Structural basis for conformational equilibrium of the catalytic spliceosome

Max E. Wilkinson[1*†], Sebastian M. Fica[1*†], Wojciech P. Galej[1,2*] & Kiyoshi Nagai[1‡]

The catalytic spliceosome exists in equilibrium between the branching (B*/ C) and exon ligation (C*/ P) conformations. Here we present the electron cryo-microscopy reconstruction of the *Saccharomyces cerevisiae* C complex at 2.8 Å resolution and identify a novel C-complex intermediate (C_i) that elucidates the molecular basis for this equilibrium. In the C_i conformation, the exon-ligation factors Prp18 and Slu7 are already bound before ATP hydrolysis by Prp16, which destabilises the branching conformation. Biochemical assays suggest these pre-bound factors prime C complex for conversion to C* by Prp16. A complete model of the Prp19-complex (NTC) shows how the NTC pre-recruits the branching factors Yju2 and Isy1 before branching. Prp16 remodels Yju2 binding after branching, allowing Yju2 to remain associated with the C* and P spliceosomes and promote exon ligation. Our results explain how Prp16 action modulates dynamic binding of step-specific factors to alternatively stabilise the C or C* conformation and establish equilibrium of the catalytic spliceosome.

Introduction

The spliceosome produces mRNA by excising introns from pre-mRNAs in two phosphoryl transfer reactions – branching and exon ligation. The spliceosome assembles *de novo* on each pre-mRNA through protein and RNA interactions that recognise conserved sequences at exon-intron junctions, called splice sites[1] (**Fig. 1a**). The U6 snRNA recognises the 5′-splice site (5′-SS), while the U2 snRNA pairs with the intron around the branch adenosine (brA), forming the branch helix[1,2]. The 5′-exon is stabilised in the active site by pairing with loop I of the U5 snRNA (refs. 3,4). The U2 and U6 snRNAs fold into a triple helix conformation that constitutes the active site and allows U6 snRNA to position two catalytic metals (refs. 1,5,6). This active site is stabilised by binding of the Prp19-associated complex[1,7] (NTC). Branching occurs in the B* complex when the 2′-hydroxyl of the brA attacks the 5′-SS. The resulting C complex is remodelled by the DEAH-box ATPase Prp16 into the C* open conformation (**Fig. 1b**). Docking of the 3′-splice site (3′-SS) at the active site (refs. 8,9) forms the closed C* complex, which then catalyses exon ligation, when the 3′-hydroxyl of the 5′-exon attacks the 3′-SS, resulting in mRNA formation and excision of the lariat-intron (**Fig. 1a,b**). Exon ligation forms the post-catalytic P complex (**Fig. 1b**), from which the ATPase Prp22 releases the mRNA.

Several step-specific splicing factors associate with the spliceosome during branching and exon ligation to promote splicing chemistry. In the budding yeast *Saccharomyces cerevisiae* the branching factors Cwc25, Yju2, and Isy1 (refs. 10-13) clamp the branch helix and dock it into the spliceosome active site[7,14,15]. The exon-ligation factors Prp18, Slu7 and Prp17, act more indirectly. Prp18 inserts a conserved loop near the active site to stabilise the docked 3′-SS, while Slu7 binds the spliceosome surface and rigidifies the C* conformation. Slu7 acts by an unknown mechanism to promote chemistry for substrates with a long distance between the brA and the 3′-SS (refs. 8,9,16-20). Prp17 binds the spliceosome before branching[21] but functions primarily during exon ligation[22] to stabilise an undocked conformation of the branch helix, as observed in the C* complex[8,9].

While a single active site catalyses both splicing steps[5,7,9], a structural rearrangement allows the 3′-SS to replace the branch helix during exon ligation. Genetic experiments have suggested a two-state model of the catalytic spliceosome where an equilibrium exists between the branching and exon-ligation conformations[23]. This equilibrium is modulated by Prp8 and by the ATPase Prp16 (refs. 24,25), whose activity promotes exon ligation in the forward direction by dissociating branching factors and undocking the branch helix from the active site (**Fig. 1b**, refs. 1,26). Indeed, under certain *in vitro* conditions, both catalytic steps of splicing are reversible, and disrupting the function of step-specific factors promotes reversal[27,28]. Intriguingly, the exon-ligation factors Slu7 and Prp18 have been suggested to bind to low affinity sites already in B*, while Prp16 action converts the binding sites to higher affinity sites in C* (ref. 29). Despite these biochemical and genetic studies, it remains unclear whether step-specific factors bind exclusively to either the branching or the exon-ligation conformation of the spliceosome.

---

[1]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, United Kingdom. [2]Present address: EMBL Grenoble, 71 Avenue des Martyrs, Grenoble 38 042, France. *These authors contributed equally to this work.
†Correspondence to mwilkin@mrc-lmb.cam.ac.uk and sfica@mrc-lmb.cam.ac.uk.

‡This work is dedicated to the memory of Kiyoshi Nagai (1949-2019).

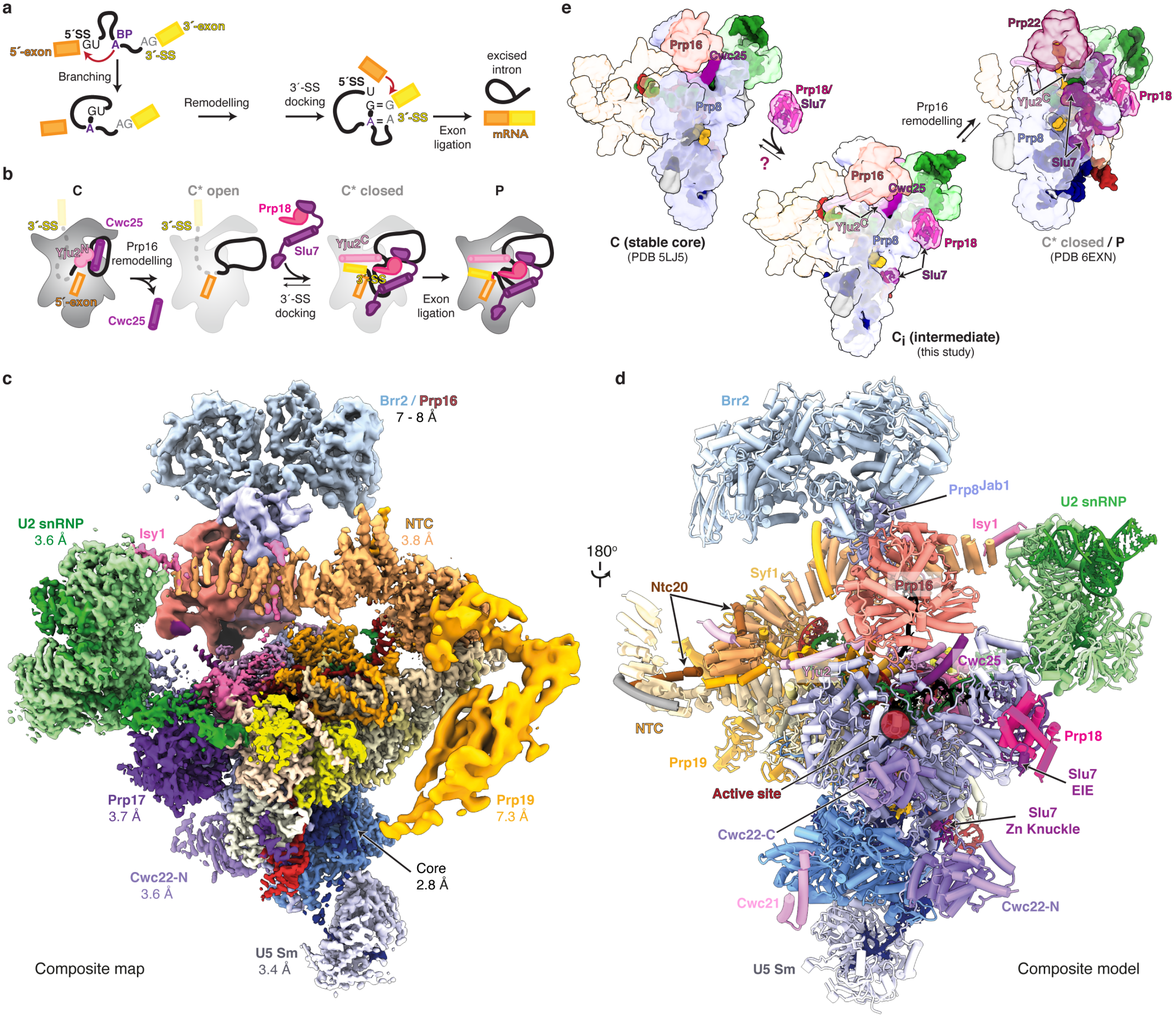

a
5´SS
BP
3´-exon
5´-exon
GU
A
AG
3´-SS
Branching
Remodelling
3´-SS docking
5´SS
U
G = G
A = A
3´-SS
excised intron
Exon ligation
mRNA
b
C
C* open
C* closed
P
3´-SS
Cwc25
Yju2N
5´-exon
Prp16 remodelling
Prp18
Slu7
Yju2C
3´-SS docking
Exon ligation
e
Prp16
Cwc25
Prp8
Prp18/ Slu7
?
C (stable core)
(PDB 5LJ5)
Prp16 remodelling
Prp22
Yju2C
Slu7
C* closed / P
(PDB 6EXN)
Ci (intermediate)
(this study)
c
Brr2 / Prp16
7 - 8 Å
U2 snRNP
3.6 Å
Isy1
NTC
3.8 Å
Prp17
3.7 Å
Prp19
7.3 Å
Cwc22-N
3.6 Å
Core
2.8 Å
U5 Sm
3.4 Å
Composite map
d
Brr2
Prp8Jab1
U2 snRNP
Isy1
180°
Ntc20
Syf1
Prp16
Cwc25
Yju2
NTC
Prp19
Active site
Prp18
Slu7 EIE
Slu7 Zn Knuckle
Cwc22-C
Cwc21
Cwc22-N
U5 Sm
Composite model

Figure 1 | A novel C-complex spliceosome intermediate. a, Pre-mRNA splicing mechanism. **b,** Canonical pathway for binding of exon ligation factors during the catalytic stage of splicing. **c,** Composite cryo-EM map of the C-complex spliceosome. Average resolution (FSC 0.143) is indicated for each focussed map. **d,** Composite model of the C-complex spliceosome with pre-bound exon ligation factors. **e,** The novel C-complex intermediate (C_i) suggests early binding of exon ligation factors before Prp16-mediated remodelling. The previously observed stable C conformation may exist in equilibrium with our novel C_i (intermediate) conformation.

We sought to determine the molecular basis for conformational equilibrium of the catalytic spliceosome by investigating whether branching and exon-ligation factors can engage the C-complex conformation at the same time. Here we present the detailed analysis of a large electron cryo-microscopy (cryo-EM) dataset of the C-complex spliceosome from budding yeast. The new cryo-EM reconstruction of the C complex at 2.8 Å resolution allowed building of the most complete atomic model of a catalytic spliceosome so far (**Fig. 1c,d**). The RNA-based active site is stabilised by both monovalent and divalent cations and uses non-Watson–Crick base pairs for splice site recognition. Most importantly, focussed classification revealed that exon-ligation factors Prp18 and Slu7 are pre-recruited to the branching conformation, creating a new C-complex intermediate (C_i) (**Fig. 1e**), while biochemical assays suggest these pre-bound factors are competent to promote exon ligation upon conversion to the C* conformation by Prp16. Finally, a complete model of the NTC shows how Syf1 acts as a recruitment hub for the branching factors Yju2 and Isy1, whose binding is remodelled in the transition from branching to exon ligation. Thus, conformational equilibrium of the catalytic spliceosome is mediated by dynamic binding of step-specific factors to the C or C* conformation.

Results

The complete structure of the yeast C-complex spliceosome. We previously reported a 3.8 Å structure of the C-complex spliceosome stalled by a 3′-SS mutation and purified with affinity tags on Prp18 and Slu7 (ref. 7). While the 3′-SS UAc mutation was expected to stall spliceosomes mainly in the C*

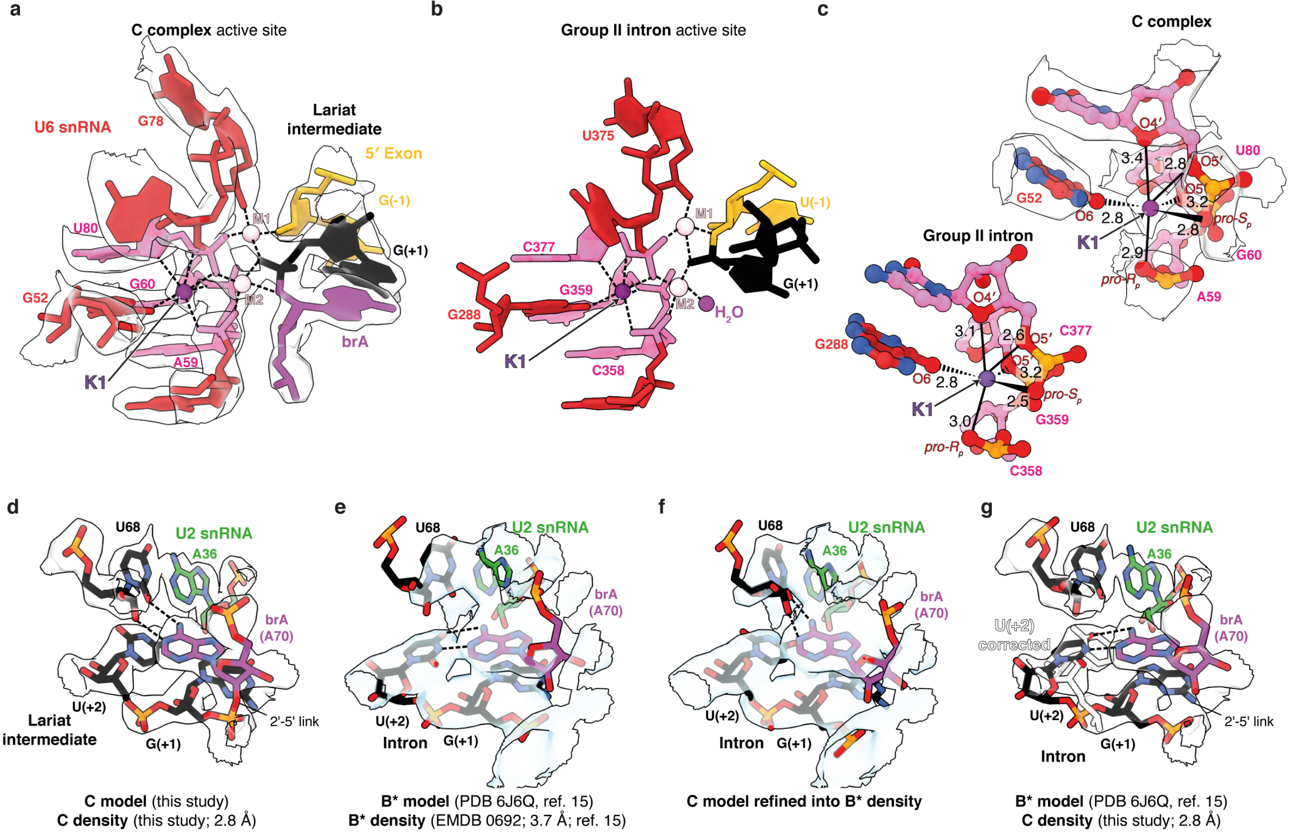

a
C complex active site
U6 snRNA
G78
Lariat intermediate
5′ Exon
G(-1)
M1
G(+1)
U80
G60
M2
G52
A59
brA
K1
b
Group II intron active site
U375
U(-1)
M1
C377
G(+1)
G359
M2
H2O
G288
K1
C358
c
C complex
O4′
U80
3.4
2.8
O5′
G52
5′
3.2
O6
2.8
pro-Sp
2.8
K1
G60
2.9
pro-Rp
A59
Group II intron
O4′
3.1
2.6
O5′
C377
G288
5′
3.2
O6
2.8
pro-Sp
2.5
K1
G359
3.0
pro-Rp
C358
d
U68
U2 snRNA
A36
brA (A70)
U(+2)
Lariat intermediate
G(+1)
2′-5′ link
C model (this study)
C density (this study; 2.8 Å)
e
U68
U2 snRNA
A36
brA (A70)
U(+2)
Intron
G(+1)
B* model (PDB 6J6Q, ref. 15)
B* density (EMDB 0692; 3.7 Å; ref. 15)
f
U68
U2 snRNA
A36
brA (A70)
U(+2)
Intron
G(+1)
C model refined into B* density
g
U68
U2 snRNA
A36
U(+2) corrected
brA (A70)
U(+2)
2′-5′ link
G(+1)
Intron
B* model (PDB 6J6Q, ref. 15)
C density (this study; 2.8 Å)

Figure 2 | Key active site elements of the C complex. a, A potassium ion stabilises the catalytic centre of the spliceosome. Metal coordination is shown as dashed lines; cryo-EM density is shown as transparent contour surface. **b,** Structure of the M1-M2-K1 catalytic metal cluster in a group IIa intron before hydrolysis (PDB 4FAQ, ref. 33). **c,** Details of the K1 coordination sphere in the spliceosome and group II intron. Distances are labelled in Ångstrom. **d-g,** Recognition of the branch adenosine (brA) during branching. Comparison of our proposed brA pairing in B* and C with the previously suggested brA interactions in B*.

conformation[30,31], accumulation of C complexes suggested that in the absence of stable 3′-SS docking the spliceosome may equilibrate back into the branching conformation (**Fig. 1b**) without complete dissociation of exon-ligation factors. Indeed, stalling the C* complex using a chemical modification of the 3′-SS also produced significant amounts of spliceosomes in the C conformation, which remained associated with Slu7 (ref. 8). These observations raised the possibility that exon-ligation factors may bind the C-complex spliceosome before Prp16 remodelling.

We merged all C-complex particles obtained by cryo-EM from various spliceosome purifications and investigated binding of Slu7 and Prp18 by focussed classification (**Extended Data Figs. 1-4**). This yielded a cryo-EM reconstruction at 2.8 Å resolution for the C-complex core, showing complete base separation, base and side chain rotamers, and allowing discrimination between adenosine and guanosine bases (**Extended Data Fig. 5**). This map allowed high-confidence modelling of every protein and RNA in the core of the catalytic spliceosome. The large number of particles facilitated classification and focussed refinement for the peripheral flexible regions of the spliceosome (**Extended Data Figs. 2, 3**). This produced improved maps for the U2 snRNP (3.6 Å resolution) and NTC (3.8 Å resolution) and allowed atomic modelling of these regions, which in all previous yeast spliceosome structures were only modelled by homology. The C-complex helicase module (Brr2 and Prp16) and Prp19 module (Prp19, Snt309, Cef1 C-terminus) were each refined to sub-nanometre resolution (6-7 Å), allowing discrimination of secondary structures and model building by molecular dynamics flexible fitting of crystal structures and homology models (**Extended Data Figs. 2-5, Table 1**). Together, these improved maps allow us to present the most complete molecular model of a catalytic spliceosome (**Fig. 1c,d**). The structure reveals novel features of the active site and unexpected binding of Prp18 and Slu7 to C complex, thus shedding light on remodelling during the catalysis.

Monovalent metal ion binding in the spliceosome active site. The improved resolution of our new C-complex map allows a finer analysis of the active site of the spliceosome. The active site is formed by the U6 snRNA, which adopts a triplex conformation to position two catalytic Mg^{2+} ions during branching and exon ligation[1,5,6,32]. The spliceosome uses the same configuration of the active site and the same two-metal ion mechanism for catalysis as observed in group II self-splicing introns[5,33], and recent structural studies support the evolution of the spliceosome from group II introns[34,35]. The active site of group II introns binds a composite, trinuclear metal cluster, in which the classical two catalytic Mg^{2+} ions (M1 and M2) are stabilised by a third monovalent ion, usually K^+ (K1) (**Extended Data Fig. 6**, refs. 33,36), which is necessary for efficient branching by some group II introns[33]. Thus, an additional positive charge may be required to stabilise M1 and M2 binding during catalysis. Although several spliceosome structures were reported with non-catalytic Mg^{2+} ions stabilising the U6 snRNA triplex conformation[15,37], no unambiguous density for K1 had been observed so far, despite several studies suggesting pre-mRNA splicing requires monovalent cations[38,39].

In our 2.8 Å map of C complex, we observed a new spherical density clearly separated from adjacent phosphates and bases in a position corresponding to K1 in group II introns (**Fig. 2a-c, Extended Data Fig. 6**). This new density has octahedral coordination geometry from surrounding oxygen ligands, with bond distances of 2.7 to 3 Å, consistent with a monovalent cation[40] (**Fig. 2c**). As our spliceosomes were purified using buffers containing K^+, we assigned this density as a potassium ion. This ion interacts with three of the four U6 snRNA residues that provide ligands for the catalytic Mg^{2+} ions, including all the ligands for M2 (A59, G60, and U80), as well as with the third strand of the triplex (G52) and is thus ideally positioned to stabilise M1 and M2 binding at the active site, as does K1 in group II introns (**Fig. 2a,b**). Intriguingly, right after formation of the active site triplex, in B^{act}, this new K1 site is partially blocked by a lysine from Prp11, a U2 snRNP SF3 sub-complex protein that is dissociated by Prp2 during catalytic activation[37] (**Extended Data Fig. 6**). In the late B* complex, before productive docking of brA at the active site, the K1 site remains unoccupied[15]. The C complex is the first spliceosome state where K1 is fully occupied and this site remains filled in P complex cryo-EM maps, provided that K^+ is present in the purification buffers[18] (**Extended Data Fig. 6**). Such progressive occupancy of the K1 site appears conserved in human spliceosomes (**Extended Data Fig. 6**) and may couple catalytic competence with docking of the branch helix at the active site. We propose that formation of the full M1-M2-K1 metal cluster is linked to docking of the brA at the active site and to catalytic activation of the spliceosome. This metal cluster is maintained throughout the catalytic stage and promotes both branching and exon ligation, similarly to its group II intron equivalent[33].

