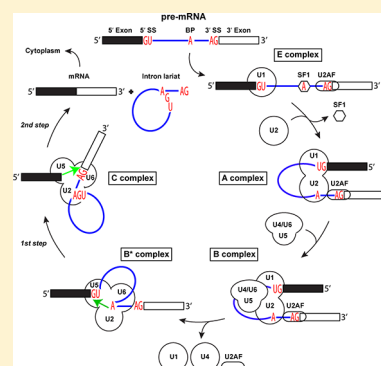


Architecture of the Spliceosome

Clarisse van der Feltz, Kelsey Anthony, Axel Brilot, and Daniel A. Pomeranz Krummel*

Department of Biochemistry, Brandeis University, 415 South Street, Waltham, Massachusetts 02454, United States

ABSTRACT: Precursor-mRNA splicing is catalyzed by an extraordinarily large and highly dynamic macromolecular assemblage termed the spliceosome. Detailed biochemical and structural study of the spliceosome presents a formidable challenge, but there has recently been significant progress made on this front highlighted by the crystal structure of a 10-subunit human U1 snRNP. This review provides an overview of our current understanding of the architecture of the spliceosome and the RNA–protein complexes integral to its function, the U snRNPs.



The simple concept of a protein-coding gene as a continuous and defined unit of DNA that gives rise to a single polypeptide was radically revised by the discovery reported in 1977 of “split genes”.^{1,2} We currently understand that most human protein-coding genes are certainly not continuous units; rather, they contain the primary protein-coding regions, exons, “split” or interrupted by introns. In constitutive precursor-mRNA splicing or splicing, introns are excised and exons spliced together to generate an mRNA. This process is catalyzed by a remarkably large and highly dynamic “machine”, the spliceosome.

Over the past decade, there has been significant progress made in understanding in detail the molecular mechanism of enzymatic assemblies critical to eukaryotic information transfer.^{3,4} In contrast, our understanding of the molecular mechanism and architecture of the spliceosome lags behind. This is due largely to challenges for biochemical and structural studies posed by the spliceosome’s very large size, highly dynamic characteristics, and a necessity thus far to use crude nuclear extract. The important role that splicing plays in metazoan development necessitates the acquisition of an improved molecular understanding. Primarily, investigators are seeking to understand how the spliceosome assembles onto its substrate (a pre-mRNA transcript) to form an active structure that will catalyze two reactions with fidelity and how this process is regulated in alternative splicing. This review presents an overview of our current understanding of the structure of the RNA–protein assemblies integral to spliceosome function (the U snRNPs) and of the intermediates in the spliceosomal reaction cycle.

SUBSTRATE

The overwhelming majority of human protein-coding genes (>90%) contain introns.⁵ Average lengths for exons and introns of human protein-coding genes are 145 and 3365 nucleotides, respectively.⁵ There is remarkably limited conservation in the primary sequence defining boundaries between exons and an

adjoining intron. More than 99% of precursor-mRNA transcripts (pre-mRNAs) share: (i) a GU at the junction between the 5′ exon and intron (5′ splice site), (ii) an adenosine within a region of the intron (branch point), and (iii) an AG at the junction between the intron and 3′ exon (3′ splice site) (Figure 1A). In metazoan pre-mRNAs, the number of nucleotides between the 5′ splice site and branch point tends to be significantly larger and the level of identity more variable than between the branch point and 3′ splice site, where there is a U-rich region of ~20–50 nucleotides, the pyrimidine tract (Py-tract) (Figure 1A). Importantly, current estimates are that more than 70% of human pre-mRNAs undergo alternative types of splicing to produce varied protein isoforms, thereby significantly increasing proteomic diversity. Guiding selection and regulation of alternative splice sites are less conserved pre-mRNA sequence(s), the structures they may form, and the auxiliary splicing proteins that may recognize them cumulatively constitute what has been termed a splicing code recognized by the spliceosome.⁶

ENZYME

The spliceosome is a nuclear assemblage that when purified in complex with a pre-mRNA is composed of at least 145 associated factors.⁷ Yeast and human spliceosomes are reported to sediment at 40–60 S and have a mass of ~4.8 MDa.^{8,9} A stable core of 45 subunits of a *Saccharomyces cerevisiae* spliceosome stalled in an intermediate state has been identified.¹⁰ While the spliceosome’s size clearly illustrates its complexity (see Table 1), it is also highly dynamic. Large subassemblies are observed in vitro to dissociate and associate in an ordered manner, and significant critical structural rearrangements take

Received: August 4, 2011

Revised: November 29, 2011

Published: April 3, 2012



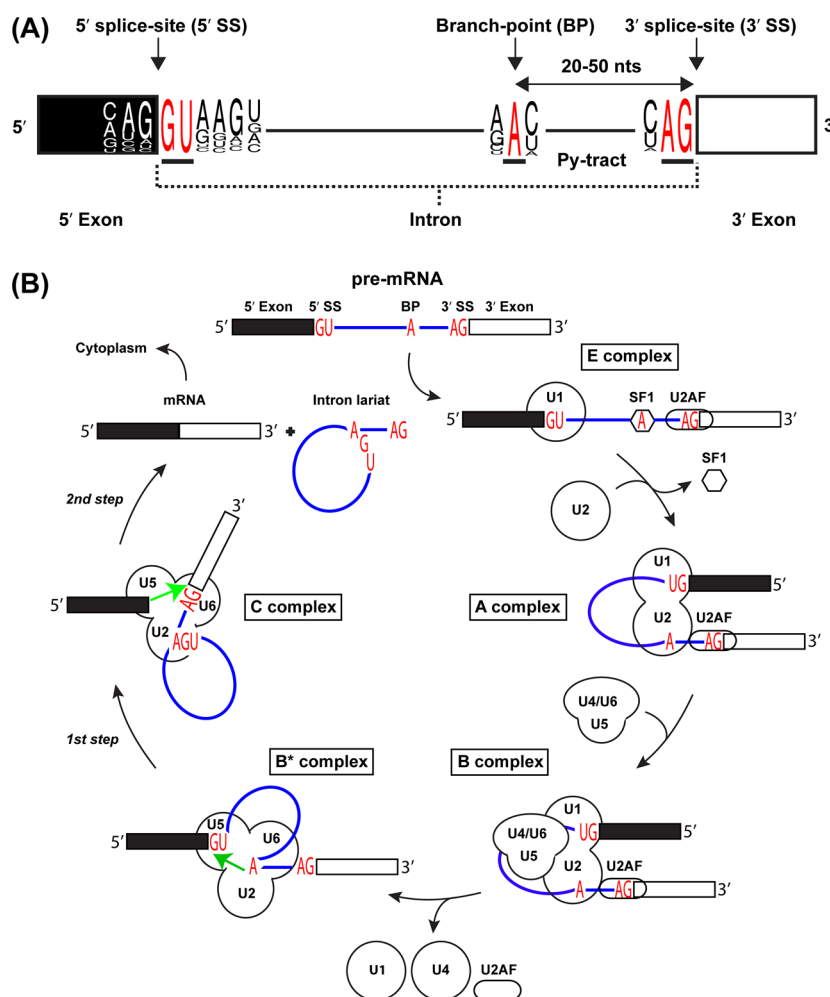


Figure 1. Precursor-mRNA recognition by the spliceosome. (A) Model metazoan precursor-mRNA transcript (pre-mRNA) consisting of 5' and 3' exons separated by an intron. Most metazoan pre-mRNAs conform to the indicated consensus sequences at the 5' exon–intron junction (5' splice site, 5' SS), branch point (BP), and 3' exon–intron junction (3' splice site, 3' SS). Invariant nucleotides at these sites are shown (underlined and colored red). Also indicated is the polypyrimidine tract (Py-tract), a sequence present in metazoan transcripts. (B) Spliceosome assembly is shown proceeding clockwise. U1 snRNP recognizes a 5' splice site (E complex). U2 snRNP associates (A complex), while the protein SF1 dissociates. The tri-snRNP (U4/U6-U5) binds (B complex). U1 and U4 snRNPs as well as the U2AF heterodimer dissociate prior to the first step or catalytic reaction (B* complex). The second step or catalytic reaction takes place (C complex), to generate an mRNA and an intron lariat. Green arrows indicate the direction of nucleophilic attack in first and second catalytic steps.

place, including between RNA subunits, to form the enzyme's active sites (Figure 1B).

Integral to spliceosome function are five RNA–protein complexes termed uridine-rich small nuclear ribonucleoprotein particles, the U1, U2, U4, U5 and U6 snRNPs. In the mammalian spliceosome, assembly intermediates are most commonly designated as E (early), A, B, B*, and C complexes and are comprised of one or more U snRNPs as well as non-U snRNP proteins in complex with a pre-mRNA (Figure 1B). Initially, the E complex is formed where U1 snRNP recognizes the 5' splice site, non-U snRNP protein splicing factor 1 (SF1) the branch point, and a heterodimer termed U2AF the Py-tract and 3' splice site. The A complex is formed as the U2 snRNP replaces SF1 to recognize the branch point. A tri-snRNP consisting of U4/U6-U5 snRNPs joins to form the B complex. After much structural and compositional rearrangement, U1 snRNP is replaced by U6 snRNP at the 5' splice site region and there is a switch to an activated B* complex that can perform the first of two catalytic reactions. A spliceosome prepared for the second step of catalysis forms the C complex.

