Structure and conformational plasticity of the U6 small nuclear ribonucleoprotein core

Eric J. Montemayor,^{a,b*} Allison L. Didychuk,^a Honghong Liao,^a Panzhou Hu,^b David A. Brow^{b*} and Samuel E. Butcher^{a*}

^aDepartment of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706, USA, and ^bDepartment of Biomolecular Chemistry, University of Wisconsin-Madison, Madison, WI 53706, USA. *Correspondence e-mail: emontemayor@wisc.edu, dabrow@wisc.edu, sebutcher@wisc.edu

Received 26 September 2016

Accepted 14 November 2016

Edited by R. J. Read, University of Cambridge, England

Keywords: splicing; RNA; crystallography; U6 snRNA; spliceosome; Prp24; ribonucleoprotein; annealing.

PDB reference: U6–Prp24 complex, 5tf6

Supporting information: this article has supporting information at journals.iucr.org/d

U6 small nuclear RNA (snRNA) is a key component of the active site of the spliceosome, a large ribonucleoprotein complex that catalyzes the splicing of precursor messenger RNA. Prior to its incorporation into the spliceosome, U6 is bound by the protein Prp24, which facilitates unwinding of the U6 internal stem-loop (ISL) so that it can pair with U4 snRNA. A previously reported crystal structure of the ‘core’ of the U6 small nuclear ribonucleoprotein (snRNP) contained an ISL-stabilized A62G mutant of U6 bound to all four RNA-recognition motif (RRM) domains of Prp24 [Montemayor *et al.* (2014), *Nature Struct. Mol. Biol.* **21**, 544–551]. The structure revealed a novel topology containing interlocked rings of protein and RNA that was not predicted by prior biochemical and genetic data. Here, the crystal structure of the U6 snRNP core with a wild-type ISL is reported. This complex crystallized in a new space group, apparently owing in part to the presence of an intramolecular cross-link in RRM1 that was not observed in the previously reported U6–A62G structure. The structure exhibits the same protein–RNA interface and maintains the unique interlocked topology. However, the orientation of the wild-type ISL is altered relative to the A62G mutant structure, suggesting inherent structural dynamics that may facilitate its pairing with U4. Consistent with their similar architectures in the crystalline state, the wild-type and A62G variants of U6 exhibit similar Prp24-binding affinities and electrophoretic mobilities when analyzed by gel-shift assay.

1. Introduction

Precursor messenger RNA (pre-mRNA) splicing requires the concerted action of five small nuclear RNAs (snRNA) and over 80 protein cofactors that together form a dynamic machine known as the spliceosome (Jurica & Moore, 2003; Fabrizio *et al.*, 2009; Will & Lührmann, 2011). The spliceosome is assembled *de novo* on pre-mRNA in order to catalyze each splicing reaction. This assembly pathway relies upon precise protein–protein, protein–RNA and RNA–RNA interactions (Will & Lührmann, 2001; Brow, 2002; Matlin & Moore, 2007), as exemplified by the recently determined structures of the U1 snRNP (Pomeranz Krummel *et al.*, 2009; Weber *et al.*, 2010; Kondo *et al.*, 2015); the U6 snRNP core (Montemayor *et al.*, 2014); tri-snRNP (Nguyen *et al.*, 2015; Wan, Yan, Bai, Wang *et al.*, 2016; Nguyen *et al.*, 2016; Agafonov *et al.*, 2016); and the B^{act}, C and ILS spliceosomes (Yan *et al.*, 2015; Wan, Yan, Bai, Huang *et al.*, 2016; Yan *et al.*, 2016; Galej *et al.*, 2016).

U6 snRNA is the most conserved of the snRNAs (Brow & Guthrie, 1988) and is located in the heart of active spliceosomes after their assembly onto pre-mRNA. U6 also exhibits highly dynamic secondary structure during its recruitment into spliceosomes: its internal stem-loop (ISL) must be unwound during assembly and subsequently re-annealed to partly

© 2017 International Union of Crystallography

compose the enzyme active site (Fica *et al.*, 2013). U6 can participate in multiple rounds of splicing through a recycling pathway in which U6 is released from a post-catalytic spliceosome and captured by the protein Prp24 (Shannon & Guthrie, 1991; Ghetti *et al.*, 1995) and the Lsm2–8 ring (Achsel *et al.*, 1999) to generate a free U6 small ribonucleoprotein (snRNP). U6 is then escorted back into a nascent spliceosome through extensive base-pairing with U4 snRNA (Siliciano *et al.*, 1987; Bringmann *et al.*, 1984; Hashimoto & Steitz, 1984). This annealing of U4 and U6 requires unwinding of the ISL so that the corresponding nucleotides can base-pair with U4 to form the U4/U6 di-snRNP. Prp24 drives this annealing process (Raghuathan & Guthrie, 1998; Didychuk *et al.*, 2016), and its action is enhanced by the Lsm2–8 ring (Rader & Guthrie, 2002; Verdone *et al.*, 2004; Didychuk *et al.*, 2016).

