Published in final edited form as:

Mol Cell. 2010 September 24; 39(6): 939–949. doi:10.1016/j.molcel.2010.08.022.

Structural basis for substrate placement by an archaeal box C/D ribonucleoprotein particle

Song Xue^{1,*}, Ruiying Wang², Fangping Yang², Rebecca M. Terns³, Michael P. Terns³, Xinxin Zhang⁴, E. Stuart Maxwell⁴, and Hong Li^{1,2}

¹Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306, USA

²Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306, USA

³Departments of Biochemistry and Molecular Biology, and Genetics, University of Georgia, Athens, GA 30602, USA

⁴Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695, USA

Abstract

Box C/D small nucleolar and Cajal body ribonucleoprotein particles (sno/scaRNPs) direct site-specific 2'-O-methylation of ribosomal and spliceosomal RNAs and are critical for gene expression. Here we report crystal structures of an archaeal box C/D RNP containing three core proteins (fibrillarin, Nop56/58, and L7Ae) and a halfmer box C/D guide RNA paired with a substrate RNA. The structure reveals a guide-substrate RNA duplex orientation imposed by a composite protein surface and the conserved GAEK motif of Nop56/58. Molecular modelling supports a dual C/D RNP structure that closely mimics that recently visualized by electron microscopy. The substrate-bound dual RNP model predicts an asymmetric protein distribution between the RNP that binds and that methylates the substrate RNA. The predicted asymmetric nature of the holoenzyme is consistent with previous biochemical data on RNP assembly and provides a simple solution for accommodating base-pairing between the C/D guide RNA and large ribosomal and spliceosomal substrate RNAs.

Box C/D sno/scaRNAs (small RNAs, or sRNAs in archaea) are among the many noncoding RNAs that guide enzymatic reactions in Eukarya and Archaea (Hannon et al., 2006; Matera et al., 2007). The large majority of box C/D RNAs guide site-specific 2'-O-methylation of ribosomal and spliceosomal RNA (Decatur and Fournier, 2003; Fatica and Tollervey, 2002; Kiss-Laszlo et al., 1996; Kiss, 2002; Terns and Terns, 2002; Weinstein and Steitz, 1999) while a small subset including the essential U3, U14, and the vertebrate U8 RNAs also guide site-specific cleavage of ribosomal RNA (Kass et al., 1990; Liang and Fournier, 1995;

© 2010 Elsevier Inc. All rights reserved.

Correspondence and requests for materials should be addressed to H.L. (hong.li@fsu.edu).

*Current address: Marine Bioproducts Engineering Group, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, 457 Zhongshan Road, Dalian 116023, China

DATA BASE ACCESSION NUMBERS:

Coordinates and structure factors have been deposited in the Protein Data Bank with accession numbers 3NMU (full complex), 3NVI (Δ_{1-9} complex), 3NVK (Δ_{33-34} complex) and 3NVM (Δ_{1cc} complex).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Peculis and Steitz, 1993; Savino and Gerbi, 1990; Venema and Tollervey, 1995). In RNA-guided processes, the RNA generally recognizes and secures the substrate RNA while partner proteins catalyze the actual chemical reaction. Regardless of type, box C/D RNAs function with a minimal set of 3-4 core proteins (Caffarelli et al., 1998; Gautier et al., 1997; Lafontaine and Tollervey, 1999; Lafontaine and Tollervey, 2000; Lyman et al., 1999; Omer et al., 2002; Schimmang et al., 1989; Tyc and Steitz, 1989; Watkins et al., 1998; Wu et al., 1998). These include fibrillarin (Nop1 in yeast), Nop56, Nop58, and 15.5kD (in yeast, Snu13p and in archaea, L7Ae). A single homologue of Nop56 and Nop58, Nop56/58, is found in archaea. There is solid evidence that fibrillarin is the catalytic subunit responsible for the methyl transfer reaction (Aittaleb et al., 2003; Aittaleb et al., 2004; Tollervey et al., 1993)

A key issue in the mechanism of RNA-guided methylation is how the large and complex rRNA or snRNA substrates access the enzyme active site through pairing with the highly restricted guide RNA. Box C/D RNAs possess a core structure that comprises an internal loop flanked by two similar box C/D and C'/D' motifs, collectively referred to as "the box C/D module" (Figure 1A). The internal loop includes 10-21 nucleotides that are complementary to up to two different substrate RNAs (Bachellerie and Cavaille, 1998). The box C/D module is conserved and contains two hallmark sequence motifs of AUGAUG (box C or C') and CUGA (box D or D') (Kiss-Laszlo et al., 1998). The combined box C (C') and D (D') sequences form a conserved RNA structural motif, called the kink-turn (K-turn), around which the RNP proteins assemble. Snu13p/15.5kD/L7Ae binds specifically to the K-turn (Kuhn et al., 2002; Szewczak et al., 2002; Watkins et al., 2000). This is then followed by binding of Nop56/58 or the Nop56/58-fibrillarin complex (Omer et al., 2002; Rashid et al., 2003; Watkins et al., 2002). The K-turn-L7Ae or K-turn-15.5kD interaction has been visualized at atomic resolution (Charron et al., 2004; Hamma and Ferre-D'Amare, 2004; Moore et al., 2004; Oruganti et al., 2005; Vidovic et al., 2000). Structures of the isolated Nop56/58-fibrillarin complex (Aittaleb et al., 2003; Oruganti et al., 2007) and its complex with the K-turn-L7Ae (Ye et al., 2009) are also available. These structural data lead to a bipartite RNP model in which a single hairpin box C/D RNA is clamped at each end by a set of Nop56/58-fibrillarin-L7Ae proteins leaving the internal loop exposed for interaction with substrate RNAs (Aittaleb et al., 2003; Oruganti et al., 2007; Ye et al., 2009). However, the bipartite model is met with several challenges. First, the length of an intact box C/D RNA is estimated to be shorter than the distance between the two K-turns in the complex, which is established by the long coiled-coil connector domains of Nop56/58. Second, it is expected to be topologically challenging for a large substrate RNA to form more than 10 base pairs with the central internal loop whose size is restricted by flanking proteins.

The box C/D RNP methyltransferase employs a remarkable strategy to specify the site of modification of the substrate RNA, the physical basis of which is unknown. 2'-O-methylation takes place strictly on the substrate RNA nucleotide paired with the 5th nucleotide upstream of the box D or D' (N+5 rule) (Cavaille et al., 1996; Kiss-Laszlo et al., 1996; Tycowski et al., 1996). Conservation of the N+5 rule in both Archaea and Eukarya places the catalytic site uniformly ~20 Å away from the first box D (or D') nucleotide. Currently, none of the available structural models is able to provide a molecular mechanism responsible for the modification site specification owing to the lack of substrate-bound structures.

