

RESEARCH ARTICLES

STRUCTURAL BIOLOGY

Structure of a yeast spliceosome at 3.6-angstrom resolution

Chuangye Yan,^{1*} Jing Hang,^{1*} Ruixue Wan,^{1*} Min Huang,²
Catherine C. L. Wong,² Yigong Shi^{1†}

Splicing of precursor messenger RNA (pre-mRNA) in yeast is executed by the spliceosome, which consists of five small nuclear ribonucleoproteins (snRNPs), NTC (nineteen complex), NTC-related proteins (NTR), and a number of associated enzymes and cofactors. Here, we report the three-dimensional structure of a *Schizosaccharomyces pombe* spliceosome at 3.6-angstrom resolution, revealed by means of single-particle cryogenic electron microscopy. This spliceosome contains U2 and U5 snRNPs, NTC, NTR, U6 small nuclear RNA, and an RNA intron lariat. The atomic model includes 10,574 amino acids from 37 proteins and four RNA molecules, with a combined molecular mass of approximately 1.3 megadaltons. Spp42 (Prp8 in *Saccharomyces cerevisiae*), the key protein component of the U5 snRNP, forms a central scaffold and anchors the catalytic center. Both the morphology and the placement of protein components appear to have evolved to facilitate the dynamic process of pre-mRNA splicing. Our near-atomic-resolution structure of a central spliceosome provides a molecular framework for mechanistic understanding of pre-mRNA splicing.

Precursor messenger RNA (pre-mRNA) splicing, which involves the removal of the noncoding introns and the ligation of neighboring exons, is a defining feature for all eukaryotes (1). Aberrant splicing contributes to numerous debilitating diseases (2). The yeast spliceosome is a multi-megadalton ribonucleoprotein (RNP) complex and consists of U1, U2, U4, U5, and U6 small nuclear RNPs (snRNPs), NTC (nineteen complex; known as the Prp19-CDC5L complex in mammals), NTC-related proteins (NTR), and numerous other enzymes and cofactors (3, 4). The spliceosome exhibits exceptional compositional dynamics and conformational flexibility, which is consistent with its function of splicing introns with diverse sequences (5–7).

At the beginning of a splicing reaction cycle, the 5'-splice site (5'SS) of an intron is recognized by U1 snRNP, and the branch point sequence (BPS) and 3'SS are bound by U2 snRNP, forming the spliceosomal A complex. Recruitment of the U4/U6.U5 tri-snRNP leads to assembly of the pre-catalytic B complex. RNP rearrangement causes displacement of U1 and U4 snRNPs and recruitment of NTC and NTR, generating the activated B complex (B^{act}). The B^{act} spliceosome is converted into the catalytically competent B* complex,

which catalyzes the first step of trans-esterification. The end product is the C complex, also known as the catalytic step I spliceosome, which contains the cleaved 5'-exon and an intron lariat-3'-exon intermediate (hereafter, intron lariat). The second step of trans-esterification results in the ligation of two exons, forming the post-catalytic P complex. Subsequently, the ligated exon is released, but the intron lariat remains associated with the intron-lariat spliceosomal (ILS) complex. Last, the intron lariat is dissociated, and the snRNPs, NTC, and NTR are recycled. The RNP rearrangements are driven by eight evolutionarily conserved DExD/H-type, RNA-dependent adenosine triphosphatases (ATPases)/helicases (8, 9).

Biochemical studies during the past few decades have provided valuable insights into the mechanism of pre-mRNA splicing by the spliceosome. Analysis of thioester substitutions in pre-mRNA identified the spliceosome as a metalloenzyme (10–12), with the intramolecular stem loop of U6 small nuclear RNA (snRNA) coordinating the catalytic magnesium ions (13, 14). The spliceosomal active site is centered around loop I of U5 snRNA, the U2/U6 duplex, and a catalytic cavity in Prp8 (15). Structural investigations of the spliceosomal components—exemplified by the crystal structures of Prp8 (16), Brr2 (17, 18), and the Sm and like-Sm (Lsm) rings (19–23)—have also improved our understanding of pre-mRNA splicing.

Elucidating the molecular mechanisms of the spliceosome and the splicing reaction requires detailed structural information on the intact spliceosome at different stages of its action. The large size of the spliceosome and its extraordi-

nary conformational and compositional diversity have made this task a challenge for structural biologists (24). These features contrast with the other central RNP complex, the ribosome, which contains two relatively stable subunits that are more amenable to structural investigation (25–27). RNA accounts for more than 50% of the total molecular mass in the bacterial or mammalian ribosome, but less than 10% in the spliceosome. The constant flux of protein factors during each cycle of the splicing reaction has prevented crystallization of the spliceosome.

In the past decade, electron microscopy (EM) has been used to visualize various spliceosomal complexes, yielding a series of structures at resolutions ranging from 20 to 50 Å (28–44). Most recently, the cryogenic EM (cryo-EM) structure of the spliceosomal U4/U6.U5 tri-snRNP from *Saccharomyces cerevisiae* was determined at 5.9 Å resolution (45). Here, we report the cryo-EM structure of a yeast spliceosome at 3.6 Å resolution. We describe the overall structure of the spliceosome and its protein components in this Research Article and discuss the structural insights into snRNA recognition and pre-mRNA catalysis in (46).

Spliceosome isolation and characterization

We sought to purify the spliceosomal C complex because of its central role of bridging the two steps of a splicing reaction and catalyzing the second step. We modified a published purification protocol for the C complex (36, 47) and obtained approximately 500 µg of spliceosome from 4 liters of *Schizosaccharomyces pombe* culture (fig. S1A and supplementary materials, materials and methods). The purified yeast spliceosome exhibited excellent solution behavior, as judged with gel filtration analysis (fig. S1B). The spliceosome contains three major RNA species—whose lengths are consistent with those of U2, U5, and U6 snRNAs from *S. pombe*—and a diverse array of RNA molecules with varying lengths, which probably includes the intron lariat (fig. S1C). In addition, the spliceosome contains a large number of protein components (fig. S1D).

To examine the identity of the spliceosome and to facilitate future structure assignment, we analyzed the sample using mass spectrometry (MS). This approach detected ~80 spliceosomal proteins, with ~50 of these in high abundance (figs. S1E and S2). Most of these abundant spliceosomal proteins were components of U2 snRNP, U5 snRNP, NTC, and NTR (fig. S1E). These proteins are shared among the spliceosomal B^{act}, B*, C, P, and ILS complexes. To differentiate among these complexes, we performed reverse transcription polymerase chain reactions (RT-PCR) on the *cut6* gene using the purified spliceosome (fig. S3). The intron lariat was present in the spliceosome, suggesting completion of the first-step reaction. However, the unspliced *cut6* gene and the ligated exon were also detected, suggesting the presence of the B^{act} and/or B* and P complexes, respectively (fig. S3). We conclude that the purified spliceosome contains a mixture of the different complexes. Last, after

¹Ministry of Education Key Laboratory of Protein Science, Tsinghua-Peking Joint Center for Life Sciences, Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China. ²National Center for Protein Science Shanghai, Institute of Biochemistry and Cell Biology, Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China.
*These authors contributed equally to this work. †Corresponding author. E-mail: shi-lab@tsinghua.edu.cn

chemical crosslinking, we uncovered by means of MS analysis 78 intermolecular interactions among the spliceosomal proteins (fig. S4), which proved important for structure assignment.

Structure determination

We first generated a low-resolution reconstruction of the yeast spliceosome by using negative staining (fig. S5). Then, we imaged the sample under cryogenic conditions with a K2 direct electron detector, mounted on a Titan Krios microscope operating at 300 kV. A total of 2246 micrographs were collected (Fig. 1A and table S1), and we subjected 224,450 particles, picked semi-automatically, to particle sorting and reference-free two-dimensional (2D) classification (Fig. 1B and figs. S6 and S7A). We performed 3D classification for 133,901 particles. The vast majority (112,795) of these particles produced a density map at an overall resolution of 3.9 Å, which was further improved by means of particle polishing to 3.6 Å on the basis of the gold-standard Fourier Shell Correlation (FSC) criteria (Fig. 1C). The actual resolution within the spliceosome ranged from 2.9 to 3.6 Å in the core region to 7 to 8 Å in the periphery (Figs. 1D, 2, and 3A). The quality and resolution of EM density for different regions of the spliceosome were improved by applying individual local masks (figs. S7B, S8, and S9).

Throughout the spliceosome, most secondary structural elements were clearly visible, and a large proportion of the amino acid side chains were well defined (figs. S10 to S17).

The spliceosome has an extended and asymmetric morphology, with the longest dimension exceeding 300 Å (Fig. 2). The bulk of the well-defined density can be attributed to a triangular Central Body (Figs. 1D and 2), which is connected to a Head group and two Arms (named I and II). The Head and Arm II in our structure appear to correspond to the “head-like” and “ridge” domains in the 29 Å cryo-EM structure of the *S. pombe* U5.U2/U6 complex (36). An elongated tubular EM density in Arm II, connected to a donut-shaped density, is visible at 20 Å resolution (Fig. 2A); at higher resolutions, this density can be assigned as a tetrameric assembly of Cwf8/Prp19, Cdc5 C-terminal fragment, and Cwf7. At 10 and 5 Å resolutions (Fig. 2, B and C), these structural features become more prominent, allowing identification of most secondary structural elements.

