Structure of a functional ribonucleoprotein pseudouridine synthase bound to a substrate RNA

Bo Liang1, Jing Zhou2, Elliot Kahen2, Rebecca M Terns3, Michael P Terns3 & Hong Li1,2

Box H/ACA small nucleolar and Cajal body ribonucleoprotein particles comprise the most complex pseudouridine syntheses and are essential for ribosome and spliceosome maturation. The multistep and multicomponent-mediated enzyme mechanism remains only partially understood. Here we report a crystal structure at 2.35 Å of a substrate-bound functional archaeal enzyme containing three of the four proteins, Cbf5, Nop10 and L7Ae, and a box H/ACA RNA that reveals detailed information about the protein-only active site. The substrate RNA, containing 5-fluorouridine at the modification position, is fully docked and catalytically rearranged by the enzyme in a manner similar to that seen in two stand-alone pseudouridine syntheses.

Structural analysis provides a mechanism for plasticity in the diversity of guide RNA sequences used and identifies a substrate-anchoring loop of Cbf5 that also interacts with Gar1 in unliganded structures. Activity analyses of mutated proteins and RNAs support the structural findings and further suggest a role of the Cbf5 loop in regulation of enzyme activity.

Pseudouridine (ψ), a rotational isomer of uridine, is the most abundant modified nucleotide and is found in virtually all transfer RNA (tRNA), ribosomal RNA (rRNA) and small nuclear RNA (snRNA)1–4. In bacteria, five families of pseudouridine synthases catalyze site-specific isomerization in tRNA and rRNA5–7. In eukaryotes and archaea, a unique class of pseudouridine synthases that depend on noncoding RNAs is responsible for site-specific isomerization of RNA and snRNA8–11. The box H/ACA small nucleolar or Cajal body ribonucleoprotein particles (sno/scRNPs) comprise a minimal set of four proteins and a box H/ACA RNA. The H/ACA RNA secures the substrate RNA by base-pairing, whereas the proteins are believed to catalyze the chemical reaction. The four proteins are Cbf5 (NAF57 in mammals and dyskerin in human), Nop10, NH2P (L7Ae in archaea) and Gar1 (refs. 11–16). Cbf5 has been identified as the putative catalytic subunit based on its sequence similarities with the TruB family of bacterial pseudouridine synthases. In addition to the nucleotide-isomerization function, some members of H/ACA snoRNPs are responsible for rRNA processing17,18, and one vertebrate scaH/ACA RNP is required for telomere maintenance19,20. Vertebrate telomerase is known to harbor a box H/ACA RNP subdomain that is crucial to its biogenesis and stability in small Cajal bodies19–21. Significantly, mutations in human Cbf5 or dyskerin22, NOP10 (ref. 23) and NH2P24 have all been linked to the rare genetic disorder dyskeratosis congenita24.

The process of uridine isomerization by Cbf5-like pseudouridine synthases is only partially understood, and what is known has come largely from studies of the stand-alone pseudouridine synthases in bacteria25–27. Cbf5 resembles the best-characterized bacterial pseudouridine synthase TruB both in sequence and structure and, in limited cases, can act as a stand-alone pseudouridine synthase on tRNA substrates28,29. It is thus believed that Cbf5 shares the same catalytic mechanism as TruB and differs only in its substrate-recognition mechanism.

The sno/scRNA-guided processes face unique challenges in substrate binding and release. The snoRNAs mediate modification of core regions of the ribosomal RNA29,30. Thus, in order for rRNA to be processed and modified in a timely manner, snoRNPs must gain access to specific nucleotides within the large and complex rRNA, carry out processing or modification reactions, and then release the mature rRNA to allow folding and assembly with ribosomal proteins. This process requires all box H/ACA complex proteins and the guide RNA via a complicated mechanism for which the molecular basis remains unknown31–33. Recent advances in structural studies of archaeal box H/ACA RNPs have provided glimpses into the architecture of the fully assembled enzyme and suggested intriguing roles for the noncatalytic subunits in substrate placement34–40. Regardless of their positions and interactions in the assembled RNP, individual protein subunits affect the placement process and the final conformation of the substrate RNA35,41.

