Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Dynamic interactions drive early spliceosome assembly

Santiago Martínez-Lumbreras[1,2], Clara Morguet[1,2] and Michael Sattler[1,2]

Splicing is a critical processing step during pre-mRNA maturation in eukaryotes. The correct selection of splice sites during the early steps of spliceosome assembly is highly important and crucial for the regulation of alternative splicing. Splice site recognition and alternative splicing depend on *cis*-regulatory sequence elements in the RNA and *trans*-acting splicing factors that recognize these elements and crosstalk with the canonical splicing machinery. Structural mechanisms involving early spliceosome complexes are governed by dynamic RNA structures, protein-RNA interactions and conformational flexibility of multidomain RNA binding proteins. Here, we highlight structural studies and integrative structural biology approaches, which provide complementary information from cryo-EM, NMR, small angle scattering, and X-ray crystallography to elucidate mechanisms in the regulation of early spliceosome assembly and quality control, highlighting the role of conformational dynamics.

Addresses
[1] Helmholtz Munich, Molecular Targets and Therapeutics Center, Institute of Structural Biology, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany
[2] Technical University of Munich, TUM School of Natural Sciences, Bavarian NMR Center and Department of Bioscience, Lichtenbergstrasse 4, 85747 Garching, Germany

Corresponding author: Sattler, Michael (michael.sattler@helmholtz-munich.de)

Current Opinion in Structural Biology 2024, **88**:102907

This review comes from a themed issue on **Protein Nucleic Acid Interactions (2024)**

Edited by **Junji Iwahara** and **David C. Williams Jr.**

For complete overview of the section, please refer the article collection - Protein Nucleic Acid Interactions (2024)

Available online 20 August 2024

https://doi.org/10.1016/j.sbi.2024.102907

Introduction

Splicing of most eukaryotic pre-mRNAs is orchestrated by the spliceosome, a megadalton complex composed of five small nuclear ribonucleoprotein particles (snRNPs) that dynamically assembles *de novo* on each pre-mRNA molecule to catalyze the removal of the intronic regions and produce the uninterrupted coding sequence for protein translation. The exon-intron boundaries are defined by the recognition of the 5′ splice site by U1 snRNP, and the binding of splicing factor 1 (SF1) and U2 auxiliary factors to the branch point (BP) site, the polypyrimidine (PY) tract and the 3′ splice site, respectively. This promotes binding of the U2 snRNP at the 3′ intronic region with subsequent assembly of the U4/U6/U5 tri-snRNP for the activation of a catalytic complex, which performs the two transesterification reactions joining two exon sequences and removing the intron in the lariat form (Box 1) [1–3].

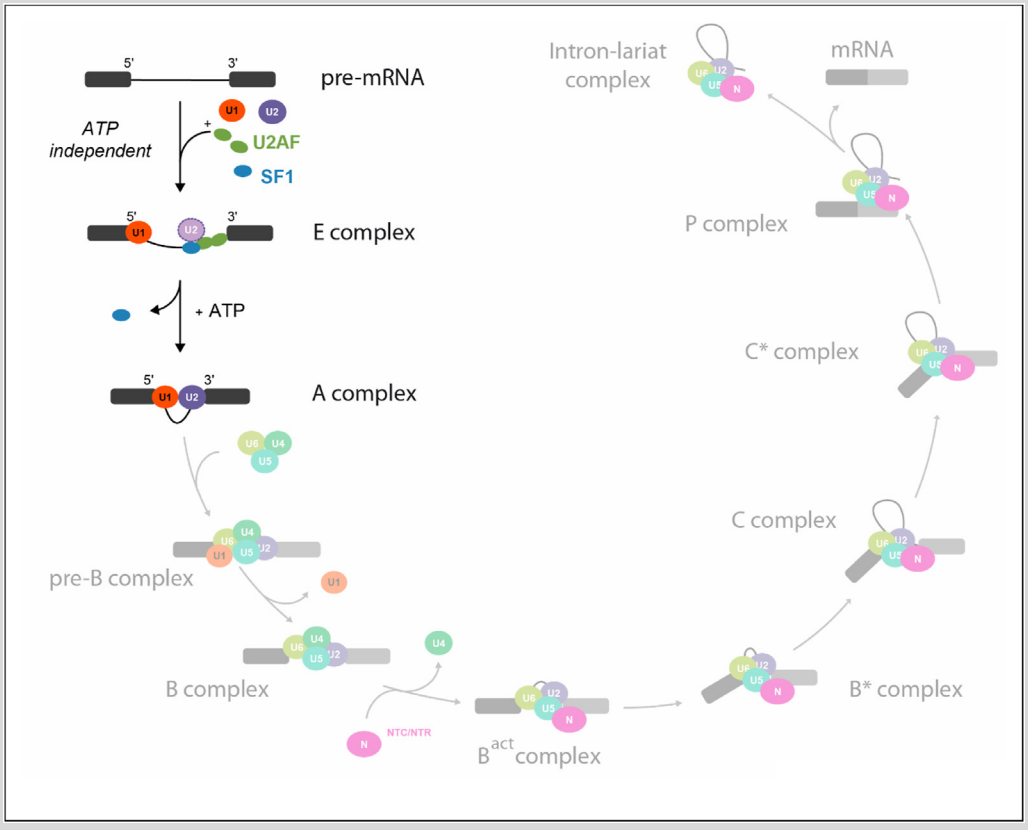
In higher eukaryotes, most pre-mRNAs can be spliced in variable ways, resulting in several isoforms encoded in a single gene, which greatly expands the diversity of the proteome over the genome [4]. This process, called alternative splicing (AS), is essential for cell differentiation and tissue identity acquisition during development [5], and its misregulation by mutations in *cis*-regulatory RNA sequence elements or *trans*-acting proteins is directly linked to several diseases, including cancer [4,6].

The resolution revolution generated by the advances in cryogenic electron microscopy (cryo-EM), together with a wealth of biochemical knowledge and availability of small molecule inhibitors that stall the spliceosome at different steps, provided unprecedented insight into structural snapshots of the splicing cycle [7]. Transitions between the individual steps can involve large rearrangements of complexes, mediated by various helicases [8]. Understanding the underlying structural transitions and mechanisms is challenging based only on the initial and final steps. While this has been alleviated by the capture of intermediate species that help to rationalize some of these transitions, characterizing dynamic states and conformational ensembles linked to spliceosome assembly remains challenging. Moreover, *trans*-acting factors involved in the very early steps of constitutive splicing and in the regulation of AS are often multidomain proteins, where individual domains are connected by flexible linkers [9], enabling conformational dynamics that are crucial for their function but escape

Box 1. Glossary of key terms and players

Spliceosome cycle

About 99% of splicing events are carried out by the major spliceosome (U2-dependent), in contrast to the small subset regulated by the minor spliceosome (U12-dependent). The spliceosome is assembled on the pre-mRNA, *de novo* for each splicing event, following the splicing cycle that involves the recognition of the splice sites during E to A complexes, the preparation for the catalytic reaction in B complexes and the performance of the two transesterification reactions, first linking the 5′ of the intron to the adenine in the branch point and then ligating the excised 3′ end of the first exon with the beginning of the second exon in complexes B* to C to remove the intron in the lariat form.



Schematic of the splicing cycle with the different spliceosome states. The early complexes (E, A) in splicing assembly are highlighted.

5′ splice site recognition

Human U1 snRNP consists of U1 snRNA and several proteins: the Sm ring (with seven Sm proteins B/B′, D1, D2, D3, E, F and G) that forms around the U1 snRNA, the 70K protein that interacts with stem loop 1, U1-A protein that interacts with stem loop 2 and U1−C protein that binds the duplex formed by the 5′ region of the U1 snRNA and the 5′ splice site. U1 snRNP is loaded during the E complex formation and remains bound in the A complex until the U1 snRNA/5′ splice site duplex is unwound by the helicase PRP28. The 5′ splice site RNA is subsequently transferred to the U6 snRNP in the B complex and U1 snRNP is released.

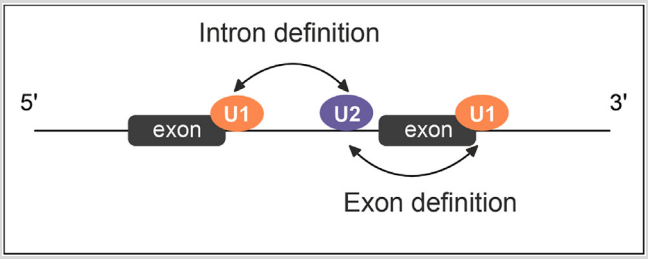
3′ splice site recognition

The canonical view of the recognition of the 3′ intronic region involves the initial identification of three *cis*-elements in the RNA by three splicing factors in the E complex of a major subset of introns. The branch point sequence is specifically recognized by SF1; upon A complex formation, SF1 is released, and the BP RNA binds the U2 snRNP by base pairing with the U2 snRNA, to later initiate the first transesterification reaction. The 3′ splice site is initially identified through recognition of the polypyrimidine tract and the 3′ splice site AG by U2AF2 and U2AF1, respectively. Protein−protein interactions involving UHM/ULM interactions further stabilize the assembly of SF1−U2AF2-U2AF1 at the 3′ intronic region. The interaction of the SF1 ULM with U2AF2 is replaced by the binding to SF3B1 ULMs in the U2 snRNP in the A complex.

Exon/intron definition

Exon or intron definition refers to mechanisms by which a strong splice site can recruit the spliceosome machinery to an adjacent weaker splice site. For example, a canonical 3′ intronic region (with a conserved 3′ splice site, a strong PY tract and a canonical BP) would promote the recognition of an adjacent 5′ splice site with a degenerated motif. This enhanced recruitment is driven by the presence of bridge proteins that

connect both sites; if the bridge is done through the exon, the mechanism is named exon definition, while if the connection is made through the intron, the mechanism is intron definition. In humans, exon definition is thought to be the main driving force for the recognition of the adjacent sites with relatively short exons and introns of variable length (up to 10^5 nt).



