

# Crystal structure of U2 snRNP SF3b components: Hsh49p in complex with Cus1p-binding domain

ANNE-MARIE M. VAN ROON,<sup>1</sup> CHRIS OUBRIDGE,<sup>1</sup> EIJI OBAYASHI,<sup>1,3</sup> BENEDETTA SPOSITO,<sup>1</sup> ANDREW J. NEWMAN,<sup>1</sup> BERTRAND SÉRAPHIN,<sup>2</sup> and KIYOSHI NAGAI<sup>1</sup>

<sup>1</sup>MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, United Kingdom

<sup>2</sup>Equipe Labellisée La Ligue, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Centre National de la Recherche Scientifique (CNRS) UMR 7104/Institut National de la Santé et de la Recherche Médicale (INSERM), U964/Université de Strasbourg, 67404 Illkirch, France

## ABSTRACT

Spliceosomal proteins Hsh49p and Cus1p are components of SF3b, which together with SF3a, Msl1p/Lea1p, Sm proteins, and U2 snRNA, form U2 snRNP, which plays a crucial role in pre-mRNA splicing. Hsh49p, comprising two RRM, forms a heterodimer with Cus1p. We determined the crystal structures of *Saccharomyces cerevisiae* full-length Hsh49p as well as its RRM1 in complex with a minimal binding region of Cus1p (residues 290–368). The structures show that the Cus1 fragment binds to the  $\alpha$ -helical surface of Hsh49p RRM1, opposite the four-stranded  $\beta$ -sheet, leaving the canonical RNA-binding surface available to bind RNA. Hsh49p binds the 5' end region of U2 snRNA via RRM1. Its affinity is increased in complex with Cus1(290–368)p, partly because an extended RNA-binding surface forms across the protein–protein interface. The Hsh49p RRM1–Cus1(290–368)p structure fits well into cryo-EM density of the B<sup>act</sup> spliceosome, corroborating the biological relevance of our crystal structure.

**Keywords:** splicing; U2 snRNP; SF3b; RNA binding; RRM

## INTRODUCTION

Splicing is a process that occurs in eukaryotes to remove non-coding sequences (introns) from pre-messenger RNA (pre-mRNA) and splice together coding sequences (exons) to obtain mature RNA transcripts, which can be translated. This process is carried out by a number of protein–RNA complexes, called small nuclear ribonucleoproteins (snRNPs), and numerous non-snRNP proteins that assemble and disassemble in a stepwise manner onto the pre-mRNA to form the spliceosome (for reviews, see Wahl et al. 2009; Will and Lührmann 2011; van der Feltz et al. 2012). At the early stage of its assembly, U1 and U2 snRNPs recognize the 5' splice site and the branch point sequence (BPS), respectively, and initiate the assembly of the spliceosome. The recent structures of an activated B-complex spliceosome (B<sup>act</sup>) (Rauhut et al. 2016; Yan et al. 2016), C-complex (Galej et al. 2016; Wan et al. 2016), and C\*-complex (Bertram et al. 2017; Fica et al. 2017; Yan et al. 2017) have provided important insights into the mechanism of pre-mRNA splicing. The branch helix

is formed when the pre-mRNA BPS pairs with U2 snRNA in U2 snRNP and is escorted into the active site of the spliceosome. In the B<sup>act</sup> complex, the branch helix is bound by U2 snRNP-specific protein complexes SF3a and SF3b. Prp2p then induces the dissociation of SF3a and SF3b, allowing Prp16p and step I factors to dock the branch helix into the active site (Ohrt et al. 2012; Galej et al. 2016) for the first transesterification reaction (branching).

U2 snRNP contains U2 snRNA, seven Sm proteins, Msl1p/Lea1p (U2B''/U2A' in humans), and two U2 specific protein complexes SF3a and SF3b. Early negative stain EM studies of isolated human U2 snRNP revealed a bipartite domain architecture (Krämer et al. 1999). U2B''/U2A' bind to stem-loop IV (Price et al. 1998) and form a large 3'-domain together with seven Sm proteins bound to the Sm site. SF3b binds toward the 5' end of U2 snRNA within 12S U2 snRNP forming the 5'-domain and allowing SF3a to bridge the two domains, forming the 17S U2 snRNP particle (Krämer et al. 1999). *Saccharomyces cerevisiae* SF3b comprises six proteins: Rse1p, Hsh155p, Cus1p, Hsh49p, Rds3p, and Ysf3p (Dziembowski et al. 2004). Human SF3b contains p14 in addition to the homologs of yeast SF3b proteins (Will et al. 2001).

<sup>3</sup>Present Address: Department of Biochemistry, Shimane University School of Medicine, Izumo 693-8501, Japan

Corresponding authors: kn@mrc-lmb.cam.ac.uk, mvroon@mrc-lmb.cam.ac.uk, cjo@mrc-lmb.cam.ac.uk

Article is online at <http://www.rnajournal.org/cgi/doi/10.1261/rna.059378.116>. Freely available online through the RNA Open Access option.

© 2017 van Roon et al. This article, published in RNA, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

CryoEM analysis of glutaraldehyde cross-linked human SF3b has been reported at better than 10 Å resolution (Golas et al. 2003). At this resolution, proteins with a known fold were fitted into the map but molecular details of their interactions were not revealed; the 22 HEAT repeats of SF3b155 were located on the outside of the complex, the RRM of p14 was fitted into density in the central cavity of the complex, and both RRMs of Hsh49p were tentatively assigned to two domains lying side-by-side on the surface of the cryoEM density (Golas et al. 2003).

High-resolution structures of the individual proteins of SF3b, and of its subcomplexes, are very valuable to gain further insights into the precise role of yeast SF3b in early stages of splicing. We reported the solution structure of Rds3p (van Roon et al. 2008), and recently the crystal structure of a core complex of human SF3b was published containing SF3b130 (Rse1p), SF3b155 (Hsh155p), SF3b14b (Rds3p), and SF3b10 (Ysf3p) (Cretu et al. 2016). In this structure the HEAT repeats of SF3b155p wrap around a bipartite scaffold comprising SF3b130, SF3b10, and SF3b14b. SF3b145 (Cus1p), SF3b49 (Hsh49p), and p14 are not present in these crystals. The solution structure of the first RRM of human SF3b49 has recently been published (Kuwasako et al. 2016).

Hsh49p is an essential protein and comprises two RNA recognition motifs (RRMs); both RRMs are required for yeast viability (Igel et al. 1998). RRM1 and RRM2 of Hsh49p show high degrees of similarity (35% identical and 55% similar) to their human counterparts. However, the proline-rich C terminus, seen in metazoans, is absent in yeast Hsh49p. Cus1p, a 50 kD protein, is also an essential protein with 43% identity and 65% similarity to the human sequence in comparable regions. It contains a domain of unknown function, DUF382 (Pfam: PF04037), and a proline-rich region (Fig. 1A). Residues 121–392 of Cus1p are required for yeast viability (Pauling et al. 2000). Cus1p was originally identified as a suppressor of cold-sensitive U2 snRNA mutations, which cause defects in spliceosome assembly (Wells et al. 1996). Yeast two-hybrid experiments indicated that Hsh49p binds Cus1p via its N-terminal RRM and that the binding region within Cus1p lies between residues 229 and 311 (Pauling et al. 2000). Recently it was shown that, in early spliceosomes, Hsh49p and Cus1p crosslink to the pre-mRNA in a region upstream of the branchpoint adenosine (Schneider et al. 2015). However, in the context of isolated U2 snRNP, cross-links were observed to the 5' end, SLI and SLIIB of U2 snRNA in humans (Dybkov et al. 2006).

In this paper, we present the crystal structures of yeast Hsh49p–Cus1(290–368)p (2.7 Å) and RRM1–Cus1(290–368)p (1.6 Å). We have also investigated the RNA-binding properties of Hsh49–Cus1(290–368)p with U2 snRNA. Finally, we fitted our complex structure into the EM density of the B<sup>act</sup> spliceosome complex published recently (Yan et al. 2016). Our study has provided new insight into the structure and function of Hsh49p–Cus1p in splicing.

