Published in final edited form as:
Nature. 2013 January 31; 493(7434): 638–643. doi:10.1038/nature11843.

Crystal structure of Prp8 reveals active site cavity of the spliceosome

Wojciech P. Galej, **Chris Oubridge**, **Andrew J. Newman**, and **Kiyoshi Nagai**
MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, England

Abstract

The active centre of the spliceosome consists of an intricate network formed by U5, U2 and U6 snRNAs, and a pre-mRNA substrate. Prp8, a component of the U5 snRNP, crosslinks extensively with this RNA catalytic core. We present the crystal structure of yeast Prp8 (residues 885-2413) in complex with the U5 snRNP assembly factor Aar2. The structure reveals new tightly associated domains of Prp8 resembling a bacterial group II intron reverse transcriptase and a type II restriction endonuclease. Suppressors of splice site mutations and an intron branchpoint crosslink map to a large cavity formed by the reverse transcriptase thumb, endonuclease-like and the RNaseH-like domains. This cavity is large enough to accommodate the catalytic core of group II intron RNA. The structure provides crucial insights into the architecture of the spliceosome's active site and reinforces the notion that nuclear pre-mRNA splicing and group II intron splicing have a common origin.

Removal of introns from nuclear pre-mRNA occurs in two consecutive trans-esterification reactions catalyzed by a multi-megadalton, dynamic RNA-protein complex - the spliceosome (reviewed in ref 1). The spliceosome is formed on pre-mRNA substrates from its five canonical subunits, the small nuclear ribonucleoprotein particles (U1, U2, U4/U6 and U5 snRNPs), and numerous non-snRNP factors. Assembly of the spliceosome begins with the recognition of the 5′-splice site (5′-SS) by U1 snRNP followed by branchpoint (BP) recognition by U2 snRNP. After the recruitment of a pre-assembled U5-U4/U6 tri-snRNP, in which U4 and U6 snRNAs are extensively base-paired, the spliceosome undergoes a major structural and compositional rearrangement including unwinding of the U4-U6 snRNA duplex and concomitant formation of a highly structured RNA network between U2, U5 and U6 snRNAs and the 5′-SS and BP sequences in the pre-mRNA. This leads to a nucleophilic attack of the BP adenosine at the 5′-SS, producing exon1 and lariat intron-exon2 intermediates. Further remodeling enables a nucleophilic attack of exon1 at the 3′-splice site (3′-SS) yielding spliced mRNA and lariat intron products. At the catalytic core of the spliceosome, the base-paired U2-U6 snRNAs provide a platform for correct positioning of the BP and 5′-SS[2-6] as well as coordinating catalytic magnesium ions[7,8] and U5 snRNA aligns the exons for the second catalytic step[9-11].

Correspondence and requests for materials should be addressed to K.N. (kn@mrc-lmb.cam.ac.uk) and A.J.N. (newman@mrc-lmb.cam.ac.uk)..

Author Contributions A.J.N. and K.N. initiated the project and worked on protein expression and purification for many years. Co-expression of Prp8 and Aar2 by A.J.N. was a crucial step of the project. W.P.G. successfully identified and expressed a stable large fragment of Prp8, crystallized the Prp8-Aar2 complex and solved and refined the structure almost single handedly with the practical support from K.N. and A.J.N. C.J.O. analysed the mercury derivative data and refined the structure of the $P2_12_12_1$ crystal form. W.P.G. and K.N. analysed the structure and wrote the paper with important input from A.J.N. and C.J.O.

Atomic coordinates and structure factors for the Prp8-Aar2 complex have been deposited in the Protein Data Bank under accession code: 4I43 ($C222_1$) and 3ZEF ($P2_12_12_1$).

The authors declare no competing financial interests.

Europe PMC Funders Author Manuscripts

Three U5 snRNP proteins, Prp8, Brr2 and Snu114, play crucial roles in the activation of the spliceosome and the formation of the catalytic core for the two trans-esterification reactions. Yeast Prp8 is a 280 kDa protein and has 61% sequence identity to its human counterpart. Human Prp8 forms a salt-stable complex with the EF2-like GTPase Snu114 and the DExD/H-box family helicase Brr2[12]. GTP-bound Snu114 activates Brr2[13,14], which unwinds the U4-U6 snRNA duplex to allow U6 snRNA to base-pair with U2 snRNA. Prp8 crosslinks with crucial snRNAs (U5 and U6) and substrate residues (5′-SS, BP and 3′-SS)[15-20]. Many Prp8 mutations suppress splicing defects caused by mutations in the 5′-SS, 3′-SS and BP (ref 21 and references therein). Hence Prp8 is located at the heart of the spliceosome. Despite the central role of Prp8 in splicing, the structures of only two small domains, the RNaseH-like and Jab1/MPN domains, have been determined so far[22-26]. Based on bioinformatic analysis it has been proposed that part of Prp8 forms an RRM[21] and recently Dlakić and Mushegian[27] reported sequence similarity between the central part of Prp8 and a catalytic domain of reverse transcriptase from bacterial group II intron encoded protein.

In the cytoplasm, Prp8 forms a large complex containing U5 snRNA, Snu114, 7 Sm proteins and the U5 snRNP assembly factor Aar2. After import of this complex into the nucleus, Aar2 is replaced by Brr2 and additional proteins are recruited to form the mature U5 snRNP[28,29]. In this study we report a crystal structure of a large C-terminal fragment of *Saccharomyces cerevisiae* Prp8 (residues 885-2413) in complex with full-length Aar2.

Overall architecture of the complex

A 176 kDa fragment of yeast Prp8 was co-crystallised with Aar2. The crystals diffracted to 1.9Å resolution and the structure was solved by molecular replacement using the crystal structures of the RNaseH-like and Jab1/MPN domains and Aar2 (3SBT, 2OG4) as search models (supplementary methods; Tables S1 and S2). A methyl mercury derivative was used to verify molecular replacement solutions (Fig S1). The structure was refined at 2 Å resolution to R free of 24.8% (Fig S2). The crystal structure revealed a new large domain of Prp8 (residues 885-1824) spanning the entire length of the complex (Figs 1a and 2a). The RNaseH-like[22-24] and Jab1/MPN domains[25,26], each connected via disordered linkers, fold back to interact with the new domain via Aar2 (Fig 2, Fig S3). This domain can be sub-divided into a large polymerase-like domain (Table S3) and a small type II restriction endonuclease-like domain (Table S4) that interact intimately through the linker domain. The polymerase-like domain is composed of three canonical sub-domains[30]: palm, fingers and thumb (Figs S4 and S5). They form a deep cleft, which accommodates the nucleic acid template and primer in polymerases. In all polymerases, the most highly conserved palm sub-domain lies at the bottom of the deep cleft and contains the catalytically important residues embedded in 4 conserved motifs (A,B,C,D)[30]. Aspartates, one in motif A and two in motif C, coordinate two Mg^{2+} ions, required for catalysis (Fig 1b), while motif B is involved in nucleotide selection. In Prp8 the palm sub-domain forms a four-stranded anti-parallel β-sheet (RTβ4, RTβ7-RTβ9) flanked by α-helices (RTα6, RTα9-RTα13). Motif C is located in the loop between RTβ7 and RTβ8 whereas motifs A, B and D in RTβ4, RTα9 and RTα13, respectively. Only one of the three aspartates (D1166), equivalent to the first aspartate in the YxDD consensus sequence in motif C[30], is conserved in Prp8 (Fig 1c, Fig S3). T1053 and R1167, which replace the two other catalytic aspartates, are neither capable of metal ion coordination, nor conserved in Prp8 from different species. Hence, the “active site” of the Prp8 polymerase-like domain is unlikely to bind divalent metal ions. Recently, Dlakić and Mushegian[27] reported that residues 1048-1182 of yeast Prp8 show significant sequence similarity to the region corresponding to the reverse transcriptase palm domain of bacterial group II intron encoded proteins (IEP). Hence, we will more appropriately refer to this domain as the reverse transcriptase (RT) domain. Residues 1058-1151 were predicted to form an RRM[21] but this region does not resemble the RRM fold and is embedded in the

finger-like sub-domain (Fig S6). The thumb sub-domain (1257-1375) forms a four-stranded anti-parallel β-sheet (RTβ10-RTβ13) followed by a 3-helix bundle (RTα14-RTα16). A significant sequence similarity between this region and the Th/X domain of fungal group II intron RT was noted and it was correctly predicted to form a helix bundle[27] (Fig S7).

