Box C/D snoRNP catalysed methylation is aided by additional pre-rRNA base-pairing

Robert Willem van Nues[1,3]**,**
Sander Granneman[2,3]**, Grzegorz Kudla**[2]**,**
Katherine Elizabeth Sloan[1]**,**
Matthew Chicken[1]**, David Tollervey**[2]
and Nicholas James Watkins[1,*]

[1]Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, UK and [2]Wellcome Trust Centre for Cell Biology, The University of Edinburgh, Edinburgh, UK

2′-O-methylation of eukaryotic ribosomal RNA (r)RNA, essential for ribosome function, is catalysed by box C/D small nucleolar (sno)RNPs. The RNA components of these complexes (snoRNAs) contain one or two guide sequences, which, through base-pairing, select the rRNA modification site. Adjacent to the guide sequences are protein-binding sites (the C/D or C′/D′ motifs). Analysis of >2000 yeast box C/D snoRNAs identified additional conserved sequences in many snoRNAs that are complementary to regions adjacent to the rRNA methylation site. This 'extra base-pairing' was also found in many human box C/D snoRNAs and can stimulate methylation by up to five-fold. Sequence analysis, combined with RNA–protein crosslinking in *Saccharomyces cerevisiae*, identified highly divergent box C′/D′ motifs that are bound by snoRNP proteins. *In vivo* rRNA methylation assays showed these to be active. Our data suggest roles for non-catalytic subunits (Nop56 and Nop58) in rRNA binding and support an asymmetric model for box C/D snoRNP organization. The study provides novel insights into the extent of the snoRNA–rRNA interactions required for efficient methylation and the structural organization of the snoRNPs.

Introduction

Three of the four eukaryotic ribosomal RNAs (rRNAs), the 18S, 5.8S and 25/28S rRNAs, are co-transcribed as a single precursor rRNA (pre-rRNA) in the nucleolus (Henras *et al*, 2008). The rRNAs undergo extensive covalent modification, including 54 2′-O-methylation and 45 pseudouridylation events in yeast. Modified nucleotides are clustered at functionally important sites in the rRNA, such as the peptidyl transferase domain, and are important for rRNA folding and ribosome function (Decatur and Fournier, 2002). The two major classes of small nucleolar RNA (snoRNA), H/ACA and C/D, guide the pseudouridylation and 2′-O-methylation of rRNA, respectively (Kiss, 2002). A few snoRNAs, including U3 and U17/snR30, do not direct modification but are important for rRNA processing and may aid rRNA folding (Hughes, 1996; Fayet-Lebaron *et al*, 2009).

Each box C/D snoRNA contains highly conserved boxes C and D, generally located at the 5′ and 3′ ends of the RNA, respectively (Reichow *et al*, 2007). Boxes C and D interact forming a stem-internal loop-stem structure known as a kink-turn (k-turn) motif. Most box C/D snoRNAs contain a second, less well-conserved copy of the C/D motif, termed the C′/D′ motif. To guide rRNA methylation, the region adjacent to the D or D′ box in the snoRNA base-pairs with the rRNA and the nucleotide 5 base-pairs from the D or D′ box is targeted for modification (Kiss-Laszlo *et al*, 1996, 1998). Box C/D snoRNAs function as small nucleolar ribonucleoprotein particles (snoRNPs) and are associated with four common core proteins, Snu13 (15.5K), Nop56, Nop58 (Nop5) and the methyltransferase Nop1 (fibrillarin) (Reichow *et al*, 2007). Snu13 binds directly to the box C/D motif, recognizing highly conserved G:A base-pairs, and is the primary RNA-binding protein that triggers recruitment of the remaining snoRNP proteins (Watkins *et al*, 2000, 2002). The human orthologue of Snu13, 15.5K, directly binds the C/D motif but not the C′/D′ motif *in vitro* even though the motifs share the same consensus sequence (Cahill *et al*, 2002; Szewczak *et al*, 2002). It is thought that many C′/D′ motifs cannot form the k-turn structure. An asymmetric distribution of core proteins was revealed through the use of a site-specific crosslinker, 4-thiouridine, inserted at specific sites in the box elements (Cahill *et al*, 2002). It was proposed that Snu13/15.5K, Nop58 and one copy of fibrillarin/Nop1 bound the C/D motif, while the C′/D′ motif was contacted by Nop56 and a second copy of fibrillarin/Nop1 (Cahill *et al*, 2002; Szewczak *et al*, 2002).

Box C/D snoRNP-like complexes (sRNPs) are also present in Archaea, but have symmetrical structures (Dennis and Omer, 2005). Both the sRNA C/D and C′/D′ motifs in archaeal sRNAs form k-loops, and each bind one molecule of the Snu13 orthologue, L7Ae, fibrillarin and the Nop56/58 orthologue, Nop5 (Kuhn *et al*, 2002; Omer *et al*, 2002; Tran *et al*, 2003). Three structural models have recently been proposed for *in vitro* assembled archaeal box C/D sRNPs. One, based

on a crystal structure, proposes a monomeric complex containing a single sRNA and two copies of the proteins, L7Ae, Nop5 and fibrillarin (Ye *et al*, 2009). The other models, based on a single-particle electron microscopy structure (Bleichert *et al*, 2009), and another crystal structure (Xue *et al*, 2010), propose that the sRNP is a dimer, containing two molecules of sRNA and four molecules each of the three proteins. Unfortunately, the RNA was not visible in the single-particle electron microscopy structure and both crystal structures used partial sRNAs. Therefore, the structural organization of these complexes has not been completely resolved. Furthermore, it is also unclear at present how this information relates to the eukaryotic complex.

In yeast, 36 of the 54 2′-*O*-methylation events are directed by sequences adjacent to a C′/D′ motif, several of which diverge significantly from the consensus (Kiss-Laszlo *et al*, 1996; Lowe and Eddy, 1999). Initial identification of the C′ box was predominantly based on analyses of the primary sequence and in most cases the accuracy of these predictions has not been tested. It was generally assumed that the only conserved elements in the snoRNA are the guide regions and

box motifs, but we considered that other conserved, functional elements might exist in the snoRNAs. To address these points, we have taken a global approach to examine the sequence conservation and RNA–protein interactions within the *S. cerevisiae* box C/D snoRNPs.

Results

Identification of unusual C′/D′ motifs in yeast box C/D snoRNAs

It has been proposed that snoRNA C′/D′ and C/D motifs are based on the same core consensus sequences, even though C′/D′ elements are generally less well conserved (Kiss-Laszlo *et al*, 1998). Surprisingly, analysis of the primary sequence failed to clearly identify C′ and/or D′ boxes in many *S. cerevisiae* box C/D snoRNAs. We, therefore, compared the sequence of each *S. cerevisiae* box C/D snoRNA across multiple yeast species (Figure 1A; Supplementary Figure S1). This enabled the identification of C′/D′ motifs in all snoRNAs (Figure 1B; Supplementary Figure S1). There was a surprising amount of variation in C′/D′ sequences between individual
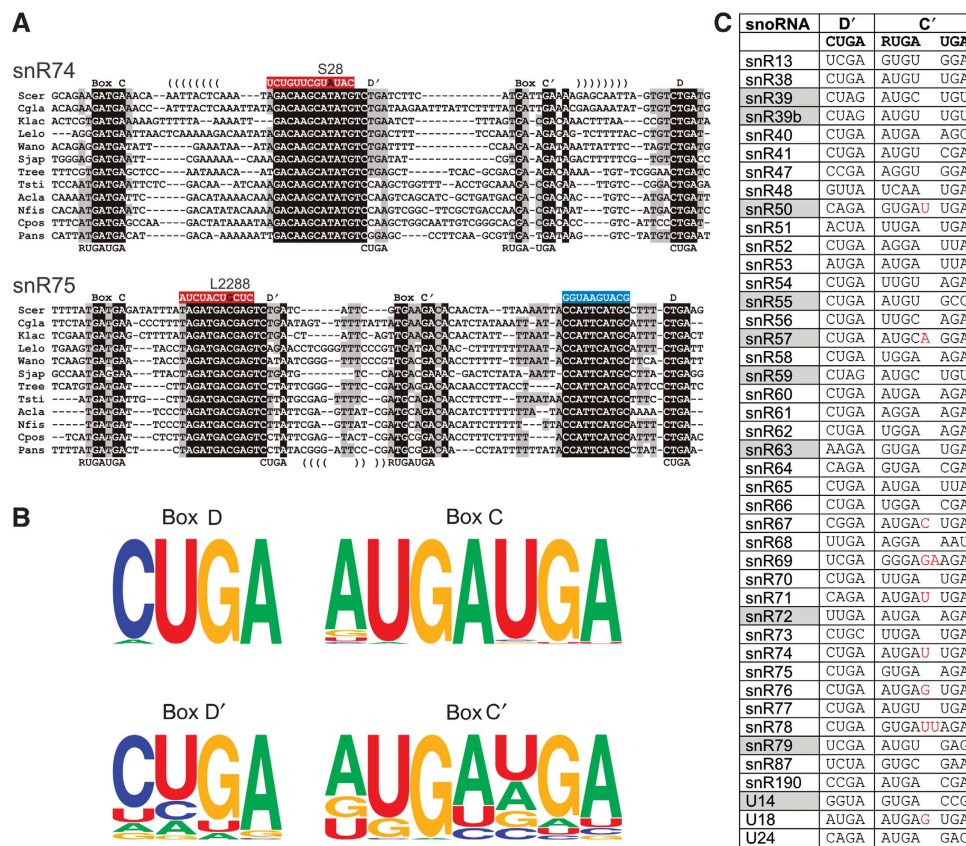