The traditional bipartite model, in which the pseudosymmetric box C/D RNA lies longitudinally along the Nop56/58-fibrillarin homodimer, has recently been challenged by the electron microscopic observations that a functional box C/D RNP comprises a pair of isologously (face-to-face) bound RNPs (diRNP) (Bleichert et al., 2009). The diRNP predicts that an intact box C/D RNA binds vertically to the Nop56/58-fibrillarin dimer and across

two parallel Nop56/58-fibrillarin dimers (Bleichert et al., 2009). The remaining Nop56/58-fibrillarin halves are occupied by an additional C/D RNA, leading to a functional particle containing four sets of proteins and two box C/D RNA molecules. It was also suggested that the diRNP potentially contains four substrate binding sites, which is believed to be an efficient means for capturing substrate RNA (Bleichert et al., 2009). These predictions, however, could not be confirmed or refuted due to the lack of atomic RNP structures. The previously determined crystal structure of *Sulfolobus solfataricus* (Ss) Nop56/58-fibrillarin bound with L7Ae and a halfmer RNA, confirmed the site of K-turn binding on Nop56/58, but was ambiguous on locations of RNA guide region and fibrillarin owing to the lack of a bound substrate RNA and unwanted crystallographic packing interactions. The processes of RNP assembly and substrate binding, therefore, remain elusive.

Here we report crystal structures of a C/D RNP methyltransferase from the archaeon *Pyrococcus furiosus* (Pf) bound with a substrate RNA at 2.7 Å resolution. The structural complex comprises Pf L7Ae residues 4-124 (full length 124) Pf Nop56/58 residues 8-369 (full length 402), Pf fibrillarin residues 1-227 (full length 227), nucleotides 1-34 of a model box C/D RNA and nucleotides 1-12 of a 13mer substrate RNA (Figure 1A). We describe previously unknown interactions involving conserved protein elements in stabilizing the guide-substrate complex. We also describe construction of a substrate-bound diRNP model that is consistent with both the overall dimension and protein distribution of the electron microscopy model. Finally, we propose mechanisms of substrate binding and catalysis of box C/D RNPs that is consistent with mutational and previously known biochemical data.

Results

Overall structure of the substrate-bound RNP

All three RNP structures were determined by molecular replacement methods using the Pf Nop56/58-fibrillarin complex coordinates (PDBcode: 2NNW). The RNA and L7Ae protein could be modelled into the electron densities phased with the Nop56/58-fibrillarin complex. Details of the structure determinations are described in Experimental Procedures. The full complex contains two sets (protomers) of the three proteins and RNA that dimerize via the coiled-coil interactions between two Nop56/58 proteins. The 13mer substrate RNA is bound to the guide region of the two box C/D RNAs (Figure 1B, 1C, & 1D). Two protomers of both RNPs superimpose well with each other (0.82 Å RMSD for all Ca atoms). Specifically, the substrate-guide duplexes of the two protomers differ by only ~3° in their helical direction, suggesting that crystal packing had little or no impact on RNP formation. The overall structure and domain orientations of the individual RNP components of the full complex are similar to those of the isolated Pf Nop56/58-fibrillarin complex but differ from those of the *Archaeoglobus fulgidus* (Af) and Ss Nop56/58-fibrillarin complexes (Aittaleb et al., 2003; Ye et al., 2009). The key difference lies in the position of fibrillarin, which results from rotations of the N-terminal domain of Nop56/58 around a peptide hinge. In the orientation shown in Figure 1B, the long coiled-coil represents a meridian that divides the bound RNA (guide and substrate) and fibrillarin. The two box C/D RNA modules fall on one side of the meridian (R face) and the two fibrillarin are on the other (F face) (Figure 1B).

It is immediately apparent from Figure 1B that the substrate binds far away from the nearest fibrillarin catalytic subunit. Although it is tempting to conclude that the structure captures an incorrectly bound substrate due to the use of catalytically inactive halfmer box C/D RNA in crystallization, detailed analyses presented below instead suggests an unexpected mechanism of modification and that the observed bound substrate likely represents a productive binding mode.

Role of the GAEK motif in specific C/D RNP assembly

Sequence alignment of Nop56/58 proteins has identified a strictly conserved motif containing Gly-Ala-Glu-Lys tripeptide (GAEK motif), suggesting its important role in box C/D RNP function (supplementary material, Figure S1 and Aittaleb et al., 2003). This region, however, has never been observed in any previously determined RNA-free or RNA-bound complex structures owing to its disorder (Aittaleb et al., 2003; Oruganti et al., 2007; Ye et al., 2009). In the presence of a guide RNA or a guide RNA-substrate complex, the GAEK motif structure unveils its remarkable roles in RNP assembly and substrate binding.

Box C/D RNP methyltransferases exhibit multi-level assembly specificity. L7Ae (Snu13p/15.5kD) binds to the K-turn specifically and this interaction has been studied extensively in isolated L7Ae-RNA complexes (Charron et al., 2004; Hamma and Ferre-D'Amare, 2004; Moore et al., 2004; Oruganti et al., 2005). The C-terminal domain of Nop56/58 subsequently binds to a composite surface of L7Ae and K-turn and facilitates another key specific interaction. The α -helix, α 9A (see secondary element labels in Figure S1) and its C-terminal extension containing the well-conserved GAEK motif (residues 293-312) (collectively termed GAEK motif) are wedged between the guide and the non-guide strands (Figure 2A). Although highly conserved, the entire GAEK motif is unstructured in all previously observed complexes involving Nop56/58-fibrillarin complexes (Aittaleb et al., 2003; Oruganti et al., 2007; Ye et al., 2009). The three terminal base pairs of the box C/D RNA are unwound by α 9A and its C-terminal extension so both strands wrap around the protein element (Figure 2A). This region is free of crystallographic contacts in all structures reported here and its crystal structure likely reflects that in solution.

The interaction between the GAEK motif and the guide RNA is non-specific for bases but highly shape complementary. The C-terminal extension interacts with the non-guide strand through both electrostatic and stacking interactions. Positively charged side chains are observed to interact with phosphate or to form cation- π stacking with nucleobases (Figure 2B). Unexpectedly, the GAEK tetrapeptide is not involved in direct interactions with either the guide or the substrate RNA. The fact that it immediately precedes helix α 9A suggests its conserved role in the RNA-induced formation of α 9A.

In order to assess the importance of the bound RNA to folding of the GAEK motif, we removed the last two nucleotides of the box C/D RNA (Δ_{33-34}) that are tightly associated with α 9A and obtained a 3.2 Å resolution crystal structure of the box C/D RNP containing Δ_{33-34} RNA and the substrate (-G33-C34, Figure 2C). Removal of the two guide RNA nucleotides completely destabilized α 9A as indicated by the disappearance of its electronic density (Figure 2C), demonstrating the mutual induced fit principle in specific box C/D RNP assembly.

These analyses indicate that the GAEK motif-C/D RNA interaction occurs through mutually induced fit and demands an open-ended non canonical (NS) stem. This secondary structure feature distinguishes box C/D RNA from the homologous U4 snRNA whose NC stem is capped by a pentaloop. Correspondingly, the Nop56/58 homolog, hPrp31, does not have the GAEK-like motif (supplementary materials, Figure S1) and the similar peptide region is not involved in binding U4 snRNA (supplementary materials, Figure S2A & S2B). It is possible that this region of hPrp31 is involved in interacting with other RNAs.

