nature structural & molecular biology

Article

https://doi.org/10.1038/s41594-024-01480-7

Structures of aberrant spliceosome intermediates on their way to disassembly

Received: 25 September 2023

Accepted: 19 December 2024

Published online: 20 January 2025

Check for updates

Komal Soni[1] ✉, Attila Horvath[2], Olexandr Dybkov[3], Merlin Schwan[1], Sasanan Trakansuebkul[2], Dirk Flemming[1], Klemens Wild[1], Henning Urlaub[3,4], Tamás Fischer[2] ✉ & Irmgard Sinning[1] ✉

Intron removal during pre-mRNA splicing is of extraordinary complexity and its disruption causes a vast number of genetic diseases in humans. While key steps of the canonical spliceosome cycle have been revealed by combined structure–function analyses, structural information on an aberrant spliceosome committed to premature disassembly is not available. Here, we report two cryo-electron microscopy structures of post-B^{act} spliceosome intermediates from *Schizosaccharomyces pombe* primed for disassembly. We identify the DEAH-box helicase–G-patch protein pair (Gih35–Gpl1, homologous to human DHX35–GPATCH1) and show how it maintains catalytic dormancy. In both structures, Gpl1 recognizes a remodeled active site introduced by an overstabilization of the U5 loop I interaction with the 5′ exon leading to a single-nucleotide insertion at the 5′ splice site. Remodeling is communicated to the spliceosome surface and the Ntr1 complex that mediates disassembly is recruited. Our data pave the way for a targeted analysis of splicing quality control.

Pre-mRNA splicing is performed by the multisubunit and highly dynamic ribonucleoprotein (RNP) particle known as the spliceosome[1–3], where two transesterification reactions, branching and exon ligation, produce the mature mRNA, as previously reviewed[4–6]. Initial steps of spliceosome assembly comprise recognition of the 5′ splice site (5′ss), branch site (BS) and 3′ss leading to the formation of an early complex (E) that transitions into the prespliceosome complex (A) and subsequently the fully assembled precursor spliceosome complex (pre-B). The spliceosome then transforms into the activated (B^{act}) and catalytically activated (B^{*}) forms before the first transesterification reaction, then into the step I (C) and step II activated complexes (C^{*}) before the second transesterification reaction and finally into the postcatalytic (P) and the intron lariat spliceosome (ILS), which lead to spliceosome disassembly for further rounds of processing.

This sequential remodeling of the spliceosome is driven by eight conserved DEAD/H RNA helicases[7,8], a subset of which (namely Prp5 (ref. 9), Prp2 (refs. 10,11), Prp16 (refs. 12,13) and Prp22 (ref. 14)) also guarantee splicing fidelity by actively promoting discard of aberrant RNA substrates as part of an internal splicing quality control or proofreading mechanism[15]. In addition to these, the RNA helicase Prp43 is responsible for dismantling the ILS to recycle the small nuclear (sn) RNPs[16–19]. Prp43 and Prp2 work together with their respective G-patch protein coactivators Ntr1 (refs. 17,20) and Spp2 (refs. 21,22). G-patch proteins are defined by a ~50-aa-long glycine-rich motif domain[23]. They are conserved in many RNA-processing proteins and serve as critical cofactors of RNA helicases[23,24]. While the action of Prp2–Spp2 triggers remodeling of the spliceosome from the B^{act} to B^{*} complex[25–27], Prp43 together with the G-patch motif of Ntr1 can disassemble the ILS complex[28]. However, Prp43 is additionally responsible for the discard of spliceosomes stalled at stages dependent on Prp16 and Prp22 (refs. 13,29–32). To selectively discard only the ILS and defective spliceosomes and prevent disassembly of properly assembled spliceosomes, Prp43 has a second associated protein, Ntr2 (ref. 17,33), which, together with the C-terminal domain (CTD) of Ntr1, acts as a

[1]Heidelberg University Biochemistry Center (BZH), Heidelberg, Germany. [2]The John Curtin School of Medical Research, The Australian National University, Canberra, Australian Capital Territory, Australia. [3]Bioanalytical Mass Spectrometry group, Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany. [4]Bioanalytics, Institute for Clinical Chemistry, University Medical Center Göttingen, Göttingen, Germany. ✉e-mail: komal.soni@uni-bayreuth.de; tamas.fischer@anu.edu.au; irmi.sinning@bzh.uni-heidelberg.de

doorkeeper for a productive disassembly of only the ILS and defective spliceosomes[34]. Prp43 and its cofactors Ntr1, Ntr2 and the stabilizing Cwc23 together form the Ntr1 complex.

After the determination of the first near-atomic cryo-electron microscopy (cryo-EM) structure of the ILS spliceosome from the fission yeast *Schizosaccharomyces pombe* (*sp*) in 2015 (ref. 35), all major functional states of the spliceosome from *Homo sapiens* (*hs*) and *Saccharomyces cerevisiae* (*sc*) have been structurally characterized providing a detailed mechanistic understanding of the canonical pre-mRNA splicing[6]. However, a high-resolution structure of a defective spliceosome intermediate committed to premature disassembly providing insights into such a splicing quality control mechanism is not available to date.

Isolation of a defective spliceosome intermediate

Recent studies found an association of the evolutionarily conserved protein Nrl1 (homologous to *hs*NRDE-2) with components of the splicing machinery[36–38], where it is part of a subcomplex known as the CNM (named after the proteins Ctr1, Nrl1 and Mtl1) that targets unspliced pre-mRNAs to the exosome for degradation during RNA surveillance[38,39]. In fact, tandem affinity purifications (TAPs) using Nrl1 as bait consistently copurified the spliceosome disassembly factors Prp43, Ntr1 and Ntr2 in high abundance[36,37,39]. A similar protein interactome was identified using reciprocal purifications with Ntr1 and Ntr2 as bait proteins[40].

We reasoned that a split-tag approach using affinity tags on Nrl1 and Prp43 might yield an enrichment of a spliceosome intermediate assembled on unspliced pre-mRNAs on its way to the discard pathway for subsequent cryo-EM analysis. Indeed, such an approach led to copurification of Nrl1-associated proteins and the Ntr1 complex (Extended Data Fig. 1a–c; the *S. cerevisiae* nomenclature is used for *S. pombe* proteins wherever homologous proteins exist; *S. pombe* protein names provided in Extended Data Table 1). We also identified the DEAH-box helicase Gih35 and the G-patch domain-containing protein Gpl1 consistent with previous reports, where Gih35 copurified with Gpl1 in TAPs[40] and other splicing factors[35,39,41]. We subjected the complex to cryo-EM, which yielded two reconstructions of a post-B^{act} spliceosome primed for discard (B^{d}) at average resolutions of 3.2 and 3.1 Å (Extended Data Fig. 1d–g, Supplementary Fig. 1 and Table 1).

Architecture of a spliceosome intermediate ready for discard

The two *sp*B^{d} structures are similar with a few compositional differences that represent two distinct states, defined here as *sp*B^{d}-I and *sp*B^{d}-II. We first describe similarities and later discuss differences between the two states. The overall architecture of the *sp*B^{d} complex (referring to both *sp*B^{d}-I and *sp*B^{d}-II) resembles a combination of B* and ILS complexes (Fig. 1 and Table 1). The *sp*B^{d} complex comprises the stable core of the spliceosome including the U5 snRNP, U6 snRNA, U2–U6 duplex, Prp19 complex (also known as the NineTeen complex (NTC)) and the NTC-related (NTR) proteins, which all remain largely unchanged in the B^{act} to ILS complexes[6] and in the *sp*B^{d} complex (Extended Data Fig. 2a). In addition to the core, we find the 5′ exon-stabilizing factors Cwc21 and Cwc22, which are recruited into the B^{act} complex and dissociate at the P-to-ILS transition[42] (Extended Data Fig. 2b). Moreover, Cwf11 (homologous to Aquarius in humans) is present in the *sp*B^{d} complex, a protein that is required for the transition from B^{act} to B* during catalytic activation in humans[43] (Extended Data Fig. 2b,c). The U2 snRNP comprises a 5′ domain (SF3b complex), a 3′ domain (core domain containing the Sm ring, Msl1 and Lea1) and the SF3a complex that bridges both domains. While the SF3a–SF3b complex is released during the B^{act}-to-B* transition, the core domain remains bound although it undergoes dramatic translocation during the B^{act}-through-ILS complex transitions. However, in our structures, the U2 core domain remains unidentified (Extended Data Fig. 2b). Most importantly, within the *sp*B^{d} complex, aided by crosslinking mass spectrometry (MS) (Extended Data Fig. 3a), we localized the helicase Gih35 and its associated G-patch-containing protein Gpl1 in our structure (Fig. 1).

Despite the overall similarity of the two *sp*B^{d} structures, there are notable differences. In the *sp*B^{d}-I state, the cryo-EM map for two components of the Ntr1 complex, namely Ntr1 and Ntr2, is very weak in comparison to the *sp*B^{d}-II complex; therefore, they are not built in the *sp*B^{d}-I structure. An exception to this is the Ntr1 CTD, which is stably bound in both states (Figs. 1 and 2a,b). Most importantly, splicing factors Bis1 (homologous to human ESS2)[44] and Saf4 are only present in

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

	spB^{d}-I (EMD-19941) (PDB 9ESH)	*sp*B^{d}-II (EMD-19942) (PDB 9ESI)
Data collection and processing		
Magnification	105,000	105,000
Voltage (kV)	300	300
Electron exposure (e^{-} per Å^2)	49.4	49.4
Defocus range (μm)	−0.6 to −1.8	−0.6 to −1.8
Pixel size (Å)	0.822	0.822
Symmetry imposed	C_1	C_1
Initial particle images (no.)	929,930	929,930
Final particle images (no.)	61,423	72,631
Map resolution (Å)	3.2	3.1
FSC threshold	0.143	0.143
Map resolution range (Å)	2.8–23	2.7–24
Refinement		
Initial model used (PDB code)	-	-
Model resolution (Å)	3.1	3.1
FSC threshold	0.143	0.143
Model resolution range (Å)	-	-
Map sharpening *B* factor (Å^2)	48.1	43.8
Model composition		
Nonhydrogen atoms	90,868	98,220
Protein residues	10,603	11,468
RNA/DNA	247	247
Ligands	13	13
B factors (Å^2)		
Protein	162	165
RNA/DNA	147	164
Ligand	183	165
Root-mean-square deviations		
Bond lengths (Å)	0.004	0.007
Bond angles (°)	0.779	0.799
Validation		
MolProbity score	1.99	1.98
Clashscore	13.53	14.32
Poor rotamers (%)	0.01	0.01
Ramachandran plot		
Favored (%)	94.88	95.40
Allowed (%)	5.00	4.42
Disallowed (%)	0.11	0.18

https://doi.org/10.1038/s41594-024-01480-7

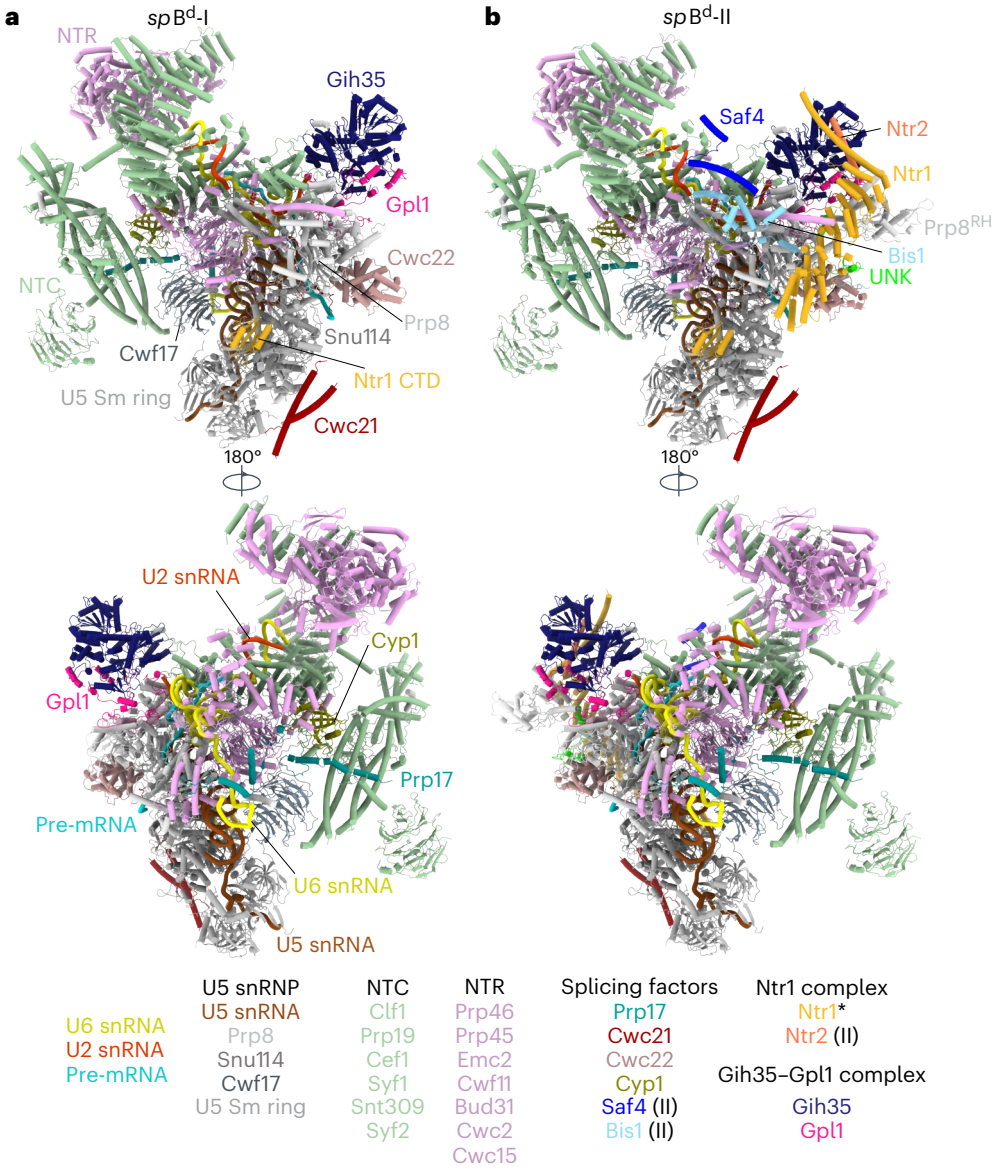


Fig. 1 | Cryo-EM structure of the *sp*B^{d} complex. a,b, Structures of the *sp*B^{d}-I (**a**) and *sp*B^{d}-II (**b**) complexes at average resolutions of 3.2 and 3.1 Å. Two views of the *sp*B^{d} complexes are shown, colored according to the different subunits, which are listed below the structure. The Ntr1 protein is marked (*) and is only partially visible in the *sp*B^{d}-I state while stabilized in the *sp*B^{d}-II state. Protein components that are only part of the *sp*B^{d}-II state are indicated.

the *sp*B^{d}-II state. Saf4 and Cwf16 both share sequence homology with the *S. cerevisiae* step I splicing factor *sc*Yju2, similar to human CCDC94 and CCDC130 (Extended Data Fig. 3b,c). We find the *sp*B^{d}-II cryo-EM map to be consistent only with Saf4, suggesting Saf4 to be the direct homolog of human CCDC94 and *sc*Yju2 rather than Cwf16, as suggested previously[45] (Extended Data Fig. 3d). In addition, two short α-helical fragments of the NTC core components Cef1 and Syf2 are destabilized in the *sp*B^{d}-II state to accommodate Saf4 and Bis1 (Fig. 2c and Extended Data Fig. 3e,f). The RNase H-like domain of Prp8 (Prp8^{RH}) is also only visible in the *sp*B^{d}-II state (Fig. 1). Lastly, well-defined density is found for multiple helices around Prp8 and Cwc22 in the *sp*B^{d}-II cryo-EM map, the identity of which could not be clarified (Fig. 2d and Supplementary Fig. 2), henceforth referred to as unknown.

Altogether, we modeled 30 (*sp*B^{d}-I) and 33 (*sp*B^{d}-II) proteins and four RNA molecules in the *sp*B^{d} cryo-EM maps (Extended Data Table 1). It is noteworthy that we do not find density corresponding to the CNM complex in either of the *sp*B^{d} cryo-EM maps although intermolecular crosslinks between Ctr1 and parts of Cwc22 are observed (Extended Data Fig. 3a), indicating that the CNM is likely located at the periphery and, therefore, flexibly linked.

The Ntr1 complex is stabilized in the *sp*B^{d}-II complex

Of the four components of the Ntr1 complex, we observe defined density only for Ntr1 and Ntr2 in the *sp*B^{d}-II cryo-EM map while Prp43 and Cwc23 remain unidentified. The positions of Ntr1 and Ntr2 proteins at the periphery of the spliceosome are similar to those described for the *sc*ILS complex (Extended Data Fig. 2b,c)[46]. Ntr1 consists of three distinct domains, namely the G-patch motif domain, a central superhelical domain and the CTD (Fig. 2a). While the Ntr1 G-patch motif domain is not found in both *sp*B^{d} states, the CTD is bound in both *sp*B^{d} states, where it is anchored on Snu114 as observed in the *sc*ILS complex (Fig. 2b and Extended Data Fig. 2b,c)[46]. In the *sp*B^{d}-II state, the superhelical domain contacts Ntr2, Prp45 and Bis1 and approaches Gih35 (Fig. 2d) while, in the *sp*B^{d}-I state, the domain has weak cryo-EM density and was, therefore, not built. This flexibility of Ntr1 superhelical domain in

https://doi.org/10.1038/s41594-024-01480-7

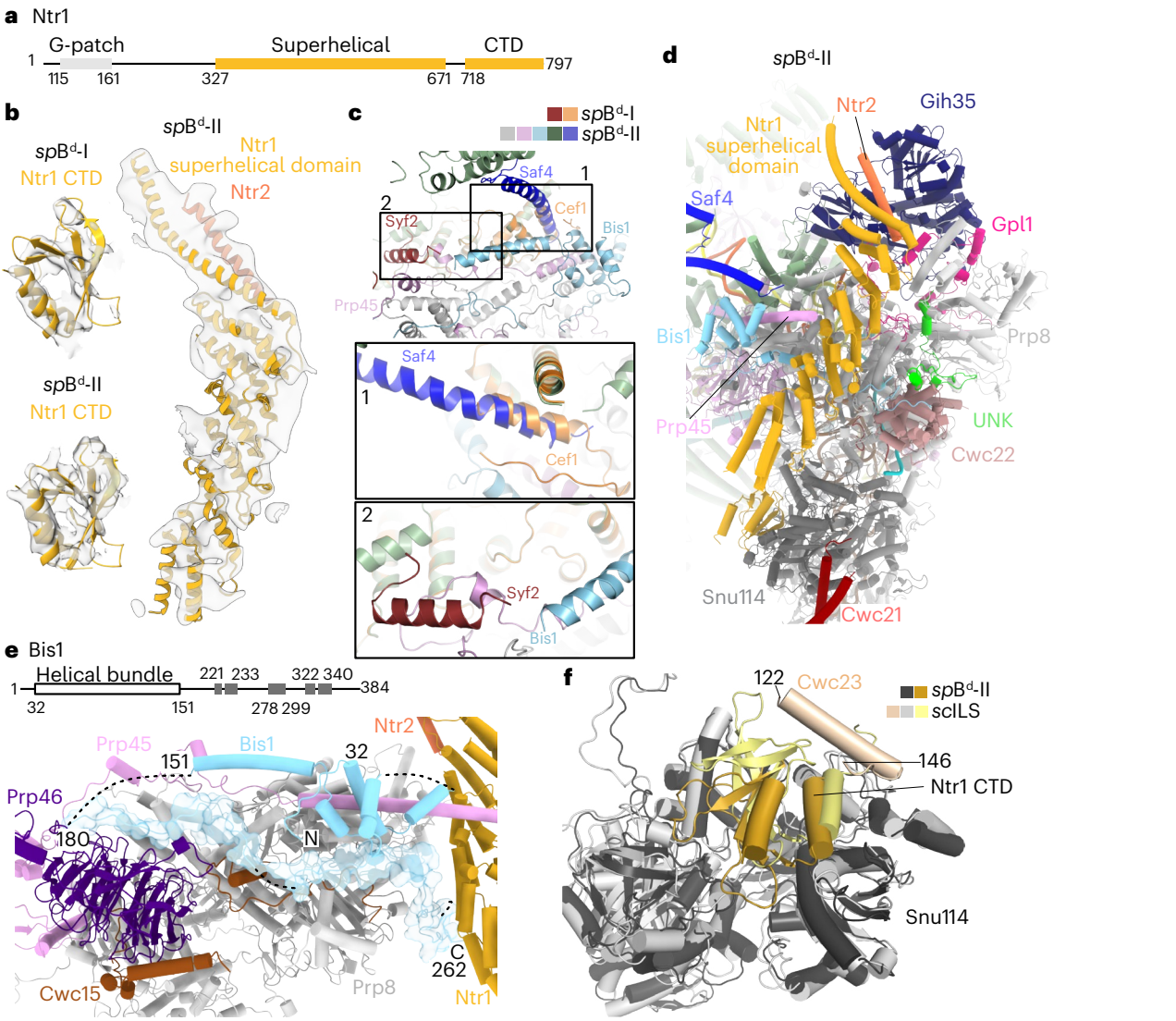


Fig. 2 | The Ntr1 complex is stabilized in the *sp*B^{d}-II state. a, Schematic representation of the domain architecture of Ntr1. **b**, Superimposition of the Ntr1 CTD from the *sp*B^{d}-I and *sp*B^{d}-II states with their respective cryo-EM maps and the Ntr1 superhelical domain and Ntr2 (residues 296–325) with the *sp*B^{d}-II cryo-EM map. **c**, Saf4 and Bis1 are incorporated in the *sp*B^{d}-II state. A structure superposition of the *sp*B^{d}-I and *sp*B^{d}-II states is shown. The color scheme of *sp*B^{d}-II state follows Fig. 1; from the *sp*B^{d}-I state, only Syf2 and Cef1 are shown (dark red and orange, respectively). Insets 1 and 2 show zoomed-in views of structural differences between the two states. Cef1 residues 379–409 in the *sp*B^{d}-I state are displaced by Saf4 in the *sp*B^{d}-II state. Syf2 residues 194–215 are destabilized by the presence of Bis1 in *sp*B^{d}-II state. **d**, The superhelical domain of Ntr1 contacts Ntr2, Prp45 and Bis1 and is positioned close to Gih35 in the *sp*B^{d}-II state. **e**, Top, domain architecture of Bis1 with the helical bundle marked and other helices shown as gray boxes. Bottom, residues 22–262 (including gaps) of Bis1 that can be traced in the *sp*B^{d}-II state. In addition to the helical bundle, ~80 residues of Bis1 can be traced (shown in surface representation). **f**, The Ntr1 CTD is anchored on Snu114. In the *sc*ILS complex, this association is stabilized by an α-helix of Cwc23 (residues 122–146).

the *sp*B^{d}-I state seems because of the absence of the two proteins Saf4 and Bis1 that stabilize the domain in the *sp*B^{d}-II state (Fig. 2d). For Bis1, only the helical bundle of its human homolog ESS2 could be localized in the *hs*C* complex (Fig. 2e)[44]. In addition to the helical bundle in Bis1, we now find a continuous stretch of ~80 residues (180–262) crawling across the surface of Prp8, Prp46 and Cwc15, part of which directly interacts with the Ntr1 superhelical domain and anchors it to Prp8 in the *sp*B^{d}-II state (Fig. 2e and Extended Data Fig. 3g).

