



# Mechanistic insights into precursor messenger RNA splicing by the spliceosome

Yigong Shi<sup>1,2</sup>

**Abstract** | Precursor messenger RNA (pre-mRNA) splicing is an essential step in the flow of information from DNA to protein in all eukaryotes. Research over the past four decades has molecularly delineated the splicing pathway, including characterization of the detailed splicing reaction, definition of the spliceosome and identification of its components, and biochemical analysis of the various splicing complexes and their regulation. Structural information is central to mechanistic understanding of pre-mRNA splicing by the spliceosome. X-ray crystallography of the spliceosomal components and subcomplexes is complemented by electron microscopy of the intact spliceosome. In this Review, I discuss recent atomic-resolution structures of the intact spliceosome at different stages of the splicing cycle. These structures have provided considerable mechanistic insight into pre-mRNA splicing and have corroborated and explained a large body of genetic and biochemical data. Together, the structural data have proved that the spliceosome is a protein-directed metalloribozyme.

**Intron lariat**  
A lasso-like intron structure formed by the first transesterification of the splicing cycle.

**Branch point sequence (BPS)**. Also known as a branch site. A conserved intron sequence, which contains an invariant adenosine nucleotide that serves as the nucleophile for the branching reaction.

<sup>1</sup>Beijing Advanced Innovation Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China.

<sup>2</sup>Institute of Biology, Westlake Institute for Advanced Study, Shilongshan Road No. 18, Xihu District, Hangzhou 310064, Zhejiang Province, China.  
[shi-lab@tsinghua.edu.cn](mailto:shi-lab@tsinghua.edu.cn)

doi:10.1038/nrm.2017.86  
Published online 27 Sep 2017

Pre-mRNA splicing was discovered 40 years ago independently by Phillip Sharp<sup>1</sup> and Richard Roberts<sup>2</sup>. Patient-derived antibodies were found to react with five uridine-rich small nuclear RNAs (snRNAs), known as U1, U2, U4, U5 and U6 snRNAs, and with seven proteins of 12–35 kDa (REF. 3), which together with the snRNAs form small nuclear ribonucleoproteins (snRNPs). The 5' end sequences of U1 snRNA were recognized to be complementary to the 5' splice site (5'SS)<sup>4,5</sup>, and the snRNPs were partially purified<sup>6</sup>. The purified U1 snRNP specifically bound to the 5'SS *in vitro*<sup>7</sup>, and the depletion of U1 snRNP inhibited *in vitro* splicing<sup>8–10</sup>. Subsequently, the involvement of multiple snRNPs in pre-mRNA splicing was confirmed<sup>11–16</sup>.

Development of the splicing assay enabled the delineation of the splicing reaction<sup>17–20</sup>. ATP and magnesium ( $Mg^{2+}$ ) were found to be indispensable for the *in vitro* splicing of adenovirus pre-mRNA<sup>18,21</sup>, and *in vitro* splicing of the mutant  $\beta$ -globin pre-mRNA derived from individuals with  $\beta$ -thalassaemia was identical to its splicing *in vivo*<sup>22</sup>. Formation of an intron lariat through linkage between an adenosine nucleotide in the branch point sequence (BPS) and the guanine nucleotide at the 5' end of the 5'SS was biochemically confirmed<sup>23–25</sup>. Each cycle of pre-mRNA splicing entails two  $S_{N}2$ -type transesterification steps. Step I (commonly known as branching) results in a free 5' exon and an

intron lariat–3' exon intermediate, and step II (exon ligation) leads to the joining of the 5' exon with the 3' exon (FIG. 1a).

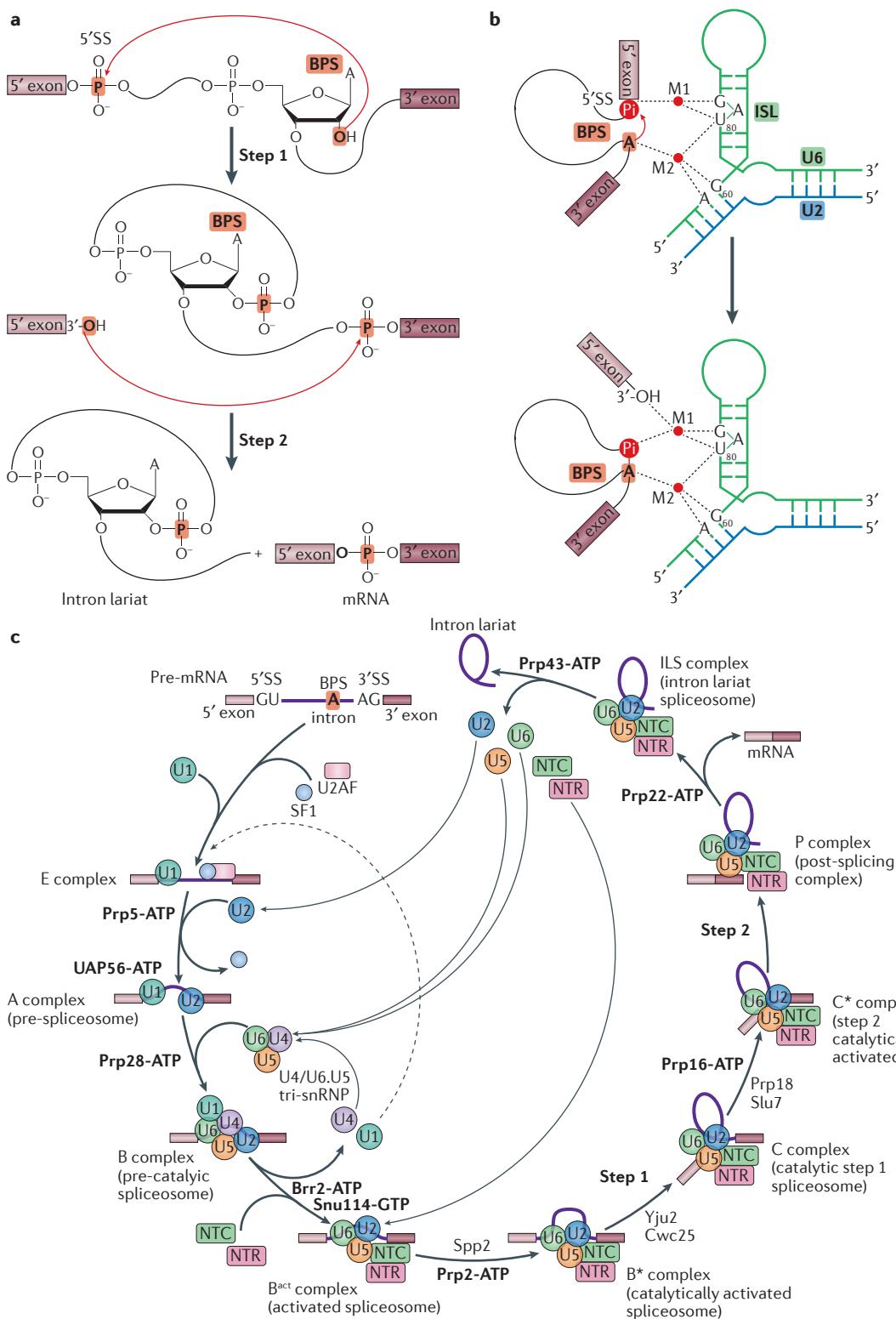
A general two-metal catalytic mechanism was proposed for pre-mRNA splicing, in which the snRNA was thought to coordinate the catalytic metal ions<sup>26</sup>. This hypothesis was supported by experiments using sulfur-substituted pre-mRNA<sup>27</sup>. Sulfur substitution of the non-bridging phosphoryl oxygen of nucleotide U80 in U6 snRNA rendered the yeast spliceosome catalytically inactive; the step I reaction was rescued by thiophilic ions such as manganese ( $Mn^{2+}$ ) in the U80( $S_p$ )-reconstituted but not the U80( $R_p$ )-reconstituted spliceosome, identifying U80 as a specific coordinating ligand of the catalytic metal<sup>28</sup> (FIG. 1b). Using a similar strategy, U6 snRNA was shown to catalyse both steps of the reaction by positioning divalent metals<sup>29</sup> (FIG. 1b). Structural comparison between the group IIC intron core and the known features of U6 snRNA enabled prediction of the molecular structure of the spliceosomal active site<sup>30,31</sup>. Consistent with the two-metal catalysis model<sup>26,31</sup>, all identified U6 metal-coordinating ligands correspond to those in the group II self-splicing introns<sup>29</sup>. These similarities suggest that the group II intron and the spliceosome share the same evolutionary origin, with the former being the ancestral type.

**S<sub>N</sub>2-type transesterification**  
Transesterification in which one ester bond is broken and another ester bond is formed synchronously; named for substitution nucleophilic and bi-molecular transesterification.

**Group II self-splicing introns**  
A large class of self-splicing catalytic RNAs found in bacteria, fungi and plants.

Pre-mRNA splicing is executed by a large ribonucleoprotein complex known as the spliceosome<sup>32–34</sup>. Assembly of the spliceosome depends strictly on the presence of the 5'SS, BPS and 3'SS, which are conserved sequence elements of introns<sup>34–38</sup>. Analysis of the *in vitro* splicing reaction by non-denaturing polyacrylamide gel electrophoresis enabled the identification

and molecular characterization of distinct spliceosomal complexes<sup>38–44</sup> (FIG. 1c). The pre-catalytic B complex is converted into an activated spliceosome ( $B^{act}$  complex), which no longer contains the U1 or U4 snRNP. Through additional remodelling, the  $B^{act}$  complex matures into a catalytically activated spliceosome ( $B^*$  complex), in which the branching reaction occurs.



**MS2-tagged**  
RNA prepared for affinity purification using a technique based on the interactions of the MS2 bacteriophage coat protein with a stem-loop RNA structure from the phage genome.

The resulting catalytic step I spliceosome (C complex) is remodelled into a step II catalytically activated spliceosome (C\* complex), where the exon ligation occurs (FIG. 1c). The post-catalytic spliceosome (P complex) sheds the ligated exon to become the intron lariat spliceosome (ILS complex), which releases the intron lariat and disassembles into individual snRNPs and associated proteins.

The five snRNPs constitute the core components of the various spliceosomal complexes. With the exception of U6, each of the snRNAs contains a trimethylated guanine nucleotide ( $m_3G$ ) at the 5' end, and a methylated adenine nucleotide ( $m^6A$ ) is present in U2, U4 and U6 snRNAs<sup>45</sup>. Use of monoclonal antibodies against  $m_3G$  and  $m^6A$  improved the purification of the individual snRNPs and identification of their protein components<sup>46</sup>. In *Saccharomyces cerevisiae*, a central spliceosomal component, pre-mRNA processing factor 8 (Prp8), was shown to stably associate with U5 snRNA<sup>47</sup>, and the *Prp8* gene of *S. cerevisiae* was shown to be required for pre-mRNA splicing<sup>48</sup>. In addition, the formation of the spliceosome and execution of splicing require many other proteins, particularly those of the yeast nineteen complex (NTC, known as the PRP19–CDC5L complex in mammals) and NTC-related (NTR) complex<sup>49,50</sup>.

Determination of the spliceosomal RNA map — the spatial placement of, and interactions between, the snRNAs and pre-mRNA — revealed that U1 and U2 snRNPs interact with the 5'SS and the BPS, respectively<sup>12</sup>, and that U2 snRNA is required for this interaction<sup>51,52</sup>. U5 snRNA appeared to select the 5'SS because a single nucleotide change in loop I of U5 snRNA allowed the selection of a cryptic 5'SS<sup>53</sup>; loop I specifically recognizes both 5' and 3' exon sequences<sup>54</sup>. The 5' end sequences of U2 snRNA form a duplex with 3' end sequences of U6 snRNA as part of the U4–U6 duplex in the tri-snRNP<sup>55</sup>. Crosslinking experiments identified

base-pairing interactions between pre-mRNA and snRNA and between snRNAs<sup>56,57</sup>. The ACAGA box of U6 snRNA forms a duplex with the 5'SS, and U5 and U6 snRNAs are both located in the splicing active site<sup>58–61</sup>. These and subsequent investigations have established a two-dimensional RNA map in the spliceosome<sup>44,62</sup>.

The conversion between the different spliceosomal complexes is driven by eight highly conserved, RNA-dependent ATPase/helicases<sup>44,63–65</sup> (FIG. 1c). Prp5, UAP56, Prp28 and Brr2 function in earlier steps leading to the formation of the B<sup>act</sup> complex. Unwinding of the U4–U6 duplex, which precedes the formation of the active spliceosome, is executed by Brr2 in yeast<sup>66</sup> or by the helicase U5-200kD (also known as SNRNP200) in humans<sup>67</sup>. By contrast, the DEAH-box ATPase/helicases Prp2, Prp16 and Prp22 have a crucial role during the catalytic cycle. Prp2, which associates with the spliceosome before and throughout step I of the splicing reaction<sup>68,69</sup>, activates the spliceosome through ATP hydrolysis<sup>70,71</sup>. Prp16 is indispensable for the step II reaction, which requires ATP binding and/or hydrolysis<sup>72,73</sup>. Prp22 is required for mRNA release<sup>74</sup>. Through these and subsequent studies, Brr2, Prp2 and Prp16 were established to be responsible for the B-to-B<sup>act</sup>, B<sup>act</sup>-to-B\* and C-to-C\* transitions, respectively (FIG. 1c). Prp16 and Prp22 are also responsible for the selection of alternative BPSs and 3'SSs, respectively<sup>75,76</sup>.