Recently, the structure of the U6 snRNP ‘core’ was determined by X-ray crystallography at 1.7 Å resolution (Montemayor *et al.*, 2014). This complex contained all four RNA-recognition motif (RRM) domains of Prp24 (lacking only the first 33 and the last 44 residues of the 444-residue protein) and nucleotides 30–101 of the 112-nucleotide U6 RNA with several point mutations: A62G, which introduces a G–C pair in the base of the ISL, and U100C/U101C, which stabilize a distal RNA helix known as the telestem. The structure revealed an unprecedented ‘interlocked’ protein–RNA topology in which the protein and RNA cooperatively fold to generate concatenated rings. The structure also revealed a striking correlation between the protein–RNA interface and mutations in Prp24 that suppress the cold-sensitive growth phenotype of the A62G mutation. However, the possibility exists that the A62G mutation traps a misfolded species both *in vitro* and *in vivo*.

We have now determined the crystal structure of the yeast U6 snRNP core with a wild-type ISL. The structure is highly similar to the complex with the A62G-stabilized ISL and also exhibits an interlocked topology, although there are differences in the orientation of the ISL relative to the protein–RNA interface. Gel-shift data show that binding affinity and electrophoretic mobility are unchanged by the A62G mutation, indicating that it does not significantly perturb the U6–Prp24 interaction. Interestingly, the crystallized core U6 snRNP containing the wild-type ISL has an apparent intramolecular chemical cross-link in RRM1 of Prp24 that was not observed in the A62G-ISL structure. This cross-link includes a highly conserved cysteine residue at position 64, and we report mutational analysis of this residue and Lys99, to which it appears to be covalently linked.

2. Materials and methods

2.1. Protein and RNA production

Protein and RNA constructs were prepared essentially as described elsewhere (Montemayor *et al.*, 2014), with the exception that the *in vitro* transcription template did not encode the A62G mutation in U6 snRNA. The final purified complex was concentrated to approximately 6 mg ml^{−1} in

Table 1

Data-collection and structure refinement for U6–Prp24 (PDB entry 5tf6).

Values in parentheses are for the highest resolution shell.

Wavelength (Å)	0.9795
Resolution range (Å)	49.1–2.30 (2.38–2.30)
Space group	<i>P</i> ₂ ₁ ₂ ₁
Unit-cell parameters (Å)	<i>a</i> = 76.9, <i>b</i> = 84.5, <i>c</i> = 255.1
Total reflections	900493 (40882)
Unique reflections	74598 (7217)
Multiplicity	12.1 (5.7)
Completeness (%)	99.71 (97.55)
Mean <i>I</i> /σ(<i>I</i>)	17.67 (1.39)
Wilson <i>B</i> factor (Å ²)	50.38
<i>R</i> _{merge}	0.09 (0.89)
CC _{1/2}	0.998 (0.626)
<i>R</i> _{work} / <i>R</i> _{free}	0.19/0.23 (0.32/0.33)
No. of atoms	
Total	9375
Macromolecules	8586
Ligands	39
Water	749
R.m.s.d., bonds	0.015
R.m.s.d., angles	1.77
Coordinate error (Luzzati) (Å)	0.33
Coordinate error (maximum likelihood) (Å)	0.35
Phase error (maximum likelihood) (°)	25.17
Ramachandran favored (%)	96.7
Ramachandran outliers (%)	0.68
Average <i>B</i> factor (Å ²)	
Overall	69.60
Protein	55.4
RNA	104.9
Ligands	69.30
Solvent	57.80

400 mM potassium chloride, 10 mM HEPES acid, 10 mM Tris base, 2 mM MgCl₂, 1 mM TCEP–HCl, 5% glycerol pH ≈ 7.5.

2.2. *In vitro* binding assay

Electrophoretic mobility shift assays were performed and analyzed as described previously (Didychuk *et al.*, 2016). Briefly, RNA was labeled at its 5′ end with ³²P and electrophoresed through a nondenaturing polyacrylamide gel in the absence and presence of varying concentrations of Prp24.

2.3. Crystallization, structure determination and refinement

High-throughput crystallization screening was performed with a Mosquito crystallization robot (TTP Labtech) using sitting-drop vapor diffusion at 4°C. Optimized crystals were grown *via* hanging-drop vapor diffusion at 4°C in 1 M lithium chloride, 0.1 M sodium MES, 0.064 M hydrochloric acid, 25% PEG 8000, 15% glycerol. The best crystals, which were used for data collection, came from prior threefold dilution of concentrated U6–Prp24 into 1 mM MnCl₂ and mixing 3 μl of the dilute complex with 1 μl of the crystallization reagent prior to vapor diffusion. Thus, MnCl₂ is present in the final crystallization mixture at a final concentration of approximately 1–3 mM.

