PROTEIN STRUCTURE REPORTS

Conserved structure of Snu13 from the highly reduced spliceosome of *Cyanidioschyzon merolae*

C. S. Black,[1] E. L. Garside,[2] A. M. MacMillan,[2] and S. D. Rader[1]*

[1]Department of Chemistry, University of Northern British Columbia, Prince George, British Columbia V2N 4Z9, Canada
[2]Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

Received 9 December 2015; Revised 16 January 2016; Accepted 19 January 2016
DOI: 10.1002/pro.2894
Published online 00 Month 2016 proteinscience.org

Abstract: Structural and functional analysis of proteins involved in pre-mRNA splicing is challenging because of the complexity of the splicing machinery, known as the spliceosome. Bioinformatic, proteomic, and biochemical analyses have identified a minimal spliceosome in the red alga *Cyanidioschyzon merolae*. This spliceosome consists of only 40 core proteins, compared to ~70 in *S. cerevisiae* (yeast) and ~150 in humans. We report the X-ray crystallographic analysis of *C. merolae* Snu13 (CmSnu13), a key component of the assembling spliceosome, and present evidence for conservation of Snu13 function in this algal splicing pathway. The near identity of CmSnu13's three-dimensional structure to yeast and human Snu13 suggests that *C. merolae* should be an excellent model system for investigating the structure and function of the conserved core of the spliceosome.

Keywords: pre-mRNA splicing; U4 snRNP; snRNP assembly; Cyanidioschyzon merolae; RNA binding; fluorescence polarization; X-ray crystallography; minimal spliceosome; thermophile

Introduction

Premessenger RNA (pre-mRNA) splicing involves two transesterifications catalyzed by the spliceosome, a multi-megaDalton-sized particle made up of 5 small, nuclear RNAs (snRNAs) and up to 250 proteins, many of which are specifically associated with the snRNAs in small, nuclear ribonucleoprotein (snRNP) particles.[1] *Cyanidioschyzon merolae* is a unicellular red alga that lives in geothermal springs at temperatures of 42–56°C and pH of 1.5–3.[2] The genome contains only 27 introns; in contrast, the *Saccharomyces cerevisiae* (yeast) genome, of similar size, has more than 250.[3,4] We recently reported a computational and biochemical assessment of the *C. merolae* splicing factor complement that revealed a substantial reduction in core splicing proteins to only 40.[5] Notably, there was no evidence for the U1 snRNA or any of its associated proteins in this organism. *C. merolae* therefore provides a minimal model for studying pre-mRNA splicing, and a source of thermostable splicing proteins for crystallographic studies.[6]

C.S. Black and E.L. Garside contributed equally to this work.

Grant sponsors: Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research, Alberta Innovates Health Solutions (AIHS).

*Correspondence to: Stephen Douglas Rader, Department of Chemistry, University of Northern British Columbia, 3333 University Way, Prince George, British Columbia V2N 4Z9, Canada. E-mail: rader@unbc.ca

This figure also includes an iMolecules 3D interactive version that can be accessed via the link at the bottom of this figure's caption.

