Published in final edited form as:

Trends Biochem Sci. 2006 June ; 31(6): 311–315. doi:10.1016/j.tibs.2006.04.002.

How A Single Protein Complex Accommodates Many Different H/ACA RNAs

U. Thomas Meier

Department of Anatomy and Structural Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461

Abstract

Over 100 mammalian H/ACA RNAs form an equal number of ribonucleoproteins (RNPs) by associating with the same four core proteins. The function of these H/ACA RNPs is essential for ribosome biogenesis, pre-mRNA splicing, telomere maintenance, and, likely, for additional cellular processes. Recent crystal structures of archaeal H/ACA protein complexes illustrate how the same four proteins accommodate more than 100 distinct though related H/ACA RNAs and reveal a spatial mutation cluster underlying the bone marrow failure syndrome dyskeratosis congenita (DC).

Keywords

H/ACA RNPs; H/ACA RNAs; crystal structure; core proteins; dyskeratosis congenita; RNA-protein interaction

The vast majority of mammalian H/ACA RNPs engage in isomerization of uridines to pseudouridines (pseudouridylation) in ribosomal and spliceosomal small nuclear RNAs. Although the function of most pseudouridines is unknown, some are essential for optimal translation and for pre-mRNA splicing [1]. Perhaps the most intriguing species of H/ACA RNPs are defined by the snoRNA U17/E1 (snR30 in yeast), telomerase RNA (which ends in an H/ACA RNA structure), and a growing number of orphan H/ACA RNAs (without complementarity to any stable RNAs) [2]. U17 is the only essential H/ACA RNA and is required for pre-ribosomal RNA processing; telomerase RNA is required for the replication of chromosome ends; and the orphan H/ACA RNAs are by definition of unknown function but potentially involved in similar important processes as U17 and telomerase RNA (Fig. 1) [3]. Recently, three groups solved the crystal structures of archaeal H/ACA RNP core complexes consisting of two or three core proteins (Fig. 1) [4-6]. For the first time, these structures provide molecular details of protein-protein interactions in the H/ACA core complex and of the pseudouridylation itself.

H/ACA RNPs

H/ACA RNAs constitute one of the two major classes of small nucleolar and Cajal body RNAs (sno/scaRNAs; Fig. 1), the other class being C/D RNAs. On average 130-140 nucleotides in length, H/ACA RNAs conform to a consensus 5'-hairpin-hinge-hairpin-tail-3' secondary structure with the characteristic ACA trinucleotide exactly three residues from the 3' end (Fig. 1). H/ACA RNAs identify the ~130 known mammalian pseudouridylation sites by base pairing to a few nucleotides flanking the target uridines. Complementarity to target RNAs lies in the upper half of a bulge (pseudouridylation pocket) of one or both of the hairpins placing the unpaired target uridine at the bottom of a helix [7, 8]. Pseudouridylation is catalyzed by one of the four H/ACA core proteins, the pseudouridylase NAP57 (also dyskerin or, in yeast and archaea, Cbf5) [9]. The small basic proteins GAR1, NOP10, and NHP2 (L7Ae in archaea) round out H/ACA RNPs. All four core proteins, except GAR1, are essential for the structural integrity of H/ACA RNPs [3].

The structures

The structure of the archaeal pseudouridylase Cbf5 was solved complexed with Nop10 [5, 6] or with Nop10 and Gar1 [4]. Based on structural and sequence homologies, pseudouridylases are classified into 5 families, RluA, RsuA, TruA, TruB, and TruD [10]. Crystal structures of bacterial and/or archaeal representatives for each family have been solved. Cbf5 belongs to the TruB family of pseudouridylases, specified by the *E. coli* enzyme responsible for the modification of uridine 55 in all elongator tRNAs. Unlike Cbf5, however, all other pseudouridylases (including TruB) function as independent enzymes that recognize and isomerize uridines without assistance from other proteins or RNAs. Therefore, the reported crystal structures of archaeal Cbf5 are the first of pseudouridylases that function in the context of a RNP and that depend on a guide RNA for target site recognition.

The pseudouridylase (Cbf5)

Although the structures are derived from three different archaea, *M. jannaschii* [5], *P. furiosus* [4] (Fig. 1), and *P. abyssi* [6], they are in good agreement. Cbf5 shares a catalytic domain that superimposes closely with that of TruB [11] and the four other pseudouridylase family members (Fig. 1, green). Matching positions of a universally conserved aspartate (Fig. 1, arrow) and a few key residues in the catalytic center suggest a conserved mode of catalysis for all pseudouridylases, RNA guided or not. In comparison to Cbf5, the catalytic domain of TruB and other pseudouridylases contains additional segments important for substrate RNA binding. Apparently, Cbf5 can function without these appendages because it associates with other proteins and interacts indirectly with its substrates via its H/ACA RNA. In addition to the catalytic domain, Cbf5 harbors a carboxy terminal pseudouridylase archaeosine tRNA-guanine transglycosylase (PUA) domain [12] (Fig. 1, cyan) that is larger and enveloped by an amino terminal extension (Fig. 1, yellow) when compared to that of the eubacterial TruB. In archaeosine transglycosylase (archaeosine is a guanine derivative specific to archaeal tRNAs), this domain is important for recognition of the CCA terminal residues of tRNA (CCA residues are essential for tRNAs to be charged with their amino

acids). Interestingly, in Cbf5 its PUA domain is similarly perched for potential recognition of the defining ACA trinucleotide at the 3' end of H/ACA RNAs.