Non-Watson–Crick base pairs recognise the branch adenosine during branching. The brA bulges out of the branch helix formed between the intron and U2 snRNA and must be precisely positioned to allow for nucleophilic attack on the 5′-SS (refs. 1,7). Thus, recognition of the brA is essential for efficient branching, and specific mutations of this adenosine inhibit branching *in vitro* and *in vivo*[41,42] during the catalytic stage[25] by destabilising the branching conformation[23]. In our original 3.8 Å structure of C complex[7], we built the branch adenosine such that its Watson–Crick face paired with the sugar edge of intron nucleotide U(+68), two nucleotides upstream (UACUA**A**C, brA in bold, U(+68) underlined). However, in a different, 3.4 Å structure of C complex[14] this pairing was not modelled. Instead, in subsequent structures of a B* complex poised for branching[15] a canonical Watson–Crick base pair was proposed between brA and the 5′-SS nucleotide U(+2), although in these structures the local resolution around the brA was limited (~ 4 Å) (**Fig. 2e**). Therefore, the structural basis for the nucleotide specificity of the branching reaction was unclear.

In our new structure of C complex, the branch helix is robustly docked by the branching factors and is visible at a local resolution of 2.8 Å around the branch adenosine. The density unambiguously demonstrates that brA pairs to the sugar edge of U(+68) in a base-triple interaction, while the 5′-SS U(+2) pairs with U2 snRNA G37 in the context of another base triple (**Fig. 2c**), as we proposed in our earlier model[7]. Furthermore, this model can be convincingly refined into density for the B* complex poised for branching, with Yju2 and Isy1 bound to the branch helix[15] (**Fig. 2d,e**). Indeed, mutations at U(+2) of the intron primarily impair exon ligation, which is inconsistent with a functional role for U(+2) in pairing to brA during branching (refs. 23,24). The previous B* model likely originates from the ambiguity of map interpretation at approximately 4 Å resolution and there is no evidence for brA base-pair remodelling between B* and C complexes (**Fig. 2e-g**). Here, we propose that the brA is recognised for branching through non-canonical pairing with U(+68), which anchors it to the branch helix while allowing the 2′-hydroxyl to be positioned for nucleophilic attack on the 5′-SS. Our model explains the effects of specific branch adenosine mutations. Substitution of brA with cytosine or uridine strongly inhibits branching, while mutation to guanosine has milder effects *in vivo* and *in vitro*[23,25,43], and these effects correlate with the level of distortion and movement of the brA 2′-OH that would result from accommodating base pairs analogous to the interaction of brA to U(+68) (ref. 44, **Extended Data Fig. 7**). These observations further underscore the importance of careful analysis and interpretation of cryo-EM maps in the context of previous functional studies of the spliceosome[45].

The exon-ligation factors Prp18 and Slu7 are recruited to a new on-pathway C_i intermediate. In C* and P complexes, the exon-ligation factor Prp18 binds to a face of the Prp8 RNaseH domain ($Prp8^{RH}$) that would also be accessible in C complexes (**Figs. 1b, 3a**), suggesting that Prp18 might bind the spliceosome even before Prp16-mediated remodelling. We wondered whether such early binding of Prp18/ Slu7 explained how we could purify C complexes using affinity tags on Prp18 or Slu7 (refs. 7,8). Indeed, by focussed

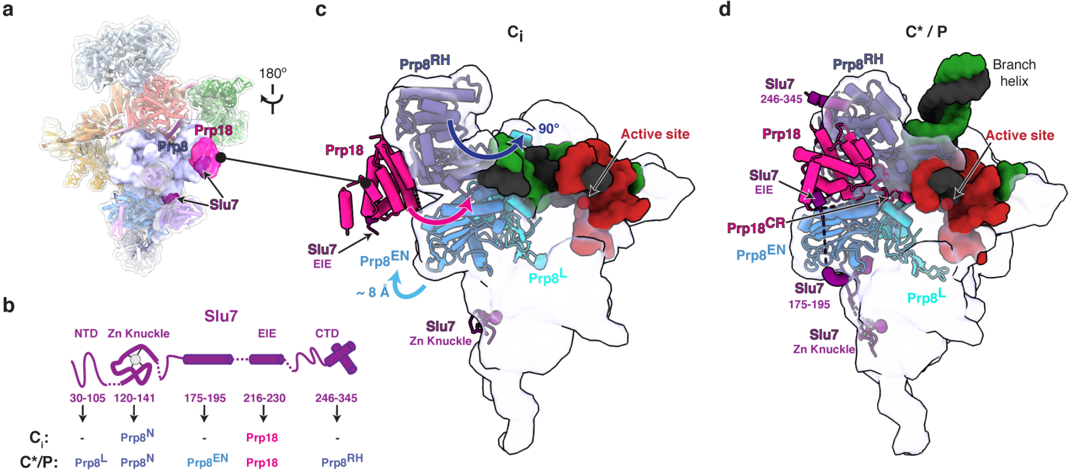


Figure 3 I Prp18 and Slu7 are recruited to a new C$_i$ conformation. a, Location of exon-ligation factors in the C-complex spliceosome. Prp8 and Prp18/ Slu7 are shown as surface representation. **b,** Domain architecture of Slu7 and the interaction partners of each domain. **c-d,** Pre-recruited Prp18 and Slu7 are remodelled from the C$_i$ to the C*/P conformation. Upward movement of the Prp8 EN domain creates additional binding sites for Slu7 in C*/P. The Prp18 conserved region (CR) is disordered in C$_i$ but engages the active site in C*/P.

classification we identified a subset of particles in the C-complex conformation with extra density on the same face of Prp8^{RH} that Prp18 binds in C* (**Figs. 1d, 3a, c, Extended Data Figs. 3, 8**). The alpha-helical domain of Prp18 visible in C* was a perfect fit for this additional density, which also accommodated a small peptide of Slu7 (termed EIE element, **Fig. 3b,c**) that is known to bind Prp18 (ref. 19). The conserved loop of Prp18 that stabilises the 3′-SS docked in the active site in C*/P is not visible in this subset of particles (Prp18^{CR}, ref. 9, **Fig. 3c,d**), suggesting it is disordered before Prp16 remodelling and 3′-SS docking. Biochemical experiments indicate that endogenous Prp18 and Slu7 may exist as a heterodimer[19,46] and thus early binding of Prp18 would also imply early binding of Slu7. Indeed, we also observed density corresponding to the Zn knuckle of Slu7 (ZnK) bound to Prp8 in this new intermediate (**Figs. 3b,c Extended Data Fig. 8**). This novel intermediate between C and C*, which we refer to as C$_i$, demonstrates that all branching and exon-ligation factors can be bound in one complex, in agreement with previous biochemical data[29].

To address the functional relevance of early binding of Prp18/ Slu7 before Prp16 remodelling, we investigated whether the C$_i$ intermediate could be purified independently of Prp18 and Slu7, using a branching factor that binds stably to the C complex before Prp16 action. We assembled native complexes on wild-type pre-mRNA and stalled the C complex with a cold-sensitive Prp16 mutation (*prp16-302*) that impairs ATP hydrolysis (C$^{prp16\text{-}302}$, **Fig. 4a**). We then purified the stalled C-complex spliceosomes with the branching factor Cwc25, which remains bound after branching when Prp16 action is blocked[12,27]. Cryo-EM reconstruction revealed that these C spliceosomes were structurally identical to the C complex assembled on the 3′-SS UAc mutant substrate (**Fig. 4b, Extended Data Figs. 2, 8**). Moreover, focussed classification identified a subset of particles in the C$_i$ conformation, in which Cwc25 and Prp18/ Slu7 were bound to the same complex (**Fig. 4c, Extended Data Figs. 3, 8**). Therefore, Prp18 and Slu7 can bind to the spliceosome before Prp16 action and the C$_i$ complex can be purified with a branching factor.

Overall, an average of 30% of C complexes resulting from various stalls during catalysis are in a C$_i$ conformation and contain Prp18 and Slu7. However, only the small EIE element of Slu7 that tightly interacts with Prp18 is observed in all C$_i$ particles while the ZnK of Slu7 engages only a subpopulation of C$_i$ complexes (**Fig. 1b**, **Extended Data Fig. 8**). Indeed, binding of Prp18 and of the Slu7 ZnK can occur independently of each other, but focussed classification indicated that binding of Prp18 significantly increases the odds for binding of the Slu7 ZnK, and vice versa (odds ratio = 1.35). This relationship was observed regardless of the stall used to obtain the C$_i$ intermediate, providing evidence that Prp18 and Slu7 interact as a heterodimer independently of their binding to the spliceosome (**Extended Data Fig. 8**). This mode of binding may allow early recruitment of Prp18 and/or Slu7 while ensuring that stable engagement of both exon-ligation factors with the spliceosome core does not occur until after Prp16 remodelling.

To determine whether C$_i$ is a functional on-pathway intermediate, we assayed whether C$^{prp16\text{-}302}$ complexes could produce mRNA when chased at the permissive temperature, as the *prp16-302* mutation allows growth and therefore splicing *in vivo* at 25°C (ref. 47). As expected, incubation of the purified C$^{prp16\text{-}302}$ complexes at 25°C without ATP or with non-hydrolysable ATP-γS did not allow exon ligation, demonstrating that these complexes are stalled before Prp16 has acted to allow remodelling to C* (**Fig. 4d,e**, ref. 30). Incubation with ATP and exogenous, recombinant Prp18/ Slu7 resulted in efficient exon ligation, showing the spliceosomes are

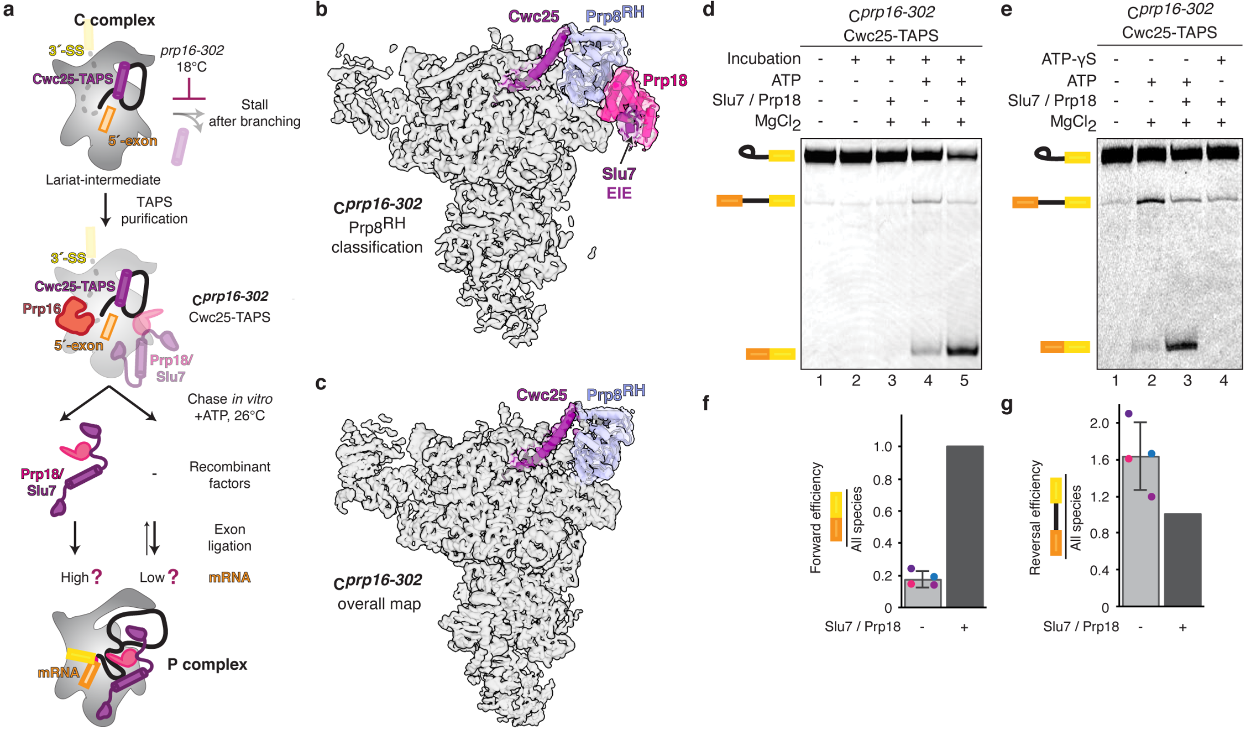

a
C complex
3′-SS
Cwc25-TAPS
5′-exon
prp16-302
18°C
Stall
after branching
Lariat-intermediate
TAPS
purification
3′-SS
Cwc25-TAPS
Prp16
5′-exon
Prp18/
Slu7
Cprp16-302
Cwc25-TAPS
Chase in vitro
+ATP, 26°C
Prp18/
Slu7
-
Recombinant
factors
Exon
ligation
High ?
Low ?
mRNA
mRNA
P complex
b
Cwc25
Prp8RH
Prp18
Slu7
EIE
Cprp16-302
Prp8RH
classification
c
Cwc25
Prp8RH
Cprp16-302
overall map
d
Cprp16-302
Cwc25-TAPS
Incubation - + + + +
ATP - - - + +
Slu7 / Prp18 - - + - +
MgCl2 - - + + +
1 2 3 4 5
e
Cprp16-302
Cwc25-TAPS
ATP-γS - - - +
ATP - + + -
Slu7 / Prp18 - - + +
MgCl2 - + + +
1 2 3 4
f
Forward efficiency
All species
1.0
0.8
0.6
0.4
0.2
0
Slu7 / Prp18 - +
g
Reversal efficiency
All species
2.0
1.6
1.2
0.8
0.4
0
Slu7 / Prp18 - +

Figure 4 | The novel C$_i$ conformation is an on-pathway intermediate. a, Biochemical assay to assess function of the C$_i$ conformation. Note that *prp16-302* impairs ATP hydrolysis at low temperatures and stalls Prp16 on the spliceosome right after branching. Chase of the purified C$^{prp16\text{-}302}$ complex at the permissive temperature is predicted to allow mRNA production even in the absence of recombinant exon ligation factors. **b-c,** Focussed classification reveals a subpopulation of C$^{prp16\text{-}302}$ spliceosomes in the C$_i$ conformation. Maps were reconstructed without refinement from the classified particles and filtered to 5Å to aid comparison of additional density for Prp18/ Slu7. **d,** Purified C$^{prp16\text{-}302}$ complexes produce mRNA when chased in the presence of ATP; the substrate was 3′-end labelled with Cy2 prior to spliceosome assembly. **e,** ATP hydrolysis is required for both exon ligation and branching reversal by C$^{prp16\text{-}302}$ spliceosomes. **f-g,** Quantification of exon ligation and branching reversal efficiency for C$^{prp16\text{-}302}$ spliceosomes. Relative ratios were normalised in each experiment to the condition where recombinant Slu7/ Prp18 was added. Values are averages of four independent experiments from three independent purifications; error bars represent s.d. (n=4). Values for individual experiments are indicated as dots overlaid on the bar graphs.

on-pathway intermediates. Importantly, incubation with ATP in the absence of exogenous Prp18/ Slu7 still produced significant levels of mRNA (**Fig. 4d-f**), demonstrating that the small C$_i$ population containing Prp18/ Slu7 (**Extended Data Fig. 8**) stalled by the *prp16-302* mutation represents a functional on-pathway intermediate. Since Prp18/ Slu7 are necessary for UBC4 splicing[26], we infer that the Prp18/ Slu7 pre-bound in C$_i$ are competent to promote exon ligation upon remodelling by Prp16. Intriguingly, in the absence of exogenous Prp18/ Slu7, C$^{prp16\text{-}302}$ complexes also catalysed reverse branching producing pre-mRNA (**Fig. 4d-g**). Thus, when Prp18 and Slu7 are limiting, the C$_i$ spliceosome can revert to the B* conformation, supporting the idea that Prp18 and Slu7 shift the equilibrium towards the C* conformation during remodelling[8,9,26].

Complete structure of the Prp19-associated complex (NTC) reveals basis for recruitment of branching factors. The NTC is one of the biggest pre-assembled components of the spliceosome and joins during formation of the active site triplex in the B^{act} complex[1,48]. The NTC stabilises the U6 triplex[6] and is essential for correct docking of the 5′-SS in the active site[49,50]. Moreover, branching factors were initially identified as proteins that interact loosely with the NTC (refs. 10,11,13,51,52). Previous yeast spliceosome structures revealed the overall architecture of the NTC as a sprawling clamp centred on a hinge formed by the tetrameric helical bundle of Prp19, from which Syf1 and Clf1 emerge as two helical arches that allow other components such as Cef1 to engage the spliceosome core (**Fig. 5a**). The Syf1 and Clf1 arches are very mobile during the C to C* transition[1,48] and this feature has reduced local resolution for these regions in previous spliceosome maps. Thus, most of Clf1 and Syf1 were previously built only as idealised alpha helices of uncertain register.

Our large C-complex dataset allowed focussed refinement that significantly improved density quality for Syf1 and Clf1 (**Extended Data Fig. 3**), showing surprisingly that several of the alpha helices previously assigned to Syf1 come from NTC proteins Syf2, Isy1, and Ntc20, and the U2 snRNP protein Lea1. We were able to build the N-terminal domain of Syf2 for the first time, revealing how Syf2 bridges Syf1 and U2/U6 helix II to anchor the NTC to active site elements in the core through Cef1 (**Fig. 5b**). Three helices on the surface of Clf1 were assigned to the N-termini of Prp45 and

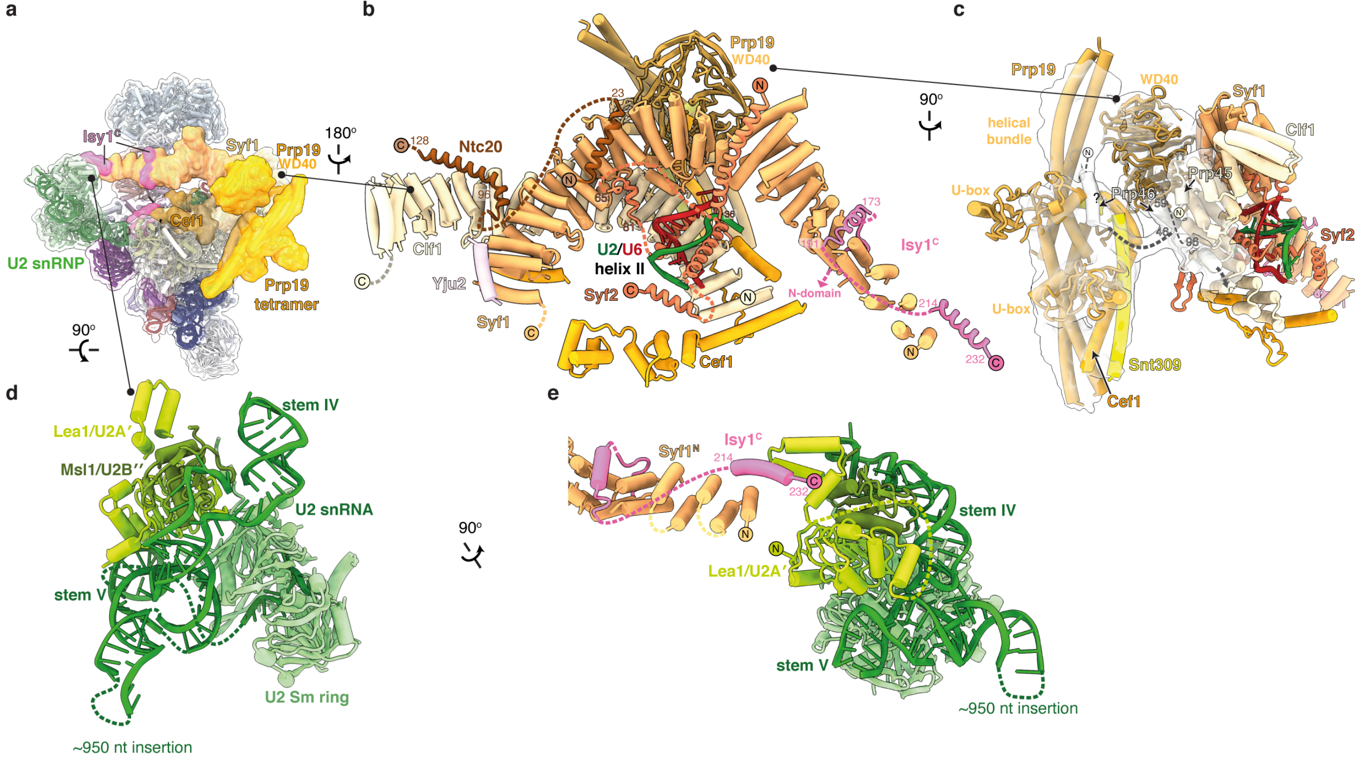


Figure 5 I Complete architecture of the NTC and U2 snRNP in C complex. a, Location of NTC and U2snRNP in C complex. **b,** Architecture of the NTC showing newly identified Ntc20, Syf2, and Isy1 elements. **c,** The N-termini of Prp45 and Prp46 mediate attachment of the Prp19 helical bundle to the spliceosome. Focus-classified Prp19 density is shown as transparent contour surface. **d,** Complete structure of the U2 snRNP in C complex. Note that yeast contain a large, flexible insertion in U2 snRNA that is not visible in our map. **e,** Isy1 bridges the NTC to U2 snRNP through Lea1.