Role of the GAEK motif in substrate placement

The α -helix α 9A and its C-terminal extension containing the GAEK tetrapeptide that forms in the presence of the guide/substrate RNA creates a protein groove with L7Ae α 2 (see secondary structure notation in Moore et al. (Moore et al., 2004) and Figure S1). This groove directs the emerging guide strand (Figure 1D) and its duplex with substrate RNA. As

a result, the guide-substrate duplex is nearly perpendicular to the NS stem but parallel to the CS stem of the K-turn. Furthermore, the first two nucleotides of the substrate RNA (sG1, sA2) stack non-specifically on the C-terminal end of α 9A that acts as a resting pole for the substrate RNA of even longer length (Figure 1D & 2A). The guide-substrate duplex follows the natural curvature of the surrounding protein surface that contacts the phosphate backbone of the substrate strand (Figure 1D). There are no specific interactions between the protein subunits and substrate nucleotides, thereby underlining the unique substrate binding mechanism in RNA-guided enzymes. Interestingly, the two guide-substrate duplexes in the asymmetric unit differ at position U12 of the guide RNA. In one of the duplexes, U12 is not paired with substrate nucleotide sA2 and instead is stabilized by stacking with Phe299 (Figure 2B), suggesting that this base pair is not required for correctly positioning the substrate RNA. Biochemical data have provided evidence for little or no substrate base pairing at this position but significant substrate base pairing at other positions (Appel and Maxwell, 2007).

The role of the substrate RNA itself and the GAEC motif interaction in directing the guide-substrate duplex is further demonstrated in two crystal structures that lack either one of the elements. In the previously described Δ_{33-34} complex, where the absence of G33 and C34 of the box C/D RNA led to destabilization of α 9A, a large portion of the guide-substrate duplex is disordered and the remaining portion follows a direction that is nearly perpendicular to that in the wild-type complex (Figure 2C). In addition, we crystallized a guide RNA lacking the guide region (Δ_{1-9}) with L7Ae and a truncated Nop56/58 (amino acids 127-372, Δ Nop56/58). The N-terminal truncation in Δ Nop56/58 removed the fibrillarin binding domain but retained the RNA binding and the coiled-coil domains. In this complex, the first and second guide nucleotides, U12 and C11, point in the opposite direction to that in the substrate-bound complex (Figure 2D). Thus, the substrate RNA has the ability to further order and direct the guide RNA.

Substrate RNA-bound diRNP model predicts cross-RNP catalysis

In order to see if the previously determined diRNP structural model by electron microscopy (EM) (Bleichert et al., 2009) accommodates the substrate-bound halfmer RNP, we duplicated the halfmer RNP structure and manually docked it to the original RNP (Figure 3A). Remarkably, the two halfmer box C/D RNPs complement each other isologously at the R face without severe steric clashes. Furthermore, the core region (Nop56/58 without its N-terminal domain, L7Ae, and guide RNA) of the manually constructed diRNP fits reasonably well to the EM density of the diRNP as a rigid body (Figure 3B). The N-terminal domain of Nop56/58 and fibrillarin could not be docked into the density in their current conformations, reflecting their widely varied locations as previously noted (Aittaleb et al., 2003; Oruganti et al., 2007; Ye et al., 2009; Bleichert et al., 2009). Thus, the close agreement of the crystal structure-derived diRNP to the EM density of the substrate-free and fully assembled diRNP further validates the atomic model of diRNP with bound substrates.

In our diRNP model, the substrate RNA cannot interact with the fibrillarin associated with the same K-turn (Figure 4, *cis* fibrillarin) but can with the fibrillarin associated with the opposing K-turn (Figure 4, *trans* fibrillarin), especially if the fibrillarin is repositioned via a hinge motion to that previously observed in Af Nop56/58-fibrillarin structure (Figure 4, Af fibrillarin). At this position, the target nucleotide 2'-hydroxyl group is 10 Å from the methyltransferase donor S-adenosyl-L-methionine (SAM) bound to fibrillarin and can be positioned for catalysis with minor adjustments (Figure 4). Structural evidence for the described hinge motion in Nop56/58 already exists (Aittaleb et al., 2003; Oruganti et al., 2007; Ye et al., 2009; Bleichert et al., 2009). This analysis leads to an attractive hypothesis that diRNPs catalyze methylation reaction by a cross-RNP mechanism.

Mutational evidence for the cross-RNP catalysis

In order to test the importance of the structural elements critical for assembly and the hinge motion of fibrillarin, we constructed Pf Nop56/58 (Figure 5A) that either shortened the coiled-coil domain or lengthened the peptide hinge and examined the ability of the mutant Nop56/58 proteins in enzyme assembly and methylation. We shortened the coiled coil domain of Pf Nop56/58 (Nop56/58- Δ 1cc) by removing one turn in each helix. We also inserted one (Nop56/58-In1) or three (Nop56/58-In3) amino acids to the Pf Nop56/58 peptide hinge region (Figure 5A).

A shortened coiled-coil is expected to change the longitudinal dimension of the Nop56/58 dimer and, therefore, orientation of the RNA binding surface. We confirmed this prediction by solving a crystal structure of Pf Nop56/58- Δ 1cc-fibrillarin complex (Figure 5B). The mutant complex also forms a homodimer mediated by the coiled-coil in a similar manner as the wild-type complex but differs from the wild-type in its overall dimension and the relative orientation of Nop56/58 domains (Figure 5B). Notably, the coiled-coiled dimerization domain of the homologous U4 snRNP complex, hPrp31-15.5kD-U4, would be similar in length to that of Nop56/58- Δ 1cc. Strikingly, while the Pf Nop56/58- Δ 1cc was able to assemble into an RNP (Figure 5C), it completely abolished methylation activity of the enzyme (Figure 5D), suggesting a stringent requirement for correct orientation of protein domains in methylation.

The peptide hinge connecting the coiled-coil and the N-terminal domains was previously predicted to facilitate fibrillarin repositioning (Oruganti et al., 2007). Accordingly, neither Nop56/58-In1 nor Nop56/58-In3 affected assembly (Figure 5C). However, both mutations severely impaired the RNP's ability in methylation, although Nop56/58-In1 maintained a substantial activity towards the substrate paired with the guide adjacent to box D (Figure 5D). These results are consistent with the occurrence of the hinge motion that most likely serves to reposition fibrillarin during dynamic methylation.

To provide further experimental support for that Af fibrillarin conformation is a functionally important state, we introduced a pair of cysteine residues in Af Nop56/58 (V73C/F288C) that are expected to lock the conformation of the N- and C-terminal domains, and therefore, that of fibrillarin (Figure 5A). We showed by the Ellman method (Creighton, 1989) that the prescribed disulfide bond is indeed formed (supplementary materials, Figure S3) and tested the ability of the cross-linked Af Nop56/58 in RNP assembly and catalysis. Consistent with the structural prediction, the Af V73C/F288C-fibrillarin complex was fully active in RNP assembly and catalysis and to some extent, had higher activity than the wild-type (Figure 5C & Figure 5D). This result further supports the model of dynamic methylation.

Substrate-bound diRNP model predicts an asymmetric holoenzyme

Earlier biochemical studies suggested the possibility that asymmetrically assembled diRNPs may represent an important functional state. Results of the *in vivo* chemical cross-linking and nucleotide analog interference mapping experiments show that the 15.5kD protein is associated with the terminal box C/D module and fibrillarin interacts more extensively with the internal box C/D module (box C'/D') where 15.5kD is absent (Cahill et al., 2002; Szewczak et al., 2002). Furthermore, *in vitro* binding studies showed that archaeal box C/D RNPs are often symmetric with respect to L7Ae binding but can be asymmetric with certain internal loop sequences (Gagnon et al., 2006). Our substrate-bound diRNP model provides some support for these important experimental observations.

