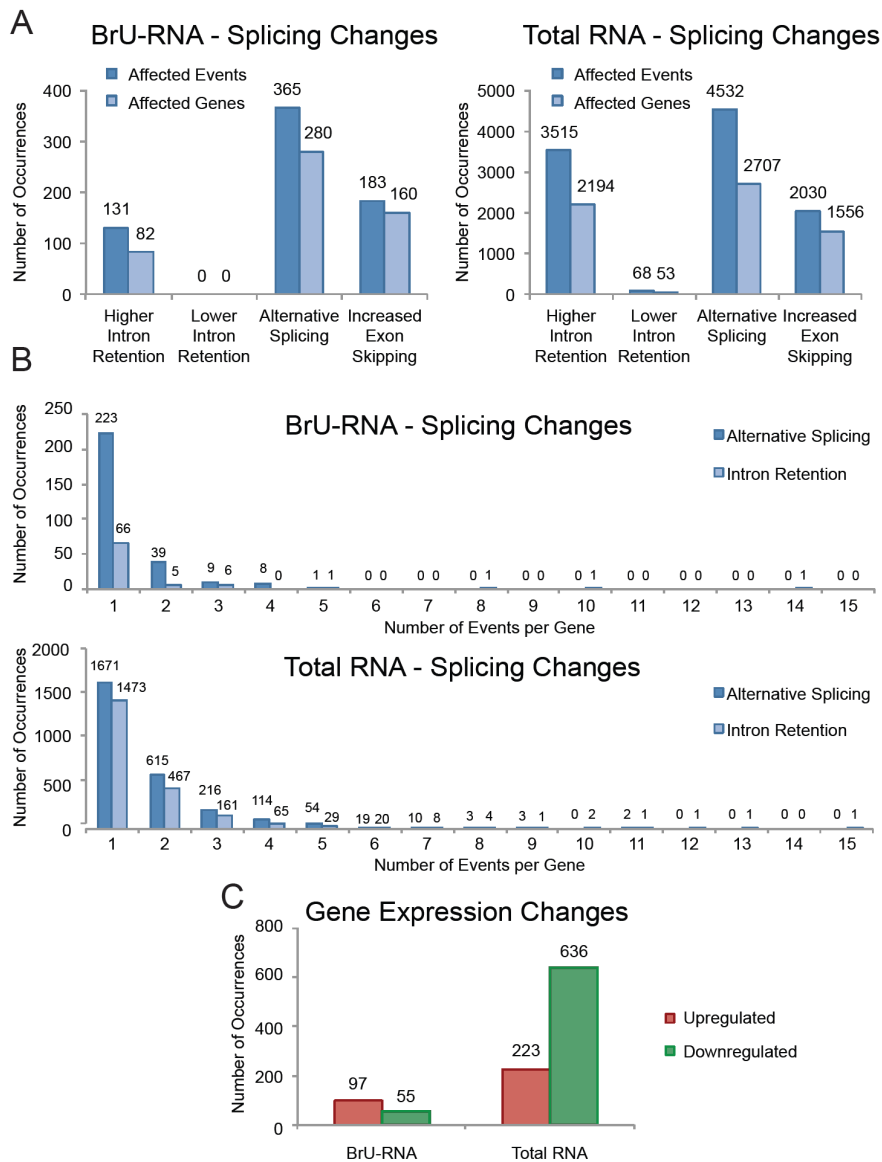


Figure II.8. Summary of RNA-Seq results. (A) Summary of different types of splicing changes detected by SANJUAN pipeline (dark blue) and the corresponding genes (light blue) in BrU-RNA and total RNA. Alternative splicing category includes alternative 3'ss and 5'ss, mutually exclusive exons and composite events. (B) Distribution of the number of occurrences (single or multiple events per gene) for intron retention and alternative splicing changes in BrU-RNA and total RNA. Instances of several retention events in the same transcript occur in genes with more than 30 introns, like *PLEC* (coding for structural protein Plectin), *FASN* (coding for fatty acid synthase) and *FLNA* (coding for actin-binding filamin A protein). (C) Summary of gene

expression changes in the two sequencing experiments. Alternative splicing analysis was performed with SANJUAN, setting a medium confidence threshold (p -value < 0.01 , Delta PSI $> 10\%$). Gene expression analysis was performed with Cuffdiff with a q -value ≤ 0.01 and FPKM ≥ 1 as thresholds.

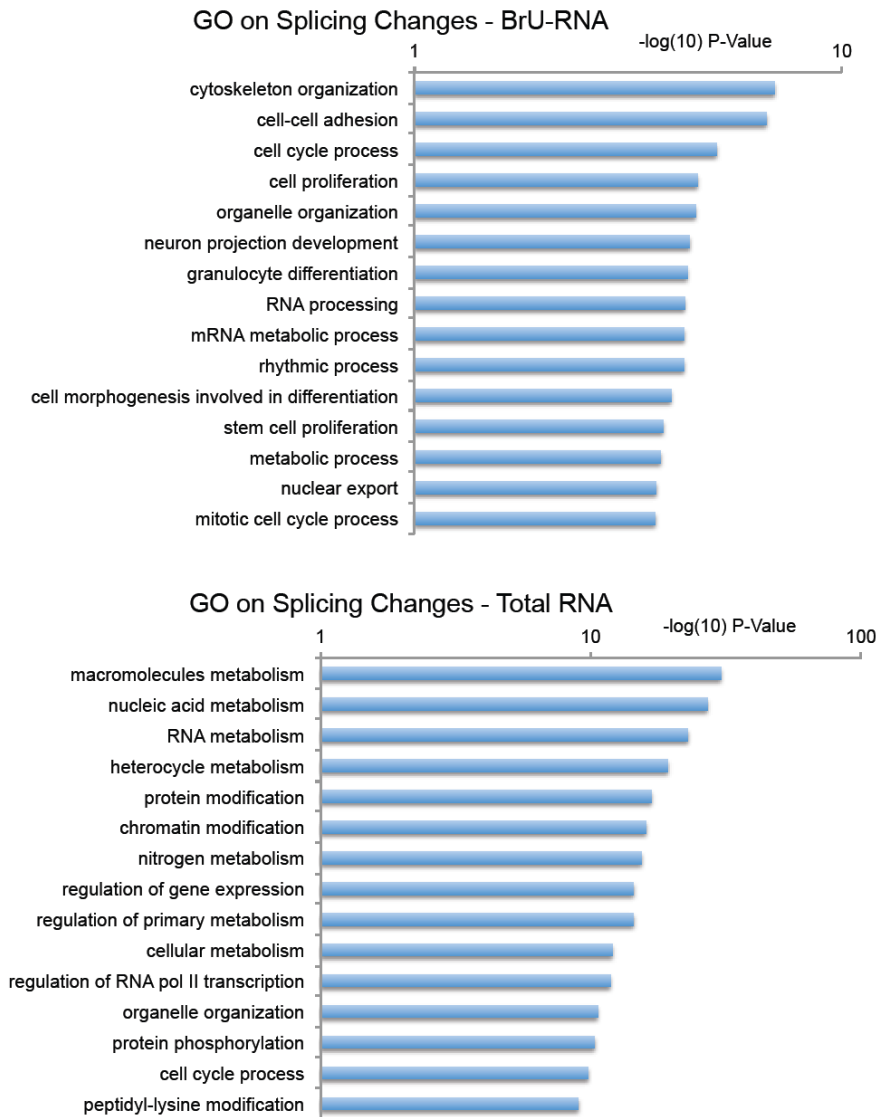


Figure II.9. Gene ontology analysis of genes displaying splicing changes upon Sudemycin C1 treatment. Only intron retention and exon skipping changes were considered for performing GOrilla gene ontology analysis using the set of all expressed genes as control list (thresholds: FPKM ≥ 1 , “ok” status). The top 15 categories of enriched processes are shown.

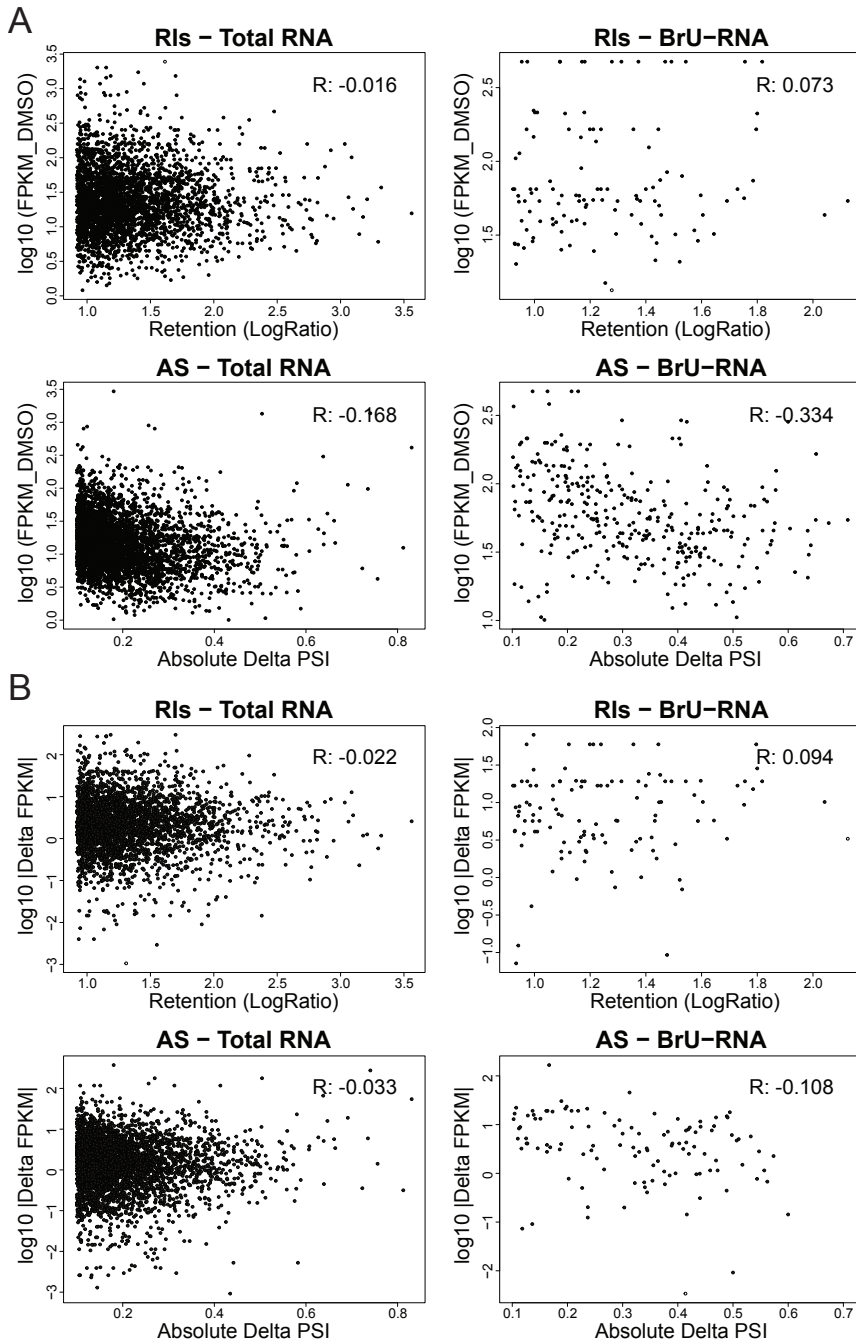


Figure II.10. Gene ontology analysis of genes displaying gene expression changes upon Sudemycin C1 treatment. Cuffdiff genes with $FPKM \geq 1$, “ok” status and $q\text{-value} \leq 0.01$ were considered for performing GOrilla gene ontology analysis against the list of all expressed genes (i.e. with $FPKM \geq 1$ and “ok” status). The top 15 categories of enriched processes are shown.

Sudemycin C1-induced splicing alterations correlate with sequence content and BP features rather than expression levels

Connections between transcription and splicing are tight and recent studies reported that elevated cMYC levels increase the sensitivity of cancer cells to splicing inhibition (Hsu et al., 2015; Hubert et al., 2013), pointing at the possibility that high levels of transcription require high levels of splicing and therefore could be one of the main reasons of sensitivity to splicing inhibition. We therefore analyzed whether there was a correlation between the levels of gene expression and the inhibitory effects of Sudemycin C1 on splicing events within that gene (distinguishing between intron retention and other alternative splicing categories). Results show that the correlation is very low with both the expression levels in DMSO (Figure II.11A) and the change in expression among treated and control (Figure II.11B). The same result was obtained from total and recently transcribed BrU-labeled RNA (Figures II.11A and II.11B), further supporting the concept that transcript-specific sequence features are the main determinants of drug sensitivity.

Figure II.11 (next page). Absence of correlation between drug-induced splicing effects and gene expression levels. (A) Cuffdiff and SANJUAN outputs were merged to calculate the correlation coefficient R (shown in every plot) between splicing changes and expression levels of the gene corresponding to each altered splicing event. Splicing changes were measured as LogRatio (natural logarithm of the ratio among retention levels in treated vs control samples) for intron retention (RIs) and as absolute Delta PSI (difference in “Percentage of Spliced In” among treated and control) for alternative splicing events (AS). Expression levels were measured in DMSO-treated cells (control dataset) in FPKM units, but similar results were obtained in Sudemycin C1-treated cells (all coefficients $|R| < 0.2$). (B) Same analysis of panel (A), but showing correlations with the absolute difference in expression between drug- and control- treated cells ($|\Delta \text{FPKM}|$), measured as log10.



In order to study the role of intronic sequence content for alternative splicing regulation upon drug treatment, we first focused on the region around the BP that is contacted by SF3B1 (Gozani et al., 1998; Rauhut et al., 2016; Yan et al., 2016). We took advantage of two available tools: an algorithm for scoring BP and Py-tracts (Corvelo et al., 2010) and a list of human BPs mapped in cultured cells (Mercer et al., 2015). We focused on the more abundant classes of affected events (i.e. retained introns and skipped exons; Figure II.8A), and overlapped them with the information about mapped and scored BPs.

For statistical analysis of affected exon skipping events, we considered the upstream introns as the regulated dataset and the downstream ones (containing 3'ss that successfully compete for U2 snRNP binding with the 3'ss affected by the drug) as the control dataset. For intron retention analysis, we compared more retained introns versus all the introns within expressed genes ($\text{FPKM} \geq 1$) once the regulated introns were removed.

Mapped BPs are only available for a subset of annotated introns. Of the 56176 mapped BPs (Mercer et al., 2015), about half of them (28672 BPs, belonging to 25687 different introns) correspond to computationally predicted BPs (Corvelo et al., 2010) and could therefore be scored (the other half of the mapped BPs could not be bioinformatically predicted mainly because they are non-A BPs, they belong to non-consensus motifs or they are above the 200 nt distance threshold set for the analysis). Therefore, focusing only on sequences with mapped and scored BPs reduced the total number of affected cases to be considered, as reported in Figure II.12A (to be compared to Figure II.8A). However, this selection was important to focus on

accurately scored functional BPs and make the downstream analysis more robust.

By comparing the sequence content around mapped BPs (from 30 nt upstream to 19 nt downstream of the BP), clear differences between affected and non-affected RNA regions were detected (Figure II.12A), mainly corresponding to a reduction of T nucleotides upstream and downstream of the BP in retained introns (and more moderately in introns preceding skipped exons) in BrU-RNA, consistent with our previous results arguing that weak Py-tracts are more susceptible to drug inhibition (Figure II.5). Similar -albeit less pronounced- differences can be detected in total RNA data (Figure II.12B).

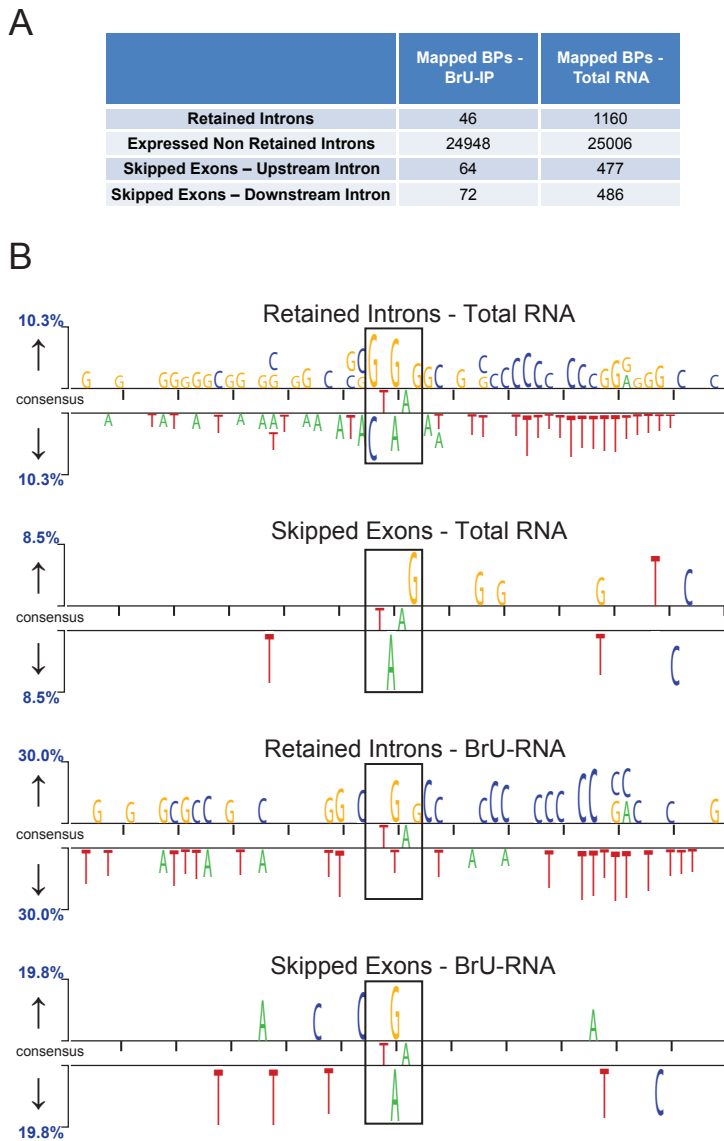


Figure II.12. Sequence content around mapped BPs for Sudemycin C1-affected and non-affected RNA regions. (A) Number of sequences with mapped and scored BPs used for the analyses in Figures II.12B and II.13A. (B) Two sample logos (Vacic et al., 2006) showing changes in nucleotides frequencies around mapped and scored BPs between retained introns and control datasets, or introns preceding skipped cassette exons and the downstream intron, in recently transcribed, BrU-labeled RNA. Positions for which the t-test gives p-values < 0.05 are shown above the line if enriched, below the line if depleted. The size of the residues is proportional to the enrichment/depletion. Consensus nucleotides (i.e. the branch A and the T two

nucleotides upstream of it) are represented between the two horizontal axes. A black box highlights the position of the BP consensus sequence.

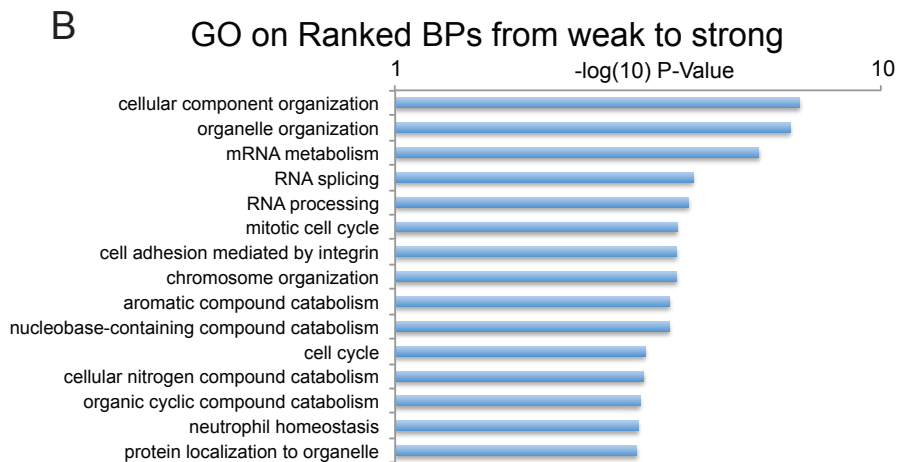
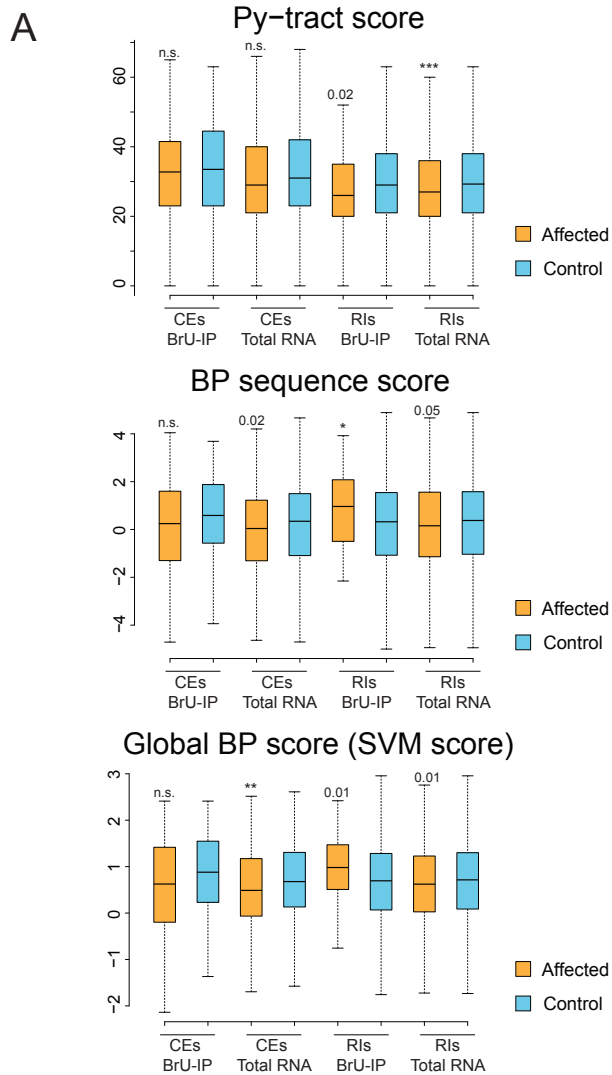
Consistently, scores corresponding to Py-tract strength, BP sequence score and global BP strength, defined by the SVM score (Corvelo et al., 2010), are generally lower in affected versus non-affected 3' ss regions in both retained introns and skipped exons from total RNA-Seq data (Figure II.13A). The statistical significance of the differences varies, however, probably because the number of sequences analyzed is relatively limited, particularly for BrU-RNA, which may be the reason why BP scores of 3' ss of retained introns appear to be stronger in BrU-labeled RNAs (Figure II.13A).