Prp46, with the latter projecting towards the Prp19 helical bundle, possibly acting as a tether for this domain (**Fig. 5c**). Ntc20 was the only unassigned NTC protein and here we show that two helices from Ntc20 bind the C-terminal domains of Syf1 and Clf1, respectively (**Fig. 5b**, **Extended Data Fig. 5**), thus linking these arches during activation and catalysis, consistent with previous genetics and biochemistry experiments[51,52].

Our new map also explains how Isy1 (also known as Ntc30) is recruited to the spliceosome to act as a branching factor. The N-terminal domain of Isy1 was previously built in C complex as a clamp element that promotes docking of the branch helix at the active site[7,14]. We identified two helices from the C-terminus of Isy1 that bind the N-terminus of Syf1, explaining previous biochemical and genetic data implicating Isy1 as a peripheral NTC component[51]. Thus, the Isy1 N-terminal domain, which promotes branching, is likely already bound in B^{act} and tethered close to its site of action after Prp2 activity, in a manner parallel to Prp18 binding before exon ligation in C_i. Our improved model of the U2 snRNP (**Fig. 5a,d**) shows that the C-terminal helix of Isy1 also bridges the extreme N-terminus of Syf1 to a newly identified 3-helix bundle of the U2 snRNP component Lea1 (**Fig. 5d,e**). Thus, Isy1 acts as a link between the NTC and the U2 snRNP and this interaction may be maintained throughout the catalytic stage and may influence C to C* remodelling, although Isy1 has yet to be identified in maps of C* or P complexes.

Discussion

Our high-resolution structure of the C complex shows how the single active site of the spliceosome positions a trinuclear Mg^{2+} / K^{+} cluster for catalysis (**Fig. 2**). The active site remains catalytically licensed from B* onwards (**Extended Data Fig. 6**) but juxtaposes different reactants for each catalytic step. Although the 5′-exon remains bound onto U5 snRNA for both steps, the branch helix and brA must be removed from the active site after branching to make space for the 3′-SS to dock near the 5′-exon during exon ligation[1]. The ATPase Prp16 drives remodelling of brA interactions after branching to allow brA pairing to the 3′-SS, while Prp8 cradles the active site and governs a dynamic equilibrium between the branching and exon ligation conformations[2,23,24]. Step-specific factors also modulate this equilibrium[1,12,27-29]. The branching factors Cwc25 and Isy1 dissociate from the C-complex active site to allow exchange of reactants in the active site in C*/ P (ref. 1), while the exon-ligation factors Slu7 and Prp18 stabilise the C*/ P conformation to promote docking of the 3′-SS (ref. 1). Unexpectedly, we found that approximately 30% of C complex particles that retained branching factors (e.g. Cwc25) also contained the exon-ligation factor Prp18 (**Extended Data Fig. 8**), demonstrating that exon-ligation factors can bind the

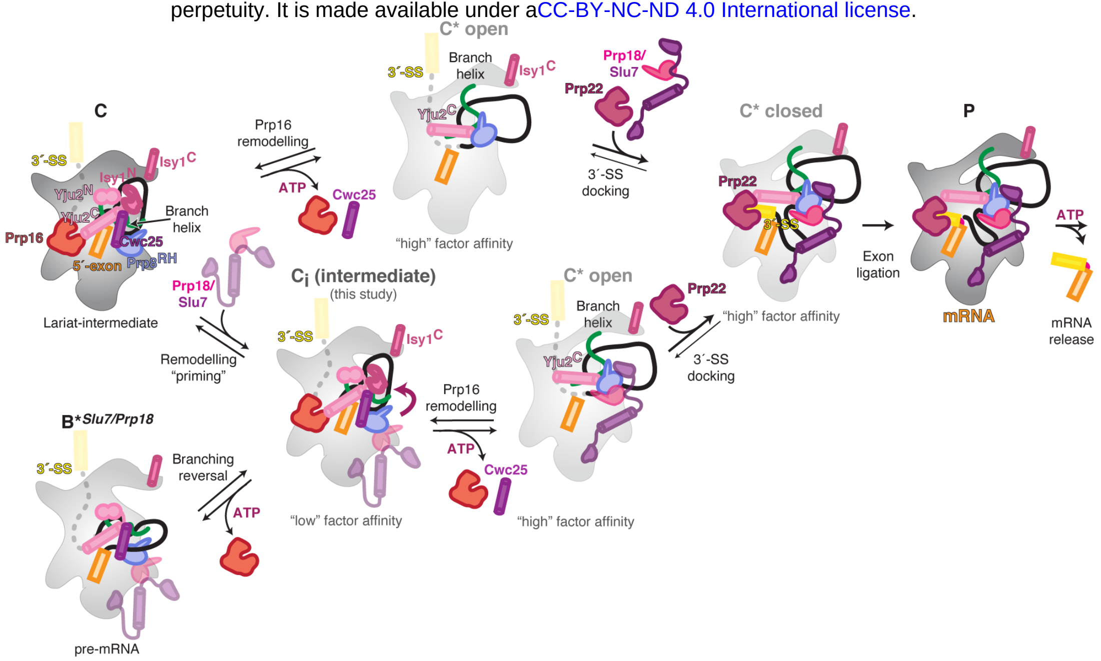

C
3′-SS
Isy1C
Isy1N
Yju2N
Yju2C
Prp16
Cwc25
Branch helix
5′-exon
Prp8RH
Lariat-intermediate
Prp16 remodelling
ATP
Cwc25
C* open
3′-SS
Branch helix
Isy1C
Yju2C
"high" factor affinity
Prp18/ Slu7
Prp22
3′-SS docking
C* closed
Prp22
3′-SS
Exon ligation
P
ATP
mRNA
mRNA release
Prp18/ Slu7
Remodelling "priming"
Ci (intermediate)
(this study)
3′-SS
Isy1C
"low" factor affinity
Prp16 remodelling
ATP
Cwc25
C* open
3′-SS
Branch helix
Yju2C
Prp22
"high" factor affinity
3′-SS docking
"high" factor affinity
B*Slu7/Prp18
3′-SS
Branching reversal
ATP
pre-mRNA

Figure 6 I Model for conformational equilibrium during the catalytic stage of *S. cerevisiae* pre-mRNA splicing. Following branching, the exon ligation factors Slu7 and Prp18 can bind the C conformation with low affinity by interacting with the Prp8 RNase H domain (Prp8^{RH}) to prime remodelling, thus forming the C$_i$ intermediate. Prp16-mediated remodelling of this C$_i$ complex leads to a C* 'open' conformation, with high affinity for Slu7/ Prp18. Dissociation of Cwc25 from C$_i$ allows the Yju2 C-terminal domain (Yju2^C) to interact with Prp8^{RH} and stabilise binding of Prp22. The undocked branch helix is locked in a C* 'closed' conformation by the stable binding of Prp18 and Slu7 and by Yju2^C. The C* closed conformation is competent for stable 3′-SS docking, allowing exon ligation. The C$_i$ complex is likely unstable and during Prp16 activity, in the absence of additional Slu7/ Prp18 binding, remodelling can also allow C$_i$ to catalyse reversal of branching, thus reverting to a B* conformation in which Slu7/ Prp18 may remain bound (B*$^{Slu7/Prp18}$). Remodelling of C complex may also occur through a pathway that does not involve C$_i$, where Prp16 action precedes, or is concomitant with, recruitment of Slu7/ Prp18 in the C* 'open' conformation.

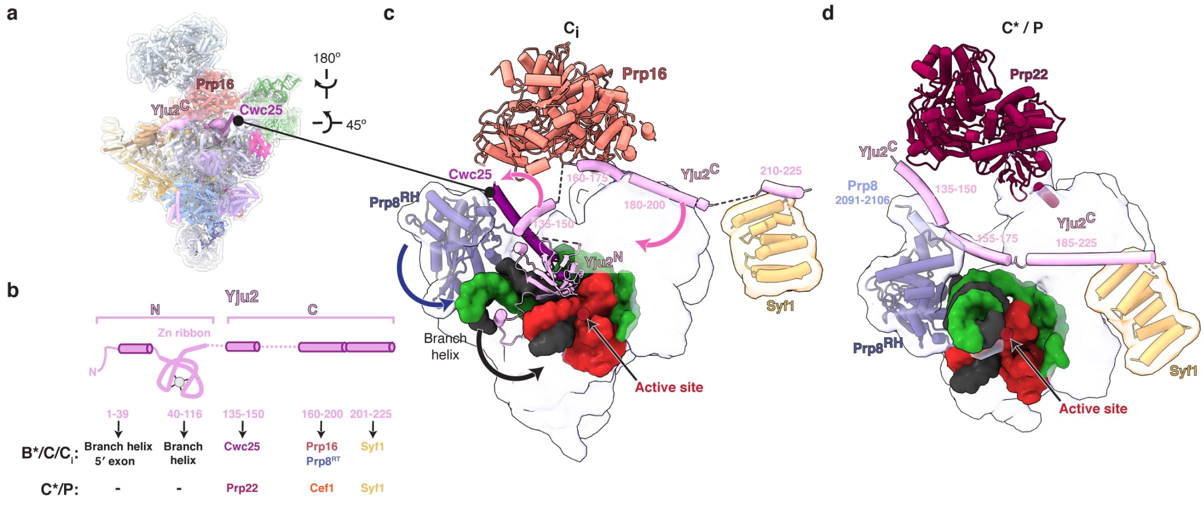

a
180°
45°
Prp16
Yju2C
Cwc25
b
Yju2
N
C
Zn ribbon
N
1-39
40-116
135-150
160-200
201-225
B*/C/Ci:
Branch helix 5′ exon
Branch helix
Cwc25
Prp16 Prp8RT
Syf1
C*/P:
-
-
Prp22
Cef1
Syf1
c
Ci
Prp16
Cwc25
Prp8RH
160-175
180-200
Yju2C
210-225
135-150
Yju2N
Syf1
Branch helix
Active site
d
C* / P
Prp22
Yju2C
Prp8 2091-2106
135-150
155-175
Yju2C
185-225
Prp8RH
Syf1
Active site

Figure 7 I Remodelling of Yju2 interactions during the C to C* transition. a, Location of the branching factor Yju2 in the C-complex spliceosome. **b,** Domain architecture of Yju2 and its interaction partners during branching and exon ligation. **c,** Yju2 supports binding of Cwc25 and Prp16 in C complex. **d,** Yju2 is remodelled in C*/P to act as exon-ligation factor. Note that dissociation of Cwc25 is required for formation of a new Prp8 helix that supports Yju2-mediated recruitment of Prp22 in C*/P.

branching conformation before Prp16 remodelling. This new C$_i$ intermediate, in which branching and exon-ligation factors bind the same complex (**Figs. 1e, 6**), provides the structural basis for how the spliceosome toggles between the branching and exon-ligation conformations.

In C$_i$, Prp18 binds the Prp8^{RH} domain before Prp16 remodelling, and potentially before branching catalysis[29]. Slu7 is recruited to C$_i$ through interactions of its EIE element with Prp18 or through binding of its zinc knuckle domain to the N-terminal domain of Prp8 (**Fig. 6**). Binding of the remainder of Slu7 is prevented in C$_i$ by Cwc25, which would clash with the C-terminal domain of Slu7, while engagement of the central region of Slu7 requires upward movement of the Prp8^{EN} domain in C* (**Fig. 3**). Our C$_i$ structure explains why previous studies have observed binding of Slu7 and Prp18 before Prp16 action at low salt concentrations and thus proposed a weaker affinity site before remodelling by Prp16 (ref. 29). In fact, our data suggest that exon-ligation factors bind at the same sites before and after Prp16 action but their affinity likely increases following remodelling by Prp16 (**Fig. 6**).

Consistently, at least 30% of C complexes stalled with the conditional *prp16-302* allele and purified at low salt concentrations contain Prp18 in the C_i state (**Extended Data Fig. 8**). These complexes can be reactivated to produce mRNA upon ATP hydrolysis by Prp16 even in the absence of exogenous Slu7 and Prp18 added following purification (**Fig. 4**). We conclude that C_i is a key on-pathway intermediate in remodelling between branching and exon ligation (**Fig. 6**): Prp18 and Slu7 pre-bound in C_i can stably engage the spliceosome to stabilise the higher-energy C* complex and promote exon ligation following ATP hydrolysis by Prp16 (**Figs. 4d-g,6**).

Binding of branching factors is also remodelled during Prp16 activity. In B*/C complexes, branching is promoted by engagement of the N-termini of Cwc25, Yju2, and Isy1 with the branch helix and brA, while the C-termini of these branching factors interact with the NTC and Prp16. Two C-terminal helices of Yju2 cross Cwc25 before interacting with the NTC, forming a unique binding platform for Prp16 at the branching stage[15,53] (**Fig. 7a-c**). These Yju2 helices are reorganised in C*/P complex to instead bridge the rotated Prp8 RNaseH domain to Prp22, while the Yju2 N-domain dissociates (**Fig. 6d**). While the N-terminus of Yju2 is essential for viability and promotes branching, deletion of the C-terminus allows only inefficient exon ligation in the absence of Prp16 (ref.54), suggesting that Yju2 stabilises both C and C* in a manner consistent with our structural model (**Figs. 6, 7**). Interestingly, we show here that the C-terminus of Isy1 also binds the NTC, gluing the interface between Syf1 and the U2 snRNP (**Fig. 5**). Although detailed structures of the NTC are not yet available in B^{act} and C*/P complexes, this interaction is likely preserved after remodelling to C* complex, while the Isy1 N-domain similarly dissociates. Thus, paralleling the primed association of exon-ligation factors with the branching conformation in C_i complex, the branching factors Isy1 and Yju2 are both associated with the exon-ligation conformation in C* and P complex. Indeed, tethering of Isy1 by the NTC throughout catalysis explains how Isy1 affects proofreading of both the branch site and 3′-SS (ref. 13).

The C_i intermediate indicates that complexes in the branching conformation (B*/ C/ C_i) differ little in composition from those in the exon-ligation conformation (C*/ P). Only Cwc25 dissociates from branching to exon-ligation, while all other factors can already bind in the C conformation and are remodelled as a result of Prp16 action (**Fig. 6**). Indeed, the C conformation is particularly stable, as complexes that cannot complete catalysis appear to revert to this state. C complexes accumulate not only when Prp16 remodelling is blocked but also when exon ligation is impaired by mutation of the 3′-SS, or when mRNA release is prevented by blocking Prp22 activity (**Extended Data Fig. 2**). Therefore, the energy of ATP hydrolysis by Prp16 is necessary to disrupt this stable conformation and establish an equilibrium between branching and exon ligation that underlies reversibility of both splicing reactions[55] and is likely necessary for proofreading during catalysis.

The C_i complex structure shows how exon-ligation factors can prime remodelling before Prp16 activity and shift the equilibrium from branching to exon ligation after Prp16 hydrolyses ATP. By stabilising the higher-energy C* conformation, exon-ligation factors act like a Brownian ratchet, similarly to how branching factors drive the equilibrium towards B* during spliceosome activation by Prp2 (ref. 56, cf. ref. 26). Specific branching factors further modulate this equilibrium through partial dissociation of their N-terminal domains as a result of Prp16 action (**Fig. 6**). Conversely, when Slu7 and Prp18 are limiting, the spliceosome cannot become 'locked' into its exon-ligation conformation and Prp16 action therefore allows the C-complex spliceosome to reverse branching and revert to the B* conformation (**Figs. 4d-g, 6**). Such reversibility may be important during proofreading of branching by Prp16 (refs. 26,57) and suggests that Slu7 and Prp18 may affect catalysis and proofreading more broadly, by modulating the equilibrium between B*/ C/ C_i and C*/ P complexes. In support of this idea, we observed the C_i conformation when spliceosomes were stalled with a 3′-SS mutant that docks poorly in the active site but can produce mRNA when proofreading is disabled[28,58], suggesting that proofreading of exon ligation by Prp22 can lead to substrate rejection and collapse to the C_i conformation without dissociation of exon-ligation factors (**Fig. 6**, **Extended Data Fig. 8**).

Overall, Prp16 acts as a classical catalyst by reducing the activation barrier to transition from the very stable branching conformation to the higher-energy exon-ligation conformation, thus allowing thermodynamic control of splicing, a feature likely necessary for proofreading of catalysis. Indeed, genetic studies suggest that the stability of RNA elements that form the active site is also disrupted during Prp16 action in a manner that influences splice site proofreading[6,59].

Our model of the catalytic stage involving the novel C_i conformation (**Fig. 6**) explains how the spliceosome uses the energy of ATP hydrolysis to remodel binding of step-specific protein factors and to modulate substrate transactions at a single RNA-based active site, thus allowing conformational equilibrium and proofreading during splicing catalysis.