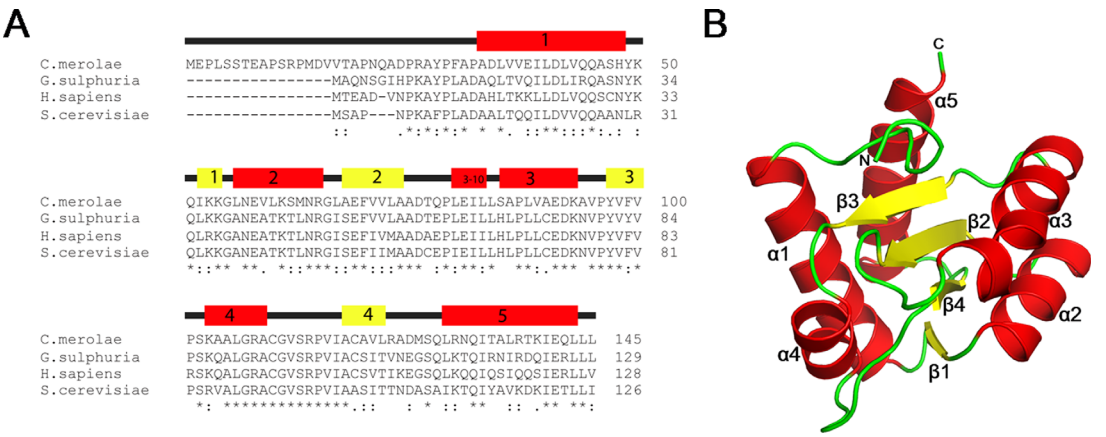

Figure 1. X-ray structure of *C. merolae* Snu13. (a) Sequence alignment[30] with sequences from *G. sulphuraria* (GsSnu13), human (15.5K), and *S. cerevisiae* (Snu13p), indicating sequence identity (*), conservation (:) and partial conservation (.). Secondary-structure diagram depicting α-helices (red) and β-strands (yellow) derived from the structure is shown above the alignment. (b) Ribbon diagram of CmSnu13 structure depicting α-helices (red) and β-strands (yellow). An interactive view is available in the electronic version of the article.

A 13–17 kDa, U4-associated RNA binding protein was identified as part of the U4/U6.U5 tri-snRNP particle in humans (15.5K) and yeast (Snu13p) with a homologous protein (ribosomal protein L7) identified in archaea.[7–9] We have identified the homolog of yeast Snu13p in *C. merolae*[5] (hereafter CmSnu13). Snu13p/15.5K has also been implicated in 2′ *O*-methylation of preribosomal RNA because of its association with box C/D snoRNPs.[10–12] It has been shown that CmSnu13 homologs recognize the kink-turn motif, a purine-rich asymmetrical (5 + 2) internal loop with flanking stems that bend to 65° of each other.[13] Upon binding, the protein acts as a nucleation factor for RNP assembly.[14,15]

Crystal structures of human, yeast, and archaeal homologs of CmSnu13 alone and bound to RNA have been reported.[16–19] Snu13 is an RNA-binding protein featuring an alpha/beta sandwich fold. Interactions with both the U4 5′ stem loop and hPrp31 have been well characterized.[11,13–18] It has also been shown that Prp3p (human 90K) and Prp4p (human 60K) associate and require Snu13p to bind to U4 snRNA prior to their own assembly into the U4 snRNP.[14,20] It is not known what proteins are associated with U4 snRNA prior to assembly of the U4/U6 di-snRNP, but they may include the core Sm and Snu13 proteins.[14,21] In the second stage of U4/U6 snRNP assembly, Prp31 and Prp3/4 join independently.[21] Notably, Prp4 and Prp31 are both missing in *C. merolae*. To study the role of Snu13 in *C. merolae*, we determined the crystal structure of free CmSnu13 and compared it to both free (yeast and human) and RNA-bound (human) structures. The *C. merolae* structure is very similar to both free Snu13 and 15.5K in complex with both U4 snRNA and hPrp31; these results argue for a conserved role of the algal protein in the *C. merolae* splicing pathway. Importantly, the conservation of Snu13 structure also demonstrates that, despite its extreme reduction in complexity, *C. merolae*'s spliceosome should provide a good model for the most highly conserved elements of spliceosome structure and function.

Table I. *Data Collection, Phasing, and Refinement Statistics*

	CmSnu13a
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	30.33, 57.58, 65.38
α, β, γ (°)	90, 90, 90
Wavelength (Å)	1.1271
Resolution (Å)	2.35
$I/\sigma I$	21.36 (11.85)
Completeness (%)	99.77 (99.39)
Redundancy	7.7 (7.6)
Refinement	
Resolution (Å)	43.21–2.35
No. reflections	158,498
R_{work}/R_{free}	0.166/0.220
No. of atoms	
Protein	933
Water	42
B-factors (Å^2)	
Protein	49.5
Water	40.0
R.M.S deviations	
Bond lengths (Å)	0.016
Bond angles (°)	1.52

a Statistics for highest resolution shell (2.43–2.35 Å) are shown in parentheses.

