ANNUAL REVIEWS

Annual Review of Biophysics

Cryo-EM Studies of Pre-mRNA Splicing: From Sample Preparation to Model Visualization

Max E. Wilkinson,* Pei-Chun Lin,* Clemens Plaschka,* and Kiyoshi Nagai

MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, United Kingdom; email: mwilkin@mrc-lmb.cam.ac.uk, pclin@mrc-lmb.cam.ac.uk, cplasch@mrc-lmb.cam.ac.uk, kn@mrc-lmb.cam.ac.uk

Annu. Rev. Biophys. 2018. 47:8.1–8.25

The *Annual Review of Biophysics* is online at biophys.annualreviews.org

https://doi.org/10.1146/annurev-biophys-070317-033410

*These authors contributed equally to this work.

Keywords

structural biology, electron microscopy, RNA–protein complexes, spliceosome

Abstract

The removal of noncoding introns from pre-messenger RNA (pre-mRNA) is an essential step in eukaryotic gene expression and is catalyzed by a dynamic multi-megadalton ribonucleoprotein complex called the spliceosome. The spliceosome assembles on pre-mRNA substrates by the stepwise addition of small nuclear ribonucleoprotein particles and numerous protein factors. Extensive remodeling is required to form the RNA-based active site and to mediate the pre-mRNA branching and ligation reactions. In the past two years, cryo-electron microscopy (cryo-EM) structures of spliceosomes captured in different assembly and catalytic states have greatly advanced our understanding of its mechanism. This was made possible by long-standing efforts in the purification of spliceosome intermediates as well as recent developments in cryo-EM imaging and computational methodology. The resulting high-resolution densities allow for de novo model building in core regions of the complexes. In peripheral and less ordered regions, the combination of cross-linking, bioinformatics, biochemical, and genetic data are essential for accurate modeling. Here, we summarize these achievements and highlight the critical steps in obtaining near-atomic resolution structures of the spliceosome.

Contents

1. INTRODUCTION

In eukaryotes, the majority of protein-coding genes are interrupted by noncoding sequences called introns. To generate a continuous protein-coding sequence, these introns must be removed from the newly synthesized transcript, the pre-messenger RNA (pre-mRNA), and the coding sequences, called exons, must be ligated (9, 25). This reaction proceeds in two transesterification reactions via a branched lariat-intron intermediate and is catalyzed by the spliceosome, a large and dynamic ribonucleoprotein enzyme (139). For step I of splicing (branching), the 2′-hydroxyl group of the branch point adenosine at the branch point sequence (BPS) performs a nucleophilic attack on the phosphorus atom of the 5′ splice site (5′SS), producing a free 5′ exon and the lariat intron–3′ exon intermediate (139) (**Figure 1*a***). For step II of splicing (exon ligation), the 3′-hydroxyl group of the last nucleotide of the 5′ exon attacks the phosphorus atom at the 3′SS. This produces the ligated exons (mRNA) and the lariat intron (139) (**Figure 1*a***). The three reactive groups required for this reaction are located within conserved sequences in the intron: at the intron 5′ end (the 5′SS), in the internal BPS, and at the intron 3′ end (3′SS) (21, 118–120) (**Figure 1*b***). The 5′SS is located in the consensus sequence N/GUAUGU in *Saccharomyces cerevisiae* (91), where the slash indicates the 5′ splice junction, N represents any nucleotide, and invariant nucleotides are underlined. In human cells, the 5′SS is more degenerate and has the consensus sequences N/GURAGN or AG/GUNNN (R represents any purine) (21, 65, 119). The BPS is located 18 to 40 nucleotides upstream of the 3′SS and comprises the consensus sequence UACUA**A**C in *S. cerevisiae* (80) and YNYUR**A**Y in humans (21, 106, 141) (the branch point adenosine is shown in bold; Y represents any pyrimidine). The 3′SS follows immediately after an invariant AG dinucleotide in the canonical yeast and human introns (21). Most human 3′SSs are preceded by a stretch of 15–20 pyrimidines—the so-called polypyrimidine tract [poly(Y) or Y_N] (21, 104).

The spliceosome assembles anew on each intron from five small nuclear ribonucleoprotein particles (snRNPs) and non-snRNP protein factors. Each snRNP contains numerous proteins

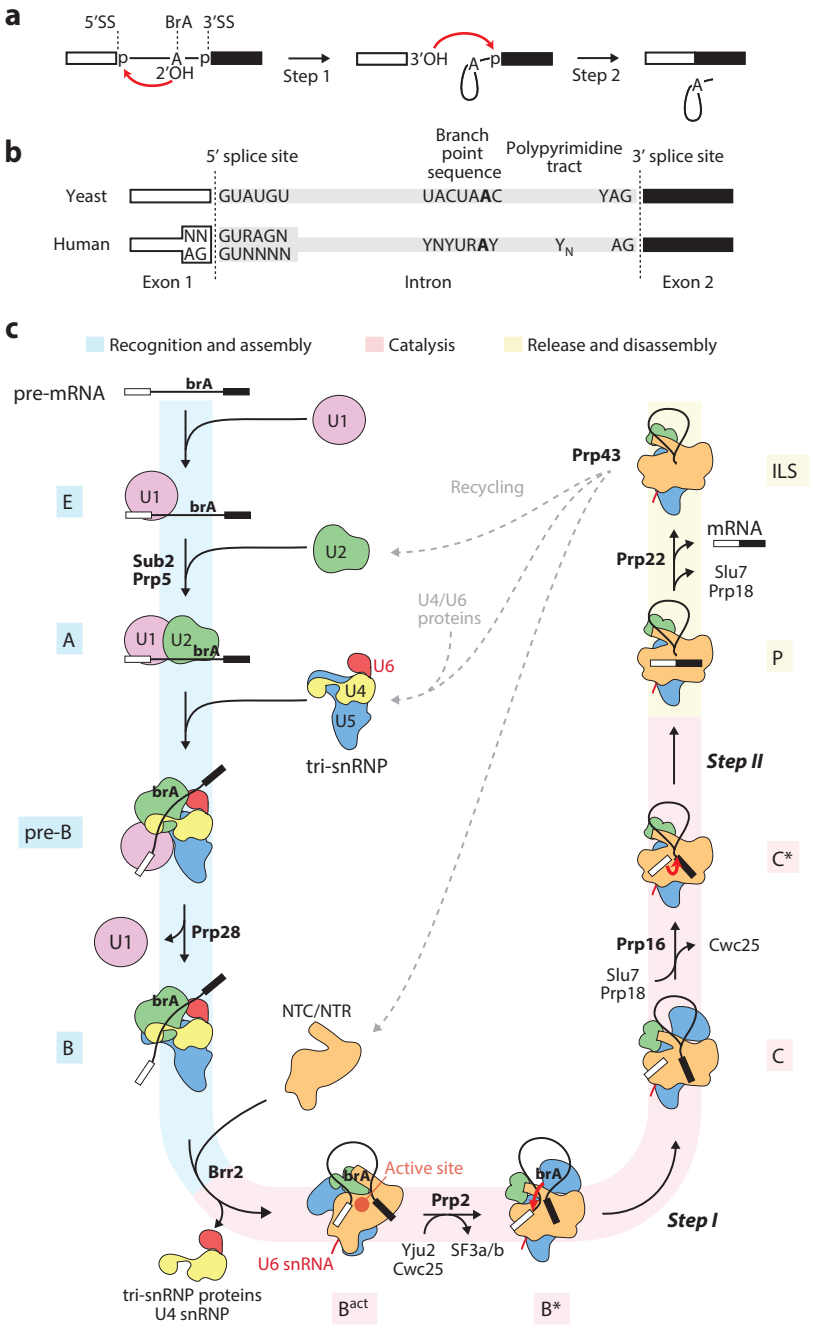


Figure 1

Splicing cycle and two-step mechanism of the splicing reaction. (*a*) Introns are excised from pre-mRNA by two consecutive transesterification reactions, step I (branching) and step II (exon ligation). (*b*) Conserved sequences in *Saccharomyces cerevisiae* and human introns. (*c*) Schematic of the spliceosome intermediates formed during the splicing reaction. The drawn spliceosome contours are based on the available cryo-electron microscopy structures (2, 12, 37, 43, 82, 95, 144, 145, 150), where E-, A-, pre-B-, and B*-complex structures are yet to be determined. Abbreviations: 3′SS, 3′ splice site; 5′SS, 5′ splice site; brA, branch point adenosine; ILS, intron-lariat spliceosome; N, any nucleotide; pre-mRNA, pre-messenger RNA; NTC, nineteen complex; NTR, nineteen-related complex; R, any purine nucleotide; snRNP, small nuclear ribonucleoprotein particles; Y, any pyrimidine nucleotide; Y_N, polypyrimidine tract.

and one of five U-rich small nuclear RNAs (U1, U2, U4, U5, or U6 snRNA). Additionally, eight conserved RNA helicases (Sub2, Prp5, Prp28, Brr2, Prp2, Prp16, Prp22, and Prp43) drive transitions between specific spliceosome intermediates by remodeling of the RNA and protein networks (27). Through these transitions, the spliceosome adopts at least ten distinct states [E, A, pre-B, B, B^{act}, B*, C, C*, P, and the intron-lariat spliceosome (ILS)]. Splicing occurs in three

phases: (1) recognition and assembly, (2) catalysis, and (3) release and disassembly (**Figure 1*c***), with each detailed as follows:

1. First, the pre-mRNA 5′SS is recognized by the U1 snRNP, forming the E complex (64, 65). The U2 snRNP is then recruited and recognizes the BPS, generating the A complex (15, 59, 92), and this depends on the RNA helicases Sub2/UAP56 (41, 57) and Prp5 (68, 92, 143). While the 3′SS is not identified during spliceosome assembly in *S. cerevisiae*, the U2-auxiliary factor, U2AF35, binds the 3′SS directly in *Schizosaccharomyces pombe* and higher eukaryotes (142). The preassembled U4/U6-U5 tri-snRNP complex (124) then associates with the A complex, resulting in the pre-B complex (17). The 5′SS is then transferred from U1 to U6 snRNA upon the action of the helicase Prp28, yielding the precatalytic B complex (121, 125). In the B complex, the eventual catalytic U6 snRNA is extensively base-paired with U4 snRNA, maintaining U6 in an inactive form. During spliceosome activation, the helicase Brr2 unwinds the U4/U6 snRNA duplex, dissociating the U4/U6 snRNP proteins and U4 snRNA (63, 102). This allows U6 snRNA to fold an intramolecular stem loop and to make new base pairs with U2 snRNA, facilitating formation of the catalytic core of the spliceosome (20, 73, 81, 123, 126). In addition, the nineteen complex (NTC) and nineteen-related complex (NTR) (23, 86) join to stabilize the catalytic core, resulting in the activated B complex, B^{act} (47).
2. B^{act} is subsequently converted to B^{*} by the helicase Prp2, juxtaposing the 5′SS and branch point adenosine for the first step of the splicing reaction (136). The resulting C complex contains the lariat–3′ exon intermediate and the cleaved 5′ exon. The C complex is subsequently remodeled to C^{*} by the helicase Prp16 (115, 128) to allow docking of the 3′SS for step II of splicing.
3. Exon ligation produces the mRNA, which is then released by the helicase Prp22 from the P complex (26, 114). The remaining spliceosome complex retains only the intron lariat and is therefore called the ILS. The ILS is disassembled by Prp43 and Brr2 helicases (4, 127), leading to recycling of the snRNPs for subsequent splicing rounds and to the degradation of the excised intron lariat (42).

