Published online 17 November 2010 *Nucleic Acids Research, 2011, Vol. 39, No. 6* ***2445–2457***
doi:10.1093/nar/gkq1131

Structural insight into the functional mechanism of Nep1/Emg1 N1-specific pseudouridine methyltransferase in ribosome biogenesis

Seth R. Thomas[1], Christopher A. Keller[1], Agnieszka Szyk[1], Joe R. Cannon[1] and Nicole A. LaRonde-LeBlanc[1,2,*]

[1]Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Organization, University of Maryland, College Park, MD 20742 and [2]University of Maryland, Marlene and Stewart Greenebaum Cancer Center, Baltimore, MD 21201, USA

Received August 20, 2010; Revised October 1, 2010; Accepted October 21, 2010

ABSTRACT

Nucleolar Essential Protein 1 (Nep1) is required for small subunit (SSU) ribosomal RNA (rRNA) maturation and is mutated in Bowen–Conradi Syndrome. Although yeast (*Saccharomyces cerevisiae*) Nep1 interacts with a consensus sequence found in three regions of SSU rRNA, the molecular details of the interaction are unknown. Nep1 is a SPOUT RNA methyltransferase, and can catalyze methylation at the N1 of pseudouridine. Nep1 is also involved in assembly of Rps19, an SSU ribosomal protein. Mutations in Nep1 that result in decreased methyl donor binding do not result in lethality, suggesting that enzymatic activity may not be required for function, and RNA binding may play a more important role. To study these interactions, the crystal structures of the scNep1 dimer and its complexes with RNA were determined. The results demonstrate that Nep1 recognizes its RNA site via base-specific interactions and stabilizes a stem-loop in the bound RNA. Furthermore, the RNA structure observed contradicts the predicted structures of the Nep1-binding sites within mature rRNA, suggesting that the Nep1 changes rRNA structure upon binding. Finally, a uridine base is bound in the active site of Nep1, positioned for a methyltransfer at the C5 position, supporting its role as an N1-specific pseudouridine methyltransferase.

INTRODUCTION

Ribosome biogenesis is a complex stepwise process that begins with the transcription of ribosomal RNA (rRNA), and continues with a coordinated processing pathway by which the rRNA is processed and ribosomal proteins are assembled [reviewed in (1–3)]. In *Saccharomyces cerevisiae* (yeast), a single 35S rRNA transcript is cleaved at several sites, and then further processed by exo- and endonucleases to produce the mature 18S, 25S and 5.8S rRNAs (2). Ribosomal RNA processing and ribosome assembly require several non-ribosomal protein factors, as well as small nucleolar RNA (snoRNA) molecules (2,4). These factors not only guide cleavage steps, but also carry out site-specific rRNA modifications that include methylation and pseudouridylation (2).

Nucleolar essential protein 1 (Nep1) is a highly conserved protein required for ribosome biogenesis and found in organisms from archaea to humans. Nep1 was first identified in fission yeast (*Schizosaccharomyces pombe*) as a high-copy number suppressor of a mating deficiency caused by a mutation in the effector domain of Ras1 and thus named Mra1 (multicopy suppressor of Ras 1) (5). It has also been designated the name Emg1 for essential for mitotic growth 1 (6). In yeast, Nep1 is localized to the nucleolus, interacts with a known nucleolar protein, Nop14, and is required for 18S rRNA maturation (6,7). To release the yeast 18S SSU rRNA, cleavage must occur at four sites, designated sites A_0, A_1, A_2 and D (1,2). Cleavage at site A_0 and A_1 truncates the 5′-end of the rRNA transcript, while cleavage at A_2 results in separation of a 20S pre-rRNA species. Final

*To whom correspondence should be addressed. Tel: +3014050462; Fax: +3013140386; Email: nlaronde@umd.edu
Present address:
Agnieszka Szyk, Cell Biology and Biophysics Unit, NINDS Porter Neuroscience Research Center Building 35, Room 3B-203, 35 Convent Drive, MSC 3700, Bethesda, MD 20892-3700, USA.

cleavage at site D produces the mature 18S SSU rRNA. Depletion of yeast Nep1 (scNep1) results in loss of cleavage at site A_2, which results in an accumulation of a 21S SSU rRNA species (7). This is a result of cleavage at site A_3, which normally happens after site A_2 cleavage in the maturation of the 5′-end of the 5.8S rRNA species (1). Nep1 has also been identified as the human gene mutated in Bowen–Conradi syndrome, a lethal genetic disease that results in low birth weight, small head, joint deformities and failure to thrive (8–10). The mutation results in a switch of D86 to glycine in the human Nep1 protein (8).

Prior structural work on Nep1 includes the X-ray crystal structures and backbone nuclear magnetic resonance (NMR) assignments of the archaeal Nep1 dimer from *Methanocaldococcus jannaschii* and a monomeric structure of scNep1 (11–13). Those structural analyses showed that Nep1 is structurally homologous to SPOUT RNA methyltransferases, provided a mutational analysis of Nep1 and allowed NMR mapping of some of the residues that interact with RNA (11,13,14). SPOUT methyltransferases are a group of proteins with an S-adenosylmethionine (SAM)-binding knotted methyltransferase fold originally defined to contain the SpoU and TrmD methyltransferases (15). This family now includes a large number of methyltransferases containing similar knotted folds (16). The members of the family each contain a SPOUT methyltransferase domain and may contain additional inserted domains required for selectivity of RNA substrates. Thus far, SPOUT methyltransferases have been implicated in the methylation of ribose 2′-OH (2′-O-ribose), guanine N1 (m1G), uridine C3 (m3U) and N3 of pseudouridine (m3Ψ) (16,17). Within the family, Nep1 proteins are most closely related to the TrmD and RsmE subfamilies, which catalyze m1G and m3U modifications, respectively (16). All SPOUT methyltransferases characterized thus far are dimeric, as is Nep1.

Nep1 plays a yet unclear role in the loading of Rps19 into the SSU. Rps19 is an SSU ribosomal protein found only in archaeal and eukaryotic ribosomes and is also known as a protein mutated in roughly 25% of cases of Diamond–Blackfan Anemia, a rare congenital disease characterized by defects in red blood cell development and predisposition to cancers (18,19). Rps19 is an essential protein, and its depletion also results in the loss of cleavage at site A2 (11,19–22). In yeast studies performed by Buchhaupt *et al*. (23) overexpression of Rps19 partially suppresses the phenotype observed with loss of Nep1 function. Surprisingly, this suppression is enhanced by deleting snr57, the C/D box snoRNA that guides 2′-O-ribose methylation at 18S G1572 (23,24). In fact, deletion of Nep1 is not lethal in an snR57 deletion mutant (23). Taken together, it appears that Nep1 aids in Rps19 loading and is essential in the presence of snr57. This function is either through direct interaction with Rps19, or alteration of the rRNA structure (23). Since the defects in rRNA processing observed in Nep1 and Rps19 depletion are identical, it is likely that some of the ribosome biogenesis defect caused by loss of Nep1 is a consequence of improper Rps19 assembly.

RNA-binding screens have indicated that Nep1 binds to the consensus sequence C/UUCAAC (23). This sequence is found in three regions of 18S rRNA in yeast including the stem loop of helix 31, bases 1188–1193, and part of the stem of helix 42, bases 1565–1570. In the RNA three-hybrid screen that identified this consensus sequence, all the binding sites were located in the unpaired region of a stem loop structure predicted by RNA secondary structure predictions (23). Both these binding sites are near helix 41, where Rps19 was shown to bind using cryo-EM reconstructions (23,25). In addition, one of the sites, between bases 1565 and 1570, overlaps with the predicted binding site of snR57 (bases 1569–1583).

Recent work has identified Nep1 as a pseudouridine N1-methyltransferase (14). Wurm *et al*. (14) showed that both human and *M. jannaschii* Nep1 are capable of catalyzing methylation on pseudouridine-containing RNA sequences, especially C/UΨCAAC. This sequence is the Nep1 site found in the stem loop of helix 31 of yeast 18S rRNA mentioned above (14). This pseuoduridine, located at position 1189, is hypermodified in yeast, with a 3-amino-3-carboxypropyl group on the N3 position and a methyl group on the N1 position (m1acp3Ψ) (26). Nep1 is the first identified example of an N1-specific pseudouridine methyltransferase (14).

Given that it plays an essential, though yet unclear, role in maturation of the 18S rRNA, we have determined the X-ray crystal structures of dimeric Nep1 from scNep1 and *Archaeoglobis fulgidus* (afNep1) at 1.85 and 1.45 Å, respectively, bound to *S*-adenosyl homocysteine (SAH), and scNep1 bound to one molecule of cognate RNA at 1.90 Å, and two molecules of cognate RNA at 3.00 Å. The structures of Nep1 with RNA have revealed recognition of a base flipped uridine in the active site of the enzyme, and a hairpin loop in the RNA-binding site. The RNA bound in the active site contains the consensus sequence UUCAA C, and does not contain pseudouridine, and therefore is a substrate analog. Consistent with the role as a pseudouridine N1-methyltransferase, structural evidence indicates that the substrate analog is positioned for acceptance of a methylation at the C5 position of the uridine moiety, which is the equivalent of the N1 position of pseudouridine. This is the first structure of a SPOUT methyltransferase bound to cognate RNA. In addition, Nep1 is the first example of a SPOUT-class methyltransferase that functions as an N1-pseudouridine-specific methyltransferase, and therefore these results provide a structural analysis of a unique new group of enzymes.