### Electron microscopy of spliceosomes

The highly dynamic nature of the spliceosome makes its crystallization practically impossible. X-ray-based investigations have been limited to individual components and subcomplexes of the spliceosome. Representative studies include the crystal structures of U1 snRNP<sup>77–79</sup>, U2 snRNP subcomplexes<sup>80–84</sup>, U4 snRNP core<sup>85</sup>, U6 snRNP subcomplexes<sup>86,87</sup>, a large fragment of Prp8 (REF. 88), Brr2 (REFS 89,90) and the splicing factor 3b (SF3b) complex<sup>91</sup>. In contrast to X-ray crystallography, electron microscopy performs better with complexes of larger molecular mass and is less dependent on conformational or compositional homogeneity. These advantages have made electron microscopy the only viable approach for structural determination of the intact spliceosome. For electron microscopy studies, spliceosomes were assembled *in vitro* on MS2-tagged pre-mRNA and successfully purified<sup>92,93</sup>. In the two decades before 2015, electron microscopy-based investigations have uncovered structural features of the spliceosome at various stages of the splicing reaction<sup>94</sup>. The mammalian spliceosome structures were elucidated for the A (REFS 95,96), B (REFS 97–99), B<sup>act</sup> (REF. 100), C (REFS 101,102) and P (REF. 103) complexes. The yeast structures include those of the B (REF. 104), B<sup>act</sup> (REF. 104), C (REFS 104,105) and ILS (REF. 106) complexes. These electron microscopy studies have moderate resolution, with the highest being around 20–29 Å, and only reveal general features of the spliceosome. The cryo-electron microscopy (cryo-EM) structure of the U4/U6.U5 tri-snRNP from *S. cerevisiae*, reported at 5.9 Å resolution<sup>107</sup>, allows identification of individual protein and RNA components but reveals few features for atomic modelling.

◀ Figure 1 | The pre-mRNA splicing reactions and the splicing cycle. **a** | Two steps of transesterification take place during pre-mRNA splicing. In step 1 (branching), the 2'-OH of the branch point sequence (BPS) adenine nucleotide attacks the phosphate of the guanine nucleotide at the 5' end of the 5' splice site (5'SS). In step 2 (ligation), the 3'-OH of the 3' end nucleotide of the 5' exon attacks the phosphate of the 5' end nucleotide of the 3' exon. **b** | Coordination of the catalytic metal ions before and after the first step of transesterification. The upper and lower panels represent the B<sup>act</sup> and C complexes, respectively. The two metals, designated as M1 and M2, are bound mainly by phosphates from U6 small nuclear RNA. In the first step of transesterification, M2 activates the nucleophile, whereas M1 stabilizes the leaving group. In the second step of transesterification, M1 activates the nucleophile, whereas M2 stabilizes the leaving group. **c** | Assembly and activation of the yeast spliceosome and the complete splicing-reaction cycle. The 5'SS, BPS and 3'SS are first recognized by the U1 small nuclear ribonucleoprotein (snRNP), splicing factor 1 (SF1; also known as branchpoint-bridging protein) and U2AF, respectively, forming an early spliceosome (known as the E complex). SF1 is displaced by the U2 snRNP to form the pre-spliceosome (A complex), which associates with the U4/U6.U5 tri-snRNP to assemble into the pre-catalytic spliceosome (B complex). The B complex represents the first fully assembled spliceosome. There are at least six additional distinct spliceosome complexes: B<sup>act</sup>, B\*, C, C\*, P and the intron lariat spliceosome (ILS). Each complex has a unique composition, and conversion between complexes is driven by highly conserved RNA-dependent ATPase/helicases (in bold). Notably, a spliceosomal complex can have distinct conformational states, which may also differ in composition. For example, the B and ILS complexes each have at least two distinct conformations.

Mechanistic understanding of pre-mRNA splicing has taken a quantum leap through recent cryo-EM studies of the spliceosome at atomic and near-atomic resolutions<sup>108–119</sup> (Supplementary information S1 (table)). The overall design principles of the spliceosome were revealed by the structure of the ILS complex from *Schizosaccharomyces pombe* at an average resolution of 3.6 Å<sup>108,109</sup> (FIG. 2a) and confirmed by subsequent structures of the *S. cerevisiae* spliceosome at different stages of the splicing cycle<sup>113–118</sup>. These principles govern organization of the splicing active site, coordination of the catalytic metal ions, recognition of the pre-mRNA sequences, assembly of the snRNPs and distribution of the snRNA elements, and the central role of protein components in shaping the splicing active site and the overall spliceosome architecture. We now have an atomic view of the spliceosomal RNA map, which includes three snRNAs (U2, U5 and U6) and the pre-mRNA (FIG. 2b,c). The bulk of the yeast splicing cycle has been visualized in atomic detail through structural elucidation of the fission or budding yeast B<sup>act</sup> complex<sup>113,118</sup>, C complex<sup>114,115</sup>, C\* complex<sup>116,117</sup> and ILS complex<sup>108,109</sup> (FIG. 2a). In the following discussion, the focus is on the mechanistic information revealed by these structures.

### Organization of the splicing active site

The splicing active site comprises the intramolecular stem-loop (ISL) of U6 snRNA, helix I of the U2–U6 duplex, the associated Mg<sup>2+</sup> ions and loop I of U5 snRNA (FIGS 2b, 3a). The central element is a stretch of 35 nucleotides from U6 snRNA, including the ISL and the 10 preceding nucleotides that form helix I. Despite known information on the base-pairing specifics, atomic modelling of these 35 nucleotides proved to be the most challenging task in building the entire spliceosome<sup>108,109</sup>. The final detailed structure of these nucleotides is distinct from any other known RNA element (FIG. 3b). The backbone of these 35 nucleotides zig-zags in a rectangular box that measures approximately 20 × 25 × 50 Å<sup>3</sup>, making five 180-degree turns. The close contacts among the RNA nucleotides are likely stabilized in part by two metal ions that were interpreted to be magnesium<sup>108,109</sup>. These two structural metals are buried in the centre of the ISL and presumably neutralize the negative charges of the RNA phosphates (FIG. 3c). The overall conformation of the splicing active site observed in the *S. pombe* ILS complex<sup>108,109</sup>, including the exact configuration of the ISL, remains largely unchanged in subsequent structures of the various *S. cerevisiae* spliceosomal complexes<sup>113–118</sup>.

Helix I is interrupted by an unpaired bulge of three nucleotides in U2 snRNA<sup>55</sup> (FIG. 2c). Within helix I, the nucleobases in the last three nucleotides AGC of U6 snRNA also make non-Watson–Crick hydrogen bonds to three nucleotides within U6 snRNA, forming a characteristic triple helix known as the catalytic triplex<sup>120</sup> (FIGS 2c, 3d). Loop I, which anchors both 5' and 3' exons<sup>54</sup>, is placed in close proximity to the ISL. The ligated exon is already released in the *S. pombe* ILS complex<sup>108,109</sup>. By contrast, the 5' exon remains bound to

loop I in the structures of the *S. cerevisiae* B<sup>act</sup>, C and C\* complexes<sup>113–118</sup>, with three nucleotides from the 3' end of the 5' exon forming a duplex with loop I<sup>54</sup>.

### Coordination of the catalytic metal ions

The two catalytic metal ions, known as M1 and M2, are coordinated by the backbone phosphates of U6 snRNA and the pre-mRNA<sup>29</sup> (FIG. 1b). The 3'-OH of the nucleotide at the 3' end of the 5' exon is stabilized by M1 as the leaving group in the first transesterification and activated by the same metal as the nucleophile in the second transesterification. The 2'-OH of an adenine nucleotide in the BPS is activated by M2 as the nucleophile in the first transesterification. Following the first reaction, the lariat junction is dislocated from the centre of the active site by approximately 20 Å to make room for the intron–3' exon sequences. The leaving group in the second reaction — the 3'-OH of the nucleotide at the 3' end of the intron — is stabilized by M2. Notably, the displacement distance of 20 Å for the pre-mRNA intron lariat is conserved in the group II intron lariat<sup>121</sup>.

In the activated but catalytically dormant B<sup>act</sup> complex<sup>113</sup>, two putative catalytic metals are already loaded onto the active site, but neither is properly coordinated (FIG. 3e). In the C complex<sup>114</sup>, M1 is coordinated by the phosphates of G78 and U80 of U6 snRNA, the 3'-OH of the 3' end nucleotide of the 5' exon, and the phosphate of the 5' end nucleotide of the 5'SS. M2 is bound by the phosphates of A59 and U80 of U6 snRNA but is approximately 6 Å away from the nucleophile. In the C\* complex<sup>116</sup>, metal coordination is very similar to that of the C complex<sup>114</sup>, except that M1 is only coordinated by three ligands and separated from the fourth one — the phosphate of the 5' end nucleotide of the 5'SS — by a distance of approximately 21 Å (FIG. 3e). Coordination of M2 remains the same as that in the C complex<sup>114</sup>. In the *S. pombe* ILS complex<sup>108,109</sup>, M1 is bound by G48, G66 and U68, which correspond to G60, G78 and U80, respectively, of the *S. cerevisiae* U6 snRNA. M2 is recognized by A47, G48 and U68 (A59, G60 and U80 in *S. cerevisiae*).

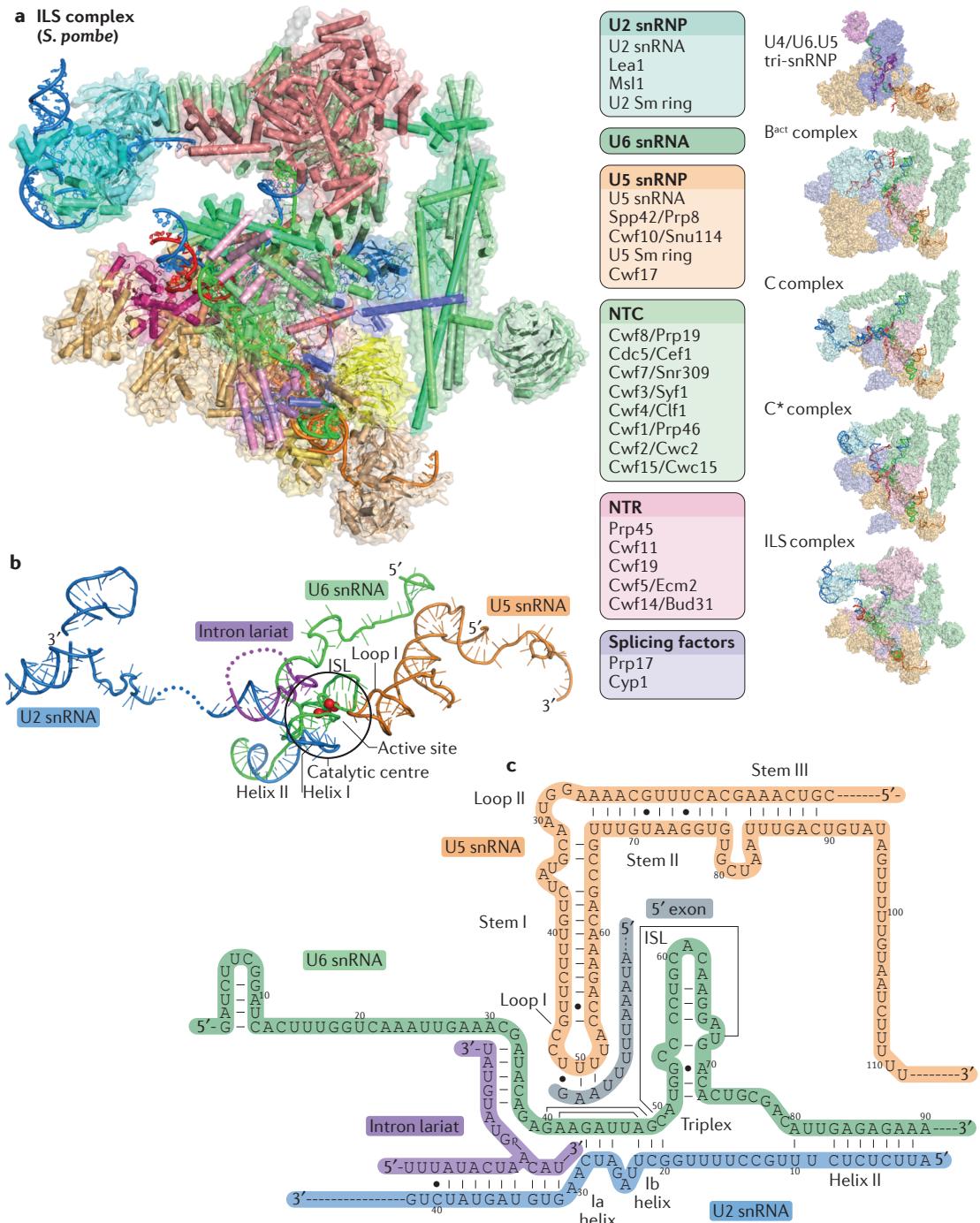
Consistent with previous analysis<sup>28,29</sup>, U80 of U6 snRNA has a central role in coordinating both M1 and M2 in the two transesterification steps. Coordination of M1 is fairly static because the ion is both a stabilizer and an activator for the same group — the 3'-OH of the nucleotide at the 3' end of the 5' exon, which serves as the leaving group in the first transesterification and the nucleophile in the second transesterification. The catalytic metals are presumably Mg<sup>2+</sup> in these cryo-EM structures, in part because the pattern of metal coordination is consistent with that for Mg<sup>2+</sup>. In addition, the Mg<sup>2+</sup> assignment is supported by structural comparison with the group II intron<sup>109</sup> and the phosphorothioate substitution experiments in the spliceosome<sup>28,29</sup>. Nonetheless, the reported local resolutions at the active site are insufficient to conclusively identify the metals.