The spliceosome requires ATP to catalyze structural rearrangements that are critical to progression through all these assembly stages.¹¹

CATALYSIS

The spliceosome is a metalloenzyme that catalyzes two consecutive nucleophilic substitutions or S_N2 phospho-transfer reactions.^{12,13} In the first reaction, the intron's branch point adenosine 2'-OH engages in a nucleophilic attack on the phosphodiester bond of the first intron nucleotide, a guanosine at the 5' splice site. A lariat-like structure forms in the intron, as the organophosphate is transferred to the branch point adenosine, and a new nucleophile is created in the now free 5' splice site 3'-OH (Figure 1B). In the second reaction, the 3'-OH at the end of the free 5' exon attacks the bridging phosphate at the 3' intron–exon junction (Figure 1B). The final two products are a spliced mRNA consisting of conjoined 5' and 3' exons and an intron in the form of a lariat-like structure.

Table 1. Core Subunits of Human U snRNPs

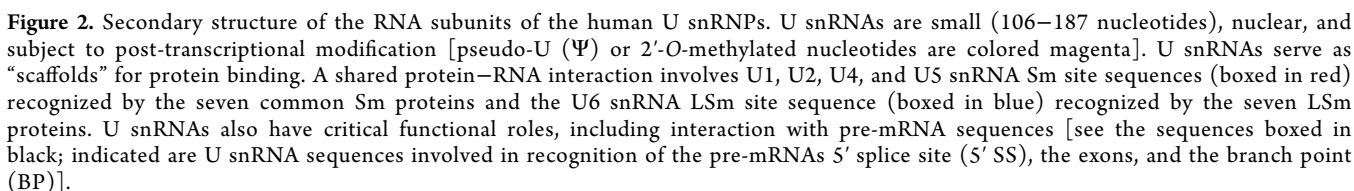
U snRNP	subunit gene	common subunit name(s) ^a	molecular mass (kDa) ^b	% U snRNP	recognizable domain/functional site ^c
U1 (248.1 kDa)	RNU1	U1 snRNA	53.5	21.6	
	SNRNPB, -B2, -D1, -D2, -D3, -E, -F, -G	seven Sm proteins	94.3	38.0	Sm
	SNRNPA	U1-A	31.3	12.6	RRM
	SNRNPA70	U1-70k	51.6	20.8	RRM; SR repeat
	SNRNPC	U1-C	17.4	7.0	Znf
U2 (987.4 kDa)	RNU2	U2 snRNA	61.2	6.2	
	SNRNPB, -B2, -D1, -D2, -D3, -E, -F, -G	seven Sm proteins	94.3	9.6	Sm
	SNRPA1	U2A'	28.4	2.9	LRR
	SNRNPB2	U2B''	25.4	2.6	RRM
	SF3A1	SF3a120	88.9	9.0	SWAP; UBQ domain
	SF3A2	SF3a66	49.3	5.0	Znf
	SF3A3	SF3a60	58.6	5.9	Znf; SAP
	SF3B1	SF3b155	145.8	14.8	HEAT repeat
	SF3B2	SF3b145	100.2	10.1	SAP
	SF3B3	SF3b130	135.5	13.7	DEH/D
	SF3B4	SF3b49	44.4	4.5	RRM
	SF3B5	SF3b10	10.1	1.0	
	SF3B14	SF3b14a; p14	14.6	1.5	RRM
	PHF5A	SF3b14b; Rds3	12.4	1.3	PHD-like
	DDX46	DDX46; hPrp5p	117.4	11.9	DEH/D; SR repeat
	SMNDC1	SPF30/SMNrp	26.7	2.7	Tudor domain
U5 (1055.7 kDa)	RNU5	U5 snRNA	37.6	3.6	
	SNRNPB, -B2, -D1, -D2, -D3, -E, -F, -G	seven Sm proteins	94.3	8.9	Sm
	TXNL4A	U5-15K	16.9	1.6	TRX
	SNRNPA40	U5-40K	39.3	3.7	WD40
	CD2BP2	U5-52K	37.6	3.6	GYF
	DDX23	U5-100K; hPrp28	95.6	9.1	DEH/D; SR repeat
	PRPF6	U5-102K; hPrp6	106.9	10.1	HAT/TPR repeats
	EFTUD2	U5-116K; hSnu114	109.4	10.4	EF2-like fold; GTPase
	SNRNPA20	U5-200K; hBrr2	244.5	23.2	DEH/D
	PRPF8	U5-220k; hPrp8	273.6	25.9	RNase H-fold; RRM; Jab1/MPN
U4/U6 (589.1 kDa)	RNU4	U4 snRNA	46.9	8.0	
	RNU6	U6 snRNA	34.6	5.9	
	SNRNPB, -B2, -D1, -D2, -D3, -E, -F, -G	seven Sm proteins	94.3	16.0	Sm
	LSM2, -3, -4, -5, -6, -7, -8	seven LSm proteins	78.9	13.4	Sm
	NHP2L1	15.5K	14.2	2.4	
	PPIH	U4/U6-20K; SnuCyp-20	19.2	3.3	cyclophilin-like
	PRPF31	U4/U6-61K; hPrp31	55.5	9.4	Nop
	PRPF4	U4/U6-60K; hPrp4	58.4	9.9	WD40
	PRPF3	U4/U6-90K; hPrp3	77.5	13.1	PWI
	SART3	p110; SART3; hPrp24	109.6	18.6	HAT repeats; RRM

^aComposition of human U snRNPs compiled from several sources.^{19,98–111} ^bRNA masses were determined using oligocalc (<http://www.basic.northwestern.edu/biotools/oligocalc.html>). Protein molecular masses were determined using ProtParam (<http://prosite.expasy.org/>). For identification of domain and functional site architecture, Prosite (<http://prosite.expasy.org/>) was used. ^cProtein domains and motifs: Sm, highly bent β -barrel structure of Sm and LSm proteins; RRM, RNA recognition motif; SR motif, sequence of Ser-Arg dipeptide repeats; Znf, zinc finger; LRR, Leu-rich repeats; SWAP, suppressor-of-white-apricot protein; UBQ, ubiquitin domain; SAP, ~35-residue motif named after proteins SAF-A/B, Acinus, and PIAS; HEAT, ~40-residue α -helical fold named after proteins Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A), and yeast PI3-kinase TOR1; DEH/D, motif characteristic of ATP-dependent helicases and nucleotide-hydrolyzing enzymes; PHD-like, structure like the PHD C₄HC₃ zinc finger; Tudor domain, ~50-residue highly bent β -barrel structure; TRX-like, thioredoxin-like fold; WD40, ~40-residue motif containing Trp-Asp dipeptide repeats; GYF, ~60-residue fold containing a Gly-Tyr-Phe conserved sequence; TPR, Tetratricopeptide repeats; HAT, half-a-TPR repeat; EF2-like, fold like that of eukaryotic translational elongation factor EF2; GTPase, motif characteristic of enzymes that bind and hydrolyze guanosine triphosphate; RNase H-fold, topologically like the ribonuclease H-fold; Jap1/MPN, five-polar residue motif resembling the active site residues of hydrolytic enzymes; cyclophilin, β -barrel topology characteristic of peptidyl-prolyl isomerases; Nop, α -helical RNA binding module; PWI, ~80-residue fold that includes a Pro-Trp-Ile tripeptide.

There are similarities in structure and catalytic mechanism between the spliceosome and the group II catalytic intron, a “ribozyme”.^{14,15} These similarities hint at the spliceosome’s possible evolutionary origins. It appears likely that RNA is the catalyst in the spliceosome, and experiments using RNAs transcribed completely in vitro have provided evidence that supports this contention.^{16,17}

■ U SNRNPS

The five U snRNPs are compositionally similar but functionally distinct. Each is composed of (i) a single uridine-rich small nuclear RNA (U snRNA), (ii) a set of seven Sm or Sm-like (LSm) proteins, and (iii) three or more U snRNP-specific proteins (see Table 1).



In the absence of a U snRNA, Sm proteins associate as stable heterodimers (SmD₁·D₂ and SmD₃·B) and a heterotrimer (SmF·E·G) (Figure 3B).²⁶ Crystal structures of SmD₃·B and SmD₁·D₂ revealed Sm fold architecture, how each Sm protein interacts with a neighbor, and led to a model for a heptameric ring.²⁵ In this model, strand β 4 of one Sm protein interacts in an antiparallel manner with strand β 5 of its neighboring Sm protein to form a continuous ring of β -sheets (Figure 3B,C). Subsequently, the crystal structure at 5.5 Å resolution of a 10-subunit human U1 snRNP revealed that in the presence of U1 snRNA a single copy of each of seven Sm proteins indeed