The difference in BP strength would be compatible with the presence of weaker, possibly rather promiscuous BPs contributing to drug sensitivity (like in the *MCL1* case, Figures II.3 and II.4), while drug-resistant sequences could be associated with multiple strong BPs (like in the *PDCD10* case, Figures II.3 and II.4). We therefore explored the possible correlation between the presence of multiple BPs and drug sensitivity, but the results were inconclusive (data not shown).

Collectively, genome-wide and minigenes analyses support the hypothesis that BP and Py-tract strength influences 3' ss sensitivity to SF3B1 inhibition. These differential effects may be at the basis of the antitumor effects of these inhibitors, as supported by GO analyses (Figures II.9 and II.10). Consistent with this, GO analysis of genes expressed in HeLa cells, ranked according to BP strength (from weakest to strongest based on their SVM score), shows that genes containing weak BPs are enriched in functional categories related to RNA processing, cell cycle, organelle organization, cell adhesion and several other vital functions (Figure II.13B). Thus, evolution might have

avored the presence of weak BPs in some regulatory transcripts to increase the possibility of regulation and/or detecting alterations of the splicing machinery (e.g. to induce apoptosis) in particular cell contexts, including tumoral transformation (Crews et al., 2016; Hsu et al., 2015).

Figure II.13 (next page). Weak BPs and Py-tracts are associated with drug sensitivity and are enriched in genes related to RNA processing and cell cycle control. (A) Scores corresponding to Py-tract sequence, BP sequence and overall BP strength (SVM score). RIs: retained introns (affected and control sets in orange and blue, respectively). CEs: cassette exons (introns upstream of the regulated exon in orange or downstream control set, in blue). N.s.: non-significant t-test (p -value > 0.05); *: p -value < 0.01 ; **: p -value < 0.001 ; ***: p -value < 0.0001 . Borderline p -values (between 0.05 and 0.01) are specified. (B) Gene ontology (GO) analysis of genes containing weak BPs. BP information was available for around 26000 mapped BPs (Mercer et al., 2015) that could be scored by the SVM scoring system (Corvelo et al., 2010). These BPs belong to 6800 different expressed genes (FPKM ≥ 1 in DMSO total RNA-Seq). BPs were ranked based on their SVM scores and corresponding genes were subject to ranked GOrilla GO analysis. Top 15 enriched processes are shown.



GC content and transcript length differ between Sudemycin C1-retained and non-retained introns

The analyses presented above revealed interesting sequence features linked to drug sensitivity and prompted us to carry out a second approach to have an independent confirmation of the results. RNA-Seq data were therefore reanalyzed using a combination of VAST-TOOLS software (Braunschweig et al., 2014) and Matt, a toolkit developed by André Gohr in Manuel Irimia's group and our group, which was used for the analysis of retained introns.

BP and Py-tracts were scored again with the SVM algorithm (Corvelo et al., 2010) and SF1/BBP binding sites (the first step in BP recognition) within the 3' 150 intronic nucleotides were also used as a measure of potential BP strength (Corioni et al., 2011). All the analyzed features are predicted to be weaker in drug-retained introns, which also display a tendency for having fewer predictable BPs with score > 0, both in BrU RNA (Figure II.14A) and in total RNA (Figure II.15B). These results are in agreement with our previous minigene and genome-wide analyses (Figures II.3, II.4, II.6 and II.13A). Of notice, several other sequence properties (including 5' and 3'ss strengths) were also found to be significantly different, e.g. with weaker splice sites generally correlating with higher drug sensitivity (data not shown).

The analyses also revealed a clear tendency for retained introns and neighboring exons being shorter and more GC-rich in both BrU-RNA (Figure II.14B) and in total RNA (Figure II.15A). Previous studies reported lower nucleosome occupancy for exons neighboring short and GC-rich introns. The latter are prone to intron retention, also in association with lower recruitment of nucleosomes-associated SF3B1

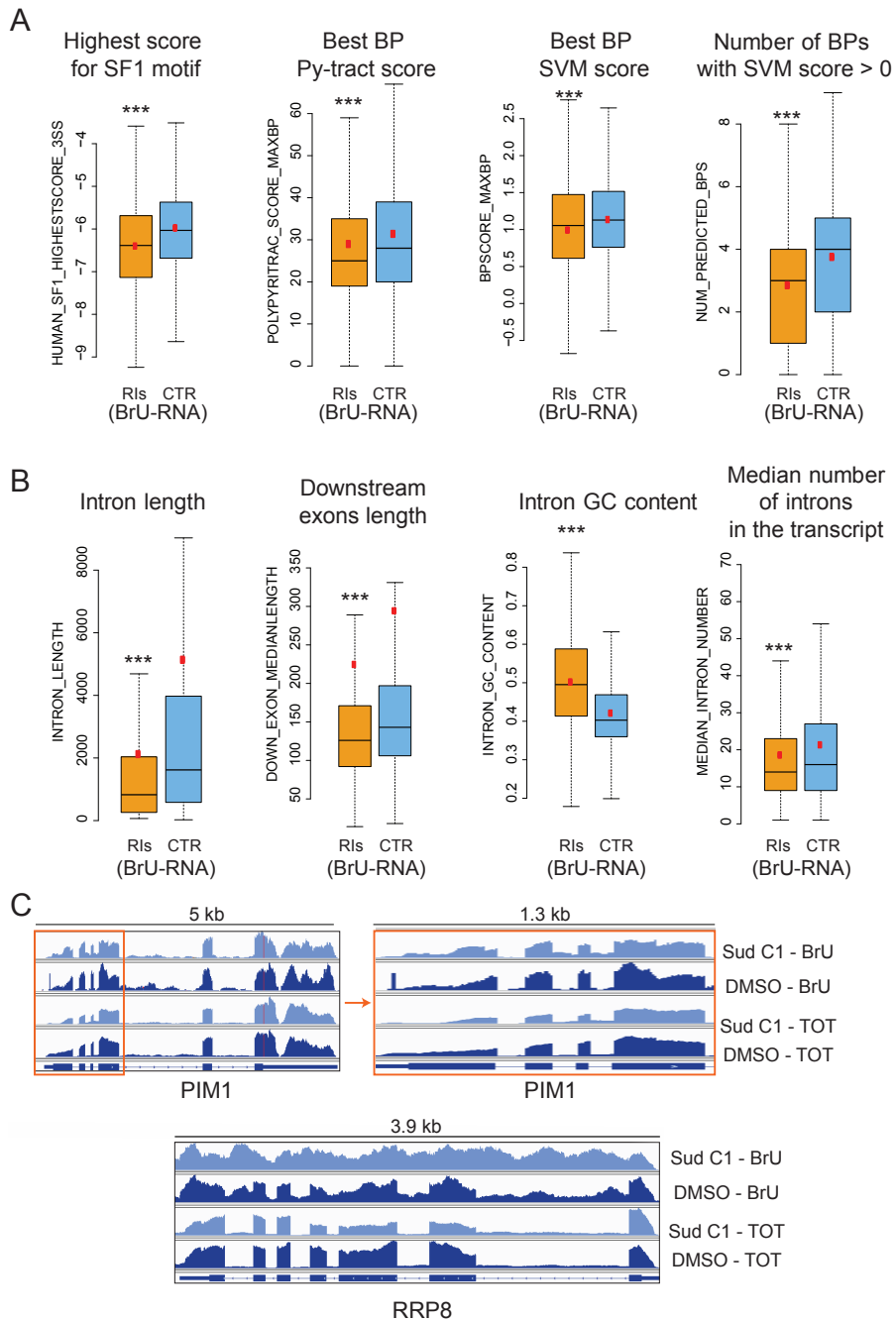
(Amit et al., 2012; Kfir et al., 2015). A different configuration of BPs and Py-tracts was also proposed to be linked to GC-content, with C-rich Py-tracts in GC-rich introns and preferential CUNAN motif for the BP (Mercer et al., 2015).

In general, drug-retained introns also occurred in transcripts with lower number of introns (Figure II.14A), suggesting that shorter genes may be more susceptible to drug inhibition. Two examples are shown in Figure II.14C: PIM1 (Pim-1 Proto-Oncogene, Serine/Threonine Kinase) belongs to a known class of short-lived oncogenes, along with MCL1, cMYC, cyclin D1 (Schatz and Wendel, 2011), while RRP8 (Ribosomal RNA Processing 8 Methyltransferase Homolog) contains one of the top regulated introns detected in the analysis. While the first displays higher intron retention for very short introns, the second displays more widespread intron retention across the transcript, which is more prominent in BrU-labeled RNA (partly due -as expected- to the isolation of nascent RNA that still has to be fully spliced, as revealed by the higher intronic signal in the DMSO control). These examples illustrate how drug-induced intron retention can occur in transcripts with high biological relevance for cancer cells, as supported also by the gene ontology analyses (Figures II.9 and II.10), and how different ranges of effects can be achieved within a general tendency for retention.

Our genome-wide data also reveal that SF3B1 inhibition affects other categories of transcripts, including a small proportion of snoRNAs (data not shown), whose processing is linked to splicing (Hirose et al., 2006), and some U12 introns, suggesting that SF3B1 can be affected by the drugs also when is part of the U12 snRNP (although indirect effects or

crosstalk between U2 and U12 introns cannot be ruled out at this point). In any case, the prevalence of intron retention and exon skipping upon drug treatment is clear from both the RNA-Seq analysis (presented in this section) and the microarray analysis (presented in the next section). Collectively, our data support an important role for intronic features in Sudemycin C1-induced regulation, with relevant consequences for proliferation and apoptosis control.

Figure II.14 (next page). Analysis of drug-induced intron retention in BrU-RNA reveals anti-correlation with BP and Py-tract strength, and roles for GC content and transcript length. (A) Boxplots corresponding to the indicated features are shown, with median values shown by the black line and mean values by the red dot. Outliers were discarded. RIs: retained introns, CTR: non-differentially spliced introns, with absolute PSI difference < 2.5%. BP features and SF1 binding motif were analyzed in the 3' 150 nucleotides of each intron (introns shorter than 150 nt were not considered), obtaining 3759 RIs and 5564 CTR. (B) Boxplots corresponding to the indicated features. Upstream and downstream exon length: median value of all possible upstream/downstream exons. Number of RIs: 4581; number of CTR: 6197 introns. Statistical significance was evaluated by permutation tests with 100'000 iterations. ***: p-value < 0.001. Because of the low sequencing coverage, reads from the two replicates were merged into one single dataset for the analysis. (C) Examples of short transcripts with short affected introns: PIM1 (Pim-1 Proto-Oncogene, Serine/Threonine Kinase) is a known short-lived oncoprotein (Schatz and Wendel, 2011) and RPM8 (Ribosomal RNA Processing 8 Methyltransferase Homolog) is among the top differentially retained introns detected in the analysis. A zoomed figure for the first three introns of *PIM1* is also shown.



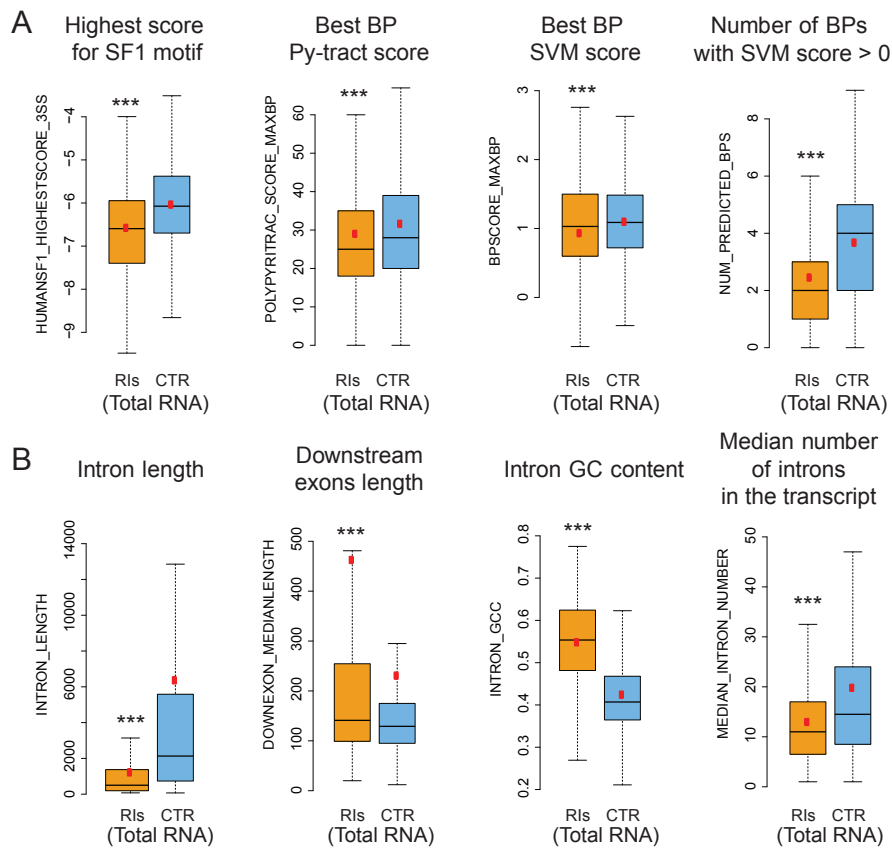


Figure II.15. Analysis of intron retention in total RNA also reveals anti-correlation with BP and Py-tract strength and roles for GC content and transcript length. (A) Boxplots corresponding to the indicated features are shown, with median values shown by the black line and mean values by the red dot. Outliers were discarded. RIs: retained introns, CTR: non-differentially spliced introns, with absolute PSI difference < 2.5%. BP features and SF1 binding motif were analyzed in the 3' 150 nucleotides of each intron (introns shorter than 150 nt were not considered), obtaining 628 RIs and 4568 CTR sequences (from a downsampled sample of 5000). (B) Boxplots corresponding to the indicated features. Upstream and downstream exon length: median value of all possible upstream/downstream exons. Number of RIs: 827; number of CTR: 43885 introns, down-samples to 5000. Statistical significance was evaluated by permutation tests with 100'000 iterations. ***: p-value < 0.001. Because of the high sequencing coverage, reads from the two replicates were kept separated and a conservative approach was followed by only considering introns passing the threshold for every cross-comparison.

Annexed Table II.1. Summary of minigenes results presented in Part II.

Minigene	Average Inclusion DMSO (%)	StDev DMSO	Average Inclusion SUDC1 (%)	StDev SUDC1	First intron length	First intron sequence (last 100nt)
PDCD10 WT	98.60	2.26	96.46	2.28	546	ATTAGAGTCTACTTAAATTTCTTCAAAATAGCTGATGTTCTTTCTCTCACAAATTTCTTCCAAATGAGCTGATATCTTTTCCAAATGTTCCCTTCCACAG
PDCD10 Δ E1	96.50	3.11	71.63	3.66	524	ATTAGAGTCTACTTAAATGATTAAGAGCTGATTAATTTCTTCAAAATAGCTGATGTTCTTTCTCTCACAAATTTCTTCCAAATGTTCCCTTCCACAG
PDCD10 Δ E1-E2	95.62	5.35	58.23	5.92	509	AAAGCTGACATTAATTAAGCTGATTAAGAGCTGATTAATTTCTTCAAAATAGCTGATGTTCTTTCTCTCACAAATTTCTTCCAAATGTTCCCTTCCACAG
PDCD10 Δ E1-E2-E3	84.44	14.62	31.52	17.28	486	AAAGCTGACATTAATTAAGCTGATTAAGAGCTGATTAATTTCTTCAAAATAGCTGATGTTCTTTCTCTCACAAATTTCTTCCAAATGTTCCCTTCCACAG
MCL1 wt	90.91	6.00	7.06	6.64	351	TAAATCCAGTAGCAGATTTCCGCGCGGCTGGCCGAGGAGTTCGAGTTCGCGGCTGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG
MCL1 Δ	97.76	0.69	10.59	12.19	275	TTTGGAAATGACAGCTCTTCTTCAAGACGGAAAGGCTGAGTCAATTTCAAGTGGCTCAACCTGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG
MCL1 Δ + upstream PDCD10	97.41	1.33	21.26	5.84	356	TCCCACTCCAAACAGTAGTCGAAAGGCTGGCTGTTCTGAGTGAATGGCTGATTAATTTCTTCAAAATAGCTGATGTTCTTTCTCTCACAAATTTCTTCCAAATGTTCCCTTCCACAG
MCL1 Δ + E3-E2-E1	96.58	0.72	85.28	10.01	356	TAGATTAAAGTGTACTTAAATTTCTTCAAAATAGCTGATGTTCTTTCTCTCACAAATTTCTTCCAAATGTTCCCTTCCACAG
MCL1 Δ + BP	100.00	0.00	69.66	7.60	282	ATGGCAGCTCTTTCTTCAAAAGACCGGAAAGGCTGGATGTCTCAATTTCAAGTGGCTCAACCTGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG
MCL1 Δ + DECOY	83.36	9.18	4.97	0.51	281	AATGGCAGCTTTCTTCAAAAGACCGGAAAGGCTGGATGTCTCAATTTCAAGTGGCTCAACCTGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG
MCL1 Δ + Y-rich seq	97.96	2.09	66.89	15.49	289	CTCTTGTCAAGAAGGCTGGATGCTCAATTTCAAGTGGCTCAACCTGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG
MCL1 Δ + Y-rich seq - BP	100.00	0.00	99.65	0.60	296	TCAAGACCGTAAGGCTGGATGCTCAATTTCAAGTGGCTCAACCTGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG
PDCD10 Δ CAATGTAG	62.55	10.39	19.39	6.62	539	TGATTAGATTAGAGCTGTACTTAAATTTCTTCAAAATAGCTGATGTTCTTTCTCTCACAAATTTCTTCCAAATGTTCCCTTCCACAG
MCL1 Δ + TTCTTTCT	99.16	0.88	16.71	5.01	284	GGCAGCTCTTTTCAAAAGACCGGAAAGGCTGGATGTCAATTTCAAGTGGCTCAACCTGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG
MCL1 Δ + AAATGT	99.68	0.08	65.31	4.91	281	AATGGCAGCTTTTGTTCAAAGACCGGAAAGGCTGGATGTCAATTTCAAGTGGCTCAACCTGAGTTCGAGTTCGAGTTCGAGTTCGAG
MCL1 Δ -26 MUT	63.94	6.94	4.04	0.52	351	TAAATCCAGTAGCAGATTTCCGCGCGGCTGGCCAGGCGCAAGTTCGCGCGGCTTAAAGACAAAGAGGCTGGCTGAGTTCGAGTTCGAGTTCGAG
MCL1 Δ -26 TACTAAC	99.41	1.03	100.00	0.00	351	TAAATCCAGTAGCAGATTTCCGCGCGGCTGGCCAGGCGCAAGTTCGCGCGGCTTAAAGACAAAGAGGCTGGCTGAGTTCGAGTTCGAG
MCL1 Δ ACCTGCA	11.85	2.42	4.35	0.40	344	AGTTCTGTAATCCAGTAGCAGATTTCCGCGCGGCTGGCCAGGCGCAAGTTCGCGCGGCTTAAAGACAAAGAGGCTGGCTGAGTTCGAGTTCGAG
PDCD10 Δ BP-25	98.42	1.37	99.88	0.10	545	GATTAAGATTTACTTAAATTTCTTCAAAATAGCTGATGTTCTTTCTCTCACAAATTTCTTCCAAATGTTCCCTTCCACAG
PDCD10 Δ A-49	99.32	0.60	93.35	2.38	545	GATTAAGATTTACTTAAATTTCTTCAAAATAGCTGATGTTCTTTCTCTCACAAATTTCTTCCAAATGTTCCCTTCCACAG
mMCL1 wt	88.59	2.17	85.24	18.57	273	TAGACCGGAGAGACAGTCCCTTCCTCTCTGCGGCTGGCGGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG
m-HMCL1 chimera 1	97.63	3.36	99.30	0.99	349	AAATCCCACTAGGCAATTTCCGCGCGGCTGGCCAGGCGCAATCTGCGCGGCTTAAAGACAAAGAGGCTGGCTGAGTTCGAGTTCGAG
m-HMCL1 chimera 2	95.46	0.62	10.99	1.54	348	TAAATCCAGTAGCAGATTTCCGCGCGGCTGGCCAGGCGCAAGTTCGCGCGGCTTAAAGACAAAGAGGCTGGCTGAGTTCGAGTTCGAG
hMCL1 G-29A	97.91	0.34	44.96	5.59	351	TAAATCCAGTAGCAGATTTCCGCGCGGCTGGCCAGGCGCAAGTTCGCGCGGCTTAAAGACAAAGAGGCTGGCTGAGTTCGAGTTCGAG
mMCL1 A-26G	100.00	0.00	97.48	4.37	273	TAGACCGGAGAGAGAGGCTCCCTTCCTCTCTGCGGCTGGCGGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG
hMCL1 mPY	97.26	2.23	92.72	1.14	355	TCCGATAGCAGATTTCCGCGCGGCTGGCCAGGCGCAAGTTCGCGCGGCTTAAAGACAAAGAGGCTGGCTGAGTTCGAGTTCGAG
hMCL1 mPY Δ ACCTGCA	84.78	2.46	80.58	3.53	348	CTGTAAATCCAGTAGCAGATTTCCGCGCGGCTGGCCAGGCGCAAGTTCGCGCGGCTTAAAGACAAAGAGGCTGGCTGAGTTCGAGTTCGAG
MCL1-FN1EDB chimera	99.82	0.05	99.28	0.27	367	TTTACTACTATTAACCTGAGTACACTGCTGCTTAAATTTCTTCAAAATAGCTGATGTTCTTTCTCTCACAAATTTCTTCCAAATGTTCCCTTCCACAG

Results

Part III. Comparison of the effects of different drug variants on alternative splicing modulation

Similar biochemical mechanism of spliceosome inhibition by Spliceostatin A and Sudemycin C1

Spliceostatin A and Sudemycin C1 have related structures, but they differ in several chemical groups (Figure III.1A). The two drugs require different concentrations to reach comparable effects, as previously described for *MDM2* splicing modulation (Fan et al., 2011). A difference in inhibitory concentrations affecting cultured cells viability was also observed (Figure III.1B).