## RESULTS

### Mapping of the Hsh49p-binding region of Cus1p and crystallization of Hsh49p–Cus1p complexes

*S. cerevisiae* Hsh49p was readily overexpressed as soluble protein in *E. coli*, while *S. cerevisiae* Cus1p was only obtained in small quantities and is highly susceptible to degradation in vivo. When the full-length Hsh49p and Cus1p are coexpressed in *E. coli*, the Hsh49p–Cus1p dimer tends to aggregate during purification. The smallest viable fragment of Cus1p spanning residues 121–392 (Pauling et al. 2000) was stably coexpressed with Hsh49p and the complex was purified to homogeneity. However, this complex did not yield any crystals despite extensive crystallization trials so the complex was subjected to limited proteolysis using trypsin. MALDI-TOF analysis and N-terminal sequencing of the proteolyzed fragments revealed that a Cus1p fragment comprising residues 285–355 could still stably bind to Hsh49p. We were able to coexpress and purify a complex of Cus1(290–350)p with full-length Hsh49p but crystals of this complex did not diffract beyond 7 Å. We then coexpressed and purified Hsh49p with several slightly longer Cus1p fragments (286–350, 290–355, 290–360, 290–363, and 290–368). The Hsh49p–Cus1(290–368)p construct yielded small needle-like trigonal crystals, which diffracted to a maximum resolution of 2.7 Å, but at this point we were unable to solve the structure by molecular replacement or isomorphous replacement, partly due to twinning. It was previously shown that RRM1 of Hsh49p binds to Cus1p (Igel et al. 1998), so we coexpressed RRM1 of Hsh49p with the same Cus1p peptide (290–368). This complex yielded crystals that diffracted to 1.6 Å. The structure was solved by molecular replacement using the RRM of cyclophilin33 as a search model (pdb 3MDF [Hom et al. 2010]) and refined to a final model with  $R/R_{\text{free}}$  of 19.5/21.7 (see Table 1). All four molecules in the asymmetric unit are very similar with average overall RMSDs for the main chain of 0.22 Å for RRM1 and 0.42 Å for Cus1(290–368)p. The structure of the complex between full-length Hsh49p and Cus1(290–368)p could then be solved by molecular replacement using the RRM1–Cus1(290–368)p structure and RRM2 of the polyadenylate-binding protein (pdb 1CVJ [Deo et al. 1999]). Restrained twin refinement led to a final model with  $R/R_{\text{free}}$  of 18.4/23.3 (see Table 1). All three molecules in the asymmetric unit are very similar, with average overall RMSDs for the main chain of 0.45 Å for Hsh49p and 0.23 Å for Cus1(290–368)p. The higher resolution crystal structure of RRM1–Cus1(290–368)p was used for the detailed analysis of the structure of Cus1(290–368)p and the interface between RRM1 and Cus1(290–368)p. Both crystal structures overlay very well with average overall main chain RMSDs for RRM1 of 0.47 Å and 0.68 Å for Cus1(290–368)p.



**TABLE 1.** Crystallographic data collection and refinement statistics

	Hsh49p–Cus1(290–368)p	RRM 1–Cus1(290–368)p
Data collection		
Space group	<i>P</i> 32	<i>P</i> 1
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	142.1, 142.1, 40.4	44.9, 47.8, 88.0
$\alpha$ , $\beta$ , $\gamma$ (°)	90.0, 90.0, 120.0	89.9, 83.6, 85.6
Wavelength (Å)	0.9310	0.9795
Resolution (outer shell) (Å)	35.52–2.70 (2.83–2.70)	44.47–1.65 (1.68–1.65)
Unique observations	25001 (3331)	84322 (4050)
Redundancy	4.2 (4.2)	2.3 (1.9)
Completeness	99.9 (100)	96.7 (95.0)
$R_{\text{merge}}^a$	0.128 (0.541)	0.063 (0.36)
$R_{\text{p.i.m.}}^b$	0.114 (0.481)	0.059 (0.36)
Mn[I]/SD[I]	13.0 (3.0)	13.3 (2.2)
Refinement		
Resolution (Å)	2.7	1.65
Number of reflections	23691	80084
$R_{\text{work}}/R_{\text{free}}$	0.184/0.233	0.195/0.217
Twinning fraction	0.64	-
Number of atoms	5799	4991
Mean B-factor	15.4	11.2
Luzatti coordinate error (Å) <sup>c</sup>	0.46	0.23
RMSD bond length (Å)	0.011	0.014
RMSD bond angle (°)	1.46	1.63
Ramachandran statistics for protein residues <sup>d</sup>		
In preferred regions	652 (95.6%)	537 (98.5%)
In allowed regions	28 (4.1%)	8 (1.5%)
Outliers	2 (0.3%)	0 (0.0%)

$$^a R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \overline{I(hkl)}|}{\sum_{hkl} \sum_i I_i(hkl)}$$

$$^b R_{\text{p.i.m.}} = \frac{\sum_{hkl} [1/(N-1)]^{1/2} \sum_i |I_i(hkl) - \overline{I(hkl)}|}{\sum_{hkl} \sum_i I_i(hkl)}$$

<sup>c</sup>Calculated in SFcheck (Collaborative Computational Project, Number 4, 1994).

<sup>d</sup>Calculated in Coot (Emsley et al. 2010).

## Structure of Hsh49p

Both RRM1s of Hsh49p display the canonical RRM structure (Fig. 1C) as reviewed in Muto and Yokoyama (2012). RRM1 has three solvent exposed conserved aromatic residues, Tyr12 in the RRM's RNP2 motif, Tyr52 and Phe54 in its RNP1 motif (Figs. 1B, 6D). The first residue of RNP1 is typically basic but is a glutamine (Gln50) in RRM1 of Hsh49p and its homologs (Fig. 1B). RRM2 has two solvent-exposed, conserved aromatic residues, Phe111 in RNP2 and Tyr152 in RNP1, but is atypical in yeast as it lacks the conserved glycine at the start of  $\beta$ 3 and the first aromatic residue of RNP1 is replaced by a cysteine (Cys150). However, there is an extra solvent-exposed tyrosine (Tyr154) within its RNP1 motif (Igel et al. 1998). Like many RRM1s, RRM2 contains a  $\beta$ -hairpin in the loop between  $\alpha$ 2 and  $\beta$ 4 (Cléry et al. 2008). The refined model contains residues 8–86 of RRM1, 106–185 of RRM2, and residues 187–205 of a C-terminal extension of RRM2. No density was observed for the first seven residues of RRM1, the linker connecting RRM1 and RRM2 and the last eight C-terminal residues of Hsh49p. In the absence of density for the linker it is not possible to conclude which RRM1 and RRM2 in the crystal lattice are connected covalently. The crys-

tals do contain full-length Hsh49p as visualized by SDS-PAGE (data not shown), and the Hsh49p–Cus1(290–368)p complex behaves as a monomer in solution as seen by analytical gel filtration (data not shown). The distance between the C terminus of RRM1 and the N terminus of RRM2 of the closest pair within the asymmetric unit is 17–18 Å; to symmetry-related RRM1s the distance is 30–39 Å. These distances can all be bridged by the disordered linker peptide so it is possible that a domain swap has occurred. In RRM2, the  $\beta$ 2/ $\beta$ 3 loop is disordered and density for one residue is missing from the linker between  $\beta$ 4 and the C-terminal extension. This C-terminal extension of RRM2 (186–205) forms an  $\alpha$ -helix, which folds back and wedges between RRM1 and RRM2 (Fig. 1C,D), interacting with RRM1 via the  $\beta$ 1/ $\alpha$ 1 and  $\alpha$ 2/ $\beta$ 4 loops and more extensively with the longer  $\beta$ 2/ $\beta$ 3 loop. The C-terminal helix binds to RRM2 via the  $\beta$ 3/ $\alpha$ 2 loop.

We overlaid Hsh49p with other proteins containing tandem RRM1s: hnRNP A1 (pdb 2LYV [Barraud and Allain 2013]), hnRNP L (RRM3 and 4, pdb 4QPT [Blatter et al. 2015]), CPEB1 (pdb 2MKH [Afroz et al. 2014]), Prp24p (pdb 2GO9, 2GHP [Bae et al. 2007]), and PTB domains 3 and 4 (pdb 2EVZ [Vitali et al. 2006]). The relative arrangement of the RRM1s of Hsh49p in the crystal is completely

different compared with these other structures. When we overlaid RRM1 of Hsh49p with RRM2 of Prp24 bound to U6 snRNA (pdb 4N0T [Montemayor et al. 2014]) on its canonical RNA binding surface, the C-terminal helix of Hsh49p would interfere with RNA binding to RRM1 in the Hsh49p–Cus1(290–368)p complex (Fig. 1D). It is possible that the arrangement of the RRM1 of Hsh49p with its C-terminal helix is stabilized by crystal packing interactions. Indeed crystal contacts are observed between the bottom of the C-terminal helix and residues in the linker between  $\alpha 1$  and  $\alpha 2$  of a symmetry-related Cus1(290–368)p (Fig. 1E).