Schematic representation of the exon and intron definition mechanisms.

detection by cryo-EM or X-ray crystallography. Integrating solution-state techniques like nuclear magnetic resonance (NMR) spectroscopy or small angle scattering (SAS) thus provides important complementary information on conformational dynamics. Regulation of alternative splicing in a cell- and tissue-dependent manner is controlled by a complex splicing code [10,11], which is governed by the presence of *cis*-regulatory motifs and pre-mRNA structure, expression levels and interactions with RNA binding proteins, depending on cell type and developmental stage. The complex interplay of these factors and conformational dynamics involved allow a graduated and fine-tuned regulation of alternative splicing, which can be impaired by disease-linked mutations in the *cis*-regulatory motifs or *trans*-acting factors involved. Understanding mechanisms of splicing regulation also enables the development of splicing modulators for therapeutic approaches for human diseases, as highlighted by recent approvals of antisense oligonucleotide and small molecule drugs [12–14]. In this review, we summarize how recent findings on structures and dynamics in early spliceosome assembly expand our understanding of splicing regulation and proofreading in health and disease.

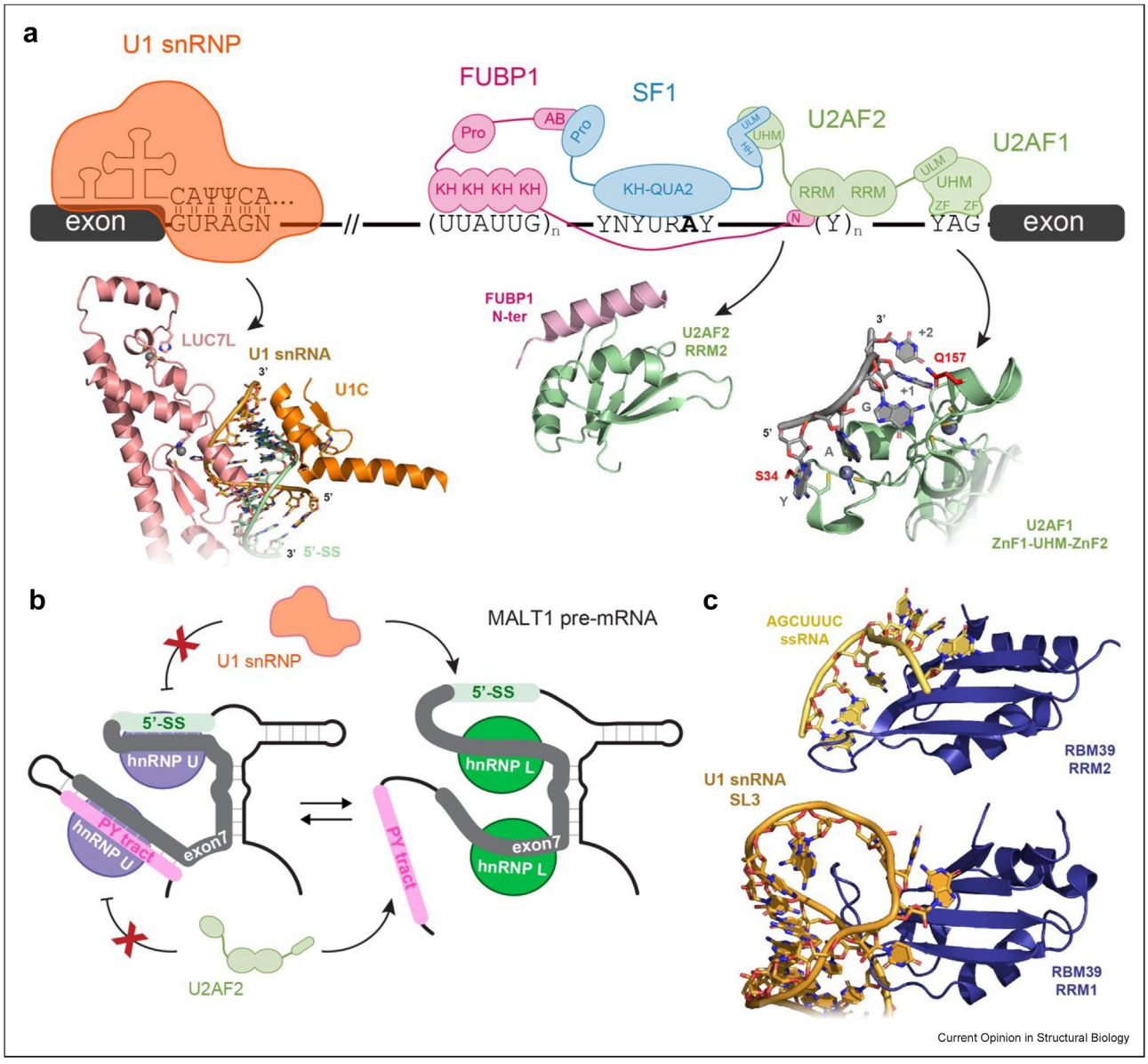
Early recognition of splice sites

Early recognition during the initiation of spliceosome assembly (E complex formation) involves the initial identification of exon-intron boundaries (Box 1). U1 snRNP directly recognizes the 5′ splice site, via base pairing of the U1 snRNA with the beginning of the intronic region. Concomitantly, splicing factor 1 (SF1) and the heterodimeric U2 snRNP auxiliary factor U2AF, comprised of U2AF2 and U2AF1, recognize the BP site, the PY-tract and the 3′ splice site of the major class of introns, respectively (Figure 1a). RNA recognition involves dedicated RNA-binding domains within these proteins: KH-QUA2 in SF1, two RNA recognition motifs (RRMs) in U2AF2 and two zinc fingers (ZnFs) in U2AF1. The recognition of the 3′ region is further modulated by protein–protein interactions involving UHM-ULM interactions [1,15,16]. Quantitative variations in the binding affinity of cognate splicing factors with *cis*-regulatory RNA motifs and conformational dynamics allow a graduated quantitative regulation of splicing [9,17]. The initial identification of intron/exon boundaries is key for accurately defining the regions that undergo splicing. In the subsequent catalytic steps, the 5′ splice site is recognized by U6 snRNP, while the 3′ splice site is defined by additional factors at later spliceosome complexes.

Modulation of 3′ splice site recognition by dynamic protein-RNA interactions

The U2AF heterodimer is essential for 3′ splice site recognition of most introns during E complex formation. Solution NMR data revealed that this initial recognition is highly dynamic, enabling a quantitative regulation of spliceosome assembly. The unbound U2AF2 RRM1-RRM2 tandem domains adopt a conformational ensemble of closed states with the RNA binding site of RRM1 occluded, and an open state that resembles the conformation when bound to strong PY-tracts [17,18]. This conformational equilibrium is shifted towards the open state depending on the "strength", i.e. binding affinity of a given PY-tract RNA sequence, thereby enabling a graduated regulation of 3′ splice site recognition, E complex formation and splicing [17]. Combining Förster resonance energy transfer (FRET) and NMR experiments, U2AF1 was shown to shift the equilibrium towards the open conformation of U2AF2 [19,20]. The PY-tract recognition by U2AF2 is further modulated by the flexible linker connecting RRM1 and RRM2, which transiently interacts with the RRM2, thereby rejecting the interaction with weak PY-tracts by an autoinhibition mechanism [21]. At the same time, the linker region mediates additional contacts within RRM1, stabilizing a compact arrangement of the RNA complex upon binding of strong PY-tracts [21], as is seen in the crystal structure of U2AF RRM1,2 bound to a strong PY-tract [22]. This region was also shown to form contacts with the central nucleotide of the PY-tract in between the binding regions of RRMs 1 and 2, slightly selecting for a uracil at this position [23].

Figure 1



(a) Canonical recognition of 5′ and 3′ splice sites including the novel factor FUBP1. A model of the 5′ site – U1 snRNA duplex is shown on the left, based on the U1C – RNA cryo-EM structure (pdb: 6qx9 [38]) and the alphafold 2 model of LUC7L [39], fitted to the RNA duplex in a similar orientation as in the yeast E complex [40]. The NMR-derived structure of FUBP1–U2AF2 interaction is shown in the middle (pdb: 8p25 [29]) and the crystallographic structure of the *S. pombe* U2AF1 ZnFs interacting with the 3′ splice site is shown on the right. Residues that are frequently mutated in cancer (Ser 34, Gln 157) and affect alternative splicing are highlighted in red (pdb: 7c07 [26]). **(b)** Regulation of exon 7 inclusion of *MALT1* pre-mRNA driven by modulation of RNA structure. hnRNP U is able to stabilize an RNA hairpin that occludes the PY-tract and the 5′ splice sites and, therefore, is not identified by U2AF2 and U1 snRNP (left), respectively. In contrast, hnRNP L binds to the complementary sequences to the PY-tract and 5′ splice site, destabilizing the duplexes and promoting U2AF2/U1 snRNP binding (right). **(c)** RNA binding modes of the RBM39 splicing factor. RRM2 recognizes single-stranded RNA motifs using an extended version of the canonical RRM binding surface, while RRM1 binds to stem loops using the RRM β-sheet and the β2- β3 loop [35].

