



# RNA

A PUBLICATION OF THE RNA SOCIETY

## Structural basis for the dual U4 and U4atac snRNA-binding specificity of spliceosomal protein hPrp31

Sunbin Liu, Homa Ghalei, Reinhard Lührmann, et al.

*RNA* 2011 17: 1655-1663 originally published online July 22, 2011

Access the most recent version at doi:[10.1261/rna.2690611](https://doi.org/10.1261/rna.2690611)

---

**Supplemental  
Material**

<http://rnajournal.cshlp.org/content/suppl/2011/07/20/rna.2690611.DC1.html>

**References**

This article cites 21 articles, 7 of which can be accessed free at:  
<http://rnajournal.cshlp.org/content/17/9/1655.full.html#ref-list-1>

**Email alerting  
service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

---

---

To subscribe to *RNA* go to:  
<http://rnajournal.cshlp.org/subscriptions>

---

## REPORT

# Structural basis for the dual U4 and U4atac snRNA-binding specificity of spliceosomal protein hPrp31

SUNBIN LIU,<sup>1,3</sup> HOMA GHALEI,<sup>1,2,3</sup> REINHARD LÜHRMANN,<sup>2</sup> and MARKUS C. WAHL<sup>1,4</sup>

<sup>1</sup>Freie Universität Berlin, Fachbereich Biologie/Chemie/Pharmazie, Abteilung Strukturbiochemie, Takustraße 6, D-14195 Berlin, Germany

<sup>2</sup>Max-Planck-Institut für Biophysikalische Chemie, Abteilung Zelluläre Biochemie, Am Faßberg 11, D-37077 Göttingen, Germany

## ABSTRACT

Human proteins 15.5K and hPrp31 are components of the major spliceosomal U4 snRNP and of the minor spliceosomal U4atac snRNP. The two proteins bind to related 5'-stem loops (5'SLs) of the U4 and U4atac snRNAs in a strictly sequential fashion. The primary binding 15.5K protein binds at K-turns that exhibit identical sequences in the two snRNAs. However, RNA sequences contacted by the secondary binding hPrp31 differ in U4 and U4atac snRNAs, and the mechanism by which hPrp31 achieves its dual specificity is presently unknown. We show by crystal structure analysis that the capping pentaloops of the U4 and U4atac 5'SLs adopt different structures in the ternary hPrp31–15.5K–snRNA complexes. In U4atac snRNA, a noncanonical base pair forms across the pentaloop, based on which the RNA establishes more intimate interactions with hPrp31 compared with U4 snRNA. Stacking of hPrp31–His270 on the noncanonical base pair at the base of the U4atac pentaloop recapitulates intramolecular stabilizing principles known from the UUCG and GNRA families of RNA tetraloops. Rational mutagenesis corroborated the importance of the noncanonical base pair and the U4atac-specific hPrp31–RNA interactions for complex stability. The more extensive hPrp31–U4atac snRNA interactions are in line with a higher stability of the U4atac compared with the U4-based ternary complex seen in gel-shift assays, which may explain how U4atac snRNA can compete with the more abundant U4 snRNA for the same protein partners *in vivo*.

**Keywords:** major and minor spliceosome; pre-mRNA splicing; RNA–protein complex; U4 snRNP and U4atac snRNP; X-ray crystallography

## INTRODUCTION

In most eukaryotic genes, coding regions (exons) are interspersed with noncoding intervening sequences (introns). Such genes give rise to mosaic precursor messenger RNAs (pre-mRNAs), from which the introns have to be removed, while the exons have to be ligated in a process termed pre-mRNA splicing in order to produce mature mRNAs for protein biosynthesis on the ribosome. Each splicing event encompasses two sequential transesterification reactions (steps 1 and 2) carried out by a spliceosome, a large and highly dynamic ribonucleoprotein (RNP) enzyme (for review, see Wahl et al. 2009).

Some metazoa and plants harbor two types of spliceosomes. The major (U2-dependent) spliceosome is responsible for the vast majority of splicing events, while the minor (U12-dependent) spliceosome carries out the excision of

a small number of specialized introns (for review, see Patel and Steitz 2003). The main building blocks of major and minor spliceosomes are compositionally distinct, but functionally analogous small nuclear ribonucleoprotein particles (snRNPs). Each snRNP consists of a unique snRNA, seven common Sm or Sm-like (LSm) proteins, and a varying number of particle-specific proteins (for review, see Will and Lührmann 2001; Khusial et al. 2005). With the exception of U5 snRNP, which is common to both machineries, the analogous major and minor spliceosomal snRNPs differ either with respect to their snRNAs and their specific proteins (U1 and U2 vs. U11/U12 snRNPs) or with respect to their snRNAs only (U4/U6 vs. U4atac/U6atac snRNPs) (for review, see Will and Lührmann 2001, 2005). Presently, it is largely unknown how some spliceosomal proteins can associate with different snRNAs.

In U4atac/U6atac and U4/U6 snRNA duplexes a primary binding protein, 15.5K (Nottrott et al. 1999), shapes and stabilizes kink-turns (K-turns) (Klein et al. 2001) in the 5'-stem loops (5'SLs) of U4 and U4atac snRNAs (Vidovic et al. 2000; Cojocararu et al. 2005; Falb et al. 2010). Binding of 15.5K is a prerequisite for subsequent binding of

<sup>3</sup>These authors contributed equally to this work.

<sup>4</sup>Corresponding author.

E-mail [mwahl@chemie.fu-berlin.de](mailto:mwahl@chemie.fu-berlin.de).

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.2690611>.

protein hPrp31 (work herein focuses on the human (h) proteins, which are paradigmatic for other systems) and of the hPrp3–hPrp4–Cyp20 complex (Nottrott et al. 2002). It has been shown that hPrp31 uses a Nop domain to recognize a composite protein–RNA platform in the primary 15.5K–U4 snRNA complex, providing an explanation for the ordered binding of 15.5K and hPrp31 to the RNA (Liu et al. 2007). A similar situation was expected in the hPrp31–15.5K–U4atac snRNA complex (Liu et al. 2007). However, while the RNA residues directly contacted by 15.5K are identical in U4 and U4atac, it is unclear how hPrp31 copes with sequence differences in the RNA regions it binds.