### Structure of Cus1(290–368)p

Residues 290–350 of Cus1(290–368)p are ordered and included in the model (Figs. 1 and 2). This region coincides with the smallest stable fragment that binds to Hsh49p, as shown by our limited proteolysis and coexpression experiments. The remaining 18 C-terminal residues present in our construct are disordered. The Cus1(290–368) domain is folded with most secondary structure elements located at its periphery, and it covers almost the entire  $\alpha$ -helical side of RRM1 (Fig. 2A). The domain has two short  $\alpha$ -helical regions,  $\alpha 1$  (Gln296–Met302) and  $\alpha 2$  (Trp312–Ile319), that are linked by a short anti-parallel  $\beta$ -sheet with a  $3_{10}$ -helix in the linker connecting the two  $\beta$ -strands ( $\beta 1$ , Leu328–Ile330 and  $\beta 2$ , Gly346–Ile348). This is further stabilized by two  $\beta$ -turns formed between Arg290–Arg293 and Lys341–Val344.

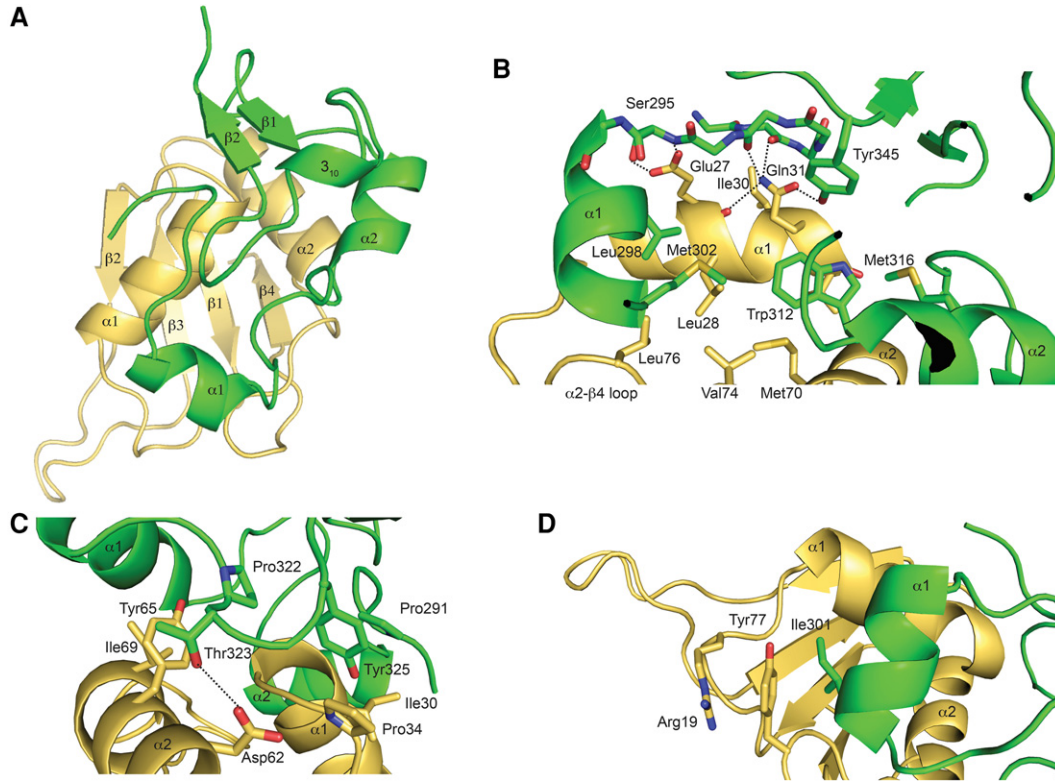
### Structure of the interface between RRM1 and Cus1(290–368)p

An extensive interface is created between Cus1(290–368)p and the  $\alpha$ -helical surface of RRM1 of Hsh49p. The  $\alpha 1$  of RRM1 is lying along a crevice on the surface of Cus1(290–368)p, making numerous hydrophobic contacts to Cus1(290–368)p via Leu28, Ile30, and Ile32. In addition, side chains of Glu27 and Gln31 form hydrogen bonds to both the main chain and side chains of Cus1(290–368)p (Fig. 2B). RRM1  $\alpha 2$  is located at the edge of the complex and in addition to the hydrophobic interactions via Ile69 (Fig. 2C) and Met70 (Fig. 2B) a hydrogen bond is present between the side chain of Asp62 and the side chain of Thr323 on Cus1(290–368)p (Fig. 2C). At the center of this interaction surface, a completely conserved tryptophan (Trp312), located at the bottom of  $\alpha 2$  of Cus1(290–368)p, lies flat in a hydrophobic cleft on the surface of RRM1 created by  $\alpha 1$ ,  $\alpha 2$ , and the extended linker between  $\alpha 2$  and  $\beta 4$  (Fig. 2B). The hydrophobic surface of Cus1(290–368)p is extended by Leu298, Met302 ( $\alpha 1$ ), and Met316 ( $\alpha 2$ ). In addition, Tyr345 is lying almost perpendicular to Trp312 and points toward RRM1, making a hydrogen bond with Gln31 on  $\alpha 1$  of Hsh49p (Fig. 2B). The interface between RRM1 and Cus1(290–368)p is stabilized by face-to-face hydrophobic stacking interactions. At the top of RRM1  $\alpha 1$ , Tyr325 of Cus1(290–368)p

is wedged in a hydrophobic pocket created by Pro291 on Cus1(290–368)p and Ile30 and Pro34 on RRM1 (Fig. 2C). On the same side of the complex, a similar interaction takes place where Tyr65 on  $\alpha 2$  of RRM1 sits in a pocket created by Pro322 and Thr323 on Cus1(290–368)p and Ile69 on RRM1 (Fig. 2C). On the other edge of the complex, Tyr77, located in the extended linker between  $\alpha 2$  and  $\beta 4$ , makes stacking interactions with Ile301 on  $\alpha 1$  of Cus1(290–368)p and Arg19 on RRM1 (Fig. 2D). All the key interacting residues on both Hsh49p and Cus1(290–368)p are highly conserved (Fig. 1B). The buried surface area of the interface is about 960  $\text{\AA}^2$  with a complex formation significance score of one, implying that this interface plays an essential role in complex formation (PDBe PISA [Krissinel and Henrick 2007]). This interface is highly hydrophobic in nature, which is consistent with the complex being resistant to salt concentrations of up to 1 M NaCl or 6% ammonium sulfate. In addition, there is an interaction between the  $\beta$ -sheet face of RRM1 and a symmetry related Cus1(290–368)p in the Hsh49p–Cus1(290–368)p crystals. The buried surface area of this interface is about 505  $\text{\AA}^2$ , but a complex formation significance of zero implies it is likely a result of crystal packing. Indeed, this interface is not present in the RRM1–Cus1(290–368)p crystals.

### Similarities to other structures

Recently, the structure of the complex of exosomal accessory factors RBM7 and ZCCHC8 proline-rich region was published (Falk et al. 2016). These proteins together with hMTR4 form the nuclear exosome targeting (NEXT) complex. It was suggested that this structure would be highly homologous to the proline-rich region of SAP145 (SAP145<sup>Pro</sup>), the human counterpart of Cus1p. We therefore expected it to be very similar to our structure as well. Indeed, when we overlaid our structure onto the RBM7–ZCCHC8 complex, the RMSD of the main chain was 2.9  $\text{\AA}$  for the RRM1 and 5.6  $\text{\AA}$  for Cus1(290–368)p between residues 290 and 324 (Fig. 3A). Interestingly, residues forming the interaction hotspot involving Hsh49p  $\alpha 1$ , the extended linker between  $\alpha 2$  and  $\beta 4$  and  $\alpha 1$  of Cus1(290–368)p (denoted patch 1 in the RBM7–ZCCHC8 structure) are highly conserved except for Leu25, which is replaced by Gln24 in Hsh49p (Fig. 3A). Its aliphatic side chain still provides hydrophobicity, but in addition a hydrogen bond with Gln297 of Cus1p further stabilizes the interaction. Another interesting difference is the replacement of the completely conserved Trp312 in Cus1p for Phe309 in ZCCHC8 (Fig. 3A). Most residues involving Hsh49p  $\alpha 2$  and the loop following  $\alpha 1$ , the N terminus of Cus1(290–368)p as well as  $\alpha 2$  and the  $3_{10}$  helix (denoted patch 2 in the RBM7–ZCCHC8 structure), are highly conserved with one remarkable exception: Phe286 in ZCCHC8 is sandwiched by Pro288, Trp322, and Pro35 of RBM7 in a face-to-face stacking array (Fig. 3B). In our structure, the N terminus of Cus1(290–368)p is pushed away by  $\beta 2$  in Cus1(290–368)p, and the position of Phe286 is taken by

