Molecular architecture of SF3B and the structural basis of splicing modulation

Dissertation

for the award of the degree

“Doctor rerum naturalium” (Dr. rer. nat.)

in the Molecular Biology Graduate Program
Division of Mathematics and Natural Sciences
of the Georg-August-Universität Göttingen

submitted by

Constantin Cretu

born in
Chisinau

Göttingen

2018

Members of the thesis committee:

Dr. Vladimir Pena (1st reviewer)

Research group Macromolecular Crystallography,

Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Patrick Cramer (2nd reviewer)

Department of Molecular Biology,

Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Henning Urlaub

Research group Bioanalytical Mass Spectrometry,

Max Planck Institute for Biophysical Chemistry, Göttingen

Further members of the Examination Board:

Prof. Dr. Reinhard Lührmann
Department of Cellular Biochemistry,
Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Ralf Ficner
Department of Molecular Structural Biology,
Institute for Microbiology and Genetics, Göttingen

Alexis Caspar (Alex) Faesen, PhD
Research group Biochemistry of Signal Dynamics,
Max Planck Institute for Biophysical Chemistry, Göttingen

Date of submission of the thesis: March 31, 2018

Date of the oral examination: June 26, 2018

Affidavit

I hereby declare that this thesis has been written independently and with no other sources and aids other than quoted. This thesis has not been submitted elsewhere for any academic degree or qualification.

Constantin Cretu

Göttingen, 2018

This work is dedicated to my beloved grandfather

“Să vezi departe e ceva, să ajungi acolo e altceva.”

“To see far is one thing, going there is another.”

(Constantin Brancusi)

Table of Contents

1 Summary

During splicing non-coding introns are excised from the transcribed pre-messenger RNA (pre-mRNA), and the protein-coding exons are ligated to generate the mature mRNA. In cells, the pre-mRNA splicing reaction is catalyzed by the spliceosome, a highly dynamic molecular machine composed of five small nuclear ribonucleoprotein particles (snRNPs) and additional non-snRNP factors (Wahl et al., 2009; Will and Luhrmann, 2011). At the earlier stages of spliceosome assembly, the U2 snRNP is recruited to the 3' region of the intron for the U2 snRNA to base-pair with the branch-site (BS), in a complex and insufficiently understood process (Wahl et al., 2009). SF3B is the largest U2 subcomplex, and several of its seven subunits, including SF3B1, contact both the U2 snRNA and the intron near the BS, stabilizing the U2/BS base-pairing interaction. Recurrent somatic cancer mutations in SF3B1, and in several related splicing factors, reduce the accuracy of BS selection and, finally, lead to aberrant splicing (Dvinge et al., 2016). The compromised function of SF3B1 affects splicing of many different transcripts and thus translates to global changes in cancer cell transcriptome (Alsafadi et al., 2016; Darman et al., 2015). SF3B is also targeted by several small-molecule splicing modulators, regarded as promising chemotherapeutic agents (Bonnal et al., 2012; Effenberger et al., 2017). At the start of this project, it was unclear how SF3B is organized prior to its incorporation in the spliceosome, what structural features of human SF3B are perturbed in cancers, and how the antitumor compounds act on SF3B to modulate splicing.

In the first part of this thesis work, we carried out a thorough structural analysis of the human SF3B complex (Cretu et al., 2016). Firstly, we have defined a structurally stable SF3B core complex (~254 kDa), composed of SF3B1's C-terminal HEAT domain, SF3B3, SF3B5, and

PHF5A, and determined its structure by X-ray crystallography. The crystal structure of the SF3B core complex revealed that the 20 HEAT repeats of SF3B1 adopt a distinctive superhelical conformation and share extensive contacts with the other three core subunits (Cretu et al., 2016). SF3B3 exhibits a triple β-propeller fold and accommodates the three alpha helices of the SF3B5 subunit in a deep, clam-shaped cleft (Cretu et al., 2016). Organized as a compact knot composed of three zinc finger motifs, PHF5A bridges the terminal repeats of SF3B1's HEAT domain, contributing to the unique conformation of the superhelix. Using a set of orthogonal mass spectrometry approaches, we showed that SF3B1-PHF5A together with the more mobile SF3B6/p14 subunit form a multipartite RNA binding platform that, in spliceosomes, stabilizes the U2/BS helix and the downstream 3' end of the intron (Fica and Nagai, 2017; Shi, 2017). Comparative analyses with recent cryo-EM structures of yeast spliceosomes (Fica and Nagai, 2017) show that the cancer-related residues of SF3B1 map to a basic groove of the HEAT superhelix where, likely, the 3' pyrimidine-rich region of the intron binds. Altogether, our analyses suggest how changes in SF3B1 structure and interactome may lead to a compromised BS selection, thus providing insights into the molecular mechanism of SF3B1-driven cancers.

In the follow-up work (Cretu et al., 2018, *unpublished data*), we determined co-crystal structures of SF3B core variants in complex with different compounds that modulate splicing, including some approved for clinical trials. Our work shows that splicing modulators from the pladienolide and herboxidiene families target SF3B at the same site and bind to an hourglass-shaped tunnel formed by SF3B1's H15-H17 repeats and PHF5A. Their molecular recognition is achieved in part by shape complementarity to the tunnel, enforced by the conjugated diene group – a moiety regarded as the common pharmacophore of SF3B modulators. Importantly, while SF3B1 exhibits a "closed" conformation in fully assembled spliceosomes, we observe a more "open" state in the

presence of splicing modulators. Structural comparisons indicate that the modulator binding site, available in the "open" conformation, is rearranged in the "close" state of SF3B1 to accommodate the invariant BS adenosine. Thus, our analyses suggest that splicing modulators interfere with a conformational rearrangement of SF3B1 and, in this respect, act as competitive BS antagonists. Overall, this work may serve as a conceptual framework for the structure-based design of next-generation splicing modulators.

2 Introduction

2.1 The central dogma of molecular biology

The DNA macromolecule is central to life as we know it. In all cellular life forms, the genetic information, required to build entire molecules, organelles, and cells, is encoded as a sequence of nucleotides in the DNA (Watson et al., 2017). However, as a sophisticated computer program, the DNA is the raw source code defined by syntax and formalisms, composed of more autonomous subroutines (or "genes"), and separated by comments (or "noncoding regions"). To read and interpret the "code" a "compiler" is needed. In cells, various molecular machines cooperate to emulate the "compiler." RNA polymerases "read" the DNA "code" and "transcribe" it into an RNA intermediate, so that other cellular machines, such as ribosomes could accurately translate ("execute") it into proteins (**Figure 2.1**). Importantly, the end-result of "program's" execution (i.e., proteins) cannot be used to generate the "source code" (DNA) or the intermediate "code" (RNA) (Watson et al., 2017). The unidirectional flow of genetic information from DNA, to RNA and proteins, has stood the test of time and is universally valid in all life forms, from *Mycoplasma* to yeast and humans (**Figure 2.1**).

2.2 Split genes and pre-mRNA splicing

In eukaryotes, different types of RNA polymerases "read" the different "subprograms" (i.e., genes) of the DNA "code". RNA polymerase I transcribes ribosomal RNA genes (rRNA 5.8S, 18S, 28S) which account for ~80% of the total RNA species in cells (Vannini and Cramer, 2012; Warner, 1999). RNA polymerase III synthesizes tRNAs, the U6 snRNA, and rRNA 5S (Khatter et al., 2017). Importantly, RNA polymerase II (Pol II) "reads" all the protein-coding genes as well as the genes coding for all the other small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), and microRNAs (miRNAs) (Hirose and Manley, 2000; Sainsbury et al., 2015). However, not all sequences of a protein-coding gene will be equally translated into the amino acid sequence of a polypeptide chain; rather, most eukaryotic genes are split/mosaic. Typically, the pre-messenger RNA (pre-mRNA) molecule synthesized by Pol II consists of coding sequences, termed exons, interspersed by noncoding regions or introns (**Figure 2.1**).

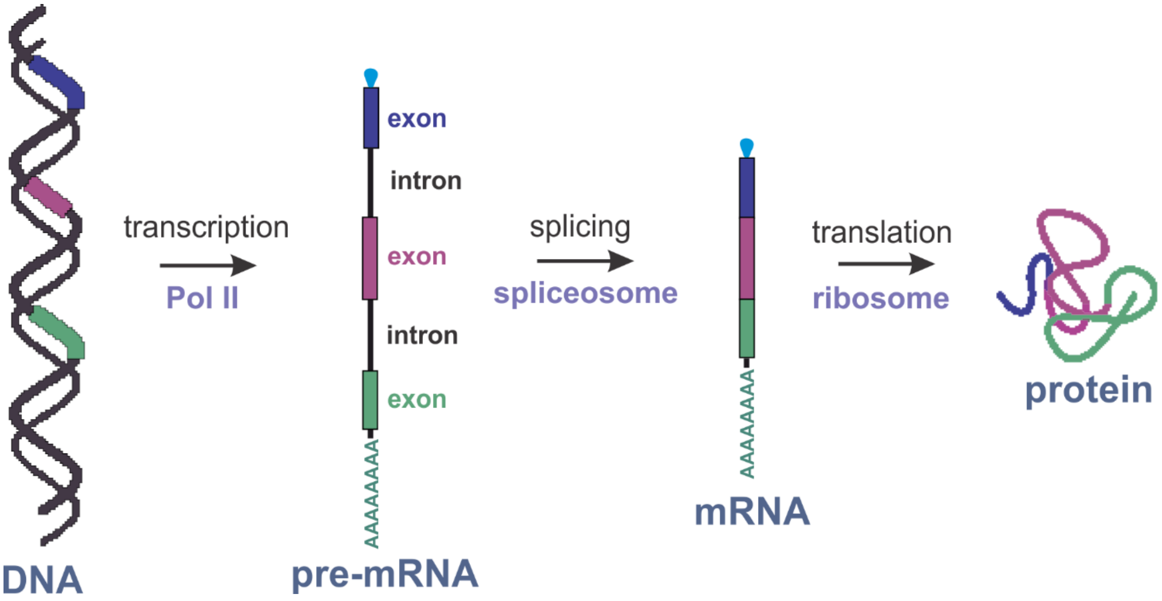


Figure 2.1. The central dogma of molecular biology.

The schematic depicts the unidirectional flow of genetic information from DNA, to RNA, to proteins. A typical eukaryotic gene is a mosaic of protein-coding sequences (exons) and non-coding regions (introns). The non-coding introns are removed from the pre-messenger RNA (pre-mRNA) by the spliceosome. At the same time, the newly transcribed pre-mRNA is polyadenylated and capped. The mature mRNA can then be translated into proteins by ribosomes in the cytoplasm.

During pre-mRNA splicing, the noncoding introns are removed from the pre-mRNA, and each pair of exons are ligated to generate the mature messenger RNA (mRNA) (Will and Luhrmann, 2011). The process is rather complicated and is catalyzed in cells by the spliceosome, a megadalton-size cellular machine composed of five uridine-rich small nuclear ribonucleoproteins (snRNPs) – so-called U1, U2, U4, U5, and U6 – and numerous associated non-snRNP factors

(Wahl et al., 2009). In contrast to other molecular machines, such as RNA polymerases or ribosomes, the spliceosomes form *de novo* and in a stepwise manner onto each intron-exon pair of the pre-mRNA and do not exist as fully assembled complexes in the absence of their substrate (Will and Luhrmann, 2011). In a larger cellular context, spliceosome assembly can be coupled to Pol II transcription (termed co-transcriptional splicing) or may occur after pre-mRNA is fully transcribed, capped, and polyadenylated (post-transcriptional splicing) (Saldi et al., 2016).

Introns account for a large part of the transcribed pre-mRNA and their length and number increase with the complexity of the organism (Irimia and Roy, 2014; Roy and Gilbert, 2006). For instance, an average human protein-coding gene has ~27 kb, and only ~5% of its entire sequence consists of coding exons (International Human Genome Sequencing, 2004; Venter et al., 2001). Thus, it is a particularly challenging task to accurately distinguish the shorter coding exons from the significantly larger introns in different sequence contexts (Wahl et al., 2009). Typically, intron-exon junctions are defined in *cis* by relatively short consensus sequences, known as splice sites (SS) (Will and Luhrmann, 2011). While in yeast (*Saccharomyces cerevisiae*) these consensus motifs are well-conserved amongst different introns, in metazoans and plants the splice sequences are more degenerated (Wahl et al., 2009; Will and Luhrmann, 2011).

The 5' exon-intron border is defined by the 5' splice site (5'SS); it has the 5'-**GU**RAGU-3' consensus sequence, where R represents any purine, in the majority of human introns (Will and Luhrmann, 2011) (**Figure 2.2**). In yeast, the 5'SS is extremely well-conserved and features the 5'-**GU**AUGU-3' sequence motif (Staley and Guthrie, 1998). The 5'SS region is complementary to the 5' end of the U1 snRNA which precedes the first stem-loop (Staley and Guthrie, 1998). The intron-3' exon border is specified by a conserved sequence motif, known as the 3' splice site (3'SS). Largely, the 3'SS is defined by the 5'-Y**AG**-3' sequence, where Y represents any pyrimidine, in both yeast and human (Wahl et al., 2009; Will and Luhrmann, 2011). A uridine-rich sequence motif usually precedes the 3'SS. This sequence is known in metazoans as the polypyrimidine tract (PPT) (Garcia-Blanco et al., 1989; Reed, 1989) (**Figure 2.2**). In higher eukaryotes, both the sequence composition and the length of the PPT are important for the accurate selection of the branch-site (BS) region (Bessonov et al., 2010; Norton, 1994; Roscigno et al., 1993). The BS sequence element (**Figure 2.2**) is located ~10-60 nucleotides (nts) upstream of the 3’SS (Black et al., 1985; Chabot et al., 1985; Taggart et al., 2017). Importantly, the BS interacts via base-pairing with the branch-site interacting region (BSiR) of the U2 snRNA (Query et al., 1994; Wu and Manley, 1989; Zhuang and Weiner, 1989). As a result, the BS adenosine (BS-A), a site which is universally conserved across different species, is bulged out from the U2/BS duplex (Query et al., 1994; Wahl et al., 2009). The degeneracy of the BS in metazoans (i.e., 5’-YUN**A**Y-3’ consensus sequence, where N represents any nucleotide) (**Figure 2.2**) explains the more stringent requirement for additional *cis* sequences, such as the PPT and exonic enhancers (Corvelo et al., 2010; Gao et al., 2008; Taggart et al., 2017). Additional *trans*-acting factors are also required to ensure accurate BS recognition (Lee and Rio, 2015). In yeast, the BS is conserved and has the 5’-UACUA**A**C-3’ consensus (Staley and Guthrie, 1998).

A minority of human introns are excised from pre-mRNA by the minor spliceosome. The U12-type introns, compared to the more common U2-type, are specified by alternative intron-exon borders: 5’-A/GU instead of 5’-GU, and AC/G-3’, instead of AG-3’ (Patel and Steitz, 2003) (**Figure 2.2**). Interestingly, the U12-type introns lack the PPT region and have a more conserved, yeast-like, BS sequence (5’-UCCUUA**A**C-3’) (Patel and Steitz, 2003). Most importantly, the U12-type 5’SS base-pairs with the U11 snRNA, whereas the BS region interacts with the U12 snRNA, both of which are part of the preassembled U11/U12 di-snRNP (Frilander and Steitz, 1999; Patel and Steitz, 2003).

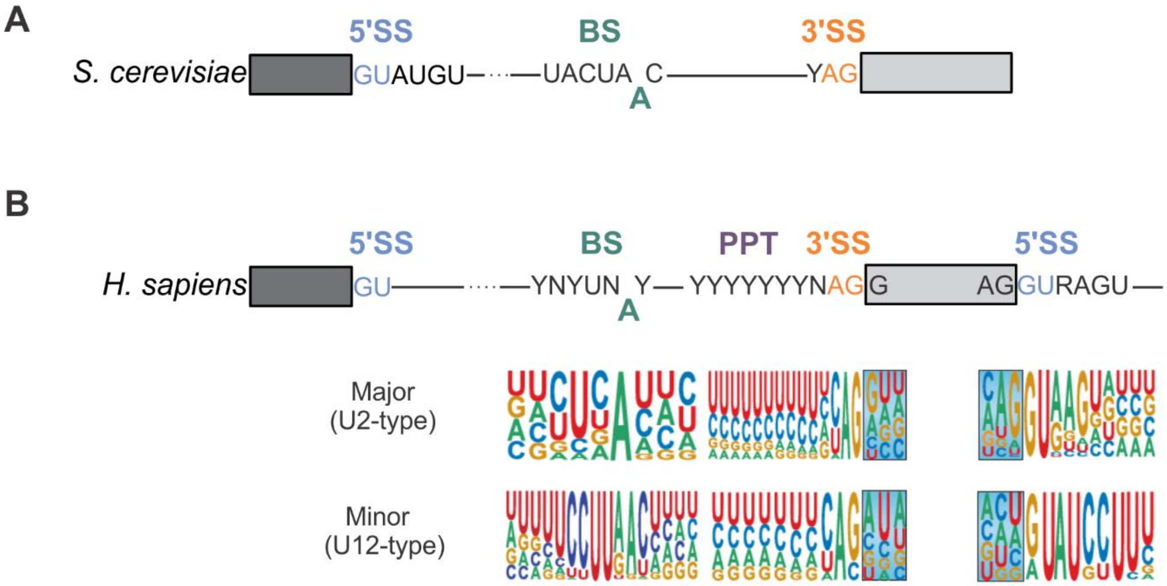


Figure 2.2. Introns are defined in *cis* by conserved splice sites.

(**A**) The typical consensus splice sequences in yeast (*S. cerevisiae*). Importantly, the 5'SS (5' splice site, light blue), the branch-site (BS, green), and the 3'SS (3' splice site, orange) are, generally, well-conserved across different introns. The invariant branch-site adenosine is colored in green. The coding exons are represented as gray boxes. (**B**) The splice site sequences are more variable in human (*H. sapiens*) introns. In metazoans, in addition to the 5'SS, the 3'SS, and the BS, introns, typically, contain a polypyrimidine tract (PPT, dark purple), located upstream of the 3'SS. The U12-type introns have largely different consensus sequences and are spliced by the minor spliceosome. The sequence logo plots for the U2-type and U12-type introns depict the relative frequencies of each nucleobase at the corresponding splice site, and are adapted from (Scotti and Swanson, 2016).

2.3 The chemistry of pre-mRNA splicing

Despite the overall complexity of the process, the general chemistry of the splicing reaction is rather simple. The excision of introns from pre-mRNA proceeds through two ordered S_N2-type transesterification reactions which involve reactive groups from the *cis*-acting splice sequences (Will and Luhrmann, 2011) **(Figure 2.3**). These chemical reactions are isoenergetic and progress through a trigonal bipyramidal transition state stabilized by two catalytic metal ions (Fedor and Williamson, 2005).

During the first transesterification reaction, called branching, the 2'-OH of the invariant BS adenosine acts as the nucleophile group and attacks the 5'SS to generate the cleaved 5'exon and an intron-3'exon lariat intermediate (Will and Luhrmann, 2011) (**Figure 2.3**). In the step 1 lariat intermediate, the BS-A is linked to the cleaved 5'SS via a 2'-5' phosphodiester bond, thus resulting in an apparent three-way junction. In the second step of splicing, known as exon ligation, the 3'-OH group of the excised 5' exon performs a nucleophilic attack on the 3'SS (Will and Luhrmann, 2011). As a result, the two consecutive exons are ligated, and the intron is liberated as a lariat intermediate (**Figure 2.3**).

For the pre-mRNA splicing catalysis to occur, the chemically reactive groups from the 5'SS, the 3'SS, and the BS have to be brought in proximity. However, eukaryotic introns lack inherent sequence conservation, except for the consensus SSs. (Will and Luhrmann, 2011). Thus, eukaryotic introns cannot adopt *per se* similar structures in isolation. In contrast, group II self-splicing introns, from which, likely, eukaryotic introns derived, form complex structures in the absence of additional factors (Pyle, 2016; Zhao and Pyle, 2017). By the intrinsic folding of the group II intron, the reactive groups of the SSs, which are strikingly similar to the eukaryotic consensus sites, are precisely positioned within an RNA-only catalytic active center (Pyle, 2016).

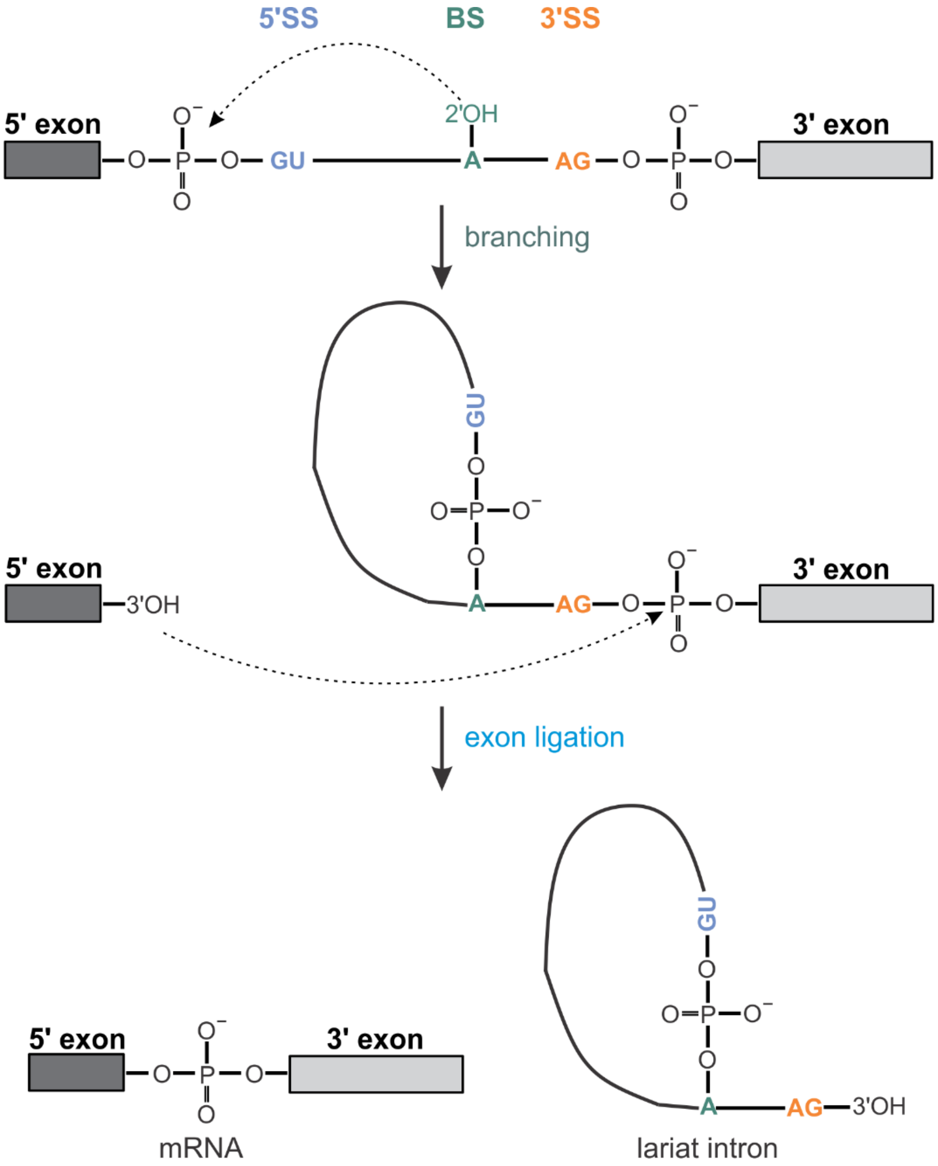


Figure 2.3. The chemistry of pre-mRNA splicing.

Introns are excised from pre-mRNA via two sequential transesterification reactions, known as branching and exon ligation, respectively. During branching, the 2'-OH group of the invariant branch-site (BS, green) adenosine performs a nucleophilic attach on the 5' splice site (5'SS, light blue). During exon ligation, the 3'-OH group of the liberated 5' exon attacks the 3' splice site (3'SS, orange), releasing the intron as a lariat intermediate and joining the two exons. Exons are depicted as gray boxes.

In eukaryotes, it is the spliceosome that coordinates, in *trans*, the accurate recognition of the SSs and the "folding" of the intron within an intricate RNA-RNA network whereby the reactive groups are activated sequentially (Wahl et al., 2009). Importantly, the dynamic organization of the spliceosome complex results in a significant degree of flexibility, so that introns of different sizes and sequences can be equally well recognized and spliced. For instance, the efficient splicing of short introns (less than ~200 nts) can be achieved through a "cross-intron" assembly of the spliceosome (Fox-Walsh et al., 2005) – the typical assembly pathway in yeast. Conversely, long introns are specified across the exons which are significantly shorter in mammals (~120 nts) (Ast, 2004; Berget, 1995; Deutsch and Long, 1999).

2.4 snRNPs – the building blocks of the spliceosome

The five uridine-rich U1, U2, U4/U6, and U5 snRNPs represent the core constituents of the U2-type major spliceosome (Wahl et al., 2009). They are composed of an RNA moiety, the U snRNA, and a specific set of proteins (**Figure 2.4**). The U12-type spliceosome, which splices a minority of mammalian introns, assembles from a specific set of snRNPs: the U11 snRNP, the U12 snRNP, the U4atac snRNP, and the U6atac snRNP (Patel and Steitz, 2003). The U5 snRNP is present in both minor and major spliceosomes. The snRNA moiety of each snRNPs has a characteristic secondary structure, is posttranscriptionally modified and capped (Matera and Wang, 2014) (**Figure 2.4**). Although the overall structures are different, U1, U2, U4, and the U5 snRNA share a common uridine-rich site, the so-called Sm site, onto which the heteroheptameric Sm ring complex is loaded (Matera and Wang, 2014). The Sm protein ring assembles sequentially around the Sm site from SmB/B', SmD1, SmD2, SmD3, SmE, SmF, and the SmG subunit (Kondo et al., 2015; Li et al., 2016; Neuenkirchen et al., 2008). In contrast, the U6 snRNA lacks the typical Sm site. Instead, a specific uridine-rich sequence at the very 3' end of the U6 snRNA serves as the docking site for the Sm-like LSm complex (Didychuk et al., 2018). The heteromeric LSm ring forms from LSm2, LSm3, LSm4, LSm5, LSm6, LSm7, and LSm8 (Zhou et al., 2014). Importantly, the U4 and U6 snRNA base-pair to each other in the U4/U6 di-snRNP and the U4/U6.U5 tri-snRNP (Didychuk et al., 2018; Wahl et al., 2009) **(Figure 2.4)**. However, upon the activation of the spliceosome, this base-pairing interaction is disrupted to liberate the U6 intramolecular stem-loop (ISL).

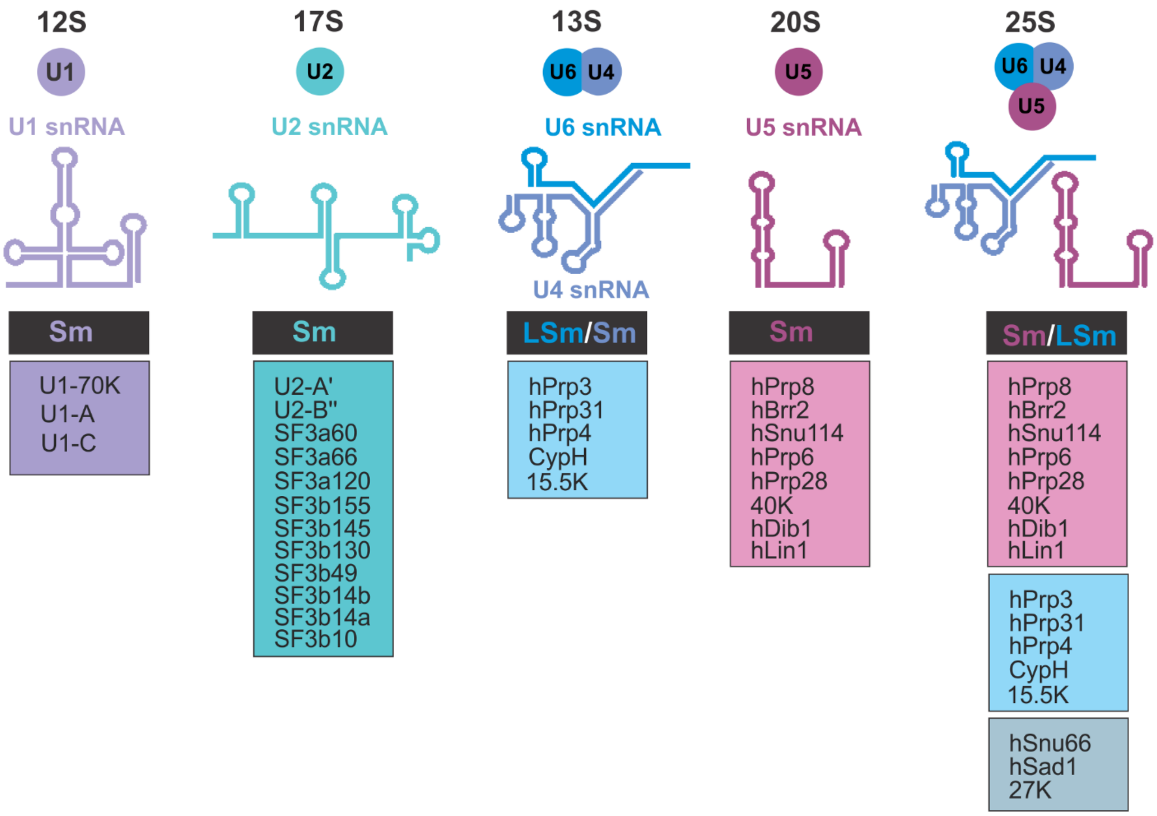


Figure 2.4. Protein-RNA composition of human snRNPs from the major spliceosome.

Each snRNP is composed of a snRNA moiety, the Sm or LSm ring complex, and a specific set of proteins. The secondary structure of each snRNA is schematically indicated. The protein nomenclature is as in (Fabrizio et al., 2009; Wahl et al., 2009).

In addition to the Sm or Lsm ring, each of the snRNPs has a characteristic set of proteins (**Figure 2.4**). The U1 snRNP specific proteins U1-70K and U1-A play mainly architectural roles and interact with the stem-loop I and stem-loop II of the U1 snRNA, respectively (Kondo et al., 2015). Instead, U1-C is positioned at the 5' end of the U1 snRNA and, likely, proofreads/tunes its interaction with the 5'SS (Kondo et al., 2015) (**Figure 2.5**).

The human 17S U2 snRNP particle has a distinctive modular architecture (**Figure 2.4**; **Figure 2.5**). The U2 snRNA moiety serves as an "extended scaffold" for the essential multimeric splicing factors SF3A, SF3B, U2-A'-U2-B'', and the Sm ring (Kramer et al., 1999; Will et al., 2002; Plaschka et al., 2017). Additional splicing factors co-purify with these major U2 components (Will et al., 2002). These include two RNA helicases (DDX46/Prp5, DDX15/Prp43), U2AF1, U2AF2, PUF60, as well as several other proteins present in substoichiometric amounts (Will et al., 2002). The heteroheptameric SF3B complex is located mainly at the 5' end of the U2 snRNA, close to the stem-loop IIa and stem-loop IIb (Dybkov et al., 2006; Plaschka et al., 2017). The conserved Sm ring and the U2-A'-U2-B'' dimer form a second distinct module and reside at the 3' extremity of the U2 snRNA (Bertram et al., 2017a; Kramer et al., 1999; Plaschka et al., 2017). The heteromeric SF3A complex (SF3A1/SF3a120, SF3A2/SF3a66, SF3A3/SF3a60) acts as a "protein bridge" between these two structural modules of the U2 snRNP (Kramer et al., 1999; Plaschka et al., 2017).

Although the 13S U4/U6 di-snRNP and the 20S U5 snRNP can be isolated as stable complexes from the HeLa nuclear extract, these snRNPs are recruited to the spliceosome as a preassembled particle – the 25S U4/U6.U5 tri-snRNP (Wahl et al., 2009; Will and Luhrmann, 2011). Recent cryo-EM analyses of yeast (Nguyen et al., 2016; Wan et al., 2016b) and human (Agafonov et al., 2016) tri-snRNP particles revealed the structural organization of this core building block of the spliceosome (**Figure 2.4**; **Figure 2.5**). The overall structure of the yeast U4/U6.U5 tri-snRNP resembles a right-angled triangle with the U5 Sm, U4 Sm, and U6 LSm protein rings positioned at its three vertices (Nguyen et al., 2016; Wan et al., 2016b) (**Figure 2.5**). The U5 snRNA occupies the bottom-half of the tri-snRNP, the “foot” domain, and adopts a complex Y-shaped tertiary structure. The U5 stem I and stem II regions form a pseudo-continuous duplex, interrupted by the variable stem region (Nguyen et al., 2016; Wan et al., 2016b). The downstream stem III region of U5 is oriented at ~90^o versus stem IV and continues with the Sm site which threads through the donut-shaped U5 Sm ring (Nguyen et al., 2016; Wan et al., 2016b). The U4/U6 di-snRNP is located in the upper-half of the tri-snRNP and forms the “head” and the “arm” domains, where the U4 snRNA and the U6 snRNA moieties are extensively base-paired along the longitudinal axis of the assembly (Nguyen et al., 2016; Wan et al., 2016b). As a result, the intermolecular stem I and stem II, formed between the U4 snRNA and the U6 snRNA, are connected to the U4 5’-stem-loop in a three-way junction (**Figure 2.5**). The U4 snRNA further extends into the “head” domain, where the 3’ region forms an intramolecular 3’-stem-loop that precedes the Sm site and the U4 Sm ring (**Figure 2.5**). Importantly, the 5’ end of the U6 snRNA projects towards the U5-stem I/loop 1 and forms two additional intramolecular stem-loops – the short ACAGAGA stem-loop and the 5’-stem-loop (Nguyen et al., 2016; Wan et al., 2016b). Thus, in the tri-snRNP, the key catalytic regions of the U6 snRNA, the ISL and the ACAGAGA sequence, are sequestered in intermolecular (U4/U6 stem II) and intramolecular (ACAGAGA stem) secondary structure elements, respectively. Moreover, the U5 loop 1 interacts with a single-stranded region of the U6 snRNA, in the proximity of the U6 5’-stem-loop. Later on, the ACAGAGA sequence base-pairs with the 5’SS and U5 loop 1 interacts with the 5’-exon. At the same time, the U6 ISL adopts a complex

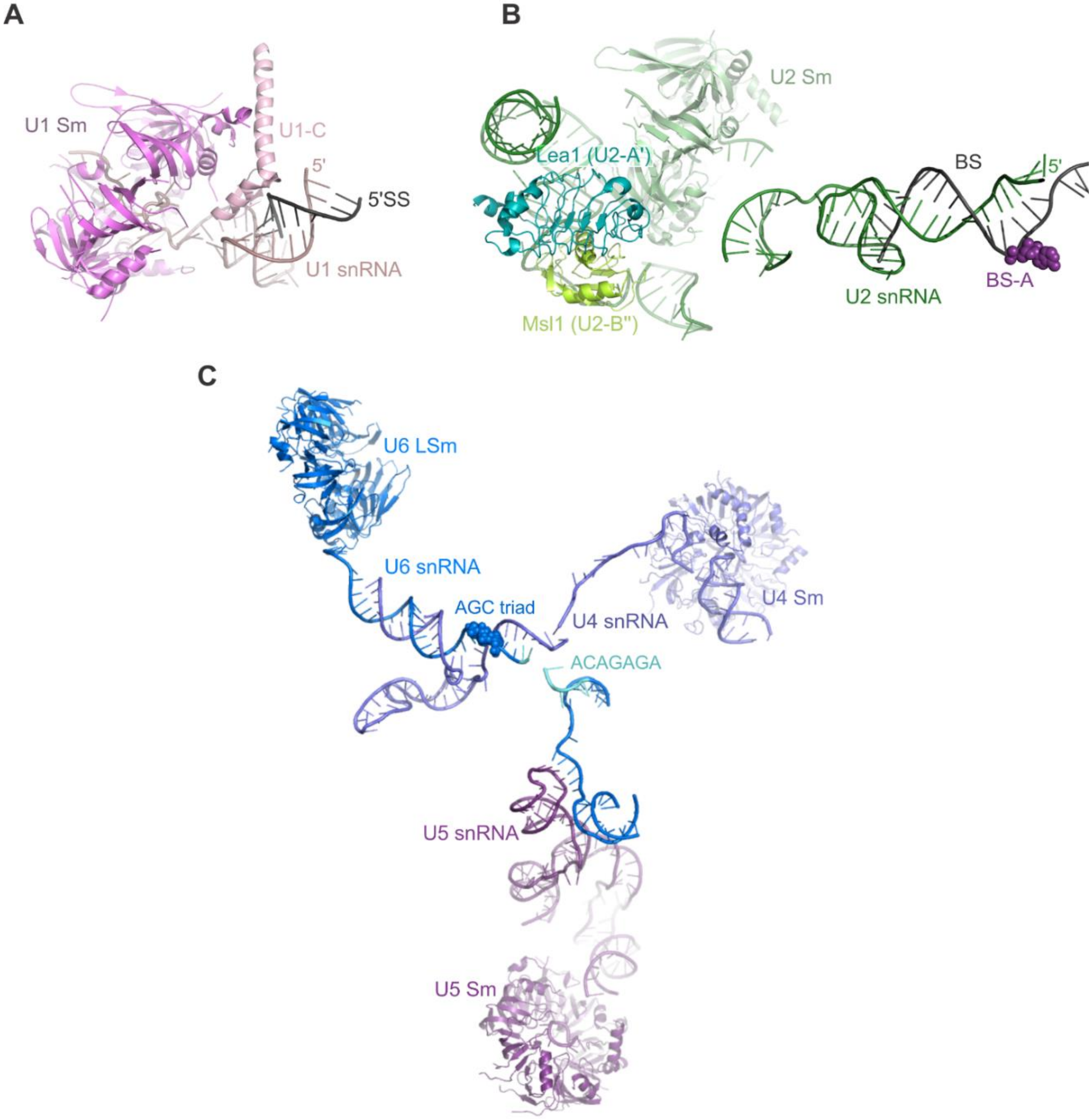


Figure 2.5. Modular structures of spliceosomal snRNPs.

(**A**) Structure of the isolated U1 snRNP core (PDB 4PJO) in complex with a 5'SS oligo. The interaction between the U1 snRNA and the 5'SS is likely stabilized/tuned by U1-C. The heptameric Sm ring is colored in magenta. (**B**) Structure of the U2 snRNP core module from the yeast B complex (PDB 5NRL) bound to the branch-site (BS) region of the intron. Note that the SF3A and SF3B complexes are not shown for simplicity. The U2/BS helix is followed by the stem-loop IIa and IIb of the U2 snRNA (green). The bulged BS adenosine (BS-A) is represented as spheres and colored in dark purple. The U2 Sm ring and the Lea1/U2-A'-Msl1/U2-B'' dimer form a distinct module at the 3' end of the U2 snRNA. (**C**) Structure of the yeast U4/U6.U5 tri-snRNP particle (PDB 5GAN). The AGC catalytic triad and the ACAGAGA sequence of the U6 snRNA are colored in blue and cyan, respectively. The U4 snRNA and the U5 snRNA are colored in light and dark purple, respectively. The tri-snRNP specific proteins are not shown for simplicity sake.

tertiary structure and resides right at the active catalytic center of the spliceosome (Galej et al., 2016; Wan et al., 2016a; Yan et al., 2016).

The cryo-EM reconstructions of the yeast particle (Nguyen et al., 2016; Wan et al., 2016b) provide additional insights into the structures and functions of the tri-snRNP specific proteins. Prp8 is the largest component of the U5 snRNP and one of the most conserved spliceosomal proteins (Galej et al., 2013; Wahl et al., 2009). It is positioned at the center of the assembly and constitutes a large part of the "body" domain. Its modular structure consists of (Nguyen et al., 2016; Wan et al., 2016b): (i) an N-terminal alpha-helical domain (N); (ii) the "Large" domain (L), composed of the helix bundle, the reverse transcriptase (RT), the linker and the endonuclease-like (EN) regions; (iii) the RNaseH-like domain (RH); (iv) the Jab1/MPN domain (Jab). In the "foot" region of the tri-snRNP, the N-terminal domain of Prp8 stabilizes the U5 snRNA and interacts extensively with Snu114 bound to GTP (Nguyen et al., 2016). The flexibly linked Jab and RH domains of Prp8 contact the N-terminal helicase domain of Brr2 and the U4/U6 helix I, respectively. Combined, these interactions bridge the "Large" domain of Prp8 and the "arm" module of U6, where also Prp4 and Snu13 reside. These tri-snRNP subunits together with the C-terminal domain of Prp3 and Prp31 stabilize the 3' end of the U6 snRNA and the U4 5'-stem-loop region (Nguyen et al., 2016; Wan et al., 2016b). The large Brr2 RNA helicase is located close to the U4 Sm ring, in the "head" domain of the tri-snRNP, and is bound to a single-stranded region of the U4 snRNA (Nguyen et al., 2016; Wan et al., 2016b). Thus, translocation of Brr2 along this region would lead to the unwinding of the U4/U6 snRNA duplex and release of the U6 ISL from the U4 snRNA "trap". Intriguingly, in the yeast tri-snRNP, the ACAGAGA stem of the U6 snRNA is "protected" by the small protein Dib1, bound to Prp8 (Nguyen et al., 2016). This, likely, provides an additional layer of regulation and, potentially, prevents the premature interaction of the ACAGAGA sequence with the 5'SS.

2.5 The splicing cycle – lessons from recent cryo-EM structures

In contrast to other molecular machines, the spliceosome assembles *de novo* onto each pre-mRNA substrate in a stepwise manner (Will and Luhrmann, 2011) (**Figure 2.6**). In metazoans, at the earliest stage of spliceosome assembly, the 5'SS is recognized by the U1 snRNP via base-pairing with the U1 snRNA moiety, in an ATP-independent manner (Liao et al., 1990; Michaud and Reed, 1991; Seraphin et al., 1988). At the same time, the splicing factors SF1/mBBP and the U2AF1-U2AF2 (also known as U2AF35-U2AF65) heterodimer bind in a cooperative manner to the 3' end of the intron to form the E complex (Abovich and Rosbash, 1997; Berglund et al., 1997) (**Figure 2.7**). The KH-QUA2 domain of SF1 interacts in a sequence-specific manner with the BS (Berglund et al., 1998; Liu et al., 2001; Peled-Zehavi et al., 2001), whereas its N-terminal ULM motif (U2AF ligand motif) contacts the C-terminal UHM (U2AF homology motif) domain of U2AF2 (Loerch and Kielkopf, 2016; Selenko et al., 2003; Wang et al., 2013) (**Figure 2.7**). Concomitantly, the two tandem RRM (RNA recognition motif) domains of U2AF2 recognize the PPT (Sickmier et al., 2006), while the N-terminal ULM motif contacts the UHM domain of U2AF1 (Kielkopf et al., 2001) (**Figure 2.7**). The UHM domain of U2AF1 is flanked by two CCCH-type zinc knuckles which recognize the 3'SS consensus sequence (Yoshida et al., 2015). Thus, while the 5'SS is specified using an RNA-RNA base-pairing interaction, the 3' intron-exon border is defined via multiple protein-RNA contacts (**Figure 2.7**). On alternatively spliced introns, a different set of 3' interacting proteins may substitute for the canonical factors (Conte et al., 2000; Handa et al., 1999; Oberstrass et al., 2005) and their recruitment is subject to regulation by additional auxiliary factors (Lee and Rio, 2015).

In the human spliceosome, the U2 snRNP is only loosely associated with the E complex (Das et al., 2000; Hong et al., 1997). However, in the presence of ATP, the E complex is converted into the next stable complex, the pre-spliceosome (or the A complex) (Bennett et al., 1992). In the pre-spliceosome, the U2 snRNA base-pairs with the BS region of the intron and the invariant BS-A is bulged out from the branch helix (Query et al., 1994; Query et al., 1996) (**Figure 2.6**). In the A complex, the displacement of SF1 from the BS is coupled to the formation of new protein-RNA interactions between the SF3A and SF3B subcomplexes of the U2 snRNP and the intron regions flanking the BS (Gozani et al., 1996; Gozani et al., 1998) (**Figure 2.7**). Additional protein-protein interactions are also established as the U2AF2 UHM domain interacts with the ULM motifs of SF3B1 (Cass and Berglund, 2006; Spadaccini et al., 2006; Thickman et al., 2006), whereas the N-terminal RS (arginine/serine-rich) region of U2AF2 stabilizes the U2/BS helix (Mackereth et al., 2005; Valcarcel et al., 1996). Importantly, the stable formation of the U2/BS duplex at the A-complex stage requires contributions from at least two RNA helicases, Sub2/UAP56 and Prp5/DDX46 (**Figure 2.6**). In yeast, Sub2/UAP56 facilitates the displacement of SF1 from the BS (Cordin et al., 2012; Libri et al., 2001), whereas Prp5/DDX46, likely, remodels the U2 snRNA itself to expose the BSiR for BS binding, in part, through the displacement of Cus2p/Tat-SF1 (Perriman and Ares, 2000; Perriman et al., 2003; Perriman and Ares, 2007; Yan et al., 1998) (**Figure 2.7**). In addition to its function as an unwindase/RNPase, Prp5/DDX46 might also bridge the U1 snRNP and the U2 snRNP across the intron (Shao et al., 2012; Xu et al., 2004).

Recruitment of the U4/U6.U5 tri-snRNP as a preassembled particle to the pre-spliceosome leads to the formation of the 37S pre-B complex, a spliceosome intermediate containing all five U snRNPs (Boesler et al., 2016). In this intermediate complex, the 5’SS is still base-paired to the U1 snRNP and the U4 snRNA interacts with the U6 snRNA via the intermolecular stem I and stem II, while the U5 loop 1 contacts the U6 snRNA, downstream of the 5’ stem-loop (Bertram et al., 2017a; Boesler et al., 2016; Plaschka et al., 2017). It is likely that at this stage, the tri-snRNP is attached to the A complex mainly via the intermolecular U2/U6 helix II and, potentially, more

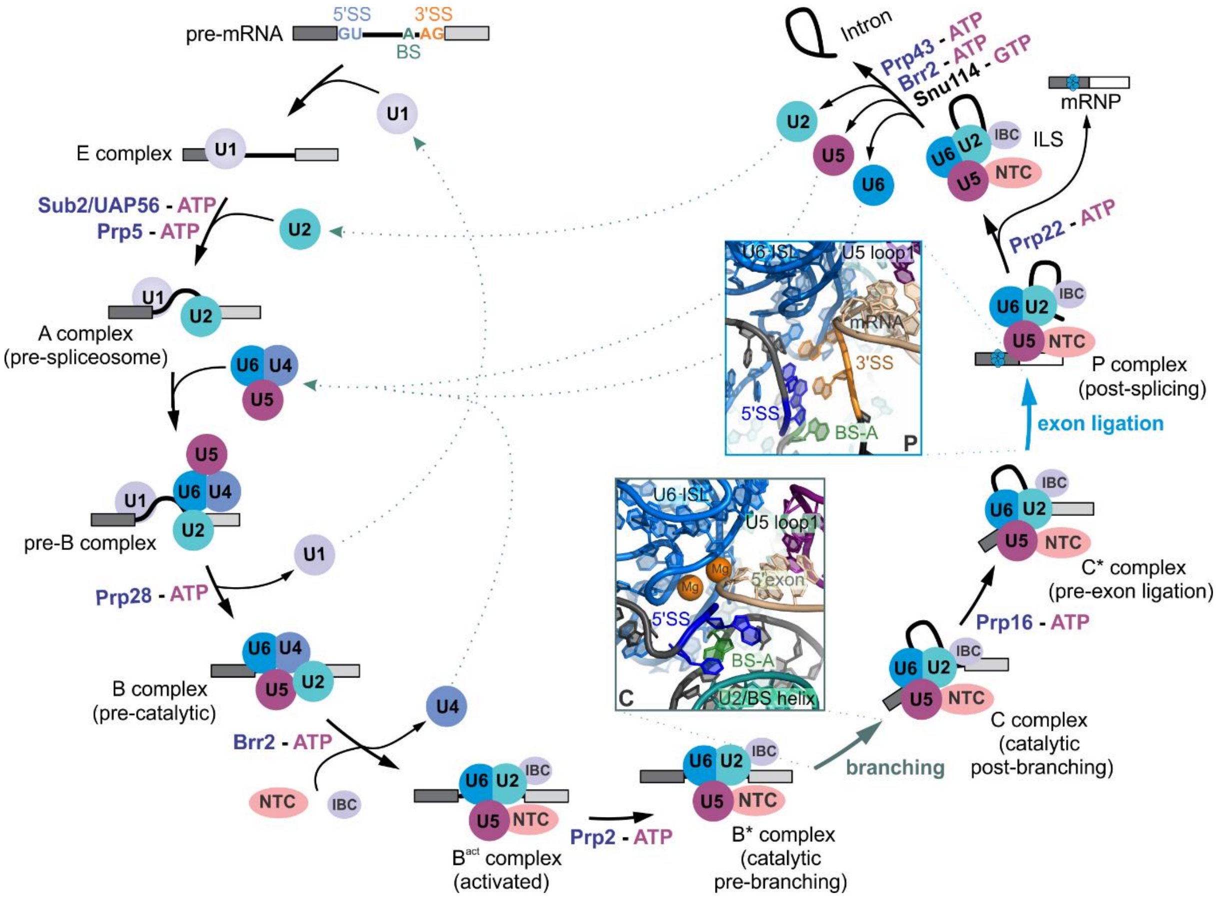


Figure 2.6. The splicing cycle and spliceosome’s dynamic composition.

The spliceosome forms onto each pre-mRNA substrate by the sequential recognition of the 5’ splice site (5’SS, light blue), the 3’ splice site (3’SS, orange), and the branch-site (BS, green) regions of each intron. The spliceosome transits through several landmark complexes: E, A, pre-B, B, B*, B^{act}, C, C*, P, and ILS (intron-lariat spliceosome). Eight spliceosomal RNA helicases (Sub2/UAP56, Prp5/DDX46, Prp28, Brr2, Prp2, Prp16, Prp22, Prp43) act at defined stages of the cycle to trigger the formation and the remodeling of the active site. After the two exons (gray boxes) are ligated, the spliceosomal snRNPs are recycled. The two insets show the organization of the catalytic active center of the spliceosome right after the branching reaction (C complex, PDB 5LJ3) and after the exon ligation step (P complex, PDB 6EXN). The two catalytic magnesium ions (M1 and M2) are indicated in the post-branching C complex and colored in orange. Only the M2 magnesium ion was observed in the post-splicing complex (Wilkinson et al., 2017) and it is not shown here for simplicity sake. NTC – NineTeen Complex, IBC – Intron-Binding Complex.

"fuzzy" protein-protein interactions (Bertram et al., 2017a; Boesler et al., 2016; Plaschka et al., 2017).

Next, the Prp28/hPrp28 RNA helicase, which in human is a core component of the U5 snRNP and the tri-snRNP (Teigelkamp et al., 1997), displaces the U1 snRNP from the 5'SS. As a result, the liberated 5'SS can interact with the ACAGAGA sequence of the U6 snRNA, to form, in human, an extended U6 ACAGAGA/5'SS helix (Bertram et al., 2017a). At the same time, the U5 loop 1 contacts the 5' exon to form the U5 loop 1/5'exon helix (Bertram et al., 2017a). Dissociation of the U1 snRNP from the pre-B complex is coupled with the recruitment of several B-specific proteins, such as RED, MFAP1, FBP21, hSmu1, hPrp38, and hSnu23, and, ultimately, leads to the formation of the 45S pre-catalytic B complex (Bertram et al., 2017a; Plaschka et al., 2017). The B-specific proteins promote a more stable integration of the tri-snRNP and reinforce the U6 ACAGAGA/5'SS helix (Bertram et al., 2017a; Plaschka et al., 2017). Importantly, at this stage, the U2/U6 catalytic core of the spliceosome is not yet formed, and the U2/BS helix is located at ~150 Å from the 5'SS (Bertram et al., 2017a; Plaschka et al., 2017).

During spliceosome's activation, the pre-catalytic B complex undergoes dramatic structural and compositional rearrangements to form the B^{act} complex (**Figure 2.6**). In part, the B to B^{act} transition is triggered by the Brr2 RNA helicase which unwinds the U4/U6 intermolecular helices. As a result, the U4 snRNP and most of the tri-snRNP and B-specific proteins dissociate from the complex. Concomitantly, the liberated U6 snRNA refolds to form the characteristic ISL. Downstream of the U6 ISL, the U6 snRNA base-pairs with the U2 snRNA to form the intermolecular helices Ia and Ib (Fica and Nagai, 2017; Shi, 2017). In yeast, the bulged nucleotides of the ISL adopt a contorted conformation and together with the AGC catalytic triad, base-paired to U2 in the U2/U6 helix Ib, coordinate the two catalytic magnesium ions (**Figure 2.6**). The RNA catalytic core of the spliceosome, formed at this stage, will perform the two phosphoryl transfer reactions (i.e., branching and exon ligation), by the two-metal ion mechanism, without further rearrangements (Fica and Nagai, 2017). In the B^{act} complex, the U6 ACAGAGA/5'SS helix has already docked into the catalytic core and the 5' exon is anchored by the U5 loop 1. However, the U2/BS helix and the reactive BS-A are located ~50 Å away from the active site, "trapped" by the SF3B complex in a molecular "clamp" (Rauhut et al., 2016; Yan et al., 2016). The rearrangements of the RNA-RNA network of the spliceosome are also coupled with the recruitment of additional B^{act} specific factors (Cwc24, Cwc27), and of the RNA helicase Prp2 with its cofactor Spp2 (**Figure 2.6**). The multimeric Prp19 complex (also known as NTC), the Prp19-related factors (Rauhut et al., 2016; Yan et al., 2016), and, in higher eukaryotes, the intron-binding complex (IBC) (De et al., 2015), are also recruited at this stage. These factors, likely, help fasten the RNA catalytic core onto Prp8's RT, EN, and N domains.

For the branching reaction to occur the U2/BS helix has to dock into the active site. The result is the formation of the B^* complex, where the first-step reaction takes place (Warkocki et al., 2009; Warkocki et al., 2015). Intriguingly, the cryo-EM structures of yeast B^{act} complexes show that the Prp2 RNA helicase is positioned at the 3' end of the intron, close to the SF3B complex (Rauhut et al., 2016; Yan et al., 2016). The strategical location of the helicase suggests that it may "act a distance" to trigger the release of the branch helix (Rauhut et al., 2016; Yan et al., 2016). In turn, the SF3A and SF3B complexes are destabilized, and the step-one factors (Cwc25, Yju2, Isy1) join to fix the U2/BS helix at the active site (Galej et al., 2016; Warkocki et al., 2009; Warkocki et al., 2015). Interestingly, the catalytic activation step might be more complicated in higher eukaryotes (Haselbach et al., 2018).

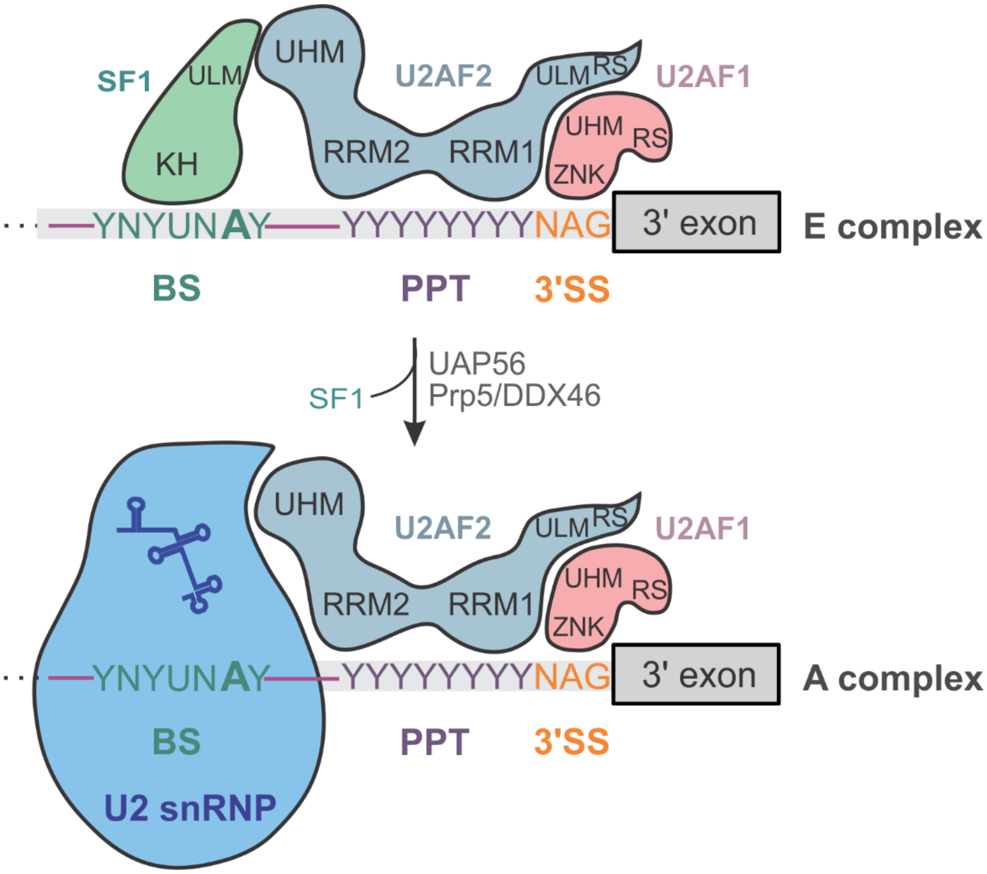


Figure 2.7. Stepwise recognition of the 3' end of the intron.

During E complex formation, the 3' end of the intron is recognized in a cooperative manner by SF1, U2AF2, and U2AF1. SF1 binds to the branch-site (BS, green), U2AF2 interacts with the polypyrimidine tract (PPT, dark purple), while U2AF1 contacts the 3' splice site (3'SS, orange). In the presence of ATP, UAP56 and Prp5/DDX46 RNA helicases, SF1 is displaced from the BS and the U2 snRNP is recruited instead to the BS region of the intron. Note that the U2 snRNP is not drawn to scale and only the human consensus sequences are shown. The schematic is, in part, adapted from (Loerch and Kielkopf, 2016). ULM – U2AF ligand motif, UHM – U2AF homology motif, ZNK – zinc knuckle, RS – arginine/serine-rich.

After the branching reaction occurs, the resulting C complex is remodeled for the 3' exon to be docked into the active site (**Figure 2.6**). Rearrangement of the C complex depends on the Prp16 RNA helicase. The conserved helicase acts downstream of the BS to induce the release of the branching factors (Galej et al., 2016; Wan et al., 2016a). In the remodeled complex, the C^* complex, the U2/BS helix is displaced away from the active site with help from the step-two factors Slu7 and Prp18, Prp17, and Prp8's RH domain (Galej et al., 2016). As a result, the 3'SS docks at the active site, where, intriguingly, its conserved AG dinucleotide engages in non-Watson-Crick interactions with both the 5’SS and the BS-A (Bai et al., 2017; Liu et al., 2017; Wilkinson et al., 2017) (**Figure 2.6**). After the exon ligation takes place, the exon-junction complex is deposited onto the mature mRNA which is then released from the post-catalytic P complex through the action of the Prp22 RNA helicase (Bai et al., 2017; Bertram et al., 2017b; Liu et al., 2017; Wilkinson et al., 2017). The resulting intron lariat spliceosome (ILS) is further dismantled by the Prp43 RNA helicase and its cofactors, and U2, U5, and the U6 snRNP are recycled (Wan et al., 2017; Yan et al., 2015).