The dynamic nature of the spliceosome has presented an enormous challenge for structural biologists (139). Fully assembled spliceosomes can be purified from endogenous sources, but the low yields as well as its conformational and compositional heterogeneity have left it intractable to X-ray crystallography. While in vitro reconstitution of the whole spliceosome from recombinant material has so far not been achieved, isolated spliceosome components as well as protein–protein and RNA–protein subcomplexes, including the fully recombinant human U1 snRNP, have been previously studied at high resolution by X-ray crystallography and nuclear magnetic resonance (97, 149). However, the large molecular size of the spliceosome makes it an ideal target for electron microscopy (EM) (122). Since the snRNPs were first isolated from HeLa-cell nuclear extract in early 1980s, human spliceosomes and snRNPs have been characterized using negative-stain EM (16, 50). Intact spliceosomes and their subcomplexes were subsequently studied by cryo-EM, whereby complexes are maintained in a native hydrated state at cryogenic temperatures. These early cryo-EM studies yielded the first low-resolution reconstructions of the yeast tri-snRNP (109), yeast ILS (87), human supraspliceosome (5), human U1 snRNP (52), human U2 snRNP SF3b subcomplex (44), human U11/U12 di-snRNP (45), human B complex (16), and human C^{*} complex (referred to as C complex in the original paper) (50). Importantly, the methods developed for these studies laid the foundation for current high-resolution cryo-EM studies of the spliceosome.

Recent revolutionary advances in cryo-EM single-particle analysis have enabled structural biologists to determine the structures of macromolecules at near-atomic resolution (46, 62, 112).

Embracing these advances, we (37, 43, 82, 83, 95, 138) and the groups of Reinhard Lührmann/Holger Stark (2, 12, 13, 103), Yigong Shi (6, 132–134, 144–146, 150), and Rui Zhao/Hong Zhou (66, 70) have purified multiple spliceosome intermediates from yeast and human sources and have determined their three-dimensional (3D) structures at intermediate to high resolution. This body of work has revealed novel mechanistic details of pre-mRNA splicing. New and powerful computational methods developed by the Scheres group (110, 111) have played a major role in identifying homogenous particle sets from compositionally and structurally heterogeneous spliceosome mixtures. The resulting cryo-EM densities generally contain a well-ordered core with flexible regions in the periphery. Using so-called focused classification schemes and refinement together with partial signal subtraction has enabled structure determination in some of these peripheral regions. To obtain complete models of the spliceosome, we draw on data from protein cross-linking coupled to mass spectrometry, bioinformatics, and biochemical data to facilitate density interpretation in comparatively poorly resolved regions. In this review, we outline the key considerations for sample preparation, cryo-EM imaging, data processing, model building, and biological interpretation as applied to the spliceosome. We hope this is a useful resource for those working on the spliceosome and other macromolecular nucleic acid–protein complexes.

2. PURIFICATION OF SPLICEOSOME INTERMEDIATES

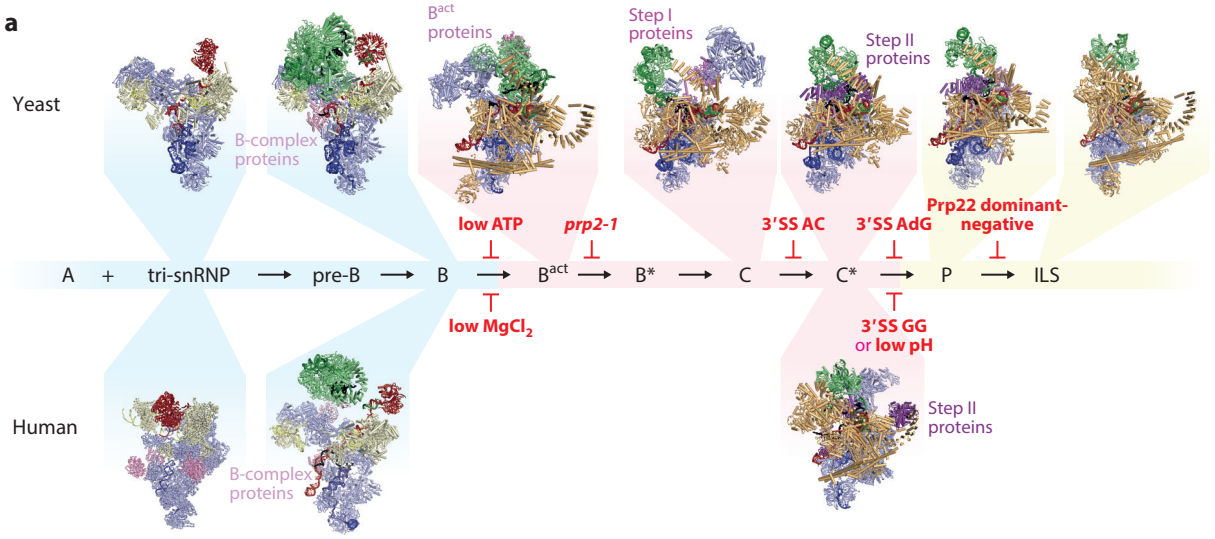
Spliceosomes can be assembled in vitro in yeast whole-cell or human nuclear-cell extract on synthetic pre-mRNA substrates (53, 69, 89) that contain multiple copies of the MS2 RNA aptamer. This aptamer can be bound by a fusion of an MS2 coat protein and maltose-binding protein for affinity purification (28, 33, 48). Unless otherwise indicated, all current purification schemes include this step. Various strategies are then used to enrich for a specific spliceosome intermediate. Here, we briefly describe the different strategies used to arrest spliceosomes at specific stages for structure determination by cryo-EM (**Figure 2**).

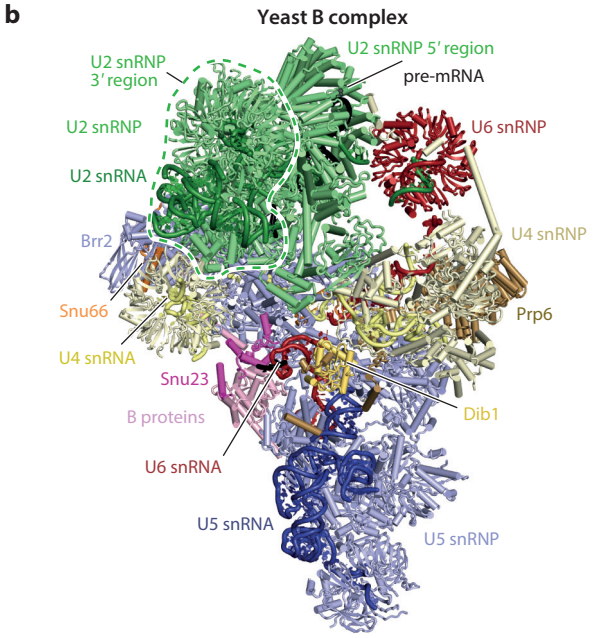
2.1. Yeast Spliceosomes

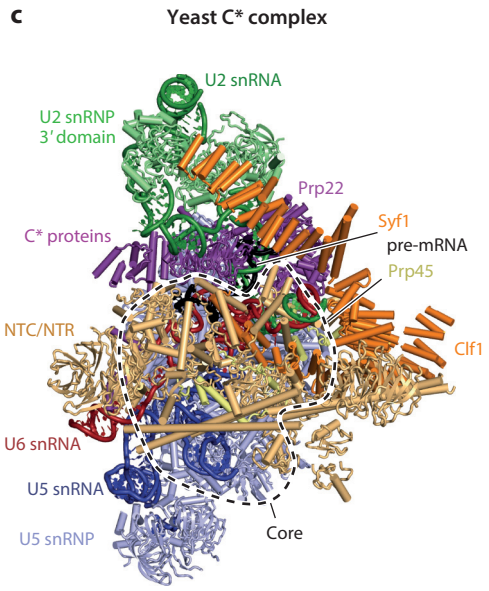
In yeast, the splicing reaction can be stalled at each of the known spliceosome states. For example, to stall the reaction prior to the first or second transesterification reaction, the nucleotides at the 5′SS, BPS, or 3′SS can be mutated (131) These stalled states can also be achieved by replacing the nucleotide at the splice site with its deoxy, 2′-O-methyl (78) or phosphorothioate derivative (32, 148). The advantage of the phosphorothioate substitution is that splicing can still occur in the presence of a thiophilic metal ion (38). Alternatively, the DExD/H-box RNA helicases that remodel the spliceosome at specific stages can stall the reaction by use of their ATPase-deficient helicase mutants, which act in a dominant-negative manner (6, 30, 70, 96, 113, 127, 138). For several of the helicases, temperature-sensitive mutations have been identified. In these cases, the yeast cells are grown at the permissive temperature and the spliceosome is then stalled in the in vitro splicing reaction at the restrictive temperature (54, 116). The spliceosome can also be stalled by changing buffer conditions or by limiting ATP concentration (33). Tagging of a protein found in a particular spliceosome state can be used to further enrich specific complexes.

To determine the B-complex structure, the concentration of ATP was reduced from the usual 2 mM down to 50 μM (33, 95), thereby blocking Brr2 helicase activity. Brr2 was then used for subsequent affinity purification (95). To purify the B^{act} complex, a strain with a temperature-sensitive mutant of the helicase Prp2 (*prp2-1*) was used to prepare splicing extract (55). At the restrictive temperature, the *prp2-1* helicase dissociates from the spliceosome, allowing the incorporation of an exogenously added dominant negative mutant of Prp2, yielding the B^{act} complex (136). With

the aim of purifying C*, a pre-mRNA substrate containing an AC substitution of AG at 3′SS was used and complexes were subsequently purified using a tag on Prp18 or Slu7 (43). Although this approach was designed to isolate the C* complex, it instead yielded predominantly the C complex. This is consistent with previous data showing that Slu7 is weakly bound already at the C-complex stage (88). The genuine C* complex could be arrested by using a pre-mRNA substrate with a modification of the guanosine immediately preceding the 3′SS from a 2′-OH to a 2′-H group. The C* complex was assembled on this substrate in splicing extract from a strain expressing a







(Caption appears on following page)

Figure 2 (*Figure appears on preceding page*)

Biochemical strategies to stall the spliceosome in various states for structural studies. (*a*) The yeast and human tri-snRNPs were purified directly from whole-cell and nuclear extract, respectively (2, 82). The yeast B complex was enriched by reducing the amount of ATP in the reaction (95), whereas its human counterpart was enriched by reducing the amount of $MgCl_2$ (12). The yeast B^{act} complex was stalled using the *prp2-1* temperature-sensitive mutant (103), the yeast C complex by using of a 3′SS AC mutation (43), the yeast C* complex by using a modification of the guanosine immediately preceding the 3′SS from a 2′-OH to a 2′-H group (AdG) (37). Yeast B^{act}, C, and C* structures were also determined by three-dimensional image classification of a large data set comprising a mixture of endogenous spliceosomes (132, 145, 146). The human C* complex was stalled using low pH during the splicing reaction (13) or using a GG mutation at the 3′SS (150). The yeast ILS was purified from Cef1-tagged *Schizosaccharomyces pombe* cells directly (144) or in silico from a mixture of Cef1- and Yju2-tagged *Saccharomyces cerevisiae* spliceosomes (133). The yeast P complex was stalled using dominant-negative mutant Prp22 protein, either added exogenously to yeast extract (138) or overexpressed to bias the population of endogenous spliceosomes toward this state (6, 70). Structures of the E, A, pre-B, and B* complexes remain to be determined (compare to **Figure 1*c***). Protein and RNA components mentioned in the text are indicated (*b*) for the yeast B complex and (*c*) for the yeast C* complex. Abbreviations: 3′SS, 3′ splice site; ILS, intron lariat spliceosome; pre-messenger RNA, pre-mRNA; snRNP, small nuclear ribonucleoprotein particle.

tagged C*-complex protein, Slu7, for subsequent affinity purification (37). To stall the P complex, the Shi (6) and Zhou/Zhao labs (70) overexpressed dominant-negative mutant Prp22 to bias the endogenous population of spliceosomes toward the P complex and then purified these by tags on Cef1 and/or Prp22. We used a similar strategy but added exogenous dominant-negative mutant Prp22 to splicing extract and specifically purified the P complex by making use of its unique RNase H protection pattern on the mRNA substrate to cleave the MS2 tag off non-P-complex spliceosomes (138). The *S. pombe* ILS complex is highly abundant in whole-cell extract and retains a mixture of all excised intron lariats in the cell. This complex was purified using solely a tag on the NTC protein Cef1 (24, 144).