Identification of protein and RNA components

We combined de novo atomic model building and homologous structure modeling to generate an atomic model for the entire *S. pombe*

spliceosome (Fig. 3 and table S2). On the basis of the EM density, we identified the heptameric Sm ring and the WD40-repeat proteins Cwf1/Prp5, Cwf8/Prp19, and Cwf17 (fig. S18). The large proteins Spp42 (Prp8 in *S. cerevisiae*), Cwf10 (Snu114 in *S. cerevisiae*), and Cwf11, each with an available homologous structure (16, 48, 49), were docked into the EM density. Model building of Spp42 allowed assignment of U5 snRNA, which is mostly bound by the N-terminal 800 amino acids of Spp42. These procedures were followed by identification of the tetrameric assembly of Cwf8/Prp19, which resides in Arm II of the spliceosome; the superhelical proteins Cwf3/Syf1 and Cwf4/Syf3, which connect the Central Body to the Head and Arm I; and several other proteins, including Cdc5, Cwf2/Prp3, Cwf14, Prp45, Cwf19, Cwf5/Ecm2, and Prp17 (fig. S18). The EM density for Arm I is weak. A local mask refinement after two rounds of 3D classification generated an improved map and enabled subsequent identification of Leal1, Msl1, a portion of U2 snRNA, and the Sm ring for U2 snRNA (fig. S18). In addition to EM density, assignment of the U2 and U6 snRNAs was facilitated by the location of known interacting proteins, predicted secondary structures, and published base-pairing specifics (table S2). After identification of U2 and U6 snRNAs, the RNA lariat was located, and the BPS and 5'SS were tentatively assigned. Last, on the basis of EM density and MS identification of crosslinked proteins (fig. S4), we assigned Cwf7, Cwf15, and Cyp1.

Overall structure

The final refined atomic model of the yeast spliceosome contains 10,574 amino acids from 37 proteins, three snRNA molecules, and an intron lariat (Fig. 3B and tables S1 and S2), with a combined molecular mass of ~1.3 MD. Among the modeled amino acids, 9312 were assigned specific side chains, and the remaining 1262 residues were built into a poly(Ala) model. U2, U5, and U6 snRNAs contain a total of 405 nucleotides, of which 314 were tentatively assigned in our atomic model. We also modeled 18 nucleotides from the intron lariat. Some areas of very weak EM density, probably reflecting dynamic components of the spliceosome, remain unassigned. The protein components in the atomic model include all 10 core proteins of U5 snRNP (except Brr2), all nine core proteins of U2 snRNP, eight of the nine core proteins of NTC, and five of the eight core proteins of NTR. Nearly complete atomic models are now available for Spp42 (residues 47 to 2030), which, as a central component of the U5 snRNP, anchors the catalytic center of the spliceosome, and for Cwf10 (residues 68 to 971), which, as the only guanosine triphosphatase (GTPase) among the spliceosomal components, regulates the splicing reaction (50).

The U2 snRNP—comprising Leal1, Msl1, the heptameric Sm ring, and U2 snRNA—constitutes Arm I of the spliceosome. Arm I is linked to the Central Body through two associations: one with the superhelical proteins Cwf3/Syf1 and Cwf4/Syf3 and another through U2 snRNA, which

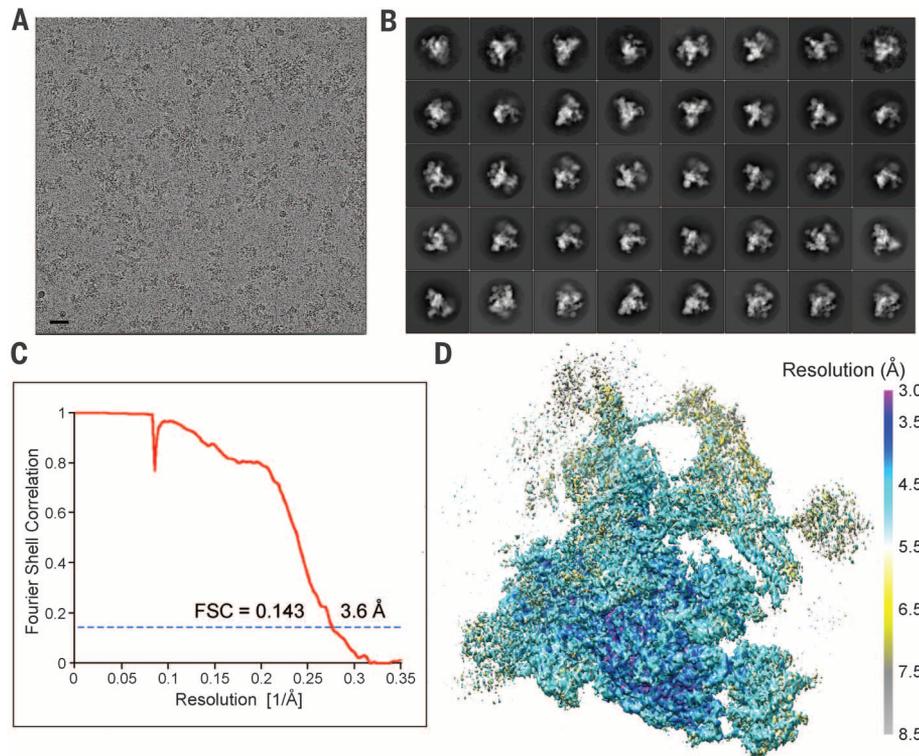


Fig. 1. Cryo-EM analysis of a yeast spliceosome. (A) A representative cryo-EM micrograph of the yeast spliceosomal complex. An entire micrograph is shown. Scale bar, 30 nm. (B) Representative 2D class averages of the yeast spliceosome. (C) The overall resolution is estimated to be 3.6 Å on the basis of the gold-standard FSC criteria of 0.143. (D) An overall view of the EM density for the yeast spliceosome. The resolution varies among different regions of the yeast spliceosome, as indicated by the different colors. The surface view of the spliceosome is shown here. The resolution reaches 2.9 to 3.2 Å in the center of the spliceosome.

enters the catalytic center (Fig. 3B). Arm II of the spliceosome consists of an elongated tetrameric assembly of Cwf8/Prp19 that is wrapped by Cwf7 and Cdc5 in their extended conformations (Fig. 3B). Cdc5, in turn, directly contacts Cwf3/Syf1, Cwf4/Syf3, the NTC component Cwf7, and the U5 snRNP protein Cwf17. The Head region comprises mainly Cwf11 and is connected to the triangular Central Body through the superhelical proteins Cwf3/Syf1 and Cwf4/Syf3. The base of the Central Body is mainly composed of U5 snRNP (Fig. 3B). All other proteins are located at the center of the Central Body, making direct or indirect contacts with the RNA molecules at the catalytic center.

The catalytic center, identified by the U2/U6 snRNA triplex, is located at the heart of the Central Body (Fig. 3B). The catalytic center is more than 100 Å away from the Head region, either Arm, and the corner of the Central Body (as defined by the 3' end of U5 snRNA). The distance between the tip of the Head region and the far corner of the Central Body measures ~335 Å, whereas the two Arms are separated by a distance of up to 320 Å (Fig. 3B). The Head and two Arms of the spliceosome are linked to the Central Body through limited contacts, which may engender conformational flexibility. Both the large size and the extended organization of the spliceosome are likely to be functionally important for proper

splicing of pre-mRNAs with varying lengths and sequences.

Structure of Spp42

At 270 kD, Spp42 is the largest and most conserved component of all spliceosomal proteins (51), displaying 63 and 73% sequence identity with its functional ortholog Prp8 in *S. cerevisiae* and humans, respectively. As a major component of U5 snRNP, Spp42 serves as a central scaffold for pre-mRNA splicing. Except for the N-terminal 46 amino acids and the C-terminal Jab1/MPN domain, most of the Spp42 sequences have a well-defined EM density (fig. S10, A to H, and fig. S11A). The structure of Spp42

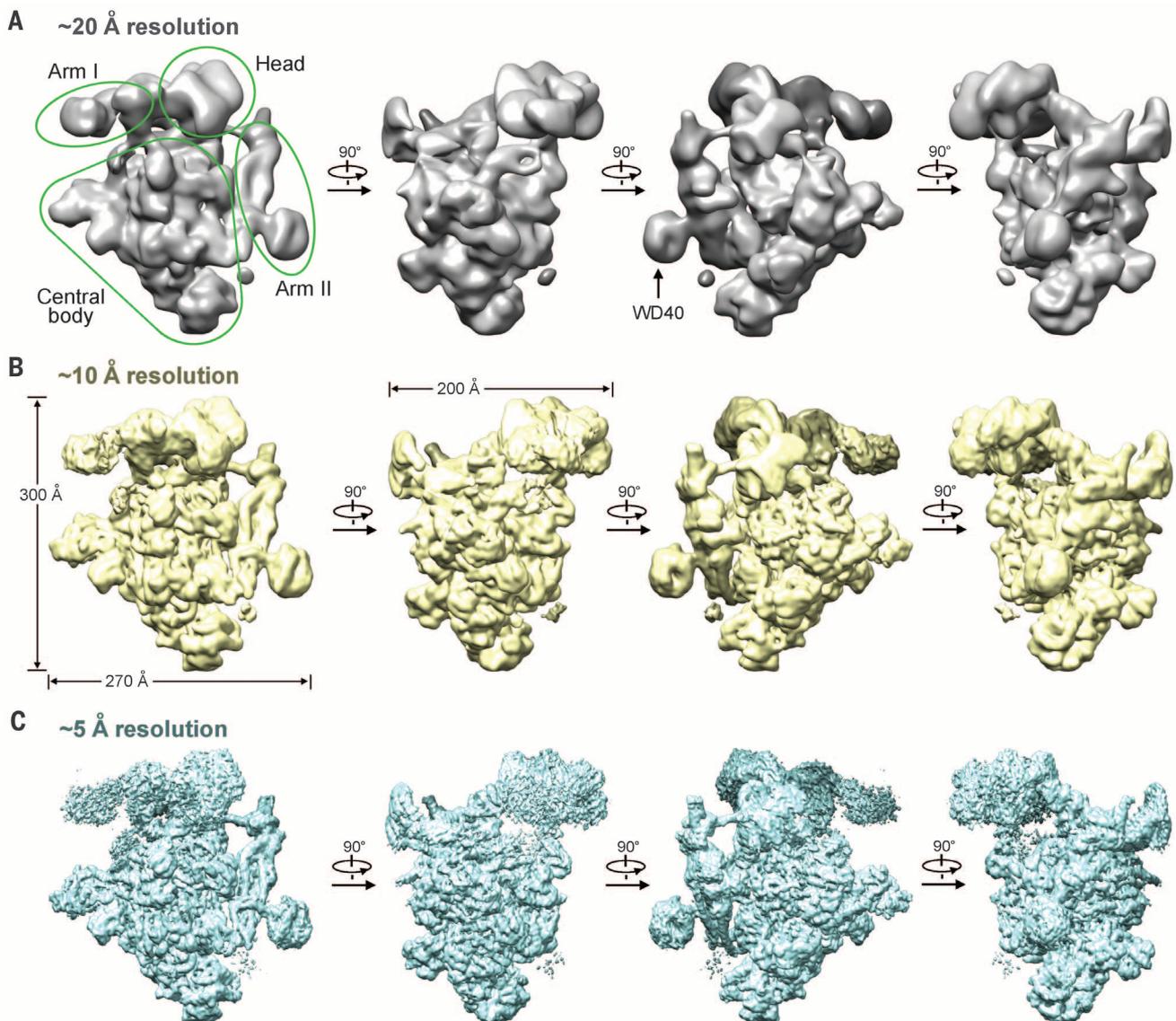


Fig. 2. Structural features of a yeast spliceosome at different resolution limits. (A) The EM density map of the yeast spliceosome at an averaged resolution of 20 Å. The spliceosome contains a triangular Central Body, a Head, and two Arms (I and II). The four perpendicular views illustrate the overall structural features. The same four views are shown in (B) and (C) for comparison. (B) The EM density map of the yeast spliceosome at an averaged