### Prp8 and its three catalytic motifs

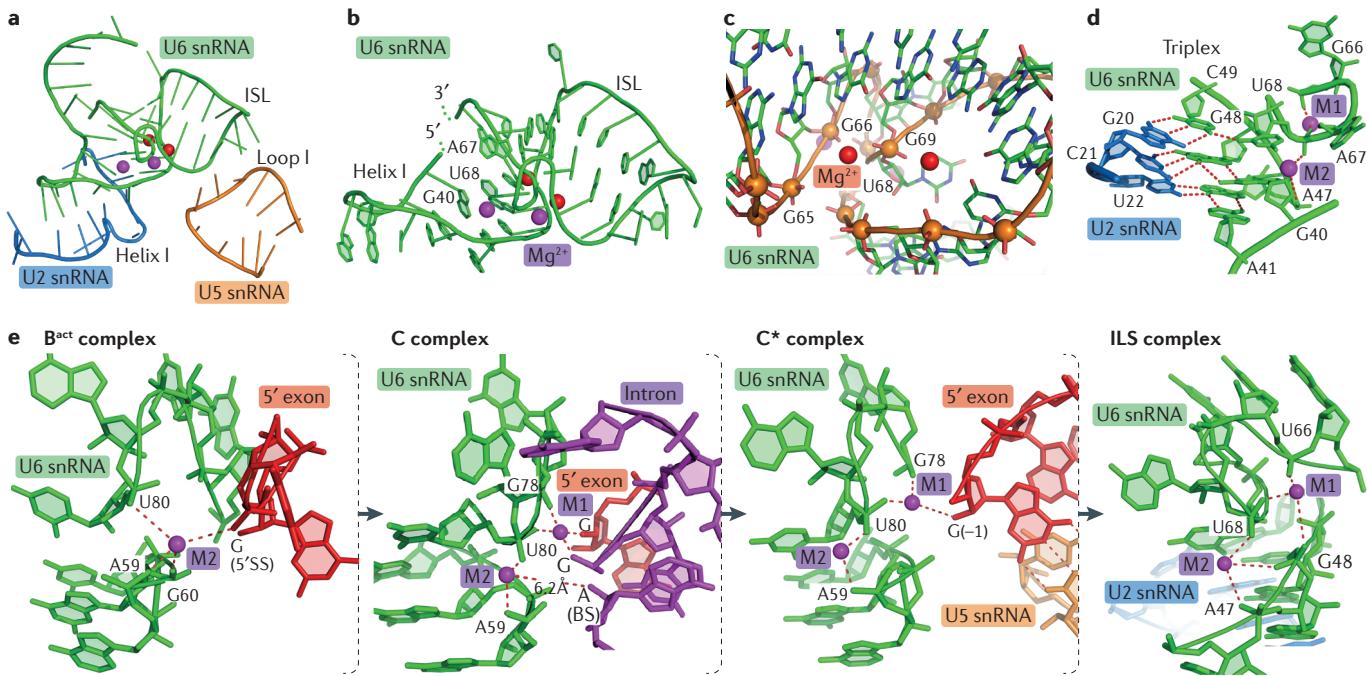
A central component of U5 snRNP, Prp8 (*Spp42* in *S. pombe*), is the largest and most highly conserved spliceosomal protein<sup>122</sup>. Prp8 comprises seven domains:

#### Lariat junction

The three-way junction in which a 2' to 5' linkage is formed between the ribose 2'-OH of an adenosine nucleotide in the intron branch site and the phosphate of a guanine nucleotide at the 5' end of the intron 5' splice site.



**Figure 2 | Atomic structures of the spliceosome.** **a |** The structure of the intron lariat spliceosome (ILS) at an overall resolution of 3.6 Å (REFS 108,109) is shown in the left panel. Colour coding for the various spliceosomal subcomplexes and components is indicated. The structures of the U4/U6.U5 tri-snRNP, the B<sup>act</sup> complex<sup>113,118</sup>, the C complex<sup>114,115</sup> and the C\* complex<sup>116,117</sup> of *Saccharomyces cerevisiae* are shown in smaller panels on the right, along with the structure of the *S. pombe* ILS complex for comparison. The orientation and the scale of the spliceosome, as defined by U5 small nuclear RNA (snRNA), are maintained exactly the same for all the complexes. **b |** Spatial arrangement of the RNA elements from the *S. pombe* ILS complex<sup>108,109</sup>. U2 snRNA, U5 snRNA, U6 snRNA and the intron lariat are coloured blue, orange, green and purple, respectively. This colour scheme is preserved in all structural figures throughout this Review unless otherwise indicated. The metal ions in the active site are shown as red spheres. **c |** The observed base-pairing specifics of the RNA map of the *S. pombe* ILS complex. Non-standard Watson–Crick interactions are marked by black dots. The 5' exon sequences (shown in grey) are absent in the ILS complex, but their interactions with loop I of U5 snRNA in other spliceosomal complexes are nonetheless indicated in this diagram. All essential features of the RNA map observed in the *S. pombe* ILS complex were confirmed by subsequent structures of the B<sup>act</sup>, C and C\* complexes from *S. cerevisiae*<sup>113–118</sup>. Part b is reproduced with permission from REF. 109, AAAS.



**Figure 3 | Organization of the splicing active site and the catalytic metal ions.** **a** | Structure of the splicing active site in the *Schizosaccharomyces pombe* intron lariat spliceosome (ILS)<sup>108,109</sup>. In addition to the intramolecular stem-loop (ISL) and the associated metals (spheres), the active site also includes helix I of the U2–U6 duplex and loop I of U5 small nuclear RNA (snRNA), which recognizes the 5' exon. **b** | Structure of the ISL and the preceding 10 nucleotides of U6 snRNA from the *S. pombe* ILS complex<sup>108,109</sup>. The overall conformation is preserved in subsequent structures of the B<sup>act</sup>, C, and C\* complexes from *Saccharomyces cerevisiae*<sup>113–117</sup>. **c** | Two structural metal ions in the active site of the *S. pombe* ILS complex stabilize the RNA fold by neutralizing negative charges of the phosphodiester backbone<sup>108,109</sup>. The phosphorus atoms are highlighted as large orange spheres. The location and coordination of these two metals are highly conserved in the other *S. cerevisiae* spliceosomal complexes<sup>113–118</sup>. **d** | A close-up view of the catalytic

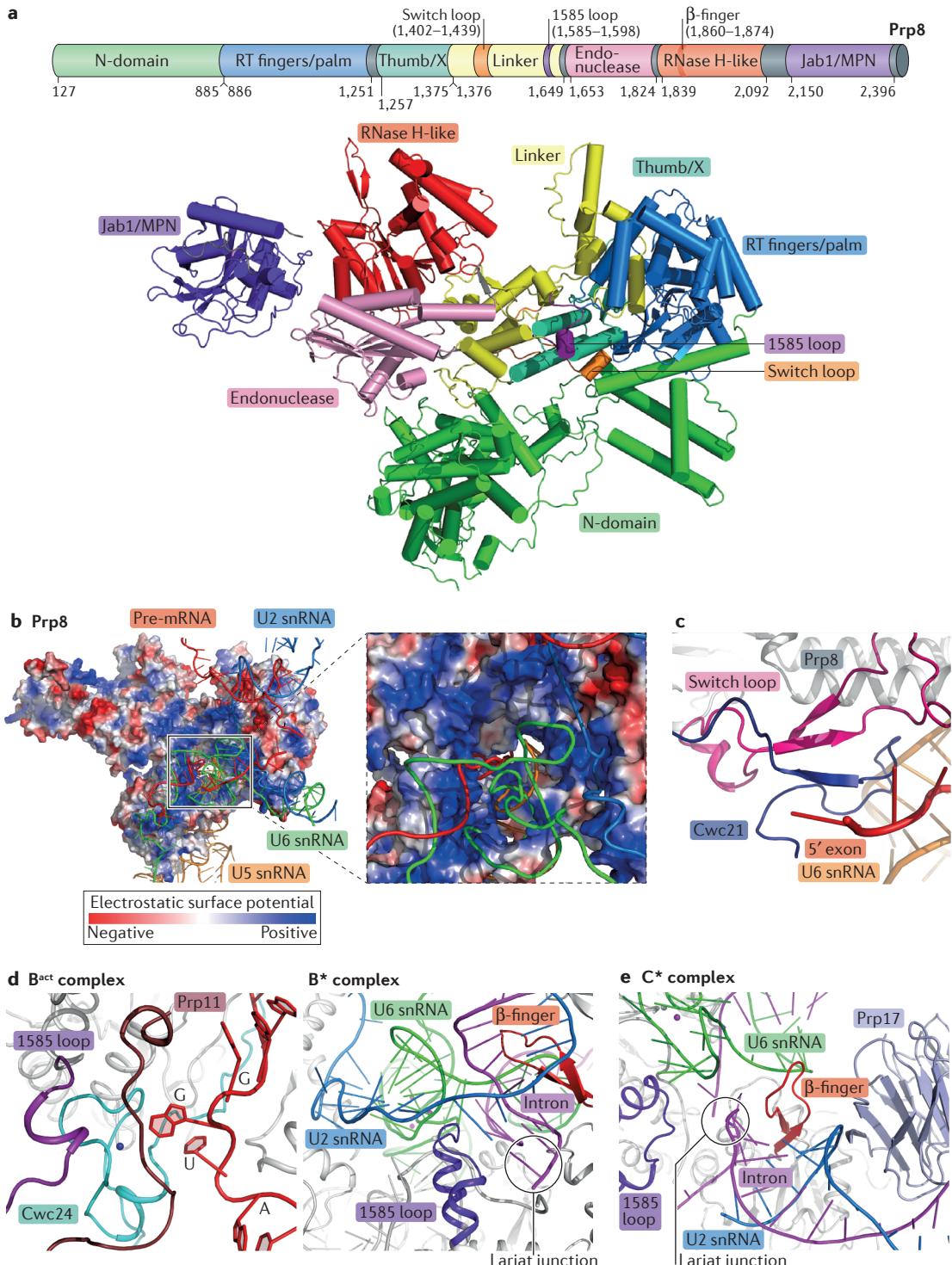
triplex in the active site of the *S. pombe* ILS complex. Hydrogen bonds are represented by red dashed lines. **e** | In the *S. cerevisiae* B<sup>act</sup> complex<sup>113</sup>, two metals are bound in the active site but neither is properly coordinated. Only one metal ion, which may correspond to M2, is shown. For comparison, the overall orientation of the active site in panels **a–d** is approximately the same. In the *S. cerevisiae* C complex<sup>114</sup>, M1 is coordinated by G78 and U80 of U6 snRNA, the 3' end nucleotide of the 5' exon, and the 5' end nucleotide of the 5'SS. M2 is recognized by A59 and U80 of U6 snRNA but is located 6.2 Å away from the nucleophile. In the *S. cerevisiae* C\* complex<sup>116</sup>, metal coordination is similar to that in the C complex, except that M1 is no longer bound by the 5' end nucleotide of the 5'SS. In the *S. pombe* ILS complex<sup>108,109</sup>, M1 is bound by the nucleotides G48, G66 and U68 of the U6 snRNA, which correspond to G60, G78 and U80, respectively, in *S. cerevisiae*. M2 is recognized by A47, G48 and U68, which correspond to A59, G60 and U80 in *S. cerevisiae*.

an amino-terminal domain (N-domain), reverse transcriptase (RT) fingers/palm, thumb/X, linker, endonuclease-like, RNase H-like and Jab1/MPN (FIG. 4a). The core of Prp8 and Spp42, also known as the large domain<sup>88</sup>, contains four domains: RT fingers/palm, thumb/X, linker and endonuclease-like. The core has a compact structure and, except for a few surface elements, maintains a rigid conformation throughout spliceosome assembly, activation and catalysis. Relative to the core, the N-domain is rotated by approximately 30 degrees during activation of the spliceosome<sup>110–112</sup>. The RNase H-like domain adopts quite different positions in the different spliceosomal complexes<sup>114</sup>. The Jab1/MPN domain, which regulates the function of Brr2 through direct interactions<sup>89</sup>, is highly mobile and has only been structurally captured with reasonable detail in the U4/U6/U5 tri-snRNP<sup>110,111</sup> and the B<sup>act</sup> complex<sup>113,118</sup>.

Prp8 and Spp42 harbour a positively charged cavity that functions as a central scaffold for splicing. The N-domain is chiefly responsible for the recognition of U5 snRNA<sup>108,109,123</sup>. The relative orientation of the N-domain with respect to the core remains unchanged

in the B<sup>act</sup>, C, C\* and ILS complexes. This arrangement yields a positively charged scaffold between the N-domain and the core; the scaffold accommodates the splicing active site<sup>108,109</sup> (FIG. 4b). Notably, the catalytic cavity is different from that predicted on the basis of the structure of the isolated Prp8 core<sup>88</sup>.

Three Prp8 structural motifs — the Switch loop, the 1585 loop and the β-finger — have important roles during splicing. The Switch loop stabilizes 5' exon binding by interacting with the splicing factor Cwc21, which directly recognizes the 5' exon (FIG. 4c). Compared with its position in the tri-snRNP<sup>110,111</sup> or the ILS complex<sup>108,109</sup>, where the 5' exon is no longer bound, the Switch loop is flipped by approximately 180 degrees in the B<sup>act</sup>, C or C\* complex to stabilize the bound 5' exon<sup>113–119</sup>. The 1585 loop may directly promote the splicing reaction<sup>124</sup>. In the B<sup>act</sup> complex<sup>113</sup>, the 1585 loop interacts with helix Ia of the U2–U6 duplex and contributes to pre-mRNA binding by closely associating with the elements of Prp11 and Cwc24 that recognize the 5' end guanine nucleotide of the 5'SS (FIG. 4d, left panel). In the C\* complex<sup>116,117</sup>, the 1585 loop directly contacts the ISL and the lariat junction, with its tip



**Figure 4 | Prp8 and its catalytic motifs.** **a** | Structure of Prp8 from *Saccharomyces cerevisiae*. The seven domains of Prp8 are colour-coded. The core comprises four domains (reverse transcriptase (RT) fingers/palm, thumb/X, linker and endonuclease-like) and maintains a rigid conformation throughout spliceosome assembly, activation and catalysis. **b** | A positively charged catalytic cavity is located between the amino-terminal (N)-domain and the core<sup>108,109</sup>. **c** | The Switch loop stabilizes the bound 5' exon sequences through direct interactions with Cwc21. **d** | The 1585 loop has a prominent role during the splicing reaction. In the B<sup>act</sup> complex<sup>113</sup>, the 1585 loop interacts with the U2–U6 small nuclear RNA (snRNA) duplex and stabilizes the 5' end nucleotide of the 5'SS through direct interactions with Cwc24 and Prp11 (left panel). In the B\* complex<sup>116</sup>, the 1585 loop binds directly to the intramolecular stem-loop of U6 (not shown) and the lariat junction, with its tip close to the 5' exon and the catalytic Mg<sup>2+</sup> ions (pink dots; right panel). **e** | In the C\* complex, the  $\beta$ -finger interacts with Prp17 near the lariat junction<sup>116</sup>. The  $\beta$ -finger and the 1585 loop are located on two sides of the lariat junction.

close to the 5' exon and the catalytic metals (FIG. 4d, right panel). The third catalytic motif is the  $\beta$ -finger<sup>125</sup> from the RNase H-like domain. In the C\* complex<sup>116,117</sup>, the  $\beta$ -finger interacts with the step II splicing factor Prp17, and the lariat junction is sandwiched between the  $\beta$ -finger and the 1585 loop (FIG. 4e). This arrangement may stabilize the active site conformation for exon ligation. The  $\beta$ -finger remains unengaged in the B<sup>act</sup> complex<sup>113</sup> and contacts the intron–U2 duplex in the C complex<sup>114,115</sup>.