This figure also includes an iMolecules 3D interactive version that can be accessed via the link at the bottom of this figure's caption.

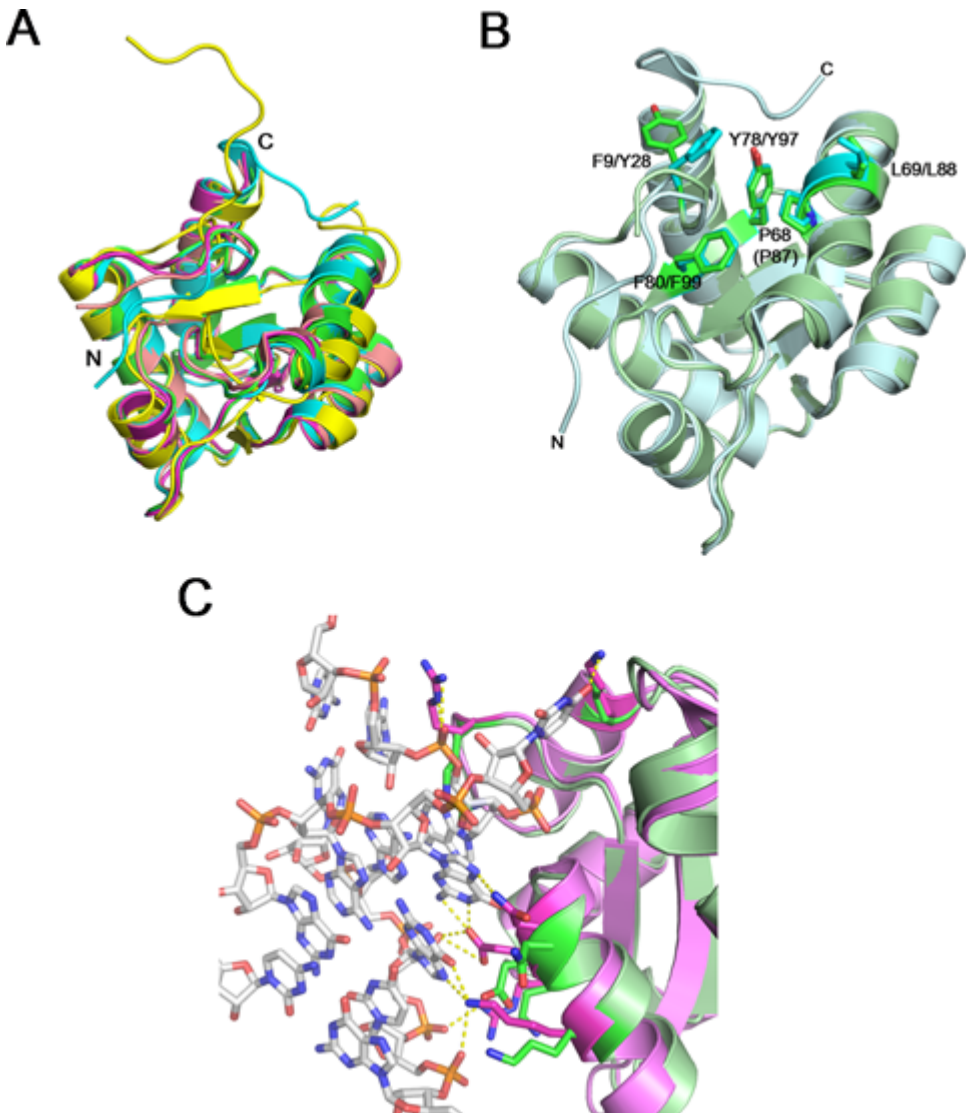

Figure 2. Structural comparison of *C. merolae* Snu13 with orthologs. (a) Structural overlay of ribbon diagrams of CmSnu13 (*C. merolae*; PDB: 5EWR; green), Snu13p (*S. cerevisiae*; PDB: 2ALE; cyan), 15.5K (human; PDB: 2JNB; yellow), 15.5K (human bound to U4atac RNA and Prp31; PDB: 3SIU; magenta), 15.5K (human bound to Prp31 and U4 snRNA; PDB: 2OZB; light pink). (b) Conservation of the hydrophobic pocket proposed to interact with Prp3. Shown are ribbon diagrams of CmSnu13 (*C. merolae*; PDB: 5EWR; green) overlaid on Snu13p (*S. cerevisiae*; PDB: 2ALE; cyan) highlighting the residues lining the hydrophonic pocket: F9 (Y28), P68 (P87), L69 (L88), Y78 (Y97), F80 (F99). Yeast numbering followed by algal in parentheses. (c) Detail of overlay of CmSnu13 (*C. merolae*; PDB: 5EWR; ribbons, green) with 15.5K bound to U4atac RNA (human; PDB: 3SIU; ribbons, magenta; RNA in stick representation). An interactive view is available in the electronic version of the article.

Results and Discussion

To evaluate the similarity of CmSnu13 to its homologs, we performed an amino acid alignment of CmSnu13 with diverse orthologs revealing high sequence conservation [Fig. 1(A)], suggesting functional conservation. The 57% identity between *C. merolae* and yeast orthologs is one of the highest levels of sequence conservation of any *C. merolae* splicing protein.[5]