resolution of 10 Å. The improved resolution of the structural features allows docking of homologous structures and assignment of many protein components. (C) The EM density map of the yeast spliceosome at an averaged resolution of 5 Å. Most secondary structural elements of the protein components can be assigned at this resolution. Connectivity between the adjacent secondary structural elements is mostly clear.

includes the following domains: N-terminal (residues 47 to 825), reverse transcriptase (RT) Palm/Finger (residues 826 to 1210), Thumb/X (residues 1211 to 1327), Linker (residues 1328 to 1602), endonuclease-like (residues 1603 to 1783), and ribonuclease (RNase) H-like (residues 1784 to 2030) (Fig. 4A).

Spp42 looks like a thick triangular plate, with the three sides measuring 175, 150, and 125 Å and a thickness of ~80 Å (Fig. 4A). The N-terminal domain of Spp42, the structure of which has not previously been described, binds the GTPase Cwf10 and recognizes the bulk of U5 snRNA (52, 53). In the structure, the N-terminal domain adopts an extended conformation, with three protruding

structural elements reaching out to contact Cwf10, Cwf17, and the RT Palm/Finger domain of Spp42 (Fig. 4B). Spp42 can be superimposed onto Prp8 (16) with a root mean square deviation (RMSD) of 1.46 Å over 859 aligned C α atoms in the RT Palm/Finger, Thumb/X, Linker, and endonuclease-like domains (Fig. 4, C and D). These four contiguous domains constitute a conserved structural scaffold between Spp42 and Prp8.

The reported Prp8 structure contains a C-terminal Jab1/MPN domain (Fig. 4C) (16). Placement of the Jab1/MPN domain depends on the U5 snRNP assembly factor Aar2, which contributes a pairing β strand between the RNase H-like and Jab1/MPN domains (16). Superposition

of Spp42 and Prp8 revealed marked differences between the RNase H-like domains (Fig. 4D). Compared with Prp8, the RNase H-like domain in Spp42 undergoes a rotation of ~25° away from the RT Palm/Finger domain, resulting in a translation of up to 28 Å for its protruding β hairpin (Fig. 4E). The space vacated by the RNase H-like domain is in turn occupied by Cwf19, which interacts closely with both the RNase H-like and the RT Palm/Finger domains (Fig. 4A). The Jab1/MPN domain, responsible for binding the ATPase/helicase Brr2 (17, 18), is flexible and disordered in our structure. Consequently, despite its presence in the spliceosome (fig. S1E), Brr2 has no visible EM density, probably reflecting its highly dynamic

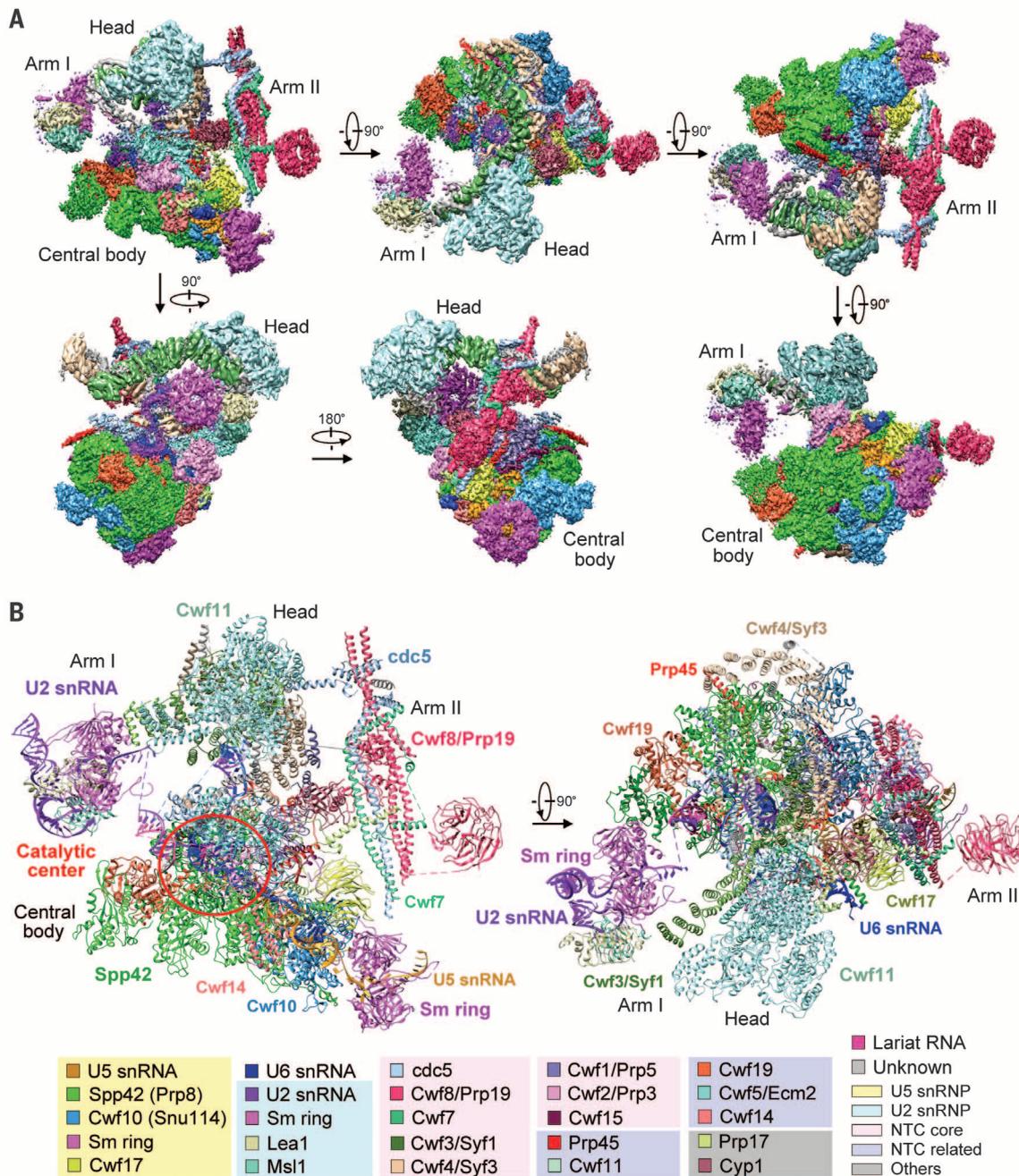


Fig. 3. Structure of a yeast spliceosome at 3.6 Å resolution.

(A) The EM density map of the yeast spliceosome at an overall resolution of 3.6 Å. Four perpendicular views around the x axis and two additional views around the y axis show the color-coded protein and RNA components. (B) A representative illustration of the yeast spliceosome from two perpendicular views. The protein and RNA components are color-coded. This structure includes 37 proteins, three snRNAs, and one RNA lariat, with a combined molecular weight of ~1.3 MD. Among the modeled 10,574 amino acids, 9312 have been assigned side chains. All 332 RNA nucleotides were tentatively assigned.

state. This structural arrangement would place Brr2 in close proximity to the catalytic center and allow ample conformational freedom for Brr2 to apply its ATP-dependent RNP remodeling activity in both steps of the splicing reaction.

