




# Structural basis of intron selection by U2 snRNP in the presence of covalent inhibitors

Constantin Cretu<sup>1,5</sup>, Patricia Gee<sup>2</sup>, Xiang Liu<sup>2</sup>, Anant Agrawal<sup>2</sup>, Tuong-Vi Nguyen<sup>2</sup>, Arun K. Ghosh<sup>3</sup>, Andrew Cook<sup>2</sup>, Melissa Jurica <sup>4</sup>, Nicholas A. Larsen<sup>2</sup> & Vladimir Pena<sup>1</sup> 

Intron selection during the formation of prespliceosomes is a critical event in pre-mRNA splicing. Chemical modulation of intron selection has emerged as a route for cancer therapy. Splicing modulators alter the splicing patterns in cells by binding to the U2 snRNP (small nuclear ribonucleoprotein)—a complex chaperoning the selection of branch and 3' splice sites. Here we report crystal structures of the SF3B module of the U2 snRNP in complex with spliceostatin and sudemycin FR901464 analogs, and the cryo-electron microscopy structure of a cross-exon prespliceosome-like complex arrested with spliceostatin A. The structures reveal how modulators inactivate the branch site in a sequence-dependent manner and stall an E-to-A prespliceosome intermediate by covalent coupling to a nucleophilic zinc finger belonging to the SF3B subunit PHF5A. These findings support a mechanism of intron recognition by the U2 snRNP as a toehold-mediated strand invasion and advance an unanticipated drug targeting concept.

<sup>1</sup>Research Group Mechanisms and Regulation of Splicing, The Institute of Cancer Research, London, UK. <sup>2</sup>H3 Biomedicine, Inc, Cambridge, MA, USA. <sup>3</sup>Departments of Chemistry and Medicinal Chemistry, Purdue University, West Lafayette, IN, USA. <sup>4</sup>Department of Molecular, Cell and Developmental Biology, University of California, Santa Cruz, CA, USA. <sup>5</sup>Present address: Cluster of Excellence Multiscale Bioimaging (MBExC), Universitätsmedizin Göttingen, Göttingen, Germany. email: [vlad.pena@icr.ac.uk](mailto:vlad.pena@icr.ac.uk)

**D**uring splicing, introns are removed from the nascent transcripts by two transesterification reactions catalyzed by the spliceosome—a molecular machine composed of five ribonucleoprotein particles (snRNPs), known as U1, U2, U4, U5, and U6, and additional non-snRNP factors<sup>1</sup>. Spliceosomes assemble on introns by the stepwise recognition of conserved sequences that provide the reactive groups for catalysis: the 5′SS (splice site), the branch site (BS), followed downstream by the polypyrimidine tract (PPT) and the 3′SS. The spliceosome transits through several landmark complexes, such as E, A, B, C, P, and ILS, during the splicing pathway<sup>1,2</sup>.

The intron's BS and 3′SS regions are defined early in the splicing cycle, in a process that often occurs in alternative ways<sup>3</sup>. The human BS sequence has a short and degenerate consensus motif YUNAY<sup>4,5</sup>, and is sometimes present in multiple, alternative copies selectively used in different cell types<sup>6</sup>. In yeast and subsets of human introns, BS and PPT-3′SS are recognized cooperatively by SF1 and the U2AF heterodimer, respectively<sup>7–9</sup>, whereas the U1 snRNP binds concomitantly to the 5′SS. These recognition events may occur across the intron (i.e., intron definition) or the exon (i.e., exon definition)<sup>7,10,11</sup>. Subsequently, displacement of SF1 enables the recruitment of the U2 snRNP—an intricate and dynamic apparatus essential for the accurate selection of the BS and 3′SS<sup>4,5,12</sup>. Following a currently unclear mechanism, the U2 snRNA component of the U2 snRNP base pairs to the BS in an ATP-dependent manner, yielding the pre-spliceosome A complex. The invariant BS adenosine (BS-A), extruded from the U2/BS helix, serves later as the reactive nucleophile during the branching reaction of splicing (reviewed in refs. <sup>13,14</sup>).

The U2 snRNP has a bipartite organization, exhibiting a smaller U2 3′ module and a larger U2 5′ module<sup>15–18</sup>. The latter contains the heptameric SF3B complex, whose core subunit SF3B1 is recurrently mutated in cancers<sup>15–19</sup>. SF3B1 exhibits a HEAT domain (SF3B1<sup>HEAT</sup>) that adopts an open conformation in the 17S U2 snRNP and closed conformation in spliceosomes where it stabilizes the extended U2/intron helix and the polypyrimidine tract, in close proximity of the 3′SS<sup>17,18,20–22</sup>.

The protein–RNA and RNA–RNA rearrangements that enable selection of the BS and 3′SS by the U2 snRNP during the formation of pre-spliceosomes are complex and insufficiently understood<sup>22–25</sup>. Initially, the U2 snRNA is folded in a compact form within the 17S U2 snRNP, with the BS-interacting region presented as the loop region of the so-called branchpoint-interacting-stem-loop (BSL)<sup>22,23</sup>. After BS binding, the BSL should unwind and interact with the intron to form the extended U2/intron duplex, whose conformation remains unchanged throughout the A to B<sup>act</sup> complexes<sup>17,18</sup>. The length of the helix is confined to 16 base pairs by a structural frame of the SF3B and SF3A proteins<sup>17,18,26,27</sup>. Concomitantly, SF3B1<sup>HEAT</sup> accommodates the branch helix like a clamp, while trapping the BS-A within a pocket<sup>17,18,20,21</sup>. The BS-A binding pocket also serves as a hinge of the SF3B1 clamp (referred to as the hinged pocket). Simultaneously, SF3B1<sup>HEAT</sup> binds the PPT several bases upstream of the 3′SS, suggesting an important contribution to the cooperative recognition of the BS and 3′SS<sup>26,27</sup>.

Pre-mRNA splicing defects have emerged as a hallmark for many cancers, and spliceosomes are viewed as promising therapeutic targets<sup>28,29</sup>. In particular, the selection of introns by the U2 snRNP is amenable to modulation by small-molecule anti-tumor compounds that share a common binding site on SF3B<sup>28,30</sup>. Referred to as splicing modulators, these compounds profoundly impact pre-mRNA splicing patterns in cells by inducing widespread exon skipping and intron retention in a manner that depends on the sequence of the inhibited substrate, modulators' structure, affinity, and dose<sup>30–33</sup>. More than 20

modulators belonging to three distinct chemotypes—FR901464 analogs, herboxidienes, and pladienolides—have been reported, and some have entered clinical trials<sup>28,29,32</sup>. The first mechanistic insight into modulators' binding to SF3B was provided by the recombinant SF3B core structure in a complex with pladienolide B (PB)<sup>34</sup>. The structure shows that the ligand occupies a hinged pocket in the open conformation of SF3B1, outcompeting the BS-A and preventing transition to the closed state of SF3B<sup>20,32,35</sup>. However, in the absence of native pre-spliceosome complexes stalled with modulators, the exact mechanism of interference in intron recognition remains unclear.

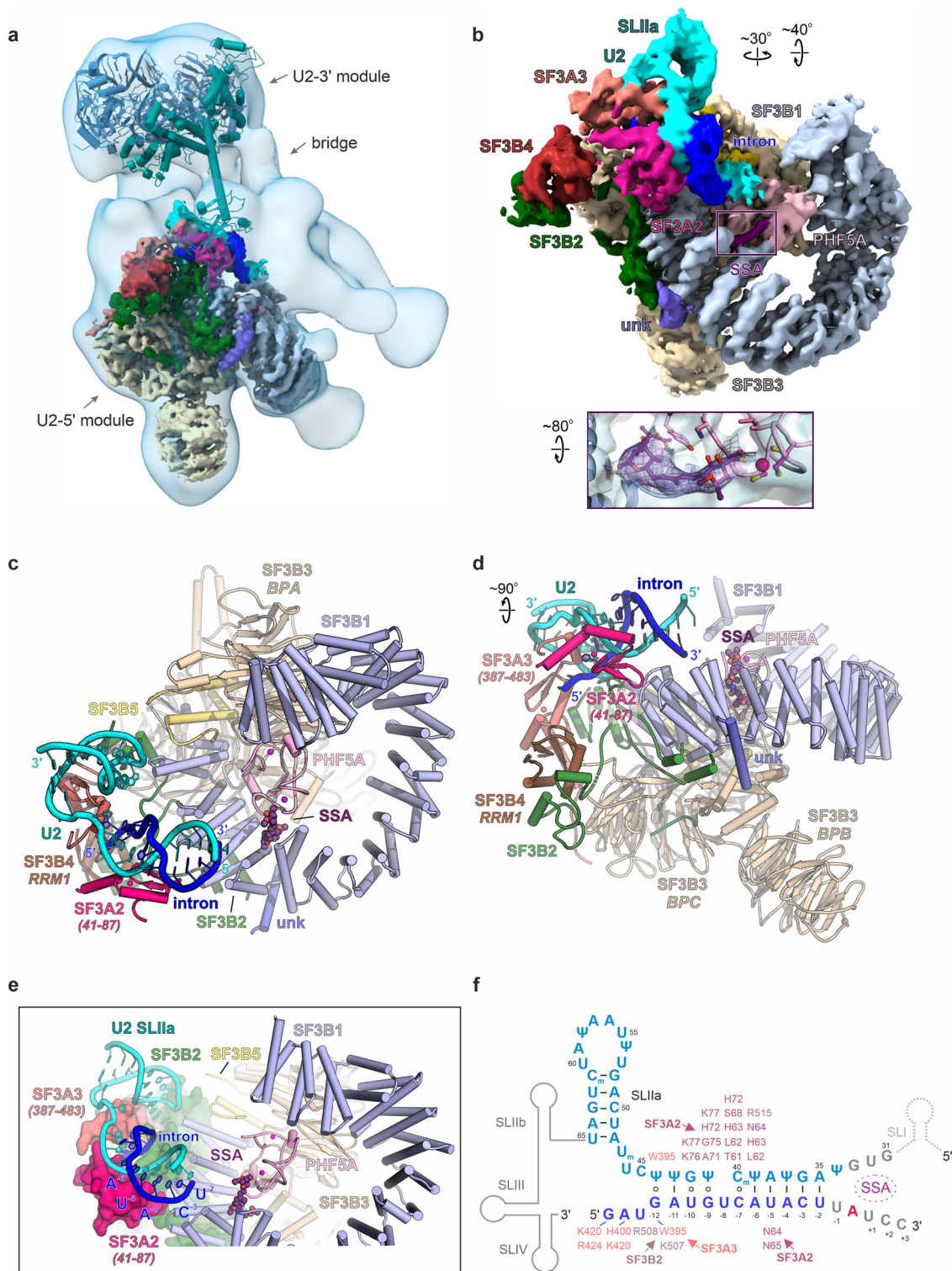
Identified more than two decades ago<sup>36</sup>, the splicing modulator FR901464 has a different chemistry than pladienolides and herboxidienes. Since then, several groups have proposed that the epoxide group carried by FR901464 analogs might react with SF3B's subunits<sup>37–40</sup>. Extensively investigated *in vitro* and *in vivo*, the FR901464 analog spliceostatin A (SSA) provides a model system for splicing modulators and a molecular tool for dissecting the assembly of pre-spliceosomes and intron selection<sup>31,41–43</sup>. Intriguingly, SSA can recapitulate the effects of SF3B1 knock-down in cells<sup>43</sup>, suggesting a possible irreversible inactivation of SF3B1 through a yet unknown mechanism. However, at lower concentrations, FR901464 analogs have a less pleiotropic effect and inhibit splicing in a manner dependent on the intron sequence<sup>31,33,42</sup>.

In this work, by employing FR901464 analogs as molecular tools to dissect the splicing pathway, we elucidate how spliceostatin/sudemycins interfere with pre-spliceosome assembly and splicing commitment by covalent coupling to a reactive zinc finger of PHF5A. In addition, the structures support a general mechanism of intron selection by the U2 snRNP and provide a mechanistic explanation for the differential inhibition of introns in cells by known SF3B modulators.