Previous work showed that Spliceostatin A (SSA) can destabilize complex A formation on 3' splice sites observed as an inhibitory effect in formation of this complex in biochemical assays which was detected only in the presence of heparin (Corrionero et al, 2011). To test whether Sudemycin C1 (Sud C1) operates by a similar mechanism, a radioactively labeled RNA corresponding to the 3' end of Adenovirus Major Later (AdML) promoter intron 1 and second exon was incubated in HeLa nuclear extracts in the absence or presence of Sudemycin C1. The drug inhibited A3' complex formation only in the presence of heparin (Figure 1C: the intensity of Sudemycin C1 A3' complex is comparable to DMSO control in absence of heparin and reduced in presence of heparin), as observed for Spliceostatin A (Corrionero et al.,

2011). Biochemical assays with both drugs in parallel confirmed the results of viability assays: Sudemycin C1 inhibitory concentration is at least 100-fold higher than for Spliceostatin A (Figure III.1D). Given that different concentrations are required for achieving similar inhibitory effects both in biochemical assays and in cultured cells, these results suggest that small chemical differences in drug structure significantly modify the activity of these compounds, as also investigated in Part I.A.

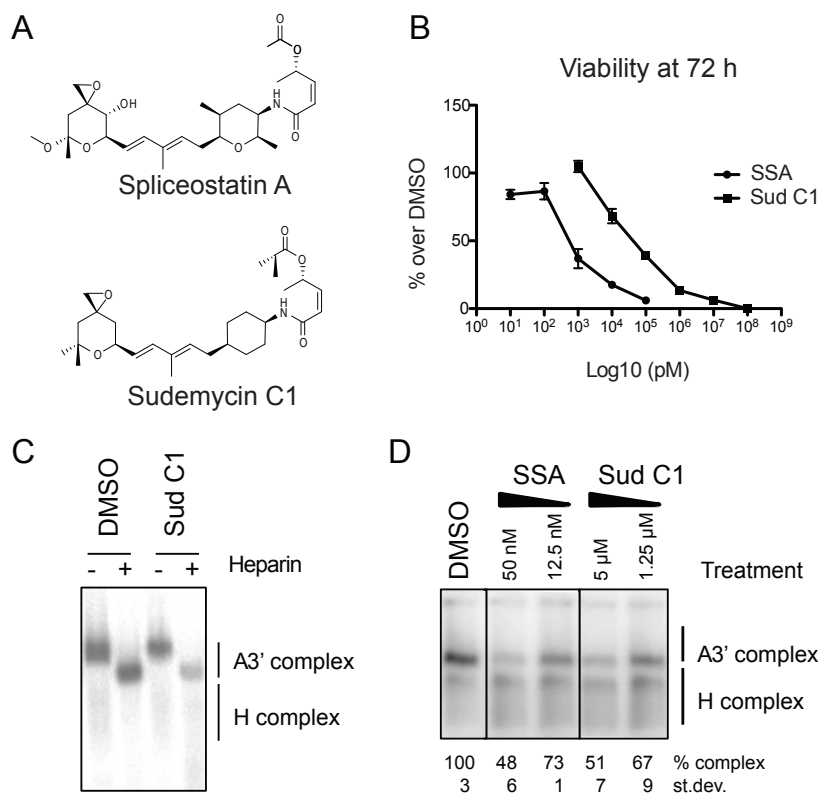


Figure III.1. Cytotoxicity and A3' complex inhibition by Sudemycin C1 and Spliceostatin A. (A) Chemical formulas of Spliceostatin A and Sudemycin C1. (B) Viability assay in HeLa cells after 72 h of incubation with the indicated amounts of drug. (C) Spliceosome assembly assay on AdML RNA containing the 3' 40 nucleotides of intron 1 and exon 2. H complex: heterogeneous ribonucleoprotein complexes. A3' complex: U2 snRNP complex on 3' AdML RNA. DMSO was used as negative control for the

treatments. 5 µg/µl heparin were added to the indicated reactions, an equivalent volume of water was added to the others. Sudemycin C1 was added at 5 µM concentration or an equivalent volume of DMSO was added as control. (D) Same assay as in (C), where both drugs were tested in parallel and 5 µg/µl heparin were added to all the reactions. Sudemycin C1 used for these and following experiments was kindly provided by Dr. Thomas Webb (St. Jude's Children Hospital) or Kamil Makowski and Mercedes Alvarez (University of Barcelona), while Spliceostatin A by Dr. Melissa Jurica (UCSC).

Evidence of differential alternative splicing modulation by Sudemycin C1 and Spliceostatin A

Alternative splicing regulation upon SSA treatment or SF3B1 depletion was previously reported (Corrionero et al., 2011). Consistently, our lab and others showed that reduced levels or activity of *core* spliceosome components affects alternative splicing without completely blocking constitutive RNA splicing (Papasaikas et al., 2015; Park et al., 2004; Pleiss et al., 2007; Saltzman et al., 2011; Wong et al., 2013).

In order to obtain a global picture of alternative splicing regulation induced by SSA and Sudemycin C1, we carried out splicing-sensitive microarray analyses of RNAs isolated from HeLa cells treated with the drugs using the HJAY Affymetrix platform. We compared these results with those corresponding to depletion of SF3B1 using two different siRNAs (Corrionero et al., 2011). The two SF3B1 depletion conditions (siRNA3 and siRNA5) resulted in very similar gene expression (60-70% overlap) and splicing changes (around 70% overlap).

The array analysis detected around 10'000 gene expression changes upon SF3B1 knockdown, 7'000 upon Sudemycin C1 treatment and 4'000 upon SSA treatment (Figure III.2A, left panels). While around 2'100 alternative splicing changes were found to be induced by SF3B1

knockdown and 1'500 upon Sudemycin C1 treatment, only 282 were observed upon SSA treatment (Figure III.2A, right panels). Therefore, under these experimental conditions, SSA induced a more limited spectrum of gene expression and -particularly- splicing alterations compared to Sudemycin C1.

Around 57% of the alternative splicing changes detected upon Sudemycin C1 treatment were also detected upon SF3B1 knockdown (Figure III.2B), while only 33% of SSA-induced changes were also observed upon SF3B1 knockdown (Figure III.2B). Remarkably, despite the structural similarity between SSA and Sudemycin C1 (Figure III.1A) and their similar biochemical mechanism of action (Figure III.1D), only 28% of the alternative splicing changes induced by SSA were also observed upon Sudemycin C1 and, similarly, only 5% of the events affected by Sudemycin C1 treatment were also observed upon SSA treatment (Figure III.2B). These results argue that structurally similar drugs can elicit a very different landscape of alternative splicing alterations.

Analysis of genes with alternative splicing changes and gene expression changes revealed a limited overlap between the two categories (Figure III.2C), consistent with the possibility that the drugs affect at least two different aspects of gene regulation, either as indirect consequences of the splicing changes or in independent ways, as previously reported (Convertini et al., 2014). At last, pathways analysis shows that SF3B inhibition induces alterations of cancer-related processes, including p53 signaling, RNA metabolism, apoptosis, cell cycle the spliceosome itself, among several affected ones (Figure III.2D), highlighting a potential role of these processes in the antitumor effects of the drugs.

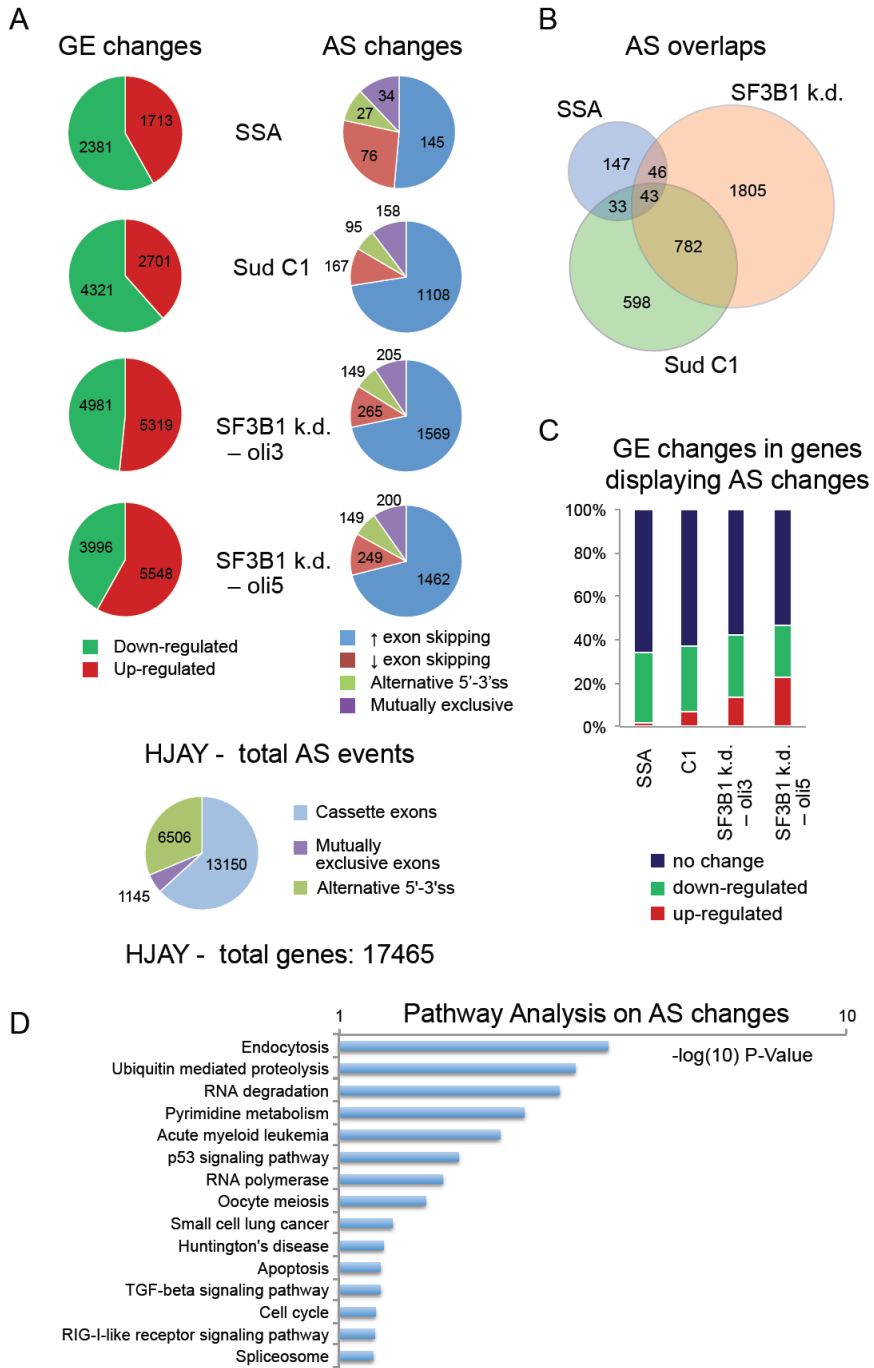


Figure III.2 (previous page). Evidence of differential Alternative Splicing regulation by Sudemycin C1 and Spliceostatin A. (A) Summary of Gene Expression (GE) and Alternative Splicing (AS) changes detected by hybridizing HJAY microarrays platforms with the different samples. Events were categorized in different classes of confidence, based on the concordance of different probes and their p-values. All events are summarized here, including high, medium and low confidence. Events with higher inclusion were generally of low confidence and prone to be artifacts, sometimes due to retention of neighboring introns rather than increased exon inclusion. Total numbers of events detectable with the HJAY microarray platform are also indicated. (B) Venn diagram showing the overlap of alternative splicing changes induced by Sud C1 (20 μ M, 8 h), SSA (260 nM, 3 h) or SF3B1 depletion (60 nM siRNA, 72 h) in HeLa cells. The union of the changes induced by the two SF3B1 siRNAs was considered for this comparison. The diagram refers to all the splicing changes types detected by the platform, without considering the direction and the intensity of the change. Exons affected by more than one type of event were counted just once for this analysis, so that total numbers of changes are a bit lower than in panel A. (C) Overlap of genes with AS changes and GE changes. The total refers to genes with detected splicing changes, of which some are also up- or down-regulated, and some have no change at the level of gene expression. (D) KEGG pathways analysis on genes affected at AS level in at least two different conditions among SSA, Sud C1 and SF3B1 k.d. Top 15 pathways are shown. Differential events were not sufficient to get differential pathways enrichments.

Validation of differential alternative splicing modulation by the two drugs

To validate the important conclusion that SSA and Sudemycin C1 can display both similar and distinct effects on alternative splicing, we carried out RT-PCR assays focusing on events predicted to display differences in each category. Thus, we could validate that skipping of *MCL1* exon 2 was similarly induced by the two drugs (Figures III.3A-B, top upper panel), while skipping of *PDCD10* exon 7 was more affected by Sudemycin C1 (Figures III.3A-B second panel from top), skipping of *ARRDC3* exon 3 was significantly more affected by SSA (Figures

III.3A-B, third panel from top), and in the case of *OGT* exon 7, SSA induced an apparent increase in exon inclusion while Sudemycin C1 induced exon skipping (Figures III.3A-B, lower panel). While *MCL1*, *OGT* and *ARRDC3* alternative splicing events were selected from changes detected with the microarrays, *PDCD10* was selected as previously reported target of SSA (Corrionero et al., 2011), further studied in Part II (although many other events with stronger response to Sudemycin C1 could be validated).

Taken together, the results of Figures III.1 to III.3 reveal that, despite their structural similarities and similar biochemical properties, SSA and Sudemycin C1 display a distinct profile of splicing effects, both in terms of the number of splicing changes elicited and in terms of the relatively limited overlap in their targets.

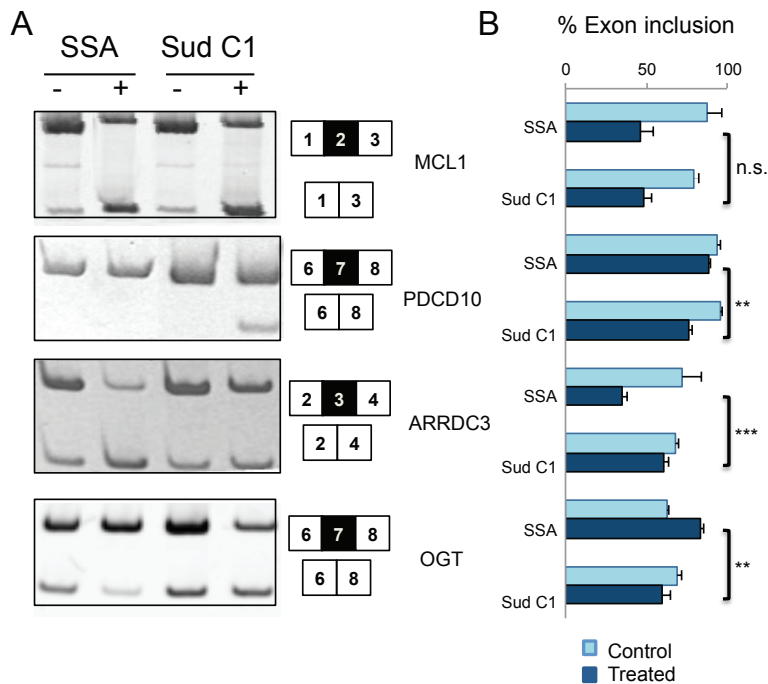


Figure 3. Validation of comparable and differential alternative splicing

events induced by SSA and Sudemycin C1. (A) RT-PCR products of a common target of Sud C1 and SSA (*MCL1*) and three differential ones (*PDCD10*, *ARRDC3* and *OGT*), run in polyacrylamide gel. (B) RT-PCR products were quantified by Image J.

Some drug differences are maintained in recently transcribed RNA

Considering that regulation of splicing isoforms in living cells is complex and their steady-state levels can be influenced by indirect effects (e.g. splicing-related changes in expression of splicing, mRNA transport or stability factors) as well as by effects of the drugs on various mechanisms of gene expression, we focused on the analysis of splicing effects in recently transcribed RNA. This was achieved by feeding cells with 5-Bromouridine (BrU) for the time of drug treatment (3 hours), isolating BrU-containing RNAs by IP (immunoprecipitation) with specific antibodies against this modified nucleotide and analyzing patterns of alternative splicing in this recently-transcribed RNAs by RT-PCR. Treatment concentrations were optimized to give a similar extent of *MCL1* alternative splicing regulation in 3 h, using this event as reference for calibrating the treatments (Figure III.4).

The results indicate that drug-induced *MCL1* exon 2 skipping is more pronounced in BrU pulse-labeled RNA than in steady state (Figure III.4), arguing that the drugs directly modulate *MCL1* splice site choice, at a comparable extent (Figure III.4).

With these treatment conditions, skipping of *OGT* exon 7 was induced by Sudemycin C1, but no change was detected with SSA (Figure III.4), while *PDCD10* transcript behaved similarly in both treatments (Figure

III.4). This result, contrasting with Figure III.3, suggests that some events (like *OGT* and *PDCD10*) might be very sensitive to drug concentration or influenced by other confounding factors. Indeed, several of these alternatively spliced isoforms (including those from *PDCD10*, *ARRDC3* and *OGT*) may undergo Nonsense-Mediated-Decay, which can contribute to differential stability of RNA isoforms. Nevertheless, differences in *ARRDC3* regulation by Sudemycin C1 and Spliceostatin A are consistent among several experimental conditions and titration experiments and are confirmed also by BrU-IP (Figure III.4).

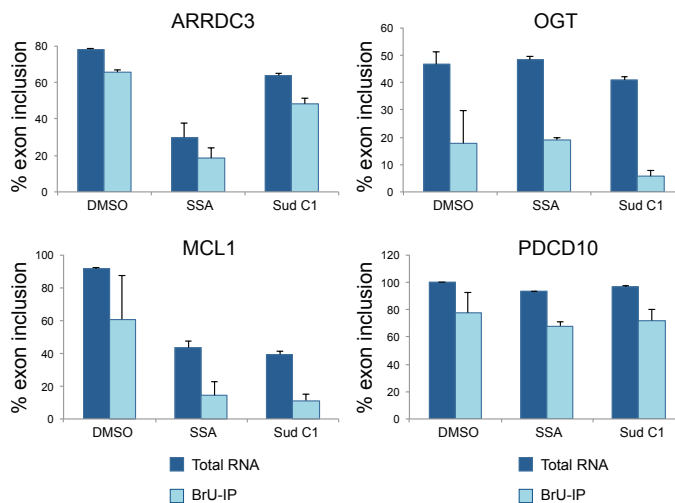


Figure III.4. Some drug differences are kept in recently transcribed RNA. RT-PCR of total RNA (Input) and recently transcribed RNA isolated by BrU RNA IP, run and quantified by capillary electrophoresis (Labchip). Treatments: 5 nM SSA, 10 μ M Sud C1 for 3 h in HeLa cells. Conditions were optimized to obtain a similar condition to start with, as seen by considering *MCL1* alternative splicing as reference event.

RNA-Seq analyses reveal distinct profiles of intron retention induced by different drugs

Given the restriction of splicing sensitive-microarrays to annotated alternative splicing patterns and the very limited detection of intron retention events, we analyzed drug-treated samples by RNA-Seq. In addition to the analysis of Spliceostatin A and Sudemycin C1, we also included Sudemycin K, the highly active structural variant described in Part IA.

Both total RNA and BrU-RNA were sequenced at a depth of 150 and 30 million reads/sample respectively and a list of retained introns was generated using VAST-TOOLS analysis software (Braunschweig et al., 2014). Heatmaps reflecting the extent of intron retention for each intron in at least one condition show clear clustering of Sudemycin C1 and Sudemycin K, while Spliceostatin A induces a higher number of changes, which are often clearly distinct to those induced by the other two drugs. Similar results are observed for BrU-RNA (Figure III.5A) and total RNA (Figure III.5B). Indeed, in both analyses, clusters of introns retained upon treatment with each of the drugs, as well as introns retained predominantly upon SSA treatment, are visible. The converse category (intron retention upon Sudemycin C1 and K and efficient splicing upon SSA treatment) is less prevalent but also detected (see heatmaps for each condition in Figure III.6).

The higher prevalence of intron retention detected with SSA might suggest that the treatment conditions with this drug were more drastic than for the other two drugs. It is important to point out, however, that each drug treatment was optimized to induce comparable effects in *MCL1* alternative splicing regulation and also similar cytotoxicity and

biochemical activities. Furthermore, the three drugs induced a similar number of exon skipping events (SANJUAN analysis detected around 4000 events for total RNA and 200-300 for BrU-RNA, with numbers coming from Sudemycin C1 or K treatments slightly higher than the ones corresponding to SSA).

Next, we compared sequence features of introns retained upon each treatment, or features of introns commonly retained by each of the three drugs (by Matt pipeline, as already employed in Part II). The results confirmed that retained introns are generally shorter, more GC-rich and contain weaker BPs compared to non-retained introns (Figure III.7), as reported in Part II for Sudemycin C1. However, these tendencies are less marked for SSA-regulated introns (Figure III.7), consistent with the more prevalent and extensive intron retention observed upon treatment with this drug. In contrast to Sudemycins C1 and K, SSA may simply display higher affinity for SF3B1 or induce a different conformation in the protein, less sensitive to the influence of neighboring sequences.

Taken together, these results support the idea that structural differences between drugs lead to distinct effects on the splicing machinery and consequently different splicing regulation outcomes. Consistent with this concept, Sudemycin K and Sudemycin C1, which differ exclusively in the replacement of an oxygen (ester -COO-) by a nitrogen-hydrogen (amide -CONH-) display far more similar effects than Spliceostatin A, which differs from Sudemycins by the presence of an oxygen in the cyclohexane moiety, two hydroxyl groups and by the presence of a number of additional chiral groups (Figure III.6C).