Figure 1 Sequence alignments of box C/D snoRNAs. (**A**) Homologues for each of the *S. cerevisiae* box C/D snoRNAs were retrieved from the fungal genomic sequence databases and aligned. Two example alignments, using a limited subset of the sequences for snR74 and snR75, are shown. The sequence is shown 5′–3′ and the position of the box sequences are indicated, with the consensus sequence shown at the bottom. The rRNA target (3′–5′) is shown in white on a red background. The extra base-pairing target of snR75 is shown in white with a blue background. Identical sequences: white with a black background; conserved sequences: black with a grey background. Brackets indicate possible intra-molecular base-pairing. Scer: *Saccharomyces cerevisiae*; Cgla: *Candida glabrata*; Klac: *Kluyveromyces lactis*; Lelo: *Lodderomyces elongisporus*; Wano: *Wickerhamomyces anomalus* (*Pichia anomala*); Sjap: *Schizosaccharomyces japonicas*; Tree: *Trichoderma reesei* (*Hypocrea jecorina*); Tsti: *Talaromyces stipitatus*; Acla: *Aspergillus clavatus*; Nfis: *Neosartorya fischeri*; Cpos: *Coccidioides posadasii*; Pans: *Podospora anserine*. (**B**) The D′ and C′ sequences of the *S. cerevisiae* box C/D snoRNAs are shown. Insertions in the C′ boxes are indicated in red. The snoRNAs containing box C′/D′ motifs that do not appear to direct methylation are indicated in grey. (**C**) A schematic representation of the conservation of the sequences of the C, D, C′ and D′ boxes of the *S. cerevisiae* box C/D. The diagram was prepared using the WebLogo software (Crooks *et al*, 2004).

snoRNAs; ranging from poorly (e.g., snR51) to highly conserved (e.g., snR53) motifs. Some snoRNAs harbour sequences that are highly conserved in yeast evolution but quite distinct from the accepted consensus, for both D′ (e.g., snR73, snR70, snR51 and snR87) and C′ (e.g., snR68 and U24). Surprisingly, nine C′ boxes in *S. cerevisiae* aligned better to the consensus with between one or two nucleotide insertions (Figure 1B; e.g., snR50 and snR69). In some cases, insertions were present in a subset of the orthologues of a single snoRNA (e.g., snR71) and in some snoRNAs there could be up to nine nucleotides inserted in the split C′ boxes (Supplementary Figure S1; snR190 and snR76). One or two nucleotide insertions were apparent in C′ boxes from vertebrate snoRNAs (e.g., rodent U15B, HBII-234, HBII-82 and mgh28S-2411; data not shown), indicating that this is not specific to yeast. C′/D′ motifs were also highly conserved in snoRNAs that do not appear to use this motif to direct methylation (Figure 1B); the 'guide' region adjacent to the D′ motif in each of these RNAs was not conserved and no target has been identified (Supplementary Figure S1). Indeed, some of these 'inactive' motifs show better sequence conservation than active motifs, suggesting that the C′/D′ motif has a key role in the overall architecture of the snoRNP. The identification of many unusual C′/D′ motif sequences raised questions about the validity of the original consensus sequence. Re-evaluation of the C′ and D′ sequences confirmed that, while the original consensus sequence was correctly identified, significant divergence is tolerated in this motif (Figure 1C). The alignments also identified highly conserved regions present in several snoRNAs (e.g., snR75 and snR70; Figure 1A; Supplementary Figure S1), which do not correspond to either the box or guide regions. We speculated that these could assist in snoRNA function through providing a protein-binding site or additional base-pairing potential (see below).

Asymmetric distribution of proteins in box C/D snoRNPs

We were next interested in defining how the core box C/D proteins contact the box C/D snoRNAs. In particular, we were interested in how the proteins interacted with some of the more divergent C′/D′ motifs and with the additional conserved regions. We therefore performed CRAC, an RNA–protein crosslinking approach, followed by Illumina/Solexa sequencing, with Nop56, Nop58 and Nop1 to identify sites within the snoRNAs that crosslink to the core box C/D proteins (Granneman *et al*, 2009). Briefly, snoRNP proteins were C-terminal tagged with a His6-TEV-Protein A (HTP) tag (Granneman *et al*, 2009) and the fusion proteins were expressed from the genome under control of the endogenous promoter. Cells were UV irradiated at 254 nm to induce RNA–protein crosslinks. RNAs associated with HTP-tagged proteins were partially fragmented, purified and identified by linker ligation, cDNA synthesis and Illumina/Solexa sequencing. Analysis of the derived data revealed that 70–90% of the reads for each protein corresponded to box C/D snoRNA sequences. Reads were identified for every snoRNA, but U3 (snR17), U14 (snR128) and snR4 sequences were particularly enriched (Figure 2A). The different proteins showed significant variation in the number of reads recovered for each individual box C/D snoRNA (Figure 2A). Low-level hits were also recovered for the H/ACA snoRNAs. The highest number of hits for this class of snoRNA were recovered for snR37, which was used as a baseline for background hits (Figure 2A). Note, however, that these reads may reflect genuine interactions taking place within pre-ribosomes.

For each of the proteins, many reads were mapped to 3′ regions of the snoRNAs, near box D, presumably reflecting protein organization on the snoRNA (see Discussion). The average sequence length of the reads in the Solexa data was between 22 and 33 nts, resulting in significant overlap of individual reads. To better localize the binding sites, we generated heatmaps (Figure 2B) and graphs (Figure 2C) of reads smaller than 19 nucleotides. Comparison of the read distribution revealed differential protein localization. Nop58 primarily crosslinked to the 5′ and 3′ ends of the snoRNA, which include boxes C and D (Figure 2B). Nop56 preferentially crosslinked to the C′/D′ motif and the guide regions, in addition to binding near the 3′ end of the snoRNA. Nop1 frequently crosslinked to the guide adjacent to the D box and, occasionally, to the C′/D′ motif and associated guide (Figure 2B). The proteins, in particular Nop56, also make contacts outside the C/D and C′/D′ motifs. The data also indicate an asymmetric distribution of the three proteins on the snoRNAs.