In contrast to the biochemical strategies above, the Shi lab has also carried out computational purification of specific intermediates from a mixture of spliceosomes found in a large cryo-EM data set, as first applied to the ribosome where specific states were isolated from a mixture by classification of single-particle images (40). Yan and colleagues (132) collected a data set of 12,000 micrographs of all Cef1-containing spliceosomes purified directly from whole *S. cerevisae* cell extract. Cef1 is thought to be present in B^{act}, B*, C, C*, P, and ILS, and several rounds of computational image sorting (see Section 4.1) revealed the structures of the B^{act}, C, and C* spliceosome complexes (132, 145, 146). By combining these data with another 4,000 micrographs of Yju2-tagged spliceosomes, the *S. cerevisiae* ILS structure could also be determined. While very successful in this case, this approach relies on generous access to EM and computation resources.

2.2. Human Spliceosomes

HeLa nuclear extract has been the main source of human spliceosomes. In vitro splicing is carried out similarly to yeast splicing reactions, often using the synthetic AdML or its derivative MINX model pre-mRNA substrate (89, 152). The extract is prepared from HeLa or other human cells using a protocol initially developed by the Roeder lab for in vitro transcription (29, 53). Human splicing requires creatine phosphate in the splicing reaction and an m7G(5′)ppp(5′)G-cap on the pre-mRNA, which can be added during the in vitro transcription reaction. The purification of human spliceosomes has thus far not involved protein tags but rather the modification of buffer conditions and of the pre-mRNA substrate. To enrich for the human precatalytic B complex, the $MgCl_2$ concentration was reduced to 0.3 mM in the splicing reaction from the normal 2.5–3.5 mM and complexes were purified by MS2 affinity selection and a sucrose gradient (12). The human C* complex was stalled by two methods to determine its cryo-EM structure (13, 150). The Lührmann group (13) accumulated C* by lowering the pH of the in vitro splicing reaction from 7.9 to 6.4,

whereas the Shi lab stalled at the C* stage by using a pre-mRNA substrate containing a 3′SS AG to GG mutation (48). The stalled complexes were then purified by MS2 affinity selection, followed by a sucrose gradient (150). In the case of human C*, unwanted complexes could be further separated from C* by RNase H-mediated digestion of the intron (48, 150).

3. ELECTRON MICROSCOPY

Once a biochemically defined spliceosome preparation is obtained, the conditions for visualization in an electron microscope can be optimized.

3.1. Sample Preparation

Cryogenic freezing of the purified complex is required for high-resolution EM analysis, preserving the complexes in their native hydrated state, and reducing radiation damage upon imaging (58, 75). To achieve this, the purified sample is commonly applied to a support grid that was glow discharged, and the excess liquid is blotted away, after which the grid is quickly plunged into liquid ethane (**Figure 3*a***) (75). This produces a thin layer of vitreous ice containing the complexes as distinguishable particles. The metal grids commonly contain either a carbon or gold foil support (108) with a regular lattice of holes over the grid mesh. The particles are suspended within these holes, in free-standing ice, and are then imaged where the ice is thinnest (without damaging particles) and the background noise is low. Alternatively, a thin (20–60 Å) layer of carbon can be deposited on top of the holey carbon foil, which can help concentrate the sample on the support. Several parameters must therefore be optimized during the EM sample preparation.

1. Grid type: Spliceosome purifications generally yield low amounts of material (<30–60 μg), particularly when rare splicing intermediates are studied. These dilute samples can be concentrated on the amorphous carbon film by an extended incubation before freezing. In our experience, spliceosomes have a high affinity for the carbon support, whereas free-standing ice tends to exclude spliceosomes from the holes. We and others (Stark, Shi, and Moore) therefore exclusively use carbon-coated grids. However, the carbon film itself reduces image contrast (**Figure 3*c***), leading to the development of new supports such as graphene oxide (18, 90) and graphene (107), which may equally keep spliceosomes intact but have not been applied so far.
2. Ice thickness: Electron microscopy images of biological samples must be recorded in ice of the optimal thickness to maximize image contrast while maintaining complex integrity. While spliceosomes appear undisturbed in thick ice, this reduces image contrast, limiting the resolution of resulting reconstructions. Thin ice has lower background noise but can damage the spliceosomes by freeze-drying the complexes (**Figure 3*b***).
3. Particle orientation distribution: For one to obtain an isotropic 3D reconstruction of the sample of interest, the particles should adopt random orientations in the ice. A poor orientation distribution can limit the final resolution and may also cause distortions in the density map. Ice that is too thin can restrict the range of orientations owing to particles being squeezed into certain views by their interaction with the hydrophobic air–water interface. An amorphous carbon film can change orientation distributions, since the sample may interact preferentially with the carbon film. For the C complex, we found that particles in free ice show strong orientation bias, whereas particles on amorphous carbon had a more uniform orientation distribution (**Figure 3*c,d***). Whether free-standing ice or carbon support improves the orientations is highly sample dependent. To determine the isolated yeast tri-snRNP structure, the grids were additionally glow discharged in the presence of a

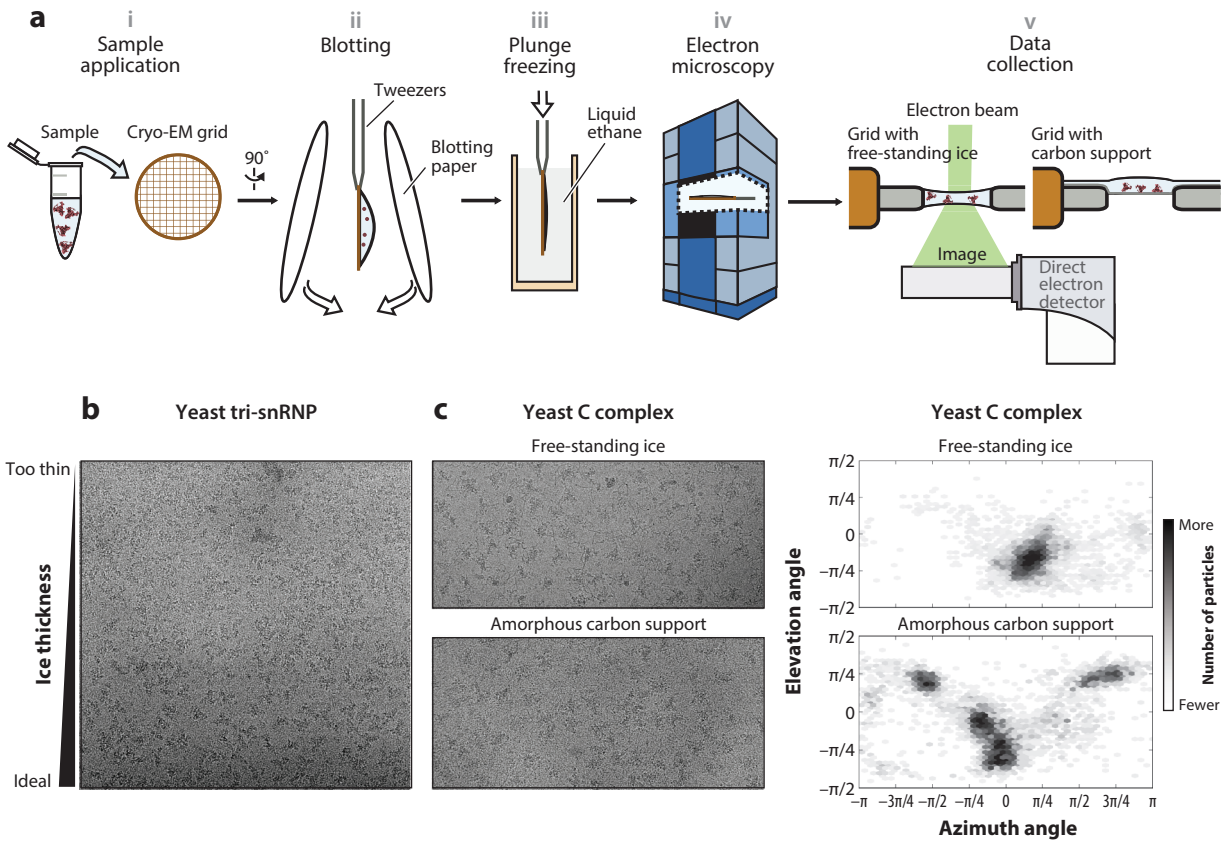


Figure 3

Cryo-EM sample preparation and data collection of the spliceosome. (*a*) Schematic depicting the procedure for cryo-EM sample preparation and data collection. Briefly, the purified sample is (*i*) applied to a metal grid, (*ii*) blotted, and (*iii*) plunged into liquid ethane (or an ethane/propanol mixture), flash cooling the sample in a thin film of vitreous ice. The spliceosome sample is then (*iv*) inserted into a transmission electron microscope and (*v*) imaged using a direct detector. In panel *v*, two grid types are depicted, showing particles in free-standing ice or on a carbon support. (*b*) Cryo-EM micrograph of the yeast tri-snRNP (82) indicating the effect of ice thickness on quality of the resultant single particles. (*c*) Comparison of the particle orientation distribution of the yeast C-complex spliceosome collected in free-standing ice and with a carbon support film. Representative cryo-EM micrographs (*left*) and the resultant particle orientations determined from 7,000 respective particles refined in CryoSPARC (100) (*right*) are shown (M.E. Wilkinson & K. Nagai, unpublished data). Abbreviations: EM, electron microscopy; snRNP, small nuclear ribonucleoprotein particle.

polyamylamine vapor, altering the surface charge of the carbon support and improving the angular distribution (82, 83).

4. Protein cross-linking: Transient and dynamic macromolecular complexes can be stabilized by cross-linking the protein components. The Stark group (51) developed the GraFix protocol, which combines glutaraldehyde cross-linking with a gradient ultracentrifugation step. In the Nagai group (37, 95), we often utilize the milder cross-linking agent bis-sulfosuccinimidylsuberate, which tends to cause less complex aggregation compared to glutaraldehyde (C. Plaschka & P. Cramer, unpublished results). Importantly, none of these cross-linking agents impede high-resolution structure determination of spliceosomes (37, 95, 150) and other macromolecular complexes (10, 11, 84, 94). Additionally, we and others have observed that cross-linking can improve the orientation distribution for spliceosome

particles (C. Plaschka, P.-C. Lin & K. Nagai, unpublished data) and other macromolecular complexes (10). Although the reason for this is unknown, we speculate that the cross-linked surface of the complex is likely to become chemically more similar, allowing for (slightly) less preferred interactions of the complex with the carbon support or the air–water interface.