Residues 1650-1810 adopt a type II restriction endonuclease–like fold, with characteristic five mixed β-strands (Enβ1, Enβ2, Enβ4-Enβ6) flanked by three α-helices (Enα1-Enα3). The Prp8 endonuclease (En) domain is structurally most similar to the PA endonuclease domain of the Influenza virus[31,32] with little apparent sequence conservation (9% identity) (Fig S8, Table S4). Two glutamates (E80, E119), one aspartate (D108) and one histidine (H41) are involved in 2-metal ion coordination essential for catalytic activity in the PA endonuclease domain[31,32]. Intriguingly, although all these residues (Fig 1d and e) are highly conserved in Prp8 except E1684 (Fig S3), replacement of these residues with uncharged amino acids (Asp→Asn, Glu→Gln), individually and in combination, had no effect on viability (Fig S9). These residues form a network of polar interactions and stabilize the polypeptide loops that block the active site (Fig S10) and hence may be conserved for a structural reason.

Domain organization

The RT/En, RNaseH-like and Jab1/MPN domains are connected by disordered linkers but form a large assembly stabilized by a network of pivotal interactions involving Aar2 (Fig 2b); this domain arrangement is undoubtedly crucial in biogenesis of U5 snRNP[28,29]. Aar2 interacts with Prp8 across the junction between the Linker and En domains burying 1230Å^2 of solvent-accessible surface (Fig S11). The C-terminal tail (G318 to P355) of Aar2 extends from its main body through the cleft between the RT fingers and Th/X domains; its very C-terminal end (residues 348-353) forms a remarkable intermolecular, parallel β-sheet, zipping together the β-hairpin of the RNaseH-like domain (residues 1860-1864) and the β-barrel of the Jab1/MPN domain (2167-2171) (Fig 2b, Fig S12). This accounts for the crucial role of the C-terminal tail of Aar2 (residues 331-354) in bringing the RNaseH-like and Jab1/MPN together[29]. Furthermore, the C-terminal helical domain (α5 and α6) of Aar2 interacts with RHα1 and RHα2 of the RNaseH-like domain burying 433 Å^2 of solvent accessible surface (Fig 2b, Fig S13). The interface between Aar2 and the RNaseH-like domain previously reported by Weber *et al.*[29] was not observed in our structure. However, one of the crystal contacts present in the Aar2-RNaseH-like domain complex structure (3SBT) is remarkably similar to the interface between the helical regions observed in our complex (Fig S14). This raises the possibility that the biological interface was incorrectly assigned previously. The Jab1/MPN domain makes contact with the RT and Linker domains (Fig S15). Individually these interactions may not be sufficient to fix the Jab1/MPN domain in position but the network of interactions between the four domains (RT/En, RNaseH-like, Jab1/MPN and Aar2) holds them together. The RNaseH-like domain itself has little if any direct contact with the RT/En domain and is positioned relative to the RT/En domain indirectly by Aar2 (Fig 2c). Hence the exact position of the RNaseH-like and Jab1/MPN domains with respect to the RT/En may be altered when Aar2 is replaced by Brr2 and Prp8 forms a complex with Snu114 and Brr2. Indeed, comparison of the $P2_12_12_1$ and $C222_1$ crystal forms revealed movements of the RNaseH-like and Jab1/MPN domains with respect to the RT/En domain (Fig S16). The domain interactions could be modulated when Snu114 binds different nucleotides.

The Active site of the spliceosome

The catalytic centre of the spliceosome includes an intricate network of interactions involving U2, U5 and U6 snRNAs and substrate pre-mRNA (Fig S17). Prp8 crosslinks to

crucial residues in U6 snRNA and in the invariant exon-binding loop 1 of U5 snRNA as well as to all 3 sites of chemistry in the pre-mRNA (5′-SS, BP and 3′-SS)[15-18,33,34]. Contacts between yeast Prp8 and catalytic core RNA residues were previously mapped by crosslinking and proteolytic cleavage; almost all of the crosslinks lie within the RT-En domain of Prp8[20] (Fig S17). We have used substrates with modified 3′-SS and captured spliceosomes via the NTC complex protein Prp19 or the step 2 factor Prp18 in order to focus on crosslinks made just before catalytic step 2 (C.M. Norman and A.J.N, unpublished results). Prp8-BP(+2) crosslinks were mapped to the region between residues 1585 and 1598. This disordered region is located between the blue spheres in Fig 3c. The crosslinking site is located in the mobile loop near the RT Th/X domain and distant from the residues corresponding to the Mg^{2+}-coordinating residues in the RT and En domains.

Splice site and branch point suppressors

Screening for suppression of splicing defects caused by mutations in the 5′-SS, 3′-SS and BP led to the isolation of many Prp8 suppressors[35,36] (references in ref 20 and Table S5). Most suppressor mutations are located on the concave surface of the Th/X and En domains facing the RNaseH-like domain (Fig 3). This space is lined with extended polypeptide chains making few contacts. These loops are part of the regions crosslinked to the 5′-SS(−1), BP(+2), U6 snRNA(U54) and the U5 snRNA loop 1 (U97)[20]. The surface of the RNaseH-like domain facing this cavity includes E1960 and E1834, which are sites of 5′-SS and 3′-SS suppressor mutations[36]; they are positioned close to the 5′SS-hPrp8 cross-link (Fig 3)[18]. These site-specific crosslinks together with the suppressor mutations unambiguously locate the active site of the spliceosome to the cavity formed by the Th/En and RNaseH-like domains.

Four suppressors of 5′- and 3′-SS and BP mutations map in the RNaseH-like domain β-finger. As in the Aar2-Prp8 complex, the β-finger may be involved in alternative interaction to position the RNaseH-like domain in the Brr2-containing mature U5 snRNP and these mutations may affect the interaction between Prp8 and the RNA network as previously proposed[24]. Some of the suppressors facilitate the first step but inhibit the second (first step alleles) and others do the opposite[35] (Table S5). It has been proposed that Prp8 has two alternative states that facilitate either the first or the second step[24,35]. It is plausible that this corresponds to alternative positions of the RNaseH-like domain with respect to the RT/En domain; the RNaseH-like domain may in turn transmit conformational changes of the RNA network after each trans-esterification reaction. Comparison of the two crystal forms confirmed flexibility of the RNaseH-like domain, revealing its significant interdomain movements with respect to RT and En domains (Fig S16).