Here, we describe crystal structures of hPrp31–15.5K–U4atac 5′SL complexes that address this question. hPrp31 interacts with the central region of the U4atac 5′SL in a sequence-independent fashion, as previously also seen in the analogous U4-based complex. However, a pentaloop region in U4atac snRNA interacts more intimately with hPrp31 side chains than observed in the U4-based complex. Site-directed mutagenesis revealed that the locally different interactions between hPrp31 and U4 or U4atac snRNA form a basis for differential stabilities of the U4 and U4atac snRNPs. These results suggest that higher

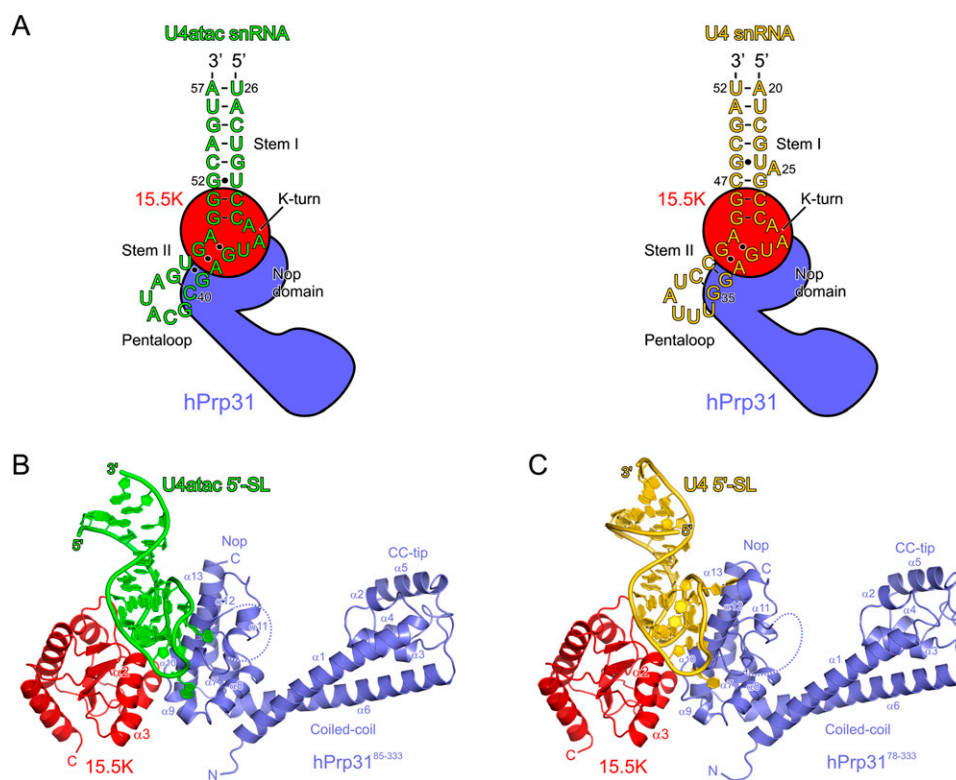
protein affinities may allow U4atac snRNA to compete with the more abundant U4 snRNA for the same protein partners.

## RESULTS

### Structure of a hPrp31<sup>85–333</sup>–15.5K–U4atac 5′SL complex

In order to compare binding of hPrp31 to 15.5K–U4 and 15.5K–U4atac complexes, we determined the crystal structure of a hPrp31–15.5K–U4atac 5′SL assembly. For crystallization, we produced a truncated hPrp31 protein that contained residues 85–333 (hPrp31<sup>85–333</sup>), lacking seven N-terminal residues that were disordered in the previous U4-based structure (Liu et al. 2007). The best diffracting crystals were obtained using an RNA oligomer comprising residues 28–55 of U4atac snRNA (Fig. 1A).

The structure of the hPrp31<sup>85–333</sup>–15.5K–U4atac 5′SL complex was solved by molecular replacement with the structure coordinates of the U4-based complex as a search model (Liu et al. 2007) (PDB ID 2OZB) and refined at 2.6 Å resolution to final  $R_{\text{work}}/R_{\text{free}}$  values of 19.9%/24.2% with good overall stereochemistry (Table 1; Supplemental Fig.



**FIGURE 1.** Overall structure of the hPrp31<sup>85–333</sup>–15.5K–U4atac complex. (A) Schematics of the 5′SL of U4 snRNA and U4atac snRNA. Binding of 15.5K (red) and hPrp31 (steel blue) is indicated schematically. (B) Overview of the hPrp31<sup>85–333</sup>–15.5K–U4atac 5′SL complex. hPrp31<sup>85–333</sup>, steel blue; 15.5K, red; U4atac RNA, green. The dashed line in hPrp31<sup>85–333</sup> indicates a disordered loop. Secondary structure elements are labeled. (C) The structure of the hPrp31<sup>78–333</sup>–15.5K–U4 5′SL complex (PDB ID: 2OZB) shown in the same orientation. Coloring as in B, except U4 RNA is shown in gold.

1). The present crystal form contained two complexes in an asymmetric unit. We traced residues 86–332 (residues 88–333) of hPrp31, residues 4–128 (4–128) of 15.5K, and all residues of the RNA in complex 1 (complex 2). In both complexes, a flexible loop in hPrp31 (residues 256–265) was disordered. All results presented below apply to both crystallographically independent assemblies.

The overall structure of the hPrp31<sup>85–333</sup>–15.5K–U4atac 5′SL complex is similar to that of the analogous U4-based complex (Fig. 1B,C; Liu et al. 2007). Each subunit interacts with the other two, giving rise to an overall triangular structure. The U4atac 5′SL comprises a K-turn in which residues A34, A35, and U36 are unpaired (Fig. 1A). A34 stacks on the terminal C33–G50 pair of stem I (the canonical stem), A35 stacks on the first G37–A49 pair of stem II (the noncanonical stem), and U36 is bulged out (Supplemental Fig. 2A) (for nomenclature of RNA elements refer to Fig. 1A). The canonical stem I is Watson–Crick paired except for a U•G wobble pair at positions 31–52 and adopts a standard A-form conformation. In stem II, two sheared G•A pairs (G37•A49 and A38•G48), which stabilize the K-turn structure by cross-strand A-stacking and A-minor interactions (Vidovic et al. 2000; Klein et al. 2001), are followed by a G39•U47 wobble pair and a Watson–Crick C40–G46 pair, to which an ordered pentaloop (nucleotides 41–45) is appended (Supplemental Fig. 2A).

