A Dimeric Structure for Archaeal Box C/D Small Ribonucleoproteins
Franziska Bleichert, Keith T. Gagnon, Bernard A. Brown II, E. Stuart Maxwell, Andres E. Leshcziner, Vinzenz M. Unger, Susan J. Baserga

Methylation of ribosomal RNA (rRNA) is required for optimal protein synthesis. Multiple 2′-O-ribosyl methylation is carried out by box C/D guide ribonucleoproteins (sRNPs) and small nuclear ribonucleoproteins (snoRNPs), which are conserved from archaea to eukaryotes. Methylation is dictated by base pairing between the specific guide RNA component of the sRNP or snoRNP and the target rRNA. We determined the structure of a reconstituted and catalytically active box C/D sRNP from the archaeon Methanocaldococcus jannaschii by single-particle electron microscopy. We found that archaeal box C/D sRNPs unexpectedly formed a dimeric structure with an alternative organization of their RNA and protein components that challenges the conventional view of their architecture. Mutational analysis demonstrated that this di-sRNP structure was relevant for the enzymatic function of archaeal box C/D sRNPs.

In contrast to their eukaryotic counterparts, enzymatically active archaean C/D sRNPs can be reconstituted in vitro (7). For efficient 2′-O-ribosyl methylation, each of the box C/D and box C′/D′ motifs in one sRNA are required to assemble symmetrically with all three core proteins into an RNP that is conventionally illustrated as containing one sRNA and two copies of each of the three core proteins (Fig. 1A) (8, 9). However, no structure of enzymatically active box C/D sRNPs containing the full-length sRNA exists.

To determine the three-dimensional (3D) structure of a reconstituted and catalytically competent methylation guide sRNP from the hyperthermophilic euryarchaeote Methanocaldococcus jannaschii, we used electron microscopy (EM) and single-particle analysis. We reconstituted box C/D sRNPs in vitro using recombinant M. jannaschii core proteins and in vitro transcribed M. jannaschii sR8 sRNA and subsequently purified the assembled RNP on glycerol gradients. All sRNP components comigrated in peak gradient fractions 10 and 11 (Fig. 1B). Consistent

1Department of Genetics, Yale University School of Medicine, New Haven, CT 06520, USA. 2Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695, USA. 3Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA. 4Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, CT 06520, USA. 5Department of Therapeutic Radiology, Yale University School of Medicine, New Haven, CT 06520, USA.

*To whom correspondence should be addressed. E-mail: susan.baserga@yale.edu
with previous results (9), assembly required the presence of the sRNA (Fig. 1C).

The purified sRNP was catalytically active for methylation of RNA substrates using either guide sequence in the sR8 sRNA, and methylation was specific to the fifth nucleotide upstream of boxes D and D’ (Fig. 1E). Comparison of the reconstituted sRNP to molecular mass markers of indicated S values (Fig. 1B) as well as to the human U1 snRNP (240 kD; Fig. 1D) showed that the catalytically active box C/D sRNP migrated at 12 S, faster than the human U1 snRNP. In contrast, the heterotetrameric complex formed by Nop5 and fibrillarin (136 kD) (8, 10, 11) migrated at ~6.5 S in glycerol gradients (Fig. 1C and fig. S1). These results indicate that the box C/D sRNP is a much larger complex than what would be expected on the basis of the conventional model of box C/D sRNP architecture (183 kD predicted). Analysis of complexes by gel filtration chromatography supported these conclusions (fig. S2, A to C). Furthermore, various biochemical experiments (fig. S2) demonstrated that the reconstituted and catalytically active box C/D sRNPs were biochemically homogeneous.

EM analysis of negatively stained complexes from peak gradient fractions showed a monodisperse population of particles (Fig. 2A and fig. S2D). Use of the random conical tilt method (12) enabled a 3D ab initio reconstruction of the box C/D sRNP by single-particle analysis (figs. S3 to S8). The EM structure was refined to a resolution of 27 Å (fig. S6B) (13). The experimental class averages and projections of the refined volume were in good agreement with each other (Fig. 2B).