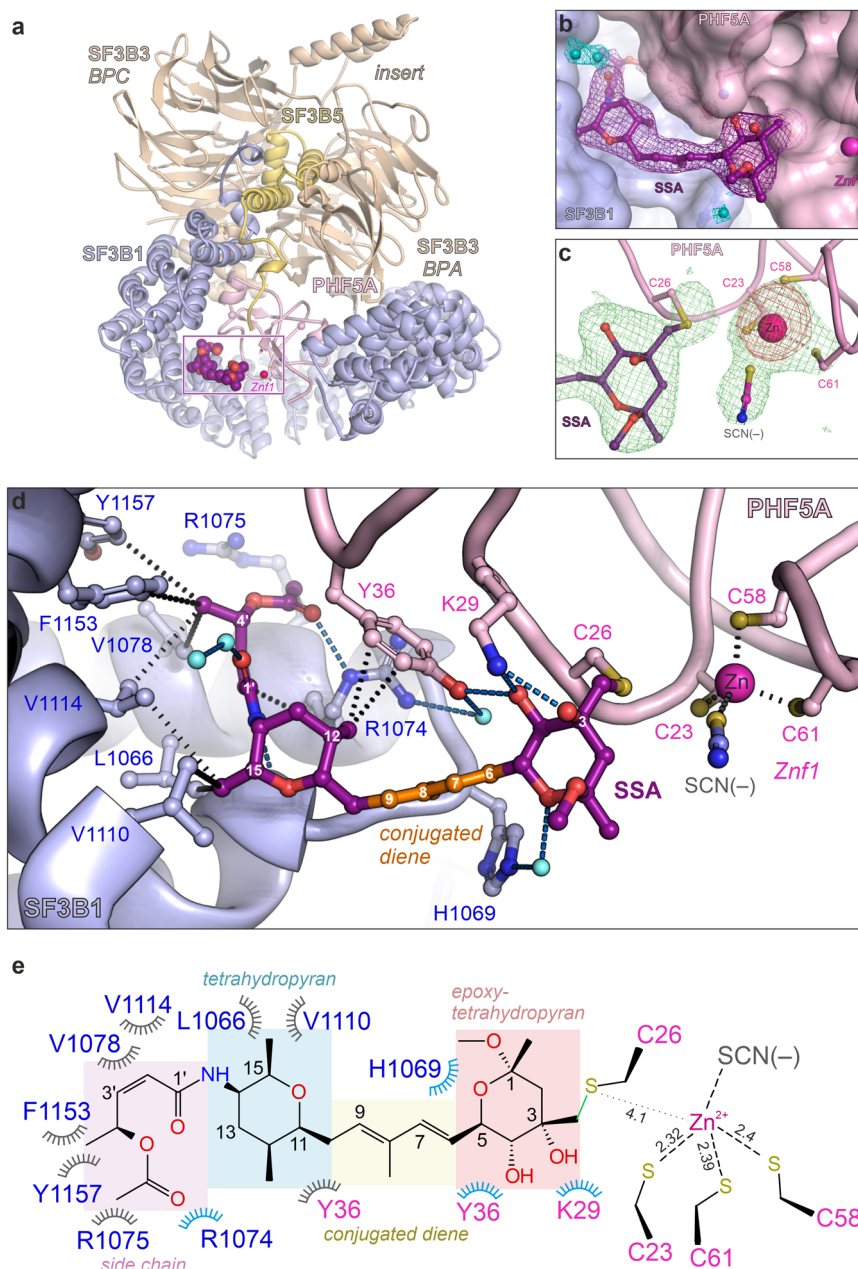
## Results and discussion

**SSA arrests spliceosomes during the selection of introns.** To elucidate the impact of anti-tumor splicing modulators on the splicing pathway and reveal how the U2 snRNP accurately selects and binds the introns, we set out to obtain the cryo-EM structure of a pre-spliceosome arrested by SSA. Attempting to avoid the potential intrinsic flexibility of cross-intron pre-spliceosomes<sup>44</sup>, we decided to assemble a U2-containing cross-exon complex<sup>10</sup>. We designed a model RNA substrate consisting of 54 intron nucleotides, the second MINX exon, the downstream 5′SS, and three MS2 aptamers for affinity purification (Supplementary Fig. 1a). This RNA construct is sufficient to recapitulate *in vitro* the effects of spliceostatins and pladienolides on spliceosome assembly, including dependence on ATP and a functional BS, as well as the sensitivity to heparin treatment<sup>43</sup>. Mass-spectrometric analysis of this cross-exon pre-spliceosome—referred to as the A3′ complex—reveals the presence of many proteins typically found in A complexes<sup>45</sup>. In contrast to cross-exon complexes purified in the absence of inhibitors<sup>10</sup>, the PRP5/DDX46 helicase (yeast Prp5p) is abundant, whereas the U4/U6.U5 tri-snRNP components are poorly represented in the A3′-SSA proteome (Supplementary Fig. 1b and Supplementary Data 1).

Single-particle cryo-EM analysis of A3′-SSA enabled the reconstruction of a ~12 Å density map that corresponds to the U2 snRNP module of the complex. The overall map of the U2 snRNP shows two distinct lobes connected by a bridge where part of the SF3A complex is present (Fig. 1a and Supplementary Figs. 1c, d, 2). Focused classification and refinement enabled us to resolve the U2 5′ module to a resolution of ~3.1 Å (Supplementary Fig. 1e–h, Fig. 2 and Supplementary Table 1). In this map, we modeled the SF3B complex, the two matrix-type zinc-finger



**Fig. 1 Cryo-EM structure of human A3' prespliceosome-like complex arrested with SSA.** **a** Overall cryo-EM density map of the U2 snRNP module of the A3' prespliceosome. The density maps of the U2 5' module and U2 snRNP are superimposed. The subunits are color-coded as in **b**. The proteins of the U2 3' module and associated SF3A domains are docked as a rigid body and depicted in cartoon representation. **b** Cryo-EM density map of the U2 5' module resolved at 3.1 Å resolution. The inset shows the density element observed in SF3B's hinged pocket and corresponds to the SSA ligand. **c, d** Overall organization of the U2 5' module in two orientations. **e** Structure of the U2 5' module, emphasizing the U2/intron duplex relative to the surface of contacting subunits and the SSA. **f** Organization of the U2/intron structure within the A3'-SSA complex. Proteins and residues that interact with the U2/intron duplex are labeled. Methylated bases are marked with subscript "m", and Ψ denotes pseudouridine. Except for **a**, all subunits are colored and labeled accordingly throughout the figure.



**Fig. 2** Crystal structure of SF3B in complex with SSA. **a** Structure overview of the SF3B<sup>ABPB</sup> core in complex with SSA. **b** Electron density of the SSA within the crystal structure of the SF3B core. The  $2mF_o - Df_c$  map is contoured at  $-1.5 \sigma$ , and is displayed around the ligand and several water molecules (colored in purple and cyan, respectively). **c** Electron density ( $mF_o - Df_c$ ,  $3 \sigma$ ) of the Znfi motif of PHF5A and the covalently bound SSA ligand contoured around the final model. An anomalous difference map is depicted around the zinc ion and is colored red. **d** Crystal structure shows that SSA is engaged in multiple contacts with residues from the SF3B1-PHF5A binding tunnel. Distances between SSA and residues involved in hydrophobic contacts are shown as black dashes, whereas polar contacts are depicted as blue dashes. **e** Schematic depiction of the SSA interaction with residues from the hinged pocket, as observed in the crystal structure. Polar contacts are depicted in blue, whereas the hydrophobic interactions are colored gray. SSA's functional groups are indicated as boxes.

domains of SF3A2 and SF3A3, part of the intron substrate paired to the U2 snRNA, and the SSA ligand used to arrest the splicing reaction. SSA has an elongated, L-shaped density and is positioned in the hinged pocket of SF3B, between its SF3B1 and PHF5A subunits (Fig. 1b and Supplementary Fig. 3). We do not observe density for the DDX46 helicase from the 17S U2 snRNP, indicating either repositioning or destabilization during the intron's pairing to U2 (Supplementary Fig. 4a, b). As expected, subunits and RNA structures specifically found in the 17S U2 snRNP (such as TAT-SF1 or U2 BSL) are not present in

the A3'-SSA complex (Supplementary Fig. 4c-e). Surprisingly, SF3B adopts the same open conformation as in the 17S U2 snRNP or the isolated recombinant complex<sup>20,22</sup>, contrasting with the closed conformation observed in A-to-B<sup>act</sup> spliceosomes (Supplementary Fig. 4f)<sup>17,18,46</sup>. Equally unexpected, U2 pairs with the intron, despite SSA having bound to the open state of SF3B. However, the U2/intron duplex observed in the presence of SSA, which we denote as the precursor U2/intron helix, is shorter than the extended U2/intron helix evidenced in the cross-intron A and later complexes<sup>18</sup> and forms upon U2 pairing to the  $-12$  to  $-2$

intron region upstream of the BS-A. Importantly, the intron's BS sequence has not fully paired U2, and the BS-A is not available for binding to SF3B's hinged pocket, which is occupied by SSA (Fig. 1e, f).

Interestingly, an  $\alpha$ -helix of unknown identity resides on the convex surface of SF3B1<sup>HEAT</sup>, on the opposite side of the hinged pocket that binds SSA (Fig. 1b–d). This location suggests a potential role of this helix in the transition from the open to the closed conformation. Furthermore, a globular density module that lacks discernable features is located at the periphery of the U2 snRNP (i.e., map M1, Supplementary Fig. 2a). It is currently unclear whether this density corresponds to the U1 snRNP or an assembly of non-snRNP proteins. Superposition of the cross-exon A3' complex, the cross-intron yeast A complex, and the human pre-B complex shows that the unassigned density does not align to any other spliceosome subunits, including the U1 snRNP (Supplementary Fig. 5)<sup>44,47</sup>. Future studies are required to clarify this density's possible relevance in the stepwise formation of cross-exon complexes.

**FR901464 analogs bind covalently to a reactive zinc finger of PHF5A.** Despite the unambiguous localization of SSA by cryo-EM, the resolution is insufficient to explain in atomic details how the pocket recognizes SSA and the reasons for its irreversible effects in cells<sup>43,48</sup>. To resolve the small-molecule compound at a higher resolution, we set out to investigate its co-crystal structure with SF3B. We have engineered a variant of the human SF3B core that lacks the  $\beta$ -propeller BPB domain of SF3B3 (SF3B<sup>ΔBPB</sup>, Supplementary Fig. 6a, b). The complex formed crystals in two alternative space-groups, belonging to the orthorhombic and trigonal symmetry systems (Supplementary Fig. 6c–e and Supplementary Table 2). Our best orthorhombic crystals diffracted X-rays to a resolution of 2.3 Å, while diffraction data for the trigonal form was limited to ~3.0 Å.

The overall organization of the SF3B core and SSA's conformation are largely the same as in the A3'-SSA complex (Fig. 2a)<sup>34</sup>. However, the level of detail is much higher at the 2.3 Å resolution (Fig. 2b–d and Supplementary Figs. 6f–i, 7). The distal half of the SSA molecule (C11–C15) establishes hydrophobic contacts and hydrogen bonds with the hinged pocket (Fig. 2d, e). The proximal half of SSA (atoms C1–C10) extends away from the tunnel towards one of the zinc fingers of PHF5A (i.e., Znf1, Fig. 2b). Its epoxy-tetrahydropyran group is recognized by a hydrogen bond network, involving SF3B1 and PHF5A residues and several solvent molecules in the binding pocket (Fig. 2d, e and Supplementary Fig. 7c, d). Strikingly, SSA's epoxide group resides close to C26-PHF5A—one of the four cysteine residues of PHF5A's Znf1 zinc cluster. The electron density is continuous between C26-PHF5A and the epoxide group, while absent between C26-PHF5A and Znf1's zinc ion, indicating an intermolecular covalent bond between SSA and C26-PHF5A (Fig. 2c and Supplementary Fig. 6f–i). Although C26-PHF5A is no longer within coordination distance from the zinc ion, the anomalous difference density shows that the metal is not ejected from PHF5A upon covalent modification of Znf1. In turn, a water molecule appears to occupy the freed zinc coordination shell in the trigonal crystal form (Supplementary Fig. 8a).

Notably, a thiocyanate ion replaces the water molecule at Znf1 in the orthorhombic SF3B<sup>ΔBPB</sup>-crystals that grow in 200 mM potassium thiocyanate. The 2.3 Å resolution of these crystals enables visualization of thiocyanate's distinctive elongated shape next to the zinc atom (Fig. 2c, d and Supplementary Figs. 6f, h, i, 7a, b, 8a). Importantly, crystals of SF3B<sup>ΔBPB</sup> in complex with pladienolide D (PD) form as well in the presence of thiocyanate. However, no thiocyanate density is observed at the Znf1 cluster in

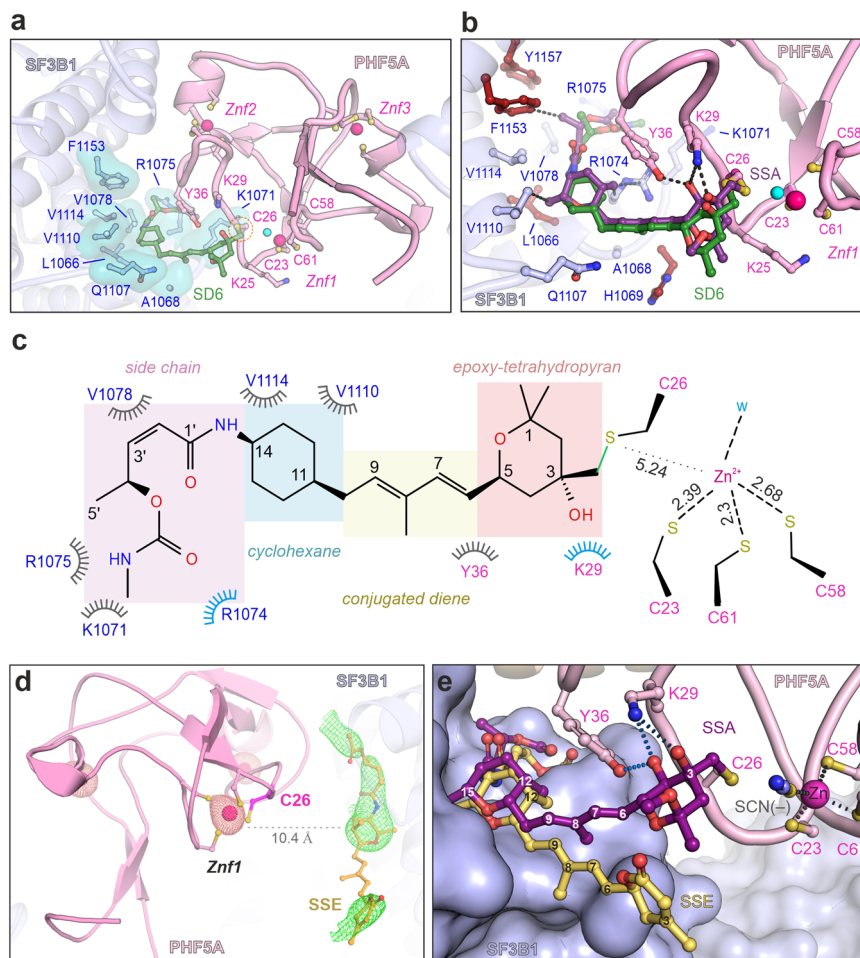
SF3B<sup>ΔBPB</sup>-PD crystals (Supplementary Fig. 8a, b), consistent with the idea that the thiocyanate ion interacts with Znf1 only when a free zinc coordination shell is being made available (i.e., after the coupling reaction between SSA and C26-PHF5A has occurred).