In order for fibrillarin to reach the target nucleotide, it is necessary for it to be positioned as that in the Af Nop56/58-fibrillarin complex. This position is supported by the activity of the cross-linked Af Nop56/58-fibrillarin complex described above. This position, however, is

also where L7Ae binds in a fully assembled RNP (Figure 4, bottom). We thus predict that L7Ae is absent from the site of catalysis where the RNA is primarily anchored through the interaction between the guide-substrate duplex and fibrillarin. In addition to accommodate the catalytic subunit, the release of L7Ae also facilitates substrate RNA binding to diRNP. As predicted by Bleichert et al. and suggested by our diRNP model, box C/D RNA do not follow the direction of the coiled-coil domain. Rather, they bind across two RNPs (Figure 6A). This mode of binding is consistent with the optimal separation of 10-12 bases between two box C/D modules (Tran et al., 2005). Our substrate-bound halfmer RNP structure further shows that it is difficult for the two K-turns associated with substrates to make a smooth connection across two halfmer RNPs due to their different paths (Figure 6B & 6C). This dilemma could be resolved by the asymmetric model discussed above in which the site lacking L7Ae releases the K-turn in order for the substrate RNA to bind to diRNP (Figure 6D).

The predicted asymmetric methylation model is already found support in previous studies of eukaryotic and archaeal box C/D RNP assembly. The eukaryotic homolog of L7Ae, the 15.5kD protein, was found to be asymmetrically associated with the box C/D and box C'/D' motifs (Cahill et al., 2002; Szewczak et al., 2002). In a more recent experiment with a mutant *Methanococcus jannaschii* L7Ae (L7Ae EL9-K26Q) that selectively binds box C/D motif only, an asymmetrically assembled archaeal box C/D RNP was constructed (Gagnon et al., 2010). Strikingly, the asymmetrically assembled box C/D RNP catalyzed methylation of the substrate paired with either D or D' guide more efficiently than the wild-type enzyme (Gagnon et al., 2010). Paradoxically, the asymmetrically assembled box C/D RNP was found to contain at least two copies of the single box C/D motif binding L7Ae EL9-K26Q protein. This result could be explained by a weak but direct association between L7Ae EL9-K26Q and Nop56/58 in the classical bipartite assembly model. However, no direct interaction between L7Ae and Nop56/58 has been observed. Thus, this result is better explained by the diRNP model without invoking a direct L7Ae-Nop56/58 interaction. Taken together, simultaneous binding of L7Ae (or 15.5kD) to both box C/D and C'/D' motifs is not necessary and, may in fact, be inhibitory to efficient methylation as predicted by the model of asymmetric methylation model.

Discussion

This work provides insights into two important aspects of the methylation RNP function. First, the high resolution halfmer RNP structures delineate a remarkable molecular mechanism for specific assembly that distinguishes the methylation RNP from the homologous U4 snRNP. Second, the substrate-bound diRNP model suggests a surprising cross-RNP catalysis strategy and predicts an asymmetric holoenzyme.

Box C/D RNP assembles differently than U4 snRNP

Previously, interesting comparisons have been drawn between the eukaryotic box C/D RNP and the spliceosomal 15.5 kD-U4 snRNA complex. In mammals, 15.5kD protein helps to form box C/D RNPs and also assemble the spliceosome by binding to U4 snRNA (Nottrott et al., 2002). The 15.5 kD-U4 complex facilitates subsequent binding of the splicing factor, Prp31 (Nottrott et al., 2002). Prp31 shares a strong homology in its RNA binding domain (or Nop domain) with the Nop56/58 protein and is believed to bind the spliceosomal 15.5kD-U4 complex in a manner similar to the way Nop56/58 binds the 15.5 kD-bound box C/D RNA. Structural and functional studies of human spliceosomal protein hPrp31 have shed light on how its Nop domain interacts with the 15.5 kD-U4 stem II complex (Liu et al., 2007) but could not explain the specificity of Nop56 or Nop58 for the 15.5kD- box C/D RNA complex. Despite the high degree of homology between hPrp31 and Nop56/58, the Nop domain of Nop56/58 interacts specifically with the composite surface formed by both box C/

D RNA and L7Ae and the guide RNA fork. The Nop domain of hPrp31 interacts, in addition to a composite surface formed by U4 RNA and 15.5kD protein, with the U4-specific pentaloop (Liu et al., 2007). This result explains the previous biochemical data on RNP assembly and provides a structural basis for distinct RNP assembly.

A cross-RNP model of assembly and catalysis

Our data is consistent with an assembly model based on the diRNP structure visualized by electron microscopy (Bleichert et al., 2009). Our data further predicts a cross-RNP catalysis mechanism based on the position of the bound substrate RNA. The guide and substrate RNAs are strategically oriented by the well conserved GAEK motif. This mode of binding is likely maintained in the full diRNP structure. The substrate-bound diRNP model shows that the target nucleotide is stably tethered by Nop56/58, and to some extent by L7Ae, to be contacted and modified by the fibrillarin associated with the opposing Nop56/58 in the other halfmer RNP. The fact that the experimentally observed Af fibrillarin can contact the target nucleotide further predicts the functional state of the catalytic subunit, thereby increasing the confidence in the cross-RNP catalysis model.

In the traditional bipartite assembly model where the box C/D RNA binds along a single RNP, the target nucleotide would be modified by the fibrillarin associated with the same Nop56/58 that also holds the substrate RNA. In this model, the protein surface surrounding the target nucleotide makes it difficult for the fibrillarin to contact it, even if the fibrillarin repositions. The cross-RNP strategy suggested by the diRNP model resolves this problem by supplying the catalytic subunit in *trans*. The substrate-bound diRNP model, therefore, provides a valuable perspective in understanding box C/D RNP function.

The observed symmetric archaeal apoenzyme and the predicted asymmetric holoenzyme raise an interesting mechanistic question, namely, does box C/D RNP dynamically change its conformation upon substrate binding or does it maintain a distribution of static conformations? Evidence is available for both archaeal and human box C/D RNPs that suggests a strong likelihood of conformational changes in box C/D RNP function. Human box C/D RNP assembly requires multiple biogenesis factors including ATPases (McKeegan et al., 2007; McKeegan et al., 2009). Temperature plays a key role in symmetric assembly of archaeal box C/D RNPs (Gagnon et al., 2006). However, experiments focusing on substrate-induced conformational changes have not been carried out. Thus, determination of the exact mechanism of the predicted symmetric to asymmetric transition requires further studies.

The model of asymmetric holoenzyme provides an opportunity to understand how box C/D RNPs capture large and complex substrate RNAs. In a symmetrically assembled box C/D RNP, the internal loop possessing two guides is closed by the bound L7Ae at both ends (Figure 7). In the asymmetric model, however, one end of the internal loop is freed, which significantly enlarges the internal loop opening (Figure 7). The rRNA or snRNA substrates must make 10-21 base pairs with one strand of the internal loop. It is clearly more difficult for rRNA or snRNA to thread through a small opening internal loop (linking number $\Delta L = +1$) than to loop through a large opening internal loop (linking number $\Delta L=0$) (Figure 7). We suggest that both formation of the RNA fork by the GAEK motif and the asymmetric distribution of L7Ae/15.5kD protein on box C/D RNA, therefore, provide a necessary means for the substrate-guide interaction to take place (Figure 7).