MATERIALS AND METHODS

Protein expression and purification

The full-length Nep1 gene was PCR amplified from *A. fulgidus* and *S. cerevisiae* genomic DNA (American Type Culture Collection) and subcloned into a plasmid vector containing an N-terminal 6× histidine tag followed by a Tobacco Etch Virus (TEV) protease

cleavage site (Protein Expression Laboratory, SAIC, Frederick, MD, USA). These constructs were transformed into *Escherichia coli* Rosetta™-DE3-pLysS cells (Novagen). Expression cultures were grown at 37°C and induced with 1 mM IPTG at OD_{600} of 0.6 for 4 h for afNep1. For scNep1, cultures were grown at 37°C until OD_{600} of 0.6, then cooled to 20°C and induced with 1 mM IPTG for overnight expression. After expression, the cells were harvested by centrifugation. For afNep1, cells were resuspended in 50 ml of 50 mM Tris pH 8.0, 300 mM NaCl, 0.2% β-mercaptoethanol, 0.2× Bugbuster™ (Novagen) and 0.1 mg/ml DNase1 (Roche) per liter of expression culture and stirred at room temperature. After 30 min, the lysate was transferred to centrifuge tubes and placed in a 75°C water bath for 15 min to denature the *E. coli* proteins then immediately centrifuged for 20 min at 18 000 rpm in a ultracentrifuge to remove insoluble material. For scNep1, cells were resuspended in 50 ml of 25 mM Tris pH 8.0, 200 mM NaCl, 0.2% β-mercaptoethanol, 0.2× Bugbuster™ and 0.1 mg/ml DNase1 (Roche) per liter of expression culture and stirred at 4°C. After 45 min, the lysate was transferred to centrifuge tubes and centrifuged for 20 min at 18 000 rpm. The supernatants were passed through a 0.22 μM filter and loaded onto a 5 ml HisTrap™ HP column (GE Healthcare) equilibrated in buffer containing 50 mM Tris pH 8.0, 300 mM NaCl and 0.2% β-mercaptoethanol for afNep1 and 25 mM Tris pH 8.0, 200 mM NaCl and 0.2% β-mercaptoethanol for scNep1. The bound Nep1 was washed in 100 mM imidazole in equilibration buffer and then eluted with a gradient from 100 mM to 1 M imidazole in equilibration buffer in 20 column volumes. The fractions containing Nep1 were pooled and dialysed overnight into 25 mM Tris pH 8.0, 300 mM NaCl, 1 mM EDTA and 0.2% β-mercaptoethanol for afNep1 and 25 mM Tris pH 8.0, 200 mM NaCl, 1 mM EDTA and 0.2% β-mercaptoethanol for scNep1. In the case of afNep1, attempts made to remove the tags by TEV protease cleavage were unsuccessful. The tag was cleaved in scNep1 and the resulting protein contained a single extra glycine residue at the N-terminus. The cleaved scNep1 protein was passed over a 5 ml HisTrap HP column (GE Healthcare) a second time to remove uncleaved protein and TEV protease. The proteins were concentrated to 5–10 mg/ml and further purified to >99% by size exclusion chromatography using a Superdex 75™ column (GE Healthcare) equilibrated in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM EDTA and 0.2% β-mercaptoethanol for afNep1, and a Superdex 200™ column (GE Healthcare) equilibrated in 20 mM Tris pH 8.0, 200 mM NaCl, 1 mM EDTA and 0.2% β-mercaptoethanol for scNep1. The protein was concentrated again to 8 mg/ml for afNep1 and 20 mg/ml for scNep1. This preparation was stored at 4°C until crystallization screening. Se-Met afNep1 was expressed in minimal media with all amino acids supplemented except for methionine, which was replaced by seleno-methionine (Se-Met). Reducing agent (0.2% β-mercaptoethanol) was added to Se-Met protein immediately prior to crystallization screening.

Nep1–RNA complex crystallization

ScNep1 was mixed in a 1:1 molar ratio with RNA oligos of varying lengths and sequences purchased HPLC-purified from IDT and resuspended in 10 mM Tris pH 8.0. The mixtures were analyzed by gel electrophoresis on a 6% TBE-polyacrylamide gel and visualized by UV after ethidium bromide staining to determine stoichiometric complex formation. A 14-base RNA oligo containing the sequence 5′-GGGCUUCAACGCCC-3′ was the most successful at producing stoichiometric complexes.

Initial crystallization conditions were obtained through utilization of several sparse matrix screens (Emerald Biostructures Inc., Nextal, Hampton Research) with the sitting drop vapor diffusion method. Crystals of the afNep1 protein were obtained from drops containing 3–4 M sodium chloride, 100 mM HEPES buffer pH 7.0–8.0 placed over 1 ml reservoirs. Crystals of free scNep1 were obtained from several conditions, including 50 mM BIS–TRIS, pH 6.5, 50 mM ammonium sulfate, 30% pentaerythritol ethoxylate (15/4 EO/OH) and 20% glycerol, used for data collection. The Nep1/1RNA complex crystallized in 100 mM HEPES, pH7.0 and 10% PEG (polyethylene glycol) 6000 and crystals of the Nep1/2RNA complex were obtained in 50 mM MES, pH 6.0, 15 mM magnesium sulfate and 7% PEG 4000. Diffraction quality crystals were obtained after 4-6 days at 20°C.

Data collection and processing

Crystals were flash frozen in mother liquor containing 20% glycerol for all crystals. Diffraction data for the native and Se-Met were collected at 100 K at the NE-CAT beamline at the Advanced Photon Source (APS), Argonne, Illinois. Data from the Se-Met crystals were collected at one wavelength (peak). All data were integrated and merged using HKL2000. Table 1 contains details regarding data statistics for all data sets.

Structure determination and refinement

For afNep1, structures were solved using single wavelength anomalous diffraction (SAD) data with Se-MET protein crystals, collected at 2.0 Å. The AutoSharp program suite was used to obtain the phases using the SAD method, apply solvent flattening and density modification to the initial electron density map, and perform automatic model building with wARP (27). The complete model was finalized by rebuilding in COOT (28) and refinement with REFMAC5 (29). The model was then used to phase the native data set using MOLREP as part of CCP4i (30–32). The scNep1 structures were solved by molecular replacement in MOLREP using PDB ID 2V3J (13). All models were subjected to several rounds of building in COOT and refinement using REFMAC5 or PHENIX (28,29,33,34). TLS groups were determined using the TLSMD server (35). R_{free} was monitored by using 5% of the reflections as a test set in each case. Refinement statistics are provided in Table 1. The coordinates and structure factors have been deposited in the

Table 1. Data collection and refinement statistics

PDB code	afNep1 Se-Met SAD	afNep1 3O7B	scNep1 3OII	scNep1 + 1RNA 3OIJ	scNep1 + 2RNA 3OIN
Data collection					
Space group	$P3_12_1$	$P3_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions					
a, *b*, *c* (Å)	70.3, 70.3, 95.3	70.1, 70.1, 95.2	46.0, 84.7, 112.5	44.8, 88.7, 115.8	83.4, 91.9, 96.9
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 90.0	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Molecules/Asym. unit	1 (1 Protein)	1 (1 Protein)	2 (2 Protein)	3 (2 Protein: 1 RNA)	4 (2 Protein: 2 RNA)
Wavelength (Å)	0.97919	0.99998	1.00931	1.00931	1.00931
Resolution (Å)	30–2.0	30–1.45	30–1.8	30–1.9	30–2.9
R_{sym} (last shell)	0.041 (0.142)	0.075 (0.517)	0.061 (0.403)	0.066 (0.392)	0.078 (0.361)
$I/\sigma I$	37.3 (16.3)	32.9 (5.8)	22.6 (1.7)	25.5 (2.3)	21.2 (2.0)
Completeness (%)	99.6 (99.9)	99.0 (100.0)	94.4 (63.1)	96.8 (80.2)	97.0 (76.5)
Redundancy	7.6 (7.6)	14.3 (10.0)	6.9 (3.2)	7.1 (4.2)	6.3 (2.9)
Refinement					
Resolution (Å)		29–1.45	20–1.85	28–1.9	30–3.0
R_{work}/R_{free} (%)		19.8/21.1	17.1/21.6	19.2/23.0	17.4/25.2
No. of protein residues monomers A/B		216	216/208	217/217	211/217
No. of non-protein molecules					
Nucleic Acid (bases)		0	0	14	28
Ligand (SAH)		1	2	1	2
Solvent		242	224	258	64
Mean B-factors $(Å)^2$ (TLS refined)		23.3	35.0	48.1	68.9
RMS deviations					
Bond lengths (Å)		0.010	0.021	0.008	0.008
Bond angles (°)		1.281	1.881	1.069	1.257
Ramachandran plot					
Most favorable (%)		94.2	93.4	92.8	87.5
Additionally allowed (%)		5.8	6.3	7.0	12.5
Generously allowed (%)		0.0	0.3	0.3	0.0
Disallowed (%)		0.0	0.0	0.0	0.0

Protein Data Bank (accession codes 3O7B for afNep1, 3OII, 3OIN and 3OIJ for scNep1 free dimer, 1RNA complex and 2RNA complex, respectively).