To investigate whether the covalent binding to PHF5A also occurs for the sudemycin analogs of FR901464<sup>49,50</sup>, we crystallized the SF3B<sup>ΔBPB</sup> core in complex with sudemycin D6 (SD6, Fig. 3a, b)<sup>33,51</sup>. Like SSA, SD6 adopts the L-shaped pose, the thioether bond is present between the C13 atom of SD6 and C26-PHF5A, indicating that the pyran-epoxide moiety is a general feature of splicing modulators that bind covalently to SF3B (Fig. 3a–c and Supplementary Fig. 9, Supplementary Table 3). The reduced number of interactions between SD6 and the SF3B1-PHF5A pocket explains its lower potency and binding affinity compared to SSA<sup>31,51,52</sup>.

Spliceostatin E (SSE) lacks the reactive epoxide group, and previous studies indicate an inhibitory activity of several orders of magnitude lower than SSA *in vitro*<sup>52</sup>. To understand how an inactive SSA analog is recognized in the absence of the epoxide warhead group, we have also determined the structure of the SF3B<sup>ΔBPB</sup>-SSE complex at 3.0 Å resolution (Supplementary Table 3). The structure shows that the binding tunnel stabilizes the distal half of SSE, largely similar to the equivalent region of SSA and SD6. Conversely, the proximal half of SSE yields weak and noisy electron density, likely due to increased mobility caused by the lack of covalent coupling to C26-PHF5A (Fig. 3d and Supplementary Fig. 10). Comparison between the structures of SSA, SSE, and SD6 highlights a bipartite functional organization of FR901464 analogs, where the distal moiety is necessary and sufficient for molecular recognition of the tunnel, while the proximal moiety is required for the irreversible, covalent binding to the SF3B complex (Fig. 3e and Supplementary Figs. 8a, 10b, c).

**Reaction mechanism of covalent coupling.** The epoxide rings are generally susceptible to attack from various nucleophiles, including thiols, especially when an appropriate base efficiently deprotonates the latter<sup>53</sup>. A notable precedent is provided by the antibiotic fosfomicin, which binds covalently to C115 residue from the catalytic center of the bacterial enzyme MurA. Although not belonging to a zinc finger, the nucleophilicity of the cysteine appears enhanced by the nearby guanidinium moieties of R120 and R397<sup>54</sup>.

However, covalent coupling between a zinc-coordinated cysteine and an epoxide ring is intriguing, as it was never reported for SSA, SD6, or other compounds. Therefore, we used several methods to show that covalent coupling is not dependent on the crystallization conditions. First, mass-spectrometric (MS) analysis of the SF3B core incubated with SSA or SD6 confirms that the molecular weight of detected PHF5A molecules increases with the expected mass of the bound modulators (Supplementary Fig. 11a). Secondly, we employed scintillation proximity assays with tritiated PB to show *in vitro* a dramatic loss of SSA and SD6 binding to the core SF3B complex harboring C26H-PHF5A (Fig. 4a). Third, we show that mutated cells harboring C26H-PHF5A can discriminate between covalent and non-covalent inhibitors, being insensitive to several orders of magnitude increase in SSA concentration. We do not observe this effect for PB, consistent with the critical importance of C26-PHF5A for the binding of covalent inhibitors to SF3B (Fig. 4b). In contrast, previously reported mutations of PHF5A (Y36A) or SF3B1 (R074H, V1078I/A) confer resistance both to pladienolide-related E7107 as well as SSA and SD6 (Supplementary Fig. 11c, d)<sup>33</sup>. As these residues are located remotely from the zinc finger (Figs. 2d, e and 3a, c), their substitutions are likely to impair the interactions with modulators' distal moiety without interfering with the covalent coupling.



**Fig. 3 Crystal structures of SF3B in complex with SD6 and SSE.** **a** Crystal structure shows that SD6 binds covalently to PHF5A and is engaged in contacts with fewer residues than those binding SSA (depicted and labeled). The water molecule and zinc ions are shown as cyan and magenta spheres, respectively. **b** Structural comparison between SSA and SD6 in co-crystal structure with the SF3B core. The interactions with chemical groups specific for SSA (i.e., absent from SD6) are shown. **c** Schematic depiction of the SD6 interaction with residues from the hinged pocket, as observed in the crystal structure. Polar contacts are depicted in blue, whereas the hydrophobic interactions are colored gray. SD6's functional groups are indicated as boxes. **d** Crystal structure of SF3B $\Delta$ BPB in complex with spliceostatin E (SSE). The polder (5.0  $\sigma$ ) electron density maps are displayed around SSE. An anomalous difference map (dark red, 6.0  $\sigma$ ) is contoured around PHF5A's three zinc ions. Note the presence of ligand density in the SF3B1-PHF5A tunnel corresponding to the distal half of SSE. In contrast, density for the proximal moiety is largely absent, likely because of its flexibility in the absence of covalent coupling to C26-PHF5A. **e** Structural comparison between SSA and SSE in co-crystal structure with the SF3B core. Note that electron density was observed for the distal moiety of SSE, while the proximal moiety (C1-C10) is likely mobile. The depicted model is based on stereochemistry and space availability in the tunnel.

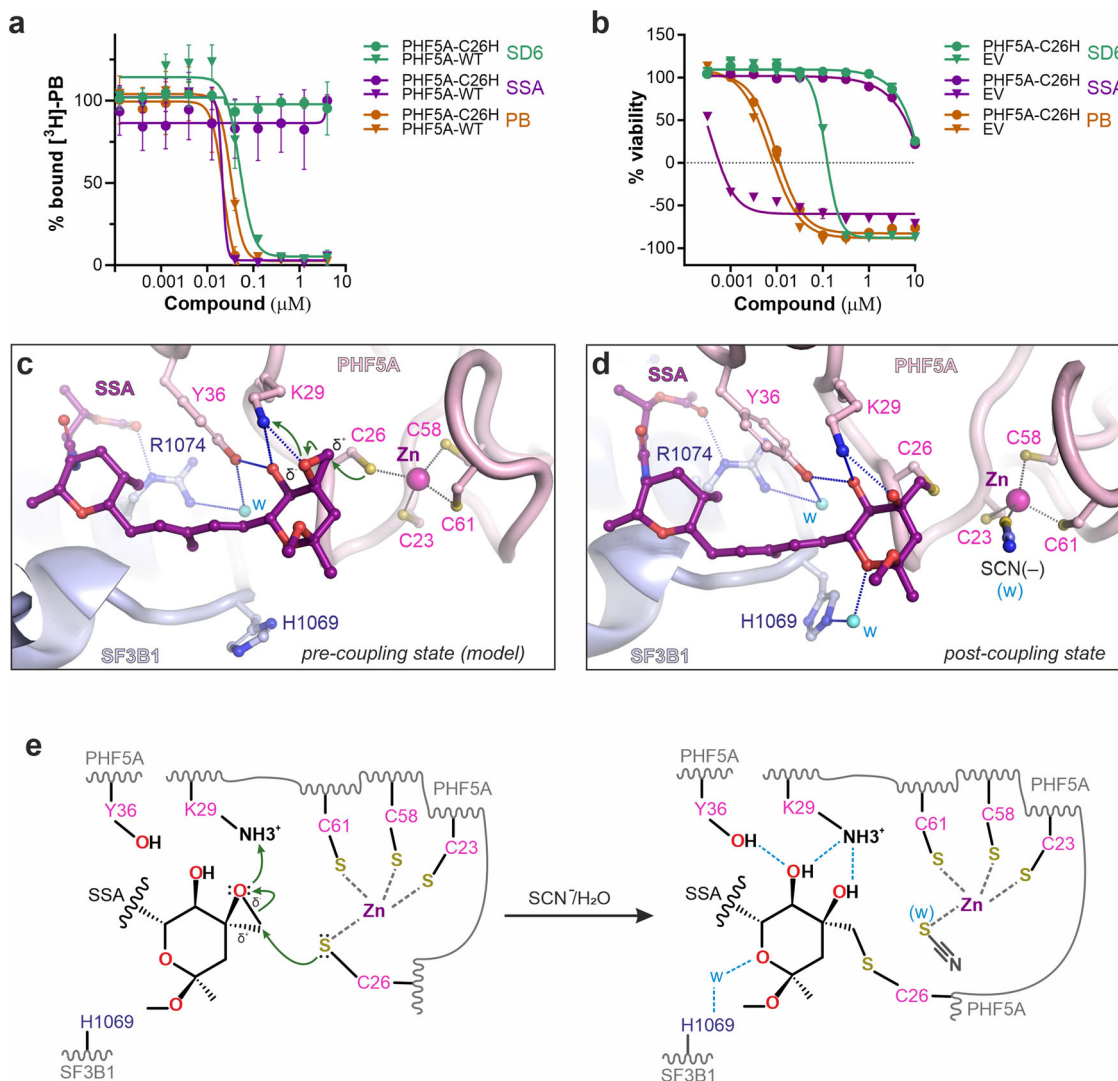
As captured in the crystal structure, the configuration of the chemical groups in the post-reaction state supports a nucleophilic substitution mechanism. The zinc atom serves a catalytic role in raising the nucleophilicity of C26-PHF5A by deprotonating the thiol group. The resulting thiolate anion attacks and opens the epoxy ring to form a thioether bond, while K29-PHF5A contributes to stabilizing the leaving group (Fig. 4c–e). Consistently, viability assays show that K29-PHF5A may have a mild impact on the coupling efficiency, with cells being less sensitive to SSA/SD6 when K29 was substituted for alanine or arginine (Supplementary Fig. 11b). Concomitantly, a water molecule likely replaces C26-PHF5A from the coordination sphere of the zinc atom, resulting in a configuration reminiscent of the active sites of zinc enzymes (e.g., carbonic anhydrase or alcohol dehydrogenase)<sup>55</sup>, where a water molecule and three side chains coordinate the zinc atom. Incidentally, thiocyanate substitutes the water in a crystal form that grows in the presence of this ion.

The proposed reaction chemistry bears some similarities with other systems, such as the Ada system in *E. coli*<sup>56,57</sup> or the zinc ejectors compounds<sup>56</sup> or fosfomycin coupling to MurA<sup>53</sup>.

However, the covalent modification of PHF5A's structural zinc-finger motif by spliceostatins/sudemycins is unique in its mechanism and shows how these modulators exploit the rich microenvironment of the SF3B binding tunnel to inactivate an early prespliceosome complex.