Catalytic cavity of the spliceosome

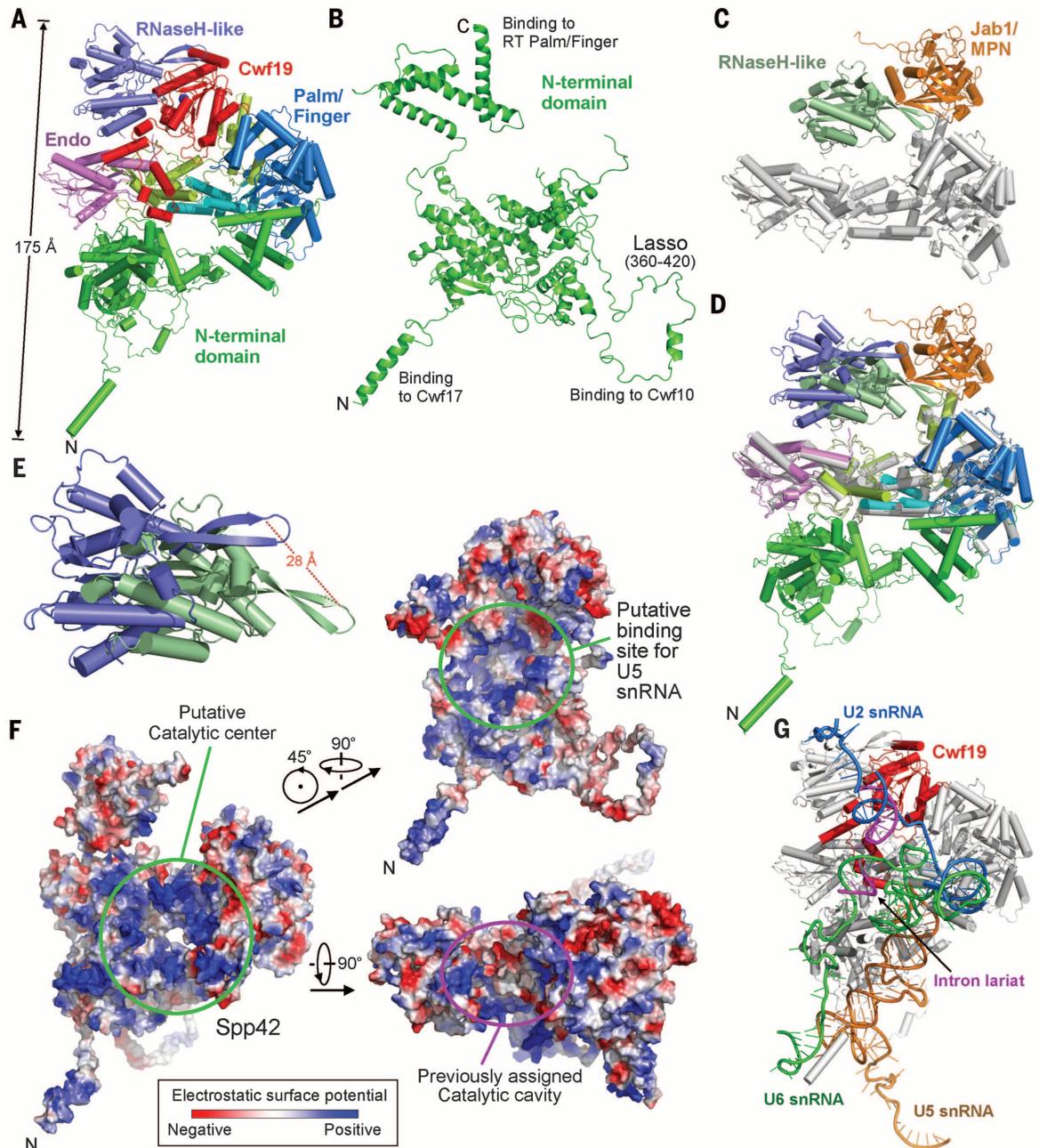
Spp42 plays a key role in pre-mRNA splicing. Splicing defects as a result of splice-site mutation

can be suppressed by rescuing mutations in Prp8 (54–56). Based on examination of these mutation-targeted residues, the presence of a catalytic cavity on Prp8, formed mainly by the Linker and RNase H-like domains, has been proposed (16). Analysis of the electrostatic surface potential of Spp42 revealed a striking cavity that is highly enriched by positively charged amino acids (Fig. 4F) but differs from the proposed catalytic cavity

(16). Instead, this cavity is formed between the N-terminal domain and the Thumb/X-Linker region of Spp42 and hence can only be recognized in the full-length Spp42 protein. The RNA triplex of U2 and U6 snRNAs and the intron lariat are located in this cavity (Fig. 4G), making close contact with the positively charged amino acids. The positively charged residues in this catalytic cavity are invariant among *S. pombe*, *S. cerevisiae*, and

Fig. 4. Structure of the central protein component Spp42 (Prp8 in *S. cerevisiae*).

(A) The overall structure of Spp42. The structure contains the following domains: N-terminal (green, residues 47 to 825), RT Palm/Finger (blue, residues 826 to 1210), Thumb/X (cyan, residues 1211 to 1327), Linker (pale green, residues 1328 to 1602), endonuclease-like (pink, residues 1603 to 1783), and RNase H-like (purple, residues 1784 to 2030). The Jab1/MPN domain (residues 2044 to 2363) is disordered, and its approximate position is occupied by Cwf19 (red). (B) The structure of the N-terminal domain of Spp42. The N-terminal helix is responsible for binding to Cwf17, and the lasso on the right corner interacts with Cwf10. (C) The structure of Prp8 (residues 885 to 2413) from *S. cerevisiae* (16). The RNase H-like and Jab1/MPN domains are colored light green and gold, respectively. (D) Structural comparison between Spp42 and Prp8. The structures from the Palm/Finger domain to the endonuclease-like domain are very similar. However, the RNase H-like domain exhibits a major conformational shift. (E) A close-up view of the RNase H-like domains of Spp42 and Prp8. The RNase H-like domain of Spp42 can be aligned to its counterpart in Prp8 by a translation of ~20 Å and a clockwise rotation of 25° around an axis perpendicular to the page. (F) Analysis of the electrostatic surface potential of Spp42 reveals a putative catalytic center. A central hole at the interface between the N-terminal domain



and the Thumb/X and Linker domains is enriched by positively charged amino acids and probably represents the catalytic center (left). In contrast, the previously identified catalytic cavity lacks such positive potential (bottom right). Another region enriched by positive electrostatic potential (top right) represents the binding site for U5 snRNA. (G) U2 snRNA, U6 snRNA, and the intron lariat are bound at the catalytic center of Spp42.

and the Thumb/X and Linker domains is enriched by positively charged amino acids and probably represents the catalytic center (left). In contrast, the previously identified catalytic cavity lacks such positive potential (bottom right). Another region enriched by positive electrostatic potential (top right) represents the binding site for U5 snRNA. (G) U2 snRNA, U6 snRNA, and the intron lariat are bound at the catalytic center of Spp42.

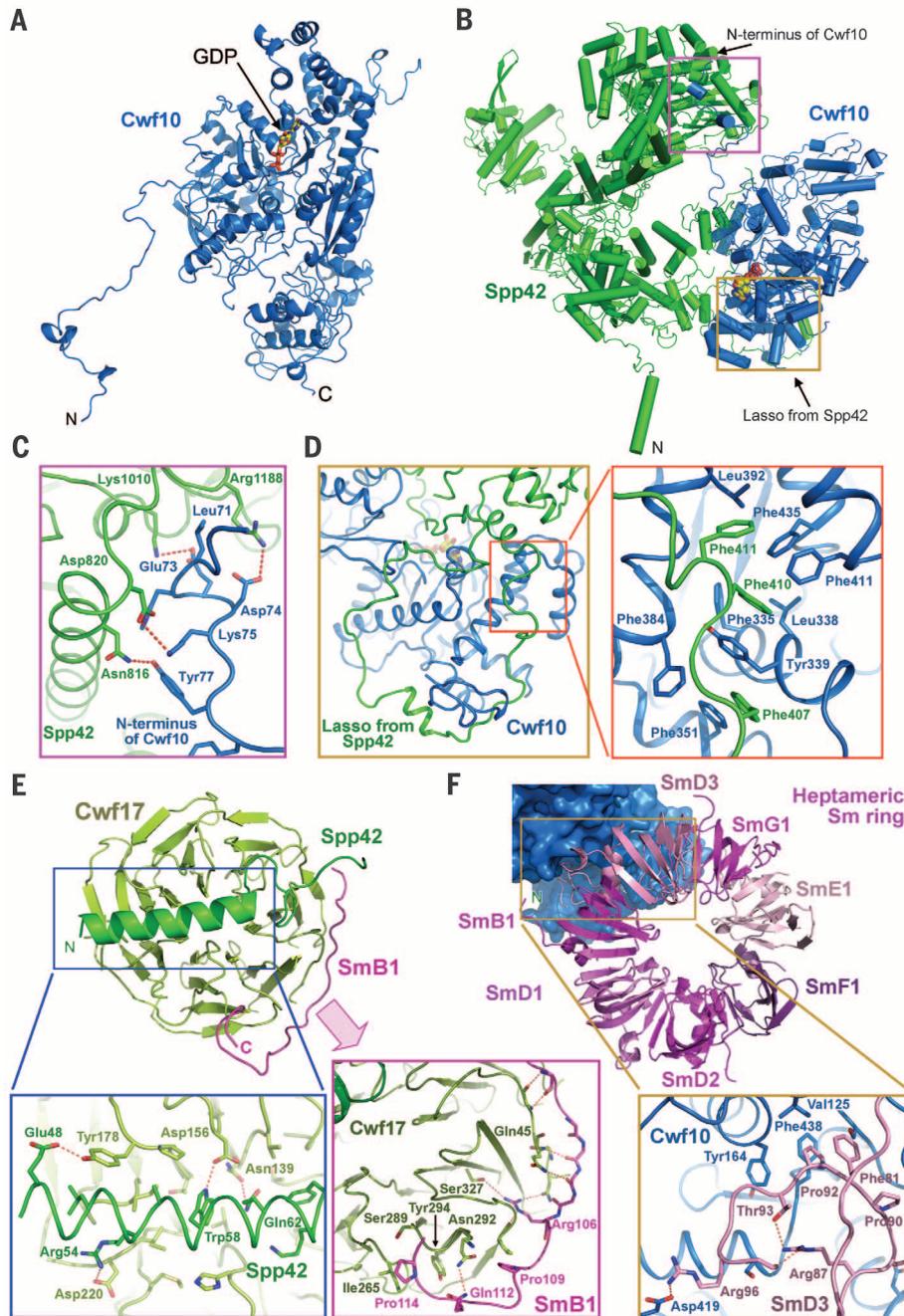


Fig. 5. Protein components in U5 snRNP. (A) Structure of the EF-2-family GTPase Cwf10. The bound GDP molecule is shown in stick representation. The extended N terminus binds Spp42. (B) An overall view of the interactions between Cwf10 and Spp42. The N-terminal sequences of Cwf10 recognize the Palm/Finger domain of Spp42, whereas the N-terminal domain of Spp42 forms a lasso over the back side of the nucleotide-binding domain of Cwf10. (C) A close-up view of the interface between the N-terminal sequences of Cwf10 and a shallow surface groove on the Palm/Finger domain of Spp42. (D) A close-up view of the Spp42 lasso over the N-terminal domain of Cwf10. In particular, three phenylalanine residues from Spp42 make a number of van der Waals contacts to hydrophobic amino acids on the surface of Cwf10. (E) Cwf17 and its interactions with Spp42 and SmB1. The N-terminal α -helix from Spp42 binds the bottom face of the Cwf17 WD40 propeller, whereas the C-terminal sequences of SmB1 interact with the outer surface of the propeller. Detailed interactions are shown in the two insets. (F) Structure of the heptameric Sm ring. The Sm ring directly interacts with the N-terminal domain of Cwf10 (blue in the background). Detailed interactions are shown in the inset. The Sm ring recognizes a stretch of U-rich RNA sequences at the 3' end of U5 snRNA.

humans (fig. S19A), allowing modeling of a similar catalytic cavity in Prp8 (fig. S19B).