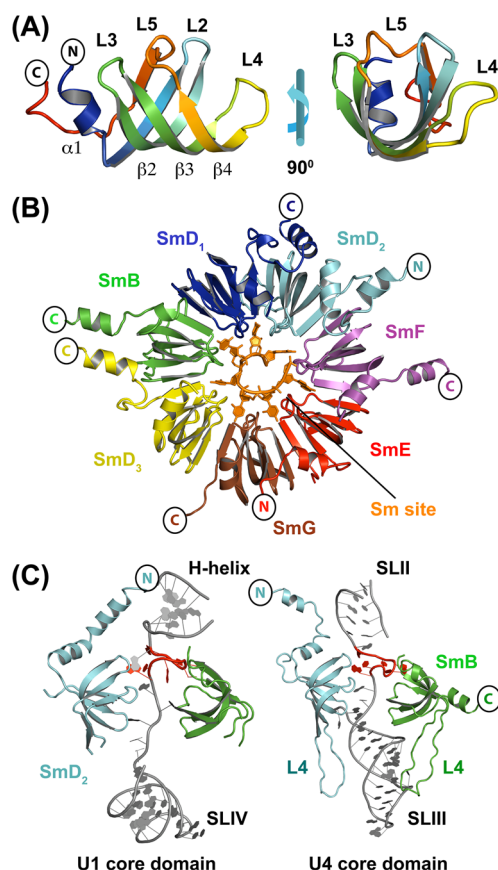


Figure 3. Structure of the U snRNP Sm core domain. (A) Structure of the Sm fold as exemplified by SmD₃. The Sm fold consists of an N-terminal α -helix followed by a highly bent five-stranded antiparallel β -sheet. Sm1 and Sm2 sequence motifs form β 1– β 3 and β 4– β 5 strands, respectively. (B) Structure of the heptameric Sm ring. The β 4 strand of one Sm protein interacts in an antiparallel manner with the β 5 strand of an Sm protein neighbor to form a continuous ring of β -sheets in the presence of a seven-nucleotide single-stranded Sm site RNA (orange). (C) Structure of U1 and U4 snRNP core domains. Shown are SmB (green) and SmD₂ (cyan) and the respective Sm sites of U1 and U4 snRNAs (red). In the U1 snRNP core domain (left), α 1 of SmD₂ points into the minor groove of an RNA duplex, while in the U4 snRNP core domain (right), α 1 of SmD₂ points away. SmD₂ and SmB L4 loops extend along the U4 snRNA 3' stem–loop structure. Structural images were created using PyMOL (<http://www.pymol.org>).

interacts pairwise to assemble into a heptameric toroidal structure or ring around the Sm site (Figure 3B).²⁷ The heptameric Sm ring, arranged in an SmE·G·D₃·B·D₁·D₂·F order, has an outer diameter of 70 Å and an inner hole of 20 Å (Figure 3B).^{27–29} The single-stranded Sm site RNA leaves through the center of this ring, the seven nucleotides assuming a striking cartwheel-like arrangement where Sm site bases radiate outward in varying orientations to interact with Sm protein residues (Figure 3B).

Details of RNA–Sm protein interactions at the Sm site have come most recently from the 3.6 Å resolution crystal structure of a U4 snRNP core domain.²⁹ Positioned into the hole of the Sm ring are L2, L3, and L5 loop residues of each Sm protein (Figure 3). L3 and L5 loop residues are located near the “top” of the hole, where they contribute to the creation of structurally and chemically distinct binding pockets to interact with seven Sm site nucleotides. The L2 loops are positioned proximal to

where the RNA emerges from the hole to form contacts with the phosphodiester backbone of the emerging RNA.

There is a second, non-sequence-specific mode of Sm protein–RNA interaction that involves N-terminal α -helices of Sm proteins. The N-terminal α -helices that mediate this interaction are not part of the canonical Sm fold but rather Sm protein-specific extensions. An N-terminal α -helix dipole of SmD₂ points into the A-form minor groove of the H helix of U1 snRNA, while the N-terminal α -helix of SmB interacts with its stem–loop 2 (SL2) motif (Figure 3C).^{27,28} These Sm–RNA interactions are proposed to stabilize and orient the Sm ring onto U1 snRNA.²⁷ Interestingly, in the U4 snRNP core domain structure, the N-terminal α -helix of SmD₂ is oriented differently than in U1 snRNP; it points away from U4 snRNA with only an asparagine of the preceding loop in position to interact with the RNA backbone (Figure 3C).²⁹ This suggests that there is a U snRNP-specific strategy for orienting the same Sm protein ring onto varying U snRNA scaffolds.

In yet a third mode of Sm protein–RNA interaction, SmD₂ and SmB loops have roles in interacting with U1 and U4 snRNAs. L4 loops of the Sm proteins are quite variable in length, with those of SmD₂ and SmB being particularly long and lysine-rich. In the U4 snRNP core domain structure, SmD₂ and SmB L4 loops radiate from the Sm ring to positions adjacent to the terminal 3' stem–loop structure of U4 snRNA (Figure 3C). Nucleotides in the terminal 3' stem–loop structure of U1 snRNA were identified as being important to the stable association of the Sm ring onto U1 snRNA,³⁰ presumably, this is also the case in the U4 core domain. SmD₂ and SmB L4 loops similarly appear to interact with the 3' stem–loop structure of U1 snRNA, as reported in the crystal structure of a proteolyzed native human U1 snRNP.²⁸ However, L4 loops in the two molecules in the asymmetric unit of this structure have different conformations, most likely because of the removal of stabilizing features as a consequence of proteolysis and/or crystal contacts. The conformation of the L4 loops in the context of the U1 snRNP therefore remains unclear.

Human U1 snRNP Structure. Structural insight into U snRNP assembly comes from crystallographic studies of the 10-subunit recombinant human U1 snRNP.²⁷ Human U1 snRNP (248 kDa) is composed of (i) a single RNA (U1 snRNA), (ii) seven Sm proteins, and (iii) three U1 snRNP-specific proteins (U1-70k, U1-C, and U1-A) (Table 1). U1 snRNA (164 nucleotides, 54 kDa) consists of four stem–loop structures (SL1–4) and a single helix H (Figures 2 and 4A). On the 5' side of the Sm site, SL1–3 and H helices emanate from a central point, a four-helix junction. Structurally, neighboring helices in the four-helix junction coaxially stack: SL1 and SL2, SL3 and H helix (Figure 4B). The lack of unpaired nucleotides at the junctions between these helices likely contributes to its overall conformation, where the coaxially stacked pair of helices cross each other at $\sim 90^\circ$. There are also two functionally significant single-stranded stretches in U1 snRNA: a strictly conserved sequence at the 5' end that base pairs with the 5' splice site and is crucial to pre-mRNA recognition³¹ and the Sm site nucleotides between SL3 and SL4, without which the particle is unable to assemble.

U1 snRNP-specific proteins U1-70k (52 kDa) and U1-A (31 kDa) bind to SL1 and SL2 of U1 snRNA, respectively, through their RNA recognition motifs (RRMs) (Figure 4A). Each is capable of binding to RNA in the absence of other proteins.^{32,33} Importantly, assembly of U1 snRNP is highly hierarchical, as

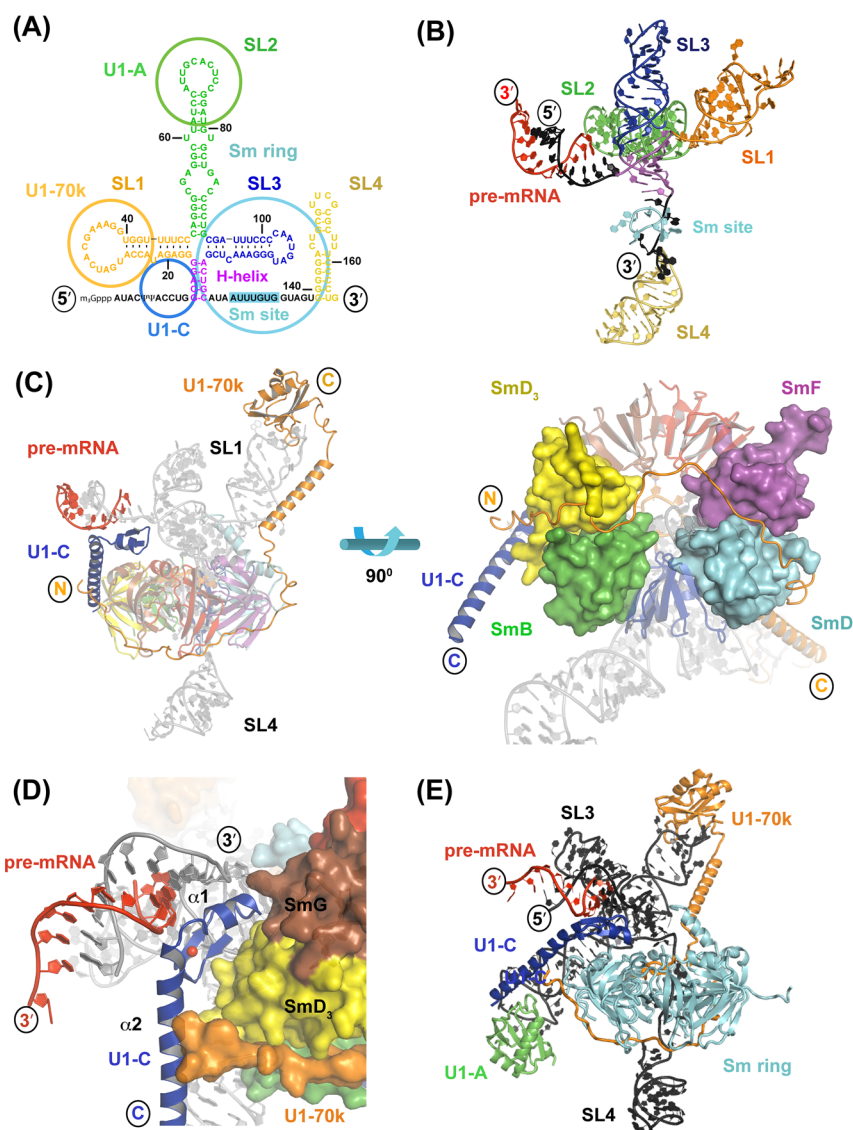


Figure 4. Structure of human U1 snRNP. (A) Secondary structure of U1 snRNA (black) and approximate location of the following: Sm proteins (cyan), recognizing the Sm site (boxed in cyan); U1-70k (orange), recognizing SL1; U1-A (green), recognizing SL2; and U1-C (blue). (B) Structure of U1 snRNA, colored as in panel A. SL1 and SL2 coaxially stack, as do SL3 and H helix to form a cruciform-like structure. (C) U1-70k (orange) extends the length of SL1 and crosses the Sm ring near the interface of SmD₂ and SmF and then that of SmD₃ and SmB to reach U1-C (blue). For the sake of clarity, U1 snRNA SL4 has been omitted from the right image. (D) The U1-C α1 helix points into the minor groove of an RNA duplex formed between the 5' end of U1 snRNA (gray) and a 5' splice site mimic (red), while the U1-C α2 helix is held in place in part by a fork created by the N- and C-termini of U1-70k (orange) and SmD₃ (yellow), respectively. (E) Nearly complete model of human U1 snRNP. SL2 of U1 snRNA and the N-terminal RRM1 (green) of U1-A are modeled onto the crystal structure of human U1 snRNP.²³ The structure is colored as in panel A, with a 5' splice site mimic displayed (red). Structural images were created using PyMOL.