#### Comparison between FR901464 and pladienolide analogs.

Although FR901464 and pladienolide analogs share the same binding pocket (Fig. 5)<sup>34,52</sup>, the determined structures highlight striking differences as well as unexpected similarities between these different families of compounds. First of all, this work establishes that FR901464 analogs are covalent inhibitors, in contrast to pladienolides. Despite this surprising difference, the existing structures suggest that both types of modulators bind SF3B in several steps, finally becoming stabilized in an L-shape pose induced by multiple constraints of the binding tunnel. One segment of the ligands' L-shape follows a common binding path in FR901464 analogs and pladienolides, while the second segment is oriented in the opposite direction (Fig. 5a, b)<sup>34</sup>. Comparison



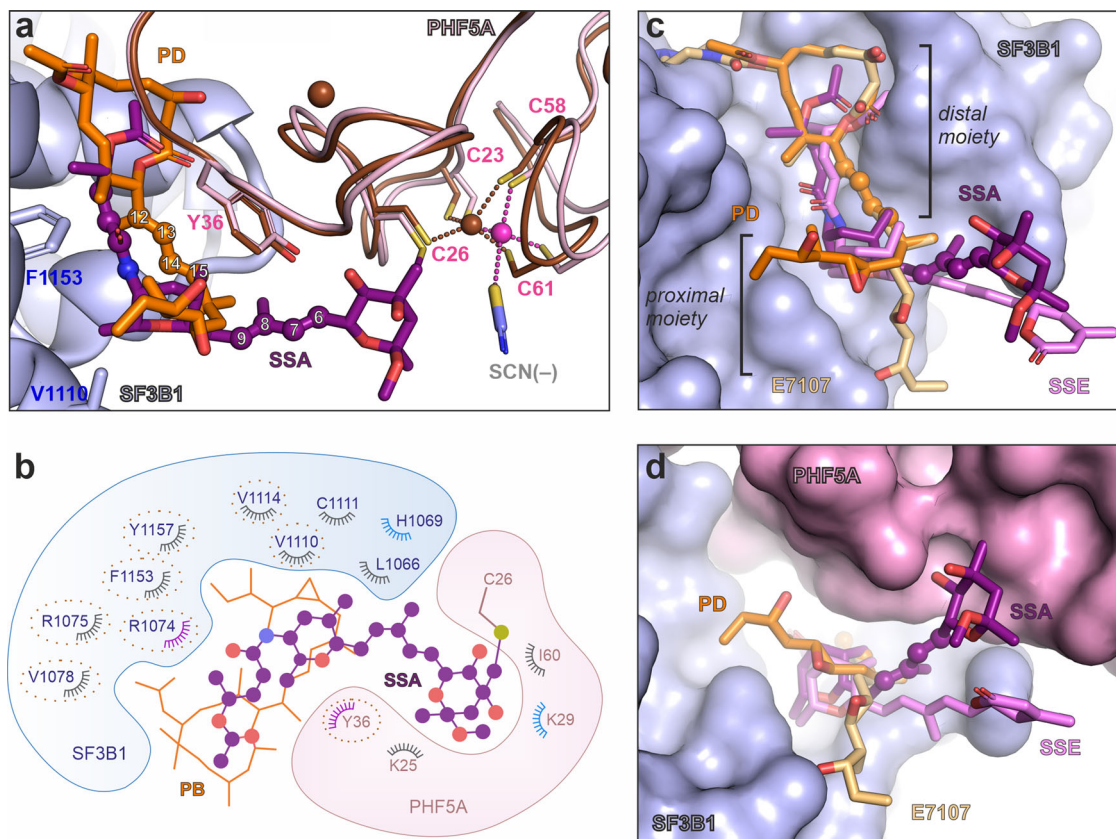
**Fig. 4** Mechanism of covalent coupling to the nucleophilic zinc finger. **a** Competitive titration of  $[^3\text{H}]$ -pladienolide B (10 nM) with SSA, SD6, and PB by scintillation proximity assays (SPA). Error bar indicates SEM,  $n = 2$  biological replicates. Source data are provided in the Source Data file. **b** Cell growth inhibition assays of HCT116 cells harboring the C26H-PHF5A point mutation. Control cells were transfected with an empty vector. Error bar indicates SEM,  $n = 2$  biological replicates. Source data are provided in the Source Data file. **c** Model for the pre-coupling state was obtained by docking the unreacted ligand into the experimental density of the covalently-linked SSA. **d** Crystal structure of SF3B $\Delta$ BPB-SSA corresponding to the post-coupling state depicted as in **c** for a side-by-side comparison. **e** Mechanism of covalent coupling of epoxy modulators to the nucleophilic zinc finger of PHF5A, as inferred from crystal structures and stereochemistry. Available structural data and functional assays favor a nucleophilic substitution mechanism where the activated thiolate group of C26 carries the nucleophilic attack, whereas K29 stabilizes the leaving hydroxyl group. The zinc-bound C26 can be replaced either by water or thiocyanate.

between SSA, SD6, and SSE suggests their distal part initiates the molecular recognition by SF3B (Fig. 3a, e). In the next step, the warhead group belonging to the proximal moiety establishes polar and hydrophobic contacts with SF3B residues near PHF5A's Zn1 motif, enabling the coupling reaction between the epoxy group and the thiolate moiety of C26-PHF5A (Fig. 4c–e).

Pladienolide derivatives appear to bind in two steps as well. Thus, a comparison between SF3B-PD $\Delta$ BPB, SF3B-PB and isolated E7107 suggests that pladienolides access the binding tunnel in an extended side-chain conformation<sup>34,35</sup>, while later bind stably by an induced fit reconfiguration of the aliphatic side-chain. This reconfiguration appears to “lock” the compound on SF3B, in the distinctive L-shaped conformation (Fig. 5a, c, d and Supplementary Fig. 8b). To some extent, the induced fit of the aliphatic side-chain and covalent coupling to Zn1 may serve

similar roles for pladienolide derivatives and FR901464 analogs, respectively.

Contrary to earlier suggestions<sup>57</sup>, a striking difference between FR901464 analogs and pladienolides is the location and function of the conjugated diene—a distinctive moiety present in all families of modulators that bind SF3B<sup>29,58</sup>. In pladienolides, this moiety occupies the narrower and central part of the binding tunnel<sup>34</sup>. In covalent inhibitors, an amide group followed by a carbon-carbon double bond (C2'–C3') replaces the role and location of the conjugated diene. (Fig. 5a, b). In contrast to pladienolides, the conjugated diene group in SSA and SD6 has a different function, serving as a rigid spacer that positions the epoxide group in the reach of PHF5A's reactive zinc finger, while its contacts with the binding tunnel are minimal. Following these findings, the construction of spacers carrying epoxy warheads and



**Fig. 5 Structural comparison between covalent and non-covalent modulators.** **a** Superposition of SF3B $\Delta$ BPB-SSA onto SF3B $\Delta$ BPB-PD. Atoms from the conjugated diene moiety are numbered. PHF5A complexed with SSA is colored brown. **b** Schematic of the SF3B1-PHF5A pocket with the PB (PDB 6EN4) and the SSA ligands superimposed. SF3B residues contacting SSA are indicated. Residues that also contact PB are circled. Residues involved in polar contacts and hydrophobic interactions with SSA are colored in blue and gray, respectively. **c** Comparison between SF3B $\Delta$ BPB-SSA, SF3B $\Delta$ BPB-SSE, and SF3B $\Delta$ BPB-PD structures. The crystal structure of isolated E7107 (DB12508) is superimposed on the macrolide ring of PD. PHF5A is not shown for clarity's sake. **d** A different view of the superpositions shown in **c**. PHF5A is depicted in pink.

their grafting on weakly binding compounds may guide the design of next-generation splicing modulators.

**A mechanism of intron's progressive recognition by the U2 snRNA during prespliceosome formation.** Often present in multiple copies within the same intron<sup>6</sup>, human BSs are to a great extent recognized before recruitment of the U2 snRNP to the intron<sup>7–9</sup>. The U2/intron duplex formation is likely essential in the BS's final selection step for splicing commitment. Far exceeding the conserved YUNAY motif of BSs<sup>4,5</sup>, the extended U2/intron duplex consists of 16 base pairs in cryo-EM structures of A to B<sup>act</sup> spliceosomes, confined by conserved residues of SF3A and SF3B complexes<sup>17,18,26,27</sup>. The duplex length is invariant between humans and yeast, suggesting that its formation mechanism is highly conserved.

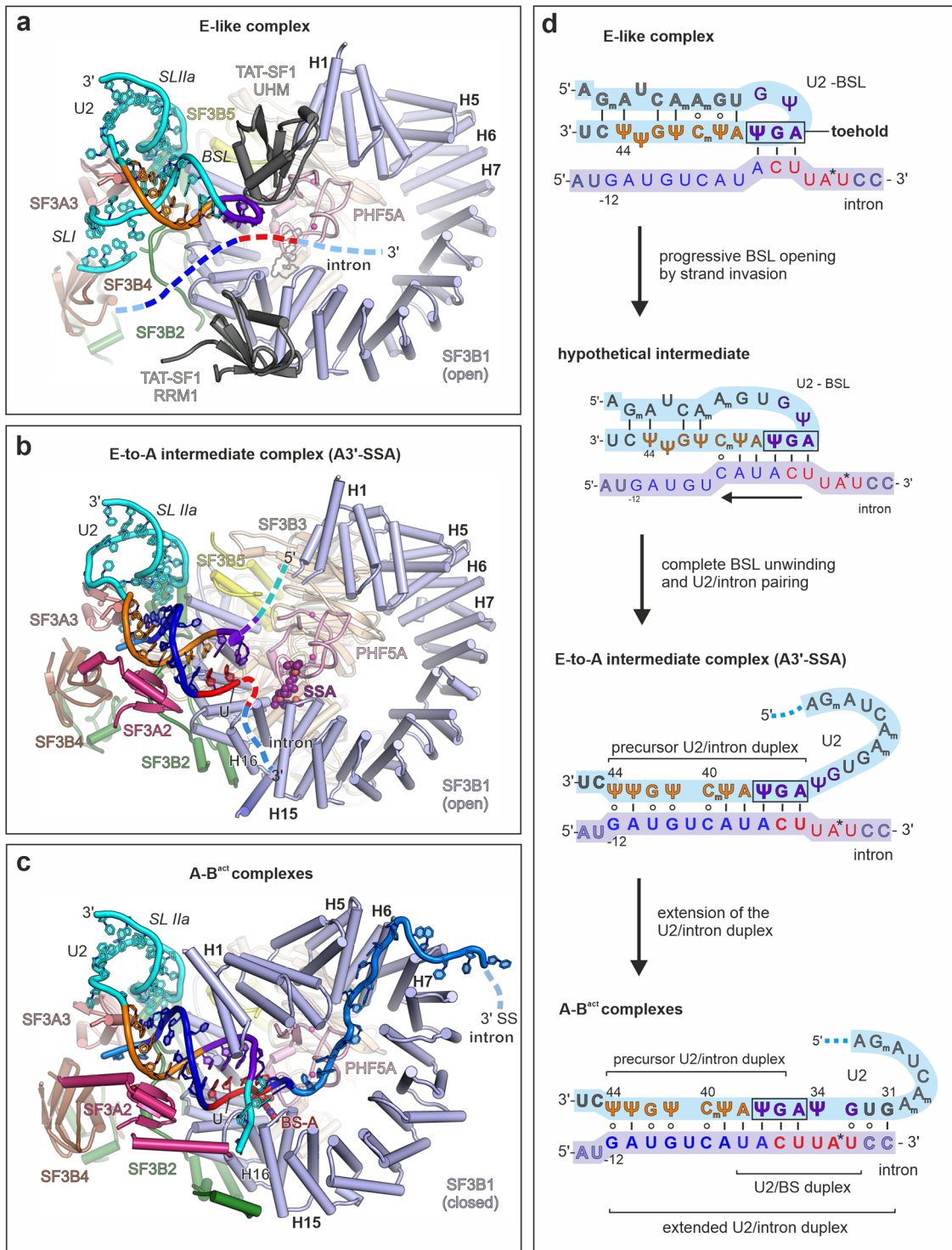
The structure of the A3'-SSA captures an E-to-A intermediate, indicating how the U2/intron duplex forms in at least two steps (Fig. 6). Initial binding of the 17S U2 snRNP requires unwinding of the U2 BSL and pairing to the intron region -12 to -2 upstream of the BS-A. The resulting precursor-U2/intron duplex, resembles BSL with respect to length and orientation (Fig. 6a, b). Remarkably, the last three base pairs of the precursor originate in the BSL's loop, consistent with previous observations<sup>22,23</sup>, indicating that the intron may establish its first contacts with BSL's loop region (A35-U37). It would then invade the BSL's stem asymmetrically, displacing the 5'-terminal moiety of the U2 snRNA from SF3B (Fig. 6). In this respect, the

mechanism of BSL's unwinding with concomitant U2/intron pairing is reminiscent of a toehold-mediated strand invasion reaction<sup>59–61</sup>, where three exposed bases of BSL act as an internal toehold (Fig. 6a, b, d). This type of strand exchange is initiated by a single-stranded oligomer (i.e., the invading strand) binding to an unpaired domain (the "toehold") of a double-stranded duplex (for a review see ref. 61). The toehold can also originate in the loop of a hairpin<sup>60</sup>, as might be the case of BSL (Supplementary Fig. 12).

The transition from the BSL to the precursor U2/intron duplex is likely facilitated by the stepwise action of PRP5/DDX46, TAT-SF1, and SF3A2 (Supplementary Fig. 4)<sup>23,24,62</sup>. Studies in yeast indicate that Prp5p's (human PRP5/DDX46) ATPase activity is required to displace Cus2p (human TAT-SF1) from SF3B1<sup>62</sup>. Consequently, it has been proposed that Prp5-dependent destabilization of Cus2p results in BSL's disruption, thereby enabling the extended pairing between U2 and the intron, during prespliceosome's formation<sup>62</sup>.