Other protein components in U5 snRNP

A major component of the U5 snRNP, Cwf10 (Snul14 in *S. cerevisiae* and U5-116K in humans), is a translation elongation factor 2 (EF-2)-like GTPase that regulates the RNP remodeling activity of Brr2 (57). The excellent EM density for Cwf10 allows atomic modeling of nearly the entire length of the protein and of a bound guanosine diphosphate (GDP) molecule (Fig. 5A and fig. S11, B and C). Cwf10 closely interacts with Spp42 through two discrete, reciprocal interfaces (Fig. 5B). Residues from the extended N-terminal sequences of Cwf10 bind to a surface groove on Spp42, making specific interactions with amino acids in the groove (Fig. 5C). In addition, an extended loop in the N-terminal domain of Spp42 forms a lasso over the nucleotide-binding domain of Cwf10 (Fig. 5D). In particular, three phenylalanine residues from the lasso (Phe⁴⁰⁷, Phe⁴¹⁰, and Phe⁴¹¹) are nestled in a greasy surface pocket formed by hydrophobic amino acids in Cwf10 (Fig. 5D).

Cwf17/Spf38 (U5-40K in humans) is a WD40-repeat protein that is a core component of U5 snRNP. The functional ortholog of Cwf17/Spf38 in *S. cerevisiae* is yet to be identified. In the structure, an N-terminal α -helix of Spp42 associates with the bottom face of the WD40 repeats through specific hydrogen bonds and van der Waals contacts (Fig. 5E). The last core component of the U5 snRNP is the heptameric Sm ring, which orients the 3' end sequences of U5 snRNA and specifically interacts with the N-terminal domain of Cwf10 (Fig. 5F). The extended C-terminal sequences of SmB1 reach out to contact the outer surface of the Cwf17 β propeller, forming a number of hydrogen bonds (Fig. 5E).

Protein components in NTC and NTR

A shared feature of the spliceosomal proteins is their extended architecture. This is best illustrated by the tetrameric assembly of Cwf8/Prp19, which measures up to 175 Å in length (Fig. 6A). Two molecules of Cwf8/Prp19 form an intertwined dimer through their elongated coiled-coil region, with the two U-boxes distant from each other. Two such Cwf8/Prp19 dimers use their respective U-boxes to further associate with each other (Fig. 6A). This structural arrangement may allow considerable conformational flexibility and markedly increase surface areas for potential interactions with other spliceosomal factors.

The extended architecture is also exemplified by Cwf3/Syf1 and Cwf4/Syf3. Both of these are superhelical proteins, containing 19 and 16 half- α -tetratricopeptide (HAT) repeats (58), respectively, and exhibiting a twisted C-shaped morphology (Fig. 6, B and C). The maximal intramolecular distance for these proteins exceeds 150 Å. The middle portion of the convex side of Cwf3/Syf1 associates with the central region of the concave surface of Cwf4/Syf3 (Fig. 6D). HAT-repeat proteins exhibit conformational plasticity (58); the presence of Cwf3/Syf1 and Cwf4/Syf3 in the

spliceosome probably ensures adequate conformational freedom for the splicing reactions and exchange of cofactors. Cwf8/Prp19, Cwf3/Syf1, and Cwf4/Syf3 are all core components of NTC,

supporting an important regulatory role for this complex.

Prp45, a core component of NTR, has been predicted to be intrinsically disordered (59). Our

structure confirms the prediction: Prp45 only contains extended secondary structural elements that span a distance of over 150 Å (Fig. 6E). Prp45 appears to promote spliceosome assembly by directly interacting with, and thus linking together, at least nine distinct proteins, including Spp42, Cwf17, and Cwf4/Syf3 (Fig. 6F and fig. S20). Prp45 also directly interacts with U2 and U6 snRNAs at the catalytic center. Such an arrangement allows the conformational changes at the catalytic center to be propagated to any distant corner of the spliceosome through the connection among Prp45, Cwf4/Syf3, and Cwf3/Syf1. Thus, similar to Cwf8/Prp19, Cwf3/Syf1, and Cwf4/Syf3, the morphology and the placement of Prp45 seem to have evolved to facilitate the dynamic process of pre-mRNA splicing.

Our structure includes a number of other NTC and NTR components. These include five NTC core components: (i) Cdc5, which plays an important role in pre-mRNA splicing by stabilizing the second-step spliceosome (60–62); (ii) Cwf7, which modulates interactions of Prp19 with other associated cofactors (63, 64); (iii) Cwf2/Prp3, which helps link NTC to the catalytic center; (iv) Cwf1/Prp5, which contains a 7-bladed β propeller; and (v) Cwf15, which interacts with U5 snRNA, Prp5, and Spp42. In addition, we modeled four NTR components: (i) Cwf5/Ecm2, which is involved in base-pairing interactions of U2/U6 helix II (65); (ii) Cwf11, which is an armadillo domain containing RNA helicase (49); (iii) Cwf14, which facilitates both steps of the splicing reaction (66); and (iv) Cwf19, the human ortholog of which is thought to be involved in the development of recessive ataxia syndrome (67). We also identified Prp17 and the prolyl isomerase Cyp1 in the EM density map. Structural mapping of these proteins in the spliceosome allows critical assessment of their functions in pre-mRNA splicing (Fig. 6, G to K, and fig. S21).

Assembly of snRNPs and NTC

The atomic model of the yeast spliceosome mainly comprises four subcomplexes—U2 snRNP, U5 snRNP, NTC, and NTR—which are intertwined with one another to form an asymmetric assembly (Fig. 7A). The U2/U6 duplex and the intron lariat are partially exposed on the surface of the spliceosome and accessible to the modifying enzymes, such as the ATPase/helicase Brr2. The U5 snRNP, located at the bottom of the triangular Central Body, serves as the base of the entire spliceosome (Fig. 7B). At the heart of the Central Body is an interconnected RNA assembly involving U2 snRNA, U6 snRNA, and the intron lariat. Loop I of U5 snRNA is located close to the U2/U6 duplex (Fig. 7B). NTC and NTR together form a supporting network above the U5 snRNP and the centrally located RNA molecules.

The NTC is characterized by a large and extended architecture, 200 Å in height and width and 170 Å in thickness (Fig. 8A). The eight identified protein components in NTC are mostly interconnected, constituting a large scaffold. In contrast, the five protein components of NTR are mostly unconnected (Fig. 8B) and only come together through interactions with the NTC