In the *sc*ILS complex, the Ntr1 CTD is further stabilized by an α-helix of Cwc23, which is not found in the *sp*B^{d} complex (Fig. 2f). However, we do observe crosslinks between Cwc23 and the linker connecting the Ntr1 superhelical domain with the CTD, suggesting that Cwc23 is flexibly bound in the *sp*B^{d} complex (Extended Data Fig. 3a). Similarly, crosslinks between the Ntr1 G-patch domain and Prp43 suggest that Prp43 is also flexibly tethered to the *sp*B^{d} complex (Extended Data Fig. 3a). We find a small α-helical fragment of Ntr2 in the *sp*B^{d}-II state (Fig. 2b,d) positioned in close proximity to Gih35, where it may have a role in stabilization of Gih35 (Extended Data Fig. 3h). Taking all interactions and components together, the *sp*B^{d} structures represent post-B^{act} spliceosome intermediates captured on their way to disassembly.

Gpl1 is anchored to the spliceosome by Prp8

Gpl1 was recently characterized as a splicing regulator in *S. pombe*, necessary for the recruitment of Gih35 (ref. 47) and proper canonical splicing[47,48]. It is a G-patch domain-containing protein (residues 137–181) comprising 534 aa, of which about 200 residues (residues 25–188, and 199–234) can be traced in our complex (Fig. 3a,b; for structural illustrations of redundant regions between the two *sp*B^{d} states, only the higher-resolution *sp*B^{d}-II state is further described unless mentioned otherwise). Our data show how Gpl1 recruits Gih35 to the spliceosome and interacts exclusively with different domains of Gih35 and Prp8 (Fig. 3c,d).

https://doi.org/10.1038/s41594-024-01480-7

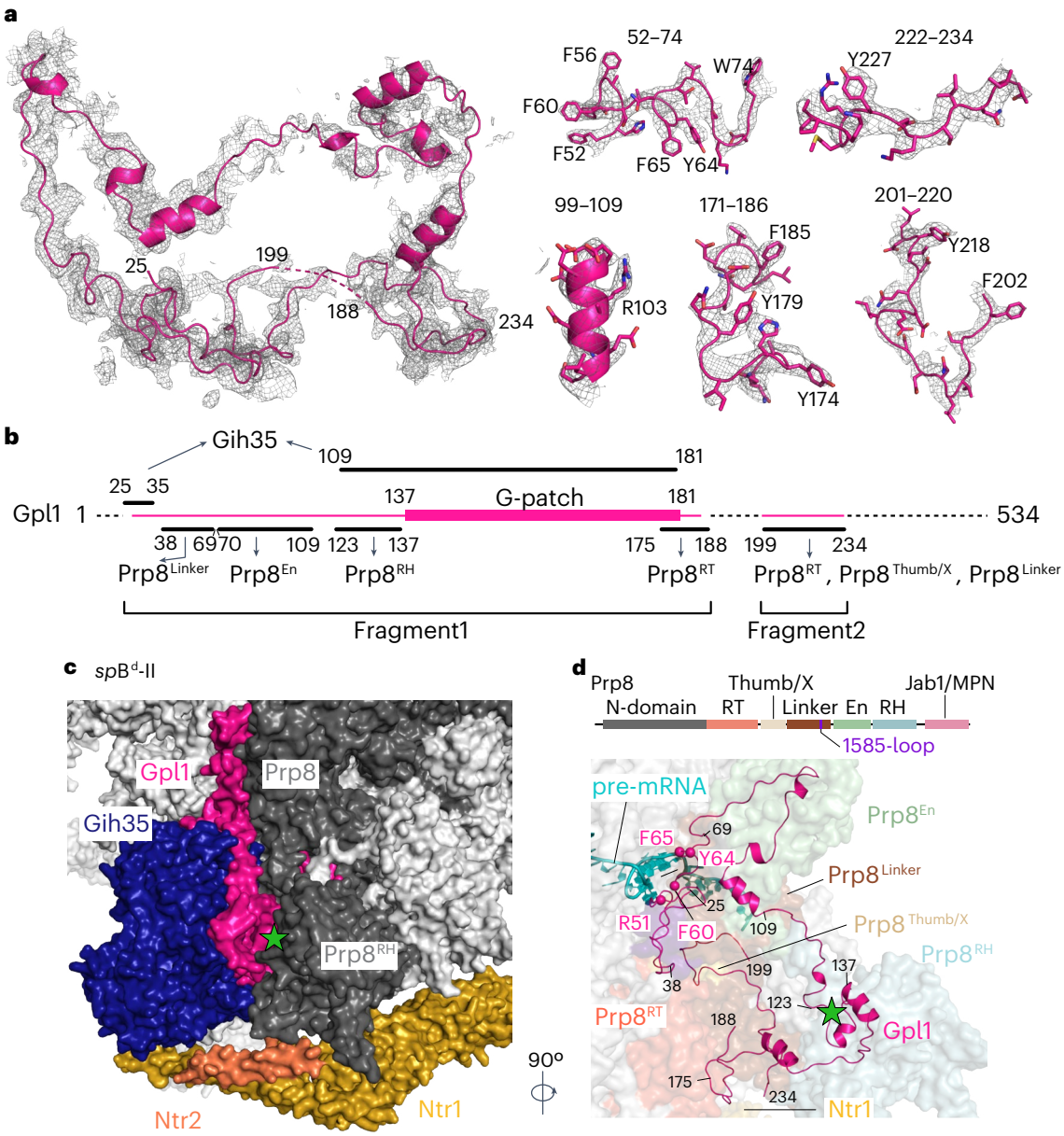


Fig. 3 | Interactions of Gpl1 and Gih35 with the spliceosome. a, Cryo-EM map for the entire Gpl1 trace visible in the *sp*B^{d}-II complex (residues 25–188, 199–234). Zoomed-in views of representative regions of Gpl1 for *sp*B^{d} interaction are shown as insets. **b**, Gpl1 forms a bridge between Prp8 and Gih35. Regions largely adjacent to the G-patch domain contribute to the interaction. The boundaries of different regions of Gpl1 interacting with the helicase Gih35 and distinct domains of Prp8 are shown (N and C termini of the two Gpl1 fragments visible in the *sp*B^{d} complex are marked with black dots). **c**, Gpl1 binds to Prp8 and Gih35 and bridges between the two proteins. **d**, Gpl1–Prp8 interaction, Top, schematic representation of the domain architecture of Prp8. Bottom, Gpl1 contacts the Prp8 RT, thumb/X, linker (including the 1585-loop), En and RH domains. Residues contacting the pre-mRNA in the active site of the spliceosome are indicated. The Gpl1 knob (residues 123–137) is marked with a green star in **c**,**d**.

Overall, the N-terminal part of Gpl1 forms a remarkably large interface with Prp8 of 5,600 Å^2. In detail, a region of Gpl1 (residues 38–69) that is fixed on the linker domain of Prp8 (Prp8^{linker}, including the 1585-loop of Prp8, also known as α-finger residues 1537–1550) also reaches into the active site of the spliceosome where it contacts the pre-mRNA with a cluster of aromatic and charged residues (region 51–65, as detailed below) (Fig. 3d). Next, Gpl1 residues 70–109 are anchored on the endonuclease domain of Prp8 (Prp8^{En}) where Gpl1 makes a sharp turn to cover a large surface area of interaction (Fig. 3d and Extended Data Fig. 4a). Furthermore, Gpl1 residues 123–137 form a knob that inserts into Prp8^{RH} (Fig. 3c,d Extended Data Fig. 4b). It is known that Prp8 rearranges throughout the splicing cycle, with the Prp8^{RH} and the Prp8$^{Jab1/MPN}$ domains displaying different positions with respect to the core, and contributes to stabilization of the exchanging spliceosome proteins and branch helix movements[49]. While Prp8$^{Jab1/MPN}$ remains invisible in both *sp*B^{d} complexes, Prp8^{RH} is observed only in the *sp*B^{d}-II state, where it adopts by far the most open conformation (Extended Data Fig. 5a,b). This extended open conformation of Prp8^{RH} is fixed by its interaction with the Gpl1 knob only in the *sp*B^{d}-II state (Extended Data Fig. 4b). Moreover, Gpl1 residues 175–188 are part of a small and rather hydrophobic interface with the Prp8 reverse transcriptase domain (Prp8^{RT}) (Fig. 3d and Extended Data Fig. 4c). Finally, Gpl1 residues 199–234 insert into a central Prp8 cavity created by its RT, thumb/X and linker domains (Fig. 3d and Extended Data Fig. 4d). This stretch of Gpl1 occupies the same general location as components of Prp45 in the *sc*B^{act} or Ntr2 in the *sc*ILS states (Extended Data Fig. 5c–f).

https://doi.org/10.1038/s41594-024-01480-7

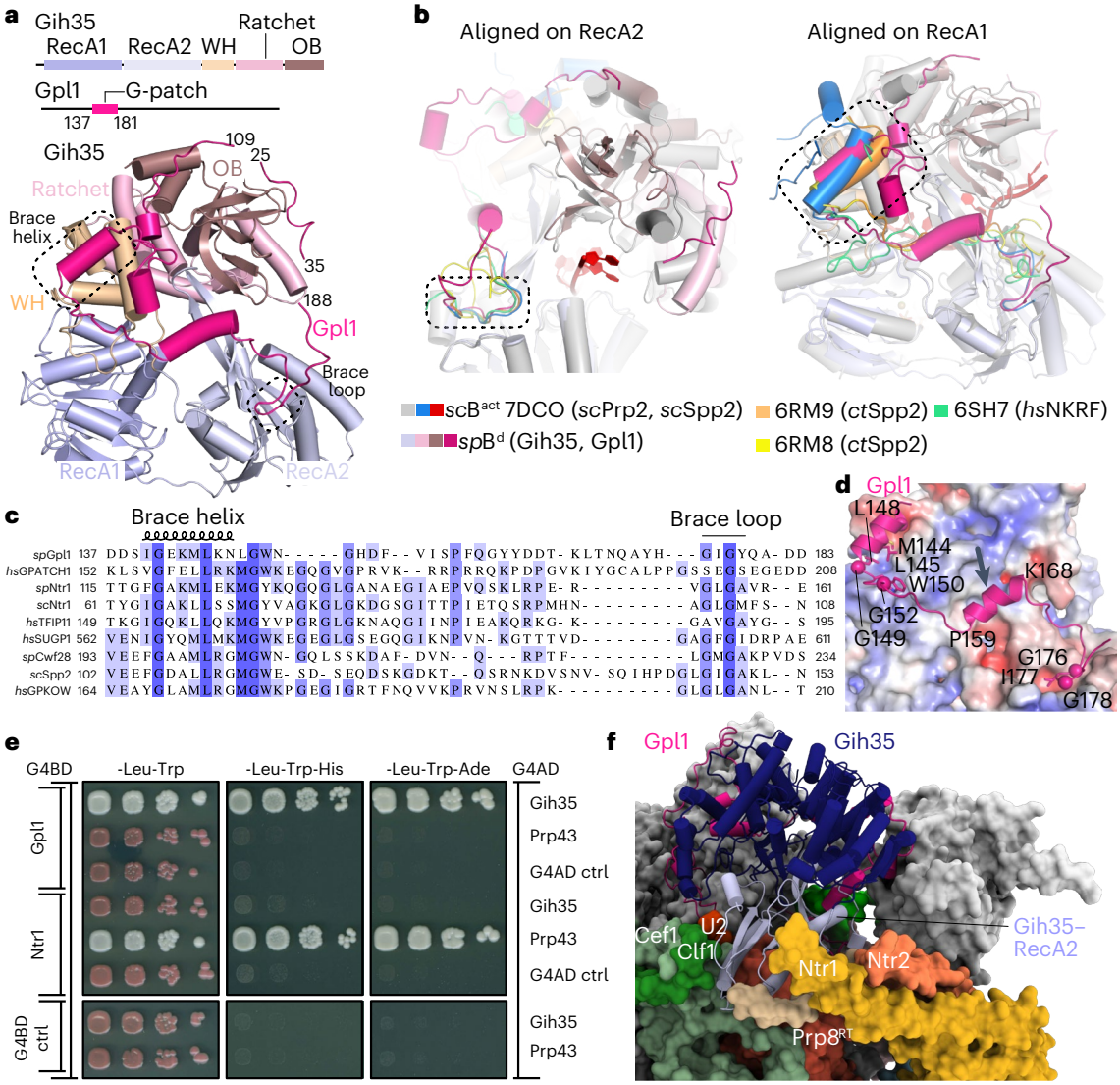


Fig. 4 | Gih35–Gpl1 is a helicase–G-patch protein pair. a, Top, schematic representation of the domain architecture of Gih35 and Gpl1. Bottom, overview of the Gih35–Gpl1 interactions. The positions of the Gpl1 brace helix and brace loop are marked. **b**, Comparison of helicase–G-patch protein interactions. Superposition of the Gih35–Gpl1 complex (*sp*B^{d}) with that of *S. cerevisiae* Prp2–Spp2, two structures of the *C. thermophilum* Prp2–Spp2 and the human DHX15–NKRF complex. The structures are aligned on the RecA2 (left) and RecA1 (right) domains of the respective helicase, showing structural conservation in the brace helix and brace loop regions, respectively. For simplicity, only the helicases Gih35 (color-coded similar to **a**) and *sc*Prp2 (gray) are shown. **c**, Sequence alignment of G-patch proteins from *S. pombe*, *S. cerevisiae* and *H. sapiens* known or predicted as coactivators of Gih35, Prp43 and Prp2 (DHX35, DHX15 or DHX16 in humans, respectively). **d**, The electrostatic surface potential (±5 kT; red, negative; blue, positive) of Gih35 is plotted. Conserved hydrophobic residues of the Gpl1 G-patch are marked and the Cα atoms of the conserved glycine residues are shown as spheres. The central α-helix in the brace linker is marked with an arrow. **e**, Y2H experiments show that Gpl1 interacts with Gih35 and Ntr1 interacts with Prp43. Full-length Gpl1 or Ntr1 constructs were fused to the Gal4 DNA-binding domain (G4BD) while full-length Gih35 and Prp43 were fused to the Gal4 activation domain (G4AD). Autoactivation controls are provided. Serial dilutions of equivalent amounts of yeast were plated on double-dropout (−Leu−Trp) and triple-dropout (−Leu−Trp−His, −Leu−Trp−Ade) media, with growth on triple-dropout media indicating an interaction between the tested proteins. **f**, In the *sp*B^{d}-II state, the Gih35 RecA2 domain (light purple) is anchored in a pocket formed by Prp8^{RT}, Ntr1, Ntr2, NTC proteins Cef1 and Clf1 and a helix of an unknown protein (light brown).

Gpl1 binds to Gih35 and tethers it to the spliceosome

Gih35 is a canonical DEAH-box RNA helicase with a conserved core comprising the RecA1 (residues 20–202) and RecA2 (residues 203–379) domains that are flanked by the C-terminal WH (residues 380–447), ratchet (residues 448–557) and OB domains (residues 558–647) (Fig. 4a). In the *sp*B^{d} complex, Gpl1 tethers Gih35 to the periphery similar to other helicases known to remodel the spliceosome, including Prp2 (refs. 50–52), Prp16 (refs. 45,53,54) and Prp22 (refs. 55–60) (Extended Data Fig. 5g). Gpl1 shares an interface of >2,500 Å^2 with Gih35, similar to Prp2, which is bound to the *sc*B^{act} complex through Spp2 (ref. 52). Two stretches of Gpl1 form distinct interactions with Gih35. While Gpl1 residues 25–35 insert in a groove between Gih35 WH and OB domains, residues 109–181 including the G-patch (137–181) traverse along the OB, WH and RecA2 domains, forming an extended interaction. The N terminus of the G-patch forms an α-helix (residues 139–148; termed as a brace helix[61]) and the C-terminal loop (termed as a brace loop[61]) inserts into a hydrophobic pocket on top of the RecA2 domain (Fig. 4a and Extended Data Fig. 4e,f). Superpositions of Gih35–Gpl1 with other known DEAH-box helicase–G-patch coactivators, such as the Prp2–Spp2 pair from *Chaetomium thermophilum* (*ct*)[62] and DHX15 with its G-patch cofactor NKRF (ref. 62), show that the

https://doi.org/10.1038/s41594-024-01480-7

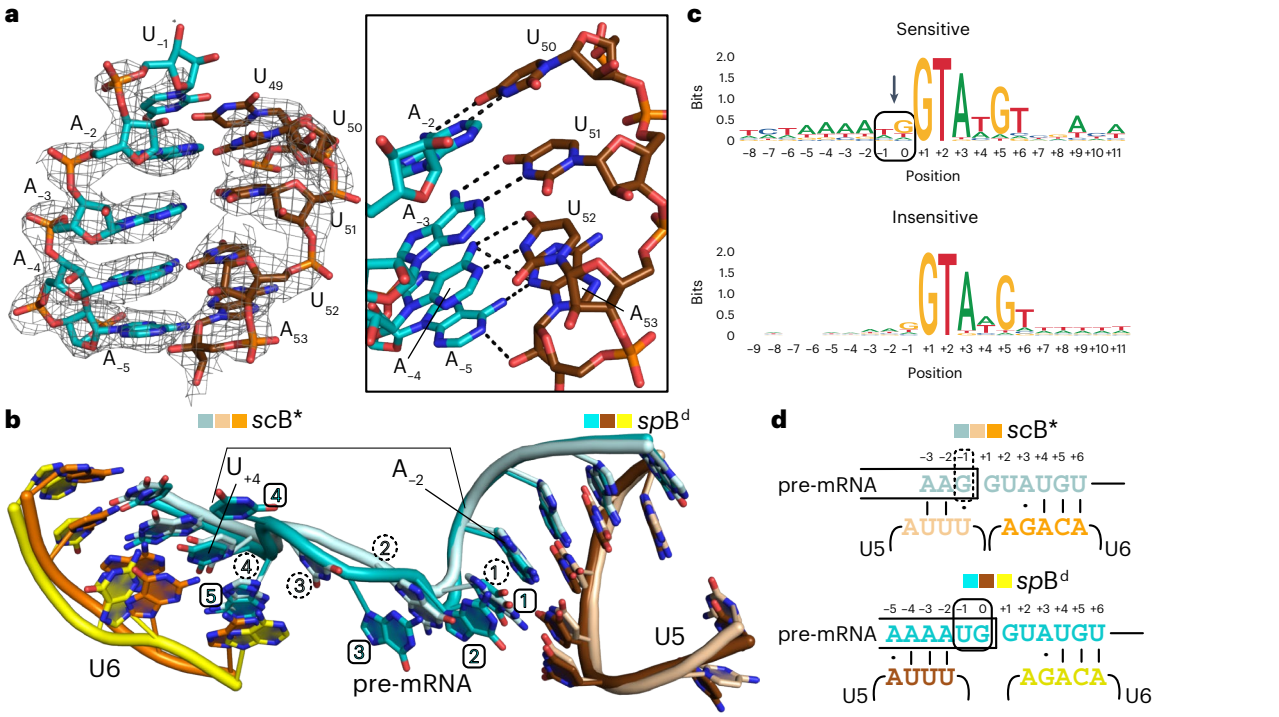


Fig. 5 | Noncanonical active site of the *sp*B^{d} complex. a, The *sp*B^{d} cryo-EM map shows that a tight RNA duplex is formed between U5 loop I and 5′ exon nucleotides $A_{-2}A_{-3}A_{-4}A_{-5}$. Inset, zoomed-in view of interactions between U5 loop I and 5′ exon with hydrogen bonds marked as dotted lines. **b**, Superposition of RNA elements (pre-mRNA, U5 and U6) at the active site of *sp*B^{d} and canonical *sc*B* (PDB 6J6Q) complexes. The bracket marks the positions of pre-mRNA bases A_{-2} and U_{+4}, which are covered by four nucleotides in *sc*B* (dashed circles) and five nucleotides in *sp*B^{d} (square boxes). **c**, The 5′ss sequence logo generated from CNM-sensitive (top) and CNM-insensitive (bottom) transcripts after normalization by gene expression. The insertion at the −1/0 position is marked with an arrow. **d**, Scheme of RNA elements at the active site of *sp*B^{d} using the 5′ss sequence logo from **c** and canonical *sc*B* (PDB 6J6Q) complexes. The insertion at the −1/0 position is marked with a continuous box in *sp*B^{d} complex while the G_{-1} position in the *sc*B* complex is marked with a dashed box.