Prp8 and spliceosome activation

In the U4-cs1 (cold sensitive) mutant the three nucleotides (AAA) adjacent to helix I of the U4-U6 snRNA duplex are replaced by UUG, extending helix I by three base-pairs and concealing part of the U6 snRNA sequence (ACAGA) which base-pairs with the 5′-SS. At the restrictive temperature (16 °C) the spliceosome stalls before the first trans-esterification because the U4-U6 snRNAs remain base-paired and the ACAGA box fails to base-pair with the 5′-SS. Screens for suppressors of U4-cs1 in Prp8 isolated many single mutations in five regions (Prp8-cat mutants) (Table S6)[36,37]. Two of the five regions (d and e) are within our crystal structure (Fig 4). Six mutations in region d map within or near the four-stranded β-sheet protruding from the reverse transcriptase fingers domain (Fig 4c), and three map on the loop connecting RTα12 and RTα13. These are the most exposed parts of the fingers and palm domain. In region d and part of region e, three of the mutations are located within or near the exposed β-hairpin of the reverse transcriptase-like domain and two on the exposed

loop in the endonuclease domain. Intriguingly all of the suppressor mutations in the RT/En domain map on one face of the RT/En domain. Five mutations of region e map within the RNaseH-like domain (Fig 4a), and four of these are located within the β-finger, which forms a continuous β-sheet with the tail of Aar2 and the β-barrel in the Jab1/MPN domain. It is possible that this β-hairpin is also involved in a protein–protein interaction crucial for positioning the RNaseH-like domain in the U5 snRNP or spliceosome.

One of the Prp8-cat mutations (V1098D) suppresses a cold sensitive mutation in the first RecA domain in Brr2 known as *brr2-1* (E610G). The residues (1022-1214) covering the entire fingers/palm domain were subjected to further mutagenesis, and five additional *brr2-1* suppressors were isolated[38] (Table S7). These mutations also map within or at the base of the four-stranded β-sheet exposed on the surface of the RT palm domain (Fig 4c). The exact mechanism of U4-cs1 suppression is unclear but the fact that all suppressor mutations map on one face of the RT/En domain suggests this is a crucial RNA or, more likely, protein binding surface. As *brr2-1* and some U4-cs1 suppressors map in the same region it may be a binding interface for Brr2 in addition to the Jab1/MPN domain. Brr2 bound on this surface would be ideally positioned to feed U6 snRNA, unwound from the U4/U6 duplex, into the active site cleft in Prp8 and may be in close proximity to the active site RNA. Ski2-type RNA helicases such as Brr2 unwind duplex RNA after loading onto a 3′ overhanging end. The extended helix I of the U4/U6 duplex in U4-cs1 mutant alters the position and orientation of the 3′ overhanging nucleotides. It is possible that U4-cs1 and *brr2-1* suppressor mutations may slightly reposition Brr2 to facilitate loading of the U6/U4-cs1 substrate.

The origin of the spliceosome

The crystal structure of Prp8$^{885\text{-}2413}$ has revealed a new large domain consisting of reverse transcriptase and endonuclease domains. The fact that the palm and Th/X domains have significant sequence and structural similarity to their counterparts in bacterial and fungal group II intron encoded proteins has very important functional and evolutionary implications. The farsighted hypothesis that nuclear pre-mRNA splicing and group II self-splicing have a common origin was based on the fact that in both processes introns are excised by two successive trans-esterification reactions via a lariat intermediate[39,40]. Self-splicing group II introns have six structurally conserved domains DI–DVI[41]. The fact that the RNA core of the spliceosome contains structural and functional counterparts of Domains V and VI and the exon binding loop of group II intron further strengthened this hypothesis[9,39,40,42-44]. However, there is still an enormous evolutionary gap between the two. Our crystal structure has provided the first experimental evidence for a link between a group II intron encoded protein (IEP) and a component of the spliceosome. Group II introns are mobile genetic elements whose self-splicing is facilitated by the maturase activity of their IEP[43]. IEP remains bound to the excised intron and targets it to a homing site in genomic DNA where intron insertion is achieved by reverse splicing. The opposite DNA strand is then cleaved by the endonuclease domain of IEP and used as a primer for reverse transcription of the intron RNA by the IEP reverse transcriptase domain. IEP usually contains an endonuclease domain of the H-N-H endonuclease family[43] whereas Prp8 contains a type II restriction endonuclease domain.

Remarkably organellar group II introns, split into independently transcribed segments[45], can undergo inefficient *trans*-splicing[41,45]. Ancestral nuclear pre-mRNA splicing activity could have evolved from the IEP open reading frame (encoding ancestral Prp8) and the group II intron RNA domains (ancestral snRNAs), which became independent transcription units. The resulting incomplete group II introns could have been excised with the help of IEP and the *trans*-acting Group II intron RNA domains and were gradually freed from the

evolutionary constraints to maintain self-splicing activity as all the RNA domains, except the BP sequence, became *trans*-acting elements. When some group II introns ceased to be mobile elements the selective pressure to maintain the catalytically active RT domain was lost, but the RT domain continued to function as a maturase and became an assembly platform for the primitive snRNAs and substrate pre-mRNA. The RT domain, particularly the fingers and thumb domains, continued to evolve and addition of the RNaseH-like and Jab1/MPN domains facilitated the evolution of ancestral snRNAs into snRNPs as they recruited more proteins.

The crystal structure of an *Oceanobacillus iheyensis* group II intron reveals a tightly packed functional centre consisting of exon-binding loops (EBS), exons and domain V organized by the surrounding RNA scaffold[46,47]. No structure of a group II intron in complex with IEP has yet been reported but the interaction between the *Lactoccocus lactis* L1LtrB intron and its IEP has been studied biochemically[48-50]. The N-terminal region of the RT domain binds intron DIVa with high affinity whereas the Th/X domain makes contact with the catalytic core of the intron including the E1-DI, DII and DVI-E2 regions to promote splicing[48,50]. The functional RNA core of the spliceosome is postulated to be similar to that of a group II intron. Prp8 and the *O. iheyensis* group II intron have remarkably similar dimensions and the Prp8 active site cavity is approximately the right size to accommodate the essential RNA domains of group II intron RNA (Fig 5a and 5b). It is tempting to suggest that Prp8 has replaced the RNA scaffold surrounding the group II intron. Notably, the spliceosomal RNA catalytic core crosslinks to the region of Prp8 between RT and En domains (C.M. Norman and A.J.N., unpublished results) and to the RNaseH-like domain[18] (Fig 3b and 3c). The surface of this region exhibits extraordinary sequence conservation (Fig S18) and is remarkably electropositive (Fig 5c). This is consistent with its role as the binding site for the RNA catalytic core. Structural analysis of a group II intron at different stages of catalysis has revealed that the intron active site can adopt two alternative conformations[46,47]. It has been proposed that Prp8 may undergo a transition between two alternative states that facilitate the first and second steps of splicing, respectively[35]. This transition may be achieved by repositioning of the RNaseH-like domain and the extended polypeptide chains (Fig 3b and 3c), which line the inner surface of the active site cavity.

It has hitherto been hard to imagine how an RNP machine as immense and complex as the spliceosome could have evolved. The structure of Prp8 has given crucial insight into the active centre of the spliceosome and its similarity to group II intron encoded protein provides a compelling additional link between group II self-splicing and the spliceosome.