Experimental Procedures

Protein purification, crystallization and structure determination

The gene encoding recombinant Pf Nop56/58 or the N-terminal domain truncated Δ Nop56/58, and Pf L7Ae and an N-terminal 6xHistidine tag were cloned into a pET28a

vector. The gene encoding recombinant Pf fibrillarin was cloned into a pET24a vector. The proteins were over-expressed in *Escherichia coli* BL21(DE3) and purified from a Ni-NTA column followed by Superdex 200 gel filtration (Amersham Pharmacia) as previously described (Oruganti et al., 2007). All synthetic RNA oligos were ordered from Integrated DNA Technologies and purified according to manufacture instructions. The complexes were similarly formed at a 1.2:1.2:1:1 (1.2:1:1) molar ratio of RNA:L7Ae:Nop56/58:fibrillarin (RNA:L7Ae:ΔNop56/58) with a total concentration of 46-108 mg ml⁻¹ before crystallization. Crystals of the three RNPs were formed at 30°C in hanging drops. The full complex was equilibrated with reservoir solutions containing 0-400 mM potassium chloride, 200 mM-1.5 M sodium chloride, 150-250 mM magnesium acetate, 200 mM ammonium acetate, 50 mM HEPES-NaOH (pH 7.0), and 0-5% PEG 4000. The Δ₁₋₉ complex was equilibrated with reservoir solutions containing 200 mM potassium chloride, 5 mM magnesium chloride, 50 mM TrisHCl (pH 7.5), and 4%-8% PEG4000 or PEG3350. The Δ₃₃₋₃₄ complex used reservoir solutions containing 1.8 M potassium chloride, 40 mM magnesium chloride, 50 mM MES pH 5.6, and 3% PEG8000. All crystals were soaked stepwise in cryosolutions containing the mother liquor plus 10% (v/v) and 15% (v/v) glycerol, respectively, before being flash cooled in a liquid nitrogen stream for data collection. Diffraction data sets were collected at the South Eastern Consortium Access Team (SER-CAT) beamline 22ID and was processed using HKL2000 (Otwinowski and Minor, 1997) (Table S1). All three complex crystals contain two RNPs in the asymmetric unit and have estimated solvent contents of 63%-68%. The full complex crystals are in C2 space group with cell dimensions $a = 293.5 \text{ \AA}$, $b = 94.1 \text{ \AA}$, $c = 96.8 \text{ \AA}$, and $\beta 101.47^\circ$. The Δ₁₋₉ complex crystals are in P2₁2₁2₁ space group with cell dimensions $a = 87.2 \text{ \AA}$, $b = 91.8 \text{ \AA}$, and $c = 155.6 \text{ \AA}$. The Δ₃₃₋₃₄ complex crystals are in P2₁2₁2 space group with cell dimensions $a = 330.0 \text{ \AA}$, $b = 94.6 \text{ \AA}$, and $c = 97.7 \text{ \AA}$. The three structures were determined by a molecular replacement method (MOLREP) (Vagin and Teplyakov, 1997) using either ΔNop56/58 or full-length Nop56/58-fibrillarin monomeric complex (Oruganti et al., 2007) as the search model. In all cases, two solutions were found that correctly reproduced the typical dimer observed in isolated Nop56/58 complex structures. Electron densities phased with proteins alone clearly indicated location of L7Ae and RNA and allowed their placement in the complex. The RNP structures were refined to satisfactory statistics (Table S1) using CNS (Brunger et al., 1998), REFMAC5 (Murshudov et al., 1997) or PHENIX (Adams et al., 2002). The following programs were employed for modelling and displaying: COOT (Emsley and Cowtan, 2004), O (Jones et al., 1991), and Pymol (DeLano).

To generate the duplicated halfmer RNP, transformation with Euler angles (°) (55.3, 163.2, 27.8) and translation vector (Å) = (33.9 6.8 9.4) was applied to the original halfmer RNP. The duplicated halfmer RNP was then fit by hand to the original halfmer RNP to generate the diRNP guided by the overall shape of the EM model. To fit the core region of the manually constructed diRNP model to EM density, we first removed the N-terminal domain of Nop56/58, fibrillarin, and the bound RNA from the manually constructed diRNP coordinates. We then fitted the remaining core protein complex as a rigid body to the EM density (EMD-1636) manually and by the “fit in map” procedure in program Chimera (<http://www.cgl.ucsf.edu/chimera>). The bound guide and substrate RNA molecules were later added to the fitted complex. No effort was made to fit the N-terminal domain of Nop56/58 and fibrillarin, which was explored in details by Bleichert et al. (Bleichert et al., 2009).

Crystals of the Pf Nop56/58-Δ1cc-fibrillarin complex were obtained similarly as those of the wild-type Pf Nop56/58-fibrillarin complex (Oruganti et al., 2007). The crystals are in the same space group as those of the wild-type complex but have a short c -axis of the unit cell ($a=100.2 \text{ \AA}$, $b=100.2 \text{ \AA}$, $c=265.4 \text{ \AA}$ versus $a=100.8 \text{ \AA}$, $b=100.8 \text{ \AA}$, $c=283.2 \text{ \AA}$). The structure

was determined by the molecular replacement method and refined to satisfactory statistics using PHENIX (Adams et al., 2002).

RNA assembly, in vitro methylation, and EMSA analysis

Plasmids carrying the genes encoding Pf Nop56/58 and Af Nop56/58 mutants were obtained using Quikchange kits and the mutant proteins were purified similarly as the wild-type proteins. RNP complexes were assembled and methyltransferase activity assessed as previously described (Tran et al., 2003; Zhang et al., 2006). Briefly, RNP complexes were assembled at 70°C for 10 mins in 80 μ l reaction volumes. Standard assembly reactions were 0.65 μ M of 5' radiolabeled sR8 box C/D sRNA ($\sim 1 \times 10^4$ cpm), 0.8 μ M L7Ae, 3.1 μ M Nop56/58, and 1.1 μ M fibrillarin. Assembled RNPs were then incubated with 6.5 μ M target RNA (IDT), 3.3 μ M S-adenosyl-L-methionine S-adenosyl-L-methionion (SAM) (Calbiochem), and 3 μ l of 3 H-SAM (55 mCi/ml; MP Biomedicals) in a final volume of 110 μ l. After incubation at 68°C for various times (10-60 minutes), 20 μ l aliquots were removed and TCA precipitated onto filter paper. Methyl incorporation into target RNAs was measured by scintillation counting. Methylation assays were performed in triplicate and reported as the average of three experiments. Electrophoretic mobility-shift analysis of assembled RNPs was carried out on native 4-6% polyacrylamide gels containing 2% glycerol and buffered with 25 mM potassium phosphate (pH 7.0). Polyacrylamide gels were dried and visualized by autoradiography or phosphorimaging.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by National Institutes of Health grants R01GM066958 (H.L.) and R01GM54682 (MT and RT) and NSF grant MCB 0543741 (ESM). S. Xue is supported by an American Heart Association Florida/Puerto Rico Affiliate postdoctoral fellow (0725583B). X-ray diffraction data were collected the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID and 22-BM beamlines at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions for APS beamlines may be found at <http://necat.chem.cornell.edu/> and www.ser-cat.org/members.html. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.