Mutants within the U2AF1 ZnFs, associated with acute myelodysplastic leukemia (AML) and Myelodysplastic Syndrome (MDS), show different preferences for the nucleotides flanking the 3′ splice site AG [12,24]. This results in a genome-wide change of the AS pattern in a complex manner but also affects the formation of stress granules in myeloid malignancies [25]. So far, the only structural studies about how U2AF1 recognizes RNA through the ZnFs and how the cancer-associated mutations may modulate AS are based on a yeast homolog [24,26]. Thus, the molecular details of the recognition of the 3′ splice site by the human protein are not fully

clear and remain a subject for future studies (Figure 1a). Several studies have reported that the AG dinucleotide recognition can be modulated by additional factors, such as DEK or hnRNPA1 [27,28]. Structural details are yet to be resolved and are expected to involve the direct recognition of the 3′ splice site AG by the ZnFs of U2AF1.

FUBP1 has been recently identified as a key factor in recognizing the 3′ splice site, especially in long introns, by binding of its four, flexibly connected KH domains to a clustered U-rich element upstream of the BP site. Intrinsically disordered regions in FUBP1 mediate additional protein–protein interactions with the U2AF2 RRM2, and with the proline-rich C-terminal region of SF1 [29]. Thereby, FUBP1 enhances 3′ splice site recognition depending on the presence of a U-rich *cis*-element (Figure 1a) and may contribute to bridging interactions with factors at the 5′ splice site (see below).

Spliceosome assembly can also be modulated by the accessibility of *cis*-elements for splicing regulators, for example, by RNA secondary structure that may occlude splice sites. RNA binding proteins can modulate the RNA structure and thereby enable an additional layer of regulation [30,31]. A recent study demonstrated that RNA secondary structure sequesters the 3′ and 5′ splice sites flanking exon 7 of the *MALT-1* pre-mRNA and thus inhibits splicing. Combining chemical probing of RNA structure, NMR and biophysical experiments revealed the mechanisms by which two RNA binding proteins, hnRNP U and hnRNP L, antagonistically modulate exon 7 splicing. While an intrinsically disordered arginine-glycine-rich (RGG) region in hnRNP U binds and stabilizes hairpin structures inhibiting exon 7 splicing, hnRNP L shifts the weak and dynamic RNA secondary structure by binding to and stabilizing single-stranded conformations, thereby promoting exon 7 splicing. This conformational shift of pre-mRNA structures exposes the PY-tract and 5′ splice site, flanking exon 7, for recognition by U2AF2 and U1 snRNP and, therefore, promotes its inclusion [32] (Figure 1b).

It is estimated that about 12 % of the functional 3′ splice sites are recognized and spliced independent of U2AF2 [33]. Recent studies showed that splicing of short introns with very short PY tracts does not require U2AF2, but depends on RBM17 (also known as SPF45) to promote U2 snRNP binding. RBM17 is essential for the splicing of these short introns via interactions with SF3B1 in the U2 snRNP [33] and by cooperation with SAP30BP involving dynamic UHM-ULM interactions [34]. The precise molecular mechanism of RBM17-dependent spliceosome assembly remains unknown. Given that RBM17 does not bind RNA, it is likely that protein–protein interactions with U2 snRNP components may bypass the requirement for direct 3′ splice site RNA binding.

Another example of non-canonical 3′ splice site recognition involves the splicing factor RBM39, where its RRM2 recognizes characteristic 3′ splice site features that are not identified by U2AF2 and promotes U2 snRNP recruitment via a UHM-ULM interaction. In addition, the RBM39 RRM1 binds to a stem-loop sequence that may be part of the pre-mRNA or in the U1 snRNP (Figure 1c) [35]. This splicing factor is considered a target for degradation in high-risk neuroblastoma, where aberrant AS seems to play a critical role [36,37].

Modulation of 5′ splice site recognition

Many *trans*-acting protein factors influence the binding of U1 snRNP in AS to modulate exon inclusion or skipping by recognizing specific *cis*-regulatory RNA elements. The human U1 snRNP is smaller and simpler than the yeast particle, but many of the proteins exclusively found in the yeast version have human homologs. Although their functions are still not fully understood, there is evidence that they could act as AS factors. For example, the dimer of human PRPF39 (homolog of the yeast heterodimer Prp39-Prp42) recruits U1 snRNP to weak 5′ splice sites close to GC-rich sequences via interaction with the U1 snRNP component U1C [41,42]. In fact, other homologs are found to be recruited to the U1 snRNP via U1C interactions. For example, RRM2 and RRM3 of TIA1 (Nam8 in yeast) recognize a PY-rich region downstream of a 5′ splice site. The TIA1 binding near the 5′ splice site is further enhanced by the direct interaction of its RRM1 and Q-rich regions to U1C [43,44].

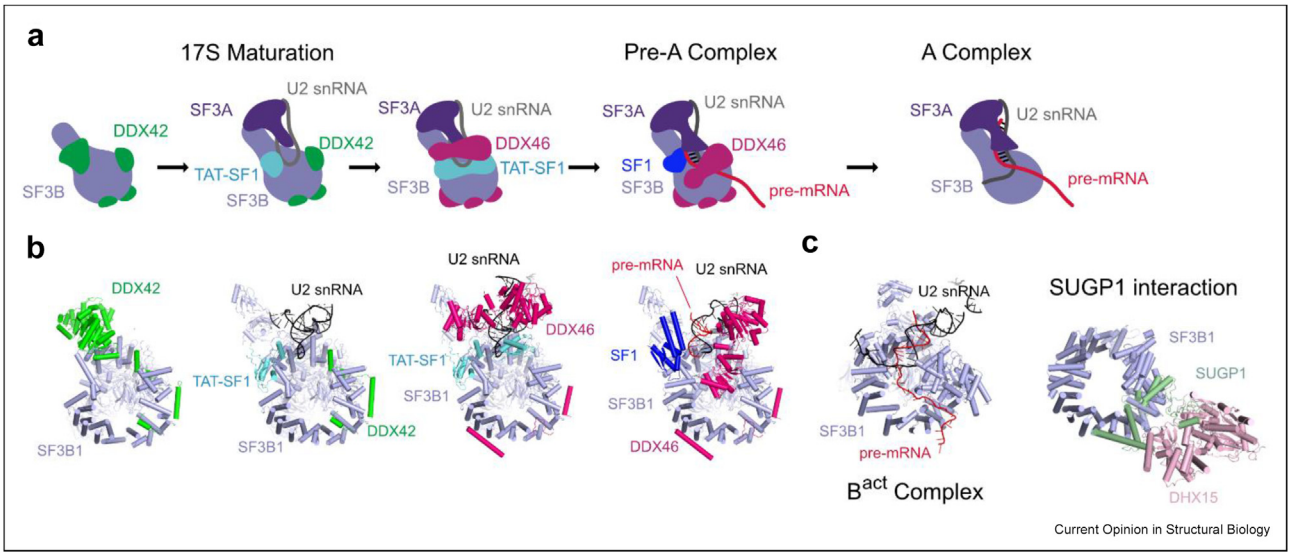
Human LUC7L paralogs (Luc7 homologs) have also been proposed to interact with U1C and other U1 snRNP-related factors to regulate alternative 5′ splice site selection. In addition, LUC7L*2* is implicated in the regulation of energy metabolism and in myeloid neoplasms [42,45,46]. A recent study suggests that LUC7L paralogs can discriminate non-canonical 5′ splice sites by contacting nucleotides that are either located in the preceding exon or the intron flanking the GU dinucleotide [47] (Figure 1a).

Towards formation of complex A

The loading of U2 snRNP to the correct BP sequence in the intron involves a major dynamic reorganization of the complete 3′ splice site region and the U2 snRNP particle itself (Figure 2a), which is driven by helicases. Here, two types of helicases, DDX46 (from the DEAD-box family) and DHX15 (DEAH-box family), with the help of G-patch containing proteins, drive dynamic rearrangements and ensure correct complex assembly and proofreading of BP selection.

A recent cryo-EM structure shows that during U2 snRNP maturation, the DEAD-box helicase DDX42

Figure 2



(a) Schematic representation of the 17S particle maturation and early steps of human spliceosome assembly at the 3′ splice. U2 snRNP maturation involves the protection of SF3B with the helicase DDX42 before loading the U2 snRNA-SF3A complex, displacing DDX42 and promoting TAT-SF1 recruitment. DDX46 is loaded, replacing DDX42, and stabilizes the UHM domain of TAT-SF1. Next, the mature 17S particle is recruited to the intron, where the BS sequence is recognized by the U2 snRNA after DDX46 has unwound the BSL and TAT-SF1 has been released. The helicase activity of DDX46 is thought to support the identification and selection of a correct BP. Subsequent rearrangement involves binding of the hinged pocket to the flipped-out adenine from the BP, creating a conformational change on SF3B1 HEAT repeat domain to a closed conformation and producing the release of DDX46. The interaction of an acidic loop of DDX46 with a region in the HEAT repeat is now replaced by the interaction with the intron RNA region corresponding to the PY-tract. SF1, which binds to the BP during early 3′ splice site recognition, is fully released and SF3A2 now strengthens the interaction with the intron-U2 snRNA duplex. **(b)** View of SF3B1 HEAT domain binding different factors during the early events: DDX42 N-terminal plug binds during 17S particle maturation (pdbs: 7evn, 8hk1 [48]) and is replaced by the DDX46 acidic loop (pdb: 7evo, 7vpx [50]). **(c)** This, in turn, is replaced by the intronic region of the RNA once the BP is locked (pdb 5z56 [61]). In a recent model, SUGP1 is proposed to connect the U2 snRNP with the helicase DHX15 interacting in the same region through the surrounding areas to the G-patch region [59].

participates in stabilizing the SF3B subunit, and upon binding of the U2 snRNA, it is replaced by another helicase, DDX46 (also known as PRP5), resulting in the matured 17S particle [48]. Further biochemical characterization of the role of the DDX42 helicase in this maturation process will be important to complement and rationalize the structural data. In this state, the branch–point interaction stem-loop (BSL) region of the U2 snRNA covers the complementary sequence of the intron BP [49,50].