Initial screening was performed on beamline 21-ID-F at the Advanced Photon Source. The final diffraction data were collected on beamline 24-ID-C at the Advanced Photon Source using an MD-2 microdiffractometer and a PILATUS 6MF detector. Diffraction data were collected at 100 K with a

shutterless ‘continuous vector scan’ across the crystal to minimize the effect of radiation damage. Data were integrated using *XDS* (Kabsch, 2010) and the *RAPD* pipeline available at beamline 24-ID-C. Space-group determination and scaling were performed in *POINTLESS* (Evans, 2011) and *AIMLESS* (Evans & Murshudov, 2013), respectively. *Phenix.xtriage* was used to assay potential twinning in the diffraction data (Adams *et al.*, 2010). Data-collection and refinement statistics are given in Table 1.

Initial phases were determined by molecular replacement using *Phaser* (McCoy *et al.*, 2007) with PDB entry 4n0t (Montemayor *et al.*, 2014), with subsequent refinement accomplished *via* iterative rounds of manual model building in *Coot* (Emsley *et al.*, 2010) and automated refinement in *PHENIX* (Chen *et al.*, 2010; Afonine *et al.*, 2012) or *REFMAC* from CCP4 (Winn *et al.*, 2011; Murshudov *et al.*, 1997, 2011). The solvent content of the complex was determined using the *phenix.f000* routine (Adams *et al.*, 2010) and an online tool provided by Bernhard Rupp (Matthews, 1968; Weichenberger & Rupp, 2014). The final model has unusually high atomic displacement parameters that arise from large dynamic regions of RNA in the crystal lattice (see Fig. 1). These regions

of RNA were subjected to TLS-restrained refinement of grouped anisotropic atomic displacement parameters to more effectively model their dynamics. The identities of bound ions and water molecules were determined using the anticipated coordination geometries and interatomic distances of the components in the crystallization mixture, particularly in the case of magnesium, which was present in the crystallization mixture but was not visible in the structure (Klein *et al.*, 2004; Petrov *et al.*, 2011). Placed solutes were validated by their ability to minimize residual $mF_o - DF_c$ densities relative to other solutes and to yield refined atomic displacement parameters that were in accordance with the local environment of the solute in the crystal. In the case of ions that possess weak

Figure 1

Structure of the U6 snRNP core with a wild-type ISL. (a) Sequences of the protein and RNA constructs used for crystallization. Regions colored white were deleted to promote crystallization. Gray regions in the Prp24 sequence correspond to linkers surrounding the RRM domains. U6 nucleotides 64–83 are disordered in one of the two complexes in the crystallographic asymmetric unit [gray complex in (b)]. (b) Packing interactions between the two complexes in the crystallographic asymmetric unit, which are related to each other by noncrystallographic twofold rotational symmetry. One U6–Prp24 complex is colored gray to enhance the contrast between the two complexes in the asymmetric unit. The electropositive groove is thought to bind double-stranded RNA to promote the formation of the U4/U6 di-snRNA (Montemayor *et al.*, 2014; Didychuk *et al.*, 2016). (c) Crystal-packing environment surrounding the asymmetric unit. Parallel cavities with cross-sections of approximately 1500 Å² traverse the crystal lattice. (d) The two complexes in the asymmetric unit are virtually identical, with the exception of the U6 ISL, which in one complex is positioned in the center of the large solvent cavity and is disordered.

Figure 2

Mutant and wild-type ISL structures in the U6 snRNP. (a) Overlay of the wild-type snRNP core with the U6-A62G snRNP core (gray with yellow A62G mutation; Montemayor *et al.*, 2014). The U6 ISL was omitted from the selection used to construct the overlay. The pivot for the altered trajectories of the two ISLs is located at approximately the A59:U88 pairing at the base of the ISL. (b) Superposition of only the ISL in the wild-type and A62G complexes. Nucleotides 59–67 and 80–88 were used to construct the overlay, and the protein is omitted for clarity. Despite the altered sequences and helical trajectories, the pairing of nucleotides in the lower ISL is unchanged in the wild-type and A62G mutant structures. All nucleotides of the wild-type ISL are visible in the crystal form presented here. Nucleotide A79 is extrahelical, unlike in the A62G structure, which was crystallized in a different space group. (c, d) Base-pairing at U6 nucleotide 62. The interatomic distances are consistent with protonation of the N1 atom of adenine in the wild-type ISL presented here. (e) Structure of the U6 ISL in the C complex spliceosome (Galej *et al.*, 2016). For clarity, only the proteins Prp8, Isy1, Cef1, Cwc15 and Clf1 are shown in gray. (f) Associated protein cofactors dramatically alter the conformation and pairing of the ISL in assembled spliceosomes, and the A62–C85 pairing is not preserved. Dashes and circles represent Watson–Crick and non-Watson–Crick base pairs, respectively, and ‘+’ denotes protonation within the A62–C85 base pair.

anomalous scattering properties at the experimental wavelength, an anomalous difference map was also used to validate their identity. All figures were prepared using *PyMOL* (<http://www.pymol.org>).