References

1. Wilkinson, M. E., Charenton, C. & Nagai, K. RNA Splicing by the Spliceosome. *Annual Review of Biochemistry* **89,** 1–30 (2019).
2. Fica, S. M. & Nagai, K. Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. *Nat Struct Mol Biol* **24,** 791–799 (2017).
3. 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).
4. Newman, A. J. & Norman, C. U5 snRNA interacts with exon sequences at 5‘ and 3’ splice sites. *Cell* **68,** 743–754 (1992).
5. Fica, S. M. *et al*. RNA catalyses nuclear pre-mRNA splicing. *Nature* **503,** 229–234 (2013).
6. 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).
7. Galej, W. P. *et al*. Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537,** 197–201 (2016).
8. Fica, S. M. *et al*. Structure of a spliceosome remodelled for exon ligation. *Nature* **542,** 377–380 (2017).
9. Wilkinson, M. E. *et al*. Postcatalytic spliceosome structure reveals mechanism of 3'-splice site selection. *Science* **358,** 1283–1288 (2017).
10. 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).
11. Liu, Y. C., Chen, H. C., Wu, N. Y. & Cheng, S. C. A Novel Splicing Factor, Yju2, Is Associated with NTC and Acts after Prp2 in Promoting the First Catalytic Reaction of Pre-mRNA Splicing. *Mol. Cell. Biol.* **27,** 5403–5413 (2007).
12. Warkocki, Z. *et al*. Reconstitution of both steps of Saccharomyces cerevisiae splicing with purified spliceosomal components. *Nat Struct Mol Biol* **16,** 1237–1243 (2009).
13. Villa, T. The Isy1p component of the NineTeen Complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing. *Genes Dev* **19,** 1894–1904 (2005).
14. 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).
15. Wan, R., Bai, R., Yan, C., Lei, J. & Shi, Y. Structures of the Catalytically Activated Yeast Spliceosome Reveal the Mechanism of Branching. *Cell* **177,** 339–351.e13 (2019).
16. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast step II catalytically activated spliceosome. *Science* **355,** 149–155 (2017).
17. Bai, R., Yan, C., Wan, R., Lei, J. & Shi, Y. Structure of the Post-catalytic Spliceosome from Saccharomyces cerevisiae. *Cell* **171,** 1589–1598.e8 (2017).
18. Liu, S. *et al*. Structure of the yeast spliceosomal postcatalytic P complex. *Science* **358,** 1278–1283 (2017).
19. 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).
20. Brys, A. & Schwer, B. Requirement for SLU7 in yeast pre-mRNA splicing is dictated by the distance between the branchpoint and the 3' splice site. *RNA* **2,** 707–717 (1996).
21. Haselbach, D. *et al*. Structure and Conformational Dynamics of the Human Spliceosomal Bact Complex. *Cell* **172,** 454–464.e11 (2018).
22. Jones, M. H., Frank, D. N. & Guthrie, C. Characterization and functional ordering of Slu7p and Prp17p during the second step of pre-mRNA splicing in yeast. *Proc Natl Acad Sci USA* **92,** 9687–9691 (1995).
23. Query, C. C. & Konarska, M. M. Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants. *Mol Cell* **14,** 343–354 (2004).
24. Liu, L., Query, C. C. & Konarska, M. M. Opposing classes of prp8 alleles modulate the transition between the catalytic steps of pre-mRNA splicing. *Nat Struct Mol Biol* **14,** 519–526 (2007).
25. Tseng, C. K., Liu, H. L. & Cheng, S. C. DEAH-box ATPase Prp16 has dual roles in remodelling of the spliceosome in catalytic steps. *RNA* **17,** 145–154 (2010).
26. 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).
27. Tseng, C.-K., Chung, C.-S., Chen, H.-C. & Cheng, S.-C. A Central Role of Cwc25 in Spliceosome Dynamics during Catalytic Phase of Pre-mRNA Splicing. *RNA* rna.059204.116 (2017). doi:10.1261/rna.059204.116
28. Tseng, C. K. & Cheng, S. C. Both Catalytic Steps of Nuclear Pre-mRNA Splicing Are Reversible. *Science* **320,** 1782–1784 (2008).
29. 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).
30. Schwer, B. & Guthrie, C. A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *EMBO J* **11,** 5033–5039 (1992).
31. Fabrizio, P. *et al*. The Evolutionarily Conserved Core Design of the Catalytic Activation Step of the Yeast Spliceosome. *Mol Cell* **36,** 593–608 (2009).
32. Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci USA* **90,** 6498–6502 (1993).
33. Marcia, M. & Pyle, A. M. Visualizing Group II Intron Catalysis through the Stages of Splicing. *Cell* **151,** 497–507 (2012).
34. Galej, W. P., Toor, N., Newman, A. J. & Nagai, K. Molecular Mechanism and Evolution of Nuclear Pre-mRNA and Group II Intron Splicing: Insights from Cryo-Electron Microscopy Structures. *Chem. Rev.* **118,** 4156–4176 (2018).
35. Haack, D. B. & Toor, N. Retroelement origins of pre−mRNA splicing. *WIREs RNA* **255,** 269 (2020).
36. 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).
37. 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).
38. 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).
39. Lin, R. J., Newman, A. J., Cheng, S. C. & Abelson, J. Yeast mRNA splicing in vitro. *J Biol Chem* **260,** 14780–14792 (1985).
40. Auffinger, P., Grover, N. & Westhof, E. in *Structural and Catalytic Roles of Metal Ions in RNA* **1,** 1–36 (Royal Society of Chemistry, 2011).
41. 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).
42. Lesser, C. F. & Guthrie, C. Mutational analysis of pre-mRNA splicing in Saccharomyces cerevisiae using a sensitive new reporter gene, CUP1. *Genetics* **133,** 851–863 (1993).
43. McPheeters, D. S. Interactions of the yeast U6 RNA with the pre-mRNA branch site. *RNA* **2,** 1110–1123 (1996).
44. Leontis, N. B., Stombaugh, J. & Westhof, E. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* **30,** 3497–3531 (2002).
45. Mayerle, M. & Guthrie, C. Genetics and Biochemistry Remain Essential in the Structural Era of the Spliceosome. *Methods* **125,** 3–9 (2017).
46. Zhang, X. & Schwer, B. Functional and physical interaction between the yeast splicing factors Slu7 and Prp18. *Nucleic Acids Res.* **25,** 2146–2152 (1997).
47. Madhani, H. D. & Guthrie, C. Genetic interactions between the yeast RNA helicase homolog Prp16 and spliceosomal snRNAs identify candidate ligands for the Prp16 RNA-dependent ATPase. *Genetics* **137,** 677–687 (1994).
48. Plaschka, C., Newman, A. J. & Nagai, K. Structural Basis of Nuclear pre-mRNA Splicing: Lessons from Yeast. *Cold Spring Harbor perspectives in biology* **11,** a032391 (2019).
49. Chan, S. P. The Prp19p-Associated Complex in Spliceosome Activation. *Science* **302,** 279–282 (2003).
50. Chan, S. P. The Prp19-associated Complex Is Required for Specifying Interactions of U5 and U6 with Pre-mRNA during Spliceosome Activation. *Journal of Biological Chemistry* **280,** 31190–31199 (2005).
51. Chen, C. H., Tsai, W. Y., Chen, H. R., Wang, C. H. & Cheng, S. C. Identification and characterization of two novel components of the Prp19p-associated complex, Ntc30p and Ntc20p. *J Biol Chem* **276,** 488–494 (2001).
52. Chang, K.-J., Chen, H.-C. & Cheng, S.-C. Ntc90 is required for recruiting first step factor Yju2 but not for spliceosome activation. *RNA* **15,** 1729–1739 (2009).
53. Zhang, X. *et al*. An Atomic Structure of the Human Spliceosome. *Cell* **169,** 918–929.e14 (2017).
54. Chiang, T.-W. & Cheng, S.-C. A weak spliceosome-binding domain of Yju2 functions in the first step and bypasses Prp16 in the second step of splicing. *Mol. Cell. Biol.* **33,** 1746–1755 (2013).
55. Tseng, C.-K. & Cheng, S.-C. Both Catalytic Steps of Nuclear Pre-mRNA Splicing Are Reversible. *Science* **320,** 1782–1784 (2008).
56. Krishnan, R. *et al*. Biased Brownian ratcheting leads to pre-mRNA remodelling and capture prior to first-step splicing. *Nat Struct Mol Biol* **20,** 1450–1457 (2013).

57. Burgess, S. M. & Guthrie, C. A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. *Cell* **73,** 1377–1391 (1993).
58. Mayas, R. M., Maita, H. & Staley, J. P. Exon ligation is proofread by the DExD/H-box ATPase Prp22p. *Nat Struct Mol Biol* **13,** 482–490 (2006).
59. Eysmont, K., Matylla-Kulińska, K., Jaskulska, A., Magnus, M. & Konarska, M. M. Rearrangements within the U6 snRNA Core during the Transition between the Two Catalytic Steps of Splicing. *Mol Cell* **75,** 538–548 (2019).

Data availability

The cryo-EM maps have been deposited in the Electron Microscopy Data Bank with accession codes EMD-NNNNN, EMD-NNNNN, EMD-NNNNN, EMD-NNNNN. The coordinates of the atomic model have been deposited in the Protein Data Bank under accession code XXXX.

For immediate access, the composite map and model, as well as a PyMol session can be downloaded from the following folder, using "Connect As: Guest" if prompted for a password:
ftp://ftp.mrc-lmb.cam.ac.uk/pub/mwilkin/Ci_spliceosome/

Acknowledgements

We thank K. Nguyen for a gift of Slu7/ Prp18 heterodimer used for preliminary tests; S. Scheres, X.C. Bai, C. Savva, S. Chen, K. R. Vinothkumar, and G. Cannone for help and advice on data collection and processing; G. McMullan, J. Grimmett, and T. Darling for smooth running of the EM and computing facilities; the staff at Diamond Light Source (DLS) for help with data collection; A. Amunts and S. Aibara for help with data collection at SciLifeLab; the mass spectrometry facility for help with protein identification; P. Emsley and G. Murshudov for help and advice with model building and refinement; the members of the spliceosome group for help and advice throughout the project. We thank J. Löwe, D. Barford and S. Scheres for their continuing support, C. Charenton, P-C. Lin, S. Lövestam, A. Newman, and S. Routh for critical reading of the manuscript, and J. Staley for a gift of reagents. The project was supported by the Medical Research Council (MC_U105184330) and a European Research Council Advanced grant (SPLICE3D). S.M.F. was supported by EMBO and Marie Skłodowska-Curie fellowships; M.E.W. was supported by a Rutherford Memorial Cambridge Scholarship.

Author contributions

S.M.F. and W.P.G. established the method for C complex preparation. S.M.F. and M.E.W. established the method for P complex preparation. S.M.F. designed and performed $C^{prp16\text{-}302}$ experiments including sample and grid preparation and biochemistry. M.E.W. and W.P.G. prepared C complex sample and grids. M.E.W prepared P complex sample and grids. M.E.W., S.M.F., and W.P.G. collected and performed initial processing of EM data. M.E.W. merged and re-processed all datasets, carried out model building and refinement and prepared the data for deposition. M.E.W. and S.M.F. wrote the manuscript with input from all authors. K.N. initiated and coordinated the spliceosome project.

Methods

Cloning and protein expression

A TAPS tag cassette was added in frame to the C-terminus of the endogenous CWC25 locus in *S. cerevisiae* yJPS983 (carrying a chromosomal *prp16-302* allele[6]), using the *kanMX6* resistance cassette[60]. *prp16-302* encodes the *prp16R456K* in motif Ib, which significantly reduces ATP hydrolysis at low temperatures and stalls spliceosomes after branching[61]. For extract preparation the resulting strain was grown normally at 30°C in YEPD media in a batch fermenter to an optical density OD_{600} ~ 2.4 – 3.2.

A construct encoding Prp18 and Slu7, which form a heterodimer[19], was used for recombinant expression in *E. coli* BL21 (DE3) RIL cells. The expressed heterodimer was purified via Ni-NTA-agarose and gel filtration chromatography.

$C^{prp16\text{-}302}$ purification and biochemistry

Spliceosomes were assembled on a modified UBC4 pre-mRNA with 25 nt exons[5] in extracts from *prp16-302* / *CWC25*-TAPS by incubation at 19°C for 30 minutes. Following splicing, reactions were incubated at 19°C for another 15 minutes in the presence of 0.2 μM of a DNA oligo directed against the 5′-splice site, to degrade unassembled pre-mRNA, and 2 mM glucose to deplete ATP and minimise Prp16 activity. Reactions were centrifuged through a 40% glycerol cushion in buffer K-75 (20 mM HEPES KOH pH 7.9, 75 mM KCl, 0.25 mM EDTA, 0.05% NP-40) and assembled complexes were affinity purified from the cushions using lgG Sepharose 6 Fast Flow (GE Healthcare) to capture the Protein A tag. Following extensive washing with buffer K-75, complexes were eluted from the beads by incubation with 25 ug/ mL TEV protease (expressed in house) by incubation at room temperature (22-23°C) for one hour in the presence of 1 mM DTT, and the protease was removed by subsequent concentration through a 50 kDa MWCO Amicon concentrator. Spliceosomes were further purified via the Strep II tag on Cwc25 using Streptactin affinity resin (GE) in buffer K-75 (20 mM HEPES KOH pH 7.9, 75 mM KCl, 0.25 mM EDTA, 5% glycerol, 0.05% NP-40) and eluted with 5 mM desthiobiotin. Eluted complexes were concentrated in a 100 kDa MWCO Amicon concentrator to ~ 5-10 nM, buffer exchanged into buffer K-75 without glycerol (20 mM HEPES KOH pH 7.9, 75 mM KCl, 0.25 mM EDTA, 0.0025% NP-40), and used for EM data collection or biochemistry. This sample was used for EM data set 4.

For biochemical assays, purified C complexes were chased at 26°C for 60 minutes in reactions containing 10% concentrated spliceosomes (~100 fmoles, 2 nM, in K-75 without glycerol), 3% PEG_{8000}, 60 mM potassium phosphate buffer (pH 7.0), 2 mM ATP (or ATP- γS), and 4 mM $MgCl_2$. Specific reactions were also supplemented with 250 nM Slu7/ Prp18 heterodimer. The data in Fig. 4 were quantified using an Amersham Typhoon imaging system and ImageQuant TL. The rolling ball algorithm was used for background subtraction. Bands for individual splicing species were first normalised to the total signal from all splicing species in each lane to obtain percentages, which were then used for calculating the efficiency of exon ligation and of pre-mRNA formation by reversal of branching from the lariat-intermediate. Quantifications in Fig. 4e,f were obtained from four independent spliceosome purifications using extracts from two independent biological replicates (independent extract preparations from two different batches of yeast fermenter growth).

For EM data collection following biochemical chase, purified C complexes were incubated at 26°C with 2 mM ATP and 2.5 mM $MgCl_2$ for 15 minutes prior to grid preparation. This sample was used for EM dataset 5.

For both un-chased and chased complexes, the purified sample was applied to Cu 300 R1.2/1.3 holey carbon grids (Quantifoil) coated with a ~6 nm homemade carbon film. Grids were glow discharged for 30 seconds before application of 3.5 µL sample, then incubated for 25 s and blotted for 2.5-3.5 s before vitrification by plunging into liquid ethane using an FEI Vitrobot MKIII operated at 100% humidity and 4°C.

P complex purification

Dataset 6 described in this paper derives from our early attempts to purify P complex. mRNA release was stalled by a dominant negative Prp22 mutant (K512A) and spliceosomes were purified first by mRNA pulldown then by Slu7-TAPS pulldown. This strategy produces a mixture of C and C*/P complex and predates our discovery that RNaseH cleavage could be used to significantly enrich P complex in the final sample[9] (**Extended Data Fig. 1**). In detail, Slu7-TAPS splicing extract was prepared as described[7]. Splicing extracts were treated on ice for 10 min with 0.04 mg/mL recombinantly-expressed Prp22 K512A. Spliceosomes were assembled in this extract on a *UBC4* pre-mRNA with two MS2 hairpins pre-bound to MS2-MBP fusion protein as described[62] by incubation at 23°C for 30 minutes. Following splicing, reactions were centrifuged through a 40% glycerol cushion in buffer K-75 (20 mM HEPES KOH pH 7.9, 75 mM KCl, 0.25 mM EDTA, 0.05% NP-40) and assembled complexes were affinity purified from the cushions using amylose resin (New England Biolabs). Following washing with buffer K-75, complexes were eluted from the beads with 12 mM maltose. Spliceosomes were further purified via the Strep-II tag on Slu7 using Streptactin affinity resin (GE) in buffer K-100 (20 mM HEPES KOH pH 7.9, 100 mM KCl, 0.25 mM EDTA, 5% glycerol, 0.01% NP-40) and eluted with 5 mM desthiobiotin. Eluted complexes were concentrated in a 100 kDa MWCO Amicon concentrator and dialysed into buffer K-75 without glycerol (20 mM HEPES KOH pH 7.9, 75 mM KCl, 0.25 mM EDTA). This sample was used for EM dataset 6: the purified sample was applied to Cu 300 R1.2/1.3 holey carbon grids (Quantifoil) coated with a ~7 nm homemade carbon film. Grids were glow discharged for 30 seconds before application of 3 µL sample, then incubated for 30 s and blotted for 2.5 s before vitrification by plunging into liquid ethane using an FEI Vitrobot MKIII operated at 100% humidity and 4°C

Cryo-EM data acquisition

Eight data sets and a total of 24,115 movies were collected manually or using EPU on various Titan Krios microscopes (Thermo Fisher) all equipped with energy filters (slit width of 20 eV) and K2 detectors operated in counting or super-resolution mode (**Extended Data Fig. 2**). Dataset 1 (2,213 movies) is described[7] and was collected manually on LMB Krios 1 in super-resolution mode with a real pixel size of 1.427 Å/pix, a defocus range of -0.5 to -4 µm, with an exposure time of 16 s fractionated into 20 frames and a total dose per micrograph of 40 $e^-/Å^2$. Datasets 2 and 3 were described previously[8]. Dataset 2 (1,571 movies) was collected manually on LMB Krios 1 in super-resolution mode with a real pixel size of 1.427 Å/pix, a defocus range of -0.5 to -4.5 µm, with an exposure time of 16 s fractionated into 20 frames and a total dose per micrograph of 40 $e^-/Å^2$.

Dataset 3 (2,025 movies) was collected using EPU on Diamond Light Source (DLS) Krios 1 in counting mode with a pixel size of 1.031 Å/pix, a defocus range of -0.5 to -3.5 µm, with an exposure time of 14 s fractionated into 28 frames and a total dose per micrograph of 42 $e^-/Å^2$.

Dataset 4 (2,944 movies) was collected using EPU on Diamond Light Source (DLS) Krios 2 in counting mode with a pixel size of 1.023 Å/pix, a defocus range of -0.5 to -3.5 µm, with an exposure time of 14 s fractionated into 28 frames and a total dose per micrograph of 44 $e^-/Å^2$.

Dataset 5 (2,523 movies) was collected using EPU on LMB Krios 1 in counting mode with a pixel size of 1.12 Å/pix, a defocus range of -0.5 to -3 µm, with an exposure time of 14 s fractionated into 28 frames and a total dose per micrograph of 56 $e^-/Å^2$

Dataset 6 (4,369 movies) was collected using EPU on a Krios at SciLifeLab (Stockholm, Sweden) in counting mode with a pixel size of 1.028 Å/pix, a defocus range of -0.5 to -3.5 µm, with an exposure time of 8 s fractionated into 20 frames and a total dose per micrograph of 39.7 $e^-/Å^2$.

Dataset 7 (2,384 movies) is described in ref. 9 and was collected using EPU on LMB Krios 1 in counting mode with a pixel size of 1.12 Å/pix, a defocus range of -0.2 to -3 µm, with an exposure time of 12 s fractionated into 20 frames and a total dose per micrograph of 47 $e^-/Å^2$.

Dataset 8 (1,614 movies) is described in ref. 63 and was collected using EPU on LMB Krios 1 in counting mode with a pixel size of 0.88 Å/pix, a defocus range of -0.5 to -3.5 µm, with an exposure time of 7 s fractionated into 35 frames and a total dose per micrograph of 45.2 $e^-/Å^2$.

Dataset 9 was of a P complex sample prepared identically to ref. 9 except that the final sample buffer contained 1 mM $MgCl_2$ instead of 0.25 mM EDTA. 4830 micrographs were collected using EPU on a Diamond Light Source Krios in counting mode with a pixel size of 1.03 Å/pix, a defocus range of -0.5 to -3 µm, with an exposure time of 12 s fractionated into 40 frames and a total dose per micrograph of 49.2 $e^-/Å^2$.

Initial cryo-EM data processing

All datasets were initially processed separately (**Extended Data Fig. 2**). Dataset 1 was processed as described previously[7] yielding a 3.8 Å reconstruction of C complex (EMD-4055). Re-refining these particles with a mask on the spliceosome core improved the resolution to 3.61 Å. Datasets 2 and 3 were processed as described previously[8], yielding both a reconstruction of C* complex at 3.8 Å resolution (EMDB-3539) and reconstructions of C complex at 4.62 Å resolution (dataset 2) and 3.61 Å resolution (dataset 3).

Datasets 4 and 5 were processed similarly: particles were picked using RELION auto-pick using representative 2D class averages of C complex as templates. After 3D classification to select good quality particles, particles were polished using the method implemented in RELION 1.3 (ref. 64, i.e. not the Bayesian polishing routine of RELION 3.0). Masked refinement then yielded C complex reconstructions at 3.34 Å resolution (dataset 4) and 3.37 Å resolution (dataset 5).

Dataset 6 was processed similarly to datasets 4 and 5 except that due to splicing being stalled prior to mRNA release, approximately half the particles were in the exon-ligation conformation (C*/P) and half were in the branching conformation (C). 3D classification was used to resolve these different populations, and each was subjected to particle polishing in RELION 2.0, producing a 3.58 Å resolution reconstruction of C*/P complexes, and 3.42 Å resolution reconstruction of C complex.

Dataset 7 (ref. 9) was reprocessed in RELION 3.1. After motion-correction with dose-weighting, particles were picked with crYOLO (ref. 65) using a model trained on the dataset. After 3D classification to select good P complex particles, particles were subject to CTF refinement and Bayesian polishing, producing a P complex reconstruction at 3.50 Å resolution.

Dataset 8 was processed as described in ref. 63, yielding a 3.56 Å reconstruction of P complex.

Dataset 9 processed in RELION 3.1. After motion-correction with dose-weighting, particles were picked with crYOLO (ref. 65) using a general model. After 3D classification to select good P complex particles, particles were subject to CTF refinement and Bayesian polishing, producing a P complex reconstruction at 3.13 Å resolution.

C complex data processing

Initial refinement. C complex data were merged using RELION 3.1, using separate optics groups for datasets 1, 2, 3, 4, 5, and 6 (**Extended Data Fig. 2**). Dataset 1 was further split into 3 optics groups as this dataset was collected over three different microscope sessions. The individual reconstructions were compared with UCSF Chimera to determine the relative pixel sizes of each dataset, and defocus values were scaled by the relative difference in pixel sizes[63]. Unscaled particles were then refined together with a mask on the ordered core, with RELION 3.1 internally scaling the particles to match a 1.12 Å/pix reference in a 400 pixel box. Two rounds of CTF refinement (refining per-particle defocus, per-micrograph astigmatism and B-factor, and per-optics group anisotropic magnification, beam tilt, trefoil, and 4th order aberrations) produced a reconstruction at 2.80 Å resolution (**Extended Data Fig. 4**). The reconstruction had a good angular distribution that was improved by merging the 6 datasets (**Extended Data Fig. 4**). Density modification in Phenix (ref. 66) further improved the map resolution to 2.69 Å.

Focussed refinements. Although the core of the C-complex spliceosome is well ordered, many of the peripheral domains are flexible and had weak EM density in the overall refinement map, and correspondingly the local resolution quickly decayed towards the edge of the spliceosome (**Extended Data Fig. 4**). For each domain, strategies including signal subtraction, classification without alignment, and focussed refinement with and without the reconstruction algorithm SIDESPLITTER (ref. 67) were systematically investigated. All signal subtractions started from a 3 Å resolution map of C complex where the refinement mask encompassed the entire complex. All focussed refinements used solvent-flattened FSCs to calculate the resolution during refinement. The following strategies produced the best maps for each domain, and are summarised in **Extended Data Fig. 3** and **Table 1**.

Density for the U2 snRNP core domain was improved first by signal subtraction with a soft mask (8 pixel hard edge, 16 pixel soft edge) encompassing the U2 snRNP core, Prp8 RNaseH domain and Syf1 N-terminus. Signal-subtracted particles were shifted to the mask centre-of-mass and cropped to a 200 pixel box. 3D classification without alignment and T=4 into 4 classes required 60 iterations for convergence and showed that 26% of particles (108,540) did not have strong U2 snRNP density. These particles were removed, and the resultant STAR file was reverted to the original, un-subtracted particles. These were focus-refined using the same mask used for signal subtraction, using the overall map low-pass filtered to 7 Å as a reference and performing local angular searches starting at 0.9 degree sampling. This produced a reconstruction at 4.07 Å resolution. Refinement was then continued using a tighter mask that only encompassed the U2 snRNP core (3 pixel hard edge, 16 pixel soft edge), using a 5 Å resolution reference and performing local angular searches at 0.5 degree sampling. This produced a map for the U2 snRNP at 3.58 Å resolution. Density modification in Phenix (ref. 66) further improved the map resolution to 3.26 Å.