Here we describe a 2.35-Å crystal structure of a functional archaeal H/ACA RNP containing Pyrococcus furiosus (Pt) Cbf5, Nop10, L7Ae and a guide RNA with a substrate RNA bound at the active site. The guide RNA used in crystallization is a composite RNA based on the previously characterized Pf9 RNA31 and another computationally identified Pf H/ACA RNA, Pf6 (Fig. 1a)42,43. The substrate RNA is a 21-nt oligomer containing 5-fluorouridine (F5U) at the target uridine position (position 10). We also report a structure of a subcomplex containing Cbf5, Nop10 and a model guide RNA46 with a bound but undocked substrate RNA at 3.65 Å.

1Institute of Molecular Biophysics, 2Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida, USA. 3Department of Biochemistry and Molecular Biology, University of Georgia at Athens, Athens, Georgia, USA. Correspondence and requests for materials should be addressed to H.L. (hong.li@fsu.edu).

Received 21 May; accepted 22 May; published online 28 May 2009; doi:10.1038/nsmb.1624
RESULTS

Overall structure

The overall guide RNA structure resembles that of the fully assembled RNP in the absence of the substrate RNA in which its lower stem is anchored by the hallmark trinucleotide ACA on the PUA domain of Cbf5 and its upper stem is anchored by the kink-turn motif bound to L7Ae (Fig. 1b and Table 1). The structure of the bound guide-substrate RNA complex in the active enzyme differs from that in the L7Ae-minus enzyme (PDB 2RFK) and more so from the protein-free complexes (PDB 2P89 and 2PCW) suggesting an important role of proteins in shaping the flexible RNA. Although crystallographic contacts are observed at the end of the lower stem of the guide RNA, the bound substrate RNA is completely free of crystallographic contacts. The proteins, Cbf5, Nop10 and L7Ae, form an elongated platform on which the guide-substrate RNA complex lies longitudinally. The RNA-protein interface buries an extensive solvent-accessible surface (6,694 Å²), of which ~30% is attributable to the substrate. The protein-RNA interface is greater in this structure than that in the L7Ae-minus RNP structure (PDB 2RFK, 5,332 Å²) and that of the substrate-minus RNP structure (PDB 2HYV, 4,488 Å²). This finding suggests that binding of the substrate RNA enhances the interaction between the protein complex and the RNA.

Active site structure

Unbiased electron density maps calculated before modeling the substrate RNA indicate that the nucleotide targeted for modification is fully docked into the active site of Cbf5 (Fig. 1c). The f5U is rearranged and also seems to be cis hydrated at position 6 to (5S,6R)-5-fluoro-6-hydroxy-pseudouridine (f5ho6Ψ), as was previously observed in the TruB-substrate RNA complex. However, the features of the electron density do not completely exclude a hydrated trans isomer of f5ho6Ψ or a nonhydrated f5U, as was suggested by MS analysis of the reaction products of a f5U-containing substrate by TruB. For simplicity, we modeled the nucleotide as the rearranged and cis hydrated product, f5ho6Ψ. We superimposed the f5ho6Ψ in the structure of this complex with those found in the TruB-RNA and RluA-RNA complex structures and compared its surrounding amino acids (Fig. 2). The three active sites share strong similarities. In addition to the strictly conserved aspartate, a tyrosine, two glycines, an arginine, a lysine and a hydrophobic residue form a nearly identical pocket that accommodates the modified nucleotide to be modified is fully docked into the active site and catalytic mechanism with TruB and possibly RluA. Our finding that Cbf5 probably shares a similar catalytic mechanism with TruB and RluA complexes suggests that the RNP lacking Gar1 is a functional pseudouridine synthase, although its efficiency is significantly compromised in the absence of Gar1 (refs. 31,33).

The active site of Cbf5 is accessible to both ordered and bulk solvent. Cross-validated and ϕA-weighted difference density maps reveal bound solvent or ion molecules. A cluster of peaks is found near the sugar phosphate moiety of f5ho6Ψ and is assigned to a hydrated potassium ion based on coordination geometries (Fig. 3a).
Gar1 binding for example, could prevent escape of the uracilate intermediate and, therefore, enhance the activity of Cbf5.