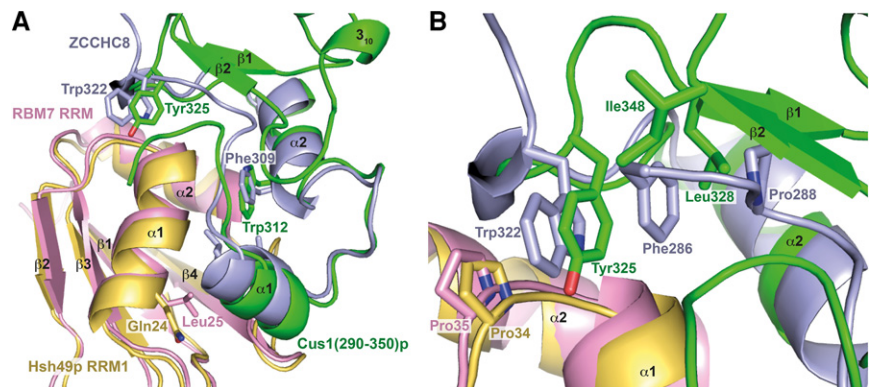


**FIGURE 2.** Overview of RRM1–Cus1(290–368)p interaction. (A) Overview of Cus1(290–368)p (green) binding to the  $\alpha$ -helical face of RRM1 (yellow). (B) Close up of the *center* region of the interaction surface. Cus1(290–368)p is lying in a hydrophobic crevice on RRM1. Key interacting residues are depicted in sticks and labeled. Hydrogen bonds are depicted as dashed lines. (C) Close up of the face-to-face arrays of Tyr65 and Tyr325 in Hsh49p and Cus1(290–368)p, respectively. (D) Close up of the face-to-face array of Tyr77.

Leu328 and Ile348. Interestingly, Phe286 is conserved in ZCCHC8 homologs but not in Cus1p from different species, where it is replaced by a basic residue (Fig. 1B). Trp322 is replaced by Tyr325 in Cus1(290–368)p, which only stacks face to face with Pro34 of Hsh49p and in addition forms a hydrogen bond with the main chain carbonyl of Ile30 creating extra stability.

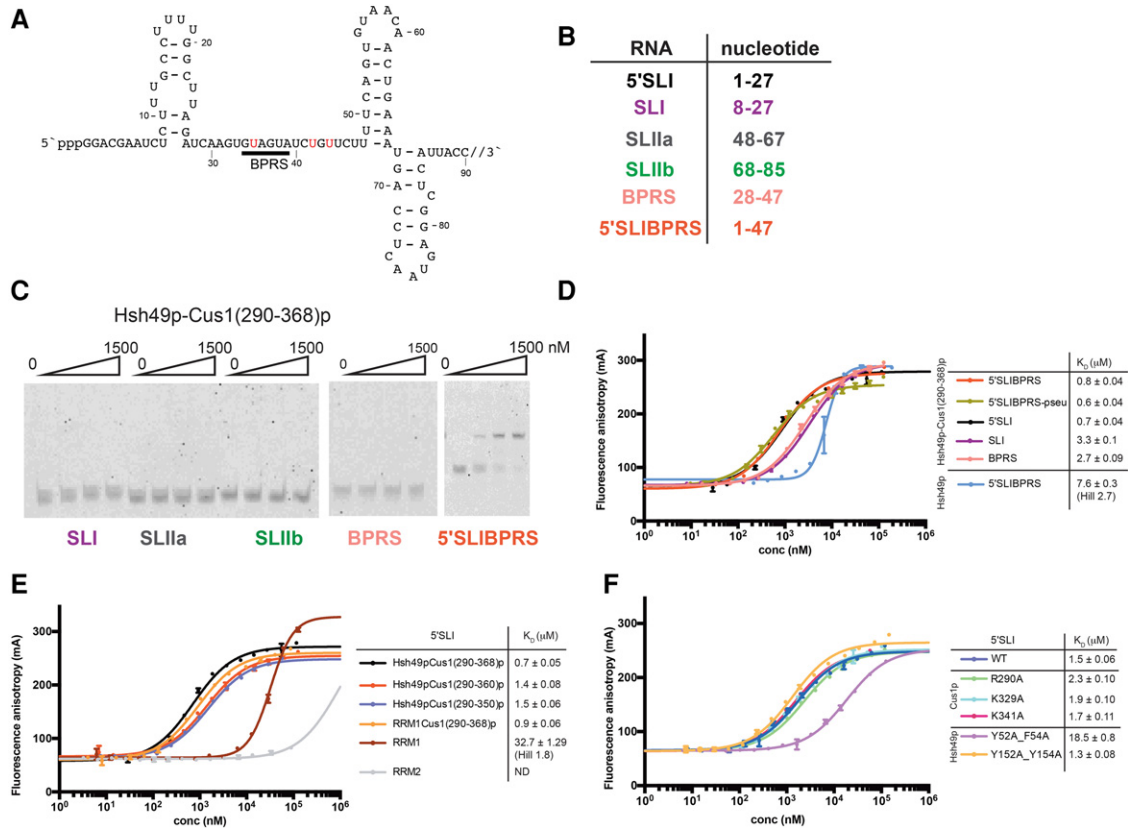
In addition, solution structures of both RRMs of the human homolog of Hsh49p, SF3b49, have been deposited to the protein databank (pdb codes 1X5U and 1X5T). The ordered regions of *S. cerevisiae* RRM1 and RRM2 align very well with their human homologs, with overall RMSDs for the main chain of 1.6 Å and 2.4 Å, respectively. In a recent paper, the mode of interaction between SF3b49 RRM1 and a fragment of the proline-rich region of SF3b145, residues 598–631 (corresponding to Cus1p residues 279–311) was studied by NMR (unreleased pdb) (Kuwasako et al. 2016). A model was proposed based on chemical shift mapping and NOESY

measurements in which SF3b145 residues 607–616 form an  $\alpha$ -helix upon interaction with RRM1, and this helix interacts with RRM1 of SF3b49 via  $\alpha$ 1 and Tyr80 (equivalent to Tyr77 in our structure). This is consistent with our structure,



**FIGURE 3.** Overlay of RRM1–Cus1(290–350)p with RBM7 RRM–ZCCHC8(286–324) ([Falk et al. 2016]; pdb 5LXR). The RMSD of the main chains of the RRMs is 2.9 Å. (A) Overview of the interface with Hsh49p in yellow, Cus1p in green, RBM7 in pink, and ZCCHC8 in blue. Some of the variant interface residues are shown in stick representation. (B) Close-up of “patch 2” showing the extensive stacking array around Phe286 in the RBM7–ZCCHC8 heterodimer compared to the less striking interactions seen in this region of the Hsh49p–Cus1p complex.





**FIGURE 5.** RNA-binding studies of Hsh49p–Cus1(290–368)p with different 5' end U2 snRNA oligos. (A) 5' end of *S. cerevisiae* U2 snRNA. The pppG at the 5' terminus is a result of the in vitro transcription, and the nucleotides denoted in red are post-transcriptionally modified to pseudouridine in vivo. (B) Summary of 5' U2 snRNA oligos used for bandshift assays. (C) Bandshift of different 5' U2 snRNA oligos. (D) Fluorescence anisotropy measurements of different RNA oligos with Hsh49p–Cus1(290–368)p or Hsh49p. In the absence of Cus1(290–368)p, Hsh49p binds significantly weaker. The error bars represent the SD of each data point calculated from three independent fluorescence anisotropy measurements. (E) Fluorescence anisotropy measurements of different Hsh49p–Cus1p constructs with 5'SLI oligo. As in D, it is immediately obvious that the presence of Cus1p proline-rich domain enhances the affinity. The data of RRM1 had to be fitted with a Hill coefficient of 1.8; in addition the higher maximum anisotropy shows that multiple RRM1 could be binding to the RNA. (F) Fluorescence anisotropy measurements of mutants of Hsh49p or Cus1(290–350)p with 5'SLI. Mutation of the canonical RNA-binding residues of RRM1 significantly impair RNA binding. In addition, the Cus1p Arg290Ala mutant also decreases the RNA-binding affinity.