positions of the brace helix and brace loop are fixed while the linker region connecting the N-terminal and C-terminal parts of the G-patch domain varies (Fig. 4b). In accordance, the sequence alignment of Gpl1 with other G-patch proteins that are known activators of Prp2/DHX16 or Prp43/DHX15 shows a high degree of conservation only in the brace helix and brace loop regions while the connecting linker is poorly conserved (Fig. 4c). Overall, apart from the eponymous glycines of the motif, only hydrophobic residues within the brace helix and brace loop are maintained. Specificity of the helicase–G-patch pair is likely to be ascribed to the nonconserved regions such as the brace linker, comprising a unique α-helix in Gpl1 (Fig. 4d), and/or N-terminal regions, which form additional contacts outside the G-patch. To ascertain whether Gpl1 binds specifically to Gih35 or does not discriminate between other DEAH-box helicases such as Prp43, we performed yeast two-hybrid (Y2H) experiments (Fig. 4e). While Gpl1 binds to Gih35, as also shown previously[47], it does not bind to Prp43. Similarly, Ntr1 (the bona fide interaction partner of Prp43) does not bind to Gih35 but interacts with Prp43 as expected. In accordance, the Ntr1 G-patch domain comigrates with Prp43 in size-exclusion chromatography experiments, indicative of complex formation, while the Gpl1 G-patch domain does not (Extended Data Fig. 4g). These data underline the specific interaction of the Gih35–Gpl1 pair.

In contrast to Gpl1, interactions of Gih35 with the spliceosome are sparse. Only the RecA2 domain of Gih35 is placed in a pocket formed by Prp8^{RT}, NTC proteins Cef1 and Clf1, a helix of an unknown protein and Ntr1 and Ntr2 (in the case of the *sp*B^{d}-II state) (Fig. 4f). In summary, the Gih35–Gpl1 pair is anchored on the spliceosome in a two-pronged manner with Gpl1 interacting extensively and reaching deep into the active site, while Gih35 is recruited in a piggyback manner by stabilizing contacts to the spliceosome only through its RecA2 domain.

Canonical and noncanonical features of the *sp*B^{d} active site

The comparison of snRNAs from canonical spliceosomes in catalytic states (B^{act} to ILS complexes) with the aberrant *sp*B^{d} complex shows that RNA elements at the active site superimpose well (Extended Data Fig. 6a). The U5–5′ exon duplex, U6 internal stem loop (ISL) and U2–U6 helices Ia, Ib and II are also well defined in the *sp*B^{d} structure, while the ACAGA helix is less defined (Extended Data Fig. 6b). We find three structural metal ions at the active site, likely Mg^{2+} ions, which are exclusively coordinated by the U6 ISL (Extended Data Fig. 6c and Supplementary Fig. 3). While the catalytic metal ions M1 and M2 are not incorporated into the *sp*B^{d} states, the K1 (potassium) site[63] is occupied (Extended Data Fig. 6c). Strikingly, we find a tight RNA duplex formed between the U5 snRNA and the 5′ exon with Watson–Crick base pairing between U5 loop I nucleotides $U_{50}U_{51}U_{52}$ and 5′ exon nucleotides $A_{-2}A_{-3}A_{-4}$ and optimized packing of A_{-5} with A_{53} (Fig. 5a). The pairing is similar to a cognate codon–anticodon mRNA–transfer RNA interaction.

Most importantly, we find that the 5′ss contains a single-nucleotide insertion (Fig. 5b). It is important to mention that, while the cryo-EM density for snRNAs at the active site is very well defined, that for the pre-mRNA at the 5′ss is weak in comparison, rendering identification of bases or the conformation beyond the −1 position inaccurate (Fig. 5a and Extended Data Fig. 6c,d). This is comprehensible, given that the *sp*B^{d} complex is an affinity-purified in vivo complex and, therefore, incorporates a mixture of RNA species rather than being assembled in vitro on a single RNA. Nevertheless, the presence of the insertion is very strongly supported (Fig. 5b and Extended Data Fig. 6d). Together with the assumption that the canonical base pairing in yeast spliceosomes between the U6 ACAGA box with the intron positions +4 to +6 is preserved, the strong U5 loop I interaction with the 5′ exonic adenine

https://doi.org/10.1038/s41594-024-01480-7

quadruple (positions −2 to −5) induces a defined register constraint on the pre-mRNA. In effect, the distance between the pre-mRNA −2 and +4 positions cannot be bridged by the canonical four nucleotides but needs five nucleotides instead (Fig. 5b). To accommodate the insertion at the active site, the pre-mRNA is squeezed, introducing a bend that alters its overall geometry (Fig. 5b and Extended Data Fig. 6e).

Insights into pre-mRNA targets of the *sp*B^{d} complex

To validate our structural finding of an insertion at the pre-mRNA 5′ss pre-mRNAs, we wanted to ascertain whether all pre-mRNAs bound in the *sp*B^{d} complex generally harbor the exonic quadruple adenines (positions −2 to −5) and whether this leads to the single-nucleotide insertion at the 5′ss, which could be a hallmark of transcripts targeted for discard by the CNM. We previously reported that, compared to the wild-type (WT) strain, strains harboring deletions of CNM components (*ctr1Δ*, *nrl1Δ*) showed accumulation of unspliced pre-mRNAs[38]. We reanalyzed the RNA-sequencing (RNA-seq) data for *ctr1Δ*, which previously showed the highest accumulation of unspliced pre-mRNAs[38] (Extended Data Fig. 7a–c). Using a sensitivity metric wherein introns with at least 20% increased accumulation in the *ctr1Δ* strain compared to the WT were defined as introns sensitive to CNM-mediated discard (Extended Data Fig. 7d), we found that an exonic quadruple adenine motif is slightly enriched in the sensitive transcripts compared to the insensitive ones (Extended Data Fig. 7e). Normalization of RNA species according to abundance led to the emergence of a more prevalent exonic quadruple adenine motif ($A_{-2}A_{-3}A_{-4}A_{-5}$) (Fig. 5c and Extended Data Fig. 7f). Strikingly, the motif shows the presence of five nucleotides between the A_{-2} and U_{+4} positions (and, therefore, an insertion, defined as position 0) in the sensitive transcripts, which is in agreement with the cryo-EM map (Fig. 5c,d). The RNA-seq analysis shows that the likely gene candidate, which is highly expressed and, therefore, majorly contributes to the structural observation of these features, is *rpl39* encoding the 60S ribosomal protein eL39 (Extended Data Fig. 7g,h).

To further validate these observations, we isolated and sequenced the coprecipitating RNA transcripts from our *sp*B^{d} spliceosome purification. These RNA immunoprecipitation (RNA-IP) experiments confirmed that our purifications strongly enrich intronic sequence coverage (Extended Data Fig. 7i,j). Most intronic reads in affected introns overlap with exon–intron boundaries, indicating that they represent unspliced and uncleaved pre-mRNAs before the first catalytic step of the splicing reaction (Extended Data Fig. 7i). Determination of the abundance of individual introns proved to be highly complex because of the large amount of contaminating RNA molecules in these samples, which constituted the majority of reads in these RNA-IP experiments. Nevertheless, after filtering for enriched intronic sequences in the IP fraction and excluding misannotated, alternatively spliced and retained introns, the most abundant intronic sequence was the intron in the *rpl39* gene (Extended Data Fig. 7k). Of note, contaminating RNA molecules that were present in our affinity purification were not incorporated into the protein complex and, therefore, did not appear or pose a problem in structural analysis.

Overall, these results further confirm that the pre-mRNA substrate observed in the *sp*B^{d} complex is likely *rpl39*, which is probably the most abundant substrate for the *sp*B^{d}-mediated splicing quality control in *S. pombe*.

The aberrant active site in the *sp*B^{d} complex

Following the validation of the insertion in the pre-mRNA at the active site, we proceeded with the investigation of additional changes and alterations at the active site. Remarkably, we found that the Gpl1 N-terminal region (residues 38–69) closely engulfs the active site (Fig. 6a). The nucleotides of the 5′ss, including U_{-1}, G_{0} and G_{+1}, stack with each other to form a continuous ladder that is further extended by Gpl1 F60, which stacks on top (Fig. 6b and Extended Data Fig. 6f,g).

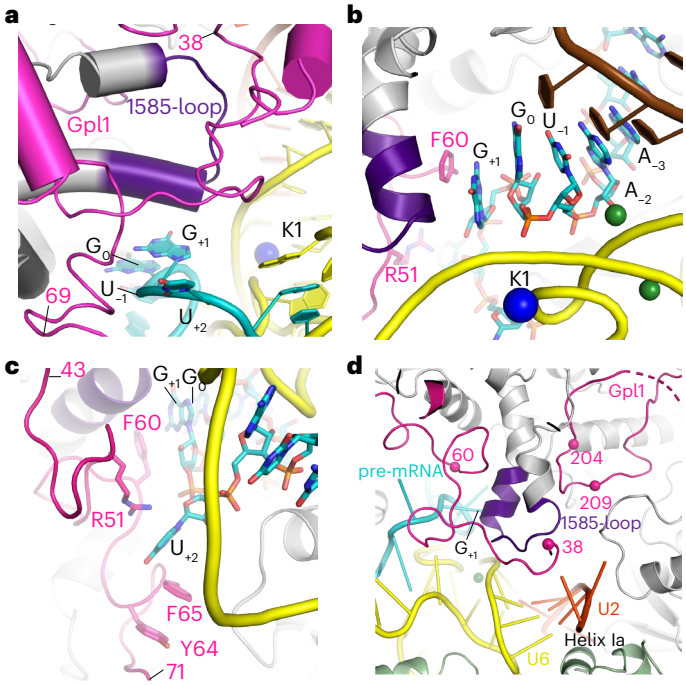


Fig. 6 | Interactions of Gpl1 at the active site. a, Gpl1 binds at the heart of the spliceosome where it interacts with the 5′ss and the 1585-loop of Prp8 (purple). The color scheme for protein and RNA elements follows Fig.1. **b**, Gpl1 residue F60 stacks with G_{+1} of the 5′ss. **c**, Rotated view showing recognition of U_{+2} of the 5′ss by Gpl1. **d**, The 1585-loop is held in position by Gpl1 residues 38–60 on one side and residues 204–209 on the other. It also directly interacts with U2–U6 helix Ia.

Because of the insertion, the canonical U_{+2} is also positioned differently and is sequestered by Gpl1 in a pocket formed by residues R51, Y64 and F65 (Fig. 6c and Extended Data Fig. 6h). The presence of Gpl1 also likely occludes the docking of the U2–BS duplex at the active site, which remains invisible in the *sp*B^{d} complex.

The positioning of Gpl1 at the active site somewhat mimics that of Prp11 and Cwc24 in the B^{act} complex, where they shield the 5′ss and, thus, maintain catalytic dormancy (Extended Data Fig. 8a). The conformation of the pre-mRNA at the active site and the simultaneous binding of Gpl1 are both mutually exclusive with Prp11 and Cwc24 accommodation (Fig. 6b and Extended Data Fig. 8a,b). The 1585-loop, which helps to maintain catalytic dormancy at the B^{act} state by interacting with the N terminus of Prp11, Cwc24 and the U2/U6 helix Ia[51] (Extended Data Fig. 8c), is also structured in the *sp*B^{d} complex (Fig. 6a and Extended Data Fig. 8d). However, in the *sp*B^{d} complex, it is located in a position similar to that observed in the *sc*C* complex[54] (Fig. 6d and Extended Data Fig. 8e). While in the *sc*C* complex, the 1585-loop stabilizes the active site by interacting with the U6 ISL and the lariat junction (Extended Data Fig. 8e), in the *sp*B^{d} complex, it is sandwiched between Gpl1 residues 38–60 on the one side and residues 204–209 on the other (Fig. 6d). Furthermore, it also interacts with U2–U6 helix Ia, the stacked intronic G_{+1} and Gpl1 F60 (Fig. 6b,d).

Not all pre-mRNAs targeted by the Nrl1-containing CNM complex harbor the quadruple adenine motif at the 3′ end of the 5′ exon; however, *rpl39* serves as an example of an aberrant splice site leading to the recruitment of Gpl1 (and Gih35). Given that the spliceosome is known to adopt different substrate-specific conformations[64], the mechanism of Gpl1 recruitment and its conformation at the active site is likely to be substrate specific and depends on the type of aberration. Likewise, the base pairing of the 5′ exon and 5′ss to U5 loop I and U6 snRNA for different CNM-targeted transcripts and the mRNA loop conformation in between may also differ from our proposed structural model.

https://doi.org/10.1038/s41594-024-01480-7

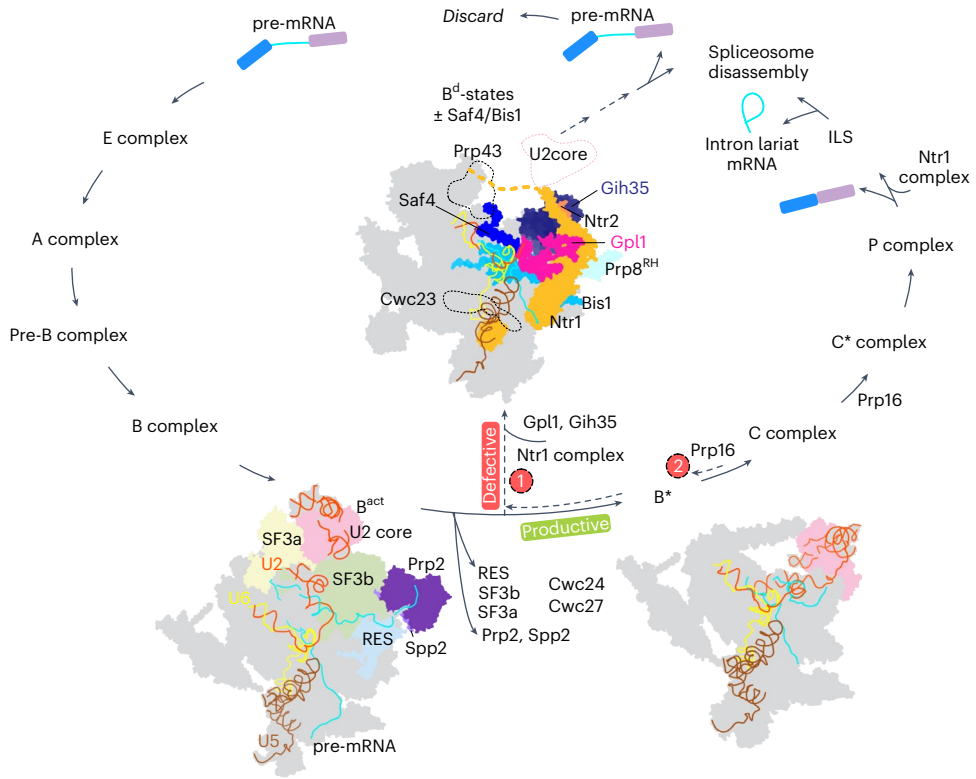


Fig. 7 | Scheme for Gpl1-mediated discard of the aberrant B^d complex. Following the action of the Prp2–Spp2 complex, the RES, SF3a and SF3b complexes and splicing factors Cwc24 and Cwc27 are released. Sensing an aberration, Gpl1, either alone or in complex with Gih35, binds to the active site, blocking splicing progression (model 1). Alternatively, the spliceosome is targeted for disassembly by Prp16-mediated proofreading before branching, triggering Gpl1–Gih35 binding at the B* state (model 2). Concomitantly, the Ntr1 complex is recruited and Saf4 and Bis1 help stabilize the Ntr1 complex in the B^d complex. Subsequently, disassembly of the defective spliceosome can be initiated by Prp43. The two discard models 1 and 2 are marked by dashed lines.

Usually, during the formation of the active site, the U5 loop I pairs with the 5′ exon sequence immediately upstream of the 5′ss[65,66], holding the 5′ exon in place. The presence of four adenines at positions −2 to −5, which pair with U5 loop I, and the ACA sequence within the U6 ACAGA box, which pairs with $U_{+4}G_{+5}U_{+6}$ of the intron[67,68], form strong anchor points, which cause insertion of a single nucleotide at the active site. Taken together, this constraint seems sufficient for formation of an aberrant active site preventing the canonical progression of splicing.

Discussion

While the recent progress in our structural and functional understanding of the splicing machinery is enormous, this information is still lacking for defective spliceosome intermediates committed to premature disassembly. Our structures provide a snapshot of a scenario, where the overstabilization of the U5 loop I interaction with the 5′ exon and the U6 interaction with the intron leads to an extra nucleotide insertion at the active site, rendering the spliceosome defective and committed to discard (Fig. 7). Normally, during transition from the B^{act} to the B* complex in a productive splicing cycle, the U2 snRNP SF3a–SF3b complex, RES complex and splicing factors Cwc24 and Cwc27 dissociate to release the 5′ss and the branch helix as a result of Prp2 action in yeast[64] or the combined action of Prp2 and Cwf11 (Aquarius) in humans[69]. At an intermediate step, before recruitment of branching factors in the B* complex, the branch helix is highly flexible and does not dock the BP adenosine into the active site[64]. The Prp8 1585-loop, which directly interacts with U2–U6 helix I, Cwc24 and Prp11 and protects the G_{+1} nucleotide at the 5′ss in the B^{act} complex[51], is likewise flexible at this intermediate step.

We propose that changes in the conformation of the active site, as observed in this study, might lead to inefficient binding of Prp11 and Cwc24 to the pre-mRNA in the B^{act} complex. Cwc24 is required for both proper interaction of U5 and U6 with the 5′ss and 5′ss selection and stable association of Prp2 with the spliceosome; accordingly, its deletion results in aberrant cleavage at the 5′ss[70,71]. Whether it is the propensity for the altered U5 loop I–5′ exon interaction at positions −5 to −2 of the pre-mRNA (because of the exonic quadruple adenine motif) that leads to a compromised Cwc24 binding capacity or the other way around, in both cases, this overstabilized 'improper' interaction leads to the noncanonical active site conformation observed in the B^d complex. Subsequently, remodeling of the spliceosome mediated by Prp2 and/or Cwf11 (ref. 69) might be faster, creating a time window for recruitment and binding of Gpl1 to the modified active site. Sensing of the 5′ss aberration at this stage involves Gpl1 binding at the active site, either as a preformed complex with Gih35 or with Gih35 being recruited sequentially. Gpl1 binding occludes the docking of the branch helix at the active site and, thus, maintains catalytic dormancy of the spliceosome as found in our B^d complex. The 1585-loop becomes ordered because of its interaction with Gpl1. Taken together, binding of Gpl1 blocks the progression of productive splicing at the B^d intermediate between the B^{act} and B* states.