Structures of Snu13 homologs have been reported alone and bound to fragments of U4 snRNA.[13,16] To determine whether similarity in primary sequence was reflected at the structural level, we solved the 2.35 Å resolution X-ray structure of CmSnu13 [Fig. 1(B), Table I]. To our knowledge, this is the first structure of a splicing protein from *C. merolae*.

The CmSnu13 structure is highly conserved with respect to the yeast and human homologs, with an overall root-mean-square deviation (rmsd) of alpha carbons of 0.014 Å among the structures shown in Figure 2(A). The role of two parts of Snu13 in splicing has been determined: a U4 snRNA-

This figure also includes an iMolecules 3D interactive version that can be accessed via the link at the bottom of this figure's caption.

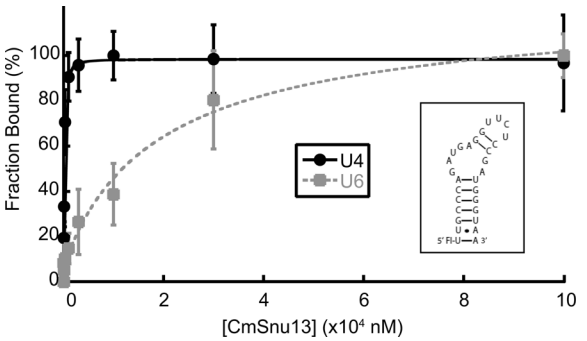

Figure 3. Analysis of CmSnu13•RNA binding. Fluorescence polarization measured by incubation of 5′-fluorescein-labeled RNAs, fragments of *C. merolae* U4 (5′ stem loop, inset) and *S. cerevisiae* U6 (5′ stem loop) snRNAs, with increasing concentration of CmSnu13. Dissociation constants, K_D, were determined by fitting these data to the Hill equation. An interactive view is available in the electronic version of the article. Interactive Image 4: An interactive view is available in the electronic version of the article.

binding region, and a hydrophobic pocket that is believed to interact with Prp3 [Fig. 2(B)]. These features are conserved in CmSnu13, and modeling of RNA from the human RNA-protein complex into the *C. merolae* structure suggests that the RNA interaction could be accommodated [Fig. 2(C)]. These observations strongly support a conserved role for CmSnu13 in U4/U6 di-snRNP structure.