RESULTS

Structures and comparison of the dimeric Nep1 from *S. cerevisiae* (yeast) and *A. fulgidus*

The overall structure of the Nep1 protein, shown in Figure 1, consists of a central β-sheet surrounded by α-helices on either side, characteristic of a Rossmann fold (11,13). The central β-sheet is comprised of six β-strands, five of which are parallel to each other and one anti-parallel (β3), with an additional two loosely ordered strands completing the sheet (β4, β5). The order in which sequential β-strands form the β-sheet is β5-β4-β3-β2-β1-β7-β6-β8. The helices within the Rossmann domain are located between β1 and β2 (αA, αAb, αd′), β3 and β4 (αc′, Hc) and after β8 (αE). Two additional helices, Hb′ and αa′ are present in a small Nep1 specific 40 amino-acid insertion between β1 and α2 (residues 48–88). This insertion and the orientation of αa′ is characteristic of Nep1 SPOUT methyltransferases, and also exhibits the most difference between the archaeal and yeast structures. This domain is moderately conserved based on sequence comparison with eukaryotic Nep1 sequences (Figure1A). Electron density was observed for all the residues except 1–27 for both monomers, 56–64 for monomer A and 54–65 and 153–157 for monomer B, of scNep1, and for all but the last three for afNep1. In both models, density was observed in the SAM-binding pocket into which SAH was modeled.

Dimerization interface of scNep1

Although the structure of Nep1 from yeast has been previously reported, the deposited structure (PDB ID 2V3J) was monomeric (13). In the crystal structure of scNep1 bound to SAH presented here, the molecule is dimeric. In both archaea and yeast, a large interacting surface is observed between Nep1 molecules that buries surface area in excess of 845 $Å^2$ for afNep1 and around 1475 $Å^2$ for scNep1. The dimer interface is formed primarily between residues on helices αA and αE and the N-terminal end of αc′. The interactions form in a 2-fold symmetric interface that can be seen in Figure 1B and C (Supplementary Figure S1 for afNep1). The residues at the dimer interface of scNep1 are highly conserved among eukaryotes, but less with archaeal organisms (Supplementary Figures S2 and S3). The scNep1 interface consists of a mixture of electrostatic interactions and hydrogen bonds on the edges and a few hydrophobic interactions at the center, also several water molecules were observed in the charged regions of the interface. This shows a less tight dimer interface than that observed for the afNep1 interface, which contains a stabilizing disulfide bond, commonly used to stabilize protein structure in thermophiles (Supplementary Figure S1) (36). Interestingly, in scNep1, the side chain of the C-terminal Leu 252, forms hydrophobic packing interactions at the interface and the carboxyl group at the end of the chain forms a hydrogen bond with

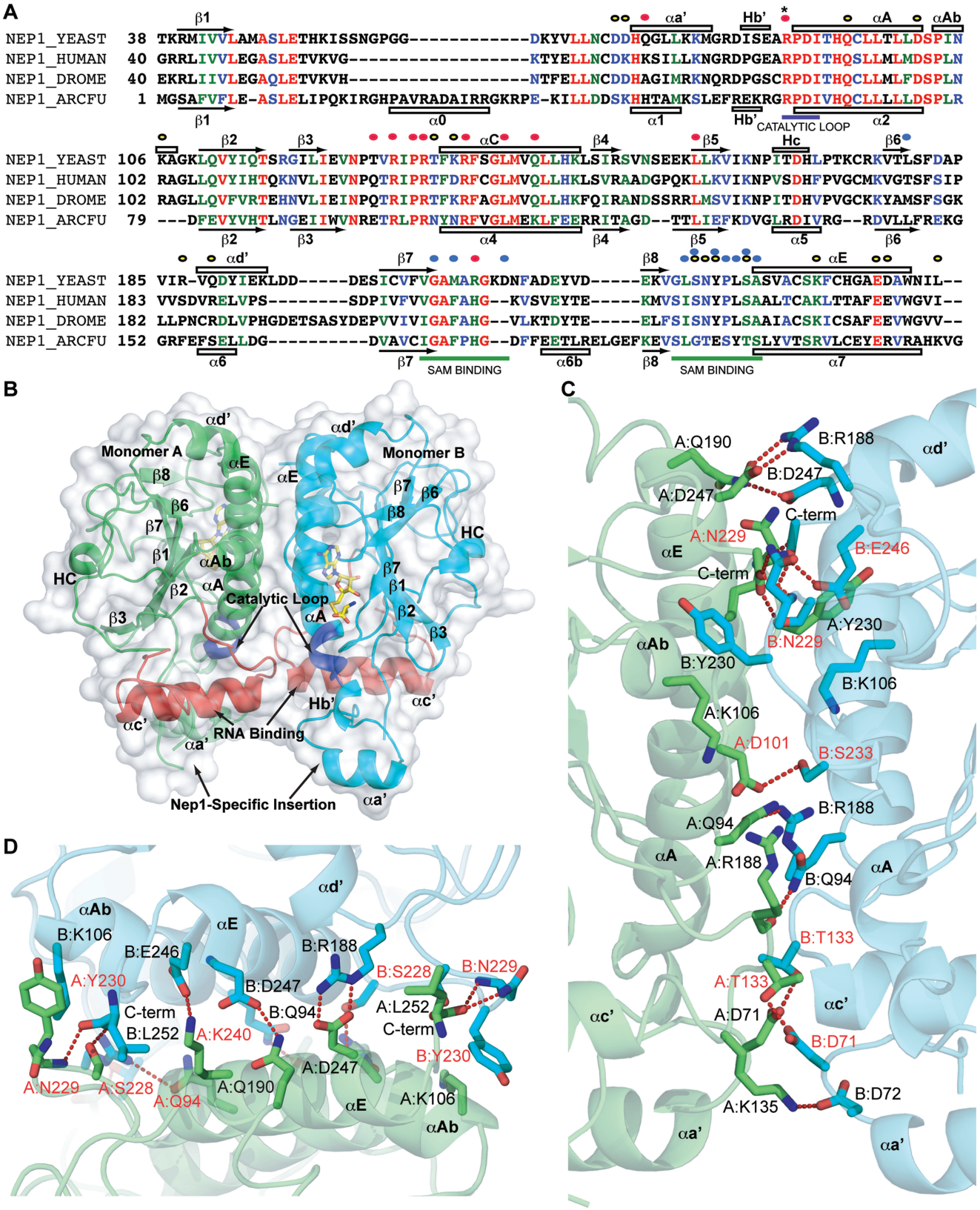

Figure 1. Overall structure of Nep1 and dimer interface. (**A**) Sequence alignment from of Nep1 from *S. cerevisiae* (NEP1_YEAST), *Homo sapiens* (NEP1_HUMAN), *Drosophila melanogaster* (NEP1_DROME) and *A. fulgidus* (NEP1_ARCFU). Yellow, red and blue dots indicate residues involved in dimerization, RNA binding and SAM binding respectively. Secondary structure for yeast (top) and *A. fulgidus* (bottom) shown as open bars for α-helices and arrows for β-strands and numbered according to Leulliot *et al.* for yeast, and Taylor *et al.* for *A. fulgidus* (11,13). (**B**) The scNep1 dimer. RNA binding and catalytic regions of the molecules are colored red and dark blue respectively. The molecules of SAH are shown in sticks. Secondary structure elements are labeled according to Leulliot *et al.* (13). (**C**) Front view of the dimerization interface of Nep1. Residues are labeled with chain names and residue numbers. (**D**) Top view of the interface showing the interaction of the C-terminus in the interface. Residues that are identical in eukaryotic sequences are labeled in red text. All 3D structure figures were generated in PyMOL (www.pymol.org).

the side chain of Ser 228 and Asn 229 and the backbone amide nitrogen of Asn 229 of the other monomer (Figure1C and D). Therefore, the C-terminus participates in dimer formation. In the previously reported structure of scNep1, the construct used contained a C-terminal histidine tag that may have disrupted this interaction and weakened the dimer to some extent, resulting in monomers in the crystals (13). In the absence of RNA, one monomer of scNep1 aligns with the other with an RMSD of 0.34 Å, and small conformational differences are observed between loop regions of the two monomers.

Interaction of scNep1 with RNA

Using the binding sequences for scNep1 obtained through RNA three-hybrid screening by Buchhaupt *et al.* (23), we designed a series of oligonucleotides and tested their binding to scNep1 in stoichiometric complexes. We determined that an ideal complex was obtained when scNep1 and a 14-base RNA fragment with the sequence $G_1G_2G_3C_4U_5U_6C_7A_8A_9C_{10}G_{11}C_{12}C_{13}C_{14}$ were mixed in a 1:1 ratio. We obtained crystals of Nep1 bound to this 14-base RNA fragment in two crystal forms (Figure 2A and B). In one crystal form, a dimer of scNep1 is bound to one RNA molecule (Figure 2A). In the second crystal form, the dimer binds two separate RNA molecules (Figure 2B). In both cases, the RNA forms a stem-loop structure in which eight bases participate in canonical base pairs in the stem, and six bases in the loop (UUCAAC) have base-specific contacts with the protein. Electron density was not observed for residues 1–27 and 56–64 of both monomers in the RNA complexes, and residues 76–82 of monomer A of the two-RNA complex. There are no significant differences between the RNA molecules bound in the two complexes. The six bases in the form