The snoRNAs use either the guide adjacent to the D or D′ box to direct methylation, with some using both guide regions (Lowe and Eddy, 1999). However, no clear differences were observed between snoRNAs with one active guide region or two (Figure 2C; data not shown). The snoRNA sequence alignments revealed some unusual C′/D′ motifs. Several contained C′ and/or D′ boxes that, while evolutionarily highly conserved, differed substantially from the consensus. CRAC data confirmed that these unusual C′/D′ motifs are contacted by core box C/D snoRNP proteins (Figure 2C; e.g., snR70, snR79 and snR51). The same is true for the snoRNAs in which the C′ motif contains an insertion or have poorly conserved C′/D′ motifs (Figure 2C; e.g., snR69). The CRAC data further showed that several snoRNAs with conserved C′/D′ motifs but inactive adjacent guides (e.g., snR39, snR57, snR72 and snR79) are also bound by core proteins. Several snoRNAs contain conserved regions that do not correspond to the box elements or the methylation guides (Figure 1A; Supplementary Figure S1). As CRAC

Figure 2 Asymmetric distribution of core snoRNP proteins on box C/D snoRNAs. (**A**) Box C/D snoRNAs are substantially enriched in core snoRNP protein CRAC Solexa data sets. Total hits for H/ACA and C/D snoRNAs were calculated in each data set, log transformed, clustered and displayed as a heatmap. Box C/D and H/ACA snoRNAs are indicated by brackets. (**B**) Heatmaps of average read densities along box C/D snoRNAs. The positions of C, D, C′ and D′ boxes (black), and the two guide regions (red), are indicated at the top. (**C**) Distribution of reads smaller than 20 nucleotides along individual snoRNAs is shown as plots. The Nop1 hits are shown in red, Nop56 hits are shown in green and the Nop58 hits are shown in blue. The number of hits for Nop58 for both snR57 and snR39 were below those recorded for the H/ACA snoRNA snR37, the baseline for these experiments, and were therefore represented using a dashed line. snoRNA genes and location of conserved sequences (blue), guide sequences (red) or C/D snoRNP boxes (black) are indicated below the *x* axis. Coverage (*y* axis) indicates a fraction and was calculated by dividing the number of times a nucleotide in a gene was found in a read by the total number of hits for the gene.

data resolution is roughly 15–20 nucleotides, we could not unambiguously determine whether core proteins bound these additional conserved regions in most snoRNAs. The con-

served regions within snR70 and snR190 are, however, well over 20 nts away from either the guide regions or the box motifs. Analysis of the crosslinking data revealed that Nop1
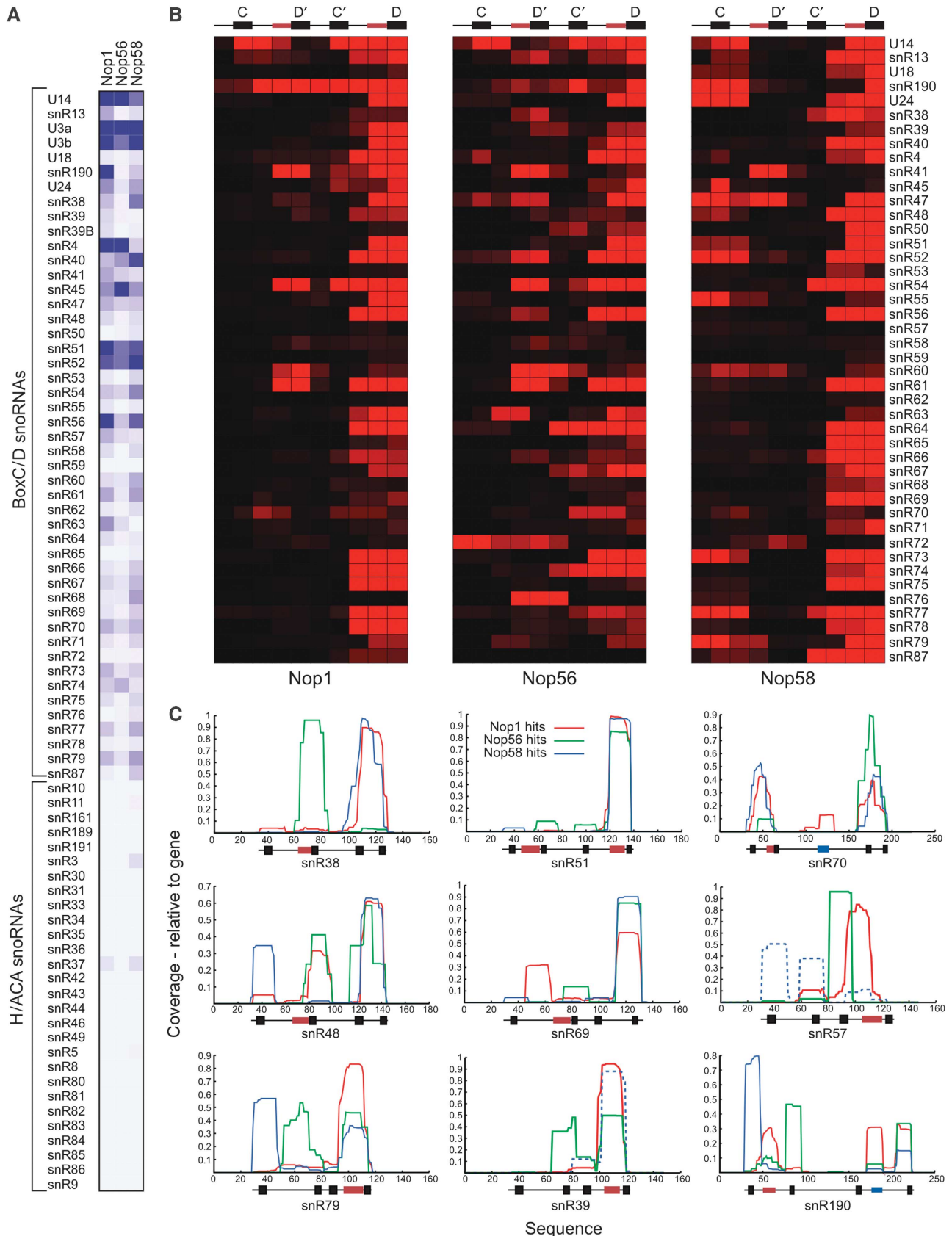
crosslinked to these conserved regions (Figure 2C), suggesting that these conserved elements are recognized by at least Nop1.

Divergent C′/D′ motifs are capable of directing efficient methylation

Since many C′/D′ motifs diverge from the consensus, we next compared the activity of several such motifs in directing rRNA methylation. To perform this, we developed an expression system for an artificial snoRNA designed to target methylation of nucleotide S1316 in the 18S rRNA (Figure 3A) and expressed under the control of a *GAL* promoter. This site is not naturally methylated but its modification does not affect growth (Decatur and Fournier, personal communication; data not shown). We used the human U24 C′/D′ sequence as a standard in the artificial snoRNA as this motif matches the consensus sequence and was previously used to characterize C′/D′ sequence function in yeast (Kiss-Laszlo *et al*, 1998; Qu *et al*, 2011). Several C′/D′ motifs were cloned into this snoRNA construct, including the divergent motifs from snR47, snR51 and snR70, the 'split' motif from snR78 and 'inactive' C′/D′ motifs from snR39,
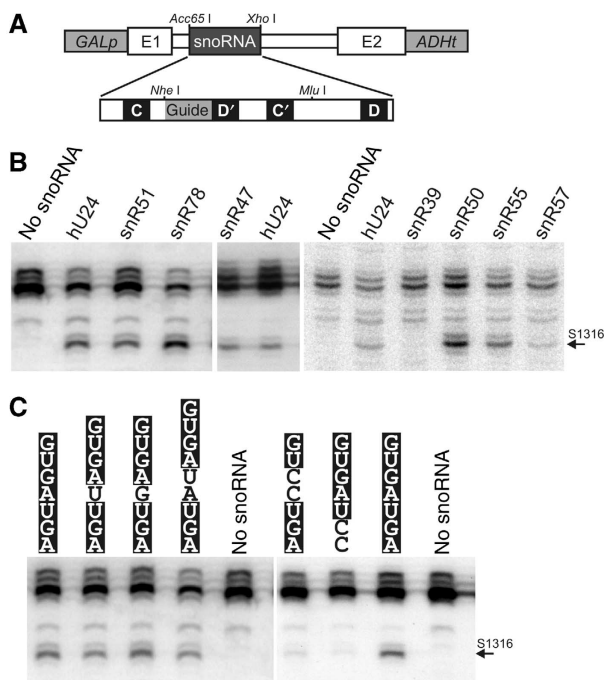