best exemplified in the case of U1-C (17 kDa), which requires for its incorporation into U1 snRNP prior assembly of the Sm core and U1-70k.³⁴

U1-70k is structurally a fascinating protein. It consists of a region at its very N-terminus (residues 2–57) that while highly conserved is proline-rich and understandably predicted to be unstructured, a region predicted to form an α-helix (residues 63–88), the RRM that mediates its interaction with SL1 (residues 92–202), and a C-terminus that is rich in repeats of Arg and Ser residues (RS domain) as well as Arg-(Asp/Glu) residues. U1-70k assumes a strikingly extended path in the structure (Figure 4C,D). The ~90 N-terminal residues of U1-70k that include the proline-rich region traverse a distance of ~180 Å. The path that it takes was unambiguously established

by site-specifically incorporating the anomalous scatterer selenomethionine incrementally along its polypeptide chain.³⁵ Specifically, N-terminal to the RRM there is a α-helix (residues 63–88) that runs parallel to SL1 of U1 snRNA followed by a series of prolines that create a significant kink, as well as rigidity, in the polypeptide chain. The protein then crosses near the interface of SmD₂ and SmF, and then past the center of the Sm ring from where the single-stranded RNA emerges, and crosses near the interface of SmD₃ and SmB to finally reach a point to assist in the association of U1-C with the U1 snRNP. At this point, the very N-terminus of U1-70k and the C-terminus of SmD₃ create a binding fork through which an extended helix of U1-C traverses. Why does U1-70k take an extended path, completely wrapping underneath the Sm ring to interact with

U1-C? One may surmise that it stabilizes U1-C not only by a direct interaction, participating in the binding fork, but also by stabilizing the Sm ring entirely through which U1-C shares a significant interface with SmD₃. We expect that other U snRNP-specific proteins may similarly stabilize the Sm ring and utilize it as a binding platform.

The zinc finger of U1-C consists of a β -hairpin and two α -helices,³⁶ distinguishing it from the canonical $\beta\beta\alpha$ fold of the classical C₂H₂-type zinc finger.³⁷ Mutations in the U1-C zinc finger have a significant effect on 5' splice site recognition,³⁸ suggesting that it might interact directly with the 5' splice site and/or other regions of a pre-mRNA. The structure of human U1 snRNP is entirely consistent with these results.²⁷ Fortuitously, two of the particles in the crystal's asymmetric unit are engaged in an RNA–RNA interaction where the single-stranded 5' end of U1 snRNA from one particle and its counterpart from an adjacent particle interact with each other such that the 5' end of one mimics a 5' splice site for the other. U1-C from each particle interacts with this RNA duplex to provide an unexpected glimpse into its importance in 5' splice site recognition. Highly conserved basic residues of helix α_1 of its zinc finger are in position to interact with the minor groove of the RNA duplex (Figure 4D). However, at the resolution of this structure (5.5 Å), it is not possible to provide details concerning specific contacts. One significant question is whether U1-C functions to exclusively stabilize this duplex, a passive role, or may it also have an active role, possibly discriminating against binding of an incorrect pre-mRNA sequence(s)? The solution NMR structure of the 60 N-terminal residues of U1-C has a $\beta\beta\alpha_1\alpha_2\alpha_3$ fold,³⁶ while in the crystal structure, the second and third helices (α_2 and α_3 , respectively) form a continuous helix α_2 .²⁷ Is it possible that U1-C undergoes a structural change when in the particle and that its mode of binding to and the orientation of the 5' end of U1 snRNA are changed when in complex with a pre-mRNA? Such a binding event may trigger other important structural changes that may facilitate binding of other U snRNPs.

Early EM studies of a native U1 snRNP revealed a central globular domain, presumably the Sm core, emanating from which were two protuberances, identified as U1-70k and U1-A.³⁹ Protein U1-A is composed of two RRM. The recent structure of a proteolyzed U1 snRNP contains the N-terminal RRM of U1-A only,²⁸ leaving unknown the location of the C-terminal RRM. RRMs not only mediate interaction with RNA but also may mediate protein–protein interactions.⁴⁰ From EM images of native human U1 snRNP,^{39,41} there is a lack of density to account for a C-terminal RRM of U1-A extending away from the Sm core. More consistent would be for the C-terminal RRM to turn back and interact in some manner within the context of U1 snRNP, possibly with an internal loop of SL2 and/or the Sm core as previously proposed.²⁷

■ HUMAN SPICEOSOMAL INTERMEDIATE STRUCTURES

As detailed above, U1 and other U snRNPs assemble in an ordered manner to form the spliceosomal E, A, B, B*, and C complexes. These intermediates have been trapped *in vitro* by using a variety of approaches, including the use of modified pre-mRNA substrate(s), addition of inhibitors, or exclusion of ATP.⁴² Significant effort has been made over the past decade to acquire EM structures of spliceosomal intermediates, primarily of human origin (Figure 5A). This is a challenging undertaking

given the size and dynamics (compositional and structural) of the spliceosome.

E Complex. All conserved pre-mRNA sequences are recognized at the time of E complex formation: (i) 5' splice site by U1 snRNP, (ii) branch point by SF1, recognizing the catalytically critical adenosine through an enlarged hydrophobic cleft,^{43–45} and (iii) Py-tract and the AG dinucleotide at the 3' splice site by U2AF-65 and U2AF-35, respectively, which form the stable heterodimer U2AF.⁴⁶ While U1 snRNP, SF1, and U2AF independently interact with pre-mRNA during formation of the E complex, they also interact with each other directly or through protein intermediaries.^{43,46,47} Illustrative of spliceosome dynamics, all complexes or proteins associated with pre-mRNA at this stage are displaced from their positions prior to the first catalytic step.

A Complex. To form the A complex, there must be a displacement of some E complex-bound factor(s). Recognition of the branch point by U2 snRNP requires ATP, presumably expended to assist in an active dissociation of non-snRNP protein(s). Important interactions are also established, between the U2AF heterodimer and U2 snRNP as well as U1 and U2 snRNPs, thus juxtaposing the 5' splice site and branch point adenosine.⁴⁸ In total, the A complex consists of at least 70 proteins.^{49,50}

Human U2 snRNP (>900 kDa) consists of (i) a single U2 snRNA, (ii) seven Sm proteins, and (iii) at least 14 particle-specific proteins. A critical function of U2 snRNA is to base pair with the pre-mRNA branch site⁵¹ (Figure 2), an interaction that leaves the conserved adenosine unpaired and bulged out of the resulting RNA duplex (Figure 5B).⁵² Among U2 snRNP proteins are those that form two large multimeric complexes that contribute to the stabilization of the conformation of the branch point adenosine: SF3a (197 kDa) and SF3b (463 kDa), which consist of three and seven proteins, respectively (Table 1). One SF3b protein in particular, SF3b14a or p14, interacts with the U2 snRNA sequence that base pairs to a branch site. Structurally, p14 and SF3b155 form a binding pocket that exposes a conserved aromatic residue that in solution cross-links to the branch point adenosine.^{53–55} Adding another level of regulation prior to binding to the intron, the U2 snRNA branch point binding sequence is structurally sequestered in an intramolecular interaction that requires the activity of a helicase Prp5 before branch point interaction.⁵⁶

There are EM structures of human SF3b and an A complex reported at 10 and 45 Å resolution, respectively.^{50,57,58} In the SF3b structure [160 Å × 150 Å × 125 Å (Figure 5A)], p14 is localized to a central cavity and SF3b155, its significantly larger interacting partner, is peripheral and positioned to enclose p14. The human A complex (2.5 MDa) appears as a large asymmetric complex (205 Å × 195 Å × 150 Å; $V \approx 2.9 \times 10^6$ Å³) consisting of a tubular head connected by a short neck to a rectangular main body possessing several irregularly shaped protuberances roughly at locations that anthropomorphically can be described as feet and arms (Figure 5A).⁵⁰

B Complexes. There have been various intermediates isolated and grouped as B complexes. It remains unclear whether all these various isolated states represent functionally active intermediates in the assembly of the spliceosome. It is generally accepted that an initial stage (B complex) is compositionally defined by an association of the tri-snRNP as well as the U1 snRNP. Defining a second stage (a Δ U1 complex), the U1 snRNP has dissociated from the intermediate, while in a third stage (B* complex), there are