The NTC TPR domains (Syf1 and Clf1 and associated proteins) was improved first by signal subtraction with a mask (1 pixel hard edge, 8 pixel soft edge) encompassing the entirety of Clf1, Syf1 and the associated proteins. Signal-subtracted particles were shifted to the mask centre-of-mass and cropped to a 300 pixel box. 3D classification without alignment and T=4 into 4 classes required 25 iterations for convergence and showed that 43% of particles (172,895) did not have strong density for the peripheral helical arches. These particles were removed, and the remaining particles were focus-refined using a soft mask (8 pixel hard edge, 16 pixel soft edge) that only encompassed the flexible peripheral helical arches, e.g. excluding the well resolved N-terminus of Clf1. The best 3D class, low-pass filtered to 8 Å, was used as a reference, local angular searches started at 0.9 degree sampling, and the external reconstruction program SIDESPLITTER was used to reduce overfitting. This produced a reconstruction at 3.82 Å resolution. Refinement was then continued for 3 more iterations with the same mask, using a 3.8 Å resolution reference and performing local angular searches with 0.9 degree sampling. This produced an improved reconstruction with smoother densities, although the resolution stayed constant at 3.82 Å. Density modification in Phenix (ref. 66) then further improved the map resolution to 3.50 Å.

Despite numerous attempts, the helicase module of C complex, consisting of Brr2, Prp16, and the Prp8 Jab1/MPN domain, could not be improved to near-atomic resolution. This module is larger than the U2 snRNP core which was successfully refined to near-atomic resolution. Therefore, this module may have internal flexibility, e.g. between the RecA domains of Brr2 and Prp16, that makes particles difficult to align. Nevertheless, densities were made more interpretable as follows. First signal subtraction was performed using a mask (1 pixel hard edge, 8 pixel soft edge) loosely encompassing Brr2, Prp16, and Prp8 Jab1/MPN. Signal-subtracted particles were shifted to the mask centre-of-mass and cropped to a 300 pixel box. 3D classification without alignment and T=4 into 4 classes required 60 iterations for convergence and showed that only 26% of particles (103,461) had strong density for the helicase module. These were focus refined with a very soft mask (16 pixel hard edge, 12 pixel soft edge), using 0.9 degree local angular sampling, and the best class low-pass filtered to 20 Å as a reference. Postprocessing with a tighter mask (1 pixel hard edge, 16 pixel soft edge) gave a reconstruction of the helicase module at 8 Å resolution, which was used for interpreting Prp16. Classification without alignment suggested that Brr2 and Prp8 Jab1/MPN was flexible relative to Prp16. Therefore, a further focussed refinement was performed using a soft mask (8 pixel hard edge, 16 pixel soft edge) just around Brr2 and Prp8 Jab1/MPN, with a 20 Å resolution reference, 0.9 degree local angular sampling, and an initial offset search range of 3 pixels, producing a map at 7.15 Å resolution. Similar attempts to perform focussed refinement on just Prp16 were not successful.

In the overall C-complex map, the WD40 domain of Prp17 was weakly defined and did not allow unambiguous docking of the model from P complex, with all 14 possible orientations of the 7-bladed beta propeller being consistent with the density. However, this domain was too small for focussed refinement. Instead, to improve the local resolution we performed multiple rounds of classification without alignment and with high T values to select for high-resolution subsets of particles with Prp17 in similar positions. First, signal subtraction was performed using a loose spherical mask (12 pixel hard edge, 8 pixel soft edge) around the WD40 domain and some neighbouring proteins. Signal-subtracted particles were shifted to the mask centre-of-mass and cropped to a 200 pixel box. 3D classification without alignment and T=50 into 4 classes required 60 iterations for convergence. The single class with the most well-defined density (24%, 96,800 particles) was selected and 3D classification without alignment was repeated using a tighter mask (1 pixel hard edge, 8 pixel soft edge), T=50, and 4 classes. Again, the class with best defined density (58%, 59,152 particles) was selected. The selected particles were then reverted to the un-subtracted particles, and half maps were reconstructed from the original Euler angles using relion_reconstruct. Postprocessing with a mask encompassing the entire C complex produced a map at overall 3.67 Å resolution. The local resolution around the Prp17 WD40 domain was still limited, but now allowed unambiguous docking of the model from P complex, with one of the 14 possible orientations giving clearly higher correlations and atom inclusion scores when docking in UCSF Chimera.

Focussed classification of step II factors

To obtain robust estimates of the occupancy of the step II factors Prp18 and Slu7 on the C complex spliceosome, we used multiple parallel 3D classifications and merged selected classes (**Extended Data Fig. 3**). To identify particles containing the Slu7 zinc knuckle domain, signal subtraction was used with a small spherical mask (8 pixel hard edge, 8 pixel soft edge) centred on Slu7. Signal-subtracted particles were shifted to the mask centre-of-mass and cropped to a 200 pixel box. Two parallel 3D classifications without alignment were performed, both into 4 classes with 60 iterations and using a tighter spherical mask (1 pixel hard edge, 8 pixel soft edge). One classification used T=100, the other used T=1000. The T=100 classification showed 29% of particles contained the zinc knuckle domain, and the T=1000 classification showed that 18% of particles contained the same domain. After merging these particles and removing duplicates, 33% (130,588) of the original particles were captured, and further classification without alignment did not result in further segregation into different classes: using T values ranging from 100 to 2000, 85 – 95% of particles would converge on a single class with strong zinc knuckle density. Therefore, these 33% were taken as the final "Slu7 zinc knuckle" class.

To identify particles containing Prp18 and the associated EIE region of Slu7, 3D classification was performed both with varying T values and with and without signal subtraction. Signal-subtracted particles were generated using a soft mask centred on Prp18 (4 pixel hard edge, 12 pixel soft edge), and were shifted to the mask centre-of-mass and cropped to a 200 pixel box. Two parallel 3D classifications without alignment were performed on the signal-subtracted particles, both into 4 classes and using the same mask as for signal subtraction. One classification used T=4 and 60 iterations, the other used T=10 and 120 iterations. The T=4 classification showed 20% (82,492) of particles with strong density for Prp18, and the T=10 classification showed that 20% (81,431) of particles contained the same domain. These subsets were merged and duplicate particles were removed, yielding 23% (91,424) of the total particles.

Two parallel 3D classifications without alignment were also performed on the non-signal-subtracted particles, both into 4 classes with 60 iterations and using the same mask as for the signal-subtracted particles. One classification used T=4 and showed 20% (82,398) of particles with strong Prp18 density, the other classification used T=20 and showed 21% (82,872) of particles with strong Prp18 density and 21% (83,389) with some weak Prp18 density. These subsets were merged with the 91,424 un-subtracted particles from the signal-subtracted classifications, and duplicate particles were merged, yielding 186,323 particles. These were subjected to a final 3D classification without alignment into 4 classes, using a looser mask (8 pixel hard edge, 8 pixel soft edge), 60 iterations, and T=20. Two of the resulting classes had strong Prp18 density and were selected as the final "strong Prp18" class, containing 120,141 particles, or 30% of the original particles.

Contingency tables for each dataset (**Extended Data Fig. 8**) were determined from the above classifications, i.e. the datasets were not individually classified.

Merging C and P complex data to improve Prp19, Cwc22, and U5 Sm reconstructions

The Prp19 module, U5 Sm domain, and the N-terminal domain of Cwc22, are each very flexible but are in similar positions in C and P complexes as they are unaffected by the conformational change between step I and step II of splicing. We therefore reasoned that the best resolutions for these regions could be obtained by merging C and P complex data to give high initial particle numbers for focussed classification.

First, all P complex data were first merged using RELION 3.1, using separate optics groups for datasets 6, 7, 8, and 9 (**Extended Data Fig. 2**). The individual reconstructions were compared with UCSF Chimera to determine the relative pixel sizes of each dataset, and defocus values were scaled by the relative difference in pixel sizes[63]. Unscaled particles were then refined together, with RELION 3.1 internally scaling the particles to match a 1.12 Å/pix reference in a 400 pixel box. Two rounds of CTF refinement (refining per-particle defocus, per-micrograph astigmatism, and per-optics group anisotropic magnification, beam tilt, trefoil, and 4th order aberrations) followed by a final masked refinement produced a reconstruction at 2.95 Å resolution from 320,770 particles. This map had unclear density for the docked 3′ splice site, so we still consider our previous reconstruction at 3.30 Å with strong 3′ splice site density as the best "reference" P complex structure[63].

Next, we merged all our C complex particles with these P complex particles described above, giving a total of 724,244 particles that when refined produced a map that largely resembled C complex (presumably due to the majority of the particles – 56% – being in the step I conformation) although with weaker density for the mobile U2 snRNP, the branch helix, and the step I factors (**Extended Data Fig. 3**). For the extremely mobile Prp19 domain of the NTC, no focussed refinements were successful. We therefore performed signal subtraction with a loose mask that would encompass most possible orientations of this domain, constructed by summing all classes from a preliminary classification without alignment, then adding a 2 pixel hard edge and 8 pixel soft edge. Signal-subtracted particles were shifted to the mask centre-of-mass and cropped to a 300 pixel box. 3D classification without alignment and T=70 into 12 classes showed there were indeed many possible orientations of this domain (**Extended Data Fig. 3**). The most stable class, containing 7% of the original particles (49,514) was selected. Half maps were reconstructed for the signal-subtracted particles using relion_reconstruct and the original Euler angles, allowing postprocessing with a soft mask (hard edge 8 pixels, soft edge 16 pixels) encompassing just the stable orientation, which produced a map at nominally 7.30 Å resolution. This resolution is probably overestimated but the map nonetheless shows some secondary structure features consistent with the expected helical bundle (**Extended Data Fig. 5**). We also reverted to the original C-complex particles (35,716) and calculated half-maps with relion_reconstruct to show how this stable orientation relates to the body of the spliceosome. After postprocessing with a mask around the entire C-complex, this produced a reconstruction at 3.97 Å resolution (local resolution around Prp19 ~ 8 Å) showing how the N-terminus of Prp46 probably projects into the Prp19 module.

A similar approach was applied for the U5 Sm domain (**Extended Data Fig. 3**), as this domain was too small for focussed refinements. Signal subtraction was performed on the combined C/C*/P particles using a soft mask encompassing the U5 Sm ring and the base of U5 snRNA (hard edge 4 pixels, soft edge 8 pixels), and re-centering particles on the mask centre-of-mass and cropping to a 200 pixel box. Two rounds of 3D classification without alignment were used to select the particles with the most stably-bound, highest resolution U5 Sm domain. The first round with T=20 eliminated 54% of particles which showed no density for the U5 Sm domain, and the second round with T=100 selected 23% of particles (65,561 particles, 9% of original particles) which were used to calculate half-maps with relion_reconstruct using the original Euler angles. After postprocessing with a mask around the entire spliceosome, this produced a reconstruction at 3.13 Å resolution with much stronger density for the U5 Sm ring than the original map. The map was further improved by density modification in Phenix (ref. 66) to 2.96 Å resolution, which smoothened some of the discontinuous densities in the peripheral Sm chains.

Finally, a similar approach was utilised for the N-terminal domain of Cwc22 (**Extended Data Fig. 3**). Signal subtraction was performed on the combined C/C*/P particles using a soft mask encompassing the Cwc22 NTD (hard edge 4 pixels, soft edge 8 pixels), and re-centering particles on the mask centre-of-mass and cropping to a 200 pixel box. Two rounds of 3D classification without alignment were used to select the most stable, high-resolution position of Cwc22 NTD. The first round with T=100 eliminated 72% of particles which showed no density for Cwc22 NTD, and the second round with T=1000 selected 46% of particles (73,802 particles, 10% of original particles) which were used to calculate half-maps with relion_reconstruct using the original Euler angles. After postprocessing with a mask around the entire spliceosome, this produced a reconstruction at 3.16 Å resolution. The map was further improved by density modification in Phenix (ref. 66) to 2.97 Å resolution.

C complex model building

To avoid being biased by our previous model of C complex[7] which was built into lower resolution density (3.8 Å average resolution), the entire ordered core of C complex was rebuilt entirely de novo in Coot (ref.68), based only on the 2.8 Å cryoEM density. This included the proteins Prp8 (except the RNaseH and Jab1/MPN domains), Snu114, Prp45, Prp46, Ecm2, Cwc2, Cwc15, Bud31, Cef1, Clf1, Syf1, Syf2, Cwc21, Cwc22, Yju2, Cwc25, Isy1, and the U2, U5, and U6 snRNAs, the intron, and the 5′ exon. To improve model geometry, we compared the scale of the map to crystal structures of Prp8 (PDB 4I43, ref.69) and the NMR structure of Bud31 (PDB 2MY1, ref.70), which showed that the true pixel size of the map should be 1.145 Å/pix. The model coordinates were all scaled by 1.145/1.12 = 1.02232 before further model building.

ISOLDE (ref.71) was then used to diagnose and fix errors and improve Ramachandran, CaBLAM, and rotamer outliers. The resultant core model largely resembled our earlier model, but with more accurate backbone geometry and rotamer assignment (**Table 1**), and some fixing of register errors including in Cef1 helix residues 230-249 and the Cwc15 N-terminus. Notable areas of improvement include the C-terminal domain of Ecm2, the interface of U2 snRNA stem IIb with Cwc2, domain IV of Snu114, assignment of the C-terminal helices of Yju2, and assignment of the very N-terminus of Prp45 and an N-terminal helix of Prp46 projecting towards Prp19.

For building into the focussed-refined maps, these maps were first all aligned and resampled in Chimera to their average positions, as determined from their most populated 3D classes. phenix.combine_focused_maps was then used to make a composite map from the resampled, density-modified maps for the core (2.69 Å), U2 snRNP (3.26 Å), NTC (3.50 Å), U5 Sm (2.96 Å) and Cwc22 NTD (2.97 Å), and models were built into this composite map.

The helical arches of Clf1 and Syf1 were built de novo into the NTC part of the composite map using Coot. The map was of sufficient quality to allow unambiguous building of the N-terminal half of Clf1 and the middle region of Syf1 (**Extended Data Fig. 5**). However, the local resolution still deteriorated towards the periphery of the focus-refined map (**Extended Data Fig. 4**), so we used secondary structure predictions generated by the GeneSilico MetaServer (ref.72) and cross-linking data from yeast B^{act} complex (ref. 73) to support our building. Multiple sequence alignment was also useful, revealing a yeast-specific insertion in Syf1 that we predicted formed a disordered loop, since the cryoEM density of human Syf1 in human P complex (ref. 74) largely resembled the cryoEM density for yeast Syf1. The completed Syf1 model left several helical densities unassigned, with no topological means of filling them with Syf1 sequence. Based on further secondary structure predictions, crosslinking in yeast B^{act} complex, and in some cases clear side-chain density, we assigned these densities to Ntc20, Isy1, and Syf2. This left two remaining helices. One was tentatively assigned to Cef1 but was built with UNK residues, and the other is left unassigned (chain X, UNK residues). The model was then improved in (ref. 71).

The Sm domain of the U2 snRNP was built by docking in the structure of the yeast U1 snRNP Sm domain from pre-B complex (ref. 75) and manually adjusting all side chains and loops to fit the density using Coot. Homology models for Lea1 and Msl1 were generated with I-TASSER (ref.76) based on the structure of human U2A′/U2B″ in complex with U2 snRNA (PDB 1A9N, ref.77) and were docked into the density and manually fixed. Density connecting the U2 snRNP to Syf1 could not be accounted for by Syf1 helices, and was assigned to Isy1 and the C-terminus of Lea1 based on side-chain density and crosslinking data from yeast B^{act} complex. Double-stranded RNA density corresponding to stems IV, V and the 950 nt yeast-specific insertion in U2 snRNA, was modelled by generating idealised structures in RNAComposer (ref. 78) and adjusting to the density in Coot. The model was then improved using ISOLDE, with adaptive distance restraints used to maintain pairing in U2 stems IV and V.

The Sm domain of U5 snRNP was built by copying our new model for the U2 snRNP Sm domain and docking into the composite map. The associated region of U5 snRNA was docked from yeast B complex (ref.79). The fit for both was improved used using ISOLDE, with adaptive distance restraints on the Sm ring using U2 snRNP as a template, and strong adaptive distance restraints (kappa value of 50) on U5 snRNA.

The Cwc22 N-terminal domain was built by docking an I-TASSER model based on the crystal structure of human Cwc22 complexed with eIF4AIII (PDB 4C9B) into the composite map. Extraneous loops were truncated and the fit was improved with ISOLDE.

Prp17, the Prp8 RNaseH domain, Prp18, and Slu7 all have stronger densities in P complex than in C complex. These were therefore rebuilt using ISOLDE into our newer 3.3 Å map (EMD-10140, ref.63), scaled up to 1.145 Å/pix from 1.12 Å/pix, starting from our original deposited 3.7 Å P-complex model (PDB 6EXN, ref. 9). Prp17, Slu7 zinc-knuckle, and Prp8 RNaseH + Prp18 + Slu7 EIE region were each docked into their respective focus-classified C-complex maps, and ISOLDE was used to fix any clashes and improve the fit to density.

For the helicase domain, a composite map was prepared with phenix.combine_focused_maps from the overall helicase focussed refinement, which had best density for Prp16, and the Brr2 focussed refinement, which had better density for Brr2 and Prp8 Jab1. The crystal structure of Brr2 associated with the Jab1/MPN domain of Prp8 (PDB 4BGD, ref.80) was docked into this composite map after minor rebuilding into the original structure factors with ISOLDE and refinement with phenix.refine to improve starting geometry. A homology model of Prp16 was generated with I-TASSER, with the top two templates being Prp43 (PDB 3KX2) (ref. 81) and MLE helicase (PDB 5AOR), and was docked into the composite map. ISOLDE with adaptive distance restraints was used to morph these starting models into the density. Very strong restraints (kappa of 20) were used for Brr2/Jab1 since the starting model was known to be of good quality. For Prp16, the individual domains (RecA1, RecA2, and CTD) were individually restrained with kappa of 10 to allow movement between them. The resulting Brr2 model was largely unchanged, but Prp16 adopted an open conformation that strongly resembled the crystal structure of *Chaetomium thermophilum* Prp22 in complex with RNA (PDB 6I3P, ref. 82). Indeed, equivalent density to the Prp22-bound RNA was observed next to the RecA2 domain, which based on distance to the modelled intron in the spliceosome core was assigned as residues 87 – 92 of the UBC4 intron (+17 to +22 relative to the branch point, -9 to -4 relative to the 3′ splice site). Prp16 and this region of the intron were then further refined in ISOLDE using strong adaptive distance restraints to the *Ct*Prp22-RNA crystal structure. Finally, the density did not allow correction of Ramachandran and CaBLAM outliers which were inherited from the source structures, but all rotamer outliers were manually corrected when possible.

The Prp19 module, consisting of the Prp19 tetramer, Snt309, and Cef1 C-terminus, was modelled by first rebuilding our earlier model of the human orthologue in human P complex (ref.74) using ISOLDE. An extra helix was built as chain X, residues UNK, but is likely the N-terminus of Prp46. Multiple sequence alignments were then used to mutate and truncate this model to fit the yeast sequences using phenix.sculptor. The U-box domains of Prp19 were modelled using crystal structures of the yeast protein (PDB 2BAY, ref. 83). This model was then docked into the 7.3 Å focussed classification maps of the Prp19 module. This map was good enough to allow improvement of the fit using ISOLDE, using adaptive distance restraints to the starting positions, except for the U-box domains which were strongly restrained (kappa=50) to the crystal structure positions. This fitting showed a pronounced bend in one of the Prp19 dimeric coiled-coils compared to the human structure. After fitting, a large portion of unassigned density remained between the Prp19 module and Syf1. Although low resolution, this density neatly accommodates the crystal structure of the Prp19 WD40 domain (PDB 3LRV, ref.83), which was docked in after rebuilding in ISOLDE and refinement into the original structure factors using phenix.refine.. The Prp19 tetramer should have four such domains, and previous cryoEM structures have shown these in a wide variety of positions, generally without strong support from density. The current assignment of the WD40 domain should also be viewed as tentative, although more well-supported than previous assignments.