Alternatively, the spliceosome could reach the B*/C state transiently and then be targeted by proofreading activity of Prp16. In the B*/C complexes, branching is promoted by engagement of the N termini of step I splicing factors Cwc25, Isy1 and Yju2 with the branch helix and branching occurs spontaneously upon Cwc25 binding[53,63,64]. However, removal of Cwc25 by Prp16 before branching reverts the spliceosome back to B* state, which is vulnerable to disassembly[72]. The changes in the 5′ss conformation as observed in the B^d complex could hinder canonical interactions with the branch helix[53,63], likely weakening the association of Cwc25 with the spliceosome, delaying branching and, therefore, promoting Prp16-mediated discard. This alternative scenario is further supported by the presence of factors usually recruited

to the spliceosome at post-B^{act} stages, such as step I factor Yju2 at B* in *S. cerevisiae*[64], ESS2 at C* in humans[44] and DHX35 and GPATCH1 at the C or C* state in humans[44,73]. In this regard, it seems more likely that the spliceosome transiently reaches the B*/C state but delays in branching promote Prp16-mediated discard, recruitment of Gih35–Gpl1 and formation of the B^d complex.

Once Gpl1 is loaded into the active site and folds on the surface of Prp8, Gih35 is positioned right at the beginning of the branch helix. Weak cryo-EM density attributable to 2–3 nucleotides in the RNA-binding channel of Gih35 is observed (Extended Data Fig. 8f). While absence of continuous cryo-EM density into the active site of Gih35 makes identification of its RNA substrate elusive, structural comparisons with *sc*B^{act} and *sc*B* complexes show its close proximity to the branch helix and indicate that pre-mRNA or U2 snRNA are plausible candidates (Extended Data Fig. 8g–j). In any case, because of steric exclusion, the branch helix must be dissolved either to allow or as a consequence of Gih35 binding. As a result of Gih35–Gpl1 binding, catalysis is blocked and the spliceosome is licensed to be disassembled by the Ntr1 complex.

Gih35 and Gpl1 are found in catalytically active spliceosomes in both humans and *Cryptococcus neoformans*[44,73,74]. Studies in *S. pombe* also reported that Gih35 and Gpl1 are required for proper canonical splicing[47,48]. Together, this suggests that Gih35–Gpl1 might be recruited not only to aberrant or defective spliceosomes undergoing discard by the CNM complex as seen in our purifications but also to productive spliceosomes. Indeed, Gih35 and Gpl1 were also detected in purifications from *S. pombe* using an affinity-tagged NTC component (Cef1) while components of the CNM complex were not reported[35]. Taken together, these observations indicate that Gih35–Gpl1 might have other functions apart from their central role in the discard of aberrant spliceosome complexes proposed here.

How would the conformation of Gih35–Gpl1 bound to a catalytically active complex compare to the B^d complex? To answer this question, we performed structural superpositions of our *sp*B^d-II complex with the human C* complex[44]. We found that Gpl1 cannot enter the active site in *hs*C* because of steric clashes with $Prp8^{RH}$, PRKRIP1 and FAM32A binding at the position of the 5′ss $G_{+1}U_{+2}$ nucleotides in the B^d complex (Extended Data Fig. 9a,b). While all crosslinks among GPATCH1, DHX35 and CDC5L observed in the *hs*C* are satisfied in the *sp*B^d complex, those between GPATCH1 and $PRP8^{RH}$ remain unsatisfied because of substantial differences in location of $Prp8^{RH}$/$PRP8^{RH}$ with respect to Gpl1–Gih35 in the *sp*B^d complex (Extended Data Fig. 9c–f). Taken together, the interaction between GPATCH1 and DHX35 remains likely conserved from the *sp*B^d complex to *hs*C*. However, Gpl1 might not be bound to the active site, being associated only at the periphery, where it could help in tethering Gih35 to the spliceosome. Moreover, Gih35–Gpl1 could function during the transition from C to C* complex, where they might associate transiently in a pre-C* complex as proposed before[44]. However, there might be species-specific differences between *S. pombe* and humans; therefore, further biochemical, structural and functional studies need to be performed to ascertain when Gih35 and Gpl1 are recruited and what role they have as part of catalytically active spliceosomes.

Our RNA-seq experiments showed that *rpl39* is an example of a gene prone to discard by the CNM complex. *Rpl39* encodes the ribosomal protein eL39 important for maturation of the nascent polypeptide exit tunnel[75]. Curiously, the first exon of *rpl39* itself encodes a stop codon and, therefore, does not in principle require proper splicing for the production of a functional transcript. It is important to note that *rpl39* mRNA is not always discarded and the majority of the *rpl39* transcripts go through productive splicing. Hence, why is a productive spliceosome sometimes assembled on this substrate as opposed to a defective spliceosome? We propose that, depending on the register of interactions between the 5′ss exon and U5–U6 snRNAs, the erroneous conformation of the active site is an aberrant complex of similar binding energy, which is then discarded through the B^d state route. Recent reports showed that fine-tuning the strength of interactions between the 5′ss and U5 loop I and between the 5′ss and U6 ACAGA box together determine the efficiency of splicing[76,77]. In *S. pombe*, the m^6A modification status of the central adenosine of the ACAGA box, which pairs with the +4 position in the intron, changes the strength of the U6–intron interaction[76]. For *rpl39* mRNA, the U_{+4} could form an energetically favorable Watson–Crick base pair with unmethylated U6 ACAGA[78], which, together with strong base pairing of U5 loop I–5′ exon, could be a detrimental combination compared to a decreased stability of the U6–intron duplex because of the m^6A modification. In addition, *trans*-acting factors that recognize *cis*-regulatory RNA elements to modulate spliceosome recognition of certain splice sites[79] might also contribute to context-specific splicing efficiency of *rpl39* mRNA. The propensity of *rpl39* mRNA to be discarded might actually have a regulatory role itself, with changes in its pre-mRNA splicing efficiency modulating the amount of the protein produced. The influence of varying the amounts of different ribosomal proteins on ribosome biosynthesis is a well-known phenomenon[80].

Taken together, the *sp*B^d complex describes a route for the recognition and discard of aberrant splicing intermediates. This route is part of a splicing quality control mechanism characterized by a shortcut of the splice cycle involving the Gih35–Gpl1 DEAH-box helicase–G-patch protein pair. As splicing errors are an increasing threat of growing intron complexity, more strategies and components of splicing quality control mechanisms await to be deciphered.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-024-01480-7.

References

1. Brody, E. & Abelson, J. The 'spliceosome': yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science* **228**, 963–967 (1985).
2. Frendewey, D. & Keller, W. Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. *Cell* **42**, 355–367 (1985).
3. Grabowski, P. J., Seiler, S. R. & Sharp, P. A. A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell* **42**, 345–353 (1985).
4. Will, C. L. & Luhrmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, a003707 (2011).
5. Shi, Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat. Rev. Mol. Cell Biol.* **18**, 655–670 (2017).
6. Wilkinson, M. E., Charenton, C. & Nagai, K. RNA splicing by the spliceosome. *Annu. Rev. Biochem.* **89**, 359–388 (2020).
7. Cordin, O. & Beggs, J. D. RNA helicases in splicing. *RNA Biol.* **10**, 83–95 (2013).
8. Cordin, O., Hahn, D. & Beggs, J. D. Structure, function and regulation of spliceosomal RNA helicases. *Curr. Opin. Cell Biol.* **24**, 431–438 (2012).
9. Xu, Y. Z. & Query, C. C. Competition between the ATPase Prp5 and branch region–U2 snRNA pairing modulates the fidelity of spliceosome assembly. *Mol. Cell* **28**, 838–849 (2007).
10. Chen, H. C., Tseng, C. K., Tsai, R. T., Chung, C. S. & Cheng, S. C. Link of NTR-mediated spliceosome disassembly with DEAH-box ATPases Prp2, Prp16, and Prp22. *Mol. Cell. Biol.* **33**, 514–525 (2013).
11. Wlodaver, A. M. & Staley, J. P. The DExD/H-box ATPase Prp2p destabilizes and proofreads the catalytic RNA core of the spliceosome. *RNA* **20**, 282–294 (2014).

12. 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).
13. Koodathingal, P., Novak, T., Piccirilli, J. A. & Staley, J. P. The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5′ splice site cleavage during pre-mRNA splicing. *Mol. Cell* **39**, 385–395 (2010).
14. 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).
15. Egecioglu, D. E. & Chanfreau, G. Proofreading and spellchecking: a two-tier strategy for pre-mRNA splicing quality control. *RNA* **17**, 383–389 (2011).
16. Fourmann, J. B. et al. Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. *Genes Dev.* **27**, 413–428 (2013).
17. Tsai, R. T. et al. Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. *Genes Dev.* **19**, 2991–3003 (2005).
18. Martin, A., Schneider, S. & Schwer, B. Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome. *J. Biol. Chem.* **277**, 17743–17750 (2002).
19. Arenas, J. E. & Abelson, J. N. Prp43: an RNA helicase-like factor involved in spliceosome disassembly. *Proc. Natl Acad. Sci. USA* **94**, 11798–11802 (1997).
20. Boon, K. L. et al. Yeast Ntr1/Spp382 mediates Prp43 function in postspliceosomes. *Mol. Cell. Biol.* **26**, 6016–6023 (2006).
21. Warkocki, Z. et al. The G-patch protein Spp2 couples the spliceosome-stimulated ATPase activity of the DEAH-box protein Prp2 to catalytic activation of the spliceosome. *Genes Dev.* **29**, 94–107 (2015).
22. Roy, J., Kim, K., Maddock, J. R., Anthony, J. G. & Woolford, J. L. Jr. The final stages of spliceosome maturation require Spp2p that can interact with the DEAH box protein Prp2p and promote step 1 of splicing. *RNA* **1**, 375–390 (1995).
23. Aravind, L. & Koonin, E. V. G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins. *Trends Biochem. Sci.* **24**, 342–344 (1999).
24. Bohnsack, K. E., Ficner, R., Bohnsack, M. T. & Jonas, S. Regulation of DEAH-box RNA helicases by G-patch proteins. *Biol. Chem.* **402**, 561–579 (2021).
25. Lardelli, R. M., Thompson, J. X., Yates, J. R. 3rd & Stevens, S. W. Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA* **16**, 516–528 (2010).
26. Bao, P., Hobartner, C., Hartmuth, K. & Luhrmann, R. Yeast Prp2 liberates the 5′ splice site and the branch site adenosine for catalysis of pre-mRNA splicing. *RNA* **23**, 1770–1779 (2017).
27. Ohrt, T. et al. Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS. *RNA* **18**, 1244–1256 (2012).
28. Fourmann, J. B. et al. The target of the DEAH-box NTP triphosphatase Prp43 in *Saccharomyces cerevisiae* spliceosomes is the U2 snRNP–intron interaction. *eLife* **5**, e15564 (2016).
29. Pandit, S., Lynn, B. & Rymond, B. C. Inhibition of a spliceosome turnover pathway suppresses splicing defects. *Proc. Natl Acad. Sci. USA* **103**, 13700–13705 (2006).
30. Koodathingal, P. & Staley, J. P. Splicing fidelity: DEAD/H-box ATPases as molecular clocks. *RNA Biol.* **10**, 1073–1079 (2013).
31. Semlow, D. R. & Staley, J. P. Staying on message: ensuring fidelity in pre-mRNA splicing. *Trends Biochem. Sci.* **37**, 263–273 (2012).
32. Mayas, R. M., Maita, H., Semlow, D. R. & Staley, J. P. Spliceosome discards intermediates via the DEAH box ATPase Prp43p. *Proc. Natl Acad. Sci. USA* **107**, 10020–10025 (2010).
33. Tsai, R. T. et al. Dynamic interactions of Ntr1–Ntr2 with Prp43 and with U5 govern the recruitment of Prp43 to mediate spliceosome disassembly. *Mol. Cell. Biol.* **27**, 8027–8037 (2007).
34. Fourmann, J. B., Tauchert, M. J., Ficner, R., Fabrizio, P. & Luhrmann, R. Regulation of Prp43-mediated disassembly of spliceosomes by its cofactors Ntr1 and Ntr2. *Nucleic Acids Res.* **45**, 4068–4080 (2017).
35. Yan, C. et al. Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* **349**, 1182–1191 (2015).
36. Lee, N. N. et al. Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. *Cell* **155**, 1061–1074 (2013).
37. Aronica, L. et al. The spliceosome-associated protein Nrl1 suppresses homologous recombination-dependent R-loop formation in fission yeast. *Nucleic Acids Res.* **44**, 1703–1717 (2016).
38. Zhou, Y. et al. The fission yeast MTREC complex targets CUTs and unspliced pre-mRNAs to the nuclear exosome. *Nat. Commun.* **6**, 7050 (2015).
39. Zhu, J. *Identification and Characterization of the Fission Yeast Exosome Targeting Complexes: The MTREC and CNM Complexes*. PhD thesis, Ruperto-Carola University of Heidelberg (2017).
40. Cipakova, I. et al. Identification of proteins associated with splicing factors Ntr1, Ntr2, Brr2 and Gpl1 in the fission yeast *Schizosaccharomyces pombe*. *Cell Cycle* **18**, 1532–1536 (2019).
41. Ren, L. et al. Systematic two-hybrid and comparative proteomic analyses reveal novel yeast pre-mRNA splicing factors connected to Prp19. *PLoS ONE* **6**, e16719 (2011).
42. Yan, C., Wan, R. & Shi, Y. Molecular mechanisms of pre-mRNA splicing through structural biology of the spliceosome. *Cold Spring Harb. Perspect. Biol.* **11**, 291–323 (2019).
43. De, I. et al. The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes. *Nat. Struct. Mol. Biol.* **22**, 138–144 (2015).
44. Dybkov, O. et al. Regulation of 3′ splice site selection after step 1 of splicing by spliceosomal C* proteins. *Sci. Adv.* **9**, eadf1785 (2023).
45. Zhan, X., Yan, C., Zhang, X., Lei, J. & Shi, Y. Structure of a human catalytic step I spliceosome. *Science* **359**, 537–545 (2018).
46. Wan, R., Yan, C., Bai, R., Lei, J. & Shi, Y. Structure of an intron lariat spliceosome from *Saccharomyces cerevisiae*. *Cell* **171**, 120–132 (2017).
47. Selicky, T. et al. Defining the functional interactome of spliceosome-associated G-patch protein Gpl1 in the fission yeast *Schizosaccharomyces pombe*. *Int. J. Mol. Sci.* **23**, 12800 (2022).
48. Larson, A., Fair, B. J. & Pleiss, J. A. Interconnections between RNA-processing pathways revealed by a sequencing-based genetic screen for pre-mRNA splicing mutants in fission yeast. *G3 (Bethesda)* **6**, 1513–1523 (2016).
49. Plaschka, C., Newman, A. J. & Nagai, K. Structural basis of nuclear pre-mRNA splicing: lessons from yeast. *Cold Spring Harb. Perspect. Biol.* **11**, a032391 (2019).
50. Rauhut, R. et al. Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. *Science* **353**, 1399–1405 (2016).
51. 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).
52. Bai, R. et al. Mechanism of spliceosome remodeling by the ATPase/helicase Prp2 and its coactivator Spp2. *Science* **371**, eabe8863 (2021).
53. Galej, W. P. et al. Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537**, 197–201 (2016).
54. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast step II catalytically activated spliceosome. *Science* **355**, 149–155 (2017).

55. Bai, R., Yan, C., Wan, R., Lei, J. & Shi, Y. Structure of the post-catalytic spliceosome from *Saccharomyces cerevisiae*. *Cell* **171**, 1589–1598 (2017).
56. Fica, S. M. et al. Structure of a spliceosome remodelled for exon ligation. *Nature* **542**, 377–380 (2017).
57. Liu, S. et al. Structure of the yeast spliceosomal postcatalytic P complex. *Science* **358**, 1278–1283 (2017).
58. Wilkinson, M. E. et al. Postcatalytic spliceosome structure reveals mechanism of 3′-splice site selection. *Science* **358**, 1283–1288 (2017).
59. 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).
60. Zhang, X. et al. Structures of the human spliceosomes before and after release of the ligated exon. *Cell Res.* **29**, 274–285 (2019).
61. Studer, M. K., Ivanovic, L., Weber, M. E., Marti, S. & Jonas, S. Structural basis for DEAH-helicase activation by G-patch proteins. *Proc. Natl Acad. Sci. USA* **117**, 7159–7170 (2020).
62. Hamann, F. et al. Structural analysis of the intrinsically disordered splicing factor Spp2 and its binding to the DEAH-box ATPase Prp2. *Proc. Natl Acad. Sci. USA* **117**, 2948–2956 (2020).
63. Wilkinson, M. E., Fica, S. M., Galej, W. P. & Nagai, K. Structural basis for conformational equilibrium of the catalytic spliceosome. *Mol. Cell* **81**, 1439–1452 (2021).
64. 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 (2019).
65. Newman, A. J. & Norman, C. U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. *Cell* **68**, 743–754 (1992).
66. 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).
67. Kandels-Lewis, S. & Seraphin, B. Involvement of U6 snRNA in 5′ splice site selection. *Science* **262**, 2035–2039 (1993).
68. Lesser, C. F. & Guthrie, C. Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science* **262**, 1982–1988 (1993).
69. Schmitzová, J., Cretu, C., Dienemann, C., Urlaub, H. & Pena, V. Structural basis of catalytic activation in human splicing. *Nature* **617**, 842–850 (2023).
70. Wu, N. Y. & Cheng, S. C. Functional analysis of Cwc24 ZF-domain in 5′ splice site selection. *Nucleic Acids Res.* **47**, 10327–10339 (2019).
71. Wu, N. Y., Chung, C. S. & Cheng, S. C. Role of Cwc24 in the first catalytic step of splicing and fidelity of 5′ splice site selection. *Mol. Cell. Biol.* **37**, e00580-16 (2017).
72. Chung, C. S., Wai, H. L., Kao, C. Y. & Cheng, S. C. An ATP-independent role for Prp16 in promoting aberrant splicing. *Nucleic Acids Res.* **51**, 10815–10828 (2023).
73. Ilagan, J. O., Chalkley, R. J., Burlingame, A. L. & Jurica, M. S. Rearrangements within human spliceosomes captured after exon ligation. *RNA* **19**, 400–412 (2013).
74. Sales-Lee, J. et al. Coupling of spliceosome complexity to intron diversity. *Curr. Biol.* **31**, 4898–4910 (2021).
75. Micic, J. et al. Ribosomal protein eL39 is important for maturation of the nascent polypeptide exit tunnel and proper protein folding during translation. *Nucleic Acids Res.* **50**, 6453–6473 (2022).
76. Ishigami, Y., Ohira, T., Isokawa, Y., Suzuki, Y. & Suzuki, T. A single m^6A modification in U6 snRNA diversifies exon sequence at the 5′ splice site. *Nat. Commun.* **12**, 3244 (2021).
77. Parker, M. T. et al. m^6A modification of U6 snRNA modulates usage of two major classes of pre-mRNA 5′ splice site. *eLife* **11**, e78808 (2022).
78. Roost, C. et al. Structure and thermodynamics of N^6-methyladenosine in RNA: a spring-loaded base modification. *J. Am. Chem. Soc.* **137**, 2107–2115 (2015).
79. Coltri, P. P., Dos Santos, M. G. P. & da Silva, G. H. G. Splicing and cancer: challenges and opportunities. *Wiley Interdiscip. Rev. RNA* **10**, e1527 (2019).
80. Petibon, C., Malik Ghulam, M., Catala, M. & Abou Elela, S. Regulation of ribosomal protein genes: an ordered anarchy. *Wiley Interdiscip. Rev. RNA* **12**, e1632 (2021).

Methods

TAP tagging in *S. pombe*

Using the Cre-recombinase-mediated cassette exchange[81] system, the respective genes were tagged at the C-termini with a loxP/M flanked ura4+ cassette, which was further replaced by a Myc-tag fused to a tobacco etch virus (TEV) protease cleavage site and a protein A tag in the case of Nrl1 and 6×Flag-tag in the case of Prp43. The transformants were replica-plated on SDC + FOA, SDC − Ura or SDC − Leu plates and selected for those only growing on SDC + FOA plates. Finally, the genomic integration of the TAP-tagged Nrl1 and Prp43 was confirmed by PCR and western blot[39].