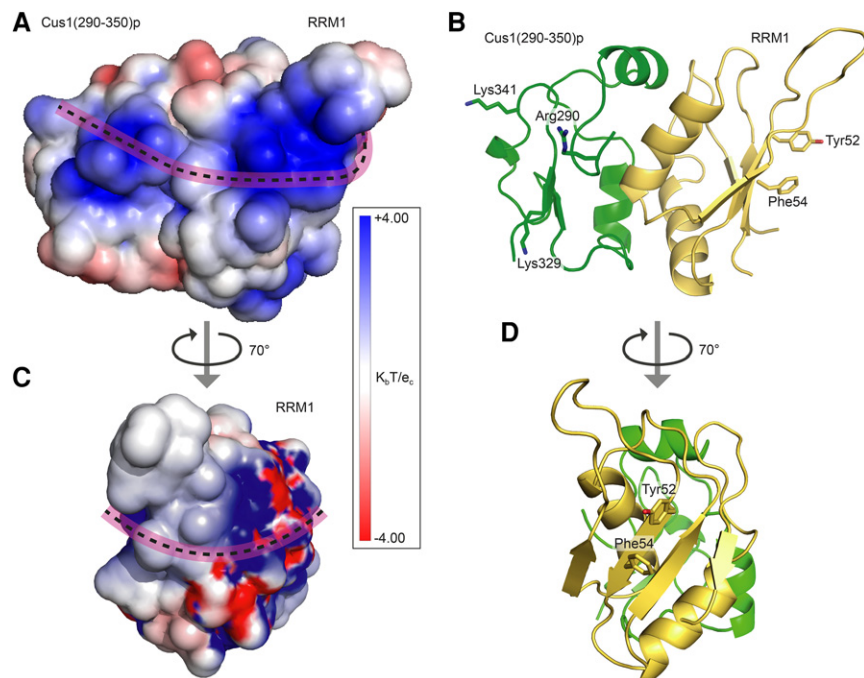
patch mutants Arg290Ala, Lys329Ala, and Lys341Ala in complex with wild-type full-length Hsh49p. From the graph in Figure 5F it is immediately clear that the RRM1 mutant has drastically decreased RNA affinity, about 10-fold compared to the wild type. As expected, the RRM2 mutant showed no effect, whereas there was a slight but significant decrease in the Cus1(290–350)p-Arg290Ala mutant, further supporting our hypothesis that Cus1(290–368)p directly contacts the RNA as well. Figure 6A and C show a possible path of the RNA, extending from the canonical RNA binding site of RRM1 to the basic patch of Cus1(290–350)p.

### Fitting of Hsh49p–Cus1(290–368)p complex into B<sup>act</sup> EM density

Recently, the coordinates and EM maps of an activated B-complex spliceosome (B<sup>act</sup>) were released (pdb 5GM6, EMD-9524) (Yan et al. 2016), and we have overlaid our crys-

tal structure with the B<sup>act</sup> model and fitted it into the electron density map. Interestingly, RRM1 fits very well into the density that was assigned as Hsh49p RRM1, but this fit would then position the Cus1(290–368)p fragment into the density that was assigned as RRM2 in the B<sup>act</sup> complex. The Cus1(290–368)p model fits this density much better than RRM2 (Fig. 7B), and in addition its N terminus now perfectly links to the C terminus of Cus1(131–289)p modeled in the EM density. This shows that the interface between Hsh49p and Cus1(290–368)p we found is indeed biologically relevant. However, this implies that RRM2 has to be placed elsewhere in the structure. There is unassigned density present to the left of RRM1 (Fig. 7C), which is close enough to be RRM2. Consequently, RRM2 is now located near stem-loop IIb of U2 snRNA, and indeed a crosslink has been observed between SAP49p and stem-loop IIb in human 17S U2 snRNP and purified A and B complexes (Dybkov et al. 2006). However, we did not detect any binding between RRM2





**FIGURE 6.** Electrostatic surface of Hsh49p RRM1–Cus1(290–350)p. Crystals contained Cus1(290–368)p but residues 351–368 were disordered. (A) The electrostatic surface of the interface displays an extended positively charged region (blue). The surface potential was calculated using the program PDB2PQ (Dolinsky et al. 2007) and APBS (Baker et al. 2001). A pink dotted line represents where RNA may bind this basic region as well as the  $\beta$ -sheet of RRM1 on the *right-hand* edge, which is the canonical RNA-binding surface, to partially explain Cus1p’s contribution to Hsh49p RNA binding. (B) Cartoon representation of the surface shown in A, upon which the residues mutated in the RNA-binding studies are indicated. (C) Rotated view of A showing the canonical RNA-binding surface. A possible RNA-binding site is indicated by the pink dotted line. (D) Cartoon representation of the surface shown in C.

and SLIIa or SLIIb by fluorescence anisotropy (data not shown). The flexible linker between RRM1 and RRM2 is about 20 amino acids long and should allow independent movement of the two RRMs, indicating that the relative orientation of both RRMs held together by the C-terminal helix in the crystal structure was indeed most likely due to crystal packing interactions.

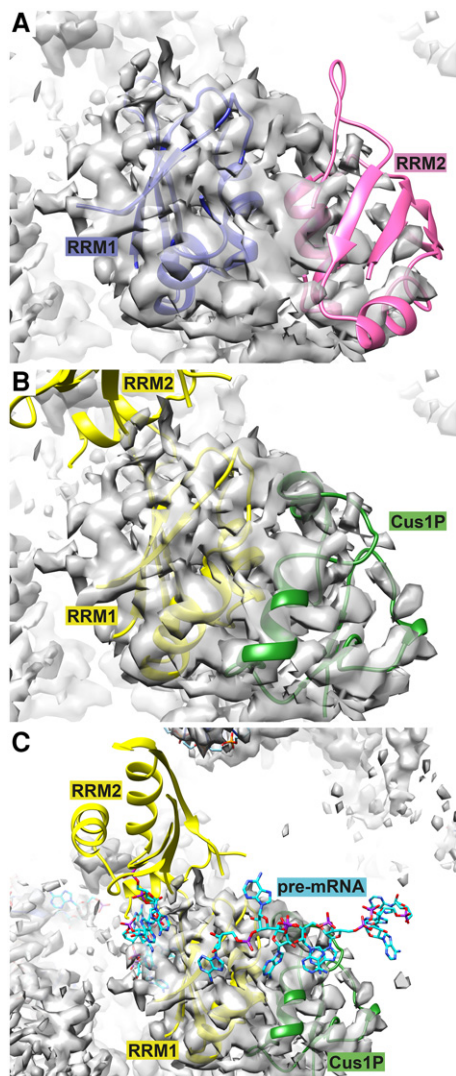
## DISCUSSION

In this article, we report the structure of Hsh49p in complex with the Hsh49p-binding domain of Cus1p. The entire  $\alpha$ -helical face of RRM1 is cradled by Cus1(290–368)p, leaving the opposite  $\beta$ -sheet side available to bind another protein partner or RNA. RRM2 is not involved in Cus1p binding, which is in agreement with previous pull-down studies (Igel et al. 1998). The ordered part of Cus1(290–368)p in the crystal, residues 290–350, corresponds to the minimum binding domain of Cus1p able to bind Hsh49p, as determined by limited proteolysis and coexpression studies.

A completely conserved tryptophan (Trp312) is located at the center of the interaction surface between Hsh49p and

Cus1(290–368)p. We have shown in our pull-down studies that Trp312 is required to bind Hsh49p. A critical role has been found for tryptophan residues in the interactions between RRMs that contain a U2AF homology motif (UHM) and their ligands, which contain a UHM ligand motif (ULM). The ULM tryptophan is inserted into a hydrophobic pocket between the two  $\alpha$ -helices of the UHM-containing RRM (Fig. 8), as reviewed in Kielkopf et al. (2004). This pocket is surrounded by negatively (or occasionally positively) charged residues (Elantak et al. 2010) with oppositely charged residues on the ligand surrounding the tryptophan. In addition, an Arg-X-Phe motif can be found at the bottom of  $\alpha$ 2 and the loop connecting  $\alpha$ 2– $\beta$ 4 on the UHM. When we overlaid our structure with the U2AF35–U2AF65 complex (pdb 1JMT [Kielkopf et al. 2001]), a canonical UHM–ULM complex, the conserved tryptophan Trp312 side chain in our structure is rotated by about 90°. It is not buried in a hydrophobic pocket as seen in the UHM–ULM interface but lies flat in a hydrophobic crevice on the surface of RRM1 (Fig. 8). Apart from the conserved tryptophan, our structure lacks the other important UHM–ULM features. Instead, it represents a recently discovered novel type of RRM protein interaction (Falk et al. 2016). In our structure an extensive and conserved hydrophobic interface is present. Mutations of residues at the center of the interface completely disrupted the complex as we found in our *in vitro* pull-down assays (Fig. 4). Our interface is similar to that recently observed in the RBM7–ZCCHC8 structure (Fig. 3; Falk et al. 2016). It was suggested that the mutually exclusive interaction of RBM7 with ZCCHC8 and SAP145<sup>Pro</sup> targets intronic sequences to the exosome. No homologs of the NEXT complex proteins were found in budding yeast, so it is unlikely that Cus1p is involved in a similar mechanism.