C complex model refinement

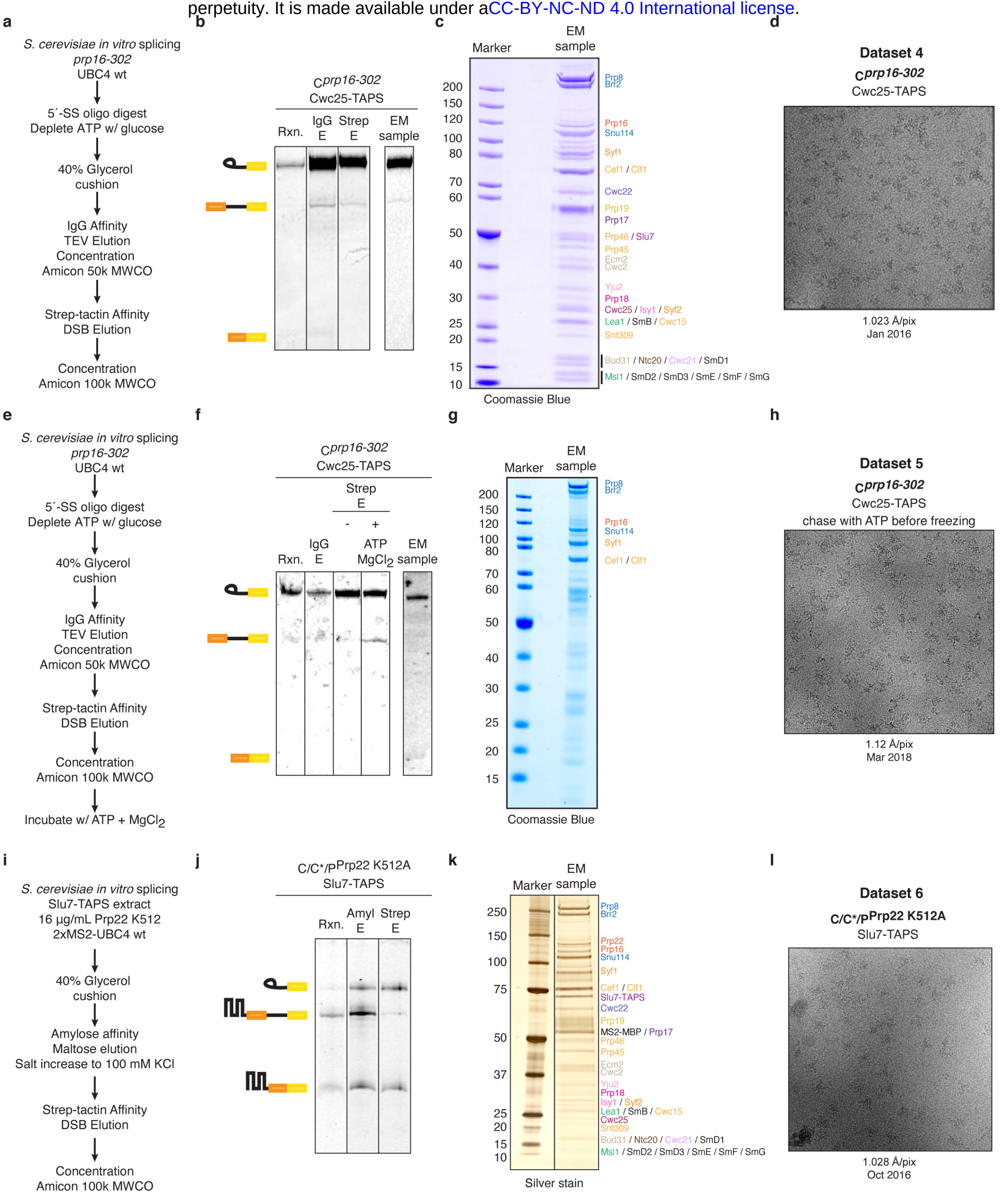
Coordinates for the core of C complex, U2 snRNP, NTC, U5 Sm, and Cwc22N were refined simultaneously in real space using phenix.real_space_refine in PHENIX (ref. 84) into a composite map of the 5 respective density-modified focussed maps. Base-pairing, base-stacking, and metal-coordination restraints were not imposed for the core, where the density was good enough to obtain better geometry and fit to density in the absence of external restraints. For refinement of the lariat-intron 2'-5' linkage a custom set of restraints was adapted from the geometry of the 3'-5' phosphodiester RNA backbone. Base-pairing and base-stacking restraints were however imposed for U5 snRNA near the U5 Sm site and U2 snRNA stems IIb/c, IV and V. We found that the default real_space_refine settings consistently worsened the excellent starting geometry of our models from ISOLDE. Systematic investigation found that performing one macro-cycle of global minimisation and ADP refinement, skipping local grid searches, with a nonbonded weight of 2000 and overall weight of 0.5, gave the best model quality statistics (**Table 1**). The Prp19 module, helicase module, and Prp17 WD40 domain were not refined in PHENIX, but ISOLDE was used to resolve any clashes between these modules and the refined structure.. The final C complex model comprises 45 protein chains, 3 snRNAs, and the intron and 5′ exon.

Model visualisation. All structural figures were generated with UCSF ChimeraX (ref. 85).

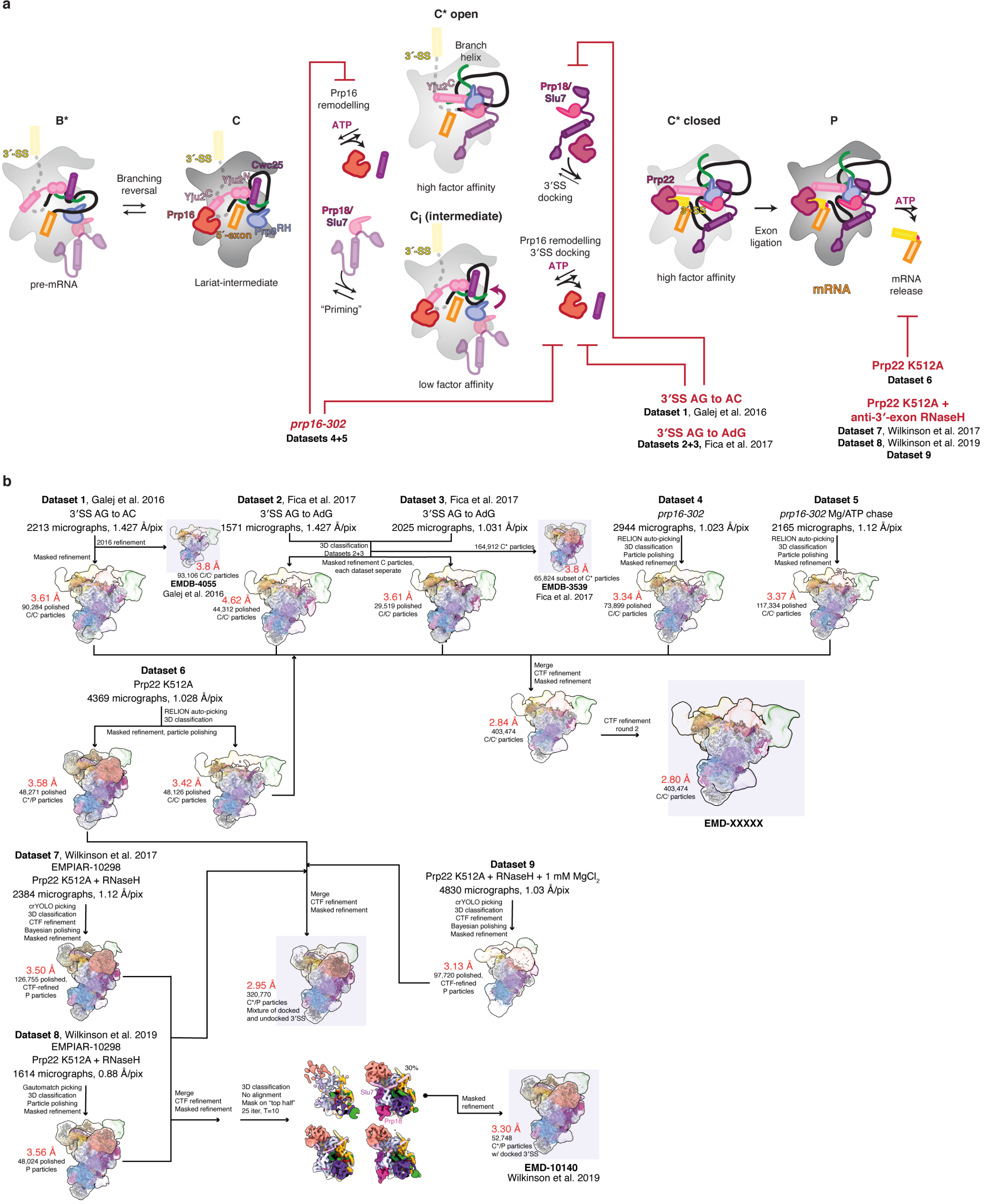
Additional References

60. Nguyen, T. H. D. *et al*. The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* **523,** 47–52 (2015).
61. Lardelli, R. M., Thompson, J. X., Yates, J. R. & Stevens, S. W. Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA* **16,** 516–528 (2010).
62. 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).
63. Wilkinson, M. E., Kumar, A., Casañal, A.IUCr. Methods for merging data sets in electron cryo-microscopy. *Acta Crystallogr D Struct Biol* **75,** 782–791 (2019).
64. Scheres, S. H. & Kühlbrandt, W. Beam-induced motion correction for sub-megadalton cryo-EM particles. *Elife* **3,** e03665 (2014).
65. Wagner, T. *et al*. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun Biol* **2,** 1–13 (2019).
66. Terwilliger, T. C., Ludtke, S. J., Read, R. J., Adams, P. D. & Afonine, P. V. Improvement of cryo-EM maps by density modification. *Nat. Methods* **17,** 923–927 (2020).
67. Ramlaul, K., Palmer, C. M., Nakane, T. & Aylett, C. H. S. Mitigating local over-fitting during single particle reconstruction with SIDESPLITTER. *J. Struct. Biol.* **211,** 107545 (2020).
68. Casañal, A., Lohkamp, B. & Emsley, P. Current developments in Coot for macromolecular model building of Electron Cryo−microscopy and Crystallographic Data. *Protein Science* **29,** 1055–1064 (2020).
69. 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).
70. van Roon, A.-M. M. *et al*. ^{113}Cd NMR experiments reveal an unusual metal cluster in the solution structure of the yeast splicing protein Bud31p. *Angew. Chem. Int. Ed. Engl.* **54,** 4861–4864 (2015).
71. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74,** 519–530 (2018).
72. Kurowski, M. A. GeneSilico protein structure prediction meta-server. *Nucleic Acids Res*. **31,** 3305–3307 (2003).

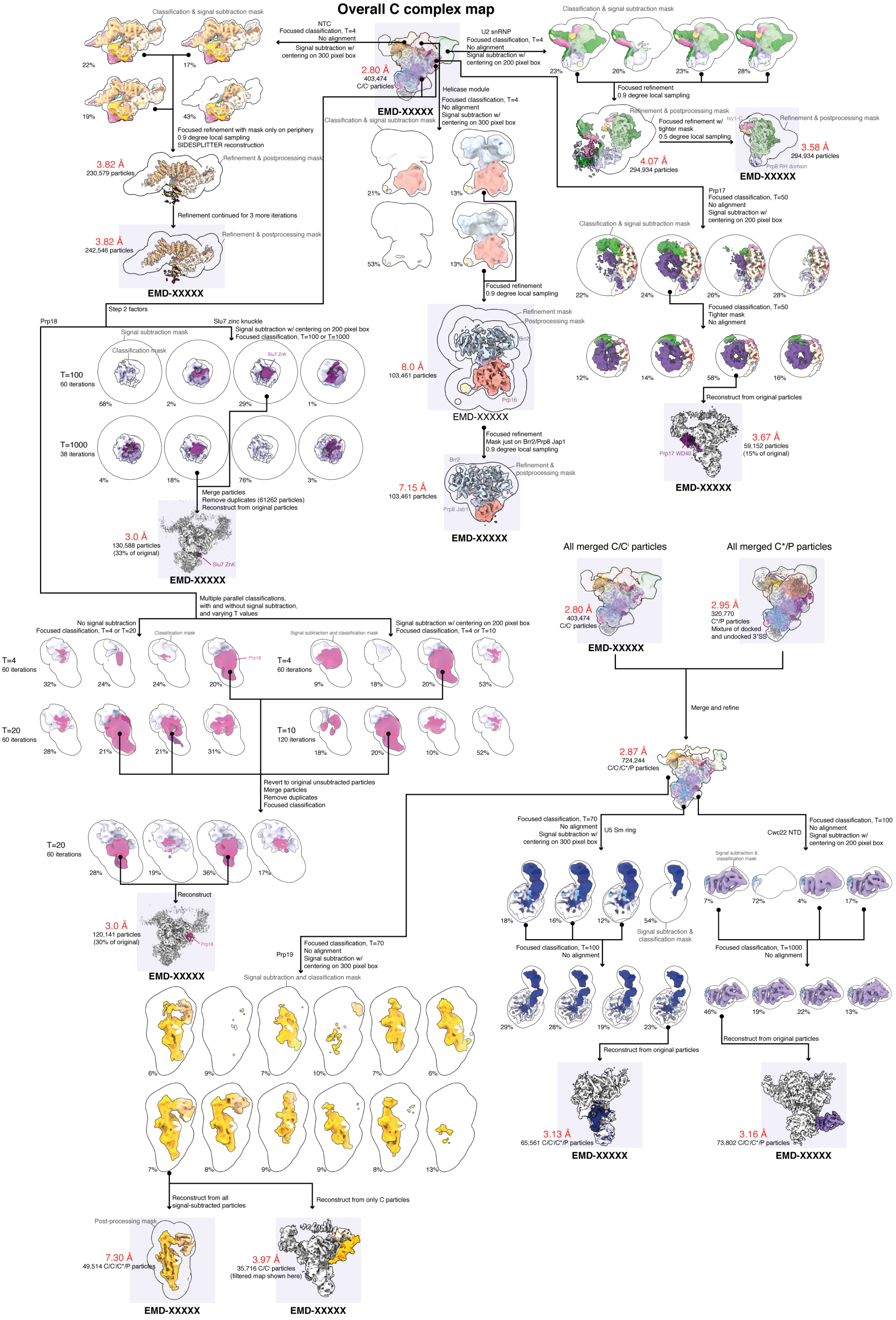
73. Rauhut, R. *et al*. Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. *Science* **353,** 1399–1405 (2016).
74. Fica, S. M., Oubridge, C., Wilkinson, M. E., Newman, A. J. & Nagai, K. A human postcatalytic spliceosome structure reveals essential roles of metazoan factors for exon ligation. *Science* **363,** 710–714 (2019).
75. Bai, R., Wan, R., Yan, C., Lei, J. & Shi, Y. Structures of the fully assembled Saccharomyces cerevisiae spliceosome before activation. *Science* **360,** 1423–1429 (2018).
76. Yang, J. *et al*. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12,** 7–8 (2015).
77. 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).
78. Popenda, M. *et al*. Automated 3D structure composition for large RNAs. *Nucleic Acids Res*. **40,** e112 (2012).
79. Plaschka, C., Lin, P.-C. & Nagai, K. Structure of a pre-catalytic spliceosome. *Nature* **546,** 617–621 (2017).
80. Nguyen, T. H. D. *et al*. Structural Basis of Brr2-Prp8 Interactions and Implications for U5 snRNP Biogenesis and the Spliceosome Active Site. *Structure* **21,** 910–919 (2013).
81. He, Y., Andersen, G. R. & Nielsen, K. H. Structural basis for the function of DEAH helicases. *EMBO Rep* **11,** 180–186 (2010).
82. Hamann, F., Enders, M. & Ficner, R. Structural basis for RNA translocation by DEAH-box ATPases. *Nucleic Acids Res*. **47,** 4349–4362 (2019).
83. Vander Kooi, C. W. *et al*. The Prp19 U-box crystal structure suggests a common dimeric architecture for a class of oligomeric E3 ubiquitin ligases. *Biochemistry* **45,** 121–130 (2006).
84. Afonine, P. V. *et al*. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74,** 531–544 (2018).
85. Goddard, T. D. *et al*. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science* **27,** 14–25 (2018).


a
S. cerevisiae in vitro splicing
prp16-302
UBC4 wt
5´-SS oligo digest
Deplete ATP w/ glucose
40% Glycerol cushion
IgG Affinity
TEV Elution
Concentration
Amicon 50k MWCO
Strep-tactin Affinity
DSB Elution
Concentration
Amicon 100k MWCO
b
Cprp16-302
Cwc25-TAPS
Rxn.
IgG E
Strep E
EM sample
c
Marker
EM sample
200
150
120
100
80
70
60
50
40
30
25
20
15
10
Prp8
Brr2
Prp16
Snu114
Syf1
Cef1 / Clf1
Cwc22
Prp19
Prp17
Prp46 / Slu7
Prp45
Ecm2
Cwc2
Yju2
Prp18
Cwc25 / Isy1 / Syf2
Lea1 / SmB / Cwc15
Snt309
Bud31 / Ntc20 / Cwc21 / SmD1
Msl1 / SmD2 / SmD3 / SmE / SmF / SmG
Coomassie Blue
d
Dataset 4
Cprp16-302
Cwc25-TAPS
1.023 Å/pix
Jan 2016
e
S. cerevisiae in vitro splicing
prp16-302
UBC4 wt
5´-SS oligo digest
Deplete ATP w/ glucose
40% Glycerol cushion
IgG Affinity
TEV Elution
Concentration
Amicon 50k MWCO
Strep-tactin Affinity
DSB Elution
Concentration
Amicon 100k MWCO
Incubate w/ ATP + MgCl2
f
Cprp16-302
Cwc25-TAPS
Strep E
-
+
Rxn.
IgG E
ATP MgCl2
EM sample
g
Marker
EM sample
200
150
120
100
80
70
60
50
40
30
25
20
15
Prp8
Brr2
Prp16
Snu114
Syf1
Cef1 / Clf1
Coomassie Blue
h
Dataset 5
Cprp16-302
Cwc25-TAPS
chase with ATP before freezing
1.12 Å/pix
Mar 2018
i
S. cerevisiae in vitro splicing
Slu7-TAPS extract
16 µg/mL Prp22 K512
2xMS2-UBC4 wt
40% Glycerol cushion
Amylose affinity
Maltose elution
Salt increase to 100 mM KCl
Strep-tactin Affinity
DSB Elution
Concentration
Amicon 100k MWCO
j
C/C*/PPrp22 K512A
Slu7-TAPS
Rxn.
Amyl E
Strep E
k
Marker
EM sample
250
150
100
75
50
37
25
20
15
10
Prp8
Brr2
Prp22
Prp16
Snu114
Syf1
Cef1 / Clf1
Slu7-TAPS
Cwc22
Prp19
MS2-MBP / Prp17
Prp46
Prp45
Ecm2
Cwc2
Yju2
Prp18
Isy1 / Syf2
Lea1 / SmB / Cwc15
Cwc25
Snt309
Bud31 / Ntc20 / Cwc21 / SmD1
Msl1 / SmD2 / SmD3 / SmE / SmF / SmG
Silver stain
l
Dataset 6
C/C*/PPrp22 K512A
Slu7-TAPS
1.028 Å/pix
Oct 2016

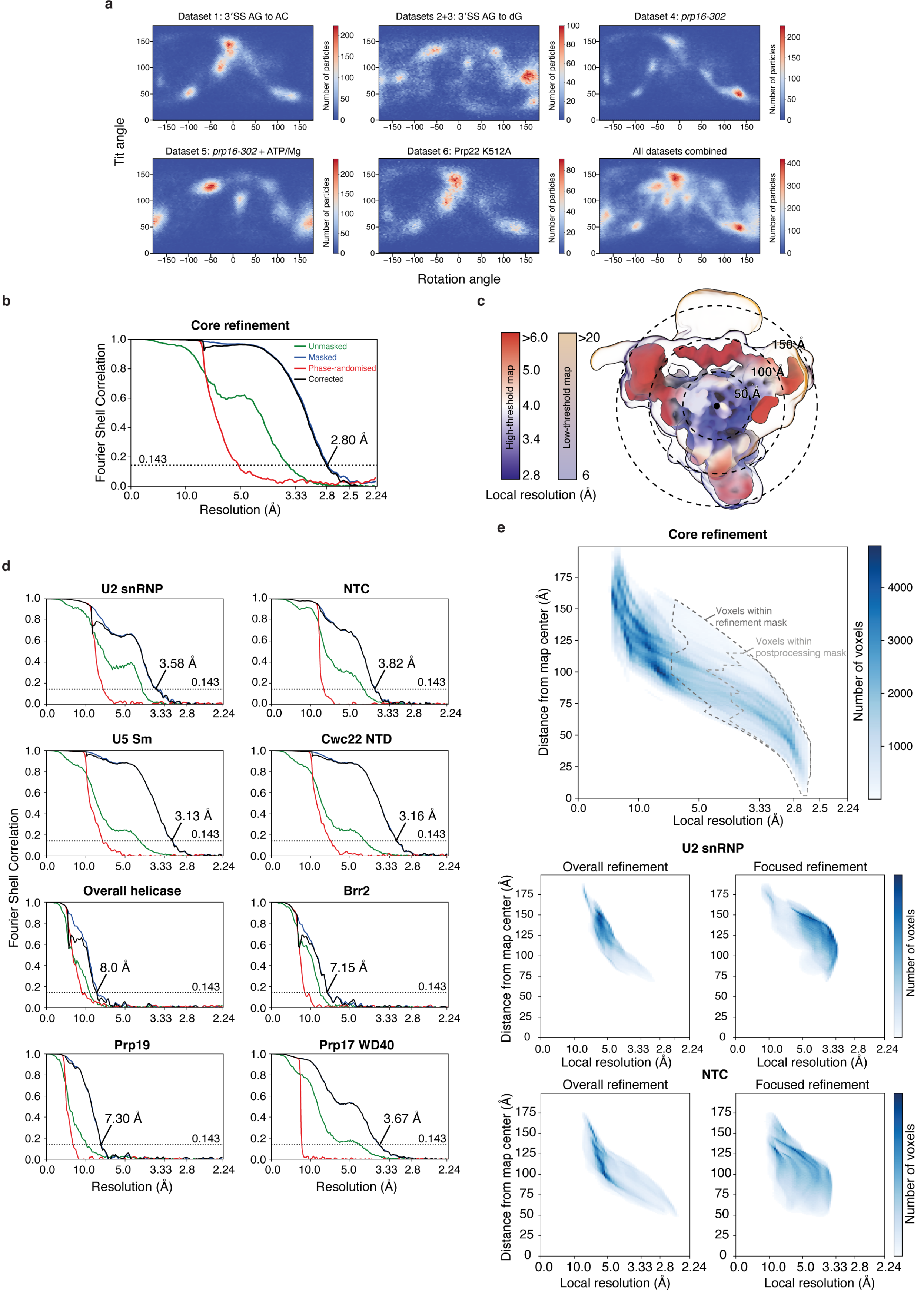
Extended Data Fig. 1 | Purification and cryo-EM of new spliceosome samples used for this study. a-d, Purification scheme, RNA and protein composition, and representative micrograph for C-complex Dataset 4. **e-h,** Purification scheme, RNA and protein composition, and representative micrograph for C-complex Dataset 5. **i-l,** Purification scheme, RNA and protein composition, and representative micrograph for the P-complex Dataset 6.