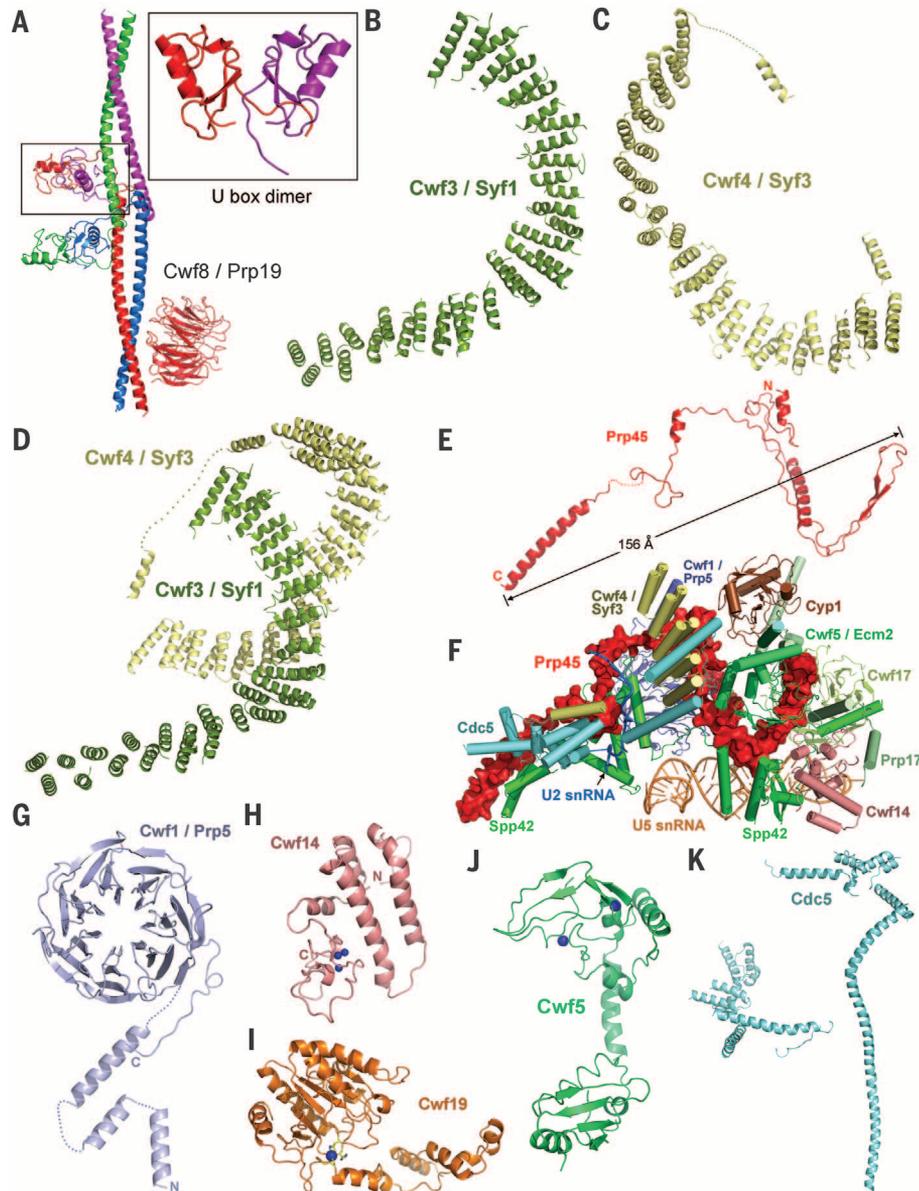


Fig. 6. Representative structures of individual protein components. (A) The tetrameric assembly of Cwf8/Prp19. The U-box dimer is highlighted in the inset. (B) The structure of the HAT-repeat superhelical protein Cwf3/Syf1. (C) The structure of the HAT-repeat superhelical protein Cwf4/Syf3. (D) Cwf3/Syf1 directly interacts with Cwf4/Syf3, both using HAT repeats in the middle portions of their respective structures. (E) The structure of Prp45. The 216 amino acids in Prp45 span 156 Å. (F) Prp45 interacts with at least nine protein components and two snRNAs. Prp45 is shown in surface view. (G) The structure of the WD40-repeat protein Cwf1/Prp5. (H) The structure of the NTR component Cwf14. Cwf14, which is thought to facilitate both steps of the splicing reaction, contains three zinc ions (blue spheres). (I) Structure of the functionally unknown NTR component Cwf19. (J) The NTR component Cwf5/Ecm2, which is thought to be involved in base-pairing interactions of U2/U6 helix II, exhibits an extended structure. (K) The structure of the NTC component Cdc5, which is thought to serve as a scaffold to recruit other protein factors. Of these proteins, no relevant structural information was previously available for Cwf19, Prp45, the N-terminal region of Cwf5, the N-terminal region of Spp42, or the C-terminal region of Cdc5. The other proteins have homologous structures, but this is the first time that they have been reported in *S. pombe*.

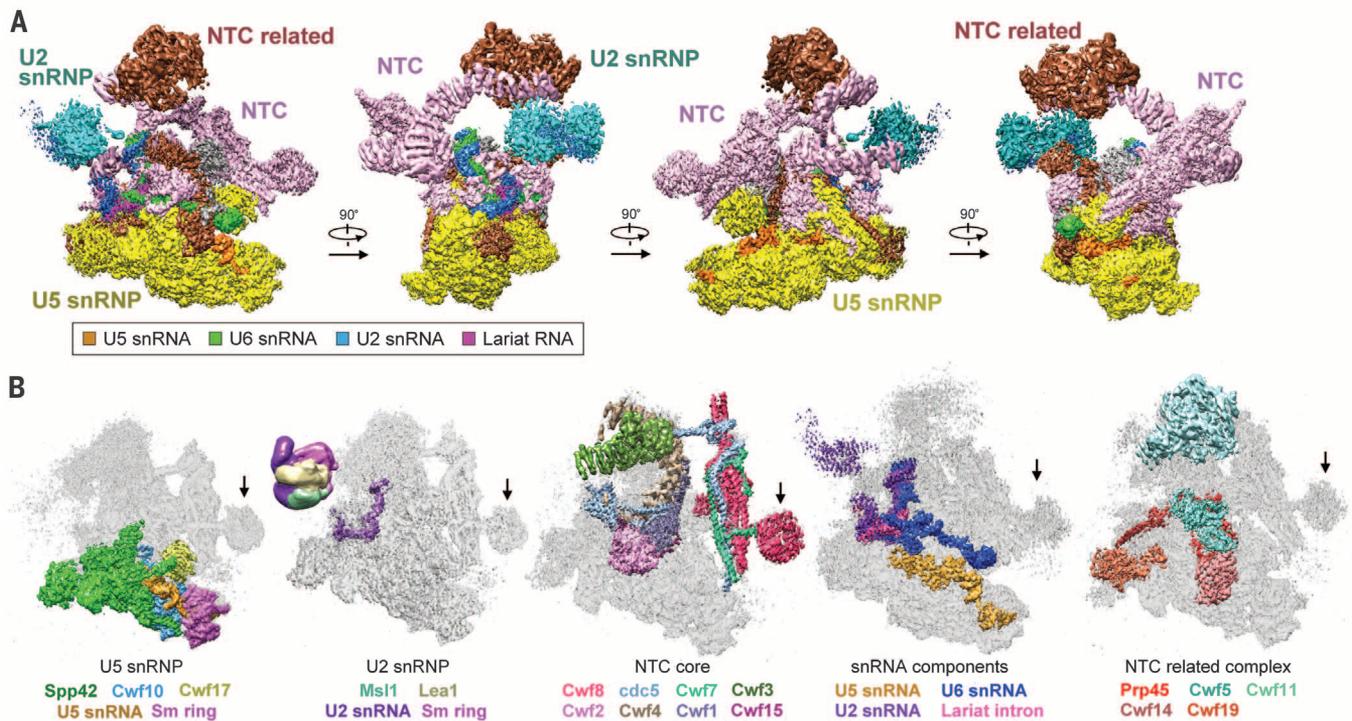


Fig. 7. The four multicomponent subcomplexes U5 snRNP, U2 snRNP, NTC, and NTR closely associate with one another to form the yeast spliceosome.

(A) U5 snRNP forms a scaffold onto which the other three subcomplexes dock around the RNA molecules and the catalytic center. The protein components within the same snRNP are colored identically: yellow for U5 snRNP, cyan for U2

snRNP, light pink for NTC, and brown for NTR. The four RNA molecules are individually colored. (B) Relative positions of the four multicomponent subcomplexes and the RNA molecules within the spliceosome. The protein components within each subcomplex are differentially color-coded. These five views of the yeast spliceosome have the same general orientation as that of the first view in (A).

components (Fig. 8C). Together, NTC and NTR define a large central space, part of which is occupied by the catalytic center (Fig. 8C).

Dynamic conformation of the spliceosome

Most of the spliceosomal particles display a well-defined conformation, allowing us to obtain a 3D structure at an overall resolution of 3.6 Å. The high-resolution regions are largely restricted to the Central Body, which appears to have a rigid conformation. The peripheral regions occupied by components of NTC and NTR exhibit considerable conformational flexibility, which poses a challenge to any effort aimed at improving the local resolution. In addition to local conformational variations, large-scale structural changes are also associated with the spliceosome. We were able to obtain two conformational states, with 37,622 and 49,079 particles, through an additional round of 3D classification. We low-pass filtered both maps to 10 Å and compared the conformational changes (fig. S22A). Superposition of these two structures revealed a major movement of the Head region toward Arm I (fig. S22B). As a consequence of this movement, much of the already flexible U2 snRNP becomes disordered.

Discussion

Structure determination by means of x-ray crystallography has been focused on the individual

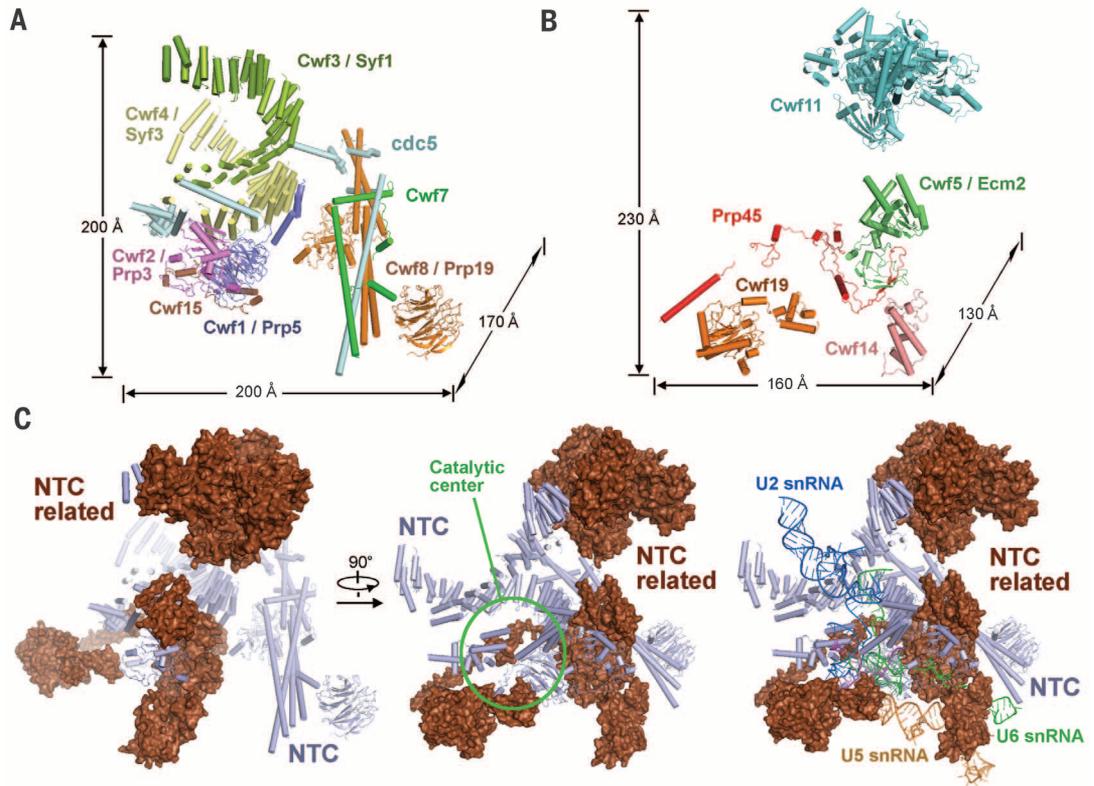
components or subcomplexes of the spliceosome, yielding valuable information on a number of important proteins—namely, the structures of the core components of U1 snRNP (21–23), U2 snRNP (68–71), and U4 snRNP (19); two subcomplexes of the U6 snRNP (20, 72); a large fragment of Prp8 (16); and Brr2 (17, 18). These structures provide individual pieces of the spliceosome jigsaw puzzle. In the past decade, EM-based studies have produced structural insights into the intact spliceosome at various stages of the splicing reaction (73). These studies were performed on the mammalian spliceosomal A complex (29), B complex (31, 38, 40), B^{*} complex (30), C complex (33, 35), and P complex (42), as well as on the yeast B complex (32), B^{act} complex (32), the U5/U2/U6 complex or C complex (32, 36), and the ILS complex (43). The highest resolution reported in these EM studies is 20 to 29 Å, which only allows description of general features of the spliceosome. After submission of this Research Article, the cryo-EM structure of the spliceosomal U4/U6/U5 tri-snRNP in *S. cerevisiae* was reported at 5.9 Å resolution (45), which allows domain identification and secondary-structure assignment for some components. The U4/U6/U5 tri-snRNP is an important complex for the assembly of the functional spliceosomes.