References

- Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr*. 2002; 58:1948–1954. [PubMed: 12393927]
- Aittaleb M, Rashid R, Chen Q, Palmer JR, Daniels CJ, Li H. Structure and function of archaeal box C/D sRNP core proteins. *Nat Struct Biol*. 2003; 10:256–263. [PubMed: 12598892]
- Aittaleb M, Visone T, Fenley MO, Li H. Structural and thermodynamic evidence for a stabilizing role of Nop5p in S-adenosyl-L-methionine binding to fibrillarin. *J Biol Chem*. 2004; 279:41822–41829. [PubMed: 15286083]
- Appel CD, Maxwell ES. Structural features of the guide:target RNA duplex required for archaeal box C/D sRNA-guided nucleotide 2'-O-methylation. *RNA*. 2007; 13:899–911. [PubMed: 17438123]
- Bachellerie, J.-P.; Cavaille, J. Small Nucleolar RNAs Guide the Ribose methylations of Eukaryotic rRNAs. In: Grosjean, H.; Benne, R., editors. *Modification and editing of RNA*. Washington, DC: ASM Press; 1998.
- Bleichert F, Gagnon KT, Brown BA 2nd, Maxwell ES, Leschziner AE, Unger VM, Baserga SJ. A dimeric structure for archaeal box C/D small ribonucleoproteins. *Science*. 2009; 325:1384–1387. [PubMed: 19745151]

- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*. 1998; 54:905–921. [PubMed: 9757107]
- Caffarelli E, Losito M, Giorgi C, Fatica A, Bozzoni I. In vivo identification of nuclear factors interacting with the conserved elements of box C/D small nucleolar RNAs. *Mol Cell Biol*. 1998; 18:1023–1028. [PubMed: 9447999]
- Cahill NM, Friend K, Speckmann W, Li ZH, Terns RM, Terns MP, Steitz JA. Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP. *EMBO J*. 2002; 21:3816–3828. [PubMed: 12110593]
- Cavaille J, Nicoloso M, Bachellerie JP. Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides. *Nature*. 1996; 383:732–735. [PubMed: 8878486]
- Charron C, Manival X, Clery A, Senty-Segault V, Charpentier B, Marmier-Gourrier N, Branlant C, Aubry A. The archaeal sRNA binding protein L7Ae has a 3D structure very similar to that of its eukaryal counterpart while having a broader RNA-binding specificity. *J Mol Biol*. 2004; 342:757–773. [PubMed: 15342235]
- Creighton, TE. Disulphide bonds between cysteine residues, in *Protein Structure—A Practical Approach*. 1. Creighton, T., editor. IRL Press at Oxford University Press; New York: 1989. p. 155–157.
- Decatur WA, Fournier MJ. RNA-guided nucleotide modification of ribosomal and other RNAs. *The Journal of biological chemistry*. 2003; 278:695–698. [PubMed: 12431975]
- DeLano, WL. The PyMOL Molecular Graphics System. DeLano Scientific LLC; San Carlos CA USA: <http://www.pymol.org>
- Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004; 60:2126–2132. [PubMed: 15572765]
- Fatica A, Tollervey D. Making ribosomes. *Curr Opin Cell Biol*. 2002; 14:313–318. [PubMed: 12067653]
- Gagnon KT, Zhang X, Agris PF, Maxwell ES. Assembly of the archaeal box C/D sRNP can occur via alternative pathways and requires temperature-facilitated sRNA remodeling. *J Mol Biol*. 2006; 362:1025–1042. [PubMed: 16949610]
- Gagnon KT, Zhang X, Qu G, Biswas S, Suryadi J, Brown BA 2nd, Maxwell ES. Signature amino acids enable the archaeal L7Ae box C/D RNP core protein to recognize and bind the K-loop RNA motif. *RNA*. 2010; 16:79–90. [PubMed: 19926724]
- Gautier T, Berges T, Tollervey D, Hurt E. Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. *Mol Cell Biol*. 1997; 17:7088–7098. [PubMed: 9372940]
- Hamma T, Ferre-D'Amare AR. Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 Å resolution. *Structure*. 2004; 12:893–903. [PubMed: 15130481]
- Hannon GJ, Rivas FV, Murchison EP, Steitz JA. The Expanding Universe of Noncoding RNAs. *Cold Spring Harb Symp Quant Biol*. 2006; 71:551–564. [PubMed: 17381339]
- Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A*. 1991; 47:110–119. [PubMed: 2025413]
- Kass S, Tyc K, Steitz JA, Sollner-Webb B. The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. *Cell*. 1990; 60:897–908. [PubMed: 2156625]
- Kiss-Laszlo Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell*. 1996; 85:1077–1088. [PubMed: 8674114]
- Kiss-Laszlo Z, Henry Y, Kiss T. Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J*. 1998; 17:797–807. [PubMed: 9451004]
- Kiss T. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*. 2002; 109:145–148. [PubMed: 12007400]

- Kuhn JF, Tran EJ, Maxwell ES. Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Res.* 2002; 30:931–941. [PubMed: 11842104]
- Lafontaine DL, Tollervey D. Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability. *RNA.* 1999; 5:455–467. [PubMed: 10094313]
- Lafontaine DL, Tollervey D. Synthesis and assembly of the box C+D small nucleolar RNPs. *Mol Cell Biol.* 2000; 20:2650–2659. [PubMed: 10733567]
- Liang WQ, Fournier MJ. U14 base-pairs with 18S rRNA: a novel snoRNA interaction required for rRNA processing. *Genes Dev.* 1995; 9:2433–2443. [PubMed: 7557394]
- Liu S, Li P, Dybkov O, Nottrott S, Hartmuth K, Luhrmann R, Carlomagno T, Wahl MC. Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science.* 2007; 316:115–120. [PubMed: 17412961]
- Lyman SK, Gerace L, Baserga SJ. Human Nop5/Nop58 is a component common to the box C/D small nucleolar ribonucleoproteins. *RNA.* 1999; 5:1597–1604. [PubMed: 10606270]
- Matera AG, Terns RM, Terns MP. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nature reviews.* 2007; 8:209–220.
- McKeegan KS, Debieux CM, Boulon S, Bertrand E, Watkins NJ. A dynamic scaffold of pre-snoRNP factors facilitates human box C/D snoRNP assembly. *Mol Cell Biol.* 2007; 27:6782–6793. [PubMed: 17636026]
- McKeegan KS, Debieux CM, Watkins NJ. Evidence that the AAA+ proteins TIP48 and TIP49 bridge interactions between 15.5K and the related NOP56 and NOP58 proteins during box C/D snoRNP biogenesis. *Mol Cell Biol.* 2009; 29:4971–4981. [PubMed: 19620283]
- Moore T, Zhang Y, Fenley MO, Li H. Molecular basis of box C/D RNA-protein interactions; cocrystal structure of archaeal L7Ae and a box C/D RNA. *Structure.* 2004; 12:807–818. [PubMed: 15130473]
- Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* 1997; 53:240–255.
- Nottrott S, Urlaub H, Luhrmann R. Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J.* 2002; 21:5527–5538. [PubMed: 12374753]
- Omer AD, Ziesche S, Ebhardt H, Dennis PP. In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proc Natl Acad Sci U S A.* 2002; 99:5289–5294. [PubMed: 11959980]
- Oruganti S, Zhang Y, Li H. Structural comparison of yeast snoRNP and spliceosomal protein Snu13p with its homologs. *Biochem Biophys Res Commun.* 2005; 333:550–554. [PubMed: 15963469]
- Oruganti S, Zhang Y, Li H, Robinson H, Terns MP, Terns RM, Yang W, Li H. Alternative conformations of the archaeal Nop56/58-fibrillarin complex imply flexibility in box C/D RNPs. *J Mol Biol.* 2007; 371:1141–1150. [PubMed: 17617422]
- Otwinowski, Z.; Minor, W. *Processing of X-ray Diffraction Data Collected in Oscillation Mode.* Vol. 276. San Diego: Academic Press; 1997.
- Peculis BA, Steitz JA. Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the *Xenopus* oocyte. *Cell.* 1993; 73:1233–1245. [PubMed: 8513505]
- Rashid R, Aittaleb M, Chen Q, Spiegel K, Demeler B, Li H. Functional requirement for symmetric assembly of archaeal box C/D small ribonucleoprotein particles. *J Mol Biol.* 2003; 333:295–306. [PubMed: 14529617]
- Savino R, Gerbi SA. In vivo disruption of *Xenopus* U3 snRNA affects ribosomal RNA processing. *EMBO J.* 1990; 9:2299–2308. [PubMed: 2357971]
- Schimmang T, Tollervey D, Kern H, Frank R, Hurt EC. A yeast nucleolar protein related to mammalian fibrillarin is associated with small nucleolar RNA and is essential for viability. *EMBO J.* 1989; 8:4015–4024. [PubMed: 2686980]
- Szewczak LB, DeGregorio SJ, Strobel SA, Steitz JA. Exclusive interaction of the 15.5 kD protein with the terminal box C/D motif of a methylation guide snoRNP. *Chem Biol.* 2002; 9:1095–1107. [PubMed: 12401494]
- Terns MP, Terns RM. Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. *Gene Expr.* 2002; 10:17–39. [PubMed: 11868985]