Supplementary Methods

Protein expression and purification

The Prp8$^{885-2413}$ fragment fused to an N-terminal Calmodulin binding peptide (CBP) and a full-length Aar2 with a C-terminal 8xHis-tag were cloned into pUC18 vectors containing the expression cassette described by Wagenbach et al.[51] These expression cassettes were transferred to pRS426 (Prp8) and pRS424 (Aar2) plasmids[52]. BCY123 cells (*MATa pep4::HIS3 prb1::LEU2 bar1::HIS6 lys2::GAL1/10GAL4 can1 ade2 trp1 ura3 his3 leu23,112*) harbouring both plasmids were grown on –URA –TRP selective medium with 1% raffinose to A_{600}=0.8-1.0. Protein expression was induced by addition of galactose to the final concentration of 2% and cells were grown at 30 °C for 12-16h. Cell pellets were re-suspended in one volume of 2×CAL350 buffer (700 mM NaCl, 100 mM Tris-Cl pH 8.5, 4 mM $CaCl_2$, 2 mM Mg acetate, 2 mM Imidazole, 20 mM 2-mercaptoethanol, 0.2% Igepal CA-630, EDTA-free protease inhibitor cocktail (Roche)) and frozen in liquid nitrogen. Solid phase cell disruption was performed with Freezer Mill 6870 (Spex) and the crude extract was adjusted to pH 8.0 with Tris base, then centrifuged at 48,000 × g at 4°C for 30min. The

supernatant was incubated with calmodulin-sepharose (recombinant calmodulin coupled to cyanogen bromide-activated sepharose (GE)) overnight at 4°C. Resin was washed with CAL500W buffer (500 mM NaCl, 20 mM Tris-Cl pH 8.0, 2 mM $CaCl_2$, 1 mM MgAc, 1 mM Imidazole, 10 mM 2-mercaptoethanol) and eluted with CAL500E buffer (500 mM NaCl, 20 mM Tris-Cl pH 8.0, 2 mM EGTA, 1 mM MgAc, 1 mM Imidazole, 10 mM 2-mercaptoethanol). After dialysis against Ni-dialysis buffer (500 mM NaCl, 20 mM Tris-Cl pH 8.0, 10 mM 2-mercaptoethanol, 5 mM imidazole) at 4°C, the sample was incubated with Ni-NTA agarose for 3-6 hours. The Ni-NTA agarose was packed into a small column, washed with Ni500W buffer (500 mM NaCl, 20 mM Tris-Cl pH 8.0, 20 mM imidazole, 10 mM 2-mercaptoethanol) and the protein was eluted with Ni500E buffer (500 mM NaCl, 20 mM Tris-Cl pH 8.0, 250 mM imidazole, 10 mM 2-mercaptoethanol). The eluate was dialysed against: 300 mM KCl, 20 mM K-HEPES pH 7.8, 1 mM DTT and diluted with 1/3 volume of: 20 mM K-HEPES pH 7.8, 1 mM DTT. The sample was applied to a MonoQ ion exchange column (10/100 GL) equilibrated with a 0.85:0.15 mixture of buffer A (50 mM KCl, 20 mM K-HEPES pH 7.8, 1 mM DTT) and buffer B (1 M KCl, 20 mM K-HEPES pH 7.8, 1 mM DTT) and eluted with a linear gradient of 15-50% of buffer B over 10 column volumes. Typically, 24 liter culture yielded in 4-8 mg of purified protein.

Crystallization

Crystals of the $Prp8^{885-2413}$-Aar2 complex were obtained by sitting-drop vapor diffusion technique at 293K. The protein solution (10-25 mg ml^{-1} in 300 mM KCl, 20 mM K-HEPES pH 7.8, 1 mM DTT) was mixed with an equal amount of reservoir solution (7-9% PEG 8000, 100 mM sodium citrate, 50-200 mM ammonium sulfate) and equilibrated against reservoir solution for 1h before streak-seeding with a feline whisker. Crystals suitable for data collection appeared within 1-3 days, reaching maximum dimensions of 0.6 × 0.2 × 0.1 mm^3. Cryo-protection was achieved in 3 steps by in-well buffer exchange with (1) 20% PEG 8000, 100 mM sodium citrate, for 5 min; (2) 30% PEG 8000, 100 mM sodium citrate, for 5 min and (3) 30% PEG 8000, 5% PEG 400, 100 mM sodium citrate, for 60 min. For mercury derivative crystals, the last step of cryo-protection was extended to 16h and the buffer was supplemented with 1 mM methyl mercury nitrate. Cryo-protected crystals were flash frozen in liquid nitrogen for data collection.

Crystals in both space groups (Table S1) appeared under the same condition. Initial diffraction of the mercury derivative crystals was limited to 8-10 Å, regardless of the concentration of the mercury compound, soaking time and which mercury compound was used. Close inspection of the 3SBT crystal structure29 revealed 2 cysteines in Aar2 (C251 and C292) that were in close proximity (~3.5 Å apart). Successful derivatisation of both cysteines would certainly lead to a steric clash and subsequent disintegration of the crystal lattice, which could explain the observed loss of diffraction. An Aar2 double mutant (C251S, C292S) was produced to overcome this limitation. Mutant protein crystallized under the same condition and the diffraction of the derivative crystals was improved from 8 to 3.7 Å.

Data collection and processing

Data for the $C222_1$ crystals was collected using an ADSC Q315R detector at ESRF (beamline ID23-1) at 0.91 Å wavelength with 0.1° oscillation range. Data for the $P2_12_12_1$ crystals were collected at Diamond Light Source (beamlines: I02, I03) with a Pilatus 6M detector. A mercury derivative SAD dataset was collected at the Hg LIII peak wavelength (1.00726 Å) with 0.2° oscillation range. Images were indexed and integrated in Mosflm[53] or XDS[54] and then scaled and analysed in Aimless[55].

Structure determination

An initial solution for the $C222_1$ crystals was obtained by molecular replacement with Phaser[56] and Molrep[57]. Search models were obtained from PDB: 2OG4 – Jab1/MPN domain; 3SBT – RNaseH-like domain and Aar2. Domains were located one by one in the following order: Aar2, RNaseH, Jab1/MPN. The solution was verified by a comparison with anomalous density peak positions in methyl mercury nitrate derivative data in the $P2_12_12_1$ space group (Figure S1). Molecular replacement solution was refined in Refmac5[58] (100 cycles of jelly-body refinement) and boosted by 50 cycles of automated model building/density modification in SHELXE[59]. The resulting improved phases were used to extend the existing model in ARP/wARP 7.3[60]. This procedure was repeated twice. Structure was manually rebuilt and modified in Coot[61] and refined in Refmac5. Structure validation was performed with Coot and Molprobity server[62].

The $P2_12_12_1$ crystal form was solved by molecular replacement with Phaser using the refined $C222_1$ structure as a search model. The density map from Phaser indicated that some regions of the Prp8$^{885\text{-}2413}$-Aar2 complex that were ordered in the $C222_1$ crystals were disordered in the $P2_12_12_1$ crystals and that other parts of the complex were orientated a little differently, particularly the Jab/MPN and RNaseH-like domains. The structure was modified by real-space rigid body refinement and manual rebuilding in Coot and refined in Refmac5[58].

Structure analysis

Structural similarity between the Prp8$^{885\text{-}2413}$-Aar2 complex and other proteins in PDB databases was assessed by secondary structure matching as implemented in PDBeFold web server[63]. Pairwise superpositions were made using DaliLite server[64]. Surface electrostatic potential was calculated with Adaptive Poisson-Boltzmann Solver[65] implemented in Pymol. Surface conservation analysis was performed with ConSurf web server[66]. Structure was visualized in Pymol (www.pymol.org). Multiple sequence alignment was prepared with ClustalW[67]. Secondary structure assignment was carried out using DSSP[68] and STRIDE[69].