Here, we report the cryo-EM structure of an intact *S. pombe* spliceosome at a near-atomic resolution of 3.6 Å. The resolution exceeds 3.2 Å in

the core region of the spliceosome, which covers the bulk of the U5 snRNP and a number of other proteins from NTC and NTR. On the basis of the EM density, we have generated the first atomic model of an intact, functional spliceosome. We have provided an overall and preliminary analysis of the spliceosome structure. The enormous amount of information contained therein requires additional analysis. There are over 100 discrete protein-protein and protein-RNA interfaces among the 37 proteins and four RNA molecules of the spliceosome, involving a substantial amount of buried surface area. There are numerous other details that warrant attention; for example, Cwf5, Cwf14, and Cwf19 contain three different zinc-binding motifs (fig. S23). Determination of a near-atomic-resolution structure of an intact spliceosome represents a major milestone. Insights gained from structural analysis will greatly improve the mechanistic understanding of pre-mRNA splicing.

Although we intended to purify the spliceosomal C complex, the purified sample included a mixture of different spliceosomes, as judged by use of RT-PCR (fig. S3). Because Cdc5, which is recruited into the spliceosomal B^{act} complex, was used as the affinity tag for purification, the purified sample should be restricted to the B^{act}, B^{*}, C, P, and ILS complexes. This finding is corroborated by the preponderance of NTC and NTR components (which are also recruited into the B^{act}

Fig. 8. Structures of the NTC and NTR. (A) The overall structure of the NTC. The color-coded NTC components are mostly interconnected through direct interactions. (B) Overall structure of the NTR. In contrast to NTC, components of the NTR make no direct contacts with each other and are mostly unconnected. (C) NTC and NTR surround the catalytic center. The catalytic center is identified by three intertwined RNA molecules: U6 snRNA (green), U2 snRNA (blue), and the intron lariat (magenta). The U5 snRNA (orange) is located next to the intertwined RNAs.



complex) (figs. S2 and S4). The EM density is well defined for the intron lariat, suggesting that the structure probably reflects the C, P, or ILS complex, or a mixture of the three complexes. The EM density for 5'-exon is weak, suggesting that either the ILS complex was a major species in the final EM sample or 5'-exon was lost during purification. These complexes share a large set of stable spliceosomal components, allowing determination of the cryo-EM structure at 3.6 Å resolution. The different spliceosomal complexes in the EM sample, major or minor, will probably be resolved upon collection and analysis of a larger set of micrographs.

Pre-mRNA splicing and protein translation, two central processes in eukaryotes, are each carried out by supramolecular protein-RNA machineries: spliceosome and ribosome, respectively. The 80S eukaryotic ribosome consists of two well-organized subcomplexes: the 60S large subunit and the 40S small subunit. In both subunits, the core protein components and the RNA molecules, in a mass ratio of approximately 1:1, form relatively stable RNP complexes. A flux of protein factors regulates the initiation, elongation, and termination of protein synthesis. In contrast to the ribosome, the spliceosome has a mass ratio of at least 10:1 in favor of the protein components and exhibits extreme dynamism in both composition and conformation. In the ribosome, the protein components are mainly located on the exterior, away from the catalytic center for peptide bond formation. In the spliceosome, the protein components surround and support the RNA-based catalytic center. The ribosome exhibits a generally isometric shape,

whereas the spliceosome has a highly asymmetric morphology, with numerous surface cavities of varying sizes and cut-through spaces. Apparently, through evolution, two highly divergent strategies have been adopted for assembling RNPs to achieve complex functions in gene expression.

REFERENCES AND NOTES

- C. B. Burge, T. Tuschl, P. A. Sharp, in *The RNA World, Second Edition: The Nature of Modern RNA Suggests a Prebiotic RNA World*, R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999), pp. 525–560.
- T. A. Cooper, L. Wan, G. Dreyfuss, *Cell* **136**, 777–793 (2009).
- D. A. Wassarman, J. A. Steitz, *Science* **257**, 1918–1925 (1992).
- C. L. Will, R. Lührmann, *Cold Spring Harb. Perspect. Biol.* **3**, a003707 (2011).
- M. C. Wahl, C. L. Will, R. Lührmann, *Cell* **136**, 701–718 (2009).
- W. Chen, M. J. Moore, *Curr. Opin. Struct. Biol.* **24**, 141–149 (2014).
- A. Hegele et al., *Mol. Cell* **45**, 567–580 (2012).
- O. Cordin, D. Hahn, J. D. Beggs, *Curr. Opin. Cell Biol.* **24**, 431–438 (2012).
- J. P. Staley, C. Guthrie, *Cell* **92**, 315–326 (1998).
- T. A. Steitz, J. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6498–6502 (1993).
- E. J. Sontheimer, S. Sun, J. A. Piccirilli, *Nature* **388**, 801–805 (1997).
- P. M. Gordon, E. J. Sontheimer, J. A. Piccirilli, *RNA* **6**, 199–205 (2000).
- S.-L. Yean, G. Wuenschell, J. Termini, R.-J. Lin, *Nature* **408**, 881–884 (2000).
- S. M. Fica et al., *Nature* **503**, 229–234 (2013).
- T. W. Nilsen, in *RNA Structure and Function*, R. W. Simons, M. Grunberg-Manago, Eds. (Cold Spring Harbor Monograph vol. 35, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1998), pp. 279–307.
- W. P. Galej, C. Oubridge, A. J. Newman, K. Nagai, *Nature* **493**, 638–643 (2013).
- S. Mozaffari-Jovin et al., *Science* **341**, 80–84 (2013).
- T. H. Nguyen et al., *Structure* **21**, 910–919 (2013).
- A. K. Leung, K. Nagai, J. Li, *Nature* **473**, 536–539 (2011).
- L. Zhou et al., *Nature* **506**, 116–120 (2014).
- G. Weber, S. Trowitzsch, B. Kastner, R. Lührmann, M. C. Wahl, *EMBO J.* **29**, 4172–4184 (2010).
- D. A. Pomeranz Krummel, C. Oubridge, A. K. Leung, J. Li, K. Nagai, *Nature* **458**, 475–480 (2009).
- Y. Kondo, C. Oubridge, A. M. van Roon, K. Nagai, *eLife* **4**, e04986 (2015).
- W. P. Galej, T. H. Nguyen, A. J. Newman, K. Nagai, *Curr. Opin. Struct. Biol.* **25**, 57–66 (2014).
- N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz, *Science* **289**, 905–920 (2000).
- F. Schluenzen et al., *Cell* **102**, 615–623 (2000).
- B. T. Wimberly et al., *Nature* **407**, 327–339 (2000).
- M. Azubel, S. G. Wolf, J. Sperling, R. Sperling, *Mol. Cell* **15**, 833–839 (2004).
- N. Behzadnia et al., *EMBO J.* **26**, 1737–1748 (2007).
- S. Bessonov et al., *RNA* **16**, 2384–2403 (2010).
- D. Boehringer et al., *Nat. Struct. Mol. Biol.* **11**, 463–468 (2004).
- P. Fabrizio et al., *Mol. Cell* **36**, 593–608 (2009).
- M. M. Golas et al., *Mol. Cell* **40**, 927–938 (2010).
- M. Grote et al., *Mol. Cell. Biol.* **30**, 2105–2119 (2010).
- M. S. Jurica, D. Sousa, M. J. Moore, N. Grigorieff, *Nat. Struct. Mol. Biol.* **11**, 265–269 (2004).
- M. D. Ohi, L. Ren, J. S. Wall, K. L. Gould, T. Walz, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 3195–3200 (2007).
- B. Sander et al., *Mol. Cell* **24**, 267–278 (2006).
- E. Wolf et al., *EMBO J.* **28**, 2283–2292 (2009).
- E. M. Makarov et al., *Science* **298**, 2205–2208 (2002).
- J. Deckert et al., *Mol. Cell. Biol.* **26**, 5528–5543 (2006).
- M. S. Jurica, L. J. Licklider, S. R. Gygi, N. Grigorieff, M. J. Moore, *RNA* **8**, 426–439 (2002).

