

The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes

Inessa De¹, Sergey Bessonov², Romina Hofele³, Karine dos Santos⁴, Cindy L Will², Henning Urlaub^{3,5}, Reinhard Lührmann² & Vladimir Pena¹

Aquarius is a multifunctional putative RNA helicase that binds precursor-mRNA introns at a defined position. Here we report the crystal structure of human Aquarius, revealing a central RNA helicase core and several unique accessory domains, including an ARM-repeat domain. We show that Aquarius is integrated into spliceosomes as part of a pentameric intron-binding complex (IBC) that, together with the ARM domain, cross-links to U2 snRNP proteins within activated spliceosomes; this suggests that the latter aid in positioning Aquarius on the intron. Aquarius's ARM domain is essential for IBC formation, thus indicating that it has a key protein-protein-scaffolding role. Finally, we provide evidence that Aquarius is required for efficient precursor-mRNA splicing *in vitro*. Our findings highlight the remarkable structural adaptations of a helicase to achieve position-specific recruitment to a ribonucleoprotein complex and reveal a new building block of the human spliceosome.

The spliceosome is a dynamic ribonucleoprotein (RNP) machine whose conformation and composition change throughout its assembly and reaction cycle¹. Spliceosome assembly is initiated by binding of U1 and U2 small nuclear ribonucleoproteins (snRNPs) to the precursor mRNA (pre-mRNA) 5' splice site and branch site (BS), respectively, to yield the A complex. Subsequently, the U4/U6•U5 tri-snRNP joins, thus generating the precatalytic B complex. Spliceosome activation, yielding the activated (B^{act}) complex, involves major RNP rearrangements, including displacement of U1 and U4, and exchange of more than 30 proteins^{2–4}. The B^{act} complex is transformed into the catalytically activated (B^{*}) complex, which catalyzes step I of pre-mRNA splicing, to generate the C complex. After recruitment of additional factors, the spliceosome catalyzes step II, generating spliced mRNA.

The spliceosome's structural rearrangements are driven by eight conserved DExH/D-box RNA helicases that contain two RecA-like domains, which form a motor module required for ATP hydrolysis, RNA unwinding and coupling of these two processes². RNA helicases can have very diverse functions^{5–7}, including 'proofreading' specific steps of pre-mRNA splicing⁸. Most helicases also contain specific accessory domains important for their regulation or task-specific function⁶. The human spliceosome contains five additional RNA helicases (Aquarius, SF3b125, eIFAI1 and DDX35 and Abstrakt) whose function in splicing is unclear⁹. This suggests a more complex spliceosome assembly pathway in humans compared to yeast, potentially involving additional ATP-dependent remodeling steps.

Aquarius (also known as intron-binding protein 160, IBP160) is the only spliceosomal helicase belonging to the SF1 superfamily of

RNA helicases, whereas all others belong to SF2. The distinction between these superfamilies is based on the phylogenetic relationship between their RecA domains and the arrangement of their canonical sequence motifs¹⁰. Aquarius is stably integrated into spliceosomes during formation of the B^{act} complex, and it remains associated during all subsequent steps of splicing^{4,11}. It binds introns in a sequence-independent manner but interacts –33 to –40 nt upstream of the BS, i.e., directly adjacent to where the U2 snRNP SF3a and SF3b proteins are bound^{12,13}. However, Aquarius's function, if any, in pre-mRNA splicing is unclear.

Aquarius is crucial for the assembly of box C and box D (box C/D) small nucleolar RNPs (snoRNPs)¹³, whose small nucleolar RNAs (snoRNAs) are found in introns ~50 nt upstream of the BS, i.e., 10 nt upstream of Aquarius's binding site¹³. Although the mechanism whereby Aquarius aids in snoRNP maturation is unclear, it is thought to promote folding of the box-C/D snoRNA within the intron to facilitate the association of box-C/D snoRNP proteins¹³. Aquarius also appears to aid in recruitment of the exon junction complex to the spliceosome and its subsequent deposition on spliced mRNA¹⁴. Sequence analysis has indicated that some of Aquarius's domains share similarity with the RNA helicase Upf1 (ref. 15), which has an essential role in nonsense-mediated mRNA decay¹⁶. However, except for the two RecA domains, the sequence conservation of the remaining 70% of Aquarius is very poor, and it is difficult to reliably assess the fold that it adopts in the absence of a high-resolution structure.

We set out to determine the crystal structure of Aquarius and to subsequently analyze its enzymatic activities *in vitro*, its recruitment to the spliceosome and its role in pre-mRNA splicing. Our work

¹Macromolecular Crystallography, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany. ²Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany. ³Bioanalytical Mass Spectrometry, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany. ⁴Department of Structural Biochemistry, Free University, Berlin, Germany. ⁵Bioanalytics, Department of Clinical Chemistry, University Medical Center, Göttingen, Germany. Correspondence should be addressed to V.P. (vladimir.pena@mpi-bpc.mpg.de) or R.L. (reinhard.luehrmann@mpi-bpc.mpg.de).

Received 22 July 2014; accepted 11 December 2014; published online 19 January 2015; doi:10.1038/nsmb.2951

provides insight into how an RNA helicase is specifically targeted to its physiological location within an RNP complex and how it has structurally evolved to achieve this.

RESULTS

Crystal structure of Aquarius in complex with an ATP analog

To determine its structure, we produced recombinant human Aquarius in insect cells and purified it by several chromatographic steps (Supplementary Fig. 1a). The protein formed well-diffracting crystals in the presence of the nonhydrolyzable ATP analog AMP-PNP. We determined the structure by single-wavelength anomalous dispersion with a 2.3-Å data set collected from a selenomethionine (SeMet)-substituted crystal. The final model comprises residues 19–1381 and exhibits good stereochemistry (Table 1). The structure reveals a multidomain architecture organized in a relatively compact unit (Fig. 1a,b). According to structure-similarity searches, we divided Aquarius into two parts: one (designated the SF1 core) encompassing the center of the structure and sharing notable similarity to Upf1 (Supplementary Fig. 2a), the prototype of SF1 RNA helicases, and another (denoted Aquarius specific) consisting of three domains not previously observed in other helicases (Fig. 1a, Supplementary Fig. 2b and Supplementary Fig. 3).

The SF1 core of Aquarius consists of two RecA-like domains, an α -helical stalk and a β -barrel domain (Fig. 1a,b and Supplementary Fig. 2a). The canonical sequence motifs of RNA helicases are conserved (Fig. 1c and Supplementary Fig. 3), and the ATP analog binds with the adenine sandwiched between Leu1157 and Ile800 (Fig. 1d). The amino group of the adenine is recognized by Gln806 from the Q motif (Fig. 1d)—a distinctive feature of helicases able to use only ATP¹⁷.

The space between the RecA1, RecA2, β -barrel and a domain that we denote as an armadillo (ARM) domain is filled by the so-called stalk (Fig. 1a and Supplementary Fig. 2c)—a structural unit encompassing 125 residues (416–492 and 660–710) exhibiting helices, extensive coils and loops. The large surfaces that the stalk shares with other domains (2,200 Å² with RecA1, 1,400 Å² with the β -barrel and 1,230 Å² with the ARM) suggest that it has a key architectural role (Fig. 1a and Supplementary Fig. 2c). Notably, numerous large hydrophobic residues that mediate these contacts are highly conserved (Supplementary Fig. 3), thus suggesting a similar orientation of these domains among Aquarius's orthologs. The β -barrel domain (493–661; Fig. 1a,b) is located on top of the stalk and has several insertions that radiate toward the periphery, to give the appearance of a propeller (Supplementary Fig. 2d).

Aquarius also possesses three domains not found in other helicases; these domains comprise about half of the protein, and their folding could not be predicted from their sequence. They include a large N-terminal domain partially composed of helical repeats from the armadillo family (16–416), a globular domain (681–804) situated between the stalk and RecA1, and an elongated helical bundle protruding from the RecA1 domain (881–1034); these are denoted the ARM, thumb and pointer, respectively, and the latter two resemble the corresponding parts of a left hand (Fig. 1a,b).

Aquarius's ARM domain has the appearance of an elongated and straight body (50 × 84 Å) that interacts extensively with the RecA domains and the stalk, sharing interfaces of 600 Å² and 1,234 Å², respectively. The first three-quarters of the ARM domain (16–310) contains five ARM repeats similar to those of β -catenin (Supplementary Fig. 2b), whereas the remaining quarter (311–416) contains helices that are not organized as ARM repeats (Supplementary Fig. 2b). Another distinctive feature of Aquarius's ARM domain is a protrusion located between the third and fourth ARM repeat, which contains three α -helices (Supplementary Fig. 2b).

Table 1 Data collection and refinement statistics

	Native	SeMet
Data collection		
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	95.1, 140.6, 145.4	94.9, 140.4, 144.8
Wavelength	1.0	0.98
Resolution (Å)	50.0–2.6 (2.6)	47.5–2.3 (2.3)
<i>R</i> _{sym}	5.2 (35.6)	9.3 (54.3)
<i>I</i> / σ <i>I</i>	25.6 (5.0)	18.6 (4.2)
Completeness (%)	100 (100)	100 (100)
Redundancy	6.7 (6.8)	7.8 (7.8)
Refinement		
Resolution (Å)		47.5–2.3
No. reflections		84,028
<i>R</i> _{work} / <i>R</i> _{free}		19.0 / 22.6
No. atoms		
Protein		11,598
Ligand		1
Water		555
<i>B</i> factors		
Protein		36.4
Ligand		30.1
Water		34.5
r.m.s. deviations		
Bond lengths (Å)		0.003
Bond angles (°)		0.739

One crystal was used for each data set. Values in parentheses are for the highest-resolution shell.

The thumb and pointer domains are located on the side opposite the ARM domain, and they do not show similarity to other proteins (Fig. 1a). The thumb (681–804) adopts a globular fold composed of a four-stranded β -sheet packed against an α -helix. The pointer (881–1034) exhibits an elongated and arched shape, mainly defined by three long helices, two of which run parallel to each other and form an angle of about 120° with the third (Fig. 1a).

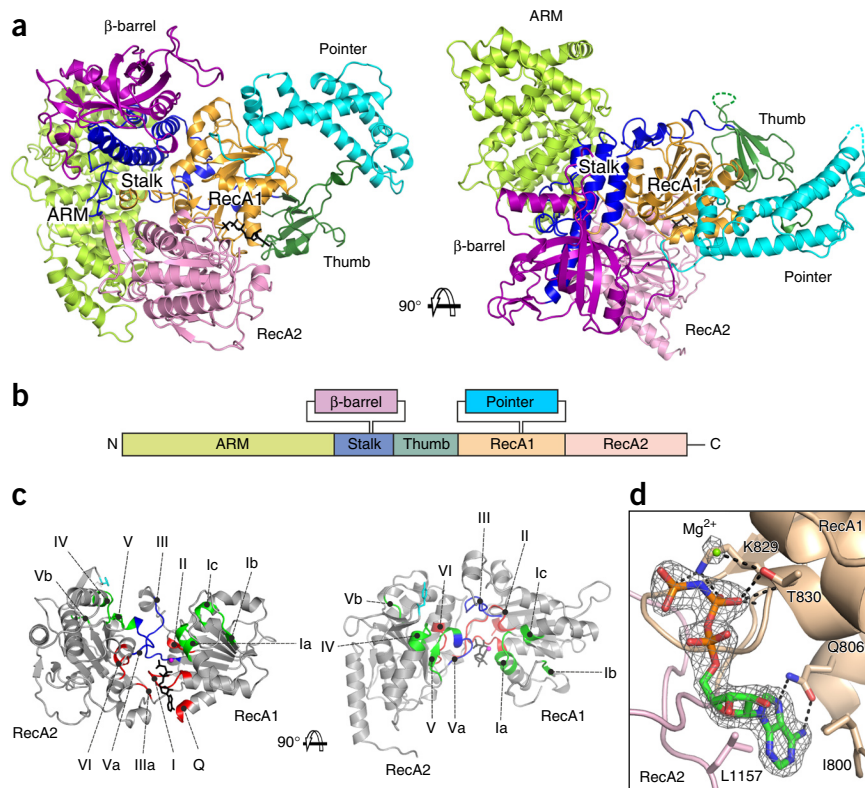
Aquarius is present in a pentameric complex

ARM domains are well-known protein-protein scaffolds, and therefore we next searched for interaction partners of Aquarius that could explain the need for such a large protein-protein-interaction domain. Aquarius was previously detected in a complex (termed XAB2) together with several other spliceosomal proteins¹⁸, but it remained unclear whether most of Aquarius is present in this form. We therefore immunoaffinity purified Aquarius-containing complexes from HeLa nuclear extract and analyzed them on a glycerol gradient (Fig. 2a). Four spliceosomal proteins—hSyf1 (also called Xab2), CCDC16, hIsy1 and CypE—comigrated together with Aquarius in fractions 14–16, results indicating that they form an ~11S complex. Importantly, we detected very little Aquarius in other gradient fractions. Prp19, an XAB2 component¹⁸, was neither detected by MS in our purified complexes (Fig. 2a) nor coimmunodepleted from extract together with Aquarius, as observed for hSyf1, CCDC16, hIsy1 and CypE (Fig. 2b); this is possibly because of our more stringent purification conditions. Thus, nearly all Aquarius in HeLa nuclear extract was present in a pentameric complex, henceforth designated the intron-binding complex (IBC).

Aquarius is integrated into the spliceosome as part of the IBC

To determine whether Aquarius is recruited to the spliceosome as part of the IBC, we isolated recombinant IBC from insect cells. After

Figure 1 Structure of Aquarius in complex with AMP-PNP. (a) Ribbon representation of human Aquarius bound to AMP-PNP (black sticks) in two perpendicular orientations. (b) Schematic of Aquarius's domain arrangement. (c) Conserved sequence motifs of the RecA domains of Aquarius. The conserved sequence motifs (Q and Roman numerals) are colored according to their predominant biochemical function: red, binding and hydrolysis of ATP; green, nucleic acid binding; blue, coordination between the nucleic acid- and ATP-binding sites. The Y1196 residue of Aquarius is shown in sticks and is colored cyan. (d) Detailed view of the AMP-PNP (black sticks)-binding site showing conserved Aquarius residues that recognize the phosphates (K829), magnesium ion (T830) and adenine base (Q806) sandwiched between L1157 and I800. The electron density map ($2F_o - F_c$, gray) covering the AMP-PNP and Mg^{2+} ion is contoured at the 1.7σ level. (Additional data are in **Supplementary Figs. 2 and 3.**)



purification by several chromatographic steps, it eluted as a single peak during size-exclusion chromatography (**Fig. 2c** and **Supplementary Fig. 1b**). The molar mass of the recombinant IBC determined by multiangle light scattering was 378 kDa, compared to a theoretical mass of 384 kDa (**Supplementary Fig. 1c**), thus demonstrating that it contains only one copy of each protein. We then added recombinant histidine-tagged Aquarius either alone or as part of the IBC to a splicing reaction at an ~15-fold excess over the endogenous protein or complex. We then affinity purified spliceosomes formed on MS2-tagged PM5 pre-mRNA, which lacks a 3' exon and stalls assembly at the C-complex stage¹¹. Immunoblotting with anti-histidine antibodies revealed that Aquarius was efficiently integrated into spliceosomes only when added as part of the IBC (**Fig. 2d**, lanes 2 and 3), thus indicating that the latter is a new, preformed spliceosome building block.

The ARM domain is essential for IBC formation

Because ARM repeats are often found in protein scaffolds, the ARM domain of Aquarius might have a key architectural role within the IBC. To test this, we expressed a truncated form of Aquarius lacking the

ARM domain (Δ ARM, residues 417–1485; **Supplementary Fig. 4a**) and prepared an insect-cell lysate containing it and the other four IBC proteins. Affinity selection with Ni-NTA beads and subsequent size-exclusion chromatography revealed that Δ ARM Aquarius does not support formation of the IBC (**Fig. 3a**). Hence, the ARM domain is crucial for the structural integrity of the IBC, thus underscoring its pivotal role as a protein-protein-interaction scaffold.

To identify interaction partners of the ARM domain and other regions of Aquarius within the IBC, we performed protein-protein cross-linking with the homobifunctional cross-linking reagent BS3. MS detected multiple cross-links between Aquarius and hSyf1, hlsy1, CCDC16 or CypE, with the majority of the cross-linked residues located in domains or structural insertions specific for Aquarius, such as the ARM, pointer and thumb domains, and in the large insertions of the β -barrel (**Fig. 3b**, **Supplementary Data Set 1** and **Supplementary Table 1**).

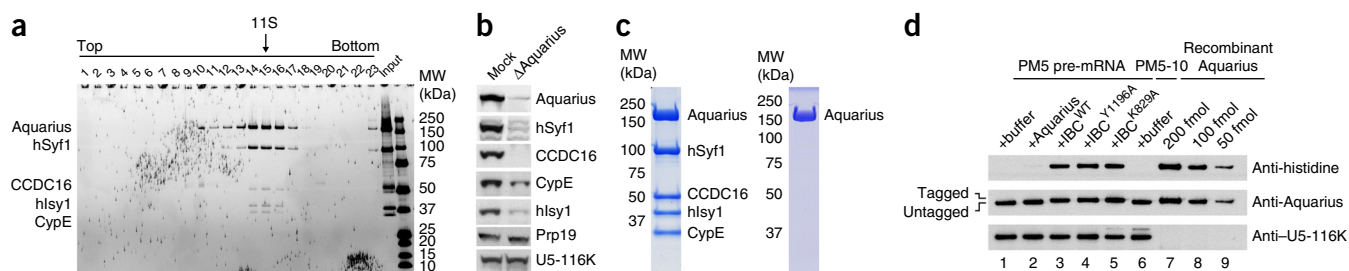


Figure 2 Aquarius is preassembled in a pentameric IBC complex. (a) Glycerol gradient fractionation of Aquarius and associated proteins affinity purified from HeLa nuclear extract. SDS-PAGE of proteins in the indicated gradient fractions with identities shown on the left. Molecular-weight markers (MW) are indicated on the right. (b) Western blotting analysis of nuclear extract depleted (Δ) of Aquarius. Immunostaining of the indicated proteins is shown. (c) SDS-PAGE showing the purified recombinant IBC (left) and Aquarius (right). Molecular-weight markers are indicated on the left. (d) Western blotting analysis of the affinity-purified spliceosomes with anti-histidine and anti-Aquarius antibodies. Samples are spliceosomal complexes (100 fmol, lanes 1–6) assembled on the PM5 pre-mRNA in nuclear extract in the absence or presence of Aquarius alone or the wild-type (WT) recombinant IBC or IBC with Aquarius K829A or Y1196A, as indicated. B^{act} complexes and 50–200 fmol of recombinant Aquarius (lanes 7–9) are controls; U5-116K is a loading control. (Additional data are in **Supplementary Fig. 1**. Uncropped blot images are shown in **Supplementary Data Set 4.**)

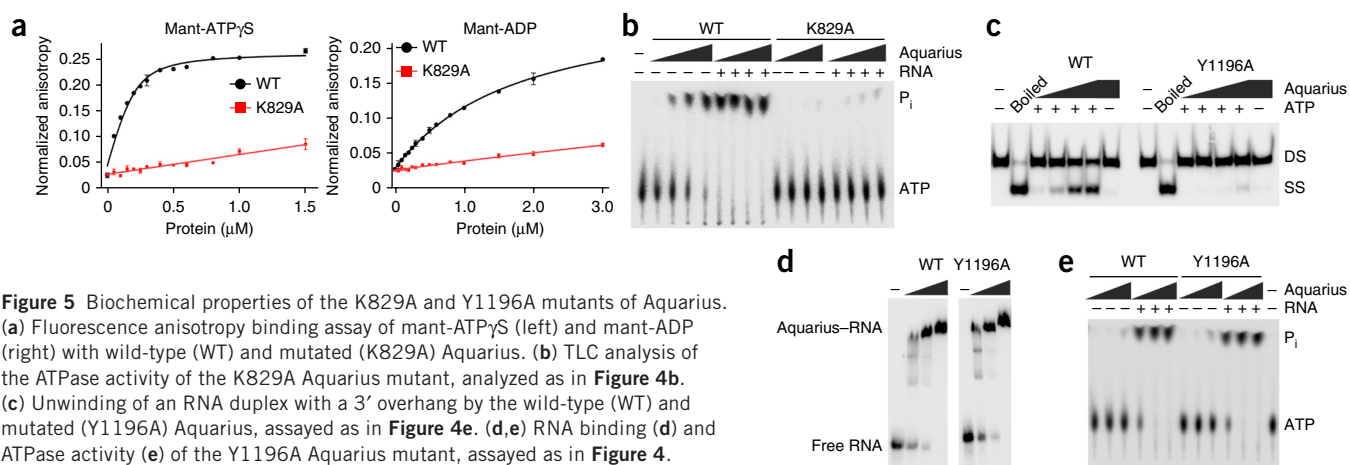


Figure 5 Biochemical properties of the K829A and Y1196A mutants of Aquarius. (a) Fluorescence anisotropy binding assay of mant-ATP γ S (left) and mant-ADP (right) with wild-type (WT) and mutated (K829A) Aquarius. (b) TLC analysis of the ATPase activity of the K829A Aquarius mutant, analyzed as in **Figure 4b**. (c) Unwinding of an RNA duplex with a 3' overhang by the wild-type (WT) and mutated (Y1196A) Aquarius, assayed as in **Figure 4e**. (d, e) RNA binding (d) and ATPase activity (e) of the Y1196A Aquarius mutant, assayed as in **Figure 4**.

or two blunt ends (**Fig. 4e**). Aquarius was unable to unwind RNA duplexes with no single-stranded overhang or with a 5' overhang, but it did unwind the model RNA duplex with a 3' overhang, thus demonstrating that it possesses 3'-to-5' RNA-unwinding activity (**Fig. 4e**). We observed the latter in the presence of ATP but not with AMP-PNP or ADP, results indicating that ATP hydrolysis is required for strand separation (**Fig. 4f**). Deletion of the ARM domain had only a modest effect on the ability of Aquarius to hydrolyze ATP, to bind RNA or to unwind RNA duplexes (**Supplementary Fig. 4b**), thus suggesting that the ARM domain only mildly affects Aquarius's RecA functions. The IBC also exhibited similar levels of RNA-stimulated ATPase and 3'-to-5' RNA-unwinding activity (**Fig. 4g,h**) and did not bind blunt-ended dsRNA (**Fig. 4c**). Because hSyf1, hIsy1, CypE and CCDC16 do not possess ATPase or RNA helicase motifs, the observed IBC activities can be ascribed solely to Aquarius. Thus, other IBC proteins do

not substantially alter Aquarius's enzymatic activities or change the directionality of its RNA-unwinding activity. However, we cannot rule out that under splicing conditions or in the context of the spliceosome the IBC may have more pronounced effects on Aquarius's RNA binding and/or enzymatic activities.

An ATPase-deficient Aquarius mutant hinders splicing

Immunodepletion of Aquarius from nuclear extract only moderately reduced splicing of PM5 pre-mRNA (~50% less step I), and complementation with recombinant wild-type IBC only partially restored splicing activity (**Supplementary Fig. 6**). Thus, as an alternative to test whether Aquarius is required for pre-mRNA splicing, we isolated recombinant IBC complexes containing histidine-tagged Aquarius with Lys829 substituted with alanine (K829A) or Tyr1196 substituted with alanine (Y1196A). Mutation of the invariant lysine residue (K829A)

from motif I involved in ATP binding and hydrolysis drastically reduced Aquarius's ability to bind ATP and ADP and to hydrolyze ATP (**Fig. 5a,b**). Mutation of a conserved aromatic residue (Y1196A) located near the putative RNA-binding surface of the RecA2

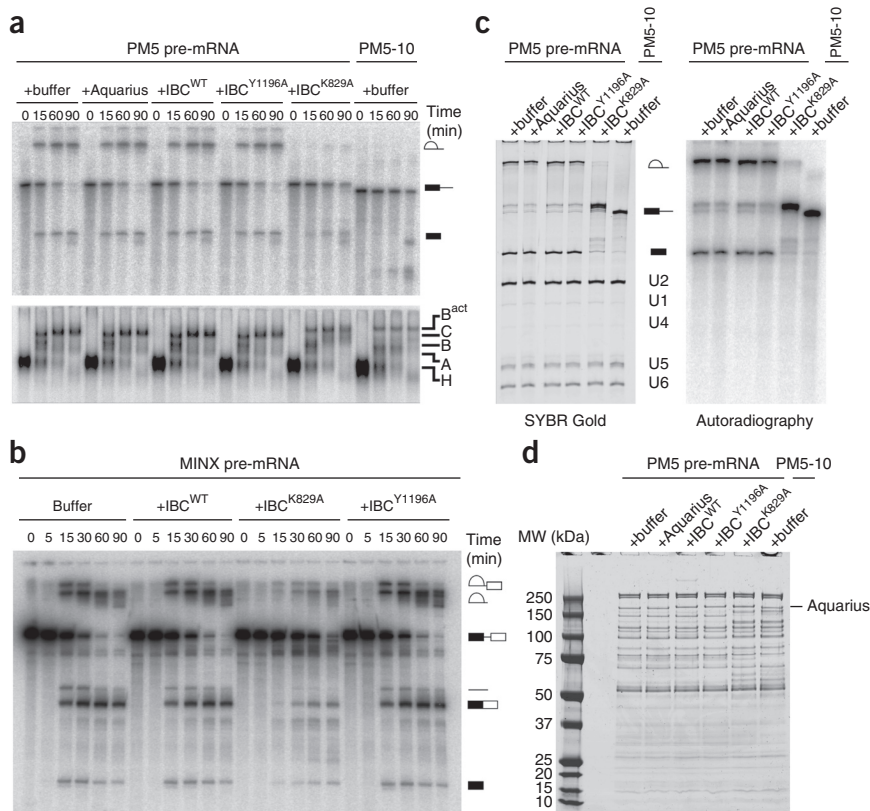


Figure 6 ATP hydrolysis by Aquarius is required for efficient pre-mRNA splicing. (a) Denaturing PAGE (top) and native agarose gel (bottom) analysis of splicing of PM5 pre-mRNA in the absence (buffer) or presence of recombinant Aquarius or the IBC containing either wild-type Aquarius (WT) or an ATPase (K829A)- or helicase (Y1196A)-deficient mutant. Pre-mRNA and splicing intermediates or spliceosomal complexes are indicated on the right. To better visualize the B^{act} complex, we also performed splicing with PM5-10 pre-mRNA. (b) Denaturing PAGE analysis of splicing of MINX pre-mRNA, analyzed as in a. Splicing intermediates and products are indicated on the right. (c) Denaturing PAGE showing RNA composition of affinity-purified spliceosomes assembled as in a. RNA is visualized by SYBR Gold staining (left) and autoradiography (right). The positions of the snRNAs, pre-mRNA and splicing intermediates are indicated in the middle. (d) SDS-PAGE showing the protein composition of affinity-purified spliceosomes assembled as in a. Molecular weight markers (MW) are indicated on the left. (Additional data are in **Supplementary Fig. 6**.)

domain (Fig. 1c) inhibited Aquarius's RNA-unwinding activity without changing its RNA-binding and ATPase properties (Fig. 5c–e).

Both mutant complexes exhibited the same protein stoichiometry and chromatographic behavior as recombinant wild-type IBC, thus indicating that these single amino acid changes did not affect IBC formation. We added an ~15-fold excess of wild-type or mutant IBC to *in vitro* splicing reactions containing PM5 pre-mRNA and then affinity purified the spliceosomal complexes. Immunoblotting with anti-histidine antibodies or with antibodies recognizing endogenous and recombinant Aquarius revealed that the majority of Aquarius in spliceosomes was histidine tagged (on the basis of comparison of the signals obtained with known amounts of histidine-tagged Aquarius) (Fig. 2d, lanes 4, 5 and 7–9). Thus, the mutant IBCs are efficiently incorporated into spliceosomes, and neither ATPase nor helicase activity is required for IBC integration.

Addition of the IBC containing Aquarius^{Y1196A} or recombinant wild-type IBC had no effect on the splicing of PM5 pre-mRNA, which undergoes only step I (Fig. 6a), or MINX pre-mRNA, which undergoes both steps of splicing (Fig. 6b). Thus, Aquarius's RNA-unwinding activity does not appear to be required for splicing *in vitro*. Both steps of splicing were strongly reduced when IBC with Aquarius^{K829A} was added (Fig. 6a,b), thus indicating that this mutant acts in a dominant-negative fashion. In the presence of Aquarius^{K829A}, spliceosomal complexes migrating similarly to B^{act} accumulated (Fig. 6a). Affinity purification of these Aquarius^{K829A} spliceosomes and analysis of their RNA and protein composition (Fig. 6c,d and Supplementary Data Set 3) indicated that they are stalled after B^{act} formation but before formation of the step I spliceosome (C complex), because they (i) contained predominantly unspliced pre-mRNA and just the U2, U5 and U6 snRNAs (Fig. 6b), (ii) lacked all U4/U6 snRNP proteins and (iii) contained U2 snRNP proteins and all other proteins typically found in B^{act} complexes, including the IBC proteins and Prp2, in abundances similar to those in wild-type B^{act} spliceosomes (Supplementary Data Set 3). Together, these results suggest that for efficient splicing in higher eukaryotes there is an additional requirement for an ATPase before step I.

DISCUSSION

Structural divergence of evolutionarily related helicases

Aquarius possesses several domains not observed in other RNA helicases, grafted onto a central Upf1-like core, denoted SF1, which contains the two RecA, stalk and β -barrel domains (Fig. 1a,b). The SF1 core covers ~50% of Aquarius's size and is substantially larger than the equivalent region in Upf1, because of the presence of much larger insertions within the stalk and β -barrel (Supplementary Fig. 2a,c,d and Supplementary Fig. 3). The pointer and the thumb domains do not bear any clear resemblance to other known structures and appear to be unique to Aquarius (Fig. 1a,b). However, Aquarius's pointer and Upf1's 1C domain have an equivalent location at the primary sequence level, thus suggesting that they have diverged from a common fold as an adaptation to fulfill distinct functions. Indeed, the pointer and/or other Aquarius-specific domains are prime candidates for reversing the RNA-unwinding polarity of Aquarius relative to that of Upf1 (i.e., 3'-5' versus 5'-3' for Upf1). Cross-linking analyses suggest that the pointer and thumb domains, and also Aquarius-specific insertions in the β -barrel, are involved in interactions within the IBC (Fig. 3b) and thus might have also arisen during evolution as adaptors for protein-protein interactions.

An ARM domain as an adaptation for recruitment

Like all structural motifs that form helical solenoids, Armadillo repeats are found in protein scaffolds, and most often such proteins lack additional large domains^{19,20}. Aquarius is the only known helicase

to possess a scaffold-like domain that contains ARM repeats (Fig. 1a and Supplementary Fig. 2b). Moreover, we present the first crystal structure, to our knowledge, of an enzyme that exhibits a helical solenoid domain. Deletion of the ARM domain had little effect on the RNA-binding or ATPase and RNA-unwinding activities of Aquarius (Supplementary Fig. 4b), thus indicating that it does not regulate Aquarius's RecA functions. Instead, it acts as a protein-protein scaffold of the IBC (Fig. 3a). Moreover, the ARM domain appears to help guide Aquarius to the correct position on the intron via its interactions with SF3a and SF3b proteins within the spliceosome (Fig. 3c). Our studies provide a fascinating example of how an RNA helicase is specifically targeted to its functionally important location within an RNP complex and how it has structurally evolved to achieve this.

The IBC is a new spliceosome building block

Our data reveal that Aquarius is recruited to the spliceosome solely as a pentameric complex (the IBC) that also contains hSyf1, hIsy1, CypE and CCDC16 (Fig. 2). Thus, the IBC is an important spliceosome building block that ensures simultaneous recruitment of its components during B^{act}-complex formation. In yeast, orthologs of two IBC proteins, hSyf1 and hIsy1, are conserved components of the so-called NineTeen complex (NTC), which acts as a scaffold for the recruitment of B^{act} proteins and stabilizes the rearranged RNA network in the B^{act} spliceosome^{21,22}. In *Saccharomyces cerevisiae*, all NTC proteins, including Syf1 and Isy1, are recruited together to the spliceosome as a preassembled building block^{9,23}. In contrast, human NTC proteins are recruited from separate building blocks, including the IBC (containing hIsy1 and hSyf1) and the hPrp19–CDC5L complex (containing homologs of five other yeast NTC proteins²⁴). Because numerous proteins are exchanged during the B-to-B^{act} transition in humans^{3,9}, this transition might involve multiple steps, including sequential addition of NTC proteins. This stepwise handover of proteins would allow for additional points of regulation during spliceosome activation in higher eukaryotes.

IBC proteins facilitate positioning of Aquarius

Most RNA helicases do not exhibit RNA sequence specificity, and thus much remains to be learned about how helicases are targeted to their physiological substrates. Because Aquarius binds introns in a position-specific manner, contacts that it or the IBC makes with other spliceosomal proteins interacting near the BS are likely to have important roles in guiding Aquarius to its proper position on the intron. Cross-linking studies of B^{act} complexes (Fig. 3c and Supplementary Table 2) revealed multiple cross-links between IBC and SF3a and SF3b proteins, which contact nucleotides just upstream of the BS, adjacent to Aquarius's binding site, consistently with SF3a and SF3b playing a central part in recruiting the IBC to the spliceosome and ensuring proper positioning of Aquarius on the intron, as previously proposed¹³. Further, the U2-SF3b130 and Isy1 cross-link is consistent with recent data implicating Isy1 in proofreading the pre-mRNA's BS in yeast²⁵. Because the IBC binds much later than SF3 during spliceosome assembly, formation of the IBC binding site must involve other spliceosomal proteins and/or RNP rearrangements after SF3 binding. Indeed, we also observed cross-links between the IBC and proteins that join the spliceosome during or after B-complex formation, such as Prp8 and Prp19–CDC5 complex proteins, thus suggesting that they also contribute to IBC recruitment.

Aquarius is required for efficient pre-mRNA splicing

Aquarius plays an important part in linking splicing to other RNA-processing events, such as snoRNA formation and deposition of the exon junction complex^{13,14}. However, Aquarius's role in splicing has remained unclear. RNA-interference knockdown of Aquarius¹³

or immunodepletion of Aquarius from HeLa nuclear extract (Supplementary Fig. 6) only moderately inhibited splicing *in vitro*, thus suggesting that Aquarius is not essential but has a role in splicing, e.g., by potentially enhancing splicing efficiency. However, these moderate effects could potentially be due to sufficient amounts of Aquarius remaining after RNA interference and immunodepletion, because there is an overabundance of some splicing factors in the cell²⁶.

We show that Aquarius exhibits RNA-stimulated ATPase and dsRNA-unwinding activity, both alone or as part of the IBC (Fig. 4). These enzymatic activities were required for neither the formation of the IBC nor its incorporation into spliceosomes (Fig. 2d), and its RNA helicase activity did not appear to be required for pre-mRNA splicing *in vitro* (Fig. 6 and Supplementary Fig. 6). However, we cannot exclude that Aquarius's helicase activity is required for spliceosome disassembly (which is not analyzed by our *in vitro* splicing assay) or for splicing a subset of pre-mRNAs with secondary structures that require unwinding to allow proper accommodation of the IBC on the intron. Furthermore, Aquarius's 3'-to-5' unwinding polarity and downstream localization relative to snoRNAs in some introns are consistent with the use of its motor activity to disrupt base-pairing or to displace bound proteins to allow proper folding of snoRNAs, as previously proposed¹³.

An IBC-containing ATPase-deficient Aquarius (K829A) blocked pre-mRNA splicing before step I, to result in accumulation of B^{act}-like spliceosomes (Fig. 6). Our data thus clearly demonstrate that Aquarius has a role in splicing, and they specifically suggest that its ATPase activity is required after B^{act}-complex formation but before splicing catalysis. However, because Aquarius^{K829A} was also deficient in binding ATP and ADP, we cannot rigorously exclude that ATP or ADP binding by Aquarius as opposed to ATP hydrolysis is required for efficient catalytic activation and step I of splicing. Conceivably, ATP binding and hydrolysis or nucleotide-product release could change the conformation of Aquarius, leading to rearrangements in, or release of, its interaction partners and/or to a remodeled structure that supports catalytic activation of the spliceosome. Given that Aquarius extensively contacts SF3a and SF3b proteins within B^{act} spliceosomes (Fig. 3c), and RNP interactions involving SF3a and SF3b are remodeled during catalytic activation^{3,27}, Aquarius may facilitate this remodeling process. However, future studies are needed to dissect Aquarius's exact role in splicing. Nonetheless, our data support the idea that spliceosome assembly in humans is more complex compared to that in yeast, and it may involve additional ATP-dependent remodeling steps.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. Coordinates and structure factors have been deposited in the Protein Data Bank under accession code PDB 4PJ3.

Note: Any Supplementary Information and Source Data files are available in the [online version of the paper](#).

ACKNOWLEDGMENTS

We are grateful to M. Raabe for assisting with peptide sequencing; K. Gencalp for help with multiangle light scattering measurements; A. Draycheva and M. Thommen for help with fluorescence anisotropy experiments; T.R. de Moura, J. Schmitzová, B. Kastner and K. Hartmuth for advice and helpful discussions; and the teams of beamlines 14.2 (BESSY, Berlin, Germany) and PXII (SLS, Villigen, Switzerland) for support during diffraction data collection.

AUTHOR CONTRIBUTIONS

The crystallographic research, purification and biochemical characterization of recombinant proteins and the IBC were performed by I.D. under the supervision

of V.P.; the biochemical experiments involving analysis of the IBC components in nuclear extracts, *in vitro* splicing, purification and analysis of spliceosomes were performed by S.B. under the supervision of R.L.; cross-linking experiments and MS analyses were performed by R.H. under the supervision of H.U. K.d.S. and C.L.W. participated in recombinant expression of Aquarius and in experiments with purified spliceosomes, respectively. All authors participated in the interpretation of the experiments and the writing of the paper. This study was supported by the Max-Planck-Society (V.P. and R.L.) and by the German Research Foundation (V.P., Deutsche Forschungsgemeinschaft grant PE 2079/2-1).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Wahl, M.C., Will, C.L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).
2. Staley, J.P. & Guthrie, C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315–326 (1998).
3. Bessonov, S. *et al.* Characterization of purified human B^{act} spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA* **16**, 2384–2403 (2010).
4. Agafonov, D.E. *et al.* Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol. Cell. Biol.* **31**, 2667–2682 (2011).
5. Bleichert, F. & Baserga, S.J. The long unwinding road of RNA helicases. *Mol. Cell* **27**, 339–352 (2007).
6. Pyle, A.M. Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu. Rev. Biophys.* **37**, 317–336 (2008).
7. Chakrabarti, S. *et al.* Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol. Cell* **41**, 693–703 (2011).
8. Semlow, D.R. & Staley, J.P. Staying on message: ensuring fidelity in pre-mRNA splicing. *Trends Biochem. Sci.* **37**, 263–273 (2012).
9. Fabrizio, P. *et al.* The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol. Cell* **36**, 593–608 (2009).
10. Fairman-Williams, M.E., Guenther, U.P. & Jankowsky, E. SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.* **20**, 313–324 (2010).
11. Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H. & Luhrmann, R. Isolation of an active step I spliceosome and composition of its RNP core. *Nature* **452**, 846–850 (2008).
12. Gozani, O., Feld, R. & Reed, R. Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev.* **10**, 233–243 (1996).
13. Hirose, T. *et al.* A spliceosomal intron binding protein, IBP160, links position-dependent assembly of intron-encoded box C/D snoRNP to pre-mRNA splicing. *Mol. Cell* **23**, 673–684 (2006).
14. Ideue, T., Sasaki, Y.T., Hagiwara, M. & Hirose, T. Introns play an essential role in splicing-dependent formation of the exon junction complex. *Genes Dev.* **21**, 1993–1998 (2007).
15. Korneta, I., Magnus, M. & Bujnicki, J.M. Structural bioinformatics of the human spliceosomal proteome. *Nucleic Acids Res.* **40**, 7046–7065 (2012).
16. Chang, Y.F., Imam, J.S. & Wilkinson, M.F. The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* **76**, 51–74 (2007).
17. Tanner, N.K., Cordin, O., Banroques, J., Doere, M. & Linder, P. The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol. Cell* **11**, 127–138 (2003).
18. Kuraoka, I. *et al.* Isolation of XAB2 complex involved in pre-mRNA splicing, transcription, and transcription-coupled repair. *J. Biol. Chem.* **283**, 940–950 (2008).
19. Andrade, M.A., Petosa, C., O'Donoghue, S.I., Muller, C.W. & Bork, P. Comparison of ARM and HEAT protein repeats. *J. Mol. Biol.* **309**, 1–18 (2001).
20. Tewari, R., Bailes, E., Bunting, K.A. & Coates, J.C. Armadillo-repeat protein functions: questions for little creatures. *Trends Cell Biol.* **20**, 470–481 (2010).
21. Chan, S.P., Kao, D.I., Tsai, W.Y. & Cheng, S.C. The Prp19p-associated complex in spliceosome activation. *Science* **302**, 279–282 (2003).
22. McGrail, J.C., Krause, A. & O'Keefe, R.T. The RNA binding protein Cwc2 interacts directly with the U6 snRNA to link the nineteen complex to the spliceosome during pre-mRNA splicing. *Nucleic Acids Res.* **37**, 4205–4217 (2009).
23. Hogg, R., McGrail, J.C. & O'Keefe, R.T. The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. *Biochem. Soc. Trans.* **38**, 1110–1115 (2010).
24. Makarova, O.V. *et al.* A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *EMBO J.* **23**, 2381–2391 (2004).
25. Villa, T. & Guthrie, C. The Isy1p component of the NineTeen complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing. *Genes Dev.* **19**, 1894–1904 (2005).
26. Baserga, S.J., Yang, X.D. & Steitz, J.A. An intact Box C sequence in the U3 snRNA is required for binding of fibrillar, the protein common to the major family of nucleolar snRNPs. *EMBO J.* **10**, 2645–2651 (1991).
27. Warkocki, Z. *et al.* Reconstitution of both steps of *Saccharomyces cerevisiae* splicing with purified spliceosomal components. *Nat. Struct. Mol. Biol.* **16**, 1237–1243 (2009).

ONLINE METHODS

Cloning and mutagenesis. Recombinant proteins were produced from synthetic genes (GeneArt, Life Technologies). Full-length and truncated (Δ ARM) Aquarius with a C-terminal His₆ tag were cloned into the vector pFL. For coexpression of hSyl1, hSyl1, CCDC16 and CypE, the hSyl1 and hSyl1 genes cloned into pFL, and CCDC16, and CypE inserted into pSPL, were fused with site-specific *Cre-loxP* recombination *in vitro*. The K829A and Y1196A mutations were introduced with the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies) and were verified by DNA sequencing. The expression constructs were transformed into DH10MultiBacY *Escherichia coli* cells, and recombinant bacmids were isolated with a QIAprep Spin Miniprep Kit (Qiagen) and used for transfection of Sf9 insect cells.

Protein production. Recombinant proteins were produced in insect cells with the baculovirus expression system, essentially as previously described²⁸. Virus production (initial V_0 and amplified V_1) was performed in Sf9 or Sf21 cells. Large-scale protein production was conducted in High Five cells.

For production of selenomethionine-substituted Aquarius, cells were transferred to methionine-free medium (921 Delta Series, Expression Systems) and grown for two passages. Subsequently, cells were infected with V_1 viral stock and, after 8 h, DL-selenomethionine (Sigma) was added to 100 mg/L.

The same purification protocol was used for the wild-type and selenomethionine-substituted Aquarius (1–1485), Δ ARM (417–1485) and the mutants. Cells were resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 150 mM KCl, 10% glycerol, and 2 mM DTT) supplemented with EDTA-free protease inhibitors (Roche) and lysed by sonication. Cell debris was removed by centrifugation, and the soluble fraction was loaded onto a 50-ml anion-exchange Q-Sepharose column (GE Healthcare) equilibrated with lysis buffer. The bound proteins were eluted with a linear gradient from 150 mM to 1 M KCl and analyzed by SDS-PAGE. The peak fractions were pooled and loaded on a 5-ml Protino Ni-NTA column (Macherey-Nagel) equilibrated with 50 mM HEPES-KOH, pH 7.5, 500 mM KCl, 10% glycerol, 2 mM DTT and 15 mM imidazole. The bound proteins were eluted with a linear imidazole gradient of 15 to 250 mM and analyzed by SDS-PAGE. The peak fractions were pooled, concentrated to 2 ml with centrifugal filters (Millipore) and loaded on a gel-filtration column (Superdex 200, GE Healthcare) equilibrated with 20 mM HEPES-KOH, pH 7.5, 100 mM KCl, 5% glycerol, 2 mM DTT and 2 mM MgCl₂. The peak fractions were analyzed by SDS-PAGE and concentrated with centrifugal filters. Purified protein was flash frozen in liquid nitrogen and stored at -80°C .

The recombinant IBC was reconstituted with two populations of High Five cells: one population contained the histidine-tagged Aquarius (wild-type or mutated), and the second one contained hSyl1, hSyl1, CCDC16 and CypE. The two populations of cells were resuspended in lysis buffer (25 mM HEPES-KOH, pH 7.5, 200 mM KCl, 10% glycerol, 2 mM DTT, and 15 mM imidazole) and were colysed by sonication. Cell debris was removed by centrifugation, and the soluble fraction was loaded on a 5-ml Protino Ni-NTA column equilibrated with lysis buffer. The proteins were eluted with a linear imidazole gradient of 15 to 250 mM and analyzed by SDS-PAGE. The IBC-containing fractions were loaded on an anion-exchange MonoQ 5/50 GL column (GE Healthcare) equilibrated with 20 mM HEPES-KOH, pH 7.5, 150 mM KCl, 10% glycerol and 5 mM DTT. Proteins were eluted with a linear gradient of 150 mM to 1 M KCl. The peak fractions were pooled, concentrated with centrifugal filters to 2 ml and loaded on a Superose 6 column equilibrated with 20 mM HEPES-KOH, pH 7.5, 150 mM KCl, 10% glycerol, 5 mM DTT and 2 mM MgCl₂. The purified IBC was concentrated with centrifugal filters, flash frozen in liquid nitrogen and stored at -80°C . The mass of the IBC was determined by multiangle light scattering (Wyatt Technology).

NTPase assays. NTPase activity was assayed in a 20- μl reaction containing 5–100 nM protein, 10 μM cold NTP, 1 μl [γ -³²P]NTP or [α -³²P]NTP (0.01 $\mu\text{Ci}/\mu\text{l}$) in buffer R (50 mM HEPES-NaOH, pH 7.5, 50 mM NaCl, 0.5 mM MgCl₂, 0.4 U/ μl RNasin, 100 $\mu\text{g}/\text{ml}$ BSA, 5% glycerol, and 2 mM DTT) with or without 50 nM RNA. The reactions were incubated 1 h at 30 $^\circ\text{C}$. Subsequently, 2 μl was spotted onto a PEI TLC plate (Macherey-Nagel) ~1 cm from the bottom of a plate, which was then developed for 2 h in 0.3 M KH₂PO₄. The plate was air dried, and spots were visualized with a phosphorimager.

Unwinding assays. RNA duplexes were prepared with synthetic RNA oligonucleotides (Thermo Fisher Scientific). Prior to use, all RNA oligonucleotides were

2' deprotected according to the manufacturer's instructions. Duplexes were prepared with the same 12-nt top RNA strand and a corresponding complementary bottom RNA strand. The top RNA strand was labeled at its 5' end with [γ -³²P]ATP and was subsequently mixed with a 1.2-fold molar excess of the corresponding unlabeled bottom strand. Annealing of the strands was performed by heating to 95 $^\circ\text{C}$ and then slowly cooling to 20 $^\circ\text{C}$ over 90 min.

A typical unwinding reaction (20 μl) contained 0.5 nM RNA duplex and 50 nM–1 μM protein in buffer R and was initiated by addition of 2 mM ATP/MgCl₂ and then incubated for 1 h at 30 $^\circ\text{C}$. The reaction was stopped with 5 μl of quenching buffer containing 2.5% SDS, 50 mM EDTA, 25% glycerol, 50 mM Tris-HCl, pH 8.0, and 0.1% bromophenol blue. The reaction products were resolved by 15% nondenaturing PAGE (acrylamide/bisacrylamide ratio 19:1) in 0.5 \times TBE buffer at 4 $^\circ\text{C}$. Bands were visualized with a phosphorimager.

RNA binding assays. RNA binding was assayed by electrophoretic mobility shift assays. Single-stranded RNA corresponding to the PM5 pre-mRNA intron (158 nt, GUAUGUAUCAAGCUUACAAGACAGCUUUAAGGACACCAUAGAAAC UGGCAUGUGGAGACAGAGAAGACUCUUGGCCUCGAGAAACCGUUA ACUGGAAUGUGUGGAGUGUGACUGAUAGAACACUACCUGAUUC UUAUGUAUUUACUGACCUGUG) was produced by *in vitro* transcription and 5' labeled with [γ -³²P]ATP. A typical reaction (10 μl) contained 1 nM 5'-³²P-labeled RNA and 5–200 nM protein in buffer R. Reactions were incubated at 30 $^\circ\text{C}$ for 20 min and then directly loaded onto a 4% polyacrylamide (acrylamide/bisacrylamide ratio 37.5:1) nondenaturing gel. The gel was run in 0.5 \times Tris-glycine buffer at 4 $^\circ\text{C}$. Bands were visualized with a phosphorimager.

Fluorescence anisotropy. Fluorescence anisotropy measurements were performed at 25 $^\circ\text{C}$ by incubation of a fixed amount of Cy3-labeled RNA (20 nM) or intrinsically fluorescent mant-ATP γ S/ADP (200 nM) with increasing concentrations of the protein in binding buffer (20 mM HEPES-NaOH, pH 7.5, 100 mM NaCl, 1 mM MgCl₂ and 5% glycerol). The excitation and emission wavelengths were 540 nm and 570 nm for Cy3-RNA and 355 nm and 448 nm for mant-ATP γ S/ADP. The equilibrium measurements of fluorescence anisotropy were performed with a FluoroMax-4 (Horiba Scientific) with the standard conditions provided by FluorEssence V3.5. The data were analyzed with the quadratic equation

$$A = A_0 + (A_{max} - A_0) \frac{(P_t + L_t + K_d) - \sqrt{(P_t + L_t + K_d)^2 - 4 \times P_t \times L_t}}{2}$$

where A is the fluorescence anisotropy change, A_0 the initial anisotropy, A_{max} the final anisotropy level, L_t the total ligand concentration (Cy3-RNA or mant-ATP γ S/ADP), P_t the added protein concentration and K_d the dissociation constant of the complex between P and L .

Crystallization and diffraction data collection. Crystallization of Aquarius was performed at 20 $^\circ\text{C}$ with the sitting-drop vapor-diffusion method. Crystals were obtained by mixture of 100 nl protein solution with 100 nl reservoir solution. To obtain crystals of Aquarius in complex with AMP-PNP, 10 mg/ml protein was mixed with 2 mM AMP-PNP (Sigma) and incubated on ice for 15 min. Crystals were grown from drops formed by combining equal volumes of the Aquarius-AMP-PNP complex and crystallization solution (100 mM MES monohydrate, pH 6.5, 15% PEG MME 550). Crystals of selenomethionine-substituted Aquarius in complex with AMP-PNP were obtained according to the same procedure and were grown under the same crystallization condition.

Prior to data collection, all crystals were cryoprotected by addition of 15–20% ethylene glycol (Fluka) prepared with the corresponding crystallization solution. Diffraction data were collected at beamline PXII of SLS (Paul Scherrer Institute, Villigen, Switzerland) and beamline 14.2 of BESSY II (HZB, Berlin, Germany) and processed with XDS²⁹.

Structure determination, model building and refinement. The structure of Aquarius in complex with AMP-PNP was determined by single-wavelength anomalous dispersion (SAD). The location of selenomethionine sites, calculation of the initial phases and density modification were performed with the SHELX suite³⁰. The initial model was built with an automated protein model building program, Buccaneer³¹. The model was subsequently built and corrected manually with COOT³². Structure refinement was performed with Phenix³³.

In vitro splicing reactions. Nuclear extracts were prepared from HeLa cells essentially as previously described³⁴. Uniformly ³²P-labeled pre-mRNA was produced by *in vitro* transcription. A typical splicing reaction contained 10 nM pre-mRNA, 30% HeLa nuclear extract, 3 mM MgCl₂, 65 mM KCl, 20 mM HEPES-KOH, pH 7.9, 2 mM ATP and 20 mM creatine phosphate. To assay the effect of ATPase or helicase-deficient Aquarius on splicing, wild-type or mutant IBC was added to the splicing reaction to 100 ng/μl. The reaction mixture was incubated at 30 °C for 0–90 min. RNA was resolved on an 8 M urea–10% polyacrylamide gel. Spliceosome assembly was analyzed on a 1.5% native agarose gel after addition of 0.5 μl of heparin (5 mg/ml) to 20-μl aliquots of the splicing reaction.

Purification of the spliceosomal B^{act} and C complexes. Spliceosomal B^{act} and C complexes were isolated as previously described^{3,11}. Briefly, PM5-10, which has the same sequence as PM5 but has a shortened polypyrimidine tract that stalls splicing at the B^{act}-complex stage, or PM5 pre-mRNA (which stalls splicing at the C-complex stage) was incubated with a 20-fold molar excess of purified MS2-MBP fusion protein. A splicing reaction containing 10 nM of ³²P-labeled PM5 pre-mRNA (for C complex) or PM5-10 pre-mRNA (for B^{act} complex) was incubated for 60 min at 30 °C. To assay for recruitment of Aquarius or the IBC, protein/IBC was added to a final concentration of 100 ng/μl. Subsequently, pre-mRNAs not incorporated into B^{act} and C complexes were digested by DNA oligonucleotide-directed RNase H digestion³. Complexes were separated on 10–30% glycerol gradients containing 150 mM NaCl. Spliceosomes were affinity purified from the peak gradient fractions with amylose beads³. RNA was recovered from the eluted B^{act} or C complexes, resolved by denaturing 10% PAGE. Proteins from the purified complexes were analyzed on a 4–12% NuPAGE gel (Invitrogen). The identity of the spliceosome-associated proteins was determined by LC-MS/MS.

For BS3-cross-linking, PM5-10 pre-mRNA was incubated with a 15-fold molar excess of MS2-MBP fusion protein. The splicing reaction and RNase H digestions were performed as described above. After incubation, the splicing mixture was centrifuged for 10 min at 10,000g. The supernatant was loaded on a 1-ml MBPTrap HP column (GE Healthcare) at 0.5 ml/min after the salt concentration was adjusted to 150 mM of total monovalent cations. Subsequently, the column was washed with 40 ml of G-150 buffer (1.5 mM MgCl₂, 150 mM NaCl, and 20 mM HEPES-KOH, pH 7.9) at 1 ml/min. Spliceosomal complexes were eluted with 20 mM maltose in G-150 buffer at 0.5 ml/min and subjected to BS3 cross-linking.

Immunodepletion of nuclear extract and purification of the endogenous IBC.

HeLa nuclear extract was immunodepleted of IBC with anti-peptide antibodies against Aquarius²⁴. Antibodies were immobilized on protein A-Sepharose beads (Sigma) at 2 μg of antibody per 1 μl beads. 100 μl beads were subsequently washed five times with G-400 buffer (400 mM NaCl, 1.5 mM MgCl₂, and 20 mM HEPES-KOH, pH 7.9). HeLa nuclear extract was incubated with two portions of the beads sequentially for 1.5 h at 4 °C and then dialyzed against buffer containing 100 mM KCl, 1.5 mM MgCl₂, and 20 mM HEPES-KOH, pH 7.9. To isolate endogenous IBC, HeLa nuclear extract was incubated with anti-Aquarius antibodies coupled to 1 ml protein A-Sepharose at 4 °C for 3 h. The beads were extensively washed with G-400 buffer supplemented with 0.05% NP-40 and then with G-150 buffer. Complexes were eluted with 0.66 mg/ml of the cognate peptide in G-150 buffer and then were subjected to BS3 cross-linking or were separated on a 5–20% (v/v) glycerol gradient containing G-150 buffer, by centrifugation in a TH660 rotor (Sorvall) for 17 h at 37,000 r.p.m. The sedimentation coefficient of the IBC was determined by comparison of its migration behavior with that of protein-RNP complexes with known S values analyzed in parallel. Purified endogenous and recombinant IBC exhibited identical S values (not shown).

Immunoblotting. Proteins were resolved on a 4–12% SDS-polyacrylamide gel, transferred to a nitrocellulose Hybond-P membrane (GE Healthcare) and immunostained with an enhanced chemiluminescence (ECL) detection kit (GE Healthcare). Antibodies (experimental dilution 1:5,000) against the following human proteins were used: Aquarius/KIAA0560 (ref. 24), hSyf1 (B4-1; raised against aa 6–31; **Supplementary Data Set 4 and Fig. 2b**), hIsy1 (Sigma, HPA016995), CCDC16 (Novus Biological, NB100-68229), CypE (CE1; raised against aa 104–118; **Supplementary Data Set 4 and Fig. 2b**), hPrp19 (ref. 24) and U5-116K³⁵. Validation of commercial antibodies is provided on the manufacturers' websites.

Protein-protein cross-linking of Bact and IBC complexes. Purified B^{act} complexes were cross-linked with 150 μM bisulfosuccinimidyl suberate (BS3) for

20 min at 30 °C and were loaded onto a linear 10–30% (v/v) glycerol gradient containing G-150 buffer. The gradient was centrifuged at 23,000 r.p.m. for 13 h in a TST 41.14 rotor (Sorvall) and was harvested from the bottom. The B^{act} complexes from the peak fractions were pelleted at 60,000 r.p.m. for at least 5 h in a S100AT4 rotor (Thermo Fisher Scientific). Pelleted complexes were then subjected to mass spectrometric analysis. The purified IBC was cross-linked with 75 μM BS3 for 20 min at 30 °C and was subsequently loaded onto a linear 5–20% glycerol gradient containing G-150 buffer. The gradient was centrifuged at 40,000 r.p.m. for 18 h in a TH660 rotor (Sorvall) and was harvested manually from the top. Gradient fractions were analyzed by 4–12% gradient SDS-PAGE. The IBC from the peak fractions was pelleted at 100,000 r.p.m. in a S100AT4 rotor. Pelleted complexes were then analyzed by MS.

Identification of protein cross-links by mass spectrometry. In-solution digestion, reduction and alkylation of cross-linked B^{act} and IBC complexes was performed as previously described³⁶, with minor modifications. Cross-linked complexes were dissolved in 25 μl 8 M urea in buffer (100 mM ammonium bicarbonate, pH 7.8, 30 min, RT). Reduction and alkylation were achieved by addition of 25 μl 10 mM DTT (Calbiochem)/buffer (50:50, 30 min, RT) and 25 μl of 60 mM IAA/buffer (50:50, 30 min, RT). Modified trypsin (Promega) was added in a 1:20 w/w ratio for overnight hydrolysis. All reagents were purchased from Sigma-Aldrich unless otherwise indicated.

IBC peptides were enriched by strong cation-exchange (SCX) chromatography as reported previously³⁷. Enrichment of B^{act} cross-linked peptides by size-exclusion chromatography (SEC) was performed as previously reported³⁸. The peptide mixture was desalted with C18 chromatography as described³⁹. Fractions were reconstituted in 12 μl 5% v/v ACN, 0.1% v/v formic acid.

LC-MS/MS analysis was performed on an Orbitrap Velos instrument as described³⁶, with the following modifications. Peptides were separated on a 12-cm, 75-μm inner-diameter C18 (120 Å, 5 μm, Dr. Maisch) analytical column with a 97-min, 7–38% buffer-B gradient (95% ACN, 0.1% FA) and a flow rate of 300 μl/min. A TOP8 DDA method (MS1 = 300–1,800 m/z at 60,000 resolution, MS2 = CID activation (NCE = 30), charge state 3 and above, dynamic exclusion = 30 s, repeat count 1, at 7,500) was used.

Data analysis on the cross-linked IBC was performed with MassMatrix as described previously³⁶. The resulting matches were validated manually. Identification of B^{act} cross-links with pLink was performed according to previous publications⁴⁰, with the following parameters: carbamidomethylation of cysteine, fixed; methionine oxidation, variable; FDR = 5%. Spectra were searched against a forward and reverse database containing the UniProt sequences of the protein complex components. Spectra validation was performed with pLabel.

28. Trowitzsch, S., Bieniossek, C., Nie, Y., Garzoni, F. & Berger, I. New baculovirus expression tools for recombinant protein complex production. *J. Struct. Biol.* **172**, 45–54 (2010).
29. Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
30. Sheldrick, G.M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479–485 (2010).
31. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
32. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
33. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
34. Dignam, J.D., Lebovitz, R.M. & Roeder, R.G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**, 1475–1489 (1983).
35. Fabrizio, P., Lagerbauer, B., Lauber, J., Lane, W.S. & Luhrmann, R. An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. *EMBO J.* **16**, 4092–4106 (1997).
36. Christian, H., Hofele, R.V., Urlaub, H. & Ficner, R. Insights into the activation of the helicase Prp43 by biochemical studies and structural mass spectrometry. *Nucleic Acids Res.* **42**, 1162–1179 (2014).
37. Chen, Z.A. *et al.* Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J.* **29**, 717–726 (2010).
38. Leitner, A. *et al.* Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Mol. Cell. Proteomics* **11**, M111 014126 (2012).
39. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
40. Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nat. Methods* **9**, 904–906 (2012).