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

Data collection	Form 1	Form 2
Wavelength (Å)	0.9800	0.9841
Temperature (K)	100	100
Space group	P2 <sub>1</sub>	C2
Unit cell parameters (Å, °)	a = 49.5 b = 111.3 c = 110.9 β = 101.8	a = 255.4 b = 105.3 c = 188.6 β = 127.5
Resolution (Å)	30.0–2.6 (2.7–2.6) <sup>a</sup>	30.0–3.3 (3.4–3.3)
Reflections		
Unique	34,765 (3849)	59,255 (4586)
Completeness (%)	100.0 (99.7)	98.9 (92.5)
Redundancy	5.7 (4.9)	4.8 (4.7)
I/σ(I)	19.5 (1.5)	7.9 (1.4)
R <sub>sym</sub> (I) <sup>b</sup>	0.055 (0.652)	0.112 (0.412)
Refinement	Form 1	Form 2
Resolution (Å)	30.0–2.6 (2.70–2.6)	30.0–3.30 (3.36–3.30)
Reflections		
Number	34,700 (2292)	59,249 (2519)
Test set (%)	5.0	5.0
R <sub>work</sub> <sup>c</sup>	19.8 (35.2)	19.7 (36.2)
R <sub>free</sub> <sup>c</sup>	24.1 (41.7)	23.8 (38.9)
ESU (Å) <sup>d</sup>	0.41	0.42
Refined residues/atoms		
Total	836/6896	1584/14,114
Protein	721/5627	1456/11,374
RNA	56/1202	128/2740
Water	57/57	-
Ligand (SO <sub>4</sub> <sup>2-</sup> )	2/10	-
Mean B-factors (Å <sup>2</sup> )		
Wilson	66.9	118.9
Protein	80.4	138.5
RNA	90.9	134.6
Water	72.0	-
Ligand (SO <sub>4</sub> <sup>2-</sup> )	120.6	-
Ramachandran plot <sup>e</sup> (%)		
Favored	98.45	88.9
Outliers	0	0.77
RMSD <sup>f</sup> from target geometry		
Bond lengths (Å)	0.008	0.009
Bond angles (°)	1.098	1.232
PDB ID	3SIU	3SIV

<sup>a</sup>Data for the highest resolution shell in parentheses.

<sup>b</sup>R<sub>sym</sub>(I) =  $\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ ; for n independent reflections and i observations of a given reflection;  $\langle I(hkl) \rangle$ —average intensity of the i observations.

<sup>c</sup>R =  $\sum_{hkl} |F_{obs} - F_{calc}| / \sum_{hkl} F_{obs}$ ; R<sub>work</sub>—hkl ∉ T; R<sub>free</sub>—hkl ∈ T; T—test set.

<sup>d</sup>ESU: estimated overall coordinate error based on maximum likelihood.

<sup>e</sup>Calculated with MolProbity (<http://molprobity.biochem.duke.edu/>).

<sup>f</sup>RMS: root-mean-square deviation.

### 15.5K–U4atac snRNA interactions

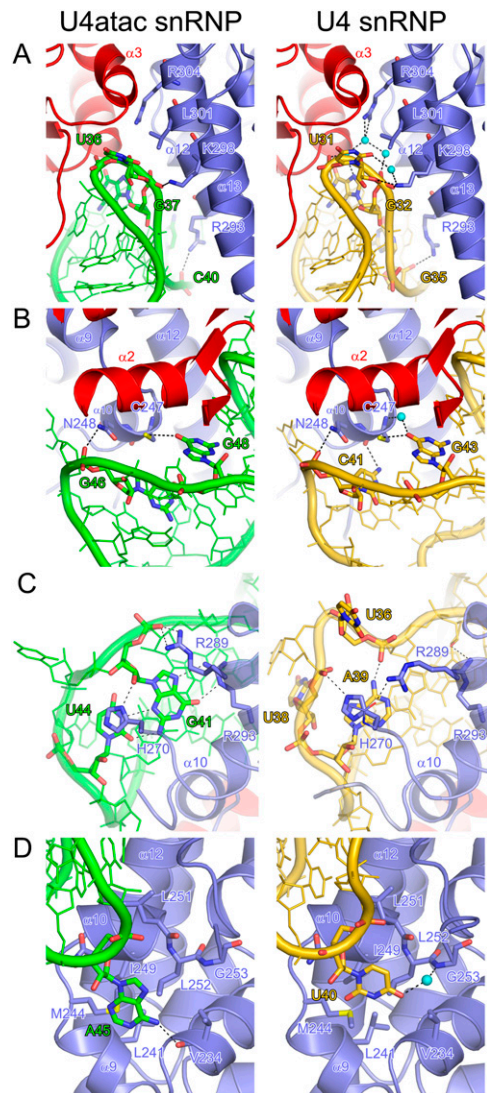
Protein 15.5K is a member of a family of K-turn-binding proteins (Koonin et al. 1994). As seen in previous structural analyses (Vidovic et al. 2000; Liu et al. 2007), the protein adopts a globular fold with a central, four-stranded, mixed β-sheet that is covered by two helices (α2 and α3) on one side and three (α1, α4, and α5) on the other (Supplemental Fig. 2A,B). 15.5K uses three loops (between strand β1 and helix α2, between strand β2 and helix α3, and between helix α4 and strand β4) to latch onto the RNA K-turn (Supplemental Fig. 2A,B). The backbone NH and carbonyl of Glu61 form hydrogen bonds with the O4 and N3 atoms of the bulged U36, respectively. Helix α2 is placed across the major groove of stem II of the RNA, allowing its N-terminal half to interact with the Hoogsteen edges of G37 and G48 of the two sheared G•A pairs. All nucleobases contacted by 15.5K are invariant between U4 and U4atac snRNAs, explaining how 15.5K can act as an assembly-initiating protein in both RNPs.