### Recognition of the intron sequences

The 5'SS and the BPS are recognized through duplex formation with U6 and U2 snRNAs, respectively, in the structures of the various yeast spliceosomes<sup>108,109,113–118</sup>.

In all cases, five consecutive nucleotides of the intron, including the last three nucleotides UGU from the 5'SS, form Watson–Crick interactions with the nucleotides AUACA of U6 snRNA (FIG. 2c). Notably, only three nucleotides of the U6 snRNA ACAGA box are directly involved in 5'SS recognition. This base-pairing arrangement is likely necessary for both catalytic steps of splicing. In step I, the guanine nucleotide at the 5' end of the 5'SS is flipped out of base-pairing register (FIG. 5a, top left), thereby disallowing duplex formation between the GUA nucleotides of the 5'SS and U6 snRNA. In step II, delivery of the 3'SS–3' exon sequences into the active site requires the dislocation of the lariat junction away from the centre of the active site, which requires that the U6 sequences between those recognizing the 5'SS and those forming helix I are sufficiently flexible.

The first three nucleotides of the 5'SS (GUA) are differentially recognized by surrounding proteins in different spliceosomal complexes. In the *S. cerevisiae* B<sup>act</sup> complex<sup>113</sup>, the guanine nucleotide is recognized by three hydrogen bonds from Cwc24 and Prp11, with its base sandwiched between the aromatic side chains of Tyr155 and Phe161 of Cwc24 (FIG. 5a, top right). The uridine nucleotide also interacts with Cwc24 through hydrogen bonds, whereas the adenine nucleotide stacks against A51 of U6 snRNA. In the C complex<sup>114,115</sup>, the guanine nucleotide is bonded to the BPS, with its ribose accepting a hydrogen bond from Arg4 of Yju2 (also known as Cwc16; FIG. 5a, bottom left). The uridine nucleotide interacts with U2 snRNA and Arg4 of Yju2 through hydrogen bonds, and the adenine nucleotide remains stacked with A51 of U6 snRNA. In the C\* complex<sup>116,117</sup>, the guanine nucleotide is hosted by the 1585 loop and the RNase H-like domain of Prp8, whereas the uridine and adenine nucleotides are close to Prp8 and U6 snRNA (FIG. 5a, bottom right).

In the B<sup>act</sup>, C, C\* and ILS complexes<sup>108,109,113–118</sup>, the consensus BPS UACUAAC and the preceding nucleotides are similarly recognized through duplex formation with U2 snRNA. In the B<sup>act</sup> complex<sup>113,118</sup>, the intron–U2 duplex is held in place mainly by the SF3b proteins Hsh155 and Rds3 and the SF3a protein Prp11 and to a lesser extent by Cus1 and the SF3b protein Ysf3 (also known as RDS3 complex subunit 10). The protein–RNA interactions are mostly hydrogen bonds directed at the RNA phosphate backbone and stabilized by the positive

electrostatic surface potential of the proteins. Duplex formation between the BPS UACUAAC (lower case marks the adenosine nucleophile) and the U2 snRNA sequence GUAGUA leaves the adenosine nucleophile unpaired and flipped out of register (FIG. 5b). The adenine base is sandwiched by two residues from Hsh155; the ribose, the base and the phosphate each accept a hydrogen bond from Hsh155 (FIG. 5c). This pattern of recognition is analogous to that of the 5' end guanine nucleotide in the 5'SS<sup>114</sup>. The catalytic dormancy of the B<sup>act</sup> complex is ensured by the structural feature that the nucleophile for the first transesterification — the ribose 2'-OH — is separated from the scissile bond by approximately 49 Å, with the two splicing factors Cwc24 and Prp11 located in between<sup>113,118</sup>.

The proteins that hold the intron–U2 duplex in the B<sup>act</sup> complex are replaced in the C complex by Prp8, Isy1 and the splicing factors Cwc25 and Yju2 (REFS 114,115). Owing to the formation of the intron lariat, the adenosine nucleophile is caged by three surrounding intron nucleotides and the GU nucleotides at the 5' end of the 5'SS (FIG. 5d). In the C\* complex<sup>116,117</sup>, the intron–U2 duplex is recognized by Cef1, Prp8 and the splicing factor Prp17 (also known as Cdc40). Unlike in the B<sup>act</sup> or C complex, in the C\* complex, the adenosine nucleophile is no longer recognized by surrounding elements. The lariat junction is located close to the RNase H-like domain of Prp8. In the *S. pombe* ILS complex<sup>108,109</sup>, the adenine base is uncoordinated by the protein components; the lariat junction contacts Cwf19, a homologue of the *S. cerevisiae* debranching-enzyme cofactor Drn1 (REF. 126).

In the B<sup>act</sup>, C and C\* complexes<sup>113–118</sup>, five consecutive nucleotides downstream of the 5'SS closely associate with the NTR component Cwc2 (FIG. 5e). Cwc2 may help orient the 5'SS into the active site through interactions with downstream intron sequences<sup>127</sup>. In the B<sup>act</sup> complex<sup>113,118</sup>, the pyrimidine-rich sequences downstream of the BPS are recognized by Hsh155, Rds3 and the retention and splicing (RES) complex (FIG. 5f). Human HSH155 (also known as SF3B1) is frequently mutated in several malignancies<sup>128</sup>. Most of the mutated residues are at the HEAT repeats 3 and 4 of HSH155, which bind to the pyrimidine-rich sequences but not to the BPS. The mutations that abolish recognition of the BPS are presumably lethal; the milder mutations that sabotage binding of the pyrimidine-rich sequences might compromise the splicing fidelity by promoting selection of alternative BPSs, thus contributing to oncogenesis<sup>129,130</sup>. In the C complex<sup>114,115</sup>, these pyrimidine-rich sequences are recognized by the fingers/palm, thumb/X and linker domains of Prp8.

### Recognition of the exon sequences

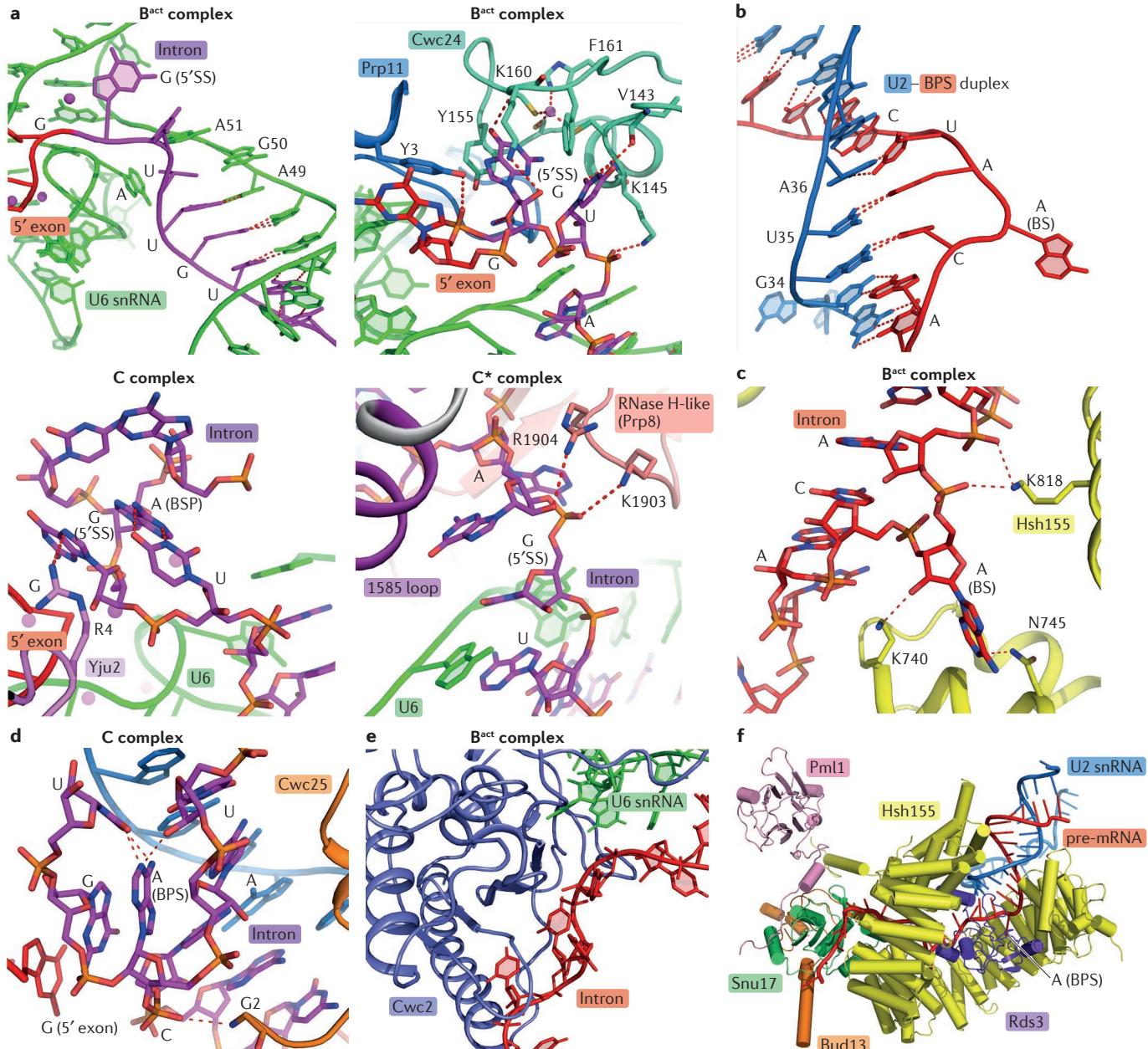
In the B<sup>act</sup>, C and C\* complexes<sup>113–118</sup>, the 5' exon is anchored identically in the active site, with three consecutive nucleotides (modelled as A<sub>3</sub>A<sub>2</sub>G<sub>1</sub>) at the 3' end forming a duplex with U<sub>96</sub>U<sub>97</sub>U<sub>98</sub> of the U5 snRNA loop I (Supplementary information S2 (figure)). The phosphodiester backbone of the three 5' exon nucleotides is in close contact with the N-domain of

#### HEAT repeats

Repetitive motifs frequently found in signalling proteins; each HEAT repeat comprises a pair of anti-parallel  $\alpha$ -helices, linked by a short loop.

Prp8 in all three complexes. The N-terminal three residues Glu3-Arg4-Lys5 of Yju2 in the C complex<sup>114,115</sup> and Lys1637 of the Prp8 linker domain in the C\* complex<sup>116,117</sup> interact directly with the backbone of these three nucleotides. Seven preceding nucleotides are also recognized similarly in all three spliceosomal complexes.

Nucleotides -10 to -8 of the exon are bound between Cwc21 and the Prp8 N-domain, with Cwc21 being oriented by the Prp8 Switch loop; nucleotides -7 to -4 interact with the N-domain and the linker domain of Prp8 (Supplementary information S2 (figure)). In the C and C\* complexes<sup>114-117</sup>, nucleotides -13 to -11 are close



**Figure 5 | Recognition of the intron sequences.** **a** | Recognition of the 5' splice site (5'SS). In all structurally characterized spliceosomal complexes, the UGU nucleotides of the 5'SS and two ensuing intron nucleotides form a duplex with U6 small nuclear RNA (snRNA). In the B<sup>act</sup> complex<sup>113</sup>, the guanine nucleotide at the 5' end of the 5'SS is flipped out of base-pairing register and recognized by Cwc24 and Prp11 (upper panels). After branching, the guanine nucleotide and the ensuing nucleotides are bound by Yju2 and by snRNAs in the C complex and by the 1585 loop and RNase H-like domain of Prp8 in the C\* complex (lower panels). **b** | Duplex formation between the branch point sequence (BPS) and U2 snRNA. Two bases of the intron are unpaired. Similar to the 5' end guanine nucleotide of the 5'SS,

the adenine nucleophile of the BPS is flipped out of base-pairing register with the BPS-U2 duplex. **c** | Recognition of the adenine nucleophile of the BPS in the B<sup>act</sup> complex<sup>113</sup>. The adenine nucleophile interacts closely with residues of Hsh155. **d** | Recognition of the adenine nucleophile of the BPS in the C complex<sup>114</sup>. The base of the adenine nucleophile is surrounded by intron nucleotides and the GU nucleotides at the 5' end of the 5'SS. **e** | Recognition of intron sequences downstream of the 5'SS by Cwc2 in the B<sup>act</sup> complex<sup>113</sup>. These structural features are preserved in the C and C\* complexes<sup>114,116</sup>. **f** | In the B<sup>act</sup> complex, the pyrimidine-rich sequences downstream of the BPS are recognized by Hsh155, Rds3 and the retention and splicing complex proteins Pml1, Snu17 and Bud13 (REF. 113).

to Cwc21, Cwc22 and the Prp8 N-domain. In contrast to the observed recognition of the 5'SS and the 5' exon, how the 3'SS and the 3' exon are recognized and selected during splicing remains to be structurally characterized. U6 snRNA is thought to be involved in recognizing the 3'SS through duplex formation<sup>131–133</sup>, and the spacing between the BPS and 3'SS matters in this process<sup>134,135</sup>. The step II factor Slu7 may have an essential role in the selection of the 3'SS<sup>136–139</sup>.