2.4. Yeast strains and growth assays

Mutations in the *PRP24* gene were introduced by inverse PCR and self-ligation of the plasmid pRS313-PRP24(S288C), which harbors a 2152 bp *SpeI/XhoI* fragment of the *Saccharomyces cerevisiae* S288C genome that includes the *PRP24* ORF plus 689 bp of upstream DNA and 136 bp of downstream DNA (Vidaver *et al.*, 1999). For the K99A/K102A substitution, both AAA lysine codons were changed to GCG alanine codons. For the C64A and C64S substitutions, the TGT cysteine codon was changed to a GCT alanine codon or an AGT serine codon, respectively. Mutant and wild-type *PRP24* plasmids were transformed into yeast strain LL101 (*MATa*, *prp24-ΔClaSnaB::ADE2, ade2-1, can1-100, his3-11, leu2-3,112, lys2-Δ2, met2-Δ1, trp1-1, ura3-52*, [pUN50-PRP24]; Vidaver *et al.*, 1999), selecting for histidine prototrophy. Three colonies from each transformation were grown overnight in liquid synthetic complete medium lacking histidine (SC –his; MP Biomedicals) and were then plated onto solid SC –his medium containing 0.75 mg ml^{−1} 5-fluoroorotic acid (5-FOA) to select for the loss of the *URA3*-marked plasmid containing a wild-type *PRP24* allele. Colonies from 5-FOA plates were picked into liquid YEPD medium and grown to an OD₆₀₀ of 1.0 for serial dilution and plating onto solid YEPD medium.

3. Results and discussion

3.1. Core structure of the U6 snRNP with a wild-type ISL

The core structure of the yeast U6 snRNP with a wild-type ISL was determined *via* X-ray crystallography in order to assess the structural implications of the A62G mutation in our previous structure (Montemayor *et al.*, 2014). U6 nucleotides 30–101 with telestem mutations U100C/U101C and Prp24 residues 34–400 (Fig. 1*a*) crystallized in space group $P2_12_12_1$. The crystals diffracted X-rays to a maximum resolution of 2.3 Å and exhibit a crystal-packing arrangement that is distinct from the previous A62G structure in space group $P2_1$ that had only one complex in the asymmetric unit. In the new structure, there are two complexes in the asymmetric unit that exhibit twofold rotational symmetry (Fig. 1*b*) about an axis nearly parallel to the *y* axis of the unit cell, thus raising the possibility of a misassigned crystallographic space group. However, the two complexes within the asymmetric units are each in distinct crystal-packing environments, precluding an alternative space group of higher symmetry (Fig. 1*c*).

The protein components of the two complexes in the asymmetric unit superimpose with an r.m.s.d. of 0.47 Å (Fig. 1*d*). While this high degree of structural similarity extends to RNA located at the protein–RNA interface (nucleotides 37–58 and 91–92), it is not possible to characterize the similarity of the entire RNA, particularly nucleotides 59–88, as the majority of one ISL in the asymmetric unit is disordered and protrudes into a linear solvent cavity that spans the crystal lattice parallel to the crystallographic *x* axis (Fig. 1*c*). Electron density is visible for all nucleotides in the other ISL and shows that it packs against the surface of RRM1 from the other U6–Prp24 complex in the asymmetric unit and RRM1 and oRRM4 (or ‘occluded RRM’) domains belonging to two different symmetry-related protomers in the crystal lattice. The packing arrangement places the ISL near to, but not directly inside, an electropositive groove

along the surface of Prp24 that has been proposed to function in U4/U6 annealing by binding duplex RNA structures present in the annealing pathway (Fig. 1*b*; Montemayor *et al.*, 2014; Didychuk *et al.*, 2016). Interestingly, the ISL–RRM1 interaction surface contains an apparent intramolecular cross-link in RRM1 that was not observed previously (see §3.3).