Figure 3 Methylation activity of C′/D′ motifs. (**A**) Schematic representation of the galactose-inducible snoRNA expression cassette. The positions of the GAL promoter (GALp), ADH terminator sequence (ADHt) and exons 1 and 2 of the actin gene (E1 and E2) are shown. The positions of the *Nhe* I and *Mlu* I restriction sites, used in the cloning of the various C′/D′ fragments, are indicated. The C′/D′ sequences cloned into this cassette are shown in Supplementary Figure S10. (**B**, **C**) snoRNAs containing wild-type and mutant C′ boxes (as indicated above each lane) were transformed into yeast cells. RNA was extracted from the cells and analysed by primer extension, using primer Map1316 (upper panel), to detect rRNA methylation, and by northern hybridization (Supplementary Figure S2) to detect the expression of the snoRNA. The position of the stop corresponding to methylation of the target nucleotide, S1316 in the 18S rRNA, is indicated on the right. The snoRNA containing the C′/D′ motif from hU24 was used in all experiments to enable the comparison of the relative methylation activity of the various C′/D′ motifs.

snR50, snR55 and snR57, that are not naturally adjacent to a guide sequence. The resulting plasmids were transformed into *S. cerevisiae* grown on galactose medium to induce snoRNA expression, and RNA was extracted. Methylation status was determined by primer extension (Figure 3B; Figure 5B (snR70)) and snoRNA levels were monitored by northern blotting (Supplementary Figure S2).

Expression of the artificial snoRNA containing the consensus hU24 C′/D′ motif resulted in methylation of S1316 (Figure 3B and C). This was not observed in the absence of the artificial snoRNA. The artificial snoRNAs, containing the C′/D′ motifs from snR51, snR78, snR50 and snR55, all methylated the rRNA to approximately the same level as seen with the hU24 C′/D′ motif. Slightly lower (50%) and higher (180%) methylation was seen, relative to the consensus motif, with the C′/D′ motifs of snR57 and snR47. This indicated that divergent C′/D′ motifs, including those with insertions in the C′ box (snR78), are roughly as active as the consensus C′/D′ motif (hU24). In contrast, a snoRNA carrying the C′/D′ motif of snR39 did not direct detectable methylation activity. The C′/D′ motif of snR39, although very conserved within the gene family, does not appear to naturally direct methylation and may be inactive.

The data on the 'split' C′ motifs predict that between 1 and 3 nt insertions in the C′ box will not block the function of the C′/D′ motif. Insertion of a U or G into the hU24 C′/D′ motif (Figure 3C) directed methylation at site S1316 to 85 and 140% of the levels seen with the wild-type motif. Insertion of 2 nts reduced methylation activity of the snoRNA to 60%. Therefore, inserting 1 or 2 nts into the C′ motif did not significantly affect the activity of the snoRNA. Several C′ boxes contain only half of the C′ consensus sequence, suggesting that only one half of the sequence element may be sufficient for methylation, potentially explaining why insertions are tolerated. However, mutation of either the first or second GA dinucleotide in the C′ box (Figure 3C) reduced methylation at site S1316 by 5- and 10-fold, respectively. Thus, both halves of box C′ used in the artificial snoRNA are essential for modification.

Many box C/D snoRNAs contain conserved elements that are complementary to sequences adjacent to the rRNA target site

We have found that many snoRNAs contain conserved regions that do not correspond to C/D or C′/D′ box elements, or to methylation guides (Figure 1A; Supplementary Figure S1). In some cases, part or all of the conserved sequence appears to support secondary structures in the snoRNA (Supplementary Figure S1; e.g., snR66, snR70 and snR74). However, highly conserved regions identified in 13 yeast snoRNAs (snR13, snR39, snR47, snR48, snR61, snR64, snR70, snR73, snR75, snR76, snR87, snR190 and U18) are complementary to sequences immediately upstream or downstream of the rRNA methylation site (Figure 4A and B; Supplementary Figures S3 and S4). A further 11 yeast snoRNAs (snR38, snR40, snR54, snR55, snR56, snR60, snR62, snR68, snR69, snR71 and snR79) contained complementarity to regions flanking the rRNA methylation site, which was present in most but not all yeasts. Interestingly, these regions were mainly located either between the D′ and C′ boxes (e.g., snR70), or, where there is only one guide, in the second guide region (e.g., snR75; Figure 4B; Supplementary Figures
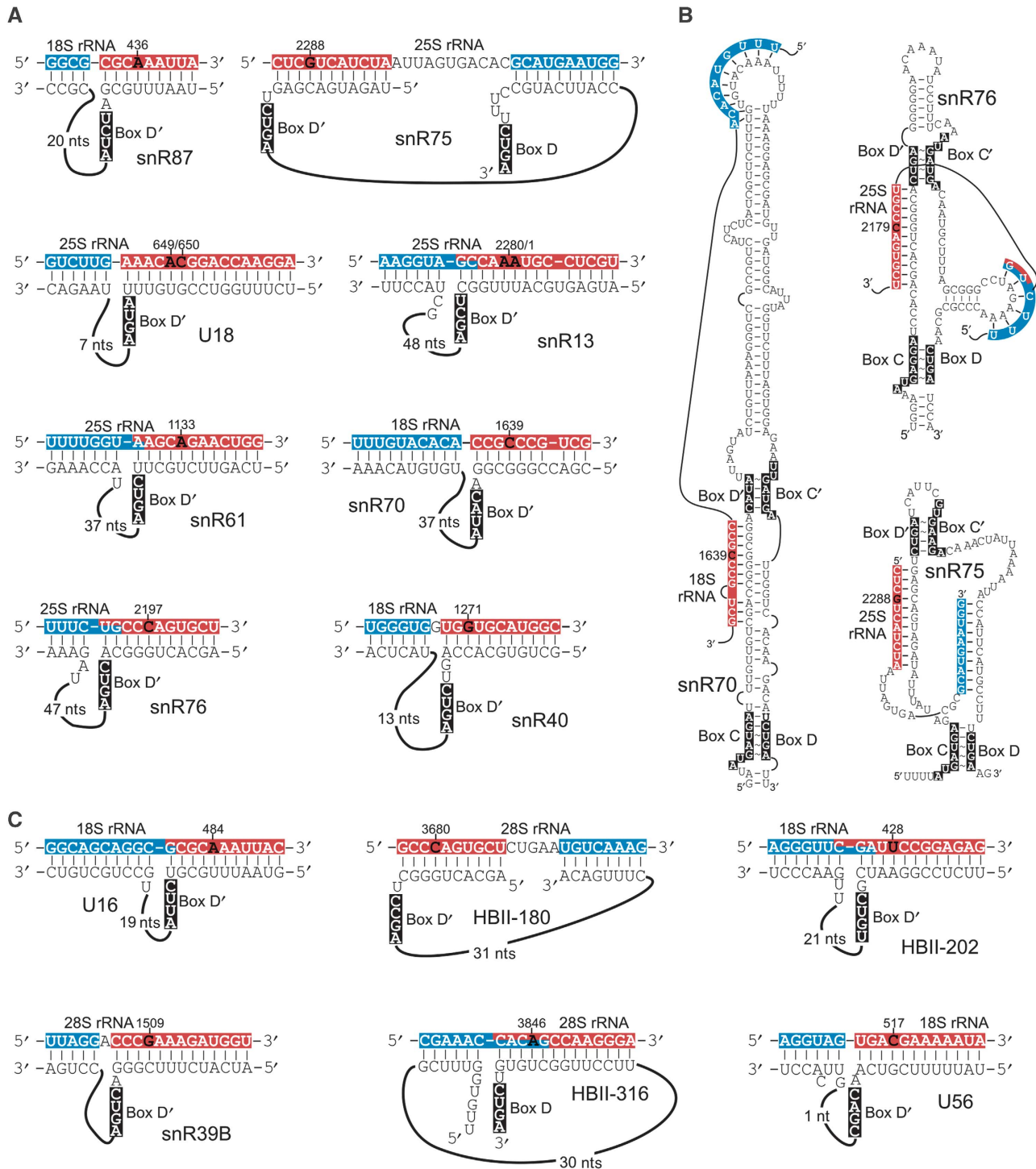
Figure 4 Extra conserved snoRNA sequences are complementary to rRNA target sites. (**A**) rRNA (upper) and snoRNA (lower) sequences, with both conventional guide-rRNA interactions (red) and novel extra base-pairing (blue) interactions for *S. cerevisiae* snoRNAs, are shown. Where sequences are shaded both red and blue, this indicates an overlap between the conventional and extra base-pairing. The D or D′ sequences are shown in white with a black background. (**B**) Schematic representations of *S. cerevisiae* snoRNA secondary structures with rRNA target sequence interactions. The regions base-paired to the guide and the extra base-pairing region are indicated using a red and blue background, respectively. Conserved boxes are indicated and the sequence shown in white on a black background. (**C**) Human snoRNA–rRNA interactions are schematically represented as in (**A**).