Plasmid shuffling

Viability of the Prp8 mutants was assessed by plasmid shuffling method. Prp8 deletion mutant strain (SC261Δ8B1)[19], carrying wt PRP8 on pRS316 (URA, centromeric replication origin)[70] was transformed with mutant Prp8 on pRS314 (TRP, centromeric replication origin)[70] and transformants were selected on -TRP plates. Cells were transferred onto plates containing 5-fluoro-orotic acid (5-FOA), to test cell growth after loss of the URA plasmid. Viability was assessed visually after 3-5 days of incubation at 30°C. Plasmids from the 5-FOA resistant strains were rescued and sequenced to eliminate possibility of the observed phenotype being a result of a recombination event at the earlier stages of experiment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Garib Murshudov for his invaluable help and guidance with crystallography, Kelly Nguyen, Yasushi Kondo, Marike van Roon, John Hardin, Jade Li, Chris Norman, Antonina Andreeva, Alexey Murzin and Minmin Yu for discussion and help, Tijana Ignjatovic and Hiroyuki Oshikane for their essential contribution at the early stage of the project, Lori Passmore and Luca Jovine for critical reading of the manuscript, Mitsu Ikura for the gift of a calmodulin clone. We are grateful to the beamline staff at Diamond light source and European Synchrotron Radiation Facility for their help and Elaine Stephens and the LMB mass spec facility for their help. W.P.G. thanks the Cambridge European Trust and Downing College for scholarship. This project was funded by Medical Research Council of the UK.

References

1. Wahl MC, Will CL, Lührmann R. The spliceosome: design principles of a dynamic RNP machine. Cell. 2009; 136:701–718. [PubMed: 19239890]
2. Wassarman DA, Steitz JA. Interactions of small nuclear RNA's with precursor messenger RNA during in vitro splicing. Science. 1992; 257:1918–1925. [PubMed: 1411506]
3. Madhani HD, Guthrie C. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. Cell. 1992; 71:803–817. [PubMed: 1423631]
4. Kandels-Lewis S, Seraphin B. Involvement of U6 snRNA in 5′ splice site selection. Science. 1993; 262:2035–2039. [PubMed: 8266100]
5. Lesser CF, Guthrie C. Mutations in U6 snRNA that alter splice site specificity: Implications for the active site. Science. 1993; 262:1992–1998.
6. Sun JS, Manley JL. A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. Genes Dev. 1995; 9:843–854. [PubMed: 7705661]
7. Yean SL, Wuenschell G, Termini J, Lin RJ. Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. Nature. 2000; 408:881–884. [PubMed: 11130730]
8. Steitz TA, Steitz JA. A general two-metal-ion mechanism for catalytic RNA. Proc. Natl. Acad. Sci. USA. 1993; 90:6498–6502. [PubMed: 8341661]
9. Newman AJ, Norman C. U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. Cell. 1992; 68:743–754. [PubMed: 1739979]
10. Sontheimer EJ, Steitz JA. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. Science. 1993; 262:1989–1996. [PubMed: 8266094]
11. O'Keefe RT, Norman C, Newman AJ. The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast. Cell. 1996; 86:679–689. [PubMed: 8752221]
12. Achsel T, Ahrens K, Brahms H, Teigelkamp S, Lührmann R. The human U5-220kD protein (hPrp8) forms a stable RNA-free complex with several U5-specific proteins, including an unwindase, a homologue of ribosomal elongation factor EF-2, and a novel WD-40 protein. Mol. Cell. Biol. 1998; 18:6756–6766. [PubMed: 9774689]
13. Bartels C, Urlaub H, Lührmann R, Fabrizio P. Mutagenesis suggests several roles of Snu114p in pre-mRNA splicing. J. Biol. Chem. 2003; 278:28324–28334. [PubMed: 12736260]
14. Small EC, Leggett SR, Winans AA, Staley JP. The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. Mol. Cell. 2006; 23:389–399. [PubMed: 16885028]
15. Teigelkamp S, Newman AJ, Beggs JD. Extensive interactions of PRP8 protein with the 5′ and 3′ splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. EMBO J. 1995; 14:2602–2612. [PubMed: 7781612]
16. Dix I, Russell CS, O'Keefe RT, Newman AJ, Beggs JD. Protein-RNA interactions in the U5 snRNP of Saccharomyces cerevisiae. RNA. 1998; 4:1239–1250. [PubMed: 9769098]
17. Vidal VP, Verdone L, Mayes AE, Beggs JD. Characterization of U6 snRNA-protein interactions. RNA. 1999; 5:1470–1481. [PubMed: 10580475]
18. Reyes JL, Gustafson EH, Luo HR, Moore MJ, Konarska MM. The C-terminal region of hPrp8 interacts with the conserved GU dinucleotide at the 5′ splice site. RNA. 1999; 5:167–179. [PubMed: 10024169]
19. MacMillan AM, et al. Dynamic association of proteins with the pre-mRNA branch region. Genes Dev. 1994; 8:3008–3020. [PubMed: 8001820]
20. Turner IA, Norman CM, Churcher MJ, Newman AJ. Dissection of Prp8 protein defines multiple interactions with crucial RNA sequences in the catalytic core of the spliceosome. RNA. 2006; 12:375–386. [PubMed: 16431982]
21. Grainger RJ, Beggs JD. Prp8 protein: at the heart of the spliceosome. RNA. 2005; 11:533–557. [PubMed: 15840809]
22. Pena V, Rozov A, Fabrizio P, Lührmann R, Wahl MC. Structure and function of an RNase H domain at the heart of the spliceosome. EMBO J. 2008; 27:2929–2940. [PubMed: 18843295]