Comparison of the new wild-type and previous A62G structures confirms that mutation of the ISL does not alter the overall architecture of the U6–Prp24 complex. The protein components of these two complexes superimpose with a main-chain r.m.s.d. of 1.0 Å and exhibit a virtually identical protein–RNA interface with unchanged secondary structure in the RNA. However, the conformation of the ISL distal to the protein–RNA interface is distinct in the two structures, involving an approximate 20° bend in the ISL near A62 that alters the trajectory of the ISL relative to the core of the complex (Fig. 2*a*). Despite this difference in orientation relative to the core, the wild-type and A62G ISLs remain quite similar in the ‘lower ISL’ region spanning bases 59–67 and 80–88 (Fig. 2*b*). The 62–85 base pair is retained, but with an altered hydrogen-bonding pattern that suggests protonation of either the N1 atom of A62 or the N3 atom of C85 (Fig. 2*c*). Protonation of A62 is consistent with its measured pK_a of 6.5 (Sashital *et al.*, 2004) in the isolated ISL and crystallization of the U6–Prp24 complex in MES buffer at an approximate pH of 5.9 (see §2). The upper ISL region spanning nucleotides 68–79 is entirely visible in this new crystal form and shows that A79 is not stacked in the RNA helix (Fig. 2*b*). This is unlike the previous A62G structure in space group $P2_1$, where nucleotides 71–76 were disordered and A79 was stacked in the helix between the adjacent nucleotides G78 and U80.

We propose that these differences in the upper ISL reflect its highly dynamic nature and by extension its great sensitivity to different crystal-packing environments in the $P2_12_12_1$ and $P2_1$ crystal structures. These differences are compatible with known dynamics in the ISL that have been observed

Figure 3

Mutation of the U6 ISL does not substantially affect binding between U6 and Prp24. (a) An electrophoretic mobility shift assay (EMSA) was used to assess the binding affinity of Prp24 towards U6 snRNA nucleotides 30–101 with a U100C/U101C mutation or the same RNA with an additional A62G mutation. The total concentration of labeled U6 is 0.5 nM in all samples. (b) Plot of bound RNA as a function of protein concentration. The A62G mutant exhibits a higher fraction of bound RNA, suggesting that stabilization of the ISL increases the fraction of U6 RNA that is competent to bind Prp24. Data points denoted by red arrows are likely to correspond to weak binding of misfolded U6 and were not used for calculation of the binding constants shown here. All data are derived from four independent technical replicates, except for those at 2400 nM, which were performed in duplicate.

previously by NMR when studying truncated variants of U6 in the absence of Prp24 (Reiter *et al.*, 2004; Blad *et al.*, 2005; Venditti *et al.*, 2009). We also note that the ISL is among the least structured regions of the crystal lattice, with nucleotides 68–79 exhibiting elevated atomic displacement parameters relative to the entire structure. These dynamics are likely to play a functional role in U4/U6 annealing (see §3.4). In contrast, the U6 ISL in catalytic spliceosomes (Galej *et al.*, 2016; Yan *et al.*, 2015; Wan, Yan, Bai, Huang *et al.*, 2016) is stabilized by extensive RNA–RNA and protein–RNA contacts (Figs. 2*e* and 2*f*). These contacts are likely to constrain the motion of the ISL and its bound magnesium ions.

3.2. Mutational analysis of U6–Prp24 binding affinity

We performed an electrophoretic mobility shift assay (EMSA) to further evaluate the structural implications of the A62G mutation. The crystallizable variants of U6 and Prp24

(Fig. 1*a*) interact with an apparent dissociation constant of 2.9 ± 0.2 nM, which is effectively unchanged by the A62G mutation. Furthermore, the electrophoretic mobility of U6 and U6–Prp24 is also unchanged (Fig. 3*a*). These observations support the notion that the U6 snRNP secondary structure and overall topology are not dramatically affected by the A62G mutation and, together with the structural data, directly argue against the alternative U6–Prp24 fold proposed elsewhere that lacks a paired ISL (Dunn & Rader, 2010).

The low nanomolar binding affinity is tighter than that observed previously under similar conditions with full-length wild-type U6 and Prp24, where K_d was approximately sixfold higher (Didychuk *et al.*, 2016). We propose that these differences lie in the U100C/U101C mutations used to facilitate the crystallization of the truncated RNA. These mutations replace two putative G–U wobble pairs in the telestem with two G–C base pairs. The telestem acts as a linchpin of the interlocked topology, and its stabilization has been shown to increase

binding affinity by reducing k_{off} (Didychuk *et al.*, 2016). A telestem-stabilized variant of full-length U6 that included the U100C/U101C and other mutations also exhibited a low nanomolar K_d (2.2 ± 0.2 nM; Didychuk *et al.*, 2016).

We note that a fraction of U6 is recalcitrant to binding submicromolar concentrations of Prp24, as shown by the presence of free RNA at protein concentrations far above the K_d and a maximal binding fraction of less than one (Figs. 3*a* and 3*b*). We therefore did not constrain the maximal fraction of bound RNA to 100% when extracting the K_d from the data by nonlinear regression over a concentration range of 0.4–270 nM. We hypothesize that this incomplete binding is owing to an alternatively folded subpopulation of U6 that is incapable of tightly binding to Prp24. We note that stabilization of the ISL in the A62G mutant RNA exhibits more complete binding, suggesting that stabilization of the lower ISL favors the conformation of U6 that binds Prp24 with high affinity. Since full-length U6 does not exhibit the same degree of incomplete binding as seen here (Didychuk *et al.*, 2016), we further speculate that nucleotides 1–29 and 102–112 in full-length U6 also help to stabilize this conformation.