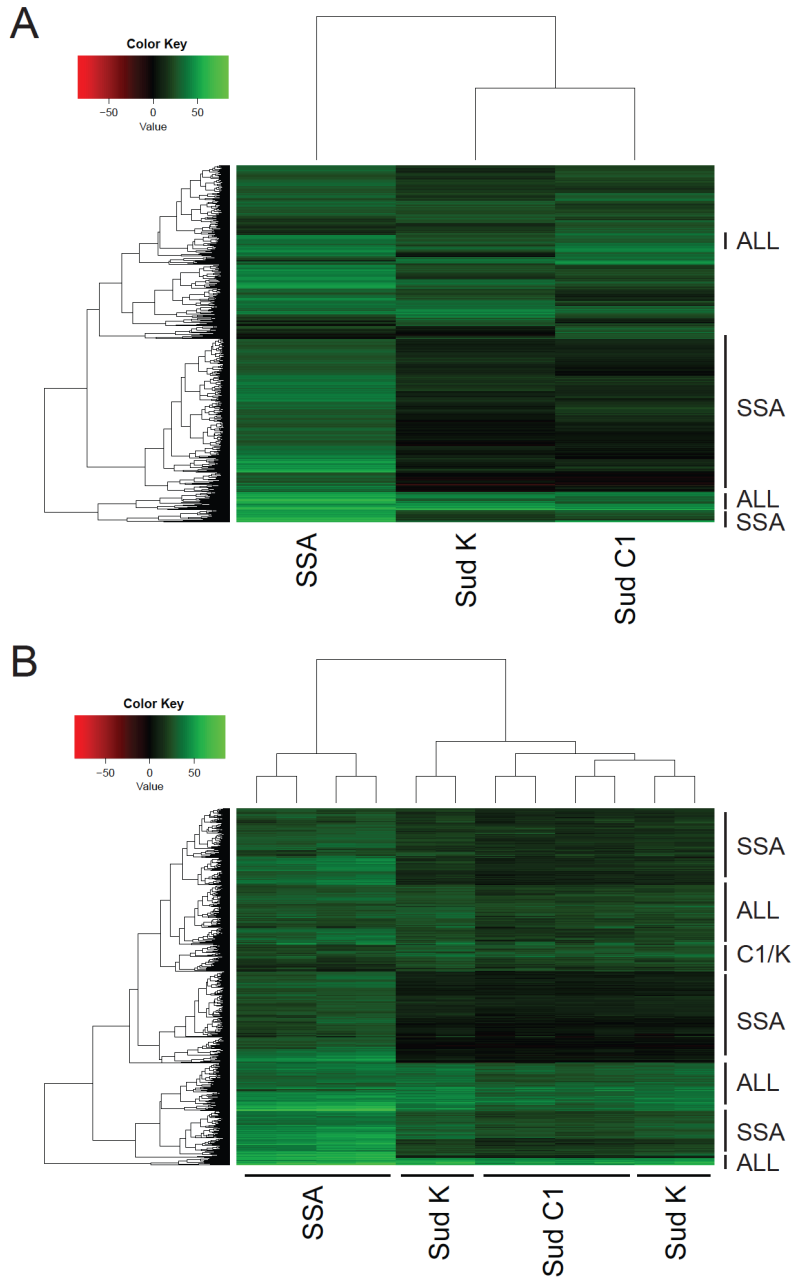


Figure III.5. Distinct profiles of drug-induced intron retention events by Spliceostatin A and Sudemycins C1 and K. Only introns with Delta PSI \geq 25% in at least one condition were considered for the analysis. Drug treatments: 5 nM SSA, 10 μ M Sud C1 and 2 μ M Sud K. All treatments were performed in parallel and lasted 3 h. (A) Heatmap from BrU-RNA data. Only

one value of intron retention per treatment is presented because the low coverage forced us to merge reads from duplicates. Categories of introns behaving similarly (“ALL”) or with distinct response to specific drugs (“SSA” or “C1/K”) are indicated to the right of the heatmaps. (B) Heatmap as in (A) for total RNA data. Intron retention changes involving different cross-comparisons between drug-treated replicas and control replicas are shown.

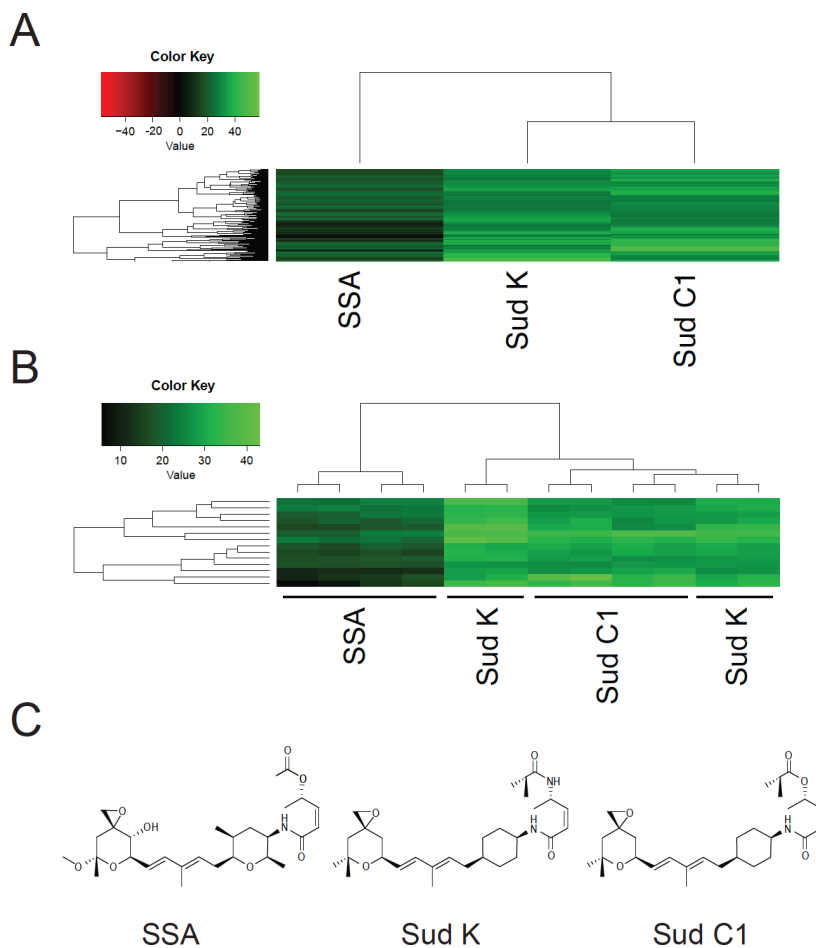


Figure III.6. Examples of distinct effects of Spliceostatin A vs Sudemycins on intron retention. Heatmaps corresponding to introns retained upon treatment with Sud K and Sud C1 and less affected upon treatment with SSA (with Delta PSI \geq 25% upon Sud K/C1 treatment and $<$ 25% upon SSA treatment) are shown for BrU-RNA (A) and for total RNA (B). (C) Comparison of drugs' structures.

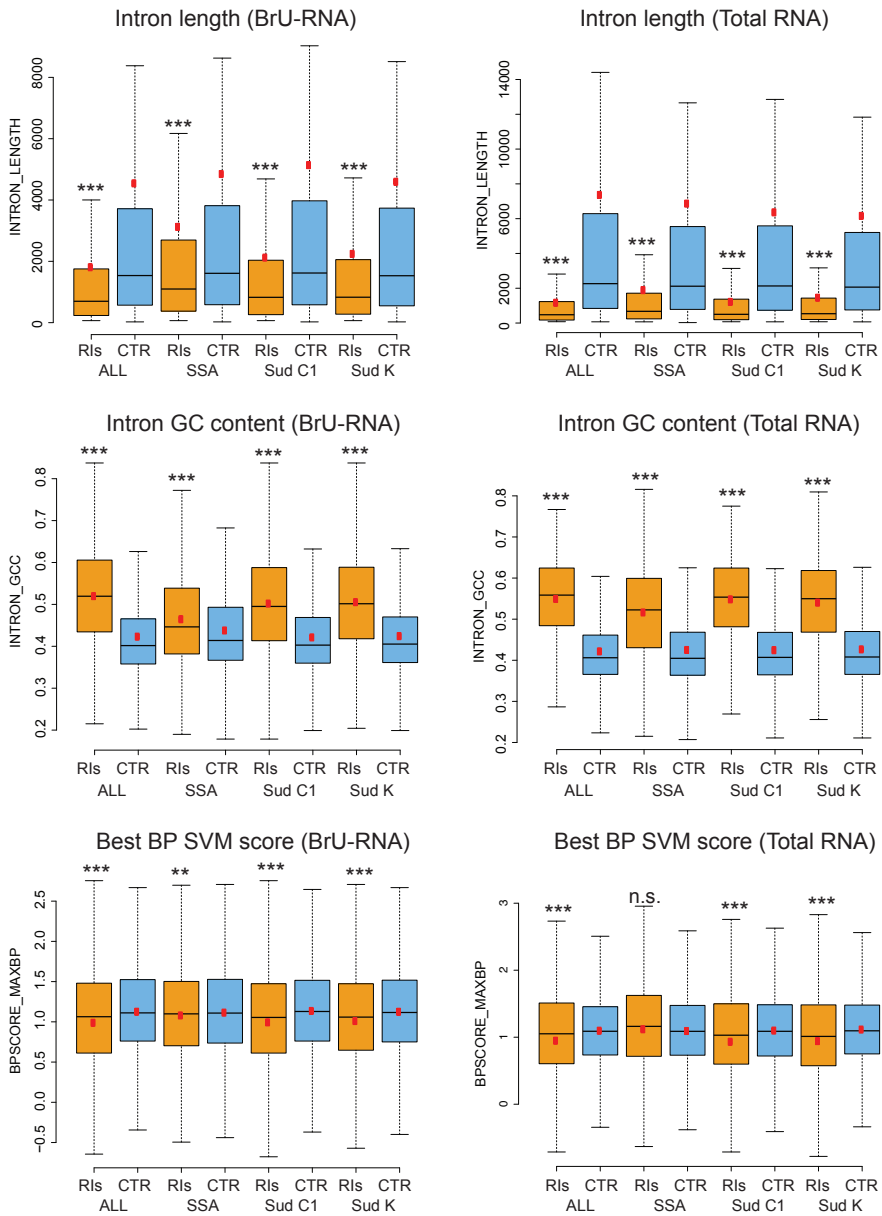


Figure III.7. Analysis of sequence features of drug-retained introns in BrU-RNA and total RNA using VAST-TOOLS and Matt. Boxplots corresponding to the indicated features are shown, with median values indicated by the black line and mean values by the red dot. Outliers were discarded. Introns retained upon each of the different treatments (“SSA”, “Sud C1” or “Sud K”) or in all of the treatments (“ALL”) were considered for the analysis. RIs: retained introns, with PSI difference $\geq 25\%$, CTR: non-differentially spliced introns, with absolute PSI difference $< 2.5\%$ (or 5% for “ALL” in BrU-RNA analysis). BP features and SF1 binding motif were analyzed in the 3' 150 nucleotides of each intron (introns shorter than 150 nt were not considered). Statistical significance was evaluated by permutation tests with 100.000 iterations. ***: p-value < 0.001 ; **: p-value < 0.01 ; n.s.: p-value > 0.05 . Because of the low sequencing coverage, reads from the two replicates were merged into one single dataset for the analysis of BrU-RNA. Number of retained introns from total RNA: 609, 3367, 827, 1424 from all treatments, SSA, Sud C1, Sud K respectively (454, 2744, 628 and 1114 for BP analysis); number of control introns from total RNA: 26462, 32526, 43885, 42862 for the four categories, down-sampled to 5000 (obtaining 4597, 4605, 4568 and 4574 introns for BP analysis). Number of retained introns from BrU-RNA: 2137, 11311, 4581 and 3862 from all treatments, SSA, Sud C1, Sud K respectively (1714, 9791, 3759 and 3196 for BP analysis); number of control introns from total RNA: 2098, 2999, 6197 and 7252 for the four categories (1860, 2671, 5564 and 6486 for BP analysis).

Materials and Methods

Drug treatments

Meayamycin was kindly provided by Dr. Kazunori Koide (University of Pittsburgh) Spliceostatin A by Dr. Melissa Jurica (UCSC), Sudemycins C1, E and F by Dr. Thomas Webb (St Jude's Children Hospital). These and other Sudemycin variants were also synthesized and provided by our collaborators Drs. Kamil Makowski, Fernando Albericio and Mercedes Álvarez (Universidad de Barcelona). TG003 and Cyclosporin A were purchased from Sigma (T5575 and 30024) and Isoginkgetin from Millipore (416154). Drugs were dissolved in DMSO and added in the culture medium or in the reaction mixes at the concentrations and times indicated for each experiment.

Cell culture

HeLa and NIH-3T3 cells were cultured in Glutamax Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and penicillin/streptomycin antibiotics (500 u/ml penicillin; 0.5 mg/ml streptomycin). Cells were maintained at 37°C in BINDER incubators under 5% CO₂.

According to the experimental needs, different seeding conditions were followed. In general, HeLa and NIH-3T3 cells were plated the day before the experiment in the following amounts: 800'000 in 6 cm dishes, 250'000-300'000/well in 6-wells plates, 80'000/well in 24-wells plates, 30'000/well 48-wells plates, 10'000/well in 96-wells plates for HeLa cells and 400'000/well in 6-well plates, 100'000/well in 24-wells plates for NIH-3T3 cells.

siRNA and plasmids transfection

For RNA silencing experiments, cells were reverse transfected with specific siRNAs. Briefly, cells were added to a previously incubated mixture containing the siRNA of interest (for SF3B1 depletion: 60 nM stealth siRNA #3, 5'-GACAGCAGAUUUGCUGGAUACGUGA-3' or stealth siRNA #5, 5'-CCCUGUGGCAUUGCUUAAUGAUAU-3' from Invitrogen; for MCL1 depletion: 50 nM Mission siRNA SASI_Hs01_00162657-5'-GUGUUAAGAGAAGCACUAA[dT][dT]-3' from Sigma-Aldrich) and Lipofectamine RNAiMAX (Life

Technologies) following the manufacturer's recommendations. Plasmids were reverse transfected following a similar procedure, but using Lipofectamine 2000 reagent (Life Technologies). RNA was extracted 72 and 24 h post-transfection, respectively. Before samples collection, cells were washed once with PBS and frozen.

RNA extraction and RT-PCR

mRNA was extracted using oligo-dT-coated 96-well plates (mRNA catcher PLUS, Life Technologies) and total RNA using Maxwell RNAeasy kit (Promega), following the manufacturers' instructions.

One third of purified poly(A) mRNAs was reverse transcribed in a 40 μ l volume with 0.25 μ l of Superscript III (Life Technologies), while 100 ng of total RNA were reverse transcribed in a 20 μ l volume and 0.5 μ l of Superscript III. Reverse transcription was carried out with 2.5 μ M oligo-dT (Sigma-Aldrich) and 250 ng of random primers (Life Technologies) following the manufacturer's recommendations.

PCR reactions were carried out using GoTaq enzyme (Promega) in 25 μ l reactions with 1-5 μ l of previously synthesized cDNA and 30-40 cycles of amplification (final reactions contain 1X GoTaq Green Reaction buffer, 1.75 mM MgCl₂, 1 μ M primers, 100 μ M dNTPs, 0.5 u of enzyme). PCR cycling was performed in Dyad 2 thermal cyclers (Bio-Rad) with the following parameters: initial incubation at 95°C (3'), 30-40 cycles of 95°C (30''), 60°C (30''), 72°C (60-90'') and a final incubation at 72°C (1').

Primers for semi-quantitative RT-PCR amplification were designed to have a melting temperature around 60°C and to amplify endogenous targets in a range between 120 and 700 bp (primers are listed in Table 1). They were designed to anneal to constitutive exons flanking the alternative one. For amplification of minigene-specific transcripts, oligonucleotide primers complementary to transcribed regions within the vector were used (PT1 and PT2, specified in the primers list). In these cases, specific amplification of endogenous transcripts was achieved using primers complementary to regions not included within the minigene (e.g. constitutive exons adjacent to the sequence included in the minigene).

Cytotoxicity assays

1'000 or 2'500 HeLa cells/well were seeded in 96-wells transparent plates the day before treatment with drugs, control vehicle or any other experimental condition to be tested. 20, 44 and 68 h after treatment, cell medium was replaced with medium containing 10 μ M Resazurin and

cells were incubated at 37°C for 4 h. Fluorescence was measured with an Infinite 200 PRO series multi-plate reader (TECAN) with 530 nm and 590 nm, as fluorescence excitation and emission wavelengths, respectively. Measurements were taken from the bottom of the plate, with an optimal gain and one value per well.

IC₅₀ values were calculated from the dose-response curve analysis using GraphPad Prism 6.01 by interpolation of 50% viability values.

Minigene transfection assays

Genomic sequences of interest were cloned under a Cytomegalovirus Promoter in a pCMV56 expression vector (Clontech) between KpnI and NotI restriction sites. Inserts span the whole regulated cassette (upstream exon and intron, regulated exon, downstream intron and exon and possibly 25 additional intronic nucleotides) flanked by two specific sequences (PT1 upstream of the cassette and PT2 reverse complement downstream), which serve to amplify minigene-specific transcripts with PT1 and PT2 primers because they do not amplify any sequence from the mammalian cells lines used. The middle part of very long introns was deleted for improving cloning efficiency.

Mutations were introduced by Gibson cloning (Gibson et al., 2009), following the instructions from Gibson Assembly Master Mix kit (NEB). In short, overlapping PCR amplicons containing the mutation at their boundaries (which was included as a mismatch in the amplification primers) were inserted into a vector prepared by restriction enzyme digestion or by PCR, containing overlapping flanking sequences. Gibson reaction master mixes, provided by the CRG Protein Technologies Unit, were then used: briefly, 50-100 ng of vector and 1-5 µl of DpnI-digested, unpurified PCR-products were mixed with 10 µl of 2X Gibson mix to get a final reaction volume of 20 µl. Reactions were incubated for 1 h at 50°C, shortly frozen and transformed in XL1-Blue competent bacteria. Assays were performed using HeLa cells seeded in 48-wells plates (30'000 cells/well) or 96-wells plates (10'000 cells/well) and transfected with 3 or 1.5 ng of minigene/well and 0.2 or 0.1 Lipofectamine 2000/well in 200 or 100 µl Opti-MEM/well, respectively. 100 ng/well of pBluescript plasmid were co-transfected with the minigenes to increase transfection efficiency. NIH-3T3 cells, were transfected in 24-wells plates (with 25 ng of minigene, 500 ng of pBluescript and 0.5 µl of Lipofectamine 2000 per well). Drug treatments were performed after ON transfection in 200, 100 or 50 µl of total medium for 24-, 48- ad 96-wells plates respectively.

Polyacrylamide gel electrophoresis (PAGE) and capillary electrophoresis

PCR products were resolved by vertical electrophoresis in 6% non-denaturing polyacrylamide gels (gels size: 8.3 x 7.3 x 0.075 cm) run in 1X TBE (89 mM Tris base, 89 mM Boric acid, 2 mM EDTA) at 200 V for 25'. Gels were stained with 1X GelRed stain (Biotium) and visualized with a GelDoc transilluminator (Bio-Rad). Band intensities were quantified using ImageJ software (<https://imagej.nih.gov/ij/>) and subtracted of the blank value coming from an empty part of the image. Alternatively, high throughput capillary electrophoresis measurements for the different splicing isoforms were performed using a Labchip GX Caliper workstation (Caliper, Perkin Elmer) and a HT DNA 5K LabChip chip (Perkin Elmer). Briefly, PCR products were diluted in a minimum of 40 μ l in twin.tec 96-wells plates (Eppendorf) and separated with high sensitivity and resolution in a semi-automatized way. In this case, bands were quantified based on estimation of their nanomolar content with LabChip GX software (version 3.0).

In Vitro Transcription

DNA templates for in vitro transcription were generated by PCR, separated in 1.5% agarose gels containing 1X GelRed (run in TBE 1X at 100 V for 45') and gel-purified with QIAquick Gel Extraction Kit (Qiagen). PCR forward primers contain a T7 promoter followed by three Gs to enhance transcription efficiency at the beginning of the transcript (gctaatacgaactcactataggg), followed by the transcript-specific forward primer sequence (primers are listed in Table 1).

Radiolabeled RNAs were transcribed using T7 RNA Polymerase (Promega) under the following conditions: 4 μ l 5X Transcription Buffer, 2 μ l DTT, 0,5 μ l (20 u) RNase inhibitor RNasin (Promega), 4 μ l 2,5 mM ATP/GTP/CTP, 2,4 μ l 100 μ M UTP, 200 ng template DNA, 1 μ l enzyme (20 u) and 1-5 μ l of radiolabeled alpha-32P-UTP. Reactions were performed in a final volume of 20 μ l and incubated at 37°C for 2 h. RNAs were purified by gel filtration using Sephadex G-50 columns (GE Healthcare).

After quantifying the cpm/ μ l using a liquid scintillator (Beckman beta counter), the RNA specific activity was calculated using the following formula:

$$\text{cpm/fmol} = (\mu\text{Ci/pmol}) \times ([\text{radioactive UTP}]/[\text{cold UTP}]) \times (\text{number of Us}),$$
 where the ($\mu\text{Ci/pmol}$) is the specific activity of the radionuclide used and the cpm were estimated to correspond approximately to 1 million per μCi .

Non-labeled RNAs were transcribed similarly but with 4 μ l of 2.5 mM mix of the four NTPs.

Products were mixed with a 2X loading dye (95% formamide, 0.025% xylene cyanol, 0.025% bromophenol blue, 18 mM EDTA, and 0.025% SDS) and boiled for 1' at 90°C.

Products up to 500 nt were checked in denaturing 8% urea-polyacrylamide vertical gels (gels size: 8.3 x 7.3 x 0.075 cm) run in 1X TBE at 180 V for 30'. Larger products were separated in 1% agarose gel (run in TBE 1X at 100 V for 45'). Cold transcripts were visualized by GelRed staining, while radioactive transcripts were visualized by Phosphorimager screens (after a drying step for agarose gels).