TAT-SF1 is present in the cryo-EM structure of the 17S U2 snRNP near the BSL<sup>22</sup>. For structural reasons, displacement of TAT-SF1 is required before the precursor U2/intron duplex can form in the A3'-SSA complex (Fig. 6a, b, d). Indeed, TAT-SF1 is absent from A3', suggesting that PRP5/DDX46 has already acted (Supplementary Fig. 4b–e). Consistently, while DDX46 is visible in the 17S U2 snRNP, we do not detect its density at the equivalent location of A3', suggesting relocation or destabilization (Supplementary Fig. 4a). Thus, DDX46 may induce the dissociation of TAT-SF1, which in turn liberates the BSL to create space





**Fig. 6 Mechanism of the stepwise intron selection by U2 snRNP.** **a** The E-like complex model is based on the 17S U2 snRNP structure (PDB 6Y50), which is positioned in the proximity of an intron schematic. The equivalent location of SSA is shown as a silhouette, for orientation's sake. **b** Structure of the U2 5' module of the A3' complex arrested with spliceostatin A (SSA). **c** Structure of the U2 5' module from the A/B<sup>act</sup> complexes. The depicted U2 5' model is derived from the human B<sup>act</sup> coordinates (PDB 5Z58). Complexes from **a** to **c** were superimposed over equivalent residues from PHF5A and SF3B1. **d** Schematic RNA transitions during intron's recognition by U2, as suggested by cryo-EM structures. Nucleotides resolved in cryo-EM structures are shown in bold letters and colored as in **a**. Ψ and the "m" subscript denotes pseudouridine and methylated nucleobases, respectively. Throughout the entire figure, the BS consensus and nucleotides from BSL's loop are shown in red and purple, respectively. The remaining sequence of the intron and of the U2 snRNA are colored in deep blue and orange, respectively. RNA strands not resolved in cryo-EM structures are dashed. All the other subunits and elements are color-coded as in Fig. 1.

for the strand exchange (Fig. 6)<sup>22</sup>. Next, SF3A2 would bind the precursor helix mostly via the sugar-phosphate backbone of U2, likely to aid in its formation and stabilization. Importantly, as SF3B1 remains in the open conformation and does not appear to contact RNA, SF3A2 may act as a placeholder to direct the extension of the precursor U2/intron duplex towards the hinged pocket of SF3B1 (Fig. 6b–d).

In the second step—the transition from A3′ to the A complex—the precursor extends with the U2/BS duplex, the bulged BS-A occupies the hinged pocket, SF3B1 clamps on the U2/BS duplex and the PPT binds firmly a deep channel framed by H4–H7 repeats (Fig. 6b, c). The U2/BS duplex extension is likely to precede SF3B1 clamping, given that strong BS can outcompete weaker modulators<sup>32,34</sup>. However, SF3B1’s function may still be required to stabilize the base-pairing interactions between U2’s G31–U34 and four intron nucleotides flanking the BS-A (−1 to +3), especially for “weak”, more degenerate intron substrates<sup>32,34</sup>.

By capturing a snapshot between E and A complexes, the structure of A3′-SSA indicates how introns are recognized and selected stepwise by the U2 snRNP. Our analyses suggest that following the initial U2 snRNP recruitment to the intron, the BSL unwinds via a strand-invasion mechanism to generate the precursor U2/intron duplex. Extension of this duplex enables subsequent selection of the BS-A and SF3B1’s clamping, thereby marking the intron’s commitment for splicing (Fig. 6).

**Structural basis of splicing modulation as a differential inactivation of branch sites.** The A3′-SSA structure reveals that FR901464 analogs, such as SSA or SD6, act at the latest stages of intron selection by the U2 snRNP by “locking” the SF3B1 subunit in an open state and preventing the formation of the extended U2/intron duplex. As covalent inhibitors, SSA and SD6 likely bind to and irreversibly modify any complex where SF3B1 exhibits the open conformation, including the isolated SF3B, the 17S U2 snRNP, or A3′-like complexes (Fig. 7). Once the coupling occurs, the covalently modified complexes are likely discarded from the splicing pathway. However, covalent inhibitors can be displaced from the binding pocket before the irreversible coupling occurs (i.e., during the initial binding of inhibitors to the target, when the two form a short-lived reversible intermediate)<sup>63,64</sup>. The competition between BS-A and inhibitors may occur during the conversion of A3′-like to A complexes, explaining why, similar to the pladienolide derivatives, SSA/SD6 elicits differential inhibition of splicing in cells (i.e., splicing modulation; Fig. 7 and Supplementary Fig. 13)<sup>31,32,42</sup>.

The differential response of splicing events to modulators *in vivo* correlates with the intron’s sequence and length, the conservation/strength of its BS and PPT, as well as the presence of additional motifs upstream of the BS<sup>31</sup>. Sequence-dependent effects were also observed *in vitro* for less potent SF3B modulators<sup>32,34</sup>. The structure of A3′-SSA shows that SSA’s binding to the hinged pocket allows the initial U2-intron pairing in the form of the precursor duplex while preventing its further extension (Fig. 6). This configuration can explain why degenerate BS motifs are more sensitive *in vivo* to SSA<sup>43</sup> or pladienolide derivatives<sup>31,34,35</sup>, as their reduced complementarity to U2 lowers the capacity to outcompete the modulator from the hinged pocket. Furthermore, our model postulates that the efficient extension of the precursor U2/intron duplex is facilitated by PPT-SF3B interactions, rationalizing why weaker/degenerate PPTs increase sensitivity to splicing modulators (Fig. 6b, c)<sup>31</sup>.

Very intriguingly, SSA can induce the U2 snRNA base-pairing upstream of the BS<sup>43</sup>, and the presence of alternative BS motifs upstream of the bonafide BS can lead to drug resistance<sup>31</sup>. Deletion of the BS-A from the upstream motifs eliminates drug resistance,

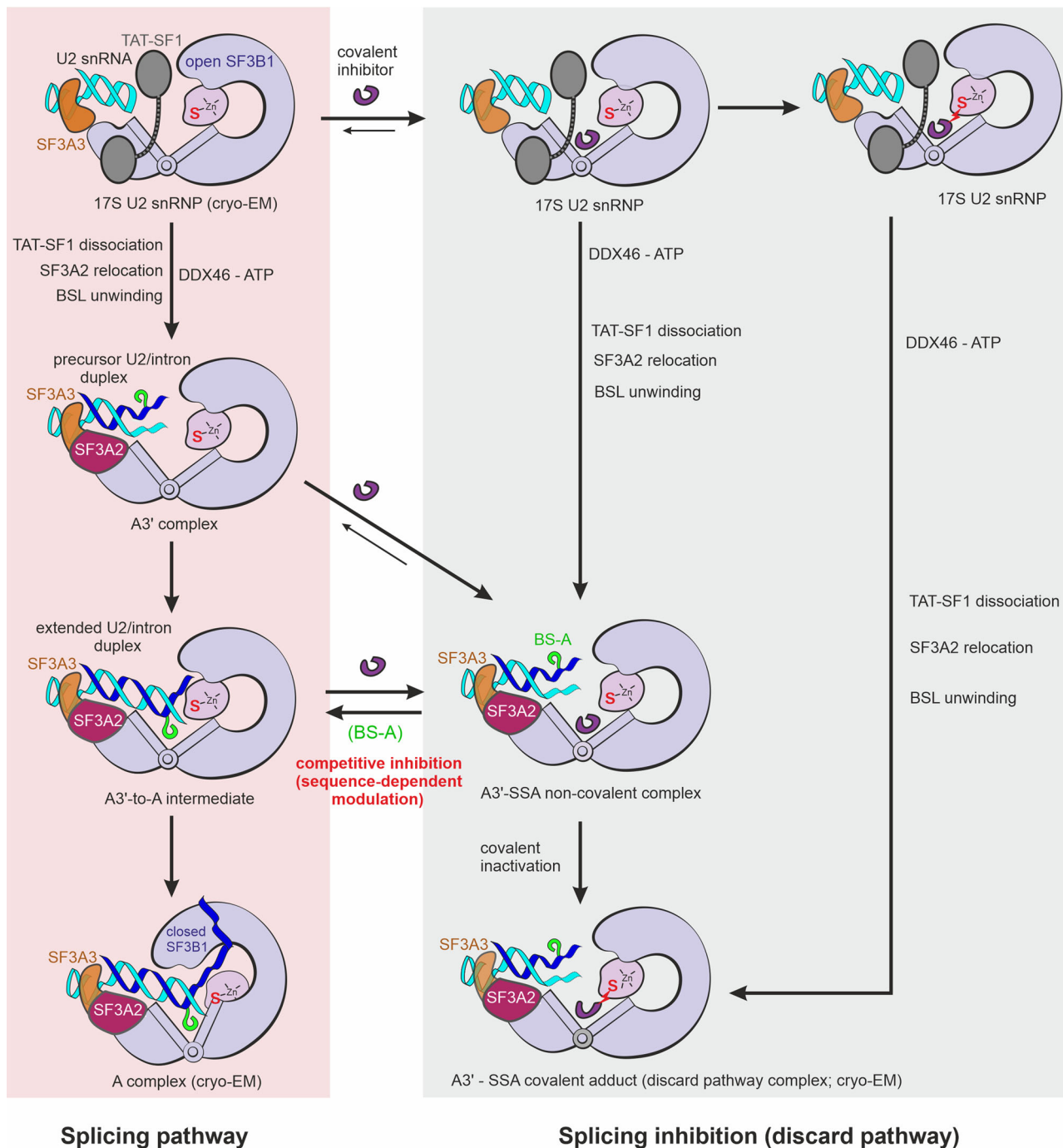
indicating a ligand-induced relocation of the U2 snRNP by an unknown mechanism<sup>31,43</sup>. The strand invasion mechanism, which we propose here, can explain how the U2 snRNP may relocate while maintaining permanent contact with the intron. Thus, when the bonafide BS is too weak/degenerate to outcompete a modulator, the U2 snRNP might slide back in an A3′-like state that continuously exchanges the intron sequence in a “scanning” process (Supplementary Fig. 12c). Once a “stronger BS” motif is detected upstream, the U2/BS helix extension occurs, with the BS-A being extruded from the RNA duplex and able to outcompete the modulator. Finally, the SF3B1 clamps on the U2/intron duplex to “lock” the U2 snRNP in place and signal commitment for splicing from the selected BS. From this perspective, the U2 snRNP would glide along the intron as a pulley device with auto-locking capabilities (Supplementary Fig. 12c). Further investigations are required to identify the molecular basis of this relocation and its potential significance for alternative splicing.

The A3′-SSA structure explains the primary mechanism of modulation as a complementarity-dependent competition between BS-A and the SF3B ligands (Fig. 7). The mechanism might be largely valid for all modulators that bind SF3B, including pladienolide and herboxidiene derivatives. Future comparative studies of prespliceosomes arrested with different compounds and on different introns might shed light on possible additional features that tune splicing modulation.

## Methods

**Engineering, expression, and purification of a minimized SF3B core complex lacking SF3B3’s BPB domain.** The first generation of SF3B core constructs encompassed the HEAT domain of SF3B1 (residues 453–1304), PHF5A (residues 1–98), SF3B3 (lacking the internal 1068–1085 residues), and full-length SF3B5, and retained an intact modulator binding pocket<sup>20,34</sup>. This minimal SF3B core packed into orthorhombic, plate-like crystals, which diffracted X-rays anisotropically to a resolution of ~2.9–3.1 Å<sup>34</sup>. As our attempts to crystallize this SF3B core variant in the presence of epoxy modulators (i.e., spliceostatin, sudemycins) or soak crystals with spliceostatins/sudemycins were not successful, we have engineered a second-generation construct lacking the more flexible BPB domain of SF3B3 (residues 442–772). SF3B3’s BPB domain does not establish contacts with the other SF3B subunits<sup>20</sup> and adopts different poses in various SF3B structures<sup>20,26,27</sup>. By making use of both the multiple cloning and of the Cre recombinase cassettes<sup>34</sup>, we, thus, constructed a single Multibac acceptor plasmid harboring all four SF3B core genes<sup>65</sup>, synthesized initially as codon-optimized genes for expression in insect cells<sup>20</sup>. Like in our previous SF3B constructs<sup>20,34</sup>, we dual-tagged the SF3B3 and the PHF5A subunits of the complex with cleavable 10xHIS and GST (Glutathione S-transferase) tags, respectively, and kept similar domain borders, except for SF3B3<sup>34</sup>. In the case of the SF3B3 subunit, in addition to the deletion of the internal 1068–1085 region, we substituted the entire BPB domain with a GGNGNSG linker by round-the-horn PCR mutagenesis (Supplementary Table 4).