Finally, U100C/U101C and analogous telestem-stabilizing mutations in full-length U6 have been shown to increase the rate of Prp24-mediated annealing of U4 and U6 in a minimal *in vitro* system (Didychuk *et al.*, 2016). This finding

Figure 4

An apparent intramolecular protein cross-link in RRM1. (a) Electron-density map spanning amino acids His63, Cys64 and Lys99 in RRM1 of Prp24. Similar density was visible for both protein molecules in the crystallographic asymmetric unit. (b) Annealed OMIT map for the same region as depicted in (a). The positions of all atoms in His63, Cys64 and Lys99 were fixed during a round of simulated annealing as implemented in *PHENIX* (Afonine *et al.*, 2012) and then omitted from the calculation of the map shown here. The depicted interatomic distances are between the side-chain N atom of Lys99 and the S atom of Cys64 and between the side-chain N atom of Lys99 and the NE2 ring N atom of His63. (c) The cross-link changes the orientation of His63 to be amenable for crystal-packing contacts with a neighboring ISL. (d) The same ISL–protein packing contacts are not possible in the absence of the cross-link, where His63 adopts a different conformation (Montemayor *et al.*, 2014). (e) Mutation of residues in the apparent cross-link does not cause a discernible growth phenotype in yeast.

provides further evidence that the U100C/U101C mutation used in the crystallization constructs does not induce a nonfunctional conformation of the U6 snRNP and that the interlocked architecture observed here is indeed an on-pathway species during annealing of U4 and U6.

3.3. An unexpected intramolecular protein cross-link

Inspection of the final $2mF_o - DF_c$ electron-density map revealed the presence of an apparent intramolecular linkage between Prp24 residues His63, Cys64 and Lys99 (Fig. 4*a*). This density remains after the calculation of a simulated-annealing OMIT map to remove phase bias from all atoms in the associated linkage (Fig. 4*b*), and calculation of an anomalous difference map shows weak density for only the SG of Cys64 (data not shown). Thus, no components of the linkage possess strong anomalous scattering properties at the experimental wavelength, precluding inter-residue density owing to the chelation of heavier ions present in the crystallization mixture. The calculated maps lack sufficient resolution to unambiguously identify the position or number of atoms comprising the linkage, and mass spectrometry failed to identify a cross-linked peptide (data not shown). Highly oxidized states of cysteine, such as sulfonic or sulfinic acid, are too large to fit in the linkage, although it is feasible that a single oxygen adduct in the form of sulfenic acid is present (Fig. 4*b*). Alternatively, a single methylene moiety of unknown origin may bridge the lysine and cysteine residues, as recently observed in a small number of other proteins in the PDB (Ruszkowski & Dauter, 2016). However, refinement with this linkage did not clear all residual density in the resulting $mF_o - DF_c$ maps and neither did refinement with methylated His63 (data not shown).

This linkage is located at a packing interface between copies of U6–Prp24 in the crystal lattice. Transition-metal ions are known to interact with cysteine and histidine residues (Dokmanić *et al.*, 2008), and the inclusion of manganese in the crystallization mixture improved the size and diffracting power of the crystals used for structure determination. We therefore propose that manganese may have acted as a catalyst in promoting cross-link formation and thus crystal growth. This hypothesis is supported by the lack of a similar linkage in the previous A62G structure (Montemayor *et al.*, 2014), which was obtained from crystals grown in the absence of manganese and shows a markedly different conformation of His63 (Fig. 4*c* and 4*d*).

Although the composition of this linkage remains unknown, it was brought to our attention that a crystal structure of a putative ubiquitin-conjugating enzyme from *Toxoplasma gondii* (PDB entry 2f4z; Vedadi *et al.*, 2007) exhibits a similar intramolecular linkage between lysine and cysteine residues (Christopher Lima, Memorial Sloan Kettering Cancer Center, personal communication). Ubiquitination pathways typically employ activated lysine and cysteine residues to covalently bind and transfer ubiquitin to target substrates (Strieter & Korasick, 2012; Streich & Lima, 2014), and the human homolog of Prp24 has been shown to play functional roles in regulating the ubiquitylation of histones and components of

the splicing machinery (Song *et al.*, 2010; Long *et al.*, 2014; Timani *et al.*, 2014). We therefore considered the possibility that the cross-link reflects an inherent chemical reactivity in Prp24 that is related to the function of the protein *in vivo*, particularly since Cys64 is conserved (Montemayor *et al.*, 2014). Lys99 is less highly conserved, but the nearby Lys102 is strongly conserved. In contrast, position 63 is poorly conserved and is rarely a histidine. Targeted mutagenesis of Cys64 and Lys99 (in combination with Lys102) was performed to assess their importance to the viability of yeast cells. Neither the mutation of Cys64 to alanine or serine nor the mutation of both Lys99 and Lys102 to alanine conferred a discernible growth defect at a range of temperatures (Fig. 4*e*). Thus, if a covalent bond to any of these side chains forms *in vivo* it is not essential or has a redundant function.