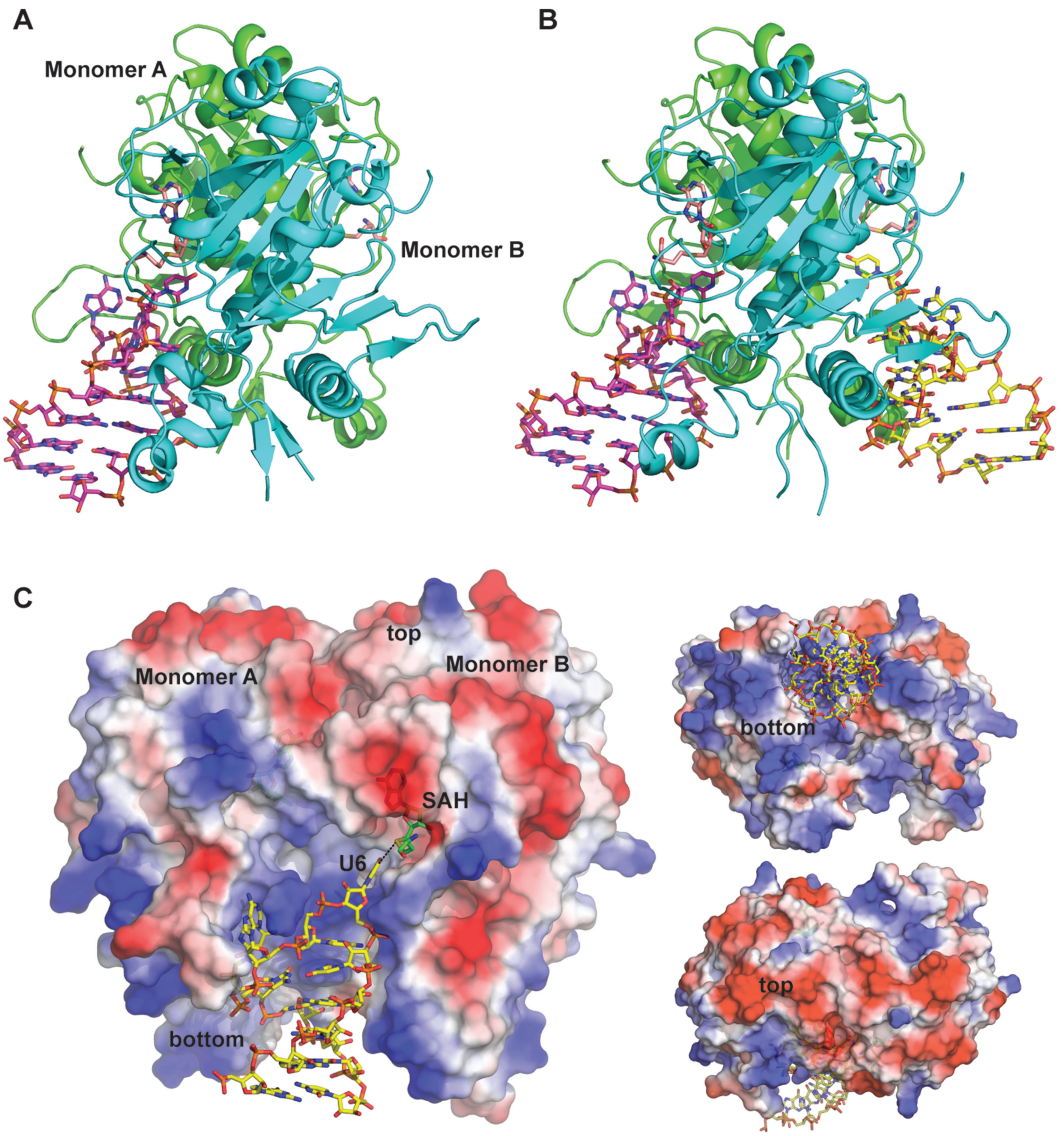

Figure 2. Overall structure of the Nep1/RNA complexes. Structures of the (**A**) one RNA and (**B**) two RNA complexes with SAH (pink) and RNA (magenta and yellow) shown in sticks. (**C**) Electrostatic surface of the Nep1 complex showing front, top and bottom views. The surface is colored red to blue, negative to positive charge (calculated using vacuum electrostatics in PyMOL).

the basis for specific recognition by the protein. In the comparison of the sequences of RNA found to bind Nep1 by RNA three-hybrid screens, all the sequences contained the consensus sequence C/UUCAAC, and the most tightly binding were those sequences that contained the consensus sequence within the loop of a predicted stem-loop (23). In the crystal containing the one RNA per dimer complex, the empty RNA-binding site is involved in four hydrogen bonds with the RNA backbone from a symmetry-related complex.

In both complexes, the RNA binds on highly positively charged surfaces of the dimer as predicted in previous reports (11,13). However, the protein contains significant positively charged surface that is not contacted by RNA in these structures (Figure 2C). This positively charged surface is conserved in the afNep1 as well (Supplementary Figure S3). This is consistent with the idea that Nep1 binds to RNA on the pre-ribosome, which makes it likely that it will contact significantly more RNA than is observed with these minimal constructs. The regions that are not contacted by RNA in these structures contain many conserved residues as well, indicating that this surface is important for Nep1 function (Supplementary Figure S2).

ScNep1 provides several key interactions that confer the observed specificity for a UUCAAC sequence (Figure 3,

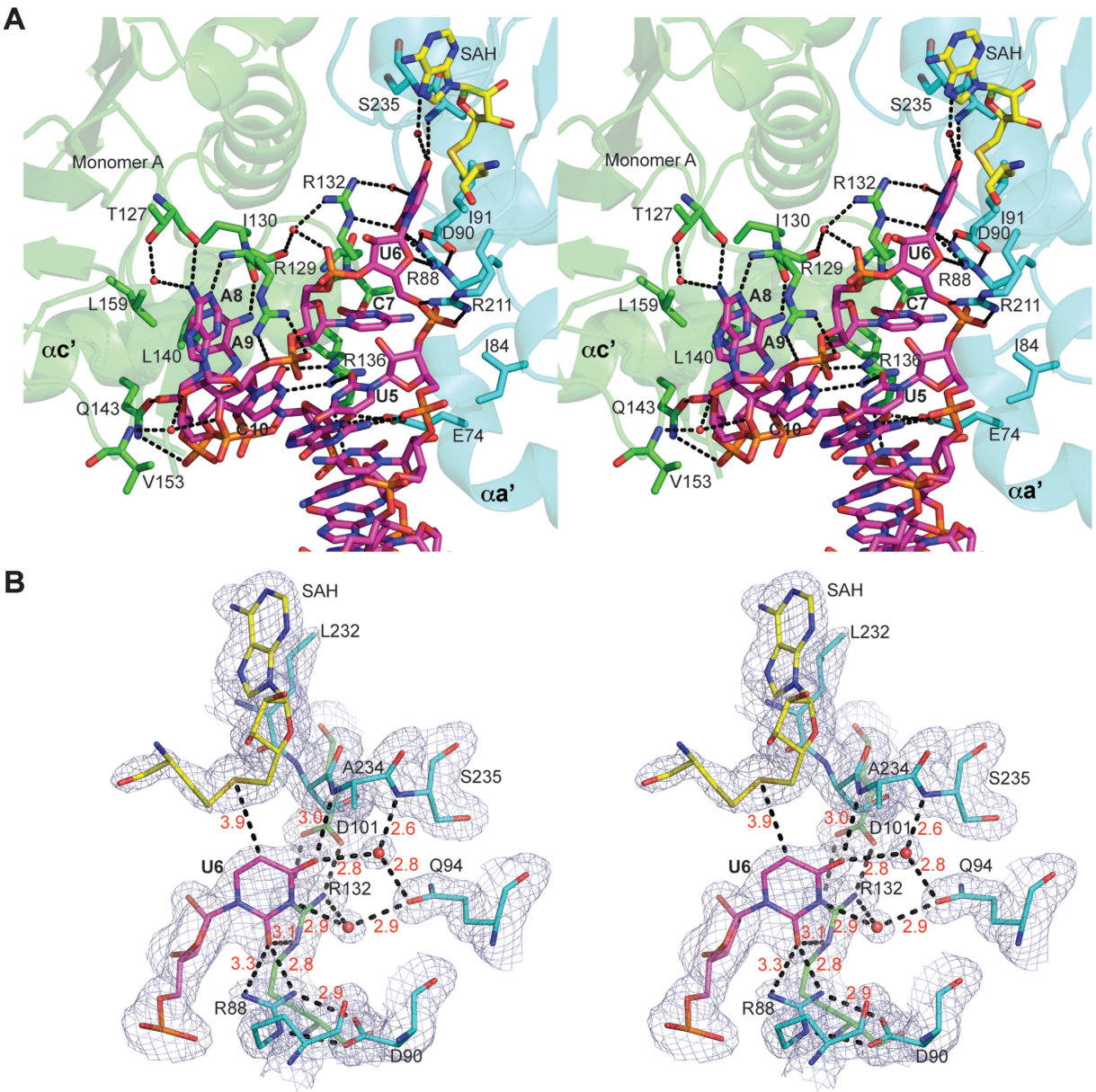

Figure 3. Nep1-RNA interactions. (**A**) Stereo figures of RNA bound in the cleft between the two monomers. Residues are colored according to monomer from which they originate (green-monomer A; cyan-monomer B). Dash lines represent connections between atoms within hydrogen bonding distance. RNA (yellow) truncated to show only the bases that contact the protein for simplicity. (**B**) Stereo view of a close-up of the uridine base interactions with 2Fo–Fc map contoured at 1σ. Protein residues are colored as in (A). U6 is shown in magenta. Distances between atoms are indicated by red text.

schematic in Supplementary Figure S4). The first base of the consensus sequence, U_5, is contacted through base-specific contacts through bidentate hydrogen bond interactions between the NH1 and NH2 of conserved Arg 136 of monomer A (A:Arg 136 NH1 and NH2) and the C2 carbonyl oxygen (O2) of the uridine base (Figure 3A and Supplementary Figure S4). There are no contacts between the protein and C4 carbonyl oxygen (O4), which explains the observation that a cytosine, which differs from uracil solely by an amine group at C4, can substitute at this position (23).