In trial expression and protein purification experiments, we observed an apparent tendency of the SF3B<sup>ΔBPB</sup> core complex to aggregate at concentrations required for crystallization. We adapted our purification procedure (see ref. <sup>34</sup> for additional details on the previous purification protocol) by carrying out the final size-exclusion chromatography at a higher salt concentration and substituted the reducing agent for TCEP (Tris (2-carboxyethyl) phosphine). Briefly, Sf9 or High Five insect cells were infected with recombinant baculoviruses and cultured, as previously described<sup>34</sup>. The harvested cells were then gently resuspended with a Dounce homogenizer in the lysis buffer (50 mM HEPES-KOH pH 7.9, 600 mM KCl, 15% (v/v) glycerol, 4 mM DTT), supplemented with the cOMplete cocktail of protease inhibitors (Roche), and lysed by sonication on ice. The crude lysate was subsequently cleared by centrifugation and passed through a 0.45 μM sterile syringe filter before incubation with ~25–30 mL (bed volume) Glutathione Sepharose HP resin (GE Healthcare) for ~2 h at 4–8 °C. The affinity resin was subsequently collected in a gravity-flow column and washed thoroughly with the lysis buffer. The bound SF3B<sup>ΔBPB</sup> complex was eluted from the affinity resin with the elution buffer (50 mM HEPES-KOH pH 7.9, 500 mM KCl, 10% (v/v) glycerol, 2 mM DTT, 30 mM L-glutathione reduced) and digested with TEV (Tobacco Etch Virus) and HRV (Human Rhinovirus) 3 °C proteases overnight, at 4–8 °C. The salt concentration was then reduced by slow dilution with the dilution buffer (50 mM HEPES-KOH pH 7.9, 25 mM KCl, 20% (v/v) glycerol, 2 mM DTT) and the sample applied onto a 5 mL HiTrap Q Sepharose HP column (GE Healthcare) equilibrated in the buffer A (20 mM HEPES-KOH pH 7.9, 200 mM KCl, 10% (v/v) Glycerol, 1 mM TCEP). The sample was eluted from the column using a linear gradient (0–30%) formed between buffer A and buffer B (20 mM HEPES-KOH pH 7.9, 1 M KCl, 5% (v/v) glycerol, 1 mM TCEP). SF3B<sup>ΔBPB</sup> peak fractions devoid of nucleic



**Fig. 7 Model of the early splicing pathway under normal and drug-induced conditions.** The two pathways are presented in two columns, on distinct backgrounds. Note that modulators can enter the splicing pathway at various stages upon binding the hinged pocket in the open conformation of SF3B1. The competitive inhibition depends on the intron sequence at the BS and occurs before the chemical coupling of covalent inhibitors, thereby leading to differential inhibition of splicing (i.e., modulation). After covalent coupling, the complexes are irreversibly inactivated, resulting in global inhibition of splicing. Except for the indicated covalently inactivated states, the proposed schematic might be largely valid for non-covalent inhibitors as well, including pladienolide derivatives or SSE. The structures determined by cryo-EM are indicated. The SSA-bound 17S U2 snRNP complexes are inferred from the accessibility of the inhibitor's binding pocket within the cryo-EM structure of the 17S U2 snRNP (PDB 6Y50). The putative transitions between the SSA-bound 17S U2 snRNP and A3'-SSA are inferred from the apparent absence of steric interference between the bound inhibitor, the intron, and exchanged prespliceosome subunits.

acid contaminants were then concentrated by ultrafiltration using the Amicon Ultra-15 centrifugal filter (50 kDa MWCO, Merck) and further applied to a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) pre-equilibrated in the gel filtration buffer (20 mM HEPES-KOH pH 7.9, 400 mM KCl, 5% (v/v) glycerol, 1 mM TCEP). SF3B<sup>ΔBPB</sup> peak fractions were then concentrated by ultrafiltration to ~8–10 mg/mL, aliquoted and flash-frozen in liquid nitrogen, and stored at -80 °C.

As in the case of our previous recombinant SF3B core, SF3B<sup>ΔBPB</sup> had an apparent 1:1:1:1 stoichiometry and, as expected, eluted later from the size-exclusion column (with a retention volume of ~11.02 mL vs. ~10.63 mL)<sup>34</sup>. Overall, this indicates that the removal of the BPB domain of SF3B3 did not significantly affect the structural organization and the solution behavior of the remaining subunits or weaken/destabilize the protein-protein interaction interfaces of SF3B<sup>ΔBPB</sup>.

**Crystallization and structure determination of SF3B<sup>ΔBPP</sup> in complex with spliceostatin/sudemycin and pladienolide modulators.** To obtain structures of SF3B<sup>ΔBPP</sup> in complex with sudemycin/spliceostatin and pladienolide modulators, we have assembled the protein-ligand complexes in vitro by mixing the purified recombinant complex at 4–5 mg/mL final concentration with an excess of the small-molecule compounds. The splicing modulatory compounds used for structural studies (spliceostatin A (SSA), sudemycin D6 (SD6), spliceostatin E (SSE), and pladienolide D (PD)) were synthesized as previously described<sup>33,35,66,67</sup> and dissolved in DMSO (dimethyl sulfoxide). SSA and PD were added in a ~5-fold molar excess from the 10 mM DMSO stock, whereas the significantly weaker modulators SD6 and SSE were added in a ~10-fold excess over SF3B<sup>ΔBPP</sup>. The samples were then incubated on ice for ~1–12 h, centrifuged, and subjected to extensive crystallization trials. To this end, we explored more than ~2000 commercially available crystallization conditions and different crystallization setups. We have succeeded in further optimizing and collecting data from two new SF3B crystal forms belonging to the orthorhombic (*P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>*) and trigonal (*P3<sub>1</sub>21*) symmetry groups. Orthorhombic crystals of SF3B<sup>ΔBPP</sup> in complex with SSA, SD6, SSE, and PD were grown by vapor diffusion at 4 °C in 1 + 1 μL hanging drops equilibrated over a reservoir solution consisting of 0.1 M Bis-Tris propane pH 7.15–7.45, 0.2 M KSCN, 11–12% PEG-3350 (w/v), and 10–15% glycerol (v/v). These crystals were harvested at 4 °C and cryoprotected by stepwise transfer to the reservoir solution supplemented with 25% (v/v) ethylene glycol before flash-freezing in liquid nitrogen. Trigonal crystals of SF3B<sup>ΔBPP</sup> in complex with SSA and SD6 were obtained at 4 °C in 1 + 1 μL hanging drops placed over a reservoir solution containing 0.1 M HEPES pH 7.17–7.42, 0.2 M MgCl<sub>2</sub>, 25–27.5% (v/v) PEG-400. Before being snap-cooled in liquid nitrogen, the trigonal SF3B<sup>ΔBPP</sup> crystals were sequentially transferred to the corresponding reservoir solution containing 40% (v/v) PEG-400. Diffraction data were collected from single crystals at 100 K on the PILATUS 6M-F or EIGER2 16M detectors (Supplementary Tables 2 and 3), indexed, integrated, and scaled with XDS<sup>68</sup>, and further merged in AIMLESS<sup>69</sup>. The trigonal crystals of SF3B<sup>ΔBPP</sup> diffracted X-rays up to ~2.9–3.0 Å, as our previous SF3B core crystals<sup>34</sup>. Importantly, diffraction data collected from our best orthorhombic crystals extended to a resolution of ~2.3 Å. To our knowledge, this represents the highest resolution structural data available for the core of the human SF3B complex. The co-crystal structures of SF3B<sup>ΔBPP</sup> in the two different space-groups were phased by molecular replacement with Phaser<sup>70</sup> using a search model lacking the BPP domain of SF3B3. The models were then iteratively rebuilt in Coot<sup>71</sup> and refined with phenix.refine<sup>72</sup>. Data collection and refinement statistics are provided in Supplementary Tables 2 and 3. Geometric restraints for the refinement of spliceostatin/sudemycin and pladienolide modulators were generated with Grade (Global Phasing Limited) and edited in phenix.reel<sup>73</sup>. The SF3B modulators were located and modeled largely based on the residual *mFo-DFc* and polder omit maps. In addition, anomalous difference maps, calculated from diffraction data collected at the Zn K-edge (Supplementary Tables 2 and 3), as well as composite omit maps<sup>74</sup> were used to accurately rebuild the Zn1 motif of PHF5A (Supplementary Fig. 7b). Unexpectedly, the initial refinement of the SF3B<sup>ΔBPP</sup> structures in complex with SSA and SD6 (i.e., epoxy modulators) revealed a strong density element between the C26 residue of the PHF5A subunit and the epoxy-tetrahydropyran moiety of the small-molecule compounds. We interpreted the continuous positive density as an intermolecular covalent bond resulting from the reaction of C26 and with the epoxy group of SSA/SD6. We located the remaining tetrahydropyran/cyclohexane group and the aliphatic side-chain of the epoxy modulators in the BS-A binding pocket of SF3B, whereas the diene group is more exposed to the solvent (Supplementary Fig. 6). The terminal acetate group of SSA has weaker density and higher temperature factors compared to its neighboring groups, consistent with it being likely more susceptible to hydrolysis during crystallization and/or radiation damage. Pladienolide D (PD) was modeled and refined as detailed in our previous work<sup>34</sup>. Structural analyses were carried out in PyMOL version 2.3.3 (Schrodinger LLC) and the protein-ligand interactions were calculated with Arpeggio<sup>75</sup>. For crystallization, we used the same batch of SSA and SD6 characterized by Teng and co-workers<sup>33</sup>.

**Cloning and in vitro transcription of the A3'-exon pre-mRNA substrate.** The A3'-5'SS exon construct was synthesized as a gBLOCK (IDT, Supplementary Table 4) and cloned into a LIC compatible vector (438-C, Dr. Scott Gradia). A *Xba*I restriction site, downstream of the MS2 aptamer sequences, was used to linearize the plasmid and generate in vitro transcription templates. The final construct includes 54 nts of the MINX intron, upstream of the 3'SS, the second MINX exon, a "strong" 5'SS, followed by three consecutive MS2 aptamer sequences:

5'-GGGCGCAGUAGUCCAGGGUUUCCUUGAUGAUGUCAUACUUUUCG CUGUCCUUUUUUUCCACAGCUCGCGGUUGAGGACAAACUCUUCGC GGUCUUCCACAGGUAAGUUGGAAGCAUGAAGAACUUGAUGAUGCAU AUCCGUACACCAUACAGGGUACGAGCAGCCUAGCCUAGCCUACACCAUAG GGUACGACUAGUAGUACUUGCAUCCUAGGUAUGGUAUGGAAUUUCU-3'

MS2-tagged RNA substrates used for the assembly of A3'-5'SS exon complexes were prepared by T7 run-off transcription of the corresponding linearized DNA templates (i.e., obtained by digestion with *Xba*I). In the in vitro transcription reactions were set up and carried out as previously described<sup>17</sup>, except omitting the m<sup>7</sup>GpppG cap analog. The in vitro transcribed RNA was recovered using the MEGAclear kit (ThermoFisher Scientific) following the manufacturer's protocol.

The quality of the transcribed RNA templates was assessed on a denaturing PAGE gel stained with SYBR Gold, largely as described in ref. <sup>17</sup>.

#### Assembly and purification of the A3' complex arrested in the presence of SSA.

HeLa S3 cells were cultured in a bioreactor<sup>17</sup> and the HeLa nuclear extract was prepared as previously described, aliquoted, and stored in Roeder C buffer at -80 °C<sup>17</sup>. The A3'-exon complex was allowed to form in vitro in the HeLa nuclear extract upon the addition of MS2-tagged pre-mRNA substrates, while the progression of the splicing reaction was prevented by spliceostatin A (SSA).

Both trial and large-scale splicing reactions were set up on ice largely as previously described<sup>17</sup>. A typical splicing reaction contained 20 mM HEPES-KOH pH 7.9, 3 mM MgCl<sub>2</sub>, 2 mM ATP, 20 mM CP (Creatine Phosphate), 20% (v/v) HeLa nuclear extract (in Roeder C buffer) and 10 nM pre-mRNA substrate. SSA was added to the nuclear extract prior to the initiation of the splicing reaction to 500 nM final concentration, followed by incubation of the extract for 1 h on ice. In parallel, the in vitro transcribed pre-mRNA substrate was mixed with a ~20-fold molar excess of recombinant MBP-MS2, followed by incubation for 1 h on ice, as described in ref. <sup>17</sup>. Prior to the start of the reaction, the pre-mRNA substrate (10 nM final concentration) complexed with MBP-MS2 was added to the splicing mix and the splicing reaction was transferred immediately to a water bath equilibrated at 30 °C and incubated for 15'. The splicing reaction was then placed on ice for 5' and centrifuged at 4 °C for 20 min at 4000 rpm (2890×g). The supernatant was carefully decanted to a pre-chilled glass cylinder and loaded on a 5 mL MBPtrap column (GE Healthcare) at 0.5 mL/min equilibrated in the binding buffer (20 mM HEPES-KOH pH 7.9, 75 mM KCl, 1.5 mM MgCl<sub>2</sub>, 5% (v/v) glycerol, 1 mM DTT). The column was then washed extensively with the binding buffer and then with the washing buffer (20 mM HEPES-KOH pH 7.9, 75 mM KCl, 1.5 mM MgCl<sub>2</sub>, 1 mM DTT). The A3' exon complexes were eluted off the affinity column with the elution buffer (20 mM HEPES-KOH pH 7.9, 75 mM KCl, 1.5 mM MgCl<sub>2</sub>, 2 mM l-maltose). The A3'-exon affinity fractions were loaded separately onto a 14 mL linear 5–20% (w/v) sucrose gradient prepared in G-75 buffer (20 mM HEPES-KOH pH 7.9, 75 mM KCl, 1.5 mM MgCl<sub>2</sub>) and ultracentrifuged in a TST41.14 (Kontron) rotor at 22,000 rpm (~85,852×g) for 13 h 30' at 4 °C. The sucrose gradients were harvested manually from top to bottom in 0.5 mL fractions or bottom to top in 0.68 mL fractions and analyzed on a denaturing PAGE gel, followed by staining with SYBR Gold and Coomassie (Supplementary Fig. 1b). We used the previously described methods for mass-spectrometry sample preparation, digestion, measurement, and analysis<sup>17,27</sup> (see also Supplementary Data 1).