Spliceosome A3' complex formation assay

This assay was carried out as previously described (Corrionero et al., 2011). Each splicing reaction contained 1 μ l of RNA mix (premix for 10 reactions: 100 fmol RNA (corresponding to transcripts covering the 3' end of intronic regions and some nucleotides of the downstream exon), 4 μ l creatine phosphate 0.5M, 1 μ l ATP 100 mM, 1 μ l MgCl₂ 270 mM, up to 10 μ l with Buffer D with 0.1M KCl and 1mM DTT (freshly added), 3 μ l of HeLa nuclear extracts (Cilbiotech), 1 μ l of drug or DMSO, 2 μ l of 15% polyvinyl alcohol, prewarmed at 30°C, and Buffer D 0.1M KCl up to 9 μ l, to obtain standard splicing conditions (3 mM MgCl₂, 1.1 mM ATP, 22 mM creatine phosphate, 1.67%polyvinyl alcohol). ATP and creatine phosphate were replaced with Buffer D 0.1 M KCl for the -ATP control.

Buffer D contains 20 mM HEPES-KOH pH 7.9; 0.2 mM EDTA, 20% glycerol, 1 mM DTT, 0.01% NP40, complemented with 0.1 M KCl and 1 mM DTT (freshly added).

The reactions were set up in a 48-wells microplates and incubated at 30°C for 15'. Subsequently, 5 mg/ml heparin and 2.2 μ l of 6x DNA loading dye (20 mM Tris-HCl at pH 7.5, 0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol) were added and the reactions were incubated for 10' at room temperature. The products were subsequently loaded on a 1.5% low-melting agarose (Invitrogen) gel in 50 mM Tris base and 50 mM glycine buffer for 90' at 4°C and run at 75 V. Gels were fixed in 10% methanol and 10% acetic acid for 10' at room temperature, dried for at least 3 h at 50°C and exposed overnight to a Phosphorimager screen. The intensity of the band corresponding to complex A3' over the signal of the whole well was measured by ImageJ and normalized to the control condition (after subtracting the background signal from each measurement).

Microarray analyses

Quality control by Bioanalyzer (Agilent) was performed to ensure that the RNA integrity number (RIN) was higher than 7.6. RNAs from triplicates of each condition were hybridized to Affymetrix Human Exon Junction Arrays (HJAY). These platforms carry 5'235'274 probes gathered in 315'137 exon probesets and 260'488 junction probesets. They cover 13'150 cassette exons, 6'506 alternative 5'/3'ss, 1'145 mutually exclusive exons. RNA hybridization and data analysis were performed by GenoSplice (Paris).

BrU-IP and RNA-Seq analysis

HeLa cells (0.8 million in a 6 cm plate) were co-treated with 2 mM BrU (5-Bromouridine, Sigma-Aldrich) and 10 μ M Sudemycin C1 for 3 hours. RNA was extracted with the Maxwell RNAeasy kit (Promega). Total RNA was kept as input and for standard RNA sequencing. 10 μ g of it were immunoprecipitated with a specific anti-BrU mouse antibody (B2531, Sigma), as previously described (Lin et al., 2008) but without tRNA addition.

Briefly, 2 μ l of antibody were nutated with 20 μ l Protein G Dynabeads (Invitrogen) per sample for 1 h at 4°C in 1 ml RSB-100 buffer (10 mM Tris-HCl, pH 7.4, 100 mM NaCl, 2.5 mM MgCl₂ and 0.4% (v/v) Triton X-100). After three washes of the beads with 1 ml of RSB-100 buffer, beads were resuspended in 150 μ l RSB-100 with 40 U RNasin (Promega). RNA was incubated for 1' at 80°C, added to the beads and nutated for 1 h at 4°C. After three washes with 1 ml of RSB-100 buffer, the RNA bound to the beads was eluted by direct addition of 300 μ l RLT buffer (Qiagen RNeasy mini kit). The mix was heated at 80°C for 10' and the supernatant was purified with the RNeasy mini kit (Qiagen). RNA-Seq was performed at CRG Genomics Unit by Illumina HiSeq. Stranded mRNA-seq libraries were prepared and 2 samples/lane were sequenced with 2x125 bp paired-end reads (to get around 150 million read pairs per sample). Duplicates were sequenced for each condition. For BrU-IP RNA, no poly(A) selection was performed before library preparation and the 8 samples were run in the same sequencing lane with 2x125 bp paired-end reads (to get around 30 million read pairs per sample). Duplicates were sequenced for each condition.

Bioinformatic analyses

Overlap representations and Venn diagrams were obtained through R (VennDiagram package) or with the online Venny tool:

<http://bioinfogp.cnb.csic.es/tools/venny/index.html>.

Splicing analyses of RNA-Seq data were performed using SANJUAN (Splicing Analysis & Junction Annotation tool) developed in our group. Unlike other splicing analysis tools, this pipeline is not limited by existing annotation and can identify novel events absent from transcript structure databases, improving the analysis of conditions that cause widespread aberrant splicing:

<https://github.com/ppapasaikas/SANJUAN>.

SANJUAN thresholds are the following for MC (medium confidence) and HC (high confidence) analyses for alternative splicing analysis, which is based on junction reads:

	HC	MC
Minimum delta PSI	15	10
Minimum number of junction reads	7	5
Minimum ratio neighbor reads/junction reads:	0.05	0.01
Minimum ratio junction reads/neighbor reads:	0.004	0.002
Maximum length of junction:	100'000	100'000
Minimum length of junction:	50	50
Minimum ln of reads fold change:	0.1	0.05
Maximum p-value of Hypergeometric test:	0.001	0.01

Intron retention analysis is based on the amount of intronic reads compared to neighboring regions and corroborated by a concordant change in the corresponding junction. Threshold is set to 0.92 ln fold change (corresponding to a 2.5 fold change).

As independent approach, splicing analysis was performed with VAST-TOOLS (Vertebrate Alternative Splicing and Transcription Tools): <https://github.com/vastgroup/vast-tools> (Braunschweig et al., 2014) and downstream analyzed with Matt, a toolkit for analyzing biological sequences and alternative splicing events being developed from André Gohr (CRG).

Gene expression analyses of RNA-Seq data were performed with Cuffdiff (Trapnell et al., 2013).

Gene Ontology analyses were performed using GOrilla (Gene Ontology enRIchment anaLysis and visualizAtion tool) (Eden et al., 2009):

<http://cbl-gorilla.cs.technion.ac.il/>.

Branch point sequence prediction and scoring were performed using the SVM-BPfinder tool (Corvelo et al., 2010):

<https://github.com/RegulatoryGenomicsUPF/svm-bpfinder>.

Splice Sites strengths were evaluated using MaxEntScan tools for 3' and 5' splice sites scoring (Yeo and Burge, 2004):

http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html;

http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html.

Sequence logos, intersections and overlaps of genomic regions (e.g. between mapped BPs and affected regions) were performed through Galaxy platform (<https://usegalaxy.org/>) (Afgan et al., 2016), WebLogo (<http://weblogo.berkeley.edu/logo.cgi>) (Crooks et al., 2004) and Two Sample Logos (<http://www.twosamplelogo.org/index.html>) (Vacic et al., 2006).

Heatmaps were obtained through heatmap.2 function from R package gplots.

Branch point Mapping from total RNA

The method was adapted from previous descriptions of branch point (BP) mapping using RT-PCR across the intronic lariat (Conklin et al., 2005; Vogel et al., 1997). cDNA was synthesized using 500 ng - 1 µg of total cellular RNA with random hexamers and Superscript III reverse transcriptase. PCR amplification was performed using a forward primer annealing towards the 3' end of the intron (but 5' of potential BPs) and a reverse primer complementary to the 5' end of the intron, in order to get products only if amplification occurs across the 2'-5' lariat structure (primers are listed in Table 1). 30 cycles of PCR were performed with 2 µl of cDNA and 15'' of elongation time. Products were diluted 1:500, re-amplified with nested primers (following the same PCR conditions) and loaded on 6% polyacrylamide gels. Bands of interest were excised from the gel, eluted by adding 20 µl of water and shaking at 50°C for 30'. Supernatant was used for sequencing reactions using BigDye mix and the primers of the second PCR (nested primers). Results were analyzed with CLC Main Workbench software.

Target	Aim of the Primer Pair	Forward primer	Reverse primer	Products size (bp)
ARRDC3	Exon skipping assay	AATTCGGAAGAAGGCTTCCA	GTATAGCCCTTCCTTTCAAT	320 - 172
OGT	Exon skipping assay	ATGTCTTGAAAGAGGCACGC	TACGATACAAGCGAACTGCCT	357 - 161
PDCD10	Exon skipping assay	CCTAAACGAAAAGGCACGAG	CAGAGTATCACTGAAACTTTTGG	229 - 150
FN1	Exon skipping assay	GGCCTGGAGTACAATGTCAGT	CATGGTGTCTGGACCAATGT	405 - 132
MCL1	Exon skipping assay	AGACCTTACGACGGTTGG	ACCAGCTCCTACTCCAGCAA	401 - 153
MCL1 - minigene	Minigene AS	GTCGACGACACTTGCTCAAC	AAGCTTGCATCGAATCAGTAG	488 - 230
MCL1 - endogenous	Endogenous AS in minigenes assay	ATCTGGTAATAACACCAGTACGGAC	ACCAGCTCCTACTCCAGCAA	586 - 338
mMCL1 - minigene	Minigene AS	GTCGACGACACTTGCTCAAC	AAAGCCAGCAGCACATTCT	641 - 393
PDCD10 - minigene	Minigene AS	GTCGACGACACTTGCTCAAC	AAGCTTGCATCGAATCAGTAG	350 - 271
PDCD10 - endogenous	Endogenous AS in minigenes assay	CTTCGTATGGCAGCTGATGA	CAGAGTATCACTGAAACTTTTGG	294 - 215
PDCD10 3'	T7 template for <i>in vitro</i> transcription	<u>GCTAATACGACTCACTATAGGGGC</u> GCTTAACATAATTAAGAGTGT	CTAGCAAGGCTTCTACTAAACGTAC C	238 (wt), 157 (ΔE1-E2-E3)
AdML 3'	T7 template for <i>in vitro</i> transcription	<u>GCTAATACGACTCACTATAGGGTGA</u> TGATGTCATACTTATC	CCCCTGGAAAGACC CGGAAGA	81
AdML full length	T7 template for <i>in vitro</i> transcription	<u>GCTAATACGACTCACTATAGGGAAT</u> ACACGGAATTCGAGCTCG	CCCCTGGAAAGACC CGGAAGA	219
MCL1 - external primers	BP mapping	TTTTGGAATGGCAGCTCTT	GTGAGTCCGGGGAGAGATG	<239
MCL1 - nested primers	BP mapping	AGGGTGGGATGCAATTTCA	AAAAAGGGAGTGAGGCCTTG	<183
PDCD10 - external primers	BP mapping	TAAATCCCCACTCCAACCA	TCTGAAACCAAACGCCATAA	<355
PDCD10 - nested primers	BP mapping	TGCGCTTAACATAATTAAGAGTG	TCGGAAGTACTTTTAAGAAAAGAAG AA	<183

Table MM.1. List of primers used in this study. For minigene experiments, endogenous-specific primers are designed to anneal to regions that are not included in the minigene. PT1 and PT2 primers are used for minigenes-specific PCRs, with the exception of *mMCL1*, for which PT1 is combined with a primer complementary to *mMCL1* exon 3 to avoid aspecific bands. Primers for amplification of templates for *in vitro* transcription contain a T7 promoter sequence followed by two Gs at their 5' end (underlined in the table).

Discussion

1.A. Structural requirements of SF3B-targeting drugs

In collaboration with Kamil Makowski, Fernando Albericio and Mercedes Álvarez (University of Barcelona), we performed structure-activity studies of new Sudemycin variants, with the aim of expanding the current understanding of their pharmacophore and identifying active compounds that might be easily conjugated with other chemical moieties.

We found that some structural flexibility in the stereochemistry of the conjugated diene is allowed. While at significant cost in activity, *ZEZ* stereoisomers provide ample possibilities for further modification through Diels-Alder reactions. Most important, derivatives replacing the key pharmacophore oxycarbonyl by an amide display improved activity compared to previously reported Sudemycins, both in biochemical as well as in cellular assays. While the source of the observed differences requires further studies, improved drug solubility and stability -by resistance to diesterases known to be important for determining the half life of this class of compounds (Laizure et al., 2013)- are likely to contribute, as discussed in Part IA. Higher activity in biochemical assays also suggests improved affinity for the target and/or improved capacity to affect conformational changes in the target (Effenberger et al., 2016c) (Part IA, Figure 2).

Our findings can pave the way to the synthesis of other variants with improved activities as well as to more refined drug-target interaction studies (that were so far technically limited because of the lack of robust

experimental set-ups). Advances in the structural understanding of SF3B complex (Rauhut et al., 2016; Yan et al., 2016) will be also instrumental to provide rational basis for the design of compounds with improved activity and selectivity, by modulating the affinity as well as different ways of interacting with SF3B1, as we speculate in Part III.

1.B. Drug-ASO conjugates

The interesting observation that drug-ASO conjugates have increased effects *in vitro* confirms the hypothesis that sequence-specific delivery of splicing inhibitors is achievable. While several studies aim at the identification of transcript-specific molecules (Naryshkin et al., 2014; Palacino et al., 2015), our results suggest that the specificity of general inhibitors can be greatly improved by conjugation, and dictated by the sequence of the conjugated ASO. This idea was recently explored for modulating *SMN2* splicing by conjugating ASOs with a phosphatase inhibitor (Kwiatkowski et al., 2016), but results showed limited efficacy. Possibly, the strategy of regulating a phosphatase involved in splicing regulation (Kwiatkowski et al., 2016) may be less direct and therefore less efficient than targeting a *core* spliceosome component (like we tested in Part IB). We cannot exclude, however, that conjugation of a small molecule to the ASO in the proximity of the BP can enhance inhibition simply by steric obstruction of U2 snRNP recruitment, without the need of any pharmacological specificity for its components. Conjugation with inactive or unrelated compounds would be an appropriate control for ruling out this point.

In any case, our initial results argue that careful design is required, because the distance between BP and ASO-complementary region

greatly affects the conjugates' efficiency. Based on this, it seems advisable to test several alternative options, including the conjugation of the drug at the 3' end of the ASO instead of at the 5' (as we tested). In this case, the ASO sequence would target the region downstream of the BP, so that the 3' conjugation would bring the drug in proximity of the BP. This approach may have the advantage of targeting additional contacts of SF3B1 with the pre-mRNA (Gozani et al., 1998; Rauhut et al., 2016; Yan et al., 2016), as well as of targeting an important *cis*-acting splicing signal like the Py-tract, but the reduced sequence diversity of these sequences may complicate the design of ASOs of unique specificity.

It is even conceivable that the conjugation of SF3B inhibitors with RNA molecules would facilitate or expand drug-target interactions, for example if the drugs target positively charged residues within SF3B1 HEAT repeats that play an important role in protein-RNA contacts (Rauhut et al., 2016; Yan et al., 2016). On the basis of these recent structural findings, it can even be speculated that adding a 20 nucleotide RNA sequence to a SF3B1 inhibitor could contribute to increase inhibitory effects.

Finally, improving compounds' stability and pharmacokinetics remains a key goal to improve and facilitate *in vivo* assays.

Although we focused on the model Adenovirus Major Late promoter as a proof of principle, the strategy should be tested for cancer-related targets within the human transcriptome. For example, *NUMB*-regulating ASOs affect lung cancer cell proliferation (Bechara et al., 2013) and *NUMB* alternative splicing is affected by SF3B-targeting drugs as well (Papasaikas et al., 2015) (Figure 5, Part II), making this

event an interesting candidate to target using this strategy. *MCL1* could also be an interesting event to test, based on its strong response to SF3B1 inhibition (Papasaikas et al., 2015) (investigated in Part II) and its important role in cancer cells (Larrayoz et al., 2016; Opferman, 2016).

Finally, similar conjugation strategies could possibly be applied to other molecules and mechanisms, including other RNA processing mechanisms for which inhibitors are available, that could also be modulated in a transcript-specific fashion by ASO-drug conjugation.

2. Sudemycin-responsive and Sudemycin-resistant sequences

Several studies have used SF3B-targeting drugs as general splicing inhibitors and have reported increased intron retention events (Kaida et al., 2010; Martin et al., 2016; Martins et al., 2011; Nojima et al., 2015). Nevertheless, results presented in this thesis suggest that at low drug concentrations, splicing inhibition does not cause a general shut down of splicing (that would lead to retention of every newly transcribed intron). Instead, SF3B inhibition can be associated with alternative splicing regulation, as previously reported (Convertini et al., 2014; Corriero et al., 2011), although it seems likely that certain level of intron retention coexists with differential effects in splice site selection. These observations are consistent with the hypothesis that a partial inhibition of SF3B complex regulates alternative splicing of some transcripts more than others, and our goal was to evaluate whether primary RNA sequence can be responsible for these differential effects.

Our findings show that specific sequence elements can determine the relative sensitivity of transcripts to drug treatment, with dramatic

differences associated with even single (or few) nucleotide changes. In particular, we correlate drug resistance with the presence of possible alternative functional BP sequences located 5' of the branch site. Previous results from our group and others argued for drug-dependent destabilization of U2 snRNP recruitment and drug-induced allowance to recognize sequences 5' of the branch site by U2 snRNA base-pairing leading to formation of splicing non-productive complexes at BP decoy sites (Corrionero et al., 2011; Folco et al., 2011). It is conceivable that the drug occupies a space within the SF3B complex and induces a conformation that might favor contacts with sequences upstream of the normal branch site, as reported for Spliceostatin A (Corrionero et al., 2011). The presence of more than one sequence with base-pairing potential with U2 snRNA that could serve as a functional branch site 5' of the normal one could warrant a functional interactions even if the drug causes ambiguity in BP recognition (Figure D.1).

Similarly, mutations in SF3B1 induce the usage of upstream BPs with A-rich motifs, associated with splicing to upstream cryptic 3'ss (Alsafadi et al., 2016; Darman et al., 2015). Although the consequences of drug treatment and SF3B1 mutation seem to be different, both situations may have in common the disabling of proofreading functions that ensure U2 snRNP recruitment to the normal BP under standard conditions.

Another relevant observation is that weak BPs increase the sensitivity of a 3'ss to the drug (Figure D.1), indicating that the extent of base-pairing between U2 snRNA and the BP region is an important determinant of the efficacy with which the drug alters BP recognition. Indeed, drug sensitivity of a highly sensitive substrate was lost when replacing its BP

by a consensus (yeast-like) BP, able to establish all possible base-pairs with the BP recognition sequence in U2 snRNA (Figure II.6). Sequences around the BP could also influence the regulation, as they play a role in U2 snRNP recruitment through better or worse interaction with SF3B proteins, which are known to bind the anchoring site 5' of the BP and/or to interact with U2AF65. Indeed, a strong Py-tract also contributes to decrease drug effects, as shown by comparing the responses to the drug of the human and mouse *MCL1* (Figures II.7, II.13, II.14 and II.15). U2AF65's more stable binding to strong Py-tracts may therefore help in the recruitment of drug-destabilized U2 snRNPs, since SF3B1 directly interacts with U2AF65 (Gozani et al., 1998).

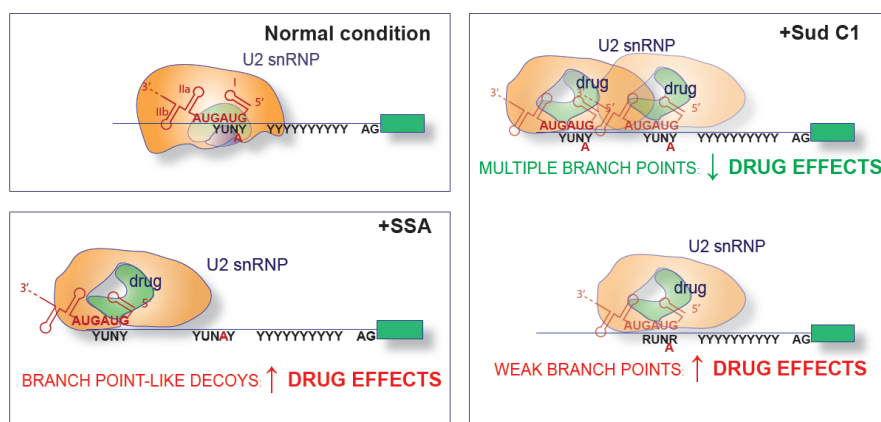


Figure D.1. Model of U2 snRNP recruitment upon Sudemycin C1 treatment. Sequences around the BP can influence the sensitivity of transcripts to the treatment: “decoy” BP-like sequences sequester U2 snRNA from functional interactions and lead to increased inhibition, multiple BP consensus sequences favour functional U2 snRNA–BP interactions and therefore decrease the inhibitor effects, while weak BPs disfavour the interactions and increase the inhibitor's effects.

Once again, structural studies -which may become possible in the near future thanks to recent progress in Cryo-electron microscopy (Rauhut et al., 2016; Yan et al., 2016)- may be instrumental to define the molecular

basis for the relative contributions of the multiple interactions that occur during U2 snRNP assembly and how they are affected by drug binding.

Interestingly, recent evidence argues for a role of pseudouridines within U2 snRNA in modulating the activity of Prp5 helicase and BP recognition (Wu et al., 2016). Prp5 was already hypothesized to be an important factor involved in the drug effects (Corrionero et al., 2011; Folco et al., 2011) and these results highlight the interest of further exploring the mutual influence between this helicase and SF3B1 and the possible interplay with the drugs.

Our conclusions are also consistent with some observations reported by other groups. For example, a mutually exclusive exon event affects the ketohexokinase (*KHK*) gene, involved in fructose metabolism. Recent evidence described that hypoxia can induce SF3B1 overexpression in cardiomyocytes, leading to a change in *KHK* mutually exclusive isoforms and therefore in cells' metabolism, with relevant consequences for heart disease (Mirtschink et al., 2015). The SF3B1-induced isoform has a stronger BP than the competing one, suggesting that alternative splice site choice can be determined by the levels of SF3B1 in the cell, such that the interplay between limiting factors for U2 snRNP recruitment is changed under different cellular conditions, with consequences for alternative splice site utilization.