U2 snRNP appears to be already loosely associated with 3′ splice site components in the E complex by multivalent interactions [51], promoting U2 snRNP proximity to the BP sequence. Only through ATP-hydrolysis the complementary sequence of the BP in the U2 snRNA (BSL) is exposed by a conformational change promoted by DDX46. This enables the formation of a preliminary duplex with the BP, resulting in the pre-A complex arrangement [50]. When the flipped-out BP adenine is specifically recognized via integration in a pocket created by SF3B1 and PHF5A, the U2 snRNA/BP RNA duplex is locked and further stabilized by SF3B6, another major reorganization step of the U2 snRNP takes place to form the A-complex [52]. The SF3B1 HEAT-repeat switches to a closed conformation, forcing the DDX46 release. The interaction of an acidic loop of DDX46 with a region in the HEAT repeat is now replaced by the intron RNA region corresponding to the PY-tract [50]. A similar reorganization of the U2 snRNP has been observed for the yeast complexes [53,54]. Unfortunately, to date, no structure is available for the human A complex, with the exception of an A-like U2 snRNP particle [52].

Proofreading of the branch point sequence

The pre-A complex constitutes an initial checkpoint for the correct identification of BP sequences. While the BP RNA is already specifically recognized in the E complex by SF1 [55], this interaction has mainly a kinetic role for 3′ splice site recognition [56] and may ensure that the BP is prepared for recognition by U2 snRNP in pre-A and A complexes. Recent studies identified the roles of helicases in proofreading and quality control by various dynamic interactions of G-patch containing proteins. It is proposed that proofreading for complementarity to the U2 snRNA happens through the action of DDX46, which unwinds suboptimal duplexes at the pre-A stage, similarly to its helicase function with the BSL [50].

More recently, the DEAH-box helicase DHX15 has been suggested to regulate quality control during pre-A/A complex assembly. To enable this function, the splicing regulator SUGP1 bridges DHX15 and the U2 snRNP through interactions with its G-patch to the helicase [57,58] and the surrounding regions of the G-patch to the HEAT domain of SF3B1 [59]. Intriguingly, the binding region of SUGP1 in SF3B1 is virtually identical to the interaction site of DDX46 and the RNA in the SF3B1 HEAT repeat domain (Figure 2b). In fact, several cancer-associated mutations occur in this region of SF3B1, highlighting its importance in splicing regulation [50]. In this disease context, the splicing factor GPATCH8 has been found to be required for the aberrant splicing induced by the mutant SF3B1 through a competing mechanism with SUGP1 for the binding to the DHX15 helicase [60].

Another recent report has also identified DHX15 in chromatin-associated purified spliceosome complexes but with limited presence of SUGP1. This suggests that other G-patch containing RBPs, including the AS factors RBM5/10 link the helicase function of DHX15 to U2 snRNPs [62]. RBM5 and RBM10 are multidomain proteins known to specifically recognize repressive *cis*-elements in the RNA through RRM1-ZnF-RRM2 domains in the N-terminal regions [63–65], promoting exon skipping. Additionally, they mediate interactions with proline-rich sequences of Sm proteins in the spliceosomal snRNPs [66]. The recent data show that these G-patch containing splicing factors are found in all isolated BP complexes regardless of the presence of *cis*-elements [62], suggesting more complex mechanisms for splicing modulation, which involve dynamic conformations and rearrangements [56].

Splice-site bridging during early spliceosome assembly

The identification of exon-intron boundaries during the initial steps of the splicing cycle is not limited to direct recognition of *cis*-regulatory motifs on the pre-mRNA but also involves crosstalk between the 3′ and 5′ splice sites by bridging proteins. The binding of these bridging factors at one splice site promotes the recognition of the adjacent splice site across the exon (exon definition) or intron (intron definition) (Box 1). Given that exons in humans are rather small and homogenous in size (200–300 nt) while introns can extend up to 10^5 nt, exon definition may be an efficient way to initially recognize 5′ and 3′ splice sites. Regardless of the mechanism of how the splice sites are recognized and bridged in the early steps, an initial exon definition complex requires rearrangement to a cross-intron complex for a productive splicing reaction (the opposite situation would result in circular RNA). The mechanisms of this shift from exon definition to cross-intron bridging remain still unknown, although based on recent structural data distinct possible solutions for this transition have been proposed [67,68].

Cross-intron connection in yeast

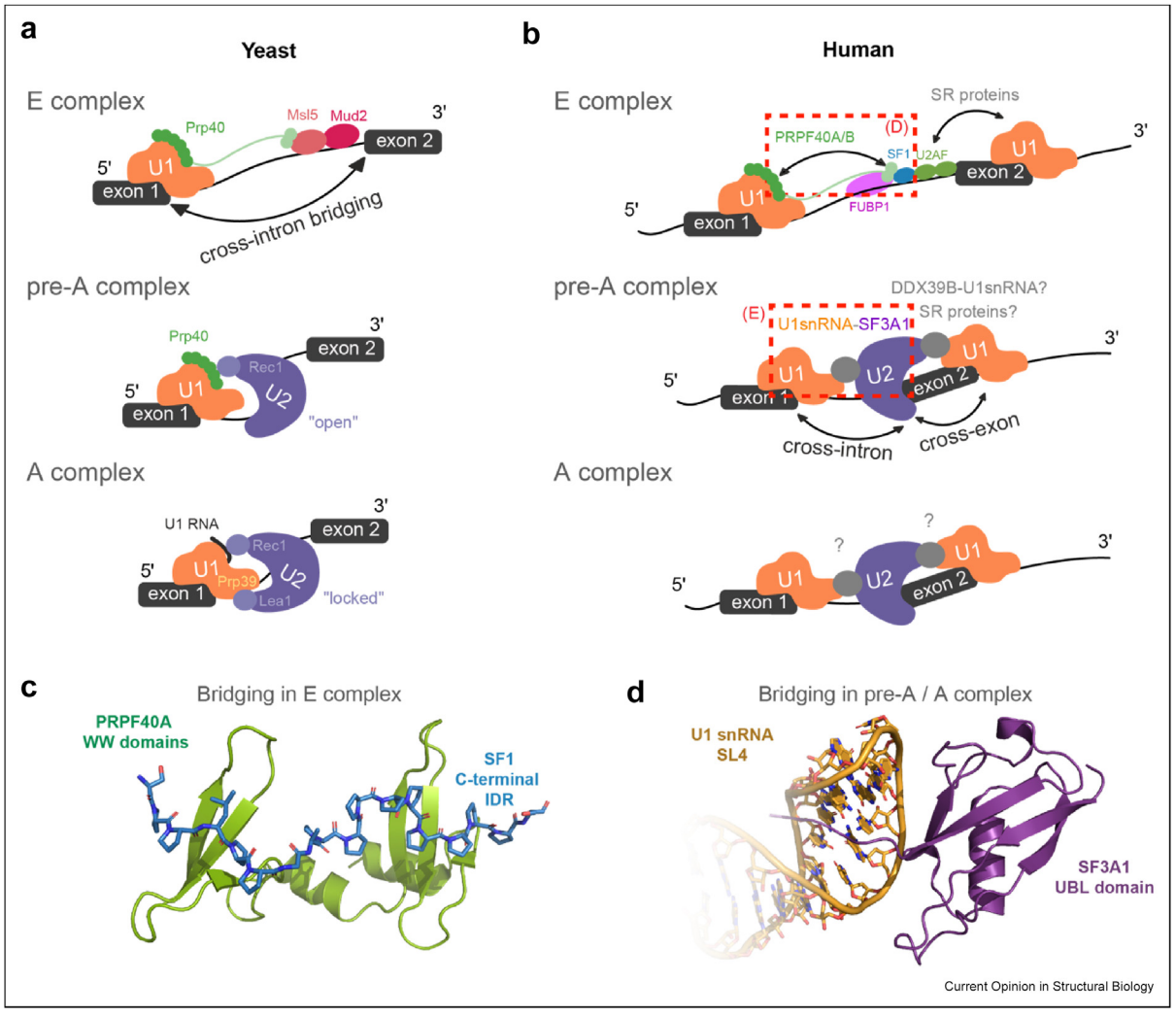
In yeast, only bridging interactions across the intron have been reported, which likely reflects the reduced complexity of the gene structures in this organism with typically one intron per mRNA that undergoes splicing [69]. These cross-intron connections are already established in the E complex, involving Prp40 interacting with Msl5 (the homolog of human SF1) at the BP and with U1 snRNP [40,70]. The bridging is maintained in the pre-A complex where Prp40 now interacts with the SF3B3 homolog Rse1, while probably Msl5 is destabilized [53]. In the A complex, Prp40 still copurifies but is not identified in cryo-EM maps; whether it still acts as a bridging factor is unknown (Figure 3a) [54]. Instead, after remodeling of the complex, a contact is detected between Prp39 in U1 snRNP and Lea1 (homolog of human U2A′) in the 3′ region of the U2 snRNP, and a weaker interaction, only detected in some of the particles, between the SL3-3 of U1 snRNA, the β-propellers of Rse1 (SF3B3) and the C-terminal region of Prp9 (SF3A3). These interactions change the conformation of the complex and reorient the U2 snRNP to promote the 5′ splice site transfer from the U1 snRNP to U6 snRNP in the pre−B complex [41] (Figure 3a).