**Cryo-EM sample preparation.** To account for the relative lability of the A3'-SSA complex (i.e., compared to the later B, C, or P stage complexes), we reconstituted the early pre-spliceosomes under low salt conditions (50–75 mM KCl) and reduced the number of purification steps prior to cryo-EM sample preparation. In addition, a chemical crosslinking strategy was employed to stabilize the more labile samples prior to freezing them in vitreous ice.

The A3'-SSA complex was prepared in a similar manner as for mass-spectrometry analysis, except that the sample was subjected to GraFix after the amylose affinity step. That is, after the amylose affinity selection step, used to capture the assembled complex from the splicing reaction, individual fractions of the sample were applied to a 14 mL 5–20% (v/v) sucrose gradient in the presence of 0–0.1% glutaraldehyde and subjected to ultracentrifugation, as described<sup>76</sup>. In the next step, the sucrose gradient was harvested, and the remaining crosslinking agent was quenched with 50 mM l-aspartate<sup>76</sup>. A3'-SSA fractions containing the U1 and U2 snRNP in an apparent equimolar amount were gently rebuffered to G-75 buffer by ultrafiltration, concentrated to ~0.24–0.26 mg/mL, and then used directly for cryo-EM grid preparation. Although we screened different types of cryo-EM grids (e.g., R2/2, R1.2/1.3, C3.5/1) and grid freezing setups, a compromise between the ice thickness, density, distribution of particles, and image contrast was observed when the A3'-SSA sample was frozen in pure ice on UltrAufoil R1.2/1.3 gold grids (Plano). In this case, the A3'-SSA complex was frozen in vitreous ice by applying ~2.2 μL crosslinked sample to both sides of a glow-discharged grid mounted in a Vitrobot Mark IV (ThermoFisher Scientific) which was operated at 4 °C and 100% humidity. The excess sample was blotted away for 2 s using a blot force of 7 and the grids were snap-frozen in liquid ethane cooled by liquid nitrogen.

**Cryo-EM data collection and image processing.** Cryo-EM images of the A3'-SSA complex were collected using SerialEM<sup>77</sup> on a high-end Titan Krios transmission electron microscope (ThermoFisher Scientific) operated at 300 kV in EFTEM mode and equipped with a BioQuantum GIF energy filter (slit width 20 eV). We acquired 10,494 "good" movie stacks of the A3'-SSA complex from the same grid in two separate sessions. The cryo-EM micrographs were recorded on a K3 detector (Gatan) in counting mode at a magnification of 81,000× corresponding to a pixel size of 1.05 Å at specimen level (see also Supplementary Fig. 1 and Supplementary Table 1). Each of the A3'-SSA micrographs was acquired over an exposure time of ~2 s and was dose-fractionated into 40 frames, resulting in a total dose of ~41.41 e<sup>-</sup>/Å<sup>2</sup> and a dose rate of ~1.04 e<sup>-</sup>/Å<sup>2</sup>/frame.

The raw cryo-EM movie stacks were preprocessed on-the-fly in Warp (i.e., motion correction, dose-weighting, and CTF estimation) and particles were picked using a custom BoxNet neural net<sup>78</sup>, which was separately retrained against each of

the two A3'-SSA data sets. To account for the elongated nature of the complex and accommodate its different conformational states, we extracted the A3'-SSA particle images in a large 680 /680 px (714/714 Å) box. Data processing was largely performed in RELION-3.0 or RELION-3.1-beta<sup>79</sup> and cryoSPARC v2<sup>80</sup>.

The two data sets of the A3'-SSA particle images (in total, ~1.05 mln "particles") were initially binned 2× and subjected separately to 2D classification, followed by supervised 3D classification in cryoSPARC. Next, the cleaned subset of particles (351,004 particles) was combined and re-extracted in RELION and further classified in 2D. Importantly, 2D class averages of A3'-SSA show clear secondary structure features for the more rigid U2 5' module of the complex (Supplementary Fig. 1). The U2 3' module of the U2 snRNP appears dynamic and its density is blurred in 2D, as it was also recently observed for the isolated U2 17S particle<sup>22</sup>. An additional peripheral density element is located on the opposing side from U2 3' module. This density is less defined and, likely, contains other proteins detected in the proteome of the A3'-SSA complex. The subset of "good" particle images (i.e., displaying secondary structure features) resulting from 2D classification (164,107 particles) was further refined in RELION to obtain a reference map for the A3'-SSA, which was then low-passed to 60 Å and applied to all subsequent 3D refinement and classification steps. To potentially include less abundant views of the A3'-SSA, we went back to our initial 2× binned particle sets and performed 3D classification in RELION with 8 classes and the low-passed A3'-SSA reference volume obtained in the previous step. Particles representing the A3'-SSA complex were merged and then subjected to another round of 3D classification with image alignment and a soft mask applied on the U2 5' module of the complex. The resulting subset of 111,972 particles was refined in 3D to obtain a ~8.4 Å overall map of the A3'-SSA complex, as estimated using the gold-standard criterion of FSC (Fourier shell correlation) = 0.143 in RELION (Supplementary Fig. S1E). The quality of the overall reconstruction (map M1) was limited by the intrinsic structural flexibility of the complex and, likely, the biased orientation of particles in ice. To improve the local density of the individual modules of the complex, we resorted to focused classification and refinement with soft solvent masks applied to the U2 3' and U2 5' modules. After re-extracting, re-centering, and refining the particles at 1.05 Å/px in a 480 px box, we performed an additional 3D classification without image alignment with 6 classes and a soft mask applied on the U2 5' region. This classification approach led to a more conformational homogenous subset of particle images (78,262 particles) which was further subjected to iterative CTF refinement (per-particle defocus, per-micrograph astigmatism) and Bayesian polishing in RELION-3.1-beta<sup>79</sup>. The resulting ~3.1 Å map of the U2 5' module (Supplementary Fig. 2, map M2) allowed us to model a large part of the SF3B and SF3A subcomplexes, the precursor U2/intron duplex, and to locate the SSA inhibitor bound to an endogenous spliceosome complex. Refinement of the same subset of particles with a "loose" mask resulted in map M3 of the U2-5' module, which also includes the "top region" of the complex. 3D classification with a soft mask applied to the U2 core and further refinement led to a map (map M4) with improved features for the U2 3' domain (Supplementary Fig. 2) which allowed us to further extend our model.

**Cryo-EM model building and refinement.** To enable model building, the U2 5' module map was automatically sharpened in RELION by applying a negative B factor of approximately  $-54 \text{ \AA}^2$ . In addition, local map sharpening with LocScale<sup>81</sup>, as implemented in CCP-EM, and phenix.auto\_sharpen<sup>82</sup> provided alternative maps for interpreting and modeling of the more peripheral, low-resolution regions of the complex. Global and local resolution of the A3'-SSA maps were estimated in RELION (Supplementary Fig. 1).

We built an initial model for the U2-5' module of A3'-SSA, first by rigid-body fitting the ~2.3 Å crystal structure of the SF3B<sup>ABPB</sup>-SSA (Fig. 2 and Supplementary Fig. 5). The more peripheral SF3B4 (residues 12–89, RRM1) and SF3B2 (458–532, 566–666, 680–687 residues) subunits we modeled based on the cryo-EM structures of the human pre-B and B<sup>act</sup> complexes<sup>27,47</sup>, whereas the BPB domain of SF3B3 was built based on our previous structure of the SF3B core complex<sup>20</sup> (PDB 5IFE). The SF3A2 (residues 42–85, zinc-finger domain) and SF3A3 (residues 392–482, zinc-finger domain) subunits were docked separately with their initial poses obtained from the available models of the human B<sup>act</sup> complex<sup>27</sup> (PDB 6FF4/6FF7). U2 snRNA and the U2/intron precursor duplex were modeled based on the human B<sup>act</sup> complex (PDB 6FF4/6FF7) and manually adjusted to fit the cryo-EM density in Coot and ISOLDE. The model of the U2-5' module was refined with phenix.real\_space\_refine using manually curated base-pairing and stacking restraints, rebuilt, and further extended in Coot. The U2 3' model was obtained from the human pre-B complex<sup>47</sup> and fitted as a rigid body in Chimera in the corresponding map (map M4).

**Mass spectrometry and data analysis.** The SF3b complex (1 μM) was incubated in 20 mM HEPES pH 8.0, 200 mM KCl, 5% glycerol, and 1 mM TCEP with a 2-fold excess of compound (2 μM) at 4 °C overnight. Mass analyses were carried out on a Thermo Scientific Q-Exactive HRM (ESI source, 3.5 kV ionization voltage, 300 °C capillary temp., 55 arb sheath gas, aux gas flow rate at 5 L/min, S-lens RF level 50) coupled with Accela Open AS 1250. Samples (5 μL) were desalted on a C18 column (Thermo Scientific Accucore 2.1 × 150 mm, 2.6 μm) for 5 min prior to gradient run. Gradient started from 20 to 65% eluent B for 5 min. Eluent A consisted of 0.1% formic acid in water and eluent B consisted of 0.1% formic acid in acetonitrile. The

flow was set to 400 μL/min. All solvents were LC/MS grade (Thermo Scientific). The mass spectrometer was run in a positive mode collecting full scan at  $R = 70,000$  from  $m/z$  400 to  $m/z$  2000. Data collected with Xcalibur 3.1 software.

Xcalibur raw files were processed using BioPharma Finder 2.0 (Thermo Scientific) with Xtract deconvolution algorithm. Peak averaged over selected retention time to generate source spectra from TIC chromatogram trace, chromatogram parameters set to  $m/z$  400 to 2000. Outputs from the deconvolution algorithm include a mass range from 10,000 to 160,000 with a mass tolerance of 20 ppm and a charge state range from 10 to 100. The target mass is the estimated mass of protein or protein + compounds with a noise rejection of 95% confidence.

**Cell viability assays.** HCT116 PHF5A-C26H, HCT116 PHF5A-K29A, HCT116 PHF5A-K29R, HCT116 EV (empty vector) cell lines were generated by lentiviral overexpression as described<sup>33</sup>. The overexpression of PHF5A in blasticidin-selected cells was confirmed by western blotting. For CellTiter-Glo analysis, 2000 cells (100 μL) were seeded in each well of a 96-well plate. The next day, cells were treated with compounds at 10-point three-fold serial dilution starting with a top dose of 10 μM. 72 h post compound addition, CellTiter-Glo reagent (100 μL) was added to the cell medium, incubated, and the signal was measured on EnVision 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 8 (GraphPad) and fit using non-linear regression analysis.

**Scintillation proximity assays for competitive binding.** SF3B core complexes (wild-type and PHF5A-C26H) used for SPA were purified as described before<sup>34</sup>. For SPA, the anti-FLAG antibody (Sigma-Aldrich) was immobilized to anti-mouse PVT SPA beads (PerkinElmer) as described<sup>34</sup>. 100 μL binding reactions were prepared with 50 μL bead slurry and 25 nM purified SF3B core complex in buffer (20 mM HEPES pH 8, 200 mM KCl, 5% glycerol), and different compounds in a 10-point three-fold serial dilution with the top concentration of 4 μM were used. After the mixture was pre-incubated for 30 min, 10 nM [<sup>3</sup>H]-probe ([<sup>3</sup>H]-labeled pladienolide B) was added to the mixture and incubated for 30 min, and luminescence signals were read using a MicroBeta2 Plate Counter (PerkinElmer). Prism 8 (GraphPad) was used for the non-linear regression curve fitting of the data.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Coordinates and structure factors were deposited in the Protein Data Bank (PDB) accession codes: 7B9C, 7B0I, 7B9I, 7B9Z, 7OMF, 7OPI. The cryo-EM maps and the associated coordinates were deposited in the Electron Microscopy Data Bank (EMDB) accession code: EMD-12994 and the Protein Data Bank (PDB) accession code: 7ONB. Requests for samples, materials, and data should be addressed to vlad.pena@icr.ac.uk. Source data are provided with this paper.

Received: 22 December 2020; Accepted: 1 July 2021;

Published online: 23 July 2021

## References

- Wahl, M. C., Will, C. L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).
- Will, C. L. & Luhrmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, <https://doi.org/10.1101/cshperspect.a003707> (2011).
- Lee, Y. & Rio, D. C. Mechanisms and regulation of alternative pre-mRNA splicing. *Annu. Rev. Biochem.* **84**, 291–323 (2015).
- Gao, K., Masuda, A., Matsuura, T. & Ohno, K. Human branch point consensus sequence is yUnAy. *Nucleic Acids Res.* **36**, 2257–2267 (2008).
- Taggart, A. J. et al. Large-scale analysis of branchpoint usage across species and cell lines. *Genome Res.* **27**, 639–649 (2017).
- Pineda, J. M. B. & Bradley, R. K. Most human introns are recognized via multiple and tissue-specific branchpoints. *Genes Dev.* **32**, 577–591 (2018).
- Abovich, N. & Rosbash, M. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89**, 403–412 (1997).
- Berglund, J. A., Rosbash, M. & Schultz, S. C. Crystal structure of a model branchpoint-U2 snRNA duplex containing bulged adenosines. *RNA* **7**, 682–691 (2001).
- Liu, Z. et al. Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294**, 1098–1102 (2001).
- Schneider, M. et al. Exon definition complexes contain the tri-snRNP and can be directly converted into B-like pre-catalytic splicing complexes. *Mol. Cell* **38**, 223–235 (2010).