### hPrp31–15.5K interactions

hPrp31<sup>85–333</sup> exhibits an all-helical fold with three domains (Fig. 1B), including a coiled-coil region (residues 86–120 and 181–215), a globular insertion between the two branches of the coiled-coil (the CC-tip; residues 121–180), and a C-terminal Nop domain (residues 216–333). A flat surface of the Nop domain formed by helices  $\alpha 9$ ,  $\alpha 10$ ,  $\alpha 12$ , and  $\alpha 13$  interacts with a composite protein–RNA platform of the 15.5K–U4atac 5'SL complex (Fig. 1B). The interactions between hPrp31 and 15.5K are conserved between the U4 and the U4atac-based complexes (Supplemental Fig. 2C,D). Helices  $\alpha 2$  and  $\alpha 3$  of 15.5K lie across the  $\alpha 9$ – $\alpha 10$  loop and helix  $\alpha 12$  of hPrp31. The proteins interact via hydrophobic contacts, electrostatic complementarity, and a limited number of hydrogen bonds. For example, Lys9 of 15.5K engages in an ionic interaction with Glu310 of Prp31, the carboxamide group of Asn40 of 15.5K interacts with the guanidinium group of Arg304, and Leu67 of 15.5K forms a hydrophobic contact with Phe308 of hPrp31 (Supplemental Fig. 2C,D).

### Similarities in hPrp31–snRNA interactions in U4 and U4atac snRNPs

Residues of helices  $\alpha 12$  and  $\alpha 13$  of Prp31 bind the K-turn at the side not occupied by 15.5K, engaging in sequence-independent electrostatic and van-der-Waals contacts to the RNA as previously seen in the U4-based complex (Fig. 2A). Water molecules that were seen to mediate interactions in the hPrp31<sup>78–333</sup>–15.5K–U4 5'SL complex were not revealed at the somewhat lower resolution of the present structure. The N-terminal half of helix  $\alpha 12$  of hPrp31 additionally interacts with the backbone of the descending (5') branch of stem II of U4atac snRNA (Fig. 2B). The major groove of stem II and the pentaloop curl around the short helix  $\alpha 10$  (Fig. 2B–D). Helix  $\alpha 10$  fits snugly to the RNA, but only a single hydrogen bond forms between the side chain of Asn248 and O3' of A45 (Fig. 2B). One sequence-specific contact between the backbone carbonyl of Cys247 of hPrp31 and C41 of U4 snRNA (Liu et al. 2007) is lost in the U4atac-based complex, in which C41 is replaced by G46 (Fig. 2B). The loop following helix  $\alpha 10$  of hPrp31 reaches out to the tip of the RNA pentaloop, positioning the side chain of His270 on top of U44 (Fig. 2C). The His270–U44 stack explains a previously observed UV-induced cross-link between these residues (Kühn-Hölsken et al. 2005). In addition, His270 laterally hydrogen bonds to backbone portions of the pentaloop, thereby stabilizing its conformation (Fig. 2C). The terminal nucleotide of the pentaloop, A45, is rotated outward and lies on a hydrophobic patch formed by residues of helices  $\alpha 8$ ,  $\alpha 9$ , and  $\alpha 10$  of hPrp31 (Fig. 2D). Its N6 atom hydrogen bonds to the backbone carbonyl of Val234, replacing a water-mediated hydrogen bond seen for the equivalent U40 of the U4 snRNA-based complex (Fig. 2D).



**FIGURE 2.** Comparison of Prp31–U4atac and Prp31–U4 interactions. Ribbon plots showing details of the interaction of hPrp31 with the U4atac (*left*) and U4 snRNA (*right*). (A) Interaction at the K-turn region. (B) Interaction at the stem II region. (C) Interaction at the terminus of the pentaloop. (D) Interaction at the terminus of the pentaloop. Coloring as before. Selected interacting residues are shown as sticks, colored by atom type and labeled. Carbon and phosphorus—as the respective molecule; nitrogen, blue; oxygen, red; sulfur, yellow. Dashed lines indicate hydrogen bonds or salt bridges. Cyan spheres, water molecules.

### A noncanonical base pair in the pentaloop of the U4atac 5'SL gives rise to novel hPrp31–snRNA contacts in U4atac snRNP

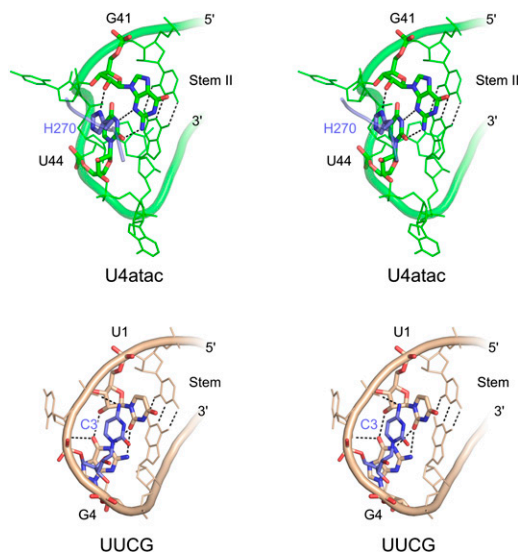
The major and minor spliceosomal complexes adopt substantially different structures at the capping pentaloops of the RNAs upon hPrp31 binding. In the U4-based structure, the penultimate nucleotide of the RNA pentaloop, A39, is turned inward and engages in a single cross-loop hydrogen bond to the phosphate of U36 (Fig. 2C). In the U4atac-

based structure, the corresponding U44 adopts a similar conformation. However, in addition, the sugar edge of the first loop nucleotide, G41, engages in noncanonical base-pairing with the Watson–Crick face of U44 (Fig. 2C). This configuration leaves the Watson–Crick and Hoogsteen edges of G41 available for additional interactions (Fig. 2C). The Watson–Crick side of G41 is in van-der-Waals contact with the surface of helix  $\alpha$ 10 of the Nop domain (Fig. 2C), consistent with our modeling studies on the hPrp31–15.5K–U4 5′SL complex, which had suggested that helix  $\alpha$ 10 would limit the number of Watson–Crick base pairs of stem II (Liu et al. 2007). Via its Hoogsteen edge, G41 interacts with the guanidinium group of Arg293, which originates from helix  $\alpha$ 12 (Fig. 2C). In addition, a second arginine, Arg289, from helix  $\alpha$ 12, engages in a salt bridge to the phosphate group of G41 (Fig. 2C). In the U4-based complex, the first loop nucleotide fails to engage in base-pairing analogous to G41 of U4atac snRNA, and the corresponding interactions with Arg289 and Arg293 are consequently lacking (Fig. 2C).