Moreover, *in vitro* splicing assays showed a different sensitivity of different transcripts to SF3B1 pharmacological inhibition: although the strength of BP and Py-tract were not found to correlate with sensitivity to Spliceostatin, titration of Pladienolide did have a different effect on model RNAs containing different BPs, with weaker ones being more

sensitive and stronger ones more resistant to the effects of the drug (Effenberger et al., 2016c), consistent with our observations using Sudemycin C1. Taken together, these results argue that different drugs differentially alter the balance between factors and interactions involved in complex A formation, leading to diverse effects in splice site selection and alternative splicing outcomes.

We cannot exclude that other features, like RNA stability or the more or less efficient export from the nucleus to the cytoplasm can also play an important role in determining the steady state accumulation of splicing isoforms, and even differential protein stability can influence the biological effects. For example, MCL1 is very sensitive to translation and transcription inhibition because of its fast turnover (Wei et al., 2012). We also observed defective splicing for other short-lived small oncoproteins, possibly related to the higher toxicity of splicing inhibitors in tumor cells versus normal cells.

Even other factors, like chromatin environment during transcription, may influence the differential effects of the drugs. Indeed, the Ast group observed that SF3B1 is positioned on nucleosomes on top of GC-rich exons flanked by long introns (Kfir et al., 2015). We report that drug-induced intron retention is more frequent for GC-rich, short introns: this observation would be compatible with a model in which the lower recruitment of SF3B1 to chromatin around these regions would make introns more susceptible to drug inhibition. Also relevant to these points, interplay between SF3B-targeting drugs and H3K36 methylation marks has been reported (Convertini et al., 2014; Kim et al., 2011) and SF3B complex components were found to associate to the epigenetic complexes Polycomb and SAGA (Isono et al., 2005;

Martinez et al., 2001; Stegeman et al., 2016). Finally, splicing of *MCL1* itself has been linked to H3K4 methylation (Khan et al., 2014; Khan et al., 2016), further supporting the interconnection between different layers of gene regulation. Although with our minigenes analyses we found clear sequence-dependent effects, it is possible that drug-induced changes in transcription kinetics and chromatin modifications also contribute to regulation of the endogenous human *MCL1* gene.

Another well-established link is the interplay within transcription kinetics and splicing (Kornblihtt et al., 2013), whereby elongation rates determine the temporal frame in which competing splice sites commit themselves for the splicing process with other splice sites. The relevance of these aspects is further enhanced by recent evidence that splicing can take place as soon as the 3' ss exits from RNA pol II exit tunnel (Carrillo Oesterreich et al., 2016). It is interesting, in this regard, that SF3B pharmacological inhibition leads to defects in RNA pol II Ser2 phosphorylation, 3' end processing and transcripts release (Koga et al., 2015; Koga et al., 2014; Martins et al., 2011), suggesting that mechanisms of drug effects should be also considered within the frame of co-transcriptional splicing choices. It has been reported that constitutive splicing tends to be co-transcriptional, while alternative splicing can be often post-transcriptional, and weakening the Py-tract induces a switch from co- to post- transcriptional splicing (Vargas et al., 2011). It will be interesting in the future to study the extent to which SF3B1 inhibition affects co- and post- transcriptional splicing.

Drug treatment has been also associated with enlargement of speckles (considered splicing factor storage sites) and pre-mRNA leakage to the cytoplasm (Girard et al., 2012; Kaida et al., 2007; Kotake et al., 2007),

which may be linked to the overload of unspliced RNAs in drug-treated cells.

In conclusion, 3'ss recognition is a very complex and sensitive process that can be regulated at several steps of U2 snRNP assembly and through the interplay between spliceosome assembly and transcription. These molecular events can severely affect cell viability and apoptosis.

3. Similarities and differences in drugs' activities as alternative splicing modulators

So far, a relatively small number of splicing inhibitors have been reported (Bonnal et al., 2012; Salton and Misteli, 2016; Zaharieva et al., 2012) and compounds targeting SF3B complex have often been considered as a homogeneous group of drugs sharing a common pharmacophore, in spite of their structural diversity (Bonnal et al., 2012; Effenberger et al., 2016c; Lagisetti et al., 2008; Lee and Abdel-Wahab, 2016). Nevertheless, our results highlight that related drugs sharing the same target are not completely equivalent because they can induce a different spectrum of splicing changes. This possibility has to be considered carefully when drugs are developed and studied across a variety of pharmacophores, and not only in the splicing field, since structure variability can have significant implications for drug activity, toxicity, secondary effects, etc.

We speculate that in the case of SSA and Sudemycin C1, their different effects on particular targets may be due to their slightly different interactions with the SF3B complex, which might thus differentially affect spliceosome assembly depending on specific configurations of

splicing signals and nearby sequences. We cannot exclude that also other intracellular mechanisms, like NMD, RNA stability or export might be differentially affected by the two drugs, although we favor the direct explanation involving the splicing machinery itself, rather than indirect effects on related processes.

These considerations lead to the speculation that AS event-specific drugs can be discovered. Indeed, recent studies described compounds with a therapeutic potential for SMA (Spinal Muscular Atrophy) that selectively influence *SMN2* alternative splicing and only few other targets in the cell (Naryshkin et al., 2014; Palacino et al., 2015). One of these drugs enhances U1 snRNA base-pairing with a small subset of 5' splice sites in a sequence-dependent manner, including *SMN2* intron 7 5'ss (Palacino et al., 2015). Consistent with this possibility, results presented in Part II and recent publications on splicing factor mutants also revealed sequence-specific effects of *core* spliceosomal components on alternative splicing (Alsafadi et al., 2016; Darman et al., 2015; Kim et al., 2015; Shirai et al., 2015), further supporting the possibility that the *core* splicing machinery displays sophisticated mechanisms of regulation by fine-tuning of splice site selection.

The structural dissection of drug-target interactions will be key to compare different drug variants. So far, competition assays showed that different small molecules are likely to affect similarly SF3B complex conformations in order to exert their inhibitory activity. In contrast, inactive drugs can bind but not modulate SF3B activity (Effenberger et al., 2016c), suggesting a single binding site but that occupancy of a binding pocket is not sufficient to elicit drug effects. Few cases of

different intron retention regulation by different drugs have been reported (Kumar et al., 2016).

SF3B complex was shown to have a shell-like structure closing around its component p14, the small protein contacting the BP (Golas et al., 2003). Because of their tight structure, SF3B1 HEAT repeats were hypothesized to undergo a structural remodeling between more open and more closed conformations, in order to allow BP recognition by U2 snRNA (possibly within the internal cavity of the shell) and consequently splicing catalysis (Golas et al., 2003). This model is supported by cryo-EM studies of fission yeast spliceosome (Rauhut et al., 2016; Yan et al., 2016). It is therefore conceivable that different small molecules, in spite of being able to compete with each other for SF3B1 binding (Effenberger et al., 2016c), could find a slightly different accommodation within SF3B complex that could in some cases lead to different target-specific modulation. While it is still not demonstrated that these inhibitors bind within the inner surface of SF3B complex shell, it was postulated that they might act at the interface between SF3B1 and SF3B3, where the drug-resistant SF3B1 mutation R1074H is thought to be located (Yokoi et al., 2011). Intriguingly, recent structural advances show that this residue is very proximal to the branch adenosine in activated spliceosomes (Rauhut et al., 2016).

4. General considerations and future perspectives

It is interesting to remark that SF3B inhibition in general leads to alterations of AS in genes related to the regulation of apoptosis and cell cycle, cancer-related genes and genes encoding RNA processing factors, among others (Figure III.2D). This suggests that SF3B may play a

pivotal role in the control of cell division and cancer progression, an area in need of more functional and clinically relevant data. For example, similarly to recent studies on MYC-driven cancer (Hsu et al., 2015), screenings for factors increasing or decreasing sensitivity to SF3B inhibitors may lead to the discovery of important SF3B interactors and downstream effectors with a possible role in cancer and also possible targets for antitumor activities. The concept that splicing modulation may have synthetic lethal interactions with other cancer-relevant processes opens a very interesting area of research with both basic and clinical implications.

Another interesting observation made in various experiments is the quantitative (but not qualitative) dependence of drugs' effects on cell densities and status, which can be relevant to explain variability across experiments. This observation is compatible with the cell cycle status influencing the splicing inhibition effects of the drugs. Indeed, alternative splicing and cell cycle are clearly connected (Dominguez et al., 2016; Welch et al., 2016). Moreover, splicing inhibitors have antitumor effects and they target preferentially tumor cells, especially if hyperproliferation is driven by the Myc oncogene (Hsu et al., 2015; Hubert et al., 2013). Hence, the relationship between cell status and response to splicing inhibition is likely to be tight and understanding its molecular basis will be important to understand drug effects.

Results of *MCL1* regulation show that this is a prominent target of splicing inhibitors in cancer (Gao and Koide, 2013; Larrayoz et al., 2016; Wei et al., 2012). This apoptotic suppressor is often overexpressed in tumor cells and associated with multidrug resistance and therefore significant efforts are made to identify new ways to target

it (Opferman, 2016). Interestingly, AS of *MCL1* seems to occur in human, but not mouse cells. Although a careful analysis is needed to assess the presence of AS transcripts in other organisms, it seems likely that evolution has favored new ways of regulating the apoptotic pathway and in this context *MCL1* may have evolved as a very responsive sensor to physiological alterations, which has adopted AS as a prominent regulatory mechanism. Figure D.2 represents the sequence similarity of *MCL1* along evolution. Sequence differences among *mMCL1* and *bMCL1* were instrumental for finding a role for Py-tracts in Sudemycin-induced regulation. However, if *MCL1* acts as prominent target in human cells (Gao and Koide, 2013; Larrayoz et al., 2016), its lack of response in mouse cells reveals that other targets (even non conserved among organisms) can also play an important role in the drugs' antitumoral effects, as recently proposed (Lee et al., 2016).

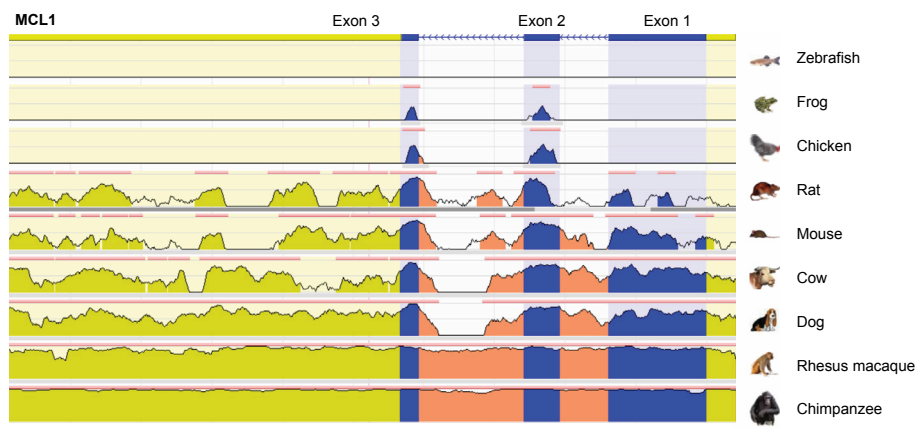


Figure 12. Sequence conservation of *MCL1* gene. Peaks' size is proportional to the sequence similarity with the human gene. UTRs are represented in yellow, coding exons in blue, introns in salmon. The information was retrieved from <https://ecrbrowser.dcode.org/> (Ovcharenko et al., 2004).

Mechanistic and physiological understanding of the effects of these drugs, with its basic and translational implications, will likely require the combination of a variety of emerging technologies.

It is not clear to which extent SF3B inhibitors might affect U2 snRNP recruitment in relation to chromatin environment. SF3B1-ChIP profiles with and without drugs may greatly help to understand if some of the effects are due to presence or absence of SF3B1 on exonic nucleosomes, as reported in recent publications (Kfir et al., 2015). A possibly complementary approach would be to assess whether the activity of SF3B1 in Polycomb complex (Isono et al., 2005) can be influenced by SF3B inhibitors.

The comparison of iCLIP profiles of SF3A and SF3B subunits in the presence and absence of the drug will allow a detailed dissection of the changes in RNA-protein contacts afforded by the drug, as suggested by previous *in vitro* experiments showing that SF3B1 interactions with the pre-mRNA are lost in presence of Spliceostatin, while SF3A1 and SF3A3 interactions are either not affected or even apparently stabilized (Corrionero et al., 2011).

Importantly, drug-target interaction studies have been limited so far by the lack of SF3B1 structures, a deficiency attenuated by the reports from *S. cerevisiae* disclosing important details of SF3B complex (Rauhut et al., 2016; Yan et al., 2016). However, even if SF3B1 is the most conserved SF3B protein (Wang et al., 1998), it is not known to which extent budding yeast *S. cerevisiae* spliceosome is affected by splicing inhibitors, while fission yeast *S. pombe* is known to be affected (Lo et al., 2007). Structures of mammalian spliceosome will help to elucidate the more complex interactions occurring in these organisms. SF3B1

R1074H mutation confers drug resistance (Lee et al., 2016; Rauhut et al., 2016; Yan et al., 2016; Yokoi et al., 2011) and it is located in very close proximity to the branch adenosine in budding yeast activated spliceosome (Rauhut et al., 2016). This information may already be interpreted in more depth and allow a much deeper understanding of the interactions affected by the drug and the discovery and even design of more specific drug variants. It is not unlikely that such drugs could be designed *in silico* to target with reasonable specificity the splicing of selected transcripts.

Transcriptome-wide techniques also allowing to study RNA-RNA interactions in large-scale or for selected RNAs of interest (Engreitz et al., 2014; Sharma et al., 2016). These approaches may also prove of great importance to elucidate the mechanism of actions of these compounds by finely mapping the U2 snRNA-BP base-pairings most altered by the drugs.

Finally, a fascinating question remains: what is the reason for several bacterial species to produce compounds targeting a key mechanism of eukaryotic gene expression like splicing? (Bonnal et al., 2012; Webb et al., 2013). This phenomenon could be related to the fact that a specific portion of SF3B1 HEAT repeats exerts toxicity for *E. coli* (Wang et al., 1998). It is conceivable that bacteria have developed these compounds to target a process essential for their natural eukaryotic competitors or predators like fungi. Since SF3B1 is present in major, minor and reduced spliceosomes (Hudson et al., 2015; Stark et al., 2015; Will and Luhrmann, 2005), it will be interesting to check to which extent all of these machineries are affected by SF3B1 inhibitors.

Conclusions

Part I.A.

- An amide derivative of Sudemycin that replaces the pharmacophore's oxycarbonyl moiety displays improved activity *in vitro* and *ex vivo*. This derivative (Sudemycin K) offers novel opportunities for further modification.
- A conjugated triene and a conjugated diene in *ZEZ* configuration decrease but do not completely suppress Sudemycin activity.
- Replacement of cyclohexane or dioxane rings by piperazine, or replacement of conjugated diene by a single double bond completely suppress Sudemycins' splicing inhibitory activity.

Part I.B.

- Sudemycin-ASO conjugates display enhanced splicing inhibitory effects compared to unconjugated ASOs.

Part II.

- Multiple BP-like sequences 5' of BPs attenuate Sudemycin C1-induced exon skipping.
- Optimal BP sequences also repress the drug-induced exon skipping effects.
- Strong polypyrimidine tracts attenuate the sensitivity of a 3' splice site to the drugs.
- Intron retention preferentially occurs in short, GC-rich introns.

Part III.

- Spliceostatin A and Sudemycin C1 display both similarities and differences in their alternative splicing modulation activities.

References

- Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Cech, M., [...] and Goecks, J. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* *44*, W3-W10.
- Albert, B.J., McPherson, P.A., O'Brien, K., Czaicki, N.L., Destefino, V., Osman, S., [...] and Koide, K. (2009). Meayamycin inhibits pre-messenger RNA splicing and exhibits picomolar activity against multidrug-resistant cells. *Mol Cancer Ther* *8*, 2308-2318.
- Alexander, R.D., Innocente, S.A., Barrass, J.D., and Beggs, J.D. (2010). Splicing-dependent RNA polymerase pausing in yeast. *Mol Cell* *40*, 582-593.
- Allende-Vega, N., Dayal, S., Agarwala, U., Sparks, A., Bourdon, J.C., and Saville, M.K. (2013). p53 is activated in response to disruption of the pre-mRNA splicing machinery. *Oncogene* *32*, 1-14.
- Alsafadi, S., Houy, A., Battistella, A., Popova, T., Wassef, M., Henry, E., [...] and Stern, M.H. (2016). Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat Commun* *7*, 10615.
- Amit, M., Donyo, M., Hollander, D., Goren, A., Kim, E., Gelfman, S., [...] and Ast, G. (2012). Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. *Cell Rep* *1*, 543-556.
- Bae, J., Leo, C.P., Hsu, S.Y., and Hsueh, A.J. (2000). MCL-1S, a splicing variant of the antiapoptotic BCL-2 family member MCL-1, encodes a proapoptotic protein possessing only the BH3 domain. *J Biol Chem* *275*, 25255-25261.
- Barash, Y., Calarco, J.A., Gao, W., Pan, Q., Wang, X., Shai, O., [...] and Frey, B.J. (2010). Deciphering the splicing code. *Nature* *465*, 53-59.
- Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., [...] and Blencowe, B.J. (2012). The evolutionary landscape of alternative splicing in vertebrate species. *Science* *338*, 1587-1593.
- Bechara, E.G., Sebestyen, E., Bernardis, I., Eyra, E., and Valcarcel, J. (2013). RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. *Mol Cell* *52*, 720-733.
- Berg, M.G., Wan, L., Younis, I., Diem, M.D., Soo, M., Wang, C., and Dreyfuss, G. (2012). A quantitative high-throughput in vitro splicing assay identifies inhibitors of spliceosome catalysis. *Mol Cell Biol* *32*, 1271-1283.
- 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.

- Bernier, F.P., Caluseriu, O., Ng, S., Schwartzentruber, J., Buckingham, K.J., Innes, A.M., [...] and Parboosingh, J.S. (2012). Haploinsufficiency of SF3B4, a component of the pre-mRNA spliceosomal complex, causes Nager syndrome. *Am J Hum Genet* *90*, 925-933.
- Beyer, A.L., and Osheim, Y.N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev* *2*, 754-765.
- Bonnal, S., Vigevani, L., and Valcarcel, J. (2012). The spliceosome as a target of novel antitumour drugs. *Nat Rev Drug Discov* *11*, 847-859.
- Braunschweig, U., Barbosa-Morais, N.L., Pan, Q., Nachman, E.N., Alipanahi, B., Gonatopoulos-Pournatzis, T., [...] and Blencowe, B.J. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res* *24*, 1774-1786.
- Braunschweig, U., Gueroussov, S., Plocik, A.M., Graveley, B.R., and Blencowe, B.J. (2013). Dynamic integration of splicing within gene regulatory pathways. *Cell* *152*, 1252-1269.
- Brody, E., and Abelson, J. (1985). The "spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science* *228*, 963-967.
- Brody, Y., Neufeld, N., Bieberstein, N., Causse, S.Z., Bohnlein, E.M., Neugebauer, K.M., [...] and Shav-Tal, Y. (2011). The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. *PLoS biology* *9*, e1000573.
- Bueno, R., Stawiski, E.W., Goldstein, L.D., Durinck, S., De Rienzo, A., Modrusan, Z., [...] and Seshagiri, S. (2016). Comprehensive genomic analysis of malignant pleural mesothelioma identifies recurrent mutations, gene fusions and splicing alterations. *Nat Genet* *48*, 407-416.
- Buljan, M., Chalancon, G., Dunker, A.K., Bateman, A., Balaji, S., Fuxreiter, M., and Babu, M.M. (2013). Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. *Curr Opin Struct Biol* *23*, 443-450.
- Carrillo Oesterreich, F., Herzel, L., Straube, K., Hujer, K., Howard, J., and Neugebauer, K.M. (2016). Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell* *165*, 372-381.
- Chakradhar, S. (2016). Splicing solutions: Companies explore new techniques to fix splicing errors. *Nat Med* *22*, 967-969.
- Chathoth, K.T., Barrass, J.D., Webb, S., and Beggs, J.D. (2014). A splicing-dependent transcriptional checkpoint associated with prespliceosome formation. *Mol Cell* *53*, 779-790.
- Cheng, J., Zhou, T., Liu, C., Shapiro, J.P., Brauer, M.J., Kiefer, M.C., [...] and Mountz, J.D. (1994). Protection from Fas-mediated apoptosis by a soluble form of the Fas molecule. *Science* *263*, 1759-1762.
- 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.
- Christofk, H.R., Vander Heiden, M.G., Harris, M.H., Ramanathan, A., Gerszten, R.E., Wei, R., [...] and Cantley, L.C. (2008). The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* *452*, 230-233.