The Snu13 hydrophobic pocket, located within α-helix 3 and β-strand 3, is conserved between *C. merolae* and *S. cerevisiae*. Dobbyn *et al*. proposed this hydrophobic region to be a site of protein–protein interaction with either Prp3p or Prp4p.[16] Our catalog of splicing factors in *C. merolae* revealed no Prp4p/60K homolog.[5] Thus, we propose that Prp3 (90K in humans) is the most likely candidate for a protein interacting with the hydrophobic pocket of Snu13. Structurally, this interaction would mean that Prp3 bridges Stem II of the U4/U6 complex and the 5′ stem loop of U4 via Snu13. This is consistent with a recent crystal structure of Prp3 in association with portions of U4/U6.[20]

We compared the CmSnu13-RNA interaction[5] with reported observations for Snu13 orthologs. The dissociation constant, K_D, for for CmSnu13 binding U4 snRNA was measured using fluorescence polarization with a 5'-fluorescein labeled RNA corresponding to U4 stem II,[5] nucleotides 22–50 (Fig. 3). An increase in concentration of CmSnu13 caused an increase in fluorescence polarization, consistent with RNA-protein binding, so we fit a Hill curve to the binding data to determine the K_D. In testing a variety of buffer conditions, neither pH, nor NaCl/$MgCl_2$ concentration had an effect on binding; $MgCl_2$ increased fluorescence anisotropy, perhaps due to the stabilization of U4's kink-turn.[23] Under optimal conditions, we measured a K_D of ∼160 n*M* (Fig. 3). In a control, we measured a K_D of ∼16 μ*M* for CmSnu13-U6 snRNA association, demonstrating that the interaction with U4 RNA is specific (Fig. 3). These observations are consistent with the conservation of CmSnu13's role in splicing via its association with U4 snRNA.

We compared our findings with those reported in the literature for homologous proteins and their cognate RNA partners (Table II). Snu13 homologs were found to have a K_D ranging from 8 to 150 n*M* (Table II). We note that these measurements were made with a variety of techniques (EMSA, FP, surface plasmon resonance, etc.) so the values may not be directly comparable.

CmSnu13 has a similar affinity for the U4 5′ stem-loop as its homologs from humans and yeast, with a measured K_D of ∼160 n*M* (Table II). This is consistent with the finding that Snu13 requires a purine-rich (5 + 2) asymmetrical loop for optimal binding.[8,13,14] In particular, positions 31–33 and 42–43 of the loop are conserved and positions 29 and 30 must be purines.[8] The Cm U4 stem loop is identical to the human version except at position 29 (G29→A).[8] The extensive conservation between *C.*

Table II. *Comparison of Protein-RNA Affinities*

Protein[a]	RNA target	Methodb	Apparent K_D (n*M*)
15.5K[24]	U3 snoRNA	EMSA	30
15.5K[24]	U8 snoRNA	EMSA	40
15.5K[24]	U14 snoRNA	EMSA	8
15.5K[24]	U4 snRNA	EMSA	20
15.5K[25]	U3 snoRNA	EMSA	130 ± 13
Snu13p[16]	U3 snoRNA	EMSA	75
Snu13p[16]	U4 snRNA	EMSA	150
15.5K[26]	U4 5′ stem loop	SPR	27
CmSnu13[b]	U4 5′ stem loop	FP	160 ± 10

EMSA: electrophoretic mobility shift assay; SPR: surface plasmon resonance; FP: fluorescence polarization.
[a]References indicated as superscript.
[b]This study.

merolae and humans, seen in both the protein and RNA structures, would suggest that U31 is sequestered in the RNA binding pocket when CmSnu13 and U4 snRNA interact.[13] Furthermore, the K_D observed with respect to CmSnu13 and U4 stem loop interaction is consistent with Snu13's role as a nucleation factor for binding of other U4/U6 di-snRNP components, such as Prp3.[16]

Together, the structural and functional results reported here strongly suggest a conserved role for CmSnu13 in the U4 snRNP within the *C. merolae* splicing pathway, and demonstrate that *C. merolae*'s spliceosome structure is likely to be highly conserved.