Split-tag TAP and RNA-IP from *S. pombe*

The split-tagged RNP complex was isolated from ~10–12-L yeast extract agar (YEA) cultures of the yeast strain grown to an optical density at 600 nm of 1.8–2.2. Cell pellets were snap-frozen in liquid nitrogen and subsequently mechanically lysed by a cryogenic cell mill (Retsch MM400). The cell lysate was resuspended in lysis buffer containing 20 mM HEPES pH 7.5, 100 mM NaCl and 1.5 mM $MgCl_2$ supplemented with 0.05% NP-40 (Sigma-Aldrich), 1 mM DTT, 1 mM phenylmethylsulfonyl fluoride, protease inhibitor cocktail (EDTA-free, Roche) and RiboLock RNase inhibitor (Thermo Fisher Scientific). The cell lysate was cleared by centrifugation and the supernatant was loaded onto IgG Sepharose 6 fast flow beads (Cytiva) for overnight incubation at 4 °C. Nonspecifically bound proteins were removed by washing with lysis buffer supplemented with 0.01% NP-40, followed by TEV cleavage for 5 h at 4 °C. For the second affinity purification, the eluate from IgG beads was loaded onto Flag beads (anti-Flag M2 affinity gel, Sigma-Aldrich) for overnight incubation at 4 °C. The beads were washed and eluted with buffer containing 3×Flag peptide at a final concentration of 0.4 mg ml^{-1}. The elution buffer for cryo-EM sample preparation contained 20 mM HEPES pH 7.5, 100 mM NaCl, 1.5 mM $MgCl_2$, 0.01 % NP-40 and 1 mM TCEP. In RNA-IP experiments, the RNA was isolated from the Flag eluate using an RNA clean and concentrator kit (Zymo Research), treated with 20 U of Turbo DNase (Thermo Fisher Scientific) at 37 °C for 30 min. The RNA clean and concentration kit was used once again to remove DNase. RNA-seq libraries were prepared using the NEBNext Ultra II directional RNA library prep kit for Illumina (New England Biolabs) following the manufacturer's instructions.

EM sample preparation and data acquisition

Negative staining was performed as a quality control before cryo-EM grid preparation and after crosslinking with BS3 for MS analysis to control for sample aggregation. For cryo-EM grid preparation, a 5-µl aliquot of the sample was applied to Quantifoil R 2/1 grids coated with a thin layer of homemade carbon film, glow-discharged for 45 s using PELCO easiGlow. After ~8–10-min incubation of the sample on grids, blotting was performed for 5 s using a blot force of 0 at 100% humidity using Vitrobot Mark IV (FEI) operated at 4 °C and the grids were immediately plunge-frozen in liquid ethane cooled with liquid nitrogen. Cryo-EM data were acquired at the cryo-EM platform of the European Molecular Biology Laboratory (EMBL) in Heidelberg on a 300-kV FEI Titan Krios EM instrument equipped with a K3 detector (Gatan) in counting mode at 0.822 Å per pixel with 1.235 e^- per pixel per frame over 40 frames and 1.491 s of total exposure. In total, 13,096 images were collected.

Image processing

The images were processed using the RELION 3.1 software package[82]. Video stacks were motion-corrected using MotionCor2 with 5 × 5 as the number of patches[83] and estimation of contrast transfer function (CTF) was performed with Gctf (version 1.06)[84] on the motion-corrected micrographs. After manual curation of particles picked with Warp (version 1.0.9)[85] on ten micrographs, 929,930 particles were automatically picked from all micrographs. Motion correction and CTF estimation were simultaneously performed in cryoSPARC (version 3.1)[86,87]. Using particle coordinates from Warp, particles were extracted in cryoSPARC and subjected to two-dimensional classification, initial ab initio three-dimensional (3D) construction and heterogenous refinement, which was subsequently low-pass-filtered to 60 Å and used as the initial model for 3D classification of all particles (929,930) in RELION. The two best classes were merged (256,572 particles) and subjected to 3D autorefinement. These particles were then imported in cisTEM (version 1.0.0)[88], where they were first subjected to 3D autorefinement (six cycles) and subsequently to focused classification using manual refinement (three classes, 40 cycles) by applying a soft circular mask around Gih35. This yielded a single class average with cryo-EM density for Gih35 (153,834 particles, 59.2%), which was further subjected to another round of focused classification with a soft mask around the Ntr1 complex (three classes, 40 cycles). Two good classes were obtained, of which one showed a decreased signal for Ntr1 (*sp*B^d-I; 40.1%, 61,423 particles) compared to the other (*sp*B^d-II; 47.1%, 72,631 particles). These particle stacks were then imported into RELION and subjected to 3D autorefinement, CTF refinement and Bayesian polishing to give final reconstructions with overall resolutions of 3.2 Å (*sp*B^d-I) and 3.1 Å (*sp*B^d-II). The cryo-EM maps were sharpened in PHENIX (version 1.19.2)[89] using the autosharpen map tool and local resolution estimation in was performed in RELION.

For the *sp*B^d-II state, further rounds of 3D classification without image alignment with separate masks around the $Prp8^{RH}$, Ntr1 complex and Cwf11 were performed, which invariably led to better local reconstructions in the flexible regions. After 3D autorefinement of the subset of particles for $Prp8^{RH}$ (20,173), Ntr1 complex (21,292) and Cwf11 (9,361), the cryo-EM maps were low-pass-filtered to 7, 8 and 10 Å, respectively, to obtain more continuous cryo-EM density in these regions. For Gih35, another round of 3D autorefinement was performed in RELION, which served as the basis for multibody refinement. For this, the bulk of the spliceosome was treated as one component and the region encompassing $Prp8^{En}$, $Prp8^{RH}$ and Gih35 was treated as the second component, moving independently of each other. The final cryo-EM densities after multibody refinement were subjected to post-processing including automatic *B*-factor sharpening and local resolution estimation in RELION. The particle stack for *sp*B^d-II was imported in cryoSPARC, where local refinement was performed with a mask around the NTC, leading to an improved resolution of 4.3 Å in this region. For the Ntr1 CTD, focused 3D classification (in cryoSPARC) with masks around Snu114 and Ntr1 CTD followed by local refinement led to a local resolution of ~6-8 Å in the *sp*B^d-I (6,500 particles) and *sp*B^d-II (8,560 particles) states. A detailed summary of the image-processing workflow is illustrated in Extended Data Fig. 1e and Supplementary Fig. 1.

Structural modelling, refinement and analysis

We used a combination of de novo model building, rigid-body docking of known structures, homology models obtained using SWISS-MODEL[90] and models deposited in the AlphaFold protein structure database[91] for different components of the *sp*B^d complexes depending on the local resolution. Identification and docking of different components of the *sp*B^d complexes were facilitated by the *sp*ILS complex (Protein Data Bank (PDB) 3JB9)[35], *sc*ILS complex (PDB 5Y88)[46], *sc*C complex (PDB 5LJ3)[53], *sc*B* complex (PDB 6J6Q)[64] and *hs*C* complex (PDB 5MQF)[92]. Identification of Saf4 and Bis1 in the *sp*B^d-II cryo-EM map was performed using ModelAngelo[93]. The structures of the individual components were first fitted into cryo-EM density using UCSF Chimera (version 1.12)[94] and then manually adjusted in Coot (version 0.8.9.3-pre)[95].

Gpl1 was built de novo in the cryo-EM map. The Gpl1 fragment contacting Gih35 suffers from limited local resolution. Here, high-resolution structures of other DEAH-helicase–G-patch protein pairs (PDB 6HS6 (ref. 61) and PDB 7DCP (ref. 52)), together with Clustal Omega[96]-based sequence alignments, visualization with ESPript (version 3.0)[97] and secondary-structure prediction using PSIPRED[98],

aided in model building and identification of the sequence register. For Gih35, the AlphaFold model was rigid-body docked in the cryo-EM map in Chimera and then individual domains were manually adjusted to best fit the cryo-EM density in Coot. All components were first built in the *sp*B^{d}-II cryo-EM map and were then adjusted for differences in the *sp*B^{d}-I cryo-EM map in Coot. The pre-mRNA sequence was built according to the consensus sequence logo from CNM-sensitive genes from positions −5 to +6 and uracil or adenine bases were placed where a pyrimidine or purine satisfied the cryo-EM map better.

While the models of individual components of the *sp*B^{d} complex were built in and refined against focused maps (only for *sp*B^{d}-II), the final model was real-space refined against the overall map in PHENIX (version 1.19.2)[89] with secondary-structure and geometry restraints. The model qualities were assessed using MolProbity[99] within PHENIX. A detailed summary of model building for *sp*B^{d}-I and *sp*B^{d}-II complexes is presented in Extended Data Table 1. Structures and cryo-EM maps were visualized with PyMOL (version 2.5.3), UCSF Chimera (version 1.12) and UCSF ChimeraX (version 1.2.5)[100].

MS analysis of the *sp*B^{d} complex

Approximately 5 pmol of the *sp*B^{d} complex was purified from ~24-L YEA culture as described above, crosslinked with 1 mM BS3 (Thermo Fisher Scientific) for 30 min at 30 °C, subsequently quenched with 50 mM ammonium bicarbonate and pelleted by ultracentrifugation. The pelleted complex was solubilized in 50 mM ammonium bicarbonate (pH 8.0) supplemented with 4 M urea, reduced with dithiothreitol and alkylated with iodoacetamide. After dilution to 1 M urea with 50 mM ammonium bicarbonate, crosslinked complexes were digested with trypsin (Promega) in a 1:20 enzyme-to-protein ratio (w/w) at 37 °C overnight. Peptides were reverse-phase extracted using SepPak Vac tC18 1 cc/50 mg (Waters) and eluted with 50% acetonitrile and 0.1 % TFA. The eluate was lyophilized. Dried peptides were dissolved in 40 µl 2% acetonitrile and 20 mM ammonium hydroxide and reverse-phase fractionated at basic pH using a Vanquish high-performance liquid chromatography (HPLC) system (Thermo Fisher Scientific) with an xBridge C18 column (3.5 µm, 1 × 150 mm; Waters) applying a 4–36% acetonitrile gradient over 45 min at a flow rate of 60 µl min^{-1}. Then, 1-min fractions of 60 µl were collected, pooled in a step of 13 min (resulting in 13 pooled fractions total), vacuum-dried and dissolved in 5% acetonitrile and 0.1% TFA for subsequent ultra (u)HPLC–electrospray ionization MS/MS analysis.

Fractionated peptides were measured in triplicate on an Orbitrap Exploris 480 (Thermo Fisher Scientific). The MS instrument was coupled to a Dionex UltiMate 3000 uHPLC system (Thermo Fisher Scientific) with a custom 35-cm C18 column (75-µm inner diameter packed with ReproSil-Pur 120 C18-AQ beads, 3-µm pore size; Dr. Maisch). MS1 and MS2 resolutions were set to 120,000 and 30,000, respectively. Only precursors with a charge state of 3–8 were selected for MS2. Protein composition was determined by MaxQuant (version 1.6.17.0). The entire *S. pombe* proteome or a set of 45 abundant spliceosomal proteins identified by MaxQuant were used as a database to search for crosslinked peptides. Protein–protein crosslinks were searched using pLink (version 2.3.9; http://pfind.org/software/pLink/) according to the recommendations of the developer[101]. A list of the top 200 proteins identified in the sample, ranked according to label-free quantitation (LFQ) intensity, as well as interprotein and intraprotein crosslinks of these proteins, is provided in Supplementary Tables 1–3, respectively.

RNA-seq data analysis and assessing intron retention in the *ctr1*Δ mutant and RNA-IP experiments

Paired-end RNA-seq reads were aligned to the *S. pombe* reference genome (GCF_000002945.1) using HISAT (version 2.2.1)[102] allowing a maximum intron length of 2,000 bp. Alignment files were sorted and indexed using SAMtools (version 1.21)[103]. Unnormalized bigwig files were generated with bamCoverage from deepTools (version 3.5.5)[104] with 1-bp resolution. To avoid any bias from ribosomal RNAs transcribed on chromosome III, raw signals were normalized by the sum of the coverage of chromosomes I and II using in-house bash and R scripts. Intron annotation was adopted from a previous study[105] and only introns longer than 10 bp were used. For RNA expression analysis, introns with normalized scores ≥ 30 in both biological replicates were included in downstream analyses. The sensitivity score in the *ctr1Δ* mutant was calculated by subtracting the normalized RNA-seq signal over the introns of the WT strain from the *ctr1Δ* strain, divided by the averaged signal of the neighboring exons. The 5′ss and 3′ss with upstream and downstream flanking regions of 9 and 10 nt, respectively, were inferred from the reference genome using the seqinr[106] and bedr[107] R packages. The occurrences of 5′ss and 3′ss sequences were then scaled to be proportional to the average signal of the upstream exon measured in the WT strain. Introns were considered as sensitive with at least a 20% increase in abundance in the *ctr1Δ* strains compared to WT. The probability weight matrices of the 5′ss and 3′ss of the sensitive introns were plotted as motif logos using the gglogo R package. Lists of all CNM-sensitive and CNM-insensitive sites are provided in Supplementary Tables 4 and 5, respectively. Absolute intron coverage in RNA-IP experiments was determined by the average sequence coverage at the 5′ end (5% of the total intron length) of all annotated introns. To exclude alternatively spliced and misannotated introns, introns with less than 25% spliced reads and introns with less than 50% coverage at their 3′ ends, compared to the coverage at their 5′ ends, were excluded from absolute intron coverage analysis. Intron enrichments were calculated in the *sp*B^{d} RNA-IP sample over the no-tag RNA-IP control. Enriched introns in the RNA-IP were defined by minimum 1.5× enrichment compared to background.

Cloning, protein expression and in vitro complex assembly

The DNA sequences encoding Prp43 (residues 1–735), Ntr1 (residues 107–170) and Gpl1 (residues 109–181) were cloned and ligated into a modified pET24d vector containing an N-terminal His–thioredoxin tag upstream of a TEV cleavage site before the corresponding protein. The plasmids were transformed in BL21(DE3) Rosetta electrocompetent cells, grown at 37 °C up to an optical density of 1.4–1.6 in autoinduction medium[108] and subsequently expressed at 20 °C for ~16 h. Cells were lysed in buffer containing 20 mM Tris pH 7.5, 500 mM NaCl, 1 M urea, 20 mM imidazole and 2 mM β-mercaptoethanol. The proteins were purified over a 2-ml His-Trap FF column (GE Healthcare) with elution buffer containing 250 mM imidazole. Next, overnight tag cleavage using TEV protease and simultaneous dialysis of the proteins into buffer containing 20 mM HEPES pH 7.5, 150 mM NaCl and 2 mM β-mercaptoethanol was carried out. The proteins were further purified over His-Trap FF columns to remove uncleaved proteins and fusion tags from the cleaved proteins and the flowthrough samples were collected. Finally, the proteins were polished using gel filtration (Superdex 200 16/60) columns (GE Healthcare) equilibrated with size-exclusion chromatography buffer containing 20 mM HEPES pH 7.5, 150 mM NaCl and 1 mM DTT. For monitoring complex formation, Ntr1 and Gpl1 cofactors were added in twofold molar excess over Prp43, incubated at 4 °C for 30 min and purified using gel filtration (Superdex 200 Increase 10/300 GL) columns (GE Healthcare).

Y2H assays

The respective sequences of full-length proteins were cloned into low-copy plasmids pG4BDN22 (Gpl1 and Ntr1) and pG4ADAHAN111 (Gih35 and Prp43) that harbor the DNA-binding domain and activation domain at the N termini, respectively. The interaction pairs were analyzed by cotransformation into the PJ69-4A strain. After a tenfold serial dilution, colonies were spotted on SDC (SDC–Leu–Trp), SDC–His (SDC–Leu–Trp–His) and SDC–Ade (SDC–Leu–Trp–Ade) plates, incubated at 30 °C and analyzed after 3 days. The strength of the interaction

was assessed by growth achieved on SDC−His and SDC−Ade as weak and strong, respectively.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The structures of the *sp*B^{d}-I and *sp*B^{d}-II states were deposited to the PDB under accession codes 9ESH and 9ESI and cryo-EM maps are available from the EM Data Bank under accession codes EMD-19941 and EMD-19942, respectively. The RNA-seq data were deposited to the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus under accession code GSE235589. The *S. pombe* reference genome (GCF_000002945.1) is accessible from NCBI. Source data are provided with this paper.

Code availability

All in-house R and Bash scripts are available from GitHub (https://github.com/ahorvath/Soni-et-al.-2023.git).

References

81. Watson, A. T., Garcia, V., Bone, N., Carr, A. M. & Armstrong, J. Gene tagging and gene replacement using recombinase-mediated cassette exchange in *Schizosaccharomyces pombe*. *Gene* **407**, 63–74 (2008).
82. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
83. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
84. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
85. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019).
86. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
87. Punjani, A., Brubaker, M. A. & Fleet, D. J. Building proteins in a day: efficient 3D molecular structure estimation with electron cryomicroscopy. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 706–718 (2017).
88. Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *eLife* **7**, e35383 (2018).
89. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in PHENIX. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 (2019).
90. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
91. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
92. Bertram, K. et al. Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* **542**, 318–323 (2017).
93. Jamali, K. et al. Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450–457 (2024).
94. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
95. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
96. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
97. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
98. Buchan, D. W. A. & Jones, D. T. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 (2019).
99. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
100. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
101. Chen, Z. L. et al. A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat. Commun.* **10**, 3404 (2019).
102. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
103. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
104. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
105. Burke, J. E. et al. Spliceosome profiling visualizes operations of a dynamic RNP at nucleotide resolution. *Cell* **173**, 1014–1030 (2018).
106. Charif, D. & Lobry, J. R. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations* (eds Bastolla, U., Porto, M., Eduardo Roman, H. & Vendruscolo, M.) (2007).
107. Haider, S. et al. A bedr way of genomic interval processing. *Source Code Biol. Med.* **11**, 1–7 (2016).
108. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
109. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).

Acknowledgements

We are thankful to EMBL Heidelberg for seamless cryo-EM data collection through the iNEXT Discovery program, to HDCryoNet for access to our inhouse cryo-EM setup for screening and to N. Dobrev and M. McDowell for their inputs on the project. We are grateful to J. Hennig for supporting completion of final experiments by K.S. in his lab. We acknowledge the data storage service SDS@hd and bwHPC supported by the Ministry of Science, Research and the Arts Baden-Württemberg and the German Research Foundation (DFG) through grants INST 35/1314-1 FUGG, INST 35/1503- 1 FUGG and INST 35/1597-1 FUGG. This work was supported by the Australian Research Council's Discovery Projects (DP190100423 and DP240102611) to T.F and A.H., the DFG through the Leibniz program (SI 586/6-1) and TRR 319 (project 439669440, TP B03) to I.S. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

K.S. purified the *sp*B^{d} complex, prepared the cryo-EM samples, collected and processed the EM data and performed the in vitro protein expression and complex assembly. O.D. performed the proteome analysis and the crosslinking MS. O.D. and H.U. interpreted the MS results. M.S. performed the Y2H experiments. S.T. prepared the samples for RNA-seq. A.H. and T.F. performed the RNA-seq analysis. D.F. performed the negative staining and helped with the cryo-EM grid screening. K.S. and K.W. built the structural models. K.S., K.W., T.F. and I.S. interpreted the data. K.S., K.W. and I.S. wrote the manuscript. K.S., T.F. and I.S. planned the study and designed the experiments. All authors contributed to the final version of this manuscript.

Competing interests
The authors declare no competing interests.