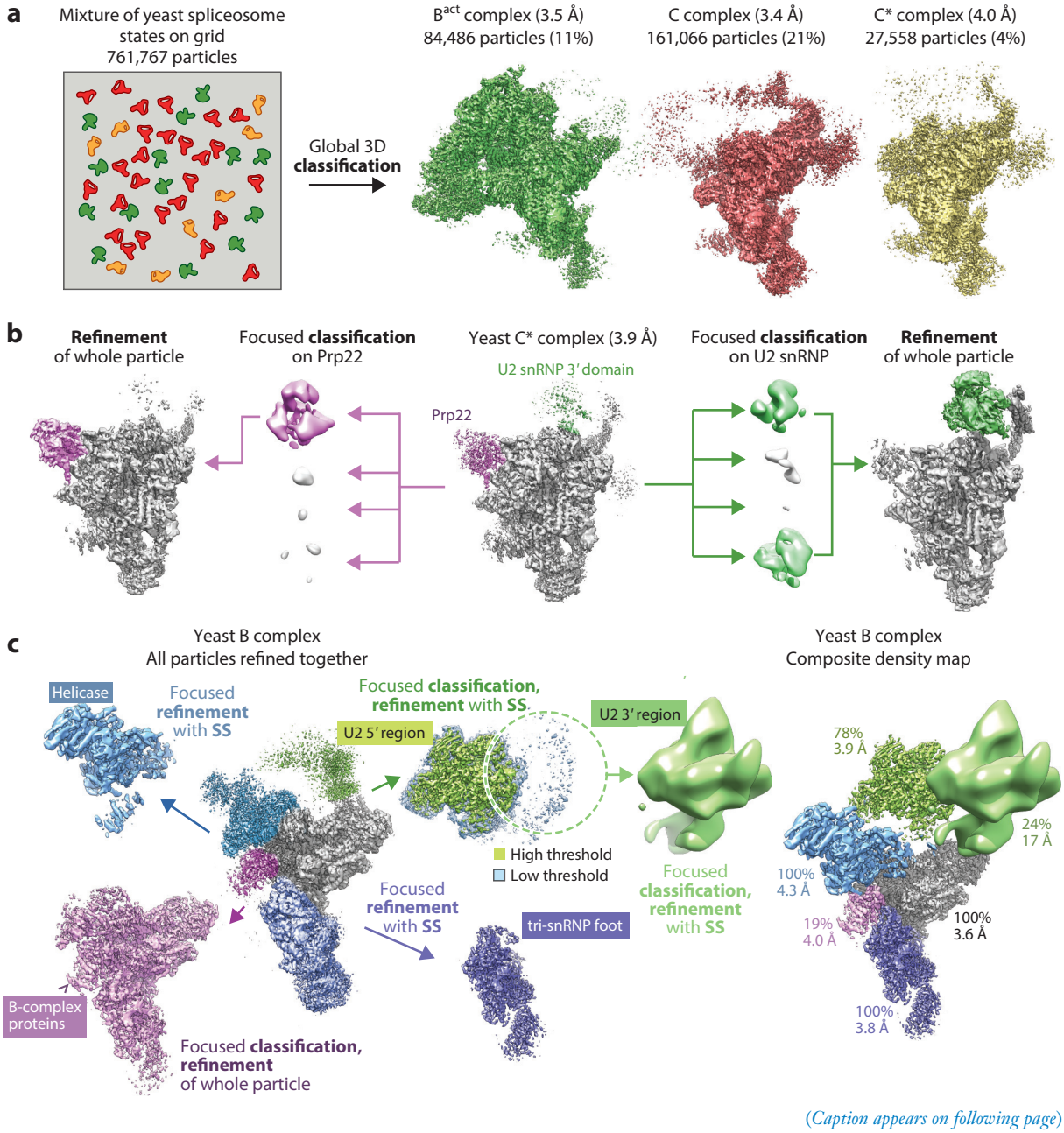
3.2. Data Collection

For one to determine high-resolution cryo-EM structures of the spliceosome, the grid is inserted into a high-energy 300-keV transmission electron microscope operated at cryogenic temperatures. While the microscope setup requires manual intervention, subsequent data acquisition is generally facilitated by automated acquisition software [e.g., EPU (FEI, Oregon), SerialEM (74), TOM (60), Leginon (99)]. A key advance in the cryo-EM resolution revolution was the development of direct electron detectors, which directly measure the energies of electron collisions with semiconductor pixels (34, 35). These provide better signal-to-noise ratio transmission than film or CCD cameras do. All subnanometer-resolution spliceosome structures were obtained using direct detectors. Direct detectors also have very fast readout rates, conferring three distinct advantages. First, images can be collected at high throughput, saving time at the microscope. Second, images can be collected as a series of frames (or so-called movies), which can subsequently be aligned to one another to correct for beam-induced motion that can blur high-resolution information (7, 19, 67). Biological samples are highly radiation sensitive, and movie frames additionally allow the down-weighting of later frames, where the high-resolution signal has deteriorated owing to radiation damage. Third, direct detectors can be operated in counting mode, whereby individual electron collisions are identified and treated equally instead of pixel intensity being proportional to the randomly varying energy of a colliding electron (as in integrating mode) (67, 77). Counting mode consequently improves signal-to-noise transmission in the final image, especially at lower spatial frequencies. This improvement in signal is particularly helpful for the alignment of small particles (<500 kDa). In our experience, it also helps in the alignment of spliceosome particles and the focused refinement of smaller subcomplexes (**Figure 4**). So far, all spliceosome maps higher than 4.5-Å resolution have been determined using the Gatan K2 summit direct detector operated in counting or super-resolution mode (12, 112). The recently developed Falcon III detector (FEI, Oregon), especially when operated in counting mode, may also produce high-resolution densities of the spliceosome.

4. IMAGE PROCESSING

The spliceosome is a dynamic macromolecule and contains multiple flexible regions. While the spliceosome has proven recalcitrant to X-ray crystallography, likely because of its conformational and compositional heterogeneity, EM image processing allows us to overcome these challenges. The published spliceosome data sets generally comprise 100,000–300,000 single particles, and these were generally all processed using the empirical Bayesian framework within the software RELION (110, 111). Briefly, after a number of micrographs have been collected, individual particle images are picked and extracted from the micrographs. To remove erroneously picked particles, such as carbon or other contaminants, the particle images are usually subjected to reference-free alignment and classification in 2D. High-quality particle images align to give class averages with secondary-structure features, while poor-quality particles do not align well and are discarded. The selected particle images can then be aligned in 3D to a low-pass filtered reference, which may be obtained from negative-stain, cryo-EM single-particle, or tomography data or from X-ray crystal structures (111). The particles can then be refined to high resolution. During the refinement

procedure, the data set is split into two random halves and refined independently to avoid overfitting. This will then reveal a well-aligned core with high-resolution density, where the peripheral regions can be of comparatively lower resolution, as exemplified by the yeast B complex described below in Section 4.2. To improve these peripheral densities, the data are further classified in 3D. Three-dimensional classification can separate populations of particles that correspond to different



(Caption appears on following page)

Figure 4 (*Figure appears on preceding page*)

Multiple image processing strategies are used to analyze spliceosome cryo-electron microscopy (cryo-EM) data. (*a*) Global three-dimensional (3D) image classification. Shi and colleagues (132, 145, 146) selected 762,000 single-particle images of a mixture of yeast spliceosome complexes. Global classification and global refinement allowed in silico purification of the B^{act}, C, and C* complexes (EMD-9524, -9525, and -6684, respectively). (*b*) Focused 3D image classification. The overall map for the yeast C* complex (EMD-3539) showed flexible peripheral components, corresponding to Prp22 (*purple*) and the U2 small nuclear ribonucleoprotein particle (snRNP) 3′ domain (*green*) (37, 95). Focused classification without image alignment produced classes with stronger density for these components. Particles from these classes were then globally refined, improving the map quality around Prp22 and the U2 snRNP 3′ domain (EMD-3541 and -3542). (*c*) Focused 3D classification and focused refinement improved the map quality for various regions of the yeast B complex. The tri-snRNP helicase and foot regions of the B complex were focused refined with signal subtraction (SS). A map with strong density for the B-complex proteins (Prp38, Snu23, Spp381) was obtained by focused classification around this region, followed by refinement of the whole tri-snRNP. The U2 snRNP 5′ region (SF3b) could be improved by focused classification and focused refinement with SS. The U2 snRNP 3′ region (SF3a and 3′ domain) was then improved by further focused classification and focused refinement with SS. The resulting maps can be combined into a composite map (*right*) to visualize the whole assembly. The percentage of particles from the polished data set, which contribute to the final densities, and the respective nominal resolution are indicated. Densities are from EMD-3682-8.

states of a feature of interest, which may represent different conformations or subcomplexes or entirely different spliceosome intermediates within the purified sample (**Figure 4**).

4.1. Three-Dimensional Classification

A major use for 3D classification in spliceosome structural biology is in resolving different functional states present in a purified sample. The most significant example of this comes from the Shi group (132, 145, 146), in which the structures of four different intermediates were determined from a mixture of spliceosomes isolated from *S. cerevisiae* (see Section 2.1) (**Figure 4*a***). Analysis of over 12,000 micrographs collected on a Titan Krios yielded approximately 760,000 particles. Several rounds of 3D classification yielded reconstructions of the B^{act}, C, and C* complexes. The abundant B^{act}- and C-complex particles produced maps at 3.5-Å and 3.4-Å resolution, respectively, and the less abundant C* could be determined to 4-Å resolution. By combining these data with an additional 4,000 micrographs of Yju2-tagged spliceosomes, Wan et al. (133) also determined the structure of the *S. cerevisiae* ILS at 3.5-Å resolution. It may be possible to capture rare splicing intermediates with an even larger data set. A potential drawback of this method is that the different identified complexes cannot be analyzed biochemically. For example, B^{act}, C, C*, and ILS complexes each contain one of the four homologous DEAH-box helicases. By this method, formally proving which helicase is bound to a particular complex is not possible. However, with significantly more data and thereby an improved local resolution, one may eventually be able to assign each helicase by its unique amino acid sequence. The Lührmann/Stark (12, 13, 103) and Nagai groups (37, 43, 95, 138) have used more biochemically defined samples, simplifying protein and RNA component assignments.

4.2. Focused Classification and Refinement

The above use of 3D classification resolves the heterogeneity of different spliceosome states. However, individual spliceosome states are also heterogeneous: proteins and RNA, particularly around the periphery, are highly flexible. Although many of the currently published spliceosome densities approach near-atomic resolution in their core, the local resolution often drops dramatically toward the periphery. In the high-resolution B^{act}, C and, C* structures, the core densities are uniformly sub-4-Å resolution (**Figure 5*a***), while peripheral elements like the U2 snRNP 3′

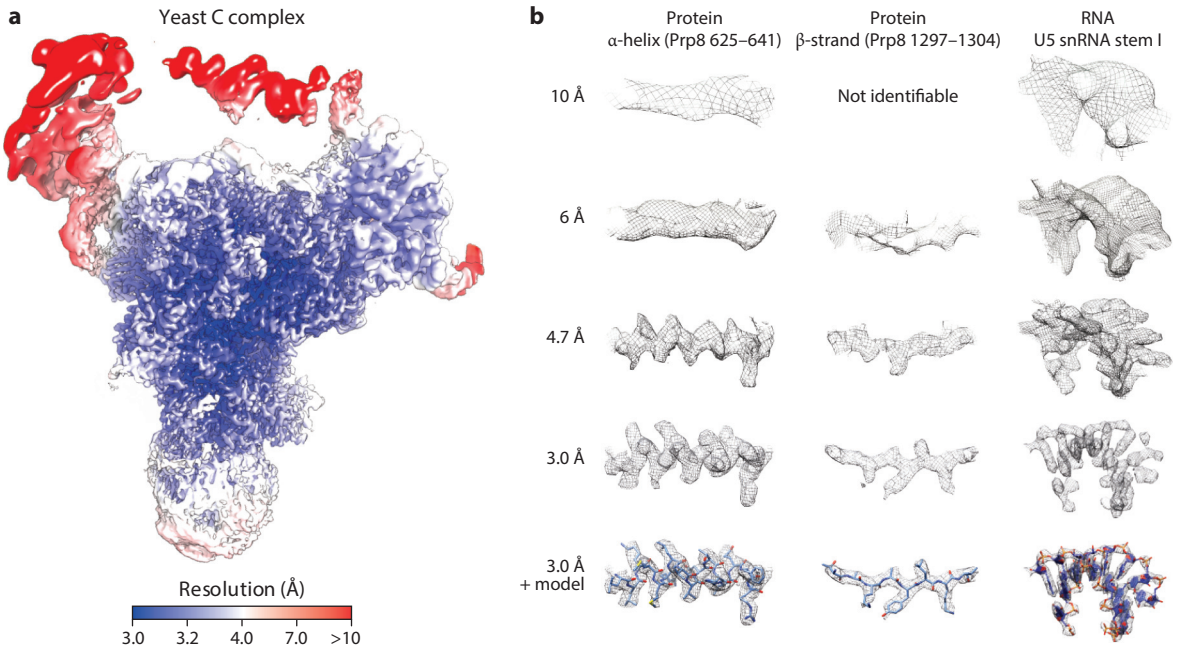


Figure 5

Local resolution of spliceosome cryo-electron microscopy (cryo-EM) densities. (*a*) The overall density of the spliceosome C complex at a nominal resolution of 3.0 Å was filtered by the local resolution in RELION (56, 110) using a 25-Å radius sphere centered around each voxel and colored (M.E. Wilkinson & K. Nagai, unpublished data). High-resolution features are apparent in the core of the complex, whereas the density in peripheral regions can drop in resolution owing to disordered regions, flexibility, and domains with low occupancy. (*b*) The C-complex cryo-EM density is shown with the fitted coordinate model and filtered to varying resolutions (3.0 Å, 4.7 Å, 6 Å, and 10 Å). Density is shown for a protein α-helix (Prp8 residues 625–641), a protein β-strand (Prp8 residues 1297–1304), and an RNA helix (U5 stem I).