Notably, the stacking of hPrp31–His270 on U44 (see above) in combination with the noncanonical G41–U44 base pair at the base of the pentaloop resemble stabilizing principles known from the UUCG and GNRA families of thermodynamically very stable RNA tetraloops (Cheong et al. 1990; Heus and Pardi 1991). In these tetraloops, the sugar edge of the first loop residue is base paired with the Watson–Crick face (UUCG) or the Hoogsteen face (GNRA) of the penultimate loop residue (Fig. 3). In addition, an intervening loop nucleotide stacks on top of the non-canonical base pair in these tetraloops, equivalent to the stacking of hPrp31–His270 in the U4atac complex (Fig. 3). These comparisons suggest that the hPrp31–U4atac pentaloop contacts also represent a thermodynamically very stable interaction.

### The noncanonical cross-loop base pair and its interactions with hPrp31 are retained in an RNA domain-swapped duplex

Reconstitution of hPrp31<sup>85–333</sup>–15.5K–U4atac 5′SL complexes invariably led to two fractions, which migrated differently in gel-filtration runs (Supplemental Fig. 3A). The slower migrating fraction corresponded to the structure described above. We also were able to determine the crystal structure of a faster migrating species (obtained with an oligomer comprising residues 26–57 of U4atac snRNA) (Table 1). In the latter structure, we observed a duplicated complex with a domain swap in the U4atac 5′SL. Instead of folding back onto itself to form a structure with a capping pentaloop, two RNA molecules were aligned in an antiparallel fashion in this complex, forming a duplex with an internal loop (Supplemental Fig. 3B). Significantly, the structure of each half-side of the internal loop closely resembles the structure of the apical pentaloop of the mono-



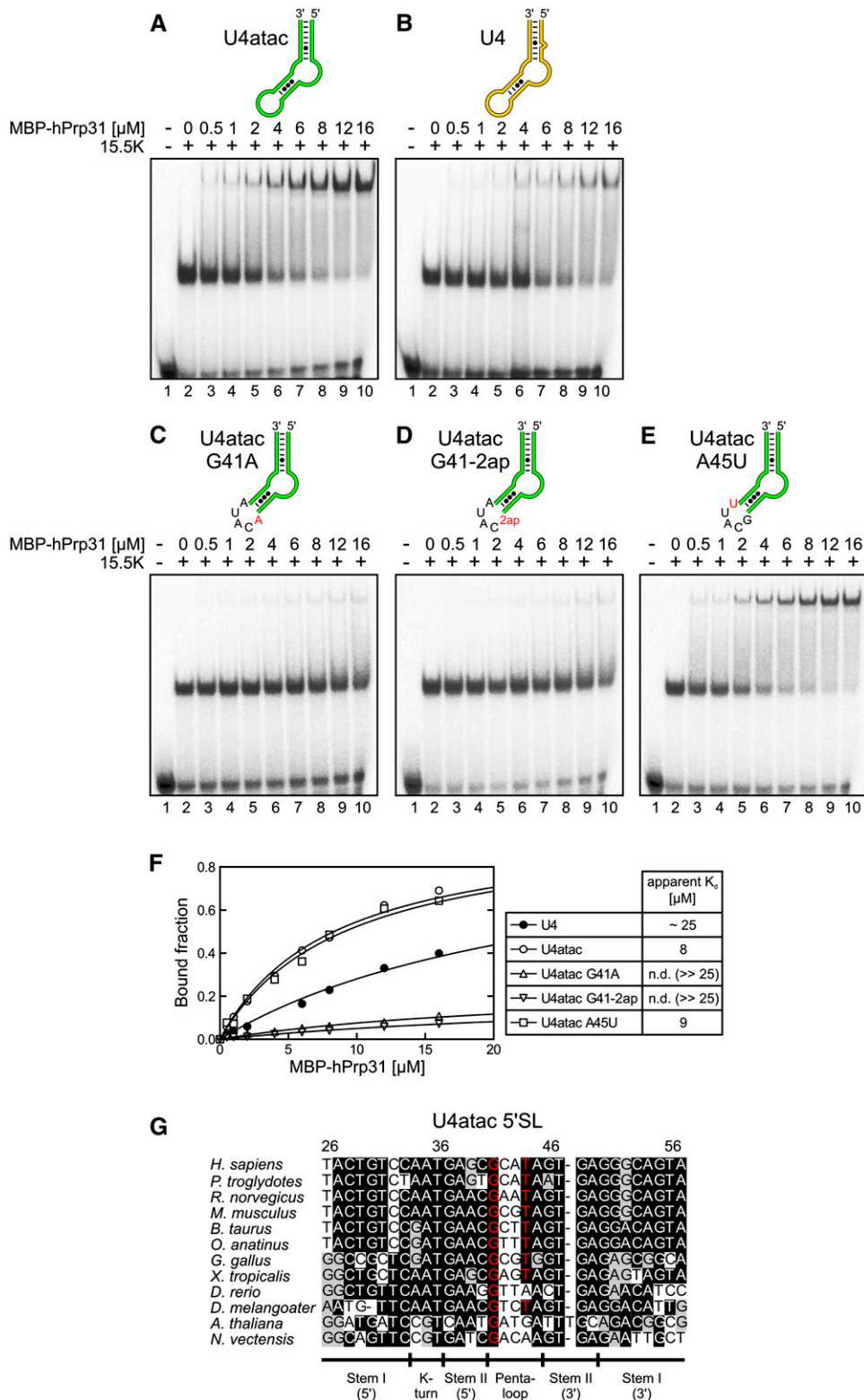
**FIGURE 3.** Comparison of the U4atac pentaloop and a UUCG tetraloop. Stereo ribbon plot of the U4atac pentaloop and His270 of hPrp31 (*top*) compared with the organization of a UUCG RNA tetraloop (*bottom*; PDB ID: 1F7Y). Coloring as before; tetraloop RNA, beige. G41 and U44 of the U4atac RNA, His270 of hPrp31, and the equivalent elements of the tetraloop are shown as sticks, colored by atom type and labeled. The C3 nucleotide of the tetraloop, which takes the equivalent position of His270, is also colored steel blue. Dashed lines indicate hydrogen bonds. Rotated 120° about the vertical axis compared with Figure 1B.