3.4. Implications for U4/U6 assembly

Reversible annealing of U4 and U6 during the splicing cycle requires an exquisite balance between the stability of free and bound RNAs, so that association may be achieved without falling into a permanent kinetic trap. Multiple crystal structures of the U6 snRNP core now provide a clue as to how this balance is achieved. During annealing, the U6 ISL must be unwound so that it can pair with U4. Prp24 specifically recognizes U6 by almost exclusively interacting with nucleotides outside the ISL. This allows U4/U6 annealing to exploit native dynamics in the ISL, as shown by structural plasticity in the conformation of the ISL when comparing a wild-type and an A62G complex with Prp24 (Fig. 2*a*), alternate base stacking of A79G (Fig. 2*b*) and additional dynamics in U80 as reported elsewhere (Reiter *et al.*, 2004; Blad *et al.*, 2005; Venditti *et al.*, 2009). These dynamics allow the ISL to explore conformational space near to the electropositive groove in Prp24 that has been proposed to function as a binding surface for intermediate duplex RNA structures generated during the annealing process (Montemayor *et al.*, 2014; Didychuk *et al.*, 2016). However, important aspects of U4/U6 annealing remain unresolved. For example, it is not known whether ISL unwinding precedes or follows the initial interaction with U4, nor is it understood in what order the two helices in U4/U6 anneal. Additionally, it is not clear how the interlocked topology is resolved to ultimately displace Prp24 from U4/U6 after annealing. Future structural and mechanistic investigation will better clarify these points and further explain the mechanism by which stable yet dynamic particles drive pre-mRNA splicing.

Acknowledgements

We are grateful to Chris Lima for drawing our attention to the Cys–Lys cross-link in PDB entry 2f4z, Ron Raines for helpful discussions regarding intramolecular cross-linking, Basudeb Bhattacharyya, Jon Schuermann and Kay Perry for assistance with the collection of diffraction data, Craig Bingman for advice on structure refinement, Grzegorz Sabat for analysis of mass-spectrometric data and members of the Brow and Butcher laboratories for critical reading of the manuscript.

Use of the Advanced Photon Source, an Office of Science User Facility operated for the US Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the US DOE under contract No. DE-AC02-06CH11357. Use of the NE-CAT Sector 24 was supported by NIH grant P41 GM103403. This work was supported by grants from the US National Institutes of Health (R01 GM065166 to SEB and DAB, R35 GM118131 to SEB and R35 GM118075 to DAB).