Methods

Cloning, expression, and protein purification

We amplified the CmSnu13 gene from genomic DNA and subcloned it into SspI-digested pMCGS7 for expression with an N-terminal His tag and a TEV protease site.[27] Protein was expressed in *E. coli* and purified by two rounds of Ni^{2+} affinity chromatography before and after cleavage with TEV protease; CmSnu13 was further purified on a Mono S column (GE Healthcare) and concentrated for crystallization/binding studies.

Fluorescence polarization

We performed binding studies with oligonucleotides (IDT) derived from the *C. merolae* U4 snRNA and the *S. cerevisiae* U6 snRNA:

Cm U4 RNA oligo: 5′ – Fluorescein – UUGCC CAGAUGAGGUUCUCCGAUGGGUAA – 3′.

Sc U6 RNA oligo: 5′ – UUCCCCUGCAUAAG GAU – Fluorescein – 3′.

We incubated CmSnu13 (0–100 μ*M*) with 5 n*M* U4 or U6 RNA oligos in 100 μL binding reactions with buffer conditions varied to observe the sensitivity of the protein-RNA interaction to pH and salt. We measured the anisotropy using a Synergy 2 Multi-Mode reader (BioTek) with black 384-well microplates (Nunc Thermo Scientific).

Structure determination

We concentrated the protein to 10 mg mL^{-1} and grew crystals in 31% PEG 3350 with 100 m*M* sodium acetate pH 4.4. We transferred the crystals to precipitant containing 20% glycerol and froze them in liquid nitrogen. We collected diffraction data remotely from the Stanford Synchrotron Radiation Light Source, and used HKL-2000[28] and CCP4[29] software suites for data processing. We used PHENIX[22] for structure refinement, with yeast Snu13p (PDB 2ALE) as the search model for molecular replacement. We have deposited the atomic coordinates and structure factors for CmSnu13 in the Protein Data Bank with the accession code 5EWR.

References

1. Green MR (1991) Biochemical mechanisms of constitutive and regulated PRE-mRNA splicing. Ann Rev Cell Biol 7:559–599.
2. Ferris MJ, Sheehan KB, Kühl M, Cooksey K, Wigglesworth-Cooksey B, Harvey R, Henson JM (2005) Algal species and light microenvironment in a low-pH, geothermal microbial Mat community. Appl Environ Microbiol 71:7164–7171.
3. Matsuzaki M, Misumi O, Shin-I T, Maruyama S, Takahara M, Miyagishima SY, Mori T, Nishida K, Yagisawa F, Nishida K, Yoshida Y, Nishimura Y, Nakao S, Kobayashi T, Momoyama Y, Higashiyama T, Minoda A, Sano M, Nomoto H, Ogasawara N, Kohara Y, Kuroiwa T (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. Nature 428:67–65.
4. Spingola M, Grate L, Haussler D, Ares M, Jr (1999) Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. RNA 5:221–234.
5. Stark MR, Dunn EA, Dunn WSC, Grisdale CJ, Daniel AR, Halstead MRG, Fast NM, Rader SD (2014) A dramatically reduced spliceosome in *Cyanidioschyzon merolae*. Proc Natl Acad Sci USA 112:E1191–E1200.
6. Dalhus B, Saarinen M, Sauer U, Eklund P, Johansson K, Karlsson A, Ramaswamy S, Bjørk A, Synstad B, Naterstad K, Sirevåg R, Eklund H (2002) Structural basis for thermophilic protein stability: structures of thermophilic and mesophilic malate dehydrogenases. J Mol Biol 318:707–721.
7. 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.
8. Stevens SW, Abelson J (1999) Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. Proc Natl Acad Sci USA 96:7226–7231.
9. Kühn JF, Tran EJ, Maxwell ES (2002) Archaeal ribosomal protein L7 is a functional homologue of the eukaryotic 15.5K/Snu13p snoRNP core protein. Nucleic Acids Res 30:931–941.
10. Oruganti S, Zhang Y, Li H (2005) Structural comparison of yeast snoRNP and spliceosomal protein Snu13p with its homologs. Biochem Biophys Res Commun 333: 550–554.
11. Marmier-Gourrier N, Clery A, Senty-Segault V, Charpentier B, Schlotter F, Leclerc F, Fournier R, Branlant C (2003) A structural, phylogenetic, and functional study of 15.5kD/Snu13 protein binding on U3 small nucleolar RNA. RNA 9:821–838.
12. Maxwell ES, Fournier MJ (1995) The small nucleolar RNAs. Ann Rev Biochem 64:897–934.
13. 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.
14. 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.
15. Watkins NJ, Dickmanns A, Lührmann R (2002) Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the