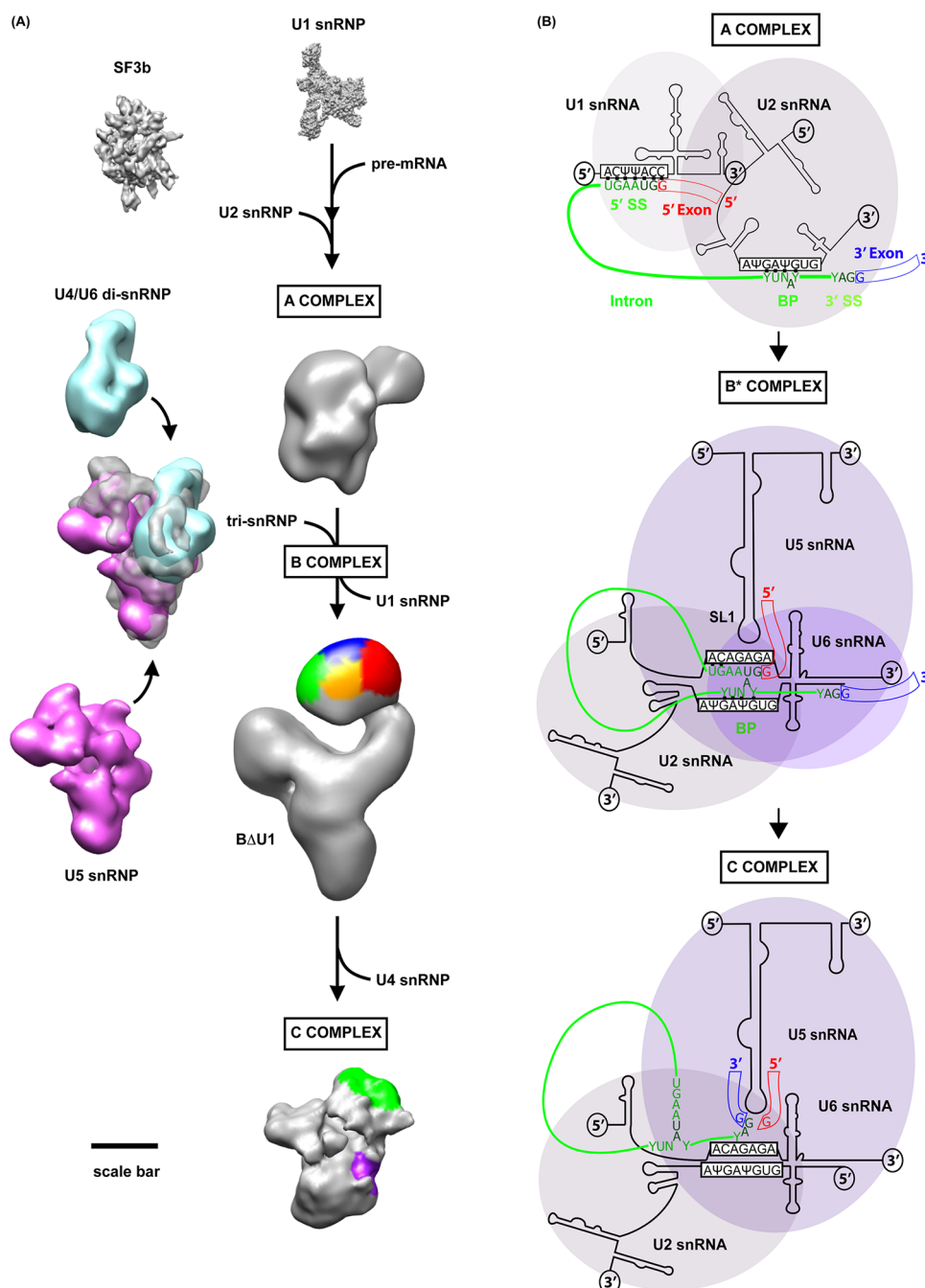


Figure 5. Spliceosomal intermediate structures and their pre-mRNA–U snRNA interactions. (A) Electron microscopic (EM) structures of human spliceosomal complexes and assembly intermediates: heteromeric SF3b complex,⁵⁶ A complex,⁵⁰ U4/U6–U5 tri-snRNP,⁶⁴ U5 snRNP,⁶⁴ U4/U6 di-snRNP,⁶⁴ BΔU1 complex,⁷⁷ and C complex.⁹³ A surface representation of the crystal structure of human U1 snRNP is shown.²⁷ The U4/U6–U5 tri-snRNP is composed of subassemblies, U4/U6 di-snRNP (cyan) and U5 snRNP (magenta). The tri-snRNP is recruited to the A complex to form the B complex. An initial B complex is defined by the presence of the tri-snRNP and U1 snRNP. Following, U1 snRNP is displaced to give rise to a BΔU1 complex, a prism-like structure similar to that of the tri-snRNP. We have modeled onto the BΔU1 complex head domain structure⁷⁷ localization studies conducted using a B complex⁷⁸: 5' exon (blue), 3' exon (red), intron (green), and SF3b155 (orange). Following loss in mass and first catalytic step, the C complex is formed. On one face of this structure is localized an intron (green) as well as exons (purple). (B) Interactions between U snRNAs and a pre-mRNA in A, B*, and C complexes (5' exon in red, 3' exon in blue, and intron in green). For the A complex, U1 and U2 snRNAs base pair with a 5' splice site (5' SS) and branch point (BP), respectively, while the 3' splice site (3' SS) is bound by protein. For the B* complex, U2 and U6 snRNAs are extensively base paired, including an interaction of U6 snRNA with nucleotides proximal to the duplex between the BP and U2 snRNA. U6 snRNA and U5 snRNA loop 1 (SL1) recognize sequences at the 5' exon–intron junction, while U2 snRNA interacts with the pre-mRNA region around the BP. These interactions contribute to nucleophilic attack of the intron's bulged branch point adenosine (BP, dark green) on the phosphodiester bond of the first intron nucleotide. A lariat-like structure forms in the intron as a product of this first reaction, and a new nucleophile is created in the now free 5' splice site 3'-OH (C complex). For the C complex, U5 snRNA SL1 contacts sequences in both 5' and 3' exons, contributing to the alignment for their ligation. The 3'-OH attacks the bridging phosphate at the 3' intron–exon junction. EM envelopes were obtained from the EM Data Bank (<http://www.ebi.ac.uk/pdbe/emdb>) and structural images generated and volumes derived using Chimera (<http://www.cgl.ucsf.edu/chimera/>). The scale bar in panel A corresponds to 100 Å.

significant rearrangements: U4/U6 snRNA base pairing (as illustrated in Figure 2) is disrupted, thereby contributing to U4 snRNP dissociation, and U2/U6 snRNA base pairing is established (Figure 4B). This last intermediate is thought to represent an intermediate that is close to the intermediate that catalyzes the first catalytic reaction.

B Complex. A defining point in the transition from the A complex to the B complex is recruitment of the U4/U6 di-snRNP (human, >580 kDa) and U5 snRNP (human, >1 MDa) (Table 1). These U snRNPs interact as a stable 25S U4/U6-U5 tri-snRNP.^{59–61} Within the tri-snRNP, U4 and U6 snRNAs form highly conserved base pairing interactions.^{62,63} EM structures of human U5 snRNP, U4/U6 di-snRNP, and a tri-snRNP have been reported at 29, 40, and 21 Å resolution, respectively.⁶⁴ U5 snRNP exhibits a crescent-shaped morphology (265 Å × 150 Å × 120 Å), while U4/U6 di-snRNP has a cylindrical shape that is 170–210 Å in length. The tri-snRNP appears as a relatively elongated tetrahedron (305 Å × 200 Å × 175 Å) (Figure 5A). Interestingly, the tri-snRNP and downstream assembly intermediates [B(Δ)U1 and C complex] all possess a triangular prism-like shape.

There are >100 proteins associated with the early B complex.⁶⁵ Among its proteins, three large U5 snRNP constituents in particular have important roles: (i) highly conserved subunit Prp8 (274 kDa), (ii) helicase Brr2 (245 kDa), and (iii) GTPase Snu114 (109 kDa) (Table 1). Brr2 is a member of a family of proteins that contains a DExD/H-box domain, a motif involved in ATP-dependent restructuring of nucleic acids.⁶⁶ Studies that aimed to localize these three massive proteins in an EM envelope of a *S. cerevisiae* tri-snRNP revealed that Prp8 and Snu114 are localized centrally while Brr2 is localized more distally.^{65,67} Once the tri-snRNP is integrated into the A complex to undergo the transition to the B complex, interactions of Prp8 with both pre-mRNA and other proteins appear to be quite extensive.^{68,69} As an indication of Prp8 having a role in ensuring catalytic activation, it cross-links at both 5' and 3' splice sites as well as the branch site, interacts with both U5 and U6 snRNAs, and can activate helicase Brr2 to catalyze base pair unwinding between U4 and U6 snRNAs.^{68–70}

Adding to our functional understanding of Prp8 and its relative position in the context of the spliceosome are crystal structures of various domains.^{71–73} The structure of a region near the C-terminus of *S. cerevisiae* Prp8 reveals a β-hairpin finger that protrudes out of the overall fold to mediate U4/U6 unwinding by Brr2.⁷¹ The structure of another domain of Prp8 revealed that it forms an RNase H-fold, although equivalent RNase catalytic residues are absent.⁷² The RNase H-fold forms a binding surface that may function to contribute to stabilizing a 5' exon during catalysis.⁷²

BΔU1 Complex. A primary function of the U1 snRNP is to initiate spliceosome assembly. In metazoans, U1 snRNP is also a target for alternative splicing factors and may serve to interact with nonspliceosomal macromolecular assemblies.^{74–76} U1 snRNP has no direct contribution to spliceosome catalysis and is displaced from B complex prior to the first catalytic step, presumably to allow for U6 snRNA binding (Figure 5B).