S3 and S4). With one exception (snR69), however, the conserved regions were not found adjacent to a D or D′ box and would not be predicted to guide methylation. Occasionally, the regions were found in the spacer region between the guide and C or C′ box (e.g., snR47 and snR60). The potential base-pairing interactions ranged from 4 to 11 base pairs

(Figure 4A). While some of these interactions are relatively short, these regions are likely functioning as an extension of the normal guide region. For snR87, this potential interaction is highly conserved in evolution and the human orthologue, U16 could form 11 consecutive base pairs with the rRNA (Figure 4C). Further analysis revealed an additional

Figure 5 Extra base-pairing sequences are important for efficient methylation. (**A**) snR75 and snR76 interactions with the 25S rRNA are shown. The regions bound by the guide and extra base-pairing sequence are indicated using a red and blue background, respectively. Conserved boxes are indicated and shown in white on a black background. The sequence of the mutated extra base-pairing regions is shown in lower case. (**B**) A *S. cerevisiae* strain, in which the snR72–snR78 cluster was deleted, was transformed with plasmids expressing the snR72–snR78 cluster containing the wild-type (wt) or mutant (mut) snR75 and snR76 snoRNA-coding sequences or the vector alone (−). RNA was extracted from the cells and analysed by site-specific RNase H cleavage, to detect rRNA methylation, and by northern hybridization (Supplementary Figure S2), to detect the expression of the snoRNA. The cleaved RNAs were separated on a glyoxal/agarose gel, stained with ethidium bromide and visualized using a transilluminator. The positions of the full-length rRNAs and the 25S (arrows) and 18S (asterisk) cleavage products are indicated on the right. Reactions were performed in the presence (+) or absence (−) of RNase H as indicated. The oligonucleotides used for the analysis of the snR75 (upper panel) and snR76 (lower panel) modification sites are indicated on the left. (**C**) The region between the D′ and C′ boxes of the snR70 snoRNA was cloned into the artificial snoRNA (Figure 3A) to target the site S1315 in the 18S rRNA (snR70C′/D′). The extra guide region was then mutated (sequence shown in lower case) so that it was complementary to the region just upstream of the 18S rRNA target site. (**D**) Plasmids expressing the snoRNAs and a snoRNA containing the human U24 C′/D′ motif (targeting S1316) were transformed into yeast. RNA was extracted and analysed by primer extension using primer Map1316 and by northern blotting (Supplementary Figure S2). The positions of the stop corresponding to methylation of the target nucleotides, S1316 and S1315 in the 18S rRNA, are indicated on the right.

17 potential snoRNA–rRNA base-pairing interactions for human box C/D snoRNAs that were evolutionarily conserved in higher eukaryotes (Figure 4C; Supplementary Figures S5–S7; U15, U21, U46, U49, U56, U103, snR39b, HBII-180, HBII-202, HBII-429, HBII-142, HBII-210, HBII-99, HBII-296, HBII-316, HBII-82 and mgh28S-2411). Interestingly, in both yeast and human snoRNAs, some of the extra potential base-pairing interactions overlap with the canonical target-snoRNA interaction by one or two nucleotides. In the case of HBII-316, however, the overlap is five nucleotides (Supplementary Figure S6). The data, therefore, suggest that the additional base-pairing might either enhance or regulate rRNA methylation.

Extra base-pairing stimulates rRNA methylation

The proximity of the base-pairing to the methylation target site suggests a role in rRNA methylation, either stabilizing or facilitating formation of snoRNA–rRNA interactions. To test this possibility, we generated a yeast strain in which the snR72–snR78 cluster was deleted. This strain was then transformed with a plasmid expressing the wild-type snoRNA cluster, or a cluster in which the additional base-pairing region of either snR75 or snR76 was mutated (Figure 5A). RNA was extracted from each strain and the methylation levels were monitored by site-specific RNase H cleavage

(Yu *et al*, 1997) directed by chimeric 2′-*O*-methyl RNA/DNA oligonucleotides to sites L2288 (snR75) and L2197 (snR76). This approach was used as primer extension analysis of methylation at site L2197 proved unreliable. The snoRNA expression levels were analysed by northern blotting (Supplementary Figure S2). In the absence of snR75 and snR76, both oligonucleotides directed RNase H-mediated cleavage of >90% of the 25S rRNA (Figure 5B). Each oligonucleotide also resulted in the non-specific cleavage of 18S rRNA (Figure 5B; asterisk). Importantly, this was not affected by the presence or absence of the snoRNAs. Expression of wild-type snR75 and snR76 rendered the 25S rRNA resistant to RNase H cleavage at the two methylation sites (<5% cleaved). In contrast, significant levels of cleavage were seen at sites L2288 and L2197 in 25S rRNA derived from cells expressing mutant snR75 (snR75mut) and snR76 (snR76mut), respectively. Mutation of the conserved extra base-pairing region of snR76 resulted in a 2–3-fold reduction in methylation (35–50% 25S rRNA uncleaved). Similarly, mutation of the conserved region of snR75 resulted in a 4–5-fold reduction in methylation (20–25% 25S rRNA uncleaved). A similar reduction in rRNA methylation was observed for the snR75mut, relative to the wild-type snoRNA, when the RNA was analysed by primer extension (Supplementary Figure S8). The data, therefore, indicate

that the putative extra base-pairing regions in both snR75 and snR76 are important for efficient methylation activity.

If the conserved regions function through base-pairing, then they should only enhance methylation activity at a natural target site in the rRNA. Therefore, if the snoRNA is mutated to modify a new target site, the conserved region should only enhance methylation when it is mutated to be complementary to the region adjacent to the new target site. We, therefore, cloned the C′/D′ region of snR70, including the intervening stem structure and extra base-pairing region, into the artificial snoRNA system so that it would direct methylation at site S1315 in the 18S rRNA (Figure 5C). A construct was generated in which the conserved loop region, containing the extra base-pairing sequence, was mutated to be complementary to the sequence immediately upstream of the new target site. The snR70 construct directed methylation at site S1315 in the 18S rRNA (Figure 5D) at a level comparable to that seen with the hU24 consensus C′/D′ motif (at site S1316). Mutation of the loop sequence to generate a sequence complementary to the flanking region immediately upstream of the target site resulted in a reproducible five-fold increase in methylation at site S1315. This, therefore, provides strong evidence that the conserved regions can function to enhance methylation through base-pairing adjacent to the target site.