The next uracil in the consensus sequence, U_6, is the site of modification by the enzyme. It is flipped out from the loop and bound in a pocket between the two monomers, near the bound SAH molecule, in the enzyme active site (Figure 3A and B). It is recognized by a direct hydrogen bond with A:Arg 132 NE (Nε of arginine 132 of monomer A) at the U_6 O2 (C2 carbonyl oxygen of U_6), and water-mediated hydrogen bond interactions between the U_6 O4 and the backbone amide of B: Ser 233 and the U_6 O4 and N3 and OE (side chain carbonyl oxygen) of conserved B: Gln 94. Arg 132 is held firmly in position by hydrogen bonding between Arg 132 NH1 and NH2 and conserved A: Asp 101 OD1 and OD2. In addition, hydrogen bonds are observed between the U_6 O2 and NH1 and NH2 of B:Arg 88 and between the ribose O4 and the NH1 of B:Arg 88. This would hold U_6 firmly in position, and may serve as a means to stabilize negative charge on the base during methyl transfer. Arg 88, in turn, is held in place by interactions between its NH2 and NE and the OD1 and OD2 of conserved A:Asp 90. Asp 90 is the equivalent of the human Asp 86 that is mutated to glycine in Bowen–Conradi syndrome. That mutation would have the effect of destabilizing the interaction with the U_6 substrate both through disruption of the interaction with the Arg 88 and hydrophobic contacts with Ile 91 through repositioning. Finally C5 of the uridine base is positioned a mere 3.9 Å away from the sulfur atom of SAH, in perfect position for a methyl transfer to the C5 position of the uracil (Figure 3). If a pseudouridine were in its place, then N1, which is an equivalent position to C5, would be the target of methylation.

In the remaining consensus sequence, C_7 is recognized through backbone interactions between the carbonyl oxygen, O2, and the backbone amides on either side of A: Arg 132, and a base stacking interaction with A: Arg 136. A_8 is sandwiched between a stacking interaction with conserved A: Arg 129 and van der Waal's contact with conserved A: Ile 159, and donates a hydrogen bond from N6 to the backbone carbonyl oxygen of A: Thr 127. A_8 N6 is also involved in a water-mediated hydrogen bond interaction with the side chain OH of Thr 127 and A_8 N1 is in hydrogen bonding distance of the backbone amide of A: Arg 129. A_9 is recognized through hydrogen bonds between the backbone carbonyl of A: Arg 129 and the N6, a van der Waal's interaction between the side chain of A: Val 128 and the C2, and a stacking hydrophobic interaction with conserved A: Leu 140. Although Val 128 is not highly conserved, it is Val, Thr or Cys in other sequences, which would be able to maintain the interaction with A_9 C2. Finally C_{10} is specified by two hydrogen bonds with conserved A: Arg 136, one between the Arg NH2 and the C_{10} N3, and the other between the Arg NE and C_{10} O2. There is also an interaction between the C_{10} ribose O2′ and the backbone carbonyl of A: Arg136. Arg 136 contacts C_{10} via three hydrogen bonds, U_5 via two hydrogen bonds, and stacks in between A_9 and C_7. Consequently, Arg 136 is invariant in all sequences. In a functional analysis by Taylor *et al.* (11) mutations of Arg 136, as well as Arg 88, Arg 129 and Arg 132 to Ala all result in diminished RNA-binding activity as measured by yeast three-hybrid assays with consensus sequences.

Non-specific interactions

Of the eight bases that form the stem in canonical Watson–Crick base pairs, only one, G_{11}, is contacted at the base by a protein residue. B: Glu 74 accepts a hydrogen bond from G_{11} N2, and contacts the G_{11} ribose O2′ via a water-mediated interaction (Figure 3 and Supplementary Figure S4). Other backbone interactions between the protein and the RNA include contacts between A: Gln 143 and C_{10} O2′ and phosphate between C_{10} and A_9, A: Arg 129 and C_{11} O2P, A: Arg 211 and the phosphate between U_5 and U_6. Of these, Arg 129 and Gln 143 are highly conserved, while Arg 211 is only weakly conserved (Figure 1A). Based on this data, we can conclude that very little specificity is determined in the stem-loop portion of the RNA.

Catalytic residues in the Nep1 active site

The methylation reaction that Nep1 catalyzes would require the deprotonation of, and transfer of the methyl group from SAM to, the N1 of the pseudouridine. Stabilization of a negative charge on the uridine ring would result in a decrease of the N1 pK_a, which was measured at 8.97 for free pseudouridine (37). The likely candidate for this stabilization would be invariant Arg 88, for which both primary amine nitrogens are within hydrogen bonding distance of U_6 O2, the C2 carbonyl oxygen (3.1 and 2.7 Å; Figure 3B and Supplementary Figure S5), and which is stabilized by bidentate hydrogen bonds with invariant Asp 90, which is mutated in Bowen–Conradi Syndrome (8). Mutation of Arg 88 to Asp resulted in a significant loss of Nep1 RNA-binding activity (13). Arg 88 and Asp 90 are located at the beginning of αA, and are surrounded by Pro 89 and Ile 91, which are also invariant. Ile 91 is in van der Waals contact distance of the uridine ring in the active site. This region is labeled as the 'Catalytic Loop' in Figure 1A and B. Arg 132 may also provide further stabilization, since it is also within hydrogen bonding distance (3.0 Å) of O2. These interactions would increase deprotonation at the N1 through localization of the negative charge to O2 where it would be stabilized by interaction or proton donation by Arg 88 (Supplementary Figure S5). This would in turn enable N1 to accept the methyl group from SAM.

Conformational changes in the scNep1 protein when bound to RNA

When a monomer of free Nep1 is compared to monomers from the RNA Nep1 complex, an RMSD of 0.76 Å is observed for 415 out of 426 residues. Most of the interaction between the RNA and the protein are mediated by the helix αc′ and adjacent loops on the monomer that does not provide the SAM to the active site (monomer A in Figures 1, 2 and 3). However, the most significant conformational changes are seen in the small Nep1 specific inserted domain on the SAM-binding side of the dimer interface (Figure 4A and B). This domain, consisting of a 3–10 helix (Hb′), small α-helix (αa′) and a β-hairpin, undergoes a shift in positioning towards the dimer interface when RNA is bound. In addition, there is alteration of overall secondary structure, as the structure changes to accommodate interactions with the RNA (Figure 4A and B). The domain, which is packed against αc', houses Arg 88, which may be important for catalysis, and Glu 74 that interacts with G11 (Figure 3A). The buried surface area in the RNA bound structure (1338 $Å^2$) is slightly decreased, due to an opening of the interface to accommodate the RNA, which buries an additional 721 $Å^2$ on the monomer containing the methyl donor and 351 $Å^2$ on the other monomer.

Two-RNA versus One-RNA

Both the 1:1 and 1:2 complexes for the complexes between Nep1 and RNA were obtained using the same preparation of Nep1 to RNA in a 1:1 ratio. The isolation of these two