The refined sRNP volume measured 14.8 nm by 13 nm by 9 nm and exhibited two-fold pseudo-symmetry (Fig. 2C). Consistent with our biochemical analyses (Fig. 1 and figs. S1 and S2), the volume was larger than anticipated on the basis of existing atomic-resolution structures of box C/D core proteins and the conventional box C/D sRNP model (10, 11). Docking of the crystal structures into the EM volume revealed that it could accommodate not one but two Nop5-fibrillarin heterotetramers, placing fibrillarin in the corners of the complex (Fig. 2D). Consistent with a relative 1:1:1 stoichiometry of L7Ae, Nop5, and fibrillarin as determined by quantitative amino acid analysis (13), four L7Ae molecules could be fit into the remaining density, placing L7Ae in proximity to the C-terminal domain of Nop5 (Fig. 2D). Thus, the sRNP EM volume contained four copies of each core protein.

Previous biochemical data (8, 9, 14), together with the presence of four sets of each of the core proteins (L7Ae, Nop5, and fibrillarin) in the sRNP volume, strongly suggest that the box C/D sRNP contains two sRNA molecules. The box C/D and C’/D’ motifs are most likely positioned in the remaining density in proximity to the sRNA binding protein L7Ae (15) and the C-terminal RNA binding domain of Nop5 (10). This topology is consistent with the recent crystal structure of hPrp31 (a Nop5 homolog) complexed with 15.5K (an L7Ae homolog) and a U4 snRNA fragment (fig. S9) (16).

Although the resolution of the EM structure did not allow the sRNA molecules to be unambiguously localized, the distribution of the remaining density in the EM volume not occupied by the core box C/D proteins implies that the two sRNAs connect the two Nop5-fibrillarin heterotetramers along the short sides of the sRNP volume and follow a different directionality than the Nop5 coiled-coil domains. Consistent with this interpretation, ribonuclease treatment of the reconstituted sRNP yielded smaller complexes (14 nm by 6 nm in projection), which were about half the size of untreated sRNPs (fig. S10).

Taken together, these results were not compatible with the conventional model of box C/D sRNP architecture (Fig. 2F) but could be explained by an alternative model: a di-sRNP containing four copies of each core protein and two sRNAs (Fig. 2E).
To verify the di-sRNP model, we analyzed sRNPs lacking both fibrillarin and the N-terminal, fibrillarin-interacting domain of Nop5 (Nop5∆N/minus fibrillarin). Docking of the crystal structures into the EM volume (Fig. 2D) predicted that these RNP components occupied the density in the four corners of the sRNP volume and that Nop5∆N/minus fibrillarin sRNPs should lack those four regions of density. Results from both biochemical and EM experiments were consistent with a smaller size of these particles relative to the fully assembled sRNP (Fig. 3, A to C, and figs. S11 and S12). Class averages of the Nop5∆N/minus fibrillarin sRNP indeed lacked the strong density in the four corners (Fig. 3, D to G). These results provide further evidence that the box C/D sRNP contained four copies of fibrillarin, and likewise four copies of L7Ae and Nop5, and consequently support the di-sRNP model (Fig. 2E).

The di-sRNP model predicted a molecular mass of 366 kD for the *M. jannaschii* box C/D sRNP. This size was consistent with the results obtained both by glycerol gradient centrifugation (Fig. 1B) and by gel filtration chromatography (fig. S2A). Furthermore, the S value calculated from the EM volume was 11.7 S (13), very close to the observed value of 12 S (Fig. 1B). Di-sRNP formation was also apparent with a different EM staining technique, the GraFix method of glycerol gradient centrifugation (17), and with different constructs for protein expression (figs. S13 to S16).

**Fig. 2.** EM and 3D reconstruction of the archaeal box C/D sRNP. (A) Electron micrograph of negatively stained sRNPs from peak fractions after glycerol gradient centrifugation. (B) Experimental class averages and corresponding 2D projections of the reconstructed 3D volume. Scale bar, 10 nm. (See also figs. S3 and S4.) (C) Isodensity map of the reconstructed 3D volume. (D) Docking of the crystal structures of *Pyrococcus furiosus* Nop5-fibrillarin [PDB 2nnw (11)] and *M. jannaschii* L7Ae [1xbi (25)] in the isodensity map. Blue, Nop5; orange, fibrillarin; yellow, L7Ae. The volumes were thresholded to 118% of the molecular mass of the di-sRNP. (E and F) Proposed di-sRNP model (E) contrasted with the conventional model of archaeal box C/D sRNP architecture (F). Colors are as in (D) and the RNA is shown in gray. The orientation of the sRNA ends was purposefully left ambiguous.