domain (**Figures 2*b*** and **4*c***) and the large HAT-repeat-containing NTC proteins Syf1 and Clf1 (**Figures 2*c*** and **4*c***) are closer to 6–10-Å resolution. Other proteins, such as the helicases Brr2 and Prp16 in the C complex or the U2 snRNP in the B complex, are so flexible that, in overall density maps, they are blurred to the point of being initially unrecognizable. To overcome such flexibility, a powerful extension to 3D classification can be employed, where masks are placed around regions of interest. This is called focused classification because the classification is focused around one particular flexible region so that a specific heterogeneity is resolved. An advantage of this approach is that particle orientations can be maintained from refinement of the whole complex. This means that very small masks and small numbers of particles can be used since images do not need to be further aligned. For the C* complex, masks were placed around U2 snRNP and the helicase Prp22. Subsequent global refinement of the selected particle subsets yielded intermediate resolution maps of 4.2 Å and 4.6 Å, respectively (**Figure 4*b***), into which crystal structures, earlier cryo-EM structures, and homology models could be reliably docked (37). However, in none of these cases did the local resolution improve enough to allow atomic modeling.

If cryo-EM data are of sufficient quality, focused classification can be extended further to the technique of focused refinement. Here, the mask around the region of interest is maintained during refinement of the particle orientations so that the region is effectively treated as a single particle. Focused refinement can be performed in combination with partial signal subtraction (8), where signal outside the masked region is subtracted from the individual particle images.

Thereby, signal outside the masked region should interfere less with refinement of the particle inside the masked region. This is especially helpful when the masked region is small relative to the whole complex. Focused refinement allows the high-resolution refinement of a region sampling a continuum of positions, whereas focused classification in such a case would identify several average positions, which may still be blurred. Focused refinement generally requires more particles than a classification does, and the masked region must be large enough for subsequent image alignment. For the spliceosome, this approach was first applied to the *S. pombe* ILS and the *S. cerevisiae* tri-snRNP (82, 144). For focused classification of the C-complex Brr2 helicase and NTC modules, subsequent focused refinement was unsuccessful owing to insufficient signal in the particle images. However, by selecting for the highest-quality particles and increasing the data set size, both could be refined to higher resolution (M.E. Wilkinson & K. Nagai, unpublished results).

The yeast tri-snRNP and the yeast B complex structures demonstrate the power of focused classification and refinement (82, 95). In collaboration with the Scheres group, we (82) could extend the standard focused refinement to a multi-body refinement using the yeast tri-snRNP, where four flexible domains could be refined at once rather than in isolation, improving the overall alignment accuracy. The B complex contains the tri-snRNP and additionally the U2 snRNP, which is flexible relative to tri-snRNP and contains two domains, the U2 snRNP 5′ and 3′ regions (**Figure 2*b***). These are also flexible relative to each other. As such, when all B-complex particles are refined together, alignment is dominated by the largest rigid component, the body domain of the tri-snRNP, while all other components blur out (**Figure 4*c***). Focused refinement then improved the resolution of the helicase (Brr2) and foot domains of tri-snRNP to 4.3-Å and 3.8-Å resolution, respectively. Focused classification identified particles where the U2 snRNP is more stably attached, and subsequent focused refinement resolved the U2 snRNP 5′ region to 3.9-Å resolution. A further-focused classification and focused refinement from the aligned U2 5′-region particles yielded a subset in which the U2 3′ region could be visualized at 17 Å, allowing docking of previously determined structures (95). Additional focused classification on the tri-snRNP allowed unambiguous identification of the B-complex-specific proteins. Thus, the whole complex was best represented not with a single density but with a composite of different cryo-EM maps that each highlighted a specific part of the complex (**Figure 4*c***). Note that such focused refinements require generous soft masks so that regions overlapping between the different refinements are still partially included. In the extreme case, such as the U2 snRNP in the B complex, we could determine a defined B-complex structure from a small subset of particles to 7.2-Å resolution, allowing the fitting of the individually refined U2 snRNP relative to tri-snRNP in a single defined position. Taken together, image classification allows initially unresolved substructures to be visualized, yielding insight into the dynamic architecture of the spliceosome.

5. DENSITY INTERPRETATION

Recent advances in cryo-EM now enable routine structure determination of macromolecular complexes at 4-Å resolution, including the spliceosome (36, 85). Cryo-EM maps are, however, characteristically nonuniform in resolution (**Figure 5*a***), requiring a range of strategies for model building. Protein α-helices appear as tubular densities between 5 and 10 Å, revealing sidechain details at ∼4 Å or better, while β-strands are visible as continuous sheetlike densities until 4.7 Å, when strands become gradually more separated as the resolution increases (**Figure 5*b***). For densities with a local resolution better than 4 Å, models can be built de novo for both protein and RNA components. However, when resolution is low or density quality is poor (4–20 Å), we draw on orthogonal biochemical data and bioinformatics for interpretation. Central regions (core) of the spliceosome densities tend to be of high quality (<4 Å), whereas peripheral regions are often lower

in resolution (>4 Å). While map quality is correlated to refined B factors of the atomic model, B-factor refinement is not always performed on published atomic models, and direct inspection of the cryo-EM densities may be required.

The challenge in modeling a spliceosome comes from its compositional complexity (139). The spliceosome structures determined thus far contain only 40–60% of their total theoretical molecular weight. Given that 30–50 proteins and 3–5 snRNAs, plus a pre-mRNA substrate, are in each of the spliceosome intermediates, prior information is important to validate and accurately interpret the resulting densities. In the following sections, we provide examples from the yeast B- and C-complex structures (43, 95) on how we approach cryo-EM density interpretation for spliceosome proteins and RNA. These approaches are not unique to the spliceosome and can be applied to other problems.

5.1. Modeling Protein

The protein composition of different spliceosome intermediates has been studied by mass spectrometry (3, 17, 33, 49, 139, 151), yielding a comprehensive list of proteins likely to be present in the corresponding cryo-EM densities. With bioinformatic tools such as the InterPro server, we first determine domain architectures for each of these proteins (39). Protein secondary-structure predictions from, for example, the GeneSilico metaserver (61) provide further information on protein architecture. To increase prediction confidence, we often determine secondary structures for homologs of the target protein from related species (i.e., other yeasts, human, and *Drosophila*). Multisequence alignments help in identifying conserved protein regions that aid in density assignment, as conserved regions are more likely to be structured. Initial models for proteins with unknown structure can be produced by homology modeling using MODELLER (137), the I-TASSER (147), or SWISS-MODEL (14) protein prediction servers. To begin density interpretation, we first place these models, together with previously determined structures from X-ray crystallography, NMR, or cryo-EM, into the density using manual or automated procedures (31, 93, 140).

For cryo-EM density with a local resolution of 10–20 Å, models derived from X-ray crystallography or NMR can be reliably fitted as rigid bodies. Examples for this include fitting of the helicase Brr2 in a peripheral yeast C-complex density (10-Å resolution) or the U2 3′ region in the periphery of the yeast B complex (17-Å resolution). At intermediate resolution (4–10-Å resolution), minor model modifications can be made: In the yeast C* complex, the three domains of a Prp22 homology model could be rigid-body fitted into a peripheral density (∼5-Å local resolution) (**Figure 4*b***). Importantly, all these fits are made independently of orthogonal biochemical data, which allows their subsequent validation. Several spliceosome proteins contain extended structures that fold over large parts of the spliceosome density, such as Prp6 (B complex), Snu66 (B complex), and Prp45 (C complex), complicating their assignment (**Figure 2*b*,*c***). For these cases, we require high-resolution densities or orthogonal information. For example, a short peptide density found next to Dib1 (B complex) could be assigned to the N-terminus of Prp6 on the basis of protein interaction analysis (134). Protein–protein cross-linking and protein pull-down data (134) allowed assignment of a helix of Snu66 bound to Brr2 (B complex). The NTC protein Prp45 (C complex) spans a distance of 130 Å and could be assigned on the basis of its high-resolution density and consistency with secondary-structure predictions. Protein cross-linking has proved essential to the interpretation of low- to intermediate-resolution spliceosome densities and continues to serve in the validation of subunit assignments at high resolution (12, 13, 134, 144). For example, cross-linking data confirmed the location of the Prp9 ZnF2 domain in the U2 snRNP SF3b subcomplex (3.9 Å) (145) as well as contacts between a long helix of Spp381 and the Brr2 helicase (4.3 Å) (134).

Cross-links between the WD40 protein Prp17 and Cef1 in B^{act} complex helped determine the correct orientation of Prp17 in the C^* complex (145).

While de novo model building can be done at resolutions lower than 4 Å, the resulting models may contain errors (i.e., register shifts) and should be viewed with caution. To indicate regions of lower confidence, we remove sidechains in these regions from the final model, leaving only a polyalanine backbone model to indicate protein architecture.

For densities with a local resolution better than ∼4 Å, we build models de novo, using Coot, and perform subsequent coordinate refinement in programs such as PHENIX (1) or REFMAC5 (130). Additionally, other programs, such as Rosetta (105), MODELLER (137), and MDFF (76), can also be used to refine model coordinates to excellent quality. Although we do not benefit from sequence markers as in X-ray crystallography [i.e., anomalous signal from native sulfur atoms (Cys, Met) or selenomethionine substitution], the pattern of particularly the large aromatic sidechains (Phe, Tyr, and Trp) and general consistency with the cryo-EM density often allow for an unambiguous protein sequence assignment. For example, a 75-residue segment of Snu66 that interacts with the Prp8 RNaseH-like domain in the B complex could be assigned on the basis of density alone. In the yeast C complex, the protein Yju2, which is required for branching (step I), was also identified based solely on its density.