References

- Achsel, T., Brahms, H., Kastner, B., Wilm, M. & Lührmann, R. (1999). *EMBO J.* **18**, 5789–5802.
- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* **D68**, 352–367.
- Agafonov, D. E., Kastner, B., Dybkov, O., Hofele, R. V., Liu, W.-T., Urlaub, H., Lührmann, R. & Stark, H. (2016). *Science*, **351**, 1416–1420.
- Blad, H., Reiter, N. J., Abildgaard, F., Markley, J. L. & Butcher, S. E. (2005). *J. Mol. Biol.* **353**, 540–555.
- Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, H. & Lührmann, R. (1984). *EMBO J.* **3**, 1357–1363.
- Brow, D. A. (2002). *Annu. Rev. Genet.* **36**, 333–360.
- Brow, D. A. & Guthrie, C. (1988). *Nature (London)*, **334**, 213–218.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Didychuk, A. L., Montemayor, E. J., Brow, D. A. & Butcher, S. E. (2016). *Nucleic Acids Res.* **44**, 1398–1410.
- Dokmanić, I., Šikić, M. & Tomić, S. (2008). *Acta Cryst.* **D64**, 257–263.
- Dunn, E. A. & Rader, S. D. (2010). *Biochem. Soc. Trans.* **38**, 1099–1104.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Evans, P. R. (2011). *Acta Cryst.* **D67**, 282–292.
- Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* **D69**, 1204–1214.
- Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H. & Lührmann, R. (2009). *Mol. Cell*, **36**, 593–608.
- Fica, S. M., Tuttle, N., Novak, T., Li, N. S., Lu, J., Koodathingal, P., Dai, Q., Staley, J. P. & Piccirilli, J. A. (2013). *Nature (London)*, **503**, 229–234.
- Galej, W. P., Wilkinson, M. E., Fica, S. M., Oubridge, C., Newman, A. J. & Nagai, K. (2016). *Nature (London)*, **537**, 197–201.
- Ghetti, A., Company, M. & Abelson, J. (1995). *RNA*, **1**, 132–145.
- Hashimoto, C. & Steitz, J. A. (1984). *Nucleic Acids Res.* **12**, 3283–3293.
- Jurica, M. S. & Moore, M. J. (2003). *Mol. Cell*, **12**, 5–14.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Klein, D. J., Moore, P. B. & Steitz, T. A. (2004). *RNA*, **10**, 1366–1379.
- Kondo, Y., Oubridge, C., van Roon, A.-M. M. & Nagai, K. (2015). *Elife*, **4**, e04986.
- Long, L., Thelen, J. P., Furgason, M., Haj-Yahya, M., Brik, A., Cheng, D., Peng, J. & Yao, T. (2014). *J. Biol. Chem.* **289**, 8916–8930.
- Matlin, A. J. & Moore, M. J. (2007). *Adv. Exp. Med. Biol.* **623**, 14–35.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Montemayor, E. J., Curran, E. C., Liao, H. H., Andrews, K. L., Treba, C. N., Butcher, S. E. & Brow, D. A. (2014). *Nature Struct. Mol. Biol.* **21**, 544–551.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240–255.
- Nguyen, T. H. D., Galej, W. P., Bai, X.-C., Oubridge, C., Newman, A. J., Scheres, S. H. W. & Nagai, K. (2016). *Nature (London)*, **530**, 298–302.
- Nguyen, T. H. D., Galej, W. P., Bai, X.-C., Savva, C. G., Newman, A. J., Scheres, S. H. W. & Nagai, K. (2015). *Nature (London)*, **523**, 47–52.
- Petrov, A. S., Bowman, J. C., Harvey, S. C. & Williams, L. D. (2011). *RNA*, **17**, 291–297.
- Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K. W., Li, J. & Nagai, K. (2009). *Nature (London)*, **458**, 475–480.
- Rader, S. D. & Guthrie, C. (2002). *RNA*, **8**, 1378–1392.
- Raghuathan, P. L. & Guthrie, C. (1998). *Science*, **279**, 857–860.
- Reiter, N. J., Blad, H., Abildgaard, F. & Butcher, S. E. (2004). *Biochemistry*, **43**, 13739–13747.
- Ruszkowski, M. & Dauter, Z. (2016). *Protein Sci.* **25**, 1734–1736.
- Sashital, D. G., Cornilescu, G. & Butcher, S. E. (2004). *Nature Struct. Mol. Biol.* **11**, 1237–1242.
- Shannon, K. W. & Guthrie, C. (1991). *Genes Dev.* **5**, 773–785.
- Siliciano, P. G., Brow, D. A., Roiha, H. & Guthrie, C. (1987). *Cell*, **50**, 585–592.
- Song, E. J., Werner, S. L., Neubauer, J., Stegmeier, F., Aspdén, J., Rio, D., Harper, J. W., Elledge, S. J., Kirschner, M. W. & Rape, M. (2010). *Genes Dev.* **24**, 1434–1447.
- Streich, F. C. Jr & Lima, C. D. (2014). *Annu. Rev. Biophys.* **43**, 357–379.
- Strieter, E. R. & Korasick, D. A. (2012). *ACS Chem. Biol.* **7**, 52–63.
- Timani, K. A., Liu, Y., Suvannasankha, A. & He, J. J. (2014). *Int. J. Biochem. Cell Biol.* **54**, 10–19.
- Vedadi, M. *et al.* (2007). *Mol. Biochem. Parasitol.* **151**, 100–110.
- Venditti, V., Clos, L., Niccolai, N. & Butcher, S. E. (2009). *J. Mol. Biol.* **391**, 894–905.
- Verdone, L., Galardi, S., Page, D. & Beggs, J. D. (2004). *Curr. Biol.* **14**, 1487–1491.
- Vidaver, R. M., Fortner, D. M., Loos-Austin, L. S. & Brow, D. A. (1999). *Genetics*, **153**, 1205–1218.
- Wan, R., Yan, C., Bai, R., Huang, G. & Shi, Y. (2016). *Science*, **353**, 895–904.
- Wan, R., Yan, C., Bai, R., Wang, L., Huang, M., Wong, C. C. L. & Shi, Y. (2016). *Science*, **351**, 466–475.
- Weber, G., Trowitzsch, S., Kastner, B., Lührmann, R. & Wahl, M. C. (2010). *EMBO J.* **29**, 4172–4184.
- Weichenberger, C. X. & Rupp, B. (2014). *Acta Cryst.* **D70**, 1579–1588.
- Will, C. L. & Lührmann, R. (2001). *Curr. Opin. Cell Biol.* **13**, 290–301.
- Will, C. L. & Lührmann, R. (2011). *Cold Spring Harb. Perspect. Biol.* **3**, a003707.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Yan, C., Hang, J., Wan, R., Huang, M., Wong, C. C. L. & Shi, Y. (2015). *Science*, **349**, 1182–1191.
- Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. (2016). *Science*, **353**, 904–911.