**Fig. 3.** Localization of fibrillarin and the N-terminal domain of Nop5 in the EM structure. (A and B) Sedimentation profile of the Nop5∆N/minus fibrillarin sRNP (A) and L7Ae and Nop5∆N in the absence of the sRNA (B) analyzed by glycerol gradient centrifugation. (C and D) Electron micrograph (C) and representative class average of negatively stained Nop5∆N/minus fibrillarin sRNP particles from peak glycerol gradient fractions (D) (see also fig. S11). (E) Class average of the reconstituted box C/D sRNP particles containing fibrillarin and the N terminus of Nop5. (F) Difference map between the class averages of the two different RNPs as shown in (D) and (E). (G) Statistically significant region as calculated with the *t* test in SPIDER with a *P* value of *P* ≤ 0.001. Scale bars, 10 nm [(D) and (E)].
Furthermore, the di-sRNP model predicted that the Nop5 proteins are critical for maintaining this conformation, as they would orchestrate the positioning of the two different sRNAs (Fig. 2, D and E). Previous biochemical experiments showed that some mutations in the coiled-coil domain of Nop5 (Nop5 CC) abrogated methylation activity of the sRNP, whereas others were tolerated (8, 14). We analyzed these Nop5 CC mutant sRNPs and found that both Nop5mut2 and Nop5mut4 sRNPs, which were previously shown to methylate substrate RNAs (14), had sizes and dimensions similar to those of sRNPs containing the wild-type Nop5, whereas Nop5ΔCC sRNPs, which are methyllation-deficient (14), were smaller (Fig. 4, A to C, and figs. S17 and S18). Collectively, our results indicate that the formation of the di-sRNP structure correlates with efficient methylation activity of the box C/D sRNP.

Our work provides a 3D structure of a catalytically active box C/D sRNP, revealing an unexpected di-sRNP topology. Multimeric RNPs composed of individual RNPs are not unprecedented. Examples include the U4/U6 di-snRNP, the U4/U6:U5 tri-snRNP, the U11/U12 di-snRNP, and telomerase (18–23). Since the first discovery of a box C/D snoRNA 40 years ago (24), box C/D sRNPs and snoRNPs have been assumed to consist of one sRNA or snoRNA molecule and one or two sets of each core protein. Contrary to this original assumption, the structure of a reconstituted archaeal box C/D sRNP presented here argues that these complexes form a functionally relevant, unanticipated di-sRNP structure with an alternative organization of the RNA and protein components. We do not know whether this structure forms in vivo. If it does, then box C/D di-sRNPs could be composed of two different box C/D sRNAs. If multiple guide sequences are used at the same time, a di-sRNP (or di-snoRNP) architecture may be an efficient means by which these RNP chaperones can participate in folding of the long pre-rRNA.

References and Notes
13. See supporting material on Science Online.
26. We thank M. Golas, P. Moore, K. Reisch, J. Steitz, T. Steitz, H. Wang, and S. Wolin for helpful discussions and/or critical reading of the manuscript, and the Yale University Biomedical High Performance Computing Center (supported by NIH grant 2R01GM69699). This work was supported by a Boehringer Ingelheim Fonds Ph.D. scholarship (F.B.), NIH grants R01GM69699 (B.A.B.) and R01GM52581 (S.J.B.), and NSF grant MCB0543741 (E.S.M.). The EM map has been deposited into EMDB (accession code EMD-1636).

Supporting Online Material
www.sciencemag.org/cgi/content/full/325/5946/1384/DC1
Materials and Methods
Figs. S1 to S18
References
11 May 2009; accepted 20 July 2009
10.1126/science.1176099