meric complex. In particular, the noncanonical U44–G41 base pair is retained as an interstrand base pair in the domain-swapped dimer (Supplemental Fig. 3B), which engages in identical interactions with hPrp31. This observation further supports the notion that the combined intra- and intermolecular interactions at the U4atac pentaloop are thermodynamically very stable.

### Different interactions of hPrp31 at the U4 and U4atac pentaloops are reflected in differential stabilities of U4 and U4atac-based snRNPs

Since a more intimate hPrp31–snRNA interaction is achieved in the U4atac-based complex, we predicted that the overall stability of this complex should be enhanced compared with the U4-based complex and tested the prediction using electrophoretic mobility-shift assays (EMSA) (Fig. 4A–F). Indeed, these experiments revealed a higher affinity of hPrp31 to the 15.5K–U4atac 5′SL complex (Fig. 4A,F; apparent  $K_d \sim 8 \mu\text{M}$ ) compared with the U4-based complex (Fig. 4B,F; apparent  $K_d \sim 25 \mu\text{M}$ ). This result is in agreement with the previous observation that truncations of stem I, which were deleterious to the stability of the hPrp31–15.5K–U4 5′SL complex, had little effect in the analogous U4atac-based complex (Schultz et al. 2006).

We next used mutational analysis to directly test whether the particular interactions at the RNA pentaloop indeed



**FIGURE 4.** Differential RNP stabilities. (A,B) Binding of a maltose-binding protein (MBP)–hPrp31 fusion to binary complexes of protein 15.5K and U4atac (A) or U4 (B) 5'SLs monitored by EMSA. Increasing amounts of MBP–hPrp31 (0, 0.5, 1, 2, 4, 6, 8, 12, or 16  $\mu\text{M}$  final concentration) were added to [ $^{32}\text{P}$ ]-5'-end labeled RNA oligonucleotides and 1  $\mu\text{M}$  15.5K protein. The higher stability of the U4atac-based ternary complex is apparent from the hPrp31-dependent super-shift and the disappearance of the binary complex at lower hPrp31 concentrations compared with the U4-based complex. (C–E) Testing of the effects of mutations in the U4atac snRNA pentaloop on the stability of the ternary complex using EMSA. Conditions as in A and B. Disruption of the noncanonical G41–U44 cross-loop base pair (G41A mutation) strongly attenuates ternary complex formation (C). Removal of one G41–R293 hydrogen bond by deletion of the O6 atom (G41-2ap mutation) severely weakens ternary complex formation (D). Change of the terminal, outward oriented nucleotide (A45U mutation) has no effect on the stability of the ternary complex (E). 2ap, 2-amino purine. Mutated positions are shown in red in the schematics of the RNAs above the gel. (F) Quantification of the band shifts shown in A–E, and deduction of the apparent  $K_d$ 's. (n.d.) Not determined. (G) Multiple sequence alignment of U4atac 5'SLs. Nucleotides engaged in a noncanonical cross-loop base pair in human U4atac snRNP are shown in red. Numbering above the alignment refers to the human U4atac sequence. Structural elements of the 5'SLs are shown below the alignments. Species: *Homo sapiens*, *Pan troglodytes*, *Rattus norvegicus*, *Mus musculus*, *Bos taurus*, *Ornithorhynchus anatinus*, *Gallus gallus*, *Xenopus tropicalis*, *Danio rerio*, *Drosophila melanogaster*, *Arabidopsis thaliana*, and *Nematostella vectensis*.

contributed significantly to the stability of the U4atac-based complex. Replacing G41 by an adenine is predicted to abolish a hydrogen bond to U44 and to reduce or abolish the interaction with Arg293 (Fig. 2C). Consistently, strongly reduced ternary complex formation was seen with an RNA bearing a G41A mutation (Fig. 4C,F; apparent  $K_d \gg 25 \mu\text{M}$ ). We specifically tested the contribution of the G41(O6)–Arg293 hydrogen bond to complex stability by replacing G41 with 2-amino purine (2ap), which is lacking an exocyclic O6 oxygen. Again, strongly reduced ternary complex formation was seen with RNAs bearing 2ap at position 41 (Fig. 4D,F; apparent  $K_d \gg 25 \mu\text{M}$ ). Finally, we tested the role of the terminal pentaloop nucleotide (U40 of U4 snRNA; A45 of U4atac snRNA), which, in both the U4 and U4atac-based structures, stacks on one flank of the hPrp31 Nop domain (Fig. 2D). Exchanging A45 of the U4atac pentaloop for a uridine as seen in U4 had little effect on hPrp31 binding (Fig. 4E,F; apparent  $K_d \sim 9 \mu\text{M}$ ). Thus, the water-mediated hydrogen bond to U40 of U4 snRNA (see above) apparently compensates for the larger stacking surface and the direct hydrogen bond of A45 of U4atac snRNA.

To test whether the noncanonical base pair is a common feature in U4atac snRNPs, we constructed a multiple-sequence alignment of U4atac 5'SLs (Fig. 4G). G41 is invariant in all tested U4atac snRNAs. U44 is conserved in the vast majority of higher animals (except *Danio rerio* in our alignment), but not in *Cnidaria* (*Nematostella vectensis*) or plants (*Arabidopsis thaliana*). Notably, the two intervening loop nucleotides, which do not engage in specific interactions in our structures, are not similarly conserved. This analysis suggests that the noncanonical base pair we observed in the framework of human U4atac snRNP is conserved in the majority of U4atac snRNPs and confers increased stability as an important property on these complexes.