23. Ritchie DB, et al. Structural elucidation of a PRP8 core domain from the heart of the spliceosome. Nat Struct. Mol. Biol. 2008; 15:1199–1205. [PubMed: 18836455]
24. Yang K, Zhang L, Xu T, Heroux A, Zhao R. Crystal structure of the beta-finger domain of Prp8 reveals analogy to ribosomal proteins. Proc. Natl. Acad. Sci. USA. 2008; 105:13817–13822. [PubMed: 18779563]
25. Pena V, Liu S, Bujnicki JM, Lührmann R, Wahl MC. Structure of a multipartite protein-protein interaction domain in splicing factor prp8 and its link to retinitis pigmentosa. Mol. Cell. 2007; 25:615–624. [PubMed: 17317632]
26. Zhang L, et al. Crystal structure of the C-terminal domain of splicing factor Prp8 carrying retinitis pigmentosa mutants. Protein Sci. 2007; 16:1024–1031. [PubMed: 17473007]
27. Dlakić M, Mushegian A. Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase. RNA. 2011; 17:799–808. [PubMed: 21441348]
28. Boon KL, et al. Prp8 mutations that cause human retinitis pigmentosa lead to a U5 snRNP maturation defect in yeast. Nat. Struct. Mol. Biol. 2007; 14:1077–1083. [PubMed: 17934474]
29. Weber G, et al. Mechanism for Aar2p function as a U5 snRNP assembly factor. Genes Dev. 2011; 25:1601–1612. [PubMed: 21764848]
30. Joyce CM, Steitz TA. Function and structure relationships in DNA polymerase. Ann. Rev. Biochem. 1994; 63:777–822. [PubMed: 7526780]
31. Dias A, et al. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. Nature. 2009; 458:914–918. [PubMed: 19194459]
32. Yuan P, et al. Crystal structure of an avian influenza polymerase PAN reveals an endonuclease active site. Nature. 2009; 458:909–913. [PubMed: 19194458]
33. Wyatt JR, Sontheimer EJ, Steitz JA. Site-specific cross-linking of mammalian U5 snRNP to the 5′ splice site before the first step of pre-mRNA splicing. Genes Dev. 1992; 6:2542–2553. [PubMed: 1340469]
34. Urlaub H, Hartmuth K, Kostka S, Grelle G, Lührmann R. A general approach for identification of RNA-protein cross-linking sites within native human spliceosomal small nuclear ribonucleoproteins (snRNPs). Analysis of RNA-protein contacts in native U1 and U4/U6.U5 snRNPs. J. Biol. Chem. 2000; 275:41458–41468. [PubMed: 11006293]
35. Query CC, Konarska MM. Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants. Mol. Cell. 2004; 14:343–354. [PubMed: 15125837]
36. Umen JG, Guthrie C. Mutagenesis of the yeast gene PRP8 reveals domains governing the specificity and fidelity of 3′ splice site selection. Genetics. 1996; 143:723–739. [PubMed: 8725222]
37. Kuhn AN, Brow DA. Suppressors of a cold-sensitive mutation in yeast U4 RNA define five domains in the splicing factor Prp8 that influence spliceosome activation. Genetics. 2000; 155:1667–1682. [PubMed: 10924465]
38. Kuhn AN, Reichl EM, Brow DA. Distinct domains of splicing factor Prp8 mediate different aspects of spliceosome activation. Proc. Natl. Acad. Sci. USA. 2002; 99:9145–9149. [PubMed: 12087126]
39. Sharp PA. On the origin of RNA splicing and introns. Cell. 1985; 42:397–400. [PubMed: 2411416]
40. Cech TR. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. Cell. 1986; 44:207–210. [PubMed: 2417724]
41. Michel F, Umesono K, Ozeki H. Comparative and functional anatomy of group II catalytic introns – a review. Gene. 1989; 82:5–30. [PubMed: 2684776]
42. Sharp PA. Five easy pieces. Science. 1991; 254:663. [PubMed: 1948046]
43. Lambowitz AM, Zimmerly S. Mobile group II introns. Ann. Rev. Genet. 2004; 38:1–35. [PubMed: 15568970]
44. Pyle, AM.; Lambowitz, AM. Group II introns: ribozymes that splice RNA and invade DNA. In: Gesteland, RF.; Cech, TR.; Atkins, editors. The RNA world. 3rd ed.. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, N. Y.: 2006. p. 469-505.

45. Qui Y-L, Palmer JD. Many different origins of trans splicing in a plant mitochondrial group II intron. J. Mol. Evol. 2004; 59:80–89. [PubMed: 15383910]
46. Toor N, et al. Tertiary architecture of the Oceanobacillus iheyensis group II intron. RNA. 2010; 16:57–69. [PubMed: 19952115]
47. Mancia M, Pyle AM. Visualizing group II intron catalysis through the stages of splicing. Cell. 2012; 151:497–507. [PubMed: 23101623]
48. Matsuura K, Noah JW, Lambowitz AM. Mechanism of maturasepromoted group II intron splicing. EMBO J. 2001; 20:7259–7270. [PubMed: 11743002]
49. Rambo RP, Doudna JA. Assembly of an active group II intronmaturase complex by protein dimerization. Biochemistry. 2004; 43:6486–6497. [PubMed: 15157082]
50. Gu SQ, et al. Genetic identification of potential RNA-binding regions in a group II intron-encoded reverse transcriptase. RNA. 2010; 16:732–747. [PubMed: 20179150]
51. Wagenbach M, et al. Synthesis of wild type and mutant human hemoglobins in Saccharomyces cerevisiae. Biotechnology (N Y). 1991; 9:57–61. [PubMed: 1367213]
52. Christianson TW, Sikorski RS, Dante M, Shero JH, Hieter P. Multifunctional yeast high-copy-number shuttle vectors. Gene. 1992; 110:119–122. [PubMed: 1544568]
53. Leslie AGW, Powell HR. Processing Diffraction Data with Mosflm. Evolving Methods for Macromolecular Crystallography. 2007; 245:41–51.
54. Kabsch W. XDS. Acta Cryst. 2010; D66:125–132.
55. Evans PR. Scaling and assessment of data quality. Acta Cryst. 2006; D62:72–82.
56. McCoy AJ, et al. Phaser crystallographic software. J. Appl. Cryst. 2007; 40:658–674. [PubMed: 19461840]
57. Vagin A, Teplyakov A. Molecular replacement with MOLREP. Acta Cryst. 2010; D66:22–25.
58. Murshudov GN, Vagin AA, Dodson EJ. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. Acta Cryst. 1997; D53:240–255.
59. Sheldrick GM. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Cryst. 2010; D66:479–485.
60. Langer G, Cohen SX, Lamzin VS, Perrakis A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nature Protocols. 2008; 3:1171–1179.
61. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Cryst. 2004; D60:2126–2132.
62. Chen VB, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Cryst. 2010; D66:12–21.
63. Krissinel E, Henrick K. Secondary-structure matching (PDBeFold), a new tool for fast protein structure alignment in three dimensions. Acta Cryst. 2004; D60:2256–2268.
64. Holm L, Park J. DaliLite workbench for protein structure comparison. Bioinformatics. 2000; 16:566–567. [PubMed: 10980157]
65. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA. 2001; 98:10037–10041. [PubMed: 11517324]
66. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 2010; 38:529–33.
67. Larkin MA, et al. ClustalW and ClustalX version 2. Bioinformatics. 2007; 23:2947–2948. [PubMed: 17846036]
68. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983; 22:2577–2637. [PubMed: 6667333]
69. Heinig M, Frishman D. STRIDE: a Web server for secondary structure assignment from known atomic coordinates of proteins. Nucl. Acids Res. 2004; 32:500–502.
70. Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics. 1989; 122:19–27. [PubMed: 2659436]