11. Wongpalee, S. P. et al. Large-scale remodeling of a repressed exon ribonucleoprotein to an exon definition complex active for splicing. *Elife* **5**, <https://doi.org/10.7554/eLife.19743> (2016).
12. van der Feltz, C. & Hoskins, A. A. Structural and functional modularity of the U2 snRNP in pre-mRNA splicing. *Crit. Rev. Biochem. Mol. Biol.* **54**, 443–465 (2019).
13. Kastner, B., Will, C. L., Stark, H. & Luhrmann, R. Structural Insights into nuclear pre-mRNA splicing in higher eukaryotes. *Cold Spring Harb. Perspect. Biol.* **11**, <https://doi.org/10.1101/cshperspect.a032417> (2019).
14. Wilkinson, M. E., Charenton, C. & Nagai, K. RNA splicing by the spliceosome. *Annu. Rev. Biochem.* **89**, 359–388 (2020).
15. Kramer, A., Gruter, P., Groning, K. & Kastner, B. Combined biochemical and electron microscopic analyses reveal the architecture of the mammalian U2 snRNP. *The J. Cell Biol.* **145**, 1355–1368 (1999).
16. Will, C. L. et al. 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 (2002).
17. Bertram, K. et al. Cryo-EM structure of a pre-catalytic human spliceosome primed for activation. *Cell* **170**, 701–713.e711 (2017).
18. Plaschka, C., Lin, P. C. & Nagai, K. Structure of a pre-catalytic spliceosome. *Nature* **546**, 617–621 (2017).
19. Darman, R. B. et al. Cancer-associated SF3B1 hotspot mutations induce cryptic 3' splice site selection through use of a different branch point. *Cell Rep.* **13**, 1033–1045 (2015).
20. Cretu, C. et al. Molecular architecture of SF3b and structural consequences of its cancer-related mutations. *Mol. Cell* **64**, 307–319 (2016).
21. Fica, S. M. & Nagai, K. Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. *Nat. Struct. Mol. Biol.* **24**, 791–799 (2017).
22. Zhang, Z. et al. Molecular architecture of the human 17S U2 snRNP. *Nature* **583**, 310–313 (2020).
23. Perriman, R. & Ares, M. Jr. Invariant U2 snRNA nucleotides form a stem loop to recognize the intron early in splicing. *Mol. Cell* **38**, 416–427 (2010).
24. Liang, W. W. & Cheng, S. C. A novel mechanism for Prp5 function in prespliceosome formation and proofreading the branch site sequence. *Genes Dev.* **29**, 81–93 (2015).
25. Tang, Q. et al. 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 (2016).
26. Rauhut, R. et al. Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. *Science* **353**, 1399–1405 (2016).
27. Haselbach, D. et al. Structure and conformational dynamics of the human spliceosomal B(act) complex. *Cell* **172**, 454–464.e411 (2018).
28. Bonnal, S., Vigevani, L. & Valcarcel, J. The spliceosome as a target of novel antitumor drugs. *Nat. Rev. Drug Discov.* **11**, 847–859 (2012).
29. Effenberger, K. A., Urabe, V. K. & Jurica, M. S. Modulating splicing with small molecular inhibitors of the spliceosome. *Wiley Interdiscip. Rev. RNA* **8**, <https://doi.org/10.1002/wrna.1381> (2017).
30. Salton, M. & Misteli, T. Small molecule modulators of pre-mRNA splicing in cancer therapy. *Trends Mol. Med.* **22**, 28–37 (2016).
31. Vigevani, L., Gohr, A., Webb, T., Irimia, M. & Valcarcel, J. Molecular basis of differential 3' splice site sensitivity to anti-tumor drugs targeting U2 snRNP. *Nat. Commun.* **8**, 2100 (2017).
32. Seiler, M. et al. H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. *Nat. Med.* **24**, 497–504 (2018).
33. Teng, T. et al. Splicing modulators act at the branch point adenosine binding pocket defined by the PHF5A-SF3b complex. *Nat. Commun.* **8**, 15522 (2017).
34. Cretu, C. et al. Structural basis of splicing modulation by antitumor macrolide compounds. *Mol. Cell* **70**, 265–273.e268 (2018).
35. Finci, L. I. et al. The cryo-EM structure of the SF3b spliceosome complex bound to a splicing modulator reveals a pre-mRNA substrate competitive mechanism of action. *Genes Dev.* **32**, 309–320 (2018).
36. Nakajima, H. et al. New antitumor substances, FR901463, FR901464 and FR901465. I. Taxonomy, fermentation, isolation, physico-chemical properties and biological activities. *J. Antibiot.* **49**, 1196–1203 (1996).
37. Thompson, C. F., Jamison, T. F. & Jacobsen, E. N. FR901464: total synthesis, proof of structure, and evaluation of synthetic analogues. *J. Am. Chem. Soc.* **123**, 9974–9983 (2001).
38. Osman, S. et al. Structural requirements for the antiproliferative activity of pre-mRNA splicing inhibitor FR901464. *Chemistry* **17**, 895–904 (2011).
39. Nicolaou, K. C., Rhoades, D. & Kumar, S. M. Total syntheses of thailanstatins A–C, spliceostatin D, and analogues thereof. Stereodivergent synthesis of tetrasubstituted dihydro- and tetrahydropyrans and design, synthesis, biological evaluation, and discovery of potent antitumor agents. *J. Am. Chem. Soc.* **140**, 8303–8320 (2018).
40. Yoshikawa, Y. et al. Design and synthesis of 1,2-deoxy-pyranose derivatives of spliceostatin A toward prostate cancer treatment. *ACS Med. Chem. Lett.* **11**, 1310–1315 (2020).
41. Kaida, D. et al. Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat. Chem. Biol.* **3**, 576–583 (2007).
42. Yoshimoto, R. et al. Global analysis of pre-mRNA subcellular localization following splicing inhibition by spliceostatin A. *RNA* **23**, 47–57 (2017).
43. Corriero, A., Minana, B. & Valcarcel, J. Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. *Genes Dev.* **25**, 445–459 (2011).
44. Plaschka, C., Lin, P. C., Charenton, C. & Nagai, K. Prespliceosome structure provides insights into spliceosome assembly and regulation. *Nature* **559**, 419–422 (2018).
45. Behzadnia, N. et al. Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO J.* **26**, 1737–1748 (2007).
46. Zhang, X. et al. Structure of the human activated spliceosome in three conformational states. *Cell Res.* **28**, 307–322 (2018).
47. Charenton, C., Wilkinson, M. E. & Nagai, K. Mechanism of 5' splice site transfer for human spliceosome activation. *Science* **364**, 362–367 (2019).
48. Albert, B. J. et al. Meayamycin inhibits pre-messenger RNA splicing and exhibits picomolar activity against multidrug-resistant cells. *Mol. Cancer Ther.* **8**, 2308–2318 (2009).
49. Fan, L., Lagiseti, C., Edwards, C. C., Webb, T. R. & Potter, P. M. Sudemycins, novel small molecule analogues of FR901464, induce alternative gene splicing. *ACS Chem. Biol.* **6**, 582–589 (2011).
50. Webb, T. R., Joyner, A. S. & Potter, P. M. The development and application of small molecule modulators of SF3b as therapeutic agents for cancer. *Drug Discov. Today* **18**, 43–49 (2013).
51. Makowski, K. et al. A synthetic antitumor splicing inhibitor variant with improved activity and versatile chemistry. *ACS Chem. Biol.* **12**, 163–173 (2017).
52. Effenberger, K. A., Urabe, V. K., Prichard, B. E., Ghosh, A. K. & Jurica, M. S. Interchangeable SF3B1 inhibitors interfere with pre-mRNA splicing at multiple stages. *RNA* <https://doi.org/10.1261/rna.053108.115> (2016).
53. Konuray, A. O. Analysis of the reaction mechanism of the thiol-epoxy addition initiated by nucleophilic tertiary amines. *Polym. Chem.* <https://doi.org/10.1039/c7py01263b> (2017).
54. Skarzynski, T. et al. Structure of UDP-N-acetylglucosamine enolpyruvyl transferase, an enzyme essential for the synthesis of bacterial peptidoglycan, complexed with substrate UDP-N-acetylglucosamine and the drug fosfomycin. *Structure* **4**, 1465–1474 (1996).
55. Coleman, J. E. Zinc enzymes. *Curr. Opin. Chem. Biol.* **2**, 222–234 (1998).
56. Rice, W. G. et al. Azodicarbonamide inhibits HIV-1 replication by targeting the nucleocapsid protein. *Nat. Med.* **3**, 341–345 (1997).
57. Lagiseti, C. et al. Antitumor compounds based on a natural product consensus pharmacophore. *J. Med. Chem.* **51**, 6220–6224 (2008).
58. Leon, B. et al. A challenging pie to splice: drugging the spliceosome. *Angew. Chem. Int. Ed. Engl.* **56**, 12052–12063 (2017).
59. Sulc, P., Ouldrige, T. E., Romano, F., Doye, J. P. & Louis, A. A. Modelling toehold-mediated RNA strand displacement. *Biophys. J.* **108**, 1238–1247 (2015).
60. Green, S. J., Lubrich, D. & Turberfield, A. J. DNA hairpins: fuel for autonomous DNA devices. *Biophys. J.* **91**, 2966–2975 (2006).
61. Guo, Y. et al. Recent advances in molecular machines based on toehold-mediated strand displacement reaction. *Quant. Biol.* **5**, 25–41 (2017).
62. Talkish, J. et al. Cus2 enforces the first ATP-dependent step of splicing by binding to yeast SF3b1 through a UHM-ULM interaction. *RNA* **25**, 1020–1037 (2019).
63. Baillie, T. A. Targeted covalent inhibitors for drug design. *Angew. Chem. Int. Ed. Engl.* **55**, 13408–13421 (2016).
64. Gehringer, M. & Laufer, S. A. Emerging and re-emerging warheads for targeted covalent inhibitors: applications in medicinal chemistry and chemical biology. *J. Med. Chem.* **62**, 5673–5724 (2019).
65. Fitzgerald, D. J. et al. Protein complex expression by using multigene baculoviral vectors. *Nat. Methods* **3**, 1021–1032 (2006).
66. Effenberger, K. A. et al. 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 (2014).
67. Ghosh, A. K. et al. Enantioselective synthesis of spliceostatin E and evaluation of biological activity. *Org. Lett.* **16**, 6200–6203 (2014).
68. Kabsch, W. Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).
69. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
70. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 32–41 (2007).

71. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
72. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **66**, 213–221 (2010).
73. Moriarty, N. W., Draizen, E. J. & Adams, P. D. An editor for the generation and customization of geometry restraints. *Acta Crystallogr. D. Struct. Biol.* **73**, 123–130 (2017).
74. Terwilliger, T. C. et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64**, 61–69 (2008).
75. Jubb, H. C. et al. Arpeggio: a web server for calculating and visualising interatomic interactions in protein structures. *J. Mol. Biol.* **429**, 365–371 (2017).
76. Kastner, B. et al. GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55 (2008).
77. Mastrorarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
78. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019).
79. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, <https://doi.org/10.7554/eLife.42166> (2018).
80. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
81. Jakobi, A. J., Wilmanns, M. & Sachse, C. Model-based local density sharpening of cryo-EM maps. *Elife* **6**, <https://doi.org/10.7554/eLife.27131> (2017).
82. Terwilliger, T. C., Sobolev, O. V., Afonine, P. V. & Adams, P. D. Automated map sharpening by maximization of detail and connectivity. *Acta Crystallogr. D.* **74**, 545–559 (2018).

## Acknowledgements

We thank Anuschka Pauluhn from the X10SA beamline (Swiss Light Source, PSI Villigen, Switzerland) and Christian Dienemann from the EM facility of the Max Planck Institute for Biophysical Chemistry for assistance with data collection and processing. We thank Klaus Hartmuth and Claudia Höbartner for their advice and helpful discussions. Funding for this study was provided by the German Research Foundation, DFG grant PE 2079/4-1 (C.C. and V.P.), NIGMS grant 122279 (M.J. and A.K.G.), the Institute of Cancer Research (C.C. and V.P.), and the Max-Planck Society (V.P.).

## Author contributions

C.C. designed and performed the reconstitution and purification of complexes for structural studies and performed all steps of X-ray crystallography and cryo-EM work in

the laboratory of V.P. C.C. and V.P. performed structural analyses. T.-V.N. performed mass-spectrometric experiments, X.L. under the supervision of A.C. synthesized SSA, P. G. under the supervision of N.A.L. purified the wild-type SF3B core and the mutant variants for SPA binding studies and intact mass spectrometry. A.K.G. and M.S.J. synthesized the SSA and SSE compounds used in structural studies. V.P. supervised the project. Correspondence for the manuscript should be addressed to V.P. (vlad.pena@icr.ac.uk).

## Competing interests

C.C., A.K.G., M.J., and V.P. declare no competing interests. T.-V.N., X.L., A.A., A.C., P.G., and N.A.L. are employees of H3 Biomedicine, Inc.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-24741-1>.

**Correspondence** and requests for materials should be addressed to V.P.

**Peer review information** *Nature Communications* thanks Kazunori Koide and other, anonymous, reviewers for their contributions to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© Crown 2021