The bracket (Nop10)

The 60 amino acid Nop10 lines the oblong catalytic domain of Cbf5 with a random coil central segment that separates the Nop10 amino terminal zinc binding and carboxy terminal alpha helical domains (Fig. 1, red). Although Nop10 alone appears to be an intrinsically disordered protein [13], it becomes structured upon Cbf5 binding. This tight interaction with Cbf5 may be important to freeze the catalytic domain in the most favorable position for catalysis and to extend the positive surface potential of Cbf5 for H/ACA and substrate RNA binding [4-6]. In fact, Nop10 and its C-terminal alpha helix alone promote binding of substrate RNA in gel shift assays and are sufficient for reconstituting RNA-guided pseudouridylase activity in the context of Cbf5 and L7Ae [6]. Nop10 also may be involved in docking of L7Ae to the complex as modeled by Rashid et al. [4] and as demonstrated for its mammalian counterpart NHP2 [14]. By mutating one of the four zinc-coordinating cysteines, which are conserved in archaeal but not eukaryal Nop10, Manival et al. show that zinc binding is not required for RNP assembly and activity [6]. Therefore, zinc seems important for the structural integrity of specifically archaeal Nop10 (which must withstand challenging temperatures).

The outsider (Gar1)

The Gar1 crystal structure in the Gar1-Cbf5-Nop10 complex is the first solved of any Gar1 homolog and shows it to belong to the reductase/isomerase/elongation factor fold superfamily [4] (Fig. 1, blue). Gar1 binds to one end of the Cbf5 catalytic domain without contacting Nop10. This is in good agreement with biochemical data demonstrating an independent interaction of archaeal, yeast, and mammalian Gar1 with Cbf5 homologs [14-17]. However, Gar1, which has been tied to the catalytic core by crosslinking in mammalian H/ACA RNPs [14], is situated too far from the active site aspartate to make contact with it. The missing one fourth of archaeal Gar1 and/or of the defining glycine/arginine-rich amino and carboxy termini of eukaryal GAR1 may account for this discrepancy. The location of Gar1 in the complex is of further interest because it may prove to be identical to that of Naf1 in eukaryal H/ACA RNPs. Naf1, required for the biogenesis of H/ACA RNPs, shares homology with the domain of Gar1 that contacts Cbf5 [18].

How do H/ACA and target RNAs bind this protein complex?

The structure of the related TruB in association with its substrate tRNA provided the basis for identifying the probable RNA binding surface in the complex [4, 5]. Specifically, the PUA domain of Cbf5, its catalytic domain, and Nop10 form a platform with positively charged surface potential ideal for RNA binding (Fig. 1, grey dashed oval and line). Indeed, mutation of two conserved basic and surface exposed amino acid residues in the catalytic domain abolished archaeal H/ACA RNA binding to the Cbf5-Nop10 dimer [5]. Moreover, the PUA domain proved essential for RNP assembly and activity as (in the presence of target RNA) it supported both, even when added in trans [6]. Using the information of the TruB-tRNA complex and of a previously solved L7Ae-guide RNA complex, Rashid et al.

modeled a fully assembled H/ACA RNP hybridized to a target RNA [4]. The three-junction helix, formed by base pairing between the substrate RNA and the pseudouridylation pocket and by the upper stem of the guide RNA, folds into an extended structure that places the target uridine next to the catalytic aspartate. The tip of the H/ACA RNA hairpin extends beyond Nop10 where it binds L7Ae, whereas the 3' ACA and both ends of the substrate RNA point beyond the PUA domain. Therefore, and consistent with all studies, H/ACA RNPs appear to be bipartite, one side protein, one side RNA. This separation between RNA and protein in H/ACA RNPs contrasts that of other RNPs, e.g., the large ribosomal subunit where proteins adorn the surface all over the RNA core and send extensions deep into the interior [19]. In the case of H/ACA RNPs, this pasting of the RNAs to one side of the core protein complex appears exquisitely suited to accommodate the over 100 different but related H/ACA and their various target RNAs. It will be interesting to see if C/D RNPs, the other major class of snoRNPs, adopted a similar mechanism to bind their many C/D RNAs.