2.6 The multimeric SF3B complex and the recognition of the branch-site region

The U2 snRNP is critical for the accurate selection of the BS region of the intron and changes in the function of its components, and related splicing factors are associated with many common diseases (Wahl et al., 2009; Wang and Cooper, 2007). SF3B is the largest U2 subcomplex and plays important roles in the recognition and stabilization of BS during constitutive and alternative splicing (Gozani et al., 1996; Gozani et al., 1998; Will et al., 2002) (**Figure 2.8**). It consists of seven different subunits (SF3B1/SF3b155, SF3B2/SF3b145, SF3B3/SF3b130, SF3B4/SF3b49, SF3B5/SF3b10, SF3B6/p14, and PHF5A/SF3b14b) and has an overall molecular weight of ~ 450 kDa (Golas et al., 2003; Will et al., 2002). An additional subunit, the DDX42/SF3b125 DEAD-box RNA helicase, is substoichiometrically associated with the complex prior to its incorporation into the U2 snRNP (Uhlmann-Schiffler et al., 2006; Will et al., 2002).

As the heteromeric complex encompasses long hydrophilic and repetitive sequences that are structurally disordered (Cass and Berglund, 2006; Thickman et al., 2006), and which account for approximately 39% of its total mass (**Figure 2.9**), it represents a challenge for structural analysis by X-ray crystallography and cryo-EM. Previous EM analyses using cryo-negative staining showed that the heptameric complex has a compact architecture, with several protruding elements projecting away from a hollow “bivalve-like” shell (Golas et al., 2003). Although the overall quality of the reconstruction precluded accurate model building, three of the SF3B subunits, SF3B1, SF3B4, and SF3B6/p14, were tentatively assigned to several defined map regions (Golas et al., 2003).

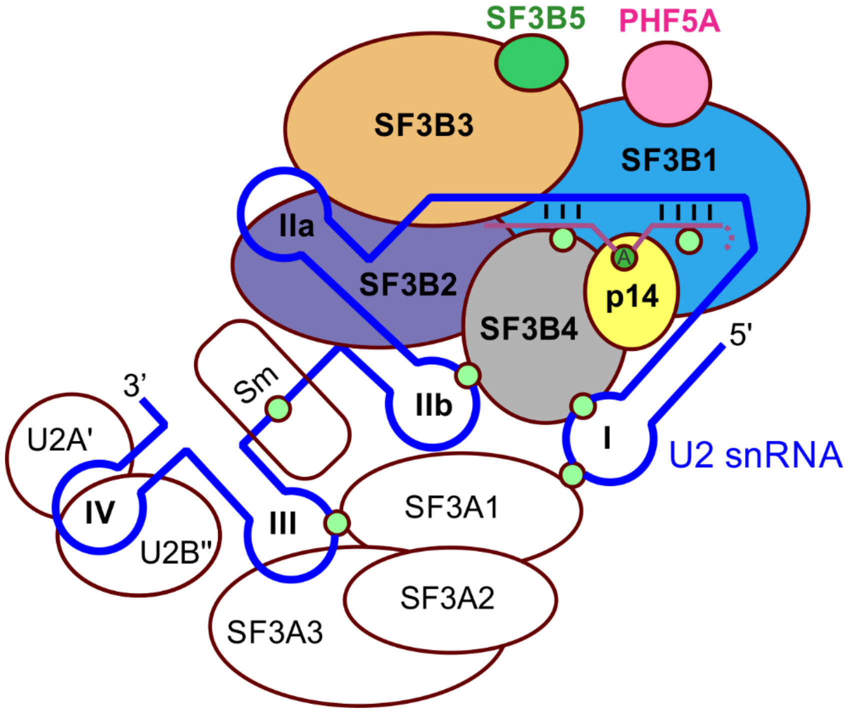


Figure 2.8. SF3B contacts both the U2 snRNA and the intron at the branch-site.

The schematic depicts an early view of the roles of SF3B subunits in the recognition of the branch-site (BS) and of the U2/BS helix. The model was derived mainly from protein-RNA crosslinking experiments (the crosslinked regions are marked with green circles). According to this model, the p14/SF3B6 subunit interacts directly with the BS adenosine (BS-A, dark green), whereas SF3B1 contacts intron regions flanking the BS-A. IIa – stem-loop IIa, IIb – stem-loop IIb; III – stem-loop III; IV – stem-loop IV; Sm – the Sm ring complex.

SF3B1, the largest subunit of the SF3B complex, comprises an N-terminal extended region (NTD) and a conserved C-terminal HEAT (Huntingtin, Elongation Factor 3, Protein phosphatase 2A, Target of rapamycin 1) domain (Wang et al., 1998) (**Figure 2.9**). Although essential for pre-mRNA splicing, the NTD of SF3B1 (SF3B1^{NTD}) is largely unstructured in isolation (Cass and

Berglund, 2006; Thickman et al., 2006). The extended conformation of the domain facilitates the multiple interactions of its linear, tryptophan-rich, ULM sequence motifs with different UHM domain containing splicing factors, such as U2AF2/U2AF65 (Cass and Berglund, 2006; Spadaccini et al., 2006; Thickman et al., 2006), SPF45/RBM17 (Corsini et al., 2007), and CAPERα/RBM39 (Loerch et al., 2014). The small SF3B subunit SF3B6/p14 also interacts with a sequence motif in SF3B1^{NTD} (residues 373-415), the so-called p14-interacting region (pir) (Schellenberg et al., 2006; Spadaccini et al., 2006), located downstream of the ULM motifs (**Figure 2.9**).

SF3B1^{NTD} also undergoes transient phosphorylation during spliceosome's catalytic activation (Bessonov et al., 2010; Wang et al., 1998). The posttranslational modification of SF3B1^{NTD} occurs after the formation of the pre-catalytic spliceosome (Bessonov et al., 2010) and is, likely, mediated by its interaction with Cyclin E-Cdk2 (Shi et al., 2006). The precise Ser/Thr phosphorylation sites were mapped by mass spectrometry to a repetitive linear motif (residues 207-374) within SF3B1^{NTD} (Girard et al., 2012; Wang et al., 1998). However, as this region could not be captured in the structures of human spliceosomes, its relation to other SF3B subunits or other splicing factors is presently unclear. Intriguingly, the hyperphosphorylated SF3B1 appears to associate with chromatin in an RNA-dependent manner (Girard et al., 2012). Recently, direct interactions between SF3B1 and the exonic chromatin have been reported (Kfir et al., 2015). Thus, it is plausible that, in cells, phosphorylation of SF3B1^{NTD} triggers the recruitment of the U2 snRNP from exonic nucleosomes to the nascent transcripts (Hollander et al., 2016). In turn, this might promote the co-transcriptional assembly of spliceosomes (Girard et al., 2012; Hollander et al., 2016; Kfir et al., 2015).

Previous sequence predictions indicate that SF3B1's C-terminal domain is composed of 22 degenerated HEAT repeats (Wang et al., 1998) (**Figure 2.9**). In the low-resolution EM reconstruction of the endogenous SF3B complex, the 22 HEAT repeats were tentatively mapped on the outer surface of the assembly, where they form an "S-like" structure (Golas et al., 2003). However, the distinct α-helical features of the HEAT domain could not be discerned in this density map. A typical HEAT repeat (~30-50 residues) consists of two antiparallel amphiphilic α helices (denoted A-helix and B-helix) linked via a short turn (Conti et al., 2006; Stewart, 2006; Yoshimura and Hirano, 2016). The individual HEAT repeats are then connected by variable linker regions to form ring-like, hook-like or solenoid-like structures, as individual coils of a flexible mechanical spring (Conti et al., 2006). This arrangement induces the inherent helical curvature of the domain: the inward, concave surface of the α-α superhelix is lined by the B-helices, whereas the HEAT A-helices face the convex side (Conti et al., 2006; Stewart, 2006). The topological properties of a given HEAT solenoid (curvature, helical pitch, and diameter) depend on the particular state of the protein and cannot be accurately predicted from sequence alone by computational methods (Yoshimura and Hirano, 2016). Importantly, the "molten globule-like" hydrophobic core confers more dynamic properties to HEAT domains compared to globular domains (Grinthal et al., 2010; Kappel et al., 2010; Zachariae and Grubmuller, 2008).

In the low-resolution EM reconstruction of SF3B, the small globular subunit SF3B6/p14 was modeled inside the "bivalve" shell of SF3B, like a "pearl in an oyster" (Golas et al., 2003). SF3B6/p14 (~14 kDa) was previously linked with the recognition of the branch helix and of the invariant BS-A during pre-spliceosome assembly (MacMillan et al., 1994; Query et al., 1994; Query et al., 1996; Will et al., 2001). Its overall structure consists of an unusual RRM motif whose RNA recognition surface is occluded by a C-terminal helical region (Schellenberg et al., 2006;

Spadaccini et al., 2006) (**Figure 2.9**). Photo-crosslinking experiments indicate that SF3B6/p14, which also interacts with SF3B1^{NTD}, contacts directly the bulged BS-A in the pre-spliceosome (A complex) and the later complexes (MacMillan et al., 1994; Query et al., 1996; Will et al., 2001). In these experiments, the pre-mRNA substrates had the BS-A labeled with benzophenone, a photo-reactive agent which crosslinks to proteins within ~15 Å (MacMillan et al., 1994; Will et al., 2001), or were UV irradiated for detection of direct protein-RNA interactions (Query et al., 1996). Importantly, as the precise subunit composition of the human SF3B complex was not clear at that time, the identities of the BS-A crosslinked proteins were assigned mainly based on their relative migration in denaturing gels (Query et al., 1996). In apparent contradiction with the earlier studies, later experiments showed that SF3B6/p14 does not exhibit sequence-specificity for the BS, the U2/BS duplex, or for the BS-A and, overall, interacts weakly with RNA (Perea et al., 2016; Spadaccini et al., 2006). However, SF3B6/p14 could be crosslinked close to the BSiR in the isolated 17S U2 snRNP (Dybkov et al., 2006) and was co-crystallized in complex with an adenine base (Schellenberg et al., 2011).

The essential SF3B4 subunit encompasses two flexibly linked RRM domains, followed by a variable proline-rich region, predicted to be disordered (Champion-Arnaud and Reed, 1994; Igel et al., 1998) (**Figure 2.9)**. In the early EM reconstruction of human SF3B, the two RRM domains of SF3B4 were tentatively modeled on the periphery of the particle (Golas et al., 2003). In both yeast and human, SF3B4 (yeast Hsh49p) forms a stable complex with the SF3B2 subunit (yeast Cus1p) (Champion-Arnaud and Reed, 1994; Pauling et al., 2000). In yeast, this interaction is mediated by the N-terminal RRM domain of SF3B4/Hsh49p and the proline-rich C-terminal domain of SF3B2/Cus1p (Pauling et al., 2000; van Roon et al., 2017), recently captured in a crystal structure (van Roon et al., 2017). Importantly, in the purified 17S U2 snRNP, SF3B4 crosslinks to

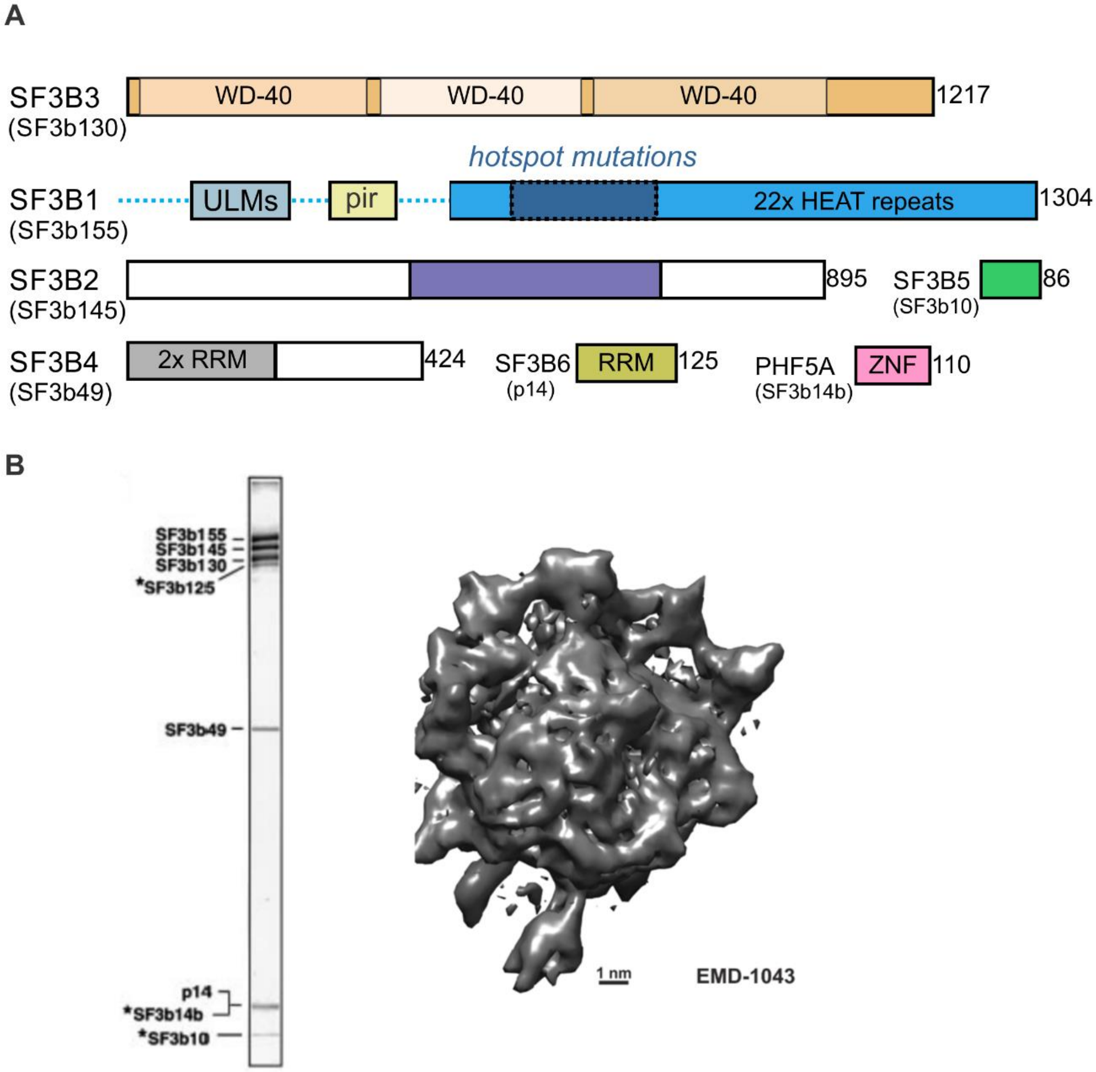


Figure 2.9. Composition and organization of the human SF3B complex.

(**A**) Domain composition of the SF3B subunits. Well-folded regions and known domains are colored and labeled. WD-40 – WD-40-like β-propeller, ULM – U2AF ligand motif; pir – p14-interacting region; RRM – RNA-recognition motif; ZNF – zinc finger. (**B**) Left, protein composition of an endogenous human SF3B complex (Will et al., 2002). Right, the early, ~10 Å EM density map of the human SF3B complex (Golas et al., 2003).

the U2 stem-loop I and the stem-loop IIb (Dybkov et al., 2006). Moreover, SF3B4 crosslinks to intron regions upstream of the BS were detected in both human and yeast spliceosomes (Champion-Arnaud and Reed, 1994; Schneider et al., 2015). Overall, these observations suggest that SF3B4, together with SF3B2, may tether the SF3B complex to the U2 snRNA in the isolated U2 snRNP and in fully assembled spliceosomes.

All the other SF3B subunits, including SF3B3/SF3b130, SF3B5/SF3b10, and PHF5A/SF3b14b, could not be located in the early EM map, owing to the mostly unknown folds and/or flexibility (Golas et al., 2003) (**Figure 2.9**). In the more recent structures of human and yeast PHF5A/Rds3p (Teng et al., 2017; van Roon et al., 2008), the small SF3B subunit adopts a knotted fold, dubbed the “triquetra motif”. The unusual structure forms from three conserved GATA-like zinc clusters positioned like at the vertices of an equilateral triangle (**Figure 2.9**). However, it is unclear how the essential subunit interacts with the other SF3B components (Wang et al., 2005; Wang and Rymond, 2003). Although the large SF3B3 (yeast Rse1p) subunit (**Figure 2.9**) is predicted to encompass three WD40-like β-propellers (Li et al., 2006), none of these distinctive, “donut-shaped”, domains could be identified in the published SF3B density map (Golas et al., 2003). Moreover, nothing is known of the structure of SF3B5/SF3b10 (yeast Ysf3p).

2.7 Splicing factor mutations in cancers

Sequencing of genomes from patients with hematopoietic malignancies (Papaemmanuil et al., 2011; Quesada et al., 2011; Yoshida et al., 2011) and several types of solid tumors (Yoshida and Ogawa, 2014) revealed frequent somatic mutations in four different splicing factor genes: SF3B1, U2AF1, SRSF2/SC35 (serine/arginine-rich splicing factor), and ZRSR2/URP (zinc finger, RNA-binding motif and serine/arginine-rich 2). Interestingly, the missense substitutions of SF3B1, U2AF1, and SRSF2 are always detected in heterozygous form, are mutually exclusive, and cluster at defined sequence hotspots, consistent with a gain-of-function/change-of-function phenotype (Dvinge et al., 2016). Instead, mutations in ZRSR2, a minor spliceosome component, disrupt the coding frame and are distributed across the entire gene (Madan et al., 2015). All these factors act early in the splicing cycle, prior to or during pre-spliceosome formation (see also 2.5). Thus, mutations in these splicing factors, likely, impact the early stages of the spliceosome assembly cycle when the precise intron-exon borders are defined (**Figure 2.7**).

SF3B1 and U2AF1 are constitutive splicing factors involved in the selection/recognition of the BS region and the 3'SS consensus sequence, respectively (see 2.5) (**Figure 2.7**). Recurrent mutations of U2AF1 affect the S34 and Q157 residues, located, correspondingly, in the N-terminal and the C-terminal zinc knuckles (Jenkins and Kielkopf, 2017). As these residues are predicted to interact with the -3 and +1 sites of the 3'SS consensus sequence (Dvinge et al., 2016; Jenkins and Kielkopf, 2017), the hotspot mutations S34F and Q157P/R, likely, alter the sequence specificity of U2AF1. Thus, as observed by RNA sequencing, the affected exons are differentially included into the mature mRNA or skipped in a mutant U2AF1-dependent manner (Ilagan et al., 2015; Jenkins and Kielkopf, 2017; Shirai et al., 2015).

SRSF2/SC35 is an SR (serine/arginine-rich) auxiliary splicing factor that binds to exonic splicing enhancer elements to promote the efficient exon definition (Wu and Maniatis, 1993) (**Figure 2.10**). That is, by interacting with U1 snRNP components and with U2AF1, SRSF2 couples the selection of the 5'SS to the recognition of the 3'SS across the exon (Shen et al., 2004; Wu and Maniatis, 1993). SRSF2's structure consists of an N-terminal RRM domain and a C-terminal RS (arginine-serine-rich) region (Daubner et al., 2012; Wu and Maniatis, 1993). SRSF2's RRM domain typically recognizes exonic enhancers with the CCNG or GGNG consensus sequence, where N represents any nucleotide (Daubner et al., 2012). Mutations of the P95 hotspot residue, frequent in CMML (chronic myelomonocytic leukemia, ~30-50%), induce minor structural changes which increase the affinity and specificity for CCNG over GGNG (Daubner et al., 2012; Kim et al., 2015).

This mechanism may explain, in part, the resulting transcriptome-wide splicing changes observed in animal models and patients (Daubner et al., 2012; Kim et al., 2015).

SF3B1 is one the most frequently mutated splicing factor in hematological malignancies, such MDS (myelodysplastic syndromes, ~50-75%), CLL (chronic lymphocytic leukaemia, ~15%), AML (acute myeloid leukemia, 3-7%), and other myeloid leukemias (Dvinge et al., 2016; Yoshida and Ogawa, 2014). Recurrent SF3B1 mutations are also detected in uveal melanomas (~15-20%), pancreatic and bladder cancers (~4%), and several other solid tumors (Dvinge et al., 2016; Yoshida and Ogawa, 2014). The hotspot mutations of SF3B1 are always heterozygous and are not concomitantly observed with other splicing factor mutations, indicative of synthetic lethality effects (Dvinge et al., 2016). SF3B1's residues mutated in cancer cluster at the C-terminal HEAT domain predicted to form a superhelical structure (Darman et al., 2015; Quesada et al., 2011), and their relative frequencies appear to be linked to a particular cancer type. Thus, the K700E, K666, H662, E622 and G740 substitutions are recurrent in myeloid malignancies, while the G742 missense mutations are detected in CLL (Dvinge et al., 2016; Yoshida and Ogawa, 2014). Interestingly, missense substitutions of R625 are relatively frequent in solid tumors, including uveal melanoma, but are rare in myeloid cancers (Dvinge et al., 2016; Yoshida and Ogawa, 2014).

Whereas mutations in U2AF1 and SRSF2 induce differential exon inclusion/skipping events, the cancer-linked mutations of SF3B1 lead to the selection of alternative 3'SSs and, ultimately, to noncanonical splicing (Alsafadi et al., 2016; Darman et al., 2015). Under a mutant SF3B1 background, the misspliced mRNAs, are, subsequently, either targeted for degradation via nonsense-mediated decay or translated into aberrant proteins (Darman et al., 2015) (**Figure 2.10**). Several recent studies (Alsafadi et al., 2016; Darman et al., 2015; DeBoever et al., 2015) showed that the recurrent mutations in SF3B1 affect a particular subset of pre-mRNA substrates. These

RNA substrates are characterized by the presence of (**Figure 2.10**): (i) a “cryptic” 3’SS (3’SS*) located ~15-24 nts upstream from the canonical 3’SS; (ii) a relatively “weak” and short PPT tract preceding the 3’SS*; (iii) a “cryptic” and “strong” BS (BS*) located ~11-14 nts from the “cryptic” 3’SS, exhibiting higher complementarity to the U2 snRNA than the canonical BS. Importantly, the noncanonical splicing by mutant SF3B1 (**Figure 2.10**) requires the presence of a functional PPT, downstream from the 3’SS*, suggesting that the alternative selection occurs in a “U2AF-dependent manner” (Darman et al., 2015). Although the exact molecular mechanisms are not entirely clear, selection of the “cryptic” sites may occur either during pre-spliceosome formation or the exon ligation step (Alsafadi et al., 2016; Carrocci et al., 2017; Darman et al., 2015).

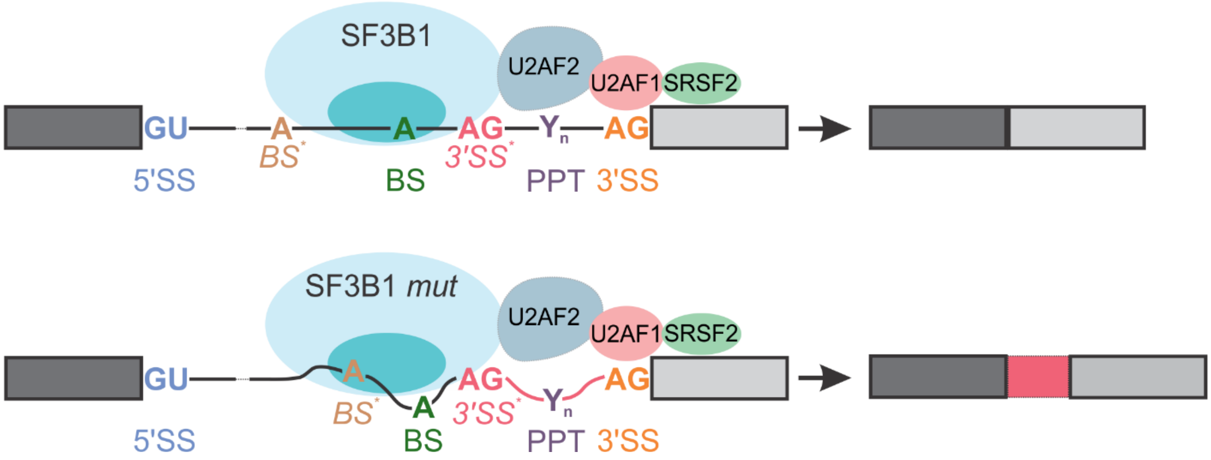


Figure 2.10. Mutations in SF3B1 induce an alternative branch-site usage in cancer cells.

The 3’ end of the intron is defined by the presence of the branch-site (BS, green), the polypyrimidine tract (PPT, dark purple), and the 3’ splice site (3’SS, orange). The concomitant binding of the U2 snRNP, including the SF3B1 subunit (light blue), U2AF2, and U2AF1 to the 3’ end of the intron leads to the selection of the canonical BS and 3’SS (top panel). At the same time, SRSF2 interacts with exonic splicing enhancer elements. Cancer-related mutations of the SF3B1 subunit (SF3B1 *mut*) promote the selection of the alternative, “cryptic” BS* and 3’SS* (bottom panel). Overall, this results in noncanonical splicing of the pre-mRNA substrate. The misspliced product may be targeted for degradation by nonsense-mediated decay or translated into an aberrant protein (Alsafadi et al., 2016; Darman et al., 2015). The PHF5A subunit, contacting the BS adenosine in spliceosomes (Rauhut et al., 2016; Yan et al., 2016), is colored in teal. The two exons are depicted as gray boxes. Adapted with modifications from (Alsafadi et al., 2016; Darman et al., 2015).

2.8 Therapeutic targeting of the spliceosome with small-molecule compounds

Since the initial discovery of FR901464, a *Pseudomonas sp* fermentation product, more than 25 different antitumor small-molecule compounds that modulate spliceosome assembly were described, engineered and synthesized (Bonnal et al., 2012; Effenberger et al., 2017; Salton and Misteli, 2016) (**Figure 2.11**). All these compounds, known as splicing modulators, target the multimeric SF3B complex at the same site (Effenberger et al., 2016; Teng et al., 2017) and compete with the stable recruitment of the U2 snRNP to canonical BSs (Corrionero et al., 2011; Folco et al., 2011). Binding of modulators to SF3B translates *in vitro* in a stalled splicing assembly (Corrionero et al., 2011; Effenberger et al., 2016; Folco et al., 2011; Roybal and Jurica, 2010). However, in cells, splicing modulators induce global changes in alternative splicing patterns (i.e., intron retention, exon skipping), and, finally, the cell-cycle arrest in G1 and G2/M, and apoptosis (Kaida et al., 2007; Kotake et al., 2007; Teng et al., 2017; Vigevani et al., 2017). Importantly, splicing modulators exhibit stronger cytostatic effects on cancer cells compared to normal cells (Lagisetti et al., 2013; Obeng et al., 2016). This effect can be exploited as a novel chemotherapeutic route in SRSF2, U2AF1, and SF3B1-driven hematological malignancies, and, possibly, in other cancers (Lee et al., 2016; Obeng et al., 2016; Seiler et al., 2018; Shirai et al., 2017).

Based on their chemical structure, the SF3B modulators are classified into four distinct families of compounds (Bonnal et al., 2012; Effenberger et al., 2017) (**Figure 2.11**): (i) pladienolides (A-G, E7107); (ii) spliceostatins (A-G, FR901464, thainlanstatins, meayamycin); (iii) herboxidiene (GEX1A); (iv) sudemycins (C1, D6, E, F1). Importantly, it was proposed that all SF3B modulators share a common pharmacophore, defined by a distinctive set of functional groups required for activity (Lagisetti et al., 2013; Lagisetti et al., 2008; Lagisetti et al., 2014). Indeed, although the overall chemistry is rather diverse, all modulators have in common the conjugated diene group.

The diene group serves as the link between two chemically complex moieties, specific for each family of splicing modulators (**Figure 2.11**). In pladienolides, the diene group connects the 12-membered functionalized macrocycle (macrolide group) to an extended aliphatic chain which also includes an epoxide group (Kotake et al., 2007). Herboxidiene (GEX1A) has a similar overall organization, except that the macrolide cycle is substituted by a functionalized pyran core (Hasegawa et al., 2011). Conversely, spliceostatins and sudemycins have significantly different structures (Kaida et al., 2007; Lagisetti et al., 2013). That is, in these compounds, the diene group connects a functionalized pyran ring to a variable 6-membered cycle, further derivatized with a compound-specific side chain (**Figure 2.11**). The pyran ring of spliceostatins also contains a reactive epoxide group, as pladienolide's aliphatic chain, important for the activity of this family of compounds (Kaida et al., 2007). Thus, it is possible that at least certain modulators might interact covalently with SF3B, via the epoxide group (**Figure 2.11**).

Although SF3B is the sole cellular target of splicing modulators, it remains unclear how the structurally different compounds are recognized by the same pocket (Effenberger et al., 2016). Early *in vitro* pulldown assays using affinity and photo-affinity probes showed that pladienolide B binds to the SF3B3 subunit (Kotake et al., 2007), whereas spliceostatin A interacts with both SF3B1 and SF3B3 (Kaida et al., 2007). Consistently, herboxidiene was crosslinked to the SF3B1 subunit (Hasegawa et al., 2011), suggesting that SF3B1 is the common target of modulators. More recent biochemical and chemogenomic analyses suggest that splicing modulators bind in the proximity of the SF3B1 and PHF5A subunits and close to the BS-A binding pocket (Teng et al., 2017). However, it is not entirely clear how the modulator binding site on SF3B relates in structural terms to the BS-A pocket, observed in later spliceosomal complexes. Most importantly, the molecular mechanism of action of these promising chemotherapeutic agents remains elusive.

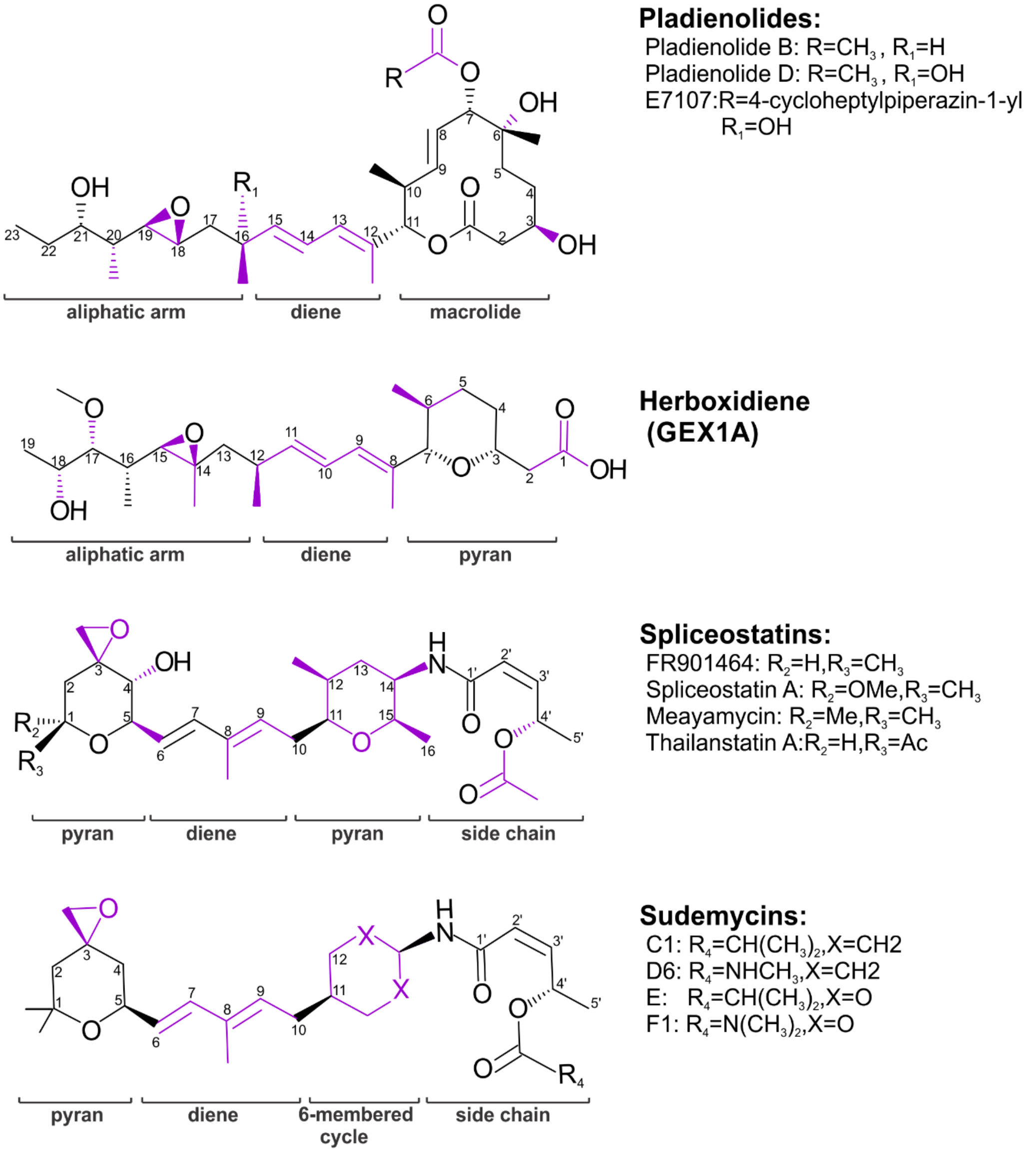


Figure 2.11. Chemistry of splicing modulators targeting SF3B.

The small-molecule compounds targeting SF3B have a modular organization. The common diene group links different chemical moieties, specific for each family of modulators. The chemical groups important for the activity of the compounds are colored in purple. The SAR (structure-activity relationship) data for pladienolides and spliceostatins is based on (Effenberger et al., 2017). The SAR data for herboxidiene is adapted from (Lagisetti et al., 2014), whereas the SAR data for sudemycins is based on (Lagisetti et al., 2013; Makowski et al., 2017).

2.9 About this work

The SF3B complex is the largest building block of the U2 snRNP and an essential component of both major and minor spliceosomes (Golas et al., 2003, 2005; Will et al., 2002). The heptameric human complex was one of the first spliceosomal components to be purified in an intact form and characterized by electron microscopy (Golas et al., 2003). At the beginning of this thesis project, detailed high-resolution structural data were available only for individual domains of SF3B6/p14 (Schellenberg et al., 2011; Schellenberg et al., 2006), PHF5A/SF3b14b (yeast Rds3p) (van Roon et al., 2008), and SF3B4/SF3b49 (yeast Hsh49p) (PDB 5GVQ). Little was known in structural terms of the large HEAT-domain containing protein SF3B1/SF3b155 (yeast Hsh155p) or of SF3B3/SF3b130 (yeast Rse1p), SF3B5/SF3b10 (yeast Ysf3p), and SF3B2/SF3b145 (yeast Cus1p). Most importantly, it was unclear how all seven subunits assemble into a compact globular particle (Golas et al., 2003) and contribute to the specific recognition of the branch-site region of the intron as part of the U2 snRNP (Gozani et al., 1996; Gozani et al., 1998; Query et al., 1996). Thus, in the first part of this thesis work, we use orthogonal structural biology approaches (X-ray crystallography, mass spectrometry) to uncover the molecular architecture of the human SF3B complex.

Abnormal changes in pre-mRNA splicing patterns, detected in different forms of cancer (Dvinge and Bradley, 2015), have been recently linked to mutations in several splicing factors (Dvinge et al., 2016), such as SF3B1, making the spliceosome complex a potential therapeutic target (Bonnal et al., 2012; Salton and Misteli, 2016). SF3B1 mutations are always detected in heterozygous form, are mutually exclusive, and induce the aberrant splicing of a selected subset of transcripts (Alsafadi et al., 2016; Darman et al., 2015). This global “spliceosome sickness” further drives/favors the proliferation of cancer cells through a still unknown mechanism (Darman et al., 2015).

Independently, several antitumor compounds (Bonnal et al., 2012) and promising chemotherapeutic agents (Eskens et al., 2013; Seiler et al., 2018), belonging to the pladienolide, spliceostatin/sudemycin, and herboxidiene families, were shown to target the very same subunit of the SF3B complex (Effenberger et al., 2016; Teng et al., 2017). These splicing modulatory compounds have a complex chemistry and knowledge of the precise features of their binding site on SF3B and of their mode of action shall provide support for ongoing structure-activity, *in vitro*, and *in vivo* studies (Seiler et al., 2018; Teng et al., 2017), and, thus, facilitate the transition from “bench to bedside”. In the second part of this work, we sought to determine crystal structures of optimized SF3B constructs in complex with different splicing modulators. Combined with biochemical and functional studies, this may reveal the structural basis of their modulatory effects on pre-mRNA splicing.

3 Results

The results section of this thesis work is largely based on two first-authored publications:

(i) "Molecular architecture of SF3b and structural consequences of its cancer-related mutations" (**Cretu et al.,** ***Molecular Cell*****, 2016**);

(ii) "Structural basis of splicing modulation by antitumor macrolide compounds" (**Cretu et al.,** ***Molecular Cell*****, 2018**)

The two publications are attached below in their original form (i.e., as available on publisher's website).

3.1 Molecular architecture of SF3b and structural consequences of its cancer-related mutations (*published manuscript*)

Constantin Cretu[1], Jana Schmitzová[1], Almudena Ponce-Salvatierra[1,2], Olexandr Dybkov[3], Evelina I. De Laurentiis[3], Kundan Sharma[4,5], Cindy L. Will[3], Henning Urlaub[4,5], Reinhard Lührmann[3], Vladimir Pena[1,6,*]

[1]Research Group Macromolecular Crystallography, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

[2]Max Planck Research Group Nucleic Acid Chemistry, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

[3]Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

[4]Bioanalytical Mass Spectrometry, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

[5]Bioanalytics Group, Institute for Clinical Chemistry, University Medical Center Göttingen, 37075 Göttingen, Germany

[6]Lead Contact

[*]Correspondence: vladimir.pena@mpibpc.mpg.de

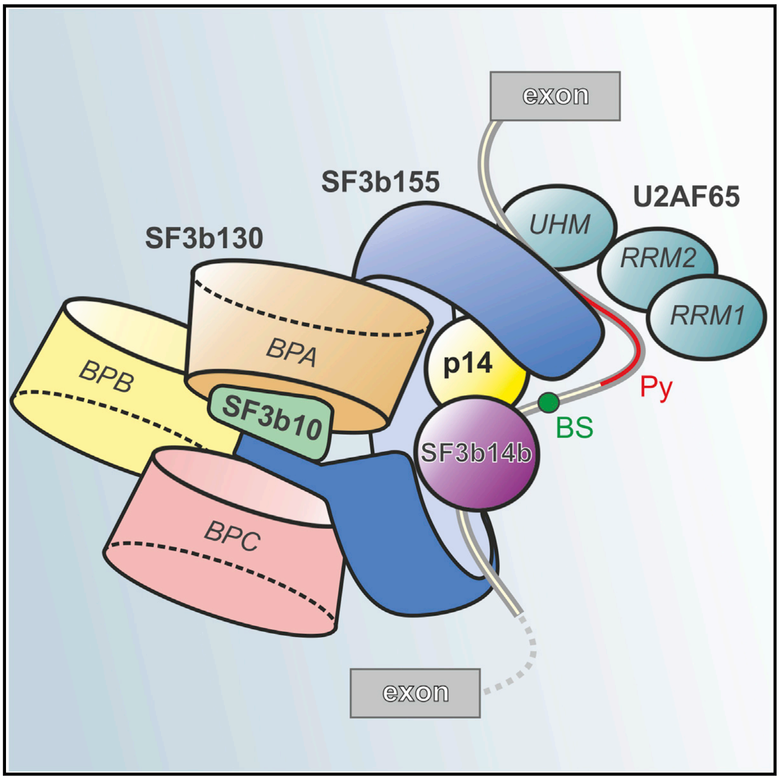
Molecular Cell*, 2016, Volume 64, pg 307-319,* https://doi.org/10.1016/j.molcel.2016.08.036

Article

Molecular Cell

Molecular Architecture of SF3b and Structural Consequences of Its Cancer-Related Mutations

Graphical Abstract



Authors

Constantin Cretu, Jana Schmitzová, Almudena Ponce-Salvatierra, ..., Henning Urlaub, Reinhard Lührmann, Vladimir Pena

Correspondence

vladimir.pena@mpibpc.mpg.de

In Brief

Cretu et al. report the crystal structure of the human SF3b complex, which is required for branch site selection during pre-mRNA splicing. The structure provides insight into how alterations in the SF3b complex caused by cancer-related mutations can affect branch site selection.

Highlights

- SF3b core contributes to the positioning of the BS-binding proteins p14 and U2AF65
- SF3b155 conformation is induced by multiple contacts in the SF3b complex
- SF3b155 surrounds p14, generating a composite platform for binding RNA
- Residues mutated in cancer cells affect the structure of the RNA-binding platform

Accession Numbers

5IFE
5HY7

Cretu et al., 2016, Molecular Cell *64*, 307–319
October 20, 2016 © 2016 Elsevier Inc.
http://dx.doi.org/10.1016/j.molcel.2016.08.036

Molecular Cell
Article

Molecular Architecture of SF3b and Structural Consequences of Its Cancer-Related Mutations

Constantin Cretu,[1] Jana Schmitzová,[1] Almudena Ponce-Salvatierra,[1,2] Olexandr Dybkov,[3] Evelina I. De Laurentiis,[3] Kundan Sharma,[4,5] Cindy L. Will,[3] Henning Urlaub,[4,5] Reinhard Lührmann,[3] and Vladimir Pena[1,6,*]
[1]Research Group Macromolecular Crystallography
[2]Max Planck Research Group Nucleic Acid Chemistry
[3]Department of Cellular Biochemistry
[4]Bioanalytical Mass Spectrometry
Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany
[5]Bioanalytics Group, Institute for Clinical Chemistry, University Medical Center Göttingen, 37075 Göttingen, Germany
[6]Lead Contact
*Correspondence: vladimir.pena@mpibpc.mpg.de
http://dx.doi.org/10.1016/j.molcel.2016.08.036

SUMMARY

SF3b is a heptameric protein complex of the U2 small nuclear ribonucleoprotein (snRNP) that is essential for pre-mRNA splicing. Mutations in the largest SF3b subunit, SF3B1/SF3b155, are linked to cancer and lead to alternative branch site (BS) selection. Here we report the crystal structure of a human SF3b core complex, revealing how the distinctive conformation of SF3b155's HEAT domain is maintained by multiple contacts with SF3b130, SF3b10, and SF3b14b. Protein-protein crosslinking enabled the localization of the BS-binding proteins p14 and U2AF65 within SF3b155's HEAT-repeat superhelix, which together with SF3b14b forms a composite RNA-binding platform. SF3b155 residues, the mutation of which leads to cancer, contribute to the tertiary structure of the HEAT superhelix and its surface properties in the proximity of p14 and U2AF65. The molecular architecture of SF3b reveals the spatial organization of cancer-related SF3b155 mutations and advances our understanding of their effects on SF3b structure and function.

INTRODUCTION

During splicing, non-coding introns are excised from pre-mRNA by the spliceosome, a multi-megadalton protein-RNA enzyme that orchestrates the accurate recognition and excision of introns, and concomitant ligation of exons to form mature mRNA (Wahl et al., 2009). The spliceosome assembles de novo and in a stepwise manner onto each pre-mRNA intron from five small nuclear ribonucleoproteins (snRNPs) (U1, U2, U4/U6, U5) and additional non-snRNP factors that often are organized as heteromeric complexes (Wahl et al., 2009). Splicing catalysis involves two transesterification steps. In the first step, the branchpoint adenosine from the branch site (BS) carries out a nucleophilic attack at the 5′ splice site (SS) of the intron, leading to 5′ SS cleavage and intron lariat formation. In the second step, the 3′ overhang (OH) of the cleaved 5′ exon acts as the nucleophile, and the 3′ SS is cleaved with concurrent ligation of the two exons (Wahl et al., 2009).

At the earliest stage of spliceosome assembly, the BS, 3′ SS, and the intervening polypyrimidine (Py) tract are recognized cooperatively by several splicing factors. Thus, splicing factor 1 (SF1) binds the BS, the U2 auxiliary factor U2AF65 binds the Py tract and makes additional contacts to the BS, while U2AF35 binds the 3′ SS (Arning et al., 1996; Valcárcel et al., 1996; Zamore et al., 1992). In the next step, the U2 snRNP replaces SF1 in an ATP-dependent manner, and the U2 snRNA forms a duplex with the BS (MacMillan et al., 1994). The branchpoint adenosine in the U2/BS helix is bulged out, enabling it to act as the nucleophile in the first transesterification step of splicing (Berglund et al., 2001; MacMillan et al., 1994). Interaction of the U2 snRNA with the BS is rather weak and requires additional factors for stabilization.

SF3b is a multi-protein component of the U2 snRNP that is essential for pre-mRNA splicing. Human SF3b can be isolated as a stable complex that contains seven subunits: SF3b155, SF3b130, SF3b145, SF3b49, SF3b14b, p14/SF3b14a, and SF3b10 (Will et al., 2002). In the spliceosome, SF3b proteins contact the pre-mRNA at or near the BS, reinforcing the U2 snRNA/BS base-pairing interaction (Gozani et al., 1996, 1998; Query et al., 1997; Will et al., 2001), and thereby play a key role in BS recognition and selection during constitutive and alternative splicing (Alsafadi et al., 2016; Corrionero et al., 2011; Darman et al., 2015).

SF3b155 is the largest subunit and encompasses a long, intrinsically unstructured N-terminal domain (NTD) and a C-terminal HEAT-repeat domain (HD) (Cass and Berglund, 2006; Wang et al., 1998). NTD plays important roles at different stages of the splicing cycle, and several discrete regions of functional importance were narrowed down. SF3b155^{NTD} contains linear sequence motifs named ULM (UHM-ligand motif) that interact with the UHM (U2AF-homology motif) domains of U2AF65 (Cass and Berglund, 2006; Spadaccini et al., 2006; Thickman et al., 2006) and of the alternative splicing regulator

SPF45/RBM17 (Corsini et al., 2007). An adjacent short region of SF3b155^{NTD} (373–415) interacts with p14 (Schellenberg et al., 2006; Spadaccini et al., 2006), an SF3b subunit that directly contacts the BS adenosine of the intron in the prespliceosome and in subsequently formed spliceosomal complexes (MacMillan et al., 1994; Query et al., 1996). SF3b155 can be crosslinked to nucleotides directly upstream and downstream of the BS, suggesting that SF3b155, together with its BS-binding partners p14 and U2AF65, acts cooperatively in the recognition of the BS (Gozani et al., 1998; Will et al., 2001).

Erroneous recognition of the BS by the U2 snRNA can lead to the accumulation of aberrant transcripts in the cell and to cancer (Alsafadi et al., 2016; Darman et al., 2015; DeBoever et al., 2015). SF3b155 is the most frequently mutated spliceosomal protein in a variety of cancer cells and thus a major drug target for therapy (Yoshida and Ogawa, 2014). Point mutations in the HD of SF3b155 detected recurrently in patients lead, via an unknown mechanism, to the selection of alternative BSs and, as a consequence, to the use of cryptic pre-mRNA 3′ SSs and the generation of aberrantly spliced mRNA (Alsafadi et al., 2016; Darman et al., 2015; DeBoever et al., 2015). The mechanism of this alternative selection is not understood in structural terms (Darman et al., 2015). Moreover, alterations in the contacts between SF3b155 and the pre-mRNA by tumor inhibitors such as spliceostatin A result in reduced fidelity of BS and 3′ SS recognition by U2 snRNP and thus in changes in alternative splicing (Corrionero et al., 2011). How SF3b is organized as a composite module and the structural basis whereby it recognizes the BS and proofreads the U2 snRNA/BS interface is presently unclear.

The structure of a late-stage spliceosome from *S. pombe* was recently reported at near atomic resolution (Yan et al., 2015). This spliceosome lacks SF3b proteins, and aside from a low-resolution electron microscopic reconstruction of human SF3b (Golas et al., 2003), little is known about SF3b's spatial organization. Here, we crystallized and determined the structure of the human SF3b155-SF3b130-SF3b14b-SF3b10 complex (i.e., SF3b core) and *Chaetomium thermophilum* (*ct*) *ct*SF3b130-SF3b10 dimer. The SF3b structure reveals how the distinctive conformation of SF3b155 is induced by the multiple interfaces that it shares with other SF3b core components, which act as a scaffold. Protein-protein crosslinking data indicate that the SF3b core adopts the same fold in solution and enabled us to position p14/SF3b14a and U2AF65, which in the spliceosome contact the BS adenosine, within the SF3b155 HEAT superhelix. The structure of the SF3b core reveals the location of numerous cancer-related SF3b155 mutations and indicates how they affect the structure of SF3b and thereby its function in splicing.

RESULTS AND DISCUSSION

Structure Determination and Overall Fold of the Human SF3b Core

We reconstituted human SF3b by co-expression of its seven proteins (SF3b155, SF3b130, SF3b145, SF3b49, SF3b14b, p14/SF3b14a, and SF3b10) in insect cells (Figure 1A). The purified complex restored splicing activity to HeLa nuclear extract depleted of endogenous SF3b (Figure 1B). On the basis of bioinformatics predictions and limited proteolysis, we removed intrinsically disordered regions from SF3b (Figure S1A, available online). The resulting SF3b core complex formed crystals and enabled de novo structure determination by combining single isomorphous replacement with anomalous scattering (SIRAS) and molecular replacement-single-wavelength anomalous dispersion (MR-SAD) (Figures S1B–S1E). The final model was refined against a native data set to 3.1 Å resolution and exhibits excellent stereochemistry (Table 1). One molecule is present in the asymmetric unit and comprises SF3b155 residues 463–1,304 (the HEAT domain), SF3b130 residues 1–1,217, SF3b14b residues 7–95, and SF3b10 residues 15–80 (Figures 1C–1F). The structure reveals a compact and bipartite architecture, dominated by the intimate association of SF3b130 and the HEAT domain of SF3b155 (Figures 1C–1E). The overall structure is further consolidated by the presence of the smaller subunits SF3b10 and SF3b14b (Figures 1C and 1D).

SF3b130 Exhibits a β-Propeller Structure Similar to DDB1 and Accommodates SF3b10 and the C Terminus of SF3b155 in Its V Cleft

Human SF3b130 in the SF3b core contains three intertwined seven-bladed β-propeller domains (BPA, BPB, and BPC from the N to C terminus), followed by an α-helical C-terminal domain (CTD; 1,133–1,217) that contacts the other three domains (Figures 2A and S2A). The overall fold of the protein closely resembles that of DDB1, a multifunctional protein scaffold and core component of the cullin4A-RING E3 ubiquitin ligase complexes (CRL4A) (Angers et al., 2006; Li et al., 2006). The similarity extends also to the manner how SF3b130 and DDB1 proteins bind other interacting partners, revealing a general interaction principle for members of this protein family. That is, SF3b130 binds SF3b10 by way of a clam-shaped cavity (i.e., a V cleft), framed by the tightly connected BPA (11–404) and BPC (1–10, 405–441, and 773–1,132) domains (Figure 2A). SF3b10 interacts mainly with BPA and is folded as a conserved hairpin-like element comprised of three α helices (Figure 2B). This is highly analogous to the way in which DDB1 interacts with its different binding partners, including DDB2 (Scrima et al., 2008). However, in contrast to DDB1, which interacts with its ligands by way of BPC, SF3b130 interacts with SF3b10 via BPA. Noteworthy, the interaction between SF3b10 and BPA is stabilized by a hairpin element that protrudes from the fourth blade of the BPA β-propeller (Figures 2A and 2B). This insert is present in SF3b130 and absent from DDB1. In the human SF3b core, the V cleft also accommodates the C-terminal region of SF3b155 (denoted the anchor) that together with SF3b10 occupies most of the cavity (Figures 2B and 2C). Intriguingly, SF3b130 alone is a stable subunit of the chromatin remodeling TFTC (Brand et al., 2001) and STAGA complexes (Martinez et al., 2001), suggesting that the protein, similarly to DDB1, represents a versatile scaffold acting in various cellular contexts. The notable functional and structural similarities between SF3b130 and DDB1 suggest evolutionary divergence from a common ancestor.

We also determined the structure of SF3b130 in complex with solely SF3b10. For this purpose, we crystallized the *ct*SF3b130-SF3b10 dimer and solved its structure by SAD. The final model contains two molecules in the asymmetric unit and was refined against a native data set to 2.9 Å resolution (Table S1), and

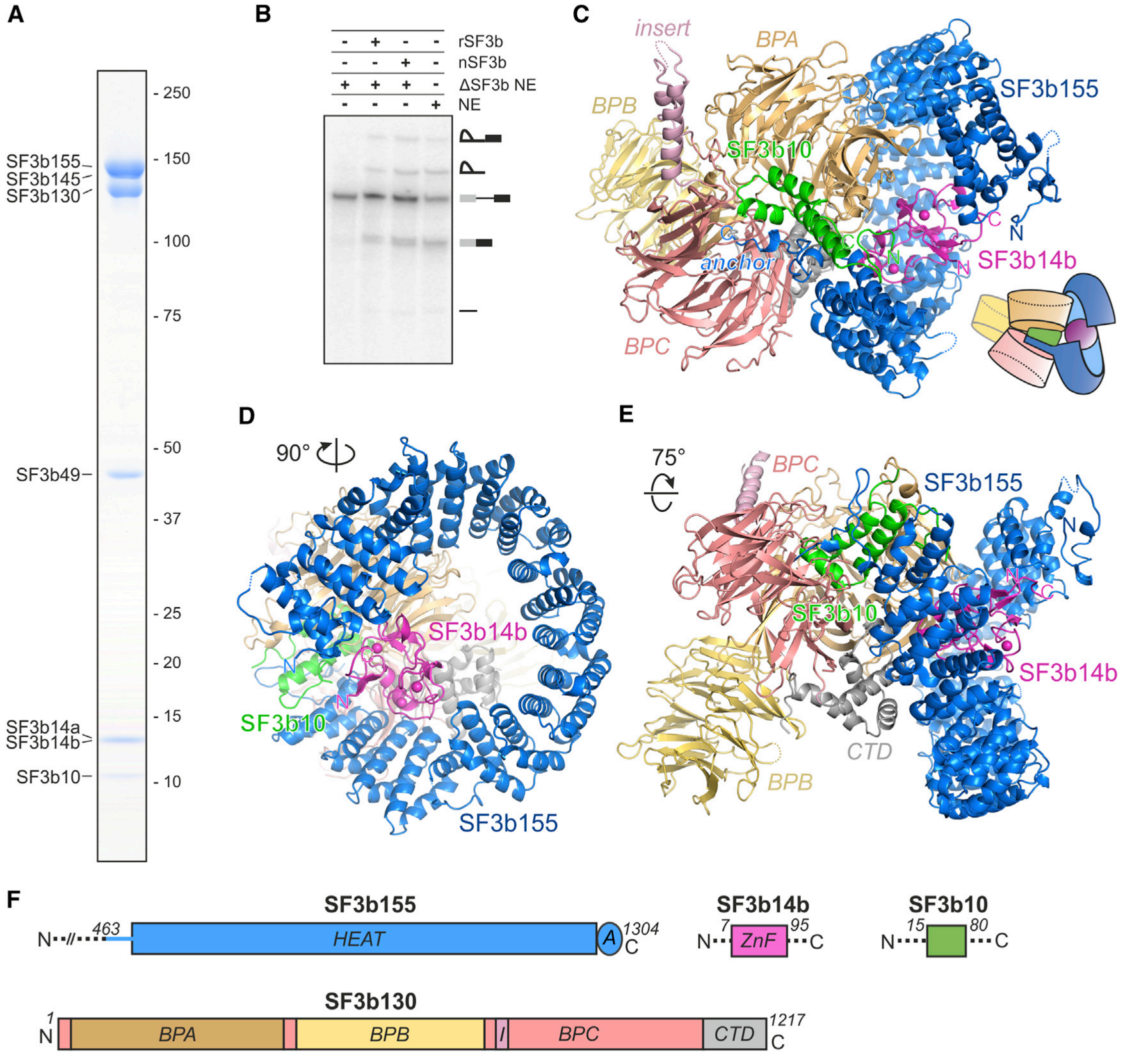


Figure 1. Structure of the Human SF3b Core Complex

(A) SDS-PAGE of the recombinant full-length human SF3b complex after affinity purification and size exclusion chromatography. Note that SF3b155 and SF3b145, as well as p14/SF3b14a and SF3b14b, co-migrate on the gel.

(B) Recombinant SF3b complex (rSF3b), like native SF3b (nSF3b), supports splicing in nuclear extract (NE) depleted (Δ) of SF3b.

(C–E) Three views of a ribbon representation of the SF3b core (middle views rotated relative to top right view). Proteins are shown in different colors. The small inset depicts the overall architecture.

(F) Schematic of the domain arrangement of the crystallized SF3b core proteins. The α-helical insert from the BPC domain and the zinc-finger domain of SF3b14b are labeled *I* and *ZnF*, respectively; *A*, anchor.

See also Figure S1.

comprises *ct*SF3b130 residues 7–1,213 and *ct*SF3b10 residues 27–81 (Figures S1F and S2A). The former exhibits the typical fold of its homolog DDB1, and also accommodates *ct*SF3b10 in its inner cavity (Figure S2B). Comparison of *ct*SF3b130-SF3b10 with its human orthologs within the SF3b core shows that the structures and relative orientations of BPA, BPC, CTD, and of SF3b10 are virtually identical (Figure S2A; Data S1). Conversely, the BPB domain (442–772) is oriented differently in SF3b130 orthologs, suggesting that, as in DDB1 structures (Li et al., 2006; Scrima et al., 2008), it is flexibly connected to BPC (Figure S2A). In CRL4A complexes the BPA-BPC module of DDB1 serves as an adaptor for various substrate receptors while the BPB domain tethers the E3 ligase subunits (Angers et al., 2006). By analogy, the mobile BPB domain of SF3b130 could sample other SF3b interactors and thus recruit additional splicing factors to U2 snRNP at later stages of the splicing cycle.

SF3b155 Conformation Is Achieved by Multiple Contacts with SF3b130, SF3b14b, and SF3b10

The HEAT domain of SF3b155 is composed of 20 tandem repeats (Figure 2D), in contrast to 22 repeats as previously proposed based solely on its sequence (Wang et al., 1998). It is organized as a right-handed superhelix with a pitch of

Table 1. X-Ray Data Collection and Refinement Statistics

SF3b Core	Ta6Br12 Derivatized—Ta Peak	Native I—SIRAS	Native II—Zn K-Edge	Native III
Data Collection				
Space group	P 21 21 21	P 21 21 21	P 21 21 21	P 21 21 21
Cell dimensions: a, b, c (Å)	105.68, 153.39, 210.32	105.25, 154.01, 209.81	104.61, 154.37, 210.56	105.05, 154.44, 210.42
Cell dimensions: α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Wavelength	1.254	1.0	1.282	0.918
Resolution (Å)	49.74–4.27 (4.56–4.27)	48.45–3.39 (3.51–3.39)	49.54–3.09 (3.17–3.09)	47.04–3.1 (3.18–3.1)
R_{merge} (%)	20.1 (>100)	15.9 (>100)	14.3 (>100)	21.3 (>100)
$I/\sigma I$	9.9 (2.4)	9.9 (1.5)	11.2 (1.4)	10.5 (1.2)
$CC_{1/2}$	0.998 (0.736)	0.997 (0.727)	0.998 (0.465)	0.995 (0.412)
Completeness (%)	99.3 (96.7)	99.5 (95.0)	98.4 (87.2)	99.9 (98.9)
Redundancy	12.4 (12.2)	6.6 (6.2)	11.1 (10.1)	8.3 (8.4)
Refinement				
Resolution (Å)	–	–	–	47.04–3.1
No. reflections	–	–	–	62,754
R_{work}/R_{free}	–	–	–	0.230/0.259
No. atoms: protein	–	–	–	17,099
No. atoms: ligand/ion	–	–	–	4
B-factors: protein	–	–	–	77.64
B-factors: ligand/ion	–	–	–	72.73
RMS deviations: bond lengths (Å)	–	–	–	0.002
RMS deviations: bond angles (°)	–	–	–	0.55
Ramachandran plot: favored (%)	–	–	–	95.43
Ramachandran plot: outliers (%)	–	–	–	0.05

Values for the highest-resolution shell are given in parentheses. RMS, root-mean-square. See also Figure S1.