We found that the RNA-binding properties of Hsh49p originate from RRM1. Mutation of the canonical RNA-binding residues of RRM1 (Tyr52 and Phe54) severely diminished binding (Fig. 5F). Cus1p proline-rich domain enhances the binding to RNA and so does the presence of the 5' end of U2 snRNA. Neither the disordered linker between RRM1 and RRM2 nor the C-terminal  $\alpha$ -helix contribute to U2 snRNA oligo binding as Hsh49p–Cus1(290–368)p and RRM1–Cus1(290–368)p behave similarly in the fluorescence anisotropy experiments (Fig. 5E). An extended positive surface is present beyond the canonical RNA binding



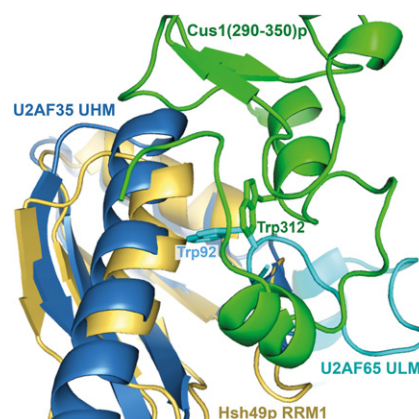
**FIGURE 7.** Comparison of Hsh49 modeled into EM density of yeast activated spliceosome ( $B^{\text{act}}$ ) ([Yan et al. 2016], EMD-9524) with the crystal structure presented here. (A) Originally, RRM1 (blue) and RRM2 (pink) of Hsh49 were modeled side-by-side into the map ([Yan et al. 2016], pdb 5GM6). (B) The Hsh49p–Cus1(290–368)p structure was fitted into the EM density by superposition on RRM1 of the model of Yan and colleagues (RMSD 1.23 Å). Cus1(290–368)p (green) fits well into the density ascribed to RRM2, but the RRM2 from the crystal structure (yellow) would locate in an empty region of the map. (C) RRM2 may fit in unassigned density seen in the *bottom left* of this panel. Some of the pre-mRNA has been modeled (Yan et al. 2016) and the nucleotides seen to the *left* of RRM1 partially occupy unassigned density across the face of its  $\beta$ -sheet. There is no clear density for the nucleotides that lie across the edge of RRM1 and Cus1(290–368)p.

site of RRM1 when it is in complex with Cus1(290–350)p (Fig. 6A). Based on our fluorescence anisotropy data from different Hsh49p–Cus1p constructs and mutants, we suggest that enhancement of the RNA interaction by the Cus1(290–368)p domain is due to its direct interaction with the RNA. Indeed, deletion of the last 18 residues of Cus1(290–368)p lowered the affinity twofold (Fig. 5E), and mutation of con-

served basic residue Arg290 to Ala of Cus1p(290–350)p had an additional small, but still significant effect (Fig. 5F). We suggest a possible path of the RNA along the surface of RRM1 and Cus1p (Fig. 6A,C). Enhancement of RNA interaction of an RRM bound to a protein on its  $\alpha$ -helical surface has also been observed in the case of U2B''–U2A', where the presence of a U2A' leucine rich repeat with a basic patch was required for U2B'' RRM to bind its cognate RNA (Price et al. 1998).

We were able to fit the RRM1–Cus1(290–368)p part of the Hsh49p–Cus1(290–368)p structure into the cryoEM map of the  $B^{\text{act}}$  complex (Yan et al. 2016), showing that the structure of our complex is of biological significance and that the relative arrangement of the RRM is due to crystal packing (Fig. 7). Unfortunately, the EM structure did not provide better insight into the RNA-binding region of Hsh49p–Cus1(290–368)p. Pre-mRNA was modeled along the top of RRM2 [which we have now reassigned as Cus1(290–368)p]; however, there is no density present for this RNA (Fig. 7C), so no conclusions can be drawn. In addition, there appears to be some extra density across the RNA-binding surface of RRM1; pre-mRNA, upstream of the branchpoint adenosine, has been modeled into it (Yan et al. 2016). Although it is only roughly fitted, this would be in agreement with the recent crosslinking study (Schneider et al. 2015). It is very likely that U2 snRNP proteins will interact differently with U2 snRNA in the 17S U2 snRNP and the spliceosome. The 5' end of U2 snRNA forms a duplex with U6 snRNA (stem-loop II) upon integration into the spliceosome. Hsh49p–Cus1p might play a role in protecting the 5' end of U2 snRNA until it is ready to make this interaction.

CryoEM has provided unprecedented insight into the molecular mechanism of pre-mRNA splicing (for review, see SH



**FIGURE 8.** Comparison of Hsh49p RRM1–Cus1(290–350)p interface with the interaction seen in the UHM–ULM complex of U2AF35–U2AF65 (pdb 1JMT). The structures are superposed on the RRM1 with an RMSD of 1.91 Å. In both cases a tryptophan is interacting with the  $\alpha$ -helical side of the RRM but the Cus1 Trp312 side chain is not inserted in a cleft between the  $\alpha$ -helices like the ULM Trp92 of U2AF65. Apart from the general proximity of the tryptophans, there is no resemblance between Cus1(290–350)p and U2AF65 ULM.

Scheres and K Nagai, unpubl.). Crystal structures of its components were either essential or greatly facilitated initial interpretation of the cryoEM density maps of the spliceosome. As exemplified here, crystal structures and functional studies of the components still prove invaluable in building complete atomic models of large complexes and understanding the interactions between the components that make up the entire assemblies.

## MATERIALS AND METHODS

### Cloning and overexpression of Hsh49p and Hsh49p–Cus1p complexes

Coding sequences for yeast *Hsh49* and *Cus1* were amplified by PCR from yeast genomic DNA. The *Hsh49*, *RRM1*, or *RRM2* sequences were cloned between the BamHI and EcoRI sites of the pRK172 vector (McLeod et al. 1987), which contained either a hexa-histidine and tobacco etch virus (TEV) protease site or a glutathione hexa-histidine and TEV site in frame and just upstream of the BamHI site. An NheI site was created just upstream of EcoRI, allowing the transfer of *Cus1* gene fragments, including a Shine–Dalgarno sequence from a pUC2 vector cut with XbaI and EcoRI. For the pull-down studies, a *Cus1*(290–350) gene fragment was cloned in frame with a glutathione (GST) hexa-histidine and TEV site and untagged *Hsh49*. Hsh49p and Hsh49p–Cus1(290–368)p complexes were expressed in BL21(DE3)RIL CodonPlus cells (Stratagene). The His-tagged and GST-His-tagged protein or complexes were expressed for 4 h at 37°C after induction with 0.5 mM IPTG. Harvested cells were resuspended in Nickel-NTA buffer A (20 mM Tris-HCl, pH 7.4, 500 mM Urea, 500 mM NaCl, 25 mM imidazole, and 5 mM β-mercaptoethanol), containing cComplete protease inhibitor cocktail (Roche) and lysed by sonication. The lysate was cleared by centrifugation, the supernatant loaded on Nickel-NTA agarose (Qiagen) equilibrated with Nickel-NTA buffer A, and the protein was eluted by a 25–500 mM linear gradient of imidazole (Ni-NTA buffer B). Peak fractions were incubated overnight with His-tagged TEV protease at room temperature while dialyzing against Ni-NTA buffer A. After complete cleavage the sample was reloaded on Ni-NTA agarose to remove His-tagged TEV protease, cleaved His-tag, and minor protein contaminants. The flowthrough containing Hsh49p or Hsh49p–Cus1(290–368)p was dialyzed into heparin buffer A (20 mM Tris-HCl, pH 7.4, 200 mM NaCl, and 5 mM β-mercaptoethanol), loaded onto heparin sepharose (GE Healthcare), and eluted with a 0.2–1 M linear gradient of NaCl. Peak fractions were pooled, and Hsh49p or Hsh49p–Cus1(290–368)p was concentrated and buffer exchanged into heparin buffer A using an Amicon-Ultra-15 concentrator (Millipore) with an exclusion size of 10 kDa.

### Limited proteolysis

HisHsh49p–Cus1(114–392)p (40 μM) was incubated with trypsin (10 μg/mL final concentration) for 10 min at 22°C, PMSF (1 mM final concentration) was added to block the reaction, and 50 μL Ni-NTA resin was added to pull-down His-tagged Hsh49p with the digested Cus1p fragments. After elution with Ni-NTA buffer B, the protein mixture was run on a 4%–20% SDS-PAGE gradient

gel and blotted for N-terminal sequencing, or the Coomassie-stained bands were cut out for MALDI-TOF analysis.

### Crystallization and structure determination

Crystals of Hsh49p–Cus1(290–368)p and RRM1–Cus1(290–368)p were grown using the sitting drop vapor diffusion method at 20°C. Protein solution at a concentration of 10–15 mg/mL was mixed with one volume of reservoir solution containing 2.45–2.65 M NaCl in 0.1 M sodium acetate pH 4.8–4.9 for Hsh49p–Cus1(290–368)p or 18% PEG4K, 0.1 M Tris-HCl pH 8.5, 45 mM Li<sub>2</sub>SO<sub>4</sub> for RRM1–Cus1(290–368)p. Crystals suitable for diffraction studies grew within 10 d after streak seeding for Hsh49p–Cus1(290–368)p and 3 d for RRM1–Cus1(290–368)p. A single crystal was transferred to a cryoprotectant solution, which contained the respective reservoir solution and 15%–20% glycerol, prior to flash-cooling by plunging into liquid nitrogen. Crystallographic data were collected at the European Synchrotron Radiation Facility beamline ID14-3 for Hsh49p–Cus1(290–368)p and beamline IO2 at Diamond Light Source for RRM1–Cus1(290–368)p. Data were indexed, scaled, and merged using the automated data reduction system xia2 using CCP4, distl, labelit, pointless, scala, and xds (Collaborative Computational Project, Number 4 1994; Evans 2006; Zhang et al. 2006; Kabsch 2010; Winter 2010).