a
B*
3´-SS
pre-mRNA
Branching reversal
C
3´-SS
Cwc25
Yju2N
Yju2C
Prp16
5´-exon
Prp2
RH
Lariat-intermediate
Prp16 remodelling
ATP
Prp18/ Slu7
"Priming"
C* open
3´-SS
Branch helix
Yju2C
high factor affinity
Ci (intermediate)
3´-SS
low factor affinity
Prp18/ Slu7
3′SS docking
Prp16 remodelling 3′SS docking
ATP
C* closed
Prp22
3´-SS
high factor affinity
Exon ligation
P
ATP
mRNA
mRNA release
prp16-302
Datasets 4+5
3′SS AG to AC
Dataset 1, Galej et al. 2016
3′SS AG to AdG
Datasets 2+3, Fica et al. 2017
Prp22 K512A
Dataset 6
Prp22 K512A + anti-3′-exon RNaseH
Dataset 7, Wilkinson et al. 2017
Dataset 8, Wilkinson et al. 2019
Dataset 9
b
Dataset 1, Galej et al. 2016
3′SS AG to AC
2213 micrographs, 1.427 Å/pix
2016 refinement
Masked refinement
3.8 Å
93,106 C/C′ particles
EMDB-4055
Galej et al. 2016
3.61 Å
90,284 polished C/C′ particles
Dataset 2, Fica et al. 2017
3′SS AG to AdG
1571 micrographs, 1.427 Å/pix
Dataset 3, Fica et al. 2017
3′SS AG to AdG
2025 micrographs, 1.031 Å/pix
3D classification Datasets 2+3
Masked refinement C particles, each dataset seperate
164,912 C* particles
4.62 Å
44,312 polished C/C′ particles
3.61 Å
29,519 polished C/C′ particles
3.8 Å
65,824 subset of C* particles
EMDB-3539
Fica et al. 2017
Dataset 4
prp16-302
2944 micrographs, 1.023 Å/pix
RELION auto-picking
3D classification
Particle polishing
Masked refinement
3.34 Å
73,899 polished C/C′ particles
Dataset 5
prp16-302 Mg/ATP chase
2165 micrographs, 1.12 Å/pix
RELION auto-picking
3D classification
Particle polishing
Masked refinement
3.37 Å
117,334 polished C/C′ particles
Dataset 6
Prp22 K512A
4369 micrographs, 1.028 Å/pix
RELION auto-picking
3D classification
Masked refinement, particle polishing
3.58 Å
48,271 polished C*/P particles
3.42 Å
48,126 polished C/C′ particles
Merge
CTF refinement
Masked refinement
2.84 Å
403,474 C/C′ particles
CTF refinement round 2
2.80 Å
403,474 C/C′ particles
EMD-XXXXX
Dataset 7, Wilkinson et al. 2017
EMPIAR-10298
Prp22 K512A + RNaseH
2384 micrographs, 1.12 Å/pix
crYOLO picking
3D classification
CTF refinement
Bayesian polishing
Masked refinement
3.50 Å
126,755 polished, CTF-refined P particles
Merge
CTF refinement
Masked refinement
2.95 Å
320,770 C*/P particles
Mixture of docked and undocked 3′SS
Dataset 9
Prp22 K512A + RNaseH + 1 mM MgCl2
4830 micrographs, 1.03 Å/pix
crYOLO picking
3D classification
CTF refinement
Bayesian polishing
Masked refinement
3.13 Å
97,720 polished, CTF-refined P particles
Dataset 8, Wilkinson et al. 2019
EMPIAR-10298
Prp22 K512A + RNaseH
1614 micrographs, 0.88 Å/pix
Gautomatch picking
3D classification
Particle polishing
Masked refinement
3.56 Å
48,024 polished P particles
Merge
CTF refinement
Masked refinement
3D classification
No alignment
Mask on "top half"
25 iter, T=10
30%
Slu7
Prp18
Masked refinement
3.30 Å
52,748 C*/P particles w/ docked 3′SS
EMD-10140
Wilkinson et al. 2019

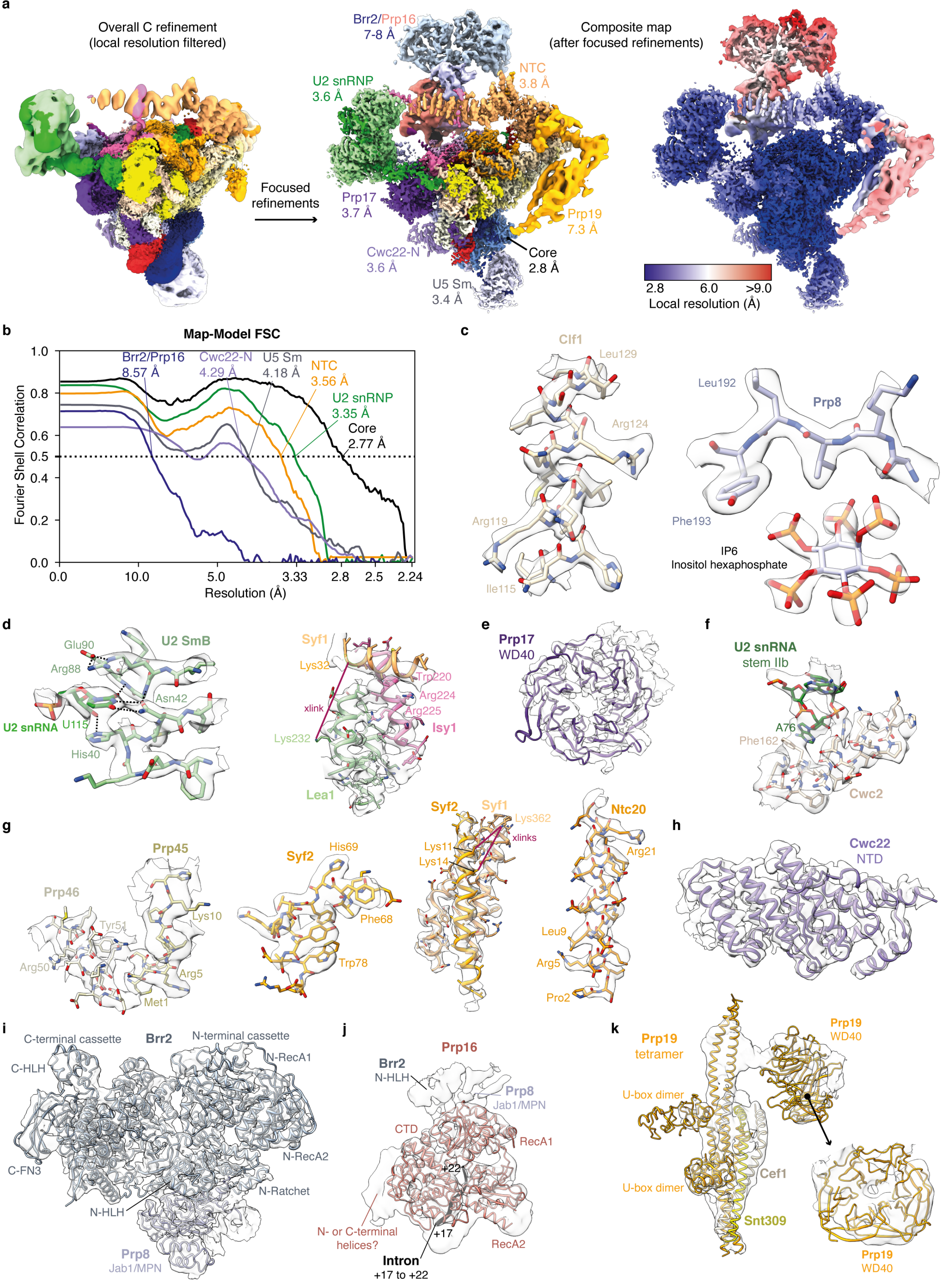
Extended Data Fig. 2 | Cryo-EM data processing scheme for dataset merging. a, Biochemical scheme for stalling of spliceosomes used for specific datasets. **b,** Dataset merging scheme for producing overall C- and P-complex structures. Final maps deposited to the EMDB are highlighted.


Overall C complex map
2.80 Å
403,474
C/C' particles
EMD-XXXXX
NTC
Focused classification, T=4
No alignment
Signal subtraction w/
centering on 300 pixel box
Classification & signal subtraction mask
22%
17%
19%
43%
Focused refinement with mask only on periphery
0.9 degree local sampling
SIDESPLITTER reconstruction
3.82 Å
230,579 particles
Refinement & postprocessing mask
Refinement continued for 3 more iterations
3.82 Å
242,546 particles
Refinement & postprocessing mask
EMD-XXXXX
U2 snRNP
Focused classification, T=4
No alignment
Signal subtraction w/
centering on 200 pixel box
Classification & signal subtraction mask
23%
26%
23%
28%
Focused refinement
0.9 degree local sampling
Refinement & postprocessing mask
Focused refinement w/
tighter mask
0.5 degree local sampling
4.07 Å
294,934 particles
Isy1-C
Refinement & postprocessing mask
3.58 Å
294,934 particles
Prp8 RH domain
EMD-XXXXX
Helicase module
Focused classification, T=4
No alignment
Signal subtraction w/
centering on 300 pixel box
Classification & signal subtraction mask
21%
13%
53%
13%
Focused refinement
0.9 degree local sampling
Refinement mask
Postprocessing mask
Brr2
8.0 Å
103,461 particles
Prp16
EMD-XXXXX
Focused refinement
Mask just on Brr2/Prp8 Jap1
0.9 degree local sampling
Brr2
Refinement &
postprocessing mask
7.15 Å
103,461 particles
Prp8 Jab1
EMD-XXXXX
Prp17
Focused classification, T=50
No alignment
Signal subtraction w/
centering on 200 pixel box
Classification & signal subtraction mask
22%
24%
26%
28%
Focused classification, T=50
Tighter mask
No alignment
12%
14%
58%
16%
Reconstruct from original particles
Prp17 WD40
3.67 Å
59,152 particles
(15% of original)
EMD-XXXXX
Step 2 factors
Prp18
Slu7 zinc knuckle
Signal subtraction w/ centering on 200 pixel box
Focused classification, T=100 or T=1000
Signal subtraction mask
Classification mask
Slu7 ZnK
T=100
60 iterations
68%
2%
29%
1%
T=1000
38 iterations
4%
18%
76%
3%
Merge particles
Remove duplicates (61262 particles)
Reconstruct from original particles
3.0 Å
130,588 particles
(33% of original)
Slu7 ZnK
EMD-XXXXX
All merged C/C' particles
2.80 Å
403,474
C/C' particles
EMD-XXXXX
All merged C*/P particles
2.95 Å
320,770
C*/P particles
Mixture of docked
and undocked 3'SS
Merge and refine
2.87 Å
724,244
C/C/C*/P particles
Multiple parallel classifications,
with and without signal subtraction,
and varying T values
No signal subtraction
Focused classification, T=4 or T=20
Classification mask
Signal subtraction and classification mask
Signal subtraction w/ centering on 200 pixel box
Focused classification, T=4 or T=10
T=4
60 iterations
32%
24%
24%
20%
Prp18
T=4
60 iterations
9%
18%
20%
53%
T=20
60 iterations
28%
21%
21%
31%
T=10
120 iterations
18%
20%
10%
52%
Revert to original unsubtracted particles
Merge particles
Remove duplicates
Focused classification
T=20
60 iterations
28%
19%
36%
17%
Reconstruct
3.0 Å
120,141 particles
(30% of original)
Prp18
EMD-XXXXX
Prp19
Focused classification, T=70
No alignment
Signal subtraction w/
centering on 300 pixel box
Signal subtraction and classification mask
6%
9%
7%
10%
7%
6%
7%
8%
9%
9%
8%
13%
Reconstruct from all
signal-subtracted particles
Reconstruct from only C particles
Post-processing mask
7.30 Å
49,514 C/C/C*/P particles
EMD-XXXXX
3.97 Å
35,716 C/C' particles
(filtered map shown here)
EMD-XXXXX
Focused classification, T=70
No alignment
Signal subtraction w/
centering on 300 pixel box
U5 Sm ring
18%
16%
12%
54%
Signal subtraction &
classification mask
Focused classification, T=100
No alignment
29%
28%
19%
23%
Reconstruct from original particles
3.13 Å
65,561 C/C/C*/P particles
EMD-XXXXX
Cwc22 NTD
Focused classification, T=100
No alignment
Signal subtraction w/
centering on 200 pixel box
Signal subtraction &
classification mask
7%
72%
4%
17%
Focused classification, T=1000
No alignment
46%
19%
22%
13%
Reconstruct from original particles
3.16 Å
73,802 C/C/C*/P particles
EMD-XXXXX

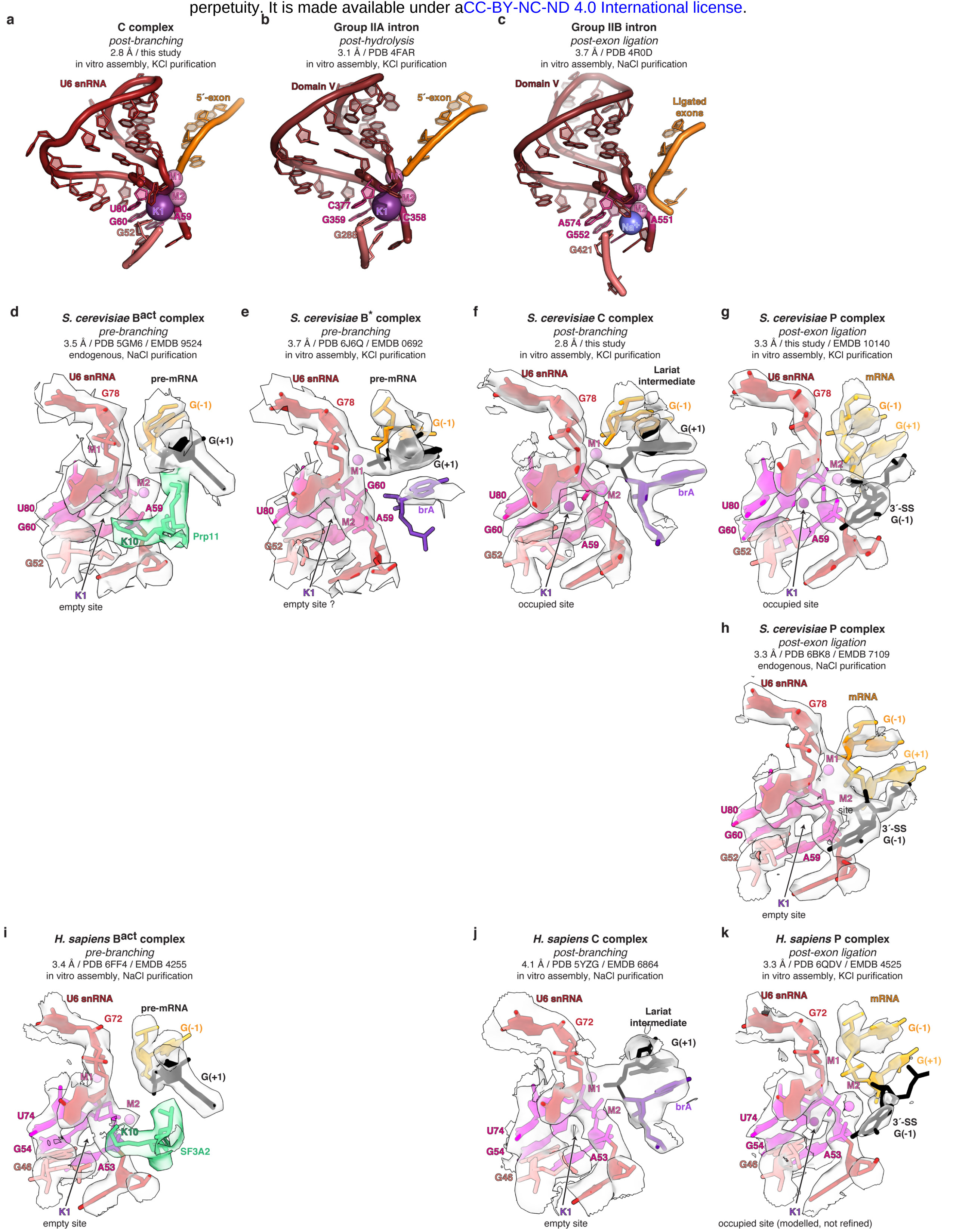
Extended Data Fig. 3 | Focussed classification scheme for peripheral regions of the C complex. Final maps deposited to the EMDB are highlighted.



Extended Data Fig. 4 | Statistics for the cryo-EM datasets and the resulting maps. a, Orientational distribution for the C-complex cryo-EM datasets. Merging datasets improves the overall angular distribution. **b,** Gold-standard Fourier-Shell Correlation curves for the overall C-complex reconstruction and the focus-refined or classified maps. **c,** Overall C-complex reconstruction coloured by local resolution. The map is shown at two thresholds coloured with different palettes to show the high core resolution and low peripheral resolution. **d,** Heatmap of local resolution against distance from map center, showing how in the overall C-complex reconstruction the local resolution decreases towards the map periphery. **e,** Local resolution heatmaps for the U2 snRNP and NTC before and after focussed refinement. Focussed refinement improves the resolution distribution.


a
Overall C refinement
(local resolution filtered)
Focused refinements
Brr2/Prp16
7-8 Å
U2 snRNP
3.6 Å
NTC
3.8 Å
Prp17
3.7 Å
Prp19
7.3 Å
Cwc22-N
3.6 Å
Core
2.8 Å
U5 Sm
3.4 Å
Composite map
(after focused refinements)
2.8
6.0
>9.0
Local resolution (Å)
b
Map-Model FSC
Fourier Shell Correlation
1.0
0.8
0.6
0.5
0.4
0.2
0.0
Brr2/Prp16
8.57 Å
Cwc22-N
4.29 Å
U5 Sm
4.18 Å
NTC
3.56 Å
U2 snRNP
3.35 Å
Core
2.77 Å
0.0
10.0
5.0
3.33
2.8
2.5
2.24
Resolution (Å)
c
Clf1
Leu129
Arg124
Arg119
Ile115
Leu192
Prp8
Phe193
IP6
Inositol hexaphosphate
d
Glu90
U2 SmB
Arg88
Asn42
U2 snRNA
U115
His40
Syf1
Lys32
Trp220
Arg224
xlink
Arg225
Isy1
Lys232
Lea1
e
Prp17
WD40
f
U2 snRNA
stem IIb
A76
Phe162
Cwc2
g
Prp45
Prp46
Tyr51
Lys10
Arg50
Arg5
Met1
Syf2
His69
Phe68
Trp78
Syf2
Syf1
Lys362
xlinks
Lys11
Lys14
Ntc20
Arg21
Leu9
Arg5
Pro2
h
Cwc22
NTD
i
C-terminal cassette
Brr2
N-terminal cassette
N-RecA1
C-HLH
C-FN3
N-RecA2
N-Ratchet
N-HLH
Prp8
Jab1/MPN
j
Prp16
Brr2
N-HLH
Prp8
Jab1/MPN
CTD
RecA1
+22
N- or C-terminal
helices?
+17
RecA2
Intron
+17 to +22
k
Prp19
tetramer
Prp19
WD40
U-box dimer
U-box dimer
Cef1
Snt309
Prp19
WD40

Extended Data Fig. 5 | Representative cryo-EM densities for the C complex. a, The overall C-complex reconstruction and a composite map of all focussed refinements. Local resolution for each focus-refined map was first calculated in RELION. Individual local-resolution maps were then resampled in Chimera on the overall reconstruction, and at each voxel the local-resolution map with the lowest value (i.e. best resolution) was taken. **b,** FSC curves calculated in PHENIX for each subcomponent of the model (see **Table 1**) against its respective focus-refined map. **c,** Representative densities for proteins and ligands in the C-complex core. Density-modified map is shown. **d-h,** Densities for the focus-refined density-modified maps of the U2 snRNP (**d**), Prp17 WD40 domain, which unambiguously identifies orientation (**e**), Cwc2 (**f**), NTC components (**g**), and for the Cwc22 NTD (**h**). All crosslinks shown are from the B^{act} complex (ref. 73). **i,** Focus-refined Brr2 map. **j,** Focus-refined overall helicase module map around Prp16. **k,** Focus-classified map of the Prp19 tetramer.