42. J. O. Ilagan, R. J. Chalkley, A. L. Burlingame, M. S. Jurica, *RNA* **19**, 400–412 (2013).
43. W. Chen *et al.*, *RNA* **20**, 308–320 (2014).
44. Z. Zhou, J. Sim, J. Griffith, R. Reed, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12203–12207 (2002).
45. T. H. Nguyen *et al.*, *Nature* **523**, 47–52 (2015).
46. J. Hang, R. Wan, C. Yan, Y. Shi, *Science* (2015).
47. M. D. Ohi *et al.*, *Mol. Cell. Biol.* **22**, 2011–2024 (2002).
48. R. M. Voorhees, I. S. Fernández, S. H. Scheres, R. S. Hegde, *Cell* **157**, 1632–1643 (2014).
49. I. De *et al.*, *Nat. Struct. Mol. Biol.* **22**, 138–144 (2015).
50. P. Fabrizio, B. Lagerbauer, J. Lauber, W. S. Lane, R. Lührmann, *EMBO J.* **16**, 4092–4106 (1997).
51. R. J. Grainger, J. D. Beggs, *RNA* **11**, 533–557 (2005).
52. S. Liu, R. Rauhut, H. P. Vornlocher, R. Lührmann, *RNA* **12**, 1418–1430 (2006).
53. I. Dix, C. S. Russell, R. T. O’Keefe, A. J. Newman, J. D. Beggs, *RNA* **4**, 1239–1250 (1998).
54. J. G. Urmen, C. Guthrie, *Genetics* **143**, 723–739 (1996).
55. C. A. Collins, C. Guthrie, *Genes Dev.* **13**, 1970–1982 (1999).
56. L. Liu, C. C. Query, M. M. Konarska, *Nat. Struct. Mol. Biol.* **14**, 519–526 (2007).
57. E. C. Small, S. R. Leggett, A. A. Winans, J. P. Staley, *Mol. Cell* **23**, 389–399 (2006).
58. P. J. Preker, W. Keller, *Trends Biochem. Sci.* **23**, 15–16 (1998).
59. I. Korneta, J. M. Bujnicki, *PLOS Comput. Biol.* **8**, e1002641 (2012).
60. W. H. McDonald, R. Ohi, N. Smelkova, D. Frenkewey, K. L. Gould, *Mol. Cell. Biol.* **19**, 5352–5362 (1999).
61. C. G. Burns, R. Ohi, A. R. Krainer, K. L. Gould, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13789–13794 (1999).
62. C. C. Query, M. M. Konarska, *RNA* **18**, 1001–1013 (2012).
63. H. R. Chen *et al.*, *Mol. Cell. Biol.* **18**, 2196–2204 (1998).
64. H. R. Chen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5406–5411 (1999).
65. D. Xu, J. D. Friesen, *Mol. Cell. Biol.* **21**, 1011–1023 (2001).
66. D. Saha, P. Khandelia, R. T. O’Keefe, V. Vijayraghavan, *J. Biol. Chem.* **287**, 5390–5399 (2012).
67. R. Burns *et al.*, *Neurology* **83**, 2175–2182 (2014).
68. S. R. Price, P. R. Evans, K. Nagai, *Nature* **394**, 645–650 (1998).
69. E. A. Sickmier *et al.*, *Mol. Cell* **23**, 49–59 (2006).
70. P. C. Lin, R. M. Xu, *EMBO J.* **31**, 1579–1590 (2012).
71. J. L. Jenkins, A. A. Agrawal, A. Gupta, M. R. Green, C. L. Kielkopf, *Nucleic Acids Res.* **41**, 3859–3873 (2013).
72. E. J. Montemayor *et al.*, *Nat. Struct. Mol. Biol.* **21**, 544–551 (2014).
73. R. Lührmann, H. Stark, *Curr. Opin. Struct. Biol.* **19**, 96–102 (2009).

ACKNOWLEDGMENTS

We thank the Tsinghua University Branch of the China National Center for Protein Sciences (Beijing) for providing the facility support. The computation was completed on the “Explorer 100” cluster system of the Tsinghua National Laboratory for Information Science and Technology. This work was supported by funds from China’s Ministry of Science and Technology (grant 2014ZX09507003006) and the National Natural Science Foundation of China (grants 31130002 and 31321062). The atomic coordinates have been deposited in the Protein Data Bank with the accession code 3JB9. The EM maps have been deposited in the Electron Microscopy Data Bank with the accession codes EMD-6413 for the overall map and EMD-6414/6415/6416/6417/6418/6419/6420/6421 for the eight local maps. The authors declare no competing financial interests.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6253/1182/suppl/DC1
Materials and Methods
Figs. S1 to S23
Tables S1 and S2
References (74–89)

10 June 2015; accepted 10 August 2015
Published online 20 August 2015
10.1126/science.aac7629

STRUCTURAL BIOLOGY

Structural basis of pre-mRNA splicing

Jing Hang,* Ruixue Wan,* Chuangye Yan,* Yigong Shi†

Splicing of precursor messenger RNA is performed by the spliceosome. In the cryogenic electron microscopy structure of the yeast spliceosome, U5 small nuclear ribonucleoprotein acts as a central scaffold onto which U6 and U2 small nuclear RNAs (snRNAs) are intertwined to form a catalytic center next to Loop I of U5 snRNA. Magnesium ions are coordinated by conserved nucleotides in U6 snRNA. The intron lariat is held in place through base-pairing interactions with both U2 and U6 snRNAs, leaving the variable-length middle portion on the solvent-accessible surface of the catalytic center. The protein components of the spliceosome anchor both 5′ and 3′ ends of the U2 and U6 snRNAs away from the active site, direct the RNA sequences, and allow sufficient flexibility between the ends and the catalytic center. Thus, the spliceosome is in essence a protein-directed ribozyme, with the protein components essential for the delivery of critical RNA molecules into close proximity of one another at the right time for the splicing reaction.

In eukaryotic cells, the coding exons in a freshly transcribed precursor messenger RNA (pre-mRNA) are interdispersed by non-coding introns that must be removed before protein translation. Removal of the introns is carried out by the spliceosome, a dynamic ribonucleoprotein (RNP) machine that comprises more than 100 protein components and five small nuclear RNAs (snRNAs) (*1*). Pre-mRNA splicing occurs through two steps, which are both $S_{\alpha}2$ -type transesterification reactions (*2, 3*).

In the first step of pre-mRNA splicing, the 2′-OH group of a conserved RNA adenine nucleotide in the branch point sequence (BPS) of an intron initiates a nucleophilic attack on the phosphorous atom of the guanine nucleotide at the 5′ end of the intron, resulting in the release of 5′-exon and formation of an intron lariat-3′-exon intermediate (hereafter, intron lariat) (Fig. 1A). In the second step, the 3′-OH group of the RNA nucleotide at the 3′ end of the 5′-exon unleashes a second nucleophilic attack on the phosphorous atom of the RNA guanine nucleotide at the 5′ end of the 3′-exon, leading to joining of two exons and release of the intron lariat (Fig. 1A) (*3*). Although the nature of the two-step reaction has been clearly defined for decades, how the spliceosome facilitates such reaction remains largely enigmatic. How are the reacting pieces placed into close proximity of one another in the correct temporal order? Given the varying lengths of the exons and introns, how does the spliceosome accommodate pre-mRNA and hold the 5′-exon for both steps of the reaction?

The spliceosome is a metalloenzyme, and a number of conserved RNA nucleotides directly coordinate at least two magnesium ions (Mg^{2+}) that collectively catalyze the two-step reaction (*2, 4–7*). The reaction mechanism of pre-mRNA

splicing is thought to closely resemble that of the group IIA or IIB self-splicing intron, each involving formation of an intron lariat, and differ from that of the group IIC intron, which splices by hydrolysis through a linear intron (*8, 9*). An in vitro reconstitution of U2 and U6 snRNAs revealed RNA splicing-like activity in the absence of the protein components, which strongly supports the ribozyme hypothesis (*10–13*). But the protein components of the spliceosome are obviously indispensable for pre-mRNA splicing to proceed because, for example, defective splicing due to a mutated RNA sequence can be rescued by mutations in Prp8 from *Saccharomyces cerevisiae* (Spp42 from *Schizosaccharomyces pombe*) (*14–18*). What are the functions of the protein components of the spliceosome during the two-step reaction? This Research Article addresses these interrelated questions through structural analysis of the yeast spliceosome from *S. pombe*, with a focus on the RNA components. The overall structure of the spliceosome and the features of the protein components are reported in (*19*).

Organization of the RNA components

The yeast spliceosome contains four distinct RNA molecules: U2 snRNA, U5 snRNA, U6 snRNA, and an intron lariat, which have been unambiguously located in the electron microscopy (EM) density map (Fig. 1B and figs. S1 and S2). U2, U5, and U6 snRNAs contain 186, 120, and 99 nucleotides (nt), respectively. In the structure, a nine-nucleotide fragment at the 3′ end of U6 snRNA (nt 91 to 99) and the middle and 3′ end portion of U2 snRNA (nt 44 to 92, 146 to 152, and 178 to 186) are flexible, exhibit poor EM density, and remain to be assigned. In addition, six nucleotides at the 5′ end and nine nucleotides at the 3′ end of U5 snRNA are disordered. All other 314 nucleotides from the snRNAs have been specifically assigned, representing ~80% of the total snRNA sequences. In addition, 18 nucleotides have been tentatively modeled for the intron lariat on the basis of conserved sequences for yeast (*1*).

Among the four RNA molecules, U5 snRNA is mostly buried in the structure (Fig. 1B). A large

Ministry of Education Key Laboratory of Protein Science, Tsinghua-Peking Joint Center for Life Sciences, Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China.

*These authors contributed equally to this work. †Corresponding author. E-mail: shi-lab@tsinghua.edu.cn

Structure of a yeast spliceosome at 3.6-angstrom resolution

Chuangye Yan, Jing Hang, Ruixue Wan, Min Huang, Catherine C. L. Wong and Yigong Shi

Science **349** (6253), 1182-1191.

DOI: 10.1126/science.aac7629originally published online August 20, 2015

Structure and function of the spliceosome

When RNA is transcribed from DNA in the eukaryotic cell nucleus, the initial transcript includes noncoding introns that must be spliced out. This splicing is done by a complex macromolecular machine, the spliceosome, which comprises five small nuclear RNAs and more than 100 associated proteins. Now, two papers reveal insights into the structure and function of the yeast spliceosome. Yan *et al.* describe a high-resolution structure determined by electron microscopy of a spliceosome complex comprising four RNAs and 37 proteins. Hang *et al.* focus on the catalytic site and show how protein components anchor the transcribed RNA and allow sufficient flexibility to deliver RNA components involved in catalyzing the splicing reaction.

Science, this issue pp. 1182 and 1191

ARTICLE TOOLS

<http://science.sciencemag.org/content/349/6253/1182>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2015/08/19/science.aac7629.DC1>

RELATED CONTENT

<http://science.sciencemag.org/content/sci/349/6253/1191.full>

REFERENCES

This article cites 87 articles, 33 of which you can access for free
<http://science.sciencemag.org/content/349/6253/1182#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)