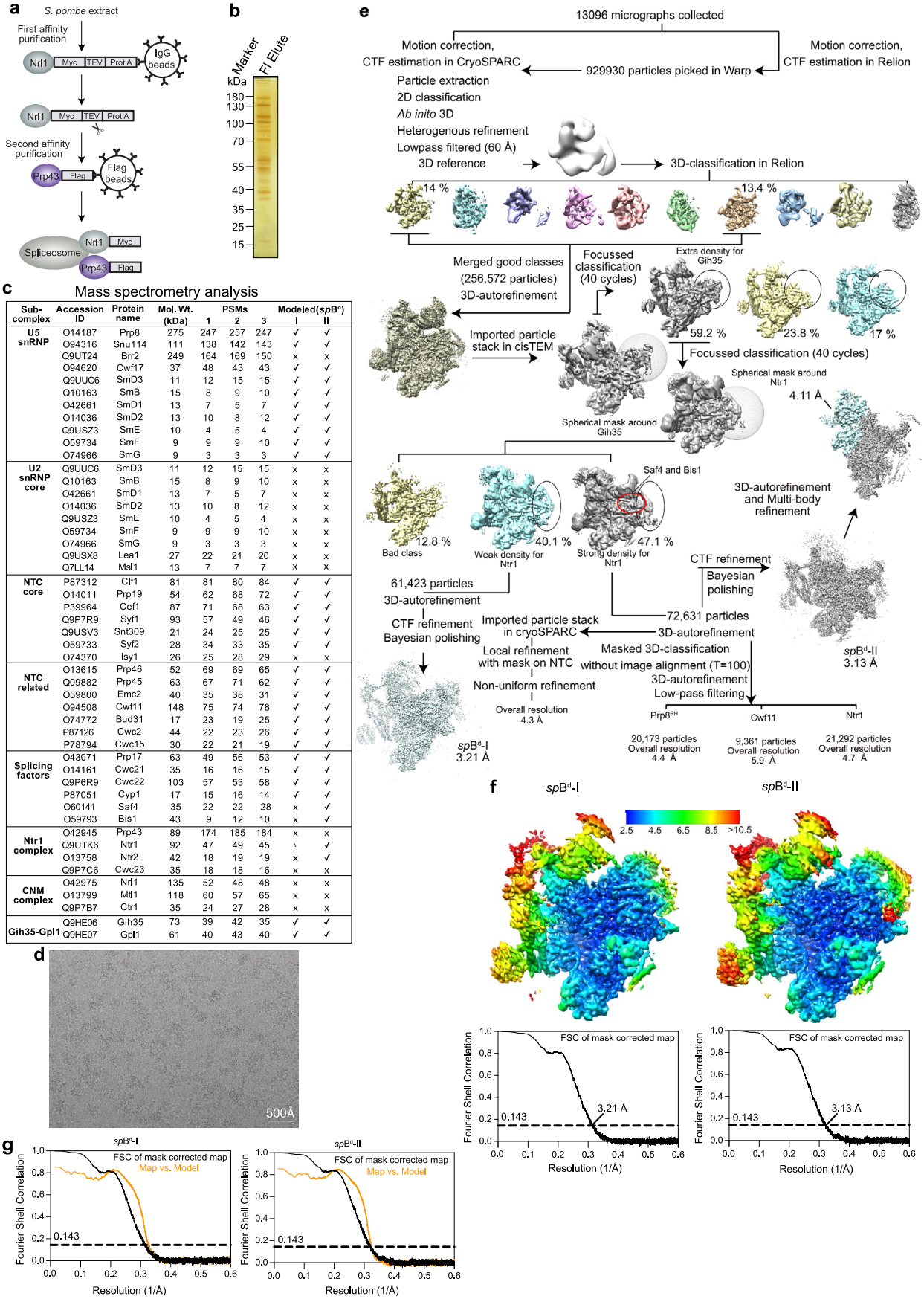
Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41594-024-01480-7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41594-024-01480-7.

Correspondence and requests for materials should be addressed to Komal Soni, Tamás Fischer or Irmgard Sinning.

Peer review information *Nature Structural & Molecular Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Dimitris Typas, in collaboration with the *Nature Structural & Molecular Biology* team.

https://doi.org/10.1038/s41594-024-01480-7



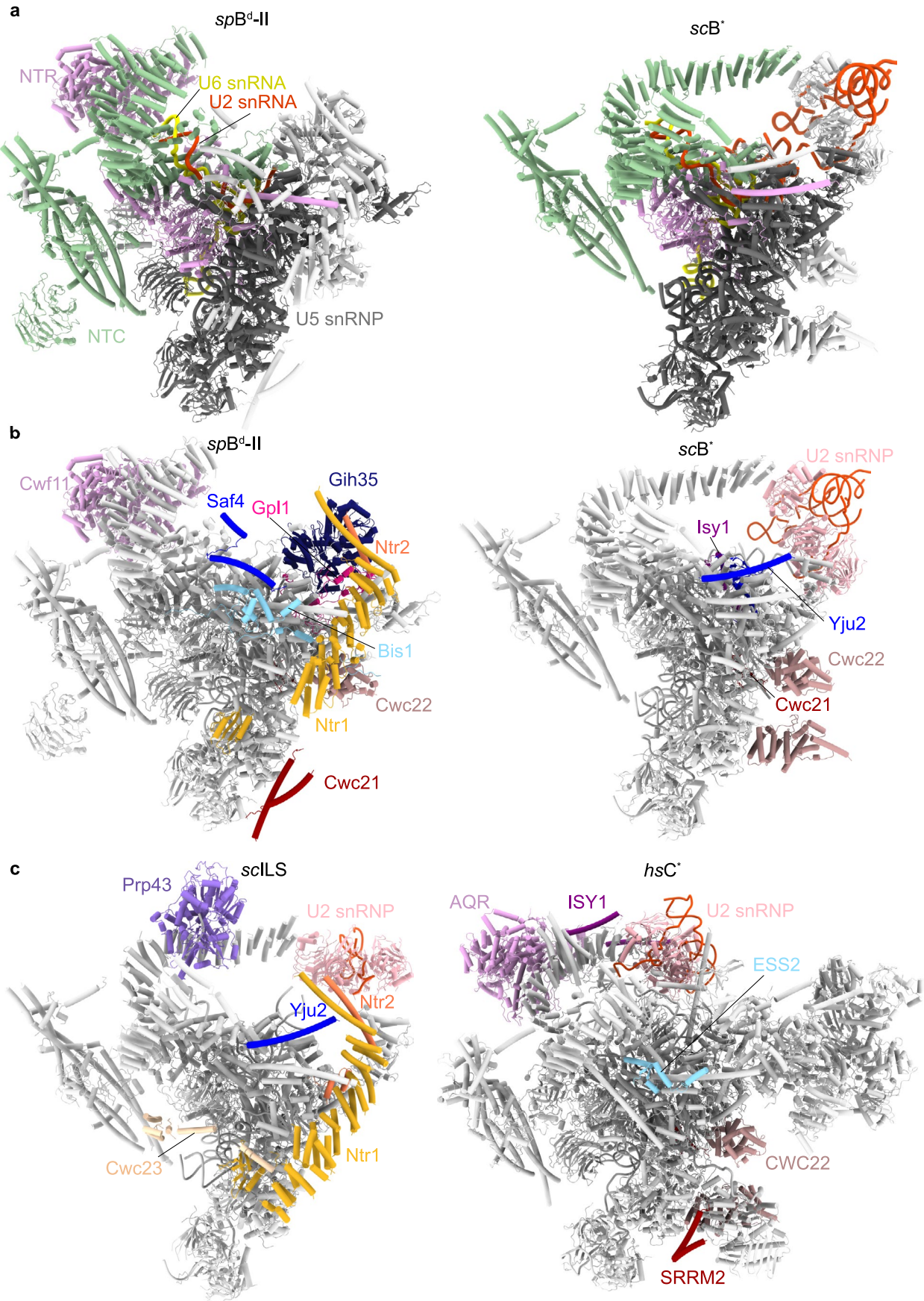
Sub-complex	Accession ID	Protein name	Mol. Wt. (kDa)	PSMs 1	2	3	Modeled(spBd) I	II
U5 snRNP	O14187	Prp8	275	247	257	247	✓	✓
	O94316	Snu114	111	138	142	143	✓	✓
	Q9UT24	Brr2	249	164	169	150	x	x
	O94620	Cwf17	37	48	43	43	✓	✓
	Q9UUC6	SmD3	11	12	15	15	✓	✓
	Q10163	SmB	15	8	9	10	✓	✓
	O42661	SmD1	13	7	5	7	✓	✓
	O14036	SmD2	13	10	8	12	✓	✓
	Q9USZ3	SmE	10	4	5	4	✓	✓
	O59734	SmF	9	9	9	10	✓	✓
	O74966	SmG	9	3	3	3	✓	✓
U2 snRNP core	Q9UUC6	SmD3	11	12	15	15	x	x
	Q10163	SmB	15	8	9	10	x	x
	O42661	SmD1	13	7	5	7	x	x
	O14036	SmD2	13	10	8	12	x	x
	Q9USZ3	SmE	10	4	5	4	x	x
	O59734	SmF	9	9	9	10	x	x
	O74966	SmG	9	3	3	3	x	x
	Q9USX8	Lea1	27	22	21	20	x	x
	Q7LL14	Msl1	13	7	7	7	x	x
NTC core	P87312	Clf1	81	81	80	84	✓	✓
	O14011	Prp19	54	62	68	72	✓	✓
	P39964	Cef1	87	71	68	63	✓	✓
	Q9P7R9	Syf1	93	57	49	46	✓	✓
	Q9USV3	Snt309	21	24	25	25	✓	✓
	O59733	Syf2	28	34	33	35	✓	✓
	O74370	Isy1	26	25	28	29	x	x
NTC related	O13615	Prp46	52	69	69	65	✓	✓
	Q09882	Prp45	63	67	71	62	✓	✓
	O59800	Emc2	40	35	38	31	✓	✓
	O94508	Cwf11	148	75	74	78	✓	✓
	O74772	Bud31	17	23	19	25	✓	✓
	P87126	Cwc2	44	22	23	26	✓	✓
	P78794	Cwc15	30	22	21	19	✓	✓
Splicing factors	O43071	Prp17	63	49	56	53	✓	✓
	O14161	Cwc21	35	16	16	15	✓	✓
	Q9P6R9	Cwc22	103	57	53	58	✓	✓
	P87051	Cyp1	17	15	16	14	✓	✓
	O60141	Saf4	35	22	22	28	x	✓
	O59793	Bis1	43	9	12	10	x	✓
Ntr1 complex	O42945	Prp43	89	174	185	184	x	x
	Q9UTK6	Ntr1	92	47	49	45	*	✓
	O13758	Ntr2	42	18	19	19	x	✓
	Q9P7C6	Cwc23	35	18	18	16	x	x
CNM complex	O42975	Nrl1	135	52	48	48	x	x
	O13799	Mtl1	118	60	57	65	x	x
	Q9P7B7	Ctr1	35	24	27	28	x	x
Gih35-Gpl1	Q9HE06	Gih35	73	39	42	35	✓	✓
	Q9HE07	Gpl1	61	40	43	40	✓	✓

Extended Data Fig. 1 | See next page for caption.

https://doi.org/10.1038/s41594-024-01480-7

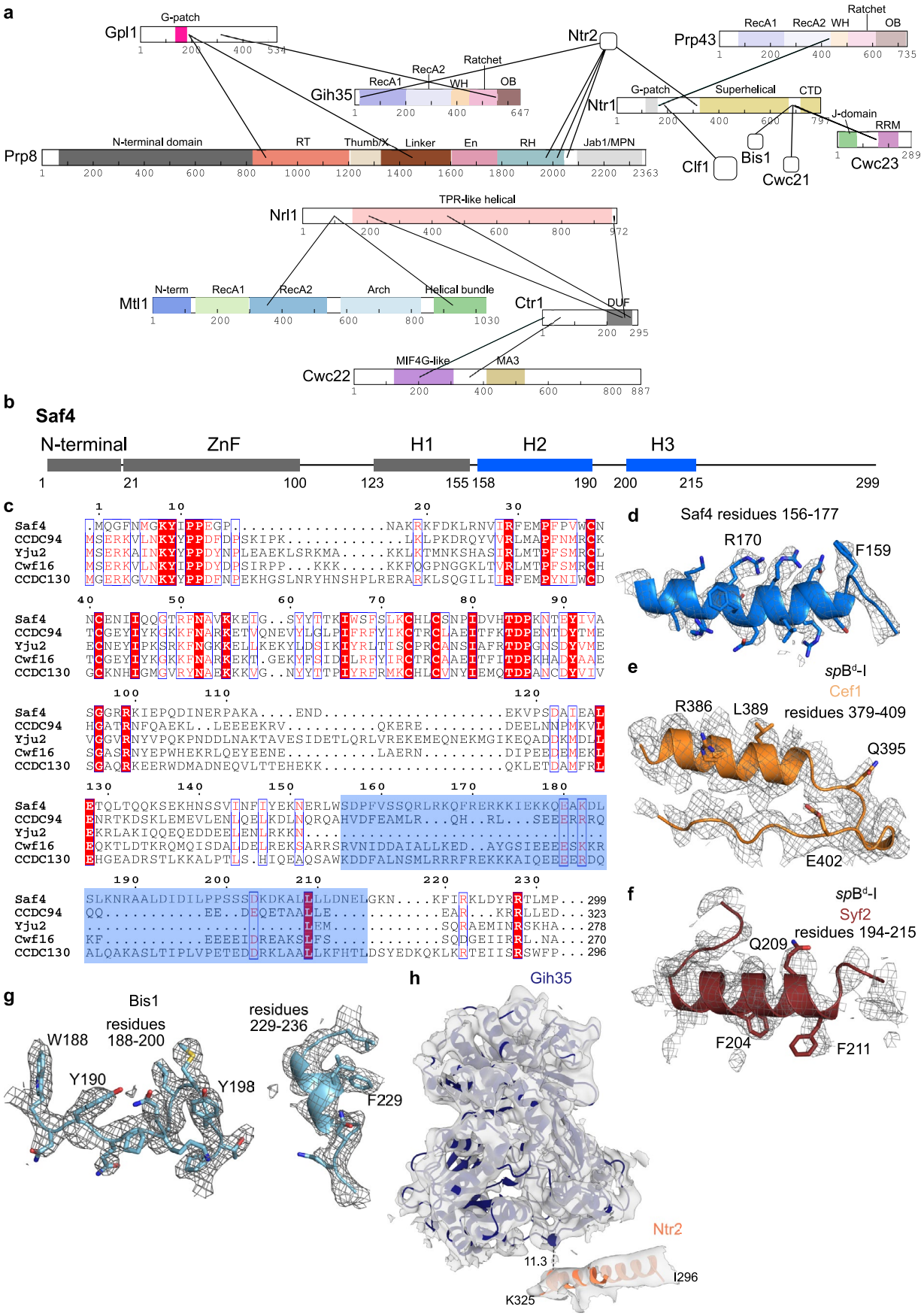
Extended Data Fig. 1 | *sp*B^{d} complex purification, characterization and cryo-EM processing. a, Purification scheme for the *sp*B^{d} complex using a split-tag approach. **b**, After elution of the *sp*B^{d} complex from the anti-FLAG beads, the sample was analyzed using SDS-PAGE and visualized by silver staining. A representative gel from two independent experiments is shown. **c**, The purified *sp*B^{d} complex was crosslinked with BS3 and subjected to mass spectrometric analysis. A list of the components of the U5 snRNP, U2 snRNP core, NTC core, NTC related proteins, other splicing factors, Ntr1 complex, CNM complex and the associated Gih35-Gpl1 proteins identified in the *sp*B^{d} complex is provided. The PSM (Peptide-spectrum match) values obtained from three technical replicates are indicated. A list of the top 200 proteins identified in the sample, ranked according to label-free quantitation (LFQ) intensity, are provided in Supplementary Table 1. The proteome analysis of the *sp*B^{d} complex was also performed on non-crosslinked samples which yielded a similar protein composition as shown here. Proteins modelled in the *sp*B^{d}-I and II complexes are marked. * represents that only the C-terminal domain of Ntr1 was modelled in the *sp*B^{d}-I state. **d**, A representative cryo-EM micrograph from a total of 13,096 is shown. Scale bar represents 500 Å. **e**, Cryo-EM processing pipeline for the *sp*B^{d} complex. **f**, Local resolution and Fourier shell correlation curves (FSC) (0.143 cut-off) along with resolutions are reported for the consensus refinements of *sp*B^{d}-I and *sp*B^{d}-II states. **g**, A superposition of the FSC curves of the consensus cryo-EM map (black) and the atomic model (orange) is shown for *sp*B^{d}-I and *sp*B^{d}-II states.

https://doi.org/10.1038/s41594-024-01480-7



Extended Data Fig. 2 | Comparison of *sp*B^{d}-II state with other spliceosome complexes. a, Comparison of overall structures of the *sp*B^{d}-II state and *sc*B* (PDB:6J6Q) shows the stable core of the spliceosome that remains largely unchanged. **b**, Comparison of overall structures of the *sp*B^{d}-II state and *sc*B* (PDB:6J6Q). **c**, Comparison of overall structures of the *sc*ILS (PDB:5Y88) and *hs*C* (PDB:8C6J) complexes. Components differing in the four states in panels b and c are shown in color.

https://doi.org/10.1038/s41594-024-01480-7

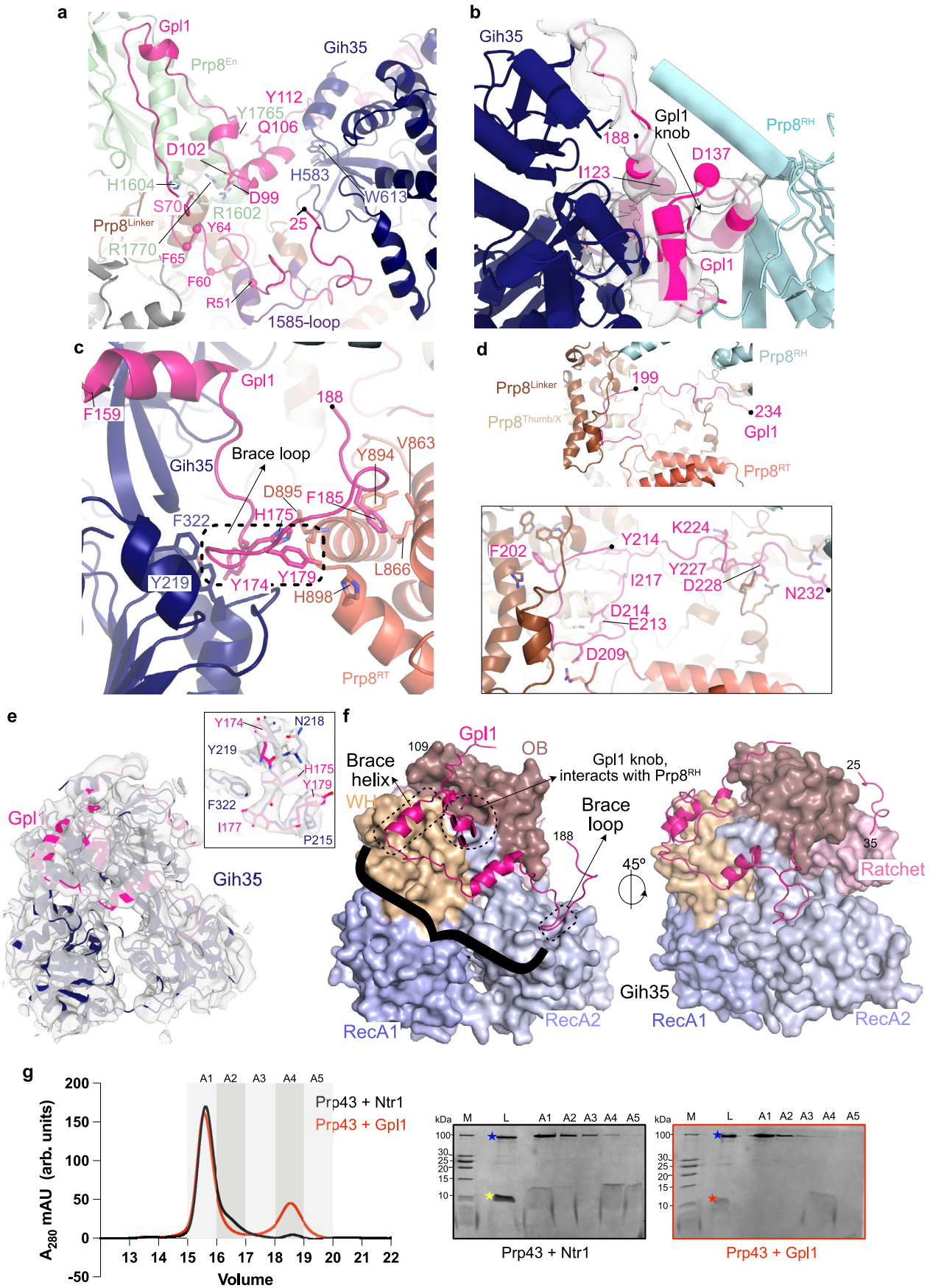


Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s41594-024-01480-7

Extended Data Fig. 3 | Inter-protein crosslinks in the *sp*B^{d} complex and differences between the two *sp*B^{d} states. a, Schematic representation of a subset of inter-protein crosslinks originating from Gpl1, Gih35, Ntr1 complex (Ntr1, Ntr2, Prp43, Cwc23) and the CNM complex (Ctr1, Nrl1, Mtl1) obtained using the entire *S. pombe* proteome as a search database. See Supplementary Tables 2 and 3 for a complete list of all inter-and intra-protein crosslinks identified, respectively. Domain annotations in the proteins are shown. **b**, Domain architecture of Saf4. **c**, Homologs of *sc*Yju2. Saf4 and Cwf16 in *S. pombe*, CCDC94 and CCDC130 in humans share sequence homology with *sc*Yju2. The two α-helices of Saf4 visible in the *sp*B^{d}-II state are highlighted in blue. **d**, Superposition of Saf4 residues 156-177 with the *sp*B^{d}-II cryo-EM map. **e**, Superposition of Cef1 residues 379-409 with the *sp*B^{d}-I cryo-EM map. **f**, Superposition of Syf2 residues 194-215 with the *sp*B^{d}-I cryo-EM map. **g**, Superposition of Bis1 residues 188-200 and 229-236 with the *sp*B^{d}-II cryo-EM map is provided. **h**, Superimposition of Gih35 and Ntr2 with their respective focused cryo-EM maps. The distance (11.3 Å) between Cα atoms of closest residues Pro 206 (Gih35) and Ile 323 (Ntr2) is shown with a dotted line.