Substrate RNA binding to Cbf5

The substrate RNA interacts exclusively with Cbf5 residues (Fig. 3) at the tip of the V-shape that ends with f^5ho^6Ψ (G7–C12). The rearranged nucleotide f^5ho^6Ψ establishes the most extensive interactions with a polar pocket of Cbf5 (Fig. 3). The carboxylate group of the putative catalytic residue Asp85 establishes two hydrogen bonds with f^5ho^6Ψ: Oδ1 with N3 (2.9 Å) and Oδ2 with 2′-OH (2.4 Å). Two amide nitrogen atoms (Ile183 and Arg184) contact N1 and O6, respectively (Fig. 3a). The amide nitrogen of Gly180 and the guanidinium group of Arg205 further enhance the interaction by contacting two nonbridging oxygen atoms (Fig. 3a).

The potassium ion is within coordination distance (~2.4–2.9 Å) of two water molecules, the carbonyl oxygen of Thr181 and the O5′ atom of f^5ho^6Ψ. It is also within 3.3 Å to O4′ and O6 of f^5ho^6Ψ and the hydroxyl group of Tyr113. In both the TruB and RluA complex structures, this site also binds a solvent molecule.46,48 This solvent molecule(s) can potentially hydrate the rearranged f°U and may therefore have a role in the isomerization reaction. A second solvent site is 2.7 Å away from N1, which mediates recognition of the nucleobase of f^5ho^6Ψ by the N terminus of the α5 helix (Fig. 3a). Unexpectedly, the nucleobase is also accessible to bulk solvent, with a notable solvent-accessible area (7.7 Å²). In contrast, the nucleobase bound to TruB or RluA has a negligibly small solvent-accessible area (<0.8 Å²)46,48. The difference in solvent accessibility was found not to be due to the insertion regions in TruB (resides 83–101) and RluA (resides 175–195) that block the back entrance to the active site46,48. Rather, the difference is attributable to the less compact active site of Cbf5. Compression of the active site, through Gly82 Gly45 Ala46 Gly180 Tyr113 Arg181 Ile183 Arg184

Figure 3 Interactions of the RNA with Cbf5. (a) Interactions between the substrate RNA (red) and Cbf5 (green). Dashed lines indicate polar atom contacts within 3.4 Å. Red spheres indicate solvent molecules, and the purple sphere indicates bound potassium. (b) Schematic interactions between the guide (yellow) and substrate RNA (red) with Cbf5 (green).
Two Cbf5 regions interact with nucleotides flanking $f^9_{\text{ho}}$: an amino acid cluster comprising His63, His80, Gly81, Gly82 and Thr83 (His-Gly-Thr cluster) and the loop connecting the $\beta 7$ and $\beta 10$ strands ($\beta 7_{-10}$ loop) (Fig. 3 and Supplementary Fig. 1). The His-Gly-Thr cluster interacts with the major groove, whereas the $\beta 7_{-10}$ loop interacts with the minor groove side of the substrate RNA. These interactions are largely nonspecific, consistent with the flexibility observed in the identity of substrate RNA nucleotides in this region, and they involve hydrophobic and electrostatic contacts to the backbone of the RNA. The sugar phosphate backbone of substrate RNA nucleotides $G^7$--$G^9$ directs the approach of the target uridine to the active site and is stabilized by interactions with protein backbone atoms and positively charged arginine residues of the $\beta 7_{-10}$ loop ($\text{Arg154 and Arg156}$). On the strand leaving the active site, the G11 nucleobase stacks with the aliphatic chain of Arg146. In Cbf5 proteins from other organisms, this position is often substituted by a hydrophobic residue (Supplementary Fig. 1). Thr83 interacts with both G9 and C12 on either side of $f^9_{\text{ho}}$ by forming a hydrogen bond with its amide nitrogen to the 2′-hydroxyl oxygen of G9 and a water-mediated interaction with its hydroxyl oxygen to the nonbridging oxygen atoms of C12. This interaction is important for maintaining the $V$-shape of the substrate RNA that places the substrate uridine in the active site.

The impact of observed structural features on substrate binding

We further examined the importance of certain structural features of the composite Pf9-Pf6 RNA in substrate binding using a previously developed fluorescence assay41. In this assay, the nucleotide immediately downstream of the target uridine is substituted with fluorescent 2-aminopurine (2AP), which produces a high fluorescence intensity as a result of substrate docking. Furthermore, titrating the substrate RNA with a protein (or a protein complex) yields an apparent $K_d$ that measures the free energy for the last-assembled protein (or protein complex) to place the substrate at its fully docked position. We used the $K_d$ of docking a 2′-amino substituted substrate (amU, Fig. 4) as an upper limit for the wild-type $K_d$, which is experimentally difficult to obtain owing to slow isomerization of the substrate or release of the product.