A 45S BΔU1 complex (4.8–5.5 MDa) has been isolated and visualized by EM.⁷⁷ The structure of this large complex at ~40 Å resolution reveals a rhomboid-shaped intermediate (370 Å × 270 Å × 170 Å; $V \approx 5.5 \times 10^6 \text{ Å}^3$), characterized by a headlike globular region attached to a larger body by a linker (Figure 5A). The head region exhibits more flexibility than the main body.

The 5' and 3' exons as well as the intron all are localized to this head region, like SF3b155 (Figure 5A).⁷⁸ On the basis of these localization studies, might it be that this seemingly distinct globular domain contains all the elements necessary for the first step of catalysis? Clearly, the C complex that catalyzes the second catalytic reaction lacks such a distinct globular domain. Might this suggest that the active site for the second reaction is significantly different in locale from the first reaction?

B* Complex. The B* complex is proposed to represent the B complex series closest to the catalytically activated form of the spliceosome.⁷⁹ Indeed, it lacks U1 and U4 snRNPs, implying that it has undergone significant RNA rearrangements. A catalytically inactive complex has also been isolated, termed B^{act}, which is compositionally very similar to the B* complex, including the absence of the U1 and U4 snRNPs as well as most U6 snRNP proteins.⁸⁰ However, this complex possibly reflects a state earlier than the B* complex because of the presence of a number of helicases necessary for rearrangements that precede catalysis.

Formation of the B* complex involves significant structural rearrangements to reposition the U2, U5, and U6 snRNPs for catalysis: (i) U5 snRNA is positioned to interact with 5' and 3' exons through its stem-loop I structure^{81,82} (Figure 2); (ii) U2 and U6 snRNAs base pair;⁸³ and (iii) conserved nucleotides of U6 snRNA are positioned to interact in the region of the 5' splice site⁸⁴ (Figures 2 and 5B). Some interactions remain the same; for example, U2 snRNA is still in contact with the branch site. The interaction between U2 and U6 snRNAs, following unwinding of U4/U6 snRNAs, is proposed to juxtapose the 5' splice site and branch point adenosine and position metal cation(s) for catalysis.⁸⁵

Among the many RNA–RNA and RNA–protein interactions destabilized or stabilized in the transition from the A complex to the B complex as well as during the existence of the B complex, the addition of the Prp19–CDC5L protein complex to transition to the B* complex is particularly important.⁸⁶ This hetero-oligomeric complex (600 kDa) acts to facilitate dissociation of the U6 snRNP LSM ring as well as stabilize U5 and U6 snRNAs.⁸⁷ Another prominent change in the transition to the B* complex involves the apparent loss of U2 snRNP sub-complexes SF3a and SF3b, which may act to free the branch point adenosine for catalysis.^{88,89}

C Complex. Following the first catalytic reaction, the C complex contains the 5' exon and the intron lariat attached to the 3' splice site (Figure 5B). The C complex must establish the necessary interactions for nucleophilic attack by the 3' splice site's 3'-OH (Figures 1B and 5B). Indeed, there are dramatic structural rearrangements that occur at this stage, with at least 50 proteins either associating or dissociating in the transition to the C complex.^{88,89} This extensive loss of mass is clear upon examination of the BΔU1 and C complex EM structures (Figure 5A). In addition, RNA interactions present in the activated B complex are significantly restructured in the C complex. In particular, the important U5 snRNA stem-loop 1 motif is restructured to contact both 5' and 3' exons (Figure 5B, reviewed in ref 90), and U2·U6 snRNA base pairing is unwound around an AGC triad in U6 snRNA.⁹¹ Cross-linking studies with a U2·U6 snRNA complex with 5' and 3' exon oligonucleotides indicate that a 5' exon contacts U6 snRNA at a conserved sequence, 5'-ACAGAGA (Figure 2), as well as

the AGC triad located at the base of its 3' stem-loop structure.⁹²

A 40S human C complex (3 MDa) has been isolated using a noncleavable substrate.⁹³ Its structure ($270 \text{ \AA} \times 240 \text{ \AA} \times 220 \text{ \AA}$; $V \approx 3.0 \times 10^6 \text{ \AA}^3$) is composed of primarily three domains: (i) a large ovoid-shaped main body domain, (ii) a protruding arm emanating from the body, and (iii) a small domain slightly detached from the rest of the structure reminiscent of the head from the $\Delta U1$ complex intermediate (Figure 5A). The intron has been localized to the top of the small domain, whereas both exons are in the large domain just below it.^{93–95} More recently, a human C complex structure at $\sim 25 \text{ \AA}$ resolution has been reported⁹⁶ that is larger ($5\text{--}5.5 \text{ MDa}$, $360 \text{ \AA} \times 340 \text{ \AA} \times 270 \text{ \AA}$) than both the earlier structure and that of a *Schizosaccharomyces pombe* U2/U5/U6 complex.⁹⁷ This C complex structure appears to exhibit internal density variation.⁹⁶ The reason(s) for size discrepancy between these different C complex structures is unclear, but it may be a consequence of differences in purification and/or methods used to determine the structures.

CONCLUDING REMARKS

There has been significant progress the past decade in defining and characterizing assembly intermediates of the spliceosome's reaction cycle. However, the spliceosome's size and dynamics have presented a formidable challenge for detailed biochemical and structural studies. Progress has been made on this front; the recent crystal structures of completely recombinant human U1 snRNP and the Sm core of U4 snRNP detailed above have provided significant insight into the assembly and function of U snRNPs. Much headway has also been made in elucidating by EM the gross morphology of the large spliceosomal intermediates. While the EM structures are limited to a quite low resolution as a consequence of their compositional and structural heterogeneity, recent studies that aimed to localize regions of both pre-mRNA and protein(s) in these structures are providing insightful detail.^{65,67,78,95} In particular, these studies are revealing architectural similarities between intermediate structures. As we advance on this front, it is important to recognize that the analysis and interpretation of the EM structures may be complicated by ambiguity with respect to which intermediate state it represents. Indeed, EM structures from different groups are not entirely consistent.^{93,96} As we look to the future, there needs to be progress made in acquiring (i) intermediates stalled at discrete, well-defined steps in the spliceosome assembly cycle and (ii) more data from labeling studies so as to confidently localize RNA and proteins as well as to place higher-resolution structures into low-resolution EM maps. While there is much to do, the community is making progress in illuminating how this remarkably large and dynamic macromolecular machine, the spliceosome, is assembled to splice split genes.

AUTHOR INFORMATION

Corresponding Author

*E-mail: dapk@brandeis.edu. Telephone: (781) 736-2359. Fax: (781) 736-2349.

Author Contributions

K.A. and A.B. contributed equally to this work.

ACKNOWLEDGMENTS

We thank K. Saha and A. Cheng for their support. D.A.P.K. thanks J. Li, K. Nagai, C. Oubridge, and N. Grigorieff for continued support and comments. We thank A. K. W. Leung for supplying a PyMOL session of the U4 Sm core domain structure prior to its publication.

ABBREVIATIONS

pre-mRNA, precursor-mRNA; U snRNP, uridine-rich small nuclear ribonucleoprotein; EM, electron microscopy; Py-tract, pyrimidine tract.

REFERENCES

- (1) Berget, S. M., Moore, C., and Sharp, P. A. (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3171–3175.
- (2) Chow, L. T., Gelin, R. E., Broker, T. R., and Roberts, R. J. (1977) An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**, 1–8.
- (3) Kornberg, R. D. (2007) The molecular basis of eukaryotic transcription. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12955–12961.
- (4) Schmeing, T. M., and Ramakrishnan, V. (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461**, 1234–1242.
- (5) International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945.
- (6) Barash, Y., Calarco, J. A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B. J., and Frey, B. J. (2010) Deciphering the splicing code. *Nature* **465**, 53–59.
- (7) Zhou, Z., Licklider, L. J., Gygi, S. P., and Reed, R. (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**, 182–185.
- (8) Brody, E., and Abelson, J. (1985) The "spliceosome": Yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science* **228**, 963–967.
- (9) Müller, S., Wolpensinger, B., Angenitzki, M., Engel, A., Sperling, J., and Sperling, R. (1998) A supraspliceosome model for large nuclear ribonucleoprotein particles based on mass determinations by scanning transmission electron microscopy. *J. Mol. Biol.* **283**, 383–394.
- (10) Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2009) The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol. Cell* **36**, 593–608.
- (11) Staley, J. P., and Guthrie, C. (1998) Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell* **92**, 315–326.
- (12) Moore, M. J., and Sharp, P. A. (1993) Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* **365**, 364–368.
- (13) Sontheimer, E. J., Sun, S., and Piccirilli, J. A. (1997) Metal ion catalysis during splicing of premessenger RNA. *Nature* **388**, 801–805.
- (14) Sashital, D. G., Cornilescu, G., McManus, C. J., Brow, D. A., and Butcher, S. E. (2004) U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction. *Nat. Struct. Mol. Biol.* **11**, 1237–1242.
- (15) Keating, K. S., Toor, N., Perlman, P. S., and Pyle, A. M. (2010) A structural analysis of the group II intron active site and implications for the spliceosome. *RNA* **16**, 1–9.
- (16) Valadkhan, S., and Manley, J. L. (2001) Splicing-related catalysis by protein-free snRNAs. *Nature* **413**, 701–707.
- (17) Valadkhan, S., Mohammadi, A., Jaladat, Y., and Geisler, S. (2009) Protein-free small nuclear RNAs catalyze a two-step splicing reaction. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11901–11906.
- (18) Hamm, J., Darzynkiewicz, E., Tahara, S. M., and Mattaj, I. W. (1990) The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal. *Cell* **62**, 569–577.