65 Å and a mean diameter of 100 Å (Figures 1C and 1D), which represents a unique conformation for a HEAT-repeat superhelix. Generally, HDs possess structural plasticity, with their conformation influenced by their binding partners (Grinthal et al., 2010; Zachariae and Grubmüller, 2008). Indeed, the distinctive conformation of SF3b155's HD appears to be induced by multiple interfaces that it shares with SF3b130, SF3b10, and SF3b14b (Figure 2E), consistent with its lack of solubility when expressed on its own (data not shown). HEAT repeats H3 and H5–H6 contact SF3b130 BPA, H18–H20 contact the CTD of SF3b130, and H20 binds both the BPA and BPC β-propellers (Figure 2D). The C-terminal anchor region of the HD lacks α-helical structure and is deeply inserted into the V cleft, where it makes extensive contacts with SF3b130 BPC and with SF3b10 (Figures 2B and 2C). The latter acts as an adaptor that reinforces the tight packing of the anchor within the V cleft. The structure is consistent with a yeast two-hybrid interaction between yeast SF3b10 and SF3b155 (Wang et al., 2005) and explains SF3b10's essential role in the assembly and stability of yeast SF3b and in prespliceosome formation (Wang et al., 2005). Moreover, SF3b10 residues 33–85, essential for yeast viability (Wang et al., 2005), correspond to those residues of the human ortholog that are located inside the V cleft and are closely associated with SF3b155 (Data S1).

SF3b155 interacts extensively also with SF3b14b, a highly conserved protein required for proper assembly of yeast U2 snRNP and stable spliceosomes (Wang and Rymond, 2003). Human SF3b14b is folded as a triple knot, similar to the previously reported nuclear magnetic resonance model of the yeast homolog (van Roon et al., 2008) (Figure S2C). In a manner reminiscent of Ran bound to HEAT-repeat nuclear transport receptors, SF3b14b contacts the concave surface of SF3b155 in two highly conserved places, bridging diametrically opposed residues from H2–H3 (contact site I) at the N terminus and H15, H17, and H18 (contact site II) at the C terminus of the superhelix (Figures 2D, S2D, and S2E). By the way of these contacts, SF3b14b appears to contribute inducing an ~90° tilt between H1–H3 and H4 at contact site I as well as between H15 and H17–H20 at contact site II (Figure S2F). These observations suggest that SF3b14b contacts the superhelix at two dynamic hinge regions (H3–H4 and H15–H16) and thus fixes the HD of SF3b155 in a defined conformation. As in other HEAT-repeat proteins (Zachariae and Grubmüller, 2008), the presence of dynamic hot spots in the HD superhelix would permit additional conformational rearrangements of SF3b upon transiting different stages of the splicing cycle (Golas et al., 2005). Taken together, the structure of SF3b indicates that the distinctive conformation of SF3b155 is induced and maintained by a multipartite scaffold composed of SF3b130, SF3b10, and SF3b14b.

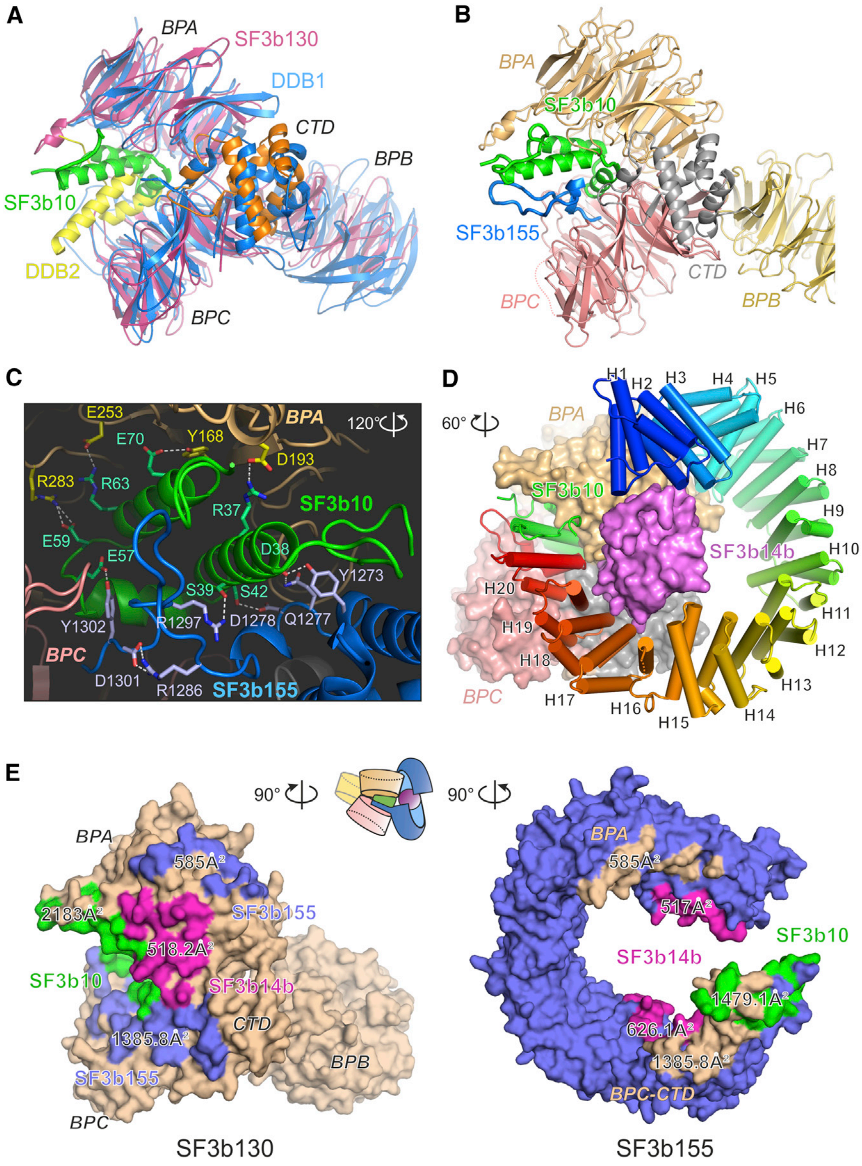


Figure 2. Intermolecular Interactions in the SF3b Core

(A) Superposition of DDB1-DDB2 (PDB: 3EI4) complex on SF3b130-SF3b10. For simplicity, only the DDB2 region that interacts with DDB1 is shown. Note that while SF3b10 contacts the BPA domain of SF3b130, DDB2 interacts with the BPC domain of DDB1. Interaction between SF3b10 and BPA is further stabilized by an insertion that protrudes from the fourth blade of the BPA β-propeller.

(B) Overview of the V cleft organization and the manner in which it mediates binding of SF3b10 and the SF3b155 anchor.

(C) Polar contacts between SF3b10, the SF3b155 anchor, and the interior of the V cleft.

(D) Overview of the interaction between the HEAT repeats (H1–H20) and other SF3b core components.

(E) Molecular interfaces between SF3b130, SF3b155, and the other core components. Interfaces are computed with PISA and colored like their interaction partners. The corresponding buried surface areas (Å^2) are indicated on the surface.

See also Figure S2 and Data S1.

The HEAT-Repeat Superhelix Defines a Composite RNA-Binding Platform for BS Recognition

To validate the architecture of the SF3b core, we treated full-length SF3b complexes with BS3 (bissulfosuccinimidyl suberate) and identified crosslinked pairs of lysines by mass spectrometry (MS) (Table S2). A total of 365 intermolecular unique crosslinks were identified for the entire complex. More than 95% of crosslink-assigned spectra were consistent with the crystal structure, indicating that the SF3b core adopts the same fold in solution (Figure 3A). SF3b49 crosslinks were limited to its binding partner SF3b145 (Champion-Arnaud and Reed, 1994), which was mainly crosslinked to SF3b155 (Figure 3B), suggesting that SF3b145 mediates the interaction of SF3b49 with the SF3b complex. Because of its largely unknown structure, we could not model SF3b145 into the core complex.

SF3b14a/p14 is regarded as a marker of the branchpoint adenosine, to which it crosslinks within prespliceosomes and subsequent spliceosomal complexes (Query et al., 1997; Will et al., 2001). BS3 crosslinking data for p14 (Figure S3A; Table S2) allowed us to model its average position in the SF3b core. The p14 structure was previously determined in complex with a short SF3b155 fragment (residues 373–415) from its unstructured N-terminal region, which assumes an ensemble of conformations in solution (Figure S4). SF3b155$^{373–415}$ is necessary and sufficient for stable binding of p14 (Schellenberg et al., 2006; Spadaccini et al., 2006). Surprisingly, in addition to several expected crosslinks with SF3b155$^{373–415}$ (Figure S3A), 56 unique crosslinks were detected between p14 and the HD, indicating that p14 is located within the circular perimeter of the latter, close to SF3b14b (which is also crosslinked to p14 by 13 unique crosslinks; Figures 3B and 3C). A similar p14 crosslinking pattern could be detected in the 17S U2 snRNP particles, suggesting that the p14-HD interface does not undergo major rearrangements upon integration of SF3b into U2 snRNP (data not shown).

As an independent means of demonstrating that p14 does indeed share interfaces with SF3b155^{HD}, we performed crosslinking of SF3b with the carbodiimide EDC, a compound that enables direct formation of covalent bonds between carboxy and amino groups of exposed residues (Table S2). Importantly, the distance restraints derived from the EDC crosslinking experiment, as in the case of BS3, enabled us to again localize p14

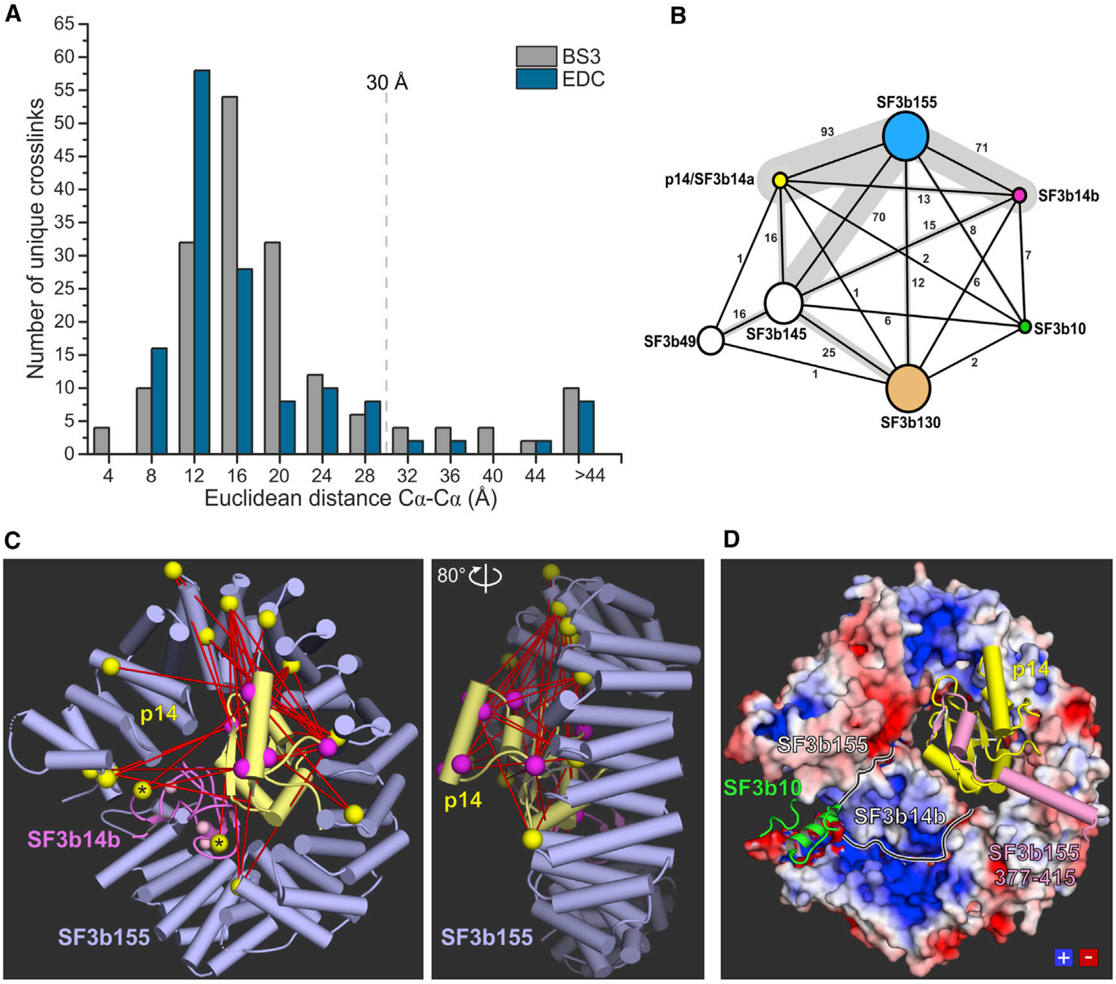


Figure 3. Position of p14 in the SF3b Core

(A) Distribution of the Cα-Cα distances between crosslinked residues in the SF3b core complex. Euclidean distances were computed between the Cα atoms of BS3 or EDC crosslinked residues using the atomic coordinates of the SF3b core complex. Chemical crosslinking data at FDR = 1% were not further filtered.

(B) Protein-protein interaction network of the seven-subunit SF3b complex as revealed by BS3 crosslinking experiments. Subunits of the human SF3b complex are shown as nodes of the graph (SF3b155, light blue; SF3b130, pale orange; SF3b14b, magenta; SF3b10, green; p14/SF3b14a, yellow). The total number of unique intermolecular crosslinks (FDR = 1%, unfiltered) between the subunits of the complex is indicated on the connecting lines.

(C) Two orientations of the model derived from the BS3 crosslinking data, showing the average location of p14 relative to SF3b155 and SF3b14b. Crosslinked residues from p14, SF3b155, and SF3b14b are depicted as magenta, yellow, and yellow starred spheres, respectively.

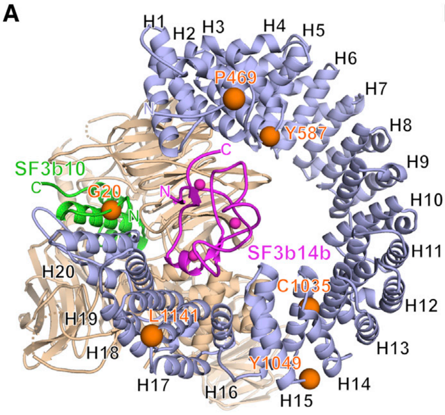
(D) Surface-charge distribution of SF3b155 and SF3b14b around p14. The crystal structure of p14 (yellow) complexed with a short SF3b155 NTD fragment (PDB: 2F9J) is shown.

See also Figures S3 and S4 and Table S2.

in the perimeter of the HEAT superhelix (Figure S3B). The presence of direct contacts between p14 and SF3b155^{HD}-SF3b14b is further supported by in vitro binding assays showing that p14 interacts with a SF3b core complex devoid of SF3b155$^{373–415}$ (Figures S3C and S3D). These data indicate that although SF3b155$^{373–415}$ suffices for tethering of p14 to SF3b, additional contacts with the HD are required to position p14 in a structurally defined location.

The location of the BS-binding, RNA recognition motif (RRM) domain of p14 in the center of the SF3b155^{HD} (Figure 3C) suggests that the latter might form a composite RNA-binding platform together with SF3b14b. To identify RNA-binding regions within the SF3b core complex we exposed to UV irradiation in vitro-assembled RNP complexes and detected crosslinked peptide-RNA conjugates by MS (Figures 4A–4C). Notably, four of the photo-crosslinked residues were located in the H3, H14, and H17 repeats of SF3b155, two in the amino-terminal region of SF3b10 (Y5, G20), and one residue (Y100) in the flexible C terminus of SF3b14b (Figures 4A and S5). Consistent with previous data (Gozani et al., 1998), we could not detect crosslinked peptides derived from SF3b130, suggesting that, most likely, the protein scaffold does not contribute to RNA binding (Figure 4A). Although these experiments are not performed in the context of the spliceosome, they show that SF3b155^{HD}-SF3b14b

A



B

Protein	Peptide Sequence
SF3b155	M^{468}KPDDIQYFDK^{477}
	^{574}ILVVIEPLLIDEDYYAR^{590}
	^{1031}VQENCIDLVGR^{1041}
	^{1046}GAEYVSAR^{1053}
	^{1138}VPELNVQNGVLK^{1149}
SF3b14b	^{96}TDLFYER^{102}
SF3b10	^{5}YTIHSQLEHLQSK^{17}
	^{18}YIGTGHADTTK^{28}

C

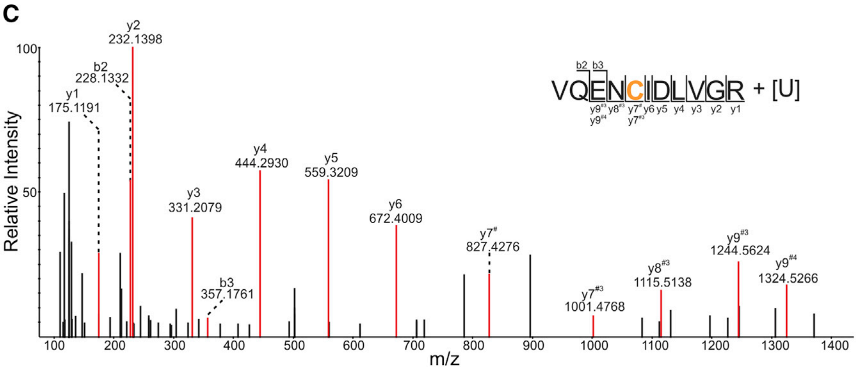


Figure 4. Protein-RNA Photo-Crosslinks Identified in the SF3b Core Complex

(A) Overview of residues in the SF3b core complex crosslinked to RNA upon UV irradiation. RNA-crosslinked residues are mapped onto the structure of the SF3b core and are shown as colored spheres.

(B) UV-induced peptide-RNA conjugates detected in the SF3b core complex. The sequence of the crosslinked peptides as identified by MS is shown, and the crosslinked amino acids are highlighted in orange.

(C) Representative MS/MS fragmentation spectrum of a photo-crosslinked peptide. The crosslinked peptide sequence and its corresponding y- and b-type fragment ions are indicated at the top, referring to the ions that retain the charge on the N or C terminus, respectively. #Observed as an adduct with [fragment of U = -C3O = 52]; #3observed as an adduct with [U-H_3PO_4]; #4observed as an adduct with [U-H_2O].

See also Figure S5.

possesses the biophysical properties required for RNA accommodation. In agreement with previously observed SF3b155 crosslinks to intron nucleotides flanking the BS adenosine (Gozani et al., 1998), p14 is flanked by positively charged patches of the HD and SF3b14b (Figure 3D).

U2AF65 Is Located on the SF3b Core and in the Proximity of p14

Another important interaction partner of SF3b155 that participates in the BS recognition is U2AF65, an essential and conserved splicing factor (Zamore et al., 1992; Valcárcel et al., 1996). U2AF65 has a modular architecture: an arginine/serine (RS)-rich domain that contacts the BS (Valcárcel et al., 1996), two tandem RRM domains that bind cooperatively the Py tract sequence (Sickmier et al., 2006), and a UHM domain that recognizes SF3b155^{NTD} (Gozani et al., 1998; Wang et al., 2013). To examine how U2AF65 is positioned relative to the SF3b core and p14, we reconstituted the SF3b-U2AF65 complex and subjected it to crosslinking and mass-spectrometric analysis (Figure 5). The overall crosslinking pattern of the SF3b core remained the same in the presence of U2AF65, indicating a similar structural organization.

The N-terminal region of U2AF65 that contains the functional RS dipeptides (20–63) was crosslinked to residues located in the immediate vicinity of p14 and belonging to SF3b155^{HD} (H14–H15), SF3b130 CTD and BPA, and SF3b14b (Table S2; Figures 5C and 5D). The preferential crosslinking of the RS region is consistent with its described ability to contact the BS and promote the BS-U2 snRNA interaction (Valcárcel et al., 1996). The RRM domains of U2AF65 were crosslinked to the p14-binding side of the HEAT superhelix (Figure 5C). A reliable docking was difficult given that relatively few pairs of crosslinks were detected (Figures S6A and S6B). Most crosslinks detected between UHM and SF3b155 were established with the ULM-rich region of NTD (188–357), confirming the previously demonstrated interactions between these regions (Cass and Berglund, 2006; Gozani et al., 1998) (Figures 5B and 5C). Remarkably, eight and seven high-confidence crosslinks were detected between the UHM domain SF3b155^{HD} and p14, respectively, indicative of its proximity to HEAT-repeats H5–H8, the area that contains the most hot spot mutations (Figure 5D).

The crosslinking pattern of the SF3b-U2AF65 complex is consistent with previously reported interactions and indicates that, via its contacts with the SF3b core, U2AF65 adopts a structurally defined location. The revealed role of the HEAT superhelix as a scaffold element that bridges p14 and U2AF65 underlines its importance in the stable recruitment of U2 snRNP and accurate selection of the BS.

Residues Mutated in Cancer Cells Are Important for the Tertiary Structure and Surface Properties of SF3b155

SF3b155 is the most frequently mutated spliceosomal protein in several types of cancer (Yoshida and Ogawa, 2014). Recurrent point mutations have the highest incidence in myelodysplastic syndrome (MDS) and chronic lymphocytic leukemia (CLL) and lower frequency in various types of solid cancers (Yoshida and Ogawa, 2014). Although more than 40 residues reported to be mutated in human cancer cells are distributed along the entire superhelix (Figures 6A–6C), 33 are clustered in HEAT repeats H4–H7, in the immediate vicinity of the UHM domain of U2AF65, and of p14 (Figure 5D). Except for E592, none are located at interfaces with other components of the SF3b core,

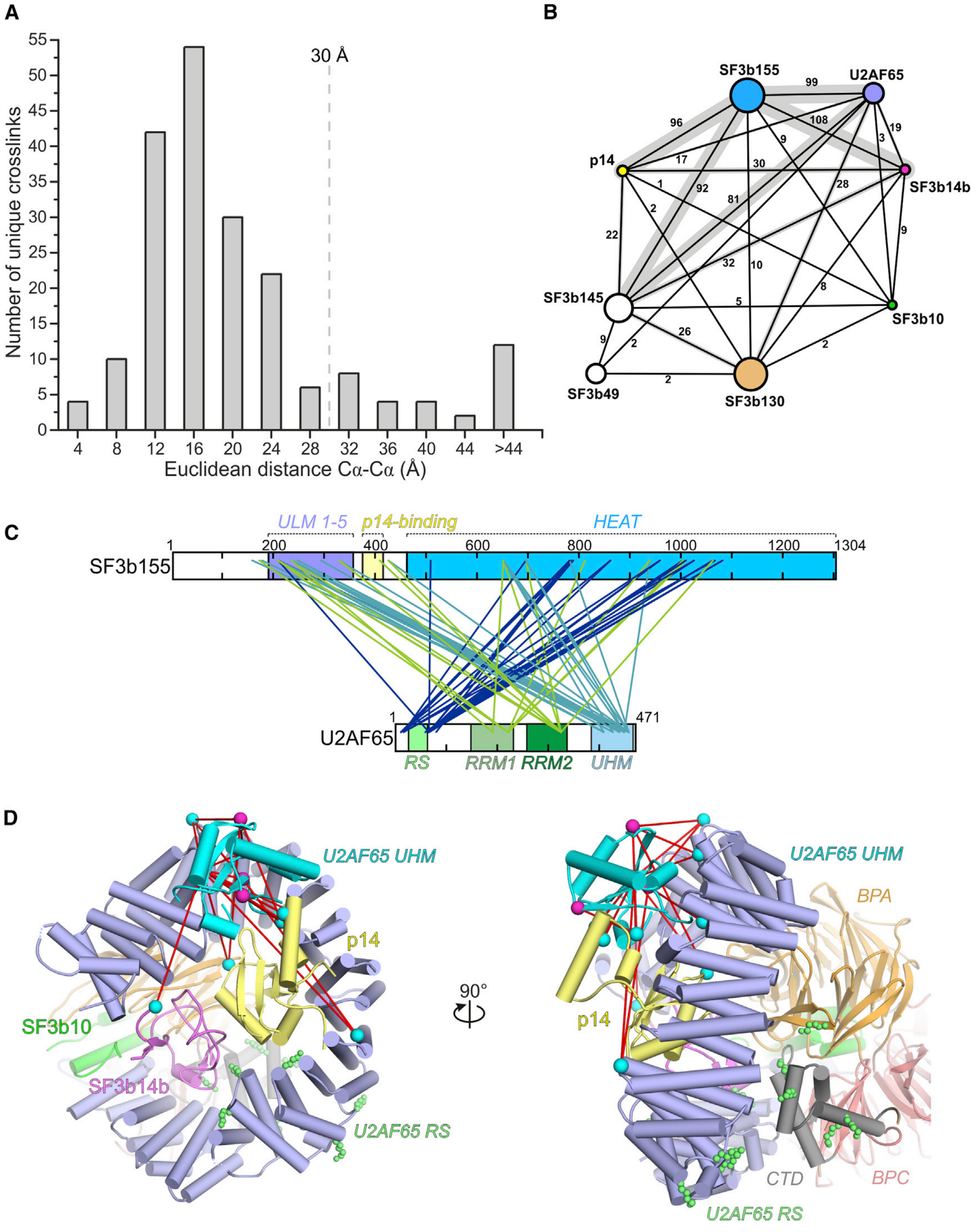


Figure 5. Location of U2AF65 in the SF3b Core

(A) Distribution of the Cα-Cα distances between the observed crosslinked residues in the SF3b-U2AF65 complex. Euclidean distances were computed between the Cα atoms of BS3 crosslinked residues. BS3 crosslinking data at FDR = 1% was filtered ($\text{Score}_{\text{max}} \geq 2$, spectral count ≥ 3; Table S2).

(B) Protein-protein crosslinking network of the SF3b-U2AF65 complex. Subunits of the human SF3b complex are shown as nodes of the graph. The number of high-confidence intermolecular crosslinks is indicated on the connecting lines.

(C) Crosslink map depicting the interaction between SF3b155 and U2AF65. The known domains are indicated, and only high-confidence inter-protein crosslinks are shown ($\text{Score}_{\text{max}} \geq 2$, spectral count ≥ 3). Intermolecular crosslinks between the N-terminal region of U2AF65 (1–94) and SF3b155 are in deep blue. Inter-protein crosslinks between the two RRM domains of U2AF65 (RRM1, RRM2) and SF3b155 are in green, and intermolecular crosslinks between U2AF65 UHM and SF3b155 are shown in cyan.

(legend continued on next page)

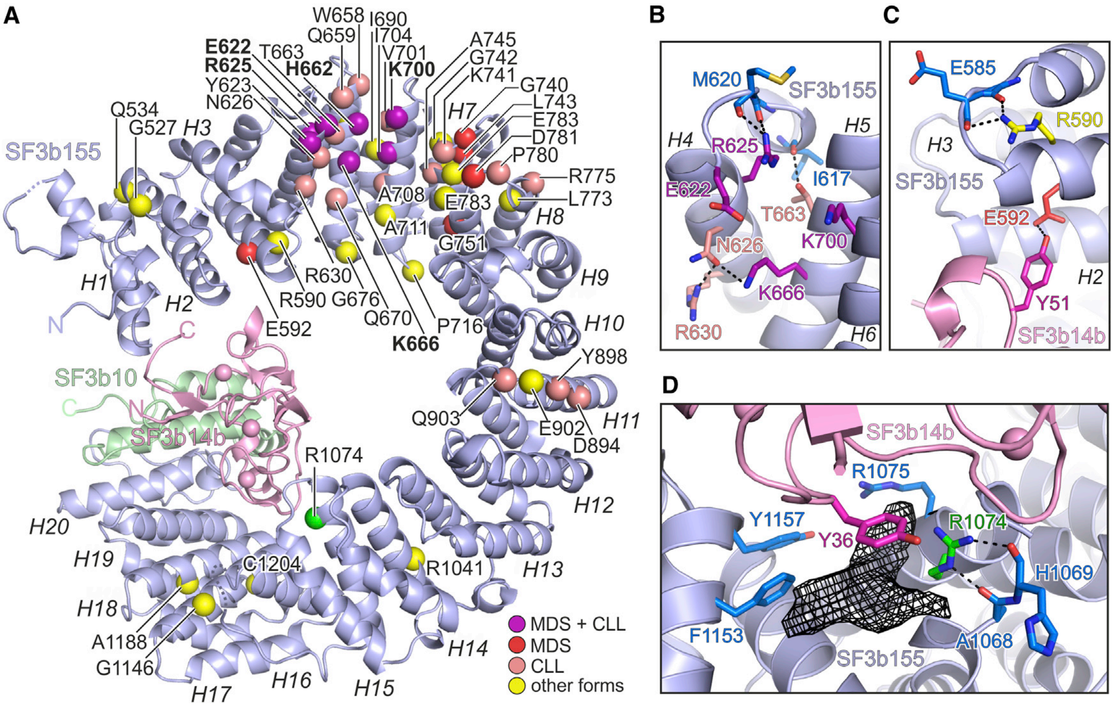


Figure 6. Residues Mutated in Cancer and Pladienolide-Resistant Cells

(A) Overview of residues mutated in various cancers (colored spheres) with mutations associated most frequently with MDS (red), CLL (pink), MDS and CLL (purple), or other cancers (yellow). Mutational hot spot residues are bold, and R1074 is green.

(B) Selected polar residues mutated in human cancers, some are solvent exposed, and some are involved in intermolecular contacts.

(C) Selected polar residues mutated in human cancers, involved in inter- and intramolecular contacts. Other residues are depicted in blue.

(D) Polar contacts between R1074 (green) and other residues of SF3b155. Residues that form the pocket between SF3b155 and SF3b14b are indicated. The unassigned electron density (black mesh) from the pocket is contoured at 3.0σ and may potentially originate in components of the crystallization cocktail or peptides generated by limited proteolysis (data not shown).

See also Figure S7.

whose disruption may destabilize entirely the SF3b155 fold and lead to the total arrest of splicing (Figure 6C).

Many residues whose mutation is linked to cancer are involved in the tertiary structure of SF3b155. Thus, 15 residues are involved in intramolecular hydrogen bonds (e.g., R625, N626, R630, T663, K666; Figures 6A–6C). Twelve are tightly packed in hydrophobic interfaces and thus substitutions would result in sterical clashes (e.g., I704, G742, L747). Three residues are located in α helices, and thus mutation to proline would not be structurally tolerated (Q534, A744, A745). Substitution of these residues likely induces changes in the conformation and specific curvature of the HEAT superhelix. This in turn could diminish the interaction of SF3b155 with the pre-mRNA and other spliceosomal proteins such as p14 or U2AF65, leading to the selection of alternative BS sequences that exhibit greater complementarity to the U2 snRNA and are thus less dependent on the stabilizing effect of SF3b (Alsafadi et al., 2016; Darman et al., 2015). Notably, some of the residues important for the tertiary structure, such as K666 and R625, are among the most frequently mutated in splicing-related cancers (i.e., mutational hot spots) (Yoshida and Ogawa, 2014).

In addition to the residues that contribute to the structure, nine residues in H4–H8, including the hot spot residue K700, are exposed to the solvent and are clustered in or close to the positively charged patch in the immediate vicinity of p14 and U2AF65 (Figures 3C and 5D), suggesting they may affect SF3b155 interactions with the intron at the BS and/or spliceosomal proteins. For instance, it was suggested that hot spot mutations lead to the cryptic 3′ SS selection by altering the contacts with U2AF65 (DeBoever et al., 2015). Indeed, according to our crosslinking, the UHM domain of U2AF65 is located in the proximity of the basic surface patch that includes K700 (Figure 5D). However, we showed by protein-protein crosslinking and in vitro binding assays that the K700E mutation does not affect the stability of the SF3b-U2AF65 complex (Figures S7E and S7F). At the same time, we found that K700E does not decrease the affinity to RNA (Figures S7A and S7B) compared with the wild-type, suggesting that the mechanism

---

(D) Two orientations of the model derived from the BS3 crosslinking data, showing the average location of the UHM domain (375–475; PDB: 4FXW) relative to the SF3b core. Inter-protein crosslinks are depicted as red lines and crosslinked residues as spheres. Side chains of the residues crosslinked to the RS domain are colored light green.

See also Figure S6.

is more complex and may involve additional splicing factors (Figures S7C and S7D).

The SF3b155 mutation R1074H impairs binding of the tumor suppressor pladienolide B (PB) (Yokoi et al., 2011). R1074H forms polar contacts with the backbone of A1068 and H1069, suggesting that it may indirectly impair PB binding by propagated distortions in the HEAT domain. On the other hand, R1074 together with other residues from H15–H17 and Y36 from SF3b14b form a pocket where an unassigned density is visible (Figure 6D). Future investigations are required to fully understand the relevance of this pocket for binding anti-tumor drugs.

Conclusions

SF3b is a multimeric module of the U2 snRNP that plays an essential role in BS recognition and selection, stabilizing the short U2 snRNA-BS RNA duplex that forms at an early stage of spliceosome assembly. The fact that the SF3b subunits p14 and SF3b155 interact directly with one another and also contact adjacent positions at or near the BS has led to the suggestion that the two proteins function during BS recognition in a cooperative manner (Gozani et al., 1998; MacMillan et al., 1994; Query et al., 1996). The structural basis of the latter and the role of other SF3b subunits in the function of SF3b as an integrated unit remained largely unknown. The crystal structure of the SF3b core shows that SF3b130, SF3b10, and SF3b14b form a multipartite scaffold that induces and maintains the specific superhelical conformation of SF3b155. The SF3b155's HD and SF3b14b exhibit RNA-binding properties, and the HD accommodates the BS-binding p14 protein at its center, indicating that the three proteins form a composite RNA-binding platform for BS recognition.

In addition to p14, U2AF65 is another splicing factor that contacts the BS and plays an important role in its selection (Valcárcel et al., 1996). Both proteins bind the intrinsically unfolded NTD of SF3b155, in a protein-protein interaction network required for the early steps of BS recognition. Complementary to this view, our crosslinking experiments indicate that the SF3b core, owing to its structured nature, acts as a scaffold to define the positions of both U2AF65 and p14.

The architecture of SF3b provides structural insight into how cancer-related mutations of SF3b155 or the binding of splicing inhibitors alter BS recognition by the U2 snRNP. That is, most residues mutated in cancer cells contribute to the tertiary structure of SF3b155's HD, while others are located on the outer side of SF3b, where SF3b155 is associated with p14 and U2AF65. Therefore, substitution of SF3b155 residues likely affects not only the conformation of the HD and its surface properties but may also have consequences for positioning p14 and U2AF65 at the BS. The position of other splicing factors that contact SF3b155 at a later stage may also be affected. Changes in the conformation and specific curvature of the HEAT superhelix could diminish the interaction of SF3b with the pre-mRNA and thus destabilize the U2/BS interaction. This in turn may lead to the selection of alternative BS sequences that exhibit greater complementarity to the U2 snRNA (Alsafadi et al., 2016; Darman et al., 2015). Thus, the structural plasticity of the SF3b155 HEAT domain tolerates discrete changes induced by point mutations and tumor inhibitors, at the cost of a loss of BS selection fidelity, aberrant splicing, and cancer.

EXPERIMENTAL PROCEDURES

Detailed protocols can be found in Supplemental Experimental Procedures.

Crystallization and Structure Determination of the Human SF3b Core Complex

The protease-resistant SF3b core complex (see Supplemental Information) was concentrated by ultrafiltration to ~6.5 mg/mL and subjected to crystallization screening. Initial needle-like crystals appeared after 5–7 days at 20°C. Larger single plates were grown at 20°C in 3 μL hanging drops over reservoirs containing 50 mM HEPES-NaOH (pH 6.91–7.05), 200 mM KCl, and 40% (v/v) pentaerythritol propoxylate 5/4. Diffraction data were collected at 100 K on PILATUS 6M detectors at the X10SA (Swiss Light Source) or BL14.1 (BESSY II) beamlines. Data were indexed, integrated and scaled with XDS (Kabsch, 2010), and reduced with POINTLESS, AIMLESS, and CTRUNCATE in CCP4 (Evans, 2011). Initial phases of the SF3b core complex were obtained by SIRAS with autoSHARP (Vonrhein et al., 2007) using data collected at the Ta peak wavelength from a crystal derivatized with Ta_6Br_{12} heavy-atom cluster (see Supplemental Information) and a native data set (native I). Density modification with SOLOMON (Abrahams and Leslie, 1996) yielded an initial electron density map where a partial model could be built with the Autobuild module in Phenix (Adams et al., 2010). Next, the partial model was used to solve the Zn-substructure by MR-SAD (as implemented in Autosol) using weaker data collected at the K edge from a native crystal (native II). The improved experimental map enabled automatic model building with Buccaneer (Cowtan, 2006), followed by iterative rebuilding in COOT (Emsley and Cowtan, 2004) and refinement with Phenix.refine (Afonine et al., 2012). The final model was refined against a more complete data set collected from a native crystal (native III) to an R_{free} value of 25.9% (see Table 1). The quality of the final model was assessed with MOLPROBITY (Chen et al., 2010), and statistics are provided in Table 1. One molecule is present in the asymmetric unit, and the model comprises the C-terminal HEAT domain of SF3b155 (residues 463–1,304), SF3b130 (residues 1–1,217), SF3b10 (residues 15–80), and SF3b14b (residues 7–95). SF3b145 (439–706) and SF3b49 (1–203) were not visible in the electron density, possibly because of low occupancy (caused by partial dissociation during crystal formation) or relative mobility. The macromolecular interfaces between the subunits of the SF3b core complex were analyzed with PISA (Krissinel and Henrick, 2007), and the figures were generated with PyMOL (version 1.7.6, Schrödinger). Multiple sequence alignments were produced with Clustal Omega (Sievers et al., 2013) and annotated with ESPript (Gouet et al., 2003). The sequence conservation was mapped onto the protein structures with ConSurf (Ashkenazy et al., 2010). The electrostatic potentials were computed with APBS (Unni et al., 2011) and visualized in PyMOL.

Crystallization and Structure Determination of the *ct*SF3b130-ctSF3b10 Complex

SF3b130-SF3b10 dimer was expressed in High Five insect cells (see Supplemental Information). Crystals of the proteolysed complex were grown at 4°C in sitting drops by mixing 0.2 μL of protein with 0.2 μL of reservoir solution (0.1 M HEPES-NaOH [pH 7.5], 0.1 M NaCl, 0.1 M $CaCl_2$, 8% [v/v] PEG 4000). Initial crystals appeared after 5 days and grew to full size after 3 weeks. Crystals were cryo-protected in reservoir solution supplemented with 25% (v/v) 2-methyl-2,4-pentanediol and flash-frozen in liquid nitrogen. The structure of the complex was determined by SAD using a crystal derivatized with Ta_6Br_{12}. The final model was refined against a native data set to an R_{free} of 25% (Table S1).

Immunodepletion of SF3b

HeLa nuclear extract was prepared essentially as described (Dignam et al., 1983). Immunodepletion of the SF3b complex was performed using anti-peptide antibodies specific for SF3b155 (Will et al., 2001). Antibodies (10 mg) were coupled to protein A-Sepharose (1 mL) beads using dimethyl pimelimidate (Sigma). Two hundred microliter beads equilibrated in G-750 buffer (750 mM

KCl, 1.5 mM $MgCl_2$, and 20 mM HEPES-KOH [pH 7.9]) were divided into two equal portions and sequentially incubated with HeLa nuclear extract in G-750 buffer for 2 hr at 4°C with end-over-end rotation. The extract was subsequently dialyzed for 5 hr against buffer containing 100 mM KCl, 1.5 mM $MgCl_2$, 20 mM HEPES-KOH (pH 7.9), and 10% (v/v) glycerol. Mock depleted extract was treated in the same manner except the antibody was omitted.

In Vitro Splicing Reactions

Uniformly ^{32}P-labeled MINX pre-mRNA was produced by in vitro transcription. Splicing reaction contained 6 nM pre-mRNA, 40% (v/v) HeLa nuclear extract, 3 mM $MgCl_2$, 60 mM KCl, 20 mM HEPES-KOH (pH 7.9), 2 mM ATP, and 20 mM creatine phosphate. Two hundred nanograms per microliter of recombinant SF3b or native purified SF3b was added to immunodepleted nuclear extract to restore splicing. The reaction mixture was incubated at 30°C for 90 min. RNA was resolved on an 8 M urea-10% polyacrylamide gel.

RNA-Binding Assays

Electrophoretic mobility shift assays were performed using 3′-Cy3 fluorescently labeled probes (see Supplemental Information).

In Vitro Binding Assays

In vitro binding assays were performed with GST-p14 or His_{14}-MBP-U2AF65 fusion proteins and recombinant SF3b complexes. Bound proteins were eluted from the affinity resins under native conditions and analyzed by SDS-PAGE (see Supplemental Information).

Chemical Crosslinking and Crosslink Identification by Mass Spectrometry

Purified SF3b was crosslinked either with 150 μM BS3 (pH 7.9) or 4 mM EDC/1 mM NHS (pH 6.0–6.5) and purified further by size exclusion chromatography using a Superose 6 10/300 column (GE Healthcare). Approximately 100 pmol of crosslinked SF3b from the peak fraction was concentrated to 40 μL using Amicon Ultra with Ultracell-100 membrane (Millipore) and analyzed essentially as described before (Leitner et al., 2012), with the following modifications: the complexes were denatured in 4 M urea/50 mM ammonium bicarbonate, reduced with DTT, alkylated with iodoacetamide, diluted to 1 M urea, and digested with trypsin (1:20 w:w). Peptides were reverse-phase extracted and fractionated by gel filtration on Superdex Peptide PC3.2/30 column (GE Healthcare). Fifty microliter fractions corresponding to an elution volume of 1.2–1.8 mL were analyzed in duplicates on Orbitrap Fusion Tribrid (data set 1) and Q Exactive HF (data set 2) mass spectrometers (Thermo Fisher Scientific). Protein-protein crosslinks were identified by pLink1.22 search engine and filtered at a false discovery rate (FDR) of 1% according to the recommendations of the developers (Yang et al., 2012). Euclidean distances between Cα atoms of crosslinked residues were visualized, measured in PyMOL and analyzed with Mathematica 9 (Wolfram) and OriginPro 9.1 (OriginLab). The BS3 or EDC crosslinks were mapped on the amino acid sequences with xiNET (Combe et al., 2015).

Protein-RNA Crosslinking and Mass Spectrometric Analysis

Protein-RNA crosslinking was performed using UV irradiation at 254 nm followed by the enrichment of crosslinked peptide–RNA heteroconjugates (see Supplemental Information).

ACCESSION NUMBERS

The accession numbers for the structural models and structure factors reported in this paper are PDB: 5IFE (SF3b155-SF3b130-SF3b14b-SF3b10) and 5HY7 (ctSF3b130-SF3b10).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, two tables, and one data file and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2016.08.036.

AUTHOR CONTRIBUTIONS

C.C. established the expression and purification of all human complexes and proteins, and crystallized, determined, and refined the structure of the SF3b core complex. J.S. contributed to the initial expression of human complexes. C.C. and V.P. performed structural analyses. J.S. expressed and purified the *ct*SF3b130-SF3b10 complex and produced the first well-diffracting crystals. J.S. and A.P.S collected diffraction data from the *ct*SF3b130-SF3b10 crystals. A.P.-S. derivatized the crystals and determined the structure. O.D. and C.C. performed protein-protein crosslinking. K.S. and C.C. performed protein-RNA crosslinking. O.D., K.S., and H.U. performed MS and contributed to structural analyses. E.I.D. and C.L.W. performed splicing assays. V.P. supervised the entire project. V.P., C.C., R.L., and C.W. wrote the paper, with contributions from the other authors.

ACKNOWLEDGMENTS

We thank the beamline staff from the X10SA beamline (Swiss Light Source) and the BL14.1 beamline (BESSY II) for assistance with data collection. We are grateful to G. Heyne, H. Kohansal, and T. Conrad for technical assistance and to C. Kibedi, M. Matthes, and A. Sidarovich for providing recombinant expression constructs. C.C. was supported by the PhD program ''Molecular Biology'' from the International Max Planck Research School at the Georg August University Göttingen. The work was funded by the Max Planck Society (V.P. and R.L.) and the German Research Foundation (V.P.).

Received: May 5, 2016
Revised: July 25, 2016
Accepted: August 30, 2016
Published: October 6, 2016

REFERENCES

Abrahams, J.P., and Leslie, A.G. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallogr. D Biol. Crystallogr. *52*, 30–42.

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. D Biol. Crystallogr. *68*, 352–367.

Alsafadi, S., Houy, A., Battistella, A., Popova, T., Wassef, M., Henry, E., Tirode, F., Constantinou, A., Piperno-Neumann, S., Roman-Roman, S., et al. (2016). Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. Nat. Commun. *7*, 10615.

Angers, S., Li, T., Yi, X., MacCoss, M.J., Moon, R.T., and Zheng, N. (2006). Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. Nature *443*, 590–593.

Arning, S., Grüter, P., Bilbe, G., and Krämer, A. (1996). Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. RNA *2*, 794–810.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. *38*, W529–W533.

Berglund, J.A., Rosbash, M., and Schultz, S.C. (2001). Crystal structure of a model branchpoint-U2 snRNA duplex containing bulged adenosines. RNA *7*, 682–691.

Brand, M., Moggs, J.G., Oulad-Abdelghani, M., Lejeune, F., Dilworth, F.J., Stevenin, J., Almouzni, G., and Tora, L. (2001). UV-damaged DNA-binding protein in the TFTC complex links DNA damage recognition to nucleosome acetylation. EMBO J. *20*, 3187–3196.

Cass, D.M., and Berglund, J.A. (2006). The SF3b155 N-terminal domain is a scaffold important for splicing. Biochemistry *45*, 10092–10101.

Champion-Arnaud, P., and Reed, R. (1994). The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. Genes Dev. *8*, 1974–1983.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *66*, 12–21.

Combe, C.W., Fischer, L., and Rappsilber, J. (2015). xiNET: cross-link network maps with residue resolution. Mol. Cell. Proteomics *14*, 1137–1147.

Corrionero, A., Miñana, B., and Valcárcel, J. (2011). Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. Genes Dev. *25*, 445–459.

Corsini, L., Bonnal, S., Basquin, J., Hothorn, M., Scheffzek, K., Valcárcel, J., and Sattler, M. (2007). U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. Nat. Struct. Mol. Biol. *14*, 620–629.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr. D Biol. Crystallogr. *62*, 1002–1011.

Darman, R.B., Seiler, M., Agrawal, A.A., Lim, K.H., Peng, S., Aird, D., Bailey, S.L., Bhavsar, E.B., Chan, B., Colla, S., et al. (2015). Cancer-associated SF3B1 hotspot mutations induce cryptic 3′ splice site selection through use of a different branch point. Cell Rep. *13*, 1033–1045.

DeBoever, C., Ghia, E.M., Shepard, P.J., Rassenti, L., Barrett, C.L., Jepsen, K., Jamieson, C.H., Carson, D., Kipps, T.J., 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.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. *11*, 1475–1489.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Evans, P.R. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr. D Biol. Crystallogr. *67*, 282–292.

Golas, M.M., Sander, B., Will, C.L., Lührmann, R., and Stark, H. (2003). Molecular architecture of the multiprotein splicing factor SF3b. Science *300*, 980–984.

Golas, M.M., Sander, B., Will, C.L., Lührmann, R., and Stark, H. (2005). Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. Mol. Cell *17*, 869–883.

Gouet, P., Robert, X., and Courcelle, E. (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res. *31*, 3320–3323.

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.

Grinthal, A., Adamovic, I., Weiner, B., Karplus, M., and Kleckner, N. (2010). PR65, the HEAT-repeat scaffold of phosphatase PP2A, is an elastic connector that links force and catalysis. Proc. Natl. Acad. Sci. U S A *107*, 2467–2472.

Kabsch, W. (2010). Xds. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. *372*, 774–797.

Leitner, A., Reischl, R., Walzthoeni, T., Herzog, F., Bohn, S., Forster, F., and Aebersold, R. (2012). Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. Mol. Cell. Proteomics *11*, M111.014126.

Li, T., Chen, X., Garbutt, K.C., Zhou, P., and Zheng, N. (2006). Structure of DDB1 in complex with a paramyxovirus V protein: viral hijack of a propeller cluster in ubiquitin ligase. Cell *124*, 105–117.

MacMillan, A.M., Query, C.C., Allerson, C.R., Chen, S., Verdine, G.L., and Sharp, P.A. (1994). Dynamic association of proteins with the pre-mRNA branch region. Genes Dev. *8*, 3008–3020.

Martinez, E., Palhan, V.B., Tjernberg, A., Lymar, E.S., Gamper, A.M., Kundu, T.K., Chait, B.T., 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.

Query, C.C., Strobel, S.A., and Sharp, P.A. (1996). Three recognition events at the branch-site adenine. EMBO J. *15*, 1392–1402.

Query, C.C., McCaw, P.S., and Sharp, P.A. (1997). A minimal spliceosomal complex A recognizes the branch site and polypyrimidine tract. Mol. Cell. Biol. *17*, 2944–2953.

Schellenberg, M.J., Edwards, R.A., Ritchie, D.B., Kent, O.A., Golas, M.M., Stark, H., Lührmann, R., Glover, J.N., and MacMillan, A.M. (2006). Crystal structure of a core spliceosomal protein interface. Proc. Natl. Acad. Sci. U S A *103*, 1266–1271.

Scrima, A., Konícková, R., Czyzewski, B.K., Kawasaki, Y., Jeffrey, P.D., Groisman, R., Nakatani, Y., Iwai, S., Pavletich, N.P., and Thomä, N.H. (2008). Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. Cell *135*, 1213–1223.

Sickmier, E.A., Frato, K.E., Shen, H., Paranawithana, S.R., Green, M.R., and Kielkopf, C.L. (2006). Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. Mol. Cell *23*, 49–59.

Sievers, F., Dineen, D., Wilm, A., and Higgins, D.G. (2013). Making automated multiple alignments of very large numbers of protein sequences. Bioinformatics *29*, 989–995.

Spadaccini, R., Reidt, U., Dybkov, O., Will, C., Frank, R., Stier, G., Corsini, L., Wahl, M.C., Lührmann, R., and Sattler, M. (2006). Biochemical and NMR analyses of an SF3b155-p14-U2AF-RNA interaction network involved in branch point definition during pre-mRNA splicing. RNA *12*, 410–425.

Thickman, K.R., Swenson, M.C., Kabogo, J.M., Gryczynski, Z., and Kielkopf, C.L. (2006). Multiple U2AF65 binding sites within SF3b155: thermodynamic and spectroscopic characterization of protein-protein interactions among pre-mRNA splicing factors. J. Mol. Biol. *356*, 664–683.

Unni, S., Huang, Y., Hanson, R.M., Tobias, M., Krishnan, S., Li, W.W., Nielsen, J.E., and Baker, N.A. (2011). Web servers and services for electrostatics calculations with APBS and PDB2PQR. J. Comput. Chem. *32*, 1488–1491.

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

van Roon, A.M., Loening, N.M., Obayashi, E., Yang, J.C., Newman, A.J., Hernández, H., Nagai, K., and Neuhaus, D. (2008). Solution structure of the U2 snRNP protein Rds3p reveals a knotted zinc-finger motif. Proc. Natl. Acad. Sci. U S A *105*, 9621–9626.

Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007). Automated structure solution with autoSHARP. Methods Mol. Biol. *364*, 215–230.

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

Wang, Q., and Rymond, B.C. (2003). Rds3p is required for stable U2 snRNP recruitment to the splicing apparatus. Mol. Cell. Biol. *23*, 7339–7349.

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, Q., He, J., Lynn, B., and Rymond, B.C. (2005). Interactions of the yeast SF3b splicing factor. Mol. Cell. Biol. *25*, 10745–10754.

Wang, W., Maucuer, A., Gupta, A., Manceau, V., Thickman, K.R., Bauer, W.J., Kennedy, S.D., Wedekind, J.E., Green, M.R., and Kielkopf, C.L. (2013). Structure of phosphorylated SF1 bound to U2AF^{65} in an essential splicing factor complex. Structure *21*, 197–208.

Will, C.L., Schneider, C., MacMillan, A.M., Katopodis, N.F., Neubauer, G., Wilm, M., Lührmann, R., and Query, C.C. (2001). A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. EMBO J. *20*, 4536–4546.

Will, C.L., Urlaub, H., Achsel, T., Gentzel, M., Wilm, M., and Lührmann, R. (2002). Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. EMBO J. *21*, 4978–4988.

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.

Yang, B., Wu, Y.J., Zhu, M., Fan, S.B., Lin, J., Zhang, K., Li, S., Chi, H., Li, Y.X., Chen, H.F., et al. (2012). Identification of cross-linked peptides from complex samples. Nat. Methods *9*, 904–906.

Yokoi, A., Kotake, Y., Takahashi, K., Kadowaki, T., Matsumoto, Y., Minoshima, Y., Sugi, N.H., Sagane, K., Hamaguchi, M., Iwata, M., 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.

Zachariae, U., and Grubmüller, H. (2008). Importin-beta: structural and dynamic determinants of a molecular spring. Structure *16*, 906–915.