Exon definition and cross-intron bridging

In vertebrates and especially in humans, the presence of genes with multiple long introns separating the shorter exons is challenging for the identification of intron/exon junctions. Here, exon definition offers an efficient mechanism where the identification of strong splice sites of one intron helps the recognition of the adjacent, potentially weaker sites across the exon (Figure 3b) [71]. These connections are established between U2 snRNP and U1 snRNP particles with the help of SR proteins as typical bridge factors. These proteins would recognize SR regions at the 3′ splice site (U2AF1/2) and in U1 snRNP (70K) at the 5′ splice site and, at the same time, would interact with conserved RNA sequence motifs in the exon (exonic splicing enhancers, ESEs); although it has been shown that such an interaction is dispensable for SRSF1 bridging [72]. Unfortunately, structural details of the interaction and the affinity and specificity that the SR regions would promote during the exon definition mechanism are scarce.

Available cryo-EM structures of the early complexes of human spliceosome do not show clear evidence of bridging [50]. This likely reflects that the intrinsic flexibility of these complexes prevents high-resolution cryo-EM analysis, as seen for the yeast E complex, where only the 5′ region is observed in the cryo-EM maps [40]. For dynamic complexes, complementary structural techniques are required to study multivalent

Figure 3



(a) Cross-intron contacts in yeast: Prp40 interacts with U1 snRNP and Msl5 in complex E, facilitating cross-intron bridging already early in spliceosome assembly. The Prp40 binding to Msl5 is replaced by an interaction with Rec1 (SF3B1) in the pre-A complex and in the A-complex, Rec1 directly interacts with the U1 snRNA. The whole complex then accommodates the tri-snRNP in the pre−B complex. **(b)** Early complexes in humans display multivalent contacts: already in E complex, SR proteins connect through the exon the two splice sites, while other factors like PRPF40 A/B or FUBP1 are able to establish connections, in principle, through the intron. In the following stages, other connections are established: U1 snRNA stem-loops are recognized by SF3A1 from the U2 snRNP and DDX39B also connected to U2AF2. **(c)** High-resolution structures of human bridging factors. The PRPF40A tandem WW domains recognize a proline-rich peptide in the C-terminal intrinsically disordered region of SF1 (pdb: 8pxx [73]). **(d)** The UBL domain of SF3A1 (a component of the U2 snRNP) specifically recognizes the stem loop 4 of the U1 snRNA (pdb: 7p0v [74]).

interactions between the U1 snRNP and 3′ splice site components that could take place through the intron or exon in E and or A complexes.

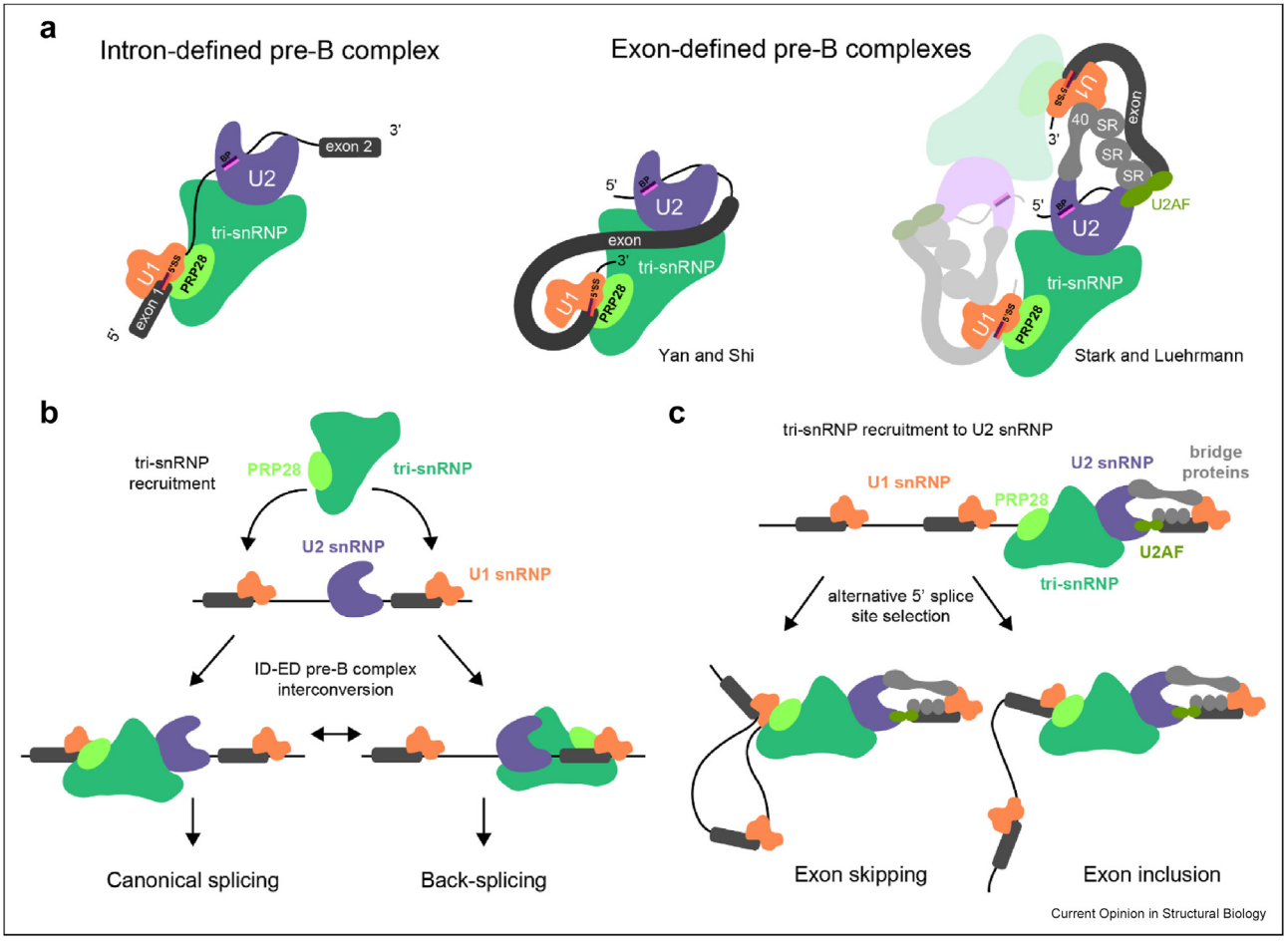
Prp40 is conserved in humans with two paralogs: PRPF40A and B, which could have a similar role as observed in yeast. A recent study presents the structural basis for the interaction of the PRPF40A tandem WW-domains with a proline-rich peptide motif in the long and disordered C-terminal region of SF1. The specificity of the cognate proline-rich peptide recognition is enhanced by an autoinhibitory mechanism of proline-rich sequences flanking the WW domains in PRPF40A, preventing interaction with sub-optimal binding motifs in other proteins (Figure 3c) [73]. The study provides the first structural view of any PRPF40A homolog (human or yeast) interacting with a BP binding protein, and has not been visible in cryo-EM studies. A cross-intron bridging mediated by PRPF40 A/B requires also the interaction of its FF domains with U1 snRNP components, for which complete structural details remain elusive.

U1 snRNA stem loops 3 and 4 (SL3, SL4) play roles in establishing interactions with the 3′ splice sites in early spliceosome assembly (pre-A and A complexes). The C-terminal UBL-like domain of SF3A1 (U2 snRNP) directly interacts with SL4 of the U1 snRNA, establishing a direct connection between the two particles (Figure 3d) [74]. In addition, the SL3 is recognized by the helicase DDX39B (UAP56), which at the same time interacts with U2AF2 [75]. The recently described 3′ splice site component, FUBP1, is also proposed to establish direct contact with U1 snRNP proteins [29].

Except for the connections orchestrated by SR proteins, functional bridging interactions were suggested to occur mainly across the intron. An initial exon definition arrangement must change to cross-intron bridging during the splicing cycle to prevent back-splicing, although it is not clear when and how this takes place (Figure 3b). Unfortunately, the observed bridging type seen in cryo-EM structures is defined by the pre-mRNA sequence that is provided for the complex assembly. Typically, exon definition is arranged in early assembly steps (pre-A and A), while for structural studies of complex B and successive stages, a cross-intron arrangement is prepared.

In two recent studies, structures of pre-B and B complexes were obtained via spliceosome assembly in an exon definition arrangement [67,68]. Notably, the topology of the new exon-defined pre-B complexes resembles the previous intron-defined structures (Figure 4a). In the work by Yan and Shi [67], the length

Figure 4



(a) Schematic representation of the intron-defined (left) and exon-defined (right) human pre-B complexes; some of the regions were not properly resolved in the cryo-EM structures but suggested by additional experiments, like the U1 snRNPs and the bridge proteins (40 = PRPF40A and SR= SRSF proteins) in the exon-defined complexes. **(b)** Postulated mechanism for the exon to intron definition switch by Zhang et al. [67], where the tri-snRNP is able to interact with either U1 snRNP up or downstream the U2 snRNP depending on the local context. In addition, while in the pre-B state, the tri-snRNP would be able to switch from one U1 snRNP to the other. **(c)** In the model from Zhang et al. [68], the tri-snRNP is loaded to the U2 snRNP to form the pre−B complex, the downstream U1 snRNP would interact with U2 snRNP through the bridging factors (e.g. SRSF proteins and PRPF40) and the PRP28 interface would be accessible for another U1 snRNP located upstream of the U2 snRNP, allowing for alternative 5′ splice site selection.

of the selected exon allows the placement of the downstream U1 snRNP in the position that occupies the upstream U1 snRNP particle in the cross-intron pre−B complex [76] via interactions with PRP28. In the study reported by Stark and Lührmann [68], the exon is significantly shorter, promoting dimerization of the complexes, in which the downstream U1 snRNP is now located near SF3B in the U2 snRNP. This resembles the position seen in the cross-exon pre-A complex [50], while another U1 snRNP binds in trans to the PRP28 interface, i.e. the position of the upstream U1 snRNP in intron-defined complexes. Interestingly, many of the bridging connections described for E, pre-A, and A complexes are also detected by cross-linking/mass spectrometry experiments in the later exon-defined pre−B complex, such as those involving SRSF proteins or PRPF40A, which interact with the SF3A1 proline-rich region as proposed in the absence of SF1 [73]. In this case, Stark and Lührmann propose that all these bridging interactions occur across the exon.