- (19) Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., and Lührmann, R. (1999) A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation *in vitro*. *EMBO J.* 18, 5789–5802.
- (20) Massenet, S., Mougin, A., and Branlant, C. (1998) Posttranscriptional modification in the U small nuclear RNAs. In *Modification and Editing of RNA* (Grosjean, H. A. B. R., Ed.) pp 201–227, American Society for Microbiology Press, Washington, DC.
- (21) Yu, Y.-T., Shu, M.-D., and Steitz, J. A. (1998) Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *EMBO J.* 17, 5783–5795.
- (22) Newby, M. L., and Greenbaum, N. L. (2001) A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. *RNA* 7, 833–845.
- (23) Hermann, H., Fabrizio, P., Raker, V. A., Foulaki, K., Hornig, H., Brahms, H., and Lührmann, R. (1995) snRNP SM proteins share two evolutionary conserved sequence motifs which are involved in Sm protein-interactions. *EMBO J.* 14, 2076–2088.
- (24) Seraphin, B. (1995) Sm and Sm-like proteins belong to a large family: Identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J.* 14, 2089–2098.
- (25) Kambach, C., Walke, S., Young, R., Avis, J. M., de la Fortelle, E., Raker, V. A., Lührmann, R., Li, J., and Nagai, K. (1999) Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* 96, 375–387.
- (26) Raker, V. A., Plessel, G., and Lührmann, R. (1996) The snRNP core assembly pathway: Identification of stable core protein heteromeric complexes and an snRNP subcore particle *in vitro*. *EMBO J.* 15, 2256–2269.
- (27) Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K. W., Li, J., and Nagai, K. (2009) Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* 458, 475–480.
- (28) Weber, G., Trowitzsch, S., Kastner, B., Lührmann, R., and Wahl, M. C. (2010) Functional organization of the Sm core in the crystal structure of human U1 snRNP. *EMBO J.* 29, 4172–4184.
- (29) Leung, A. K., Nagai, K., and Li, J. (2011) Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature* 473, 536–539.
- (30) McConnell, T. S., Lokken, R. P., and Steitz, J. A. (2003) Assembly of the U1 snRNP involves interactions with the backbone of the terminal stem of U1 snRNA. *RNA* 9, 193–201.
- (31) Zhuang, Y., and Weiner, A. M. (1986) A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* 46, 827–835.
- (32) Patton, J. R., and Pederson, T. (1988) The Mr 70,000 protein of the U1 small nuclear ribonucleoprotein particle binds to the 59 stem-loop of U1 RNA and interacts with Sm domain proteins. *Proc. Natl. Acad. Sci. U.S.A.* 85, 747–751.
- (33) Scherly, D., Boelens, W., Dathan, N. A., van Venrooij, W. J., and Mattaj, I. W. (1990) Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B' and their cognate RNAs. *Nature* 345, 502–506.
- (34) Nelissen, R. L., Will, C. L., van Venrooij, W. J., and Lührmann, R. (1994) The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. *EMBO J.* 13, 4113–4125.
- (35) Oubridge, C., Pomeranz Krummel, D. A., Leung, A., Li, J., and Nagai, K. (2009) Interpreting a low resolution map of human U1 snRNP using anomalous scatterers. *Structure* 17, 930–938.
- (36) Muto, Y., Pomeranz Krummel, D. A., Oubridge, C., Hernandez, H., Robinson, C. V., Neuhaus, D., and Nagai, K. (2004) The structure and biochemical properties of the human spliceosomal protein U1C. *J. Mol. Biol.* 341, 185–198.
- (37) Miller, J., McLachlan, A. D., and Klug, A. (1985) Repetitive zinc-binding domains in the proteins transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* 4, 1609–1614.
- (38) Will, C. L., Rumpel, S., Gunnewiek, J. K., van Venrooij, W. J., and Lührmann, R. (1996) *In vitro* reconstitution of mammalian U1 snRNPs active in splicing: The U1-C protein enhances the formation of early (E) spliceosomal complexes. *Nucleic Acids Res.* 24, 4614–4623.
- (39) Kastner, B., Kornstadt, U., Bach, M., and Lührmann, R. (1992) Structure of the small nuclear RNP particle-U1: Identification of the two structural protuberances with RNP-antigen-A and RNP-antigen-70k. *J. Cell Biol.* 116, 839–849.
- (40) Fribourg, S., Gatfield, D., Izaurralde, E., and Conti, E. (2003) A novel mode of RBD-protein recognition in the Y14–Mago complex. *Nat. Struct. Biol.* 10, 433–439.
- (41) Stark, H., Dube, P., Lührmann, R., and Kastner, B. (2001) Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* 409, 539–542.
- (42) Jurica, M. S., and Moore, M. J. (2002) Capturing splicing complexes to study structure and mechanism. *Methods* 22, 336–345.
- (43) Berglund, J. A., Abovich, N., and Rosbash, M. (1998) A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev.* 12, 858–867.
- (44) Berglund, J. A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* 89, 781–787.
- (45) Liu, Z., Luyten, I., Bottomley, M. J., Messias, A. C., Houngrinou-Molango, S., Sprangers, R., Zanier, K., Kramer, A., and Sattler, M. (2001) Structural Basis for Recognition of the Intron Branch Site RNA by Splicing Factor 1. *Science* 294, 1098–1103.
- (46) Ruskin, B., Zamore, P. D., and Green, M. R. (1988) A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* 52, 207–219.
- (47) Abovich, N., and Rosbash, M. (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* 89, 403–412.
- (48) Dönmez, G., Hartmuth, K., Kastner, B., Will, C. L., and Lührmann, R. (2007) 5' end of U2 snRNA is in close proximity to U1 and functional sites of the pre-mRNA in early spliceosomal complexes. *Mol. Cell* 25, 399–411.
- (49) Hartmuth, K., Urlaub, H., Vornlocher, H.-P., Will, C. L., Gentzel, M., Wilm, M., and Lührmann, R. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc. Natl. Acad. Sci. U.S.A.* 99, 16719–16724.
- (50) Behzadnia, N., Golas, M. M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C. L., Urlaub, H., Stark, H., and Lührmann, R. (2007) Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO J.* 26, 1737–1748.
- (51) Parker, R., Siliciano, P. G., and Guthrie, C. (1987) Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell* 49, 229–239.
- (52) Query, C. C., Moore, M. J., and Sharp, P. A. (1994) Branch nucleophile selection in pre-mRNA splicing: Evidence for the bulged duplex model. *Genes Dev.* 8, 587–597.
- (53) MacMillan, A. M., Query, C. C., Allerson, C. R., Chen, S., Verdine, G. L., and Sharp, P. A. (1994) Dynamic association of proteins with the pre-mRNA branch region. *Genes Dev.* 8, 3008–3020.
- (54) Will, C. L., Schneider, C., MacMillan, A. M., Katopodis, N. F., Neubauer, G., Wilm, M., Lührmann, R., and Query, C. C. (2001) A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO J.* 20, 4536–4546.
- (55) Schellenberg, M. J., Edwards, R. A., Ritchie, D. B., Kent, O. A., Golas, M. M., Stark, H., Lührmann, R., Mark Glover, J. N., and MacMillan, A. M. (2006) Crystal structure of a core spliceosomal protein interface. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1266–1271.
- (56) Golas, M. M., Sander, B., Will, C. L., Lührmann, R., and Stark, H. (2003) Molecular architecture of the multiprotein splicing factor SF3b. *Science* 300, 980–985.
- (57) Golas, M. M., Sander, B., Will, C. L., Lührmann, R., and Stark, H. (2005) Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. *Mol. Cell* 17, 869–883.