- Conklin, J.F., Goldman, A., and Lopez, A.J. (2005). Stabilization and analysis of intron lariats in vivo. *Methods* *37*, 368-375.
- Convertini, P., Shen, M., Potter, P.M., Palacios, G., Lagisetti, C., de la Grange, P., [...] and Stamm, S. (2014). Sudemycin E influences alternative splicing and changes chromatin modifications. *Nucleic Acids Res* *42*, 4947-4961.
- Corioni, M., Antih, N., Tanackovic, G., Zavolan, M., and Kramer, A. (2011). Analysis of in situ pre-mRNA targets of human splicing factor SF1 reveals a function in alternative splicing. *Nucleic Acids Res* *39*, 1868-1879.
- Corrionero, A., Minana, B., and Valcarcel, J. (2011). Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. *Genes Dev* *25*, 445-459.
- Corvelo, A., Hallegger, M., Smith, C.W., and Eyras, E. (2010). Genome-wide association between branch point properties and alternative splicing. *PLoS Comput Biol* *6*, e1001016.
- Crews, L.A., Balaian, L., Delos Santos, N.P., Leu, H.S., Court, A.C., Lazzari, E., [...] and Jamieson, C.H. (2016). RNA Splicing Modulation Selectively Impairs Leukemia Stem Cell Maintenance in Secondary Human AML. *Cell Stem Cell*.
- Crick, F. (1970). Central dogma of molecular biology. *Nature* *227*, 561-563.
- Crick, F. (1979). Split genes and RNA splicing. *Science* *204*, 264-271.
- Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome Res* *14*, 1188-1190.
- Daguenet, E., Dujardin, G., and Valcarcel, J. (2015). The pathogenicity of splicing defects: mechanistic insights into pre-mRNA processing inform novel therapeutic approaches. *EMBO Rep* *16*, 1640-1655.
- Darman, R.B., Seiler, M., Agrawal, A.A., Lim, K.H., Peng, S., Aird, D., [...] and Buonamici, S. (2015). Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point. *Cell Rep* *13*, 1033-1045.
- David, C.J., and Manley, J.L. (2010). Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev* *24*, 2343-2364.
- DeBoever, C., Ghia, E.M., Shepard, P.J., Rassenti, L., Barrett, C.L., Jepsen, K., [...] and Frazer, K.A. (2015). Transcriptome sequencing reveals potential mechanism of cryptic 3' splice site selection in SF3B1-mutated cancers. *PLoS Comput Biol* *11*, e1004105.
- Dhar, S., La Clair, J.J., Leon, B., Hammons, J.C., Yu, Z., Kashyap, M.K., [...] and Burkart, M.D. (2016). A Carbohydrate-Derived Splice Modulator. *J Am Chem Soc* *138*, 5063-5068.
- Disney, M.D. (2008). Short-circuiting RNA splicing. *Nat Chem Biol* *4*, 723-724.
- Dominguez, D., Tsai, Y.H., Weatheritt, R., Wang, Y., Blencowe, B.J., and Wang, Z. (2016). An extensive program of periodic alternative splicing linked to cell cycle progression. *Elife* *5*.

- Dominski, Z., and Kole, R. (1994). Identification and characterization by antisense oligonucleotides of exon and intron sequences required for splicing. *Mol Cell Biol* *14*, 7445-7454.
- Dvinge, H., Kim, E., Abdel-Wahab, O., and Bradley, R.K. (2016). RNA splicing factors as oncoproteins and tumour suppressors. *Nat Rev Cancer* *16*, 413-430.
- Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R., and Hood, L. (1980). Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. *Cell* *20*, 313-319.
- Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* *10*, 48.
- Effenberger, K.A., James, R.C., Urabe, V.K., Dickey, B.J., Linington, R.G., and Jurica, M.S. (2015). The Natural Product N-Palmitoyl-L-leucine Selectively Inhibits Late Assembly of Human Spliceosomes. *J Biol Chem* *290*, 27524-27531.
- Effenberger, K.A., Urabe, V.K., and Jurica, M.S. (2016a). Modulating splicing with small molecular inhibitors of the spliceosome. *Wiley Interdiscip Rev RNA*.
- Effenberger, K.A., Urabe, V.K., Prichard, B.E., Ghosh, A.K., and Jurica, M.S. (2016b). Interchangeable SF3B1 inhibitors interfere with pre-mRNA splicing at multiple stages. *RNA* *22*, 350-359.
- Ellis, J.D., Barrios-Rodiles, M., Colak, R., Irimia, M., Kim, T., Calarco, J.A., [...] and Blencowe, B.J. (2012a). Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol Cell* *46*, 884-892.
- Ellis, M.J., Ding, L., Shen, D., Luo, J., Suman, V.J., Wallis, J.W., [...] and Mardis, E.R. (2012b). Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* *486*, 353-360.
- Engreitz, J.M., Sirokman, K., McDonel, P., Shishkin, A.A., Surka, C., Russell, P., [...] and Lander, E.S. (2014). RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell* *159*, 188-199.
- Fan, L., Lagisetti, C., Edwards, C.C., Webb, T.R., and Potter, P.M. (2011). Sudemycins, novel small molecule analogues of FR901464, induce alternative gene splicing. *ACS Chem Biol* *6*, 582-589.
- Ferreira, P.G., Jares, P., Rico, D., Gomez-Lopez, G., Martinez-Trillos, A., Villamor, N., [...] and Guigo, R. (2014). Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. *Genome Res* *24*, 212-226.
- Fica, S.M., Tuttle, N., Novak, T., Li, N.S., Lu, J., Koodathingal, P., [...] and Piccirilli, J.A. (2013). RNA catalyses nuclear pre-mRNA splicing. *Nature* *503*, 229-234.
- Folco, E.G., Coil, K.E., and Reed, R. (2011). The anti-tumor drug E7107 reveals an essential role for SF3b in remodeling U2 snRNP to expose the branch point-binding region. *Genes Dev* *25*, 440-444.

- Furney, S.J., Pedersen, M., Gentien, D., Dumont, A.G., Rapinat, A., Desjardins, L., [...] and Marais, R. (2013). SF3B1 mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov* 3, 1122-1129.
- Furumai, R., Uchida, K., Komi, Y., Yoneyama, M., Ishigami, K., Watanabe, H., [...] and Yoshida, M. (2010). Spliceostatin A blocks angiogenesis by inhibiting global gene expression including VEGF. *Cancer Sci* 101, 2483-2489.
- Gao, K., Masuda, A., Matsuura, T., and Ohno, K. (2008). Human branch point consensus sequence is γ UnAy. *Nucleic Acids Res* 36, 2257-2267.
- Gao, Y., and Koide, K. (2013). Chemical perturbation of Mcl-1 pre-mRNA splicing to induce apoptosis in cancer cells. *ACS Chem Biol* 8, 895-900.
- Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6, 343-345.
- Gilbert, W. (1978). Why genes in pieces? *Nature* 271, 501.
- Girard, C., Will, C.L., Peng, J., Makarov, E.M., Kastner, B., Lemm, I., [...] and Luhrmann, R. (2012). Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat Commun* 3, 994.
- Glaser, S.P., Lee, E.F., Trounson, E., Bouillet, P., Wei, A., Fairlie, W.D., [...] and Strasser, A. (2012). Anti-apoptotic Mcl-1 is essential for the development and sustained growth of acute myeloid leukemia. *Genes Dev* 26, 120-125.
- Golas, M.M., Sander, B., Will, C.L., Luhrmann, R., and Stark, H. (2003). Molecular architecture of the multiprotein splicing factor SF3b. *Science* 300, 980-984.
- 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.
- Grabowski, P.J., Seiler, S.R., and Sharp, P.A. (1985). A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell* 42, 345-353.
- Graveley, B.R., Hertel, K.J., and Maniatis, T. (2001). The role of U2AF35 and U2AF65 in enhancer-dependent splicing. *RNA* 7, 806-818.
- Graveley, B.R., and Maniatis, T. (1998). Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. *Mol Cell* 1, 765-771.
- Groves, M.R., and Barford, D. (1999). Topological characteristics of helical repeat proteins. *Curr Opin Struct Biol* 9, 383-389.

- Guth, S., Martinez, C., Gaur, R.K., and Valcarcel, J. (1999). Evidence for substrate-specific requirement of the splicing factor U2AF(35) and for its function after polypyrimidine tract recognition by U2AF(65). *Mol Cell Biol* *19*, 8263-8271.
- Guth, S., Tange, T.O., Kellenberger, E., and Valcarcel, J. (2001). Dual function for U2AF(35) in AG-dependent pre-mRNA splicing. *Mol Cell Biol* *21*, 7673-7681.
- Harbour, J.W., Roberson, E.D., Anbunathan, H., Onken, M.D., Worley, L.A., and Bowcock, A.M. (2013). Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. *Nat Genet* *45*, 133-135.
- Hasegawa, M., Miura, T., Kuzuya, K., Inoue, A., Won Ki, S., Horinouchi, S., [...] and Mizukami, T. (2011). Identification of SAP155 as the target of GEX1A (Herboxidiene), an antitumor natural product. *ACS Chem Biol* *6*, 229-233.
- Havens, M.A., and Hastings, M.L. (2016). Splice-switching antisense oligonucleotides as therapeutic drugs. *Nucleic Acids Res* *44*, 6549-6563.
- Hesselberth, J.R. (2013). Lives that introns lead after splicing. *Wiley Interdiscip Rev RNA* *4*, 677-691.
- Hirose, T., Ideue, T., Nagai, M., Hagiwara, M., Shu, M.D., and Steitz, J.A. (2006). A spliceosomal intron binding protein, IBP160, links position-dependent assembly of intron-encoded box C/D snoRNP to pre-mRNA splicing. *Mol Cell* *23*, 673-684.
- Horowitz, D.S., Lee, E.J., Mabon, S.A., and Misteli, T. (2002). A cyclophilin functions in pre-mRNA splicing. *EMBO J* *21*, 470-480.
- Hsu, T.Y., Simon, L.M., Neill, N.J., Marcotte, R., Sayad, A., Bland, C.S., [...] and Westbrook, T.F. (2015). The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature* *525*, 384-388.
- Hua, Y., Sahashi, K., Hung, G., Rigo, F., Passini, M.A., Bennett, C.F., and Krainer, A.R. (2010). Antisense correction of SMN2 splicing in the CNS rescues necrosis in a type III SMA mouse model. *Genes Dev* *24*, 1634-1644.
- Hubert, C.G., Bradley, R.K., Ding, Y., Toledo, C.M., Herman, J., Skutt-Kakaria, K., [...] and Paddison, P.J. (2013). Genome-wide RNAi screens in human brain tumor isolates reveal a novel viability requirement for PHF5A. *Genes Dev* *27*, 1032-1045.
- Hudson, A.J., Stark, M.R., Fast, N.M., Russell, A.G., and Rader, S.D. (2015). Splicing diversity revealed by reduced spliceosomes in *C. merolae* and other organisms. *RNA Biol* *12*, 1-8.
- Iannone, C., and Valcarcel, J. (2013). Chromatin's thread to alternative splicing regulation. *Chromosoma* *122*, 465-474.
- Ilagan, J.O., Ramakrishnan, A., Hayes, B., Murphy, M.E., Zebari, A.S., Bradley, P., and Bradley, R.K. (2015). U2AF1 mutations alter splice site recognition in hematological malignancies. *Genome Res* *25*, 14-26.
- Irimia, M., and Roy, S.W. (2014). Origin of spliceosomal introns and alternative splicing. *Cold Spring Harb Perspect Biol* *6*.

- Irimia, M., Weatheritt, R.J., Ellis, J.D., Parikshak, N.N., Gonatopoulos-Pournatzis, T., Babor, M., [...] and Blencowe, B.J. (2014). A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* *159*, 1511-1523.
- Isono, K., Mizutani-Koseki, Y., Komori, T., Schmidt-Zachmann, M.S., and Koseki, H. (2005). Mammalian polycomb-mediated repression of Hox genes requires the essential spliceosomal protein Sf3b1. *Genes Dev* *19*, 536-541.
- Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., [...] and Sharpless, N.E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* *19*, 141-157.
- Jia, Y., Mu, J.C., and Ackerman, S.L. (2012). Mutation of a U2 snRNA gene causes global disruption of alternative splicing and neurodegeneration. *Cell* *148*, 296-308.
- Jurica, M.S. (2008). Searching for a wrench to throw into the splicing machine. *Nat Chem Biol* *4*, 3-6.
- Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* *468*, 664-668.
- Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., [...] and Yoshida, M. (2007). Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat Chem Biol* *3*, 576-583.
- Kalsotra, A., and Cooper, T.A. (2011). Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet* *12*, 715-729.
- Kanada, R.M., Itoh, D., Nagai, M., Niiijima, J., Asai, N., Mizui, Y., [...] and Kotake, Y. (2007). Total synthesis of the potent antitumor macrolides pladienolide B and D. *Angew Chem Int Ed Engl* *46*, 4350-4355.
- Kashyap, M.K., Kumar, D., Villa, R., La Clair, J.J., Benner, C., Sasik, R., [...] and Castro, J.E. (2015). Targeting the spliceosome in chronic lymphocytic leukemia with the macrolides FD-895 and pladienolide-B. *Haematologica* *100*, 945-954.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res* *12*, 996-1006.
- Kesarwani, A.K., Ramirez, O., Gupta, A.K., Yang, X., Murthy, T., Minella, A.C., and Pillai, M.M. (2016). Cancer-associated SF3B1 mutants recognize otherwise inaccessible cryptic 3' splice sites within RNA secondary structures. *Oncogene*.
- Kfir, N., Lev-Maor, G., Gleich, O., Alajem, A., Datta, A., Sze, S.K., [...] and Ast, G. (2015). SF3B1 association with chromatin determines splicing outcomes. *Cell Rep* *11*, 618-629.
- Khan, D.H., Gonzalez, C., Cooper, C., Sun, J.M., Chen, H.Y., Healy, S., [...] and Davie, J.R. (2014). RNA-dependent dynamic histone acetylation regulates MCL1 alternative splicing. *Nucleic Acids Res* *42*, 1656-1670.

- Khan, D.H., Gonzalez, C., Taylor, N., Hamedani, M.K., Leygue, E., and Davie, J.R. (2016). Dynamic Histone Acetylation of H3K4me3 Nucleosome Regulates MCL1 Pre-mRNA Splicing. *J Cell Physiol* *231*, 2196-2204.
- Kim, E., Ilagan, J.O., Liang, Y., Daubner, G.M., Lee, S.C., Ramakrishnan, A., [...] and Abdel-Wahab, O. (2015). SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. *Cancer Cell* *27*, 617-630.
- Kim, S., Kim, H., Fong, N., Erickson, B., and Bentley, D.L. (2011). Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proc Natl Acad Sci U S A* *108*, 13564-13569.
- Koga, M., Hayashi, M., and Kaida, D. (2015). Splicing inhibition decreases phosphorylation level of Ser2 in Pol II CTD. *Nucleic Acids Res* *43*, 8258-8267.
- Koga, M., Satoh, T., Takasaki, I., Kawamura, Y., Yoshida, M., and Kaida, D. (2014). U2 snRNP is required for expression of the 3' end of genes. *PLoS One* *9*, e98015.
- Kojima, S., Hyakutake, A., Koshikawa, N., Nakagawara, A., and Takenaga, K. (2010). MCL-1V, a novel mouse antiapoptotic MCL-1 variant, generated by RNA splicing at a non-canonical splicing pair. *Biochem Biophys Res Commun* *391*, 492-497.
- Kole, R., Krainer, A.R., and Altman, S. (2012). RNA therapeutics: beyond RNA interference and antisense oligonucleotides. *Nat Rev Drug Discov* *11*, 125-140.
- Kornblihtt, A.R., Schor, I.E., Allo, M., Dujardin, G., Petrillo, E., and Munoz, M.J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol* *14*, 153-165.
- Kotake, Y., Sagane, K., Owa, T., Mimori-Kiyosue, Y., Shimizu, H., Uesugi, M., [...] and Mizui, Y. (2007). Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nat Chem Biol* *3*, 570-575.
- Kumar, D., Kashyap, M.K., La Clair, J.J., Villa, R., Spaanderman, I., Chien, S., [...] and Castro, J.E. (2016). Selectivity in Small Molecule Splicing Modulation. *ACS Chem Biol*.
- Kwiatkowski, S., Sviripa, V.M., Zhang, Z., Wendlandt, A.E., Hobartner, C., Watt, D.S., and Stamm, S. (2016). Synthesis of a norcantharidin-tethered guanosine: Protein phosphatase-1 inhibitors that change alternative splicing. *Bioorg Med Chem Lett* *26*, 965-968.
- Laetsch, T.W., Liu, X., Vu, A., Sliozberg, M., Vido, M., Elci, O.U., [...] and Hogarty, M.D. (2014). Multiple components of the spliceosome regulate Mcl1 activity in neuroblastoma. *Cell Death Dis* *5*, e1072.
- Lagiseti, C., Pourpak, A., Goronga, T., Jiang, Q., Cui, X., Hyle, J., [...] and Webb, T.R. (2009). Synthetic mRNA splicing modulator compounds with in vivo antitumor activity. *J Med Chem* *52*, 6979-6990.
- Lagiseti, C., Pourpak, A., Jiang, Q., Cui, X., Goronga, T., Morris, S.W., and Webb, T.R. (2008). Antitumor compounds based on a natural product consensus pharmacophore. *J Med Chem* *51*, 6220-6224.

- Laizure, S.C., Herring, V., Hu, Z., Witbrodt, K., and Parker, R.B. (2013). The role of human carboxylesterases in drug metabolism: have we overlooked their importance? *Pharmacotherapy* *33*, 210-222.
- Lareau, L.F., and Brenner, S.E. (2015). Regulation of splicing factors by alternative splicing and NMD is conserved between kingdoms yet evolutionarily flexible. *Mol Biol Evol* *32*, 1072-1079.
- Larrayoz, M., Blakemore, S.J., Dobson, R.C., Blunt, M.D., Rose-Zerilli, M.J., Walewska, R., [...] and Steele, A.J. (2016). The SF3B1 inhibitor spliceostatin A (SSA) elicits apoptosis in chronic lymphocytic leukaemia cells through downregulation of Mcl-1. *Leukemia* *30*, 351-360.
- Lee, S.C., and Abdel-Wahab, O. (2016). Therapeutic targeting of splicing in cancer. *Nat Med* *22*, 976-986.
- Lee, S.C., Dvinge, H., Kim, E., Cho, H., Micol, J.B., Chung, Y.R., [...] and Abdel-Wahab, O. (2016). Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. *Nat Med* *22*, 672-678.
- Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., [...] and et al. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* *80*, 155-165.
- 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, X., Biswas, S., Berg, M.G., Antapli, C.M., Xie, F., Wang, Q., [...] and Cheng, Y.Q. (2013). Genomics-guided discovery of thailanstatins A, B, and C As pre-mRNA splicing inhibitors and antiproliferative agents from *Burkholderia thailandensis* MSMB43. *J Nat Prod* *76*, 685-693.
- Lo, C.W., Kaida, D., Nishimura, S., Matsuyama, A., Yashiroda, Y., Taoka, H., [...] and Yoshida, M. (2007). Inhibition of splicing and nuclear retention of pre-mRNA by spliceostatin A in fission yeast. *Biochem Biophys Res Commun* *364*, 573-577.
- Lorson, C.L., Hahnen, E., Androphy, E.J., and Wirth, B. (1999). A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc Natl Acad Sci U S A* *96*, 6307-6311.
- Madhani, H.D. (2013). The frustrated gene: origins of eukaryotic gene expression. *Cell* *155*, 744-749.
- Malcovati, L., Papaemmanuil, E., Bowen, D.T., Boulton, J., Della Porta, M.G., Pascutto, C., [...] and of the Associazione Italiana per la Ricerca sul Cancro Gruppo Italiano Malattie, M. (2011). Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood* *118*, 6239-6246.
- Martin, R.M., Rino, J., de Jesus, A.C., and Carmo-Fonseca, M. (2016). Single-Molecule Live-Cell Visualization of Pre-mRNA Splicing. *Methods Mol Biol* *1358*, 335-350.
- Martin, W., and Koonin, E.V. (2006). Introns and the origin of nucleus-cytosol compartmentalization. *Nature* *440*, 41-45.

- Martinez, E., Palhan, V.B., Tjernberg, A., Lymar, E.S., Gamper, A.M., Kundu, T.K., [...] and Roeder, R.G. (2001). Human STAGA complex is a chromatin-acetylating transcription coactivator that interacts with pre-mRNA splicing and DNA damage-binding factors in vivo. *Mol Cell Biol* *21*, 6782-6795.
- Martins, S.B., Rino, J., Carvalho, T., Carvalho, C., Yoshida, M., Klose, J.M., [...] and Carmo-Fonseca, M. (2011). Spliceosome assembly is coupled to RNA polymerase II dynamics at the 3' end of human genes. *Nat Struct Mol Biol* *18*, 1115-1123.
- Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. *Nat Rev Mol Cell Biol* *15*, 108-121.
- Mercer, T.R., Clark, M.B., Andersen, S.B., Brunck, M.E., Haerty, W., Crawford, J., [...] and Mattick, J.S. (2015). Genome-wide discovery of human splicing branchpoints. *Genome Res* *25*, 290-303.
- Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* *338*, 1593-1599.
- Miller-Wideman, M., Makkar, N., Tran, M., Isaac, B., Biest, N., and Stonard, R. (1992). Herboxidiene, a new herbicidal substance from *Streptomyces chromofuscus* A7847. Taxonomy, fermentation, isolation, physico-chemical and biological properties. *J Antibiot (Tokyo)* *45*, 914-921.
- Mirtschink, P., Krishnan, J., Grimm, F., Sarre, A., Horl, M., Kayikci, M., [...] and Krek, W. (2015). HIF-driven SF3B1 induces KHK-C to enforce fructolysis and heart disease. *Nature* *522*, 444-449.
- Mizui, Y., Sakai, T., Iwata, M., Uenaka, T., Okamoto, K., Shimizu, H., [...] and Asada, M. (2004). Pladienolides, new substances from culture of *Streptomyces platensis* Mer-11107. III. In vitro and in vivo antitumor activities. *J Antibiot (Tokyo)* *57*, 188-196.
- Mollet, I.G., Ben-Dov, C., Felicio-Silva, D., Grosso, A.R., Eleuterio, P., Alves, R., [...] and Carmo-Fonseca, M. (2010). Unconstrained mining of transcript data reveals increased alternative splicing complexity in the human transcriptome. *Nucleic Acids Res* *38*, 4740-4754.
- Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* *136*, 688-700.
- Moore, M.J., Wang, Q., Kennedy, C.J., and Silver, P.A. (2010). An alternative splicing network links cell-cycle control to apoptosis. *Cell* *142*, 625-636.
- Mupo, A., Seiler, M., Sathiaselan, V., Pance, A., Yang, Y., Agrawal, A.A., [...] and Vassiliou, G.S. (2016). Hemopoietic-specific Sf3b1-K700E knock-in mice display the splicing defect seen in human MDS but develop anemia without ring sideroblasts. *Leukemia*.
- Muraki, M., Ohkawara, B., Hosoya, T., Onogi, H., Koizumi, J., Koizumi, T., [...] and Hagiwara, M. (2004). Manipulation of alternative splicing by a newly developed inhibitor of Clks. *J Biol Chem* *279*, 24246-24254.
- Nakajima, H., Hori, Y., Terano, H., Okuhara, M., Manda, T., Matsumoto, S., and Shimomura, K. (1996a). New antitumor substances, FR901463, FR901464 and FR901465. II. Activities

against experimental tumors in mice and mechanism of action. *J Antibiot (Tokyo)* *49*, 1204-1211.