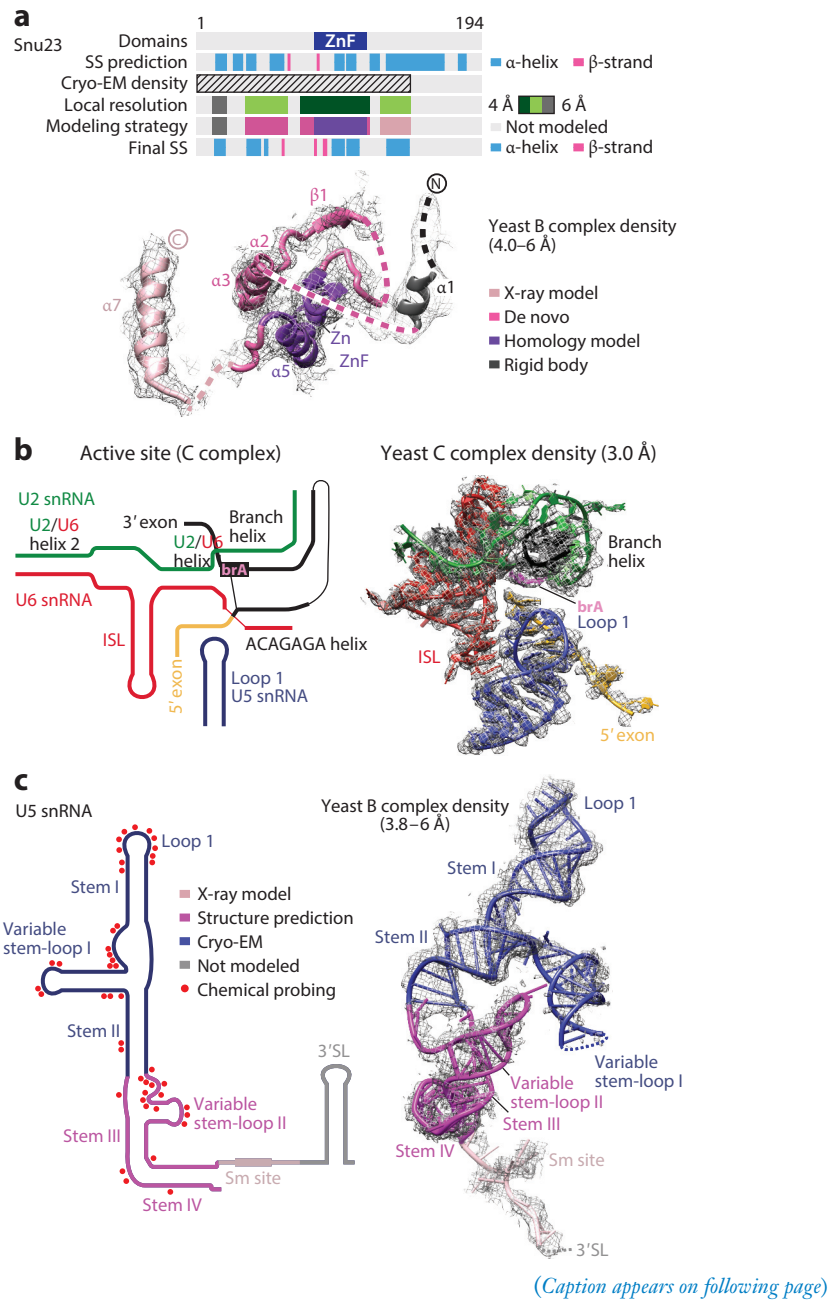
5.2. Protein Example: Snu23 (Local Resolution Between 4 and 6 Å)

Spliceosome activation requires three conserved B-complex-specific proteins, Spp381 (human MFAP1), Prp38, and Snu23 (72). In the yeast B-complex structure, we observed unassigned density near Prp8, Brr2, and U6 snRNA, which was improved by focused classification (see Section 4.2). Manual fitting in UCSF Chimera of the 2.8-Å crystal structure of the *Chaetomium thermophilum* Prp38-MFAP1-Snu23 heterotrimer from the Wahl group provided a very good fit, consistent with protein–protein cross-linking of Prp38 in the isolated yeast tri-snRNP (129, 134). Here, we focus on Snu23, an extended protein that we modeled according to the local resolution (4–6 Å) (**Figure 6*a***). From domain and secondary-structure predictions as well as prior biochemical studies (129), we knew that Snu23 is a mostly helical protein containing a central Zn-finger (ZnF) domain. A single α-helix (α7) of the *C. thermophilum* Snu23 was modeled in the crystal structure (129), providing a starting point for modeling and suggesting polarity of the connecting density. We first adjusted this helix (α7) at ∼5-Å local resolution. Toward the N-terminal side of helix α7, we could trace density into a nearby globular domain, which we identified as a canonical ZnF of expected size and sequence polarity (local resolution of ∼4 Å). We observed—consistent with secondary-structure prediction—a nearby β-strand (β1) and two α-helices (α2, α3) that could be built de novo in Coot (local resolution of 4 Å). This connected to a weak density of α-helical shape next to the helicase Brr2, suggesting that a short predicted helix in the Snu23 N-terminus binds Brr2 (6 Å) in the B complex. Consistent with our sequence assignment, larger sidechain density was observed in α-helices 2 and 3, where large hydrophobic sidechains reside (Tyr43, Tyr47), and additionally, the positively charged surface of the ZnF interacted with U6 snRNA. The Snu23 structure was of particular importance, as this protein may position the helicase Brr2 on its U4 snRNA substrate and stabilize the U6 snRNA ACACAGA stem–pre-mRNA interaction with its ZnF (95).

5.3. Modeling RNA

Over the years, increasing evidence has accumulated to support the notion that the spliceosome is a ribozyme consisting of U2, U5, and U6 snRNAs that form the catalytic RNA core for the branching (step I) and ligation (step II) reactions (22, 101, 117). Covariation analysis of the snRNAs

as well as genetics and cross-linking studies derived an RNA interaction network for the active site RNA (**Figure 6*b***) (139). Visualization of the 3D organization of the catalytic center by cryo-EM has provided new insights into the splicing mechanism. While RNA helices are readily apparent even at 15-Å resolution in cryo-EM densities, phosphate bumps become apparent around 4-Å resolution (**Figure 5*b***). Because of the small electric potential of the RNA phosphate backbone



(*Caption appears on following page*)

Figure 6 (*Figure appears on preceding page*)

Strategies for modeling protein and RNA components into spliceosome cryo-electron microscopy (cryo-EM) densities at varying resolution. (*a*) Modeling protein at 4–6-Å resolution: the B-complex protein Snu23. The tracks (*top*) show the different sources of information used to build and validate the final model: domain and secondary-structure (SS) predictions, the cryo-EM density, local resolution, the modeling strategy, and the determined SS. The model is shown below, superimposed on the cryo-EM density (92; EMD-3688, PDB 5NRL) and colored according to the modeling strategy for regions at different resolution (compare to the tracks above). (*b*) Modeling RNA at 3.0-Å resolution: the RNA active site in the yeast C complex (43). A secondary-structure diagram of the RNA interactions in the active site are shown as observed in the C-complex structure (*left*) next to the corresponding cryo-EM density (M.E. Wilkinson & K. Nagai, unpublished data) superimposed on the refined coordinate model. The active site model is entirely consistent with genetic and biochemical data. (*c*) Modeling RNA at 3.8–6-Å resolution: U5 snRNA in the yeast B complex (95). By combining chemical probing data (79), RNA tertiary-structure prediction (71, 98), and de novo modeling, we could determine a complete model for the short form of U5 snRNA [lacking the 3′ stem loop (3′SL)]. This revealed a new stem III and stem IV and the variable stem-loop II. A secondary-structure diagram (*left*) depicts the various modeling strategies for the different parts of U5 and the chemical probing data (79) used to validate the final model. The final U5 snRNA coordinate model is shown superimposed on the corresponding cryo-EM density (EMD-3686, PDB 5NRL). Additional abbreviations: ISL, internal stem loop, ZnF, zinc finger.

(135), the nucleotide bases are more readily apparent at moderate resolution (>4 Å). Both backbone and nucleotide base positions can be interpreted unambiguously at resolution better than 4 Å. At this resolution, purines (A, G) can be distinguished from pyrimidines (U, C), and as the resolution approaches 3 Å, adenosine and guanosine bases can be distinguished and stacked bases become individually resolved. In the yeast C complex at 3.8-Å, 3.4-Å, and, recently, 3.0-Å resolution (43, 132) (M.E. Wilkinson & K. Nagai, unpublished data), most of the U2/U6 active site is in the high-resolution core of the structure, enabling de novo modeling in Coot (**Figure 6*b***). As the 5′ and 3′ ends of U6 snRNA exit the core of the density, the resolution drops, and the extreme 5′ and 3′ ends (the 5′ stem loop and U2/U6 helix II, respectively) were modeled as idealized A-form helices. The model for yeast U5 snRNA variable stem loop II and stems III and VI could not be determined de novo in any of the spliceosome densities, owing to flexibility. However, through the use of RNA tertiary-structure prediction software combined with chemical probing data (71, 79, 98), one could obtain an accurate model in the yeast B-complex spliceosome at a local resolution of 3.8–6 Å (**Figure 6*c***), which was then docked into other structures (138).

5.4. Modeling at Intermediate Resolution

In every spliceosome reconstruction determined thus far, some unassigned density remains and some parts of the structure may not be connected to the core region. This is due to poor local resolution, a lack of density connectivity, or a lack of orthogonal information. When these regions are of α-helical appearance, we model them with idealized polyalanine helices to indicate their location, although their polarity sometimes remains unknown on the basis of density alone. The structures of the yeast tri-snRNP, B^{act}, C, C*, and P complexes and human C* complex were determined independently by at least two groups. These structures are in good agreement in the functionally important core of the spliceosome, but some differences in interpretation exist in the peripheral regions (6, 13, 37, 43, 70, 103, 132, 138, 145, 150). In these peripheral regions, the model remains tentative and is usually indicated in the deposited coordinate files by truncating protein sidechains, setting residues to polyalanine or so-called unknown sidechain groups. For example, for the human tri-snRNP determined at 7-Å resolution, only the protein Cα and nucleic acid phosphorus coordinates were deposited, whereas for human C* complex determined at 5.9-Å

resolution, the protein main-chain atom coordinates and the full RNA atom coordinates were deposited. Even though we can now obtain near-atomic densities of large parts of the spliceosome, care must still be taken when interpreting the data. The atomic coordinates deposited in the Protein Data Bank (PDB) are a best interpretation of the densities deposited with the Electron Microscopy Data Base (EMDB). These models could still be further improved by using new methods or by collecting more data, and the PDB file may be updated when appropriate.

6. MODEL VISUALIZATION

Preparing a coordinate model of the spliceosome requires extensive user input, partly due to the broad range in resolution of cryo-EM densities. These structural snapshots of the spliceosome allow for mechanistic insights, and we hope that these structures will aid in the design of new functional experiments. We encourage users of these structures to independently assess the quality of the local density and model in regions of interest. For the published structures, this can be gathered from figures in the publications or supplemental movies, where the densities and the model may be overlaid. Independently, the refined coordinates as well as associated densities are available from the PDB and the EMDB, respectively, and can be readily visualized using free software such as PyMOL and UCSF Chimera (93). We appreciate that the large size of the spliceosome, with more than 30 proteins and 3–6 RNAs, makes it daunting even to find out which chain name corresponds to which protein. To improve accessibility to these large structures, in the Nagai group, we therefore always publish interactive and annotated PyMOL sessions of each structure in the supplemental data of the associated publication, and a guide to using PyMOL to explore these sessions is available at the Nagai Lab website (**http://www2.mrc-lmb.cam.ac.uk/groups/nagai/**).