Finally, two models of cross-exon to cross-intron arrangement are suggested by these two recent studies. Yan and Shi propose a model, in which the loading of the tri-snRNP selects which U1 snRNP particle would form the pre−B complex, e.g. involving the upstream (intron-defined) or the downstream (exon-defined) U1 snRNP. The switch from exon to intron definition is possible by releasing one U1 snRNP to engage with the upstream one (Figure 4b). In contrast, Stark and Lührmann propose that the downstream U1 snRNP particle would be placed near the U2 snRNP, leaving the PRP28 interface free to engage with upstream U1 snRNP particles, a model compatible with alternative 5′ splice site selection (Figure 4c).

Conclusions and future perspectives

Recent developments in cryo-EM methods have greatly advanced our knowledge of the splicing mechanism and provided some glimpses into the molecular details behind early events in splice site selection and proofreading. Complementary solution techniques, such as NMR or fluorescence spectroscopy, are required to study the dynamic interactions that underlie these mechanisms and are especially important during early spliceosome assembly and (alternative) splicing regulation. In fact, while cryo-EM structures provide exciting structural insights with complexes assembled and purified in native conditions, a significant fraction of the expected protein and RNA components are not observed in the cryo-EM maps due to conformational flexibility and dynamics, involving extended intrinsically disordered regions, which can be readily studied using NMR spectroscopy (see Box 2).

A large number of alternative splicing factors that modulate splicing of specific pre-mRNAs with key roles for many cellular processes, are known, but the specific underlying molecular and structural mechanisms remain poorly understood. The interactions of these *trans*-acting factors with *cis*-regulatory motifs in the pre-mRNA typically involve dynamic RNA interactions of multiple RNA binding domains together with additional protein−protein interactions that are mediated by intrinsically disordered regions. Moreover, helicases are key for the dynamic remodeling of RNPs during the splicing cycle and are essential for proofreading during early spliceosome assembly. Much is still to be learned on the existing conformational states and transitions with the help of solution techniques, including NMR, fluorescence spectroscopy, and small angle scattering,

Box 2. Emerging topics and future challenges for understanding splicing regulation.

While truly impressive progress has been made in our understanding of the molecular and structural mechanisms of splicing catalysis during the spliceosome cycle, detailed and predictive understanding of cell and development-dependent alternative splicing remains poorly understood.

- ***What are the structural mechanisms that direct early spliceosome assembly and alternative splicing regulation?*** How do combinatorial interactions of splicing factors with *cis*-regulatory RNA elements and spliceosome complexes modulate the utilization of splice sites? What are the structural mechanisms underlying exon and intron definition? Also, most splicing factors contain intrinsically disordered regions (IDRs), which can mediate protein−protein interactions or have been implicated in the formation of biomolecular condensates that may modulate splicing. Potential roles of these interactions, as well as posttranslational modifications of splicing factors in splicing regulation are just emerging.
- ***What are the principles that underlie the splicing code?*** The complex interplay of variations in constitutive and alternative *cis*-regulatory RNA elements, and alternative splicing factors in a tissue and developmental-stage dependent manner is still not well understood. Moreover, the mechanisms involving modulating pre-mRNA structure, RNA modifications, and co-transcriptional splicing are still poorly characterized. Combining machine learning and artificial intelligence analysis of transcriptome and proteome-wide data with detailed mechanistic and structural studies will be required to reveal the splicing code.
- ***How is quality control of splicing achieved?*** The fidelity of the splicing is crucial, given that single mutations can have dramatic consequences. Some quality control checkpoints have been identified along the splicing cycle, such as the locking of the BP sequence by the U2 snRNP. But general mechanisms and roles of helicases in splicing quality control are still emerging.
- ***How is misregulation of splicing linked to human disease?*** Mutations in *cis*-elements or trans-acting RNA binding proteins are linked to genetic and neurodegenerative diseases, and splicing plays a role in viral infections. Misregulation of splicing is strongly implicated in several cancers, where also aberrant levels of splicing factors affect cancer treatment and prognosis. This suggests the modulation of splicing for the development of innovative therapeutic approaches, as has been demonstrated with first therapies to treat spinal muscular atrophy.

combined with molecular dynamics simulations to understand the regulation of AS and proofreading.

The current view on splice site selection during early steps of spliceosome assembly needs to be expanded by considering the crosstalk of the different complexes. How the initial exon definition connection results in a productive splicing reaction in favor of the cross-intron arrangement remains unclear. Whether there is a continuous connection of U1 snRNP – U2 snRNP – U1 snRNP with different or similar bridging mechanisms through the exon and the intron, or there is a particle reorganization that leads to a shift of the interaction of U2 snRNP from the downstream U1 snRNP to the upstream particle is not known. In addition, the role of RNA polymerase II in splicing is well appreciated, but detailed molecular mechanisms still remain to be determined. Many splicing events take place co-transcriptionally [77], where the polymerase acts as a hub to load the splicing factors into the nascent RNA once their *cis*-elements are transcribed. A recent cryo-EM study showed one of these interactions between the U1 snRNP particle and the RNA polymerase II [78].

Finally, many of the splicing factors comprise intrinsically disordered regions and can form liquid condensates, such as nuclear speckles. A role of liquid–liquid phase separation for splicing mechanism and regulation has been proposed but remains a subject for further research [79].

In conclusion, pre-mRNA splicing is an essential aspect of gene regulation in metazoans and humans. It is regulated by complex and dynamic interactions between proteins and RNAs during early spliceosome assembly [1–3]. Cryo-EM has greatly advanced our knowledge of the structural mechanisms underlying splicing. Significant progress has been made in understanding the dynamic recognition of *cis*-regulatory elements by multidomain RNA binding proteins. These interactions are crucial for the regulation of spliceosome assembly and AS and are revealed by solution techniques, such as NMR and FRET. To further advance structural biology of splicing and guide the design of splicing modulators as therapeutics [6,80,81], integrative structural biology combining cryo-EM, X-ray and solution techniques, including NMR, in combination with structural models predicted from deep learning [39,82] will be most promising.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Data availability

No data was used for the research described in the article.

Acknowledgments

We would like to thank Hyun Seo Kang for comments and discussions. M.S. acknowledges funding by the Deutsche Forschungsgemeinschaft (SFB1035, Project number 201302640; SPP1935, 419138605) and the European Union (ERC SyG, UNLEASH, project number 101071936). S.M.-L. is supported by an European Union HORIZON 2020 Marie Skłodowska-Curie grant, No. 792692 (SplicEcomplex).

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Will CL, Lührmann R: **Spliceosome structure and function**. *Cold Spring Harbor Perspect Biol* 2011, **3**.
2. Shi Y: **Mechanistic insights into precursor messenger RNA splicing by the spliceosome**. *Nat Rev Mol Cell Biol* 2017, **18**: 655–670.
3. Wilkinson ME, Charenton C, Nagai K: **RNA splicing by the spliceosome**. *Annu Rev Biochem* 2020, **89**:359–388.
4. Gallego-Paez LM, Bordone MC, Leote AC, Saraiva-Agostinho N, Ascensão-Ferreira M, Barbosa-Morais NL: **Alternative splicing: the pledge, the turn, and the prestige: the key role of alternative splicing in human biological systems**. *Hum Genet* 2017, **136**:1015–1042.
5. Baralle FE, Giudice J: **Alternative splicing as a regulator of development and tissue identity**. *Nat Rev Mol Cell Biol* 2017, **18**:437–451.
6. Bonnal SC, López-Oreja I, Valcárcel J: **Roles and mechanisms of alternative splicing in cancer — implications for care**. *Nat Rev Clin Oncol* 2020, **17**:457–474.
7. Tholen J, Galej WP: **Structural studies of the spliceosome: bridging the gaps**. *Curr Opin Struct Biol* 2022, **77**.
8. Wahl MC, Will CL, Lührmann R: **The spliceosome: design principles of a dynamic RNP machine**. *Cell* 2009, **136**:701–718.
9. Mackereth CD, Sattler M: **Dynamics in multi-domain protein recognition of RNA**. *Curr Opin Struct Biol* 2012, **22**:287–296.
10. Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ, Frey BJ: **Deciphering the splicing code**. *Nature* 2010, **465**:53–59.
11. Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, Yuen RKC, Hua Y, Gueroussov S, Najafabadi HS, Hughes TR, *et al.*: **The human splicing code reveals new insights into the genetic determinants of disease**. *Science* 2015, **347**.
12. Bonnal S, Vigevani L, Valcárcel J: **The spliceosome as a target of novel antitumour drugs**. *Nat Rev Drug Discov* 2012, **11**: 847–859.
13. Naryshkin NA, Weetall M, Dakka A, Narasimhan J, Zhao X, Feng Z, Ling KKY, Karp GM, Qi H, Woll MG, *et al.*: **SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy**. *Science* 2014, **345**: 688–693.
14. Palacino J, Swalley SE, Song C, Cheung AK, Shu L, Zhang X, Van Hoosear M, Shin Y, Chin DN, Keller CG, *et al.*: **SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice**. *Nat Chem Biol* 2015, **11**:511–517.
15. Kielkopf CL, Lucke S, Green MR: **U2AF homology motifs: protein recognition in the RRM world**. *Genes Dev* 2004, **18**: 1513–1526.