- Tollervey D, Lehtonen H, Jansen R, Kern H, Hurt EC. Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. *Cell*. 1993; 72:443–457. [PubMed: 8431947]
- Tran E, Zhang X, Lackey L, Maxwell ES. Conserved spacing between the box C/D and C'/D' RNPs of the archaeal box C/D sRNP complex is required for efficient 2'-O-methylation of target RNAs. *RNA*. 2005; 11:285–293. Epub 2005 Jan 2020. [PubMed: 15661846]
- Tran EJ, Zhang X, Maxwell ES. Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs. *EMBO J*. 2003; 22:3930–3940. [PubMed: 12881427]
- Tyc K, Steitz JA. U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. *EMBO J*. 1989; 8:3113–3119. [PubMed: 2531075]
- Tycowski KT, Smith CM, Shu MD, Steitz JA. A small nucleolar RNA requirement for site-specific ribose methylation of rRNA in *Xenopus*. *Proc Natl Acad Sci U S A*. 1996; 93:14480–14485. [PubMed: 8962077]
- Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. *J Appl Crystallogr*. 1997; 30:1022–1025.
- Venema J, Tollervey D. Processing of pre-ribosomal RNA in *Saccharomyces cerevisiae*. *Yeast*. 1995; 11:1629–1650. [PubMed: 8720068]
- Vidovic I, Nottrott S, Hartmuth K, Luhrmann R, Ficner R. Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol Cell*. 2000; 6:1331–1342. [PubMed: 11163207]
- Watkins NJ, Dickmanns A, Luhrmann R. 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*. 2002; 22:8342–8352. [PubMed: 12417735]
- Watkins NJ, Newman DR, Kuhn JF, Maxwell ES. In vitro assembly of the mouse U14 snoRNP core complex and identification of a 65-kDa box C/D-binding protein. *RNA*. 1998; 4:582–593. [PubMed: 9582099]
- Watkins NJ, Segault V, Charpentier B, Nottrott S, Fabrizio P, Bachi A, Wilm M, Rosbash M, Branlant C, Luhrmann R. A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell*. 2000; 103:457–466. [PubMed: 11081632]
- Weinstein LB, Steitz JA. Guided tours: from precursor snoRNA to functional snoRNP. *Curr Opin Cell Biol*. 1999; 11:378–384. [PubMed: 10395551]
- Wu P, Brockenbrough JS, Metcalfe AC, Chen S, Aris JP. Nop5p is a small nucleolar ribonucleoprotein component required for pre-18 S rRNA processing in yeast. *J Biol Chem*. 1998; 273:16453–16463. [PubMed: 9632712]
- Ye K, Jia R, Lin J, Ju M, Peng J, Xu A, Zhang L. Structural organization of box C/D RNA-guided RNA methyltransferase. *Proc Natl Acad Sci U S A*. 2009; 106:13808–13813. [PubMed: 19666563]
- Zhang X, Champion EA, Tran EJ, Brown BA 2nd, Baserga SJ, Maxwell ES. The coiled-coil domain of the Nop56/58 core protein is dispensable for sRNP assembly but is critical for archaeal box C/D sRNP-guided nucleotide methylation. *RNA*. 2006

Figure 1.

Schematics of the box C/D RNA and the overall structures of the substrate-bound halfmer RNP. **A.** Schematic drawing of a generic box C/D RNA and the model halfmer RNA used in crystallization. K-turn nucleotides are outlined in red and the substrate nucleotides are colored in yellow. The complex contains guide RNA nucleotides 1-34 and substrate RNA 1-12 or 1-13 for the two protomers respectively. **B.** Structure of the halfmer complex. Both protomers of Nop56/58 are colored green, fibrillarin is in blue, L7Ae is in cyan, guide RNA is in red and substrate RNA is in yellow. The strictly conserved Nop56/58 GAEEK motif is highlighted in pink and SAM molecules are in magenta. “R face” denotes the side of RNA binding and “F face” denotes that of fibrillarin binding. **C.** Electron density (3Fo-2Fc) map at 1σ around the RNA complex. **D.** Close-up view of the RNA binding region.

Figure 2.

Structural basis for box C/D RNP specificity and the roles of the GAEK motif. The same coloring scheme as Figure 1 is used here. **A.** Structural features around the GAEK motif. Positions of GAEK (293-296) tetrapeptide are indicated by grey spheres. **B.** Schematic illustration of the contacts between Nop56/58 residues and the guide RNA in the halfmer RNP complex. Protein residues are indicated by colored oval and those in lighter color fall outside the GAEK motif. The asterisk indicates that the stacking interaction between F299 and U12 is observed in one of the protomers. **C.** Structure of the Δ_{33-34} RNP complex (-G33-C34) in the same orientation as that in A. The GAEK motif is completely disordered. **D.** Structure of the Δ_{1-9} RNP complex (-substrate) in the same orientation as that in A. Note that C11 and U12 are flipped towards α 9A in the absence of the bound substrate.

Figure 3.