Discussion

We have used an extensive analysis of the sequence conservation of the snoRNAs, together with a high-throughput analysis of the RNA–protein contacts in the box C/D snoRNPs to better understand the structure and function of these complexes. The consensus sequence originally identified for this motif is, however, generally correct (Figure 1C; Kiss-Laszlo *et al*, 1998), but a quite unexpected degree of sequence diversity in box C′/D′ motifs is tolerated. The first GA dinucleotide in box C′ (RUGAUGA) and the GA dinucleotide in box D (CUGA) are the most conserved elements. The equivalent nucleotides also form the most conserved part of the box C/D motif (Xia *et al*, 1997). In both k-turn (C/D) and k-loop (archaeal C′/D′) structures, these GA dinucleotides form sheared GA base-pairs and comprise the core binding site for Snu13 (C/D) and L7Ae (C/D and C′/D′), suggesting that Snu13 directly binds the C′/D′ motif (Moore *et al*, 2004; Oruganti *et al*, 2005; Suryadi *et al*, 2005). We were unable to generate the tagged construct required for CRAC on Snu13 to determine whether it contacts the C′/D′ motif. It has, however, recently been shown that Snu13 is recruited to the C′/D′ motif *in vivo* (Qu *et al*, 2011). Furthermore, a L7Ae mutant, which cannot bind alone to the C′/D′ motif but still associates with the C/D motif, is recruited to the C′/D′ motif of *in vitro* assembled archaeal snoRNPs (Gagnon *et al*, 2010). From, this it was proposed that protein–protein contacts in the snoRNP contribute to Snu13 recruitment to the C′/D′ motif.

We have shown that even highly divergent C′ and D′ sequences can bind core snoRNP proteins and direct efficient 2′-*O*-methylation, including C′ sequences with one or two nucleotide insertions. These changes are obviously tolerated but in some cases it is hard to rationalize how the proteins recognize and bind these divergent elements. We were particularly surprised to observe that the C′ box in snR190 in some species contained insertions of up to nine nucleotides.

We have not experimentally tested such large insertions and cannot exclude the possibility that these C′ boxes are inactive. We were, however, unable to find any good candidates for an alternative C′ box in these snoRNAs. All C′/D′ elements tested were functional in our artificial snoRNA system, with the exception of the motif from snR39. The C′/D′ motif of snR39 does not have a naturally active guide sequence making it difficult to determine whether this motif has the potential to direct methylation. It is possible that some C′/D′ motifs are only active in the context of the parent snoRNA.

Our data indicate that snoRNAs that use only the C/D motif for methylation, also contain a C′/D′ motif that binds the core box C/D proteins and, in most cases, can drive methylation in our artificial snoRNA. This implies that the methylation guide snoRNPs have a conserved architecture regardless of whether both motifs function or not. The CRAC data are consistent with the asymmetric snoRNP model proposed by Steitz and colleagues; Nop56 and Nop1 contact the C′/D′ motif whereas all three proteins contact the box C/D motif (Cahill *et al*, 2002). Indeed, for all proteins, the sequences in the CRAC experiments are biased towards the D box. One possible explanation for this is that the core proteins form a very stable complex with the box C/D motif and bind less stably to the C′/D′ motif. This is consistent with the difference in the sequence conservation between the two. However, we cannot completely exclude a bias in the CRAC cloning protocol. It was recently suggested, from a crystal structure of an incomplete archaeal box C/D sRNP, that the fibrillarin bound to the C′/D′ motif catalyses methylation of the rRNA bound at the D box guide and vice versa (Xue *et al*, 2010). While we cannot exclude the model proposed for the archaeal sRNPs, our data strongly suggest that the proteins bound to the C′/D′ motif direct methylation at the D′ guide. In addition to contacting the box regions, all three proteins also made significant contacts to the guide regions in the snoRNAs. While expected for Nop1, this was somewhat surprising for Nop56 and Nop58. Interestingly, the novel extra base-pairing sequences, identified by sequence analysis, were also contacted by Nop1. These data are consistent with contacts made by these proteins with the rRNA (Granneman *et al*, 2009), and suggest that all three proteins have a role in substrate binding and/or release. Consistent with our results, recent analysis of archaeal box C/D sRNPs, using UV cross-linking analysis (Ghalei *et al*, 2010) and structure determination (Xue *et al*, 2010), identified important contacts between the protrusion in the NOP domain of the Nop56/Nop58 homologue, Nop5, and the spacer/guide region of the box C/D sRNA.

Sequence comparisons identified novel, phylogenetically conserved elements in individual yeast and human box C/D snoRNAs. Many of these are complementary to the sequence adjacent to the methylation target site in the rRNA and some are conserved from yeast to humans (e.g., snR87/U16). These interactions previously escaped notice, probably because they are generally more evolutionarily divergent than 'traditional' guide-target regions. We showed that these sequences can be important for efficient *in vivo* methylation by two endogenous yeast snoRNAs and that additional base-pairing to a region adjacent to the target site stimulates methylation. Some of these extra base-pairing interactions are quite short (e.g., 4 bp for snR87) and should be viewed as extensions to the natural guide base-pairing. Strikingly, snoRNAs with the
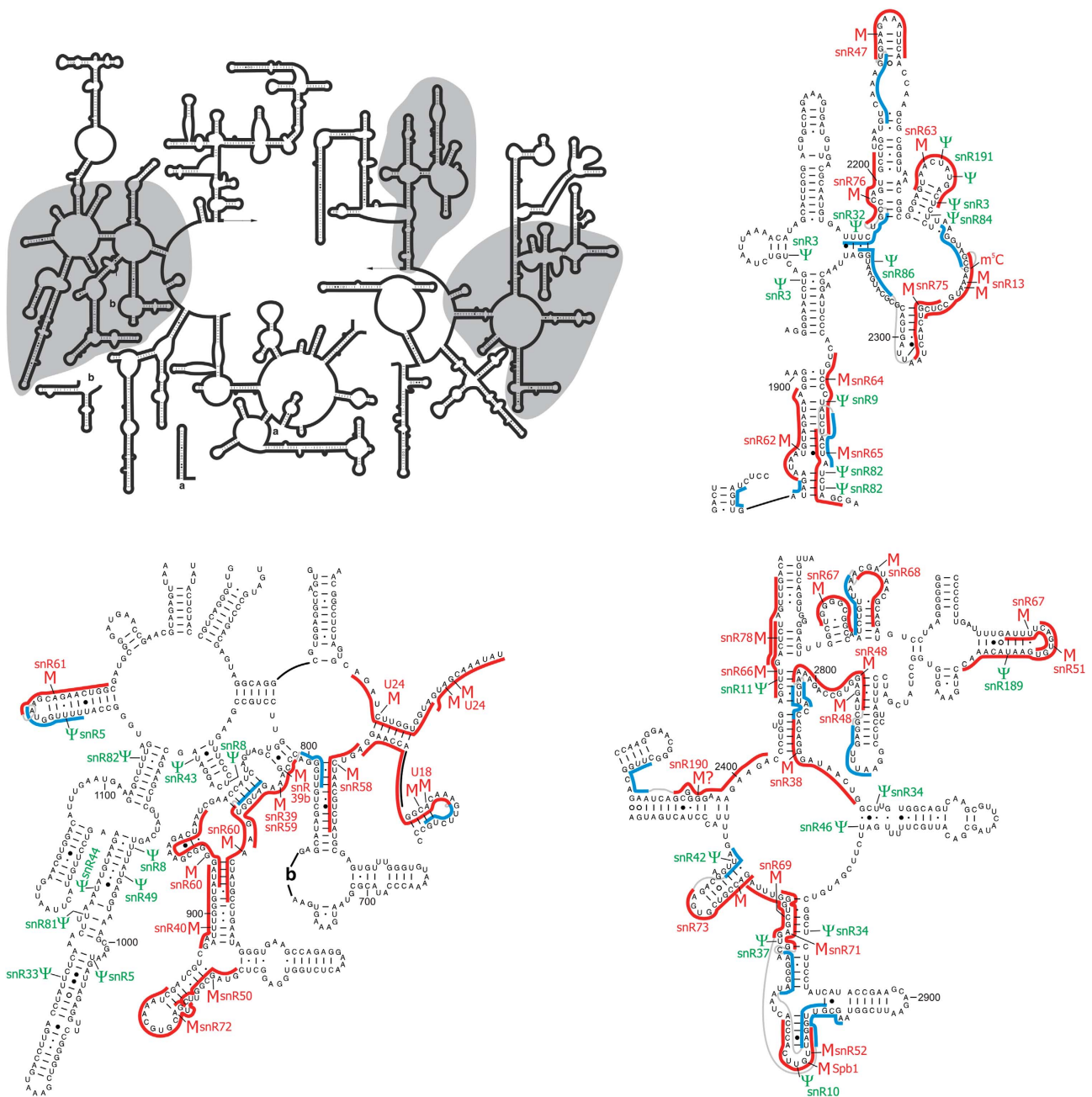
Figure 6 snoRNA base-pairing with the 25S rRNA. A line drawing of the secondary structure of the *S. cerevisiae* 25S and 5.8S rRNAs is shown at the top. The three regions containing modifications are shaded grey. The detailed secondary structures (obtained from http://www.rna.ccbb. utexas.edu/) of the three modified regions are also shown with the methylation (M) and pseudouridylation (Ψ) sites, and modifying snoRNAs, indicated in red and green, respectively. The methylation guide and extra base-pairing interaction sites in the rRNA are indicated by red and blue lines, respectively. Grey lines connect the lines for conventional guide and extra base-pairing interactions from one snoRNA.