### The core of the spliceosome is rigid

Despite frequent remodelling of the spliceosome during splicing, the structural core around the active site in the B<sup>act</sup>, C, C\* and ILS complexes is rigid. This core comprises U5 and U6 snRNA, a portion of U2 snRNA and at least 17 associated proteins. The entire U6 snRNA, the bulk of U5 snRNA and a stretch of 30 consecutive nucleotides of the 5' end of U2 snRNA are perfectly aligned in these spliceosomal complexes (FIG. 6a). This portion of U2 snRNA forms helices I and II with U6 snRNA. Unlike U5 snRNA, which forms several internal stems and loops, U6 snRNA adopts an extended conformation that is maintained by spliceosomal proteins. These proteins include nine of U5 snRNP (Spp42 (*S. pombe*) or Prp8 (*S. cerevisiae*), Cwf10/Snu114 and the heptameric Sm complex), two of NTC (Cdc5/Cef1 and Syf2) and six of NTR (Cwf1/Prp46, Cwf2/Cwc2, Cwf5/Ecm2, Cwf14/Bud31, Cwf15/Cwc15 and Prp45) (FIG. 6b). The structurally resolved portions of these 17 proteins remain largely static in the B<sup>act</sup>, C, C\* and ILS complexes, which, given the overall compositional and conformational variability of the spliceosome, is unanticipated and has important ramifications for understanding the mechanism of splicing.

Of the five snRNPs, only U5 is well represented in the core and serves as the scaffold in all spliceosomal complexes. This structural core appears to be nucleated from a rigid centre that is already present in the U4/U6.U5 tri-snRNP and includes the bulk of U5 snRNA, the N-domain of Spp42/Prp8 and Cwf10/Snu114 (REF. 110). Notably, components of the NTC and NTR heavily contribute to the core by closely interacting with U5 snRNP and among themselves. These proteins have an essential role in shaping the splicing active site. Among all three snRNAs in these spliceosomal complexes, the only mobile elements are the U2 sequences that are downstream of nucleotide 30 (FIG. 6c). A well-documented function of this mobile U2 element is to bind and deliver the BPS to the active site. During the B<sup>act</sup>-to-B\* transition, the duplex between this element and the BPS is translocated by nearly 50 Å, thereby bringing the nucleophile into close proximity with the leaving group for the first-step transesterification. During the C-to-C\* transition, the lariat junction, which remains duplexed with this element, is moved away from the active site by approximately 20 Å to accommodate the 3'SS. Notably, the 5' and 3' ends of all three snRNAs are placed away from the active site, usually anchored by spliceosomal proteins<sup>109</sup>. For example, the 3' end sequences of U2 and U6 are bound by the Sm heptamer and by Syf2, respectively.

#### HAT repeats

Repetitive motifs implicated in RNA processing; each HAT repeat comprises two α-helices.

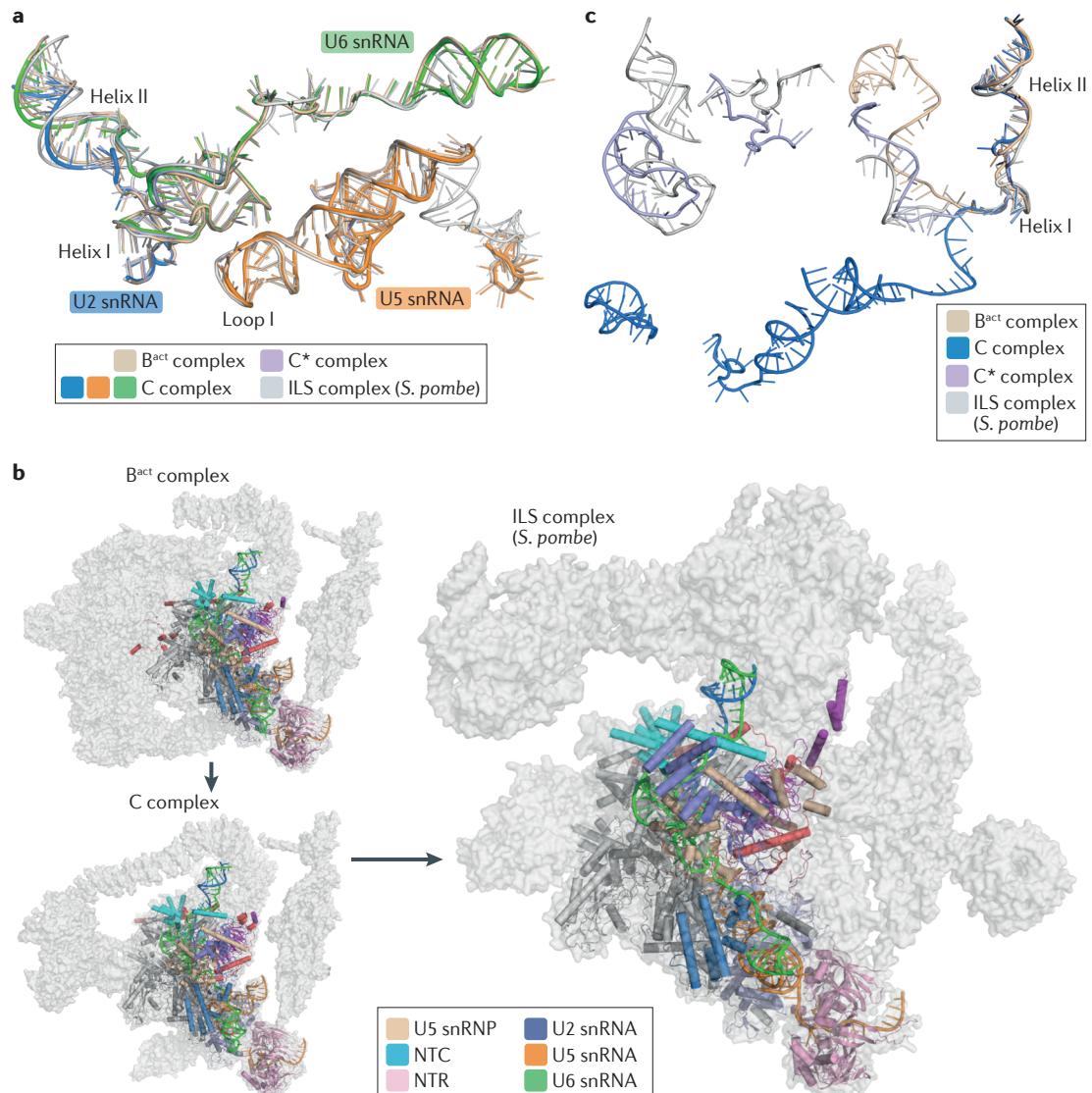
### Distinct classes of NTC and NTR proteins

The spliceosome must maintain structural stability around the active site while allowing sufficient flexibility so that pre-mRNAs of varying lengths and sequences can be bound, spliced and released<sup>140</sup>. The stable yet flexible organization of the spliceosome is ensured by myriad inter-molecular interactions. Among the 37 proteins and four RNA molecules of the *S. pombe* ILS complex, there are more than 100 pairs of bi-molecular interactions<sup>108,109</sup>. For example, each of 14 spliceosomal components of the *S. pombe* ILS complex interacts with five or more other components. Spp42 interacts with 12 other spliceosomal components and is the only protein that simultaneously binds to all three snRNAs and the pre-mRNA.

Two classes of components help to ensure the structural integrity and functional flexibility of the spliceosome: the rope-like proteins, which glue multiple components together, and the spring-like proteins, which allow large conformational shifts. All these proteins are components of the NTC and NTR, which are recruited to the B<sup>act</sup> complex following the dissociation of U1 and U4 snRNPs<sup>50,104,141</sup>. The rope-like proteins are exemplified by the NTR protein Prp45 and the NTC protein Cdc5/Cef1 (Supplementary information S3 (figure)). Prp45 spans more than 150 Å and interacts with at least nine spliceosomal components. Extended sequences in each of these rope-like proteins are likely intrinsically disordered in isolation and assume a specific conformation only upon association with other spliceosomal components. The spring-like proteins are represented by the NTC proteins Syf1 and Syf3 (also known as Cwf4)/Clf1 (Supplementary information S3 (figure)), which contain multiple HAT repeats. The inherent elasticity of these superhelical proteins may enable the spliceosome to bend and stretch in order to facilitate the splicing reaction. In the C-to-C\* transition, U2 snRNP is translocated by 60–100 Å, resulting in alteration of the overall appearance of the spliceosome. Together, these proteins and the Prp19 tetramer constitute a binding platform for other spliceosomal components (Supplementary information S3 (figure)).

### The roles of splicing factors

Splicing factors regulate the splicing reaction at least in part by stabilizing the conformation of the active site. In the B<sup>act</sup> complex<sup>113,118,142</sup>, the splicing factor Cwc24 helps maintain catalytic dormancy by insulating the 5' end nucleotide of the 5'SS, whereas Cwc21 stabilizes the binding of the 5' exon<sup>143</sup>. The role of Cwc21 is preserved in the C and C\* complexes and likely in the B\* and P complexes as well. The step I factors Cwc25 and Yju2 are both located in the centre of the active site of the C complex<sup>114,115</sup> (FIG. 7a). The N-terminal sequences of both Cwc25 and Yju2 are inserted into the active site; Cwc25 interacts with the intron–U2 duplex, ISL, helix I and a β-sheet domain of Yju2; Yju2 binds to the 5' exon, a portion of the intron–U2 duplex and both the N-domain and the core of Prp8. An α-helix from Yju2 intercalates between the intron–U2 duplex and the α1 and α2 helices of Isy1. These interactions place the nucleophile close to the scissile phosphodiester bond<sup>144–146</sup>.



**Figure 6 | A shared rigid core of the various spliceosomal complexes.** **a** | The overall conformation of U5 and U6 small nuclear RNAs (snRNAs) and part of U2 snRNA remains largely the same in the B<sup>act</sup>, C, C\* and intron lariat spliceosome (ILS) complexes, as shown here in an overlay of these RNA elements. For clarity, only the RNA elements from the C complex are colour-coded; for each of the other three complexes, the RNA elements are shown in the same colour. **b** | Approximately 20 protein and snRNA components constitute a rigid core of the spliceosome. The structures and relative orientations of these components remain largely unchanged in the B<sup>act</sup>, C, C\* and ILS complexes. Shown is a comparison of these components between the *Saccharomyces cerevisiae* B<sup>act</sup> and C complexes (left panels) and the *Schizosaccharomyces pombe* ILS complex (right panel). The components include ten U5 snRNP components (U5 snRNA, SPP42 (*S. pombe*)/Prp8 (*S. cerevisiae*), Cwf10/Snu114 and the heptameric Sm ring), U6 snRNA, a portion of U2 snRNA, the NTC proteins Cdc5/Cef1 and Syf2, and six NTR proteins (Prp45, Cwf5/Ecm2, Cwf1/Prp46, Cwf2/Cwc2, Cwf15/Cwc15 and Cwf14/Bud31). **c** | The U2 snRNA sequences downstream of nucleotide 30 are mobile during the two steps of transterification. Although the 5' end 30 nucleotides of U2 snRNAs are well aligned, the spatial placements of the ensuing nucleotides are vastly different in the yeast B<sup>act</sup>, C, C\* and ILS complexes.

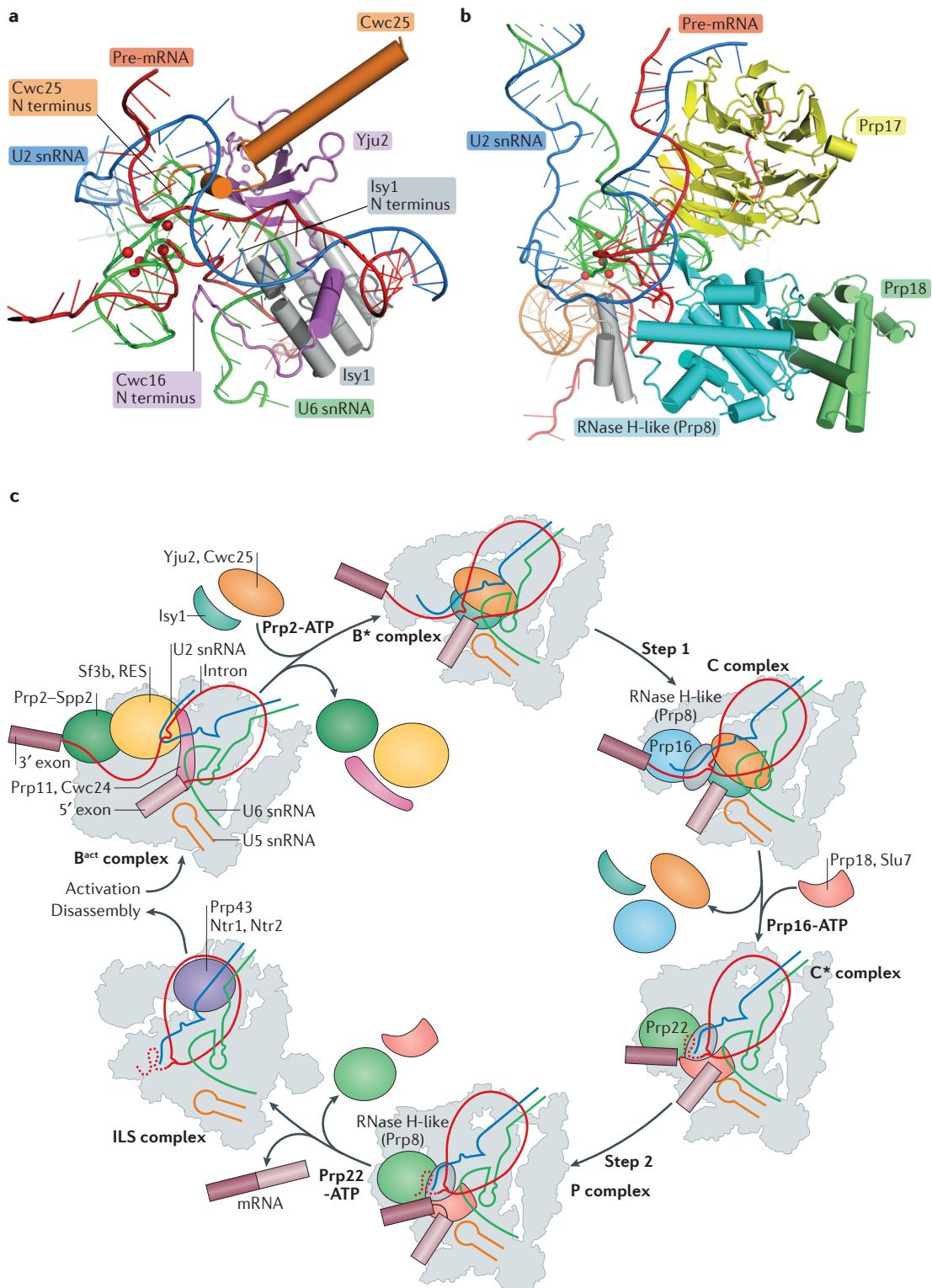
**WD40 domain**  
Also known as β-propeller or WD repeats. A type of circular solenoid protein domain that consists of 5–8 short structural motifs of approximately 40 amino acids each, often terminating in a tryptophan-aspartic acid (WD) dipeptide.