The structure of RRM1 within the RRM1–Cus1(290–368)p complex was determined by molecular replacement using the program BALBES (Long et al. 2008) with the RRM of cyclophilin33 as a search model (pdb 3MDF). This solution was verified with SHELXE (Thorn and Sheldrick 2013), followed by automated building of Cus1(290–368)p by ARP/wARP into the electron density map (Langer et al. 2008); model building was carried out in Coot (Emsley et al. 2010). The structure was refined using Refmac5 (Murshudov et al. 2011) and the maps were improved using ARP/wARP (Perrakis et al. 1997). The structure of Hsh49p–Cus1(290–368)p was solved by molecular replacement using RRM1–Cus1(290–368)p as a search model with PHASER (McCoy et al. 2007), searching for three molecules in the asymmetric unit. After a partial solution was obtained, the three RRM2 portions could be located one by one using RRM2 of the polyadenylate binding protein (pdb 1CVJ [Deo et al. 1999]) as a search model. In between each molecular replacement step the model was refined with Refmac5 using jelly body with twin refinement. The model was optimized using restrained and twin refinement with Refmac5, map improvement with ARP/wARP, and manual rebuilding. Both structures were submitted to the PDB-redo server (Joosten et al. 2011) to finalize the refinement.

### Pull-down experiments

Primers to introduce mutations in the *Cus1*(290–350) and *Hsh49* gene fragments were designed and used as described in García-Nafria et al. (2016). Glutathione hexa-histidine Cus1(290–350)p constructs were coexpressed with Hsh49p as described above except that cells were grown at 15°C after induction. After harvesting, cells were frozen, thawed, and resuspended in Pulldown buffer (20 mM Tris-HCl, pH 7.4, 0.5 M NaCl, 5 mM 2-mercaptoethanol with one cComplete EDTA-free protease inhibitor cocktail tablet [Roche] per 100 mL) then lysed by sonication. After centrifugation at 72,000g for 30 min at 4°C, 800 μL of supernatant was mixed with 10 μL 2 M imidazole-HCl, pH 7.4 and 30 μL Ni-NTA resin. After 2 h incubation

with gentle mixing at 4°C, the resin was pelleted by centrifugation and washed twice with 300 µL Wash buffer (Pulldown buffer with 25 mM imidazole-HCl), then resuspended in 100 µL SDS-PAGE loading buffer, heated for 2 min at 90°C, and the protein released was analyzed by SDS-PAGE.

### In vitro RNA transcription and purification

Synthetic DNA templates for RNA transcription were ordered from Sigma-Aldrich and RNA was transcribed by standard methods (Milligan et al. 1987). RNA products were purified on a Resource Q column (GE Healthcare) equilibrated in 20 mM Tris-HCl pH 8.0, 50 mM NaCl, 8 mM MgCl<sub>2</sub>, 1 mM DTT and eluted with a linear gradient of 50–700 mM NaCl. Peak fractions were visualized on a 15% polyacrylamide denaturing gel containing 8 M Urea in TBE (89 mM Tris-Borate, 2 mM EDTA, pH 8.3), pooled and exchanged and concentrated in water. Alternatively, the RNA products were purified on a 20% polyacrylamide denaturing gel containing 8 M Urea in TBE, visualized by UV shadowing, excised from the gel, passively eluted, and exchanged and concentrated in water.

### RNA protein-binding assays

RNAs used in the binding studies were 3' end labeled with fluorescein, as described in Hardin et al. (2015). Electrophoretic mobility shift assay (EMSA) was used to qualitatively study the interaction of U2 snRNA-related RNAs with Hsh49p and Hsh49p–Cus1(290–368)p. Labeled RNA (10 nM) was mixed with increasing amounts of protein (0–1.5 µM) in RNA-binding buffer (20 mM K-Hepes, pH 7.5, 125 mM KCl, 1 mM EDTA, 5 mM DTT, 0.01% Igepal) in 10 µL final reaction volume in the presence of *E. coli* tRNA (500 ng) as the competitor. Binding reactions were incubated at 22°C for 20 min. Protein–RNA complexes were resolved on 12% polyacrylamide native gels in 0.5× TBE at 5 W for 75–90 min at 4°C. The gels were visualized on a Typhoon imager.

### Fluorescence anisotropy

Quantitative binding studies were carried out by fluorescence anisotropy. Labeled RNA (5 nM) was mixed with increasing amounts of protein (0–100 µM) and *E. coli* tRNA (100 ng) in RNA-binding buffer in 20 µL final reaction volume. The binding reactions were incubated at 22°C for 20 min, and the change in fluorescence anisotropy with increasing protein concentration was measured on a Pherastar with filter settings FP 485 520 520 for fluorescein. The data were analyzed by nonlinear regression curve fitting using Prism 6 software. Measurements were carried out in triplicate.

### DATA DEPOSITION

Atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.wwpdb.org [pdb codes 5LSB for Hsh49p–Cus1(290–368)p and 5LSL for RRM1–Cus1(290–368)p].

### ACKNOWLEDGMENTS

We thank Yasushi Kondo for collecting data on the RRM1–Cus1(290–368)p complex crystals and LMB mass spectrometry facility for MALDI-TOF analysis of proteolytic fragments, Dr. Stephen McLaughlin for helping with fluorescence anisotropy data analysis, and the staff at Diamond Light Source beamline I02 and European Synchrotron Radiation Facility beamline ID14-3 for their assistance. We also thank Elena Conti and Sebastian Falk for communicating their unpublished results. We are grateful to Lisa Strittmatter and Pei-chun Lin for critical reading of the manuscript. This project was supported by the Medical Research Council (MC\_U105184330), a Wellcome Trust Travelling fellowship, and an MRC career development fellowship (E.O.). B.Sp. was supported by Humanitas Research Foundation. B.Se. was supported by the CNRS, the CERBM-IGBMC and INSERM, the grant ANR-10-LABX-0030-INRT, a French State fund managed by the Agence Nationale de la Recherche under the frame program Investissements d'Avenir ANR-10-IDEX-0002-02, and the Ligue Contre le Cancer (Equipe Labellisée 2014).

Received September 23, 2016; accepted March 17, 2017.

### REFERENCES

- Afroz T, Skrisovska L, Belloc E, Guillén-Boixet J, Méndez R, Allain FHT. 2014. A fly trap mechanism provides sequence-specific RNA recognition by CPEB proteins. *Genes Dev* **28**: 1498–1514.
- Bae E, Reiter NJ, Bingman CA, Kwan SS, Lee D, Phillips GN, Butcher SE, Brow DA. 2007. Structure and interactions of the first three RNA recognition motifs of splicing factor prp24. *J Mol Biol* **367**: 1447–1458.
- Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci* **98**: 10037–10041.
- Barraud P, Allain FHT. 2013. Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: how the relative orientation between RRM5 influences the nucleic acid binding topology. *J Biomol NMR* **55**: 119–138.
- Bertram K, Agafonov DE, Liu W-T, Dybkov O, Will CL, Hartmuth K, Urlaub H, Kastner B, Stark H, Lührmann R. 2017. Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* **542**: 318–323.
- Blatter M, Dunin-Horkawicz S, Grishina I, Maris C, Thore S, Maier T, Bindereif A, Bujnicki JM, Allain FHT. 2015. The signature of the five-stranded vRRM fold defined by functional, structural and computational analysis of the hnRNP L protein. *J Mol Biol* **427**: 3001–3022.
- Cléry A, Blatter M, Allain FHT. 2008. RNA recognition motifs: boring? Not quite. *Curr Opin Struct Biol* **18**: 290–298.
- Collaborative Computational Project, Number 4. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**: 760–763.
- Cretu C, Schmitzová J, Ponce-Salvatierra A, Dybkov O, De Laurentis EI, Sharma K, Will CL, Urlaub H, Lührmann R, Pena V. 2016. Molecular architecture of SF3b and structural consequences of its cancer-related mutations. *Mol Cell* **64**: 307–319.
- Deo RC, Bonanno JB, Sonenberg N, Burley SK. 1999. Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**: 835–845.
- Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA. 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* **35**: W522–W525.

