Molecular Cell *21*, 249–260, January 20, 2006 ©2006 Elsevier Inc. DOI 10.1016/j.molcel.2005.11.017

Crystal Structure of a Cbf5-Nop10-Gar1 Complex and Implications in RNA-Guided Pseudouridylation and Dyskeratosis Congenita

Rumana Rashid,[1] Bo Liang,[1] Daniel L. Baker,[2] Osama A. Youssef,[2] Yang He,[1] Kathleen Phipps,[1] Rebecca M. Terns,[2] Michael P. Terns,[2] and Hong Li[1,*]
[1] Department of Chemistry and Biochemistry
Institute of Molecular Biophysics
Florida State University
Tallahassee, Florida 32306
[2] Department of Biochemistry and Molecular Biology
Department of Genetics
University of Georgia
Athens, Georgia 30602

Summary

H/ACA RNA-protein complexes, comprised of four proteins and an H/ACA guide RNA, modify ribosomal and small nuclear RNAs. The H/ACA proteins are also essential components of telomerase in mammals. Cbf5 is the H/ACA protein that catalyzes isomerization of uridine to pseudouridine in target RNAs. Mutations in human Cbf5 (dyskerin) lead to dyskeratosis congenita. Here, we describe the 2.1 Å crystal structure of a specific complex of three archaeal H/ACA proteins, Cbf5, Nop10, and Gar1. Cbf5 displays structural properties that are unique among known pseudouridine synthases and are consistent with its distinct function in RNA-guided pseudouridylation. We also describe the previously unknown structures of both Nop10 and Gar1 and the structural basis for their essential roles in pseudouridylation. By using information from related structures, we have modeled the entire ribonucleoprotein complex including both guide and substrate RNAs. We have also identified a dyskeratosis congenita mutation cluster site within a modeled dyskerin structure.

Introduction

Box H/ACA RNA-protein complexes (H/ACA RNPs) pseudouridylate ribosomal and small nuclear RNAs (Decatur and Fournier, 2003; Kiss, 2002; Meier, 2005; Terns and Terns, 2002; Yu et al., 2005). Pseudouridine, a rotation isomer of uridine (ψ), is often found in highly conserved regions of functional RNAs such as tRNA, rRNA, and snRNA. Pseudouridylation of RNA increases thermal stability of RNA helices (Arnez and Steitz, 1994; Davis, 1995; Newby and Greenbaum, 2002a; Yarian et al., 1999) and extends the repertoire of RNA bases that can interact to form new RNA structure and protein binding sites (Newby and Greenbaum, 2002b; Yu et al., 2005). The functional importance of RNA pseudouridylation has been demonstrated in ribosome function (King et al., 2003), spliceosome function (Donmez et al., 2004; Valadkhan and Manley, 2003; Yang et al., 2005a; Zhao and Yu, 2004), and small nuclear ribonucleoprotein assembly (Yu et al., 1998).

*Correspondence: hongli@sb.fsu.edu

H/ACA RNPs are RNA-protein assemblies that achieve RNA pseudouridylation by an RNA-guided mechanism. Specifically, the RNA component of each H/ACA RNP contains regions of complementarity to a target RNA, which confer site-specific modification (Balakin et al., 1996; Ganot et al., 1997a; Ganot et al., 1997b; Ni et al., 1997). The complementarity is distributed across two strands of an internal loop of the H/ACA RNA termed the pseudouridylation pocket. A conserved sequence element termed box ACA or box H (ANANNA) is located at the base of the hairpin housing the pseudouridylation pocket and is essential for association of the RNAs with a common set of proteins (Meier, 2005; Yu et al., 2005). The core H/ACA RNP proteins include Cbf5p (Dyskerin in human, NAP57 in rodents), Nop10p, Nhp2p, and Gar1p (Watkins et al., 1998; Bousquet-Antonelli et al., 1997; Dragon et al., 2000; Henras et al., 1998; Lafontaine et al., 1998; Wang and Meier, 2004). Archaeal homologs of all four core proteins (Cbf5, Nop10, L7Ae, and Gar1, respectively) as well as H/ACA RNAs, have been identified (Henras et al., 2004; Rozhdestvensky et al., 2003; Tang et al., 2002; Watanabe and Gray, 2000) and have been used to reconstitute RNA-guided pseudouridylation in vitro (Baker et al., 2005; Charpentier et al., 2005).

Pseudouridine is the most abundant single nucleotide modification detected in a diverse range of functional RNAs (Charette and Gray, 2000; Grosjean and Benne, 1998). Correspondingly, the families of enzymes that catalyze isomerization of uridine exhibit extraordinarily conserved active sites but a wide range of RNA binding strategies (del Campo et al., 2004; Foster et al., 2000; Hoang and Ferre-D' Amare, 2004; Kaya et al., 2004; Pan et al., 2003; Sivaraman et al., 2002). Cbf5 exhibits strong sequence homology to the TruB family of pseudouridine synthases (Koonin, 1996) that catalyze isomerization of uridine 55 in most tRNAs and, thus, is predicted to be the catalytic subunit of the H/ACA RNP. However, unlike all other known pseudouridine synthases, which directly recognize and modify their cognate RNA target sequences, H/ACA RNP-associated Cbf5 requires a guide RNA and three accessory proteins to efficiently bind and modify its substrate (Baker et al., 2005; Charpentier et al., 2005). The archaeal Cbf5 protein interacts independently with the H/ACA guide RNA, Nop10 and Gar1 (Baker et al., 2005; Charpentier et al., 2005). L7Ae interacts directly with the guide RNA and does not interact with the other proteins in the absence of the guide RNA (Baker et al., 2005; Charpentier et al., 2005). All four proteins and the guide RNA are necessary for efficient modification of the substrate by Cbf5 (Baker et al., 2005; Charpentier et al., 2005). These studies indicate that Nop10, Gar1, and L7Ae (NHP2) influence catalysis, perhaps by modulating target RNA-guide RNA interaction or positioning the target RNA within the catalytic site of Cbf5.

The H/ACA proteins are also essential components of telomerase. Telomerase is an RNA-protein enzyme that catalyzes telomere addition and is responsible for maintenance of chromosome termini (Blackburn, 2005). Mammalian telomerase contains an H/ACA motif in its

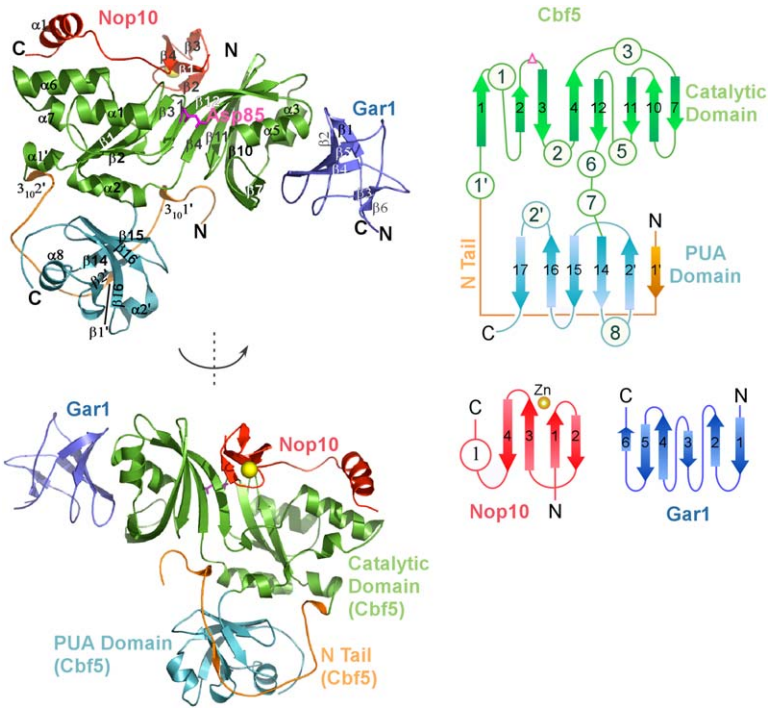

Figure 1. Overview of the Cbf5-Nop10-Gar1 Complex Structure in Two Orthogonal Views and Secondary Structure Descriptions of the Individual Proteins

Color designation is as follows: Cbf5, green (catalytic domain); turquoise, PUA domain; and orange, N tail; Nop10, red; Gar1, blue. Shown as a stick model is the conserved catalytic residue Asp85 of Cbf5. In two-dimensional (2D) topology diagrams, arrows represent β strands and circles represent helices. Note that conserved Cbf5 secondary structure elements have the same numbering as those of TruB and unique secondary structure elements are numbered independently with primes.

RNA component (Chen and Greider, 2004; Mitchell et al., 1999a) and is bound by all four core H/ACA proteins (Dragon et al., 2000; Mitchell et al., 1999b; Pogacic et al., 2000; Wang and Meier, 2004). Although telomerase is not thought to function in pseudouridylation, association of the H/ACA proteins with telomerase is important for the processing, stability, and trafficking of telomerase RNA in vivo (Chen and Greider, 2004; Lukowiak et al., 2001; Mitchell et al., 1999b).

Mutations in the human Cbf5 gene, DKC1, cause X-linked dyskeratosis congenita (DC) that is characterized by abnormal skin pigmentation and nail dystrophy, bone marrow failure, and a predisposition to epithelial cancers (Heiss et al., 1998; Marrone and Mason, 2003; Meier, 2005). The molecular basis for DC is not yet clear. Both dysfunction of ribosomal RNA (Mochizuki et al., 2004; Ruggero et al., 2003) and shortening of telomeres (Mitchell et al., 1999b; Mochizuki et al., 2004) have been observed in DC patients, suggesting that DKC1 mutations can affect both ribosomal RNA biogenesis and telomerase function.

Here, we describe the crystal structure of the *Pyrococcus furiosus* (Pf) Cbf5-Nop10-Gar1 complex at 2.1 Å resolution. The trimeric complex provides insights into H/ACA RNP assembly and the RNA-guided pseudouridylation mechanism. In addition, knowledge of the previously determined structure of a Cbf5 homolog (TruB) bound to RNA allowed us to model the interaction of the Cbf5-Nop10-Gar1 complex with an H/ACA guide RNA-target RNA duplex and to suggest possible functional roles for each H/ACA protein. Finally, our observation that DC-causing mutations, which are generally concentrated at the 5′ and 3′ termini of the DKC1 gene, affect amino acids that cluster to a specific location within the folded protein indicates the importance of this region in normal dyskerin function and identifies the functional domain that is disrupted in most DC patients.