15.5K protein, for the hierarchical assembly of the box C/D snoRNP. Mol Cell Biol 22:8342–8352.
16. Dobbyn HC, McEwan PA, Krause A, Novak-Frazer L, Bella J, O'Keefe RT (2007) Analysis of pre-mRNA and pre-rRNA processing factor Snu13p structure and mutants. Biochem Biophys Res Commun 360:857–862.
17. Hamma T, Ferre DAmareAR (2004) Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 Å resolution. Structure 12: 893–903.
18. Liu S, Ghalei H, Lührmann R, Wahl MC (2011) Structural basis for the dual U4 and U4atac snRNA-binding specificity of spliceosomal protein hPrp31. RNA 17: 1655–1663.
19. Nguyen TH, Galej WP, Bai XC, Savva CG, Newman AJ, Scheres SH, Nagai K (2015) The architecture of the spliceosomal U4/U6.U5 tri-snRNP. Nature 523:47–52.
20. Liu S, Mozaffari-Jovin S, Wollenhaupt J, Santos KF, Theuser M, Dunin-Horkawisz S, Fabrizio P, Bujnicki JM, Lührmann R, Wahl MC (2015) A composite double-/single-stranded RNA-binding region in protein Prp3 supports tri-snRNP stability and splicing. eLife 4: e07320.
21. 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.
22. Goody TA, Melcher SE, Norman DG, Lilley DMJ (2004) The kink-turn motif in RNA is dimorphic, and metal ion-dependent. RNA 10:254–264.
23. Kapust RB, Tözsér J, Fox JD, Anderson DE, Cherry S, Copeland TD, Waugh DS (2001) Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Eng 14:993–1000.
24. Otwinowski ZM, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276:307–326.
25. Potterton E, Briggs P, Turkenburg M, Dodson E (2003) A graphical user interface to the CCP4 program suite. Acta Cryst D59:1131–1137.
26. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Jung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Cryst D66:213–221.
27. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: a novel method for multiple sequence alignments. J Mol Biol 302:205–217.
28. Watkins NJ, Ségault V, Charpentier B, Nottrott S, Fabrizio P, Bachi A, Wilm M, Rosbash M, Branlant C, Lührmann R (2000) A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. Cell 103: 457–466.
29. Weinstein Szewczak LB, DeGregorio SJ, Strobel SA, Steitz J (2002) Exclusive interaction of the 15.5 kD protein with the terminal box C/D motif of a methylation guide snoRNP. Chem Biol 9:1095–1107.
30. Soss SE, Flynn PF (2007) Functional implications for a prototypical K-turn binding protein from structural and dynamical studies of 15.5K. Biochemistry 46: 14979–14986.