- Dybkov O, Will CL, Deckert J, Behzadnia N, Hartmuth K, Lührmann R. 2006. U2 snRNA-protein contacts in purified human 17S U2 snRNPs and in spliceosomal A and B complexes. *Mol Cell Biol* **26**: 2803–2816.
- Dziembowski A, Ventura AP, Rutz B, Caspary F, Faux C, Halgand F, Laprevote O, Séraphin B. 2004. Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. *EMBO J* **23**: 4847–4856.
- Elantak L, Wagner S, Herrmannová A, Karásková M, Rutkai E, Lukavský PJ, Valásek L. 2010. The indispensable N-terminal half of eIF3j/HCR1 cooperates with its structurally conserved binding partner eIF3b/PRT1-RRM and with eIF1A in stringent AUG selection. *J Mol Biol* **396**: 1097–1116.
- Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**: 486–501.
- Evans P. 2006. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**: 72–82.
- Falk S, Finogenova K, Melko M, Benda C, Lykke-Andersen S, Jensen TH, Conti E. 2016. Structure of the RBM7-ZCCHC8 core of the NEXT complex reveals connections to splicing factors. *Nat Commun* **7**: 13573.
- Fica SM, Oubridge C, Galej WP, Wilkinson ME, Bai X-C, Newman AJ, Nagai K. 2017. Structure of a spliceosome remodelled for exon ligation. *Nature* **542**: 377–380.
- Galej WP, Wilkinson ME, Fica SM, Oubridge C, Newman AJ, Nagai K. 2016. Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537**: 197–201.
- García-Nafria J, Watson JF, Greger IH. 2016. IVA cloning: a single-tube universal cloning system exploiting bacterial In Vivo Assembly. *Sci Rep* **6**: 27459.
- Golas MM, Sander B, Will CL, Lührmann R, Stark H. 2003. Molecular architecture of the multiprotein splicing factor SF3b. *Science* **300**: 980–984.
- Hardin JW, Warnasooriya C, Kondo Y, Nagai K, Rueda D. 2015. Assembly and dynamics of the U4/U6 di-snRNP by single-molecule FRET. *Nucleic Acids Res* **43**: 10963–10974.
- Hom RA, Chang P-Y, Roy S, Musselman CA, Glass KC, Selezneva AI, Gozani O, Ismagilov RF, Cleary ML, Kutateladze TG. 2010. Molecular mechanism of MLL PHD3 and RNA recognition by the Cyp33 RRM domain. *J Mol Biol* **400**: 145–154.
- Igel H, Wells S, Perriman R, Ares MJ. 1998. Conservation of structure and subunit interactions in yeast homologues of splicing factor 3b (SF3b) subunits. *RNA* **4**: 1–10.
- Joosten RP, Joosten K, Cohen SX, Vriend G, Perrakis A. 2011. Automatic rebuilding and optimization of crystallographic structures in the Protein Data Bank. *Bioinformatics* **27**: 3392–3398.
- Kabsch W. 2010. XDS. *Acta Crystallogr D Biol Crystallogr* **66**: 125–132.
- Kielkopf CL, Rodionova NA, Green MR, Burley SK. 2001. A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell* **106**: 595–605.
- Kielkopf CL, Lücke S, Green MR. 2004. U2AF homology motifs: protein recognition in the RRM world. *Genes Dev* **18**: 1513–1526.
- Krämer A, Grüter P, Gröning K, Kastner B. 1999. Combined biochemical and electron microscopic analyses reveal the architecture of the mammalian U2 snRNP. *J Cell Biol* **145**: 1355–1368.
- Krissinel E, Henrick K. 2007. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**: 774–797.
- Kuwasaki K, Nameki N, Tsuda K, Takahashi M, Sato A, Tochio N, Inoue M, Terada T, Kigawa T, Kobayashi N, et al. 2016. Solution structure of the first RNA recognition motif domain of human spliceosomal protein SF3b49 and its mode of interaction with a SF3b145 fragment. *Protein Sci* **26**: 280–291.
- Langer G, Cohen SX, Lamzin VS, Perrakis A. 2008. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **3**: 1171–1179.
- Long F, Vagin AA, Young P, Murshudov GN. 2008. BALBES: a molecular-replacement pipeline. *Acta Crystallogr D Biol Crystallogr* **64**: 125–132.
- Massenet S, Motorin Y, Lafontaine DL, Hurt EC, Grosjean H, Branlant C. 1999. Pseudouridine mapping in the *Saccharomyces cerevisiae* spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. *Mol Cell Biol* **19**: 2142–2154.
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *J Appl Crystallogr* **40**: 658–674.
- McLeod M, Stein M, Beach D. 1987. The product of the mei3+ gene, expressed under control of the mating-type locus, induces meiosis and sporulation in fission yeast. *EMBO J* **6**: 729–736.
- Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* **15**: 8783–8798.
- Montemayor EJ, Curran EC, Liao HH, Andrews KL, Treba CN, Butcher SE, Brow DA. 2014. Core structure of the U6 small nuclear ribonucleoprotein at 1.7-Å resolution. *Nat Struct Mol Biol* **21**: 544–551.
- Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**: 355–367.
- Muto Y, Yokoyama S. 2012. Structural insight into RNA recognition motif: versatile molecular Lego building blocks for biological systems. *WIREs RNA* **3**: 229–246.
- Ohr T, Prior M, Dannenberg J, Odenwälder P, Dybkov O, Rasche N, Schmitzová J, Gregor I, Fabrizio P, Enderlein J, et al. 2012. Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS. *RNA* **18**: 1244–1256.
- Pauling MH, McPheeters DS, Ares MJ. 2000. Functional Cus1p is found with Hsh155p in a multiprotein splicing factor associated with U2 snRNA. *Mol Cell Biol* **20**: 2176–2185.
- Perrakis A, Sixma TK, Wilson KS, Lamzin VS. 1997. wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr D Biol Crystallogr* **53**: 448–455.
- Price SR, Evans PR, Nagai K. 1998. Crystal structure of the spliceosomal U2B'-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* **394**: 645–650.
- Rauhut R, Fabrizio P, Dybkov O, Hartmuth K, Pena V, Chari A, Kumar V, Lee C-T, Urlaub H, Kastner B, et al. 2016. Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. *Science* **353**: 1399–1405.
- Schneider C, Agafonov DE, Schmitzová J, Hartmuth K, Fabrizio P, Lührmann R. 2015. Dynamic contacts of U2, RES, Cwc25, Prp8 and Prp45 proteins with the pre-mRNA branch-site and 3' splice site during catalytic activation and step 1 catalysis in yeast spliceosomes. *PLoS Genet* **11**: e1005539.
- Thorn A, Sheldrick GM. 2013. Extending molecular-replacement solutions with SHELXE. *Acta Crystallogr D Biol Crystallogr* **69**: 2251–2256.
- van der Feltz C, Anthony K, Brilot A, Pomeranz Krummel DA. 2012. Architecture of the spliceosome. *Biochemistry* **51**: 3321–3333.
- van Roon AMM, Loening NM, Obayashi E, Yang J-C, Newman AJ, Hernández H, Nagai K, Neuhaus D. 2008. Solution structure of the U2 snRNP protein Rds3p reveals a knotted zinc-finger motif. *Proc Natl Acad Sci* **105**: 9621–9626.
- Vitali F, Henning A, Oberstrass FC, Hargous Y, Auweter SD, Erat M, Allain FHT. 2006. Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling. *EMBO J* **25**: 150–162.
- Wahl MC, Will CL, Lührmann R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**: 701–718.
- Wan R, Yan C, Bai R, Huang G, Shi Y. 2016. Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* **353**: 895–904.
- Wells SE, Neville M, Haynes M, Wang J, Igel H, Ares M. 1996. CUS1, a suppressor of cold-sensitive U2 snRNA mutations, is a novel yeast splicing factor homologous to human SAP 145. *Genes Dev* **10**: 220–232.

- Will CL, Lührmann R. 2011. Spliceosome structure and function. *Cold Spring Harb Perspect Biol* **3**: a003707.
- Will CL, Schneider C, MacMillan AM, Katopodis NF, Neubauer G, Wilm M, Lührmann R, Query CC. 2001. A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO J* **20**: 4536–4546.
- Winter G. 2010. xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* **43**: 186–190.
- Yan C, Wan R, Bai R, Huang G, Shi Y. 2016. Structure of a yeast activated spliceosome at 3.5 Å resolution. *Science* **353**: 904–911.
- Yan C, Wan R, Bai R, Huang G, Shi Y. 2017. Structure of a yeast step II catalytically activated spliceosome. *Science* **355**: 149–155.
- Zhang Z, Sauter NK, van den Bedem H, Snell G, Deacon AM. 2006. Automated diffraction image analysis and spot searching for high-throughput crystal screening. *J Appl Crystallogr* **39**: 112–119.