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

Molecular Cell, Volume *64*

Supplemental Information

Molecular Architecture of SF3b and Structural Consequences of Its Cancer-Related Mutations

Constantin Cretu, Jana Schmitzová, Almudena Ponce-Salvatierra, Olexandr Dybkov, Evelina I. De Laurentiis, Kundan Sharma, Cindy L. Will, Henning Urlaub, Reinhard Lührmann, and Vladimir Pena

Figure S1

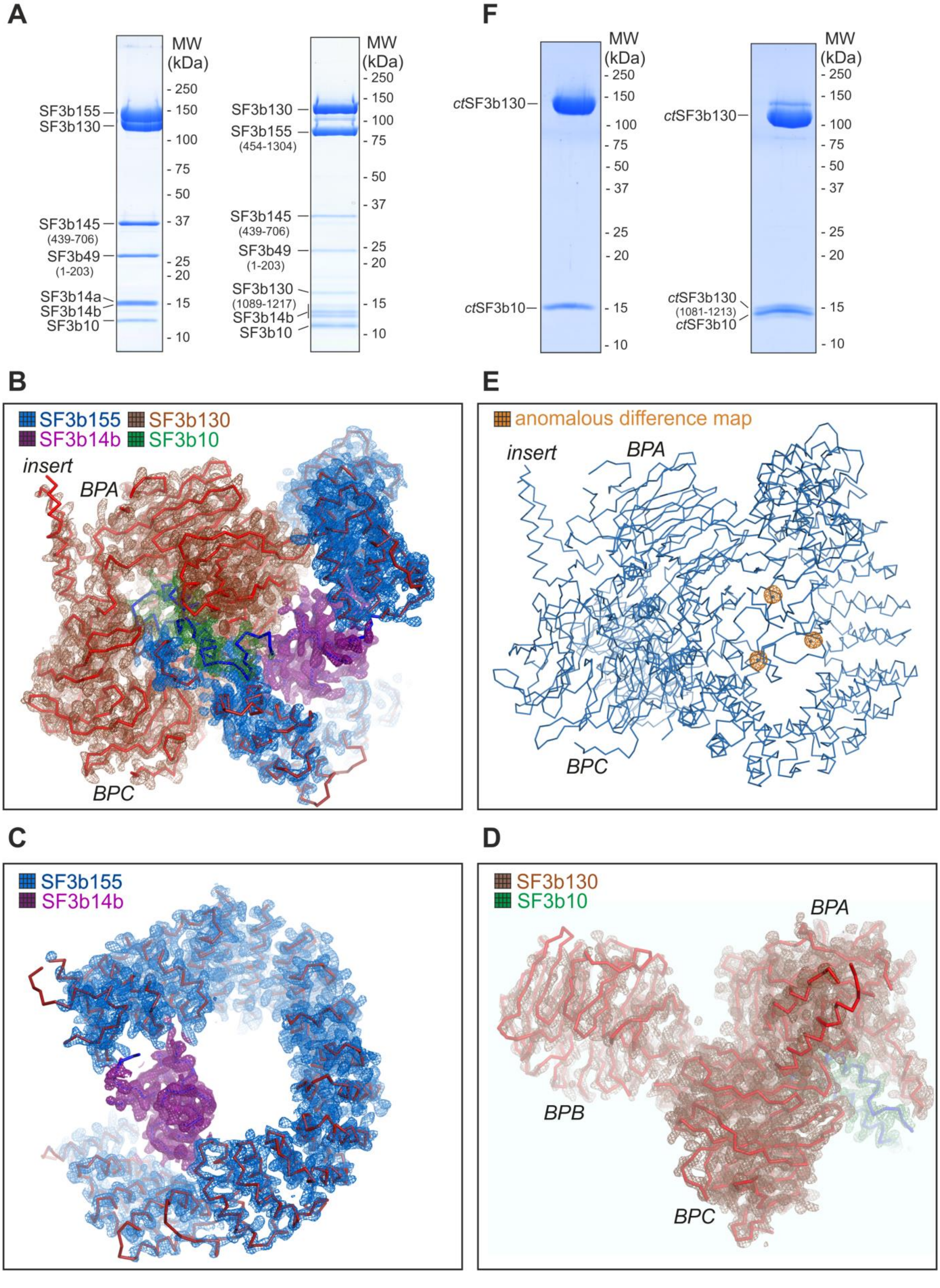

A
MW (kDa)
250
150
100
75
50
37
25
20
15
10
SF3b155
SF3b130
SF3b145 (439-706)
SF3b49 (1-203)
SF3b14a
SF3b14b
SF3b10
SF3b130
SF3b155 (454-1304)
SF3b145 (439-706)
SF3b49 (1-203)
SF3b130 (1089-1217)
SF3b14b
SF3b10
F
ctSF3b130
ctSF3b10
ctSF3b130
ctSF3b130 (1081-1213)
ctSF3b10
B
SF3b155
SF3b130
SF3b14b
SF3b10
insert
BPA
BPC
E
anomalous difference map
insert
BPA
BPC
C
SF3b155
SF3b14b
D
SF3b130
SF3b10
BPA
BPB
BPC

Figure S2

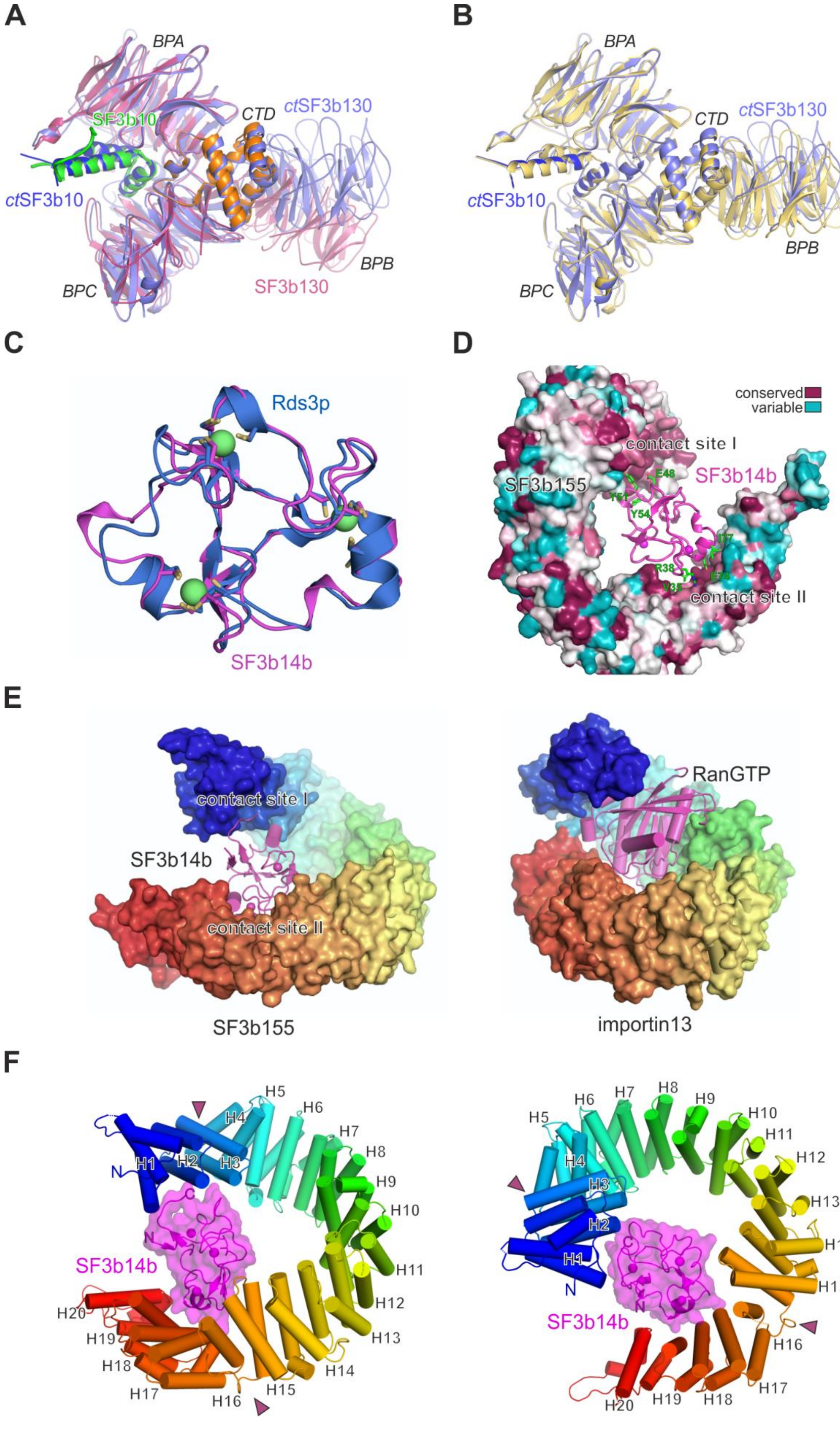

A
BPA
SF3b10
CTD
ctSF3b130
ctSF3b10
BPC
SF3b130
BPB
B
BPA
CTD
ctSF3b130
ctSF3b10
BPB
BPC
C
Rds3p
SF3b14b
D
conserved
variable
contact site I
SF3b155
E48
SF3b14b
Y51
Y54
I77
R38
E74
V35
contact site II
E
contact site I
SF3b14b
contact site II
SF3b155
RanGTP
importin13
F
H1
H2
H3
H4
H5
H6
H7
H8
H9
H10
H11
H12
H13
H14
H15
H16
H17
H18
H19
H20
N
C
SF3b14b
H1
H2
H3
H4
H5
H6
H7
H8
H9
H10
H11
H12
H13
H14
H15
H16
H17
H18
H19
H20
N
SF3b14b

Figure S3

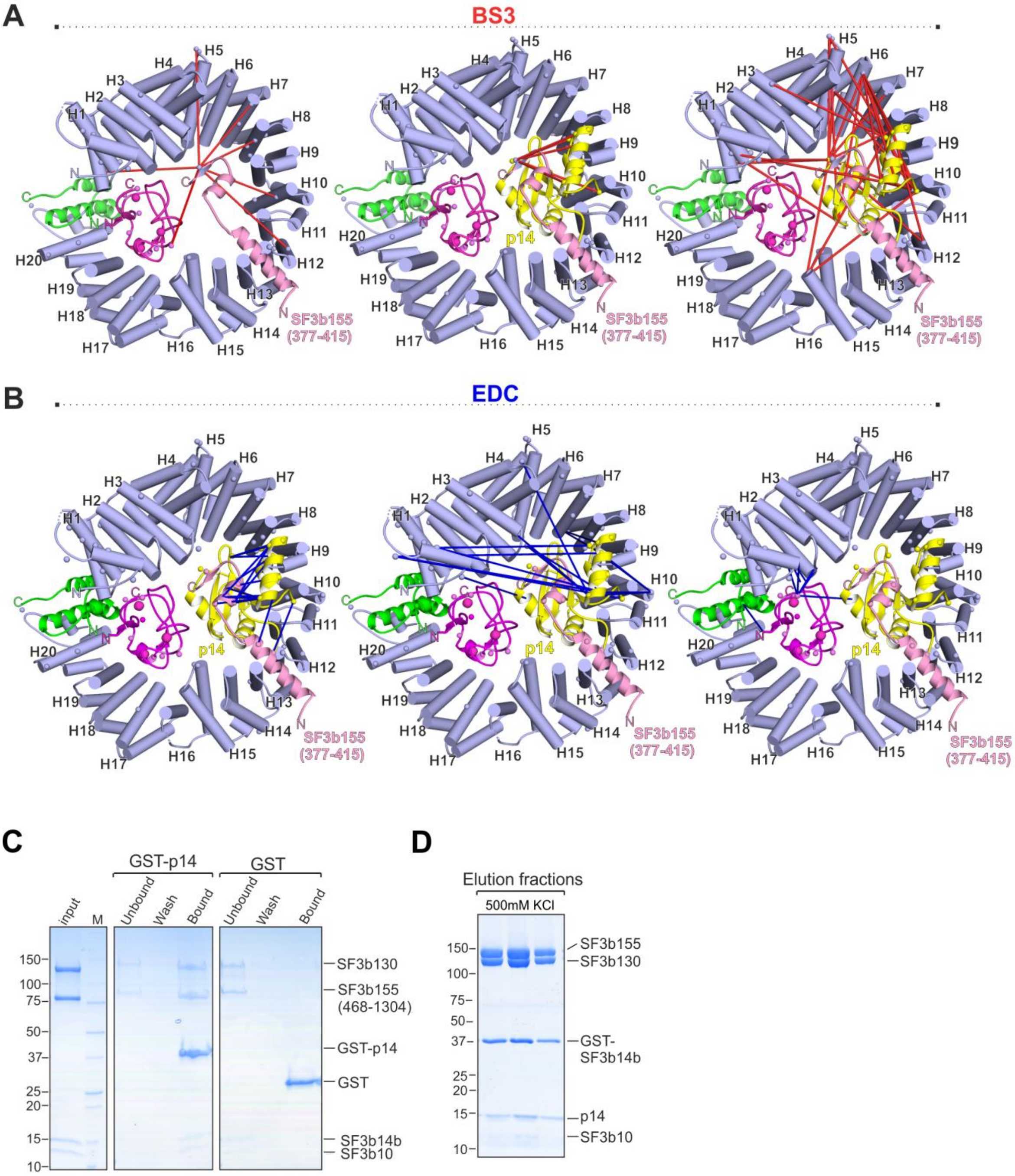
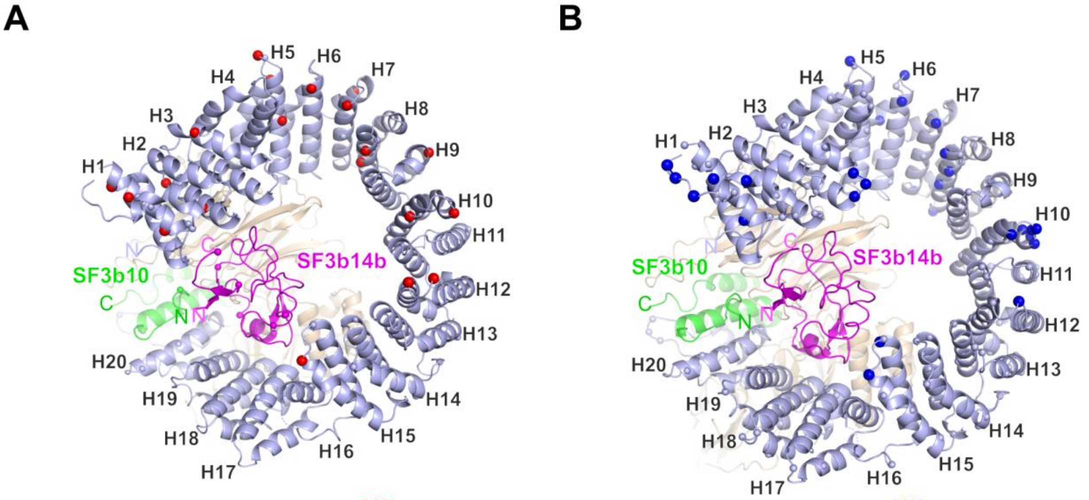
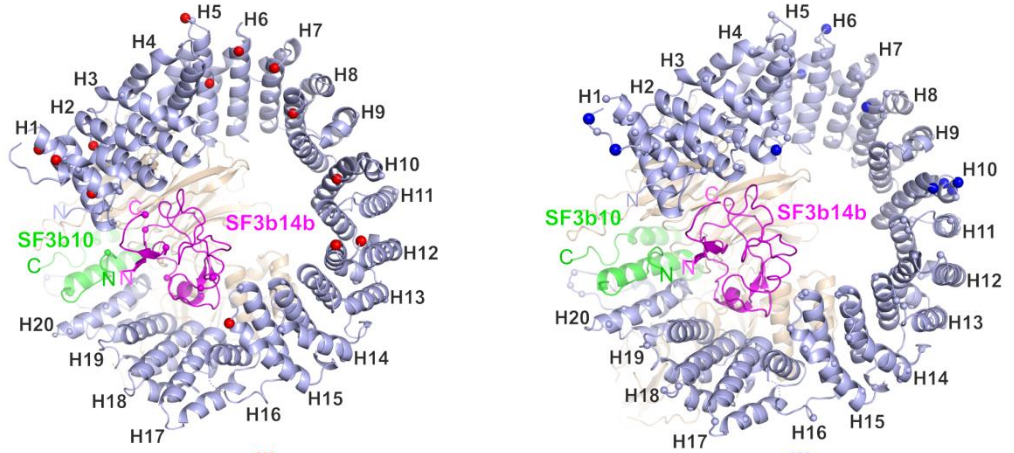

A
BS3
H1
H2
H3
H4
H5
H6
H7
H8
H9
H10
H11
H12
H13
H14
H15
H16
H17
H18
H19
H20
N
C
p14
SF3b155 (377-415)
B
EDC
C
input
M
GST-p14
GST
Unbound
Wash
Bound
150
100
75
50
37
25
20
15
10
SF3b130
SF3b155 (468-1304)
GST-p14
GST
SF3b14b
SF3b10
D
Elution fractions
500mM KCl
SF3b155
SF3b130
GST-SF3b14b
p14
SF3b10

Figure S4



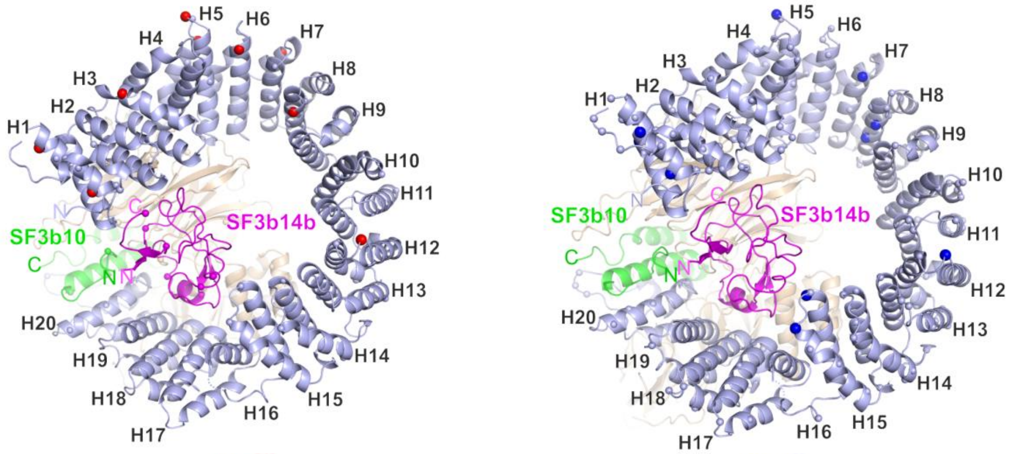
SF3b155 $^{1-376}$

SF3b155 $^{1-376}$



Lys^{333}

Lys^{333}



Lys^{240}

Asp^{69}

Figure S5

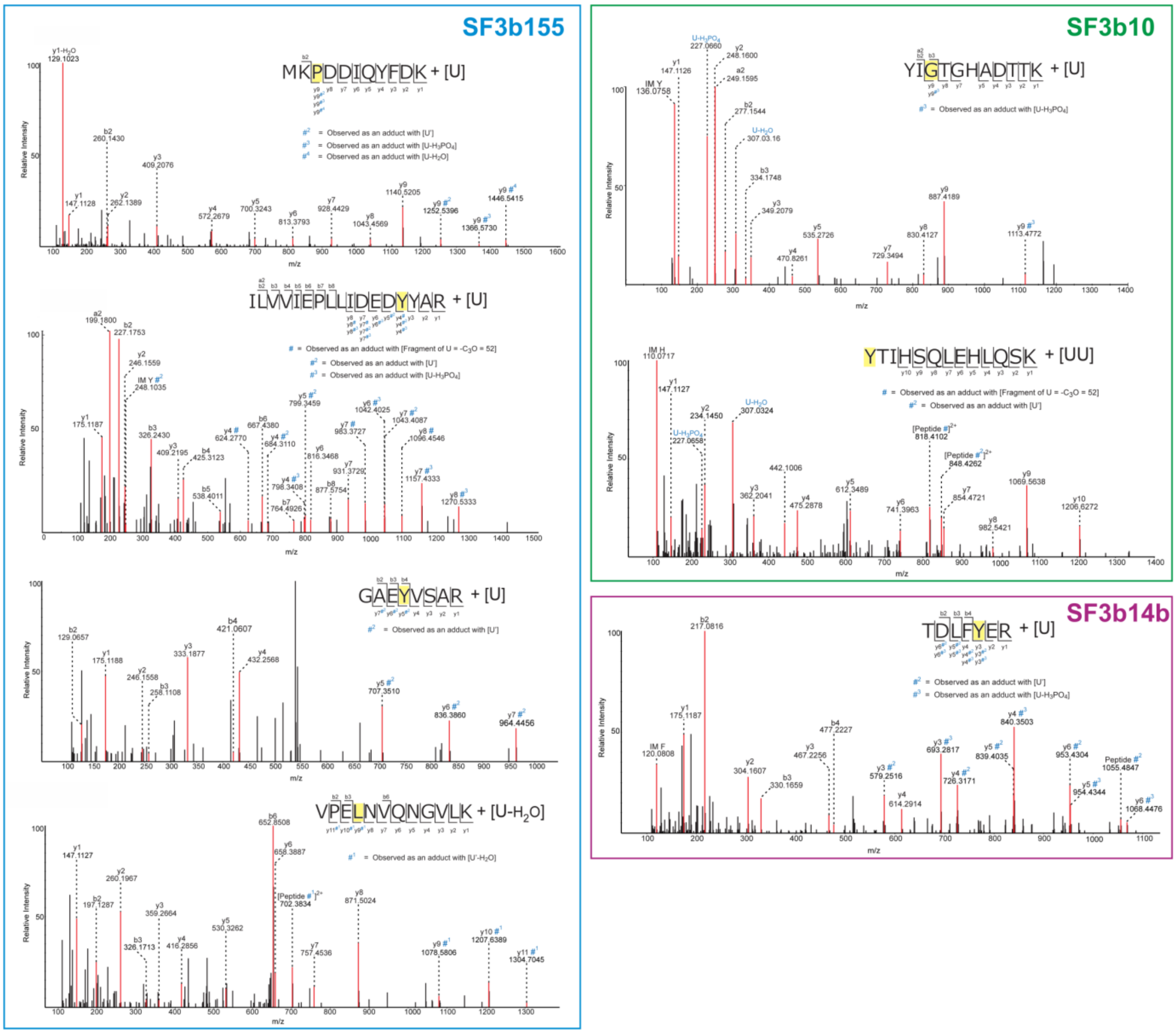

SF3b155
MKPDDIQYFDK + [U]
#2 = Observed as an adduct with [U']
#3 = Observed as an adduct with [U-H3PO4]
#4 = Observed as an adduct with [U-H2O]
ILVVIEPLLIDEDYYAR + [U]
= Observed as an adduct with [Fragment of U = -C3O = 52]
#2 = Observed as an adduct with [U']
#3 = Observed as an adduct with [U-H3PO4]
GAEYVSAR + [U]
#2 = Observed as an adduct with [U']
VPELNVQNGVLK + [U-H2O]
#1 = Observed as an adduct with [U'-H2O]
SF3b10
YIGTGHADTTK + [U]
#3 = Observed as an adduct with [U-H3PO4]
YTIHSQLEHLQSK + [UU]
= Observed as an adduct with [Fragment of U = -C3O = 52]
#2 = Observed as an adduct with [U']
SF3b14b
TDLFYER + [U]
#2 = Observed as an adduct with [U']
#3 = Observed as an adduct with [U-H3PO4]
Relative Intensity
m/z

Figure S6

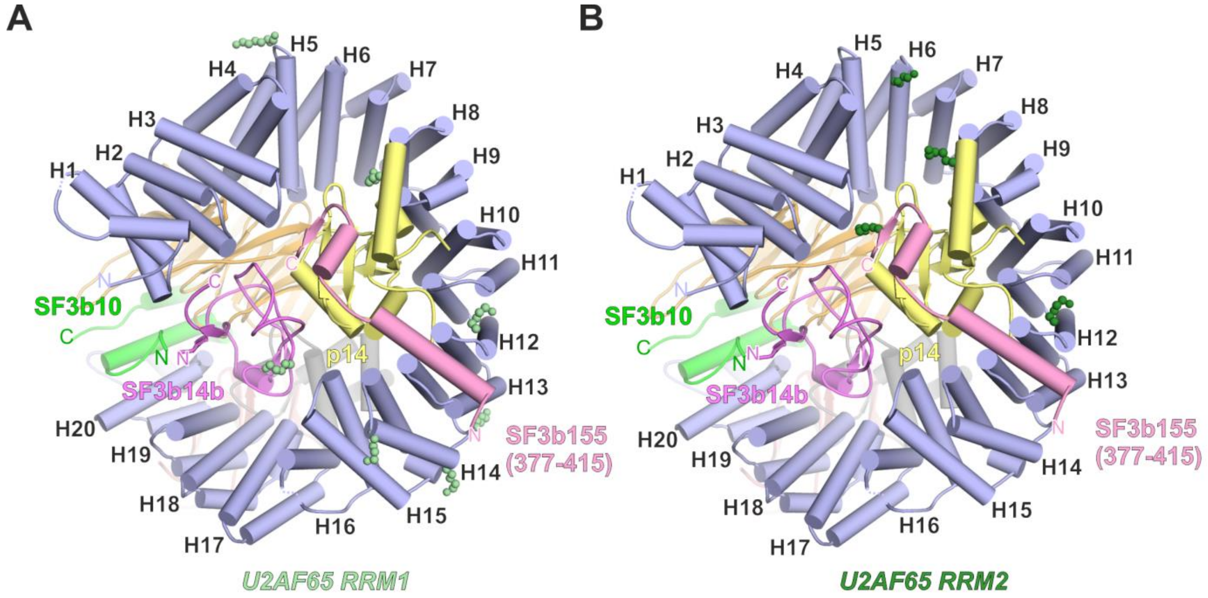
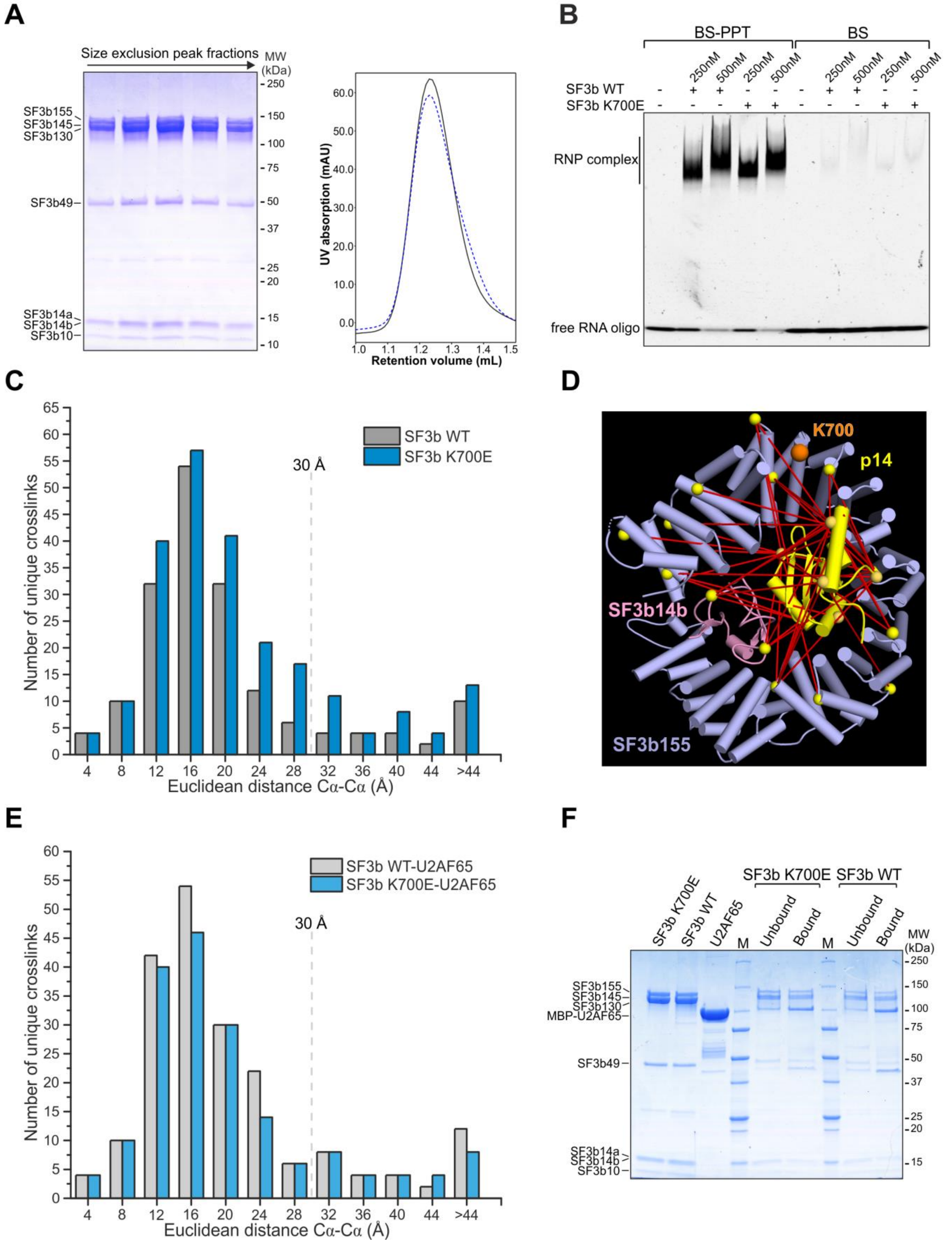

A
H1
H2
H3
H4
H5
H6
H7
H8
H9
H10
H11
H12
H13
H14
H15
H16
H17
H18
H19
H20
N
C
SF3b10
C
N
N
C
SF3b14b
C
p14
N
SF3b155 (377-415)
U2AF65 RRM1
B
H1
H2
H3
H4
H5
H6
H7
H8
H9
H10
H11
H12
H13
H14
H15
H16
H17
H18
H19
H20
N
C
SF3b10
C
N
N
C
SF3b14b
C
p14
N
SF3b155 (377-415)
U2AF65 RRM2

Figure S7


A
Size exclusion peak fractions
MW (kDa)
250
150
100
75
50
37
25
20
15
10
SF3b155
SF3b145
SF3b130
SF3b49
SF3b14a
SF3b14b
SF3b10
UV absorption (mAU)
60.0
50.0
40.0
30.0
20.0
10.0
0.0
1.0
1.1
1.2
1.3
1.4
1.5
Retention volume (mL)
B
BS-PPT
BS
250nM
500nM
250nM
500nM
250nM
500nM
250nM
500nM
SF3b WT
SF3b K700E
RNP complex
free RNA oligo
C
Number of unique crosslinks
SF3b WT
SF3b K700E
30 Å
Euclidean distance Cα-Cα (Å)
4
8
12
16
20
24
28
32
36
40
44
>44
D
K700
p14
SF3b14b
SF3b155
E
Number of unique crosslinks
SF3b WT-U2AF65
SF3b K700E-U2AF65
30 Å
Euclidean distance Cα-Cα (Å)
F
SF3b K700E
SF3b WT
U2AF65
M
Unbound
Bound
SF3b K700E
SF3b WT
MW (kDa)
SF3b155
SF3b145
SF3b130
MBP-U2AF65
SF3b49
SF3b14a
SF3b14b
SF3b10

Supplemental Figures

Figure S1. Electron density maps of the human SF3b core complex. Related to Figure 1.

A. SDS-PAGE of the truncated human SF3b complex before and after (left and right, respectively) limited proteolysis with chymotrypsin.

B. Initial electron density map of the SF3b core complex contoured at 1.2σ level. The final model is depicted as a Cα backbone trace and is superimposed on the map.

C. View of the experimental map of the SF3b core (contour level 1.2σ) with the HEAT domain of SF3b155 and SF3b14b superimposed (red and blue ribbons, respectively).

D. View of SF3b130 and SF3b10 (red and blue ribbons, respectively) superimposed on the initial electron density map (contour level 1.2σ).

E. Anomalous difference Fourier map of the SF3b core complex showing the location of three Zn atoms (orange mesh, contour level 5σ). The difference map was computed from data collected at the Zn K-edge. The final structure is superimposed on the map (blue, ribbon).

F. SDS-PAGE analysis of *Chaetomium thermophilum* (*ct*) SF3b130-SF3b10 before and after digestion with thermolysin (left and right, respectively).

Figure S2. Interactions and conservation of SF3b proteins. Related to Figure 2.

A. Superposition of human SF3b130-SF3b10 and *ct*SF3b130-SF3b10 (r.m.s.d=2.42Å, 877 Cα for SF3b130; r.m.s.d=0.7Å, 50 Cα for SF3b10).

B. Superposition of the two *ct*SF3b130-SF3b10 dimers from the asymmetric unit (r.m.s.d=0.47Å, 865 Cα for *ct*SF3b130; r.m.s.d=0.11Å, 48 Cα for *ct*SF3b10). Note that the BPB domain of *ct*SF3b130 is slightly tilted (~9.5°), suggesting relative mobility.

C. Superposition of human SF3b14b on Rds3p (PDB entry 2K0A, r.m.s.d=0.79Å, 74 Cα).

D. Interactions between SF3b14b and SF3b155. The latter is depicted with surface coloured according to the sequence conservation. Conserved residues of SF3b14b that contact SF3b155 are coloured in green.

E. Side-by-side comparison between SF3b14b-SF3b155 (left) and another HEAT-repeat protein importin13 in complex with RanGTP (PDB: 2X19, right). The surfaces of the HEAT repeats are coloured as a rainbow from the N- (blue) to the C-termini (red). SF3b14b and Ran are coloured magenta.

F. Distinctive conformation of the HEAT domain of SF3b155. The HEAT repeats H1-H20 are coloured as a rainbow from the N- (blue) to the C-termini (red). SF3b14b is coloured magenta. Note the unusual open cavity framed by SF3b14b and the HEAT domain of SF3b155 that accommodates p14/SF3b14a and the two pivoting points in the HEAT superhelix (arrows) at the two contact sites with SF3b14b. For simplicity, SF3b130 and SF3b10 are not shown.

Figure S3. Chemical crosslinking of the human SF3b complex. Related to Figure 3.

A. BS3 crosslinks between SF3b155$^{377-415}$ and the HEAT domain of SF3b155 (left) or p14 (middle). Intermolecular crosslinks between p14 and the HEAT domain of SF3b155 are shown for comparison. BS3 crosslinks are depicted as red lines. Proteins are coloured as in Figure 3.

B. Intermolecular EDC crosslinks between SF3b155$^{377-415}$ and p14 (left), p14 and the HEAT domain of SF3b155 (middle), and between p14 and SF3b14b (right). EDC crosslinks are depicted as blue lines.

C. Direct interaction between the SF3b core complex and p14 shown by GST pull-down assays. The purified SF3b core complex (SF3b155 (468-1304)-SF3b130-SF3b14b-SF3b10) was incubated with either GST-p14 or

GST. After extensive washing, the bound proteins were eluted from the beads and analysed by SDS-PAGE. The input sample is shown for comparison. M – protein marker.

D. GST-SF3b14b can pull-down both SF3b155 and p14 under high salt conditions. The pentameric SF3b155-SF3b130-GST-SF3b14b-p14-SF3b10 was purified on a glutathione resin and eluted by competition. Several elution fractions are shown. Under these conditions, full-length SF3b155 and p14 elute with the other subunits as part of a stoichiometric complex.

Figure S4. Chemical crosslinking pattern of the N-terminal domain of SF3b155. Related to Figure 3.

A. SF3b155 residues crosslinked with BS3 to SF3b155$^{1\text{-}376}$. The BS3 crosslinking patterns of SF3b155^{K333} and SF3b155^{K240} are shown below. BS3 crosslinked lysines are depicted as red spheres (FDR=1%, Score$_{max}$≥2, Spectral count≥3, Table S2). The HEAT repeats of SF3b155 are labelled H1-H20. SF3b core subunits are colour-coded as in Figure 1. Note that individual lysine residues of the NTD crosslink to multiple lysines of SF3b155 HD, resulting in similar overall patterns. This crosslinking pattern indicates that the unstructured NTD assumes an ensemble of numerous conformations in solution.

B. SF3b155 residues crosslinked with EDC to SF3b155$^{1\text{-}376}$. The EDC crosslinking patterns for SF3b155^{K333} and SF3b155^{D69} are shown below. EDC crosslinked residues are depicted as blue spheres (FDR=1%, Score$_{max}$≥2, Spectral count≥3, Table S2).

Figure S5. MS/MS fragmentation spectra of the photo-crosslinked peptides identified in the SF3b core complex. Related to Figure 4.

In each spectrum, the crosslinked peptide sequence and its corresponding y- and b- type fragment ions are indicated at the top, referring to the ions which retain the charge on the N- or C- terminus, respectively. The cross-linked amino-acid residues are highlighted in yellow. Some of the b- and y- ions were observed with a mass shift of #, #^{1}, #^{2}, #^{3}, #^{4} and #^{5} corresponding to -C_3O (a fragment of Uracil), U'-H_2O, U', U-H_3PO_4, U-H_2O and U respectively. IM: Immonium ions; U': Base of U, 112.0273 Da.

Figure S6. Crosslinks of U2AF65 with the SF3b core complex and p14. Related to Figure 5.

A. Crosslinks of the RRM1 domain of U2AF65 (149-231) with the SF3b core complex and p14. BS3 crosslinked residues are shown as spheres and coloured green (FDR=1%, Score$_{max}$≥2, Spectral count≥3, Table S2).

B. Crosslinks of the RRM2 domain of U2AF65 (259-337) with the SF3b core complex and p14. BS3 crosslinked residues are shown as spheres and coloured deep green (FDR=1%, Score$_{max}$≥2, Spectral count≥3, Table S2).

Figure S7. Biochemical and structural analysis of the K700E mutant of SF3b. Related to Figure 6.

A. Reconstitution of the human SF3b complex carrying the K700E mutation. SDS-PAGE analysis (left) of the K700E mutant of SF3b after size exclusion chromatography (right). The size exclusion profile of the wild-type complex is shown for comparison (dashed blue line).

B. Electrophoretic mobility shift assay of BS-PPT (left) or BS (right) RNA oligomers with increasing concentrations of SF3b or the K700E mutant of SF3b. The BS-PPT ssRNA oligomer contains both the BS and the Py tract regions of the intron, while the BS oligomer includes only the BS sequence. Both SF3b and SF3b K700E form stable RNP complexes with BS-PPT, but not with the BS RNA oligo.

C. Distribution of the Cα-Cα distances between crosslinked residues in the SF3b or SF3b K700E complexes. Euclidean distances were computed between the Cα atoms of BS3 crosslinked residues using the atomic coordinates of the SF3b core complex (Table S2). Chemical crosslinking data at FDR=1% were not further filtered.

D. K700E mutation does not induce re-localization of p14. Intermolecular crosslinks between p14 and the HEAT domain of SF3b155 and between p14 and SF3b14b are depicted as red lines (FDR=1%, Score$_{max}$≥2, Spectral count≥3, Table S2). Crosslinked residues are shown as spheres. K700 is coloured orange.

E. Distribution of the Cα-Cα distances between crosslinked residues in the wild-type SF3b-U2AF65 or SF3b K700E-U2AF65 complexes. Euclidean distances were computed between the Cα atoms of BS3 crosslinked residues. Chemical crosslinking data at FDR=1% was filtered ($\text{Score}_{max} \geq 2$, Spectral count≥3, Table S2).

F. K700E mutation does not affect the stability of the SF3b-U2AF65 complex. Purified SF3b or SF3b K700E mutant were incubated with MBP-tagged U2AF65 and amylose beads. The bound proteins were eluted by competition with maltose and analysed by SDS-PAGE. The input samples are shown for comparison. M – protein marker.

Supplemental Tables

Table S1. X-ray data collection and refinement statistics. Related to Experimental Procedures.

***ct*SF3b130-SF3b10**	**Native**	**Ta$_6$Br$_{12}$-derivatized**
Data collection		
Space group	P 61 2 2	P 61 2 2
Cell dimensions		
a, b, c (Å)	197.1,197.1,446.4	198.5, 198.5, 447.9
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Wavelength	1.0	1.254
Resolution (Å)	49.27–2.90 (3.00 – 2.90)*	49.02–3.48 (3.60 – 3.48)
R$_{merge}$ (%)	16 (>100)	21 (>100)#
I/σI	19.41 (0.89)	22.02 (2.45)
CC$_{1/2}$	0.99 (0.36)	1.00 (0.77)
Completeness (%)	100 (100)	100 (97)
Redundancy	20.1 (20.0)	39.0 (37.6)#
Refinement		
Resolution (Å)	2.90	
No. reflections	113502	
R$_{work/}$ R$_{free}$	0.22/ 0.25	
No. atoms		
Protein	19201	
Ligand/ion	9	
B-factors		
Protein	91.63	
Ligand/ion	147.99	
R.m.s deviations		
Bond lengths (Å)	0.006	
Bond angles (º)	0.73	
Ramachandran Plot		
Favored (%)	94%	
Outliers (%)	0.2%	

*Highest resolution shell is shown in parenthesis;
#Friedel mates treated separately

Table S2 (separate file). Descriptive statistics for the crosslinking data of the human SF3b and SF3b-U2AF65 complexes. Related to Figure 3 and Figure 5.

Residue 1 and Residue 2 columns refer to the absolute position of the crosslinked residues in Protein 1 and Protein 2, respectively. Euclidean distances were computed between the Cα atoms of the crosslinked residues present in the crystal structure of the SF3b core complex. "Inter" and "Intra" denote intermolecular and intramolecular crosslinks, respectively. Higher score indicates a higher confidence of crosslink identification.

Supplemental Data S1 (separate file). Multiple sequence alignments for the SF3b core subunits. Related to Figure 2.

SF3b130 (A), SF3b10 (B), SF3b14b (C), and SF3b155 (D) regions present in the crystal structure of the human SF3b core are marked with a coloured line. Secondary structure elements, as present in the crystal structure of the human SF3b core complex, are depicted above the sequences. Triangles indicate residues of the HEAT domain of SF3b155 mutated in human cancers or pladienolide-resistant cells. H1-H20 - HEAT repeats 1-20 of SF3b155; BPA, BPB, BPC - β-propeller domains of SF3b130; CTD - carboxy-terminal domain of SF3b130; *Hs - Homo sapiens*; *Xl - Xenopus laevis*; *Xt - Xenopus tropicalis*; *Tr - Takifugu rubripes*; *Dm - Drosophila melanogaster*; *At - Arabidopsis thaliana*; *Nc - Neurospora crassa*; *Sp - Schizosaccharomyces pombe*; *Sc - Saccharomyces cerevisiae*; *Ct - Chaetomium thermophilum*; *Ce - Caenorhabditis elegans*

Supplemental Experimental Procedures

Expression and purification of the human SF3b complex

The human SF3b complex was produced by co-expression in High Five or SF-9 cell lines using optimized synthetic genes (GeneArt, Life Technologies). To enable the purification of the complex, SF3b130 was cloned in frame with a cleavable N-terminal poly-His_{10} tag and with a C-terminal FLAG tag. SF3b14b was fused to a cleavable GST tag, whereas SF3b145 was cloned in frame with a C-terminal Strep-II tag. The cells overexpressing SF3b were harvested 72-84h after infection and resuspended in ice-cold buffer A1 (50 mM HEPES-KOH, pH 7.9, 600 mM KCl, 15% (v/v) glycerol, 5 mM 2-mercaptoethanol (2-ME), 25 mM imidazole) supplemented with a cocktail of protease inhibitors (Roche) and disrupted by sonication. The cleared lysate was loaded on a HisTrap HP 5 mL column (GE Healthcare) equilibrated with buffer A2 (50 mM HEPES-KOH, pH 7.9, 600 mM KCl, 10% (v/v) glycerol, 5 mM 2-ME, 25 mM imidazole). Proteins were eluted by a gradient formed between buffer A2 and buffer B (50 mM HEPES-KOH, pH 7.9, 600 mM KCl, 10% (v/v) glycerol, 5 mM 2-ME, 500 mM imidazole). Fractions containing SF3b were re-buffered in buffer C (50 mM HEPES-KOH, pH 7.9, 500 mM KCl, 10% (v/v) glycerol, 2 mM DTT (dithiothreitol)) using a HiPrep 26/10 column (GE Healthcare). Afterwards, the His-tag was removed by overnight cleavage with TEV protease (1:100 (w:w), 4-8° C). Next, the sample was loaded on Glutathione Agarose 4B (Macherey-Nagel) equilibrated in buffer C. The recombinant complex was eluted with buffer D (50 mM HEPES-KOH, pH 7.9, 500 mM KCl, 10% (v/v) glycerol, 2 mM DTT, 30 mM L-glutathione reduced) and the GST affinity tag removed by overnight cleavage [1:100 (w:w), 4-8°C] with PreScission Protease. Next, the sample was re-buffered in buffer E (20 mM HEPES-KOH, pH 7.9, 200 mM KCl, 10% (v/v) glycerol, 2 mM DTT) using a HiPrep 26/10 column (GE Healthcare) and further purified on MonoQ (GE Healthcare). The complex was eluted from the anion-exchanger using a linear gradient formed between buffer E and buffer F (20 mM HEPES-KOH, pH 7.9, 1 M KCl, 10% (v/v) glycerol, 2 mM DTT). Fractions containing the SF3b complex were concentrated by ultrafiltration using Amicon Ultra-15 (Millipore) and further applied to a sizing column (Superose 6 10/300 GL, GE Healthcare) equilibrated with buffer G (20 mM HEPES-KOH, pH 7.9, 200 mM KCl, 5% (v/v) glycerol, 2 mM DTT). Peak fractions were pooled, concentrated by ultrafiltration to ~1.5-2mg/mL (when used in splicing assays) or ~4-6 mg/mL (for crystallization trials) and snap frozen in liquid nitrogen. The identities of the purified proteins were confirmed by mass spectrometry and Western blot analyses (data not shown). The K700E mutation was introduced with the QuickChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies) and verified by DNA sequencing and mass spectrometry. The human SF3b complex harbouring the K700E mutation was overexpressed in High Five insect cell line and purified using the same procedure established for the wild-type complex. Analytical gel filtration analyses were performed on a Superose 6 PC 3.2/30 sizing column (GE Healthcare) equilibrated with buffer G.

Crystallization and structure determination of the human SF3b core complex

The full-length SF3b complex did not form crystals. Guided by proteolysis experiments and bioinformatics, we removed regions predicted to be disordered from SF3b145 (439-706) and SF3b49 (1-203). This truncated complex was overexpressed and purified using the same procedure established for the full-length complex. As the truncated complex still did not crystallize, we treated it with chymotrypsin (1:250 (w:w), on ice for 4h), stopped the reaction with a cocktail of protease inhibitors (Roche) and further purified the protease-resistant core complex by size-exclusion chromatography on Superdex 200 10/300 GL (GE Healthcare) equilibrated with buffer G (20 mM HEPES-KOH, pH 7.9, 200 mM KCl, 5% (v/v) glycerol, 2 mM DTT). Peak fractions were concentrated by ultrafiltration using Amicon Ultra-15 (Millipore) to ~6.5mg/mL and subjected to crystallization screening. The crystals of the SF3b core complex were harvested after 10 days at 20^0C and transferred sequentially to the reservoir solution containing 42% (v/v) pentaerythritol propoxylate 5/4 before flash cooling in liquid nitrogen. For experimental phasing, native crystals were transferred to reservoir solution containing 2 mM Ta_6Br_{12} (Jena Bioscience). After incubation for 90 min at 20°C the derivatized crystals were back-soaked in reservoir solution, and then transferred to a reservoir solution containing 42% (v/v) pentaerythritol propoxylate 5/4 and snap frozen in liquid nitrogen.

Expression and purification of the *Chaetomium thermophilum* SF3b130-SF3b10 complex

Optimized synthetic genes (GeneArt, Life Technologies) coding for the proteins SF3b130 and SF3b10 from *Chaetomium thermophilum (ct)* were cloned in frame with a cleavable N-terminal poly-His_{10} tag and a cleavable N-terminal GST tag, respectively and expressed in High Five insect cells. The cells were resuspended in lysis buffer A1 (50 mM Tris-HCl, pH 7.7, 400 mM NaCl, 5 mM imidazole, 2 mM 2-ME, 15% (v/v) glycerol), supplemented with an EDTA-free cocktail of protease inhibitors (Roche), and disrupted by sonication and loaded onto a Ni-NTA column. After elution with lysis buffer supplemented with 300 mM imidazole (buffer B1), the eluate was loaded onto GSH-sepharose 4B column (GE Healthcare). The proteins were eluted with buffer B2 (50 mM Tris-HCl, pH 7.7, 200 mM NaCl, 2 mM DTT, 15% (v/v) glycerol, 30 mM L-glutathione reduced). The His_{10} and GST tags were cleaved by TEV and PreScission proteases, respectively and removed by applying the sample to a Q Sepharose HP column (50 ml) equilibrated in buffer A2 (50 mM Tris-HCl, pH 7.7, 150 mM NaCl, 2 mM DTT, 5% glycerol). The complex was eluted using a linear gradient between A2 and B3 buffer (50 mM Tris-HCl, pH 7.7, 1 M NaCl, 2 mM DTT, 5% (v/v) glycerol). The protein complex was preparatively proteolysed with thermolysin (1:225 (w:w) for 22 hours at 4°C) and the reaction was stopped with EDTA. The protease stable complex was further purified by size exclusion chromatography on Superdex 200 16/600 (GE Healthcare) equilibrated with buffer A3 (20 mM HEPES-NaOH, pH 7.5, 150 mM NaCl, 2 mM DTT, 5% (v/v) glycerol). The fractions containing monomeric complexes were pooled, concentrated to ~4.5 mg/ml, flash frozen and stored at -80°C.

Crystallization and structure determination of the *ct*SF3b130-ctSF3b10 complex

The crystals of the full-length complex diffracted poorly X-rays. Crystals of the proteolysed complex were grown at 4°C in sitting drops by mixing 0.2 µl of protein with 0.2 µl of reservoir solution (0.1M HEPES-NaOH, pH 7.5, 0.1 M NaCl, 0.1 M $CaCl_2$, 8% (v/v) PEG 4000). Diffraction data were collected at the X10SA beamline (SLS, Villigen, Switzerland) and processed with XDS (Kabsch, 2010). The structure of the complex was solved by single anomalous dispersion (SAD) using a crystal derivatized with Ta_6Br_{12}. Determination of the heavy atom substructure was performed with SHELXD (Sheldrick, 2010). Initial SAD phase calculation was carried out with SHELXE and initial models were built with Buccaneer (Cowtan, 2006) and with the Autobuild module in Phenix. The model was further improved by iterative manual rebuilding in COOT (Emsley and Cowtan, 2004) and refinement with Phenix.refine (Afonine et al., 2012). The final model was refined against a native dataset to an R_{free} of 25% (Table S1). Two nearly identical molecules are present in the asymmetric unit, and the model comprises *ct*SF3b130 residues 7-1213 and *ct*SF3b10 residues 27-81.

RNA binding assays

RNA binding of the wild-type SF3b complex and of SF3b^{K700E} was assayed by electrophoretic mobility shift assays. Single-stranded 3'-Cy3 labelled RNA oligomers corresponding to the BS (CGAUACUAACACCG) or BS-PPT (GGGUGCUGACUGGCUUCUUCUCUCUUUUUCCCUC) regions of the intron were obtained from Sigma. A typical reaction (10µL) contained 50nM labelled probe and 50-2000nM protein in Binding buffer (20 mM HEPES-KOH, pH 7.9, 150 mM KCl, 2mM DTT, 1.5 mM $MgCl_{2,}$ 5% (v/v) glycerol). Reactions were incubated for 1h on ice and then loaded onto a 3-8% polyacrylamide nondenaturing gel (NuPAGE, Thermo Fisher Scientific). Gels were run in 1x Tris-glycine buffer for 3h at 4°C and visualized with a Typhoon Trio+ imager (GE Healthcare).

***In vitro* binding assays**

GST-tagged SF3b14a/p14 was overexpressed in *E. coli* and purified as previously described (Spadaccini et al., 2006) with minor modifications: the fusion protein was concentrated to ~5mg/ml and stored in buffer A (20 mM HEPES-KOH, pH 7.9, 100 mM KCl, 2mM DTT and 5% (v/v) glycerol). SF3b155 (468-1304)-SF3b130-SF3b14b-SF3b10 complex was expressed in High Five insect cell line and purified using the procedure established for the full-length assembly. The purified core complex was incubated with GST-p14 or GST and glutathione-agarose resin (Macherey-Nagel) for 1h at 4°C in buffer A. Bound proteins were eluted by competition with 20mM L-glutathione in binding buffer and analysed by SDS-PAGE and mass spectrometry. The SF3b155-SF3b130-GST-SF3b14b-p14-SF3b10 complex was expressed in High Five insect cells and purified first on HisTrap HP (GE Healthcare) and then on GSH-agarose (Macherey-Nagel), as described for the 7-subunit SF3b complex. The pentameric complex was eluted from GSH-agarose under high salt conditions with buffer D (50 mM HEPES-KOH, pH 7.9, 500 mM KCl, 10% (v/v) glycerol, 2 mM DTT, 30 mM L-glutathione reduced) and analysed by SDS-PAGE and mass spectrometry.

U2AF65 isoform 2 was overexpressed in BL21(DE3)pLysS *E. coli* cells as a His_{14}-MBP-fusion protein. The cells were resuspended in lysis buffer (50mM HEPES-K, pH 7.9, 600mM KCl, 15% (v/v) Glycerol, 5mM 2-ME), supplemented with an EDTA-free cocktail of protease inhibitors (Roche) and disrupted by sonication. The cleared lysate was incubated with 3mL amylose resin (NEB) for 2h at 4-8°C and the bound proteins were eluted with elution buffer (20mM HEPES-KOH, pH 7.9, 200mM KCl, 10% (v/v) glycerol, 2mM DTT, 50mM maltose). MBP-tagged U2AF65 was further purified on Q Sepharose HP (GE Healthcare) and by size exclusion chromatography on Superdex 200 Increase 10/300 (GE Healthcare) equilibrated in buffer G (20 mM HEPES-KOH, pH 7.9, 200 mM KCl, 5% (v/v) glycerol, 2 mM DTT). For the reconstitution of the SF3b-U2AF65 complex, the His_{14}-MBP tag was removed by cleavage with TEV protease. Peak fractions were pooled and concentrated by ultrafiltration to ~5.5 mg/mL. Purified MBP-U2AF65 was incubated with the wild-type SF3b or SF3b^{K700E} in buffer G containing 150mM KCl for 1h on ice in 50µL. The samples were then incubated with 25µL amylose resin (NEB) in 200µL for 1h at 4-8°C. The bound proteins were eluted with 50µL elution buffer containing 150mM KCl and analysed by SDS-PAGE. The identities of the bound proteins were verified by mass spectrometry (data not shown).

Chemical crosslinking of SF3b-U2AF65 and crosslink identification by mass spectrometry

The SF3b-U2AF65 (or SF3b^{K700E}-U2AF65) complex was assembled *in vitro* by mixing SF3b (or SF3b^{K700E}) and U2AF65 in a 1:2 molar ratio, and further purified on Superose 6 10/300 (GE Healthcare) equilibrated in buffer G (20 mM HEPES-KOH, pH 7.9, 200 mM KCl, 5% (v/v) glycerol, 2 mM DTT). The purified SF3b^{K700E}, SF3b-U2AF65 and SF3b^{K700E}-U2AF65 complexes were crosslinked with 150µM BS3 and purified by size-exclusion chromatography using Superose 6 10/300 (GE Healthcare). The samples were processed and analyzed in duplicates on Q Exactive HF (Thermo Fisher Scientific), essentially as described above. For simplicity, the crosslink Score is represented as a negative value of the common logarithm of the original pLink Score, that is Score=$-\log_{10}$(“pLink Score”).

Protein-RNA crosslinking and mass spectrometric analysis

Protein-RNA crosslinking was performed essentially as described in (Kramer et al., 2014; Sharma et al., 2015). The SF3b core (SF3b155 (468-1304)-SF3b130-SF3b14b-SF3b10)- polyU$_{25}$ RNP complex was assembled *in vitro* and then UV irradiated at 254 nm for 10 minutes as described in (Sharma et al., 2015). The samples were ethanol precipitated and the pellet was dissolved in 4 M Urea and 50 mM Tris-HCl, pH 7.9. The final concentration of urea was then adjusted to 1 M with 50 mM Tris-HCl, pH 7.9 and the RNA was hydrolysed using 1 µg RNase A (Ambion, Applied Biosystems) for 2 hours at 52 °C. The proteolysis was carried out with trypsin (Promega) at 37 °C, overnight. The sample was desalted to remove non cross-linked RNA fragments using an in-house prepared C18 (Dr. Maisch GmbH) column and the cross-linked peptides were enriched on an in-house prepared TiO_2 (GL Sciences) column using the protocol described in (Sharma et al., 2015). Following the enrichment of peptide-RNA cross-linked heteroconjugates, the samples were dried and resuspended in 5% v/v ACN, 1% v/v FA for mass spectrometry analysis. The sample was injected onto a nano-liquid chromatography system (Dionex, Ultimate 3000, Thermo Fisher Scientific) coupled with Q Exactive HF instrument (Thermo Fisher Scientific) (Sharma et al., 2015). Online ESI-MS was performed in data-dependent mode using a TOP20 HCD method. All precursor ions as well as fragment ions were scanned in the Orbitrap, the resulting spectra were measured with high accuracy (< 5 ppm) both in the MS and MS/MS level. Data analysis was done using a dedicated data base search tool as described in (Kramer et al., 2014) and by implementing the freely available plugin for RNP^{xl} in the recently released version 2.1 of Proteome Discoverer^{TM} (manuscript in preparation).

Supplemental References

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. Acta crystallographica. Section D, Biological crystallography *68*, 352-367.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta crystallographica. Section D, Biological crystallography *62*, 1002-1011.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Kabsch, W. (2010). Xds. Acta crystallographica. Section D, Biological crystallography *66*, 125-132.

Kramer, K., Sachsenberg, T., Beckmann, B.M., Qamar, S., Boon, K.L., Hentze, M.W., Kohlbacher, O., and Urlaub, H. (2014). Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. Nat Methods *11*, 1064-1070.

Sharma, K., Hrle, A., Kramer, K., Sachsenberg, T., Staals, R.H., Randau, L., Marchfelder, A., van der Oost, J., Kohlbacher, O., Conti, E., *et al.* (2015). Analysis of protein-RNA interactions in CRISPR proteins and effector complexes by UV-induced cross-linking and mass spectrometry. Methods *89*, 138-148.

Sheldrick, G.M. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta crystallographica. Section D, Biological crystallography *66*, 479-485.

Spadaccini, R., Reidt, U., Dybkov, O., Will, C., Frank, R., Stier, G., Corsini, L., Wahl, M.C., Luhrmann, R., and Sattler, M. (2006). Biochemical and NMR analyses of an SF3b155-p14-U2AF-RNA interaction network involved in branch point definition during pre-mRNA splicing. RNA *12*, 410-425.

3.2 Structural basis of splicing modulation by antitumor macrolide compounds (*published manuscript*)

Constantin Cretu[1], Anant A. Agrawal[2], Andrew Cook[2], Cindy L. Will[3], Peter Fekkes[2], Peter G. Smith[2] , Reinhard Lührmann[3], Nicholas Larsen[2], Silvia Buonamici[2], Vladimir Pena[1*]

[1]Research Group Macromolecular Crystallography, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

[2]H3 Biomedicine, Inc., Cambridge, MA 03129, USA

[3]Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

Lead Contact: Vladimir Pena

[*] Correspondence to vladimir.pena@mpibpc.mpg.de

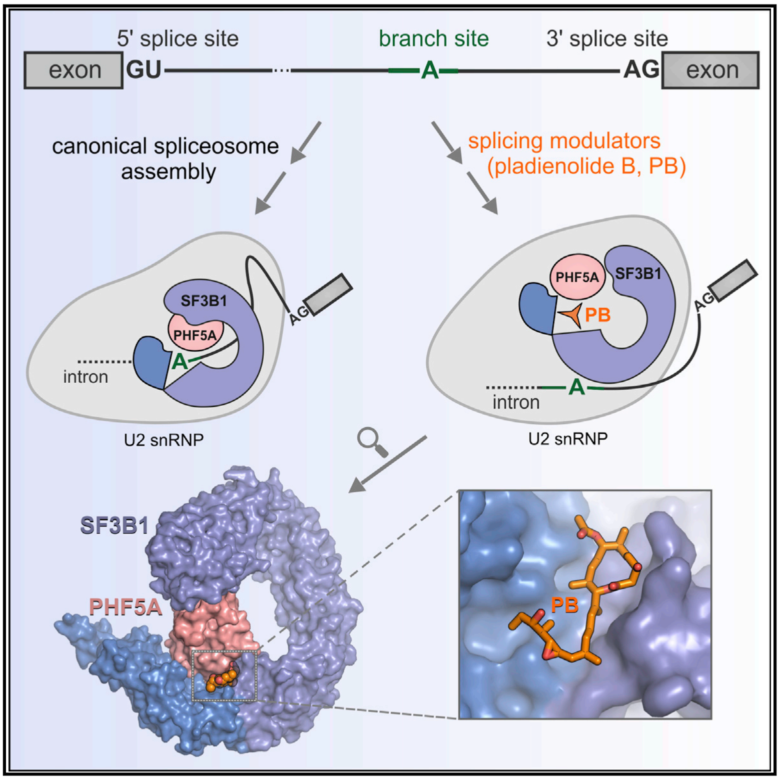
Molecular Cell, *2018*, *Volume 70, pg 265-273.e8*, https://doi.org/10.1016/j.molcel.2018.03.011

Molecular Cell

Article

Structural Basis of Splicing Modulation by Antitumor Macrolide Compounds

Graphical Abstract



Authors

Constantin Cretu, Anant A. Agrawal,
Andrew Cook, ..., Nicholas Larsen,
Silvia Buonamici, Vladimir Pena

Correspondence

vladimir.pena@mpibpc.mpg.de

In Brief

Cretu et al. report the crystal structure of the multi-protein complex SF3B in complex with pladienolide B, a potent antitumor splicing modulator. The structure explains in molecular terms how the small-molecule compound and, by extension, the related splicing modulators interfere with the proper recognition of the branch site region of the intron.

Highlights

- Splicing modulators bind a hinge of SF3B1 and regulate its conformational transition
- Modulator binding regulates formation of the branch site adenosine-binding pocket
- The common pharmacophore of modulators is crucial for their molecular recognition
- Structure of pladienolide B in complex with SF3B provides a framework for drug design

Cretu et al., 2018, Molecular Cell *70*, 265–273
April 19, 2018 © 2018 Elsevier Inc.
https://doi.org/10.1016/j.molcel.2018.03.011

CellPress

Molecular Cell
Article

Structural Basis of Splicing Modulation by Antitumor Macrolide Compounds

Constantin Cretu,[1] Anant A. Agrawal,[2] Andrew Cook,[2] Cindy L. Will,[3] Peter Fekkes,[2] Peter G. Smith,[2] Reinhard Lührmann,[3] Nicholas Larsen,[2] Silvia Buonamici,[2] and Vladimir Pena[1,4,*]
[1]Research Group Macromolecular Crystallography, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany
[2]H3 Biomedicine, Inc., Cambridge, MA 03129, USA
[3]Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany
[4]Lead Contact
*Correspondence: vladimir.pena@mpibpc.mpg.de
https://doi.org/10.1016/j.molcel.2018.03.011

SUMMARY

SF3B is a multi-protein complex essential for branch site (BS) recognition and selection during pre-mRNA splicing. Several splicing modulators with antitumor activity bind SF3B and thereby modulate splicing. Here we report the crystal structure of a human SF3B core in complex with pladienolide B (PB), a macrocyclic splicing modulator and potent inhibitor of tumor cell proliferation. PB stalls SF3B in an open conformation by acting like a wedge within a hinge, modulating SF3B's transition to the closed conformation needed to form the BS adenosine-binding pocket and stably accommodate the BS/U2 duplex. This work explains the structural basis for the splicing modulation activity of PB and related compounds, and reveals key interactions between SF3B and a common pharmacophore, providing a framework for future structure-based drug design.

INTRODUCTION

Pre-mRNA splicing is an essential step of gene expression required for the removal of non-coding introns and ligation of coding exons that generate a mature mRNA. Splicing of each pair of exons occurs in two consecutive chemical reactions, both of which involve reactive groups from three locations of the intron: an adenosine from the so-called branch site (BS) and nucleotides from the 5′ and 3′ splice sites (SS, corresponding to the exon/intron junctions at the two ends of an intron). The entire process occurs within the spliceosome, a multi-megadalton protein-RNA complex that assembles by the stepwise recruitment of five ribonucleoprotein particles (snRNPs), named U1, U2, and U4/U6.U5, plus numerous non-snRNP proteins, to the pre-mRNA (Wahl et al., 2009). During its cyclic assembly, activation, and splicing catalysis pathway, which is driven by ATP, the spliceosome passes through several landmark stages, referred to as the E, A, B, B^{act}, B*, C, C*, P, and intron-lariat spliceosomal complexes (Fica and Nagai, 2017; Shi, 2017; Wahl et al., 2009).

U2 snRNP plays an important role in the recognition of the BS during constitutive and alternative splicing (Lee and Rio, 2015; Will and Lührmann, 2011). It is composed of U2 snRNA, the multimeric splicing factors SF3A and SF3B, and several other associated proteins (Will et al., 2002). In higher eukaryotes, SF3B is a hetero-heptameric complex containing SF3B1/SF3B155, SF3B2/SF3B145, SF3B3/SF3B130, SF3B4/SF3B49, SF3B5/SF3B10, SF3B6/SF3B14a, and PHF5A/SF3B14b (Will et al., 2002). In the pre-spliceosomal A complex, U2 snRNA interacts via base pairing with the BS (Wu and Manley, 1989; Zhuang and Weiner, 1989). At the same time, the U2/BS interaction is stabilized by subunits of the SF3B complex that interact with the intron, including the SF3B1 protein that contacts intron nucleotides flanking the BS adenosine (BS-A) (Gozani et al., 1996, 1998).

Several compounds with antitumor properties, such as pladienolide B (PB) (Kotake et al., 2007), herboxidiene (GEX1A) (Hasegawa et al., 2011), and spliceostatin A (SSA) (Kaida et al., 2007), bind the SF3B complex and inhibit normal pre-mRNA BS usage, inducing changes in alternative splicing patterns (Corrionero et al., 2011; Folco et al., 2011; Teng et al., 2017). Thus, SF3B is a promising target for anti-cancer therapy (Bonnal et al., 2012; Salton and Misteli, 2016). The aforementioned compounds, which are splicing inhibitors/modulators, are derived from bacterial sources (*Pseudomonas sp* and *Streptomyces sp*) and have served as templates for the synthesis of various analogs, some of which have been used in clinical trials (Bonnal et al., 2012; Lagisetti et al., 2013; Webb et al., 2013). Although their chemistry is rather diverse, all modulators from these families contain a central pharmacophore represented by a conjugated diene, flanked by two variable moieties (Bonnal et al., 2012; Effenberger et al., 2017; Lagisetti et al., 2014). However, the significance of this common organization is presently not clear.

Splicing modulators that bind SF3B (Hasegawa et al., 2011; Kaida et al., 2007; Kotake et al., 2007) arrest spliceosome assembly prior to the ATP-dependent formation of a stable A complex, leading to an ‘‘A-like’’ complex where U2 snRNP is less stably bound and the base pairing interactions between U2 and the intron are altered (Corrionero et al., 2011; Effenberger et al., 2014; Roybal and Jurica, 2010). By modulating BS recognition and selection in cells, these compounds lead to intron retention and exon skipping (Corrionero et al., 2011; Teng

© 2018 Elsevier Inc.