The substrate-bound halfmer RNP structure is consistent with diRNP structure. **A.** Construction of the atomic diRNP model by fitting two identical crystal structures face-to-face without steric clashes. **B.** The core region of the manually constructed diRNP model (Nop56/58 is in green, L7Ae is in cyan, guide RNA is in red and substrate RNA is in yellow) was compared to the EM density map (clear). Single capital letters denote the protein subunits (F=fibrillarin, N=Nop56/58, L=L7Ae). Note that the guide-substrate duplexes are not in EM density, which highlights the functional difference between the two structural models. **C.** The substrate-bound diRNP in three different orientations. Note the similar projection dimensions as those of the electron microscopy structural model in B. Top, surface views; Bottom, ribbon diagrams;

Figure 4.

The substrate-bound diRNP model suggests cross-RNP catalysis. Structures of the modelled diRNP are shown on the left column and the corresponding cartoon representations are on the right column. A single guide-substrate RNA complex is shown for clarity. Fibrillarin is colored in blue, Nop56/58 is green and L7Ae is cyan. The guide RNA is in red and the substrate is in yellow. The target nucleotide is indicated by an orange sphere. Top shows the two fibrillarin positions of the diRNP and that of Af fibrillarin obtained by superimposing Af complex to the Pf complex. The fibrillarin associated with the same Nop56/58 bound to the guide RNA is labelled “*cis* fibrillarin” and the opposing fibrillarin is labelled “*trans* fibrillarin”. The previously determined Af Nop56/58-fibrillarin structure was superimposed to the opposing Pf RNP and the Af fibrillarin is labelled “Af fibrillarin”. Note that the “Af fibrillarin” is positioned to bind the guide-substrate duplex and the arrow indicates the predicted transition from the “*trans*” to “Af” position. Middle shows the same structure with fibrillarin molecules removed and the associated SAM molecules shown. SAM molecules are labelled as “*trans* SAM”, “Af SAM” and “*cis* SAM” respectively and is indicated by the red sphere in the cartoon representation. Note that “Af SAM” is the closest to the target

methylation site and the arrow indicates the same transition as in top panel. Bottom panel shows the positional competition between Af fibrillarin and the bound L7Ae and suggests the release of L7Ae during substrate binding and methylation.

Figure 5.

Mutational data supporting cross-RNP catalysis and asymmetric assembly. **A.** Mutational scheme in cartoon representation. **B.** Crystal structure of the Pf Nop56/58-Δ1cc-fibrillarin complex (Nop56/58 is in green and fibrillarin is in blue) in comparison with that of the wild-type (purple). **C.** Gel mobility shift assay of RNP assembly with *M. jannaschii* sR8 RNA. Positions of shifted complexes are labelled on the left and proteins added are indicated on top. The lanes with Nop56/58-fibrillarin complexes all contain L7Ae. **D.** Methylation activities of the wild-type and mutant complexes. Activity is measured by saturated ^3H counts on substrate RNA for various RNPs due to the transfer of methyl group from ^3H -labeled S-adenosyl-L-methionion. “D” indicates methylation by box D guide and “D’” indicates methylation by box D’ guide.

Figure 6.

Illustration of four possible guide RNA paths (red) in the diRNP model. Panels are of the same view as the first in Figure 4B and $\pm$ L7Ae indicates the presence and absence of L7Ae at each K-turn. **A.** Substrate-free guide RNA (from the Δ_{1-9} structure) binds across RNP with sufficient flexibility in the internal loop region and with L7Ae bound (+L7Ae). **B. & C.** Doubly or singly bound substrate RNA interrupts connectivity between two K-turns that are bound with L7Ae. **D.** Singly bound substrate to each guide RNA (from the full complex structure) accommodates cross-RNP binding model if one K-turn is free of L7Ae (-L7Ae) that allows a slight rotation of the guide RNA.

Figure 7. Schematics of the observed symmetric substrate-free diRNP structure and the proposed asymmetric holoenzyme. Possible substrate interactions with the symmetric and asymmetric assemblies are illustrated by yellow substrate RNA.

Table 1

Data collection and refinement statistics (values in parentheses refer to those of the highest resolution shell)

	full complex (PDBid: 3NMU)	Δ_{1-9} complex* (PDBid: 3NVI)	Δ_{33-34} complex* (PDBid: 3NVK)	Δ_{1cc} complex* (PDBid: 3NVM)
<i>diffraction data</i>				
Space group	C2	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2	P3 ₁ 12
Unit-cell parameters (Å)				
<i>a</i>	293.5	87.2	330	100.6
<i>b</i>	94.1	91.8	94.548	100.6
<i>c</i>	96.8	155.6	97.729	265.4
Resolution range (Å)	50.0-2.7 (2.8-2.7)	50.0-2.7 (2.8-2.7)	40.0-3.20 (3.3-3.2)	50-3.4 (3.5-3.4)
No. of unique reflections	61447 (3987)	32071 (2671)	42641 (3238)	32901(1098)
Redundancy	5.8 (4.5)	6.9 (4.1)	8.8 (4.1)	5.2(2.6)
Completeness (%)	89.2 (58.1)	99.2 (92.1)	83.4 (64.3)	80.0(26.8)
<i>I</i> / σ (<i>I</i>)	26.2 (2.2)	32.1 (2.0)	21.1 (2.2)	20.1(1.5)
<i>R</i> _{sym} (%)	7.4 (51.4)	9.6 (61.8)	10.7 (53.3)	9.3(44.5)
<i>Refinement statistics</i>				
Resolution Range (Å)	50.0-2.7 (2.8-2.7)	50.0-2.7 (2.8-2.7)	40.0-3.2 (3.3-3.2)	50.0-3.4(3.5-3.4)
<i>R</i> _{work} (%)	22.0 (32.8)	21.5 (31.9)	23.1 (32.8)	26.5 (41.1)
<i>R</i> _{free} (%)	28.6 (39.2)	26.4 (37.0)	28.6 (35.2)	32.3 (43.2)
Model information				
No. of protein-RNA complexes	2	2	2	1
No. of amino-acid/nucleotide/ligand	1426/93/2	734/48/0	1420/54/2	575/0/0
No. of protein/RNA atoms	11440/1994/52	5814/1026	11374/1162/52	4598/0/0
R.m.s. deviations of the model				
Bond length (Å)	0.09	0.007	0.009	0.012
Bond angle (°)	1.35	1.21	1.523	1.791
Average B-factors				
Nop56/58:L7Ae :fibrillarin:RNA	82.8:120:73.2:170.5	95.6:144.7:82.1:193.4	86.4:100.7:92.8	161.0:154.1
<i>Ramachandran plot</i>				
Residues in most favored region	1169 [81.9%]	654 [89.0%]	1046[73.6%]	414[72.1%]
Residues in additionally allowed region	185 [13.0%]	74 [10.2%]	330 [23.3%]	121[21.2%]
Residues in generously allowed region	72 [5%]	6 [0.08%]	44 [3.1%]	40[6.7%]
Residues in disallowed region	0 [0%]	0 [0%]	0 [0%]	0 [0%]

* Δ_{1-9} complex refers to Pf Nop56/58-fibrillarin complex bound to L7Ae and box C/D RNA 10-33 without the substrate RNA; Δ_{33-34} complex refers to Pf Nop56/58-fibrillarin complex bound to L7Ae and box C/D RNA 1-32 with the substrate RNA; Δ_{1cc} complex refers to Pf Nop56/58- Δ_{1cc} mutant (deletion of one turn of coiled-coil) bound to fibrillarin without L7Ae or RNA.