7. CONCLUSIONS

The resolution revolution in cryo-EM has allowed structural biologists to bring the spliceosome into focus over the past two years. The continued development of EM methodology, with, for example, new and improved sample grid supports (108) and better direct detectors (77) will allow us to determine the near-atomic structures of very transient and dynamic spliceosome intermediates. Software developments, from improved processing speeds on graphics processing units (56) to multi-body refinement (82, 111), will facilitate the analysis of difficult data sets. This is an exciting time to study pre-mRNA splicing. We look forward to the next few years as functional and structural biology come together to describe the splicing reaction and its regulatory mechanisms in atomic detail.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors would like to thank Sjors Scheres, Andy Newman, Lori Passmore, Ben Luisi, and Lisa Strittmatter for critical reading of the manuscript. The authors thank Kelly Nguyen, Xiao-chen Bai, Wojtek Galej, Sebastian Fica, Chris Oubridge, Chris Norman, Marike van Roon, Clément Charenton, and Christos Savva for their contributions to our cryo-EM project. This work was supported by the Medical Research Council (MC_U105184330) and European Research Council

Advanced Grant (693087-SPLICE3D). C.P. was supported by an EMBO fellowship, and M.E.W was supported by a Cambridge-Rutherford Memorial Scholarship.

LITERATURE CITED

1. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, et al. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66:213–21
2. Agafonov DE, Kastner B, Dybkov O, Hofele RV, Liu W-T, et al. 2016. Molecular architecture of the human U4/U6.U5 tri-snRNP. *Science* 351:1416–20
3. Agafonov DE, van Santen M, Kastner B, Dube P, Will CL, et al. 2016. ATPγS stalls splicing after B complex formation but prior to spliceosome activation. *RNA* 22:1329–37
4. Arenas JE, Abelson JN. 1997. Prp43: an RNA helicase-like factor involved in spliceosome disassembly. *PNAS* 94:11798–802
5. Azubel M, Wolf SG, Sperling J, Sperling R. 2004. Three-dimensional structure of the native spliceosome by cryo-electron microscopy. *Mol. Cell* 15:833–39
6. Bai R, Yan C, Wan R, Lei J, Shi Y. 2017. Structure of the post-catalytic spliceosome from *Saccharomyces cerevisiae*. *Cell* 171:1589–98.e8
7. Bai X-c, Fernandez IS, McMullan G, Scheres SHW. 2013. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife* 2:e00461
8. Bai X-c, Rajendra E, Yang G, Shi Y, Scheres SHW. 2015. Sampling the conformational space of the catalytic subunit of human γ-secretase. *eLife* 4:e11182
9. Berget SM, Moore C, Sharp PA. 1977. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *PNAS* 74:3171–75
10. Bernecky C, Herzog F, Baumeister W, Plitzko JM, Cramer P. 2016. Structure of transcribing mammalian RNA polymerase II. *Nature* 529:551–54
11. Bernecky C, Plitzko JM, Cramer P. 2017. Structure of a transcribing RNA polymerase II–DSIF complex reveals a multidentate DNA–RNA clamp. *Nat. Struct. Mol. Biol.* 24:809–15
12. Bertram K, Agafonov DE, Dybkov O, Haselbach D, Leelaram MN, et al. 2017. Cryo-EM structure of a pre-catalytic human spliceosome primed for activation. *Cell* 170:701–13.e11
13. Bertram K, Agafonov DE, Liu W-T, Dybkov O, Will CL, et al. 2017. Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* 542:318–23
14. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, et al. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 42:W252–58
15. Black DL, Steitz JA. 1986. Pre-mRNA splicing in vitro requires intact U4/U6 small nuclear ribonucleoprotein. *Cell* 46:697–704
16. Boehringer D, Makarov EM, Sander B, Makarova OV, Kastner B, et al. 2004. Three-dimensional structure of a pre-catalytic human spliceosomal complex B. *Nat. Struct. Mol. Biol.* 11:463–68
17. Boesler C, Rigo N, Anokhina MM, Tauchert MJ, Agafonov DE, et al. 2016. A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity. *Nat. Commun.* 7:11997
18. Boland A, Martin TG, Zhang Z, Yang J, Bai X-c, et al. 2017. Cryo-EM structure of a metazoan separase-securin complex at near-atomic resolution. *Nat. Struct. Mol. Biol.* 24:414–18
19. Brilot AF, Chen JZ, Cheng A, Pan J, Harrison SC, et al. 2012. Beam-induced motion of vitrified specimen on holey carbon film. *J. Struct. Biol.* 177:630–37
20. Brow DA, Guthrie C. 1988. Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature* 334:213–18
21. Burge CB, Tuschl T, Sharp PA. 1999. Splicing of precursors to mRNAs by the spliceosomes. In *The RNA World*, ed. RF Gesteland, TR Cech, JF Atkins, pp. 525–60. Cold Spring Harbor, NY: Cold Spring Harb. Lab. Press. 2nd ed.
22. Cech TR. 1986. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell* 44:207–10

23. Chan S-P, Kao D-I, Tsai W-Y, Cheng S-C. 2003. The Prp19p-associated complex in spliceosome activation. *Science* 302:279–82
24. Chen W, Shulha HP, Ashar-Patel A, Yan J, Green KM, et al. 2014. Endogenous U2·U5·U6 snRNA complexes in *S. pombe* are intron lariat spliceosomes. *RNA* 20:308–20
25. Chow LT, Gelinas RE, Broker TR, Roberts RJ. 1977. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* 12:1–8
26. Company M, Arenas J, Abelson J. 1991. Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature* 349:487–93
27. Cordin O, Beggs JD. 2013. RNA helicases in splicing. *RNA Biol*. 10:83–95
28. Das R, Zhou Z, Reed R. 2000. Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol. Cell* 5:779–87
29. Dignam JD, Lebovitz RM, Roeder RG. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res*. 11:1475–89
30. Edwalds-Gilbert G, Kim DH, Kim SH, Tseng YH, Yu Y, Lin RJ. 2000. Dominant negative mutants of the yeast splicing factor Prp2 map to a putative cleft region in the helicase domain of DExD/H-box proteins. *RNA* 6:1106–19
31. Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 60:2126–32
32. Fabrizio P, Abelson J. 1992. Thiophosphates in yeast U6 snRNA specifically affect pre-mRNA splicing in vitro. *Nucleic Acids Res*. 20:3659–64
33. Fabrizio P, Dannenberg J, Dube P, Kastner B, Stark H, et al. 2009. The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol. Cell* 36:593–608
34. Faruqi AR, Henderson R. 2007. Electronic detectors for electron microscopy. *Curr. Opin. Struct. Biol.* 17:549–55
35. Faruqi AR, McMullan G. 2011. Electronic detectors for electron microscopy. *Q. Rev. Biophys.* 44:357–90
36. Fernandez-Leiro R, Scheres SH. 2016. Unravelling biological macromolecules with cryo-electron microscopy. *Nature* 537:339–46
37. Fica SM, Oubridge C, Galej WP, Wilkinson ME, Bai X-C, et al. 2017. Structure of a spliceosome remodelled for exon ligation. *Nature* 542:377–80
38. Fica SM, Tuttle N, Novak T, Li N-S, Lu J, et al. 2013. RNA catalyses nuclear pre-mRNA splicing. *Nature* 503:229–34
39. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, et al. 2017. InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res*. 45:D190–99
40. Fischer N, Konevega AL, Wintermeyer W, Rodnina MV, Stark H. 2010. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* 466:329–33
41. Fleckner J, Zhang M, Valcárcel J, Green MR. 1997. U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev*. 11:1864–72
42. Fourmann JB, Schmitzova J, Christian H, Urlaub H, Ficner R, et al. 2013. Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. *Genes Dev*. 27:413–28
43. Galej WP, Wilkinson ME, Fica SM, Oubridge C, Newman AJ, Nagai K. 2016. Cryo-EM structure of the spliceosome immediately after branching. *Nature* 537:197–201
44. Golas MM, Sander B, Will CL, Lührmann R, Stark H. 2003. Molecular architecture of the multiprotein splicing factor SF3b. *Science* 300:980–84
45. Golas MM, Sander B, Will CL, Lührmann R, 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* 17:869–83
46. Henderson R. 1995. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* 28:171–93
47. Hoskins AA, Rodgers ML, Friedman LJ, Gelles J, Moore MJ. 2016. Single molecule analysis reveals reversible and irreversible steps during spliceosome activation. *eLife* 5:e14166
48. Jurica MS, Licklider LJ, Gygi SR, Grigorieff N, Moore MJ. 2002. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA* 8:426–39

49. Jurica MS, Moore MJ. 2003. Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* 12:5–14
50. Jurica MS, Sousa D, Moore MJ, Grigorieff N. 2004. Three-dimensional structure of C complex spliceosomes by electron microscopy. *Nat. Struct. Mol. Biol.* 11:265–69
51. Kastner B, Fischer N, Golas MM, Sander B, Dube P, et al. 2008. GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* 5:53–55
52. Kastner B, Kornstadt U, Bach M, Lührmann R. 1992. Structure of the small nuclear RNP particle U1: identification of the two structural protuberances with RNP-antigens A and 70K. *J. Cell Biol.* 116:839–49
53. Kataoka N, Dreyfuss G. 2008. Preparation of efficient splicing extracts from whole cells, nuclei, and cytoplasmic fractions. *Methods Mol. Biol.* 488:357–65
54. Kim D-H, Edwalds-Gilbert G, Ren C, Lin R-J. 1999. A mutation in a methionine tRNA gene suppresses the *prp2*-1 Ts mutation and causes a pre-mRNA splicing defect in *Saccharomyces cerevisiae*. *Genetics* 153:1105–15
55. Kim SH, Lin RJ. 1996. Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Mol. Cell. Biol.* 16:6810–19
56. Kimanius D, Forsberg BO, Scheres SH, Lindahl E. 2016. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* 5:e18722
57. Kistler AL, Guthrie C. 2001. Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. *Genes Dev.* 15:42–49
58. Knapek E, Dubochet J. 1980. Beam damage to organic material is considerably reduced in cryo-electron microscopy. *J. Mol. Biol.* 141:147–61
59. Konarska MM, Sharp PA. 1986. Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. *Cell* 46:845–55
60. Korinek A, Beck F, Baumeister W, Nickell S, Plitzko JM. 2011. Computer controlled cryo-electron microscopy–TOM^2 a software package for high-throughput applications. *J. Struct. Biol.* 175:394–405
61. Kozlowski LP, Bujnicki JM. 2012. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinform.* 13:111
62. Kuhlbrandt W. 2014. Biochemistry. The resolution revolution. *Science* 343:1443–44
63. Laggerbauer B, Achsel T, Lührmann R. 1998. The human U5–200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro. *PNAS* 95:4188–92
64. Legrain P, Seraphin B, Rosbash M. 1988. Early commitment of yeast pre-mRNA to the spliceosome pathway. *Mol. Cell. Biol.* 8:3755–60
65. Lerner MR, Boyle JA, Mount SM, Wolin SL, Steitz JA. 1980. Are snRNPs involved in splicing? *Nature* 283:220–24
66. Li X, Liu S, Jiang J, Zhang L, Espinosa S, et al. 2017. CryoEM structure of *Saccharomyces cerevisiae* U1 snRNP offers insight into alternative splicing. *Nat. Commun.* 8:1035
67. Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, et al. 2013. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* 10:584–90
68. Liang W-W, Cheng S-C. 2015. A novel mechanism for Prp5 function in prespliceosome formation and proofreading the branch site sequence. *Genes Dev.* 29:81–93
69. Lin R-J, Newman AJ, Cheng S-C, Abelson J. 1985. Yeast mRNA splicing in vitro. *J. Biol. Chem.* 260:14780–92
70. Liu S, Li X, Zhang L, Jiang J, Hill RC, et al. 2017. Structure of the yeast spliceosomal postcatalytic P complex. *Science* 358:1278–83
71. Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, et al. 2011. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6:26
72. Lybarger S, Beickman K, Brown V, Dembla-Rajpal N, Morey K, et al. 1999. Elevated levels of a U4/U6.U5 snRNP-associated protein, Spp381p, rescue a mutant defective in spliceosome maturation. *Mol. Cell. Biol.* 19:577–84
73. Madhani HD, Guthrie C. 1992. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* 71:803–17
74. Mastronarde DN. 2005. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* 152:36–51