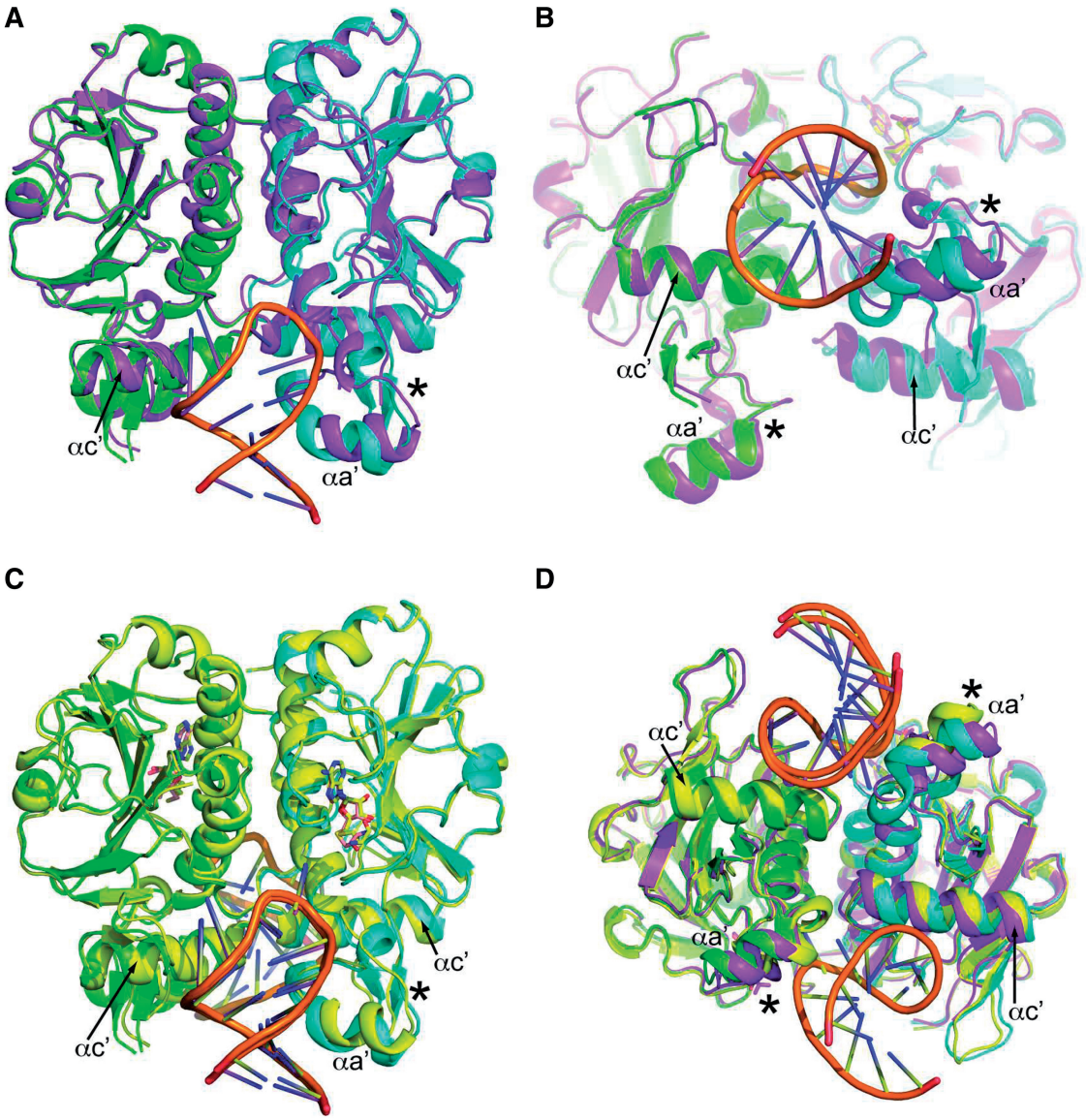

Figure 4. Conformational changes upon RNA binding. (**A**) Front and (**B**) Bottom views of the Nep1/1 RNA complex (green-monomer A, cyan-monomer B) superposed on the free dimer (magenta). (**C**) Front view of the alignment of the 1 RNA and 2 RNA (yellow) complexes. (**D**) Bottom view of both RNA complexes and the free dimer. Monomer B, shown here on the right, was used to align the monomers. Asterisk indicates the Nep1 specific domain where significant structural change occurs, and αc' indicates the RNA-binding helix that shifts relative to the aligned monomer.

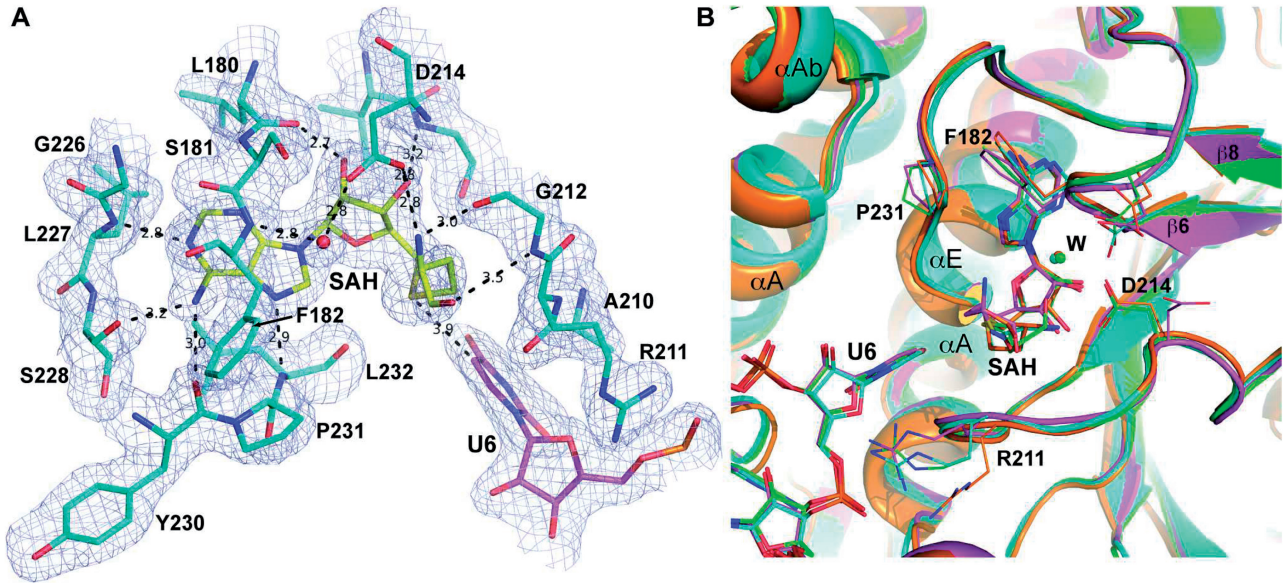

Figure 5. Comparisons of SAH ligand binding. (**A**) SAH bound in the SAM-binding pocket in the Nep1/1RNA complex, showing electron density 2*Fo*–*Fc* map contoured at 1σ. (**B**) Alignment of the SAM-binding pockets of Nep1/1RNA complex monomer B (cyan), Nep1/2RNA monomer B (green), Nep1/2RNA monomer A (magenta) and free Nep1 dimer (orange) showing side chains of residues that show the most conformational differences between the complexes. Spheres indicate bound waters, and are labeled 'W' in B.

complexes in the crystal structures suggests that both complexes exist in solution, which implies differences in the binding affinities between the two sites. Comparisons of the structures of these complexes reveal differences in the structures of the RNA-binding sites as a function of occupancy with RNA (Figure 4C and D). The RMSD for the alignment of the protein molecules in the complexes averages 0.85 Å for 421 residues of the protein chains, with the largest difference being between chain A of the 1:1 complex and chain B of the 2:1 complex. This number is larger than the RMSD for the comparison between the Nep1/1RNA complex and the free enzyme, indicating a larger conformational difference between the Nep1/1RNA and Nep1/2RNA complexes than between the free enzyme and the Nep1/1RNA complex. Upon closer inspection, we observed that the positioning of one monomer relative to the other is slightly different between the two complexes. In the Nep1/2RNA complex, the two monomers are slightly closer together near the RNA-binding site, indicating that RNA binding pulls the two monomers closer together. As a result, the SAH molecule is pulled closer to the RNA. The SAH molecule is contacted mostly by backbone interactions, but one residue, Asp 214, hydrogen bonds the ribosyl 3′OH via one of its side-chain carbonyl oxygens (Figure 5A). In the Nep1/2RNA complex, for one monomer site the Asp 214 side-chain is rotated away from the SAM-binding site and no longer contacts the bound SAM (Figure 5B and Supplementary Figure S5A). This results in a slightly altered positioning of the homocysteinyl moiety in the active site, with the sulfur atom positioned farther away (4.5 Å) from the uridine base (Figure 5B and Supplementary Figure S5A). In addition, the SAM-binding pocket on the side of the dimer not bound to RNA in the Nep1/1RNA complex is not visibly occupied by SAH. Based on these observations, and the presence of additional conserved positive charged surface (Figure 2C), it is possible that both RNA-binding sites would be occupied on rRNA, perhaps with varied binding affinity or specificity, and additional binding surface would be used to interact on the surface of the rRNA.

DISCUSSION

Prior to this work, the structures of the scNep1 in monomeric form, and the archaeal *M. jannaschii* Nep1 dimer, allowed for the classification of Nep1 as a SPOUT methyltransferase, characterization of the Nep1 specific structural elements and analysis of the SAM-binding pocket, and, in combination with mutagenesis and NMR mapping on the *M. jannaschii* Nep1, allowed for a prediction of the RNA-binding site (11–14). This report provides a view of dimeric scNep1 in apo form and in complex with RNA, allowing a first look at a SPOUT methyltransferase bound to RNA, a high resolution analysis of the interactions of the only known N1-specific pseudouridine methyltransferase with a substrate analog, and a picture of the RNA structure that Nep1 stabilizes upon binding. These data confirm that the RNA binds to the cleft between the two monomers, which agrees with previously reported NMR mapping data and mutational analysis that reported mutation of the *M. jannaschii* equivalents of R88, R129, R132 and R136 significantly reduced RNA-binding activity (11,14). In Nep1, the binding site accommodates a stem-loop structure, with the majority of binding observed between the RNA loop and the protein. Both the RNA three-hybrid screening of scNep1 and the specificity determination on *M. jannaschii* Nep1 indicate that Nep1 binds more tightly to a predicted stem-loop