shortest guide-rRNA base-pairing interactions (e.g., snR70, snR13 and snR87) all contained extra base-pairing, which likely increases the specificity of the snoRNA–rRNA interaction (Supplementary Figure S9). Of course, these snoRNAs might have shorter guides because the extra base-pairing reduces the evolutionary pressure to maintain long guide-rRNA base-pairing interactions. We found that Nop1 cross-linked to the extra base-pairing region of snR70 and snR190. It is possible that Nop1 interacts with other extra base-pairing regions but due to their proximity to the box regions in other snoRNAs, we could not clearly determine this. The targets of the extra base-pairing regions are in close proximity to the

methylation site. If the two guide and extra base-pairing regions base-pair to the rRNA simultaneously it is possible that Nop1, which as the catalytic subunit recognizes the guide-rRNA interaction, also contacts the extra base-pairing region in the snoRNA.

Not all snoRNAs require additional base-pairing to function and it is likely that this reflects the nature of the rRNA target site. The additional base-pairing presumably stabilizes the snoRNA–rRNA interaction, which could aid access to highly structured regions of the rRNA—where snoRNA interactions predominantly occur (see Figure 6). It might be envisaged that the extra base-pairing assists in

opening strong secondary structure at the modification site (e.g., U18, snR61, snR47 and snR75) by initially docking the snoRNP close to the modification target, allowing the guide region to then compete with the local secondary structure. In addition, extended rRNA base-pairing might aid in competition between snoRNAs with overlapping target sites (see Figure 6). This is particularly striking over the inter-subunit bridge (helix 69) in the 25S rRNA. Other examples include snR40 and snR55 that modify nucleotides S1269 and 1271 in the 18S rRNA (Supplementary Figure S3). This situation is likely to be even more significant in humans where there are about twice the number of modifications. In several snoRNAs, for example, snR13, snR76 and snR64, the rRNA targets for the extra base-pairing region and the methylation guide region overlap. This may indicate that these two base-pairing interactions do not occur simultaneously. In the case of human HBII-316, the overlap is five nucleotides and includes the methylation target site. It is conceivable that this extra base-pairing interaction might regulate methylation activity. Interestingly, snR47 contains a complex series of antisense sequences (Supplementary Figures S1 and S3) and modifies sites in both 18S and 25S rRNA. It is, however, unclear whether the snoRNA simultaneously base-pairs to both sites. It is likely that the extra base-pairing regions will also influence the timing of snoRNP association with the rRNA, the involvement/requirement for RNA helicases, and rRNA folding.

Materials and methods

snoRNA alignments
Genomic DNA sequences of Ascomycota (Saccharomycotina, Schizosaccharomycetes and Pezizomycotina) were searched iteratively for homologues using blastn with setting 'expect 100' (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi?organism=fungi or http://www.yeastgenome.org/cgi-bin/blast-fungal.pl). Pezizomycotina snoRNAs were also identified by initiating blast searches with sequences of snoRNAs identified in a *Neurospora Crassa* cDNA library (Liu et al, 2009). Searching EMBL EST Fungi cDNA libraries (http://www.ebi.ac.uk/Tools/sss/wublast/nucleotide.html) yielded further snoRNAs from Basidiomycota, which were not found in the systematic blast searches. Sequences were aligned using Clustal W and annotated on the basis of visually identified conserved motifs and phylogenetically supported secondary structure. A similar approach, using just the basic nr/nt nucleotide collection database, was used to identify and align sequences for vertebrate U15 and U16 snoRNAs. All other vertebrate snoRNA alignment data were derived from the snoRNABase database (http://www-snorna.biotoul.fr/).

CRAC experiments and bioinformatics
Nop1, Nop56 and Nop58 CRAC experiments were described elsewhere (Granneman et al, 2009). Reads were aligned against the *S. cerevisiae* genome using novoalign 2.05 (http://www.novocraft.com; settings –r Random, -s –h190 –a) and processed using in-house python scripts (paper in preparation). Heatmaps of log-transformed data shown in Figure 2A were generated using Java TreeView and Cluster3.0 with default settings. To calculate the densities of reads on the snoRNAs, we analysed reads between 15 and 19 nucleotides in length (after trimming linkers). This allows precise identification of protein-binding sites, but is long enough to map reads uniquely to the yeast genome. Each snoRNA was divided into 10 regions: 'before C box', 'C box', 'between C and guide', 'guide 1', 'D' box', 'between D' and C'', 'C' box', 'between C' and guide 2', 'guide 2' and 'D box'. For snoRNAs that lacked certain features, relevant regions were merged. The numbers of hits per million mapped reads were calculated separately for all nucleotide positions, and then averaged to yield read densities for each region. The densities were converted to a heatmap using Java TreeView,

with 'Contrast' set to 5.0, 0.7 and 1.5 for Nop1, Nop56 and Nop58, respectively.

Analysis of rRNA methylation
Methylation activity of various C'/D' motifs was analysed *in vivo* in W303 (MATa/MATá; leu2-3112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15; [phi+]) using an artificial snoRNA construct inserted in the intron of the actin gene. The actin/snoRNA cassette (a PCR amplified 1-kb *Bam* HI–*Xba* I fragment derived from pFL45/ACT/XK; Kiss-Laszlo et al, 1996) was placed under the control of the *GAL1* promoter (amplified as a 1-kb *Eco* R1–*Bam* HI fragment from pBL143; Liu and Fournier, 2004) and cloned into pRS416 in which the *Acc65* I and *Xho* I sites of the multiple cloning site had been deleted. Artificial snoRNAs were assembled from oligonucleotides and cloned in between the unique *Xho* I and *Acc65* I sites present within the actin intron (Figure 3A; Supplementary Figure S10; Kiss-Laszlo et al, 1996). C'/D' regions, and the target site guide, were subsequently assembled from oligonucleotides and cloned into the *Nhe* I and *Mlu* I sites in the snoRNA-coding sequence (Figure 3B; Supplementary Figure S10). The target site (S1316 or S1315) was chosen as a site detectable by reverse transcription that is not naturally modified. The expression levels of the artificial snoRNA were confirmed by northern blotting (Supplementary Figure S2).

Wild-type and mutant snR75 and snR76 were expressed in their natural polycistronic context of the snR72–snR78 gene cluster. The cluster was PCR amplified using primers snR72r (5'-AAAAGG TACCGTTATCCGTACACTTGACCTC-3') and snR78f (5'-AAAACTC GAGAAGCATGAGGTATTATAGCGAC-3') and was cloned into the *Acc65* I/*Xho* I sites of pRS416 and transformed into YPH499 (MATa, ura3-52, lys2-801, ade2-101, leu2Δ1, his3-Δ200, trp1-Δ63) in which the non-essential snoRNA gene cluster (Qu et al, 1999) was replaced by a natNT2 cassette (Janke et al, 2004). The snR75 and snR76 mutants were generated by site-directed mutagenesis.