75. McDowall AW, Chang J-J, Freeman R, Lepault J, Walter CA, Dubochet J. 1983. Electron microscopy of frozen hydrated sections of vitreous ice and vitrified biological samples. *J. Microsc.* 131:1–9
76. McGreevy R, Singharoy A, Li Q, Zhang J, Xu D, et al. 2014. xMDFF: molecular dynamics flexible fitting of low-resolution X-ray structures. *Acta Crystallogr. D Biol. Crystallogr.* 70:2344–55
77. McMullan G, Faruqi AR, Henderson R. 2016. Direct electron detectors. *Methods Enzymol.* 579:1–17
78. Moore MJ, Sharp PA. 1992. Site-specific modification of pre-mRNA: the 2′-hydroxyl groups at the splice sites. *Science* 256:992–97
79. Mougin A, Gottschalk A, Fabrizio P, Luhrmann R, Branlant C. 2002. Direct probing of RNA structure and RNA-protein interactions in purified HeLa cell's and yeast spliceosomal U4/U6.U5 tri-snRNP particles. *J. Mol. Biol.* 317:631–49
80. Newman AJ, Lin R-J, Cheng S-C, Abelson J. 1985. Molecular consequences of specific intron mutations on yeast mRNA splicing in vivo and in vitro. *Cell* 42:335–44
81. Newman AJ, Norman C. 1992. U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. *Cell* 68:743–54
82. Nguyen THD, Galej WP, Bai X-c, Oubridge C, Newman AJ, et al. 2016. Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* 530:298–302
83. Nguyen THD, Galej WP, Bai X-c, Savva CG, Newman AJ, et al. 2015. The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* 523:47–52
84. Nogales E, Louder RK, He Y. 2016. Cryo-EM in the study of challenging systems: the human transcription pre-initiation complex. *Curr. Opin. Struct. Biol.* 40:120–27
85. Nogales E, Scheres SHW. 2015. Cryo-EM: a unique tool for the visualization of macromolecular complexity. *Mol. Cell* 58:677–89
86. Ohi MD, Gould KL. 2002. Characterization of interactions among the Cef1p-Prp19p-associated splicing complex. *RNA* 8:798–815
87. Ohi MD, Ren L, Wall JS, Gould KL, Walz T. 2007. Structural characterization of the fission yeast U5.U2/U6 spliceosome complex. *PNAS* 104:3195–200
88. Ohrt T, Odenwalder P, Dannenberg J, Prior M, Warkocki Z, et al. 2013. Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. *RNA* 19:902–15
89. Padgett RA, Hardy SF, Sharp PA. 1983. Splicing of adenovirus RNA in a cell-free transcription system. *PNAS* 80:5230–34
90. Pantelic RS, Meyer JC, Kaiser U, Baumeister W, Plitzko JM. 2010. Graphene oxide: a substrate for optimizing preparations of frozen-hydrated samples. *J. Struct. Biol.* 170:152–56
91. Parker R, Guthrie C. 1985. A point mutation in the conserved hexanucleotide at a yeast 5′ splice junction uncouples recognition, cleavage, and ligation. *Cell* 41:107–18
92. Perriman R, Ares M Jr. 2010. Invariant U2 snRNA nucleotides form a stem loop to recognize the intron early in splicing. *Mol. Cell* 38:416–27
93. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. 2004. UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25:1605–12
94. Plaschka C, Hantsche M, Dienemann C, Burzinski C, Plitzko J, Cramer P. 2016. Transcription initiation complex structures elucidate DNA opening. *Nature* 533:353–58
95. Plaschka C, Lin P-C, Nagai K. 2017. Structure of a pre-catalytic spliceosome. *Nature* 546:617–21
96. Plumpton M, McGarvey M, Beggs JD. 1994. A dominant negative mutation in the conserved RNA helicase motif 'SAT' causes splicing factor PRP2 to stall in spliceosomes. *EMBO J.* 13:879–87
97. Pomeranz Krummel DA, Oubridge C, Leung AK, Li J, Nagai K. 2009. Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* 458:475–80
98. Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, et al. 2012. Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* 40:e112
99. Potter CS, Chu H, Frey B, Green C, Kisseberth N, et al. 1999. Leginon: a system for fully automated acquisition of 1000 electron micrographs a day. *Ultramicroscopy* 77:153–61
100. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* 14:290–96
101. Pyle AM. 2016. Group II intron self-splicing. *Annu. Rev. Biophys.* 45:183–205

102. Raghunathan PL, Guthrie C. 1998. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr. Biol.* 8:847–55
103. Rauhut R, Fabrizio P, Dybkov O, Hartmuth K, Pena V, et al. 2016. Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. *Science* 353:1399
104. Reed R. 1989. The organization of 3′ splice-site sequences in mammalian introns. *Genes Dev.* 3:2113–23
105. Rohl CA, Strauss CE, Misura KM, Baker D. 2004. Protein structure prediction using Rosetta. *Methods Enzymol.* 383:66–93
106. Ruskin B, Greene JM, Green MR. 1985. Cryptic branch point activation allows accurate in vitro splicing of human beta-globin intron mutants. *Cell* 41:833–44
107. Russo CJ, Passmore LA. 2014. Controlling protein adsorption on graphene for cryo-EM using low-energy hydrogen plasmas. *Nat. Methods* 11:649–52
108. Russo CJ, Passmore LA. 2014. Electron microscopy: ultrastable gold substrates for electron cryomicroscopy. *Science* 346:1377–80
109. Sander B, Golas MM, Makarov EM, Brahms H, Kastner B, et al. 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* 24:267–78
110. Scheres SHW. 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* 180:519–30
111. Scheres SHW. 2016. Processing of structurally heterogeneous cryo-EM data in RELION. *Methods Enzymol.* 579:125–57
112. Scheres SHW, Nagai K. 2017. CryoEM structures of spliceosomal complexes reveal the molecular mechanism of pre-mRNA splicing. *Curr. Opin. Struct. Biol.* 46:130–39
113. Schneider S, Hotz HR, Schwer B. 2002. Characterization of dominant-negative mutants of the DEAH-box splicing factors Prp22 and Prp16. *J. Biol. Chem.* 277:15452–58
114. Schwer B. 2008. A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Mol. Cell* 30:743–54
115. Schwer B, Guthrie C. 1992. A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *EMBO J.* 11:5033–39
116. Schwer B, Guthrie C. 1992. A dominant negative mutation in a spliceosomal ATPase affects ATP hydrolysis but not binding to the spliceosome. *Mol. Cell. Biol.* 12:3540–47
117. Sharp PA. 1985. On the origin of RNA splicing and introns. *Cell* 42:397–400
118. Sheth N, Roca X, Hastings ML, Roeder T, Krainer AR, Sachidanandam R. 2006. Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res.* 34:3955–67
119. Sibley CR, Blazquez L, Ule J. 2016. Lessons from non-canonical splicing. *Nat. Rev. Genet.* 17:407–21
120. Spingola M, Grate L, Haussler D, Ares M Jr. 1999. Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* 5:221–34
121. Staley JP, Guthrie C. 1999. An RNA switch at the 5′ splice site requires ATP and the DEAD box protein Prp28p. *Mol. Cell* 3:55–64
122. Stark H, Lührmann R. 2006. Cryo-electron microscopy of spliceosomal components. *Annu. Rev. Biophys. Biomol. Struct.* 35:435–57
123. Steitz TA, Steitz JA. 1993. A general two-metal-ion mechanism for catalytic RNA. *PNAS* 90:6498–502
124. Stevens SW, Abelson J. 1999. Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. *PNAS* 96:7226–31
125. Strauss EJ, Guthrie C. 1994. PRP28, a 'DEAD-box' protein, is required for the first step of mRNA splicing in vitro. *Nucleic Acids Res.* 22:3187–93
126. Sun JS, Manley JL. 1995. A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes Dev.* 9:843–54
127. Tsai R-T, Fu R-H, Yeh F-L, Tseng C-K, Lin Y-C, et al. 2005. Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. *Genes Dev.* 19:2991–3003
128. Tseng C-K, Liu H-L, Cheng S-C. 2011. DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. *RNA* 17:145–54
129. Ulrich AK, Seeger M, Schütze T, Bartlick N, Wahl MC. 2016. Scaffolding in the spliceosome via single α helices. *Structure* 24:1972–83

130. Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, et al. 2004. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D Biol. Crystallogr.* 60:2184–95
131. Vijayraghavan U, Parker R, Tamm J, Iimura Y, Rossi J, et al. 1986. Mutations in conserved intron sequences affect multiple steps in the yeast splicing pathway, particularly assembly of the spliceosome. *EMBO J.* 5:1683–95
132. Wan R, Yan C, Bai R, Huang G, Shi Y. 2016. Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* 353:895–904
133. Wan R, Yan C, Bai R, Lei J, Shi Y. 2017. Structure of an intron lariat spliceosome from *Saccharomyces cerevisiae*. *Cell* 171:120–32.e12
134. Wan R, Yan C, Bai R, Wang L, Huang M, et al. 2016. The 3.8 Å structure of the U4/U6.U5 tri-snRNP: insights into spliceosome assembly and catalysis. *Science* 351:466–75
135. Wang J, Moore PB. 2017. On the interpretation of electron microscopic maps of biological macromolecules. *Protein Sci.* 26:122–29
136. Warkocki Z, Odenwalder P, Schmitzova J, Platzmann F, Stark H, et al. 2009. Reconstitution of both steps of *Saccharomyces cerevisiae* splicing with purified spliceosomal components. *Nat. Struct. Mol. Biol.* 16:1237–43
137. Webb B, Sali A. 2014. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinform.* 54:5.6.1–37
138. Wilkinson ME, Fica SM, Galej WP, Norman CM, Newman AJ, Nagai K. 2017. Postcatalytic spliceosome structure reveals mechanism of 3′-splice site selection. *Science* 358:1283–88
139. Will CL, Lührmann R. 2011. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 3:a003707
140. Wriggers W. 2012. Conventions and workflows for using Situs. *Acta Crystallogr. Sect. D Biol. Crystall.* 68:344–51
141. Wu J, Manley JL. 1989. Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. *Genes Dev.* 3:1553–61
142. Wu S, Romfo CM, Nilsen TW, Green MR. 1999. Functional recognition of the 3′ splice site AG by the splicing factor U2AF35. *Nature* 402:832–35
143. Xu Y-Z, Newnham CM, Kameoka S, Huang T, Konarska MM, Query CC. 2004. Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. *EMBO J.* 23:376–85
144. Yan C, Hang J, Wan R, Huang M, Wong CC, Shi Y. 2015. Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* 349:1182–91
145. Yan C, Wan R, Bai R, Huang G, Shi Y. 2016. Structure of a yeast activated spliceosome at 3.5 Å resolution. *Science* 353:904–11
146. Yan C, Wan R, Bai R, Huang G, Shi Y. 2017. Structure of a yeast step II catalytically activated spliceosome. *Science* 355:149–55
147. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* 12:7–8
148. Yu YT, Maroney PA, Darzynkiwicz E, Nilsen TW. 1995. U6 snRNA function in nuclear pre-mRNA splicing: a phosphorothioate interference analysis of the U6 phosphate backbone. *RNA* 1:46–54
149. Zhang L, Li X, Zhao R. 2013. Structural analyses of the pre-mRNA splicing machinery. *Protein Sci.* 22:677–92
150. Zhang X, Yan C, Hang J, Finci LI, Lei J, Shi Y. 2017. An atomic structure of the human spliceosome. *Cell* 169:918–29.e14
151. Zhou Z, Licklider LJ, Gygi SP, Reed R. 2002. Comprehensive proteomic analysis of the human spliceosome. *Nature* 419:182–85
152. Zillmann M, Zapp ML, Berget SM. 1988. Gel electrophoretic isolation of splicing complexes containing U1 small nuclear ribonucleoprotein particles. *Mol. Cell. Biol.* 8:814–21