The step II splicing factors Prp17, Prp18 and Slu7 facilitate exon ligation<sup>139,147–149</sup>. In the C\* complex<sup>116,117</sup>, both Prp17 and Prp18 are intimately involved in the formation of the active site (FIG. 7b). Prp18 directly interacts with the RNase H-like domain of Prp8. Although the RNase H-like domain remains bound to the intron–U2 duplex in the C-to-C\* transition, its location is markedly shifted because of spliceosomal remodelling. Consequently, the β-finger is placed close to the lariat junction and directly

contacts Prp17. The WD40 domain of Prp17, which is positioned between the BPS–U2 duplex and the 5'SS–U6 duplex, is similarly shifted during the C-to-C\* transition.

#### RNP remodelling by the ATPase/helicases

The available structures provide insights into how the RNA-dependent ATPase/helicases remodel the spliceosome. In the yeast U4/U6.U5 tri-snRNP<sup>110,111</sup>, the ATPase/helicase Brr2 constitutes one arm of the



**Figure 7 | Step I and II splicing factors facilitate transesterifications by participating in the formation of the active site.** **a** | The step I splicing factors Cwc25 and Yju2 support the conformation of the active site in the C complex<sup>114,115</sup>. The conformation of the active site is supported also by lsy1. **b** | The step II splicing factors Prp17, Prp18 and Slu7 (not shown) are located near the active site in the C\* complex<sup>116,117</sup>. **c** | A mechanistic view of the two steps of transesterification in a cycle of pre-mRNA splicing. This model is based on the high-resolution cryo-electron microscopy structures of the yeast spliceosome and a summary of prior biochemical studies<sup>44,108,109,113–118</sup>. The three ATPases/helicases Prp2, Prp16 and Prp22 (in bold text) have essential roles in the conversion of B<sup>act</sup> to B\*, C to C\* and P to ILS, respectively. The step I factors Cwc25 and Yju2 stabilize the active site conformation for the first transesterification reaction, whereas the step II factors Prp18 and Slu7 facilitate the second transesterification reaction.

triangular-shaped complex, with a heptameric Sm ring binding the 3' end sequences of U4 snRNA at the far end of the arm. The Jab1/MPN domain of Prp8 binds to the other end of Brr2 and connects it to the centre of the complex<sup>90</sup>. The single-stranded RNA sequences between the 3'-stem-loop and stem I of the U4–U6 duplex are placed into the active site of the N-terminal helicase domain of Brr2 through a channel formed by its RecA1, RecA2, ratchet and WH domains<sup>111</sup>. In the B-to-B<sup>act</sup> transition, ATP binding and hydrolysis by Brr2 presumably generates a pulling force on the U4–U6 duplex, thereby dissociating the surrounding proteins and unwinding the U4–U6 duplex<sup>66,150,151</sup>. Intriguingly, in human tri-snRNP<sup>112</sup>, the position of Brr2 differs from that in yeast tri-snRNP by as much as 200 Å. This arrangement in yeast likely causes constant binding by Sad1, which functions to stabilize the tri-snRNP<sup>152</sup>.

The ATPase/helicase Prp2 is located at the periphery of the B<sup>act</sup> complex<sup>113,118</sup>. Prp2 interacts with the superhelical protein Hsh155 and is placed near the RES complex, which binds to the single-stranded intron sequences downstream of the BPS<sup>153,154</sup>. Prp2 presumably associates with the intron sequences downstream of those bound by the RES complex<sup>155–158</sup>. The adenosine nucleophile, together with the neighbouring intron–U2 duplex, is surrounded by proteins of the SF3b complex. In the B<sup>act</sup>-to-B<sup>\*</sup> transition, Prp2 may pull the 3' end sequences of the intron, thereby dissociating the RES complex and the SF3b complex<sup>142,156,157</sup>. These changes likely result in the exposure of the nucleophile in the BPS<sup>144,159</sup>. In the C complex<sup>115</sup>, a lobe of density was assigned to the ATPase/helicase Prp16, and another lobe was interpreted to be Brr2. The last ordered nucleotide at the 3' end of the intron is more than 60 Å away from the RNA entry site of Prp16. In the C-to-C<sup>\*</sup> transition, Prp16 would presumably pull the intron sequences from the 3' end towards the BPS, thereby dissociating the step I factors Cwc25 and Yju2 (REFS 75,147,160,161). In the C<sup>\*</sup> complex<sup>117</sup>, an ATPase/helicase tentatively assigned as Prp22 associates with the linker and RT fingers/palm domains of Prp8. In the P-to-ILS transition, Prp22 is required to dissociate the ligated exon in the 3' to 5' direction<sup>162,163</sup>. Notably, the ATPases/helicases in the B<sup>act</sup>, C and C<sup>\*</sup> complexes are located in the peripheral regions of the spliceosome and display inadequate density for atomic modelling. Consequently, the proposed functional mechanisms of these ATPase/helicases are speculative and await further experimental investigation.

### Comparison with a self-splicing intron

Pre-mRNA splicing is thought to share the same catalytic mechanism as that of the group IIB self-splicing introns<sup>26,31,164,165</sup>, both involving two catalytic metal ions and an intron lariat intermediate. Although the overall structure of the spliceosome is very different from that of the group IIB introns, they do share many common features at the active site. Structural comparison between the *S. pombe* ILS complex and the group IIB intron *Pli.LSUI2* from the brown alga *Pyliella littoralis*<sup>121</sup> reveals similar structural features between the ISL of U6 snRNA and the Domain V of the group IIB

intron, with the metal-coordinating U68 of the ILS complex replaced by the nucleotide A574 in the group IIB intron<sup>108,109</sup> (*Supplementary information S4* (figure)). The BPS–U2 duplex of the ILS complex is analogous to Domain VI of the group IIB intron, with the adenosine nucleophile in the BPS corresponding to A615 in *Pli.LSUI2* (REFS 108,109). The catalytic triad of nucleotides A47–G48–C49 in the spliceosome corresponds to A551–G552–C553 in the group IIB intron (*Supplementary information S4* (figure)). Finally, the formation of the catalytic triplex in the spliceosome is nearly identical to that in the self-splicing intron.

### U4/U6.U5 tri-snRNP and assembly of the snRNPs

Cryo-EM structures of the *S. cerevisiae* U4/U6.U5 tri-snRNP at near-atomic resolution provide insights into spliceosome assembly and activation<sup>110,111</sup>. Within the tri-snRNP, the ISL is sequestered through duplex formation with U4 snRNA, and the U4–U6 duplex is coated by Prp8 and the tri-snRNP-specific components Prp3, Prp6, Prp31 and Snu13 (also known as 13 kDa ribonucleoprotein-associated protein). In particular, the U6 nucleotides that coordinate the catalytic metals during the splicing reaction are protected by Prp3. In the tri-snRNP structure reported in one study<sup>110</sup>, the 5'SS already forms a duplex with the ACAGA box of U6, and four nucleotides at the 3' end of the exon are anchored to loop I. This observation is supported by prior biochemical characterizations<sup>57,166</sup>. By contrast, in the tri-snRNP structure reported in another study<sup>111</sup>, the 5' sequences of U6 snRNA occupy the same general location as that occupied by the pre-mRNA<sup>110</sup>. This portion of U6 forms a previously unknown short internal stem-loop, and a few consecutive nucleotides are anchored to loop I of U5 snRNA. The cryo-EM density maps in both studies have comparable overall quality. However, these two studies used different purification methods, and the tri-snRNP samples appear to represent different states of the tri-snRNP. Comparison of the two structures suggests a model in which the 5' sequences of U6 snRNA may serve as a decoy to compete with pre-mRNA binding in an early tri-snRNP. In this model, the 5' sequences of U6 snRNA may be bound to loop I of U5 snRNA in the tri-snRNP, and the recruitment of pre-mRNA may competitively displace the 5' sequences of U6 snRNA away from loop I.

### Conclusions and future perspective

During the past 2 years, the emergence of a number of atomic structures of the spliceosome by single-particle cryo-EM analysis has fundamentally advanced our mechanistic understanding of pre-mRNA splicing. The cryo-EM structure of the ILS complex from *S. pombe* enabled atomic modelling of an intact spliceosome with unprecedented structural clarity and revealed several principles that govern the function of the spliceosome and the splicing reaction<sup>108,109</sup>. This structure provided the first atomic view of the splicing active site and has enabled us to analyse the fine structural features that support the two steps of transesterification. The structure of the ISL of U6 snRNA is novel, and its coordination of catalytic metals by specific nucleotides confirms

#### Sm ring

Also known as heptameric Sm complex. A doughnut-shaped structure formed by seven distinct Sm proteins, with single-stranded RNA bound within the hollow centre.

a decades-old hypothesis<sup>26</sup> and subsequent biochemical investigations<sup>27–29,31</sup>. Unlike the mammalian ribosome, in which RNA accounts for the bulk of the molecular mass, the three snRNAs (U2, U5 and U6) of the functional human spliceosome represent no more than 5% of its total molecular mass. It is the spliceosomal protein components that organize the splicing active site, deliver the reactive moieties into the active site, and regulate the splicing process. Thus, the spliceosome is a protein-directed metalloribozyme.

The organizational principles observed in the *S. pombe* ILS complex<sup>108,109</sup> are remarkably conserved in the subsequently elucidated structures of the *S. cerevisiae* spliceosome<sup>113–118</sup>. U5 snRNP serves as the central pillar of the entire spliceosome, Prp8 anchors the splicing active site, and there is a rigid core in the spliceosome of at least 20 components. The available structural information of the yeast spliceosome enables the proposition of a mechanistic model for a pre-mRNA splicing cycle (FIG. 7c): the B<sup>act</sup> complex contains a well-formed active site but cannot yet catalyse the first transterification owing to separation of the two reactive moieties; such a structural organization is presumably suited for selection of alternative BPSs. Driven by Prp2 and its co-activator Spp2, approximately a dozen proteins in the SF3a, SF3b and RES complexes are dissociated from the B<sup>act</sup> complex<sup>142,157</sup>, and a few splicing factors are recruited to form the B\* complex, where the first transterification occurs. The resulting C complex, which contains a free 5' exon and an intron lariat–3' exon, is stabilized by the splicing factors Cwc25 and Yju2 and the NTC proteins Isy1 and Cef1. The stable C complex may enable selection of alternative 3'SS. The second transterification requires translocation of the lariat junction to vacate

space for the 3'SS–3' exon sequence, which is driven by the ATPase/helicase Prp16 and proceeds in the C\* complex. The post-catalytic P complex contains a ligated exon and a free intron lariat. The exon is released by Prp22, resulting in the ILS complex; the intron lariat is dissociated by the ATPase/helicase Prp43 with help from NTC-related protein 1 (Ntr1; also known as Spp382) and Ntr2 (REFS 167,168).

Intact spliceosomes for cryo-EM studies were either directly purified from cell nuclei or assembled *in vitro* using synthetic pre-mRNA (Supplementary information S1 (table)). The former approach more faithfully reflects cellular circumstances, whereas the latter may yield more homogeneous materials. Notably, the B<sup>act</sup>, C and C\* structures of the *S. cerevisiae* spliceosome were obtained from the same EM data set, which was collected on the same sample derived from endogenous yeast nuclear extract<sup>113,114,116</sup>. In this case, the distribution of the various spliceosomal complexes may represent that of yeast cells. Intriguingly, the B<sup>act</sup> and C complexes were the two most abundant spliceosomal species in the sample, suggesting that they represent rate-limiting regulatory steps in the splicing reaction. The selection of alternative BPSs and 3'SSs may crucially depend on the B<sup>act</sup> and C complexes, respectively.

Despite the recent structural advances<sup>108–118</sup>, important questions remain to be answered. The second transterification step, particularly the recognition and delivery to the active site of the 3'SS–3' exon sequences, remains poorly understood, and exactly how the eight conserved ATPases/helicases remodel the spliceosomal complexes remains enigmatic. Importantly, how alternative splicing occurs and how alternative BPSs and 3'SSs are selected are yet to be structurally explained.

- Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl Acad. Sci. USA* **74**, 3171–3175 (1977).
- Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**, 1–8 (1977).
- Lerner, M. R. & Steitz, J. A. Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc. Natl Acad. Sci. USA* **76**, 5495–5499 (1979).
- Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. Are snRNPs involved in splicing? *Nature* **283**, 220–224 (1980).
- Rogers, J. & Wall, R. A mechanism for RNA splicing. *Proc. Natl Acad. Sci. USA* **77**, 1877–1879 (1980).
- Hinterberger, M., Petterson, I. & Steitz, J. A. Isolation of small nuclear ribonucleoproteins containing U1, U2, U4, U5, and U6 RNAs. *J. Biol. Chem.* **258**, 2604–2613 (1983).
- Mount, S. M., Petterson, I., Hinterberger, M., Karmas, A. & Steitz, J. A. The U1 small nuclear RNA-protein complex selectively binds a 5' splice site *in vitro*. *Cell* **33**, 509–518 (1983).
- Padgett, R. A., Mount, S. M., Steitz, J. A. & Sharp, P. A. Splicing of messenger RNA precursors is inhibited by antisera to small nuclear ribonucleoprotein. *Cell* **35**, 101–107 (1983).
- Yang, V. W., Lerner, M. R., Steitz, J. A. & Flint, S. J. A small nuclear ribonucleoprotein is required for splicing of adenoviral early RNA sequences. *Proc. Natl Acad. Sci. USA* **78**, 1371–1375 (1981).
- DiMaria, P. R., Kaltwasser, G. & Goldenberg, C. J. Partial purification and properties of a pre-mRNA splicing activity. *J. Biol. Chem.* **260**, 1096–1102 (1985).
- Kramer, A., Keller, W., Appel, B. & Luhrmann, R. The 5' terminus of the RNA moiety of U1 small nuclear ribonucleoprotein particles is required for the splicing of messenger RNA precursors. *Cell* **38**, 299–307 (1984).
- Black, D. L., Chabot, B. & Steitz, J. A. U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. *Cell* **42**, 737–750 (1985).
- Krainer, A. R. & Maniatis, T. Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing *in vitro*. *Cell* **42**, 725–736 (1985).
- Berget, S. M. & Robberson, B. L. U1, U2, and U4/U6 small nuclear ribonucleoproteins are required for *in vitro* splicing but not polyadenylation. *Cell* **46**, 691–696 (1986).
- Grabowski, P. J. & Sharp, P. A. Affinity chromatography of splicing complexes: U2, U5, and U4 + U6 small nuclear ribonucleoprotein particles in the spliceosome. *Science* **233**, 1294–1299 (1986).
- Pikielny, C. W. & Rosbash, M. Specific small nuclear RNAs are associated with yeast spliceosomes. *Cell* **45**, 869–877 (1986).
- Goldenberg, C. J. & Hauser, S. D. Accurate and efficient *in vitro* splicing of purified precursor RNAs specified by early region 2 of the adenovirus 2 genome. *Nucleic Acids Res.* **11**, 1337–1348 (1983).
- Hernandez, N. & Keller, W. Splicing of *in vitro* synthesized messenger RNA precursors in HeLa cell extracts. *Cell* **35**, 89–99 (1983).
- Kole, R. & Weissman, S. M. Accurate *in vitro* splicing of human beta-globin RNA. *Nucleic Acids Res.* **10**, 5429–5445 (1982).
- Padgett, R. A., Hardy, S. F. & Sharp, P. A. Splicing of adenovirus RNA in a cell-free transcription system. *Proc. Natl Acad. Sci. USA* **80**, 5230–5234 (1983).
- Hardy, S. F., Grabowski, P. J., Padgett, R. A. & Sharp, P. A. Cofactor requirements of splicing of purified messenger RNA precursors. *Nature* **308**, 375–377 (1984).
- Krainer, A. R., Maniatis, T., Ruskin, B. & Green, M. R. Normal and mutant human beta-globin pre-mRNAs are faithfully and efficiently spliced *in vitro*. *Cell* **36**, 993–1005 (1984).
- Grabowski, P. J., Padgett, R. A. & Sharp, P. A. Messenger RNA splicing *in vitro*: an excised intervening sequence and a potential intermediate. *Cell* **37**, 415–427 (1984).
- Padgett, R. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F. & Sharp, P. A. Lariat RNAs as intermediates and products in the splicing of messenger RNA precursors. *Science* **225**, 898–903 (1984).
- Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing *in vitro*. *Cell* **38**, 317–331 (1984).
- Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl Acad. Sci. USA* **90**, 6498–6502 (1993).
- Sonthimer, E. J., Sun, S. & Piccirilli, J. A. Metal ion catalysis during splicing of premessenger RNA. *Nature* **388**, 801–805 (1997).
- Year, S. L., Wuenschell, G., Termini, J. & Lin, R. J. Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* **408**, 881–884 (2000).
- Fica, S. M. et al. RNA catalyses nuclear pre-mRNA splicing. *Nature* **503**, 229–234 (2013).
- Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* **320**, 77–82 (2008).

31. Keating, K. S., Toor, N., Perlman, P. S. & Pyle, A. M. A structural analysis of the group II intron active site and implications for the spliceosome. *RNA* **16**, 1–9 (2010).
32. Brody, E. & Abelson, J. The “spliceosome”: yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science* **228**, 963–967 (1985).
33. Grabowski, P. J., Seiler, S. R. & Sharp, P. A. A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell* **42**, 345–353 (1985).
34. Frendewey, D. & Keller, W. Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. *Cell* **42**, 355–367 (1985).
35. Aeby, M., Hornig, H., Padgett, R. A., Reiser, J. & Weissmann, C. Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. *Cell* **47**, 555–565 (1986).
36. Vijayraghavan, U. *et al.* Mutations in conserved intron sequences affect multiple steps in the yeast splicing pathway, particularly assembly of the spliceosome. *EMBO J.* **5**, 1683–1695 (1986).
37. Newman, A. J., Lin, R. J., Cheng, S. C. & Abelson, J. Molecular consequences of specific intron mutations on yeast mRNA splicing *in vivo* and *in vitro*. *Cell* **42**, 335–344 (1985).
38. Lamond, A. I., Konarska, M. M. & Sharp, P. A. A mutational analysis of spliceosome assembly: evidence for splice site collaboration during spliceosome formation. *Genes Dev.* **1**, 532–543 (1987).
39. Konarska, M. M. & Sharp, P. A. Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. *Cell* **46**, 845–855 (1986).
40. Pikielny, C. W., Rymond, B. C. & Rosbash, M. Electrophoresis of ribonucleoproteins reveals an ordered assembly pathway of yeast splicing complexes. *Nature* **324**, 341–345 (1986).
41. Konarska, M. M. & Sharp, P. A. Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell* **49**, 763–774 (1987).
42. Cheng, S. C. & Abelson, J. Spliceosome assembly in yeast. *Genes Dev.* **1**, 1014–1027 (1987).
43. Bindereif, A. & Green, M. R. An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. *EMBO J.* **6**, 2415–2424 (1987).
44. Wahl, M. C., Will, C. L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).
45. Guthrie, C. & Patterson, B. Spliceosomal snRNAs. *Annu. Rev. Genet.* **22**, 387–419 (1988).
46. Bringmann, P. & Luhrmann, R. Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. *EMBO J.* **5**, 3509–3516 (1986).
47. Lossky, M., Anderson, G. J., Jackson, S. P. & Beggs, J. Identification of a yeast snRNP protein and detection of snRNP-snRNP interactions. *Cell* **51**, 1019–1026 (1987).
48. Jackson, S. P., Lossky, M. & Beggs, J. D. Cloning of the RNA8 gene of *Saccharomyces cerevisiae*, detection of the RNA8 protein, and demonstration that it is essential for nuclear pre-mRNA splicing. *Mol. Cell. Biol.* **8**, 1067–1075 (1988).
49. Tarn, W. Y. *et al.* Functional association of essential splicing factor(s) with PRP19 in a protein complex. *EMBO J.* **13**, 2421–2431 (1994).
50. Chan, S. P., Kao, D. I., Tsai, W. Y. & Cheng, S. C. The Prp19p-associated complex in spliceosome activation. *Science* **302**, 279–282 (2003).
51. Chabot, B. & Steitz, J. A. Multiple interactions between the splicing substrate and small nuclear ribonucleoproteins in spliceosomes. *Mol. Cell. Biol.* **7**, 281–293 (1987).
52. Parker, R., Siliciano, P. G. & Guthrie, C. Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell* **49**, 229–239 (1987).
53. Newman, A. & Norman, C. Mutations in yeast U5 snRNA alter the specificity of 5' splice-site cleavage. *Cell* **65**, 115–123 (1991).
54. Newman, A. J. & Norman, C. U5 snRNA interacts with exon sequences at 5' and 3' splice sites. *Cell* **68**, 743–754 (1992).
55. Madhani, H. D. & Guthrie, C. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* **71**, 803–817 (1992).
56. Wassarman, D. A. & Steitz, J. A. Interactions of small nuclear RNA's with precursor messenger RNA during *in vitro* splicing. *Science* **257**, 1918–1925 (1992).
57. Wyatt, J. R., Sontheimer, E. J. & Steitz, J. A. Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. *Genes Dev.* **6**, 2542–2553 (1992).
58. Lesser, C. F. & Guthrie, C. Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science* **262**, 1982–1988 (1993).
59. Sontheimer, E. J. & Steitz, J. A. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* **262**, 1989–1996 (1993).
60. Kandels-Lewis, S. & Seraphin, B. Involvement of U6 snRNA in 5' splice site selection. *Science* **262**, 2035–2039 (1993).
61. Newman, A. J., Teigelkamp, S. & Beggs, J. D. snRNA interactions at 5' and 3' splice sites monitored by photoactivated crosslinking in yeast spliceosomes. *RNA* **1**, 968–980 (1995).
62. Anokhin, M. *et al.* RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. *EMBO J.* **32**, 2804–2818 (2013).
63. Jankowsky, E. RNA helicases at work: binding and rearranging. *Trends Biochem. Sci.* **36**, 19–29 (2011).
64. Cordin, O., Hahn, D. & Beggs, J. D. Structure, function and regulation of spliceosomal RNA helicases. *Curr. Opin. Cell Biol.* **24**, 431–438 (2012).
65. Staley, J. P. & Guthrie, C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315–326 (1998).
66. Raghunathan, P. L. & Guthrie, C. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr. Biol.* **8**, 847–855 (1998).
67. Lagerbauer, B., Achsel, T. & Luhrmann, R. The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplexes *in vitro*. *Proc. Natl Acad. Sci. USA* **95**, 4188–4192 (1998).
68. Chen, J. H. & Lin, R. J. The yeast PRP2 protein, a putative RNA-dependent ATPase, shares extensive sequence homology with two other pre-mRNA splicing factors. *Nucleic Acids Res.* **18**, 6447 (1990).
69. King, D. S. & Beggs, J. D. Interactions of PRP2 protein with pre-mRNA splicing complexes in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **18**, 6559–6564 (1990).
70. Kim, S. H. & Lin, R. J. Spliceosome activation by PRP2 ATPase prior to the first transterification reaction of pre-mRNA splicing. *Mol. Cell. Biol.* **16**, 6810–6819 (1996).
71. Kim, S. H., Smith, J., Claude, A. & Lin, R. J. The purified yeast pre-mRNA splicing factor PRP2 is an RNA-dependent NTPase. *EMBO J.* **11**, 2319–2326 (1992).
72. Burgess, S., Couto, J. R. & Guthrie, C. A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing. *Cell* **60**, 705–717 (1990).
73. Schwer, B. & Guthrie, C. PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature* **349**, 494–499 (1991).
74. Company, M., Arenas, J. & Abelson, J. Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature* **349**, 487–493 (1991).
75. Semlow, D. R., Blanco, M. R., Walter, N. G. & Staley, J. P. Spliceosomal DEAH-box ATPases remodel pre-mRNA to activate alternative splice sites. *Cell* **164**, 985–998 (2016).
76. Semlow, D. R. & Staley, J. P. Staying on message: ensuring fidelity of pre-mRNA splicing. *Trends Biochem. Sci.* **37**, 263–273 (2012).
77. Weber, G., Trowitzsch, S., Kastner, B., Luhrmann, R. & Wahl, M. C. Functional organization of the Sm core in the crystal structure of human U1 snRNP. *EMBO J.* **29**, 4172–4184 (2010).
78. Pomeranz Krummel, D. A., Oubridge, C., van Roon, A. M. & Nagai, K. Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* **458**, 475–480 (2009).
79. Kondo, Y., Oubridge, C., van Roon, A. M. & Nagai, K. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. *Elife* <http://dx.doi.org/10.7554/elife.04986> (2015).
80. Price, S. R., Evans, P. R. & Nagai, K. Crystal structure of the spliceosomal U2B'-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* **394**, 645–650 (1998).
81. Sickmier, E. A. *et al.* Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol. Cell* **23**, 49–59 (2006).
82. Lin, P. C. & Xu, R. M. Structure and assembly of the SF3a splicing factor complex of U2 snRNP. *EMBO J.* **31**, 1579–1590 (2012).
83. Jenkins, J. L., Agrawal, A. A., Gupta, A., Green, M. R. & Kielkopf, C. L. U2AF65 adapts to diverse pre-mRNA splice sites through conformational selection of specific and promiscuous RNA recognition motifs. *Nucleic Acids Res.* **41**, 3859–3873 (2013).
84. Yoshida, H. *et al.* A novel 3' splice site recognition by the two zinc fingers in the U2AF small subunit. *Genes Dev.* **29**, 1649–1660 (2015).
85. Leung, A. K., Nagai, K. & Li, J. Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature* **473**, 536–539 (2011).
86. Zhou, L. *et al.* Crystal structures of the Lsm complex bound to the 3' end sequence of U6 small nuclear RNA. *Nature* **506**, 116–120 (2014).
87. Montemayor, E. J. *et al.* Core structure of the U6 small nuclear ribonucleoprotein at 1.7-A resolution. *Nat. Struct. Mol. Biol.* **21**, 544–551 (2014).
88. Galej, W. P., Oubridge, C., Newman, A. J. & Nagai, K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* **493**, 638–643 (2013).
89. Mozaffari-Jovin, S. *et al.* Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8. *Science* **341**, 80–84 (2013).
90. Nguyen, T. H. *et al.* Structural basis of Brr2–Prp8 interactions and implications for U5 snRNP biogenesis and the spliceosome active site. *Structure* **21**, 910–919 (2013).
91. Cretu, C. *et al.* Molecular architecture of SF3b and structural consequences of its cancer-related mutations. *Mol. Cell* **64**, 307–319 (2016).
92. Zhou, Z., Sim, J., Griffith, J. & Reed, R. Purification and electron microscopic visualization of functional human spliceosomes. *Proc. Natl Acad. Sci. USA* **99**, 12203–12207 (2002).
93. Jurica, M. S., Licklider, L. J., Gygi, S. R., Grigorieff, N. & Moore, M. J. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA* **8**, 426–439 (2002).
94. Luhrmann, R. & Stark, H. Structural mapping of spliceosomes by electron microscopy. *Curr. Opin. Struct. Biol.* **19**, 96–102 (2009).
95. Behzadnia, N. *et al.* Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO J.* **26**, 1737–1748 (2007).
96. Furman, E. & Glitz, D. G. Purification of the spliceosome A-complex and its visualization by electron microscopy. *J. Biol. Chem.* **270**, 15515–15522 (1995).
97. Boehringer, D. *et al.* Three-dimensional structure of a pre-catalytic human spliceosomal complex B. *Nat. Struct. Mol. Biol.* **11**, 463–468 (2004).
98. Wolf, E. *et al.* Exon, intron and splice site locations in the spliceosomal B complex. *EMBO J.* **28**, 2283–2292 (2009).
99. Deckert, J. *et al.* Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol. Cell. Biol.* **26**, 5528–5543 (2006).
100. Bessonov, S. *et al.* Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA* **16**, 2384–2403 (2010).
101. Golas, M. M. *et al.* 3D cryo-EM structure of an active step I spliceosome and localization of its catalytic core. *Mol. Cell* **40**, 927–938 (2010).
102. Jurica, M. S., Sousa, D., Moore, M. J. & Grigorieff, N. Three-dimensional structure of C complex spliceosomes by electron microscopy. *Nat. Struct. Mol. Biol.* **11**, 265–269 (2004).
103. Ilagan, J. O., Chalkley, R. J., Burlingame, A. L. & Jurica, M. S. Rearrangements within human spliceosomes captured after exon ligation. *RNA* **19**, 400–412 (2013).
104. Fabrizio, P. *et al.* The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol. Cell* **36**, 593–608 (2009).
105. Ohi, M. D., Ren, L., Wall, J. S., Gould, K. L. & Walz, T. Structural characterization of the fission yeast U5, U2/U6 spliceosome complex. *Proc. Natl. Acad. Sci. USA* **104**, 3195–3200 (2007).
106. Chen, W. *et al.* Endogenous U2, U5, U6 snRNA complexes in *S. pombe* are intron lariat spliceosomes. *RNA* **20**, 308–320 (2014).