Methylation activity was determined by reverse transcription under limited nucleotide and enzyme concentrations (Maden, 2001). In all, 8 µg total RNA was annealed to ^{32}P-, 5'-end labelled primer and then incubated with M-MLV reverse transcriptase (40 u, Promega), 2 µl 5× RT buffer, 0.25 µl superasin and either 12.5 or 1.25 mmol dNTP's. The reactions were separated on either a 6 or 8% polyacrylamide/7 M urea gel and then visualized using a phosphorimager. Primers used for mapping were Map1316 (5'-TAGTCCCTCTAAGAAGTGGATAACC-3') and Map75 (5'-CTAGATAG TAGATAGGGACAGTGG-3'). In addition, methylation was also monitored by site-specific RNase H cleavage (Yu et al, 1997) directed by a chimeric 2'-O-methyl RNA/DNA oligonucleotides to sites L2197 (5'-mAmCmUGGGCmAmGmAmAmAmUmCmAmCmA mUmU-3') and L2288 (5'-mUmGmACGAGmGmCmAmUmUmUm GmGmCmUmAmC-3'). Cleaved RNA was separated on a 1.2% agarose/glyoxal gel, stained with ethidium bromide and visualized using a transilluminator.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Bleichert F, Gagnon KT, Brown II BA, Maxwell ES, Leschziner AE, Unger VM, Baserga SJ (2009) A dimeric structure for archaeal box C/D small ribonucleoproteins. *Science* **325:** 1384–1387

Cahill NM, Friend K, Speckmann W, Li ZH, Terns RM, Terns MP, Steitz JA (2002) Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP. *EMBO J* **21:** 3816–3828

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* **14:** 1188–1190

Decatur WA, Fournier MJ (2002) rRNA modifications and ribosome function. *Trends Biochem Sci* **27:** 344–351

Dennis PP, Omer A (2005) Small non-coding RNAs in Archaea. *Curr Opin Microbiol* **8:** 685–694

Fayet-Lebaron E, Atzorn V, Henry Y, Kiss T (2009) 18S rRNA processing requires base pairings of snR30 H/ACA snoRNA to eukaryote-specific 18S sequences. *EMBO J* **28:** 1260–1270

Gagnon KT, Zhang X, Qu G, Biswas S, Suryadi J, Brown II BA, Maxwell ES (2010) Signature amino acids enable the archaeal L7Ae box C/D RNP core protein to recognize and bind the K-loop RNA motif. *RNA* **16:** 79–90

Ghalei H, Hsiao HH, Urlaub H, Wahl MC, Watkins NJ (2010) A novel Nop5-sRNA interaction that is required for efficient archaeal box C/D sRNP formation. *RNA* **16:** 2341–2348

Granneman S, Kudla G, Petfalski E, Tollervey D (2009) Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc Natl Acad Sci USA* **106:** 9613–9618

Henras AK, Soudet J, Gerus M, Lebaron S, Caizergues-Ferrer M, Mougin A, Henry Y (2008) The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cell Mol Life Sci* **65:** 2334–2359

Hughes JM (1996) Functional base-pairing interaction between highly conserved elements of U3 small nucleolar RNA and the small ribosomal subunit RNA. *J Mol Biol* **259:** 645–654

Janke C, Magiera MM, Rathfelder N, Taxis C, Reber S, Maekawa H, Moreno-Borchart A, Doenges G, Schwob E, Schiebel E, Knop M (2004) A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* **21:** 947–962

Kiss T (2002) Small nucleolar RNAs: an abundant group of non-coding RNAs with diverse cellular functions. *Cell* **109:** 145–148

Kiss-Laszlo Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* **85:** 1077–1088

Kiss-Laszlo Z, Henry Y, Kiss T (1998) Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J* **17:** 797–807

Kuhn JF, Tran EJ, Maxwell ES (2002) Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Res* **30:** 931–941

Liu B, Fournier MJ (2004) Interference probing of rRNA with snoRNPs: a novel approach for functional mapping of RNA *in vivo*. *RNA* **10:** 1130–1141

Liu N, Xiao ZD, Yu CH, Shao P, Liang YT, Guan DG, Yang JH, Chen CL, Qu LH, Zhou H (2009) SnoRNAs from the filamentous fungus Neurospora crassa: structural, functional and evolutionary insights. *BMC Genomics* **10:** 515

Lowe TM, Eddy SR (1999) A computational screen for methylation guide snoRNAs in yeast. *Science* **283:** 1168–1171

Maden BE (2001) Mapping 2′-O-methyl groups in ribosomal RNA. *Methods* **25:** 374–382

Moore T, Zhang Y, Fenley MO, Li H (2004) Molecular basis of box C/D RNA-protein interactions; cocrystal structure of archaeal L7Ae and a box C/D RNA. *Structure* **12:** 807–818

Omer AD, Ziesche S, Ebhardt H, Dennis PP (2002) *In vitro* reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proc Natl Acad Sci USA* **99:** 5289–5294

Oruganti S, Zhang Y, Li H (2005) Structural comparison of yeast snoRNP and spliceosomal protein Snu13p with its homologs. *Biochem Biophys Res Commun* **333:** 550–554

Qu G, van Nues RW, Watkins NJ, Maxwell ES (2011) Spatial and functional coupling of Box C/D and C′/D′ RNPs are evolutionarily conserved features of the eukaryotic Box C/D snoRNP nucleotide modification complex. *Mol Cell Biol* **31:** 365–374

Qu LH, Henras A, Lu YJ, Zhou H, Zhou WX, Zhu YQ, Zhao J, Henry Y, Caizergues-Ferrer M, Bachellerie JP (1999) Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast. *Mol Cell Biol* **19:** 1144–1158

Reichow SL, Hamma T, Ferre-D'Amare AR, Varani G (2007) The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Res* **35:** 1452–1464

Suryadi J, Tran EJ, Maxwell ES, Brown II BA (2005) The crystal structure of the Methanocaldococcus jannaschii multifunctional L7Ae RNA-binding protein reveals an induced-fit interaction with the box C/D RNAs. *Biochemistry* **44:** 9657–9672

Szewczak LB, DeGregorio SJ, Strobel SA, Steitz JA (2002) Exclusive interaction of the 15.5kD protein with the terminal box C/D motif of a methylation guide snoRNP. *Chem Biol* **9:** 1095–1107

Tran EJ, Zhang X, Maxwell ES (2003) Efficient RNA 2′-O-methylation requires juxtaposed and symmetrically assembled box C/D and C′/D′ RNPs. *EMBO J* **22:** 3930–3940

Watkins NJ, Dickmanns A, Luhrmann R (2002) Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol Cell Biol* **22:** 8342–8352

Watkins NJ, Segault V, Charpentier B, Nottrott S, Fabrizio P, Bachi A, Wilm M, Rosbash M, Branlant C, Luhrmann R (2000) A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* **103:** 457–466

Xia L, Watkins NJ, Maxwell ES (1997) Identification of specific nucleotide sequences and structural elements required for intronic U14 snoRNA processing. *RNA* **3:** 17–26

Xue S, Wang R, Yang F, Terns RM, Terns MP, Zhang X, Maxwell ES, Li H (2010) Structural basis for substrate placement by an archaeal box C/D ribonucleoprotein particle. *Mol Cell* **39:** 939–949

Ye K, Jia R, Lin J, Ju M, Peng J, Xu A, Zhang L (2009) Structural organization of box C/D RNA-guided RNA methyltransferase. *Proc Natl Acad Sci USA* **106:** 13808–13813

Yu YT, Shu MD, Steitz JA (1997) A new method for detecting sites of 2′-O-methylation in RNA molecules. *RNA* **3:** 324–331