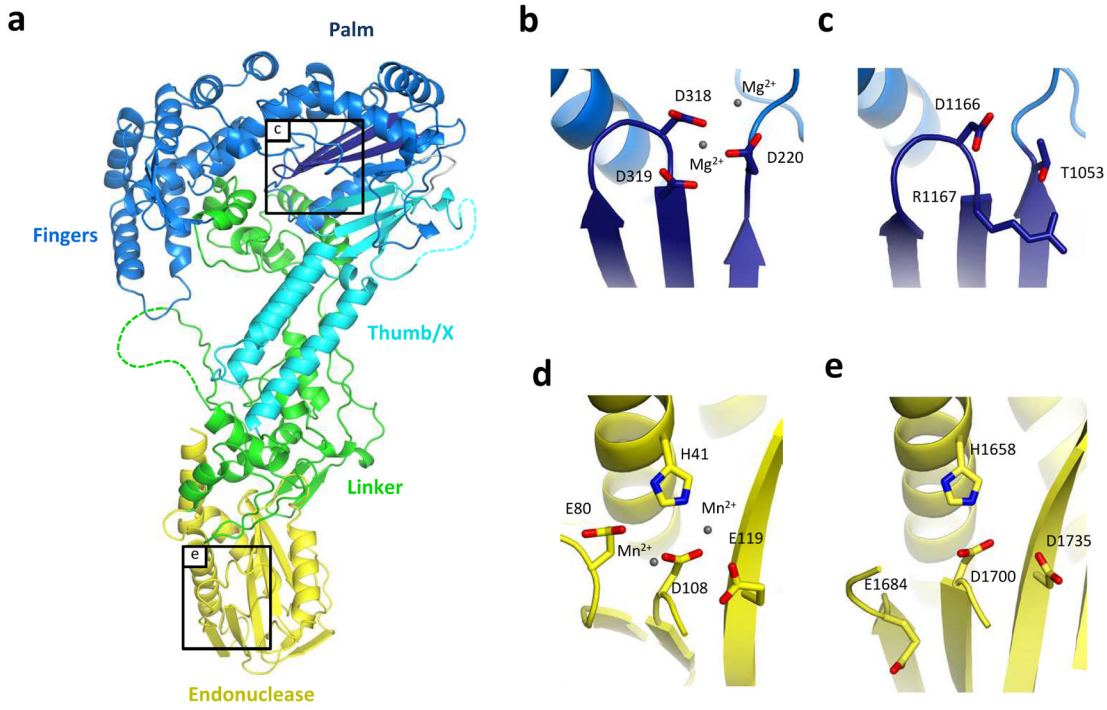

Figure 1. Structure of the large domain in yeast Prp8 (residues 885-1824)

a, The large domain consists of a group II intron reverse transcriptase-like domain and a type II restriction endonuclease-like domain; **b**, The palm sub-domain of HCV RNA-dependent RNA polymerase (1NB6). Asp residues (Asp220 in motif A, and Asp318 and Asp319 motif C) coordinate two catalytic Mg^{2+} ions; **c**, The corresponding residues in the palm sub-domain of the group II intron reverse transcriptase-like domain of Prp8; **d**, The catalytic centre of the influenza virus PA endonuclease (2W69). His-41, Glu-80, Asp-108 and Glu119 coordinate two catalytic divalent ions; **e**, The corresponding residues in the En domain of Prp8.

Europe PMC Funders Author Manuscripts

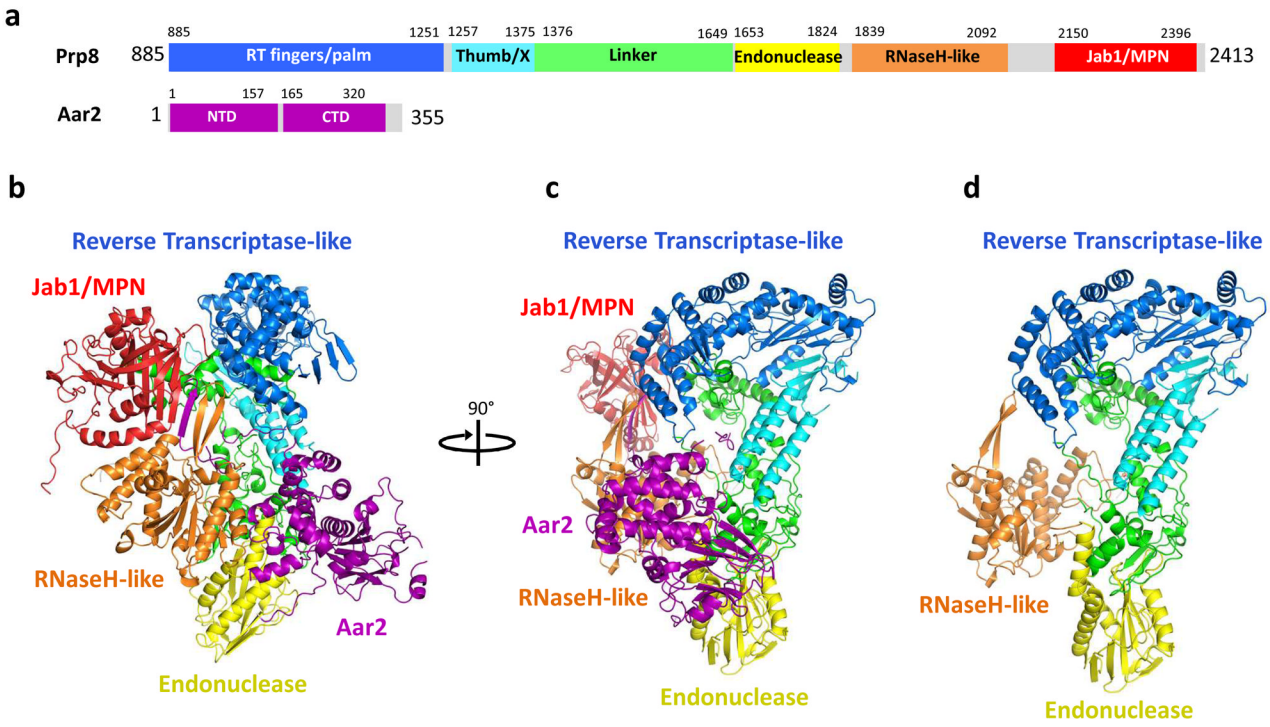

Figure 2. Overall structure of yeast Prp8$^{885\text{-}2413}$ in complex with Aar2

a, Domain architecture of Prp8$^{885\text{-}2413}$ and Aar2; **b**, Aar2 organises the arrangement of the RT/En, RNaseH-like and Jab1/MPN domains; **c**, Orthogonal view of the complex; **d**, A view (as in c) without Aar2 and the Jab1/MPN domain. The RNaseH-like domain has no direct contact with the RT/En domain.