Results and Discussion

Overview: Structure of Cbf5, Nop10, and Gar1 and Organization of the Complex

The Cbf5-Nop10-Gar1 complex was obtained by copurification of the proteins after expression in *E. coli*. Cbf5 and histidine-tagged Gar1 were coexpressed, whereas Nop10 was expressed alone. Copurified Cbf5-Nop10-Gar1 complexes bind specifically to H/ACA guide RNAs and, together with the fourth protein L7Ae, pseudouridylate target rRNA in vitro (Baker et al., 2005). The crystal structure of Cbf5-Gar1-Nop10 was determined by the multiple-wavelength anomalous diffraction method with selenomethionine-containing crystals. The final structure includes Cbf5 residues 8–336 (full-length 1–343), Nop10 residues 4–55 (full-length 1–60), and Gar1 residues 1–73 (full-length 1–97).

The Cbf5-Nop10-Gar1 complex crystallized with two trimers in each asymmetric unit. The C-terminal domains of each Cbf5 protein bury a moderate solvent-accessible surface area (1048 Å^2), suggesting the possibility of weak dimerization between Cbf5 molecules. To further assess the oligomeric state of the heterotrimer in solution, we carried out analytical ultracentrifugation experiments. Results from sedimentation velocity (SV) centrifugation with heterotrimers at concentrations as high as 172 μM revealed a species that was predominately of ~3.4 Svedberg in solution (see the Supplemental Data available with this article online), indicating that the complex exists in monomeric units in solution.

The overall structure of the Cbf5-Nop10-Gar1 complex is shown in Figure 1. There was no previous high-resolution structural information for these three H/ACA proteins from any organism. As expected, Cbf5 exhibits overall structural homology to the *E. coli* pseudouridine synthase TruB (Figure 2, Protein Data Base [PDB] number 1K8W, rmsd 1.3 Å for 191 Cα atoms; PDB number

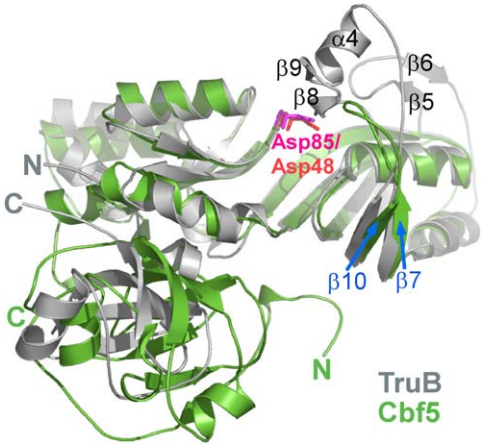

Figure 2. Structural Homology between Pf Cbf5, in Green, and *E. coli* TruB, in Gray

Secondary structure elements (comprising the thumb loop) present in TruB, but not in Pf Cbf5, are labeled in black. β7 and β10 are labeled in blue. The catalytic aspartate residues in both structures are shown as stick models (Pf Cbf5 magenta and TruB red).

1R3E, rmsd 1.4 Å for 184 Cα atoms). Cbf5 shares the two major structural domains of TruB: the catalytic domain (Cbf5 residues 43–244) and the *P*seudo*U*ridine synthase and *A*rchaeosine transglycosylase (PUA) domain (Cbf5 residues 245–336) (Figure 1, green and turquoise, respectively) (Aravind and Koonin, 1999). In addition, Cbf5 has an N-terminal tail (residues 8–42) that is not present in TruB (Figure 1, orange). An aspartate found in the catalytic domain is conserved among pseudouridine synthases and is involved in catalysis (Charpentier et al., 2005; Zebarjadian et al., 1999). Mutation of the conserved Asp85 in Pf Cbf5 eliminates in vitro pseudouridylation activity (D.L.B., R.M.T., and M.P.T., unpublished data). Asp85 of Pf Cbf5 superimposes closely in space with catalytic Asp48 of *E. coli* TruB (Figure 2).

The small Nop10 protein contains two distinct domains: an N-terminal, all β domain comprised of two closely packed β hairpins, and a C-terminal, single-helix domain (Figure 1). The domains are connected by a 10 amino acid random coil. Within the β barrel domain, four cysteines, Cys8, Cys11, Cys20, and Cys23, form a structural zinc binding site (Figure 1) (Auld, 2001) exemplified as that observed in aspartate carbamoyltransferase (Gouaux et al., 1990) (rmsd 1.1 Å for all atoms of the four cysteine residues). The four zinc ligands are highly conserved in Archaea (Cys11 is occasionally replaced by aspartate) and, thus, likely play an important role in the stability of archaeal Nop10 structure.

Gar1 forms a six-stranded β barrel structure that belongs to the reductase/isomerase/elongation factor fold superfamily (see Structural Classification of Proteins, http://scop.mrc-lmb.cam.ac.uk/scop/ (Figures 1 and 3). This fold occurs in domain 2 of the elongation factor Tu (PDB number 1EFC, Z = 5.3; for pairs of protein structures in which Z < 2.0, the pair is structurally dissimilar), in the N-terminal domain of the F1 ATPase α subunit (PDB number 1SKY, Z = 6.8), in the T-protein of the glycine cleavage system (PDB number 1WOO, Z = 6.0), in ribosomal protein L35a (PDB number 1SQR, Z = 5.1), and in translation initiation factor 2 (PDB number

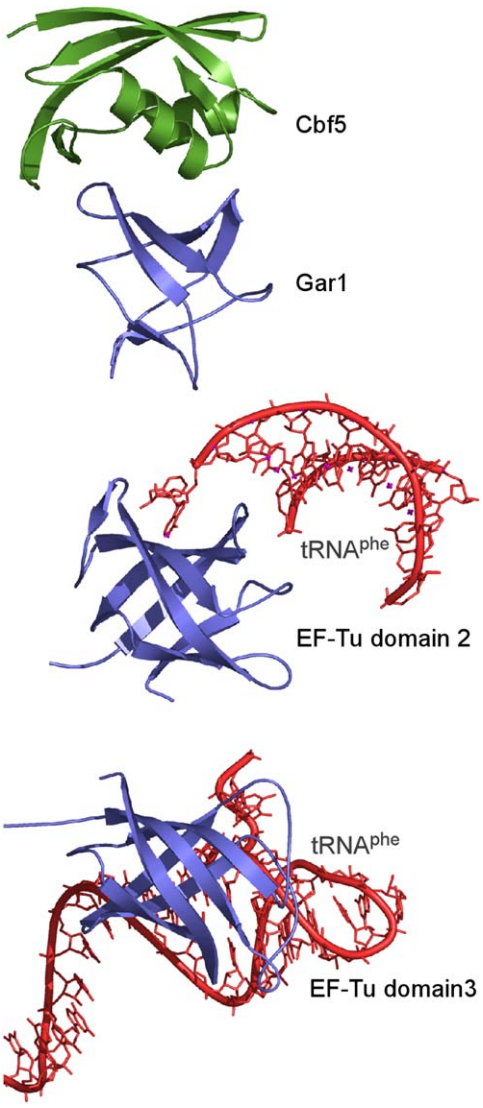

Figure 3. Gar1 Employs a Potential RNA Binding Surface to Interact with Cbf5

The complex of Gar1 (blue) and Cbf5 (green) is compared with those of EF-Tu domains (blue) and tRNA (red). The structures of Gar1, EF-Tu domain 2, and EF-Tu domain 3 are shown in the same orientation.

1G7R, Z = 5.0). In many cases, this structural fold is involved in binding RNA (e.g., in translation factors). However, it can also be a modular domain to mediate protein oligomerization (e.g., in F1 ATPase).

The organization of the Cbf5-Nop10-Gar1 heterotrimer confirms previous biochemical evidence that Nop10 and Gar1 interact independently with Cbf5 (Baker et al., 2005; Charpentier et al., 2005; Henras et al., 2004). Viewed from the top (Figure 1), Nop10 and Gar1 surround the catalytic domain of Cbf5, which measures ~58 Å across and ~28 Å in height, with Nop10 bordering to the north and Gar1 to the east. The two accessory proteins are separated by nearly 20 Å. The catalytic domain of Cbf5 is thus also the principal organizer of the three-protein complex.

The binding of Nop10 to Cbf5 is mediated by interactions along the length of Nop10, including residues in

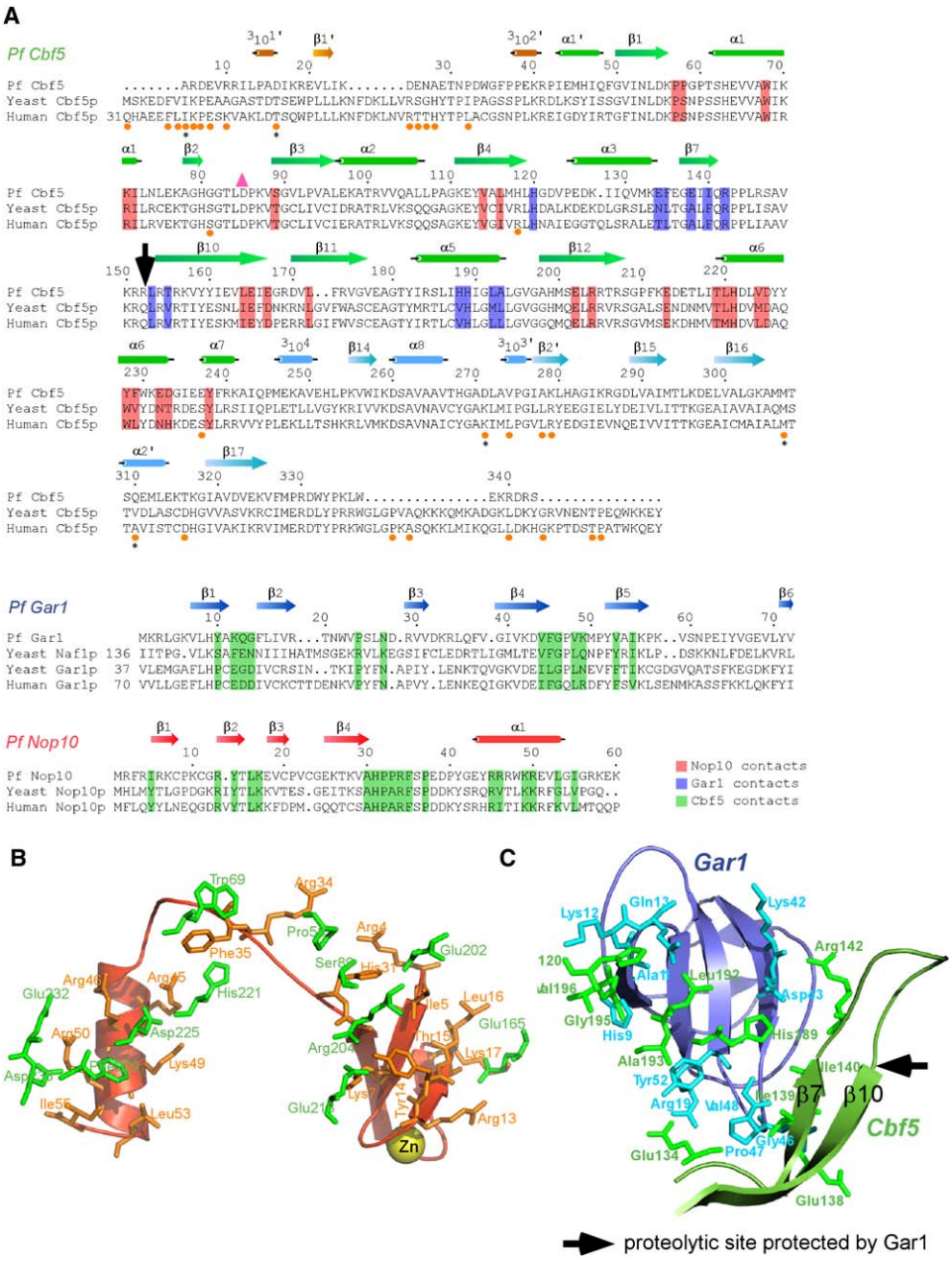

Figure 4. Residues Involved in the Interactions of Cbf5 with Nop10 and Gar1

(A) Aligned sequences of Pf, yeast, and human Cbf5 proteins overlaid with Pf Cbf5 secondary structures. Residues involved in interactions (defined by a decrease of more than 4 Å^2 on side chain solvent-accessible area due to binding of another molecule) are highlighted as follows: Cbf5 residues that interact with Nop10 in red and that interact with Gar1 in blue, and Nop10 and Gar1 residues that interact with Cbf5 are shown in green. The black arrow on the Cbf5 sequence indicates the proteolytic site of Cbf5 protected by Gar1 (see Figure S3). Dyskerin mutations are indicated under the dyskerin sequence by gold dots. Asterisks under gold dots indicate mutations that occur in multiple families (from Marrone et al., 2005).

(B) Specific contacts between Nop10 (red) and Cbf5 (green).

(C) Specific contacts between Gar1 (blue) and Cbf5 (green).

both the β barrel and helix domains, and the connecting coil (Figures 1, 4A, and 4B). The α helix of Nop10 forms a coiled-coil interaction with α6 of Cbf5. The β barrel domain of Nop10 packs against the central β sheet formed by β10-β12 and, to some extent, β3 and β4 of Cbf5. The interactions bury an extensive solvent-accessible surface area (2834 Å^2 total) and are 43% polar and electrostatic interactions and 57% nonpolar interactions.

Gar1 interacts with Cbf5 primarily via two β hairpin loops formed by β1/β2 and β4/5 of Gar1 (Figures 1, 4A, and 4C). Gar1 packs the edge of Cbf5 formed by β7, α5, and α3. The association of Gar1 and Cbf5 results in a modest buried solvent-accessible surface (1583 Å^2 total) but involves substantial nonpolar interactions (63%). Interestingly, the equivalent surface of the structurally homologous domain 2 of EF-Tu is involved in RNA

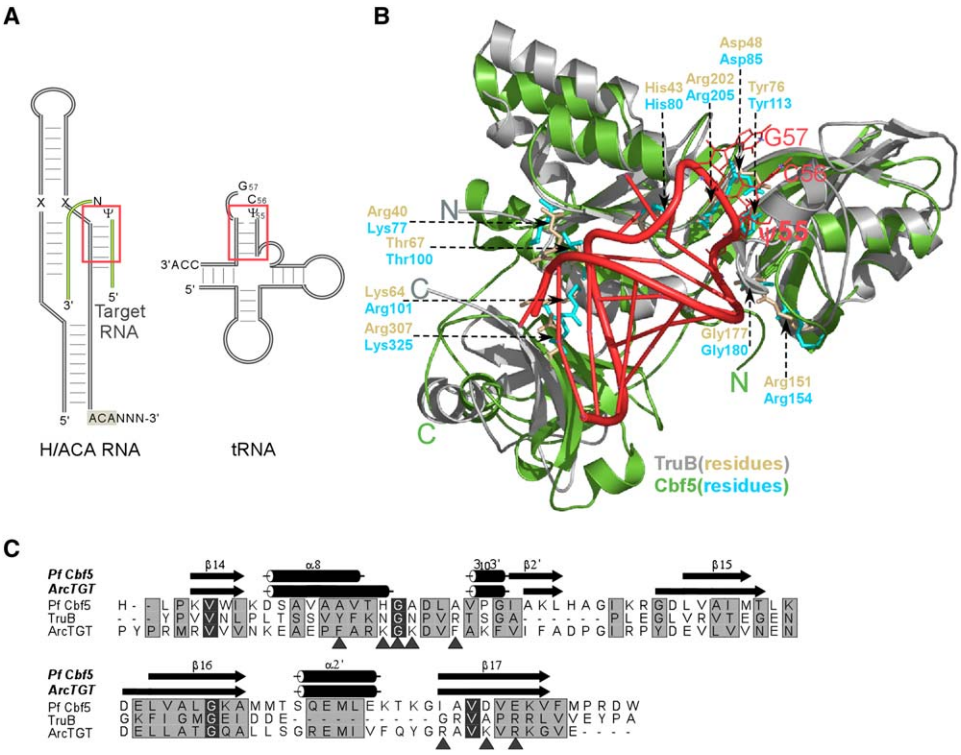

Figure 5. Comparison of Cbf5 to Regions of TruB and ArcTGT Involved in RNA Binding

(A) Comparison of the secondary structure of Cbf5 and TruB RNA substrates. H/ACA and tRNA are outlined in gray. The substrate rRNA is shown in yellow. Red boxes highlight the shared target uridine and the preceding helix in both RNAs.

(B) Cbf5 is superimposed (via catalytic domain only) with a TruB-tRNA complex (PDB number 1K8W). Colored protein residues represent conserved residues that contact the target uridine and the preceding helix. TruB residues are in wheat and Cbf5 residues are in cyan. The three bases looped out of the T stem are shown as stick models (C). Structure-assisted sequence alignment of PUA domains of Pf Cbf5, TruB, and ArcTGT overlaid with secondary structure elements of Cbf5 (top) and of ArcTGT (bottom). Solid triangles mark ArcTGT residues that specifically contact the tRNA terminal CCA trinucleotides.

binding. We superimposed EF-Tu domain 2 (rmsd 1.8 Å for 27 Cα atoms) and its bound RNA with Gar1. This comparison shows that whereas the β barrel surface of EF-Tu domain 2 forms a specific pocket for the terminal A nucleotide of tRNA, the equivalent binding surface of Gar1 is involved in contacting Cbf5 (Figure 3). Thus, Gar1 employs this fold for interaction with protein rather than with RNA. EF-Tu domain 3, which interacts with the T stem of tRNA through a set of nonspecific contacts, bears similarity to Gar1 as well (rmsd 1.9 Å for 27 Cα atoms). We superimposed EF-Tu domain 3 with Gar1. Although the equivalent surface of Gar1 is not involved in interaction with Cbf5 (Figure 3), the EF-Tu domain 3 residues involved in interaction with RNA are not conserved in Gar1 (data not shown).

The residues involved in the interactions of Nop10 and Gar1 with Cbf5 are generally conserved in eukaryotic homologs (Figure 4A), indicating that the mechanism of association of the Cbf5-Nop10-Gar1 complex will be similar in eukaryotes.

Mechanism for Interaction of Cbf5 with the Target Uridine and Adjacent Stem

E. coli TruB catalyzes the pseudouridylation of U55 on various tRNAs in the absence of guide RNAs or other proteins (Gutgsell et al., 2000). As noted by Hoang and Ferre-D' Amare, the RNA structural contexts of the uridines targeted by Cbf5 and TruB (within an rRNA-H/ACA guide RNA duplex or tRNA, respectively) possess some interesting similarities (Hoang and Ferre-D' Amare, 2001). In both cases, the target uridine is unpaired and located adjacent to a helical stem (Figure 5A). At the same time, however, the uridine isomerized by Cbf5 is present at the center of a three-way junction, whereas the tRNA uridine 55 isomerized by TruB is located in a loop (the T loop) at the end of the stem (the T stem). tRNA uridine 55 is followed by highly conserved cytosine 56.

We coanalyzed the structure of Cbf5 and the previously described structure of TruB with bound RNA (Figure 5B) and found that the Cbf5 protein shares a largely conserved RNA binding pocket with TruB but differs from TruB in regions involved in binding T loop-specific nucleotides. TruB residues involved in binding the target uridine and the adjacent stem (Ec TruB Arg40, His43, Asp48, Thr63, Lys64, Asn67, Tyr76, Arg151, Gly177, Arg202, and Arg307) are strictly conserved in Cbf5, except for Arg40, Lys64, and Arg307 (which are Lys77, Arg101, and Lys325, respectively, in Pf Cbf5) (Figure 5B). TruB His43, which is conserved in Cbf5, was proposed to play a central role in nucleotide flipping (Hoang and Ferre-D' Amare, 2001), and mutational studies support its role in catalysis (Hamilton et al., 2005). Arg202 and Tyr76 interact with the phosphate group of the target

uridine, and Arg307 interacts with the T stem (Hoang and Ferre-D' Amare, 2001). The conserved residues, which are located primarily in the catalytic domain, can be superimposed closely in space within the structures of the two proteins (Figure 5B), indicating that TruB and Cbf5 use a common mechanism to interact with the target uridine and adjacent stem.

Outside the regions involved in interacting with the target uridine and the adjacent helix, the differences in RNA recognition elements between TruB and Cbf5 are evident. Cbf5 lacks the β5/β6 hairpin, α4 helix, and short β8/β9 hairpin that comprise the "thumb loop" domain of TruB (Figures 2 and 5B). These secondary structure elements are involved in binding cytosine 56 and the rest of the T loop nucleotides in TruB (Hoang and Ferre-D' Amare, 2001). The absence of the thumb loop (β5/β6 hairpin, α4 helix, and β8/β9 hairpin) leads to a wide pocket directly above the target uridine binding site in Cbf5. This structural feature of Cbf5 could facilitate the binding of a bulkier guide RNA-target RNA complex. There are likely other elements within Cbf5 and the Cbf5-Nop10-Gar1 complex that play roles in binding specific elements of the guide and target RNAs. Although Cbf5 lacks β8 and β9, which are responsible for anchoring the tRNA T loop in TruB, the β7/β10 hairpin loop of Cbf5 is well conserved and sufficiently long to contact the RNA segment preceding the target uridine (Figures 2 and 5B). The β7/β10 hairpin very likely plays a pivotal role in binding the guide RNA-target RNA complex.

Role of the PUA Domain of Cbf5 in RNA Binding

Previous sequence analysis identified similarities between the sequences of the PUA domains of Cbf5 and other RNA processing enzymes (Koonin et al., 1994). Indeed, we find that the structure of Cbf5's PUA domain also exhibits homology to that of TruB (PDB number 1ZE1, Z = 10.4), a human homolog of yeast rRNA processing protein Nip7p (PDB number 1T5Y, Z = 10.6), and archaeosine transglycosylase (ArcTGT) (PDB number 1IQ8, Z = 11.1). Interestingly, the similarity between the PUA domains of Cbf5 and ArcTGT is even greater than that of Cbf5 and TruB.

The PUA domain of Cbf5 likely anchors the RNA helix preceding the target uridine in a manner similar to TruB (Hoang and Ferre-D' Amare, 2001; Pan et al., 2003). Accordingly, Lys325 in the PUA domain of Cbf5 is conserved both in sequence and in location for interaction with the RNA helix preceding the target uridine (Figure 5B). However, the PUA domain is significantly closer to the catalytic domain in Cbf5 than in TruB (Figures 2 and 5B). This may be brought about by a unique feature in the structure of Cbf5. The N terminus of Cbf5 encircles the entire PUA domain (formed by the C terminus of the protein) and contributes an additional β strand (β1′) to the Cbf5 PUA domain β sheet (Figure 1). The interaction of the N terminus with the C terminus buries an extensive solvent-accessible surface (2116 $Å^2$) and is mediated by residues that are well conserved among Cbf5 proteins, in particular Thr30, Gly35, and Pro38 (Figure 4A). The encirclement of the PUA domain of Cbf5 by the N-terminal tail limits its movement. In contrast, the position of the TruB PUA domain appears to change upon RNA binding (Pan et al., 2003). Tight coupling between the PUA and the catalytic domains in Cbf5 could be important for the folding stability of the catalytic domain and could narrow its specificity for guide RNA. This proposal is supported by the fact that the isolated catalytic domain of Cbf5 (residues 4–247 or 38–247) is completely insoluble (data not shown), whereas the PUA domain (comprised of residues 4–38 and 247–343) is highly soluble (see below).

Archaeosine transglycosylase catalyzes the first step in the conversion of tRNA G15 to archaeosine in archaea. The cocrystal structure of ArcTGT and its RNA substrate revealed that the ArcTGT PUA domain interacts specifically with the terminal CCA of the tRNA via both electrostatic and nucleotide-amino acid stacking forces (Ishitani et al., 2003). The strong structural homology between the Cbf5 and ArcTGT PUA domains (rmsd 1.2 Å for 73 Cα atoms) and potential similarities in the bound RNAs (3′ CCA and 3′ ACA) suggest a possible link in their RNA binding mechanisms. We further compared the amino acids of ArcTGT involved in binding CCA to their structurally equivalent residues within Cbf5 and found no physicochemical similarity among them (Figure 5C). This observation indicates that residues unique to the PUA domain of Cbf5 are likely to be responsible for its sequence-specific interaction with the guide RNA.

Cbf5 Stabilizes the Bipartite Binding of Nop10

Recent biochemical studies indicate that Nop10 and Gar1 are not required for the interaction of Cbf5 with a guide RNA but that both proteins are essential for pseudouridylation and that Nop10 may play a role in binding the target RNA (Baker et al., 2005; Charpentier et al., 2005). The interactions between Nop10 and Cbf5 observed in the structure of the Cbf5-Nop10-Gar1 complex suggest two mechanisms by which Nop10 may function in RNA-guided pseudouridylation. First, Nop10 directly contacts residues surrounding the target uridine binding pocket on Cbf5. Two conserved Nop10 residues, Tyr14 and His31, form hydrogen bonds with Arg204 and Glu202 on β12 of Cbf5 (Figures 4A and 4B). On the opposite face of Cbf5 β12, Arg205 is predicted to stabilize the phosphate group of the bound target uridine (Figure 5B). Thus, Nop10 may indirectly stabilize binding of the target uridine through a network of hydrogen bonds.

Second, the bipartite structure of Nop10 and its interaction with Cbf5 suggest that Nop10 may undergo significant conformational change or stabilization upon interaction with Cbf5. Such changes could modulate the ability of Nop10 to interact with RNA or other proteins. The N-terminal β barrel domain, the C-terminal helix domain, and the connecting loop of Nop10 all pack closely onto Cbf5 through complementary electrostatic and hydrophobic interactions (Figures 4A and 4B). Whereas the two domains of Nop10 both interact with Cbf5, they do not contact each other. This lack of intramolecular interaction within Nop10 is a strong indication that the observed structural fold of Nop10 is stabilized upon association with Cbf5. Nop10 may be an intrinsically disordered protein (Dyson and Wright, 2005). The extensive contacts made between Nop10 and Cbf5 (Figure 4B) presumably overcome the negative entropy change derived from the intrinsic flexibility in Nop10.

These observations suggest to us that Nop10 may act as a molecular switch that acquires additional function upon interaction with Cbf5.

The remarkable conformational dependence of Nop10 on Cbf5 may play a role in the ordered assembly of the H/ACA RNP. For example, whereas in archaea the fourth H/ACA protein L7Ae does not interact independently with the other H/ACA proteins, the eukaryotic homolog NHP2 does interact with NOP10, and this interaction requires prior interaction of NOP10 with the Cbf5 homolog (Wang and Meier, 2004). In addition, the interaction of Cbf5 with target RNA appears to depend on Nop10 (Charpentier et al., 2005). Although the majority of Nop10 rests below the base of the catalytic domain of Cbf5, the connecting loop and N-terminal end of the helix domain rise above and could extend the peripheral RNA binding pocket of Cbf5. Strong conservation of the Nop10 connecting loop sequence supports its potential role in RNA or protein interaction (Figure 4A). We propose that binding of Nop10 to Cbf5 stabilizes a Nop10 conformation required for interaction with substrate RNA and, in the case of mammals, with NHP2.

Role of Gar1 in Assembly and Implications in Naf1p-Mediated H/ACA RNP Biogenesis

The structural homology between Gar1 and protein domains involved in interaction with RNA suggested that Gar1 may be involved in direct RNA binding in the H/ACA RNP. However, as discussed above, in the case of Gar1, the surface of Gar1 that corresponds to the surface of EF-Tu domain 2 that mediates RNA binding is used for interaction with Cbf5 (Figure 3). Furthermore, although the exposed surface of Gar1 is oriented similarly to the RNA binding surface of EF-Tu domain 3 (Figure 3), Gar1 lacks the identified RNA binding residues.

A previous study found that mammalian Gar1 could be crosslinked to the target uridine (Wang and Meier, 2004). Given the location of Gar1 within the structure of the complex and the well-established position of the bound target uridine, our structure cannot readily accommodate this piece of important biochemical data. The observed uridine crosslink could be mediated by contact with the N- and C-terminal glycine/arginine-rich (i.e., GAR) domains that are found in the eukaryotic, but not archaeal, Gar1 proteins (Girard et al., 1992, Watanabe and Gray, 2000). Alternatively, Gar1 may transiently associate with the target uridine at some point between substrate RNA binding and product release.

Like Nop10, Gar1 contacts Cbf5 in regions that are critical for target RNA binding. Interestingly, Gar1 specifically interacts with the β7/β10 hairpin of Cbf5. As discussed above, the Cbf5 β7/β10 hairpin is likely involved in interaction with the guide RNA-target RNA complex. The constraint imposed by Gar1 on β7/β10 may stabilize the association of the target RNA with Cbf5. To test Gar1's predicted role in stabilizing the β7/β10 hairpin of Cbf5, we carried out partial proteolysis of Cbf5 in the presence and absence of Gar1. We subjected co-purified complexes of Cbf5-Nop10-Gar1 and Cbf5-Nop10 to limited trypsin digestion and analyzed the products by sodium dodecyl-sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and by N-terminal sequencing (Figure S3). Cbf5 was specifically cleaved into two discrete fragments only in the absence of Gar1, indicating that the bound Gar1 protected this proteolytic site (Figure S3). N-terminal sequencing revealed that the site of cleavage is located within the loop β7/β10 after residue Arg154 (Figure 4). This analysis confirms the binding interface of Cbf5 and Gar1 observed in our high-resolution crystal structure and supports the proposed function of Gar1 in stabilizing a potential RNA binding element of Cbf5.

In addition, it appears that the interaction of Gar1 with Cbf5 could play a specific role in biogenesis of the mature eukaryotic H/ACA RNP complexes. In yeast, the assembly of H/ACA RNPs involves Naf1p (Fatica et al., 2002; Yang et al., 2002). It is known that Naf1p interacts with Cbf5p and that the Naf1p-Cbf5p complex is recruited to the H/ACA RNPs cotranscriptionally at the H/ACA RNA gene (Yang et al., 2005b). Naf1p contains a region of ~58 amino acids that is homologous to Gar1p. We have found that the regions of Gar1 involved in contacting Cbf5 are highly conserved in Naf1p as well, particularly the β4/β5 hairpin (Figure 4A). This result supports the hypothesis that Naf1p and Gar1 occupy the same Cbf5 binding site at different stages of H/ACA RNP assembly.

Structural Model of a Functional H/ACA RNP

To gain further insight into the functional organization of the H/ACA RNP, we constructed a three-dimensional model of a complete RNA-guided pseudouridylation complex (Figure 6A). The model is based on the predicted secondary structure of a functional Pf H/ACA RNA (Baker et al., 2005) and the previously described TruB-tRNA cocrystal structure (Hoang and Ferre-D'Amare, 2001; Pan et al., 2003). A target-associated H/ACA RNA contains a three-way helical junction (Figure 5A). Guided by the three-way junction structure found in a hammerhead ribozyme (Scott et al., 1995), we manually constructed a model for the three-way junction formed by the H/ACA-guide RNA and rRNA target RNA. A K-turn motif was included within the upper stem of the guide RNA, and a lower stem was added. L7Ae was placed on the K-turn motif by using a previously determined L7Ae-K-turn RNA cocrystal structure (Moore et al., 2004). The entire model was subjected to several runs of energy minimization with the Crystallography and NMR System (CNS) (Brunger et al., 1998) to ensure reasonable stereochemical geometry. We placed the entire modeled RNA structure onto the experimentally determined Cbf5-Nop10-Gar1 complex structure, guided by the TruB-tRNA structure (Figure 6A). The target uridine and the helix formed by base-pairing of the right side of the guide RNA pseudouridylation pocket with the rRNA target are superimposed with the tRNA uridine 55 and T stem, respectively (Figure 6A).

Our structural model of an H/ACA RNP is consistent with the previous finding that archaeal Cbf5 depends on the presence of both the pseudouridylation pocket and the ACA sequence element for binding to H/ACA RNAs (Baker et al., 2005). In the model, the pseudouridylation pocket contacts the major RNA binding surface within the catalytic domain of Cbf5, and the ACA trinucleotide is located proximal to the PUA domain. Though not explicitly depicted in the model, we envision a direct interaction between the PUA domain of Cbf5 and the

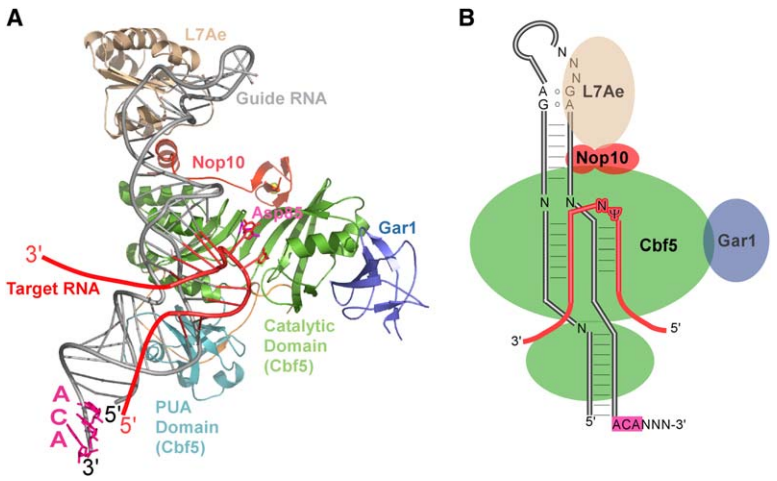

Figure 6. Structural Model of Fully Assembled H/ACA RNP

(A) A modeled H/ACA RNA-target complex is placed in the Cbf5-Nop10-Gar1 complex structure. The same coloring scheme is used as in Figure 1 with the addition of L7Ae (wheat), guide RNA (gray), and target RNA (red). The catalytic residue Asp85 of Cbf5, the target uridine, its immediate 3′ nucleotide (which mimicks the location of C56 in tRNA), and the conserved ACA motif of the guide RNA are shown as stick models.

(B) A 2D drawing of the three-dimensional model in which Nop10 and L7Ae facilitate target RNA binding by anchoring the upper stem of H/ACA guide RNA.

ACA element, perhaps by twisting of the ACA trinucleotide and/or via amino acids at the C terminus that are missing from our crystal structure. Moreover, our RNA-protein interaction model not only provides a structural basis for the observed cooperativity between the two RNA elements but also predicts cooperativity between the PUA and catalytic domains of Cbf5 with regard to RNA binding. It follows that elimination of either protein domain would abolish (or substantially reduce) the binding of Cbf5 to the guide RNA. In agreement with this prediction, we have found that the PUA domain (residues 4–38 and 247–343) does not interact with H/ACA RNA in gel shift assays at a wide range of protein concentrations (data not shown).

Our model further provides a potential explanation for the critical importance of Nop10 and L7Ae in assembly of a functional RNP. The available biochemical data indicates that Nop10 does not interact with the H/ACA RNA in the absence of Cbf5 but that it is essential for catalysis and likely facilitates binding of target RNA (Baker et al., 2005; Charpentier et al., 2005). In our model, Nop10 is located near the upper stem of the guide RNA. The upper stem also is the location of the K-turn motif where L7Ae binds the guide RNA (Baker et al., 2005; Charpentier et al., 2005; Rozhdestvensky et al., 2003), placing L7Ae close to (perhaps in contact with) Nop10 in the fully assembled H/ACA RNP. Although it is not known how (or if) Nop10 interacts with RNAs, L7Ae-induced bending of the K-turn is well established (Hamma and Ferre-D'Amare, 2004; Moore et al., 2004; Suryadi et al., 2005; Turner et al., 2005). It is possible that L7Ae and Nop10 work in concert to induce a large conformational change in the guide RNA, which facilitates the binding of target RNA. Interestingly, the eukaryotic homolog of L7Ae, NHP2, has been reported to interact with Nop10 in the presence of Cbf5 (Wang and Meier, 2004) and to interact directly with RNA, with a preference for irregular stem-loop structures (Henras et al., 2001).

Insight gained from our modeled Pf H/ACA sRNP, together with the available biochemical data, suggests that the arrangement of fully assembled H/ACA components is conserved between Archaea and Eukarya (Figure 6B). However, there is at least one apparent difference in the mechanism of assembly. Archaeal L7Ae is recruited to the complex by direct RNA-protein interaction (with the K-turn) (Baker et al., 2005), whereas specific association of the eukaryotic protein NHP2 is likely mediated primarily by protein-protein interaction (with NOP10) (Wang and Meier, 2004).

A Structural Domain Linked to Dyskeratosis Congenita

Mutations in the human DKC1 gene cause X-linked recessive DC. These mutations map to several well-separated regions on the DKC1 gene (Mason et al., 2005). The Cbf5 crystal structure provided us the opportunity to study the spatial arrangement of the amino acids affected in DC patients. Based on the Pf Cbf5 crystal structure and the sequence alignment of Pf Cbf5 with human dyskerin (Figure 4A), we constructed a three-dimensional structure of human dyskerin. This was carried out with a comparative protein structure modeling program, MODELLER (Marti-Renom et al., 2000). Dyskerin mutations were mapped on the modeled dyskerin structure (Figure 7). Clustered but distantly located in the primary structure, the majority of DC-causing mutations colocalize on one side of the PUA domain of dyskerin. Mutations that affect dyskerin residues in the N and C termini (1–35 or 359–513) could not be mapped because these residues are not included in the alignment with Pf Cbf5 and thus are not modeled (Figure 4A). However, these unmapped residues are expected to be in the same region due to the encirclement of the PUA domain by the N terminus (Figure 7).

The strikingly narrow spatial distribution of amino acids affected in DC offers insight into possible molecular mechanisms of this genetic disorder. Most of the mutations occur within or near the PUA domain, which is predicted to play a specific role in binding H/ACA and telomerase RNAs (Figure 7). Mutations within this domain would be expected to weaken the interaction between Cbf5 and its cognate RNAs, leading to decreased cellular accumulation of the RNAs. Indeed, depending upon the particular mutation investigated, significant decreases in the levels of H/ACA RNAs or telomerase RNA or both have been observed (Mitchell et al., 1999b; Mochizuki et al., 2004). Alternatively, the location where the mutations converge may delineate the binding site of a yet unidentified factor. Either possible mechanism could account for the ribosome deficiency

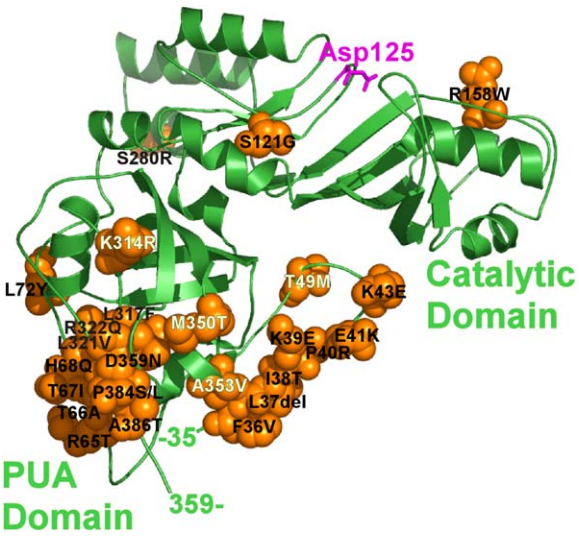

Figure 7. Clustering of Residues Affected in Dyskeratosis Congenita on a Dyskerin Model

Residues affected in DKC1 mutations (Marrone et al., 2005) are mapped on a human dyskerin structure model generated based on sequence homology between dyskerin and Pf Cbf5.

Affected residues are shown as orange spheres. Residues are labeled black or, in cases that occur in multiple families, in white.

and telomerase dysfunction associated with DKC1 mutations that lead to the disease (Meier, 2003).

Conclusion

Over the last few years, a number of pseudouridine synthase structures representing all currently known families of this enzyme have been reported. Comparison of the structures reveals a well-conserved catalytic core domain and a startling distribution of peripheral domains that are thought to be involved in binding different RNA substrates (Figure S1). Some of these RNA binding domains, such as the S4 and PUA domains, also occur in other RNA binding proteins. Interestingly, for a given RNA domain, the mode of binding can be drastically different in different enzymes. The catalytic domain of Cbf5 is fused with a PUA domain, which is likely responsible for anchoring the H/ACA-specific RNA. However, unlike any of the previously known pseudouridine synthases, Cbf5 requires partner proteins to bind its target RNA and to trigger its pseudouridylation activity. Our structural analyses suggest that both Nop10 and Gar1, in the presence of L7Ae, enhance the ability of Cbf5 to interact with the guide RNA-target RNA complex. In particular, Nop10 may prepare the guide RNA for target binding and/or stabilize RNA binding elements of Cbf5. Gar1, despite its structural homology with EF-Tu and limited evidence of its ability to associate with RNA, binds to the periphery of the active site and appears to interact with RNA binding elements in Cbf5 rather than to interact with RNA directly. Finally, the availability of the Cbf5 structure allowed us to identify a structural surface within the PUA domain of Cbf5 that is affected in DC. Our work suggests that the affected region may be involved in RNA interaction. It will be important to determine the specific role of the DC-linked structural region of Cbf5 in H/ACA and telomerase RNP biogenesis and function and dyskeratosis congenita.

Experimental Procedures

Protein Expression and Purification

Pf Gar1, Cbf5, and Nop10 were separately cloned into pET24D or pET21D vectors. Among these, only Gar1 was cloned with a 6xHis-tag at its C-terminal end. Plasmids carrying His-Gar1 and Cbf5 were transformed into BL21(DE3) cells for coexpression whereas the plasmid carrying Nop10 was individually transformed into BL21(DE3) cells. Cells coexpressing His-Gar1 and Cbf5 were mixed with those expressing Nop10 during harvesting and were disrupted by sonication in buffer A (50 mM sodium phosphate [pH 6.0], 1 M NaCl, 0.1 mM phenylmethylsulfonyl fluoride, and 14.2 mM β-mercaptoethanol). Cell lysate was then heated at 70ºC for 15 min and cleared by centrifugation at 18,000 rpm for 1 hr. The three proteins were purified as a single complex by Ni-NTA affinity chromatography followed by gel filtration (Superdex 200) in buffer B (20 mM HEPES [pH 6.0], 750 mM KCl, 5 mM β-mercaptoethanol, and 0.5 mM ethylenediamine tetra acetic acid). Purified trimeric protein complex was concentrated to 50 mg/ml and stored at −80ºC before crystallization. Selenomethionine substituted proteins (only on Cbf5 and his-Gar1) were produced according to a procedure by Ramakrishnan and Biou (1997) and were purified by using the same procedure.

Crystallization and Structure Determination

Crystals of the Cbf5-Nop10-Gar1 complex were obtained by the vapor diffusion method in hanging drops. A protein solution containing 3.5 mg of protein tricomplex in 50 mM $(NH_4)_2SO_4$, 35 mM NaCl, 100 mM $CsCl_2$, and 20 mM HEPES (pH 6.0) was mixed with 3 μl of a well solution containing 100 mM BisTris (pH 6.5), 25 mM $(NH_4)_2SO_4$, and 35% Pentaerythriotol Ethoxylate (15/4 EO/OH) (Hampton Research). Crystals of the trimeric complex grew in 5–7 days at room temperature to a maximum size of 0.3 mm × 0.6 mm × 0.6 mm. Selenomethionine-containing crystals were obtained from mixed selenomethionine and native proteins in an 8:1 mass ratio. Crystals could be frozen directly in a liquid nitrogen cryo stream in their mother liquor, and they belong to space group P1 with a = 36.087 Å, b = 75.361 Å, c = 103.524 Å, α = 95.90º, β = 96.22º, and γ = 94.98º. There are two trimers in each asymmetric unit that are related by a noncrystallographic 2-fold symmetry, resulting in 49% solvent content. A three-wavelength anomalous diffraction data set at 2.5 Å was collected with the BM-8 beamline at the Advanced Photon Source from a single selenomethionine-substituted crystal and by using the inverse beam strategy. Data were collected in 60º wedges sequentially for the peak, inflection, and a high-energy remote wavelength for a total of 360º (180º forward and 180º inverse) each. The program SOLVE (Terwilliger and Berendzen, 1999) was used to analyze the MAD data set. Twenty one out of the total of twenty six selenine sites were identified by the Single Anomalous Dispersion (SAD) method from the peak wavelength data and were used in subsequent phasing with all three wavelength data. Solvent flattened and noncrystallographically averaged experimental density was excellent and allowed tracing of the entire model for all three proteins. The initial model was built with the program O (Jones et al., 1991). Model refinement was carried out with CNS (Brunger et al., 1998) against a native data set at 2.1 Å resolution collected at the Southeast Regional Collaborative Access Team (SER-CAT) beamline. We included experimental phase probability distribution during initial stages of refinement by using the "mlhl" target function. Final refinement against experimental amplitudes, only without noncrystallographic symmetry constraint, led to an excellent model with satisfactory stereochemical geometry and crystallographic residual values (R_{free} = 25.8%, R_{work} = 21.2%). Data collection, phasing, and refinement statistics are included in Table 1. All figures were prepared with the PyMOL Molecular Graphics Program (DeLano Scientific, San Carlos, CA, USA).

Supplemental Data

Supplemental Data include three figures and Supplemental References and can be found with this article online at http://www.molecule.org/cgi/content/full/21/2/249/DC1/.

Table 1. Data Collection, Phasing, and Refinement Statistics[a]

Data Collection Statistics										
Data Sets	Wavelength (Å)	Resolution (Å)	Measured Reflections	Unique Reflections	R_{sym}	<I>/<σ(I)>	Completeness	Diffraction Ratios[b]		
								λ_1	λ_2	λ_3
λ_1 (peak)	0.9792	50.0–2.5 (2.6–2.5)	112803	32264	0.077 (0.259)	32.3 (7.5)	88.8 (21.4)	0.066	0.035	0.044
λ_2 (edge)	0.9794	50.0–2.5 (2.6–2.5)	111412	32032	0.073 (0.265)	32.1 (5.5)	93.3 (64.6)		0.048	0.049
λ_3 (remote)	0.9537	50.0–2.5 (2.6–2.5)	99467	34669	0.075 (0.355)	26 (5.1)	95.7 (86.0)			0.052
Native	1.0000	50.0–2.1 (2.2–2.1)	93272	51837	0.037 (0.182)	19.7 (3.1)	83.8 (35.5)			

Refinement Statistics						
R_{work}(%)	R_{free}(%)	B-factors (Å^2)	Number of atoms	Rmsd bond (Å)	Rmsd angle (°)	%Phi-Psi in core region (disallowed)
21.4 (37.2)	25.8 (49.6)	C, 33.5; G, 52.2; N, 49.6; W, 33.2	C, 2617; G, 612; N, 434; W,176	0.006	1.29	C, 92.5 (0.4)[c]; G, 91.0 (0); N, 92.0 (0)

Phasing Statistics	
Overall Z score	47.7
Figure of merits	0.56 (0.35)

[a] The values in parenthesis are those for the highest resolution shell. Abbreviations are as follows: C, Cbf5; G, Gar1; N, Nop10; and W, Water.
[b] Diffraction ratios are defined as $(\Delta|F|^2)^{1/2}/(|F|^2)^{1/2}$, where $\Delta|F|^2$ are taken between the current and the reference wavelength (λ_3) for dispersive ratios and between matched Bijvoet pairs for anomalous ratios.
[c] Glu97 is in a disallowed region despite a well-defined density.

Acknowledgments

This work was supported by the National Institutes of Health (NIH) grant R01 GM66958-01 (to H.L.) and NIH grant RO1 GM54682 (to M.P.T. and R.M.T.). R.R. is a predoctoral fellow of the American Heart Association, Florida/Puerto Rico Affiliate (0415201B). X-ray diffraction data were collected at both the Northeastern Collaborative Access Team (NE-CAT) BM-8 beamline and the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at http://necat.chem.cornell.edu/ and www.ser-cat.org/members.html. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract number W-31-109-Eng-38.

Received: August 18, 2005
Revised: November 2, 2005
Accepted: November 15, 2005
Published: January 19, 2006

References

Aravind, L., and Koonin, E.V. (1999). Novel predicted RNA-binding domains associated with the translation machinery. J. Mol. Evol. *48*, 291–302.

Arnez, J.G., and Steitz, T.A. (1994). Crystal structure of unmodified tRNA(Gln) complexed with glutaminyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. Biochemistry *33*, 7560–7567.

Auld, D.S. (2001). Zinc coordination sphere in biochemical zinc sites. Biometals *14*, 217–313.

Baker, D.L., Youssef, O.A., Chastkofsky, M.I., Dy, D.A., Terns, R.M., and Terns, M.P. (2005). RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. Genes Dev. *19*, 1238–1248.

Balakin, A.G., Smith, L., and Fournier, M.J. (1996). The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. Cell *86*, 823–834.

Blackburn, E.H. (2005). Telomeres and telomerase: their mechanisms of action and the effects of altering their functions. FEBS Lett. *579*, 859–862.

Bousquet-Antonelli, C., Henry, Y., G' elugne, J.P., Caizergues-Ferrer, M., and Kiss, T. (1997). A small nucleolar RNP protein is required for pseudouridylation of eukaryotic ribosomal RNAs. EMBO J. *16*, 4770–4776.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Charette, M., and Gray, M.W. (2000). Pseudouridine in RNA: what, where, how, and why. IUBMB Life *49*, 341–351.

Charpentier, B., Muller, S., and Branlant, C. (2005). Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. Nucleic Acids Res. *33*, 3133–3144.

Chen, J.L., and Greider, C.W. (2004). Telomerase RNA structure and function: implications for dyskeratosis congenita. Trends Biochem. Sci. *29*, 183–192.

Davis, D.R. (1995). Stabilization of RNA stacking by pseudouridine. Nucleic Acids Res. *23*, 5020–5026.

Decatur, W.A., and Fournier, M.J. (2003). RNA-guided nucleotide modification of ribosomal and other RNAs. J. Biol. Chem. *278*, 695–698.

del Campo, M., Ofengand, J., and Malhotra, A. (2004). Crystal structure of the catalytic domain of RluD, the only rRNA pseudouridine synthase required for normal growth of Escherichia coli. RNA *10*, 231–239.

Donmez, G., Hartmuth, K., and Luhrmann, R. (2004). Modified nucleotides at the 5′ end of human U2 snRNA are required for spliceosomal E-complex formation. RNA *10*, 1925–1933.

Dragon, F., Pogacic, V., and Filipowicz, W. (2000). In vitro assembly of human H/ACA small nucleolar RNPs reveals unique features of U17 and telomerase RNAs. Mol. Cell. Biol. *20*, 3037–3048.

Dyson, H.J., and Wright, P.E. (2005). Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. *6*, 197–208.

Fatica, A., Dlakic, M., and Tollervey, D. (2002). Naf1 p is a box H/ACA snoRNP assembly factor. RNA *8*, 1502–1514.

Foster, P.G., Huang, L., Santi, D.V., and Stroud, R.M. (2000). The structural basis for tRNA recognition and pseudouridine formation by pseudouridine synthase I. Nat. Struct. Biol. *7*, 23–27.

Ganot, P., Bortolin, M.L., and Kiss, T. (1997a). Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. Cell *89*, 799–809.

Ganot, P., Caizergues-Ferrer, M., and Kiss, T. (1997b). The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. Genes Dev. *11*, 941–956.

Girard, J.P., Lehtonen, H., Caizergues-Ferrer, M., Amalric, F., Tollervey, D., and Lapeyre, B. (1992). GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast. EMBO J. *11*, 673–682.

Gouaux, J.E., Stevens, R.C., and Lipscomb, W.N. (1990). Crystal structures of aspartate carbamoyltransferase ligated with phosphonoacetamide, malonate, and CTP or ATP at 2.8-A resolution and neutral pH. Biochemistry *29*, 7702–7715.

Grosjean, H., and Benne, R. (1998). Modification and Editing of RNA (Washington, DC: ASM Press).

Gutgsell, N., Englund, N., Niu, L., Kaya, Y., Lane, B.G., and Ofengand, J. (2000). Deletion of the Escherichia coli pseudouridine synthase gene truB blocks formation of pseudouridine 55 in tRNA in vivo, does not affect exponential growth, but confers a strong selective disadvantage in competition with wild-type cells. RNA *6*, 1870–1881.

Hamilton, C.S., Spedaliere, C.J., Ginter, J.M., Johnston, M.V., and Mueller, E.G. (2005). The roles of the essential Asp-48 and highly conserved His-43 elucidated by the pH dependence of the pseudouridine synthase TruB. Arch. Biochem. Biophys. *433*, 322–334.

Hamma, T., and Ferre-D' Amare, A.R. (2004). Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 A resolution. Structure (Camb) *12*, 893–903.

Heiss, N.S., Knight, S.W., Vulliamy, T.J., Klauck, S.M., Wiemann, S., Mason, P.J., Poustka, A., and Dokal, I. (1998). X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. Nat. Genet. *19*, 32–38.

Henras, A., Henry, Y., Bousquet-Antonelli, C., Noaillac-Depeyre, J., Gelugne, J.P., and Caizergues-Ferrer, M. (1998). Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. EMBO J. *17*, 7078–7090.

Henras, A., Dez, C., Noaillac-Depeyre, J., Henry, Y., and Caizergues-Ferrer, M. (2001). Accumulation of H/ACA snoRNPs depends on the integrity of the conserved central domain of the RNA-binding protein Nhp2p. Nucleic Acids Res. *29*, 2733–2746.

Henras, A.K., Capeyrou, R., Henry, Y., and Caizergues-Ferrer, M. (2004). Cbf5p, the putative pseudouridine synthase of H/ACA-type snoRNPs, can form a complex with Gar1p and Nop10p in absence of Nhp2p and box H/ACA snoRNAs. RNA *10*, 1704–1712.

Hoang, C., and Ferre-D' Amare, A.R. (2001). Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. Cell *107*, 929–939.

Hoang, C., and Ferre-D' Amare, A.R. (2004). Crystal structure of the highly divergent pseudouridine synthase TruD reveals a circular permutation of a conserved fold. RNA *10*, 1026–1033.

Ishitani, R., Nureki, O., Nameki, N., Okada, N., Nishimura, S., and Yokoyama, S. (2003). Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme. Cell *113*, 383–394.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for binding protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Kaya, Y., Del Campo, M., Ofengand, J., and Malhotra, A. (2004). Crystal structure of TruD, a novel pseudouridine synthase with a new protein fold. J. Biol. Chem. *279*, 18107–18110.

King, T.H., Liu, B., McCully, R.R., and Fournier, M.J. (2003). Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. Mol. Cell *11*, 425–435.

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

Koonin, E.V. (1996). Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. Nucleic Acids Res. *24*, 2411–2415.

Koonin, E.V., Bork, P., and Sander, C. (1994). A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme? Nucleic Acids Res. *22*, 2166–2167.

Lafontaine, D., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M., and Tollervey, D. (1998). The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. Genes Dev. *12*, 527–537.

Lukowiak, A.A., Narayanan, A., Li, Z.H., Terns, R.M., and Terns, M.P. (2001). The snoRNA domain of vertebrate telomerase RNA functions to localize the RNA within the nucleus. RNA *7*, 1833–1844.

Marrone, A., and Mason, P.J. (2003). Dyskeratosis congenita. Cell. Mol. Life Sci. *60*, 507–517.

Marrone, A., Walne, A., and Dokal, I. (2005). Dyskeratosis congenita: telomerase, telomerases and anticipation. Curr. Opin. Genet. Dev. *15*, 249–257.

Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000). Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. *29*, 291–325.

Mason, P.J., Wilson, D.B., and Bessler, M. (2005). Dyskeratosis congenita–a disease of dysfunctional telomere maintenance. Curr. Mol. Med. *5*, 159–170.

Meier, U.T. (2003). Dissecting dyskeratosis. Nat. Genet. *33*, 116–117.

Meier, U.T. (2005). The many facets of H/ACA ribonucleoproteins. Chromosoma *114*, 1–14.

Mitchell, J.R., Cheng, J., and Collins, K. (1999a). A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3′ end. Mol. Cell. Biol. *19*, 567–576.

Mitchell, J.R., Wood, E., and Collins, K. (1999b). A telomerase component is defective in the human disease dyskeratosis congenita. Nature *402*, 551–555.

Mochizuki, Y., He, J., Kulkarni, S., Bessler, M., and Mason, P.J. (2004). Mouse dyskerin mutations affect accumulation of telomerase RNA and small nucleolar RNA, telomerase activity, and ribosomal RNA processing. Proc. Natl. Acad. Sci. USA *101*, 10756–10761.

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

Newby, M.I., and Greenbaum, N.L. (2002a). Investigation of Overhauser effects between pseudouridine and water protons in RNA helices. Proc. Natl. Acad. Sci. USA *99*, 12697–12702.

Newby, M.I., and Greenbaum, N.L. (2002b). Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. Nat. Struct. Biol. *9*, 958–965.

Ni, J., Tien, A.L., and Fournier, M.J. (1997). Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. Cell *89*, 565–573.

Pan, H., Agarwalla, S., Moustakas, D.T., Finer-Moore, J., and Stroud, R.M. (2003). Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit. Proc. Natl. Acad. Sci. USA *100*, 12648–12653.

Pogacic, V., Dragon, F., and Filipowicz, W. (2000). Human H/ACA small nucleolar RNPs and telomerase share evolutionarily conserved proteins NHP2 and NOP10. Mol. Cell. Biol. *20*, 9028–9040.

Ramakrishnan, V., and Biou, V. (1997). Treatment of multiwavelength anomalous diffraction data as a special case of multiple isomorphous replacement. Methods Enzymol. *276*, 538–557.

Rozhdestvensky, T.S., Tang, T.H., Tchirkova, I.V., Brosius, J., Bachellerie, J.P., and Huttenhofer, A. (2003). Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. Nucleic Acids Res. *31*, 869–877.

Ruggero, D., Grisendi, S., Piazza, F., Rego, E., Mari, F., Rao, P.H., Cordon-Cardo, C., and Pandolfi, P.P. (2003). Dyskeratosis congenita and cancer in mice deficient in ribosomal RNA modification. Science *299*, 259–262.

Scott, W.G., Finch, J.T., and Klug, A. (1995). The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. Cell *81*, 991–1002.

Sivaraman, J., Sauve, V., Larocque, R., Stura, E.A., Schrag, J.D., Cygler, M., and Matte, A. (2002). Structure of the 16S rRNA pseudouridine synthase RsuA bound to uracil and UMP. Nat. Struct. Biol. *9*, 353–358.

Suryadi, J., Tran, E.J., Maxwell, E.S., and Brown, B.A., 2nd (2005). The Crystal structure of the Methanocaldococcus jannaschii multifunctional L7Ae RNA-binding protein reveals an induced-fit interaction with the box C/D RNAs. Biochemistry *44*, 9657–9672.

Tang, T.H., Bachellerie, J.P., Rozhdestvensky, T., Bortolin, M.L., Huber, H., Drungowski, M., Elge, T., Brosius, J., and Huttenhofer, A. (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. Proc. Natl. Acad. Sci. USA *99*, 7536–7541.

Terns, M.P., and Terns, R.M. (2002). Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. Gene Expr. *10*, 17–39.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. *55*, 849–861.

Turner, B., Melcher, S.E., Wilson, T.J., Norman, D.G., and Lilley, D.M. (2005). Induced fit of RNA on binding the L7Ae protein to the kink-turn motif. RNA *11*, 1192–1200.

Valadkhan, S., and Manley, J.L. (2003). Characterization of the catalytic activity of U2 and U6 snRNAs. RNA *9*, 892–904.

Wang, C., and Meier, U.T. (2004). Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. EMBO J. *23*, 1857–1867.

Watanabe, Y., and Gray, M.W. (2000). Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in Euglena gracilis, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaebacteria. Nucleic Acids Res. *28*, 2342–2352.

Watkins, N.J., Gottschalk, A., Neubauer, G., Kastner, B., Fabrizio, P., Mann, M., and Luhrmann, R. (1998). Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. RNA *4*, 1549–1568.

Yang, C., McPheeters, D.S., and Yu, Y.T. (2005a). Psi35 in the branch site recognition region of U2 small nuclear RNA is important for pre-mRNA splicing in Saccharomyces cerevisiae. J. Biol. Chem. *280*, 6655–6662.

Yang, P.K., Rotondo, G., Porras, T., Legrain, P., and Chanfreau, G. (2002). The Shq1p.Naf1p complex is required for box H/ACA small nucleolar ribonucleoprotein particle biogenesis. J. Biol. Chem. *277*, 45235–45242.

Yang, P.K., Hoareau, C., Froment, C., Monsarrat, B., Henry, Y., and Chanfreau, G. (2005b). Cotranscriptional recruitment of the pseudouridylsynthetase Cbf5p and of the RNA binding protein Naf1p during H/ACA snoRNP assembly. Mol. Cell. Biol. *25*, 3295–3304.

Yarian, C.S., Basti, M.M., Cain, R.J., Ansari, G., Guenther, R.H., Sochacka, E., Czerwinska, G., Malkiewicz, A., and Agris, P.F. (1999). Structural and functional roles of the N1- and N3-protons of psi at tRNA' s position 39. Nucleic Acids Res. *27*, 3543–3549.

Yu, Y.T., Shu, M.D., and Steitz, J.A. (1998). Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. EMBO J. *17*, 5783–5795.

Yu, Y.T., Terns, R.M., and Terns, M.P. (2005). Mechanisms and functions of RNA-guided RNA modifications. In Fine-Tuning of RNA Functions by Modification and Editing, H. Grosjean, ed. (New York, NY: Springer-Verlag Press), pp. 223–262.

Zebarjadian, Y., King, T., Fournier, M.J., Clarke, L., and Carbon, J. (1999). Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. Mol. Cell. Biol. *19*, 7461–7472.

Zhao, X., and Yu, Y.T. (2004). Pseudouridines in and near the branch site recognition region of U2 snRNA are required for snRNP biogenesis and pre-mRNA splicing in Xenopus oocytes. RNA *10*, 681–690.

Accession Numbers

The atomic coordinates have been deposited in the Protein Data Bank under accession number 2EY4.

Supplemental Data

Crystal Structure of a Cbf5-Nop10-Gar1 Complex and Implications in RNA-Guided Pseudouridylation and Dyskeratosis Congenita

Rumana Rashid, Bo Liang, Daniel L. Baker, Osama A.Youssef, Yang He, Kathleen Phipps, Rebecca M. Terns, Michael P. Terns, and Hong Li

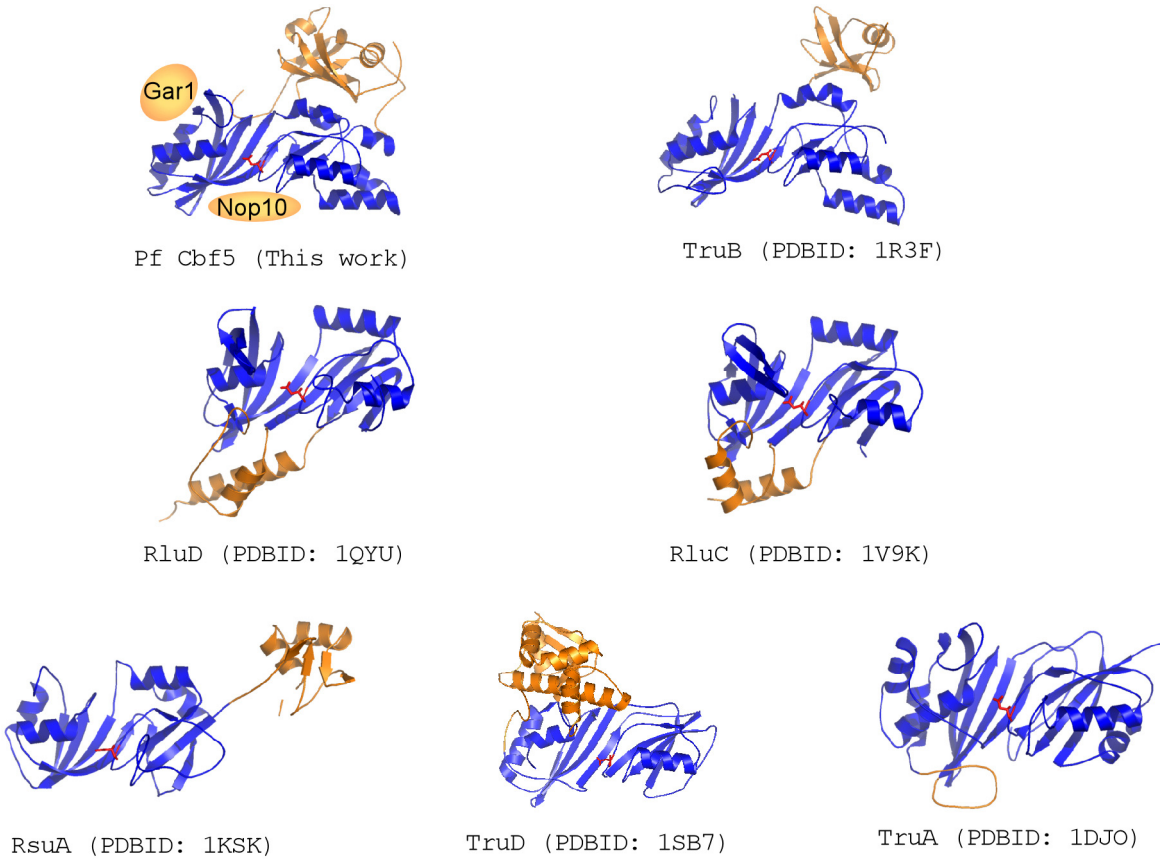

Figure S1. Comparison of Peripheral Domains of Currently Known Pseudouridine Synthase Structures

All structures were first superimposed and are displayed in the same orientation. The central domain is colored in blue and the peripheral domains are in orange. The catalytic aspartate is colored in red. The name and PDB code number of each pseudouridine synthase are indicated.

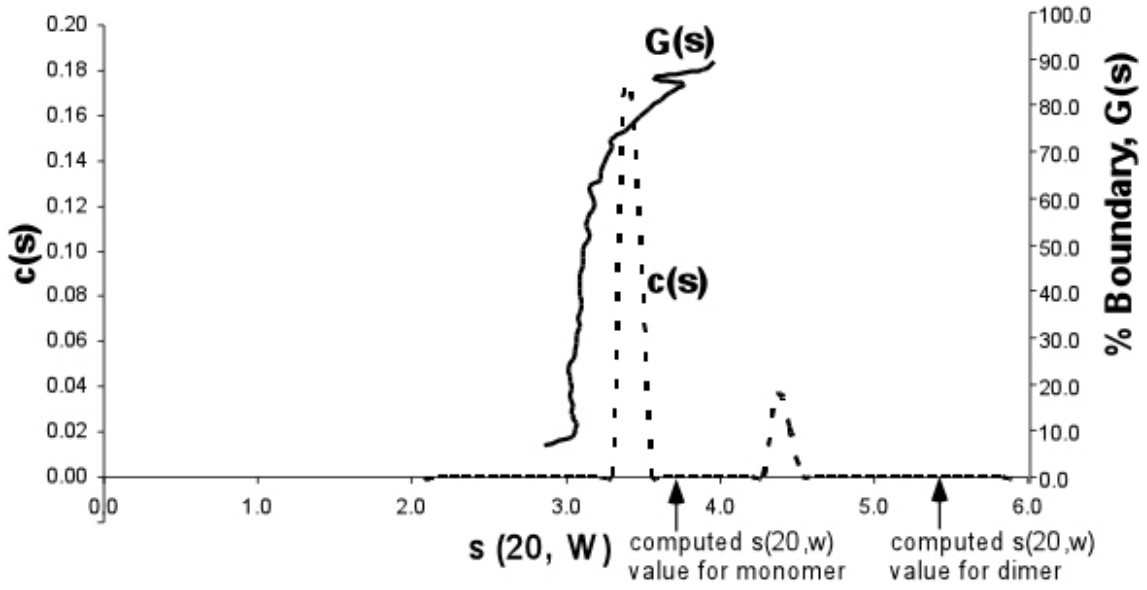

Figure S2. Analytic Centrifugation (AUC) Analysis of Pf Cbf5-Nop10-Gar1 Complex

We analyzed sedimentation velocity data using both the van Holde-Weischet (G(s), van Holde & Weischet, 1978) and the continuous sedimentation coefficient distribution (c(s), Schuck et al., 2002) methods. Both methods yielded an average sedimentation coefficient of ~3.4 Svedberg. Based on the crystal structure of the terimeric complex, we predicted sedimentation coefficients using the program HYDROPRO (Garcia De La et al., 2000). The computed sedimentation coefficient of a single trimer is 3.7 Svedberg while that of the dimer related by non-crystallographic symmetry is 5.7 Svedberg. This comparison supports a monomeric rather than a dimeric heterotrimer in solution. Sedimentation velocity experiments were conducted with a Beckman Optima XL-A analytical ultracentrifuge equipped with absorption optics at the Center for Analytic Ultracentrifugation of Macromolecular Assemblies at the University of Texas Health Science Center at San Antonio. The trimeric complex was made at 10 mg/ml (172 μM) in a buffer containing 25mM Na-phosphate (pH 6.5), and 1M KCl. Velocity data were analyzed with the UltraScan version 6.1 software (Demeler, B., 2001).

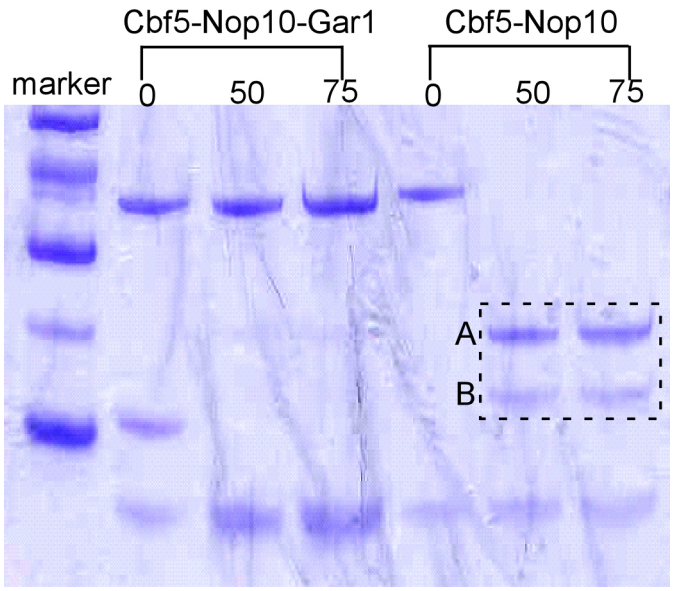

Figure S3. Mapping Interaction Surface between Gar1 and Cbf5 by Partial Proteolysis

Copurified ternary and binary complexes as indicated were subjected to limited trypsin digestion. The number in each lane indicates the trypsin unit used per milligram of protein. The box indicates the two protein fragments generated by trypsin cleavage in the absence of Gar1. The site of cleavage was identified by N-terminal sequencing. The band A starts with "LRTRKVY" and the band B starts with "ILPIDIK". The site indicated by the sequence of band A is in the loop connecting β7 and β10 and is marked in Figure 4.

Supplemental References

Demeler, B. (2001). *UltraSca 6.0-An integrated data analysis software package for sedimentation experiments.* http://www.ultrascan.uthscsa.edu, University of Texas Health Science Center at San Antonio Department of Biochemistry.

Garcia De La Torre, J., Huertas, M. L. & Carrasco, B. (2000). Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophysical Journal* **78**(2), 719–30.

Schuck, P., Perugini, M. A., Gonzales, N. R., Howlett, G. J. & Schubert, D. (2002). Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. *Biophys J* **82**(2), 1096–111.

van Holde, K. E. and W. O. Weischet. (1978). Boundary Analysis of Sedimentation Velocity Experiments with Monodisperse and Paucidisperse Solutes. Biopolymers, 17:1387–1403