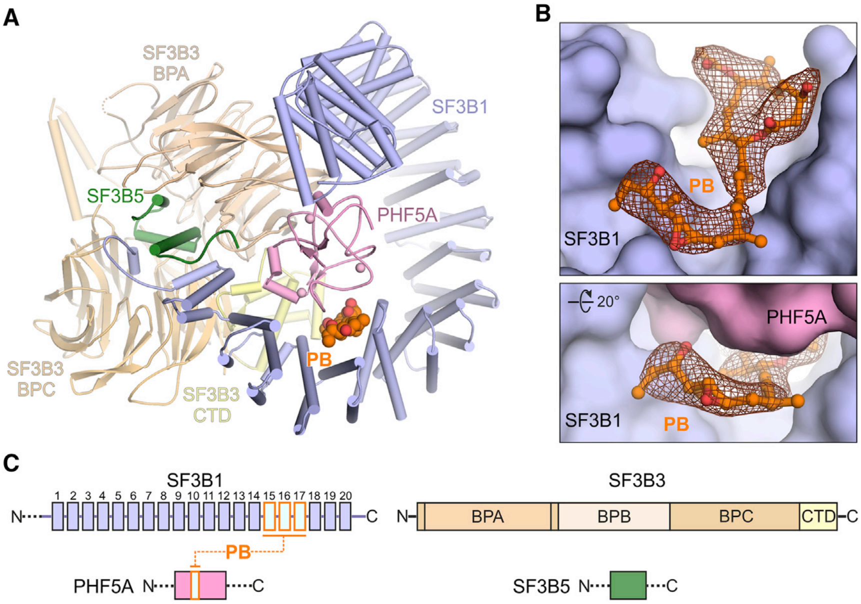


Figure 1. Overall Structure of the Human SF3B Core in Complex with PB

(A) Side view of the SF3B-PB complex. SF3B subunits are shown in different colors. The PB ligand is shown as spheres and the carbon atoms are depicted in orange.

(B) Structure of PB in the SF3B1-PHF5A pocket. PB is shown as orange sticks. The 2*mF*o-*DF*c map (brown mesh) is contoured at 1.0σ and is displayed around PB. The SF3B1 (light blue) and PHF5A (pink) subunits are shown as a surface representation.

(C) Domain composition of the SF3B core. SF3B regions forming the SF3B1-PHF5A tunnel are colored cyan and boxed in orange.

Related to Figure S1.

et al., 2017; Vigevani et al., 2017). The resulting changes in a subset of alternative splicing patterns provide an explanation for the pronounced/preferential cytotoxic activity of these compounds toward malignant versus normal cells (Lee et al., 2016; Obeng et al., 2016; Shirai et al., 2017). Cryoelectron microscopy (cryo-EM) structures of spliceosomes that are formed after the A complex, including the spliceosomal B and B^{act} complexes (Bertram et al., 2017; Plaschka et al., 2017; Rauhut et al., 2016), revealed a “closed” conformation of SF3B relative to its more “open” conformation in the isolated SF3B complex (Cretu et al., 2016). In the “closed” conformation, the terminal HEAT repeats of SF3B1 embrace the BS/U2 helix, while the reactive BS-A is shielded from the 5′ splice site in a conserved SF3B1-PHF5A pocket (Plaschka et al., 2017; Rauhut et al., 2016; Yan et al., 2016). Recent biochemical and chemogenomic analyses suggest that PB, SSA, and herboxidiene share a common binding site on SF3B, in the proximity of the BS-A-binding pocket (Effenberger et al., 2016; Teng et al., 2017). However, the precise nature of the binding site of these splicing modulators and the mechanistic consequences of their binding remain unclear.

Here we have determined the crystal structure of SF3B in complex with PB, providing a paradigm for key principles that govern molecular recognition of SF3B modulators. Together with biochemical analyses, our work reveals the mechanism whereby these compounds affect the function of SF3B and thus modulate pre-mRNA splicing.

RESULTS AND DISCUSSION

Pladienolide B Is Accommodated within an Hourglass-Shaped Tunnel that Matches the Tripartite Organization of Various SF3B Modulators

We have engineered a human SF3B core complex, comprising the HEAT domain of SF3B1/SF3B155, SF3B3/SF3B130, SF3B5/SF3B10, and PHF5A/SF3B14b, that binds a tritiated PB derivative ([^{3}H]-PB) in solution (Figures 1, S1A, and S1B). The complex forms crystals in the presence of PB under similar crystallization conditions as used for the previously reported apo form of the SF3B core (Cretu et al., 2016). The structure of the SF3B core complexed with PB was determined at 3.1 Å by molecular replacement using the apo form as a search model (Figures 1 and S1C–S1E; Table 1), and validated using selenium marker sites (Table 1; Figures S1F and S1G). One molecule is present in the asymmetric unit and, except for a few local changes (Figures S2A and S2B), the SF3B core complex binds PB without significant rearrangement (Figures S2C–S2F). In the crystal structure, PB is accommodated within an ∼15 Å-long hourglass-shaped tunnel defined by seven residues from the HEAT repeats H15-H17 of SF3B1, and two residues from the trefoil-folded protein PHF5A (Figures 2A and 2B). Each “funnel” of the hourglass accommodates one of the two large moieties of PB—the macrolide head (12-membered macrocyclic group) and the aliphatic chain—while the conjugated diene that connects the two moieties is located in the constricted neck region. Importantly, the size of the tunnel suffices to accommodate all types of known SF3B modulators, and its distinctive shape can explain the common organization of the compounds in two moieties separated by a diene (Figure S3A).

The Common Pharmacophore Represented by the Conjugated Diene Is Essential for Molecular Recognition

The diene group is a key pharmacophore feature of the pladienolide, spliceostatin, and herboxidiene families of splicing modulators (Lagisetti et al., 2008, 2014) (Figure S3A). Alteration of this group (C12-C15) or changes in its immediate vicinity lead to significant decreases in activity (Bonnal et al., 2012; Effenberger et al., 2014; Lagisetti et al., 2014). Within the neck of the SF3B1-PHF5A tunnel, the five coplanar carbon atoms of the conjugated diene (C12-C15) are within interacting distances from PHF5A-Y36 and SF3B1-R1074 (Figures 2B and 2C). The diene plane is nearly perpendicular to the aromatic ring of PHF5A-Y36, indicative of π-π interactions established between these two moieties. The required molecular geometry appears to be favored by the steric

Table 1. Data Collection and Refinement Statistics

	SF3B^{core}-PB (PDB: 6EN4)	SF3B^{core} SeMet
Data Collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Wavelength (Å)	1.0	0.979
Cell dimensions: *a*, *b*, *c* (Å)	106.07, 154.57, 210.21	105.59, 153.82, 209.30
Cell dimensions: α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	48.79-3.08 (3.16-3.08)[a]	49.80-3.25 (3.34-3.25)[a]
R_{meas} (%)	15.7 (>100)	28.1 (>100)
$I/\sigma I$	15.4 (1.0)	10.1 (0.9)
$CC_{1/2}$ (%)	99.9 (46.2)	99.7 (44.6)
Completeness (%)	100 (99.8)	99.9 (99.7)
Redundancy	13.4 (13.8)	13.3 (13.1)
Refinement		
Resolution (Å)	48.79-3.08	–
No. reflections	64,404	–
R_{work}/R_{free}	0.23/0.26	–
No. atoms: protein	17,205	–
No. atoms: ligand/ion	41	–
No. atoms: water	0	–
B factors (Å^2): protein	105.59	–
B factors (Å^2): ligand/ion	95.86	–
B factors (Å^2): water	–	–
RMSD: bond lengths (Å)	0.004	–
RMSD: bond angles (°)	0.97	–

Related to STAR Methods. RMSD, root-mean-square deviation.
[a] Statistics for the highest resolution shell are shown in parentheses

constrains exerted by R1074, which flanks the diene plane on one side (Figure 2C). This distinctive ternary arrangement, where two moieties interact in a directional manner, while a third one assists in providing the required geometry, might explain why substitution of Y36 (PHF5A) has largely the same impact on PB binding as substitution of R1074 (SF3B1), as we show by scintillation proximity assays with [^{3}H]-PB (Figures 3A and S3B). Consistent with this, we also observed in viability assays that the HCT116 cancer cell lines carrying mutations in any of these residues (PHF5A-Y36C and SF3B1-R1074H) are resistant to treatment with PB (Figure 3B) and other SF3B modulators (Teng et al., 2017). Overall, the structure shows that the invariant conjugated diene acts as a chemical rigid spacer that constrains the larger PB moieties to the hourglass-shaped tunnel, and is also an important element for molecular recognition by SF3B (Figure 2D), due to its specific interactions in the distinctive microenvironment of the neck region.

The Macrolide Ring and Aliphatic Arm Are Recognized by Shape Complementarity and a Network of Hydrophobic and Weak Polar Contacts

The 12-atom macrolide ring is largely complementary to the shape of the funnel that faces the SF3B3 subunit, where five residues provide an even distribution of contacts along its circumference (Figures 2A and 2D). Thus, several contacts are established with hydrophobic regions of SF3B1 residues F1153, K1071, and V1078, which is consistent with the detrimental effects of chemical modifications of the C3 and C6 groups, with which they interact, on the activity of PB (Effenberger et al., 2014, 2017). The C7 oxycarbonyl group of the macrolide head, regarded as an important constituent of PB's pharmacophore (Lagisetti et al., 2008), is simultaneously engaged in a hydrogen bond with the backbone amide of PHF5A-R38, and a non-polar interaction with SF3B1-Y1157 (Figures 2A and 2D). Epimerization of the C7 group leads to a strong reduction in PB activity (Müller et al., 2011), and removal of the C7 oxycarbonyl group, which yields PB-OH, results in a loss of SF3B affinity and potency in cells (Figures 3B–3D). Of note is that the relatively large size of the macrolide head and the presence of the narrow neck are indicative of the likely directionality of how PB accesses the tunnel.

The aliphatic arm of PB (C16-C23) is oriented at an ~70° angle relative to the diene plane, similar to the structure of its analog herboxidiene (GEX1A) in isolation (Edmunds et al., 2000) (Figures 4A and 4B). Although this conformation is favorable for matching the shape of the tunnel, an additional ~40° rotation about the C16-C17 bond is induced in PB by the contacts with the SF3B1 residues V1110, V1114, and F1153, and PHF5A-Y36 (Figures 2A, 4C, and 4D). The latter two residues simultaneously contact both the aliphatic arm and the other PB moieties, likely leading to additional restrictions in the conformational freedom, which enhances stable binding of the compound to SF3B. These "bridging" interactions can explain why chemical modifications at distant positions of PB reduce its activity only when they occur simultaneously (Effenberger et al., 2014, 2017). Thus, removal of the methyl groups from C16 and C20 or modification of the epoxy group (C18-C19) and of the C11-C12 bond decreases PB activity substantially (Effenberger et al., 2014, 2017).

PB Blocks a Conformational Change in SF3B, Preventing Formation of the BS-A Pocket and Accommodation of the BS/U2 Duplex

In the crystal structure of isolated complexes, SF3B has been observed in an "open" conformation, whose functional importance was unclear as it could not be clearly assigned to a precise step of the splicing cycle (Cretu et al., 2016). In the "open" conformation, the distance between the terminal HEAT repeats is too large to stably accommodate the U2/BS helix (Cretu et al., 2016). In contrast, SF3B1 adopts a "closed" conformation in fully assembled spliceosomes, where it "clamps down" on the U2/BS helix and likely stabilizes the U2/BS interaction (Bertram et al., 2017; Plaschka et al., 2017; Yan et al., 2016). The "closed" SF3B1 conformation is likely to arise within A complexes where U2 snRNA pairs with the BS and SF3B1 contacts nucleotides flanking the BS-A (Gozani et al., 1996, 1998). Importantly, in the presence of PB, SF3B1 exhibits the "open" conformation that is likely to be also present in the "A-like" complexes stalled by SF3B modulators (Corrionero et al., 2011; Folco et al., 2011; Roybal and Jurica, 2010). Structural comparison between the "open" and "closed" conformations of SF3B1 (Rauhut et al., 2016; Yan et al., 2016) revealed that the two structures differ substantially with respect to the position of HEAT repeats H1-H15

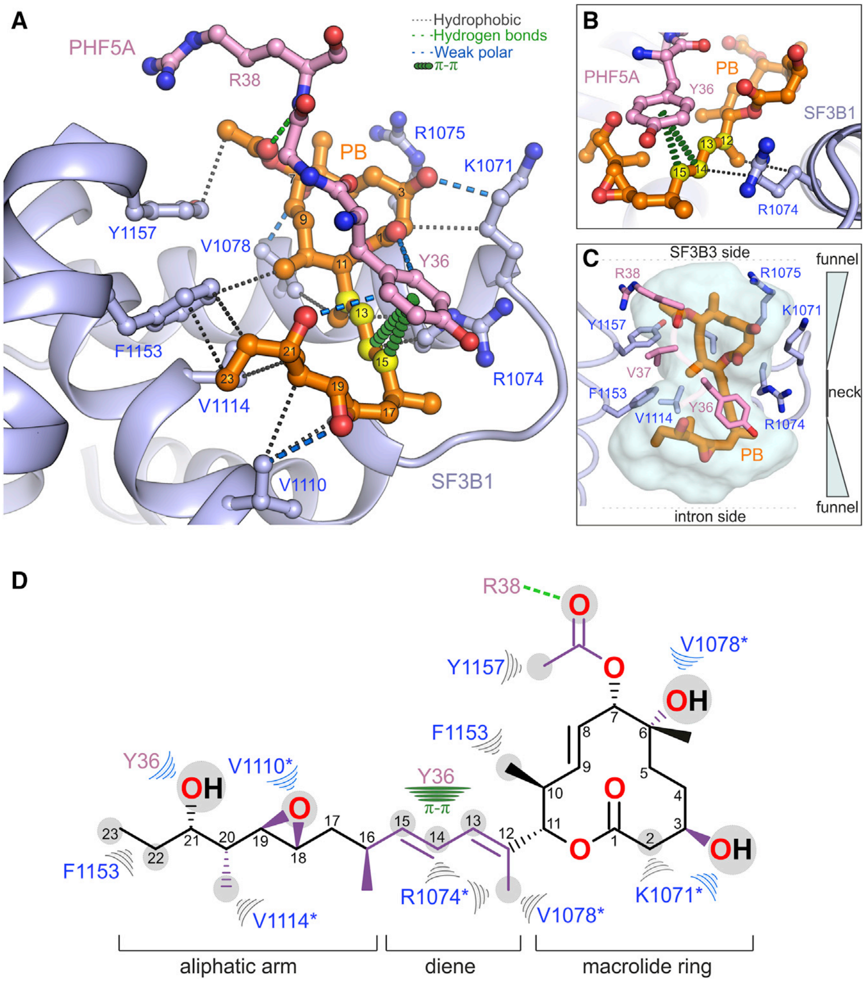


Figure 2. PB Binding to the SF3B1-PHF5A Tunnel

(A) PB is engaged in multiple contacts with residues that define the tunnel. PB (orange) is shown in stick representation and the carbon atoms of the conjugated diene moiety are shown in yellow. SF3B side chains within ~4 Å of PB are shown as sticks. V37 side chain was omitted for clarity.

(B) Overview of the PB-binding tunnel. The interface between SF3B1 H15-H17 HEAT repeats (light blue) and PHF5A (pink) is shown as a surface model and the superimposed PB ligand is represented as orange sticks.

(C) The interactions of the Y36 (PHF5A) and R1074 (SF3B1) residues with the diene moiety of PB. The π-π interactions are colored dark green and are shown as stacked disks, while hydrophobic interactions are depicted as gray dotted lines.

(D) Schematic of the PB interaction with SF3B1 (light blue) and PHF5A (pink). SF3B residues within ~4 Å of PB are shown. Chemical groups that are important for the activity of PB are colored purple. Hydrogen bonds are depicted as dotted green lines; hydrophobic and weak polar contacts are shown as gray and blue semicircles, respectively. The π-π interactions are colored dark green. In the closed SF3B conformation found in the spliceosome, PB's binding site is reconfigured, such that those residues marked with an asterisk are displaced from the unmarked ones.

Related to Figure S2.

relative to the rest of the SF3B core (including H16-H20 and the other three proteins) (Figures S4A–S4E). This large conformational difference is due to the pivoting (~23° rotation) of H1-H15 via a hinge located in the H15-H17 region of SF3B1 (Figures 5A and S4A–S4E). Intriguingly, the binding sites of both PB and the BS-A are present within the same SF3B1 hinge region, in the "open" and "closed" conformation, respectively (Figures 5A and 5B). In spite of their partial overlap, the sites are significantly different in the two states of SF3B1. Thus, SF3B1 residues that form the binding site of PB in the "open" conformation are repositioned in the "closed" state (Figures 2D, 5C, and 5D), where several of them become part of the BS-A pocket (i.e., V1114, K1071, V1078, and R1074; Figures S4F and S4G). At the same time, PHF5A-Y36 is also reoriented in the "closed" conformation to form a hydrogen bond with SF3B1-K1071, which in turn contacts the 3′-hydroxyl moiety of the reactive adenosine (Figures S4F and S4G).

Overall, the structural comparison indicates that a transition of SF3B1 from the "open" to the "closed" state, likely occurring during the assembly of the A complex, would entail the pivoting of HEAT repeats H1-H15 via the hinge region, with concomitant deconstruction of the PB-binding tunnel and formation of the BS-A pocket (Figures 5, S4F, and S4G). By occupying the tunnel, the modulator stabilizes SF3B1 in the "open" conformation, acting as a "wedge" that prevents rotation of the SF3B1 hinge. Consequently, the induced fit transition to the "closed" conformation is impaired, inhibiting formation of the BS-A pocket and stable accommodation of the BS/U2 duplex, and possibly even its formation (Corrionero et al., 2011; Folco et al., 2011) (Figure 5E). Conversely, once the "closed" conformation is achieved, SF3B should no longer be able to bind PB. Indeed, purified B complexes, which have been shown to contain SF3B with a "closed" conformation (Bertram et al., 2017; Plaschka et al., 2017), can be chased into catalytically active spliceosomes even in the presence of PB (Figure S5), consistent with the idea that PB can no longer bind SF3B.

Variations in the Affinity of Modulators for SF3B Lead to Differential Splicing Inhibition with Different Pre-mRNA Substrates

The ability of PB, as well as of other splicing modulators, to block the conformational change in SF3B1 during A complex formation explains the inhibitory effect of these compounds on pre-mRNA splicing observed *in vitro* (Corrionero et al., 2011; Effenberger et al., 2016; Folco et al., 2011). However, introns exhibit differential sensitivity to SF3B modulators in cells, resulting in intron retention and/or exon skipping rather than a general arrest of splicing (Corrionero et al., 2011; Teng et al., 2017). We thus

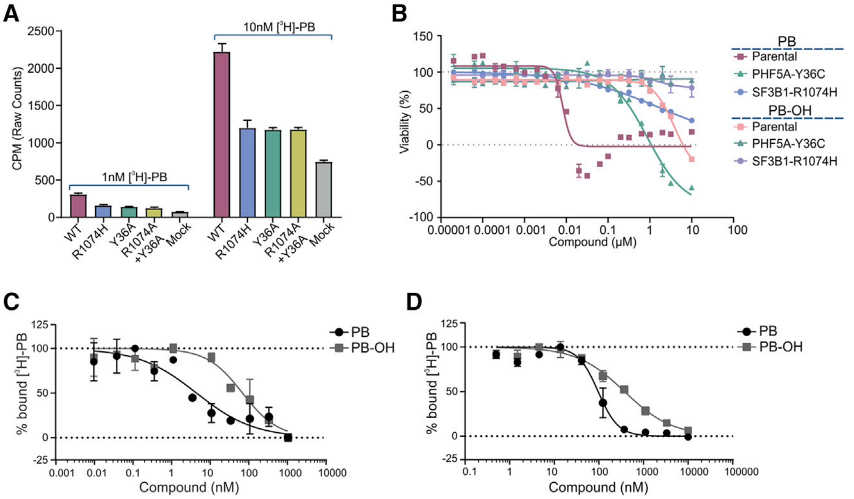


Figure 3. Structure-Guided Probing of the SF3B1-PHF5A PB-Binding Tunnel

(A) Monitoring of [^{3}H]-pladienolide B ([^{3}H]-PB) binding to the full-length wild-type (WT) SF3B complex or to recombinant complexes containing point mutations in the PB-binding tunnel (R1074H [SF3B1], Y36A [PHF5A], and R1074A/Y36A [SF3B1/PHF5A]). The scintillation proximity assays were carried out using 1 nM (left set of bars) or 10 nM (right set) [^{3}H]-PB and 10 nM recombinant FLAG-tagged SF3B complexes. Error bar indicates SD, n = 3.

(B) Cell growth inhibition assays of parental HCT116 cells or of HCT116 cells harboring the PHF5A-Y36C or the SF3B1-R1074H point mutations. HCT116 cell viability was monitored for 72 hr in the presence of PB (orange) or of the less potent PB-OH derivative (purple). Error bar indicates SD, n = 3.

(C) Competitive titration of [^{3}H]-labeled pladienolide B ([^{3}H]-PB, 1 nM) with PB (orange) or PB-OH (magenta) by scintillation proximity assay (SPA) using an endogenous SF3B complex purified by immunoaffinity from HeLa nuclear extract. PB-OH derivative, which lacks the C7 oxycarbonyl group, displays a reduced affinity toward SF3B compared to PB (IC_{50} = 78 nM versus IC_{50} = 5 nM for PB). Error bar indicates SD, n = 3.

(D) Competitive titration of [^{3}H]-PB (1 nM) with PB (orange) or PB-OH (magenta) by SPA using the recombinant full-length SF3B complex (12.5 nM). The PB-OH derivative lacks the C7 oxycarbonyl group and, as a result, exhibits reduced affinity for SF3B (IC_{50} = 378 nM versus IC_{50} = 97 nM for PB). Error bar indicates SD, n = 3.

Related to Figure S3.

next investigated whether there is a correlation between the affinity of a modulator for wild-type SF3B and its ability to inhibit the splicing *in vitro* of different pre-mRNAs. To this end, we tested splicing of substrates that differ with respect to the "strength" of the BS and/or poly-pyrimidine tract (PPT, i.e., the extent to which the BS and PPT differ from the consensus sequences). To produce the "weakest" BS variant, we incorporated an adjacent "decoy" sequence (Corrionero et al., 2011), meant to provide an alternative, unproductive site for binding of U2 snRNA. Notably, PB inhibits splicing efficiently, regardless of the "strength" of the BS or PPT, or the presence of a "decoy" sequence (Figures 6A and S6A). However, a PB derivative with reduced affinity for SF3B, such as PB-OH, is still able to inhibit splicing of pre-mRNA with a "weak" BS, but it is much less effective when a "strong" BS is present (Figures 6B and S6A). In contrast, there was no difference observed with the wild-type versus "weak" PPT (Figures 6B and S6A). This suggests that the correlation between affinity and the "strength" of the

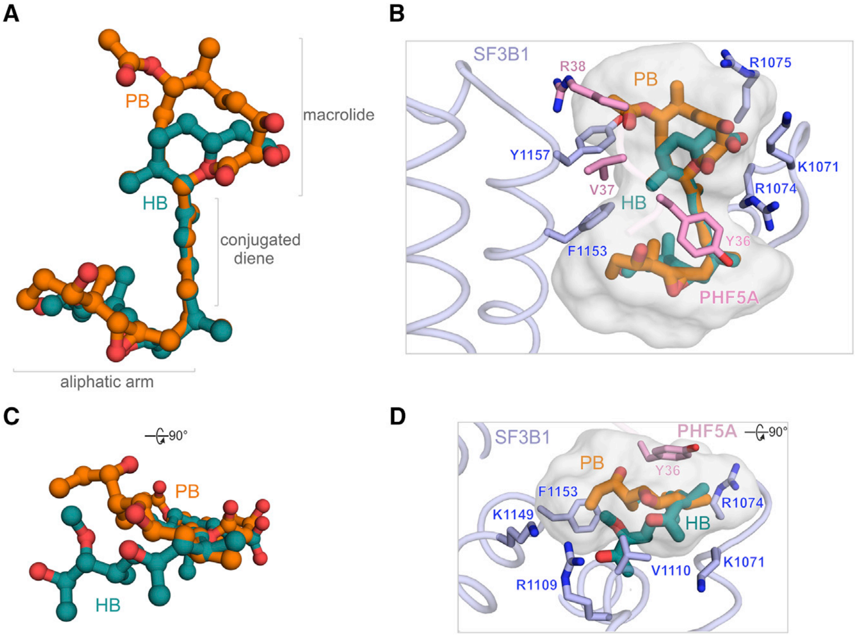


Figure 4. Comparison of PB and HB Structural Features

(A and C) Structural superposition of PB (orange) with the small molecule crystal structure of herboxidiene (HB, teal). The superposition was carried out using the common C12-C16 moiety as a reference. The aliphatic arm adopts different positions in the two structures, while the functionalized pyran ring of HB superimposes onto the PB 12-membered macrocycle. In (C), the structure is depicted as in (A) and rotated as indicated.

(B and D) Tentative location of HB in the SF3B1-PHF5A tunnel. The pyran ring of HB (teal) likely occupies the top SF3B "funnel," as observed with the PB macrocycle. The aliphatic arm of HB is likely located in the bottom "funnel" and may rearrange upon binding to SF3B1-PHF5A. SF3B1 (light blue) and PHF5A (pink) are displayed and colored as in Figure 2A. In (D), the structure is depicted as in (B) and rotated as indicated.

Related to Figure S3.

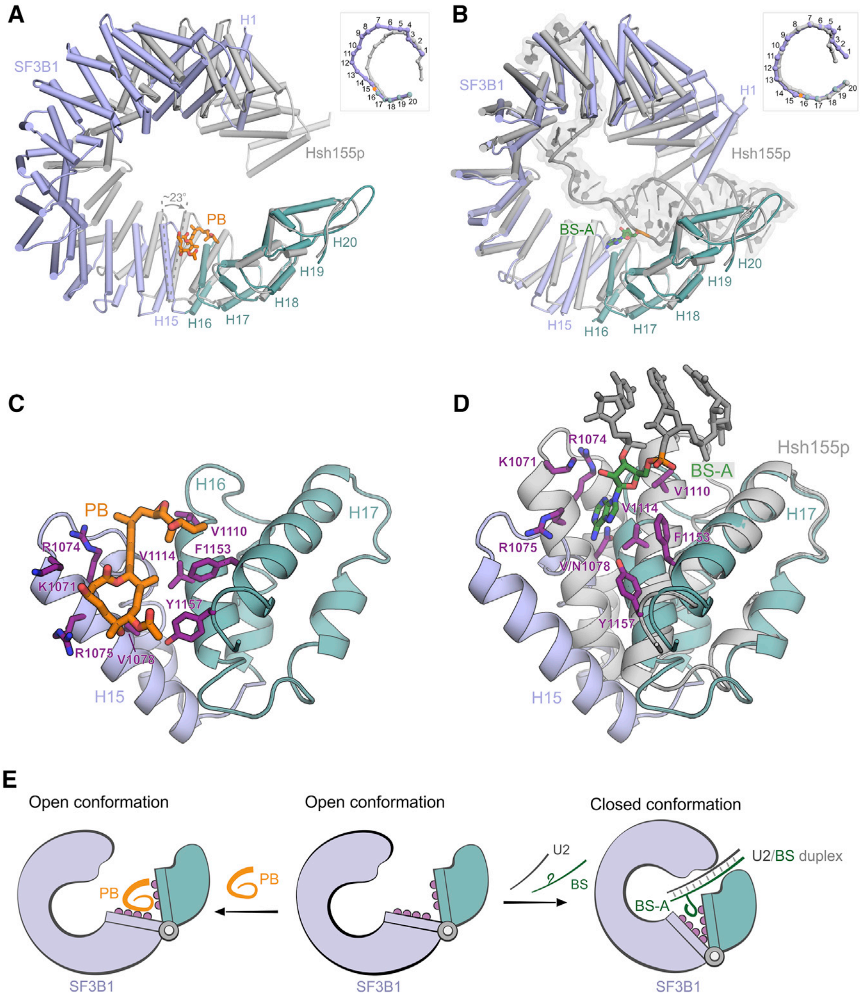


Figure 5. PB and the BS-A Bind SF3B1 in the Same Hinge Region in the Open and Closed Conformation

(A) The SF3B1 HEAT domain ("open" conformation) in complex with PB (orange sticks) and that of *S. cerevisiae* Hsh155p ("closed" conformation, gray) are aligned via the equivalent H16-H20 region (cyan). H1-H15 repeats are shown in light blue. The inset depicts a schematic of the superposition with the centroids of each HEAT repeat shown as spheres.

(B) Structural superposition of H1-H15 and H16-H20 regions of SF3B1 (as separate rigid bodies) with the equivalent regions of Hsh155p. The U2/BS helix and the downstream region of the intron, as seen in the yeast activated spliceosome, are shown in gray. The bulged BS adenosine (BS-A) is highlighted in deep green.

(C) The H15-H17 hinge region of SF3B1 (light blue) in the presence of PB.

(D) The H15-H17 hinges of SF3B1 and Hsh155p (gray) superimposed as in Figure 4A. PB is omitted for clarity and the BS-A is shown. Several residues of PB's binding pocket, as well as the equivalent residues from Hsh155p in the "closed" conformation, are shown in the left and right figures, respectively.

(E) Schematic overview of the mechanism of action of PB. The "open" conformation of SF3B1 is likely found in the U2 snRNP, prior to the stable formation of the A complex. The splicing modulator PB (orange) binds SF3B1 in the "open" state, preventing the stable accommodation of the U2/BS helix and the BS-A, which requires a "closed" conformation of the protein. The hinge (two rectangular elements connected by a ring) and the HEAT repeats H16-H20 (teal) and H1-H15 (light blue) are indicated.

Related to Figures S4 and S5.

substrate's BS enables these compounds to act as modulators rather than mere blockers of splicing (Figure S6B), consistent with their ability to selectively modulate splicing patterns in cells (Corrionero et al., 2011; Teng et al., 2017). Importantly, the model that we derived from our structural data indicates that a direct competition between the BS-A and the modulator is unlikely (Figure 5E) because the binding sites of the two ligands do not exist simultaneously (i.e., one is present in the "open" and the other in the "closed" conformation). Therefore, our model supports the idea that the intron and the modulator compete for the same conformation ("open" form) of SF3B1, rather than for the same binding site. That is, "strong" substrates exhibit higher base pair complementarity within the U2/BS duplex, favoring a shift of the equilibrium from the "open" toward the "closed" state,

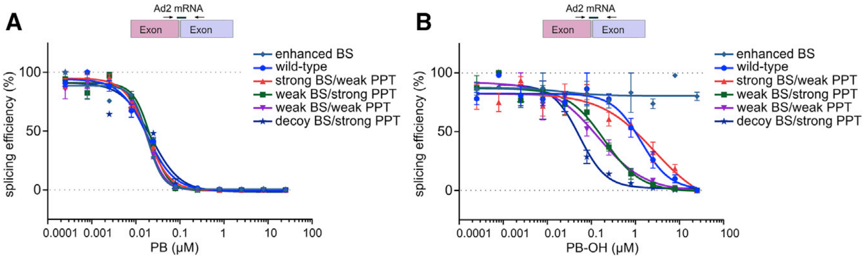


Figure 6. Differential Splicing Inhibition of Various Pre-mRNA Substrates Is Dependent on the Affinity of the Modulator for SF3B

In vitro splicing assays of substrates derived from the Ad2 model pre-mRNA, in the presence of PB (A) or PB-OH (B). Formation of spliced products was monitored by qRT-PCR and the splicing efficiency (%) was determined relative to the DMSO-treated samples. PB exerts its inhibitory action in a 3′ SS sequence-independent manner, whereas the less potent compound, PB-OH, exhibits an apparent sequence-dependent inhibition profile. Related to Figure S6.

likely via an induced fit mechanism, if the modulator binds with sufficiently weak affinity and therefore cannot act as an effective "wedge" (Figure S6B). Conversely, potent antagonists (such as PB), with high affinity for the "open" state of SF3B1, would shift the equilibrium toward a tightly bound "open" state and therefore compete efficiently with a wider range of pre-mRNAs. Similar differential effects were recently observed for the more potent modulator spliceostatin A versus sudemycins (Vigevani et al., 2017).

In light of this model, the correlation between the affinity of modulators for the SF3B-binding tunnel and their ability to selectively alter splicing of various pre-mRNA substrates suggests future strategies for manipulating/tuning their splicing modulation properties by structure-guided modification of their binding affinity.

Concluding Remarks

The crystal structure of the SF3B core in complex with PB advances a unified view on the structural definition of SF3B modulators, and on the structural basis for their inhibitory and modulatory effects on splicing. The hourglass-shaped SF3B tunnel and the contacts with PB explain the tripartite organization of SF3B modulators into two chemically variable moieties connected by an invariant diene—the key components for molecular recognition. From a mechanistic perspective, the structural and biochemical analyses reveal the significance of an early open SF3B conformation and of its conversion to the closed conformation, and indicate how modulators exploit the presence of a functional hinge to inhibit BS usage and modulate splicing.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - ○ Cell lines
- METHOD DETAILS
  - ○ Expression and purification of recombinant SF3B complexes
  - ○ Crystallization and structure determination of the human SF3B core in complex with PB
  - ○ Structural superposition of SF3B1 and Hsh155p
  - ○ Cell viability assays
  - ○ HeLa nuclear extract preparation
  - ○ In vitro splicing assays
  - ○ Scintillation proximity assays/competitive binding assay
  - ○ Affinity purification of B complexes and splicing chase experiments
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and can be found with this article online at https://doi.org/10.1016/j.molcel.2018.03.011.

ACKNOWLEDGMENTS

We thank the staff (Anuschka Pauluhn, Isabelle Martiel, Vincent Olieric, and Justyna Wojdyla) from the X10SA beamline (Swiss Light Source, PSI Villigen, Switzerland) for assistance with data collection. We are grateful to Gabi Heyne, Betty Chan, Sean Irwin, Benjamin Caleb, and Craig Karr for technical assistance. We thank H3 Biomedicine employees for their support in the presented project. This study was supported by the Max Planck Society (C.C., C.L.W., R.L., and V.P.).

AUTHOR CONTRIBUTIONS

C.C. purified, crystallized, and determined the structure of the SF3B core in complex with PB. C.C. expressed and purified SF3B mutants. C.C. and V.P. performed structural analyses. A.A.A., N.L., and P.F. performed and guided the *in vitro* biochemical assays. P.G.S., S.B., and A.C. performed and led the *in vitro* cellular assay testing. C.L.W. and R.L. performed the affinity purification of B complexes and the chase experiment. V.P. initiated and coordinated the project. The manuscript was finalized with input from all authors. Requests for cell lines and RNA substrates should be addressed to S.B. (silvia_buonamici@h3biomedicine.com). Correspondence for the manuscript should be addressed to V.P. (vlad.pena@mpibpc.mpg.de).

DECLARATION OF INTERESTS

C.C., C.L.W., R.L., and V.P. declare no competing interests. A.A.A., A.C., P.F., N.L., P.G.S., and S.B. are employees of H3 Biomedicine, Inc.

Received: November 6, 2017
Revised: February 7, 2018
Accepted: March 7, 2018
Published: April 12, 2018

REFERENCES

Abe, F., and Hiraki, T. (2009). Mechanistic role of ergosterol in membrane rigidity and cycloheximide resistance in Saccharomyces cerevisiae. Biochim. Biophys. Acta *1788*, 743–752.

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Bertram, K., Agafonov, D.E., Dybkov, O., Haselbach, D., Leelaram, M.N., Will, C.L., Urlaub, H., Kastner, B., Lührmann, R., and Stark, H. (2017). Cryo-EM structure of a pre-catalytic human spliceosome primed for activation. Cell *170*, 701–713.e11.

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

Bonnal, S., Vigevani, L., and Valcárcel, J. (2012). The spliceosome as a target of novel antitumour drugs. Nat. Rev. Drug Discov. *11*, 847–859.

Corrionero, A., Miñana, B., and Valcárcel, J. (2011). Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. Genes Dev. *25*, 445–459.

Cretu, C., Schmitzová, J., Ponce-Salvatierra, A., Dybkov, O., De Laurentiis, E.I., Sharma, K., Will, C.L., Urlaub, H., Lührmann, R., and Pena, V. (2016). Molecular architecture of SF3b and structural consequences of its cancer-related mutations. Mol. Cell *64*, 307–319.

Edmunds, A.J., Arnold, G., Hagmann, L., Schaffner, R., and Furlenmeier, H. (2000). Synthesis of simplified herboxidiene aromatic hybrids. Bioorg. Med. Chem. Lett. *10*, 1365–1368.

Effenberger, K.A., Anderson, D.D., Bray, W.M., Prichard, B.E., Ma, N., Adams, M.S., Ghosh, A.K., and Jurica, M.S. (2014). Coherence between cellular responses and in vitro splicing inhibition for the anti-tumor drug pladienolide B and its analogs. J. Biol. Chem. *289*, 1938–1947.

Effenberger, K.A., Urabe, V.K., Prichard, B.E., Ghosh, A.K., and Jurica, M.S. (2016). Interchangeable SF3B1 inhibitors interfere with pre-mRNA splicing at multiple stages. RNA *22*, 350–359.

Effenberger, K.A., Urabe, V.K., and Jurica, M.S. (2017). Modulating splicing with small molecular inhibitors of the spliceosome. Wiley Interdiscip. Rev. RNA *8*, https://doi.org/10.1002/wrna.1381.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr. D Biol. Crystallogr. *69*, 1204–1214.

Fica, S.M., and Nagai, K. (2017). Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. Nat. Struct. Mol. Biol. *24*, 791–799.

Fitzgerald, D.J., Berger, P., Schaffitzel, C., Yamada, K., Richmond, T.J., and Berger, I. (2006). Protein complex expression by using multigene baculoviral vectors. Nat. Methods *3*, 1021–1032.

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.

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.

Hasegawa, M., Miura, T., Kuzuya, K., Inoue, A., Won Ki, S., Horinouchi, S., Yoshida, T., Kunoh, T., Koseki, K., Mino, K., et al. (2011). Identification of SAP155 as the target of GEX1A (Herboxidiene), an antitumor natural product. ACS Chem. Biol. *6*, 229–233.

Jubb, H.C., Higueruelo, A.P., Ochoa-Montaño, B., Pitt, W.R., Ascher, D.B., and Blundell, T.L. (2017). Arpeggio: a web server for calculating and visualising interatomic interactions in protein structures. J. Mol. Biol. *429*, 365–371.

Kabsch, W. (2010). Xds. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., Watanabe, H., Kitahara, T., Yoshida, T., Nakajima, H., et al. (2007). Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. Nat. Chem. Biol. *3*, 576–583.

Kotake, Y., Sagane, K., Owa, T., Mimori-Kiyosue, Y., Shimizu, H., Uesugi, M., Ishihama, Y., Iwata, M., and Mizui, Y. (2007). Splicing factor SF3b as a target of the antitumor natural product pladienolide. Nat. Chem. Biol. *3*, 570–575.

Lagisetti, 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.

Lagisetti, C., Palacios, G., Goronga, T., Freeman, B., Caufield, W., and Webb, T.R. (2013). Optimization of antitumor modulators of pre-mRNA splicing. J. Med. Chem. *56*, 10033–10044.

Lagisetti, C., Yermolina, M.V., Sharma, L.K., Palacios, G., Prigaro, B.J., and Webb, T.R. (2014). Pre-mRNA splicing-modulatory pharmacophores: the total synthesis of herboxidiene, a pladienolide-herboxidiene hybrid analog and related derivatives. ACS Chem. Biol. *9*, 643–648.

Lee, Y., and Rio, D.C. (2015). Mechanisms and regulation of alternative pre-mRNA splicing. Annu. Rev. Biochem. *84*, 291–323.

Lee, S.C., Dvinge, H., Kim, E., Cho, H., Micol, J.B., Chung, Y.R., Durham, B.H., Yoshimi, A., Kim, Y.J., Thomas, M., et al. (2016). Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. Nat. Med. *22*, 672–678.

Liebschner, D., Afonine, P.V., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C., and Adams, P.D. (2017). Polder maps: improving OMIT maps by excluding bulk solvent. Acta Crystallogr D Struct Biol *73*, 148–157.

Luo, M.J., and Reed, R. (1999). Splicing is required for rapid and efficient mRNA export in metazoans. Proc. Natl. Acad. Sci. USA *96*, 14937–14942.

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

McCoy, A.J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr. D Biol. Crystallogr. *63*, 32–41.

Müller, S., Mayer, T., Sasse, F., and Maier, M.E. (2011). Synthesis of a pladienolide B analogue with the fully functionalized core structure. Org. Lett. *13*, 3940–3943.

Obeng, E.A., Chappell, R.J., Seiler, M., Chen, M.C., Campagna, D.R., Schmidt, P.J., Schneider, R.K., Lord, A.M., Wang, L., Gambe, R.G., et al. (2016). Physiologic expression of Sf3b1(K700E) causes impaired erythropoiesis, aberrant splicing, and sensitivity to therapeutic spliceosome modulation. Cancer Cell *30*, 404–417.

Pellizzoni, L., Kataoka, N., Charroux, B., and Dreyfuss, G. (1998). A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. Cell *95*, 615–624.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Plaschka, C., Lin, P.C., and Nagai, K. (2017). Structure of a pre-catalytic spliceosome. Nature *546*, 617–621.

Rauhut, R., Fabrizio, P., Dybkov, O., Hartmuth, K., Pena, V., Chari, A., Kumar, V., Lee, C.T., Urlaub, H., Kastner, B., et al. (2016). Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. Science *353*, 1399–1405.

Roybal, G.A., and Jurica, M.S. (2010). Spliceostatin A inhibits spliceosome assembly subsequent to prespliceosome formation. Nucleic Acids Res. *38*, 6664–6672.

Salentin, S., Schreiber, S., Haupt, V.J., Adasme, M.F., and Schroeder, M. (2015). PLIP: fully automated protein-ligand interaction profiler. Nucleic Acids Res. *43* (W1), W443-7.

Salton, M., and Misteli, T. (2016). Small molecule modulators of pre-mRNA splicing in cancer therapy. Trends Mol. Med. *22*, 28–37.

Shi, Y. (2017). Mechanistic insights into precursor messenger RNA splicing by the spliceosome. Nat. Rev. Mol. Cell Biol. *18*, 655–670.

Shirai, C.L., White, B.S., Tripathi, M., Tapia, R., Ley, J.N., Ndonwi, M., Kim, S., Shao, J., Carver, A., Saez, B., et al. (2017). Mutant U2AF1-expressing cells are sensitive to pharmacological modulation of the spliceosome. Nat. Commun. *8*, 14060.

Teng, T., Tsai, J.H., Puyang, X., Seiler, M., Peng, S., Prajapati, S., Aird, D., Buonamici, S., Caleb, B., Chan, B., et al. (2017). Splicing modulators act at the branch point adenosine binding pocket defined by the PHF5A-SF3b complex. Nat. Commun. *8*, 15522.

Vigevani, L., Gohr, A., Webb, T., Irimia, M., and Valcárcel, J. (2017). Molecular basis of differential 3′ splice site sensitivity to anti-tumor drugs targeting U2 snRNP. Nat. Commun. *8*, 2100.

Villa, R., Kashyap, M.K., Kumar, D., Kipps, T.J., Castro, J.E., La Clair, J.J., and Burkart, M.D. (2013). Stabilized cyclopropane analogs of the splicing inhibitor FD-895. J. Med. Chem. *56*, 6576–6582.

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

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.

Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. *3*, a003707.

Will, C.L., Urlaub, H., Achsel, T., Gentzel, M., Wilm, M., and Lührmann, R. (2002). Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. EMBO J. *21*, 4978–4988.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. *67*, 235–242.

Wu, J., and Manley, J.L. (1989). Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. Genes Dev. *3*, 1553–1561.

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.

Zhuang, Y., and Weiner, A.M. (1989). A compensatory base change in human U2 snRNA can suppress a branch site mutation. Genes Dev. *3*, 1545–1552.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-FLAG antibody	Sigma-Aldrich	Cat#F3165, lot #SLBQ7119V, clone M2
Anti-SF3B3 antibody	Abcam	Cat#AB209403, clone EPR18441
Anti-SF3B1 antibody	MBL	Cat#D221-3, clone 16
Bacterial and Virus Strains		
Escherichia coli BW23474	Dr. Imre Berger, EMBL, Grenoble	N/A
Escherichia coli DH10MultiBacY	Dr. Imre Berger, EMBL, Grenoble	N/A
Escherichia coli XL-10 Gold	Agilent Technologies	Cat#200314
Escherichia coli Rosetta 2 (DE3)	Merck	Cat#71397
Chemicals, Peptides, and Recombinant Proteins		
Pladienolide B (PB), used in crystallization trials	Santa Cruz Biotechnology	Cat#sc-391691
Pladienolide B (PB), used in viability assays, scintillation proximity assays, and *in vitro* splicing assays	This work, Kotake et al., 2007	N/A
[^{3}H]-labeled pladienolide B ([^{3}H]-PB), used in scintillation proximity assays	This work, Kotake et al., 2007	N/A
Pladienolide A (PB-OH), used in viability assays, scintillation proximity assays, and *in vitro* splicing assays	This work, Kotake et al., 2007	N/A
cOmplete Protease Inhibitor Cocktail	Roche	Cat#11836145001
Ampicillin	Carl Roth	Cat#K029.3
Gentamycin	Carl Roth	Cat#0233.4
Kanamycin	Carl Roth	Cat#T832.3
Spectinomycin	Sigma-Aldrich	Cat#S4014
Tetracycline	Sigma-Aldrich	Cat#T8032
IPTG	Carl Roth	Cat#2316.5
X-Gal	Sigma-Aldrich	Cat#B4252
Imidazole	Sigma-Aldrich	Cat#56750
Maltose	Sigma-Aldrich	Cat#M5885
L-Glutathione reduced	Sigma-Aldrich	Cat#G4251
Seleno-L-methionine	Sigma-Aldrich	Cat#S3132
Pentaerythritol propoxylate (5/4 PO/OH)	Sigma-Aldrich	Cat#418749
MS2-MBP	in-house	N/A
T7 RNA polymerase	in-house	N/A
Pfu DNA polymerase	in-house	N/A
RNasin	Promega	Cat#N2511
RQ1 RNase-Free DNase	Promega	Cat#M6101
TEV (Tobacco Etch Virus) protease	in-house	N/A
PreScission protease	GE Healthcare	Cat#27-0843-01
Cre Recombinase	New England Biolabs	Cat#M0298
Sf-900 III SFM insect cell culture medium	Thermo Fisher Scientific	Cat#12658019
ESF 921 insect cell culture medium	Expression Systems	Cat#96-001-01
ESF 921 Δ, methionine deficient insect cell culture medium	Expression Systems	Cat#96-200

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
HisTrap HP resin	GE Healthcare	Cat#17524802
Glutathione Sepharose 4B resin	GE Healthcare	Cat#17075605
HiTrap Q HP resin	GE Healthcare	Cat#17115401
HiLoad 16/600 Superdex 200 pg resin	GE Healthcare	Cat#28989335
Amylose resin	New England Biolabs	Cat#E8021L
Anti-mouse PVT SPA beads	PerkinElmer	Cat#RPNQ0017
Critical Commercial Assays		
QuickChange II XL Site-Directed Mutagenesis Kit	Agilent Technologies	Cat#200521-5
MEGAscript T7 Transcription Kit	Invitrogen	Cat#AM1334
MEGAclear Transcription Clean-Up Kit	Invitrogen	Cat#AM1908
Xtreme-GENE 9 DNA Transfection Reagent	Roche	Cat#06365787001
TaqMan RNA-to-Ct 1-Step Kit	Life Technologies	Cat#4392653
CellTiter-Glo Luminescent Cell Viability Assay	Promega	Cat#G7570
Deposited Data		
PDB coordinates of the human $\text{SF3b}^{\text{core}}$-PB complex	This work	PDB: 6EN4
PDB coordinates of the protease-resistant core of the human SF3b complex	Cretu et al., 2016	PDB: 5IFE
Cryo-EM model of the yeast B^{act} complex	Rauhut et al., 2016	PDB: 5LQW
Cryo-EM model of the yeast B^{act} complex	Yan et al., 2016	PDB: 5GM6
Cryo-EM model of the human B complex	Bertram et al., 2017	PDB: 5O9Z
Cryo-EM model of the yeast B complex	Plaschka et al., 2017	PDB: 5NRL
Experimental Models: Cell Lines		
Hive Five cells (BTI-TN-5B1-4) (*Trichoplusia ni*)	Thermo Fisher Scientific	Cat#B85502
Sf9 cells (*Spodoptera frugiperda*)	Thermo Fisher Scientific	Cat#11496015
HeLa S3 cells (isolation of spliceosomes) (*Homo sapiens*)	Helmoltz Center for Infection Research, Brunswick	N/A
HCT116 cells (*Homo sapiens*)	ATCC	Cat#CCL-247
HeLa S3 cells (*Homo sapiens*)	ATCC	Cat#CCL-2.2
Oligonucleotides		
Primer: SF3B1-fwd-del CCGAGGAGTCGACCATGAAGT CTGTCAACG	This work	N/A
Primer: SF3B1-rev-del AGGCTCTAGATTACAGGATGTA GTCCAGTTCGTAGC	This work	N/A
Primer: SF3B1-fwd-R1074H CCCACAAGAAGGCCATCC ATCGCGCTACCGTCAACAC	This work	N/A
Primer: SF3B1-rev-R1074H GTGTTGACGGTAGCGCGA TGGATGGCCTTCTTGTGGG	This work	N/A
Primer: SF3B1-fwd-R1074A CACAAGAAGGCCATCGCG CGCGCTACCGTCAA	This work	N/A
Primer: SF3B1-rev-R1074A TTGACGGTAGCGCGCGC GATGGCCTTCTTGTG	This work	N/A
Primer: PHF5A-fwd-del GATCTCGAGCCATGGCCAAGC ACCACCCTGA	This work	N/A
Primer: PHF5A-rev-del GCGCTAGCATTACAGGTCGGT CTTAGAAG	This work	N/A
Primer: PHF5A-fwd-Y36A GAAAGTGCGTCATCTGCGA CTCTGCCGTCAGGCCTTG	This work	N/A
Primer: PHF5A-rev-Y36A CAAGGCCTGACGGCAGAGTC GCAGATGACGCACTTTC	This work	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Primer: SF3B3-fwd-del TCTCAAAAGGCCGAAGTCATCATGAACTA	This work	N/A
Primer: SF3B3-rev-del GTCCACTTCGTCGTTGGTGTTAGGAGGCA	This work	N/A
Primer: Ad2-fwd ACTCTCTTCCGCATCGCTGT	This work	N/A
Primer: Ad2-rev CCGACGGGTTTCCGATCCAA	This work	N/A
Primer: Ad2-probe CTGTTGGGCTCGCGGTTG	This work	N/A
Primer: Ftz-fwd TGGCATCAGATTGCAAAGAC	This work	N/A
Primer: Ftz-rev ACGCCGGGTGATGTATCTAT	This work	N/A
Primer: Ftz-probe CGAAACGCACCCGTCAGACG	This work	N/A
Recombinant DNA		
Plasmid: pFL	Dr. Imre Berger, EMBL, Grenoble	N/A
Plasmid: pIDS	Dr. Imre Berger, EMBL, Grenoble	N/A
Plasmid: pFL-10xHIS-TEV	This work, Cretu et al., 2016	N/A
Plasmid: pIDS-GST-PP	This work, Cretu et al., 2016	N/A
Plasmid: pFL-10xHIS-TEV-SF3B3-FLAG/SF3B5/GST-PP-PHF5A/SF3B1/SF3B6 (*H. sapiens* full-length codon-optimized synthetic genes, SF3B3 has a TEV protease cleavable N-terminal 10xHIS tag and a C-terminal FLAG tag; PHF5A has an N-terminal PreScission protease (PP) cleavable GST tag)	This work, Cretu et al., 2016	N/A
Plasmid: pFL-10xHIS-TEV-SF3B2-Strep/SF3B4 (*H. sapiens* full-length codon-optimized synthetic genes, SF3B2 has a TEV protease cleavable N-terminal 10xHIS tag and a C-terminal Strep-tag II)	This work, Cretu et al., 2016	N/A
Plasmid: pFL-10xHIS-TEV-SF3B3-FLAG/SF3B5/GST-PP-PHF5A/SF3B1 (R1074H)/SF3B6 (*H. sapiens* full-length codon-optimized synthetic genes, SF3B3 has a TEV cleavable N-terminal 10xHIS tag and a C-terminal FLAG tag, PHF5A has an N-terminal PreScission protease (PP) cleavable GST tag, SF3B1 residue R1074 mutated to H)	This work	N/A
Plasmid: pFL-10xHIS-TEV-SF3B3-FLAG/SF3B5/GST-PP-PHF5A (Y36A)/SF3B1 (R1074A)/SF3B6 (*H. sapiens* full-length codon-optimized synthetic genes, SF3B3 has a TEV cleavable N-terminal 10xHIS tag and a C-terminal FLAG tag, PHF5A has an N-terminal PreScission protease (PP) cleavable GST tag, SF3B1 residue R1074 mutated to A, PHF5A residue Y36 mutated to A)	This work	N/A
Plasmid: pFL-10xHIS-TEV-SF3B3-FLAG/SF3B5/GST-PP-PHF5A (Y36A)/SF3B1/SF3B6 (*H. sapiens* full-length codon-optimized synthetic genes, SF3B3 has a TEV cleavable N-terminal 10xHIS tag and a C-terminal FLAG tag, PHF5A has an N-terminal PreScission protease (PP) cleavable GST tag, PHF5A residue Y36 mutated to A)	This work	N/A
Plasmid: pFL-10xHIS-TEV-SF3B3 (Δ 1068-1085)-FLAG/SF3B5/GST-PP-PHF5A (1-98)/SF3B1 (453-1304) (*Homo sapiens* codon-optimized synthetic genes, SF3B3 has a TEV cleavable N-terminal 10xHIS tag and a C-terminal FLAG tag, PHF5A has an N-terminal PreScission protease (PP) cleavable GST tag, SF3B1 lacks residues 1-452, SF3B3 lacks residues 1068-1085, PHF5A lacks residues 99-110)	This work	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Plasmid: pMBP-MS2	Josep Vilardell lab	Addgene Cat#65104
Plasmid: pT7-MINX-M3	Bertram et al., 2017	N/A
Plasmid: pcDNA3.1(+)	Thermo Fisher Scientific	Cat#V79020
Plasmid: pSP65-Ftz	Luo and Reed, 1999	Addgene Cat#11243
Plasmid: pcDNA3.1(+)-Ad2 (wild-type)	This work, Pellizzoni et al., 1998	N/A
Plasmid: pcDNA3.1(+)-Ad2.1 (“enhanced” BS)	This work	N/A
Plasmid: pcDNA3.1(+)-Ad2.11 (“weak” BS, “strong” PPT)	This work	N/A
Plasmid: pcDNA3.1(+)-Ad2.12 (“strong” BS, “weak” PPT)	This work	N/A
Plasmid: pcDNA3.1(+)-Ad2.15 (“weak” BS, “weak” PPT)	This work	N/A
Plasmid: pcDNA3.1(+)-Ad2.2 (“decoy” BS, “strong” PPT)	This work	N/A
Software and Algorithms		
CCP4	Winn et al., 2011	http://www.ccp4.ac.uk/
PHENIX suite	Adams et al., 2010	https://www.phenix-online.org/
XDS	Kabsch, 2010	http://xds.mpimf-heidelberg.mpg.de/
PyMOL	Schrödinger LCC	https://pymol.org/2/
UCSF Chimera	Pettersen et al., 2004	https://www.cgl.ucsf.edu/chimera/
Coot	Emsley et al., 2010	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
Grade Web Server	Global Phasing Limited	http://grade.globalphasing.org/cgi-bin/grade/server.cgi/
Arpeggio Web Server	Jubb et al., 2017	http://biosig.unimelb.edu.au/arpeggioweb/
PLIP Web Server	Salentin et al., 2015	https://projects.biotec.tu-dresden.de/plip-web/plip/
ALBULA	Dectris Ltd.	https://www.dectris.com/products/albula-software/
Prism v.6.0	GraphPad Software	https://www.graphpad.com/scientific-software/prism/

CONTACT FOR REAGENT AND RESOURCE SHARING

Requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Vladimir Pena (vlad.pena@mpibpc.mpg.de). Requests for cell lines and RNA substrates should be addressed to Silvia Buonamici (silvia_buonamici@h3biomedicine.com) and MTA will be required.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines

HCT116 and HeLa S3 cell lines were obtained from the American Type Culture Collection (ATCC). These cell lines were authenticated using ATCC and tested for mycoplasma contamination tested at IDEXX using IMPACT1 testing. HeLa S3 cells used for isolation of human spliceosomes were obtained from Helmoltz Center for Infection Research (Brunswick) and tested negative for mycoplasma contamination. HeLa S3 cells used for the isolation of spliceosomes were cultured in DMEM/F12 (1:1) medium supplemented with 5% NCS serum at a density of 6.5×10^6 cells/ml. Sf9 (*Spodoptera frugiperda*) and Hive Five (*Trichoplusia ni*) insect cell lines were purchased from Thermo Fisher Scientific and cultured in suspension in Sf-900 III SFM (Thermo Fisher Scientific) and ESF 921 medium (Expression Systems), respectively, at 27°C and 90 rpm. For cloning and bacmid production, we used the *Escherichia coli* BW23474 and DH10MultiBacY strains (gift from Dr. Imre Berger, EMBL, Grenoble), and the *Escherichia coli* XL-10 Gold strain (Agilent Technologies). MBP-MS2 fusion protein used for MS2 affinity selection of spliceosomes was expressed in the *Escherichia coli* strain Rosetta 2 (DE3) (Merck) in 2YT medium at 37°C.

METHOD DETAILS

Expression and purification of recombinant SF3B complexes

The engineered human SF3B core complex was overexpressed in insect cell lines using optimized synthetic genes (GeneArt, Life Technologies), as previously described (Cretu et al., 2016). Briefly, the SF3B3 subunit was cloned in frame with a Tobacco Etch Virus (TEV) protease cleavable N-terminal 10xHIS tag and with a C-terminal FLAG tag. The internal 1068-1085 residues of SF3B3 were removed by "round-the-horn" PCR mutagenesis (Key Resources Table). PHF5A (1-98) was amplified by PCR and fused to a PreScission protease (PP) cleavable GST tag. SF3B5 (full-length) and SF3B1 (453-1304) were cloned by PCR without addition of affinity tags. All four SF3B subunits were placed on a single acceptor vector by the sequential use of the multiplication and the Cre recombination cassettes, as implemented in the Multibac baculovirus/insect cell expression system (Fitzgerald et al., 2006). High Five insect cells (Thermo Fisher Scientific) overexpressing the human SF3B core complex were harvested 72h after infection with recombinant baculoviruses and lysed by sonication in ice-cold buffer A1 (50 mM HEPES-KOH, pH 7.9, 600 mM KCl, 15% (v/v) glycerol, 5 mM 2-mercaptoethanol (2-ME), 30 mM imidazole, pH 8.0) supplemented with a cocktail of protease inhibitors (Roche). The cleared lysate was loaded on a HisTrap HP column (GE Healthcare) equilibrated with buffer A2 (50 mM HEPES-KOH, pH 7.9, 600 mM KCl, 10% (v/v) glycerol, 5 mM 2-ME, 30 mM imidazole, pH 8.0). The bound proteins were eluted by a linear gradient formed between buffer A2 and buffer B (50 mM HEPES-KOH, pH 7.9, 600 mM KCl, 10% (v/v) glycerol, 5 mM 2-ME, 400 mM imidazole, pH 8.0) and incubated with ~20 mL glutathione resin (Glutathione Sepharose 4B, GE Healthcare) in batch for 2h at 4-8°C. The SF3B core complex was eluted from the resin with buffer C (50 mM HEPES-KOH, pH 7.9, 200 mM KCl, 10% (v/v) glycerol, 2 mM DTT, 30 mM L-glutathione reduced) and both the 10xHIS and the GST affinity tags were removed by overnight cleavage with TEV and PP (GE Healthcare) proteases, respectively. The sample was further purified on a HiTrap Q HP column (GE Healthcare) and eluted from the resin using a linear gradient (0%–30%, 50 mL) formed between buffer D (20 mM HEPES-KOH, pH 7.9, 200 mM KCl, 10% (v/v) glycerol, 2 mM DTT) and buffer E (20 mM HEPES-KOH, pH 7.9, 1 M KCl, 5% (v/v) glycerol, 2 mM DTT). SF3B peak fractions were concentrated using an Amicon Ultra-15 centrifugal filter (Millipore) and further purified by size-exclusion chromatography (HiLoad 16/600 Superdex 200 pg, GE Healthcare) in buffer SB (20 mM HEPES-KOH, pH 7.9, 200 mM KCl, 5% (v/v) glycerol, 2 mM DTT). The SF3B core fractions were subsequently concentrated and used fresh in crystallization screening or snap frozen in liquid nitrogen and stored at −80°C. The engineered SF3B core complex contained SF3B1, SF3B3, PHF5A, and SF3B5 in apparently stoichiometric amounts (Figure S1A) and displayed nanomolar affinity for pladienolide B (PB) as also observed with the full-length recombinant SF3B complex (Figure S1B). For the production of selenomethionine-derivatized SF3B core complex, Hive Five insect cells were grown for 48h in ESF 921Δ methionine deficient medium (Expression Systems) and then infected with recombinant baculoviruses. After 12h, the suspension cultures were supplemented with seleno-L-methionine (Sigma) to 0.1g/L and cultured for an additional 60h. R1074H (SF3B1), R1074A (SF3B1), and the Y36A (PHF5A) mutations in the SF3B1-PHF5A tunnel were introduced using a modified protocol based on the QuickChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies). Full-length heptameric human SF3B complexes (SF3B1-SF3B2-SF3B3-SF3B4-SF3B5-SF3B6-PHF5A), used in radioligand binding assays, were overexpressed and purified as previously described (Cretu et al., 2016). All SF3B preparations were analyzed by mass spectrometry (data not shown).