a **C complex** *post-branching* 2.8 Å / this study in vitro assembly, KCl purification

b **Group IIA intron** *post-hydrolysis* 3.1 Å / PDB 4FAR in vitro assembly, KCl purification

c **Group IIB intron** *post-exon ligation* 3.7 Å / PDB 4R0D in vitro assembly, NaCl purification

d ***S. cerevisiae* B^{act} complex** *pre-branching* 3.5 Å / PDB 5GM6 / EMDB 9524 endogenous, NaCl purification

e ***S. cerevisiae* B* complex** *pre-branching* 3.7 Å / PDB 6J6Q / EMDB 0692 in vitro assembly, KCl purification

f ***S. cerevisiae* C complex** *post-branching* 2.8 Å / this study in vitro assembly, KCl purification

g ***S. cerevisiae* P complex** *post-exon ligation* 3.3 Å / this study / EMDB 10140 in vitro assembly, KCl purification

h ***S. cerevisiae* P complex** *post-exon ligation* 3.3 Å / PDB 6BK8 / EMDB 7109 endogenous, NaCl purification

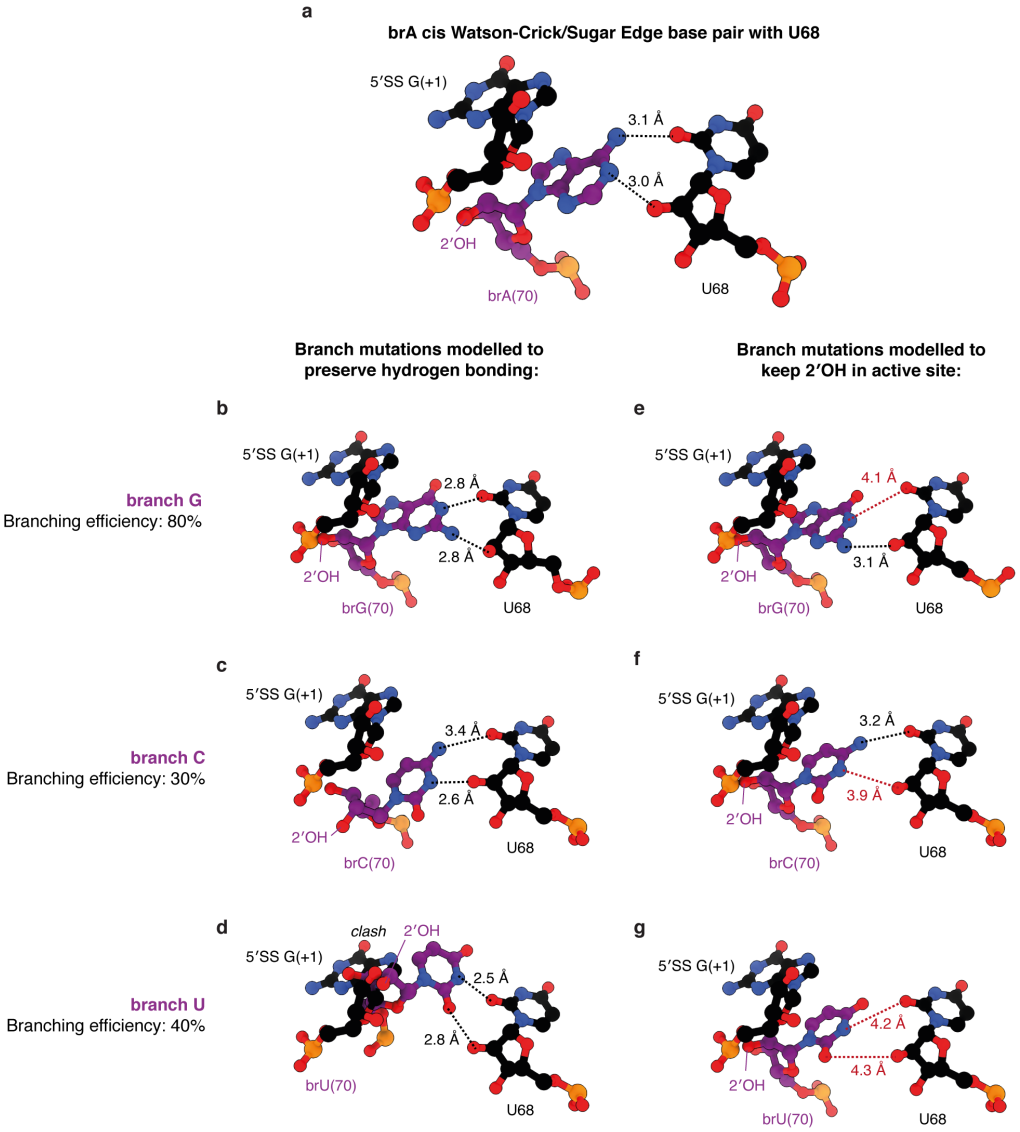
i ***H. sapiens* B^{act} complex** *pre-branching* 3.4 Å / PDB 6FF4 / EMDB 4255 in vitro assembly, NaCl purification

j ***H. sapiens* C complex** *post-branching* 4.1 Å / PDB 5YZG / EMDB 6864 in vitro assembly, NaCl purification

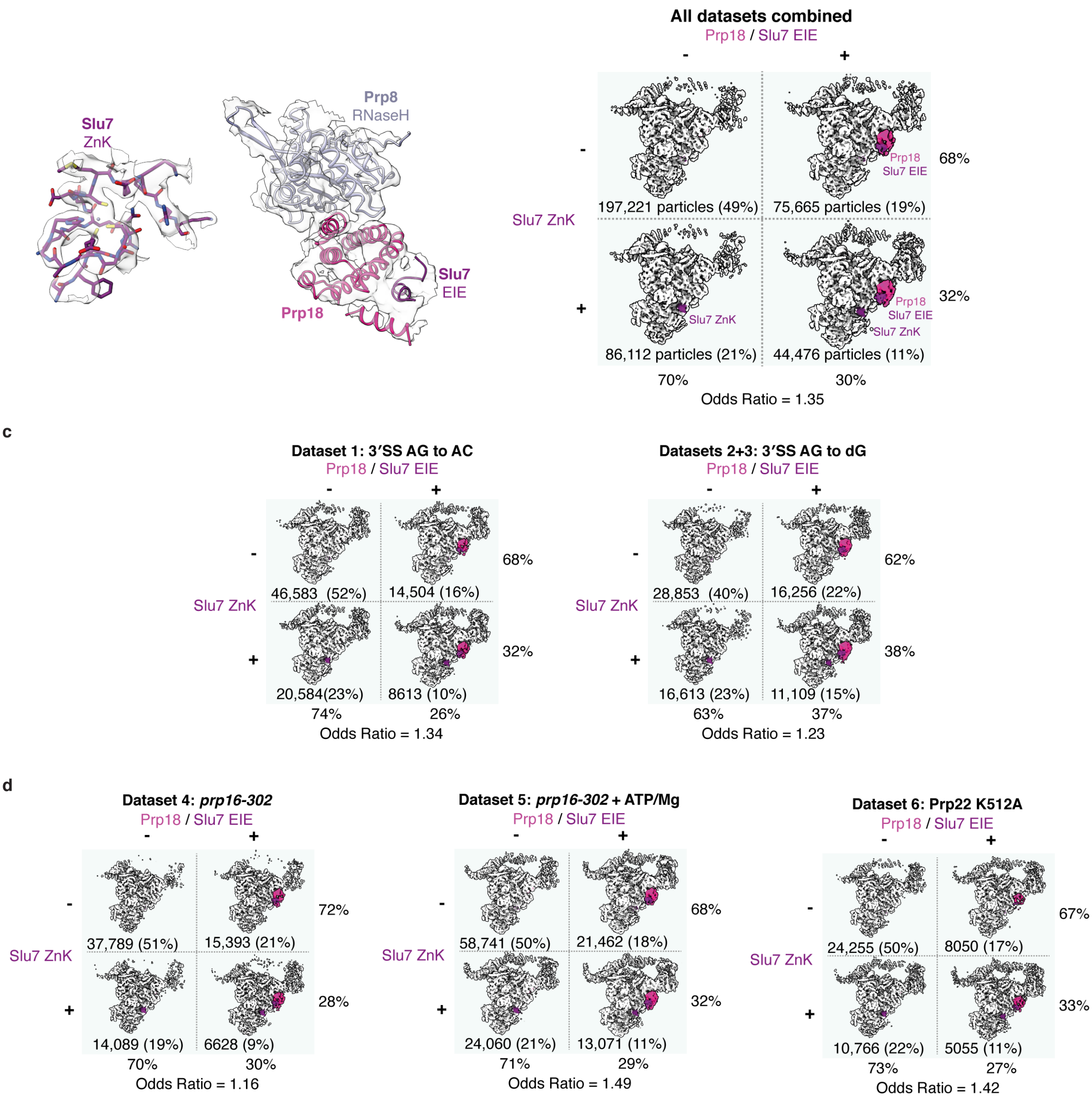
k ***H. sapiens* P complex** *post-exon ligation* 3.3 Å / PDB 6QDV / EMDB 4525 in vitro assembly, KCl purification

U6 snRNA
5´-exon
M1
M2
K1
U80
A59
G60
G52
Domain V
C377
C358
C359
G288
Ligated exons
A574
A551
G552
G421
Na+
pre-mRNA
G78
G(-1)
G(+1)
Prp11
K10
empty site
brA
empty site ?
Lariat intermediate
occupied site
mRNA
3´-SS G(-1)
M2 site
G72
U74
G54
G46
A53
SF3A2
occupied site (modelled, not refined)

Extended Data Fig. 6 | Comparison of key active site elements of the different spliceosomal complexes. a-c, Comparison between the proposed M1-M2-K1 C-complex metal cluster and the metal cluster observed in structures of group II introns. A potassium ion stabilises the catalytic centre of the spliceosome. **d-h,** Structure of the M1-M2-K1 catalytic metal cluster in various *S. cerevisiae* spliceosome complexes. The K1 metal cluster is partially blocked in B^{act} by Prp11. Potassium is likely absent at the K1 site in the structure of the P complex assembled endogenously due to the use of NaCl instead of KCl during purification. **i-j,** Structure of the M1-M2-K1 catalytic metal cluster in various *H. sapiens* spliceosome complexes. The K1 metal cluster is partially blocked in B^{act} by SF3A2, the Prp11 homolog. Clear density for potassium is likely not resolved at the K1 site in the structure of the human C complex assembled endogenously due to lower resolution and the use of NaCl instead of KCl during purification. For clarity, in all panels the potassium ion is represented using 0.5x of its Van der Waals radius to allow visualisation of the density. The magnesium ions are represented using their 1x of their van der Waals radii.


a
brA cis Watson-Crick/Sugar Edge base pair with U68
5′SS G(+1)
3.1 Å
3.0 Å
2′OH
brA(70)
U68
Branch mutations modelled to preserve hydrogen bonding:
Branch mutations modelled to keep 2′OH in active site:
b
e
branch G
Branching efficiency: 80%
5′SS G(+1)
2.8 Å
2.8 Å
2′OH
brG(70)
U68
5′SS G(+1)
4.1 Å
3.1 Å
2′OH
brG(70)
U68
c
f
branch C
Branching efficiency: 30%
5′SS G(+1)
3.4 Å
2.6 Å
2′OH
brC(70)
U68
5′SS G(+1)
3.2 Å
3.9 Å
2′OH
brC(70)
U68
d
g
branch U
Branching efficiency: 40%
clash
2′OH
5′SS G(+1)
2.5 Å
2.8 Å
brU(70)
U68
5′SS G(+1)
4.2 Å
4.3 Å
2′OH
brU(70)
U68

Extended Data Fig. 7 I Recognition of brA in C complex. a, Proposed model for brA interactions in C complex, as derived from our new high resolution map. **b-g,** Modelling of analogous interactions between mutations at the brA and U68 of the intron. Models **b-d** are based on prototypical examples of analogous base pairs as documented by Leontis and Westhof (ref. 44). Models **e-g** are simple mutations of the branch point base, keeping the sugar and phosphate in the starting position. Branch G can accommodate base pairing and correct 2′OH positioning, while branch C and U can only accommodate either pairing or correct 2′OH positioning.


a
Slu7
ZnK
Prp8
RNaseH
Slu7
EIE
Prp18
b
All datasets combined
Prp18 / Slu7 EIE
-
+
-
+
Slu7 ZnK
Prp18
Slu7 EIE
Slu7 ZnK
Prp18
Slu7 EIE
Slu7 ZnK
197,221 particles (49%)
75,665 particles (19%)
86,112 particles (21%)
44,476 particles (11%)
68%
32%
70%
30%
Odds Ratio = 1.35
c
Dataset 1: 3′SS AG to AC
Prp18 / Slu7 EIE
Slu7 ZnK
46,583 (52%)
14,504 (16%)
20,584(23%)
8613 (10%)
68%
32%
74%
26%
Odds Ratio = 1.34
Datasets 2+3: 3′SS AG to dG
Prp18 / Slu7 EIE
Slu7 ZnK
28,853 (40%)
16,256 (22%)
16,613 (23%)
11,109 (15%)
62%
38%
63%
37%
Odds Ratio = 1.23
d
Dataset 4: prp16-302
Prp18 / Slu7 EIE
Slu7 ZnK
37,789 (51%)
15,393 (21%)
14,089 (19%)
6628 (9%)
72%
28%
70%
30%
Odds Ratio = 1.16
Dataset 5: prp16-302 + ATP/Mg
Prp18 / Slu7 EIE
Slu7 ZnK
58,741 (50%)
21,462 (18%)
24,060 (21%)
13,071 (11%)
68%
32%
71%
29%
Odds Ratio = 1.49
Dataset 6: Prp22 K512A
Prp18 / Slu7 EIE
Slu7 ZnK
24,255 (50%)
8050 (17%)
10,766 (22%)
5055 (11%)
67%
33%
73%
27%
Odds Ratio = 1.42

Extended Data Fig. 8 I Focussed classification of C-complex particles for the C$_i$ conformation. a, Density quality for Prp18 and Slu7 ZnK and EIE domains in the final map obtained by classification of all datasets. **b,** Focused classification of Prp18 and the Slu7 ZnK for all combined C-complex datasets. **c,** Focused classification of Prp18 and the Slu7 ZnK for C-complex datasets resulting from impairment of exon ligation by mutation of the 3′-SS. **d,** Focused classification of Prp18 and the Slu7 ZnK for C-complex datasets resulting from impairment of exon ligation by mutation of Prp16 or Prp22.

Table 1 | Cryo-EM data collection, refinement and validation statistics

	Core (EMDB-xxxx)	U2 snRNP (EMDB-xxxx)	NTC (EMDB-xxxx)
Data processing			
Particle images (no.)	403,474	294,934	242,546
Map resolution (FSC=0.143) (Å)	2.80	3.58	3.82
Map resolution range[a] (Å)	2.71 – 4.94	3.39 – 8.20	3.64 – 8.68
Map resolution after density modification (Å)	2.69	3.26	3.50
Refinement			
Refined residues[b]	/6/E/I/C/J/K/N/M/P/L/F/a/c2:4-106/5:27-126/A:126-2078,2500/O:11-252/S:36-256/y:95-211/R:2-26/H:289-486/o:51-76/c:122-140,400/G:2-98/D:1-210	/2:107-1109/Y/W/q/p/r/n/k/l/m/G:214-232	/T/S:257-625/y:2-81/Z/G:163-191/K:1-11/J:46-59/X/O:304-327/D:209-225
Model resolution (FSC=0.5) (Å)	2.76	3.33	3.57
Map CC (around atoms)	0.68	0.75	0.69
Map sharpening *B* factor (Å^2)	-89	-138	-157
Model composition			
Non-hydrogen atoms	51,668	9959	9030
Protein residues	5463	932	1260
RNA residues	342	119	0
Ligands	17	0	0
Mean *B* factors (Å^2)			
Protein	100.3	59.0	94.2
RNA	129.7	182.9	
Ligand	104.8		
R.m.s. deviations			
Bond lengths (Å)	0.007	0.006	0.011
Bond angles (°)	1.203	1.195	1.339
Validation			
MolProbity score	0.60	1.14	0.85
Clashscore	0.26	0.58	0.48
Rotamer outliers (%)	0.14	1.77	0.27
CaBLAM outliers (%)	0.78	0.91	0.87
RNA geometry			
Correct sugar puckers (%)	97.1	96.6	
Good backbone conformations (%)	79.8	79.8	
Ramachandran plot			
Favoured (%)	98.56	96.25	96.94
Allowed (%)	1.42	3.42	3.06
Disallowed (%)	0.02	0.33	0.00

[a] Range from 5th to 95th percentiles of local resolution map within the refinement mask

[b] Shown in ChimeraX selection notation, where / indicates chain ID and : indicates residue IDs, e.g. /2:4-47 is chain 2 (U2 snRNA) residues 4 – 47.

Table 1 (cont.) | Cryo-EM data collection, refinement and validation statistics

	U5 Sm (EMDB-xxxx)	Cwc22 NTD (EMDB-xxxx)	Helicase module (EMDB-xxxx)
Data processing			
Particle images (no.)	65,561	73,802	103,461
Map resolution (FSC=0.143) (Å)	3.13 overall 3.4 around U5 Sm	3.16 overall 3.6 around Cwc22N	7.15, 8.0 (composite map)
Map resolution range[a] (Å)	3.06 – 8.41 around U5 Sm	3.40 – 5.24 around Cwc22N	6.43 – 10.56
Map resolution after density modification (Å)	2.96	2.97	
Refinement			
Refined residues[b]	/5:1-26,127-178/b/d/e/f/g/h/j	/H:10-264/X:1-20	/A:2150-2395/B/Q/I:87-92
Model resolution (FSC=0.5) (Å)	4.18	4.29	8.57
Map CC (around atoms)	0.60	0.64	0.65
Map sharpening *B* factor (Å^2)	-74	-89	-275, -494
Model composition			
Non-hydrogen atoms	6424	2166	20,741
Protein residues	602	273	2577
RNA residues	78	0	6
Ligands	0	0	0
Mean *B* factors (Å^2)			
Protein	162.9	154.39	Not refined
RNA	257.7		Not refined
Ligand			
R.m.s. deviations			
Bond lengths (Å)	0.006	0.006	0.012
Bond angles (°)	1.197	1.171	1.945
Validation			
MolProbity score	0.99	0.67	1.09
Clashscore	0.49	0.23	1.13
Rotamer outliers (%)	1.85	0.00	0.82
CaBLAM outliers (%)	0.88	0.40	1.61
RNA geometry			
Correct sugar puckers (%)	93.6		83.3
Good backbone conformations (%)	71.8		83.3
Ramachandran plot			
Favoured (%)	97.43	97.62	96.06
Allowed (%)	2.40	2.38	3.74
Disallowed (%)	0.17	0.00	0.19

[a] Range from 5th to 95th percentiles of local resolution map within the refinement mask

[b] Shown in ChimeraX selection notation, where / indicates chain ID and : indicates residue IDs, e.g. /2:4-47 is chain 2 (U2 snRNA) residues 4 – 47.

Table 1 (cont.) | Cryo-EM data collection, refinement and validation statistics

	Prp17 WD40 (EMDB-xxxx)	Prp19 (EMDB-xxxx)	Combined model (EMDB-xxxx) (PDB-NNNN)
Data processing			
Particle images (no.)	59,152	49,514	NA
Map resolution (FSC=0.143) (Å)	3.67 overall	7.30	Composite map
Map resolution range[a] (Å)	3.27 – 4.74 around Prp17	6.79 – 9.48	2.85 – 10.6
Refinement			
Refined residues[b]	/o:149-454	/O:484-536/s/t/u/v/w/X:35-51	All
Model resolution (FSC=0.5) (Å)	Not determined	Not determined	3.1
Map CC (around atoms)	0.43	0.39	Not determined
Map sharpening *B* factor (Å^2)	-90	-200	-
Model composition			
Non-hydrogen atoms	2454	6729	109,171
Protein residues	304	906	12,317
RNA residues	0	0	545
Ligands	0	0	17
Mean *B* factors (Å^2)			
Protein	Not refined	Not refined	134.12
RNA			156.50
Ligand			104.8
R.m.s. deviations			
Bond lengths (Å)	0.013	0.023	0.010
Bond angles (°)	2.180	2.566	1.513
Validation			
MolProbity score	0.95	1.09	0.82
Clashscore	0.21	2.36	0.65
Rotamer outliers (%)	1.09	0.90	0.59
CaBLAM outliers (%)	2.70	0.48	1.00
RNA geometry			
Correct sugar puckers (%)			96.3
Good backbone conformations (%)			78.3
Ramachandran plot			
Favoured (%)	95.00	97.67	97.47
Allowed (%)	5.00	2.33	2.45
Disallowed (%)	0.00	0.00	0.08

[a] Range from 5th to 95th percentiles of local resolution map within the refinement mask

[b] Shown in ChimeraX selection notation, where / indicates chain ID and : indicates residue IDs, e.g. /2:4-47 is chain 2 (U2 snRNA) residues 4 – 47.