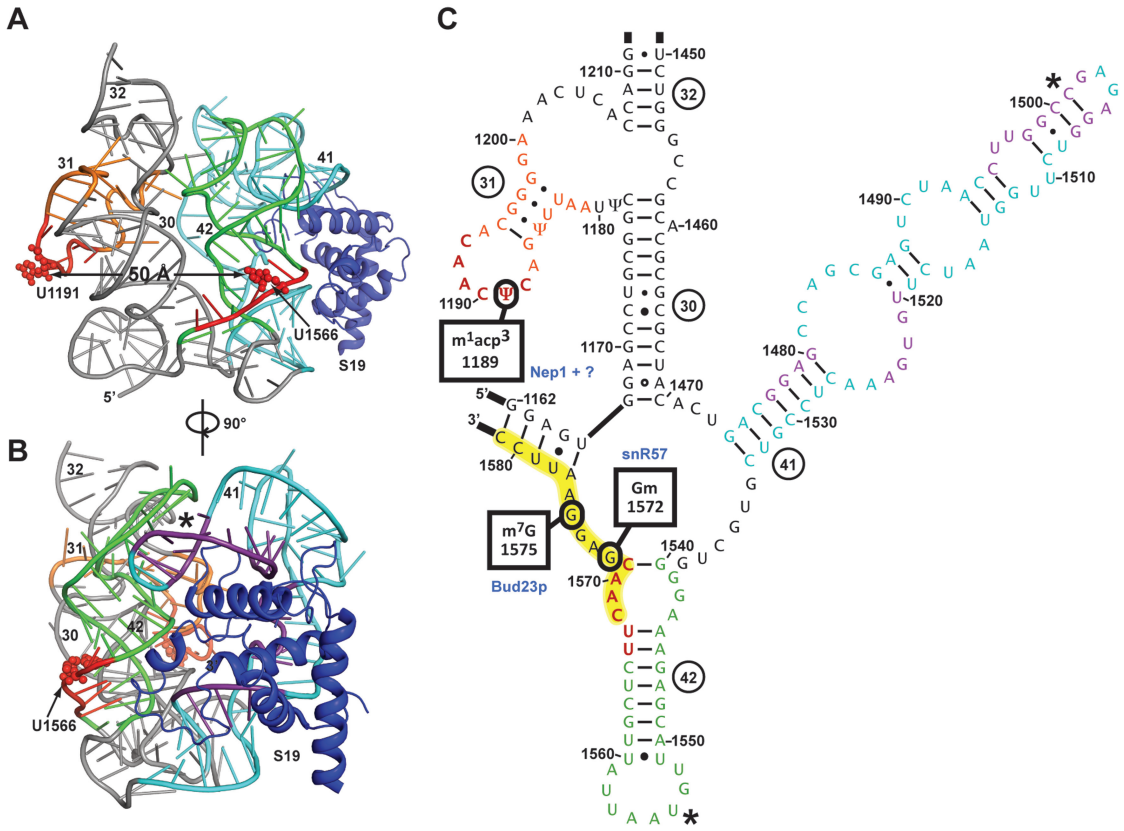

Figure 6. Possible role for Nep1 in S19 loading. (**A**, **B**) Three dimensional and (**C**) Secondary structure map of the head region of eukaryotic 18S rRNA, truncated to remove bases 1212–1439 for simplicity. (A) and (B) were generated from PDB ID 3JYV [cryo-EM model of the 40S subunit of the *Thermomyces lanuginosus* ribosome (25)]. (C) is based on the secondary structure prediction for *S. cerevisiae* 18S rRNA from the Comparative RNA Web Site (www.rna.ccbb.utexas.edu). Regions in tertiary structure and secondary are colored orange, cyan and green for helix 31, 41 and 42 respectively. Red color indicates Nep1 consensus sites, with U1189 and U1566, the bases that would bind in the Nep1 active site, in small spheres. The snR57-binding site is highlighted in yellow in (C). Purple denotes parts of rRNA within 4 Å of S19 (shown as blue cartoon). Asterisk indicates the site of interaction between helix 41 and 42.

structure, suggesting that this is the preferred substrate for Nep1 (23). However, for *M. jannaschii* Nep1, sequences that do not contain a stem can also bind (14,23). We have no information about the structure of the free RNA, and no similar loops containing this sequence were found in the structure databases, but the stem-loop is predicted to form at 37°C by RNAfold (38), although we are not sure if U6 would be flipped out in the absence of protein. The RNA seen in our structures fails to contact much of the positively charged surface of the protein, leaving open the possibility that the actual binding site is larger than the selected consensus sequence.

It has been proposed that Nep1 is the first example of an N1-specific pseudouridine methyltransferase and that it catalyzes the only such modification known in 18S rRNA, the N1-methylation on the hypermodified m1acp3Ψ at position 1189 in yeast (14). The structural evidence, seen in the specific positioning of the C5 of uridine in the enzyme active site, supports this. The uridine is held in place by several contacts that impart this specificity, and the interaction with Arg 88 suggests that this residue plays a role in catalysis, by promoting deprotonation of the ring (Supplementary Figure S5B). Asp 90, which is equivalent to Asp 86 that is mutated in Bowen–Conradi Syndrome, also plays a role by holding Arg 88 in place in the structure. There are no candidates for direct hydrogen binding near the C5 of the uridine in our structure that may allow specific selection of pseudouridine, but the absence of an activating nucleophile in the vicinity of C6 in these structures rules out the possibility that this protein functions as an m5U methyltransferase, as would be suggested if uridine was the substrate (39,40). In addition, m5U modification has not been observed in yeast 18S rRNA. However, this means that uridine is indistinguishable from pseudouridine in the Nep1 active site, but Nep1 will only catalyze methyl transfer on pseudouridine. This is supported by fluorescence binding data for *M. jannaschii* Nep1 reported by Wurm *et al.* (14) that shows no significant differences in RNA-binding K_d when pseudouridine is replaced with uridine in high affinity RNA-binding sites.

Although it is likely that the N1-methyltransferase activity is important for Nep1 function, as supported by the severe defect caused by mutations at Asp 90, it has been reported that mutations in the SAM-binding site of Nep1 that reduces its ligand-binding activity by 100-fold results in no disruption of Nep1 activity (13). Therefore it is plausible that binding of Nep1 to the rRNA in the absence of methyltransfer activity is sufficient to carry

out at least some of its role in ribosome biogenesis. In addition, overexpression of yeast ribosomal protein Rps19 compensates for deficiency in Nep1 function (23). An explanation put forth for these observations was that Nep1 may play a role that is independent of Nep1 methyltransferase activity (13,23). A possible role may be to act as a chaperone of RNA folding during a specific assembly step. The observation of a two RNA complex in which both binding sites are occupied suggests that it is plausible that rRNA occupies both sites when Nep 1 binds *in vivo*. Based on structural analysis, the two binding sites are non-equivalent, highlighted by differences in binding to SAH (Figure 5B and Supplementary Figure S5A). This idea that both sites would be occupied in the ribosome is further supported by the presence of extensive positively charged surface not contacted by the stem-loops in the structures. Two of the three possible Nep1-binding sites in yeast 18S rRNA are located within 50 Å of each other in the 3D structure of the eukaryotic ribosome (25) (Figure 6A and C). Those sites are the loop terminating helix 31 and a portion of helix 42 (Figure 6C). Based on the EM reconstruction of the eukaryotic ribosome (25), in order for Nep1 to bind the site at helix 41, significant secondary structure alteration would have to occur (Figure 6 and Supplementary Figure S6). Half of the expected stem loop that is expected to form when Nep1 binds is part of helix 42. Therefore, it is possible that when Nep1 binds, there is a rearrangement of the interactions to produce the expected stem-loop (Figure 6 and Supplementary Figure S6). This rearrangement would also result in the two Nep1 sites being moved closer together, since the distance between active sites in the Nep1 dimer is ~20 Å, while it is predicted to be 50 Å apart in the mature rRNA (Figure 6A). The loop at the end of helix 42 interacts with the loop of helix 41 (Figure 6B). Helix 41 is the site of most of the contacts between the rRNA and S19 (Figure 6B and C). Therefore, a possible model is that Nep1 interacts with the rRNA at two sites—the loop of helix 31, where it catalyzes the methylation of Ψ1189 and the stem of helix 42, where it acts as an RNA chaperone and breaks apart helix 42, rearranging the area and the interactions with helix 41 to promote S19 binding. In this scenario, more of the Nep1 surface would be expected to contact rRNA, as predicted by this structure. This is assuming that rRNA in early maturation stages have structures near the predicted structure of the mature rRNA. However, there is evidence is that pre-rRNA structure can differ from mature rRNA structure (41). In the absence of snR57, which is predicted to bind between bases 1568 and 1584, a Nep1 deletion is no longer lethal, indicating that Nep1 is only required when snR57 is present (23). Since snR57 guides the methylation of G1572 (24), a site near helix 42 (Figure 6C), this interaction between Nep1 and snR57 may be due to structural changes that occur as a result of methylation. Another possibility, however, is that the interaction between Nep1 and its binding site in that region may promote the dissociation of snR57 from rRNA. This dissociation may be necessary for proper rRNA structure, or even for N7 methylation of G1575 catalyzed by Bud23p (42). In any case, the genetic interaction between the snR57 and Nep1 supports the possibility of Nep1 binding in the region near helix 42.

Methanocaldococcus jannaschii Nep1 binds to yeast 18S rRNA sequences corresponding to the loop of helix 31 (14), and afNep1 similarly is capable of binding yeast rRNA sequences corresponding to the Nep1-binding site near helix 42 (data not shown). Residues involved in key interactions with the RNA, for example Arg 88, Arg 136, Arg 129, Leu 140 and Leu 159, are identical in afNep1. Superposition of the archaeal dimer on the yeast structures showed that the RNA interaction mode is likely to be very highly conserved between archaea and yeast (Supplementary Figure S3D). Although these structures do not provide all the answers on Nep1 function, it does provide the basis for continued studies on the ancient structural mechanisms utilized by Nep1 in rRNA processing.

ACCESSION NUMBERS

3O7B, 3OII, 3OIJ and 3OIN.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors are grateful to their colleagues Profs T. Kwaku Dayie, Jonathan Dinman and Steven Rokita (University of Maryland) for helpful discussions and Dr Jamaine Davis (National Cancer Institute) for comments on the manuscript.

FUNDING

National Science Foundation (MCB0953493, CAREER Award to N.L.); Maryland Department of Business and Economic Development Nano-Biotechnology Award (to the University of Maryland); and Achievement Rewards for College Scientists (ARCS) Fellowship (to S.T); National Center for Research Resources at the National Institutes of Health, Advanced Photon Source on the Northeastern Collaborative Access Team beamlines (award RR-15301); U.S. Department of Energy, Office of Basic Energy Sciences, Advanced Photon Source (Contract No. DE-AC02-06CH11357). Funding for open access charge: National Science Foundation (MCB0953493, CAREER Award to N.L.).

Conflict of interest statement. None declared.