Crystallization and structure determination of the human SF3B core in complex with PB

The engineered human SF3B core complex (SF3B3 (Δ1068-1085)-SF3B1 (453-1304)-PHF5A (1-98)-SF3B5) was concentrated by ultrafiltration to ~6 mg/ml, complexed with PB (purchased from Santa Cruz Biotechnology), added in a 3-fold molar excess over protein, and subjected to initial crystallization screening. Optimized single trapezoidal plates were grown at 2.5-3 mg/mL at 20°C in 1+1μl hanging drops using reservoir solutions containing 50 mM HEPES-NaOH, pH 7-7.4, 200 mM KCl, and 38%–39% (v/v) pentaerythritol propoxylate (5/4 PO/OH). The best diffracting SF3B-PB co-crystals appeared after 1 day and reached full size after ~5 days. The SF3B core crystals were harvested directly from the mother liquor at 20°C and flashed cooled in liquid nitrogen. Selenium-derivatized crystals, obtained under similar crystallization conditions, were harvested from the mother liquor at 4°C. Diffraction data were collected at 100K on PILATUS 6M-F (25 Hz) at the X10SA beamline (SLS, PSI Villigen, Switzerland). Diffraction data collected from a single crystal were indexed, integrated and scaled with XDS (Kabsch, 2010), and merged with AIMLESS (Evans and Murshudov, 2013), in CCP4 (Winn et al., 2011). The SF3B-PB co-crystal structure was solved by molecular replacement with Phaser-MR (McCoy, 2007) using the previously reported crystal structure of the protease-resistant SF3B core (PDB: 5IFE) (Cretu et al., 2016) as a search model. Like the protease-resistant core complex (Cretu et al., 2016), the ~254 kDa engineered SF3B^{core}-PB complex crystallized in the $P2_12_12_1$ space group (Table 1).

The initial search model was iteratively rebuilt in Coot (Emsley and Cowtan, 2004), followed by refinement with phenix.refine (Adams et al., 2010). As independent means of validating the residue register of the SF3B core model, we collected SAD (single-wavelength anomalous dispersion) data at the Se K-edge (12.662 keV) from selenium-derivatized crystals (Table 1). Careful inspection of the anomalous difference map revealed strong anomalous peaks for 53 out of 54 modeled methionine side chains, in good agreement with our SF3b ^{core}-PB structural model (Figures S1F and S1G). We could not assign a Se marker site for SF3B1-M620, likely because of the relative flexibility of the residue (located in the H4 repeat turn). Inspection of the residual mF_o-DF_c map of the refined SF3B core model revealed a strong positive electron density at the interface between SF3B1 and PHF5A. Importantly, the observed positive density differed significantly from the previously reported density, located close to the site (Cretu et al., 2016). The PB ligand was manually fit into the mF_o-DF_c difference electron density map in Coot with the macrolide moiety of the splicing modulator facing the SF3B3 side (Figures S1C–S1E). In agreement with previous photo-affinity labeling experiments (Kotake et al., 2007), the SF3B3 CTD domain is located ∼17Å from the macrolide cycle, within the effective distance of the reactive probe. The location of the PB ligand was confirmed using both unbiased mF_o-DF_c and polder omit maps (Liebschner et al., 2017) (Figures S1C–S1E). The Grade Web server (Global Phasing Limited) was used to generate geometric restraints for PB refinement. The final model was refined using individual B factors for protein and ligand atoms to a R_{work} value of ∼23% and a R_{free} value of ∼26% and exhibited good stereochemistry (Table 1). The higher B factors of the macrolide lactone (C1) are consistent with its likely susceptibility to radiation damage. Moreover, this observation also agrees with the described tendency of pladienolides to undergo hydrolysis of the C1 bond in aqueous media (Villa et al., 2013). The final model comprises the HEAT domain of SF3B1 (residues 463-1304), SF3B3 (residues 1-1217, lacking the 1068-1085 loop region), SF3B5 (residues 15-79) and PHF5A (residues 6-98). The 453-462 N-terminal residues of SF3B1, as well as several internal SF3B3 loop regions (646-661; 692-694; 829-832), could not be traced in the electron density, likely, owing to relative flexibility/disorder. All structural figures were generated with PyMOL (versus 1.8.4.1, Schrödinger LLC). The contact sites between PB and the SF3B core were identified with Arpeggio (Jubb et al., 2017) and PLIP (Salentin et al., 2015), and analyzed in PyMOL. The SF3B1-PHF5A tunnel was computed using a modified HOLLOW script (Abe and Hiraki, 2009) and rendered in PyMOL. Structural alignment of herboxidiene (Edmunds et al., 2000) and PB (Figure 4) was carried out in UCSF Chimera (Pettersen et al., 2004).

Structural superposition of SF3B1 and Hsh155p

Comparison of the SF3B1 HEAT helix in the PB-bound state (open) with the yeast ortholog in the B_{act} spliceosome (closed) shows major conformational differences (rmsd of 5.2Å, 710 Cα) (Figure S4A). Analyses of individual pairs of HEAT repeats (Figure S4E) showed small differences at the interfaces between H2-H3, H7-H8, H11-H12, while major variations are present at the interfaces of H15-H16 (rmsd of 2.06, 66 Cα) and H16-H17 (rmsd of 1.22, 72 Cα). In contrast, the intramolecular interfaces of H15 and H17 with the neighboring repeats remained largely unchanged (H14-H15 – rmsd of 0.67 (71 Cα), H17-H18 – rmsd of 0.64 (69 Cα). As a result, the ∼23° rotation of the H15 repeat over the H16 hinge, toward the convex side of SF3B1, would induce a counterclockwise movement of the entire H1-H15 arch. Except for the flexible BPB domain of SF3B3, all other SF3B modules behave as rigid bodies with respect to the H17-H20 repeats and superimpose onto the yeast proteins without major clashes (Figures S4D and S4E).

The large-scale reconfiguration of the SF3B1 superhelix after rotation of the H15-H17 hinge would shorten the distance between the terminal HEAT repeats to a similar extent as observed when comparing the “open” and “closed” conformation (from ∼25Å to ∼18Å). However, smaller additional rearrangements at the H2-H3 and H7-H8 interfaces – likely, induced upon interaction with the BS and polypyrimidine regions of the intron and/or with additional splicing factors – might be required to achieve a state that is virtually identical to the “closed” state.

Changes in the configuration of the SF3B1 helix are also coupled with additional rearrangements that occur at the interfaces between SF3B1 and PHF5A and between SF3B1 and the BPA domain of SF3B3. At the second contact site, the C-terminal extension of PHF5A (aa 91-98), undergoes a forward movement with respect to its globular domain. As a result, it interacts with the H2 repeat of SF3B1 replacing previous contacts of the E48, Y51 and Y54 residues (E48-H550; E48-K554; Y51-Q547; Y54-Q547); these residues are now involved in contacts with the H5-H6 repeats (E48-C677, Y51-E714). Because in isolation the C terminus of PHF5A (Rds3p in yeast) is disordered, this region likely folds to an α helix, as seen in the yeast spliceosome, upon simultaneous binding to SF3B1 and to the minor groove of the U2/BS helix. Consistent with a rotation around the H16 hinge, the interfaces between the BPA domain of SF3B3 and PHF5A are not significantly changed.

Cell viability assays

For CellTiter-Glo analysis, 500 cells (HCT116 parental, HCT116 Y36C and HCT116 R1074H generated as described in Teng et al., 2017) were seeded in each well of a 384-well plate the day before compound addition. An 11 pt serial dilution was used starting with a top final dosage of 10 mM for 10 additional doses. DMSO percentage was maintained throughout and a DMSO-only control was included. Seventy-two hours post compound addition, CellTiter-Glo reagent was added to the medium, incubated and assayed on EnVision Multilabel Reader (PerkinElmer). The luminescence value from each treatment sample was normalized to the average value of the respective DMSO control. The dosage response curve plots were generated using Prism 6 (GraphPad) and fit using nonlinear regression analysis and the log (inhibitor) versus response-variable slope (four parameters).

HeLa nuclear extract preparation

HeLa S3 cell pellets were resuspended in hypotonic buffer (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.2 mM PMSF, and 0.5 mM DTT) and the suspension was brought up to a total of 5 packed cell volume (PCV). After centrifugation, the supernatant was discarded, and the cells were brought up to 3 PCV with hypotonic buffer and incubated on ice for 10 minutes. Cells were lysed using a Dounce homogenizer and then centrifuged. The supernatant was discarded, and the pellet was resuspended with ½ packed nuclear volume (PNV) of low salt buffer (20 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 20 mM KCl, 0.2 mM EDTA, 25% glycerol, 0.2 mM PMSF, 0.5 mM DTT), followed by ½ PNV of high salt buffer (same as low salt buffer except 1.4M KCl was used). The nuclei were gently mixed for 30 minutes before centrifuging. The supernatant (nuclear extract) was then dialyzed into storage buffer (20 mM HEPES pH 7.9, 100 mM KCl, 0.2 mM EDTA, 20% glycerol, 0.2 mM PMSF, 0.5 mM DTT). Protein concentration was determined using NanoDrop 8000 UV-Vis spectrophotometer (Thermo Fisher Scientific).

In vitro splicing assays

All Ad2-derived (Pellizzoni et al., 1998) sequences were cloned into the pcDNA3.1(+) vector (Promega) using 5′ EcoRI and 3′ XbaI restriction sites. The plasmids were linearized using XbaI and used as DNA templates in the *in vitro* transcription reactions. The FtzΔi intron-less plasmid (Luo and Reed, 1999) was linearized using EcoRI. All RNAs were *in vitro* transcribed and then purified using MEGAscript T7 (Invitrogen) and MEGAclear (Invitrogen) kits, respectively. For splicing reactions using Ad2 variant pre-mRNAs, 1 μL reactions were prepared using 8 μg nuclear extracts prepared from HeLa S3, 2 ng pre-mRNA, 0.2 ng FTZΔi, and varying concentrations of PB, PB-OH or DMSO. After a 15-minute pre-incubation at 30°C, 1 μL splicing activation buffer (0.5 mM ATP, 20 mM creatine phosphate, 1.6 mM $MgCl_2$) was added, and the reactions were incubated for 90 minutes at 30°C. The reactions were then quenched with 13 μL DMSO, and 25 nL was used for RT-qPCR. RT-qPCR reactions were prepared using TaqMan RNA-to-C_T 1-step kit (Life Technologies), RNA from splicing reactions, Ad2 (forward: ACTCTCTTCCGCATCGCTGT; reverse: CCGACGGGTTTCC GATCCAA; probe: CTGTTGGGCTCGCGGTTG) and Ftz (forward: TGGCATCAGATTGCAAAGAC; reverse: ACGCCGGGTGATG TATCTAT; probe: CGAAACGCACCCGTCAGACG) mRNA primer-probe sets. Prism 6 (GraphPad) was used for non-linear regression curve fitting of the formed spliced product and normalized to the control (DMSO) sample.

Scintillation proximity assays/competitive binding assay

Batch immobilization of anti-FLAG antibody (Sigma-Aldrich) to anti-mouse PVT SPA scintillation beads (PerkinElmer) was prepared as follows. For every 1.5mg of beads, 10μg antibody was prepared in 150μl PBS. The antibody–bead mixture was incubated for 30min at RT and centrifuged at 18,000*g* for 5min. 150μl PBS was used to resuspend every 1.5mg antibody–bead mixture. The purified SF3B full-length complexes were tested for [^{3}H]-labeled pladienolide probe binding ([^{3}H]-PB), synthesized as previously described (Kotake et al., 2007). 100μl binding reactions were prepared with 50μl bead slurry and 0 or 10nM protein in buffer (20mM HEPES pH 8, 200mM KCl, 5% glycerol). The mixture was incubated for 30min, and 1 nM or 10 nM [^{3}H]-probe was added. The mixture was incubated for 30min, and luminescence signals were read using a MicroBeta2 Plate Counter (PerkinElmer). Compound competition studies were performed with the full-length heptameric SF3B and the SF3B core complex immobilized using anti-SF3B3 antibody. 100μl binding reactions were prepared with 50μl bead slurry, 12.5nM protein in buffer and compounds PB and PB-OH at varying concentrations. After a 30-min pre-incubation, 1 nM [^{3}H]-PB was added. The reactions were incubated for 30min, and luminescence signals were read. SF3B complexes from HeLa nuclear extracts were immunoprecipitated using anti-SF3B1 antibody. Batch immobilization of anti-SF3B1 antibody (MBL) to anti-mouse PVT SPA scintillation beads (PerkinElmer) was prepared as follows: for every 2.5mg of nuclear extracts, 5μg anti-SF3B1 antibody and 1.5mg of beads were mixed in 150μl PBS. The antibody–bead mixture was incubated for 30min at RT and centrifuged at 18,000*g* for 5min. The beads were suspended and added to the prepared nuclear extracts. The slurry was incubated for 2h at 4°C with gentle mixing. The beads were then collected by centrifuging at 18,000*g* for 5min and washed twice with PBS+0.1% Triton X-100. After a final centrifugation step, every 1.5mg of beads was suspended with 150μl of PBS. 100 μL binding reactions were prepared by adding varying concentrations of PB or PB-OH, and after 30min pre-incubation, 2.5nM [^{3}H]-probe was added. The mixture was incubated for 30min, and luminescence signals were read. Prism 6 (GraphPad) was used for non-linear regression curve fitting of the data.

Affinity purification of B complexes and splicing chase experiments

In vitro assembled spliceosomal B complexes were purified by gradient centrifugation, followed by MS2 affinity-selection using amylose beads (New England Biolabs) as previously described (Bessonov et al., 2008). Briefly, ^{32}P-labeled MINX containing MS2 aptamers at its 3′ was incubated with a 20-fold molar excess of MS2-MBP protein. Standard splicing reactions containing 10 nM pre-mRNA were incubated at 30°C for 6 min and chilled on ice. Splicing complexes were subsequently separated on a linear 10%–30% (v/v) glycerol gradient containing G-150 buffer (20 mM HEPES-KOH, pH 7.9, 150 mM KCl, 1.5 mM $MgCl_2$) by centrifugation for 16 h at 23000 rpm in a Sorvall Surespin 630 rotor. Complexes in the 45S peak fractions were loaded onto amylose beads (NEB), and after washing with G-75 buffer (G buffer with 75 mM KCl), bound spliceosomal complexes were eluted with 20 mM maltose in G-75 buffer. RNA was recovered from the purified complexes, separated on a denaturing polyacrylamide gel and visualized by silver staining (Figure S5A). For *in vitro* chase experiments, HeLa nuclear extract was treated with micrococcal nuclease (MN) as described previously (Makarov et al., 2002). Affinity-purified B complexes formed on ^{32}P-labeled MINX-MS2 pre-mRNA were incubated with splicing buffer alone (60 mM KCl, 3 mM $MgCl_2$, 2 mM ATP, 20 mM creatine phosphate, 20 mM HEPES-KOH, pH 7.9), or in the presence of 20% (v/v) MN-treated HeLa nuclear extract containing 0 or 6 μM PB or purified B complexes were pre-incubated with 1.5 or 6 μM PB for 30 min on ice prior to addition of MN-treated nuclear extract. A 10-fold molar excess of unlabelled MINX-MS2 pre-mRNA was added to prevent the reassembly of snRNPs (that potentially dissociate from the purified complexes) on the radiolabeled pre-mRNA. The reaction was incubated at 30°C for 0, 30 or 60 min. RNA was recovered, separated on a 14% denaturing polyacrylamide gel, and visualized with a Typhoon phosphoimager (GE Healthcare) (Figure S6B). The % mRNA formed was quantified using ImageQuantTL (GE Healthcare) and calculated by dividing the amount of mRNA by the amount of pre-mRNA, lariat-intermediate and splicing products (Figure S5B).

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were carried out with Prism 6 (GraphPad). For the biochemical (binding and *in vitro* splicing assay) and viability assays each concentration was tested in triplicate and each experiment was repeated multiple times (2 times for the biochemical assays and 3 times for the viability assay). The compound concentrations used for the assay window was estimated such that the IC_{50} was near the center of the fitted curve. All the experiments were repeated multiple times (2 times for the biochemical assays and 3 times for the viability assays). No data were excluded from the analysis. No randomization was required. Clear description of statistics (central tendency, variation, exact sample size) is included in figure legends.

DATA AND SOFTWARE AVAILABILITY

The accession number for the coordinates reported in this paper is PDB: 6EN4.

Molecular Cell, Volume *70*

Supplemental Information

Structural Basis of Splicing Modulation

by Antitumor Macrolide Compounds

Constantin Cretu, Anant A. Agrawal, Andrew Cook, Cindy L. Will, Peter Fekkes, Peter G. Smith, Reinhard Lührmann, Nicholas Larsen, Silvia Buonamici, and Vladimir Pena

Supplemental Information

Supplemental Figures

Figure S1

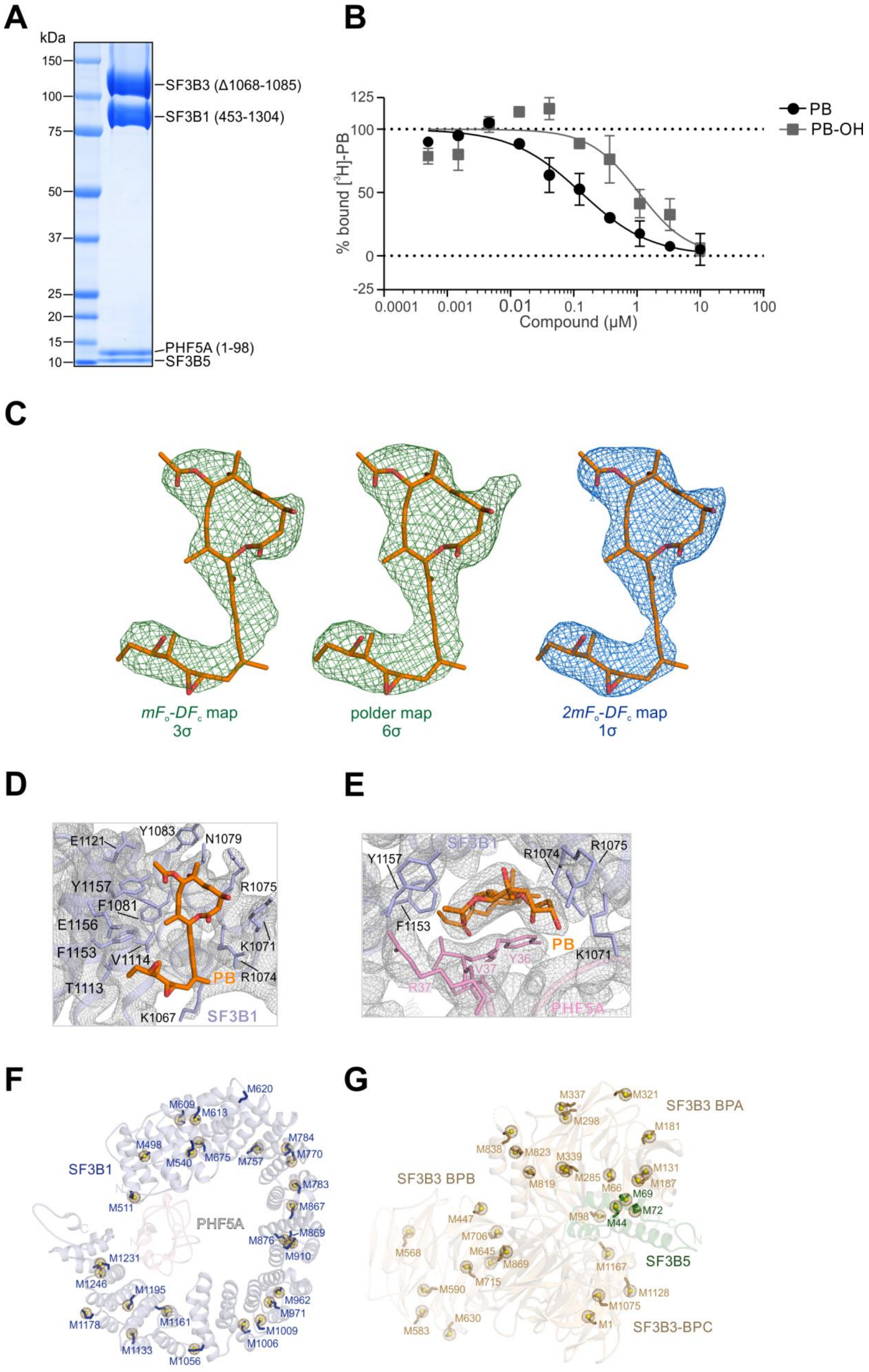


Figure S2

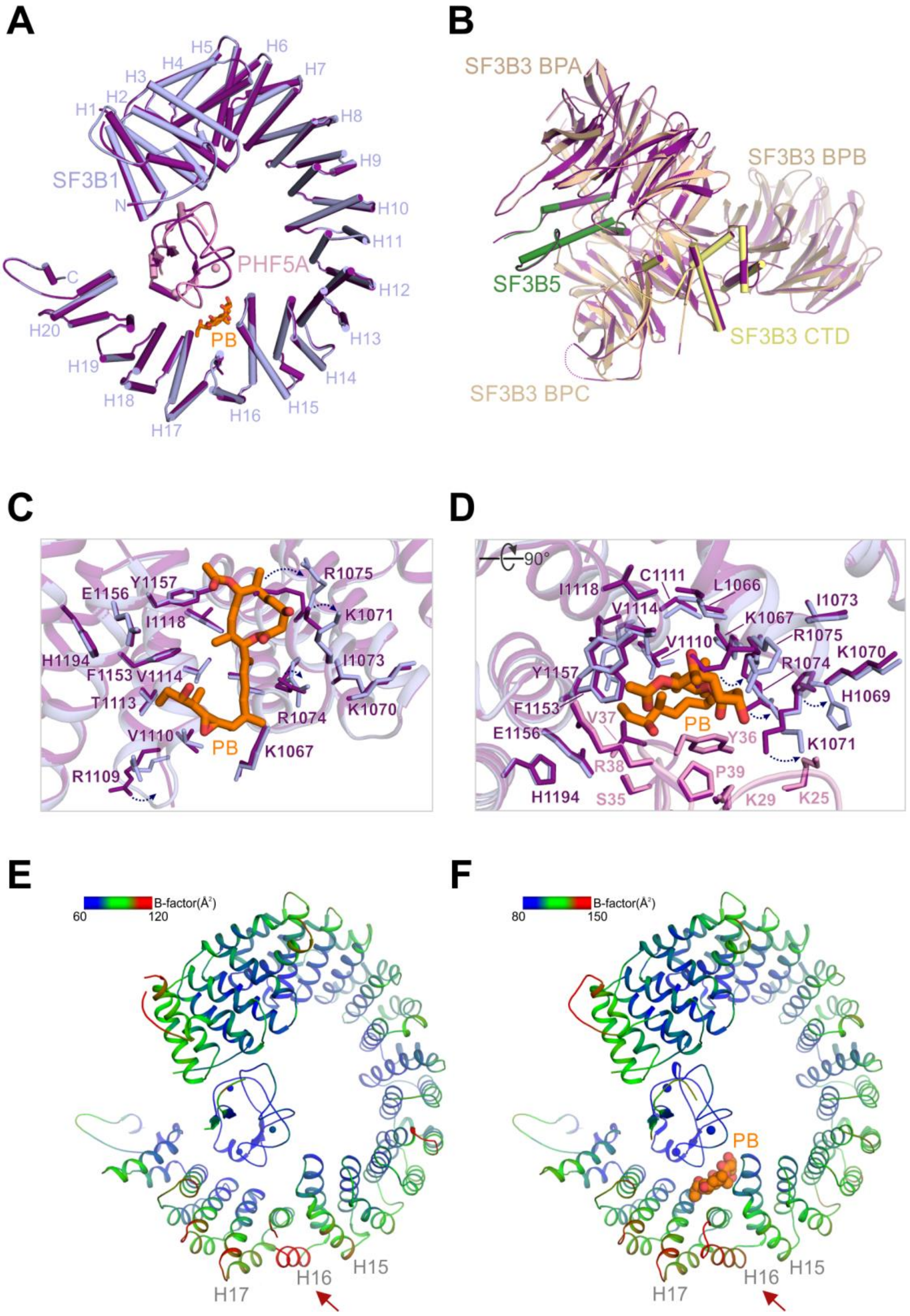

A
H5
H6
H4
H3
H7
H2
H1
H8
H9
SF3B1
N
H10
H11
PHF5A
C
H12
H20
PB
H13
H19
H14
H18
H15
H17
H16
B
SF3B3 BPA
SF3B3 BPB
SF3B5
SF3B3 CTD
SF3B3 BPC
C
R1075
E1156
Y1157
K1071
I1118
H1194
I1073
F1153
V1114
T1113
K1070
R1074
V1110
PB
K1067
R1109
D
90°
I1118
C1111
L1066
I1073
V1114
K1067
R1075
V1110
K1070
R1074
Y1157
H1069
F1153
V37
PB
E1156
Y36
K1071
R38
P39
H1194
S35
K29
K25
E
B-factor(Å²)
60
120
H15
H16
H17
F
B-factor(Å²)
80
150
PB
H15
H16
H17

Figure S3

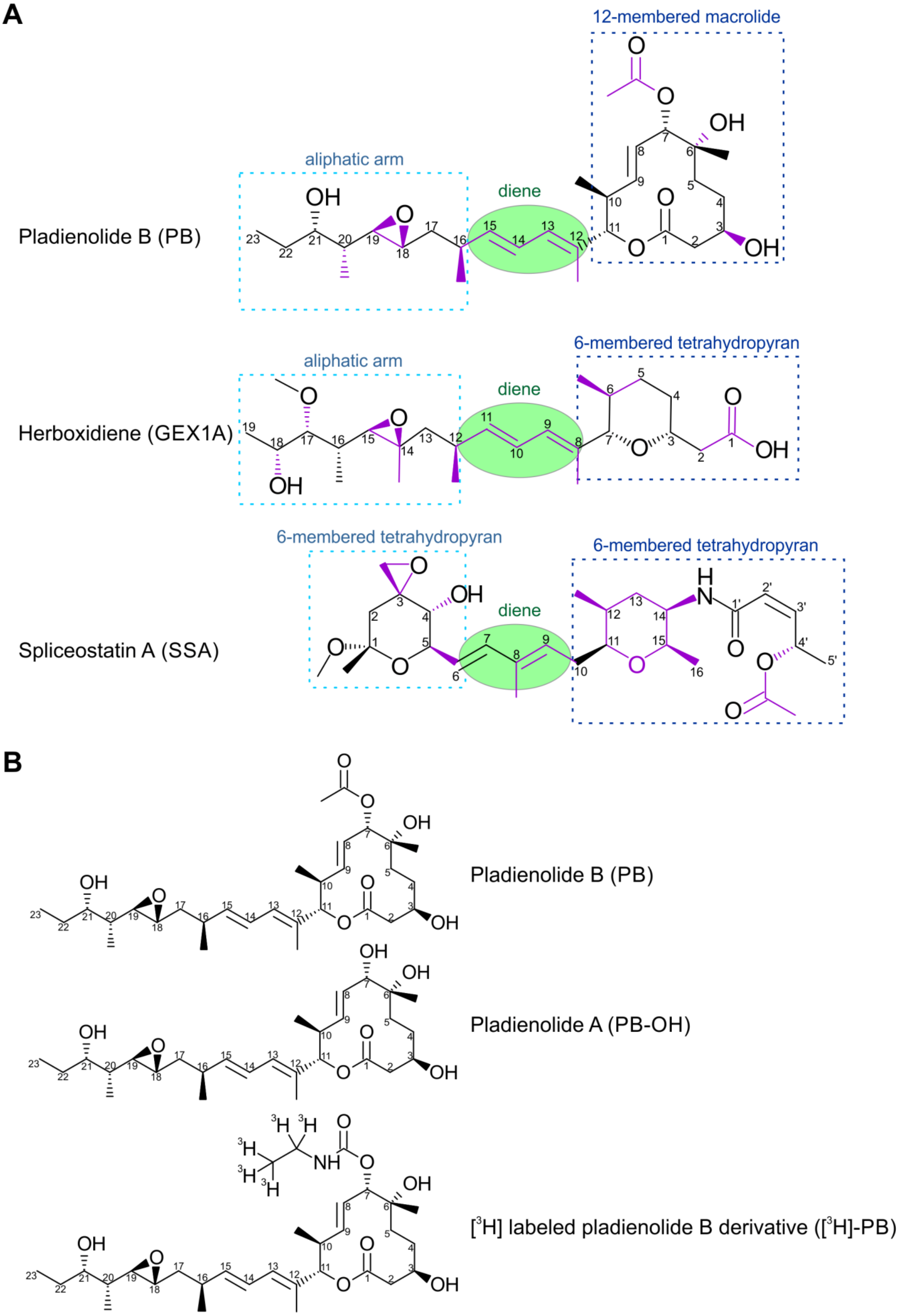

A
Pladienolide B (PB)
aliphatic arm
diene
12-membered macrolide
Herboxidiene (GEX1A)
aliphatic arm
diene
6-membered tetrahydropyran
Spliceostatin A (SSA)
6-membered tetrahydropyran
diene
6-membered tetrahydropyran
B
Pladienolide B (PB)
Pladienolide A (PB-OH)
[3H] labeled pladienolide B derivative ([3H]-PB)

Figure S4

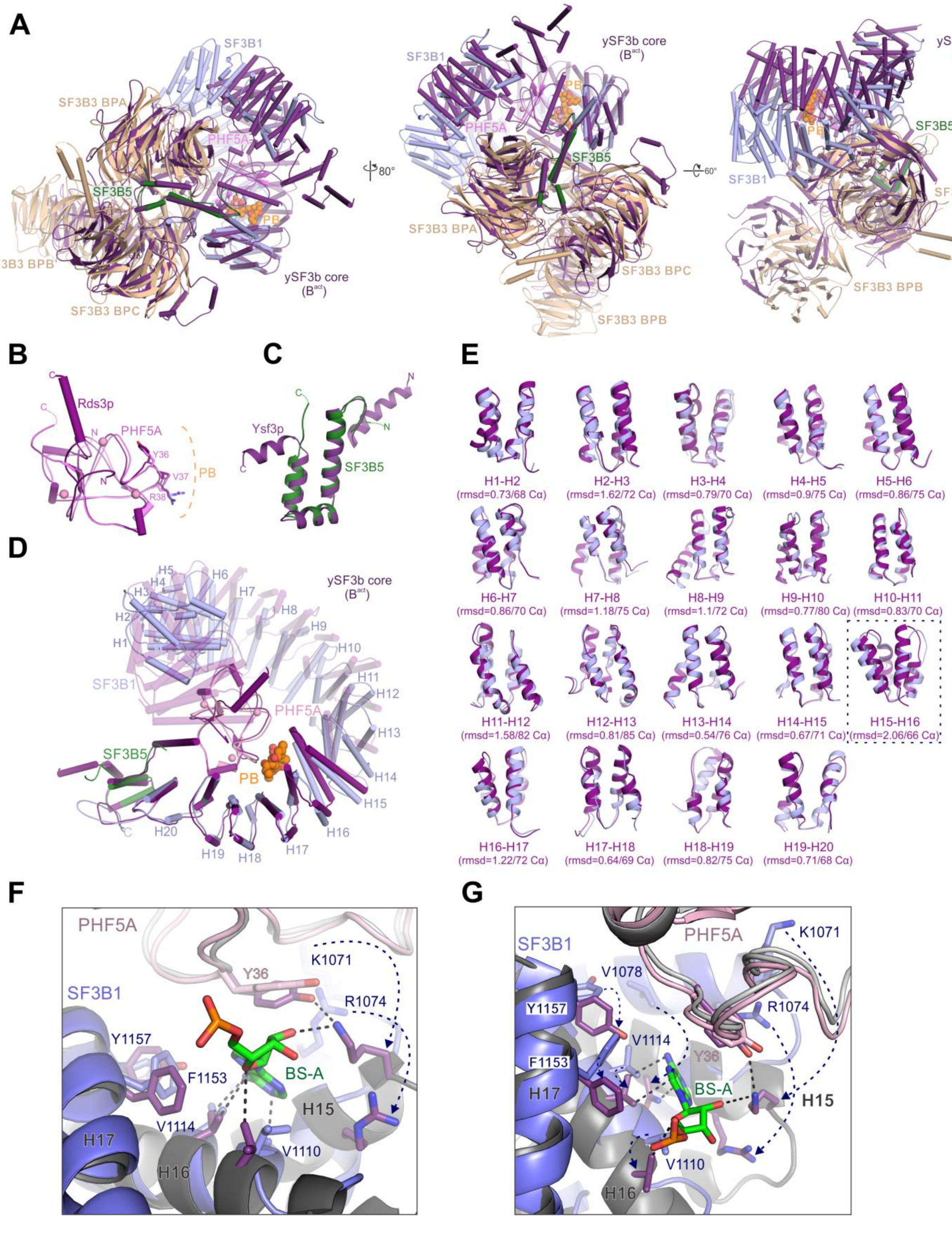

A
SF3B1
SF3B3 BPA
PHF5A
SF3B5
PB
SF3B3 BPB
SF3B3 BPC
ySF3b core
(B^{act})
80°
SF3B1
ySF3b core
(B^{act})
PB
PHF5A
SF3B5
SF3B3 BPA
SF3B3 BPC
SF3B3 BPB
60°
ySF3b core
(B^{act})
PB
SF3B5
SF3B1
SF3B3 BPA
SF3B3 BPB
B
C
Rds3p
C
N
PHF5A
Y36
V37
R38
N
PB
C
Ysf3p
C
SF3B5
N
N
D
H5
H6
H4
H7
H3
H8
H2
H9
H1
H10
ySF3b core
(B^{act})
SF3B1
H11
H12
PHF5A
H13
SF3B5
PB
H14
H15
H16
C
H20
H19
H18
H17
E
H1-H2
(rmsd=0.73/68 Cα)
H2-H3
(rmsd=1.62/72 Cα)
H3-H4
(rmsd=0.79/70 Cα)
H4-H5
(rmsd=0.9/75 Cα)
H5-H6
(rmsd=0.86/75 Cα)
H6-H7
(rmsd=0.86/70 Cα)
H7-H8
(rmsd=1.18/75 Cα)
H8-H9
(rmsd=1.1/72 Cα)
H9-H10
(rmsd=0.77/80 Cα)
H10-H11
(rmsd=0.83/70 Cα)
H11-H12
(rmsd=1.58/82 Cα)
H12-H13
(rmsd=0.81/85 Cα)
H13-H14
(rmsd=0.54/76 Cα)
H14-H15
(rmsd=0.67/71 Cα)
H15-H16
(rmsd=2.06/66 Cα)
H16-H17
(rmsd=1.22/72 Cα)
H17-H18
(rmsd=0.64/69 Cα)
H18-H19
(rmsd=0.82/75 Cα)
H19-H20
(rmsd=0.71/68 Cα)
F
PHF5A
Y36
K1071
SF3B1
R1074
Y1157
F1153
BS-A
H15
V1114
H17
H16
V1110
G
SF3B1
PHF5A
K1071
V1078
Y1157
R1074
V1114
F1153
Y36
BS-A
H17
H15
V1110
H16

Figure S5

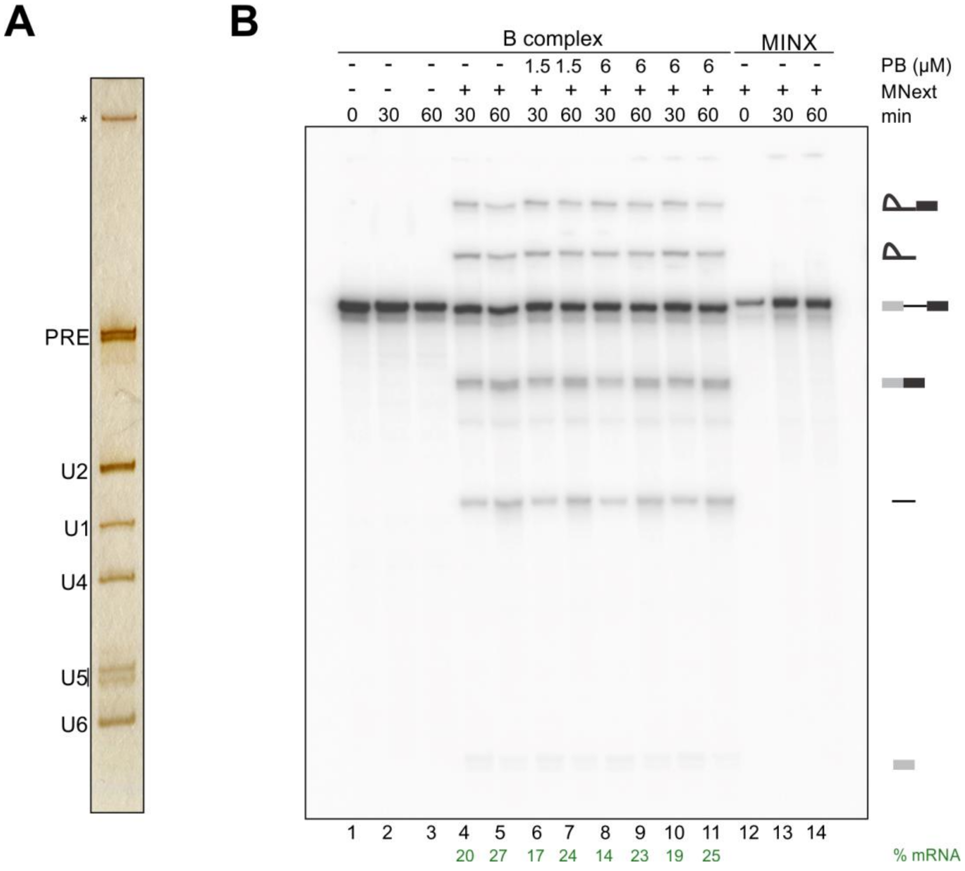

A
*
PRE
U2
U1
U4
U5
U6
B
B complex
MINX
- - - - - 1.5 1.5 6 6 6 6 - - -
PB (µM)
- - - + + + + + + + + + + +
MNext
0 30 60 30 60 30 60 30 60 30 60 0 30 60
min
1 2 3 4 5 6 7 8 9 10 11 12 13 14
20 27 17 24 14 23 19 25
% mRNA

Figure S6

A

Ad2-derived splicing substrates	Sequence (3' splice site)	PB IC$_{50}$ (nM)	PB-OH IC$_{50}$ (nM)	Fold change
Enhanced BS (Ad2.1)	-32 | -14 | -1 | TCATACT**AA**TCCTGTCCCTTTTTTTTCCACAG	19	-	>100
Wild-type (Ad2)	-32 | -14 | -1 | TCATACTT**A**TCCTGTCCCTTTTTTTTCCACAG	18	1589	88.3
Strong BS/weak PPT (Ad2.12)	-32 | -14 | -1 | TCATACTT**A**TCCTGTCCC**CCCCCCCC**CCACAG	19	1235	65
Weak BS/strong PPT (Ad2.11)	-32 | -14 | -1 | TCATA**G**TT**A**TCCTGTCCCTTTTTTTTCCACAG	22	224	10.2
Weak BS/weak PPT (Ad2.15)	-32 | -14 | -1 | TCATA**G**TT**A**TCCTGTCCC**CCCCCCCC**CCACAG	18	158	8.8
Decoy BS/strong PPT (Ad2.2)	-32 | -14 | -1 | T**AC**TACTT**A**TCCTGTCCCTTTTTTTTCCACAG	29	54	1.9

B

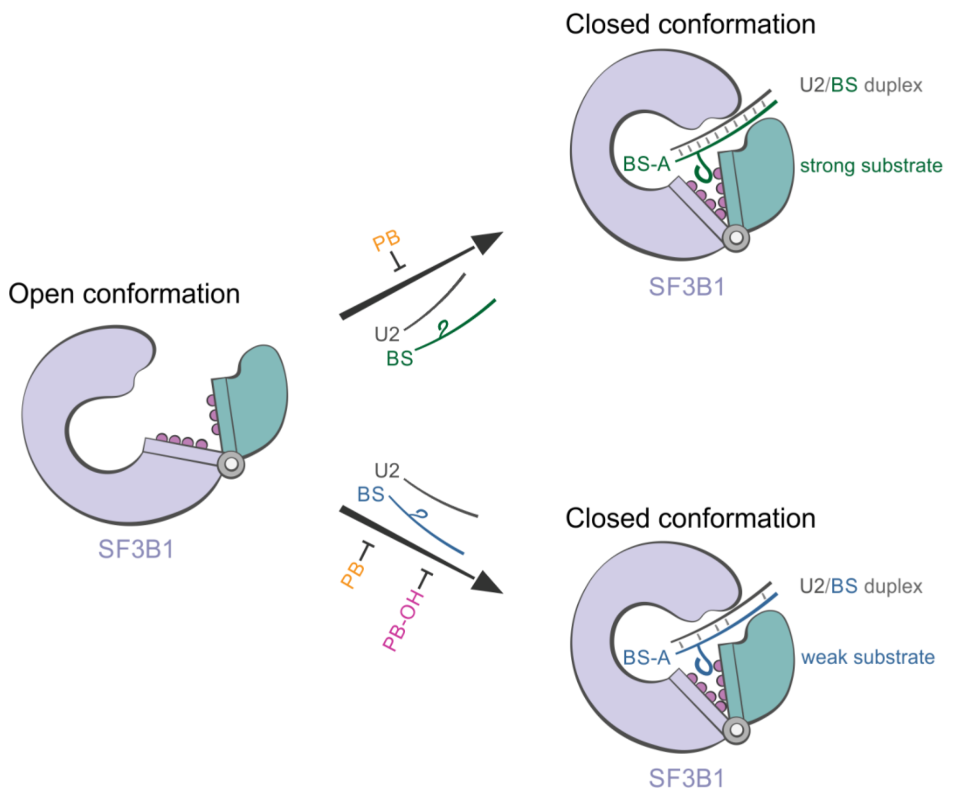


Figure S1. Crystal structure of an engineered human SF3B core in complex with pladienolide B (PB). Related to Figure 1.

(**A**) Coomassie-stained SDS-PAGE analysis of an engineered human SF3B core complex, reconstituted by co-expression in insect cells, after size-exclusion chromatography. SF3B3 (lacking residues 1068-1085), SF3B1 (residues 453-1304), PHF5A (residues 1-98), and SF3B5 (full-length) appear to be present in stoichiometric amounts. (**B**) Scintillation proximity assays of the human SF3B core in the presence of a tritiated PB derivative ([^{3}H]-PB) and unlabelled PB or PB-OH as competitors. The engineered SF3B core binds [^{3}H]-PB suggesting that the minimal complex contains an intact PB binding site (IC_{50} = 130 nM for PB and IC_{50} = 1070 nM for PB-OH). (**C**) Electron density maps of the PB ligand calculated from the refined structure. Residual mF_o-DF_c (contoured at 3σ) and polder (contoured at 6σ) omit maps are displayed next to a $2mF_o$-DF_c map (contoured at 1σ). The PB ligand is depicted as orange sticks. (**D**) Electron density map of SF3B1 at the PB binding site. The $2mF_o$-DF_c map (grey mesh, 1σ) is displayed around SF3B1 (light blue) and the PB ligand is shown as orange sticks (see also Figure 1B). (**E**), Electron density map of the SF3B1-PHF5A PB binding tunnel. The $2mF_o$-DF_c map (grey mesh, 1σ) is shown around SF3B1 (light blue), PHF5A (pink), and PB (orange). Key residues lining the SF3B1-PHF5A tunnel are depicted as sticks. (**F**), (**G**) Selenium marker sites validate the residue register of the SF3B core structure. Selenium sites were identified using SAD log-likelihood gradient maps and are shown as yellow spheres. The modeled methionine side chains are shown as sticks. SF3B1, SF3B3, SF3B5, and PHF5A are colored as in Figure 1.

Figure S2. Structural rearrangements of the SF3B core complex in the presence of pladienolide B (PB). Related to Figure 2.

(**A**) Superposition of the SF3B1 and PHF5A SF3B subunits in the presence and absence of the PB ligand. SF3B1 and PHF5A from the PB co-crystal structure are coloured light blue and pink, respectively, while the subunits from the apo SF3B core structure (PDB 5IFE) are coloured dark purple. The individual HEAT repeats of SF3B1 are labelled H1-H20. PB is depicted as sticks and is coloured orange. PHF5A's Zn atoms are indicated as spheres. (**B**) Superposition of the SF3B3-SF3B5 module of the SF3B core in the presence and absence of PB. SF3B3 and SF3B5 subunits are coloured as in Figure 1, while the SF3B3-SF3B5 module from the apo structure (PDB 5IFE) is in dark purple. The engineered SF3B3 has a shorter insertion loop in the 6th blade of the BPC β-propeller domain as a result of the removal of residues 1068-1085. (**C**), (**D**) Comparison of SF3B1 and PHF5A crystal structures at the PB binding site. The SF3B1 HEAT domain and PHF5A from the apo SF3B core structure (PDB 5IFE) are depicted as cartoons and are coloured in dark purple. The PB ligand is represented as orange sticks. SF3B1 and PHF5A subunits from the PB co-crystal structure are coloured light blue and pink, respectively. Key SF3B1 and PHF5A residues lining the PB binding tunnel are shown as sticks. SF3B1-K1071, SF3B1-R1074, and SF3B1-R1075 residues from the SF3B1-PHF5A tunnel are rearranged to accommodate the ligand (arrows), while PHF5A-Y36A and PHF5A-R38 residues adopt similar positions in both crystal structures. (**E**) Cartoon representation of SF3B1 and PHF5A subunits from the apo SF3B core crystal structure (PDB 5IFE) coloured according to the local B-factors (blue-green-red). Note that HEAT repeat H16 shows higher B factors compared to the neighbouring HEAT repeats of SF3B1. (**F**) Cartoon representation of SF3B1 and PHF5A subunits from the SF3B-PB co-crystal structure coloured according to their local B-factors (blue-green-red). The PB ligand (orange spheres) binds SF3B1 in the proximity of the H16 repeat, apparently stabilizing this more dynamic region of SF3B1.

Figure S3. Chemical structures of SF3B modulators. Related to Figures 3 and 4.

(**A**) Representative compounds belonging to the pladienolide (PB), herboxidiene (GEX1A), and spliceostatin (SSA) families of splicing modulators are shown. All compounds share a common diene group (light green) which connects two larger moieties, specific for each family of modulators. Herboxidienes have in common with pladienolides a similar aliphatic arm and with spliceostatins a 6-membered functionalized pyran ring. Chemical groups required for the activity of the compounds are highlighted in purple. (**B**) Structures of PB derivatives used in this study.

Figure S4. Structural comparison of the human SF3B-PB co-crystal structure with the yeast SF3B core. Related to Figure 5.

(**A**) Three different views of the human SF3B-PB structure superimposed onto the yeast SF3B core in the activated spliceosome. Human SF3B subunits are coloured and displayed as in Figure 1A. The yeast SF3B core (containing the Hsh155p HEAT domain, Rse1p, Rds3p, Ysf3p) is coloured dark purple. PB is depicted as orange spheres. Note that SF3B3, with the exception of its flexible BPB domain, SF3B1 H16-H20, PHF5A, and SF3B5 superimpose onto the yeast proteins without major clashes. The H1-H15 arch of SF3B1 behaves largely as a rigid body with respect to the SF3B1 (H16-H20)-PHF5A-SF3B3 (BPA-BPC)-SF3B5 module and is arranged differently in the two structures. (**B**) Superposition of PHF5A onto Rds3p as part of the SF3B1 (H16-H20)-PHF5A-SF3B3 (BPA-BPC)-SF3B5 module. Note that PHF5A residues located at the PB binding tunnel (Y36, V37, R38) face the same side, but adopt different conformations in the two structures. (**C**) Superposition of SF3B5 onto Ysf3p as part of the SF3B1 (H16-H20)-PHF5A-SF3B3 (BPA-BPC)-SF3B5 module. (**D**) Superposition of SF3B1 (H16-H20), SF3B5, and PHF5A onto the yeast SF3B core (dark purple) as part of the SF3B1 (H16-H20)-PHF5A-SF3B3 (BPA-BPC)-SF3B5 module. Note that the major difference between the two structures results from the different arrangement of the SF3B1 H1-H15 arch with respect to the SF3B1 (H16-H20)-PHF5A-SF3B3 (BPA-BPC)-SF3B5 module. SF3B3 is not shown. (**E**) Comparison of SF3B1's intramolecular interfaces between consecutive pairs of HEAT repeats in the SF3B-PB and yeast SF3B core structures. The largest difference between the two structures was detected at the H15-H16 interface (r.m.s.d. = 2.06, 66 Cα). Corroborated by the higher B-factor values (Figure S2F), this reveals the location of the hinge region. (**F**), (**G**) Structural superpositions of the H15-H17 hinge regions of SF3B1 (blue) and Hsh155p (dark grey), shown in two different orientations (left and right). PB is omitted for clarity sake and the BS-A is shown as sticks (green). PHF5A and the *S. cerevisiae* orthologue are coloured pink and light grey, respectively. SF3B1 residues that belong to the PB-binding site and the equivalent residues from Hsh155p are depicted in blue and dark purple, respectively.

Figure S5. Addition of PB to purified spliceosomal B complexes has no substantial effect on their ability to catalyse pre-mRNA splicing in the presence of MN-treated extract. Related to Figure 5.

(**A**) RNA composition of the affinity-purified B complexes. RNA was analyzed by denaturing PAGE and visualized by staining with silver. B complexes formed on ^{32}P-labelled MINX-MS2 pre-mRNA were affinity purified as described in the Online Methods. Asterisk, RNA in the loading well of the gel. The presence of the U1, U2, U4, U5, and U6 snRNAs confirms that spliceosomal B complexes were isolated. (**B**) Purified B complexes were incubated at 30°C for the indicated times (0, 30, 60 min) under splicing conditions in the presence of buffer alone (lanes 1-3) or micrococcal nuclease-treated HeLa nuclear extract (MNext) (lanes 4-14). PB (1.5 or 6 µM) was added directly to the purified B complexes, followed by a 30 min incubation on ice, prior to performing the chase with MN-treated extract (lanes 6-9) or PB (6 µM) was added to the MN-treated extract (lanes 10-11). As a control for complete MN-digestion, no splicing was observed when ^{32}P-labelled MINX-MS2 pre-mRNA was incubated with MNext (lanes 12-14). RNA was analysed by denaturing PAGE and visualized with a Phosphorimager. The positions of the pre-mRNA, splicing intermediates and products are indicated on the right. The % mRNA formed (quantitated with a Phosphorimager) is indicated below selected lanes. Addition of PB had no substantial effect on the ability of B complexes to catalyse splicing after chasing with MNext. This is consistent with the idea that once SF3B has adopted a closed conformation, as observed in the B complex, PB no longer can bind and inhibit splicing.

Figure S6. Splicing modulators may act as competitive BS antagonists. Related to Figure 6.

(**A**) Summary of the substrates and *in vitro* splicing assays carried out in the presence of PB or PB-OH. The IC_{50} values were obtained from non-linear regression curve fitting.

(**B**) The discriminatory action of splicing modulators towards various 3'SS sequences is a consequence of the competition between substrates and modulators for the “open” conformation of SF3B1 (see also Figure 5E). Potent modulators, such as PB, exhibit high affinity for the “open” conformation of SF3B1, inhibiting the transition to the “closed” conformation even for “strong” substrates. Conversely, modulators that display lower affinities for the “open” state of SF3B1, such as PB-OH, can inhibit efficiently only “weak” substrates. PB is colored in orange, while PB-OH is colored in pink. All the other elements are depicted and labeled as in the Figure 5E.

4 Discussion and Perspectives

By combining X-ray crystallography with structural mass spectrometry and biochemistry, the work presented here revealed: (i) the overall molecular architecture of the human SF3B core complex (SF3B1$^{\text{HEAT}}$-PHF5A-SF3B3-SF3B5, PDB 5IFE), one of the largest building blocks of the human spliceosome determined by X-ray crystallography (Cretu et al., 2016); (ii) the structural-consequences of SF3B1's cancer-related mutations (Cretu et al., 2016); (iii) the structural basis for the modulatory effects of splicing inhibitors from the pladienolide and herboxidiene families (Cretu et al., 2018; *unpublished data*). Also, our work provides further molecular insights into the early events of the splicing cycle and the functional dynamics of SF3B1 and p14. From a methodology perspective, we have established a pipeline for the routine determination of SF3B structures in complex with splicing modulators and, potentially, other compounds.

4.1 Molecular architecture and structural dynamics of the human SF3B complex

To facilitate crystallization, we have delineated a ~254 kDa core complex composed of the HEAT domain of SF3B1 (SF3B1$^{\text{HEAT}}$), SF3B3, SF3B5, and PHF5A and a second module composed of SF3B2 and SF3B4. We determined the structure of an initial core complex using experimentally estimated phases at a resolution of ~3.1 Å (Cretu et al., 2016). More recently, we obtained crystals of an optimized core complex (SF3B$^{\Delta\text{BPB}}$) diffracting X-rays to a resolution of ~2.8 Å (Table S1). Kiyoshi Nagai's lab has recently solved the crystal structure of the yeast SF3B2-SF3B4 core module to a resolution of ~2.7 Å (van Roon et al., 2017). Together with the previously reported crystal structure of SF3B6/p14 (Schellenberg et al., 2006) and with the recent cryo-EM structures of human spliceosomes (Bertram et al., 2017a; Haselbach et al., 2018; Zhang et al., 2018), this provides a near-complete view of human SF3B in structural terms (**Figure 4.1**).

4.1.1 The extended SF3B1's NTD domain is a protein-protein interaction hub

SF3B1 comprises an N-terminal extended region (NTD) and a conserved C-terminal HEAT domain (**Figure 2.9**). Owing to its dynamic nature, we could not capture the entire SF3B1^{NTD} in a crystal structure, and our structural mass spectrometry data indicate that the N-terminal region of SF3B1 does indeed sample an ensemble of conformations in the full-length SF3B complex (Cretu et al., 2016). However, our crosslinking experiments suggest that, in the isolated SF3B and the purified 17S U2 snRNP particle (Cindy L. Will, *personal communication*), the p14-interacting region (pir) of SF3B1^{NTD} adopts a more defined, central location with respect to SF3B1^{HEAT} (Cretu et al., 2016).

Recent cryo-EM snapshots of the human B^{act} complex (Haselbach et al., 2018; Zhang et al., 2018) revealed several additional folded regions in SF3B1^{NTD}. In the human B^{act} spliceosomes, residues 75-129 of SF3B1^{NTD} form a helix-loop-helix domain, as in yeast Hsh155p, and interact with the RT and Linker domains of Prp8, and with the RES complex (Haselbach et al., 2018; Zhang et al., 2018). In contrast to the isolated SF3B complex (Cretu et al., 2016), the pir region is positioned in the B^{act} complex more on the outer rim of SF3B1^{HEAT} and further extends towards Prp8's RT domain and SKIP (Zhang et al., 2018). Thus, during spliceosome assembly, SF3B1^{NTD}, likely, restructures upon binding to a newly available set of interacting partners.

Altogether, the available structural data suggest that the elongated topology of SF3B1^{NTD} is intimately linked to the functions of SF3B in splicing. By having an extended conformation, SF3B1^{NTD} can access concomitantly multiple interactors and thus, in turn, favor a more complex and combinatorial regulation of splicing. Hyperphosphorylation of SF3B1^{NTD} provides an additional layer of regulation, potentially coupling pre-mRNA splicing to the chromatin landscape and Pol II transcription (Girard et al., 2012; Kfir et al., 2015; Hollander et al., 2016). Overall, this is reminiscent of the highly repetitive C-terminal domain (CTD) of Pol II. Similar to SF3B1^{NTD}, Pol II CTD has low sequence complexity, adopts an extended conformation, is posttranslationally modified, and serves as a docking platform for a large number of RNA processing factors (Bentley, 2014; Cass and Berglund, 2006; Harlen and Churchman, 2017).

4.1.2 SF3B1's HEAT domain has a unique superhelical conformation

In the crystal structure of the SF3B core complex (Cretu et al., 2016), the C-terminal domain of SF3B1 folds as a right-handed superhelix composed of 20 HEAT repeats (**Figure 4.1**). The difference between our crystal structure (Cretu et al., 2016) and the previous assignment (Golas et al., 2003; Wang et al., 1998) can be attributed to the relaxed HEAT sequence motif (Cingolani et al., 1999; Conti et al., 2006) which complicates structure predictions from amino acid sequences alone. Importantly, recent cryo-EM studies of yeast and human spliceosomes (Rauhut et al., 2016; Fica and Nagai, 2017; Shi, 2017) reported the same number of repeats for SF3B1^{HEAT} as in the crystal structure of the SF3B core (Cretu et al., 2016). Corroborated by consistent chemical crosslinking-mass spectrometry data (Cretu et al., 2016), this confirms the overall topology of the HEAT domain observed in the SF3B core and rules out potential artifacts resulting from the used expression system and constructs, crystallization conditions or packing in a crystal lattice.

The sequential arrangement of individual HEAT repeats into a spring-like assembly leads to a protein fold which is notable for being flexible and elastic (Grinthal et al., 2010; Kappel et al., 2010; Monecke et al., 2014; Zachariae and Grubmuller, 2008). As the overall conformation of the domain can be readily influenced/induced by interaction with other factors, many different topological states were observed for karyopherins or other model HEAT domain proteins (Conti et al., 2006; Monecke et al., 2014). As in these model systems, the unique superhelical conformation of SF3B1^{HEAT} is, likely, a direct consequence of its multiple contacts with the other

SF3B subunits (**Figure 4.1**). That is, SF3B1$^{\mathrm{HEAT}}$ appears to fold on a molecular scaffold composed of SF3B3 and SF3B5, while the Zn-finger protein PHF5A resides on the concave side of the solenoid and interacts along the diameter with its terminal repeats (Cretu et al., 2016). However, in contrast to karyopherins, the human SF3B1$^{\mathrm{HEAT}}$ domain is not stable on its own and can be obtained in a soluble form only by coexpression with SF3B3, SF3B5, and PHF5A (*unpublished data*, see also Teng et al., 2017).