Europe PMC Funders Author Manuscripts

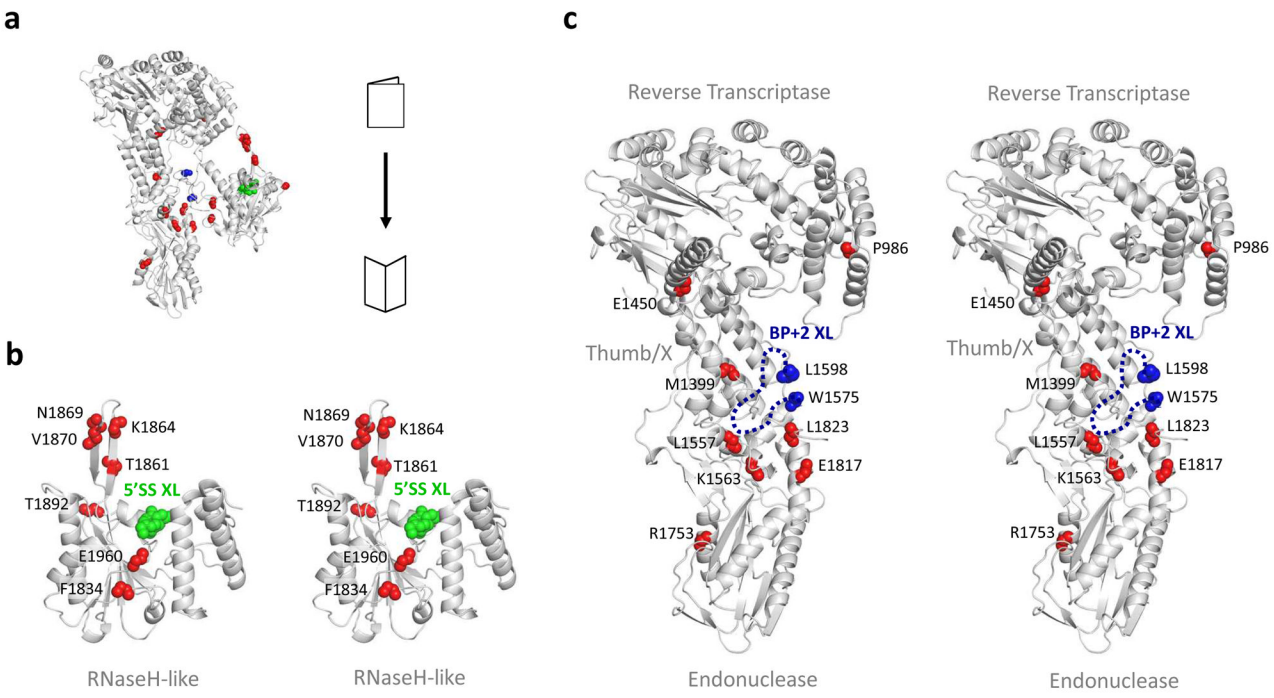

Figure 3. Overview of the Prp8 active site cavity in an open book view

a, Overview with the suppressors of splice site (5′SS, 3′SS and BP) mutations (red spheres). Green spheres indicate the sequence (1966-SAAMS-1970) corresponding to the cross-linking site of hPrp8 to the 5′ SS (ref 18); **b**, Stereo view of the RNaseH-like domain surface making up the active site cavity; **c**, Stereo view of the RT/En domain surface making up the active site cavity. Crosslink of the pre-mRNA branch point (BP+2) nucleotide is located between residues 1585-1598 in sequence (C.M. Norman and A.J.N., unpublished data). This site is found within the disordered loop (blue dotted line) between residues 1575 and 1598 (blue spheres).

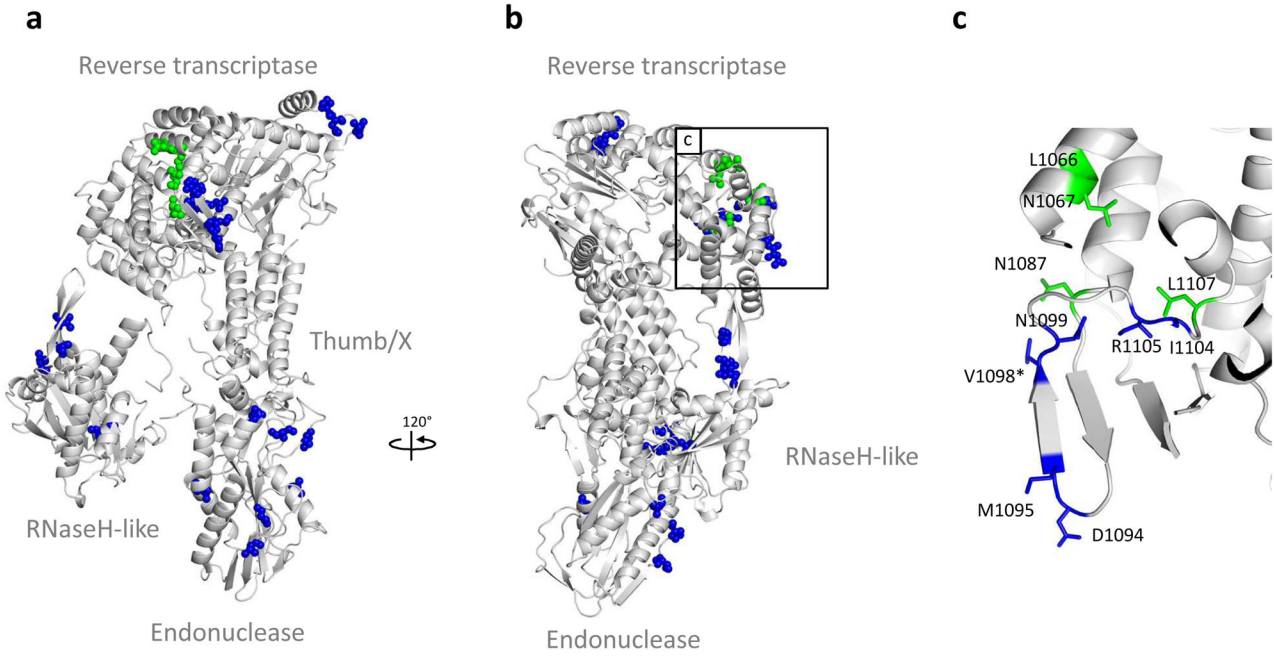

Figure 4. Suppressors of U4cs1 and *brr2-1* alleles mapped on the Prp8 structure

a, U4cs-1 (blue spheres) and *brr2-1* (green spheres) suppressor mutants map on one face of the RT-En domain of Prp8; **b**, A view rotated by 120° along y axis; **c**, Both types of suppressor mutants map to the same region of the Prp8 RT domain. Residues that suppress both alleles are marked with star (*).

Europe PMC Funders Author Manuscripts

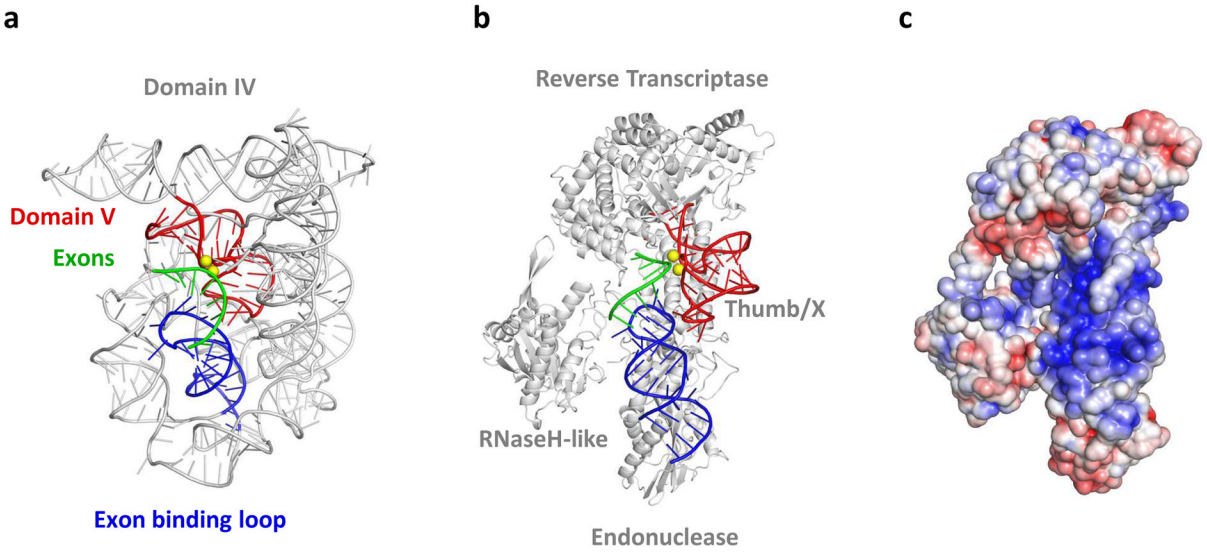

Figure 5. Comparison between the active site of group II intron and the spliceosome (Prp8)

a, Group II intron from *Oceanobacillus iheyensis* (PDB: 3IGI). Domain V, red; EBS helix (blue); spliced exons (green); Catalytic Mg^{2+} ions (yellow sphere); Scaffolding RNA (grey); **b**, The RT/En domain with the RNaseH-like domain of Prp8 with the active RNA elements of group II intron modeled on its surface for size comparison. At present there are insufficient experimental constraints for the precise position or orientation of the RNA; **c**, electrostatic potential ($\pm 5kTe^{-1}$) plotted on the solvent accessible surface of the Prp8 (calculated with APBS)[65].