https://doi.org/10.1038/s41594-024-01480-7

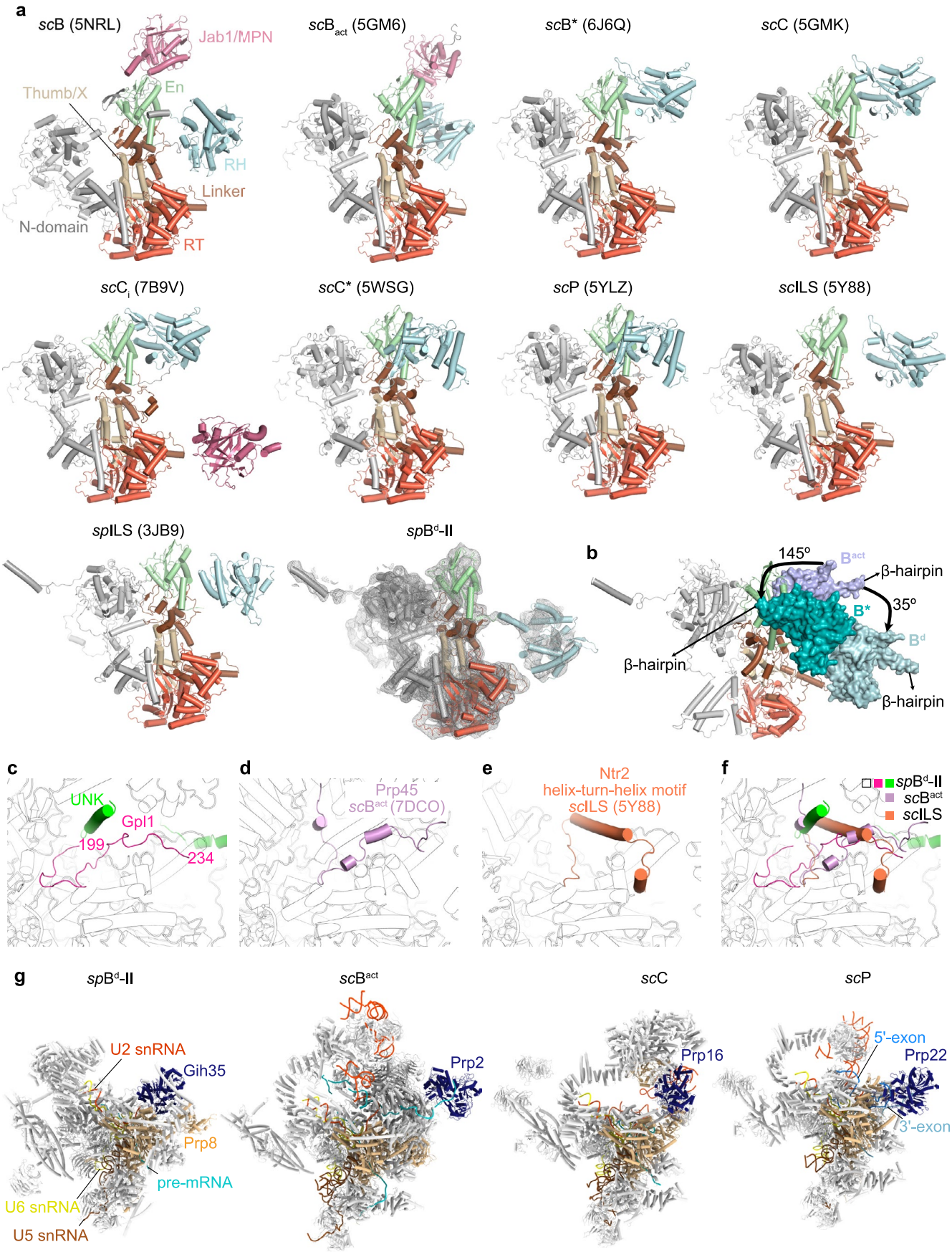


Extended Data Fig. 4 | See next page for caption.

https://doi.org/10.1038/s41594-024-01480-7

Extended Data Fig. 4 | Gpl1 interacts with Prp8 and Gih35. a, Interactions between the Gpl1 N-terminal region and $Prp8^{En}$ and $Prp8^{Linker}$ (including the 1585-loop). The Cα atoms of residues contacting the pre-mRNA in the active site of the spliceosome are shown as spheres. **b**, Gpl1 residues 123-137 form a knob and insert into $Prp8^{RH}$, thus sequestering the Prp8 domain in this extreme position. An overlay of the cryo-EM map with Gpl1 residues 109-188 is provided, and the Cα atoms of Gpl1 knob residues I123 and D137 are shown as spheres. **c**, Interactions between the Gpl1 region including the brace loop and $Prp8^{RT}$ (see also panels e and f). **d**, Interactions between Gpl1 residues 199-234 and an interface formed by the $Prp8^{RT}$, $Prp8^{Thumb/X}$ and $Prp8^{Linker}$ domains. Inset shows a zoom-in of the interactions. For simplicity, only residues from Gpl1 are labeled. **e**, Superimposition of Gih35-Gpl1 with the cryo-EM map. Inset shows a zoom-in with well-defined side chain densities of Gih35 and the region of Gpl1 encompassing the brace loop. **f**, Overview of the Gih35-Gpl1 complex with the brace helix and brace loop marked with dotted boxes. The Gpl1 knob that interacts with $Prp8^{RH}$ in the *sp*B^{d}-II state is marked by a dotted circle. A brace encompassing the brace helix and brace loop is shown in black. **g**, Size-exclusion chromatography profiles of Prp43+Ntr1 G-patch domain (black) and Prp43+Gpl1 G-patch domain (red) are shown (molar ratio Prp43: Ntr1/Gpl1 1:2). Fractions loaded on SDS-PAGE gels are marked. 'arb. units' represent arbitrary units. 1% of sample mixture before loading on size-exclusion column (L) and size-exclusion fractions A1-A5 were separated on 19% SDS-PAGE gels. Positions of Prp43 (blue), Ntr1 G-patch domain (yellow) and Gpl1 G-patch domain (red) are marked with stars. While Prp43 and Ntr1 G-patch domain co-elute in fractions A1-A2, Prp43 and Gpl1 G-patch domain elute separately in fractions A1 and A4, respectively. Excess of Ntr1 G-patch domain elutes in fractions A4-A5. A representative of two independent measurements is shown.

https://doi.org/10.1038/s41594-024-01480-7

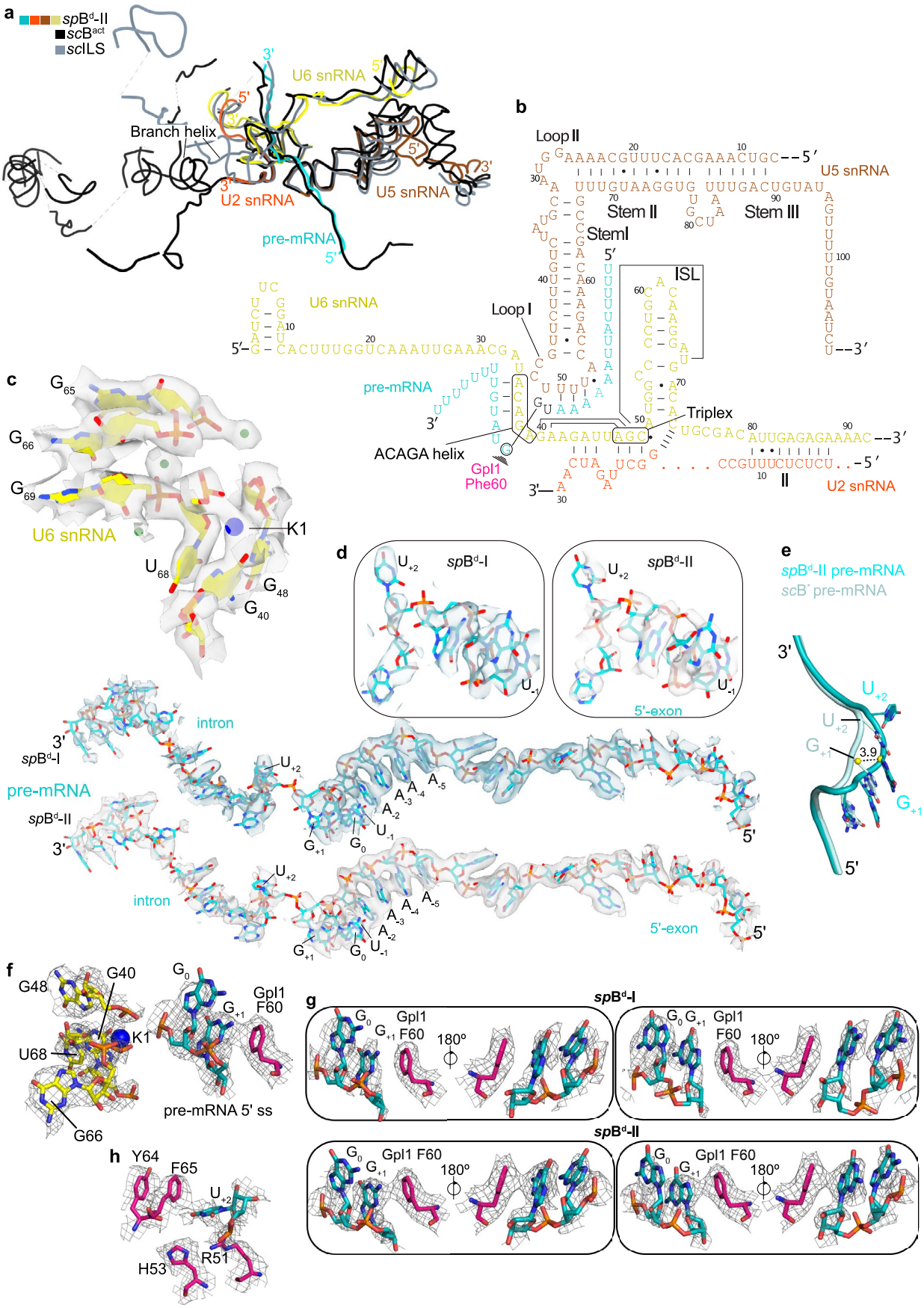


Extended Data Fig. 5 | See next page for caption.

https://doi.org/10.1038/s41594-024-01480-7

Extended Data Fig. 5 | Differences in location of proteins between the *sp*B^{d}-II state and other spliceosome complexes. a-b, Conformation sampling by the Prp8^{RH}. **a**, The domain arrangement of Prp8 from the *sc*B-*sc*/*sp*ILS and the *sp*B^{d}-II state is shown. For *sp*B^{d}-II Prp8, an overlay of the cryo-EM map is provided. The respective PDB IDs are indicated. **b**, Superposition of Prp8 from the *sc*B^{act}, *sc*B*, and the *sp*B^{d}-II complexes shows that Prp8^{RH} undergoes a 145° rotation from the *sc*B^{act} to *sc*B* state, while it undergoes a ~35° rotation in the opposite direction from the *sc*B^{act} to the *sp*B^{d}-II state. This position of the Prp8^{RH} in the *sp*B^{d}-II complex is fixed by the Gpl1 knob (Extended Data Fig. 4b). For simplicity, the Prp8$^{N\text{-}domain}$, Prp8^{RT}, Prp8$^{Thumb/X}$, Prp8^{Linker} and Prp8^{En} domains of the *sp*B^{d} complex are shown in cartoon representation, while Prp8^{RH} from *sc*B^{act} (purple), Prp8^{RH} from *sc*B* (teal blue) and Prp8^{RH} from *sp*B^{d}-II complex (pale blue) are shown as surfaces. Structure alignment is performed on the Prp8 Large domain comprising the N-domain till the En domain. **c-f**, Comparison of Gpl1 (residues 199-234) in the *sp*B^{d}-II complex with Prp45 and Ntr2 as part of the *sc*B^{act} and *sc*ILS complexes, respectively, indicates an overlapping binding site on Prp8. **c**, Positions of Gpl1 residues 199-234 (pink) and UNK (green) in the *sp*B^{d}-II complex are shown. **d**, Position of Prp45 (light magenta) in the *sc*B^{act} (PDB: 7DCO) complex is shown in the same view. **e**, Position of the helix-turn-helix motif of Ntr2 (orange) in the *sc*ILS (PDB: 5Y88) complex is shown. **f**, A superposition of the *sp*B^{d}-II, *sc*B^{act} and *sc*ILS complexes shows that Gpl1 residues 199-234 and UNK in *sp*B^{d}-II occupy the same location as stretches of Prp45 and Ntr2 in the *sc*B^{act} and *sc*ILS complexes, respectively. For simplicity, only the *sp*B^{d}-II complex is shown in black-and-white outline representation. **g**, RNA helicases Gih35, Prp2, Prp16 and Prp22 bind at similar positions at the periphery of the spliceosome. Positions of Gih35, Prp2, Prp16 and Prp22 (dark blue) at the periphery of the *sp*B^{d}-II complex, the *sc*B^{act} complex (PDB: 7DCO), the *sc*C complex (PDB: 5LJ5), and the *sc*P complex (PDB: 6BK8) are shown. The different RNA species are shown in color and Prp8 is shown in wheat.

https://doi.org/10.1038/s41594-024-01480-7

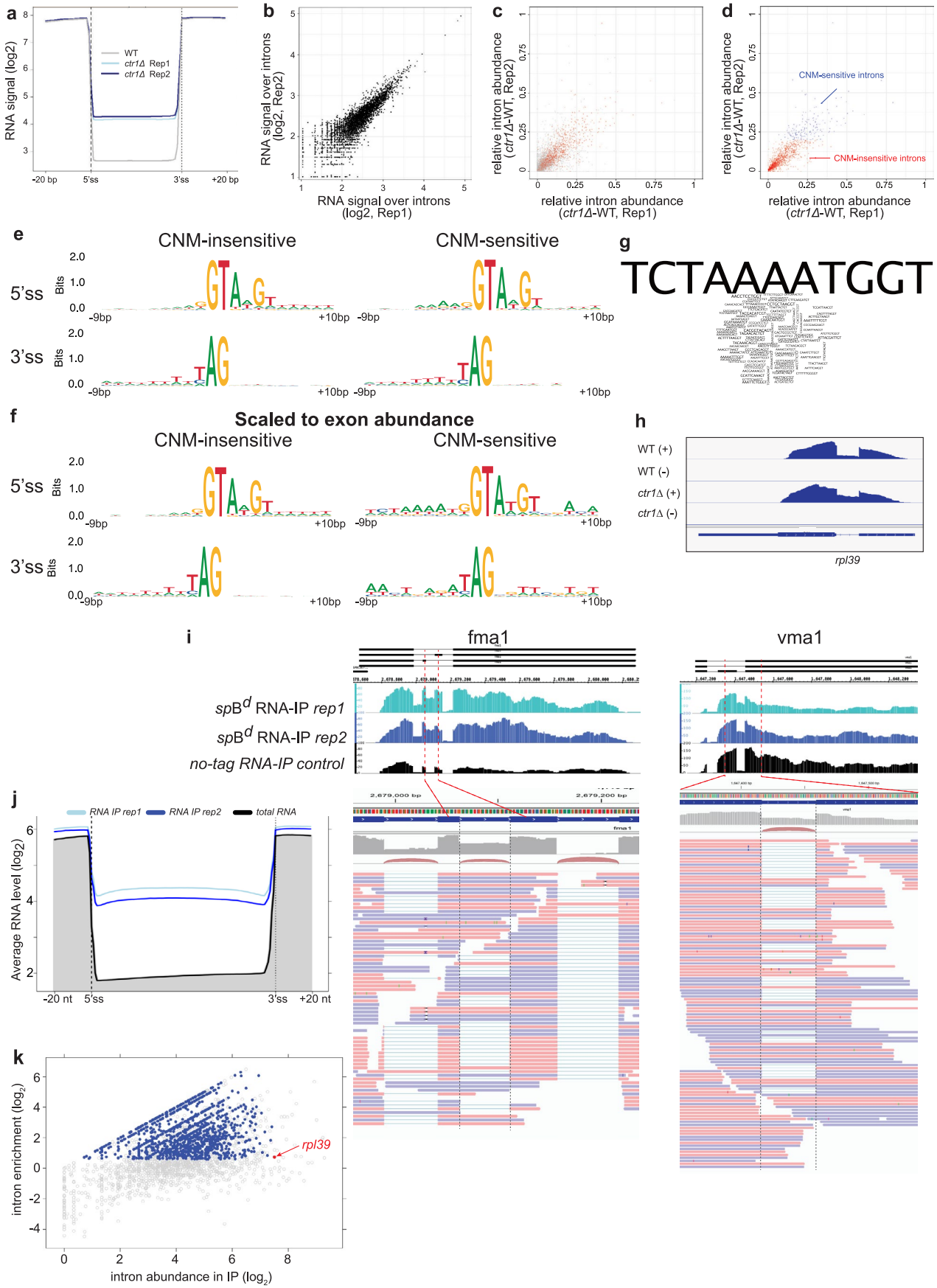


Extended Data Fig. 6 | See next page for caption.

https://doi.org/10.1038/s41594-024-01480-7

Extended Data Fig. 6 | RNA and proteins at the active site of the *sp*B^{d} complex. **a**, A comparison of the architecture of U2 snRNA, U5 snRNA, U6 snRNA and the pre-mRNA shows overall similarities of the *sp*B^{d} complex with the *sc*B^{act} (PDB: 7DCO) and *sc*ILS (PDB:5Y88) states. The structures are aligned on U6 snRNA. **b**, Schematic RNA representation for the *sp*B^{d} complex. Gpl1 Phe60, which stacks onto G_{+1} of the pre-mRNA, is shown. **c**, Cryo-EM map (*sp*B^{d}-II) shows signals for the K^{+} ion at the K1 site and three other structural Mg^{2+} ions. **d**, Overlays of *sp*B^{d}-I and *sp*B^{d}-II cryo-EM maps with the pre-mRNA. Insets show zoom-ins of nucleotides between positions -1 and +3 of the pre-mRNA. **e**, Insertion of an extra nucleotide at the active site introduces a bend in the pre-mRNA conformation (related to Fig. 5b). Yellow spheres represent centers of backbone sugar rings of *sc*B* pre-mRNA G_{+1} (PDB: 6J6Q) and *sp*B^{d}-II pre-mRNA G_{+1} and the dotted line marks the distance between them as 3.9 Å. **f**, Gpl1 Phe60 stacks on top of the 5′ss. The cryo-EM map shows density for the K^{+} ion at the K1 site. **g**, *sp*B^{d}-I and -II cryo-EM maps superimposed with the 5′ss G_{0} and G_{+1} and the Gpl1 Phe60, respectively. Since the signal of the pre-mRNA at the active site is weak, alternative conformations (180° flipped) of the G_{0} and G_{+1} can be modeled. However, in both *sp*B^{d}-I and *sp*B^{d}-II states, both the alternative conformations of G_{0} and G_{+1} stack well with Gpl1 Phe60. **h**, U_{+2} of the intron is sequestered into a pocket formed by Gpl1.