The large-scale dynamics of certain HEAT-repeat proteins can be explained by differences in the relative stability of interfaces between consecutive repeats. Molecular dynamics simulations showed the presence of defined "dynamic hotspots" where the HEAT "spring" can bend/break more readily (Kappel et al., 2010; Zachariae and Grubmuller, 2008). Intriguingly, the presence of such sites is conveniently exploited by various HEAT binding proteins. For example, by direct interactions with two such dynamic regions of importin β, RanGTP enforces a particular conformation of the "spring" which is incompatible with cargo binding (Cook et al., 2005; Lee et al., 2005; Zachariae and Grubmuller, 2008). Subsequent dissociation of RanGDP leads to a more relaxed conformation of the HEAT superhelix, to which various importin cargos bind for shuttling into the nucleus (Stewart, 2006; Zachariae and Grubmuller, 2008). Similar to Ran, the PHF5A subunit is strategically positioned between the two "arms" of the SF3B1$^{\mathrm{HEAT}}$ "clamp", in close proximity to a more dynamic region of SF3B1 – the H15-H16 repeats (Cretu et al., 2016; Cretu et al., 2018). Such "dynamic hotspots" may support the large-scale reconfiguration of the SF3B1$^{\mathrm{HEAT}}$ domain observed in the B$^{\mathrm{act}}$ spliceosome (Fica and Nagai, 2017; Shi, 2017), where the protein "clamps down" on the branch helix and exhibits a more "closed" conformation (see also 4.2.3).

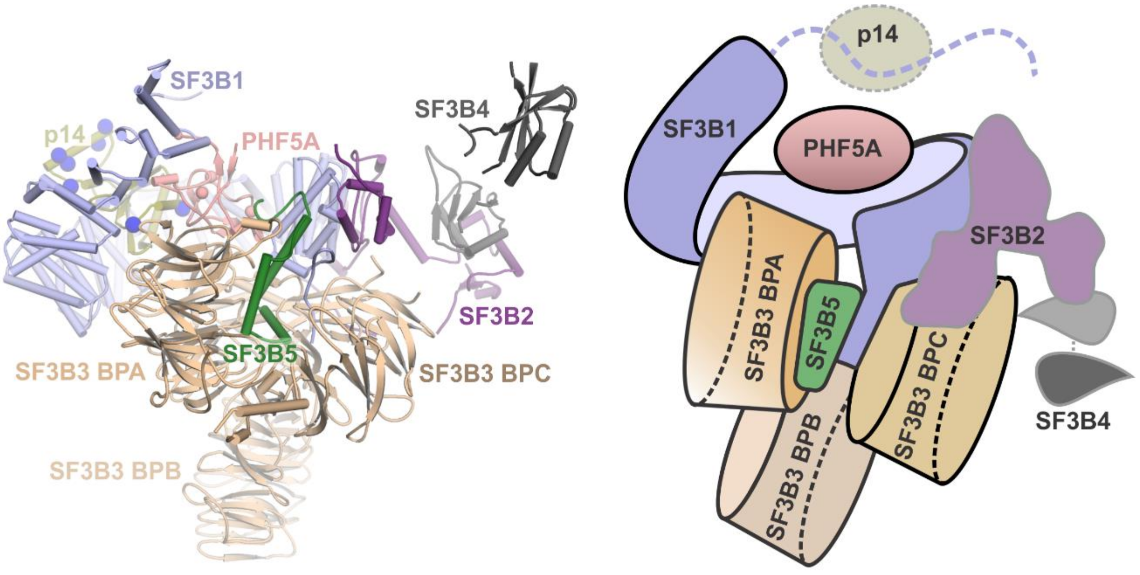


Figure 4.1. Molecular architecture of the human SF3B complex – a near-complete picture.

A near complete, pseudoatomic model of the human SF3B complex (left panel) is shown next to a schematic representation (right panel). The model was generated using the crystal structure of the SF3B core complex (SF3B1^{HEAT}-PHF5A-SF3B5-SF3B3) (Cretu et al., 2016), the crystal structure of p14/SF3B6 in complex with the pir region of SF3B1 (Schellenberg et al., 2006), and the cryo-EM structure of the yeast SF3B2-SF3B4 core module from the B complex (Plaschka et al., 2017; van Roon et al., 2017). The more flexible p14/SF3B6 subunit was modeled based on chemical crosslinking data (crosslinked residues are depicted as blue spheres) (Cretu et al., 2016).

4.1.3 SF3B1's HEAT domain and recognition of the branch-site region

In good agreement with earlier experiments (Dybkov et al., 2006; Gozani et al., 1996; Gozani et al., 1998), SF3B1^{HEAT} has the required biophysical properties for RNA binding and interacts with RNA in isolation (Cretu et al., 2016; Dybkov et al., 2006). Although we could not obtain crystals of the SF3B core in complex with single-stranded intron regions or with U2/BS duplexes, recent cryo-EM analyses show that upon stable formation of the pre-catalytic spliceosome (Bertram et al., 2017a; Plaschka et al., 2017) and of the later B^{act} complex (Haselbach et al., 2018; Rauhut et al., 2016; Yan et al., 2016) the terminal repeats of SF3B1^{HEAT} act as a "molecular pincer" to "embrace" the minor and major grooves of the branch helix. Most importantly, the conserved BS adenosine, bulged out from the branch helix, is positioned in a pocket lined by the PHF5A subunit and the H15-H17 repeats of SF3B1^{HEAT}, where it is "sequestered" prior to the first step of splicing (Rauhut et al., 2016; Warkocki et al., 2009; Yan et al., 2016).

Our mass spectrometry analyses of crosslinked peptide-RNA conjugates (Cretu et al., 2016) and earlier UV crosslinking experiments in nuclear extracts (Gozani et al., 1996; Gozani et al., 1998) indicate that SF3B1^{HEAT} interacts with several BS proximal regions as part of a composite RNA binding platform. Indeed, in the B^{act} spliceosomes, the uridine-rich region downstream of the BS (i.e., PPT) passes through the center of the HEAT superhelix, contacting both SF3B1 and PHF5A residues, and then docks on a conserved basic surface patch defined by the H4-H7 repeats (see also 5.1.5). Consistent with reported UV crosslinks in yeast (Schneider et al., 2015), the intron regions located upstream of the BS bind to the SF3B2-SF3B4 core module (yeast Cus1 (290-368)p-Hsh49p) (Plaschka et al., 2017; van Roon et al., 2017). The DUF382 domain of SF3B2, which we could not capture in a crystal structure, "folds" on the convex side of SF3B1^{HEAT} superhelix in an "extended conformation" (Plaschka et al., 2017) (**Figure 4.1**). By interacting with the H16-H20 repeats of SF3B1^{HEAT}, DUF382 serves as a bridge between the SF3B core complex and the SF3B2-SF3B4 core module (Plaschka et al., 2017; Yan et al., 2016), as we recently proposed based on chemical-crosslinking experiments (Cretu et al., 2016).

Overall, our published crystal structure and structural mass spectrometry data (Cretu et al., 2016) support the central role of SF3B1^{HEAT} in recognition of the BS and its flanking regions as part of a multipartite RNA binding platform. Recent cryo-EM structures of yeast and human spliceosomes provide unprecedented details of this essential molecular recognition event which we failed to obtain using a more reductionist approach, typical for X-ray crystallography.

4.1.4 Is SF3B6/p14 a branch-site interacting protein?

Biochemical and NMR investigations (Perea et al., 2016; Spadaccini et al., 2006) showed that p14 alone or in complex with the pir motif of SF3B1^{NTD} lacks the molecular features required for specific recognition of the branch helix or the BS-A. This partly explains why in the human B^{act} structures (Haselbach et al., 2018; Zhang et al., 2018), p14 is located on a diametrically opposed site (~60 Å) from the BS-A, on the outer rim of the SF3B1^{HEAT} domain, and close to the H1-H2 repeats. The bulged nucleobase, instead, is recognized by a different ~14 kDa SF3B subunit, PHF5A, which frames the BS-A pocket together with SF3B1's H15-H17 repeats. Importantly, the BS-A is recognized in a similar manner in yeast spliceosomes which lack p14 (Rauhut et al., 2016; Yan et al., 2016).

In the isolated SF3B (Cretu et al., 2016) and the 17S U2 snRNP (Cindy L. Will, *personal communication*), despite being somewhat dynamic, p14 samples a more central location within the circular perimeter of SF3B1^{HEAT} (**Figure 4.1**). This suggests that, at the earliest stages of spliceosome assembly, p14 might as well participate in the initial docking of the U2 snRNP to the BS, for instance, by temporarily replacing SF1 (**Figure 2.7**). In this case, its recruitment to the BS may be facilitated by the established contacts between SF3B1^{NTD} and U2AF2 bound to the PPT region of intron (Gozani et al., 1998; Sickmier et al., 2006; Spadaccini et al., 2006) and by a possible topological bend of the 3'SS (Gupta et al., 2011; Kent et al., 2003) (**Figure 2.7**). Although p14 was previously crystallized in complex with an adenine base (Schellenberg et al., 2011), it is more plausible that the BS-p14 interactions are of electrostatic nature (Perea et al., 2016; Spadaccini et al., 2006) and specific contacts between the bulged adenosine and p14 are not present in fully assembled spliceosomes. Instead, the overall surface charge of the SF3B1$^{\mathrm{HEAT}}$-p14 platform may "guide" the BS region towards the U2 snRNA (Cretu et al., 2016; Perea et al., 2016). However, upon the ATP-dependent formation of a stable U2/BS duplex, during pre-spliceosome assembly (Query et al., 1996), the BS-A will, likely, be accommodated in the PHF5A-SF3B1 pocket, as in the EM structures of the later B complex (Bertram et al., 2017a; Plaschka et al., 2017). These structural data are in apparent contradiction with previous crosslinking studies, reporting on the direct interaction of p14, and not PHF5A, with the BS-A in the A, B, and C complexes (MacMillan et al., 1994; Query et al., 1996; Will et al., 2001). It is possible, however, that the ~14 kDa protein directly crosslinked to the BS-A in spliceosomes was, in fact, PHF5A and that the different assignment is due to the comigration of these proteins on denaturing gels (Query et al., 1996; Will et al., 2001). Future cryo-EM structural studies of earlier human spliceosomes (i.e., E and A complexes) may provide a clearer view of the relative contributions of p14 to BS recognition.

4.1.5 SF3B3 – a multipurpose molecular scaffold

Human SF3B3 structure (Cretu et al., 2016), and of its related orthologs (Cretu et al., 2016; Rauhut et al., 2016; Yan et al., 2016), strikingly resembles that of DDB1 (DNA damage-binding protein 1), a multifunctional protein scaffold involved in the regulation of many cellular processes, such as DNA repair, DNA replication and transcription (Jackson and Xiong, 2009; Lee and Zhou, 2007; Scrima et al., 2011) (**Figure 4.2**). As part of the larger Cullin4 RING finger ligase (CRL4) complexes, DDB1 serves as an adaptor for specific substrate receptors while recruiting concomitantly the E3 ligase module to the substrate (Angers et al., 2006; Fischer et al., 2011).

DDB1 structure consists of three interspersed WD40-like β-propellers, followed by an all-helical domain (Angers et al., 2006; Li et al., 2006). As in the case of DDB1, SF3B3's BPA and BPC domains frame a deep clam-shaped cavity with the top surfaces of the β-propellers facing each other (Cretu et al., 2016). SF3B3's alpha helical CTD domain folds back and reinforces the double β-propeller fold by making extensive contacts with both BPA and BPC (Cretu et al., 2016). Overall, this results in a rigid module that stably accommodates the SF3B5 subunit, folded as three alpha helices, and the C-terminal region of SF3B1 while interacting, at the same time, with the "arms" of SF3B1^{HEAT} (Cretu et al., 2016).

Similar to DDB1 structures, SF3B3's BPB domain is more flexible and projects outwards making no contacts with the other SF3B subunits (Cretu et al., 2016) (**Figure 4.2**). The relative mobility of SF3B3's BPB domain indicates that the β-propeller may sample for different splicing factors interacting with SF3B and may even contribute to its recruitment to the spliceosome as part of the U2 snRNP. Indeed, in the B and B^{act} complexes, the BPB domain of SF3B3 is rearranged compared to the crystal structure (Cretu et al., 2016; Rauhut et al., 2016; Yan et al., 2016) and interacts with the Brr2 RNA helicase, a core resident of the U4/U6.U5 tri-snRNP. Thus, the BPB domain may serve as an initial docking site for the tri-snRNP upon its association with the pre-spliceosome (Bertram et al., 2017a; Plaschka et al., 2017) (**Figure 2.6**). The structural mobility of SF3B3's BPB domain, which is, likely, of Brownian nature, may increase the capture radius of the complex and thus promote the more efficient recruitment of the tri-snRNP.

The majority of DDB1 substrate receptors, such as DDB2, interact with the BPA-BPC cleft via a short alpha-helical motif (i.e., the "DDB-box") (Fukumoto et al., 2008; Lee and Zhou, 2007). This particular motif is not present in SF3B5, and the SF3B subunit interacts tightly with BPA and not with BPC as DDB1 ligands do (Li et al., 2006; Scrima et al., 2011). However, the overall fold of SF3B5 is reminiscent of the N-terminal helical region of DDB2 and related receptors, whose structures also comprise two antiparallel α-helices linked by a short turn (Fischer et al., 2011;

Scrima et al., 2011). By analogy, this indicates that SF3B5 acts as a DDB2-like “substrate receptor”, whereas SF3B3 is, functionally, the “adaptor” (**Figure 4.2**). In this respect, SF3B1-PHF5A can be viewed as the specific splicing “substrate” of the SF3B3-SF3B5 adaptor-receptor complex. Intriguingly, SF3B3 is also a stable component of the human transcriptional coactivator STAGA (Martinez et al., 2001) and TFTC (Brand et al., 2001) complexes. Recently, SF3B5 was observed in the related SAGA complex in *Drosophila* (Stegeman et al., 2016), suggesting that the entire SF3B3-SF3B5 module might be present in these chromatin-modifying complexes. SF3B3 also interacts directly with several cullins and is, likely, a DDB1-like component of different CRLs (Menon et al., 2008). We, therefore, propose that SF3B3 serves as a multipurpose molecular scaffold participating in different cellular pathways as part of larger, multi-subunit assemblies.

Interestingly, in the recent cryo-EM structure of the polymerase module of yeast CPF (cleavage and polyadenylation factor), the Cft1 subunit (human CPSF160) adopts a similar fold as SF3B3 and DDB1, and accommodates in the largely similar BPA-BPC cleft its substrate “receptor” subunit Pfs2 (human WDR33) (Casanal et al., 2017). Molecular recognition of Pfs2 by Cft1 also involves a small alpha-helical region (**Figure 4.2**). The Yth1 subunit (human CPSF30) contacts the Cft1 BPC domain and Pfs2, and comprises two zinc finger domains assembled in a fold which is reminiscent of PHF5A (Casanal et al., 2017).

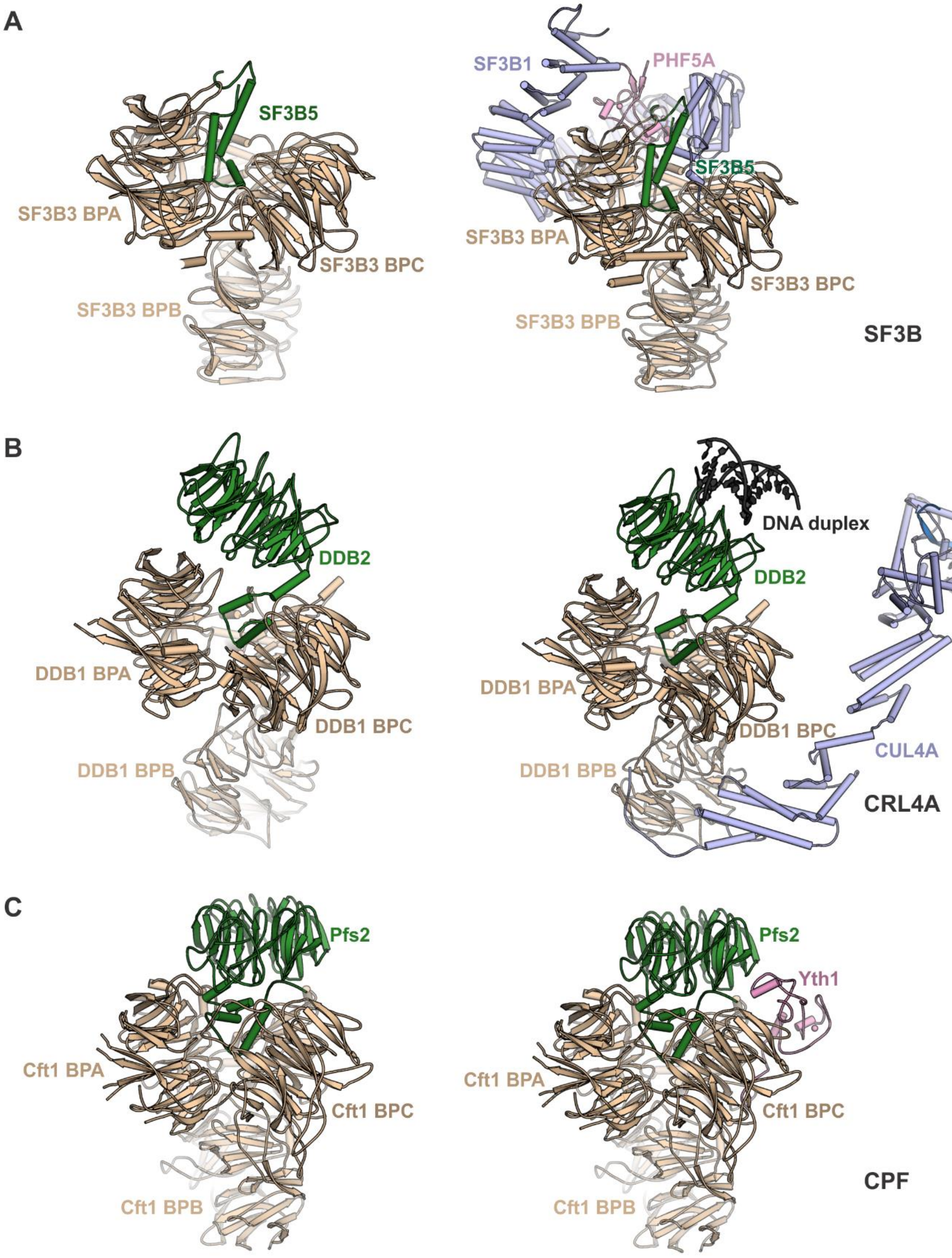

A
SF3B5
SF3B3 BPA
SF3B3 BPC
SF3B3 BPB
SF3B1
PHF5A
SF3B5
SF3B3 BPA
SF3B3 BPC
SF3B3 BPB
SF3B
B
DDB2
DDB1 BPA
DDB1 BPC
DDB1 BPB
DNA duplex
RBX1
DDB2
DDB1 BPA
DDB1 BPC
CUL4A
DDB1 BPB
CRL4A
C
Pfs2
Cft1 BPA
Cft1 BPC
Cft1 BPB
Pfs2
Yth1
Cft1 BPA
Cft1 BPC
Cft1 BPB
CPF

Figure 4.2. The similar structural organization of SF3B, CRL4A, and yeast CPF.

The SF3B core complex (**A**, this work), the CRL4A E3 ligase (**B**, PDB 4A0K), and the polymerase module of yeast CPF (**C**, PDB 6EOJ) were structurally aligned using the SF3B3 subunit as a reference (light brown). SF3B5, DDB2, and Pfs2 are colored in green. SF3B1 and CUL4A are colored in light blue, while PHF5A and Yth1 are shown in light pink. Note that the corresponding BPA and BPC domains of SF3B3, DDB1, and Cft1 form a clam-shaped cleft, where SF3B5, DDB2, and Pfs2 bind, respectively.

The similar organization of SF3B, CPF polymerase, and CRL4A are indicative of the modular architecture of these large assemblies. It also shows how a versatile protein fold, the triple β-propeller fold of SF3B3/DDB1/Cft1, can be efficiently reused as the central building “brick” of more complex molecular “legos”.

4.1.6 Structural insights into SF3B1’s cancer-related mutations

The crystal structure of the human SF3B core complex shows that most of the cancer-relevant mutations of SF3B1, including the common substitution K700E, map to the H4-H7 repeats of the HEAT superhelix in a conserved basic patch (Cretu et al., 2016) (**Figure 4.3**). Thus, substitution of these residues may target an important interaction interface between SF3B1 and pre-mRNA. Alternatively, mutations at this site may alter the direct contacts between SF3B1 and other splicing factors or trigger a structural reorganization of the entire SF3B complex. We could not detect direct UV-crosslinks between these SF3B1 residues and RNA (Cretu et al., 2016) nor we could observe a significant change in SF3B1’s RNA crosslinking pattern in the presence of K700E (*unpublished data*). In addition, the affinity of $\mathrm{SF3B}^{\mathrm{K700E}}$ towards a canonical intron sequence is, apparently, unchanged when compared to the wild-type complex (Cretu et al., 2016). Moreover, judging from the chemical-crosslinking patterns, the recombinant SF3B complex carrying the K700E hotspot mutation (Cretu et al., 2016) has a similar structural organization as the wild-type complex. Although the UHM domain of U2AF2 is located close to SF3B1’s hotspot region (Cretu et al., 2016), the K700E substitution does not affect the direct interaction of SF3B1 with U2AF2.

Altogether, these experiments suggest that the etiology of SF3B1-driven cancers is rather complex and cannot be fully captured by structural and biochemical analyses of SF3B in isolation.

Importantly, in the cryo-EM structure of the yeast B^{act} complex (Rauhut et al., 2016; Yan et al., 2016), the H4-H7 repeats of SF3B1 are in contact with the uridine-rich region of the intron which follows downstream of the BS (**Figure 4.3**). Structural superposition of human SF3B1 (Cretu et al., 2016; Haselbach et al., 2018) onto yeast Hsh155p shows that the K700 hotspot residue does interact with the phosphodiester backbone (**Figure 4.3**). Intriguingly, this contact, likely, induces a ~90^{o} "kink" in intron's structure, while the neighboring Q699 residue stabilizes the flipped-out uridine nucleobase via a hydrogen bond (**Figure 4.3**). It is unclear why such a reorientation of the intron is required or whether the equivalent region of the PPT has a similar structure in the human spliceosome. Several other SF3B1 residues mutated in cancers (e.g., E622, R625, H662, K666) map mainly on the concave side of SF3B1^{HEAT}, facing the uridine-rich region of the intron (**Figure 4.3**). These residues bind either directly to the phosphodiester backbone, before the ~90^{o} reorientation, or are involved in hydrogen-bonding interactions with neighboring SF3B1 residues (**Figure 4.3**). Altogether, the multiple contacts established by the cancer hotspot residues of SF3B1^{HEAT} appear to stabilize the intron structure downstream of the BS. We, therefore, hypothesize that SF3B1's missense substitutions, including the common K700E substitution, may interfere with the proper folding of the intron upon binding to the H4-H7 repeats (see also Jenkins and Kielkopf, 2017). In turn, this may translate into a weak and/or suboptimal interaction of the PPT with SF3B1 and/or U2AF2 (Cretu et al., 2016; Jenkins and Kielkopf, 2017). As the "cryptic" BSs can base-pair more efficiently with the U2 snRNA (Alsafadi et al., 2016; Darman et al., 2015), their selection under a mutant background may compensate for the altered interaction of the PPT (or equivalent sequences) with SF3B1.

Interestingly, in the recent cryo-EM structure of the endogenous yeast P complex (Liu et al., 2017), the uridine-rich region between the BS and the 3'SS folds into a stem-like structure. As in this late stage spliceosome, SF3B1/Hsh155p is no longer present it is possible that, at the earlier stages, the HEAT domain of SF3B1 together with U2AF2 may act as "RNA chaperones" to prevent the premature formation of this structural element. In the end, it is this plausible function of SF3B1 that might be comprised in cancer cells (Kesarwani et al., 2017).

Alternatively, the SF3B1 hotspot mutations may affect the conformation of the HEAT superhelix locally and, thus, prevent, indirectly, its interaction with the PPT. In the yeast and human B^{act} structures, the RES complex and the Prp2 RNA helicase are positioned at the 3' end of the intron, close to SF3B1's HEAT repeats mutated in cancers (Rauhut et al., 2016; Yan et al., 2016). Thus, changes in the conformation of $SF3B1^{HEAT}$ could lead to an altered interaction with these splicing complexes and, potentially, with earlier factors, such as U2AF2 (Cretu et al., 2016) and/or the Prp5/DDX46 RNA helicase (Carrocci et al., 2017; Tang et al., 2016). Finally, all the scenarios mentioned above could be valid and, in fact, the alternative BS selection in cancers may result from changes in intron's secondary structure, SF3B1's conformation, and SF3B1's interactome.

SF3B1's hotspot mutations induce global changes in cell's transcriptome with the expression levels of many genes being reduced/altered due to noncanonical splicing, a phenomenon denoted as "spliceosome sickness" (Alsafadi et al., 2016; Darman et al., 2015). However, it is possible that the aberrant splicing of certain pre-mRNAs may drive the disease or influence its outcomes. Interestingly, SF3B1 hotspot mutations have the highest incidence in RARS (refractory anemia with ring sideroblasts, ~80-90%) of all MDS subtypes (Papaemmanuil et al., 2011). A hallmark of this disease is the presence of characteristic erythroblasts bearing iron-overloaded mitochondria (i.e., ring sideroblasts) (Cazzola et al., 2003).

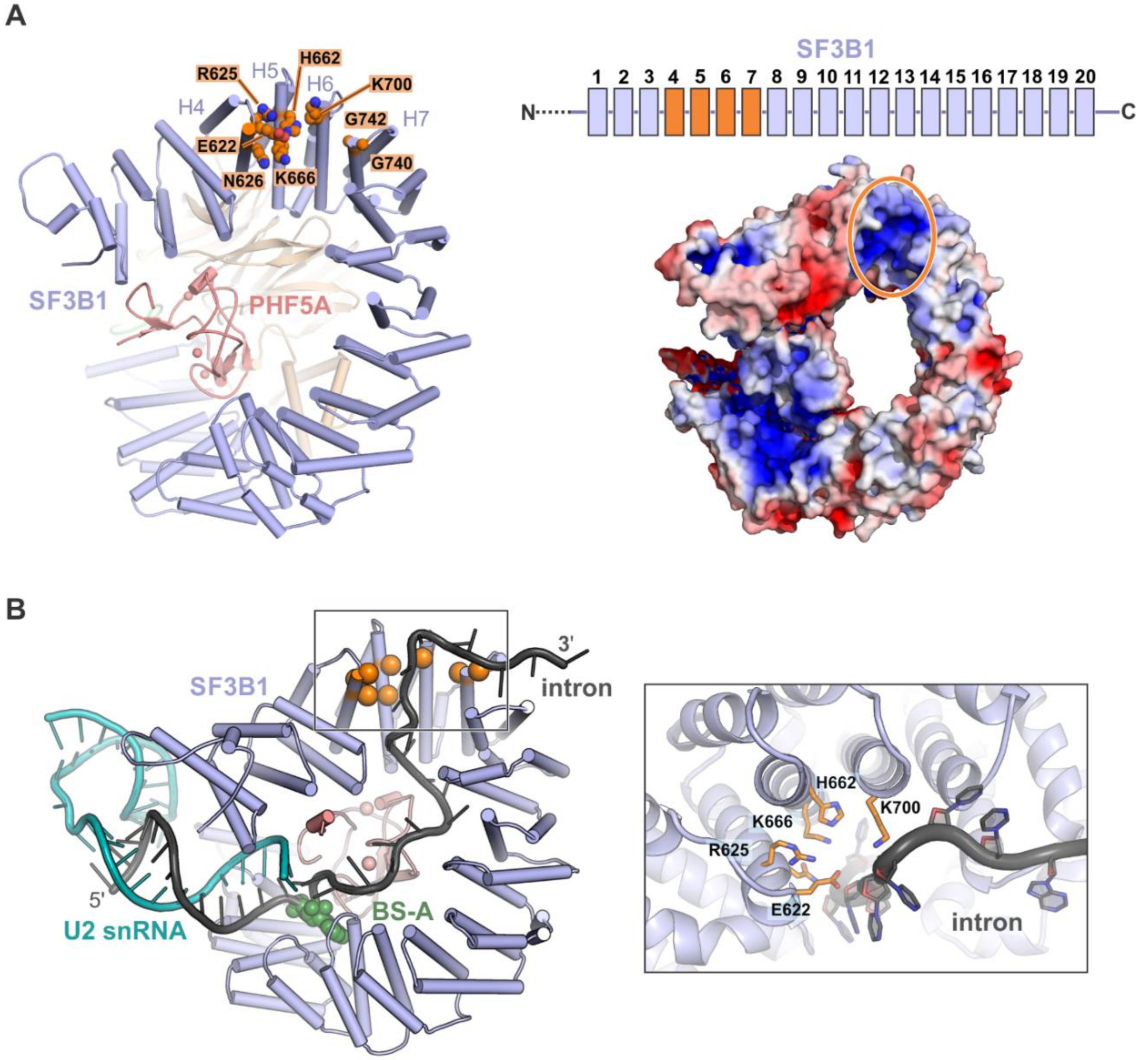


Figure 4.3. SF3B1's recurrent cancer mutations map to a conserved, intron binding site.

(**A**) Recurrent mutations in human SF3B1 cluster at the H4-H7 repeats (Cretu et al., 2016). Left, SF3B1 residues frequently mutated in hematopoietic malignancies and uveal melanoma are shown as spheres and colored in orange. SF3B1 and PHF5A are colored in light blue and light pink, respectively. Right, surface charge distribution of SF3B1-PHF5A (±7 kT/e, blue: positive, red: negative). SF3B1 hotspot residues from the H4-H7 repeats map to a basic surface patch. (**B**) Left, the structure of human SF3B1 and PHF5A (Cretu et al., 2016; Haselbach et al., 2018) bound to the 3' end of the intron. SF3B1 (light blue) and PHF5A (light pink) were modeled based on the structures of the yeast orthologs, Hsh155p and Rds3p, respectively, as observed in the B^{act} complex (PDB 5GM6). The intron is depicted in dark gray, whereas the U2 snRNA is colored in teal. SF3B1 residues are shown as orange spheres. The branch-site adenosine (BS-A) is colored in green and represented as spheres. Right, SF3B1 residues mutated in cancers, likely, interact with the uridine-rich region of the intron (equivalent to the PPT), downstream of the branch-site.

Interestingly, the ABCB7 gene, which encodes an iron transporter, is downregulated in RARS due to mutant SF3B1-dependent aberrant splicing (Darman et al., 2015; Dolatshad et al., 2016). In this example, the disease phenotype correlates well with a more targeted effect of SF3B1 mutations. However, further functional studies will be needed to understand better the relationship between SF3B1 mutations, aberrant splicing, and the pathophysiology of other related cancers.

4.2 Molecular insights into splicing modulation by antitumor SF3B inhibitors

Pladienolides, as well as the related spliceostatins, sudemycins, and herboxidiene, target the SF3B complex (Hasegawa et al., 2011; Kaida et al., 2007; Kotake et al., 2007) and interfere with the stable binding of the U2 snRNP to the BS region of the intron (Corrionero et al., 2011; Folco et al., 2011). Here, we determined crystal structures of SF3B in complex with different pladienolides and herboxidiene. Our structures show that these small-molecule compounds share a common binding site on SF3B. The modulator binding tunnel exists in the "open" conformation of the complex and is defined by the H15-H17 repeats of SF3B1 and several invariant residues of PHF5A. Our comparative structural analyses suggest that pladienolides and herboxidiene prevent a conformational change in SF3B1 which is required for the stable recognition of the branch helix and of the BS-A.

4.2.1 A pipeline for structure-based discovery of next-generation splicing modulators

Structure determination of large and multimeric complexes by X-ray crystallography is typically challenging and requires extensive engineering of the protein constructs, optimization of the crystallization conditions and data collection strategies (Bunker, 2016; Hoelz et al., 2016; McGinty and Tan, 2015). Generally, a robust high-throughput pipeline for drug discovery relies on the ability to obtain the macromolecule of interest in significant amounts, to grow high-quality crystals which diffract routinely X-rays to high resolution, and to trivially solve the phase problem (i.e., by molecular replacement) (Blundell et al., 2002; Jubb et al., 2012).

In our long-standing efforts to determine structures of SF3B in complex with splicing modulators, we have systematically optimized: (i) the heterologous overexpression of the entire complex and various core complexes in insect cells (Cretu et al., 2016); (ii) the structural homogeneity of the used protein constructs, by removing disordered regions and/or flexible domains, while preserving intact the drug binding pocket (Cretu et al., 2018, *unpublished data*); (iii) the crystallization and cryoprotection procedures (Cretu et al., 2016; Cretu et al., 2018; *unpublished data*). The result is a crystal-to-structure pipeline (*unpublished data*) which we used, as a proof of principle, to solve structures of SF3B in complex with pladienolide and herboxidiene modulators at resolutions enabling an unambiguous placement of the target ligand. In its current form, our pipeline can be further applied to determine crystal structures of SF3B in complex with other families of modulators (such as spliceostatins, meayamycin, thailanstatins or sudemycins). Moreover, our approach can be adapted for different serial crystallography routines that are becoming increasingly popular with the advent of new microfocused beamlines and of free-electron lasers (Erlanson et al., 2016; Schlichting, 2015). In addition, we think that our pipeline may support ongoing drug discovery efforts of novel compounds targeting SF3B using fragment-based methodologies (Erlanson et al., 2016; Murray et al., 2012).

4.2.2 Revisiting the common pharmacophore hypothesis

In our crystal structures, pladienolides (PB, E7107, H3B-8800) and herboxidiene (GEX1A) bind to an “hourglass-shaped tunnel” located at the interface between SF3B1’s H15-H17 repeats and PHF5A. Importantly, the small-molecules compounds are accommodated in the tunnel as such that the ring group (i.e., macrolide or pyran) resides in the top “funnel”, facing SF3B3, while the “kinked” aliphatic chain is situated in the bottom “funnel”, projecting towards SF3B1 (**Figure 4.4**). The invariant diene moiety is positioned at the “neck” region of the tunnel, where it adopts similar poses in all structures, except H3B-8800 (Cretu et al., 2018; *unpublished data*). Notably, the diene group is involved in π-π and, likely, cation-π interactions with PHF5A-Y36 and SF3B1-R1074 residues, respectively, whose mutations affect binding of modulators to the complex (Cretu et al., 2018; Teng et al., 2017) and result in drug resistance in cells (Cretu et al., 2018; Teng et al., 2017). In good agreement with the proposed pharmacophore model (Lagisetti et al., 2008; Lagisetti et al., 2014), this indicates that the diene group represents an essential molecular feature of pladienolides and herboxidiene (**Figure 4.4**).

Conversely, our co-crystal structures do not confirm the general importance of the other chemical groups from the proposed pharmacophore model (Lagisetti et al., 2008; Lagisetti et al., 2014) (**Figure 4.5**). The C7 oxycarbonyl group, attached to the macrolide ring of pladienolides, does form a crucial hydrogen bond with the backbone amide of PHF5A-R38 (Cretu et al., 2018) (**Figure 4.5**). However, a similar interaction is not observed in the case of herboxidiene. Instead, the carboxyl group of the pyran core (C3) is involved in a polar contact with the SF3B1-K1071 residue, on the opposed side of the SF3B1-PHF5A tunnel (**Figure 4.5**). Moreover, in all structures, the epoxide group is exposed to the solvent and appears to contribute very little to the overall recognition (Cretu et al., 2018; *unpublished data*), suggesting that its replacement with cyclopropane could be well tolerated (Villa et al., 2013).

When compared to pladienolides and herboxidiene, spliceostatins and the related compounds share a common diene group (**Figure 2.11**). However, the chemistry of the other moieties is significantly different as such that structural docking is not reliable. Although spliceostatins/sudemycins share a common binding site with pladienolides/herboxidiene (Effenberger et al., 2016; Teng et al., 2017), these compounds may still adopt distinct poses in the SF3B1-PHF5A tunnel or even use different principles of molecular recognition.

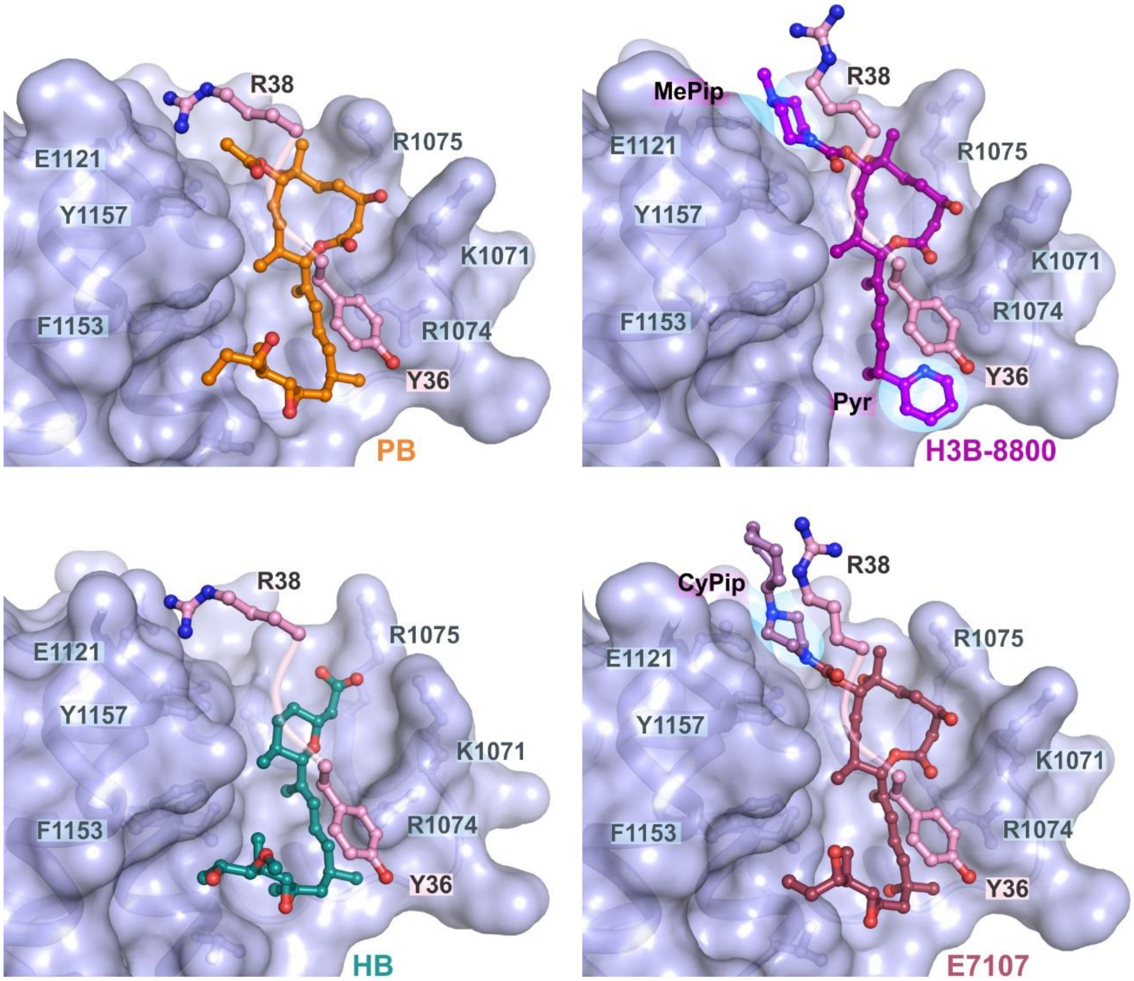


Figure 4.4. Accommodation of splicing modulators by SF3B.

Side-by-side comparison of different pladienolides and herboxidiene bound to the SF3B1-PHF5A tunnel. The structural alignment was carried out using the PB (orange) co-crystal structure as a reference. Herboxidiene (HB) is colored in teal and shown in stick representation. The lead pladienolides H3B-8800 and E7107 are colored in magenta and brown, respectively. SF3B1's side of the tunnel is depicted as a molecular surface (light blue). For the sake of clarity, only the PHF5A residues which contact the splicing modulators are depicted (sticks, colored in pink). MePip represents the N-methyl-piperazine group of H3B-8800, Pyr – the pyridine group of H3B-8800, and CyPip – the N-cycloheptyl-piperazine group of E7107.

The overall topology and rich chemistry of the SF3B1-PHF5A tunnel suggest further strategies for drug design and optimization. Derivatization of the pladienolide macrocycle at C7 with an N-cycloheptyl-piperazine (CyPip) group in E7107 is well tolerated by the binding site (**Figures 4.4 and 4.5**). Therefore, relatively large groups can be grafted at C7 without significant effects on macrolide's binding to the SF3B1-PHF5A tunnel (see also Kotake et al., 2007). The bulky moiety of E7107 is accommodated by the newly established interactions with PHF5A-R38, SF3B1-Y1157, and SF3B1-E1121 which form an extended pocket and sandwich the piperazine group (**Figures 4.4 and 4.5**). The cycloheptyl group of CyPip does not appear to interact with SF3B and is, likely, flexible (**Figures 4.4 and 4.5**). Our structure, thus, explains why its removal in H3B-8800 does not change the orientation and recognition of the piperazine group drastically.

Replacement of the entire aliphatic arm of E7107 with a pyridine (Pyr) ring in H3B-8800 results in the decrease in the affinity and potency of the compound (Seiler et al., 2018) (**Figure 4.4**). Strikingly, our crystal structure shows that the significant change in the ligand structure affects the orientation of the diene moiety, but not of the macrolide group; the macrolide core of H3B-8800 adopts a largely similar pose as in the case of the parent compound (i.e., E7107). This indicates that pladienolides and, possibly, herboxidiene bind SF3B1-PHF5A in a directional manner, primarily defined by the shape complementarity between the macrolide/pyran group and the tunnel. Therefore, it is likely that swapping the aliphatic chain for a pyridine ring in herboxidiene will not abolish entirely its binding to the SF3B tunnel. With its simplified chemistry, the herboxidiene-H3B-8000 hybrid, where the pladienolide macrocycle is replaced by a functionalized pyran ring, could be more stable in aqueous solutions and more amenable to industrial scale synthesis.

4.2.3 Splicing modulators binding site reveals a conformational switch in SF3B1

The discrete conformational transitions of SF3B have been linked to the binding and unloading of the branch helix during splicing catalysis (Golas et al., 2005; Hilliker et al., 2007; Rauhut et al., 2016). Earlier low-resolution EM analyses of the isolated SF3B (Golas et al., 2003) and of the U11/U12 di-snRNP particle (Golas et al., 2005) provided the first indications that the HEAT superhelix of SF3B1 might undergo a structural rearrangement, by way of a hinge region, during spliceosome's assembly cycle.

Our crystal structure of the SF3B core complex (Cretu et al., 2016) and recent cryo-EM structures of spliceosomal complexes (Fica and Nagai, 2017; Shi, 2017) showed that indeed SF3B1 has at least two defined conformational states. In the isolated complex, SF3B1 folds onto the SF3B3-SF3B5 protein scaffold in an "open" conformation with PHF5A positioned between the N-terminal H2-H3 and the C-terminal H15-H18 HEAT repeats (Cretu et al., 2016). However, a more "closed" state of SF3B1 was detected in the presence of a stably bound branch helix. In this conformation of the protein, the distance between the terminal HEAT repeats is reduced to ~18 Å, and PHF5A is repositioned with respect to the N-terminus of SF3B1 (Rauhut et al., 2016; Yan et al., 2016). As the branch helix makes extensive contacts with opposed HEAT repeats of SF3B1 (H1 and H15-H19), the bulged BS adenosine is accommodated in a protein pocket formed by SF3B1's H15-H17 repeats and PHF5A (Rauhut et al., 2016; Yan et al., 2016) (**Figure 4.3B**).

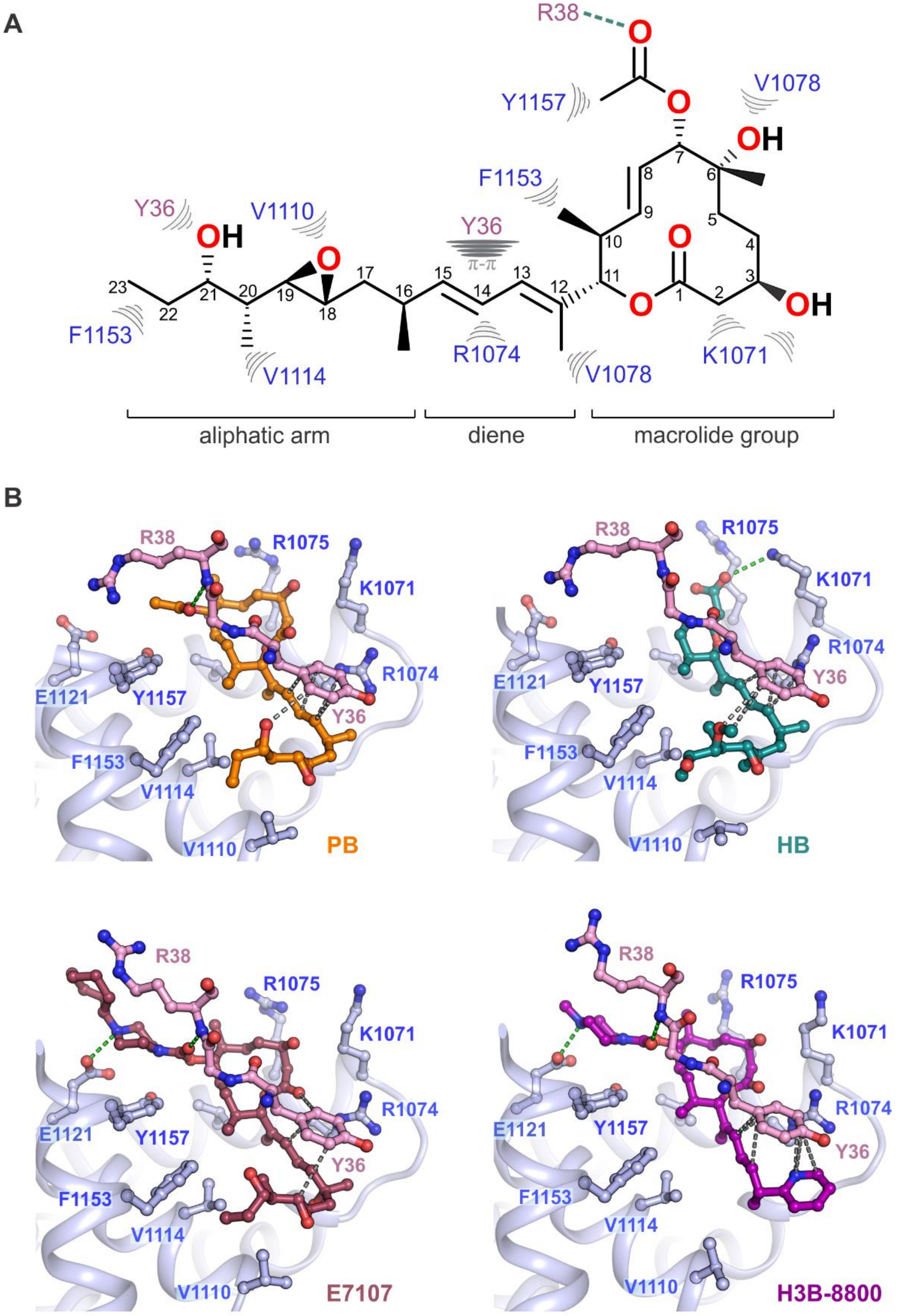

A
R38
Y1157
V1078
OH
F1153
Y36
V1110
Y36
π-π
OH
O
O
O
O
OH
F1153
V1114
R1074
V1078
K1071
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
aliphatic arm
diene
macrolide group
B
R38
R1075
K1071
R1074
E1121
Y1157
Y36
F1153
V1114
V1110
PB
R38
R1075
K1071
R1074
E1121
Y1157
Y36
F1153
V1114
V1110
HB
R38
R1075
K1071
R1074
E1121
Y1157
Y36
F1153
V1114
V1110
E7107
R38
R1075
K1071
R1074
E1121
Y1157
Y36
F1153
V1114
V1110
H3B-8800

Figure 4.5. Molecular recognition of splicing modulators: differences and similarities.

(**A**) Schematic depiction of the contacts between the SF3B1-PHF5A tunnel and pladienolide B (PB). Polar contacts are colored in green and represented as dashed lines, whereas the hydrophobic and weak polar interactions are depicted as semicircles and colored in gray. (**B**) Binding of different pladienolides and herboxidiene to the SF3B1-PHF5A tunnel. SF3B1 (light blue) and PHF5A (light pink) residues, framing the binding tunnel, are shown as sticks. PB, H3B-8800, E7107, and herboxidiene (HB) are depicted as in Fig. 4.4. Representative hydrophobic and weak polar interactions between PHF5A-Y36 and the splicing modulators are shown as dashed gray lines. For the sake of simplicity, all the other hydrophobic/weak interactions are not depicted. The polar contacts between the splicing modulators and the SF3B1-PHF5A tunnel are indicated as dashed lines, colored in green. Note that in all pladienolides, the C7 oxycarbonyl group of the macrolide core forms a hydrogen bond with the PHF5A-R38 backbone amide. However, the C3 carboxyl group of HB establishes a polar contact with SF3B1-K1071.

Our follow-up structures of SF3B bound to different splicing modulators advance a more detailed understanding of the structural dynamics of SF3B1. The co-crystal structures of SF3B in complex with different pladienolides and herboxidiene (Cretu et al., 2018, *unpublished data*) show that the small molecules are bound to the “open” state of SF3B1 in a conserved tunnel formed between the H15-H17 region and PHF5A (**Figure 4.4**). In contrast, the BS-A pocket, observed in the presence of the branch helix, lacks important structural features required for splicing modulators binding (Cretu et al., 2018) (see also 5.2.4). Thus, the BS-A and the SF3B modulators bind to mutually exclusive pockets, available in the “closed” and “open” SF3B1 conformation, respectively. Importantly, our comparisons between the “open” conformation, observed in the presence of modulators, and the “closed” state of SF3B1 indicate that the large conformational difference can be explained by the presence of a hinge located in the H15-H16 region of SF3B1, where both splicing modulators and the BS-A bind (Cretu et al., 2018). Thus, the transition from the “open” to the “closed” conformation of SF3B1, likely, entails a ~23° rotation and ~17° lateral translation of the H15 HEAT repeat over the H15-H16 hinge which leads, in the end, to the large-scale movement of the entire H1-H15 arch of SF3B1 (Cretu et al., 2018).

4.2.4 Novel structural insights into spliceosome assembly

The precise order of events leading to a stable pre-spliceosome (A complex) is not entirely known. Accurate biochemical and structural investigations were hampered by the fast assembly kinetics and the relative instability of the assembly intermediates (Das et al., 2000). This might explain a certain ambiguity in interpreting the relative effects of splicing modulators on spliceosome assembly *in vitro*. Reed lab reported that the pladienolide D derivative E7107 prevents an “ATP-dependent conformational change in the U2 snRNP” which is required for stable binding of U2 to the BS region (Folco et al., 2011). In their assay, the U2 snRNP complexes pre-treated with ATP were resistant to E7107 and could stably bind a BS oligonucleotide (Folco et al., 2011). More recently, Seiler et al. showed that a similar resistance to E7107 can be artificially induced by mutations in the SF3B1-PHF5A tunnel which inhibit splicing modulators binding (Seiler et al., 2018). Valcarcel lab and Jurica lab, however, showed that spliceostatin A and pladienolide B destabilize the A complex assembly by disrupting critical interactions between SF3B1 and the intron at the BS, even in nuclear extracts pre-incubated with ATP (Corrionero et al., 2011; Effenberger et al., 2016). Importantly, none of these results can be convincingly explained by simple binding of the small molecule compounds to SF3B1 (Effenberger et al., 2016).

In our model (**Figure 4.6**), derived from structural analyses, we show that pladienolides and related splicing modulators interfere/prevent a conformational change in SF3B1, required for branch helix binding and stabilization/formation (Cretu et al., 2018). We, therefore, propose that the very same “open” conformation of SF3B1 is also present in the A-like complexes stalled with SF3B modulators and in the isolated U2 snRNP and U11/U12 di-snRNP. It is tempting to speculate that this early A-like spliceosome may, in fact, represent a novel splicing complex which follows immediately after the E complex, also known as the E^* complex (Champion-Arnaud et al., 1995).

In this splicing complex, SF1/mBBP might have already been displaced from the BS, explaining the ATP-dependency (Champion-Arnaud et al., 1995; Corrionero et al., 2011; Effenberger et al., 2016), and, as a result, the U2 snRNP is more stably bound, compared to the E complex (Das et al., 2000). In part, this is due to the strong interactions established between SF3B1^{NTD} and U2AF2 (Cass and Berglund, 2006; Cretu et al., 2016; Thickman et al., 2006). However, as SF3B1 is trapped in an "open" state, the branch helix cannot stably anneal, as its formation requires a "closed" conformation of SF3B1 (Plaschka et al., 2017; Rauhut et al., 2016; Yan et al., 2016). Thus, the overall conformation of the U2 snRNA and the orientation of other BS-interacting factors, such as SF3A2, might be altered in the A-like complex versus a stable pre-spliceosome (Haselbach et al., 2018; Plaschka et al., 2017). Altogether, this would result in a spliceosome complex to which the U4/U6.U5 tri-snRNP cannot be recruited in a productive manner, thus explaining the stalled splicing assembly pathway.

Intriguingly, in the yeast B complex, the primary interfaces between the U2 snRNP and the tri-snRNP involve SF3B proteins: SF3B1's (yeast Hsh155p) H11-H13 repeats contact the N-terminal region of Prp3, while SF3B3's BPB domain (yeast Rse1p) and SF3B2 (yeast Cus1p) interact with the Brr2 RNA helicase of the tri-snRNP (Plaschka et al., 2017). At the same time, the U2 snRNA base-pairs with U6 snRNA to form the U2/U6 helix II (Plaschka et al., 2017). Thus, it is possible that the conformational change in SF3B1, upon the stable formation of the branch helix, "opens up" the U2 snRNP for the tri-snRNP binding. In this respect, the conformational switch in SF3B1 may serve as a checkpoint to ensure that the tri-snRNP is only recruited after the branch helix has stably formed. Current data from us (Cretu et al., 2018) and from other labs (Effenberger et al., 2016; Roybal and Jurica, 2010) suggest that this functional checkpoint in the spliceosome assembly is targeted by SF3B modulators.

Alternatively, the labile A-like spliceosome formed in the presence of SF3B modulators may represent a “dead-end” complex and not a splicing intermediate. However, this scenario is less likely to occur, given the reversibility of splicing inhibition in the presence of inactive analogs of modulators (Effenberger et al., 2016). Nevertheless, it will be interesting to compare the protein and RNA composition of the purified A-like complex with the composition of the human E and A complexes (Behzadnia et al., 2006; Das et al., 2000; Hartmuth et al., 2002).

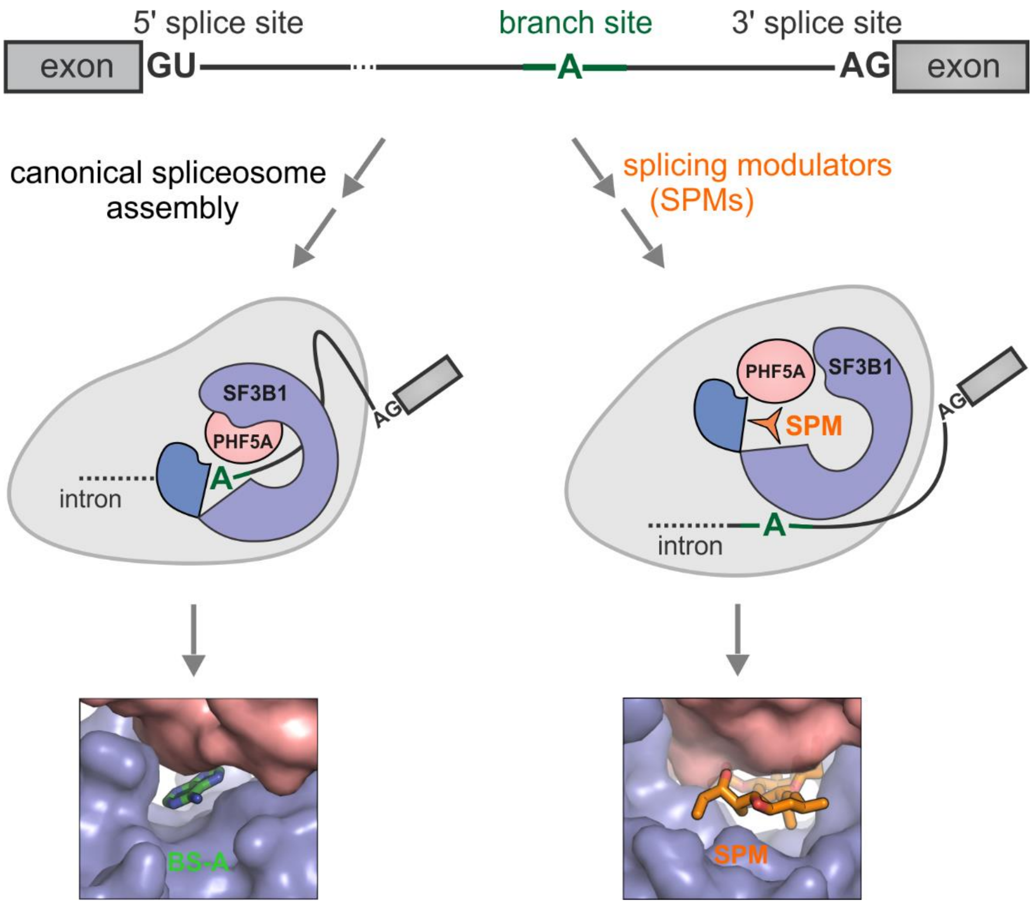


Figure 4.6. Splicing modulators may interfere with a conformational switch in SF3B1 required for the stable formation of the pre-spliceosome.

The stable formation of the branch helix requires a "closed" conformation of SF3B1 (left). During pre-spliceosome formation, it is likely that SF3B1 undergoes a conformational transition to the "closed" state, stabilizing the branch helix. In the close state of SF3B1, the branch site adenosine (BS-A) is accommodated in the SF3B1-PHF5A pocket. In the presence of splicing modulators (SPMs), SF3B1 is locked in the "open" state (right). This state of SF3B1 is incompatible with the BS recognition, thus explaining the stalled splicing assembly (right). Adapted with modifications from (Cretu et al., 2018).

The conformational change in SF3B1, which, likely, occurs during pre-spliceosome assembly, may be coupled to the Prp5-dependent rearrangements of the U2 stem II region (**Figure 2.8**), required for the BS binding in yeast (Hilliker et al., 2007; Perriman and Ares, 2010; Perriman and Ares, 2007). The stem II region of the U2 snRNA is located 3' from the BS interacting sequence (BSiR) and may adopt two mutually exclusive conformations: stem-loop IIa and stem IIc (Hilliker et al., 2007; Zavanelli et al., 1994). Genetics studies in yeast showed that the U2 stem-loop IIa promotes the stable pre-spliceosome formation (Perriman and Ares, 2007; Zavanelli et al., 1994). Conversely, the U2 stem IIc has an inhibitory effect on the assembly of the A complex (Perriman and Ares, 2007; Zavanelli et al., 1994), but favors the branching reaction (Hilliker et al., 2007).

Most recently, the stem-loop IIa configuration of the U2 snRNA was detected in the yeast and human B and B^{act} complexes (Fica and Nagai, 2017; Shi, 2017), whereas the stem IIc was observed in the later C and C^{*} spliceosomes (Fica and Nagai, 2017; Galej et al., 2016). Importantly, in the B and B^{act} complexes, where SF3B1 exhibits a "closed" conformation, the stem-loop IIa region of the U2 snRNA is stabilized by the SF3B1/Hsh155p H20 repeat, SF3A3/Prp9p, and SF3B2/Cus1p (Plaschka et al., 2017; Yan et al., 2016). During the catalytic activation of the spliceosome (i.e., the B^{act} to B^{*} transition), the SF3A and SF3B complexes are destabilized and no longer interact with the branch helix (Rauhut et al., 2016; Warkocki et al., 2009; Yan et al., 2016). As the branching reaction requires the stem IIc conformation of U2 (Hilliker et al., 2007), it is possible that the stem IIc state of U2 is linked to a more "open" conformation of SF3B1, likely to be observed in the B^* and C complexes. Instead, the "closed" state of SF3B1, exhibiting high-affinity for the branch helix, is associated with the stem-loop IIa configuration of the U2 snRNA. In this respect, the "ATP-stabilization" of the pre-spliceosome (Perriman and Ares, 2000; Perriman et al., 2003), likely, involves the Prp5-dependent remodeling of the stem-loop IIa region of U2. In turn, this might promote the stable BS binding to the U2 snRNA and an induced fit transition of SF3B1 towards a "closed" conformation.