## DISCUSSION

Cell homeostasis crucially depends on specific molecular recognition. However, some RNA-binding proteins have the ability to associate with more than one RNA target. In order to understand such plasticity in protein–RNA recognition events, structural information is required. An intriguing example, in which the same set of proteins binds to RNAs differing in sequence, is provided by the major spliceosomal U4 and minor spliceosomal U4atac snRNPs. Here, we have determined the crystal structure of a hPrp31–15.5K–U4atac 5'SL complex. Based on the comparison with the structure of the analogous U4-based complex (Liu et al. 2007), we could outline the molecular mechanisms underlying the dual-binding specificities of proteins 15.5K and hPrp31. Furthermore, we showed that the U4atac-based complex is more stable compared with the U4-based complex and traced the factors underlying this difference in stabilities.

Since U4 and U4atac snRNAs exhibit identical K-turn regions (Fig. 1A), the binding of 15.5K, which recognizes these K-turns, to either RNA poses no conundrum. As expected, we did not see any difference in the binding of either U4 or U4atac snRNAs by 15.5K. hPrp31 attaches as a secondary binding protein to these complexes. Apart from contacting RNA elements that are identical in U4 and U4atac snRNAs (K-turns), it also interacts with RNA regions that differ in sequence in the two snRNAs (stems II and pentaloops). Our results demonstrate that hPrp31 binds to RNP platforms involving U4 or U4atac snRNAs by (1) minimizing sequence-specific contacts to the K-turns and stem II regions, and (2) by inducing or recognizing different RNA-specific structures in the more flexible pentaloop regions of the snRNAs. A noncanonical G41–U44 base pair forms in the U4atac pentaloop, which stacks on stem II and allows additional hydrogen-bonding and salt-bridge interactions to hPrp31 side chains not available in the U4-based complex (Fig. 2C). The more intimate interaction of hPrp31 with U4atac snRNA is one reason for the higher stability of the U4atac snRNP.

The higher stability of the U4atac compared with the U4-based complex may be important for assembly of the minor spliceosome in vivo. In metazoan cells, the minor spliceosomal U4atac snRNA is about one order of magnitude less abundant compared with U4 snRNA ( $\sim 2000$  vs.  $\sim 20,000$  copies per cell) (Tycowski et al. 2006). The more stable protein binding revealed in the above experiments may allow the less abundant U4atac snRNA to efficiently compete with U4 snRNA for the same set of proteins.

## MATERIALS AND METHODS

### Preparation of protein and RNA samples

Human U4/U6–15.5K and MBP–hPrp31 proteins were prepared as reported previously (Liu et al. 2007). For electrophoretic mobility-shift assays, MBP–hPrp31 was dialyzed against 20 mM Tris–HCl (pH 8.0), 150 mM NaCl, and 2 mM DTT after elution from the affinity resin.

For crystallographic analyses, we used a hPrp31 fragment comprising amino acids 85–333 (hPrp31<sup>85–333</sup>). Seven residues at the N terminus of hPrp31<sup>78–333</sup> were removed, which were flexible according to the ternary hPrp31<sup>78–333</sup>–15.5K–U4 5'SL structure (Liu et al. 2007). The PCR fragment encoding hPrp31<sup>85–333</sup> was subcloned into pGEX-6P-1 via BamHI/XhoI restriction sites. Expression and purification proceeded as described for hPrp31<sup>78–333</sup> (Liu et al. 2007).

Chemically synthesized and HPLC-purified U4atac 5'SL RNA oligonucleotides were obtained commercially (Dharmacon). Before use, the 2'-protecting groups were removed by adding 2'-deprotection buffer (100 mM acetic acid, adjusted to pH 3.8 with TEMED) and incubation at 60°C for 30 min. Samples were then dried in a Speed-Vac, the RNA pellets were resuspended in 10 mM Tris–HCl (pH 7.6), 120 mM NaCl, 2 mM DTT, incubated at 75°C for 4 min, cooled slowly to room temperature, and placed on ice.



## Electrophoretic gel mobility-shift assays

[<sup>32</sup>P]-5'-end labeled RNA oligonucleotides were mixed with recombinant 15.5K (1 μM final concentration) and increasing amounts of MBP-hPrp31 (0, 0.5, 1, 2, 4, 6, 8, 12, or 16 μM final concentration) in 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2 mM DTT, 0.5 μg/μL tRNA, 1 μg/μL heparin, and incubated at 4°C for 30 min. Samples were then fractionated on a 6% (60:1) polyacrylamide gel and visualized using a PhosphorImager (Typhoon 8600, Molecular Dynamics). Bands were quantified using Quantity One software (Bio-Rad Laboratories). Saturation binding curves were fitted using the GraphPad Prism 5 program (bound fraction =  $B_{\max} * [MBP-hPrp31] / (K_d + [MBP-hPrp31])$ ); [MBP-hPrp31]—concentration of MBP-hPrp31,  $B_{\max}$ —maximum binding,  $K_d$ —apparent equilibrium dissociation constant).

## Reconstitution of RNPs and crystallographic analysis

hPrp31<sup>85–333</sup>, 15.5K, and U4atac 5'SL oligomer were mixed in equimolar amounts and incubated at 4°C for 1 h. The assembled complex was purified by size-exclusion chromatography using a Superdex 75 column (GE Healthcare) pre-equilibrated with 10 mM Tris-HCl (pH 7.6), 120 mM NaCl, 2 mM DTT. Peak fractions were analyzed by SDS-PAGE. Fractions of hPrp31<sup>85–333</sup>–15.5K–U4atac 5'SL were pooled and concentrated to 6–9 mg/mL for crystallization. A slower migrating fraction (monomeric complex) and a faster migrating fraction (dimeric, domain-swapped complex) were obtained from the gel-filtration runs with all RNA oligomers tested.