Fluorescence evidence indicates the importance of the structure of the pseudouridine pocket ($\Psi$-pocket) in substrate binding. The nucleotide immediately downstream of the target uridine is unpaired in the overwhelming majority of known substrate RNAs. This unpaired nucleotide (G11) is extruded and is free of base-specific interactions (Fig. 3). We inserted a uridine between G15 and C16 in the Pf9-Pf6 guide RNA that can potentially form a base pair with the 2AP-substituted G11 in the fluorescence substrate RNA (paired $\Psi$-pocket). The substrate no longer docks into the active site (Fig. 4). We further asked whether the substrate RNA can accommodate additional unpaired nucleotides downstream of the target uridine, and we tested this by inserting a guanine between 2AP and the target uridine (unpaired $\Psi$-pocket). We observed that the substrate RNA with the insertion was docked to the active site of the RNP enzyme, although with an increased energetic cost (Fig. 4 and Supplementary Table 1). These results are consistent with the conservation of one unpaired 3′ adjacent nucleotide while accommodating some occurrences of two unpaired 3′ adjacent nucleotides48.

We further used mutational analysis to assess the impact of the $\beta 7_{-10}$ loop. We deleted three highly conserved residues within this loop: Ala148, Val149 and Lys150 ($C_\text{ALoop}$) (Supplementary Fig. 1). This deletion led to a substantial increase in the $K_d$ of the complex (Supplementary Table 1), suggesting a defect in substrate docking. Notably, omission of Gar1 from the complex ($C_{ALoopN}$) restored binding efficiency (Supplementary Table 1). We interpret this result to indicate that Gar1 functions through the $\beta 7_{-10}$ loop of Cbf5 and that both an intact $\beta 7_{-10}$ loop and Gar1 are required for correct placement of the substrate. The structural basis for the cooperativity of Gar1 and the $\beta 7_{-10}$ loop is discussed in a later section.

Placement of substrate RNA by L7Ae

The effect of L7Ae on substrate docking has been well discussed35,36,41,49. Our structure provides unambiguous support for the proposed role of L7Ae in substrate placement. The substrate RNA binds in the absence of L7Ae but is far from the active site. Delivery to the active site upon L7Ae binding is almost entirely based on a rotation of the guide-substrate helix, SH1, as a result of anchoring the upper stem, P2, by L7Ae (Fig. 5a,b). More unexpectedly, comparison of the new substrate-bound structure with a previously determined structure lacking the substrate RNA (this work and ref. 38) suggests plasticity in the complex that could be important in the interaction of the proteins with the diverse family of H/ACA RNAs and/or with their substrates (Fig. 5a). Whereas the Cbf5 structures in complexes containing or lacking the substrate RNA are closely superimposable, in the presence of the substrate, L7Ae, Nop10 and the upper stem of the guide RNA are positioned ~6 Å closer to the back face of the catalytic domain of Cbf5, leading to further anchoring of the guide RNA (Fig. 5a). The movement of L7Ae and Nop10 could be a result of differences in the secondary structures of the distinct guide RNAs used as well as of substrate binding. The large shift suggests a

![Diagram of substrate RNA docking](Image)
Figure 5 Structure and activity analysis of variously assembled wild-type and mutant RNP complexes reveal mechanisms for plasticity and the sensitivity of β7_10 loop to substrate and Gar1 binding. (a) The RNP structures of the substrate-plus (this work) and substrate-minus (PDB 2HVY) complexes are superimposed. A large downward shift of L7Ae, Nop10 and the guide RNA is observed upon binding of substrate RNA. L7Ae and Nop10 of the substrate-plus complex are labeled in green and those of the substrate-minus complex are in pink. Note that the substrate-minus complex would have 15 base pairs between the target uridine and the ACA trinucleotide if substrate is bound, whereas the current substrate-plus structure has 14 base pairs. (b) Cbf5 superimposed structures containing a bound substrate RNA. For clarity, only the substrate RNA and the β7_10 loop are shown. Each structure is labeled, where CNL denotes the structure of the Cbf5–Nop10–L7Ae complex bound with a guide and substrate RNA (this work), CN denotes that of the Cbf5–Nop10 complex bound with a guide and substrate RNA (PDB 2RFK)36, and CGN denotes that of Cbf5–Gar1–Nop10 complex bound with a guide and substrate RNA (PDB 2EY4)38. (c) Thin-layer chromatography radiograms of reacted and nuclease-digested substrate RNA with various RNP complexes. 'gRNA' denotes the Pf9_Pf6 composite RNA. 'C', 'G', 'N', 'L' denote Cbf5, Gar1, Nop10 and L7Ae respectively. 'Cm' denotes reacted and nuclease-digested substrate RNA unable to contact either the guide or the substrate RNA 34,36. Yet both Gar1 binds to a peripheral portion of the catalytic subunit and is unable to contact either the guide or the substrate RNA 34,36. Yeast Gar1 affects substrate RNA conformation via a Cbf5 loop (Gar1-plus, substrate-minus, or 'G+,S–' conformation) Fig. 5d). In the absence of Gar1 and the presence of a fully docked substrate RNA, the β7_10 loop is shown in blue (this work). The loop is in Gar1-minus and substrate-plus conformation ('G–,S+'). The structure of both Gar1 and substrate is shown in green (PDB 2APO), and the loop (β7 and β10) adopts the Gar1-minus, substrate-minus conformation ('G–,S–'). The proposed third conformation of the loop in the presence of both Gar1 and substrate RNA ('G+,S+) is indicated by a double arrow.

β7_10 loop conformation is sensitive to binding of both substrate RNA and Gar1. However, removal of three β7_10 loop residues (CAloopNGL) did not completely abolish pseudouridylation activity under the condition of excess enzyme over substrate (Fig. 5c), suggesting that the β7_10 loop has a role in the regulation of enzyme activity. We predict that in the presence of both substrate RNA and Gar1, the Cbf5 β7_10 loop experiences an energetic ‘tug-of-war’, which may result in positioning of the loop in a third, intermediate
conformation. The third (Gar1-plus, substrate-plus, or ‘G+,S+) loop conformation may optimally position the substrate in the active site, or close the active site, or mediate substrate binding and release (Fig. 5d). This interpretation is completely consistent with the conformational behavior of the substrate RNA provoked by Gar1 in either the wild-type or a B7_10 loop mutant RNP and with the demonstrated role of the B7_10 loop in the thermodynamics of substrate RNA binding (Supplementary Table 1).

DISCUSSION
Although box H/ACA snoRNPs have long been known to be responsible for site-specific pseudouridylation of rRNA and snRNA, and box H/ACA RNPs that efficiently modify RNA have been reconstituted, the enzymatic mechanism of this unique class of pseudouridinylsyntheses has remained elusive. Box H/ACA RNP enzymes have a complex mechanism for recruiting substrate RNA and require a number of proteins in addition to the catalytic subunit for activity. The structure of an archaeal box H/ACA RNP interacting with its substrate RNA provides important insight into the processes of substrate binding and modification.

The arrangement of active site residues around  f 5ho6 is similar to that found in the previously studied bacterial pseudouridinylsyntheses. This result directly links the catalytic mechanism of the RNP pseudouridinylsynthase to that of stand-alone pseudouridinylsyntheses. More notably, the process of substrate binding to the RNP pseudouridinylsynthase is extraordinarily complex and requires all protein and RNA components. Systematic disruption of structural elements that are directly involved in the substrate–enzyme interaction, including the conserved pseudouridine pocket structure and the B7_10 loop of Cbf5, either causes misplacement of the substrate RNA or reduces binding efficiency. These findings suggest a structural plasticity of the box H/ACA RNP assembly that is optimized for the RNA-modification process.

The architecture of the archaeal and eukaryotic H/ACA RNPs is probably similar. Generally, the eukaryotic H/ACA RNP proteins contain additional sequences; however, the elements responsible for particle assembly are conserved between the two domains. The finding that mammalian Gar1 is assembled with the other H/ACA RNP components separately and at the last stage of initial biogenesis of the complex is consistent with its subtle conformational role in substrate docking. In mammalian cells, Gar1 is the last of the four core proteins to assemble on the RNP, competing off the early-binding substrate docking. In mammalian cells, Gar1 is the last of the four core proteins to assemble on the RNP, competing off the early-binding substrate docking. In mammalian cells, Gar1 is the last of the four core proteins to assemble on the RNP, competing off the early-binding substrate docking. In mammalian cells, Gar1 is the last of the four core proteins to assemble on the RNP, competing off the early-binding substrate docking. Gar1 controls the onset of the pseudouridylation activity or the release of modified RNA substrate.

Key questions remain with regard to the catalytic mechanism of Cbf5 and all pseudouridylinylsyntheses. In the absence of cofactors or external sources of energy, pseudouridylinylsyntheses break one chemical bond, rotate the uracil ring and form a different chemical bond. With new knowledge of the active site arrangement, understanding this remarkable enzymatic acrobats is now within reach.

Accession codes. Protein Data Bank: coordinates for the Cbf5—Nop1–L7Ae and Cbf5–Nop10 bound to guide and substrate RNA have been deposited with accession codes 3HJW and 3HJY, respectively.

METHODS
Methods and any associated references are available in the online version of the paper at wWw.nature.com/nsmb/.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENT
This work was supported in part by the US National Institutes of Health grants R01GM66958–01 (H.L.) and R01 GM65682 (M.T. and K.T.). B.L. (081118E) and J.Z. (081527E) are supported by the American Heart Association, Greater Southeast Affiliate. X-ray diffraction data were collected by the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source (APS), Argonne National Laboratory. Supporting institutions for APS beamlines may be found at http://ncat.chem.cornell.edu and http://www.ser-cat.org/members.html. Use of the APS was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.

AUTHOR CONTRIBUTIONS
B.L. determined the crystal structure; J.Z. prepared samples; E.K. performed fluorescence assays; M.P.T. and R.M.T. prepared the manuscript; H.L. supervised the project and prepared the manuscript.

Published online at http://www.nature.com/nsmb
Reprints and permissions information is available online at http://npg.nature.com/ reprintsandpermissions/

ONLINE METHODS

Protein and RNA preparation. We purified Pf H/ACA proteins as described with slight modification. Briefly, we purified Pf Nop10 and Pf CB5 as a binary complex by a nickel-nitritoltriacidic acid (Ni-NTA) affinity and a gel filtration procedure. We purified Pf L7Ae separately by treating cell supernatant with polyethyleneimine and precipitating with ammonium sulfate, followed by a Ni-NTA affinity purification and a gel filtration procedure. The proteins were concentrated and stored at −80 °C before crystallization. The full-length P9_P6 composite guide RNA was transcribed in vitro using T7 RNA polymerase and purified as described. The substrate RNA with sequence 5′-GAGGAGGCC(U5)GCGGUUUAUG-3′ was purchased from Dharmacon (Thermo Fisher Scientific) and purified and stored according to manufacturer instructions.

Crystallization and diffraction data. For CNL complex crystallization, we mixed the guide and target RNA at a 1:1 molar ratio and annealed them by heating the solution for 10 min at 70 °C followed by slow cooling to room temperature (25 °C). The RNA–protein complex was formed at a 1:1.2 molar ratio with a total concentration of 23 mg ml−1. We carried out crystallization using vapor diffusion methods in hanging drops against a reservoir of 0.8 M KCl, 0.15 M magnesium acetate, 0.05 M sodium cacodylate, pH 6.5, 8% (v/v) PEG 6000. Crystals, which grew to full size (0.3 mm × 0.3 mm × 0.4 mm) at 30 °C within 1 week, were soaked briefly in a solution containing 0.8 M KCl, 0.15 M magnesium acetate, 0.05 M sodium cacodylate, pH 6.5, 8% (v/v) PEG 6000 and 5% (v/v) glycerol, followed by the same solution with 10% (v/v) glycerol, before being flash cooled in a liquid nitrogen stream for data collection. The crystals (cell parameters in Table 1) of the CNL complex contained one RNP in each asymmetric unit, with a solvent content of 61.7%.

For CN complex crystallization, the two guide strands and the target RNA were annealed at a 1:1:1 molar ratio. After mixing proteins and RNAs at a 1:1:2 ratio with a total concentration of 18 mg ml−1, the full-size crystals (0.1 mm × 0.2 mm × 0.3 mm) were obtained by vapor diffusion against a reservoir solution of 0.05M sodium cacodylate, pH 6.5, 2.0 mM CoCl2, 30 mM CaCl2, 2.0 mM spermine, 2.0 M LiCl at 30 °C within 45 d. The crystals were soaked in a mother liquor plus 2.5 M LiCl for 5–8 h before being mounted in nylon loops and flash cooled in a liquid nitrogen stream.

Diffraction data were collected at beamlines 22ID and 22BM of the Southeastern Consortium Access Team (SER-CAT) at the Advanced Photon Source (APS) and were processed using HKL2000 (ref. 53).

Phase determination and structure refinement. We determined both structures by molecular replacement methods using Molrep. For the CNL complex structure, the coordinates of CB5, Nop10 and L7Ae from the substrate-minus structure (PDB 2HYY) were used as a search model. A single and outstanding solution was found in space group P212121. The initial solution was subjected to successive rigid body, energy minimization and simulated annealing refinement using CNS. Electron density computed using the protein coordinates was improved by solvent flattening. A molecular mask generated using a manually constructed RNA-protein complex was used to perform the density modification. At this stage, most RNA nucleotides for the entire guide RNA and partial substrate RNA could be built unambiguously based into the electron density map. We carried out further refinement using CNS, REFMAC5 (ref. 56) and manual model building by O and COOT iteratively until the complete model for the RNA–protein complex could no longer be improved. At the final stage of refinement, ten cycles of translation-libration-screw motion (TLS) refinement were performed in order to restrain refinement in REFMAC5. Each individual protein and RNA molecule was treated as a single ‘rigid-body’ group, and the final TLS parameters are listed in Supplementary Table 2. The real-space correlation coefficient between the final model and the composite omit 3Fo − 2Fc map are 0.888, 0.881 and 0.776 for proteins, RNA and water/ions, respectively. The final model was refined to Rfree 24.8% and Rwork 21.7% and has an r.m.s. deviation of 0.010 Å and 1.397° from ideal bond lengths and angles, respectively. The protein residues, 92.5% lie in the most favored regions of the Ramachandran plot, 7.0% in additional allowed regions, 1 residue (Lys40) in the generously allowed region, and one residue in disallowed regions (Glu97), as similarly observed in the high-resolution structure of the CB5–Nop10–Gar1 protein complex.

For the CN subcomplex structure, the coordinates of CB5 and Nop10 (PDB 2EY4) were used as search models. A single outstanding solution was obtained in the P6_12 space group. The RNA nucleotides were built from the lower stem gradually until clear density was available for the remaining structures. Simulated annealing was carried out using only the torsion angle refinement option in CNS. Similarly to what was done for the CNL complex structure, at the final stage of the refinement, we performed ten cycles of TLS refinement in addition to restrained refinement in REFMAC5 (ref. 59). Individual protein and RNA molecules were treated as single ‘rigid-body’ groups, and the final TLS parameters are listed in Supplementary Table 2. The final model was refined to Rfree 30.6% and Rwork 27.5%.

Fluorescence studies. The fluorescence assay and data fitting have been described. The wild-type and mutant guide RNAs were transcribed and purified similarly to those used for the crystallographic studies. The 2AP-labeled and 2′-substituted substrate RNAs were purchased from Dharmacon (Thermo Fisher Scientific). Titration curves were obtained in triplicate, from which the s.d. of the Kd was computed.

Pseudouridylation assay. The P9_P6 composite wild-type and mutant guide RNAs and substrate RNAs were transcribed in vitro with T7 RNA polymerase. The same conditions were used for synthesis of uniformly labeled substrate RNAs, except that 6 μCi of [α-32P]UTP (3,000 Ci mmol−1) (MP Biomedical) was added to label all six uridines in the substrate RNA. We performed pseudouridylation assays in a similar manner to those described. Briefly, we incubated 0.2 mM of [α-32P]-labeled substrate RNA, 1.2 mM guide RNA and 3 μM of indicated protein components in the reaction buffer containing 100 mM Tris-HCl, pH 8.0, 100 mM ammonium acetate, 5 mM MgCl2, 2 mM DTT, 0.1 mM EDTA for 1 h at 70 °C. The RNAs were extracted by phenol/chloroform/isoamyl alcohol (pH 4.5), purified by ethanol precipitation and digested with nuclease P1 (1 unit, United States Biological). The resulting 5′-mononucleotides were separated via thin-layer chromatography, as described.