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

17. Mackereth CD, Madl T, Bonnal S, Simon B, Zanier K, Gasch A, Rybin V, Valcárcel J, Sattler M: **Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF**. *Nature* 2011, **475**:408–413.

18. Huang JR, Warner LR, Sanchez C, Gabel F, Madl T, Mackereth CD, Sattler M, Blackledge M: **Transient electrostatic interactions dominate the conformational equilibrium sampled by multidomain splicing factor U2AF65: a combined NMR and SAXS study**. *J Am Chem Soc* 2014, **136**:7068–7076.

19. Von Voithenberg LV, Sánchez-Rico C, Kang HS, Madl T, Zanier K, Barth A, Warner LR, Sattler M, Lamb DC: **Recognition of the 3′ splice site RNA by the U2AF heterodimer involves a dynamic population shift**. *Proc Natl Acad Sci U S A* 2016, **113**:7169–7175.

20. Warnasooriya C, Feeney CF, Laird KM, Ermolenko DN, Kielkopf CL: **A splice site-sensing conformational switch in U2AF2 is modulated by U2AF1 and its recurrent myelodysplasia-associated mutation**. *Nucleic Acids Res* 2020, **48**:5695–5709.

21. Kang H-S, Sánchez-Rico C, Ebersberger S, Reymond Sutandy FX, Busch A, Welte T, Stehle R, Hipp C, Schulz L, Buchbender A, *et al.*: **An autoinhibitory intramolecular interaction proof-reads RNA recognition by the essential splicing factor U2AF2**. *Proc Natl Acad Sci U S A* 2020, **117**:7140–7149.

22. Agrawal AA, Salsi E, Chatrikhi R, Henderson S, Jenkins JL, Green MR, Ermolenko DN, Kielkopf CL: **An extended U2AF65-RNA-binding domain recognizes the 3′ splice site signal**. *Nat Commun* 2016, **7**.

23. Glasser E, Maji D, Biancon G, Puthenpeedikakkal AMK, Cavender CE, Tebaldi T, Jenkins JL, Mathews DH, Halene S, Kielkopf CL: **Pre-mRNA splicing factor U2AF2 recognizes distinct conformations of nucleotide variants at the center of the pre-mRNA splice site signal**. *Nucleic Acids Res* 2022, **50**:5299–5312.

24. Yoshida H, Park SY, Oda T, Akiyoshi T, Sato M, Shirouzu M, Tsuda K, Kuwasako K, Unzai S, Muto Y, *et al.*: **A novel 3′ splice site recognition by the two zinc fingers in the U2AF small subunit**. *Genes Dev* 2015, **29**:1649–1660.

25. Biancon G, Joshi P, Zimmer JT, Hunck T, Gao Y, Lessard MD, Courchaine E, Barentine AES, Machyna M, Botti V, *et al.*: **Precision analysis of mutant U2AF1 activity reveals deployment of stress granules in myeloid malignancies**. *Mol Cell* 2022, **82**:1107–1122.e7.

26. * Yoshida H, Park SY, Sakashita G, Nariai Y, Kuwasako K, Muto Y, Urano T, Obayashi E: **Elucidation of the aberrant 3′ splice site selection by cancer-associated mutations on the U2AF1**. *Nat Commun* 2020, **11**:1–7.
First structural analysis showing how the ZnF domains of *S. pombe* U2AF1 homolog recognize 3′ splice site RNA and the implications in cancer-associated mutations.

27. Mendes Soares LM, Zanier K, Mackereth C, Sattler M, Valcárcel J: **Intron removal requires proofreading of U2AF/3Splice site recognition by DEK**. *Science* 2006, **312**:1961–1965.

28. Tavanez JP, Madl T, Kooshapur H, Sattler M, Valcárcel J: **HnRNP A1 proofreads 3' splice site recognition by U2AF**. *Mol Cell* 2012, **45**:314–329.

29. ** Ebersberger S, Hipp C, Mulorz MM, Buchbender A, Hubrich D, Kang HS, Martínez-Lumbreras S, Kristofori P, Sutandy FXR, Llacsahuanga Allcca L, *et al.*: **FUBP1 is a general splicing factor facilitating 3′ splice site recognition and splicing of long introns**. *Mol Cell* 2023, **83**:2653–2672.
Combining genome-wide, biochemical and structural analyses, the splicing factor FUBP1 is shown to be a general factor for 3′ splice site recognition of long introns by interacting with an upstream pyrimidine-rich sequence.

30. Jiang Z, Cote J, Kwon JM, Goate AM, Wu JY: **Aberrant splicing of tau pre-mRNA caused by intronic mutations associated with the inherited dementia frontotemporal dementia with Parkinsonism linked to chromosome 17**. *Mol Cell Biol* 2000, **20**:4036–4048.

31. Singh NN, Singh RN, Androphy EJ: **Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes**. *Nucleic Acids Res* 2007, **35**:371–389.

32. ** Jones AN, Graß C, Meininger I, Geerlof A, Klostermann M, Zarnack K, Krappmann D, Sattler M: **Modulation of pre-mRNA structure by hnRNP proteins regulates alternative splicing of MALT1**. *Sci Adv* 2022, **8**:9153.
RNA secondary structure shields splice site flanking an alternative spliced exon. RNA binding proteins can antagonistically stabilize the RNA structure of shift towards single-stranded conformations to inhibit or promote splicing.

33. Fukumura K, Yoshimoto R, Sperotto L, Kang HS, Hirose T, Inoue K, Sattler M, Mayeda A: **SPF45/RBM17-dependent, but not U2AF-dependent, splicing in a distinct subset of human short introns**. *Nat Commun* 2021, **12**.

34. Fukumura K, Sperotto L, Seuß S, Kang HS, Yoshimoto R, Sattler M, Mayeda A: **SAP30BP interacts with RBM17/SPF45 to promote splicing in a subset of human short introns**. *Cell Rep* 2023, **42**.

35. Campagne S, Jutzi D, Malard F, Matoga M, Romane K, Feldmuller M, Colombo M, Ruepp MD, Allain FHT: **Molecular basis of RNA-binding and autoregulation by the cancer-associated splicing factor RBM39**. *Nat Commun* 2023, **14**.

36. Uehara T, Minoshima Y, Sagane K, Sugi NH, Mitsuhashi KO, Yamamoto N, Kamiyama H, Takahashi K, Kotake Y, Uesugi M, *et al.*: **Selective degradation of splicing factor CAPERα by anticancer sulfonamides**. *Nat Chem Biol* 2017, **13**:675–680.

37. Han T, Goralski M, Gaskill N, Capota E, Kim J, Ting TC, Xie Y, Williams NS, Nijhawan D: **Anticancer sulfonamides target splicing by inducing RBM39 degradation via recruitment to DCAF15**. *Science* 2017, **356**.

38. Charenton C, Wilkinson ME, Nagai K: **Mechanism of 5′ splice site transfer for human spliceosome activation**. *Science* 2019, **364**:362–367.

39. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al.*: **Highly accurate protein structure prediction with AlphaFold**. *Nature* 2021, **596**:583–589.

40. Li X, Liu S, Zhang L, Issaian A, Hill RC, Espinosa S, Shi S, Cui Y, Kappel K, Das R, *et al.*: **A unified mechanism for intron and exon definition and back-splicing**. *Nature* 2019, **573**:375–380.

41. De Bortoli F, Neumann A, Kotte A, Timmermann B, Schüler T, Wahl MC, Loll B, Heyd F: **Increased versatility despite reduced molecular complexity: evolution, structure and function of metazoan splicing factor PRPF39**. *Nucleic Acids Res* 2019, **47**:5867–5879.

42. Espinosa S, De Bortoli F, Li X, Rossi J, Wagley ME, Lo H-YG, Taliaferro JM, Zhao R: **Human PRPF39 is an alternative splicing factor recruiting U1 snRNP to weak 5' splice sites**. *RNA* 2022, **29**:97–110.

43. Förch P, Puig O, Martínez C, Séraphin B, Valcárcel J: **The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5′ splice sites**. *EMBO J* 2002, **21**:6882–6892.

44. Wang I, Hennig J, Jagtap PKA, Sonntag M, Valcárcel J, Sattler M: **Structure, dynamics and RNA binding of the multi-domain splicing factor TIA-1**. *Nucleic Acids Res* 2014, **42**:5949–5966.

45. Jourdain AA, Begg BE, Mick E, Shah H, Calvo SE, Skinner OS, Sharma R, Blue SM, Yeo GW, Burge CB, *et al.*: **Loss of LUC7L2 and U1 snRNP subunits shifts energy metabolism from glycolysis to OXPHOS**. *Mol Cell* 2021, **81**:1905–1919.e12.

46. Daniels NJ, Hershberger CE, Gu X, Schueger C, DiPasquale WM, Brick J, Saunthararajah Y, Maciejewski JP, Padgett RA: **Functional analyses of human LUC7-like proteins involved in splicing regulation and myeloid neoplasms**. *Cell Rep* 2021, **35**.

47. Kenny CJ, Mcgurk MP, Burge CB: **Human LUC7 proteins impact splicing of two major subclasses of 5' splice sites**. *bioRxiv* 2022, https://doi.org/10.1101/2022.12.07.519539.

48. * Yang F, Bian T, Zhan X, Chen Z, Xing Z, Larsen NA, Zhang X, Shi Y: **Mechanisms of the RNA helicases DDX42 and DDX46 in human U2 snRNP assembly**. *Nat Commun* 2023, **14**.
Cryo-EM structures visualize the maturation of the U2 snRNP, highlighting the role of helicases in 17S particle assembly.