Crystallization trials were carried out at 20°C using the sitting drop vapor diffusion method. RNA oligonucleotides of different lengths were tested in the initial cocrystallization experiments. The best crystals for a slower migrating species (monomeric complex) were obtained with complexes assembled on an RNA comprising U4atac residues 28–55 by mixing 1 μL of complex solution with 1 μL of reservoir solution containing 0.2 M Li<sub>2</sub>SO<sub>4</sub>, 0.1 M Bis-Tris (pH 7.0), 20% w/v PEG 3350. Crystals also grew from this condition in the presence of detergents such as CYPFOS-3, and additives like sucrose, methanol, and urea. The best crystals for a faster migrating species (dimeric, domain-swapped complex) were obtained with complexes assembled on an oligo comprising U4atac residues 26–57 by mixing 1 μL of complex solution with 1 μL of reservoir solution containing 0.1 M sodium cacodylate (pH 6.4), 20% (v/v) PEG 550 MME.

Prior to flash-cooling crystals of the monomeric complex in a 100 K cryogenic nitrogen stream, 2 μL of mother liquor supplemented with 50 mM NaCl and 3%, 6%, 11%, and 13% sucrose were sequentially added to a 2-μL crystallization drop and allowed to equilibrate for 5 min each with removal of 2 μL of solution in between steps. For flash freezing crystals of the dimeric complex, crystals were transferred into the mother liquor plus an additional 10% (v/v) PEG 550 MME. Diffraction data were collected at beamline PXII of the Swiss Light Source (Villigen) and processed with the HKL package (Minor et al. 2006).

## Structure solution, model building, and refinement

The structures of both monomeric and dimeric complexes were solved by molecular replacement with the program MOLREP (Vagin and Teplyakov 2010) using the structure coordinates of

a hPrp31<sup>78–333</sup>–15.K–U4 5'SL complex (Liu et al. 2007; PDB ID 2OZB). Refinement was done with PHENIX (Zwart et al. 2008), including temperature factor, positional and TLS refinement, treating each protein and RNA subunit as one TLS group. Manual rebuilding was done with COOT (Emsley and Cowtan 2004). Noncrystallographic symmetry restraints were maintained during the initial rounds of refinement of the monomeric structure and throughout the entire refinement of the dimeric structure. Water molecules were manually placed into spherical peaks of the 2F<sub>o</sub>–F<sub>c</sub> and F<sub>o</sub>–F<sub>c</sub> electron density maps, which were in hydrogen bonding distance to protein or RNA atoms.

## SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

## ACKNOWLEDGMENTS

We thank the members of our labs for fruitful discussions and the team of Clemens Schulze-Briese for support during diffraction data collection at the PXII beamline of the Swiss Light Source (Villigen, Switzerland). This work was supported by the Deutsche Forschungsgemeinschaft (SFB 740), the Fonds der Chemischen Industrie (to R.L.), and the Ernst-Jung-Stiftung (to R.L.).

Received February 22, 2011; accepted June 24, 2011.

## REFERENCES

- Cheong C, Varani G, Tinoco I Jr. 1990. Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC. *Nature* **346**: 680–682.
- Cojocaru V, Nottrott S, Klement R, Jovin TM. 2005. The snRNP 15.5K protein folds its cognate K-turn RNA: A combined theoretical and biochemical study. *RNA* **11**: 197–209.
- Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**: 2126–2132.
- Falb M, Amata I, Gabel F, Simon B, Carlomagno T. 2010. Structure of the K-turn U4 RNA: a combined NMR and SANS study. *Nucleic Acids Res* **38**: 6274–6285.
- Heus HA, Pardi A. 1991. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA tetraloops. *Science* **253**: 191–194.
- Khusial P, Plaag R, Zieve GW. 2005. LSm proteins form heptameric rings that bind to RNA via repeating motifs. *Trends Biochem Sci* **30**: 522–528.
- Klein DJ, Schmeing TM, Moore PB, Steitz TA. 2001. The kink-turn: a new RNA secondary structure motif. *EMBO J* **20**: 4214–4221.
- Koonin EV, Bork P, Sander C. 1994. A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme? *Nucleic Acids Res* **22**: 2166–2167.
- Kühn-Hölsken E, Lenz C, Sander B, Lüthmann R, Urlaub H. 2005. Complete MALDI-ToF MS analysis of cross-linked peptide-RNA oligonucleotides derived from nonlabeled UV-irradiated ribonucleoprotein particles. *RNA* **11**: 1915–1930.
- Liu S, Li P, Dybkov O, Nottrott S, Hartmuth K, Lüthmann R, Carlomagno T, Wahl MC. 2007. Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science* **316**: 115–120.
- Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. 2006. HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr* **62**: 859–866.

- Nottrott S, Hartmuth K, Fabrizio P, Urlaub H, Vidovic I, Ficner R, Lührmann R. 1999. Functional interaction of a novel 15.5kD [U4/U6-U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J* **18**: 6119–6133.
- Nottrott S, Urlaub H, Lührmann R. 2002. Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J* **21**: 5527–5538.
- Patel AA, Steitz JA. 2003. Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol* **4**: 960–970.
- Schultz A, Nottrott S, Hartmuth K, Lührmann R. 2006. RNA structural requirements for the association of the spliceosomal hPrp31 protein with the U4 and U4atac snRNP. *J Biol Chem* **281**: 28278–28286.
- Tycowski KT, Kolev NG, Conrad NK, Fok V, Steitz JA. 2006. The ever-growing world of small nuclear ribonucleoproteins. In *The RNA world* (ed. RF Gesteland et al.), pp. 327–368. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Vagin A, Teplyakov A. 2010. Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66**: 22–25.
- Vidovic I, Nottrott S, Hartmuth K, Lührmann R, Ficner R. 2000. Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol Cell* **6**: 1331–1342.
- Wahl MC, Will CL, Lührmann R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**: 701–718.
- Will CL, Lührmann R. 2001. Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol* **13**: 290–301.
- Will CL, Lührmann R. 2005. Splicing of a rare class of introns by the U12-dependent spliceosome. *Biol Chem* **386**: 713–724.
- Zwart PH, Afonine PV, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, McKee E, Moriarty NW, Read RJ, Sacchettini JC, et al. 2008. Automated structure solution with the PHENIX suite. *Methods Mol Biol* **426**: 419–435.