Nakajima, H., Sato, B., Fujita, T., Takase, S., Terano, H., and Okuhara, M. (1996b). New antitumor substances, FR901463, FR901464 and FR901465. I. Taxonomy, fermentation, isolation, physico-chemical properties and biological activities. *J Antibiot (Tokyo)* *49*, 1196-1203.

Naryshkin, N.A., Weetall, M., Dakka, A., Narasimhan, J., Zhao, X., Feng, Z., [...] and Metzger, F. (2014). Motor neuron disease. SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. *Science* *345*, 688-693.

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

Nojima, T., Gomes, T., Grosso, A.R., Kimura, H., Dye, M.J., Dhir, S., [...] and Proudfoot, N.J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* *161*, 526-540.

O'Brien, J., Wilson, I., Orton, T., and Pognan, F. (2000). Investigation of the Alamar Blue (resazurin) fluorescent dye for the assessment of mammalian cell cytotoxicity. *Eur J Biochem* *267*, 5421-5426.

O'Brien, K., Matlin, A.J., Lowell, A.M., and Moore, M.J. (2008). The biflavonoid isoginkgetin is a general inhibitor of Pre-mRNA splicing. *J Biol Chem* *283*, 33147-33154.

Obeng, E.A., Chappell, R.J., Seiler, M., Chen, M.C., Campagna, D.R., Schmidt, P.J., [...] and Ebert, B.L. (2016). Physiologic Expression of Sf3b1(K700E) Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation. *Cancer Cell* *30*, 404-417.

Okeyo-Owuor, T., White, B.S., Chatrikhi, R., Mohan, D.R., Kim, S., Griffith, M., [...] and Graubert, T.A. (2015). U2AF1 mutations alter sequence specificity of pre-mRNA binding and splicing. *Leukemia* *29*, 909-917.

Opferman, J.T. (2016). Attacking cancer's Achilles heel: antagonism of anti-apoptotic BCL-2 family members. *FEBS J* *283*, 2661-2675.

Ovcharenko, I., Nobrega, M.A., Loots, G.G., and Stubbs, L. (2004). ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. *Nucleic Acids Res* *32*, W280-286.

Padgett, R.A., Konarska, M.M., Grabowski, P.J., Hardy, S.F., and Sharp, P.A. (1984). Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science* *225*, 898-903.

Palacino, J., Swalley, S.E., Song, C., Cheung, A.K., Shu, L., Zhang, X., [...] and Sivasankaran, R. (2015). SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. *Nat Chem Biol* *11*, 511-517.

- 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.
- Papaemmanuil, E., Cazzola, M., Boulwood, J., Malcovati, L., Vyas, P., Bowen, D., [...] and Chronic Myeloid Disorders Working Group of the International Cancer Genome, C. (2011). Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *The New England journal of medicine* *365*, 1384-1395.
- Papasaikas, P., Tejedor, J.R., Vigevani, L., and Valcarcel, J. (2015). Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. *Mol Cell* *57*, 7-22.
- Papasaikas, P., and Valcarcel, J. (2016). The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends Biochem Sci* *41*, 33-45.
- Park, S.M., Ou, J., Chamberlain, L., Simone, T.M., Yang, H., Virbasius, C.M., [...] and Green, M.R. (2016). U2AF35(S34F) Promotes Transformation by Directing Aberrant ATG7 Pre-mRNA 3' End Formation. *Mol Cell* *62*, 479-490.
- Parker, R., Siliciano, P.G., and Guthrie, C. (1987). Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell* *49*, 229-239.
- Patel, A.A., and Steitz, J.A. (2003). Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol* *4*, 960-970.
- Paulsen, M.T., Veloso, A., Prasad, J., Bedi, K., Ljungman, E.A., Magnuson, B., [...] and Ljungman, M. (2014). Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA. *Methods* *67*, 45-54.
- Pawellek, A., McElroy, S., Samatov, T., Mitchell, L., Woodland, A., Ryder, U., [...] and Lamond, A.I. (2014). Identification of small molecule inhibitors of pre-mRNA splicing. *J Biol Chem* *289*, 34683-34698.
- Perriman, R., and Ares, M., Jr. (2010). Invariant U2 snRNA nucleotides form a stem loop to recognize the intron early in splicing. *Mol Cell* *38*, 416-427.
- Perriman, R.J., and Ares, M., Jr. (2007). Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing. *Genes Dev* *21*, 811-820.
- Pleasance, E.D., Cheetham, R.K., Stephens, P.J., McBride, D.J., Humphray, S.J., Greenman, C.D., [...] and Stratton, M.R. (2010). A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* *463*, 191-196.
- Przychodzen, B., Jerez, A., Guinta, K., Sekeres, M.A., Padgett, R., Maciejewski, J.P., and Makishima, H. (2013). Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. *Blood* *122*, 999-1006.
- Quesada, V., Conde, L., Villamor, N., Ordóñez, G.R., Jares, P., Bassaganyas, L., [...] and Lopez-Otin, C. (2012). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat Genet* *44*, 47-52.

- Raj, B., and Blencowe, B.J. (2015). Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. *Neuron* *87*, 14-27.
- Rauhut, R., Fabrizio, P., Dybkov, O., Hartmuth, K., Pena, V., Chari, A., [...] and Luhrmann, R. (2016). Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. *Science*.
- Robart, A.R., Chan, R.T., Peters, J.K., Rajashankar, K.R., and Toor, N. (2014). Crystal structure of a eukaryotic group II intron lariat. *Nature* *514*, 193-197.
- Rosenberg, A.B., Patwardhan, R.P., Shendure, J., and Seelig, G. (2015). Learning the sequence determinants of alternative splicing from millions of random sequences. *Cell* *163*, 698-711.
- Rossi, D., Brusca, A., Spina, V., Rasi, S., Khiabani, H., Messina, M., [...] and Gaidano, G. (2011). Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. *Blood* *118*, 6904-6908.
- Roybal, G.A., and Jurica, M.S. (2010). Spliceostatin A inhibits spliceosome assembly subsequent to pre-spliceosome formation. *Nucleic Acids Res* *38*, 6664-6672.
- Ruskin, B., Krainer, A.R., Maniatis, T., and Green, M.R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell* *38*, 317-331.
- Sakai, T., Sameshima, T., Matsufuji, M., Kawamura, N., Dobashi, K., and Mizui, Y. (2004). Pladienolides, new substances from culture of *Streptomyces platensis* Mer-11107. I. Taxonomy, fermentation, isolation and screening. *J Antibiot (Tokyo)* *57*, 173-179.
- Sakai, Y., Tsujita, T., Akiyama, T., Yoshida, T., Mizukami, T., Akinaga, S., [...] and Yoshida, T. (2002a). GEX1 compounds, novel antitumor antibiotics related to herboxidiene, produced by *Streptomyces* sp. II. The effects on cell cycle progression and gene expression. *J Antibiot (Tokyo)* *55*, 863-872.
- Sakai, Y., Yoshida, T., Ochiai, K., Uosaki, Y., Saitoh, Y., Tanaka, F., [...] and Mizukami, T. (2002b). GEX1 compounds, novel antitumor antibiotics related to herboxidiene, produced by *Streptomyces* sp. I. Taxonomy, production, isolation, physicochemical properties and biological activities. *J Antibiot (Tokyo)* *55*, 855-862.
- Salton, M., Kasprzak, W.K., Voss, T., Shapiro, B.A., Poulidakos, P.I., and Misteli, T. (2015). Inhibition of vemurafenib-resistant melanoma by interference with pre-mRNA splicing. *Nat Commun* *6*, 7103.
- Salton, M., and Misteli, T. (2016). Small Molecule Modulators of Pre-mRNA Splicing in Cancer Therapy. *Trends Mol Med* *22*, 28-37.
- Saltzman, A.L., Pan, Q., and Blencowe, B.J. (2011). Regulation of alternative splicing by the core spliceosomal machinery. *Genes Dev* *25*, 373-384.
- Satoh, T., and Kaida, D. (2016). Upregulation of p27 cyclin-dependent kinase inhibitor and a C-terminus truncated form of p27 contributes to G1 phase arrest. *Sci Rep* *6*, 27829.
- Schatz, J.H., and Wendel, H.G. (2011). Targeted cancer therapy: what if the driver is just a messenger? *Cell Cycle* *10*, 3830-3833.

- Schellenberg, M.J., Dul, E.L., and MacMillan, A.M. (2011). Structural model of the p14/SF3b155 . branch duplex complex. *RNA* *17*, 155-165.
- Schmidt, U., Basyuk, E., Robert, M.C., Yoshida, M., Villemin, J.P., Auboeuf, D., [...] and Bertrand, E. (2011). Real-time imaging of cotranscriptional splicing reveals a kinetic model that reduces noise: implications for alternative splicing regulation. *J Cell Biol* *193*, 819-829.
- Schneider-Poetsch, T., Usui, T., Kaida, D., and Yoshida, M. (2010). Garbled messages and corrupted translations. *Nat Chem Biol* *6*, 189-198.
- Scotti, M.M., and Swanson, M.S. (2016). RNA mis-splicing in disease. *Nat Rev Genet* *17*, 19-32.
- Sharma, E., Sterne-Weiler, T., O'Hanlon, D., and Blencowe, B.J. (2016). Global Mapping of Human RNA-RNA Interactions. *Mol Cell* *62*, 618-626.
- Shionyu, M., Yamaguchi, A., Shinoda, K., Takahashi, K., and Go, M. (2009). AS-ALPS: a database for analyzing the effects of alternative splicing on protein structure, interaction and network in human and mouse. *Nucleic Acids Res* *37*, D305-309.
- Shirai, C.L., Ley, J.N., White, B.S., Kim, S., Tibbitts, J., Shao, J., [...] and Walter, M.J. (2015). Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. *Cancer Cell* *27*, 631-643.
- Stalker, J., Gibbins, B., Meidl, P., Smith, J., Spooner, W., Hotz, H.R., and Cox, A.V. (2004). The Ensembl Web site: mechanics of a genome browser. *Genome Res* *14*, 951-955.
- Stark, M.R., Dunn, E.A., Dunn, W.S., Grisdale, C.J., Daniele, A.R., Halstead, M.R., [...] and Rader, S.D. (2015). Dramatically reduced spliceosome in *Cyanidioschyzon merolae*. *Proc Natl Acad Sci U S A* *112*, E1191-1200.
- Stegeman, R., Spreacker, P.J., Swanson, S.K., Stephenson, R., Florens, L., Washburn, M.P., and Weake, V.M. (2016). The Spliceosomal Protein SF3B5 is a Novel Component of Drosophila SAGA that Functions in Gene Expression Independent of Splicing. *J Mol Biol* *428*, 3632-3649.
- Supek, F., Minana, B., Valcarcel, J., Gabaldon, T., and Lehner, B. (2014). Synonymous mutations frequently act as driver mutations in human cancers. *Cell* *156*, 1324-1335.
- Te Raa, G.D., Derks, I.A., Navrkalova, V., Skowronska, A., Moerland, P.D., van Laar, J., [...] and Eldering, E. (2015). The impact of SF3B1 mutations in CLL on the DNA-damage response. *Leukemia* *29*, 1133-1142.
- Tejedor, J.R., Papasaikas, P., and Valcarcel, J. (2015). Genome-wide identification of Fas/CD95 alternative splicing regulators reveals links with iron homeostasis. *Mol Cell* *57*, 23-38.
- Tiedemann, R.E., Zhu, Y.X., Schmidt, J., Shi, C.X., Sereduk, C., Yin, H., [...] and Stewart, A.K. (2012). Identification of molecular vulnerabilities in human multiple myeloma cells by RNA interference lethality screening of the druggable genome. *Cancer Res* *72*, 757-768.
- Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakraborty, S., Djebali, S., [...] and Guigo, R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res* *22*, 1616-1625.

-
- Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* *31*, 46-53.
- Vacic, V., Iakoucheva, L.M., and Radivojac, P. (2006). Two Sample Logo: a graphical representation of the differences between two sets of sequence alignments. *Bioinformatics* *22*, 1536-1537.
- Valcarcel, J., Singh, R., Zamore, P.D., and Green, M.R. (1993). The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. *Nature* *362*, 171-175.
- Vargas, D.Y., Shah, K., Batish, M., Levandoski, M., Sinha, S., Marras, S.A., [...] and Tyagi, S. (2011). Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell* *147*, 1054-1065.
- Visconte, V., Makishima, H., Jankowska, A., Szpurka, H., Traina, F., Jerez, A., [...] and Tiu, R.V. (2012). SF3B1, a splicing factor is frequently mutated in refractory anemia with ring sideroblasts. *Leukemia* *26*, 542-545.
- Vogel, J., Hess, W.R., and Borner, T. (1997). Precise branch point mapping and quantification of splicing intermediates. *Nucleic Acids Res* *25*, 2030-2031.
- Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* *136*, 701-718.
- Wan, Y., Zheng, X., Chen, H., Guo, Y., Jiang, H., He, X., [...] and Zheng, Y. (2015). Splicing function of mitotic regulators links R-loop-mediated DNA damage to tumor cell killing. *J Cell Biol* *209*, 235-246.
- Wang, C., Chua, K., Seghezzi, W., Lees, E., Gozani, O., and Reed, R. (1998). Phosphorylation of spliceosomal protein SAP 155 coupled with splicing catalysis. *Genes Dev* *12*, 1409-1414.
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., [...] and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* *456*, 470-476.
- Wang, L., Lawrence, M.S., Wan, Y., Stojanov, P., Sougnez, C., Stevenson, K., [...] and Wu, C.J. (2011). SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *The New England journal of medicine* *365*, 2497-2506.
- Webb, T.R., Joyner, A.S., and Potter, P.M. (2013). The development and application of small molecule modulators of SF3b as therapeutic agents for cancer. *Drug Discov Today* *18*, 43-49.
- Wei, G., Margolin, A.A., Haery, L., Brown, E., Cucolo, L., Julian, B., [...] and Golub, T.R. (2012). Chemical genomics identifies small-molecule MCL1 repressors and BCL-xL as a predictor of MCL1 dependency. *Cancer Cell* *21*, 547-562.
- Welch, J.D., Hu, Y., and Prins, J.F. (2016). Robust detection of alternative splicing in a population of single cells. *Nucleic Acids Res* *44*, e73.
- Will, C.L., and Luhrmann, R. (2005). Splicing of a rare class of introns by the U12-dependent spliceosome. *Biol Chem* *386*, 713-724.

Will, C.L., and Luhrmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb Perspect Biol* 3.

Wong, J.J., Ritchie, W., Ebner, O.A., Selbach, M., Wong, J.W., Huang, Y., [...] and Rasko, J.E. (2013). Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* 154, 583-595.

Wood, L.D., Parsons, D.W., Jones, S., Lin, J., Sjoblom, T., Leary, R.J., [...] and Vogelstein, B. (2007). The genomic landscapes of human breast and colorectal cancers. *Science* 318, 1108-1113.

Wu, G., Adachi, H., Ge, J., Stephenson, D., Query, C.C., and Yu, Y.T. (2016). Pseudouridines in U2 snRNA stimulate the ATPase activity of Prp5 during spliceosome assembly. *EMBO J* 35, 654-667.

Xargay-Torrent, S., Lopez-Guerra, M., Rosich, L., Montraveta, A., Roldan, J., Rodriguez, V., [...] and Colomer, D. (2015). The splicing modulator sudemycin induces a specific antitumor response and cooperates with ibrutinib in chronic lymphocytic leukemia. *Oncotarget* 6, 22734-22749.

Xing, Y., Xu, Y., Chen, Y., Jeffrey, P.D., Chao, Y., Lin, Z., [...] and Shi, Y. (2006). Structure of protein phosphatase 2A core enzyme bound to tumor-inducing toxins. *Cell* 127, 341-353.

Xiong, H.Y., Alipanahi, B., Lee, L.J., Bretschneider, H., Merico, D., Yuen, R.K., [...] and Frey, B.J. (2015). RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science* 347, 1254806.

Yan, C., Hang, J., Wan, R., Huang, M., Wong, C.C., and Shi, Y. (2015). Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* 349, 1182-1191.

Yan, C., Wan, R., Bai, R., Huang, G., and Shi, Y. (2016). Structure of a yeast activated spliceosome at 3.5 Å resolution. *Science* 353, 904-911.

Yang, X., Coulombe-Huntington, J., Kang, S., Sheynkman, G.M., Hao, T., Richardson, A., [...] and Vidal, M. (2016). Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell* 164, 805-817.

Ye, J., and Blelloch, R. (2014). Regulation of pluripotency by RNA binding proteins. *Cell Stem Cell* 15, 271-280.

Yeo, G., and Burge, C.B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 11, 377-394.

Yokoi, A., Kotake, Y., Takahashi, K., Kadowaki, T., Matsumoto, Y., Minoshima, Y., [...] and Mizui, Y. (2011). Biological validation that SF3b is a target of the antitumor macrolide pladienolide. *FEBS J* 278, 4870-4880.

Yoshida, K., and Ogawa, S. (2014). Splicing factor mutations and cancer. *Wiley Interdiscip Rev RNA* 5, 445-459.

Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., [...] and Ogawa, S. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 478, 64-69.

Zaharieva, E., Chipman, J.K., and Soller, M. (2012). Alternative splicing interference by xenobiotics. *Toxicology* *296*, 1-12.

Zhang, J., Lieu, Y.K., Ali, A.M., Penson, A., Reggio, K.S., Rabadan, R., [...] and Manley, J.L. (2015). Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. *Proc Natl Acad Sci U S A* *112*, E4726-4734.

Zhuang, Y.A., Goldstein, A.M., and Weiner, A.M. (1989). UACUAAC is the preferred branch site for mammalian mRNA splicing. *Proc Natl Acad Sci U S A* *86*, 2752-2756.

Annexes

Bonnal S, Vigevani L, Valcárcel J.

[The spliceosome as a target of novel antitumour drugs.](#)

Nature Reviews Drug Discovery. 2012

Nov;11(11):847-59.DOI: 10.1038/nrd3823

Vigevani L, Valcárcel J.

[Molecular biology. A splicing magic](#)

[bullet](#). Science. 2014 Aug

8;345(6197):624-5. DOI: 10.1126/

science.1258444

Papasaikas P, Tejedor JR, Vigevani L, Valcárcel J.

[Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery.](#)

Mol Cell. 2015 Jan 8;57(1):7-22.DOI: 10.1016/j.molcel.2014.10.030

Acknowledgements

For these acknowledgements I could write an entire book and I would still have a lot to say! The list of names worth mentioning could be so long that I can be grateful to the beauty of life in general: I was very lucky since the very beginning in having fortunate paths in front of me and meeting wonderful people!

Doing this PhD was a dream for me...and I am really happy about this sometimes hard but amazing experience!

My biggest acknowledgement goes to Juan, who gave me the opportunity to work on a very interesting project and in a very stimulating environment: I learnt a lot on science and beyond during my time in the lab and I got very generous and positive support! Thanks for your trust and infinite patience as well!

I also thank the collaborators (in particular Kamil, André, Martina and all Jernej's group), the members of the Thesis Committee for feedback and support and the members of the tribunal for being very kind and available to arrange the defense.

Un grazie grandissimo ai miei genitori e ai miei fratelli Andrea, Maria e Irene, alle mie fantastiche nipotine Caterina, Chiara ed Elisa e al paziente cognato Michele: senza dubbio alla famiglia va il mio riconoscimento più importante per accompagnarmi in ogni piccolo o grande passo che affronto e per essere i primi tra tutti a rendermi così fortunata.

I do not have words to thank Elias and Pan, who were always there for discussing about science or life...and for some fun together as well. Pan helped me a lot with the project and gave some of the most fruitful encouragements: ευχαριστώ!

Words are even less useful to express my gratitude to Camilla, la mia ormai storica "compagna di banco" e di bicicletate, that made me feel at home in so many different places in the world, and manages to make me feel like she is always sitting next to me, no matters where she is! Grazie di cuore!

Thanks to Claudia, who also makes it much nicer to face the uphill paths everyday... and brought the sound of music into my everyday life! Muchas gracias to Elena, our sweet Rugby (but also kayak and ski) Princess and Pan con Tomate queen, as well as the best Spanish teacher I can recommend!

Cate, I am so happy that we shared so much inside and outside the lab during our parallel paths and that we arrived together also to this new finish line: enhorabona!

I acknowledge all the people from Juan's group for making the lab a nice environment to work in: Keiko (感謝 for bringing friendliness, irony and fun from far away!), Sophie (for technical tips), Anna Ribó (the angel of the lab, seriously!), Eli (thanks for your very precious collaboration on the project and for a lot of moral help!), Gwen, Sergio, Joao, MP, Belén (I had so many cakes on Mondays, mmm!), Caterina (welcome to the “traditional” rides on San Joan!), Gosia and Jordi (it was nice to support each other for the final rush!). Vamos, Splice Girls!! A special thank to Ramón.

Many thanks also to the beach volleyball teams Pronto, Pink Wigs and Splicers... and to all the people of the program and the building that contributed to make my stay a very pleasant experience, even just when walking in the corridors!

The PhD Rep experience was really fun and constructive: thanks to Vicky, Sergio... and, of course, to Imma! Álvaro and Mónica taught me to teach and helped me with the project with enthusiasm and friendliness: thanks!

I am grateful to Gaetano (for the coffees and nice chats/“inciuci napoletani”), Matteo (for his constant moral and gastronomic support!), Sarah (the perfect flat mate as well as very good friend), Joan, Valentina, Pamela and all the other flat mates that made home a nice place to go back to everyday!

Thanks a lot to Cristina, Peter, Villy, Kiana, Nadia, the Climbers, the PRBB choir and all the Italian friends that came to visit from time to time...and to my unique cousins (Elena: you were the most loyal visitor ever)! Paola, Marghe e Silvia: grazie di tutto... che bello avere un po' di Milano anche qui!

Last but not least, I am grateful to the magic of Barcelona and music, which constantly accompanied me with suitable solutions for all the different needs of every distinct moment!

Thanks to all!

You are the ones making everything worth!