https://doi.org/10.1038/s41594-024-01480-7

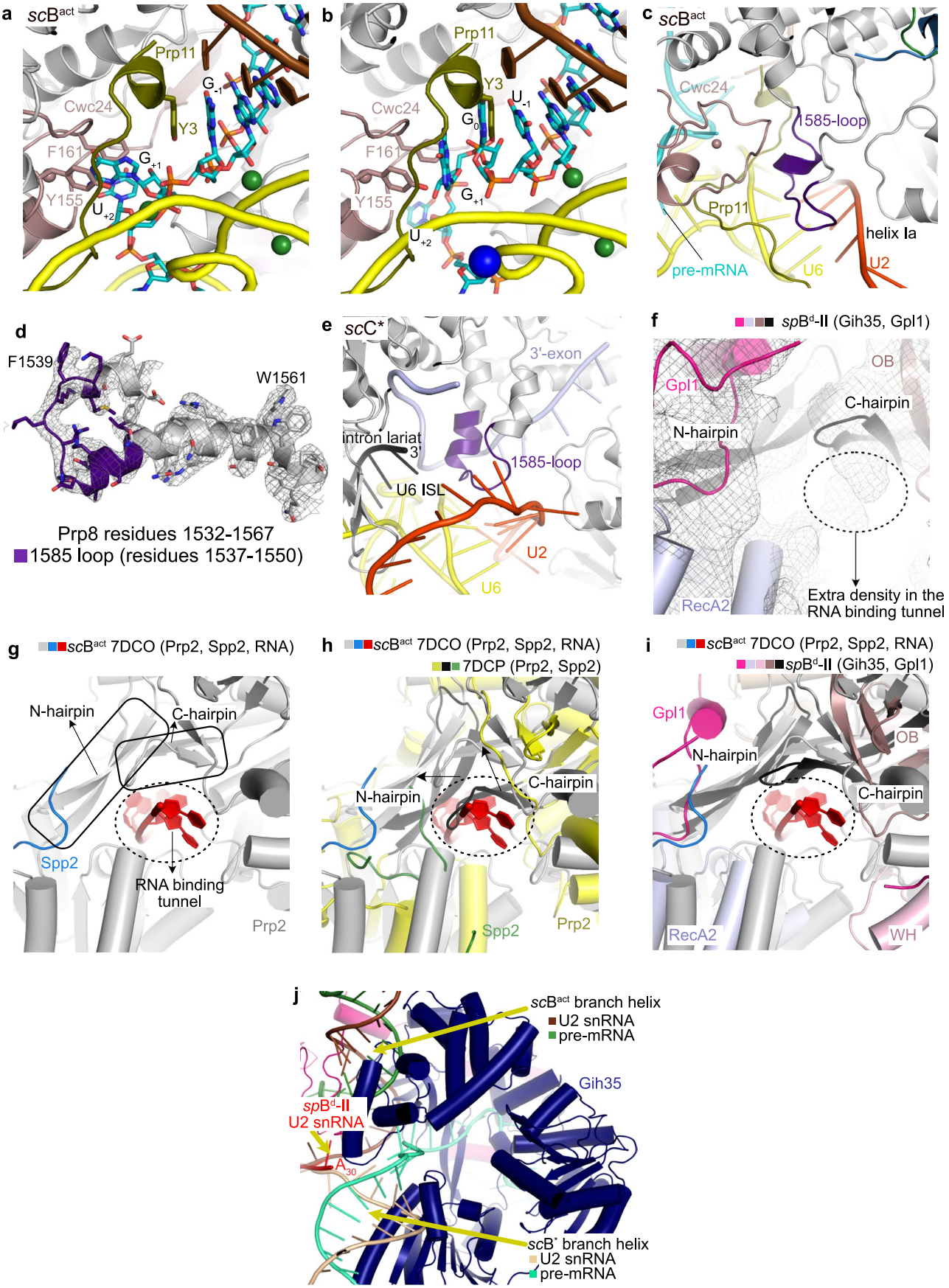


Extended Data Fig. 7 | See next page for caption.

https://doi.org/10.1038/s41594-024-01480-7

Extended Data Fig. 7 | RNA-seq analysis of *ctr1*Δ strain and *sp*B^{d} spliceosome RNA immunoprecipitation (RNA-IP). a-h, RNA-seq analysis of *ctr1Δ* strain. **a**, Metagene profile (geometric average) of sense RNA levels in the indicated strains for all annotated *S. pombe* introns from 20 bp upstream to 20 bp downstream of 5′ ss and 3′ ss of the introns. **b**, Scatter plot showing average RNA coverage of all annotated introns in two biological replicates of the *ctr1Δ* strain. **c-d**, Scatter plots representing relative intron abundance in *ctr1Δ* compared to WT strains. **c**, Introns with > 30 normalized read count in both biological replicates are shown in red. **d**, introns were further classified into CNM-sensitive sites (red) and CNM-insensitive sites (blue). **e-f**, Sequence logos of 5′ss and 3′ss of CNM-sensitive and CNM-insensitive sites shown, the height of the nucleotides expresses the information content in bits. Logos are based on unscaled ("per gene") occurrences (**e**) or being scaled to the normalized coverage of the upstream exon (**f**). **g**, Word cloud showing the 5′ss of CNM-sensitive introns. The height of the letters is proportional to normalized coverage of the upstream exon. **h**, IGV[109] snapshot of strand-specific RNA-seq coverage of the *rpl39* gene, in the indicated strains. **i-k**, *sp*B^{d} spliceosome RNA-IP. **i**, Gene browser view of strand-specific RNA levels at the *fma1* and *vma1* genes show representative examples of retained introns in the *sp*B^{d} spliceosome RNA-IP (light blue: replicate 1, dark blue: replicate 2). Black track shows the background signal (no-tag control for the RNA-IP). Lower panel shows the magnified view of individual strand-specific paired-end reads in the *sp*B^{d} RNA-IP *rep1* sample (pink: read1, blue: read2). **j**, Metagene profile (geometric average) of sense RNA levels in the *sp*B^{d} spliceosome purification (light blue: replicate 1, dark blue: replicate 2) and in total RNA of wild-type strain (black line and grey shaded area), for all annotated *S. pombe* introns (datasets were normalized for total exon coverage). **k**, Scatter plot of $\log_2$ intron coverage (X-axis) versus $\log_2$ intron enrichment, compared to no-tag RNA-IP background (Y-axis). Dots represent individual introns. Blue dots represent enriched introns in the RNA-IP, red dot represents intron 1 in *rpl39* gene.

https://doi.org/10.1038/s41594-024-01480-7

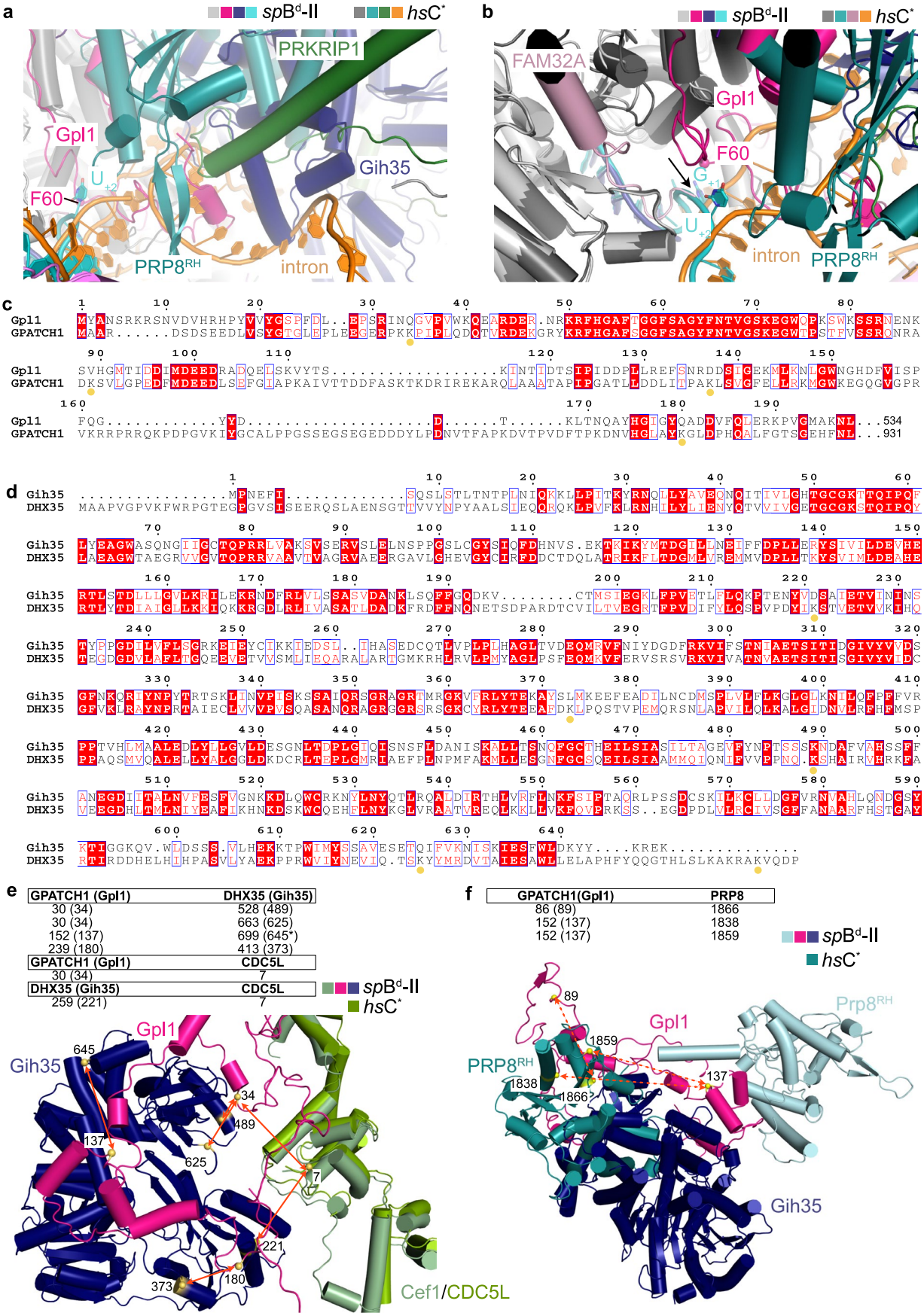


Extended Data Fig. 8 | See next page for caption.

https://doi.org/10.1038/s41594-024-01480-7

Extended Data Fig. 8 | Comparisons of *sp*B^{d} active site with that of *sc*B^{act} and scC*; RNA binding of Prp2 in the *sc*B^{act} complex and Gih35 in the *sp*B^{d} complex. a-e, Comparisons of *sp*B^{d} active site with that of *sc*B^{act} and scC*. **a**, Prp11 and Cwc24 are bound to the active site in the *sc*B^{act} complex (PDB: 7DCO). Prp11 Tyr3 stacks with G_{-1} and Tyr155, and Phe161 of Cwc24 stack with G_{+1} and U_{+2}. **b**, Superposition of Prp11 and Cwc24 from the *sc*B^{act} complex onto the *sp*B^{d} complex shows steric clashes between the pre-mRNA and Prp11/Cwc24. For simplicity, Gpl1 is not shown. **c**, The 1585-loop of Prp8 interacts with Prp11, Cwc24 and the U2/U6 duplex in the *sc*B^{act} complex (PDB: 7DCO). **d**, The 1585-loop of Prp8 comprising residues 1537-1550 (purple) in *S. pombe* is structured in the *sp*B^{d} complex as seen by the overlay with the cryo-EM map. **e**, Position of the 1585-loop in the *sc*C* complex (PDB: 5WSG) is similar to that of the *sp*B^{d} complex (Fig. 6d). **f-j**, RNA binding of Prp2 in the *sc*B^{act} complex and Gih35 in the *sp*B^{d} complex. **f**, *sp*B^{d}-II cryo-EM map shows density likely corresponding to an RNA substrate bound to Gih35. **g**, RNA-binding to Prp2. The N-and C-hairpins of Prp2 prevent the backsliding of pre-mRNA (PDB: 7DCO). The RNA binding tunnel is outlined with dashes. **h**, Structure superposition of Spp2-Prp2 in the absence (PDB: 7DCP) and presence of RNA (PDB: 7DCO) show that the N- and C-hairpins undergo large movements to accommodate the RNA. The structures are aligned on the RecA1 domains. **i**, Structure superposition of the Gih35-Gpl1 complex with Prp2-Spp2-RNA shows that the position of the N- and C-hairpins of Gih35 in the *sp*B^{d} complex are compatible with RNA binding. **j**, A close-up view of the last nucleotide of U2 snRNA visible in the *sp*B^{d} complex (A_{30}) positioned close to the RNA binding tunnel of Gih35. Superpositions with the *sc*B^{act} complex (PDB: 7DCO) and the *sc*B* complex (PDB: 6J6H) show that the branch helix from either state would clash with Gih35. Given that Gih35 is bound to an RNA substrate and the branch helix is invisible in our structure, it is likely that Gih35 acts on the branch helix.

https://doi.org/10.1038/s41594-024-01480-7



Extended Data Fig. 9 | See next page for caption.

https://doi.org/10.1038/s41594-024-01480-7

Extended Data Fig. 9 | Comparison between Gih35-Gpl1 in the *sp*B^{d} complex and DHX35-GPATCH1 in the *hs*C* complex. a, Superposition of *sp*B^{d}-II state with *hs*C* complex. For simplicity, only PRKRIP1 (green), PRP8^{RH} (teal) and intron (orange) from *hs*C* complex are shown in color. Gpl1 and Gih35 from *sp*B^{d} complex clash with PRKRIP1 and Prp8^{RH} from *hs*C* complex. For orientation, pre-mRNA U_{+2} and Gpl1 Phe60 (Cα atom as sphere) are marked. **b**, FAM32A is at the catalytic center in *hs*C* (marked by a black arrow), which is incompatible with the pre-mRNA 5′ss and Gpl1 in the *sp*B^{d} complex. **c** and **d**, Sequence alignments of Gpl1 with GPATCH1 (c) and Gih35 with DHX35 (d), respectively. Yellow dots mark the lysine residues of GPATCH1/DHX35 which crosslink to other proteins as listed in panels (e) and (f). **e**, All crosslinks between GPATCH1 and DHX35 reported for *hs*C* complex[44] are satisfied (distance between Cα atoms of crosslinked residues < 24 Å) in the *sp*B^{d} complex using corresponding amino acid positions from *S. pombe* orthologues (listed in brackets, shown as red arrows). Cef1 and CDC5L superimpose well and the crosslinks from CDC5L to Gpl1/Gih35 are also satisfied. Cα atoms for the corresponding residues from Gpl1, Gih35 and CDC5L are shown as yellow spheres. **f**, All crosslinks between GPATCH1 and Prp8^{RH} reported for *hs*C* complex[44] are unsatisfied (distance between Cα atoms of crosslinked residues > 24 Å, marked by dashed red arrows) in the *sp*B^{d}-II complex due to the differences in location of Prp8^{RH} between *sp*B^{d}-II complex and *hs*C* complex.

https://doi.org/10.1038/s41594-024-01480-7

Extended Data Table 1 | Summary of model building for the *sp*B^{d}-I (PDB ID: 9ESH) and *sp*B^{d}-II (PDB ID: 9ESI) complexes

	Protein/RNA component (*S. pombe*/ *S. cerevisiae*/*H. sapiens*)	Total length	Modeling template	Modeled region in *sp*B^{d}-I (including gaps)	Modeled region in *sp*B^{d}-II (including gaps)	*sp*B^{d}-I Chain ID	*sp*B^{d}-II Chain ID
Pre-mRNA	Pre-mRNA	-	-	29 (5'exon and intron)	29 (5'exon and intron)	1	1
U5 snRNP	U5 snRNA	120 nts	3JB9	7-108	7-108	5	5
	Spp42/Prp8/PRP8	2363		45-1781	45-2030	A	A
	Cwf10/Snu114/SNU114	984		67-984	67-984	B	B
	Cwf17/-/SNRNP40	340		36-338	36-338	C	C
	SmD3	97		2-97	2-97	D	D
	SmB	147		3-118	3-118	E	E
	SmD1	117		2-82	2-82	F	F
	SmD2	115		5-115	5-115	G	G
	SmE	84		5-84	5-84	H	H
	SmF	78		3-75	3-75	I	I
	SmG	77		3-75	3-75	J	J
U6 snRNP	U6 snRNA	99 nts	3JB9	1-92	1-92	6	6
U2 snRNP	U2 snRNA	186 nts	3JB9	3-30	3-30	2	2
NTC core	Cwf4/Clf1/SYF3	674	3JB9 + AF*	11-620	11-620	R	R
	Prp19/Prp19/PRP19	488		1-132,1-134,1-135,1-131	1-132,1-134,1-135,1-131	S,T,U,V	S,T,U,V
	Cdc5/Cef1/CDC5L	757		4-757	4-757	W	W
	Cwf3/Syf1/SYF1	790		63-735	63-735	X	X
	Cwf7/Snt309/SPF27	187		13-185	13-185	Z	Z
	Syf2/Syf2/SYF2	229	6J6Q	83-215	83-184	Y	Y
NTC related	Prp5/Prp46/PRL1	473	3JB9	83-473	83-473	K	K
	Prp45/Prp45/SKIP	557		82-332	82-332	L	L
	Cwf5/Ecm2/RBM22	354		16-305	16-305	M	M
	Cwf11/-/Aquarius	1284	AF	1-1284	1-1284	N	N
	Cwf14/Bud31/G10	146	3JB9	3-146	3-146	O	O
	Cwf2/Cwc2/RBM22	388		47-328	47-328	P	P
	Cwf15/Cwc15/CWC15	265		24-265	24-265	Q	Q
Splicing factors	Prp17/Prp17/CDC40	558	3JB9	9-160	9-160	a	a
	Cwf21/Cwc21/SRRM2	293	5LJ3 + AF	2-145	2-145	b	b
	Cwf22/Cwc22/CWC22	887	5MQF + AF	404-607	404-607	c	c
	Cyp1/-/PPIL1	155	3JB9	2-155	2-155	d	d
	Bis1/-/ESS2	384	AF	-	22-262	-	e
	Saf4/Yju2/CCDC194	299	AF	-	156-215	-	p
Ntr1 complex	Ntr1/Ntr1/TFIP11	797	5Y88 + AF	717-797	264-797	m	m
	Ntr2/Ntr2/-	361	5Y88 + AF	-	296-325	-	n
Gih35-Gpl1	Gih35/-/DHX35	647	AF	20-645	20-645	z	z
	Gpl1/-/GPATCH1	534	*de novo* + 6HS6 + 7DCP	25-234	25-234	y	y
Unknown	UNK	-	-	-	-	-	q
	Unknown (NTC)	-	-	-	-	r	r
	Unknown (around Gih35)	-	-	-	-	f	f

*AF: Alphafold2 model

nature portfolio

Corresponding author(s): Komal Soni, Tamas Fischer, Irmgard Sinning

Last updated by author(s): Nov 25, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Cryo-EM images were acquired using Serial-EM.
Data analysis	Cryo-EM images were processed with MotionCor2, Gctf v1.06, Warp v1.0.9, Relion 3.1, cisTEM 1.0.0, and cryoSPARC V3.1. Molecular models were built with Coot v0.8.9.3-pre and refined using Phenix 1.19.2. Structures and cryo-EM maps were visualized with PyMOL 2.5.3, UCSF Chimera 1.12 and ChimeraX 1.2.5 validated with Phenix 1.19.2. All protein sequence and ortholog information was obtained using PomBase database annotation v.62 and sequence analysis was performed with Clustal Omega v1.2.4, visualized with ESPript v3.0, secondary structure prediction was performed with PSIPRED and homology models were obtained from Swissmodel. Structure predictions from AlphaFold Protein Structure Database were used and identification of unknown proteins in cryo-EM map was performed with ModelAngelo. Protein-protein crosslinks were searched using pLink v.2.3.9 and protein composition was determined by MaxQuant v.1.6.17.0. Paired-end RNA-seq reads were aligned to the S. pombe reference genome (ASM294v2) using HISAT2.2.11. Alignment files were sorted and indexed using SAMtools1.21101. Un-normalized bigwig files were generated with bamCoverage from deepTools3.5.5. Normalization and further data processing was done by in-house bash and R scripts, available on the following GitHub repository: https://github.com/ahorvath/Soni-et-al.-2023.git

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All protein sequence and ortholog information was obtained using PomBase database annotation v.62 (https://www.pombase.org/). The structures of spBd-I and -II states are deposited in the PDB under accession codes 9ESH and 9ESI and cryo-EM density maps are available at the EMDB under the accession codes EMD-19941 and EMD-19942, respectively. The RNA-seq data have been deposited to NCBI GEO under the reference number: GSE235589. S. pombe reference genome (ASM294v2)is accessible at: https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000002945.1/

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Our study does not involve human research participants.
Reporting on race, ethnicity, or other socially relevant groupings	Our study does not involve human research participants.
Population characteristics	Our study does not involve human research participants.
Recruitment	Our study does not involve human research participants.
Ethics oversight	Our study does not involve human research participants.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For cryo-EM a single final dataset was collected comprising 13,096 micrographs was collected.
Data exclusions	No data were excluded from the analysis.
Replication	All experiments were repeated twice independently with similar results. n=2 for the RNA-seq experiments, n=2 for the yeast two hybrid experiments, n=2 for size exclusion chromatography co-elution experiments. For cryo-EM a single final dataset was collected comprising 13,096 micrographs. All attempts at replication were successful.
Randomization	There was no allocation of groups done in this study and so no Randomization was required.
Blinding	There was no allocation of groups done in this study and so no Randomization was required.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.