107. Nguyen, T. H. *et al.* The architecture of the spliceosomal U4/U6. U5 tri-snRNP. *Nature* **523**, 47–52 (2015).
108. Yan, C. *et al.* Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* **349**, 1182–1191 (2015).
109. Hang, J., Wan, R., Yan, C. & Shi, Y. Structural basis of pre-mRNA splicing. *Science* **349**, 1191–1198 (2015).
110. Wan, R. *et al.* The 3.8 Å structure of the U4/U6. U5 tri-snRNP: insights into spliceosome assembly and catalysis. *Science* **351**, 466–475 (2016).
111. Nguyen, T. H. *et al.* Cryo-EM structure of the yeast U4/U6. U5 tri-snRNP at 3.7 Å resolution. *Nature* **530**, 298–302 (2016).
112. Agafonov, D. E. *et al.* Molecular architecture of the human U4/U6. U5 tri-snRNP. *Science* **351**, 1416–1420 (2016).
113. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast activated spliceosome at 3.5 Å resolution. *Science* **353**, 904–911 (2016).
114. Wan, R., Yan, C., Bai, R., Huang, G. & Shi, Y. Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* **353**, 895–904 (2016).
115. Galej, W. P. *et al.* Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537**, 197–201 (2016).
116. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast step II catalytically activated spliceosome. *Science* **355**, 149–155 (2017).
117. Fica, S. M. *et al.* Structure of a spliceosome remodelled for exon ligation. *Nature* **542**, 377–380 (2017).
118. Rauhut, R. *et al.* Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. *Science* **353**, 1399–1405 (2016).
119. Bertram, K. *et al.* Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* **542**, 318–323 (2017).
120. Fica, S. M., Mefford, M. A., Piccirilli, J. A. & Staley, J. P. Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nat. Struct. Mol. Biol.* **21**, 464–471 (2014).
121. Robart, A. R., Chan, R. T., Peters, J. K., Rajashankar, K. R. & Toor, N. Crystal structure of a eukaryotic group II intron lariat. *Nature* **514**, 193–197 (2014).
122. Grainger, R. J. & Beggs, J. D. Prp8 protein: at the heart of the spliceosome. *RNA* **11**, 533–557 (2005).
123. Turner, I. A., Norman, C. M., Churcher, M. J. & Newman, A. J. Dissection of Prp8 protein defines multiple interactions with crucial RNA sequences in the catalytic core of the spliceosome. *RNA* **12**, 375–386 (2006).
124. Galej, W. P., Nguyen, T. H., Newman, A. J. & Nagai, K. Structural studies of the spliceosome: zooming into the heart of the machine. *Curr. Opin. Struct. Biol.* **25**, 57–66 (2014).
125. Yang, K., Zhang, L., Xu, T., Heroux, A. & Zhao, R. Crystal structure of the beta-finger domain of Prp8 reveals analogy to ribosomal proteins. *Proc. Natl Acad. Sci. USA* **105**, 13817–13822 (2008).
126. Garre, S. M. *et al.* A homolog of lariat-debranching enzyme modulates turnover of branched RNA. *RNA* **20**, 1337–1348 (2014).
127. Rasche, N. *et al.* Cwc2 and its human homologue Rbm22 promote an active conformation of the spliceosome catalytic centre. *EMBO J.* **31**, 1591–1604 (2012).
128. Hahn, C. N. & Scott, H. S. Spliceosome mutations in hematopoietic malignancies. *Nat. Genet.* **44**, 9–10 (2012).
129. Darman, R. B. *et al.* Cancer-associated SF3B1 hotspot mutations induce cryptic 3' splice site selection through use of a different branch point. *Cell Rep.* **13**, 1033–1045 (2015).
130. Alsafadi, S. *et al.* Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat. Commun.* **7**, 10615 (2016).
131. Jacquier, A. & Michel, F. Base-pairing interactions involving the 5' and 3'-terminal nucleotides of group II self-splicing introns. *J. Mol. Biol.* **213**, 437–447 (1990).
132. Chanfreau, G. & Jacquier, A. Catalytic site components common to both splicing steps of a group II intron. *Science* **266**, 1383–1387 (1994).
133. Umen, J. G. & Guthrie, C. The second catalytic step of pre-mRNA splicing. *RNA* **1**, 869–885 (1995).
134. Luukkonen, B. G. & Seraphin, B. The role of branchpoint-3' splice site spacing and interaction between intron terminal nucleotides in 3' splice site selection in *Saccharomyces cerevisiae*. *EMBO J.* **16**, 779–792 (1997).
135. Collins, C. A. & Guthrie, C. Genetic interactions between the 5' and 3' splice site consensus sequences and U6 snRNA during the second catalytic step of pre-mRNA splicing. *RNA* **7**, 1845–1854 (2001).
136. Frank, D. & Guthrie, C. An essential splicing factor, SLU7, mediates 3' splice site choice in yeast. *Genes Dev.* **6**, 2112–2124 (1992).
137. Umen, J. G. & Guthrie, C. Prp16p, Slu7p, and Prp8p interact with the 3' splice site in two distinct stages during the second catalytic step of pre-mRNA splicing. *RNA* **1**, 584–597 (1995).
138. Umen, J. G. & Guthrie, C. Mutagenesis of the yeast gene PRP8 reveals domains governing the specificity and fidelity of 3' splice site selection. *Genetics* **143**, 723–739 (1996).
139. Chua, K. & Reed, R. The RNA splicing factor hSlu7 is required for correct 3' splice-site choice. *Nature* **402**, 207–210 (1999).
140. Chen, W. & Moore, M. J. The spliceosome: disorder and dynamics defined. *Curr. Opin. Struct. Biol.* **24**, 141–149 (2014).
141. Ohi, M. D. & Gould, K. L. Characterization of interactions among the Cef1p-Prp19p-associated splicing complex. *RNA* **8**, 798–815 (2002).
142. Ohrt, T. *et al.* Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS. *RNA* **18**, 1244–1256 (2012).
143. Grainger, R. J., Barras, J. D., Jacquier, A., Rain, J. C. & Beggs, J. D. Physical and genetic interactions of yeast Cwc21p, an ortholog of human SRm300/SRRM2, suggest a role at the catalytic center of the spliceosome. *RNA* **15**, 2161–2173 (2009).
144. Warkocki, Z. *et al.* Reconstitution of both steps of *Saccharomyces cerevisiae* splicing with purified spliceosomal components. *Nat. Struct. Mol. Biol.* **16**, 1237–1243 (2009).
145. Chiu, Y. F. *et al.* Cwc25 is a novel splicing factor required after Prp2 and Yju2 to facilitate the first catalytic reaction. *Mol. Cell. Biol.* **29**, 5671–5678 (2009).
146. Krishnan, R. *et al.* Biased Brownian ratcheting leads to pre-mRNA remodeling and capture prior to first-step splicing. *Nat. Struct. Mol. Biol.* **20**, 1450–1457 (2013).
147. Ohrt, T. *et al.* Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. *RNA* **19**, 902–915 (2013).
148. James, S. A., Turner, W. & Schwer, B. How Slu7 and Prp18 cooperate in the second step of yeast pre-mRNA splicing. *RNA* **8**, 1068–1077 (2002).
149. Zhang, X. & Schwer, B. Functional and physical interaction between the yeast splicing factors Slu7 and Prp18. *Nucleic Acids Res.* **25**, 2146–2152 (1997).
150. Santos, K. F. *et al.* Structural basis for functional cooperation between tandem helicase cassettes in Brr2-mediated remodeling of the spliceosome. *Proc. Natl Acad. Sci. USA* **109**, 17418–17423 (2012).
151. Hahn, D., Kudla, G., Tollervey, D. & Beggs, J. D. Brr2p-mediated conformational rearrangements in the spliceosome during activation and substrate repositioning. *Genes Dev.* **26**, 2408–2421 (2012).
152. Huang, Y. H., Chung, C. S., Kao, D. I., Kao, T. C. & Cheng, S. C. Sad1 counteracts Brr2-mediated dissociation of U4/U6. U5 in tri-snRNP homeostasis. *Mol. Cell. Biol.* **34**, 210–220 (2014).
153. McPheeters, D. S. & Muhlenkamp, P. Spatial organization of protein-RNA interactions in the branch site-3' splice site region during pre-mRNA splicing in yeast. *Mol. Cell. Biol.* **23**, 4174–4186 (2003).
154. Schneider, C. *et al.* Dynamic contacts of U2, RES, Cwc25, Prp8 and Prp45 proteins with the pre-mRNA branch-site and 3' splice site during catalytic activation and step 1 catalysis in yeast spliceosomes. *PLoS Genet.* **11**, e1005539 (2015).
155. Edwards-Gilbert, G., Kim, D. H., Silverman, E. & Lin, R. J. Definition of a spliceosome interaction domain in yeast Prp2 ATPase. *RNA* **10**, 210–220 (2004).
156. Liu, H. L. & Cheng, S. C. The interaction of Prp2 with a defined region of the intron is required for the first splicing reaction. *Mol. Cell. Biol.* **32**, 5056–5066 (2012).
157. Warkocki, Z. *et al.* The G-patch protein Spp2 couples the spliceosome-stimulated ATPase activity of the DEAH-box protein Prp2 to catalytic activation of the spliceosome. *Genes Dev.* **29**, 94–107 (2015).
158. Teigelkamp, S., McGarvey, M., Plumpton, M. & Beggs, J. D. The splicing factor PRP2, a putative RNA helicase, interacts directly with pre-mRNA. *EMBO J.* **13**, 888–897 (1994).
159. Lardelli, R. M., Thompson, J. X., Yates, J. R. 3rd & Stevens, S. W. Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA* **16**, 516–528 (2010).
160. Schwer, B. & Guthrie, C. A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *EMBO J.* **11**, 5033–5039 (1992).
161. Tseng, C. K., Liu, H. L. & Cheng, S. C. DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. *RNA* **17**, 145–154 (2011).
162. Schwer, B. A conformational rearrangement in the spliceosome sets the stage for Prp2-dependent mRNA release. *Mol. Cell* **30**, 743–754 (2008).
163. Schwer, B. & Gross, C. H. Prp22, a DEH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *EMBO J.* **17**, 2086–2094 (1998).
164. Sharp, P. A. On the origin of RNA splicing and introns. *Cell* **42**, 397–400 (1985).
165. Cech, T. R. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell* **44**, 207–210 (1986).
166. Maroney, P. A., Romfo, C. M. & Nilsen, T. W. Functional recognition of 5' splice site by U4/U6. U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. *Mol. Cell* **6**, 317–328 (2000).
167. Tsai, R. T. *et al.* Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. *Genes Dev.* **19**, 2991–3003 (2005).
168. Fournier, J. B., Tauchert, M. J., Ficner, R., Fabrizio, P. & Lührmann, R. Regulation of Prp43-mediated disassembly of spliceosomes by its cofactors Ntr1 and Ntr2. *Nucleic Acids Res.* **45**, 4068–4080 (2017).

**Acknowledgements**

The author apologizes to those colleagues whose important work is not cited in this review owing to space limitation or inadvertent omission. The author thanks R. Wan, R. Bai, J. Hang, X. Zhang, X. Zhan and C. Yan for preparation of the figures and suggestions on text and references. This work was supported by funds from the Beijing Advanced Innovation Center for Structural Biology at Tsinghua University, Beijing, China.

**Competing interests statement**

The author declares no competing interests.

**Publisher's note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**SUPPLEMENTARY INFORMATION**

See online article: [S1](#) (table) | [S2](#) (figure) | [S3](#) (figure) | [S4](#) (figure)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF