Structural and functional investigation of the human snRNP assembly factor AAR2 in complex with the RNase H-like domain of PRPF8

Marco Preussner,a* Karine F. Santos,b‡ Jonathan Alles,a,b§ Christina Heroven,b§ Florian Heyd,a Markus C. Wahlb,c and Gert Weberb,c*

a Laboratory of RNA Biochemistry, Freie Universität Berlin, Takustrasse 6, 14195 Berlin, Germany, b Laboratory of Structural Biochemistry, Freie Universität Berlin, Takustrasse 6, 14195 Berlin, Germany, and c Macromolecular Crystallography, Helmholtz-Zentrum Berlin für Materialien und Energie, Albert-Einstein-Strasse 15, 12489 Berlin, Germany. *Correspondence e-mail: mpreussner@zedat.fu-berlin.de, gert.weber@helmholtz-berlin.de

Small nuclear ribonucleoprotein complexes (snRNPs) represent the main subunits of the spliceosome. While the assembly of the snRNP core particles has been well characterized, comparably little is known of the incorporation of snRNP-specific proteins and the mechanisms of snRNP recycling. U5 snRNP assembly in yeast requires binding of the the Aar2 protein to Prp8p as a placeholder to preclude premature assembly of the SNRNP200 helicase, but the role of the human AAR2 homolog has not yet been investigated in detail. Here, a crystal structure of human AAR2 in complex with the RNase H-like domain of the U5-specific PRPF8 (PRP8F RH) is reported, revealing a significantly different interaction between the two proteins compared with that in yeast. Based on the structure of the AAR2–PRPF8 RH complex, the importance of the interacting regions and residues was probed and AAR2 variants were designed that failed to stably bind PRPF8 in vitro. Protein-interaction studies of AAR2 with U5 proteins using size-exclusion chromatography reveal similarities and marked differences in the interaction patterns compared with yeast Aar2p and imply phosphorylation-dependent regulation of AAR2 reminiscent of that in yeast. It is found that in vitro AAR2 seems to lock PRPF8 RH in a conformation that is only compatible with the first transesterification step of the splicing reaction and blocks a conformational switch to the step 2-like, Mg2+-coordinated conformation that is likely during U5 snRNP biogenesis. These findings extend the picture of AAR2 PRP8 interaction from yeast to humans and indicate a function for AAR2 in the spliceosomal assembly process beyond its role as an SNRNP200 placeholder in yeast.

1. Introduction

Precursor messenger RNA (pre-mRNA) splicing is catalyzed by a highly dynamic, multi-megadalton ribonucleoprotein (RNP) machinery, the spliceosome (Will & Lührmann, 2011; Wahl et al., 2009). The small nuclear RNPs (snRNPs) U1, U2, U4, U5 and U6 are the main subunits of the major, U2-type spliceosome. Each of these snRNPs contains a particle-specific snRNA, seven common Sm proteins [or like Sm (LSm) proteins in the case of U6] and a set of particle-specific proteins (Will & Lührmann, 2001). The U5 snRNP is the only snRNP subunit that is also employed by the minor spliceosome (Will & Lührmann, 2005). Apart from the snRNPs, a multitude of proteins and protein complexes that are not stably associated with an snRNP also join the spliceosome to facilitate and regulate pre-mRNA splicing (Agafonov et al., 2011). For each round of splicing, the spliceosome assembles anew on a pre-mRNA by the stepwise association of snRNPs and non-snRNP proteins. Its catalytic center is not preformed but only emerges during assembly by repeated, extensive
remodeling of the specific RNA–protein interaction network of each assembly stage to eventually elicit intron excision and exon ligation via two transesterification reactions, referred to as step 1 and step 2 (Wahl et al., 2009; Will & Lührmann, 2011).

SnRNPs themselves are assembled via complex pathways, which in the cases of U1, U2, U4 and U5 include cytoplasmic and nuclear phases (Will & Lührmann, 2001; Matera & Wang, 2014; Gruss et al., 2017). The corresponding snRNAs are synthesized by RNA polymerase II (Pol II), modified with an m’G cap, processed by the integrator complex and exported to the cytoplasm. Here, the Sm proteins are assembled stepwise to form a ring-like structure around a U-rich Sm site in the snRNAs via the protein arginine methyltransferase 5 complex and the survival motor neuron (SMN) complex. Trimethylguanosine synthase 1 then catalyzes hypermethylation of the m’G cap to generate an m’2,7G cap. The hypermethylated cap and the assembled Sm core domain act as a composite nuclear localization signal that facilitates re-entry of the Sm core RNPs into the nucleus.

The integration of particle-specific proteins into snRNPs also requires specific assembly factors and chaperones. For example, in human cells, the adaptor protein, nuclear FMR1-interacting protein 1, and the heat-shock protein 90 (HSP90)/Rvb1–Rvb2–Tah1–Ph1 (RT2P) chaperone machinery in collaboration with a nuclear-localized SMN complex facilitate integration of the U4/U6-specific proteins NHP2-like protein 1 and pre-mRNA processing factor (PRPF) 31 into the U4/U6 di-snRNP (Bizarro et al., 2015). The HSP90/RT2P chaperone machinery also supports the assembly of a U5 snRNP module composed of the PRPF8 protein (Prp8p in yeast), the 116 kDa U5 small nuclear ribonucleoprotein component (EFTUD2; Snu114p in yeast), the U5 small nuclear ribonucleoprotein component 200 kDa helicase (SNRNP200; Br2p in yeast) and the SNRNP200 protein in the cytoplasm, thereby promoting formation of the mature U5 snRNP (Malinova et al., 2017). Several additional proteins have been implicated in U5 snRNP or U4/U6-U5 tri-snRNP assembly in metazoans (Malinová et al., 2017; Klimešová et al., 2021; Erkelenz et al., 2021; Bergfort et al., 2022).

Some snRNPs are profoundly remodeled during pre-mRNA splicing, necessitating specific recycling mechanisms to reassemble the particles for further rounds of splicing. For example, during spliceosome activation the U4 and U6 snRNAs, which are extensively base-paired in the U4/U6 di-snRNP, are unwound by the SNRNP200 helicase, and U4 snRNA and all U4/U6-associated proteins are displaced (Laggerbauer et al., 1998; Raghunathan & Guthrie, 1998; Agafonov et al., 2011). Furthermore, in human cells the U5 snRNP enters the spliceosome as a 20S particle but is released as a 35S particle after splicing due to incorporation of the PRPF19 complex and additional factors (Makarov et al., 2002). Late de novo snRNP biogenesis steps and recycling of snRNPs after pre-mRNA splicing take place in nuclear Cajal bodies (Stanček & Neugebauer, 2004, 2006; Sleeman et al., 2001; Sleeman & Lamond, 1999).

The U5-specific PRPF8 protein is one of the most conserved nuclear proteins and coordinates proteins, snRNAs and the pre-mRNA at the catalytic center of the spliceosome (Grainger & Beggs, 2005). PRPF8 harbors two regulatory pseudo-enzyme domains at its very C-terminus, comprising an RNase H-like (RH) and a Jab1/MPN-like (JM) fold (Pena et al., 2007, 2008; Zhang et al., 2007; Yang et al., 2008; Ritchie et al., 2008). In yeast, the A1 cistron-splicing factor (Aar2p) has been characterized as a U5 snRNP assembly and recycling factor that mediates the formation of pre-U5 snRNPs lacking the Br2p helicase in the cytoplasm (Gottschalk et al., 2001; Boon et al., 2007). Aar2p concomitantly binds the RH and JM domains of Prp8p (Weber et al., 2011, 2013; Galej et al., 2013). By sequestering the Prp8p JM domain, which is a major binding site of the Br2p RNA helicase, Aar2p initially prevents the integration of Br2p into U5 snRNP (Weber et al., 2013; Galej et al., 2013). Aar2p accompanies the Br2p-deficient pre-U5 snRNP particle into the nucleus, where phosphorylation of its Ser253 residue triggers refolding and release of Aar2p, allowing Br2p entry to complete U5 snRNP biogenesis (Boon et al., 2007; Weber et al., 2013).

In a previous study, we demonstrated that the human Aar2p homolog AAR2 is produced from the c20orf4 gene in HeLa cells and that it stably binds the PRPF8 RH domain in vitro (Santos et al., 2015). However, human AAR2 and yeast Aar2p exhibit only 24% sequence identity, questioning the extent to which their structures and molecular mechanisms are conserved. While proteomics studies and pull-down experiments have suggested that human AAR2 also participates in U5 snRNP assembly (Malinová et al., 2017; Klimešová et al., 2021), human AAR2 is produced in a complex with PRPF8, EFTUD2, SNRNP200 and SNRNP40 (Malinová et al., 2017), indeed indicating potential differences in the Aar2p/AAR2-mediated U5 snRNP assembly steps in yeast and humans.

Here, we report a co-crystal structure of human AAR2 in complex with the PRPF8 RH domain (PRPF8RH) and present further interaction studies of AAR2 with PRPF8 fragments and the SNRNP200 helicase in vitro. In contrast to the situation in yeast, we find that a human AAR2–PRPF8RH complex does not bind the PRPF8 JM domain and thus permits the formation of a trimeric AAR2–PRPF8–SNRNP200 complex. As in yeast, the human AAR2–PRPF8RH interaction is abrogated in vitro by a phosphomimetic S284E (S253E in yeast) mutation, indicating highly conserved regulation of AAR2 by phosphorylation. Furthermore, AAR2 seems to lock PRPF8RH in its first-step conformation and block the conformational switch to a step 2-like, Mg2+-coordinated conformation during U5 snRNP biogenesis. Our results shed the first light on the human AAR2–PRPF8RH interface and imply a different role of AAR2 in spliceosomal assembly than in yeast.

2. Materials and methods
2.1. Cloning, expression, protein purification and reconstitution of protein complexes

We employed a modified pFL vector encoding a truncated version of human SNRNP200 lacking the first 394 residues
(residues 395–2136) and containing an N-terminal, TEV-cleavable His10 tag (Santos et al., 2012), a pET-M11 plasmid encoding human PRPF8RH (residues 1747–2016) and NdeI (NEB). AAR2 constructs containing an N-terminal, TEV-cleavable His10 tag, was cloned into a modified pFL vector by restriction-enzyme cloning using EcoRI and HindIII (Pena et al., 2008; Weber et al., 2011), a pET-M11 plasmid encoding human PRPF8TM (residues 2064–2335) and containing an N-terminal TEV-cleavable GST tag (Mozaffari-Jovin et al., 2013), and pFL vectors encoding human AAR2 or AAR2Δloop (AAR2 with residues 170–200 replaced by three serines) with both AAR2 constructs containing an N-terminal, TEV-cleavable His10 tag (Santos et al., 2015), which have been described previously. The inserts were derived from codon-optimized synthetic genes. A DNA fragment encoding the C-terminal fragment of PRPF8 (encompassing the RH and JM domains; PRPF8RH-JM, residues 1760–2335) was PCR-amplified from a codon-optimized prpf8 synthetic gene and cloned into the pLIR plasmid donor vector by restriction-enzyme cloning using Xhol and NdeI (NEB). Snrnp200395–2136-pFL and prpf8760–2335-pLIR were fused by Cre-Lox recombination and used for bacmid preparation. A codon-optimized DNA fragment encoding human AAR2 residues 1–364, lacking the C-terminal 20 residues of AAR2, was cloned into a modified pFL vector by restriction-enzyme cloning using EcoRI and HindIII (NEB) to guide the production of AAR21–364 containing a C-terminal His6 tag. A codon-optimized DNA fragment encoding human AAR2 was cloned into pcDNA3.1(+) using BamHI and Xhol restriction enzymes (NEB) to guide the production of AAR2 containing an N-terminal FLAG tag. Mutations were introduced with the QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturer’s instructions. All constructs were verified by dye-terminator sequencing (Seqlab).

Purified human SNRNP200395–2136 (8 mg ml−1; Santos et al., 2012), PRPF8RH (30 mg ml−1; Pena et al., 2008), PRPF8TM (10 mg ml−1; Mozaffari-Jovin et al., 2013), AAR2 (12 mg ml−1; Santos et al., 2015), AAR2Δloop (12 mg ml−1; Santos et al., 2015), an SNRNP200395–2136–PRPF8TM complex (Mozaffari-Jovin et al., 2013; 8 mg ml−1) and an AAR2–PRPF8RH complex (20 mg ml−1; Santos et al., 2015) were obtained as described previously. AAR21–364 (11 mg ml−1) was produced and purified as described for AAR2 but with omission of the TEV protease cleavage step, leaving the His6 tag intact. SNRNP200395–2136 and PRPF8RH-JM were co-produced based on a recombinant baculovirus derived from the recombinant bacmid in Sf9 cells. For purification of the SNRNP200395–2136–PRPF8RH-JM complex, cells were resuspended in resuspension buffer [50 mM Tris–HCl pH 8.0, 300 mM NaCl, 5% (v/v) glycerol, 1 mM DTT, 0.05% (v/v) NP40] supplemented with EDTA-free protease inhibitors (Roche) and DNase I (NEB) and lysed via sonication. After centrifugation, the lysate of about 50 column volumes was filtered and passed through Ni–NTA beads (Qiagen). The beads were washed twice with ten column volumes of resuspension buffer containing 15 mM imidazole. The captured complex was eluted with two column volumes of resuspension buffer containing 500 mM imidazole, TEV protease (0.5 mg per millilitre of protein solution) was added and the mixture was dialyzed against 20 mM HEPES–NaOH pH 7.5, 200 mM NaCl, 1 mM DTT overnight at 4°C. Five column volumes of the buffer-exchanged sample were again passed through Ni–NTA beads and the flowthrough was collected. The complex was concentrated to a final concentration of 7 mg ml−1 and further purified by size-exclusion chromatography (SEC) on a Superdex 200 10/300 column (GE Healthcare) in 20 mM HEPES–NaOH pH 7.5, 200 mM NaCl, 1 mM DTT.

2.2. Analytical size-exclusion chromatography

Individual proteins and protein mixtures were analyzed by analytical SEC on a Superdex 200 Increase PC3.2/30 column (GE Healthcare) in 20 mM Tris–HCl pH 7.5, 250 mM NaCl, 0.5 mM DTT at a flow rate of 50–70 μl min−1. For analysis of complex formation, proteins (at the concentrations stated in Section 2.1) were mixed in equimolar ratios in 60 μl size-exclusion buffer and incubated for 30 min on ice. Elution fractions were supplemented with SDS–PAGE loading buffer and analyzed by SDS–PAGE.

2.3. Crystallographic analyses

Crystallization of the human AAR2Δloop–PRPF8RH complex has been described previously (Santos et al., 2015). Briefly, 1 μl purified human AAR2Δloop–PRPF8RH complex concentrated to 14 mg ml−1 in 20 mM Tris–HCl pH 7.5, 150 mM NaCl, 0.5 mM DTT was crystallized employing an equal volume of reservoir solution consisting of 0.1 M HEPES pH 7.0, 10% (v/v) PEG 6000, 5% (v/v) 2-methyl-2,4-pentanediol. Crystals were transferred to reservoir solution supplemented with 10% (v/v) 2-methyl-2,4-pentanediol and flash-cooled in liquid nitrogen. Diffraction data were collected on beamlines BL14.1, BL14.2 and BL14.3 of the BESSY II storage ring, Berlin, Germany and on beamline P14 of the PETRA III storage ring, Hamburg, Germany at 100 K and were processed with XDS (Kabsch, 2010). The structure was solved by molecular replacement with Phaser (McCoy et al., 2007) using chain A of the structural coordinates of human PRPF8RH (PDB entry 3e9l; Pena et al., 2008), omitting the water molecules. A single model of the AAR2Δloop subunit was obtained by automated model building with phenix.autobuild (Adams et al., 2010). The model was completed through alternating rounds of automated refinement using phenix.refine (Afonine et al., 2010) and manual model building using Coot (Emsley et al., 2010). The quality of the final model was assessed with MolProbity (Chen et al., 2010). Of note, a relatively large number of difference density peaks were observed in the Fo – Fc map. Structural figures were prepared with PyMOL (Schrodinger).

3. Results

3.1. Crystal structure of a human AAR2Δloop–PRPF8RH complex

Structures of yeast Aar2p in complex with the Prp8p RH and JM domains or with full-length Prp8p have been reported (Weber et al., 2011, 2013; Galej et al., 2013). In contrast, the
structure of human AAR2, which exhibits only 24% sequence identity to the yeast ortholog (Supplementary Fig. S1), and the structural basis for its interaction with PRPF8 remain unknown. We previously reported the crystallization of a structural basis for its interaction with PRPF8 RH (Santos et al., 2015). In yeast Aar2p, the corresponding internal loop was shown to hinder crystalization, to be protease-cleavable and to be irrelevant for the interaction with Prp8p C-terminal domains (Weber et al., 2011, 2013; Santos et al., 2015). Here, we report the crystal structure of the human AAR2<sup>Δ</sup>loop–PRPF8 RH complex at 2.35 Å resolution. The structure was solved by molecular replacement with Phaser using the structural coordinates of PRPF8 RH (PDB entry 3e9; Fena et al., 2008) as a molecular-replacement search model (Table 1).

| Table 1
| Crystallographic data (Santos et al., 2015) and refinement. Values in parentheses are for the highest resolution shell. |
| Data collection |
| Wavelength (Å) | 1.24 |
| Temperature (K) | 100 |
| Space group | C2 |
| a, b, c (Å) | 145.28, 57.26, 111.23 |
| α, β, γ (°) | 90, 112.74, 90 |
| Resolution (Å) | 99.90–2.35 (2.41–2.35) |
| Unique reflections | 34687 (2502) |
| Completeness (%) | 97.6 (95.6) |
| Multiplicity | 3.8 (3.9) |
| (hkl)(I) | 15.6 (1.7) |
| Rmerge(I)† | 0.05 (0.92) |
| CC1/2‡ | 99.9 (76.5) |
| Refinement |
| Resolution (Å) | 48.21–2.35 (2.42–2.35) |
| Total No. of reflections | 34678 (2667) |
| Completeness (%) | 97.6 (95.7) |
| Reflections in test set (%) | 4.96 (4.87) |
| Resolution (Å) | 99.90–2.35 (2.41–2.35) |
| Multiplicity | 3.8 (3.9) |
| (hkl)(I) | 15.6 (1.7) |
| Rmerge(I)† | 0.05 (0.92) |
| CC1/2‡ | 99.9 (76.5) |
| Contents of asymmetric unit |
| Protein molecules/residues | 2/574 |
| Water oxygen | 147 |
| Mean B factors (Å<sup>2</sup>) |
| Wilson | 55.39 |
| Protein | 77.81 |
| Solvent | 65.02 |
| Ramachandran plot‡‡ (%) |
| Favored | 95.85 |
| Allowed | 4.42 |
| Outliers | 0 |
| R.m.s.d.§§ from target geometry |
| Bond lengths (Å) | 0.009 |
| Bond angles (°) | 0.990 |
| Dihedral angles (°) | 115.920 |
| PDB code | 7pjh |

† R<sub>merge</sub>(I) = Σ<sub>i</sub> Σ<sub>k</sub> I<sub>k</sub>(hkl) − Σ<sub>i</sub> I<sub>k</sub>(hkl) / Σ<sub>i</sub> Σ<sub>k</sub> I<sub>k</sub>(hkl) for i observations of a given reflection hkl. (I(hkl))<sub>i</sub> is the average intensity of the i observations. †† CC1/2 = <sup>1</sup>(I) − (I<sup>0</sup>)<sup>2</sup> − <sub>2</sub> σ<sub>2</sub> (for the working set; no σ<sub>2</sub> cutoff was applied). § R<sub>work</sub>, the same as R<sub>work</sub> but calculated on the test set of reflections that were excluded from refinement. ††† Estimated overall coordinate error based on maximum likelihood; ‡‡ Calculated with phenix.refine (Afonine et al., 2012). §§ Root-mean-square deviation.

Despite the low sequence identity, the overall structure of human AAR2<sup>Δ</sup>loop in the AAR2<sup>Δ</sup>loop–PRPF8 RH complex is very similar to that of yeast Aar2p<sup>Δ</sup>loop in complex with the Prp8p RH and JM domains (Weber et al., 2013; Galej et al., 2013; root-mean-square deviation of 2.13 Å for 236 pairs from 330 AAR2 and 342 Aar2p C<sub>α</sub> atoms; Supplementary Fig. S3). As previously observed for Auar2p (Weber et al., 2011, 2013; Galej et al., 2013), human AAR2<sup>Δ</sup>loop exhibits an N-terminal domain (NTD; residues 10–158) mainly composed of β-strands, an α-helical C-terminal domain (CTD; residues 202–364) and a C-terminal, irregularly structured tail (residues 365–384) (Fig. 1a).

In the full-length yeast Aar2p–Prp8p structure (PDB entry 4i43; Galej et al., 2013), the C-terminal peptide of Aar2p is fully structured and contacts several other Prp8p domains. Superposition with the present human AAR2–RH complex structure suggests that due to the shorter AAR2 C-terminal peptide in humans, contacts with other PRPF8 domains may be limited. Hence, the differences in the functionality important C-terminal peptide of yeast and human Aar2p/AAR2 may hint at a somewhat different mode of action of AAR2 in U5 snRNP or U4/U6-U5 tri-snRNP assembly in humans.

Despite the overall structural similarity of both individual components, the protein interfaces between yeast and human Prp8p RH/PRPF8 RH and Aar2p/AAR2 are markedly different. As in yeast, the NTD lacks direct interactions with PRPF8 RH, while the CTD and the C-terminal tail of AAR2 establish two interfaces with PRPF8 RH (interfaces I and II, respectively; Figs. 1a–1c). In interface I, an edge of the AAR2<sup>Δ</sup>loop CTD laterally contacts PRPF8 RH (Figs. 1d and 1e). Interface II is built by the C-terminal residues 366–377 of AAR2 extending across the PRPF8 RH domain below the protruding β-finger module (Figs. 1f and 1g). Both interfaces bury a comparable surface area in the yeast and human systems (interface I, 399 and 412 Å<sup>2</sup>, respectively; interface II, 733 and 511 Å<sup>2</sup>, respectively).

Interface I is dominated by hydrophobic contacts, with only four of 12 PRPF8 RH-interacting residues conserved between yeast Aar2p and human AAR2, underlying the different organization of the interactions. The conserved core of interface I includes interactions between Ile225 of AAR2 (Ile189 in Aar2p) and Val1874 of PRPF8 RH (Val1946 in Aar2p) as well as between Met230 of AAR2 (Met195 in Aar2p) and Trp1839 of PRPF8 RH (Trp1911 in Aar2p) (Figs. 1d and 1e; Supplementary Fig. S1). Compared with yeast Aar2p, the AAR2 CTD harbors two extended helices (α11 and α12; Supplementary Fig. S3). In addition, Ile225 and Met230 are shifted by four residues (about one helical turn) along the α5 helix compared with the equivalent residues in yeast Aar2p (Figs. 1d and 1e).
and 1e), giving rise to a markedly different angle with which human AAR2^Δloop contacts the PRPF8 RH domain compared with yeast Aar2p^Δloop in the Aar2p^Δloop–Prp8p^RH–Prp8p^JM complex (Fig. 1c). Also, the AAR2^Δloop residues participating

Figure 1
(a) Crystal structure of the human AAR2^Δloop–PRPF8^RH complex. Colour scheme for this and the following figures: AAR2^Δloop, orange; PRPF8^RH, light blue; dashed orange lines indicate a flexible loop (labeled Ser3) in AAR2 connecting its two domains, which is replaced by three serine residues in AAR2^Δloop (Santos et al., 2015), and another smaller flexible loop between residues 313 and 321 (labeled flexible loop). The N- and C-termini as well as the β-finger module of PRPF8^RH are labeled. (b) Superposition of the RH domains of human PRPF8 and yeast Prp8p in complex with human AAR2^Δloop and yeast Aar2p/PRPF8^TM (PDB entry 4i43; Galej et al., 2013), respectively, to illustrate the human AAR2 in a larger PRPF8 context. Colour scheme for this and the following figures: Aar2p, maroon; Prp8p^RH, dark blue; Prp8p^TM, cyan. (c) Comparison of the human AAR2^Δloop–PRPF8^RH complex and the yeast Aar2p^Δloop–Prp8p^RH–Prp8p^JM complex (PDB entry 4ilg; Weber et al., 2013) by superposition of the RH domains. Dashed maroon line, flexible linker preceding the C-terminal tail of Aar2p (Weber et al., 2013). (d, e) Close-up views of interface regions I of the yeast Aar2p^Δloop–Prp8p^RH–Prp8p^JM complex (f) and the human AAR2^Δloop–PRPF8^RH complex (g). In (d–f) and the following figures interacting residues are shown as sticks colored by atom type, with carbon colored as the respective protein, nitrogen in blue, oxygen in red and sulfur in yellow; green spheres are water molecules, dashed black lines represent hydrogen bonds or salt bridges and rotation symbols represent orientations relative to (a).
in interface II are only partially conserved between yeast and humans (two of eight residues; Supplementary Fig. S1).

3.2. Similarities and differences in AAR2–PRPF8–SNRNP200 and Aar2p–Prp8p–Brr2p interactions in humans and yeast

The low degree of conservation of AAR2 and observed marked differences in the interface with PRPF8RH have apparent consequences for AAR2 function and likely for interactions within the spliceosome. To test the importance of the specific contacts between AAR2Δloop and PRPF8RH that are observed in our co-crystal structure, we conducted analytical SEC runs with wild-type (WT) proteins and variants. To this end, we investigated the binding of WT AAR2 to WT PRPF8RH in previous work, which is only shown here for comparison (Figs. 2a–2c; Santos et al., 2015). In yeast, the C-terminal tail of Aar2p is dispensable for Prp8pRH binding (Weber et al., 2011). Conversely, in the human system, AAR21–364, which lacks the C-terminal tail, no longer binds stably to PRPF8RH (Figs. 2a–2d). Likewise, converting Trp1839 of PRPF8RH or Met230 of AAR2, which are part of the conserved core of interface I, individually to alanine residues abrogated complex formation (Fig. 2e and 2f). Again, the situation differs in yeast, where only Trp1911 of Prp8pRH (equivalent to Trp1839 in human PRPF8RH), but not Met195 of Aar2p (equivalent to Met230 in human AAR2), is essential for the interaction (Weber et al., 2013).

The low sequence conservation and the resulting structural differences in the AAR2–PRPF8RH interface might also have consequences for the wider protein interaction network around AAR2. Concomitant binding of the Prp8p RH and JM domains by Aar2p in yeast sequesters the JM domain, preventing binding of the Br2p RNA helicase to Aar2p–pre-U5 snRNP (Weber et al., 2013; Galej et al., 2013). In yeast Aar2p–Prp8p complexes (Weber et al., 2011, 2013; Galej et al., 2013), the C-terminal tail of Aar2p runs along the protruding Prp8pRH β-finger module, stringing the β-finger and the central Prp8pJM β-sheet into an extended, intermolecular β-structure (Figs. 1b and 1c).

While the beginning of the C-terminal tail in human AAR2Δloop maintains similar interactions with PRPF8RH as in yeast, for example employing Val373–Val375 to form a short β-sheet of three hydrogen bonds to PRPF8RH, distal parts of the C-terminal tail (beyond Val374) deviate from the direction of the Aar2p C-terminal tail (Figs. 1b and 1c). In yeast, the formation of the penultimate β-strand of Aar2p and the concomitant sequestration of JM from Br2p is mediated exclusively by a series of hydrophobic residues at the very C-terminus of Aar2p, which are complementary to hydrophobic residues of the neighboring β-strands of RH and JM (Weber et al., 2013; Galej et al., 2013). A structure-based alignment revealed that the respective very C-terminal residues of AAR2, Pro378, Glu379, Gly380 and Glu382, are unlikely to support β-sheet formation with the corresponding highly conserved residues of the PRPF8RH β-finger and PRPF8JM due to their steric or polar properties (Supplementary Fig. S1; compare Figs. 1f and 1g). However, we cannot exclude that in the context of the full-length proteins the very C-terminus of hAAR2 may engage in a yeast-like interaction with the PRPF8 JM domain.

The C-terminal tail of human AAR2Δloop in the observed conformation would not be able to concomitantly bind the
PRPF8JM domain as observed in yeast. Indeed, also confirming a prior study (Malinová et al., 2017), AAR2–PRPF8RH did not stably bind PRPF8JM in analytical SEC (Figs. 2g and 2h) and failed to sequester PRPF8JM from a pre-formed SNRNP200395–2136–PRPF8JM complex (Fig. 3a and 3b).

AAR2 alone or in complex with PRPF8RH did not bind stably to SNRNP200395–2136 or to a SNRNP200395–2136–PRPF8JM complex (Figs. 3c and 3d). Instead, a stable AAR2–PRPF8RH–SNRNP200395 ternary complex was formed upon mixing the components (Fig. 3b).

3.3. Conserved Aar2 phosphorylation between humans and yeast

Aar2p can be phosphorylated at five positions in vivo (Ser253, Thr274, Tyr328, Ser331 and Thr345) and phosphomimetic S253D or S253E variants of Aar2p interfered with Aar2p–Prp8p interaction in yeast extracts (Weber et al., 2011). Structural analysis of a phosphomimetic Aar2pS253E variant suggested that phosphorylation leads to a local conformational rearrangement of the Aar2p CTD and thereby to disruption of the Prp8pRH binding site (Weber et al., 2011). A structure-based sequence alignment revealed that Ser284 in human AAR2 corresponds to Ser253 in yeast Aar2p (Supplementary Fig. S1; Fig. 3e), and AAR2 has been found to be phosphorylated at Ser284 in human liver cancer cells (Hornbeck et al., 2012). Recapitulating the situation in yeast, an AAR2S284E phosphomimetic variant failed to stably bind PRPF8RH in analytical SEC (Fig. 3f). Taken together, our interaction studies reveal differences in the relative importance of AAR2/Aar2p regions in maintaining a stable interaction with the PRPF8/Prp8p RH domain in the human and yeast.
yeast systems. Furthermore, AAR2 does not sequester the PRPF8 JM domain to intermittently prevent SNRNP200 association with the U5 snRNP. AAR2 displacement from PRPF8 may involve reversible phosphorylation of AAR2 at Ser284. Thus, the U5 snRNP assembly steps apparently differ in detail in yeast and humans.

Figure 4
AAR2-mediated blocking of a step 2 conformation in PRPF8RH. (a–d) Structure of the AAR2\textsuperscript{Δloop}–PRPF8\textsuperscript{RH} complex (a) and close-up views comparing the AAR2\textsuperscript{Δloop} C-terminal tail (sticks) traversing the PRPF8 RH domain below the protruding \textit{β}-finger module (surface views) as observed in the AAR2\textsuperscript{Δloop}–PRPF8\textsuperscript{RH} complex (b) or modeled onto the PRPF8\textsuperscript{RH} domain in the step 1 conformation (PDB entry 4jk7; Schellenberg \textit{et al.}, 2013) (c) or onto the PRPF8\textsuperscript{RH} domain in the step 2 conformation (PDB entry 4jk7; Schellenberg \textit{et al.}, 2013) (d) by superposition of the RH domains. Yellow sphere, coordinated Mg\textsuperscript{2+} ion. The AAR2 C-terminus clashes with the PRPF8RH domain in the step 2 conformation. (e–h) SDS–PAGE analyses (left) and UV elution profiles (right) of analytical size-exclusion chromatography runs monitoring the interaction between AAR2 and PRPF8\textsuperscript{RH} (e, g) or PRPF8\textsuperscript{RH,T1789P} (f, h) at 2 mM (e, f) or 40 mM (g, h) magnesium chloride. Lane M, molecular-mass standard (kDa); lane I, input samples. Protein bands are identified on the right. Elution fractions are indicated at the top of the gel and profile in (e); elution volumes are indicated at the bottom of the profile in (h). Icons are explained at the bottom left. Variants are indicated below the respective icons. Peaks labeled with transparent icons represent an excess of the respective protein.
3.4. Human AAR2 counteracts a step 2-like conformation in the PRPF8 RH domain

Mutations in the prpf8 gene can lead to retinitis pigmentosa (RP; Růžičková & Staněk, 2017), a disease that causes blindness in humans, and the corresponding PRPF8/Prp8p variants cause defects in U5 snRNP assembly (Malinová et al., 2017) and splicing (Mayerle & Guthrie, 2016; Mozaffari-Jovin et al., 2013) in humans and yeast. In baker’s yeast, two sets of prpf8 mutant alleles, corresponding to RP-related mutations in humans that disrupt either the first or the second step of splicing, cluster in the Prp8p RH domain (Grainger & Beggs, 2005).

Furthermore, the human PRPF8 RH domain can undergo a conformational switch in a protruding β-finger module, with one conformation promoting the first step and an alternative, Mg$_{2+}$-bound conformation supporting the second step of splicing (Schellenberg et al., 2013). Despite the biochemical and structural evidence reported previously, which supports this switch, a caveat of our AAR2–PRPF8$^{\text{RH}}$ structure may be that the RH β-finger module makes crystal contacts with a neighboring symmetry-related RH β-finger module. However, recent cryogenic electron-microscopy structures of spliceosomes also confirm this conformational switch, rationalize some of the effects of PRPF8 RH domain variants and demonstrate repeated, long-range repositioning of the PRPF8$^{\text{RH}}$–PRPF8RH$^{\text{C12}}$-finger module, with an important function of AAR2 in the human system. In addition, transient blocking of binding sites on the PRPF8 RH domain, possibly supported by allosteric effects due to the selective stabilization of a step 1-like conformation in the PRPF8 RH domain by AAR2, may help to order assembly steps during U5 snRNP biogenesis. The above findings and suggestions are in agreement with the previous observation of the interaction of human AAR2 with a PRPF8–EFTUD2–SNRNP200–SNRNP40 U5 submodule (Malinová et al., 2017).

As most protein-coding genes in humans contain multiple introns (Lee & Rio, 2015), pre-mRNA splicing is an inherent step in their expression. Moreover, pre-mRNA splicing predominantly occurs co-transcriptionally (Alpert et al., 2017) and splicing is physically and functionally coupled to transcription, other pre-mRNA processing steps and mRNA export (Carrocci & Neugebauer, 2019; Tellier et al., 2020). Thus, efficient splicing is a prerequisite for efficient gene expression and, due to its stabilization of the step 1 configuration of RH, a potential role of human AAR2 in pre-mRNA splicing cannot be ruled out. AAR2 may have a moonlighting function during pre-mRNA splicing independent of its role as a U5 snRNP assembly factor. By binding the PRPF8 RH domain during a stage of splicing when it is available, for example, in the pre-catalytic B complex (PDB entry 7abg; Townsend et al., 2020), AAR2 may hinder the transition to subsequent stages, thus impeding splicing and, as a consequence, gene expression. As in the case of U5 snRNP assembly, direct blocking of binding sites on PRPF8$^{\text{RH}}$ and allosteric effects due to the stabilization of a step 1 conformation in PRPF8$^{\text{RH}}$ may support such a splicing-inhibitory role of AAR2. Again, the observed high nuclear levels of AAR2 might ensure that sufficient AAR2 is available to serve multiple functions, as moonlighting is known for some splicing factors that are in excess over other splicing machinery. For example, U1 snRNP has additional roles in 3′-end processing of Pol II transcripts (telescripting; Di et al., 2019). However, AAR2 has never been found to be associated with the spliceosome at any stage of splicing (Agafonov et al., 2011), arguing against a direct effect of AAR2 on splicing. Further studies on human Aar2 in a spliceosomal context will hopefully resolve these remaining questions.

4. Discussion

We have elucidated similarities and differences in the structures and interaction profiles of yeast Aar2p and human AAR2 and have identified a putative, conserved phosphorylation event that is most likely to be involved in the functional cycle of AAR2 as a U5 snRNP assembly factor. Based on our findings, we conclude that the precise roles of Aar2p and AAR2 in U5 snRNP biogenesis differ. In yeast, an Aar2p–pre-U5 snRNP, from which the Brr2p RNA helicase is excluded, seems to constitute an important U5 snRNP assembly intermediate (Boon et al., 2007; Weber et al., 2013). In contrast, our observations of (i) human AAR2 failing to sequester the PRPF8 JM domain from SNRNP200$^{\text{995–2136}}$ and (ii) AAR2 concomitantly binding to a PRPF8 fragment encompassing the RH and JM domains and SNRNP200$^{\text{995–2136}}$ suggest that an equivalent, long-lived intermediate is not formed in the human system. Association of AAR2 with the PRPF8 RH domain as in our AAR2–PRPF8$^{\text{RH}}$ structure would prevent the PRPF8 RH domain from engaging with other regions of PRPF8, the N-terminal region of SNRNP200, the C-terminal region of PRPF31, PRPF6, U4/U6 di-snRNAs and U5 snRNA as observed in the human U4/U6-U5 tri-snRNP (Agafonov et al., 2016; Charenton et al., 2019). This finding suggests that prevention of the premature association of U4/U6 di-snRNP components with pre-U5 particles may be an important function of AAR2 in the human system.

Acta Cryst. (2022). D78, 1373–1383

Marco Preussner et al. • AAR2 complexed with the RNase H-like domain of PRPF8

1381
5. Data availability
Structure coordinates and diffraction data have been deposited in the Protein Data Bank (http://www.pdb.org) under accession code 7pjh. All other data supporting the findings of this study are described in the manuscript or in the supporting information or are available from the corresponding authors on request.

6. Related literature
The following references are cited in the supporting information for this article: Barton (1993), Kabsch & Sander (1983) and Pettersen et al. (2004).

Acknowledgements
We are grateful to Reinhard Lührmann, Max-Planck-Institut für Biophysikalische Chemie, Göttingen, Germany for the gift of rabbit αAR2 polyclonal antibody. We acknowledge access to beamlines BL14.1, BL14.2 and BL14.3 of the BESSY II storage ring, Berlin, Germany via the Joint Berlin MX-Laboratory sponsored by the Helmholtz Zentrum Berlin für Materialien und Energie, the Freie Universität Berlin, the Humboldt-Universität zu Berlin, the Max-Delbrück Centrum, the Leibniz-Forschungsinstitut für Molekulare Pharmakologie, Charité – Universitätsmedizin Berlin and the Max-Planck-Institut für Kolloid- und Grenzflächenforschung, Potsdam, Germany and to beamline P14 of the PETRA III storage ring, Deutsches Elektronen-Synchrotron, Hamburg, Germany. Author contributions are as follows. KFS, JA, CH and GW conducted molecular cloning, expression, protein purification and interaction studies. KFS, CH and GW conducted crystallographic analyses. All authors participated in data interpretation. MP, MCW, FH and GW wrote the manuscript, which was revised and approved by all authors. MP and GW initiated the study and directed the project. MCW and FH provided funding for the project. The authors declare no competing interests. Open access funding enabled and organized by Projekt DEAL.

Funding information
This work was funded by a grant from the Deutsche Forschungsgemeinschaft (TRR186/A15 to FH and MCW).

References
Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). Acta Cryst. D66, 213–221.
Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Ursprech, A., Zwart, P. H. & Adams, P. D. (2012). Acta Cryst. D68, 352–367.
Agafonov, D. E., Deckert, J., Wolf, E., Odenwidler, P., Bessonov, S., Will, C. L., Urlaub, H. & Lührmann, R. (2011). Mol. Cell. Biol. 31, 2667–2682.
Agafonov, D. E., Kastner, B., Dybkov, O., Hofele, R. V., Liu, W.-T., Urlaub, H., Lührmann, R. & Stark, H. (2016). Science, 351, 1416–1420.
Alpert, T., Herzel, L. & Neugebauer, K. M. (2017). WIREs RNA, 8, e1401.
Barton, G. J. (1993). Protein Eng. Des. Sel. 6, 37–40.
Bergfort, A., Hilal, T., Kurooka, B., Ilik, I. A., Weber, G., Aktaş, T., Freund, C. & Wühl, M. C. (2022). Nucleic Acids Res. 50, 2938–2958.
Bertram, K., Agafonov, D. E., Dybkov, O., Haselbach, D., Leeclaram, M. N., Will, C. L., Urlaub, H., Kastner, B., Lührmann, R. & Stark, H. (2017). Cell, 170, 701–713.
Bertram, K., Agafonov, D. E., Liu, W.-T., Dybkov, O., Will, C. L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H. & Lührmann, R. (2017). Nature, 542, 318–323.
Bizarro, J., Dodrè, M., Huttin, A., Charpentier, B., Schlötter, F., Branlant, C., Verheggen, C., Massenet, S. & Bertrand, E. (2015). Nucleic Acids Res. 43, 8973–8989.
Boon, K.-L., Grainger, R. J., Elsani, P., Barrass, J. D., Auchynnikava, T., Inglehearn, C. F. & Beggs, J. D. (2007). Nat. Struct. Mol. Biol. 14, 1077–1083.
Carrocci, T. J. & Neugebauer, K. M. (2019). Cold Spring Harb. Symp. Quant. Biol. 84, 11–20.
Charenton, C., Wilkinson, M. E. & Nagai, K. (2019). Science, 364, 362–367.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). Acta Cryst. D66, 12–21.
Di, C., So, B. R., Cai, Z., Arai, C., Duan, J. & Dreyfuss, G. (2019). Cold Spring Harb. Symp. Quant. Biol. 84, 115–122.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Acta Cryst. D66, 486–501.
Erkelenz, S., Stanković, D., Mundorf, J., Bresser, T., Claudius, A. K., Boehm, V., Gehring, N. H. & Uhlirova, M. (2021). Nucleic Acids Res. 49, 1688–1707.
Fica, S. M., Oubridge, C., Galew, W. P., Wilkinson, M. E., Bai, X.-C., Newman, A. J. & Nagai, K. (2017). Nature, 542, 377–380.
Galew, W. P., Oubridge, C., Newman, A. J. & Nagai, K. (2013). Nature, 493, 638–643.
Galew, W. P., Wilkinson, M. E., Fica, S. M., Oubridge, C., Newman, A. J. & Nagai, K. (2016). Nature, 537, 197–201.
Gottschalk, A., Kastner, B., Lührmann, R. & Fabrizio, P. (2001). RNA, 7, 1554–1565.
Grainger, R. J. & Beggs, J. D. (2005). RNA, 11, 533–557.
Gruss, O. J., Meduri, R., Schilling, M. & Fischer, U. (2017). Chromosoma, 126, 577–593.
Hornbeck, P. V., Kornhauer, J. M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V. & Sullivan, M. (2012). Nucleic Acids Res. 40, D261–D270.
Kabsch, W. (2010). Acta Cryst. D66, 125–132.
Kabsch, W. & Sander, C. (1983). Biopolymers, 22, 2577–2637.
Karplus, P. & Diederichs, K. (2012). Science, 336, 1030–1033.
Klimešová, K., Vojáčková, J., Radivojević, N., Vandermoere, F., Bertrand, E., Verheggen, C. & Staněk, D. (2011). Nat. Commun. 12, 3646.
Laggerbauer, B., Achsel, T. & Lührmann, R. (1998). Proc. Natl Acad. Sci. USA, 95, 4188–4192.
Lee, Y. & Rio, D. C. (2015). Annu. Rev. Biochem. 84, 291–323.
Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L., Wilm, M. & Lührmann, R. (2002). Science, 298, 2025–2028.
Malinová, A., Cvačková, Z., Matějů, D., Horčízková, Z., Abéza, C., Van de Peer, E., Bertrand, E., Staněk, D. & Verheggen, C. (2017). J. Cell Biol. 216, 1579–1596.
Matera, A. G. & Wang, Z. (2014). Nat. Rev. Mol. Cell Biol. 15, 108–121.
Mayerle, M. & Guthrie, C. (2016). RNA, 22, 793–809.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). J. Appl. Cryst. 40, 658–674.