A clearer understanding of the relations between Prp5's actions, U2 and SF3B1 dynamics, and splicing modulators binding to SF3B require additional structural, biochemical, and biophysical investigations. For example, the spliceosomal A-like complexes formed in the presence of SF3B modulators can be purified from the HeLa nuclear extract, frozen in vitreous ice and imaged using cryo-EM. The relative dynamics of the U2 snRNA and the correlative changes in the conformation of SF3B1 can be accurately traced using single-molecule spectroscopy approaches (i.e., FRET, CoSMoS), as it was elegantly shown for the yeast E complex (Larson and Hoskins, 2017). Such complementary approaches would provide additional insights into spliceosome's dynamics at the earliest stages of assembly and into the biology of SF3B modulators.

4.2.5 Splicing modulators as competitive branch-site antagonists

Splicing modulators from the pladienolide and herboxidiene families bind SF3B in a conserved SF3B1-PHF5A tunnel, close to the BS adenosine (BS-A) binding pocket (Rauhut et al., 2016; Yan et al., 2016). The BS-A pocket, likely, forms during pre-spliceosome assembly and was observed both in the later pre-catalytic and activated spliceosomes (Plaschka et al., 2017; Rauhut et al., 2016; Yan et al., 2016). Structural comparisons indicate that the two pockets are formed by a similar set of residues which have different and, most importantly, mutually exclusive topologies.

In the presence of the BS-A, the invariant PHF5A-Y36 contacts and stabilizes the K1071 and the R1075 residues of SF3B1, whereas in the presence of modulators the residue interacts with the diene group of the compounds. Consistently, the conserved SF3B1-R1074 residue contacts the diene moiety of modulators, while facing the intron side of SF3B1 in spliceosomes. Moreover, our analyses suggest that the concomitant accommodation of splicing modulators and the BS-A is not possible due to steric clashes between the extended aliphatic sidechain of pladienolides/ herboxidiene and the BS-A (Cretu et al., 2018, *unpublished data*). Corroborated by previous competition studies between SF3B modulators and their inactive analogs (Effenberger et al., 2016), this allowed us to propose that splicing modulators act as competitive antagonists (Cretu et al., 2018) (**Figure 4.7**). That is, binding of the BS-A ("agonist") to the SF3B1-PHF5A tunnel in the "open" state of SF3B1 (see also 5.2.3), likely, results in an induced fit conformational transition towards the "closed" conformation, where the nucleobase is stable accommodated in the BS-A pocket. Conversely, binding of splicing modulators ("antagonists") to the same site traps SF3B1 in an "open" state which does not engage in a productive conformational transition.

Such an "agonist-antagonist" model (**Figure 4.7**) suggests that the relative inhibition of splicing by SF3B modulators would depend on: (i) the relative affinities of the "antagonist" (i.e., splicing modulators) and "agonist" (i.e., the BS-A) for the binding site; (ii) the effective concentrations of modulators and the BS-A. In metazoans, the BS sequence of the intron is only weakly conserved and exhibits different degrees of complementarity to the U2 snRNA (Gao et al., 2008; Taggart et al., 2017) (**Figure 2.2**). "Strong" BSs, compared to "weaker" BS sequences, can engage more readily the U2 snRNA to form the U2/BS duplex, where the BS-A is bulged out. Therefore, in this context, it is the entire BS region that should be viewed as "substrate" for SF3B, rather than the BS-A alone.

Accordingly, in an "agonist-antagonist" model, a "strong" splicing substrate may compete more efficiently with modulators for SF3B binding, whereas "weaker" substrates may be more sensitive to inhibition (**Figure 4.7**). Indeed, using an *in vitro* splicing system and a model pool of RNA substrates with different "strengths", we showed that the more potent BS antagonists, such as pladienolide B, exhibit an apparent BS independent inhibition profile. In contrast, "weaker" antagonists, such PB-OH, were able to inhibit efficiently splicing of pre-mRNA substrates harboring weak BSs (Cretu et al., 2018). Altogether, these experiments support our hypothesis that splicing modulators may act as competitive BS antagonists (**Figure 4.7**) and provide an explanation for their modulatory effects in cells (see also Effenberger et al., 2017). In addition, our model suggests that by tuning the affinity of modulators ("antagonists") for the SF3B complex one could achieve a more targeted effect on splicing in cells, as it was recently reported for the lead pladienolide H3B-8800 (Seiler et al., 2018).

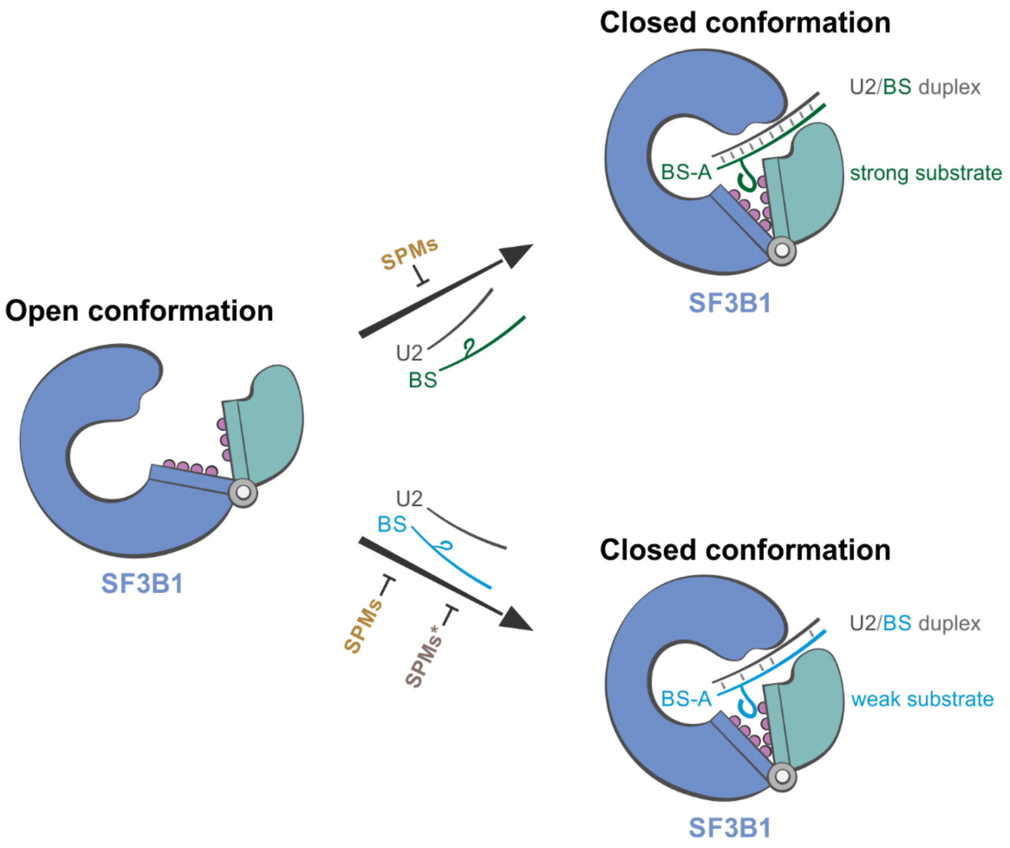


Figure 4.7. Splicing modulators as competitive branch-site antagonists.

The “agonist-antagonist” model postulates that splicing modulators (SPMs) and the BS/BS-A compete for the same “open” conformation of SF3B1. Stable binding of the branch helix results in an induced fit transition to the “closed” state. Conversely, SPM binding locks SF3B1 in an “open” conformation. The model also predicts that a “weak” antagonist (SPM*) would compete less efficiently with “strong” substrates, for SF3B1 binding, than with “weak” RNA substrates. SF3B1 is depicted as in Figure 4.6 and the two arches (i.e., H1-H15, H16-H20) of the SF3B1 HEAT superhelix are colored in blue and teal, respectively.

5 Conclusions and Outlook

The multimeric SF3B complex is an essential building block of the U2 snRNP and spliceosomes. SF3B is key to the selection and accurate recognition of the branch-site (BS) region of the intron during pre-mRNA splicing, and mutations in the SF3B1 subunit are connected to cancers. The complex is also targeted by different small-molecule compounds, “splicing modulators”, that are emerging as potential chemotherapeutic agents.

In this thesis work, we determined the crystal structures of a minimal SF3B core, comprising the HEAT domain of SF3B1, SF3B3, PHF5A, and SF3B5, in the “apo” state (Cretu et al., 2016) and in complex with different splicing modulators (Cretu et al., 2018, *unpublished data*). Our structures revealed that the SF3B3 subunit exhibits a DDB1-like triple-propeller fold. SF3B3 together with PHF5A and SF3B5 assemble a composite scaffold onto which SF3B1’s HEAT domain stably binds. SF3B1’s HEAT domain consists of 20 repeats which form a unique α-α superhelix and functions, in spliceosomes, as a “molecular clamp” to accommodate the branch duplex. Cancer-related mutations of SF3B1 affect residues from the H4-H7 repeats of the HEAT superhelix. These residues cluster around a basic surface patch and are essential for intron recognition in spliceosomes.

The crystal structures of different splicing modulators bound to SF3B revealed the general principles for their molecular recognition. That is, splicing modulators from the pladienolide and herboxidiene families bind to a conserved hourglass-shaped tunnel framed by the SF3B1 and PHF5A subunits. The chemical groups from their common pharmacophore make critical contacts with both SF3B1 and PHF5A, indicating how the specific recognition is achieved. Structural comparisons with recent cryo-EM structures suggest that these compounds bind SF3B1 in the proximity of the BS adenosine (BS-A) binding pocket, close to a tentative hinge region. As SF3B1 cannot accommodate the modulators and the BS-A at the same time, this raises the possibility that these compounds interfere with a conformational transition of SF3B1 required for the stable binding of the BS. In this respect, splicing modulators may be viewed as competitive BS antagonists. The structures also show that although these BS antagonists target the same site on SF3B, they exploit the rich chemistry of the binding tunnel in a unique manner. Overall, our crystal structures provide a framework for the future structure-based design of more stable, more soluble, and more targeted splicing modulators that could be readily available as novel therapeutic agents in cancers and, possibly, in other complex human diseases.

Even in the "structural era" of the spliceosome (Abelson, 2017), X-ray crystallography of individual modules of the complex machinery can complement and even validate the cryo-EM structures of the larger assemblies (van Roon et al., 2017). The higher resolution crystal structure of the human SF3B core complex was critical to the accurate modeling of the yeast (Rauhut et al., 2016) and human B^{act} spliceosomes (Haselbach et al., 2018) as it provided confidence in the modeled residue register (i.e., cross-validated using selenium marker sites). Individual snapshots obtained by X-ray crystallography may even provide additional context, and unexpected molecular insights as our lab recently showed for the Prp19 core complex (de Moura et al., 2018). Despite significant technical and methodological improvements (Bai et al., 2015; Nogales, 2016), it is still challenging to accurately model small-molecule ligands in the relatively low resolution and anisotropic EM density maps of spliceosomal complexes (~3.5-4.0 Å). Thus, crystallography of "druggable" spliceosomal subcomplexes, such as SF3B, may represent a better alternative.

References

Abelson, J. (2017). A close-up look at the spliceosome, at last. Proc Natl Acad Sci U S A *114*, 4288-4293.

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

Agafonov, D.E., Kastner, B., Dybkov, O., Hofele, R.V., Liu, W.T., Urlaub, H., Luhrmann, R., and Stark, H. (2016). Molecular architecture of the human U4/U6.U5 tri-snRNP. Science *351*, 1416-1420.

Alsafadi, S., Houy, A., Battistella, A., Popova, T., Wassef, M., Henry, E., Tirode, F., Constantinou, A., Piperno-Neumann, S., Roman-Roman, S., *et al.* (2016). Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. Nat Commun *7*, 10615.

Angers, S., Li, T., Yi, X., MacCoss, M.J., Moon, R.T., and Zheng, N. (2006). Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. Nature *443*, 590-593.

Ast, G. (2004). How did alternative splicing evolve? Nat Rev Genet *5*, 773-782.

Bai, R., Yan, C., Wan, R., Lei, J., and Shi, Y. (2017). Structure of the Post-catalytic Spliceosome from Saccharomyces cerevisiae. Cell *171*, 1589-1598 e1588.

Bai, X.C., McMullan, G., and Scheres, S.H. (2015). How cryo-EM is revolutionizing structural biology. Trends Biochem Sci *40*, 49-57.

Behzadnia, N., Hartmuth, K., Will, C.L., and Luhrmann, R. (2006). Functional spliceosomal A complexes can be assembled in vitro in the absence of a penta-snRNP. RNA *12*, 1738-1746.

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

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat Rev Genet *15*, 163-175.

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

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

Berglund, J.A., Fleming, M.L., and Rosbash, M. (1998). The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. RNA *4*, 998-1006.

Bertram, K., Agafonov, D.E., Dybkov, O., Haselbach, D., Leelaram, M.N., Will, C.L., Urlaub, H., Kastner, B., Luhrmann, R., and Stark, H. (2017a). Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. Cell *170*, 701-713 e711.

Bertram, K., Agafonov, D.E., Liu, W.T., Dybkov, O., Will, C.L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H., and Luhrmann, R. (2017b). Cryo-EM structure of a human spliceosome activated for step 2 of splicing. Nature *542*, 318-323.

Bessonov, S., Anokhina, M., Krasauskas, A., Golas, M.M., Sander, B., Will, C.L., Urlaub, H., Stark, H., and Luhrmann, R. (2010). Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. RNA *16*, 2384-2403.

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

Blundell, T.L., Jhoti, H., and Abell, C. (2002). High-throughput crystallography for lead discovery in drug design. Nat Rev Drug Discov *1*, 45-54.

Boesler, C., Rigo, N., Anokhina, M.M., Tauchert, M.J., Agafonov, D.E., Kastner, B., Urlaub, H., Ficner, R., Will, C.L., and Luhrmann, R. (2016). A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity. Nat Commun *7*, 11997.

Bonnal, S., Vigevani, L., and Valcarcel, J. (2012). The spliceosome as a target of novel antitumour drugs. Nat Rev Drug Discov *11*, 847-859.

Brand, M., Moggs, J.G., Oulad-Abdelghani, M., Lejeune, F., Dilworth, F.J., Stevenin, J., Almouzni, G., and Tora, L. (2001). UV-damaged DNA-binding protein in the TFTC complex links DNA damage recognition to nucleosome acetylation. EMBO J *20*, 3187-3196.

Bunker, R.D. (2016). Tackling the crystallographic structure determination of the COP9 signalosome. Acta Crystallogr D Struct Biol *72*, 326-335.

Carrocci, T.J., Zoerner, D.M., Paulson, J.C., and Hoskins, A.A. (2017). SF3b1 mutations associated with myelodysplastic syndromes alter the fidelity of branchsite selection in yeast. Nucleic acids research *45*, 4837-4852.

Casanal, A., Kumar, A., Hill, C.H., Easter, A.D., Emsley, P., Degliesposti, G., Gordiyenko, Y., Santhanam, B., Wolf, J., Wiederhold, K., *et al.* (2017). Architecture of eukaryotic mRNA 3'-end processing machinery. Science *358*, 1056-1059.

Cass, D.M., and Berglund, J.A. (2006). The SF3b155 N-terminal domain is a scaffold important for splicing. Biochemistry *45*, 10092-10101.

Cazzola, M., Invernizzi, R., Bergamaschi, G., Levi, S., Corsi, B., Travaglino, E., Rolandi, V., Biasiotto, G., Drysdale, J., and Arosio, P. (2003). Mitochondrial ferritin expression in erythroid cells from patients with sideroblastic anemia. Blood *101*, 1996-2000.

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

Champion-Arnaud, P., Gozani, O., Palandjian, L., and Reed, R. (1995). Accumulation of a novel spliceosomal complex on pre-mRNAs containing branch site mutations. Mol Cell Biol *15*, 5750-5756.

Champion-Arnaud, P., and Reed, R. (1994). The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. Genes Dev *8*, 1974-1983.

Cingolani, G., Petosa, C., Weis, K., and Muller, C.W. (1999). Structure of importin-beta bound to the IBB domain of importin-alpha. Nature *399*, 221-229.

Conte, M.R., Grune, T., Ghuman, J., Kelly, G., Ladas, A., Matthews, S., and Curry, S. (2000). Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. EMBO J *19*, 3132-3141.

Conti, E., Muller, C.W., and Stewart, M. (2006). Karyopherin flexibility in nucleocytoplasmic transport. Curr Opin Struct Biol *16*, 237-244.

Cook, A., Fernandez, E., Lindner, D., Ebert, J., Schlenstedt, G., and Conti, E. (2005). The structure of the nuclear export receptor Cse1 in its cytosolic state reveals a closed conformation incompatible with cargo binding. Mol Cell *18*, 355-367.

Cordin, O., Hahn, D., and Beggs, J.D. (2012). Structure, function and regulation of spliceosomal RNA helicases. Curr Opin Cell Biol *24*, 431-438.

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.

Corsini, L., Bonnal, S., Basquin, J., Hothorn, M., Scheffzek, K., Valcarcel, J., and Sattler, M. (2007). U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. Nat Struct Mol Biol *14*, 620-629.

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.

Cretu, C., Schmitzova, J., Ponce-Salvatierra, A., Dybkov, O., De Laurentiis, E.I., Sharma, K., Will, C.L., Urlaub, H., Luhrmann, R., and Pena, V. (2016). Molecular Architecture of SF3b and Structural Consequences of Its Cancer-Related Mutations. Mol Cell *64*, 307-319.

Cretu, C., Agrawal, A.A., Cook, A., Will, C.L, Fekkes, P., Smith, P.G., Luhrmann, R., Larsen, N., Buonamici, S., Pena, V. (2018). Structural basis of splicing modulation by antitumor macrolide compounds. Mol Cell, *in press*, DOI: 10.1016/j.molcel.2018.03.011.

Darman, R.B., Seiler, M., Agrawal, A.A., Lim, K.H., Peng, S., Aird, D., Bailey, S.L., Bhavsar, E.B., Chan, B., Colla, S., *et al.* (2015). Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point. Cell Rep *13*, 1033-1045.

Das, R., Zhou, Z., and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. Mol Cell *5*, 779-787.

Daubner, G.M., Clery, A., Jayne, S., Stevenin, J., and Allain, F.H. (2012). A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. EMBO J *31*, 162-174.

De, I., Bessonov, S., Hofele, R., dos Santos, K., Will, C.L., Urlaub, H., Luhrmann, R., and Pena, V. (2015). The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes. Nat Struct Mol Biol *22*, 138-144.

de Moura, T.R., Mozaffari-Jovin, S., Szabo, C.Z.K., Schmitzova, J., Dybkov, O., Cretu, C., Kachala, M., Svergun, D., Urlaub, H., Luhrmann, R., and Pena, V. (2018). Prp19/Pso4 Is an Autoinhibited Ubiquitin Ligase Activated by Stepwise Assembly of Three Splicing Factors. Mol Cell *69*, 979-992 e976.

DeBoever, C., Ghia, E.M., Shepard, P.J., Rassenti, L., Barrett, C.L., Jepsen, K., Jamieson, C.H., Carson, D., Kipps, T.J., 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.

Deutsch, M., and Long, M. (1999). Intron-exon structures of eukaryotic model organisms. Nucleic acids research *27*, 3219-3228.

Didychuk, A.L., Butcher, S.E., and Brow, D.A. (2018). The life of U6 small nuclear RNA, from cradle to grave. RNA *24*, 437-460.

Dolatshad, H., Pellagatti, A., Liberante, F.G., Llorian, M., Repapi, E., Steeples, V., Roy, S., Scifo, L., Armstrong, R.N., Shaw, J., *et al.* (2016). Cryptic splicing events in the iron transporter ABCB7 and other key target genes in SF3B1-mutant myelodysplastic syndromes. Leukemia *30*, 2322-2331.

Dvinge, H., and Bradley, R.K. (2015). Widespread intron retention diversifies most cancer transcriptomes. Genome medicine *7*, 45.

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.

Dybkov, O., Will, C.L., Deckert, J., Behzadnia, N., Hartmuth, K., and Luhrmann, R. (2006). U2 snRNA-protein contacts in purified human 17S U2 snRNPs and in spliceosomal A and B complexes. Mol Cell Biol *26*, 2803-2816.

Effenberger, K.A., Urabe, V.K., and Jurica, M.S. (2017). Modulating splicing with small molecular inhibitors of the spliceosome. Wiley Interdiscip Rev RNA *8*.

Effenberger, K.A., Urabe, V.K., Prichard, B.E., Ghosh, A.K., and Jurica, M.S. (2016). Interchangeable SF3B1 inhibitors interfere with pre-mRNA splicing at multiple stages. RNA *22*, 350-359.

Erlanson, D.A., Fesik, S.W., Hubbard, R.E., Jahnke, W., and Jhoti, H. (2016). Twenty years on: the impact of fragments on drug discovery. Nat Rev Drug Discov *15*, 605-619.

Eskens, F.A., Ramos, F.J., Burger, H., O'Brien, J.P., Piera, A., de Jonge, M.J., Mizui, Y., Wiemer, E.A., Carreras, M.J., Baselga, J., *et al.* (2013). Phase I pharmacokinetic and pharmacodynamic study of the first-in-class spliceosome inhibitor E7107 in patients with advanced solid tumors. Clinical cancer research : an official journal of the American Association for Cancer Research *19*, 6296-6304.

Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Luhrmann, R. (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. Mol Cell *36*, 593-608.

Fedor, M.J., and Williamson, J.R. (2005). The catalytic diversity of RNAs. Nat Rev Mol Cell Biol *6*, 399-412.

Fica, S.M., and Nagai, K. (2017). Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. Nat Struct Mol Biol *24*, 791-799.

Fischer, E.S., Scrima, A., Bohm, K., Matsumoto, S., Lingaraju, G.M., Faty, M., Yasuda, T., Cavadini, S., Wakasugi, M., Hanaoka, F., *et al.* (2011). The molecular basis of CRL4DDB2/CSA ubiquitin ligase architecture, targeting, and activation. Cell *147*, 1024-1039.

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.

Fox-Walsh, K.L., Dou, Y., Lam, B.J., Hung, S.P., Baldi, P.F., and Hertel, K.J. (2005). The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc Natl Acad Sci U S A *102*, 16176-16181.

Frilander, M.J., and Steitz, J.A. (1999). Initial recognition of U12-dependent introns requires both U11/5' splice-site and U12/branchpoint interactions. Genes Dev *13*, 851-863.

Fukumoto, Y., Dohmae, N., and Hanaoka, F. (2008). Schizosaccharomyces pombe Ddb1 recruits substrate-specific adaptor proteins through a novel protein motif, the DDB-box. Mol Cell Biol *28*, 6746-6756.

Galej, W.P., Oubridge, C., Newman, A.J., and Nagai, K. (2013). Crystal structure of Prp8 reveals active site cavity of the spliceosome. Nature *493*, 638-643.

Galej, W.P., Wilkinson, M.E., Fica, S.M., Oubridge, C., Newman, A.J., and Nagai, K. (2016). Cryo-EM structure of the spliceosome immediately after branching. Nature *537*, 197-201.

Gao, K., Masuda, A., Matsuura, T., and Ohno, K. (2008). Human branch point consensus sequence is yUnAy. Nucleic acids research *36*, 2257-2267.

Garcia-Blanco, M.A., Jamison, S.F., and Sharp, P.A. (1989). Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. Genes Dev *3*, 1874-1886.

Girard, C., Will, C.L., Peng, J., Makarov, E.M., Kastner, B., Lemm, I., Urlaub, H., Hartmuth, K., and Luhrmann, R. (2012). Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. Nat Commun *3*, 994.

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.

Golas, M.M., Sander, B., Will, C.L., Luhrmann, R., and Stark, H. (2005). Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. Mol Cell *17*, 869-883.

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. Molecular and cellular biology *18*, 4752-4760.

Grinthal, A., Adamovic, I., Weiner, B., Karplus, M., and Kleckner, N. (2010). PR65, the HEAT-repeat scaffold of phosphatase PP2A, is an elastic connector that links force and catalysis. Proc Natl Acad Sci U S A *107*, 2467-2472.

Gupta, A., Jenkins, J.L., and Kielkopf, C.L. (2011). RNA induces conformational changes in the SF1/U2AF65 splicing factor complex. J Mol Biol *405*, 1128-1138.

Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999). Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. Nature *398*, 579-585.

Harlen, K.M., and Churchman, L.S. (2017). The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. Nat Rev Mol Cell Biol *18*, 263-273.

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

Hasegawa, M., Miura, T., Kuzuya, K., Inoue, A., Won Ki, S., Horinouchi, S., Yoshida, T., Kunoh, T., Koseki, K., Mino, K., *et al.* (2011). Identification of SAP155 as the target of GEX1A (Herboxidiene), an antitumor natural product. ACS Chem Biol *6*, 229-233.

Haselbach, D., Komarov, I., Agafonov, D.E., Hartmuth, K., Graf, B., Dybkov, O., Urlaub, H., Kastner, B., Luhrmann, R., and Stark, H. (2018). Structure and Conformational Dynamics of the Human Spliceosomal B(act) Complex. Cell *172*, 454-464 e411.

Hilliker, A.K., Mefford, M.A., and Staley, J.P. (2007). U2 toggles iteratively between the stem IIa and stem IIc conformations to promote pre-mRNA splicing. Genes Dev *21*, 821-834.

Hirose, Y., and Manley, J.L. (2000). RNA polymerase II and the integration of nuclear events. Genes Dev *14*, 1415-1429.

Hoelz, A., Glavy, J.S., and Beck, M. (2016). Toward the atomic structure of the nuclear pore complex: when top down meets bottom up. Nat Struct Mol Biol *23*, 624-630.

Hollander, D., Naftelberg, S., Lev-Maor, G., Kornblihtt, A.R., and Ast, G. (2016). How Are Short Exons Flanked by Long Introns Defined and Committed to Splicing? Trends Genet *32*, 596-606.

Hong, W., Bennett, M., Xiao, Y., Feld Kramer, R., Wang, C., and Reed, R. (1997). Association of U2 snRNP with the spliceosomal complex E. Nucleic acids research *25*, 354-361.

Igel, H., Wells, S., Perriman, R., and Ares, M., Jr. (1998). Conservation of structure and subunit interactions in yeast homologues of splicing factor 3b (SF3b) subunits. RNA *4*, 1-10.

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.

International Human Genome Sequencing, C. (2004). Finishing the euchromatic sequence of the human genome. Nature *431*, 931-945.

Irimia, M., and Roy, S.W. (2014). Origin of spliceosomal introns and alternative splicing. Cold Spring Harb Perspect Biol *6*.

Jackson, S., and Xiong, Y. (2009). CRL4s: the CUL4-RING E3 ubiquitin ligases. Trends Biochem Sci *34*, 562-570.

Jenkins, J.L., and Kielkopf, C.L. (2017). Splicing Factor Mutations in Myelodysplasias: Insights from Spliceosome Structures. Trends Genet *33*, 336-348.

Jubb, H., Higueruelo, A.P., Winter, A., and Blundell, T.L. (2012). Structural biology and drug discovery for protein-protein interactions. Trends Pharmacol Sci *33*, 241-248.

Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., Watanabe, H., Kitahara, T., Yoshida, T., Nakajima, H., *et al.* (2007). Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. Nat Chem Biol *3*, 576-583.

Kappel, C., Zachariae, U., Dolker, N., and Grubmuller, H. (2010). An unusual hydrophobic core confers extreme flexibility to HEAT repeat proteins. Biophys J *99*, 1596-1603.

Kent, O.A., Reayi, A., Foong, L., Chilibeck, K.A., and MacMillan, A.M. (2003). Structuring of the 3' splice site by U2AF65. J Biol Chem *278*, 50572-50577.

Kesarwani, A.K., Ramirez, O., Gupta, A.K., Yang, X., Murthy, T., Minella, A.C., and Pillai, M.M. (2017). Cancer-associated SF3B1 mutants recognize otherwise inaccessible cryptic 3' splice sites within RNA secondary structures. Oncogene *36*, 1123-1133.

Kfir, N., Lev-Maor, G., Glaich, O., Alajem, A., Datta, A., Sze, S.K., Meshorer, E., and Ast, G. (2015). SF3B1 association with chromatin determines splicing outcomes. Cell Rep *11*, 618-629.

Khatter, H., Vorlander, M.K., and Muller, C.W. (2017). RNA polymerase I and III: similar yet unique. Curr Opin Struct Biol *47*, 88-94.

Kielkopf, C.L., Rodionova, N.A., Green, M.R., and Burley, S.K. (2001). A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. Cell *106*, 595-605.

Kim, E., Ilagan, J.O., Liang, Y., Daubner, G.M., Lee, S.C., Ramakrishnan, A., Li, Y., Chung, Y.R., Micol, J.B., Murphy, M.E., *et al.* (2015). SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer Cell *27*, 617-630.

Kondo, Y., Oubridge, C., van Roon, A.M., and Nagai, K. (2015). Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. Elife *4*.

Kotake, Y., Sagane, K., Owa, T., Mimori-Kiyosue, Y., Shimizu, H., Uesugi, M., Ishihama, Y., Iwata, M., and Mizui, Y. (2007). Splicing factor SF3b as a target of the antitumor natural product pladienolide. Nat Chem Biol *3*, 570-575.

Kramer, A., Gruter, P., Groning, K., and Kastner, B. (1999). Combined biochemical and electron microscopic analyses reveal the architecture of the mammalian U2 snRNP. J Cell Biol *145*, 1355-1368.

Lagisetti, C., Palacios, G., Goronga, T., Freeman, B., Caufield, W., and Webb, T.R. (2013). Optimization of antitumor modulators of pre-mRNA splicing. J Med Chem *56*, 10033-10044.

Lagisetti, 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.

Lagisetti, C., Yermolina, M.V., Sharma, L.K., Palacios, G., Prigaro, B.J., and Webb, T.R. (2014). Pre-mRNA splicing-modulatory pharmacophores: the total synthesis of herboxidiene, a pladienolide-herboxidiene hybrid analog and related derivatives. ACS Chem Biol *9*, 643-648.

Larson, J.D., and Hoskins, A.A. (2017). Dynamics and consequences of spliceosome E complex formation. Elife *6*.

Lee, J., and Zhou, P. (2007). DCAFs, the missing link of the CUL4-DDB1 ubiquitin ligase. Mol Cell *26*, 775-780.

Lee, S.C., Dvinge, H., Kim, E., Cho, H., Micol, J.B., Chung, Y.R., Durham, B.H., Yoshimi, A., Kim, Y.J., Thomas, M., *et al.* (2016). Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. Nat Med *22*, 672-678.

Lee, S.J., Matsuura, Y., Liu, S.M., and Stewart, M. (2005). Structural basis for nuclear import complex dissociation by RanGTP. Nature *435*, 693-696.

Lee, Y., and Rio, D.C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu Rev Biochem *84*, 291-323.

Li, J., Leung, A.K., Kondo, Y., Oubridge, C., and Nagai, K. (2016). Re-refinement of the spliceosomal U4 snRNP core-domain structure. Acta Crystallogr D Struct Biol *72*, 131-146.

Li, T., Chen, X., Garbutt, K.C., Zhou, P., and Zheng, N. (2006). Structure of DDB1 in complex with a paramyxovirus V protein: viral hijack of a propeller cluster in ubiquitin ligase. Cell *124*, 105-117.

Liao, X.L., Kretzner, L., Seraphin, B., and Rosbash, M. (1990). Universally conserved and yeast-specific U1 snRNA sequences are important but not essential for U1 snRNP function. Genes Dev *4*, 1766-1774.

Libri, D., Graziani, N., Saguez, C., and Boulay, J. (2001). Multiple roles for the yeast SUB2/yUAP56 gene in splicing. Genes Dev *15*, 36-41.

Liu, S., Li, X., Zhang, L., Jiang, J., Hill, R.C., Cui, Y., Hansen, K.C., Zhou, Z.H., and Zhao, R. (2017). Structure of the yeast spliceosomal postcatalytic P complex. Science *358*, 1278-1283.

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

Loerch, S., and Kielkopf, C.L. (2016). Unmasking the U2AF homology motif family: a bona fide protein-protein interaction motif in disguise. RNA *22*, 1795-1807.

Loerch, S., Maucuer, A., Manceau, V., Green, M.R., and Kielkopf, C.L. (2014). Cancer-relevant splicing factor CAPERalpha engages the essential splicing factor SF3b155 in a specific ternary complex. J Biol Chem *289*, 17325-17337.

Mackereth, C.D., Simon, B., and Sattler, M. (2005). Extending the size of protein-RNA complexes studied by nuclear magnetic resonance spectroscopy. Chembiochem *6*, 1578-1584.

MacMillan, A.M., Query, C.C., Allerson, C.R., Chen, S., Verdine, G.L., and Sharp, P.A. (1994). Dynamic association of proteins with the pre-mRNA branch region. Genes Dev *8*, 3008-3020.

Madan, V., Kanojia, D., Li, J., Okamoto, R., Sato-Otsubo, A., Kohlmann, A., Sanada, M., Grossmann, V., Sundaresan, J., Shiraishi, Y., *et al.* (2015). Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. Nat Commun *6*, 6042.

Makowski, K., Vigevani, L., Albericio, F., Valcarcel, J., and Alvarez, M. (2017). Sudemycin K: A Synthetic Antitumor Splicing Inhibitor Variant with Improved Activity and Versatile Chemistry. ACS Chem Biol *12*, 163-173.

Martinez, E., Palhan, V.B., Tjernberg, A., Lymar, E.S., Gamper, A.M., Kundu, T.K., Chait, B.T., 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.

Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. Nat Rev Mol Cell Biol *15*, 108-121.
McGinty, R.K., and Tan, S. (2015). Nucleosome structure and function. Chem Rev *115*, 2255-2273.
Menon, S., Tsuge, T., Dohmae, N., Takio, K., and Wei, N. (2008). Association of SAP130/SF3b-3 with Cullin-RING ubiquitin ligase complexes and its regulation by the COP9 signalosome. BMC Biochem *9*, 1.
Michaud, S., and Reed, R. (1991). An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. Genes Dev *5*, 2534-2546.
Monecke, T., Dickmanns, A., and Ficner, R. (2014). Allosteric control of the exportin CRM1 unraveled by crystal structure analysis. FEBS J *281*, 4179-4194.
Murray, C.W., Verdonk, M.L., and Rees, D.C. (2012). Experiences in fragment-based drug discovery. Trends Pharmacol Sci *33*, 224-232.
Neuenkirchen, N., Chari, A., and Fischer, U. (2008). Deciphering the assembly pathway of Sm-class U snRNPs. FEBS Lett *582*, 1997-2003.
Nguyen, T.H.D., Galej, W.P., Bai, X.C., Oubridge, C., Newman, A.J., Scheres, S.H.W., and Nagai, K. (2016). Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 A resolution. Nature *530*, 298-302.
Nogales, E. (2016). The development of cryo-EM into a mainstream structural biology technique. Nat Methods *13*, 24-27.
Norton, P.A. (1994). Polypyrimidine tract sequences direct selection of alternative branch sites and influence protein binding. Nucleic acids research *22*, 3854-3860.
Obeng, E.A., Chappell, R.J., Seiler, M., Chen, M.C., Campagna, D.R., Schmidt, P.J., Schneider, R.K., Lord, A.M., Wang, L., Gambe, R.G., *et al.* (2016). Physiologic Expression of Sf3b1(K700E) Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation. Cancer Cell *30*, 404-417.
Oberstrass, F.C., Auweter, S.D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D.L., *et al.* (2005). Structure of PTB bound to RNA: specific binding and implications for splicing regulation. Science *309*, 2054-2057.
Papaemmanuil, E., Cazzola, M., Boultwood, J., Malcovati, L., Vyas, P., Bowen, D., Pellagatti, A., Wainscoat, J.S., Hellstrom-Lindberg, E., Gambacorti-Passerini, C., *et al.* (2011). Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. The New England journal of medicine *365*, 1384-1395.
Patel, A.A., and Steitz, J.A. (2003). Splicing double: insights from the second spliceosome. Nat Rev Mol Cell Biol *4*, 960-970.
Pauling, M.H., McPheeters, D.S., and Ares, M., Jr. (2000). Functional Cus1p is found with Hsh155p in a multiprotein splicing factor associated with U2 snRNA. Mol Cell Biol *20*, 2176-2185.
Peled-Zehavi, H., Berglund, J.A., Rosbash, M., and Frankel, A.D. (2001). Recognition of RNA branch point sequences by the KH domain of splicing factor 1 (mammalian branch point binding protein) in a splicing factor complex. Mol Cell Biol *21*, 5232-5241.
Perea, W., Schroeder, K.T., Bryant, A.N., and Greenbaum, N.L. (2016). Interaction between the Spliceosomal Pre-mRNA Branch Site and U2 snRNP Protein p14. Biochemistry *55*, 629-632.
Perriman, R., and Ares, M., Jr. (2000). ATP can be dispensable for prespliceosome formation in yeast. Genes Dev *14*, 97-107.

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., Barta, I., Voeltz, G.K., Abelson, J., and Ares, M., Jr. (2003). ATP requirement for Prp5p function is determined by Cus2p and the structure of U2 small nuclear RNA. Proc Natl Acad Sci U S A *100*, 13857-13862.

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.

Plaschka, C., Lin, P.C., and Nagai, K. (2017). Structure of a pre-catalytic spliceosome. Nature *546*, 617-621.

Pyle, A.M. (2016). Group II Intron Self-Splicing. Annu Rev Biophys *45*, 183-205.

Query, C.C., Moore, M.J., and Sharp, P.A. (1994). Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. Genes Dev *8*, 587-597.

Query, C.C., Strobel, S.A., and Sharp, P.A. (1996). Three recognition events at the branch-site adenine. EMBO J *15*, 1392-1402.

Quesada, V., Conde, L., Villamor, N., Ordonez, G.R., Jares, P., Bassaganyas, L., Ramsay, A.J., Bea, S., Pinyol, M., Martinez-Trillos, A., *et al.* (2011). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nature genetics *44*, 47-52.

Rauhut, R., Fabrizio, P., Dybkov, O., Hartmuth, K., Pena, V., Chari, A., Kumar, V., Lee, C.T., Urlaub, H., Kastner, B., *et al.* (2016). Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. Science *353*, 1399-1405.

Reed, R. (1989). The organization of 3' splice-site sequences in mammalian introns. Genes Dev *3*, 2113-2123.

Roscigno, R.F., Weiner, M., and Garcia-Blanco, M.A. (1993). A mutational analysis of the polypyrimidine tract of introns. Effects of sequence differences in pyrimidine tracts on splicing. J Biol Chem *268*, 11222-11229.

Roy, S.W., and Gilbert, W. (2006). The evolution of spliceosomal introns: patterns, puzzles and progress. Nat Rev Genet *7*, 211-221.

Roybal, G.A., and Jurica, M.S. (2010). Spliceostatin A inhibits spliceosome assembly subsequent to prespliceosome formation. Nucleic acids research *38*, 6664-6672.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nat Rev Mol Cell Biol *16*, 129-143.

Saldi, T., Cortazar, M.A., Sheridan, R.M., and Bentley, D.L. (2016). Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. J Mol Biol *428*, 2623-2635.

Salton, M., and Misteli, T. (2016). Small Molecule Modulators of Pre-mRNA Splicing in Cancer Therapy. Trends Mol Med *22*, 28-37.

Schellenberg, M.J., Dul, E.L., and MacMillan, A.M. (2011). Structural model of the p14/SF3b155 . branch duplex complex. RNA *17*, 155-165.

Schellenberg, M.J., Edwards, R.A., Ritchie, D.B., Kent, O.A., Golas, M.M., Stark, H., Luhrmann, R., Glover, J.N., and MacMillan, A.M. (2006). Crystal structure of a core spliceosomal protein interface. Proc Natl Acad Sci U S A *103*, 1266-1271.

Schlichting, I. (2015). Serial femtosecond crystallography: the first five years. IUCrJ *2*, 246-255.

Schneider, C., Agafonov, D.E., Schmitzova, J., Hartmuth, K., Fabrizio, P., and Luhrmann, R. (2015). Dynamic Contacts of U2, RES, Cwc25, Prp8 and Prp45 Proteins with the Pre-mRNA Branch-Site and 3' Splice Site during Catalytic Activation and Step 1 Catalysis in Yeast Spliceosomes. PLoS Genet *11*, e1005539.

Scotti, M.M., and Swanson, M.S. (2016). RNA mis-splicing in disease. Nat Rev Genet *17*, 19-32.

Scrima, A., Fischer, E.S., Lingaraju, G.M., Bohm, K., Cavadini, S., and Thoma, N.H. (2011). Detecting UV-lesions in the genome: The modular CRL4 ubiquitin ligase does it best! FEBS Lett *585*, 2818-2825.

Seiler, M., Yoshimi, A., Darman, R., Chan, B., Keaney, G., Thomas, M., Agrawal, A.A., Caleb, B., Csibi, A., Sean, E., *et al.* (2018). H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. Nat Med.

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A., and Sattler, M. (2003). Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. Mol Cell *11*, 965-976.

Seraphin, B., Kretzner, L., and Rosbash, M. (1988). A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. EMBO J *7*, 2533-2538.

Shao, W., Kim, H.S., Cao, Y., Xu, Y.Z., and Query, C.C. (2012). A U1-U2 snRNP interaction network during intron definition. Mol Cell Biol *32*, 470-478.

Shen, H., Kan, J.L., and Green, M.R. (2004). Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. Mol Cell *13*, 367-376.

Shi, Y. (2017). Mechanistic insights into precursor messenger RNA splicing by the spliceosome. Nat Rev Mol Cell Biol *18*, 655-670.

Shi, Y., Reddy, B., and Manley, J.L. (2006). PP1/PP2A phosphatases are required for the second step of Pre-mRNA splicing and target specific snRNP proteins. Mol Cell *23*, 819-829.

Shirai, C.L., Ley, J.N., White, B.S., Kim, S., Tibbitts, J., Shao, J., Ndonwi, M., Wadugu, B., Duncavage, E.J., Okeyo-Owuor, T., *et al.* (2015). Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. Cancer Cell *27*, 631-643.

Shirai, C.L., White, B.S., Tripathi, M., Tapia, R., Ley, J.N., Ndonwi, M., Kim, S., Shao, J., Carver, A., Saez, B., *et al.* (2017). Mutant U2AF1-expressing cells are sensitive to pharmacological modulation of the spliceosome. Nat Commun *8*, 14060.

Sickmier, E.A., Frato, K.E., Shen, H., Paranawithana, S.R., Green, M.R., and Kielkopf, C.L. (2006). Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. Mol Cell *23*, 49-59.

Spadaccini, R., Reidt, U., Dybkov, O., Will, C., Frank, R., Stier, G., Corsini, L., Wahl, M.C., Luhrmann, R., and Sattler, M. (2006). Biochemical and NMR analyses of an SF3b155-p14-U2AF-RNA interaction network involved in branch point definition during pre-mRNA splicing. RNA *12*, 410-425.

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

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.

Stewart, M. (2006). Structural basis for the nuclear protein import cycle. Biochem Soc Trans *34*, 701-704.

Taggart, A.J., Lin, C.L., Shrestha, B., Heintzelman, C., Kim, S., and Fairbrother, W.G. (2017). Large-scale analysis of branchpoint usage across species and cell lines. Genome Res *27*, 639-649.

Tang, Q., Rodriguez-Santiago, S., Wang, J., Pu, J., Yuste, A., Gupta, V., Moldon, A., Xu, Y.Z., and Query, C.C. (2016). SF3B1/Hsh155 HEAT motif mutations affect interaction with the spliceosomal ATPase Prp5, resulting in altered branch site selectivity in pre-mRNA splicing. Genes Dev *30*, 2710-2723.

Teigelkamp, S., Mundt, C., Achsel, T., Will, C.L., and Luhrmann, R. (1997). The human U5 snRNP-specific 100-kD protein is an RS domain-containing, putative RNA helicase with significant homology to the yeast splicing factor Prp28p. RNA *3*, 1313-1326.

Teng, T., Tsai, J.H., Puyang, X., Seiler, M., Peng, S., Prajapati, S., Aird, D., Buonamici, S., Caleb, B., Chan, B., *et al.* (2017). Splicing modulators act at the branch point adenosine binding pocket defined by the PHF5A-SF3b complex. Nat Commun *8*, 15522.

Thickman, K.R., Swenson, M.C., Kabogo, J.M., Gryczynski, Z., and Kielkopf, C.L. (2006). Multiple U2AF65 binding sites within SF3b155: thermodynamic and spectroscopic characterization of protein-protein interactions among pre-mRNA splicing factors. J Mol Biol *356*, 664-683.

Uhlmann-Schiffler, H., Jalal, C., and Stahl, H. (2006). Ddx42p--a human DEAD box protein with RNA chaperone activities. Nucleic acids research *34*, 10-22.

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

van Roon, A.M., Loening, N.M., Obayashi, E., Yang, J.C., Newman, A.J., Hernandez, H., Nagai, K., and Neuhaus, D. (2008). Solution structure of the U2 snRNP protein Rds3p reveals a knotted zinc-finger motif. Proc Natl Acad Sci U S A *105*, 9621-9626.

van Roon, A.M., Oubridge, C., Obayashi, E., Sposito, B., Newman, A.J., Seraphin, B., and Nagai, K. (2017). Crystal structure of U2 snRNP SF3b components: Hsh49p in complex with Cus1p-binding domain. RNA *23*, 968-981.

Vannini, A., and Cramer, P. (2012). Conservation between the RNA polymerase I, II, and III transcription initiation machineries. Mol Cell *45*, 439-446.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., *et al.* (2001). The sequence of the human genome. Science *291*, 1304-1351.

Vigevani, L., Gohr, A., Webb, T., Irimia, M., and Valcarcel, J. (2017). Molecular basis of differential 3' splice site sensitivity to anti-tumor drugs targeting U2 snRNP. Nat Commun *8*, 2100.

Villa, R., Kashyap, M.K., Kumar, D., Kipps, T.J., Castro, J.E., La Clair, J.J., and Burkart, M.D. (2013). Stabilized cyclopropane analogs of the splicing inhibitor FD-895. J Med Chem *56*, 6576-6582.

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

Wan, R., Yan, C., Bai, R., Huang, G., and Shi, Y. (2016a). Structure of a yeast catalytic step I spliceosome at 3.4 A resolution. Science *353*, 895-904.

Wan, R., Yan, C., Bai, R., Lei, J., and Shi, Y. (2017). Structure of an Intron Lariat Spliceosome from Saccharomyces cerevisiae. Cell *171*, 120-132 e112.

Wan, R., Yan, C., Bai, R., Wang, L., Huang, M., Wong, C.C., and Shi, Y. (2016b). The 3.8 A structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. Science *351*, 466-475.

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

Wang, Q., He, J., Lynn, B., and Rymond, B.C. (2005). Interactions of the yeast SF3b splicing factor. Molecular and cellular biology *25*, 10745-10754.

Wang, Q., and Rymond, B.C. (2003). Rds3p is required for stable U2 snRNP recruitment to the splicing apparatus. Molecular and cellular biology *23*, 7339-7349.

Wang, W., Maucuer, A., Gupta, A., Manceau, V., Thickman, K.R., Bauer, W.J., Kennedy, S.D., Wedekind, J.E., Green, M.R., and Kielkopf, C.L. (2013). Structure of phosphorylated SF1 bound to U2AF(6)(5) in an essential splicing factor complex. Structure *21*, 197-208.

Warkocki, Z., Odenwalder, P., Schmitzova, J., Platzmann, F., Stark, H., Urlaub, H., Ficner, R., Fabrizio, P., and Luhrmann, R. (2009). Reconstitution of both steps of Saccharomyces cerevisiae splicing with purified spliceosomal components. Nat Struct Mol Biol *16*, 1237-1243.

Warkocki, Z., Schneider, C., Mozaffari-Jovin, S., Schmitzova, J., Hobartner, C., Fabrizio, P., and Luhrmann, R. (2015). The G-patch protein Spp2 couples the spliceosome-stimulated ATPase activity of the DEAH-box protein Prp2 to catalytic activation of the spliceosome. Genes Dev *29*, 94-107.

Warner, J.R. (1999). The economics of ribosome biosynthesis in yeast. Trends Biochem Sci *24*, 437-440.

Watson, J.D., Berry, A.J., and Davies, K. (2017). DNA : the story of the genetic revolution, Revised and updated.Second edition. edn (New York: Alfred A. Knopf).

Wilkinson, M.E., Fica, S.M., Galej, W.P., Norman, C.M., Newman, A.J., and Nagai, K. (2017). Postcatalytic spliceosome structure reveals mechanism of 3'-splice site selection. Science *358*, 1283-1288.

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

Will, C.L., Schneider, C., MacMillan, A.M., Katopodis, N.F., Neubauer, G., Wilm, M., Luhrmann, R., and Query, C.C. (2001). A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. Embo J *20*, 4536-4546.

Will, C.L., Urlaub, H., Achsel, T., Gentzel, M., Wilm, M., and Luhrmann, R. (2002). Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. Embo J *21*, 4978-4988.

Wu, J., and Manley, J.L. (1989). Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. Genes Dev *3*, 1553-1561.

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

Xu, Y.Z., Newnham, C.M., Kameoka, S., Huang, T., Konarska, M.M., and Query, C.C. (2004). Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. EMBO J *23*, 376-385.

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 A resolution. Science *353*, 904-911.

Yan, D., Perriman, R., Igel, H., Howe, K.J., Neville, M., and Ares, M., Jr. (1998). CUS2, a yeast homolog of human Tat-SF1, rescues function of misfolded U2 through an unusual RNA recognition motif. Mol Cell Biol *18*, 5000-5009.

Yoshida, H., Park, S.Y., Oda, T., Akiyoshi, T., Sato, M., Shirouzu, M., Tsuda, K., Kuwasako, K., Unzai, S., Muto, Y., *et al.* (2015). A novel 3' splice site recognition by the two zinc fingers in the U2AF small subunit. Genes Dev *29*, 1649-1660.

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., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., *et al.* (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. Nature *478*, 64-69.

Yoshimura, S.H., and Hirano, T. (2016). HEAT repeats - versatile arrays of amphiphilic helices working in crowded environments? J Cell Sci *129*, 3963-3970.

Zachariae, U., and Grubmuller, H. (2008). Importin-beta: structural and dynamic determinants of a molecular spring. Structure *16*, 906-915.

Zavanelli, M.I., Britton, J.S., Igel, A.H., and Ares, M., Jr. (1994). Mutations in an essential U2 small nuclear RNA structure cause cold-sensitive U2 small nuclear ribonucleoprotein function by favoring competing alternative U2 RNA structures. Mol Cell Biol *14*, 1689-1697.

Zhang, X., Yan, C., Zhan, X., Li, L., Lei, J., and Shi, Y. (2018). Structure of the human activated spliceosome in three conformational states. Cell Res *28*, 307-322.

Zhao, C., and Pyle, A.M. (2017). Structural Insights into the Mechanism of Group II Intron Splicing. Trends Biochem Sci *42*, 470-482.

Zhou, L., Hang, J., Zhou, Y., Wan, R., Lu, G., Yin, P., Yan, C., and Shi, Y. (2014). Crystal structures of the Lsm complex bound to the 3' end sequence of U6 small nuclear RNA. Nature *506*, 116-120.

Zhuang, Y., and Weiner, A.M. (1989). A compensatory base change in human U2 snRNA can suppress a branch site mutation. Genes Dev *3*, 1545-1552.

Appendix

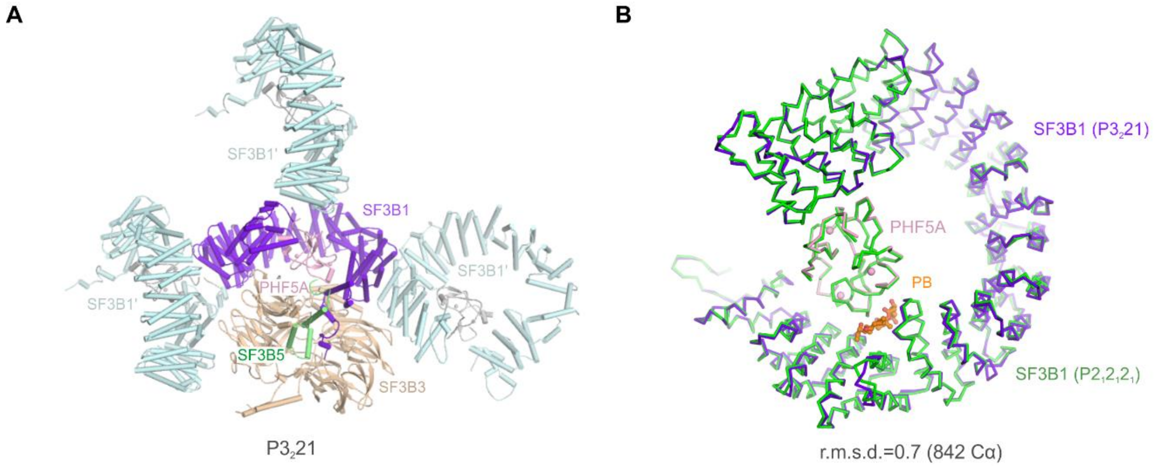


Figure S1. SF3B1 adopts the same "open" conformation in two different crystal forms.

(**A**) Crystal packing of SF3B1 in the $P3_221$ crystal form. An alternative crystal form for a SF3B core complex was obtained using the very same fragments of SF3B1 (453-1304) and PHF5A (1-98), but lacking the SF3B3 BPB domain (443-767). These crystals formed under different conditions (Cretu C., Pena V., *unpublished data*) and belonged to a higher-symmetry space group ($P3_221$). SF3B1 is colored purple and the symmetry-related SF3B1 and PHF5A are colored light blue and gray, respectively. (**B**) Structural superposition of the SF3B1-PHF5A module crystallized in two different space groups. SF3B1 (purple) exhibits the same "open" conformation in $P3_221$ as in the $P2_12_12_1$ crystal form (green), complexed with pladienolide B (PB, orange).

Table S1. Data collection and refinement statistics for the unpublished crystal structures.

	$\mathrm{SF3B}^{\Delta\mathrm{BPB}}$	$\mathrm{SF3B}^{\Delta\mathrm{BPB}}$-HB	$\mathrm{SF3B}^{\Delta\mathrm{BPB}}$-H3B-8800	$\mathrm{SF3B}^{\mathrm{core}}$-E7107
Data collection				
Space group	$P3_221$	$P3_221$	$C121$	$P2_12_12_1$
a, b, c (Å)	108.05, 108.05, 359.48	107.85, 107.85, 359.17	184.93, 110.43, 354.45	105.11, 154.46, 210.35
α, β, γ (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 92.23, 90.00	90.00, 90.00, 90.00
Wavelength	1.0	1.0	1.0	1.0
Resolution	49.25-2.8 (2.87-2.8)[*]	49.17-3.1 (3.21-3.1)[*]	49.66-3.1 (3.15-3.10)[*]	48.42-3.0 (3.07- 3.00)[*]
R_{merge} (%)	9.1 (>100)	14.7 (>100)	21.5 (>100)#	11.5 (>100)
R_{meas} (%)	9.3 (>100)	15.1 (>100)	22.4 (>100)	12.7 (>100)
R_{pim} (%)	2.1 (>100)	3.4 (69.3)	6.5 (>100)	5.3 (66.4)
$I/\sigma I$	21.4 (0.8)	15.4 (1.1)	9.6 (0.7)	12.3 (1.3)
$CC_{1/2}$	1.0 (0.438)	0.999 (0.619)	0.998 (0.289)	0.998 (0.522)
Completeness (%)	100.0 (99.9)	100.0 (100.0)	99.9 (98.9)	99.9 (100.0)
Redundancy	19.9 (19.6)	19.9 (20.4)	11.9 (10.7)	5.6 (5.9)
Refinement				
Resolution	49.25-2.8	49.17-3.1	49.66-3.1	48.42-3.1
No. reflections	60935	45076	129732	62779
$R_{\mathrm{work}}/R_{\mathrm{free}}$	0.23/0.26	0.22/0.26	0.24/0.26	0.24/0.26
No. atoms				
Protein	14756	14746	44268	17205
Ligand/ion	4	34	89	54
Water	-	-	-	-
B-factors (Å^2)				
Macromolecules	109.09	101.62	90.44	83.24
Ligand/ion	96.58	81.78	70.21	83.54
Water	-	-	-	-
R.m.s. deviations				
Bond lengths (Å)	0.002	0.004	0.004	0.004
Bond angles (°)	0.55	0.98	0.95	0.95
Ramachandran				
favored (%)	95.03	95.19	94.71	94.27
allowed (%)	4.65	4.54	5.04	5.31
outliers (%)	0.32	0.27	0.25	0.42

[*]Statistics for the highest-resolution shell are shown in parentheses; #Two datasets were merged.

List of Figures

Abbreviations

Abbreviation	**Expansion**
ATP	adenosine triphosphate
bp	base pair
ca.	circa
cryo-EM	cryo-electron microscopy
DNA	deoxyribonucleic acid
e.g.	*exempli gratia*
EM	electron microscopy
HEAT	Huntingtin, Elongation Factor 3, Protein phosphatase 2A, Target of rapamycin 1
i.e.	*id est*
kDa	kilodalton
mRNA	messenger RNA
NMR	nuclear magnetic resonance
NTP	nucleotide triphosphate
nts	nucleotides
pre-mRNA	precursor mRNA
RNA	ribonucleic acid
RRM	RNA recognition motif
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
snRNPs	small nuclear ribonucleoproteins
SR	serine/arginine-rich
SS	splice site
U snRNA	uridine-rich small nuclear RNA
U snRNPs	uridine-rich small nuclear ribonucleoproteins

Acknowledgments

First and foremost I would like to express my gratitude and deep appreciation to my supervisor Dr. Vladimir Pena for his constant help and support, and, especially, for the long, thorough and insightful discussions in front of the electron density map of the human SF3B complex. I feel so honored and privileged to be part of the "Mecca of Splicing" headed by Prof. Dr. Reinhard Lührmann. I would like to thank Reinhard for his generous support during all these years, and for the diligent advice on our SF3B manuscripts. I am also honored to have Prof. Dr. Henning Urlaub and Prof. Dr. Patrick Cramer as members of my thesis committee. Their insightful comments and suggestions were invaluable, especially at the earlier stages of the project.

I would like to express my sincere gratitude to all, present and past, members of the research group "Macromolecular Crystallography" and the department of "Cellular Biochemistry" for providing an amicable working environment. In particular, I would like to thank Dr. Olex Dybkov for the constructive discussions, accurate structural analyses, and his enormous help throughout this challenging project. Very special thanks to Dr. Cindy L. Will for the invaluable suggestions at the earlier stages of the project and for the critical contributions to both of our SF3B manuscripts. I have been fortunate to work together with Dr. Evelina de Laurentiis and Dr. Kundan Sharma, and their expertise was key to the overall success of the first part of our project. I would like to thank Dr. Berthold Kastner, Dr. Klaus Hartmuth, Dr. Reinhard Rauhut, Dr. Norbert Rigo, and Dr. Jana Schmitzova for insightful discussions and expertise help.

Many thanks to our skilled technicians: Uli Steuerwald, Jürgen Wawrzinek, Gabi Heyne, Thomas Conrad, Hossein Kohansal, Monika Raabe, Annika Kühn, and Heiko Ludwig from the IT service. I would also like to thank Juliane Moses for helping me with all the possible administrative hurdles.

In addition, I would like to thank all my lab friends: Csaba, Uli, Sebi, Sara, Sascha, Aleks, Almudena, Judit, Mateusz, Ilya, and Inessa. I would like to express my gratitude to the IMPRS "Molecular biology" program, and its faculty members, for providing an excellent scientific environment. Especially, I would like to thank Dr. Steffen Burkhardt for his help, friendly advice, and understanding.

Finally, I would like to thank my dear, dear family and close friends, for the unlimited support and love.