- (58) Perriman, R., and Ares, M. (2010) Invariant U2 snRNA Nucleotides Form a Stem Loop to Recognize the Intron Early in Splicing. *Mol. Cell* 38, 416–427.
- (59) Konarska, M. M., and Sharp, P. A. (1987) Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell* 49, 763–774.
- (60) Cheng, S. C., and Abelson, J. (1987) Spliceosome assembly in yeast. *Genes Dev.* 1, 1014–1027.
- (61) Black, D. A., and Pinto, A. L. (1989) U5 small nuclear ribonucleoprotein: RNA structure analysis and ATP-dependent interaction with U4/U6. *Mol. Cell. Biol.* 9, 3350–3359.
- (62) Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, H., and Lührmann, R. (1984) Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. *EMBO J.* 3, 1357–1363.
- (63) Brow, D. A., and Guthrie, C. (1988) Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature* 334, 213–218.
- (64) Sander, B., Golas, M. M., Makarov, E. M., Brahms, H., Kastner, B., Lührmann, R., and Stark, H. (2006) Organization of Core Spliceosomal Components U5 snRNA Loop I and U4/U6 Di-snRNP within U4/U6-U5 Tri-snRNP as Revealed by Electron Cryomicroscopy. *Mol. Cell* 24, 267–278.
- (65) Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C. L., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2006) Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol. Cell. Biol.* 26, 5528–5543.
- (66) Jankowsky, E., and Fairman, M. E. (2007) RNA helicases: One fold for many functions. *Curr. Opin. Struct. Biol.* 17, 316–324.
- (67) Hacker, I., Sander, B., Golas, M. M., Wolf, E., Karagoz, E., Kastner, B., Stark, H., Fabrizio, P., and Lührmann, R. (2008) Localization of Prp8, Brr2, Snu114 and U4/U6 proteins in the yeast tri-snRNP by electron microscopy. *Nat. Struct. Mol. Biol.* 15, 1206–1212.
- (68) Grainger, R. J., and Beggs, J. D. (2005) Prp8 protein: At the heart of the spliceosome. *RNA* 11, 533–557.
- (69) Turner, I. A., Norman, C. M., Churcher, M. J., and Newman, A. J. (2006) Dissection of Prp8 protein defines multiple interactions with crucial RNA sequences in the catalytic core of the spliceosome. *RNA* 12, 375–386.
- (70) Maeder, C., Kutach, A. K., and Guthrie, C. (2009) ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. *Nat. Struct. Mol. Biol.* 16, 42–48.
- (71) Ritchie, D. B., Schellenberg, M. J., Gesner, E. M., Raithatha, S. A., Stuart, D. T., and Macmillan, A. M. (2008) Structural elucidation of a PRP8 core domain from the heart of the spliceosome. *Nat. Struct. Mol. Biol.* 15, 1199–1205.
- (72) Pena, V., Rozov, A., Fabrizio, P., Lührmann, R., and Wahl, M. C. (2008) Structure and function of an RNase H domain at the heart of the spliceosome. *EMBO J.* 27, 2929–2940.
- (73) Yang, K., Zhang, L., Xu, T., Heroux, A., and Zhao, R. (2008) Crystal structure of the β -finger domain of Prp8 reveals analogy to ribosomal proteins. *Proc. Natl. Acad. Sci. U.S.A.* 105, 13817–13822.
- (74) Gunderson, S. I., Polycarpou-Schwarz, M., and Mattaj, I. W. (1998) U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol. Cell* 1, 255–264.
- (75) Fortes, P., Kufel, J., Fornerod, M., Polycarpou-Schwarz, M., Lafontaine, D., Tollervey, D., and Mattaj, I. W. (1999) Genetic and physical interactions involving the yeast nuclear cap-binding complex. *Mol. Cell. Biol.* 19, 6543–6553.
- (76) Das, R., Yu, J., Zhang, Z., Gygi, M. P., Krainer, A. R., Gygi, S. P., and Reed, R. (2007) SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol. Cell* 26, 867–881.
- (77) Boehringer, D., Makarov, E. M., Sander, B., Makarova, O., Kastner, B., Lührmann, R., and Stark, H. (2004) Three-dimensional structure of a pre-catalytic human spliceosomal complex B. *Nat. Struct. Mol. Biol.* 11, 463–468.
- (78) Wolf, E., Kastner, B., Deckert, J., Merz, C., Stark, H., and Lührmann, R. (2009) Exon, intron and splice site locations in the spliceosomal B complex. *EMBO J.* 28, 2283–2292.
- (79) Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L., Wilm, M., and Lührmann, R. (2002) Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* 298, 2205–2208.
- (80) Bessonov, S., Anokhina, M., Krasauskas, A., Golas, M. M., Sander, B., Will, C. L., Urlaub, H., Stark, H., and Lührmann, R. (2010) Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA* 16, 2384–2403.
- (81) Wyatt, J. R., Sontheimer, E. J., and Steitz, J. A. (1992) Site-specific cross-linking of mammalian U5 snRNP to the 5' splice-site before the 1st step of pre-messenger RNA splicing. *Genes Dev.* 6, 2542–2553.
- (82) Okeefe, R. T., Norman, C., and Newman, A. J. (1996) The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast. *Cell* 86, 679–689.
- (83) Madhani, H. D., and Guthrie, C. (1992) A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* 71, 803–817.
- (84) Kandels-Lewis, S., and Seraphin, S. (1993) Role of U6 snRNA in 5' splice site selection. *Science* 262, 2035–2039.
- (85) Valadkhan, S. (2010) Role of the snRNAs in spliceosomal active site. *RNA Biol.* 7, 345–353.
- (86) Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C. L., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2006) Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol. Cell. Biol.* 26, 5528–5543.
- (87) Hogg, R., McGrail, J. C., and O'Keefe, R. T. (2010) The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. *Biochem. Soc. Trans.* 38, 1110–1115.
- (88) Bessonov, S., Anokhina, M., Will, C. L., Urlaub, H., and Lührmann, R. (2008) Isolation of an active step I spliceosome and composition of its RNP core. *Nature* 452, 846–851.
- (89) Jurica, M. S., Licklider, L. J., Gygi, S. R., Grigorieff, N., and Moore, M. J. (2002) Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA* 8, 426–439.
- (90) Turner, I. A., Norman, C. M., Churcher, M. J., and Newman, A. J. (2004) Roles of the U5 snRNP in spliceosome dynamics and catalysis. *Biochem. Soc. Trans.* 32, 928–931.
- (91) Mefford, M. A., and Staley, J. P. (2009) Evidence that U2/U6 helix I promotes both catalytic steps of pre-mRNA splicing and rearranges in between these steps. *RNA* 15, 1386–1397.
- (92) Lee, C., Jaladat, Y., Mohammadi, A., Sharifi, A., Geisler, S., and Valadkhan, S. (2010) Metal binding and substrate positioning by evolutionarily invariant U6 sequences in catalytically active protein-free snRNAs. *RNA* 16, 2226–2238.
- (93) Jurica, M. S., Sousa, D., Moore, M. J., and Grigorieff, N. (2004) Three-dimensional structure of C complex spliceosomes by electron microscopy. *Nat. Struct. Mol. Biol.* 11, 265–269.
- (94) Stroupe, M. E., Xu, C., Goode, B. L., and Grigorieff, N. (2009) Actin filament labels for localizing protein components in large complexes viewed by electron microscopy. *RNA* 15, 244–248.
- (95) Alcidi, E. A., and Jurica, M. S. (2008) A protein-based EM label for RNA identifies the location of exons in spliceosomes. *Nat. Struct. Mol. Biol.* 15, 213–215.
- (96) Golas, M. M., Sander, B., Bessonov, S., Grote, M., Wolf, E., Kastner, B., Stark, H., and Lührmann, R. (2010) 3D cryo-EM structure of an active step I spliceosome and localization of its catalytic core. *Mol. Cell* 40, 927–938.
- (97) Ohi, M. D., Ren, L., Wall, J. S., Gould, K. L., and Walz, T. (2007) Structural characterization of the fission yeast U5-U2/U6 spliceosome complex. *Proc. Natl. Acad. Sci. U.S.A.* 104, 3195–3200.

- (98) Bringmann, P., and Lührmann, R. (1986) Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. *EMBO J.* 5, 3509–3516.
- (99) Bach, M., Winkelmann, G., and Lührmann, R. (1989) 20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition. *Proc. Natl. Acad. Sci. U.S.A.* 86, 6038–6042.
- (100) Lauber, J., Fabrizio, P., Teigelkamp, S., Lane, W. S., Hartmann, E., and Lührmann, R. (1996) The HeLa 200 kDa U5 snRNP-specific protein and its homologue in *Saccharomyces cerevisiae* are members of the DEXH-box protein family of putative RNA helicases. *EMBO J.* 15, 4001–4015.
- (101) Teigelkamp, S., Mundt, C., Achsel, T., Will, C. L., and Lührmann, R. (1997) The human U5 snRNP-specific 100-kD protein is an RS domain-containing, putative RNA helicase with significant homology to the yeast splicing factor Prp28p. *RNA* 3, 1313–1326.
- (102) Fabrizio, P., Laggenbauer, B., Lauber, J., Lane, W. S., and Lührmann, R. (1997) An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. *EMBO J.* 16, 4092–4106.
- (103) Horowitz, D. S., Kobayashi, R., and Krainer, A. R. (1997) A new cyclophilin and the human homologues of yeast Prp3 and Prp4 form a complex associated with U4/U6 snRNPs. *RNA* 3, 1374–1387.
- (104) Lauber, J., Plessel, G., Prehn, S., Will, C. L., Fabrizio, P., Gröning, K., Lane, W. S., and Lührmann, R. (1997) The human U4/U6 snRNP contains 60 and 90kD proteins that are structurally homologous to the yeast splicing factors Prp4p and Prp3p. *RNA* 3, 926–941.
- (105) Achsel, T., Ahrens, K., Brahms, H., Teigelkamp, S., and Lührmann, R. (1998) The human U5–220kD protein (hPrp8) forms a stable RNA-free complex with several U5-specific proteins, including an RNA unwindase, a homologue of ribosomal elongation factor EF-2, and a novel WD-40 protein. *Mol. Cell. Biol.* 18, 6756–6766.
- (106) Reuter, K., Nottrott, S., Fabrizio, P., Lührmann, R., and Ficner, R. (1999) Identification, characterization and crystal structure analysis of the human spliceosomal U5 snRNP-specific 15 kD protein. *J. Mol. Biol.* 294, 515–525.
- (107) Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R., and Lührmann, R. (1999) Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J.* 18, 6119–6133.
- (108) Makarov, E. M., Makarova, O. V., Achsel, T., and Lührmann, R. (2000) The human homologue of the yeast splicing factor prp6p contains multiple TPR elements and is stably associated with the U5 snRNP via protein-protein interactions. *J. Mol. Biol.* 298, 567–575.
- (109) Will, C. L., Urlaub, H., Achsel, T., Gentzel, M., Wilm, M., and Lührmann, R. (2002) Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. *EMBO J.* 21, 4978–4988.
- (110) Makarova, O. V., Makarov, E. M., Liu, S., Vornlocher, H. P., and Lührmann, R. (2002) Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6*U5 tri-snRNP formation and pre-mRNA splicing. *EMBO J.* 21, 1148–1157.
- (111) Bell, M., Schreiner, S., Damianov, A., Reddy, R., and Bindereif, A. (2002) p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. *EMBO J.* 21, 2724–2735.