REFERENCES

1. Venema,J. and Tollervey,D. (1999) Ribosome synthesis in Saccharomyces cerevisiae. *Annu. Rev. Genet.*, **33**, 261–311.
2. Fromont-Racine,M., Senger,B., Saveanu,C. and Fasiolo,F. (2003) Ribosome assembly in eukaryotes. *Gene*, **313**, 17–42.
3. Henras,A.K., Soudet,J., Gerus,M., Lebaron,S., Caizergues-Ferrer,M., Mougin,A. and Henry,Y. (2008) The

post-transcriptional steps of eukaryotic ribosome biogenesis. *Cell Mol. Life Sci.*, **65**, 2334–2359.
4. Nazar,R.N. (2004) Ribosomal RNA processing and ribosome biogenesis in eukaryotes. *IUBMB Life*, **56**, 457–465.
5. Hakuno,F., Hughes,D.A. and Yamamoto,M. (1996) The Schizosaccharomyces pombe mra1 gene, which is required for cell growth and mating, can suppress the mating inefficiency caused by a deficit in the Ras1 activity. *Genes Cells*, **1**, 303–315.
6. Liu,P.C. and Thiele,D.J. (2001) Novel stress-responsive genes EMG1 and NOP14 encode conserved, interacting proteins required for 40S ribosome biogenesis. *Mol. Biol. Cell*, **12**, 3644–3657.
7. Eschrich,D., Buchhaupt,M., Kotter,P. and Entian,K.D. (2002) Nep1p (Emg1p), a novel protein conserved in eukaryotes and archaea, is involved in ribosome biogenesis. *Curr. Genet.*, **40**, 326–338.
8. Armistead,J., Khatkar,S., Meyer,B., Mark,B.L., Patel,N., Coghlan,G., Lamont,R.E., Liu,S., Wiechert,J., Cattini,P.A. *et al.* (2009) Mutation of a gene essential for ribosome biogenesis, EMG1, causes Bowen-Conradi syndrome. *Am. J. Hum. Genet.*, **84**, 728–739.
9. Hunter,A.G., Woerner,S.J., Montalvo-Hicks,L.D., Fowlow,S.B., Haslam,R.H., Metcalf,P.J. and Lowry,R.B. (1979) The Bowen-Conradi syndrome – a highly lethal autosomal recessive syndrome of microcephaly, micrognathia, low birth weight, and joint deformities. *Am. J. Med. Genet.*, **3**, 269–279.
10. Lemire,E.G. (2002) Bowen-Conradi syndrome. *Clin. Dysmorphol.*, **11**, 149.
11. Taylor,A.B., Meyer,B., Leal,B.Z., Kotter,P., Schirf,V., Demeler,B., Hart,P.J., Entian,K.D. and Wohnert,J. (2008) The crystal structure of Nep1 reveals an extended SPOUT-class methyltransferase fold and a pre-organized SAM-binding site. *Nucleic Acids Res.*, **36**, 1542–1554.
12. Wurm,J.P., Duchardt,E., Meyer,B., Leal,B.Z., Kotter,P., Entian,K.D. and Wohnert,J. (2009) Backbone resonance assignments of the 48 kDa dimeric putative 18S rRNA-methyltransferase Nep1 from Methanocaldococcus jannaschii. *Biomol. NMR Assign.*, **3**, 251–254.
13. Leulliot,N., Bohnsack,M.T., Graille,M., Tollervey,D. and Van Tilbeurgh,H. (2008) The yeast ribosome synthesis factor Emg1 is a novel member of the superfamily of alpha/beta knot fold methyltransferases. *Nucleic Acids Res.*, **36**, 629–639.
14. Wurm,J.P., Meyer,B., Bahr,U., Held,M., Frolow,O., Kotter,P., Engels,J.W., Heckel,A., Karas,M., Entian,K.D. *et al.* (2010) The ribosome assembly factor Nep1 responsible for Bowen-Conradi syndrome is a pseudouridine-N1-specific methyltransferase. *Nucleic Acids Res.*, **38**, 2387–2398.
15. Anantharaman,V., Koonin,E.V. and Aravind,L. (2002) SPOUT: a class of methyltransferases that includes spoU and trmD RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. *J. Mol. Microbiol. Biotechnol.*, **4**, 71–75.
16. Tkaczuk,K.L., Dunin-Horkawicz,S., Purta,E. and Bujnicki,J.M. (2007) Structural and evolutionary bioinformatics of the SPOUT superfamily of methyltransferases. *BMC Bioinformatics*, **8**, 73.
17. Ero,R., Peil,L., Liiv,A. and Remme,J. (2008) Identification of pseudouridine methyltransferase in Escherichia coli. *RNA*, **14**, 2223–2233.
18. Draptchinskaia,N., Gustavsson,P., Andersson,B., Pettersson,M., Willig,T.N., Dianzani,I., Ball,S., Tchernia,G., Klar,J., Matsson,H. *et al.* (1999) The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. *Nat. Genet.*, **21**, 169–175.
19. Morimoto,K., Lin,S. and Sakamoto,K. (2007) The functions of RPS19 and their relationship to Diamond-Blackfan anemia: a review. *Mol. Genet. Metab.*, **90**, 358–362.
20. Ferreira-Cerca,S., Poll,G., Gleizes,P.E., Tschochner,H. and Milkereit,P. (2005) Roles of eukaryotic ribosomal proteins in maturation and transport of pre-18S rRNA and ribosome function. *Mol. Cell*, **20**, 263–275.
21. Idol,R.A., Robledo,S., Du,H.Y., Crimmins,D.L., Wilson,D.B., Ladenson,J.H., Bessler,M. and Mason,P.J. (2007) Cells depleted for RPS19, a protein associated with Diamond Blackfan Anemia, show defects in 18S ribosomal RNA synthesis and small ribosomal subunit production. *Blood Cells Mol. Dis.*, **39**, 35–43.
22. Leger-Silvestre,I., Caffrey,J.M., Dawaliby,R., Alvarez-Arias,D.A., Gas,N., Bertolone,S.J., Gleizes,P.E. and Ellis,S.R. (2005) Specific role for yeast homologs of the Diamond Blackfan Anemia-associated Rps19 protein in ribosome synthesis. *J. Biol. Chem.*, **280**, 38177–38185.
23. Buchhaupt,M., Meyer,B., Kotter,P. and Entian,K.D. (2006) Genetic evidence for 18S rRNA binding and an Rps19p assembly function of yeast nucleolar protein Nep1p. *Mol. Genet. Genomics*, **276**, 273–284.
24. Lowe,T.M. and Eddy,S.R. (1999) A computational screen for methylation guide snoRNAs in yeast. *Science*, **283**, 1168–1171.
25. Taylor,D.J., Devkota,B., Huang,A.D., Topf,M., Narayanan,E., Sali,A., Harvey,S.C. and Frank,J. (2009) Comprehensive molecular structure of the eukaryotic ribosome. *Structure*, **17**, 1591–1604.
26. Bakin,A. and Ofengand,J. (1995) Mapping of the 13 pseudouridine residues in Saccharomyces cerevisiae small subunit ribosomal RNA to nucleotide resolution. *Nucleic Acids Res.*, **23**, 3290–3294.
27. Vonrhein,C., Blanc,E., Roversi,P. and Bricogne,G. (2006) Automated structure solution with autoSHARP. *Methods Mol Biol*, **364**, 215–230.
28. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.*, **60**, 2126–2132.
29. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D. Biol. Crystallogr.*, **53**, 240–255.
30. Vagin,A. and Teplyakov,A. (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D. Biol. Crystallogr.*, **56**, 1622–1624.
31. Vagin,A. and Teplyakov,A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 22–25.
32. Potterton,E., Briggs,P., Turkenburg,M. and Dodson,E. (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D. Biol. Crystallogr.*, **59**, 1131–1137.
33. Adams,P.D., Gopal,K., Grosse-Kunstleve,R.W., Hung,L.W., Ioerger,T.R., McCoy,A.J., Moriarty,N.W., Pai,R.K., Read,R.J., Romo,T.D. *et al.* (2004) Recent developments in the PHENIX software for automated crystallographic structure determination. *J. Synchrotron. Radiat.*, **11**, 53–55.
34. Adams,P.D., Grosse-Kunstleve,R.W., Hung,L.W., Ioerger,T.R., McCoy,A.J., Moriarty,N.W., Read,R.J., Sacchettini,J.C., Sauter,N.K. and Terwilliger,T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D. Biol. Crystallogr.*, **58**, 1948–1954.
35. Painter,J. and Merritt,E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D. Biol. Crystallogr.*, **62**, 439–450.
36. Beeby,M., O'Connor,B.D., Ryttersgaard,C., Boutz,D.R., Perry,L.J. and Yeates,T.O. (2005) The genomics of disulfide bonding and protein stabilization in thermophiles. *PLoS Biol.*, **3**, e309.
37. Ofengand,J. and Schaefer,H. (1965) On the ionization constant of 5-ribosyluracil. *Biochemistry*, **4**, 2832–2835.
38. Hofacker,I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.*, **31**, 3429–3431.
39. Alian,A., Lee,T.T., Griner,S.L., Stroud,R.M. and Finer-Moore,J. (2008) Structure of a TrmA-RNA complex: a consensus RNA fold contributes to substrate selectivity and catalysis in m5U methyltransferases. *Proc. Natl Acad. Sci. USA*, **105**, 6876–6881.
40. Hur,S., Stroud,R.M. and Finer-Moore,J. (2006) Substrate recognition by RNA 5-methyluridine methyltransferases and pseudouridine synthases: a structural perspective. *J. Biol. Chem.*, **281**, 38969–38973.
41. Schafer,T., Maco,B., Petfalski,E., Tollervey,D., Bottcher,B., Aebi,U. and Hurt,E. (2006) Hrr25-dependent phosphorylation state regulates organization of the pre-40S subunit. *Nature*, **441**, 651–655.
42. White,J., Li,Z., Sardana,R., Bujnicki,J.M., Marcotte,E.M. and Johnson,A.W. (2008) Bud23 methylates G1575 of 18S rRNA and is required for efficient nuclear export of pre-40S subunits. *Mol. Cell Biol.*, **28**, 3151–3161.