49. Zhang Z, Will CL, Bertram K, Dybkov O, Hartmuth K, Agafonov DE, Hofele R, Urlaub H, Kastner B, Lührmann R, *et al.*: **Molecular architecture of the human 17S U2 snRNP**. *Nature* 2020, **583**: 310–313.

50. ** Zhang X, Zhan X, Bian T, Yang F, Li P, Lu Y, Xing Z, Fan R, Zhang QC, Shi Y: **Structural insights into branch site proofreading by human spliceosome**. *Nat Struct Mol Biol* 2024, https://doi.org/10.1038/s41594-023-01188-0.
The first cryo-EM structure of the human pre-A complex provides new insights into branch site proofreading involving multiple structural rearrangements, directed by the helicase DDX46.

51. Das R, Zhou Z, Reed R: **Functional association of U2 snRNP with the ATP-independent splicecosomal complex E**. *Mol Cell* 2000, **5**:779–787.

52. * Tholen J, Razew M, Weis F, Galej WP: **Structural basis of branch site recognition by the human spliceosome**. *Science* 2022, **375**:50–57.
High-resolution cryo-EM structures of U2 snRNP at different states from the 17S particle to the A-like complex provide insights for the branch point site recognition.

53. Zhang Z, Rigo N, Dybkov O, Fourmann JB, Will CL, Kumar V, Urlaub H, Stark H, Lührmann R: **Structural insights into how Prp5 proofreads the pre-mRNA branch site**. *Nature* 2021, **596**: 296–300.

54. Plaschka C, Lin PC, Charenton C, Nagai K: **Prespliceosome structure provides insights into spliceosome assembly and regulation**. *Nature* 2018, **559**:419–422.

55. Liu Z, Luyten I, Bottomley MJ, Messias AC, Houngninou-Molango S, Sprangers R, Zanier K, Krämer A, Sattler M: **Structural basis for recognition of the intron branch site RNA by splicing factor 1**. *Science* 2001, **294**:1098–1102.

56. Guth S, Valcarcel J: **Kinetic role for mammalian SF1/BBP in spliceosome assembly and function after polypyrimidine tract recognition by U2AF**. *J Biol Chem* 2000, **275**: 38059–38066.

57. Zhang J, Huang J, Xu K, Xing P, Huang Y, Liu Z, Tong L, Manley JL: **DHX15 is involved in SUGP1-mediated RNA missplicing by mutant SF3B1 in cancer**. *Proc Natl Acad Sci U S A* 2022:119.

58. Beusch I, Rao B, Studer MK, Luhovska T, Šukytė V, Lei S, Oses-Prieto J, SeGraves E, Burlingame A, Jonas S, *et al.*: **Targeted high-throughput mutagenesis of the human spliceosome reveals its in vivo operating principles**. *Mol Cell* 2023, **83**: 2578–2594.e9.

59. * Zhang J, Xie J, Huang J, Liu X, Xu R, Tholen J, Galej WP, Tong L, Manley JL, Liu Z: **Characterization of the SF3B1-SUGP1 interface reveals how numerous cancer mutations cause mRNA missplicing**. *Genes Dev* 2023, **37**:968–983.
This biochemical study identifies an interaction between SUGP1 and SF3B1, which involves the same region of the SF3B1 HEAT repeats that interacts with helicase DHX46 and the intron RNA in the pre-A and A complexes.

60. Benbarche S, Pineda JMB, Galvis LB, Biswas J, Liu B, Wang E, Zhang Q, Hogg SJ, Lyttle K, Dahi A, *et al.*: **GPATCH8 modulates mutant SF3B1 mis-splicing and pathogenicity in hematologic malignancies**. *Mol Cell* 2024, **84**:1886–1903.e10.

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

62. Damianov A, Lin C-H, Huang J, Zhou L, Jami-Alahmadi Y, Peyda P, Wohlschlegel J, Black DL: **The splicing regulators RBM5 and RBM10 are subunits of the U2 snRNP engaged with intron branch sites on chromatin**. *Mol Cell* 2024, **84**:1–16.

63. Collins KM, Kainov YA, Christodolou E, Ray D, Morris Q, Hughes T, Taylor IA, Makeyev EV, Ramos A: **An RRM-ZnF RNA recognition module targets RBM10 to exonic sequences to promote exon exclusion**. *Nucleic Acids Res* 2017, **45**: 6761–6774.

64. Soni K, Martínez-Lumbreras S, Sattler M: **Conformational dynamics from ambiguous zinc coordination in the RanBP2-type zinc finger of RBM5**. *J Mol Biol* 2020, **432**:4127–4138.

65. Soni K, Jagtap PKA, Martínez-Lumbreras S, Bonnal S, Geerlof A, Stehle R, Simon B, Valcárcel J, Sattler M: **Structural basis for specific RNA recognition by the alternative splicing factor RBM5**. *Nat Commun* 2023, **14**.

66. Mourao A, Bonnal S, Soni K, Warner L, Bordonne R, Valcarcel J, Sattler M: **Structural basis for the recognition of spliceosomal SmN/B/B' proteins by the RBM5 OCRE domain in splicing regulation**. *Elife* 2016, **5**.

67. ** Zhang W, Zhang X, Zhan X, Bai R, Lei J, Yan C, Shi Y: **Structural insights into human exon-defined spliceosome prior to activation**. *Cell Res* 2024 2024, **0**:1–12.
Several exon-defined structures from pre-B to B complexes show similar assembly for exon and the intron-defined complexes, suggest a model for switching from exon to intron definition.

68. ** Zhang Z, Kumar V, Dybkov O, Will CL, Zhong J, Ludwig SEJ, Urlaub H, Kastner B, Stark H, Lührmann R: **Structural insights into the cross-exon to cross-intron spliceosome switch**. *Nature* 2024, https://doi.org/10.1038/s41586-024-07458-1.
Exon-defined structures of pre-B to B complexes in different conditions are compatible with binding of the upstream and downstream U1 snRNPs and suggest a model for the transition from exon to intron definition.

69. Lopez PJ, Séraphin B: **Genomic-scale quantitative analysis of yeast pre-mRNA splicing: implications for splice-site recognition**. *RNA* 1999, **5**:1135–1137.

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

71. Robberson BL, Cote GJ, Berget SM: **Exon definition may facilitate splice site selection in RNAs with multiple exons**. *Mol Cell Biol* 1990, **10**:84–94.

72. Jobbins AM, Campagne S, Weinmeister R, Lucas CM, Gosliga AR, Clery A, Chen L, Eperon LP, Hodson MJ, Hudson AJ, *et al.*: **Exon-independent recruitment of SRSF1 is mediated by U1 snRNP stem-loop 3**. *EMBO J* 2022, **41**.

73. * Martínez-Lumbreras S, Träger LK, Mulorz MM, Payr M, Dikaya V, Hipp C, König J, Sattler M: **Intramolecular autoinhibition regulates the selectivity of PRPF40A tandem WW domains for proline-rich motifs**. *Nat Commun* 2024, **15**.
The structure of the PRPF40A tandem WW domains bound to a proline-rich motif in the unstructured C-terminal region of SF1 may bridge 5′ and 3′ splice sites, An autoinhibitory intramolecular interaction contributes to proline-rich peptide binding selectivity and may be relevant for many WW domain proteins.

74. * de Vries T, Martelly W, Campagne S, Sabath K, Sarnowski CP, Wong J, Leitner A, Jonas S, Sharma S, Allain FH-T: **Sequence-specific RNA recognition by an RGG motif connects U1 and U2 snRNP for spliceosome assembly**. *Proc Natl Acad Sci USA* 2022, **119**.
Structural details of a bridging interaction in pre A/ A complexes between stem-loop 3 of U1 snRNA in U1 snRNP and the UBL domain of the U2 component SF3A1.

75. Martelly W, Fellows B, Kang P, Vashisht A, Wohlschlegel JA, Sharma S: **Synergistic roles for human U1 snRNA stem-loops in pre-mRNA splicing**. *RNA Biol* 2021, **18**:2576–2593.

76. Charenton C, Wilkinson ME, Nagai K: **Mechanism of 5' splice site transfer for human spliceosome activation**. *Science* 2019, **364**:362–367.

77. Ameur A, Zaghlool A, Halvardson J, Wetterbom A, Gyllensten U, Cavelier L, Feuk L: **Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain**. *Nat Struct Mol Biol* 2011, **18**:1435–1440.

78. Zhang S, Aibara S, Vos SM, Agafonov DE, Lührmann R, Cramer P: **Structure of a transcribing RNA polymerase II-U1 snRNP complex**. *Science* 2021, **371**:305–309.

79. Liao SE, Regev O: **Splicing at the phase-separated nuclear speckle interface: a model**. *Nucleic Acids Res* 2021, **49**: 636–645.

80. Cretu C, Agrawal AA, Cook A, Will CL, Fekkes P, Smith PG, Lührmann R, Larsen N, Buonamici S, Pena V: **Structural basis of splicing modulation by antitumor macrolide compounds**. *Mol Cell* 2018, **70**:265–273.e8.

81. Jagtap PKA, Kubelka T, Soni K, Will CL, Garg D, Sippel C, Kapp TG, Potukuchi HK, Schorpp K, Hadian K, *et al.*: **Identification of phenothiazine derivatives as UHM-binding inhibitors of early spliceosome assembly**. *Nat Commun* 2020, **11**.

82. Baek M, Dimaio F, Anishchenko I, Dauparas J, Lee GR, Wang J, Cong Q, Kinch LN, Dustin R, Millán C, *et al.*: **Accurate prediction of protein structures and interactions using a 3-track neural network**. *Science* 2021, **373**:871–876.