The mutation cluster

A most interesting result of these structural analyses is the observation that the amino terminus of Cbf5 wraps around its carboxy terminal PUA domain thereby generating a single hotspot for mutations identified in the Cbf5 ortholog NAP57 of dyskeratosis congenita patients (Fig. 1, arrowheads; Box 1) [4-6]. This was dramatically illustrated by modeling the structure of NAP57 (residues 35-359 out of 514) based on that of archaeal Cbf5 and by mapping the DC mutations on this structure [4]. The tight spatial clustering of mutations that are distant from each other in the linear sequence suggests that most DC mutations impact on the same function of NAP57. The location of the mutations in and near the PUA domain indicates a possible effect on H/ACA RNA binding. Alternatively and more intriguingly, the exposed surface defined by the mutation is strategically situated at one end of the complex for interaction with a non-core H/ACA component, e.g., a telomerase specific protein or an H/ACA RNP assembly factor. Thus, interference with such an interaction could explain a specific effect of DC mutations on only a few select H/ACA RNPs.

Concluding remarks

As outlined above, this exciting flurry of archaeal H/ACA protein structures has yielded considerable insight into many aspects of H/ACA RNP biology. However, several important questions remain. In human H/ACA RNPs, how will the missing one third and one half of NAP57 and GAR1, respectively, fit into the RNP and how will they impact the overall structure? Given the often-conserved nature of the DC missense mutations, what is their molecular impact? The answers will only be known once the structure of the human H/ACA core has been solved. Ultimately, the crystal structure of the entire RNP including the H/ACA RNA (and target RNA) will be required. Meanwhile, the solution structure of one of the H/ACA RNA hairpins may aid in modeling the complete RNP [13]. The full RNA protein complex may also provide insight as to how the base pairing between guide and target RNA is released after pseudouridylation. Therefore, the structure of mammalian holo-H/ACA RNPs will answer basic as well as clinically relevant questions.

Acknowledgments

I thank Narcis Fernandez-Fuentes for help with the preparation of figure 1 and I appreciate the critical comments on the manuscript by Jon Warner, Nupur Kittur, and anonymous reviewers. The work in the author's laboratory is supported by a grant from NIH.

References

1. Yu, YT., et al. Mechanisms and functions of RNA-guided RNA modification. In: Grosjean, H., editor. *Fine-Tuning of RNA Functions by Modification and Editing*. Springer-Verlag Press; 2005. p. 223–262.
2. Bachellerie JP, et al. The expanding snoRNA world. *Biochimie*. 2002; 84:775–790. [PubMed: 12457565]
3. Meier UT. The many facets of H/ACA ribonucleoproteins. *Chromosoma*. 2005; 114:1–14. [PubMed: 15770508]
4. Rashid R, et al. Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol Cell*. 2006; 21:249–260. [PubMed: 16427014]
5. Hamma T, et al. The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nat Struct Mol Biol*. 2005; 12:1101–1107. [PubMed: 16286935]
6. Manival X, et al. Crystal structure determination and site-directed mutagenesis of the *Pyrococcus abyssi* aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Res*. 2006; 34:826–839. [PubMed: 16456033]
7. Kiss T. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*. 2002; 109:145–148. [PubMed: 12007400]
8. Decatur WA, Fournier MJ. RNA-guided nucleotide modification of ribosomal and other RNAs. *J Biol Chem*. 2003; 278:695–698. [PubMed: 12431975]
9. Lafontaine DL, Tollervey D. Birth of the snoRNPs: the evolution of the modification-guide snoRNAs. *Trends Biochem Sci*. 1998; 23:383–388. [PubMed: 9810226]
10. Kaya Y, Ofengand J. A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. *RNA*. 2003; 9:711–721. [PubMed: 12756329]
11. Hoang C, Ferre-D'Amare AR. Cocrystal structure of a tRNA Ψ 55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell*. 2001; 107:929–939. [PubMed: 11779468]
12. Ferre-D'Amare AR. RNA-modifying enzymes. *Curr Opin Struct Biol*. 2003; 13:49–55. [PubMed: 12581659]
13. Khanna M, et al. Structural study of the H/ACA snoRNP components Nop10p and the 3' hairpin of U65 snoRNA. *RNA*. 2006; 12:40–52. [PubMed: 16373493]
14. Wang C, Meier UT. Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J*. 2004; 23:1857–1867. [PubMed: 15044956]
15. Baker DL, et al. RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev*. 2005; 19:1238–1248. [PubMed: 15870259]
16. Charpentier B, et al. Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res*. 2005; 33:3133–3144. [PubMed: 15933208]
17. Henras AK, et al. 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*. 2004; 10:1704–1712. [PubMed: 15388873]
18. Fatica A, et al. Naf1p is a box H/ACA snoRNP assembly factor. *RNA*. 2002; 8:1502–1514. [PubMed: 12515383]
19. Moore PB, Steitz TA. The structural basis of large ribosomal subunit function. *Annu Rev Biochem*. 2003; 72:813–850. [PubMed: 14527328]
20. Marrone A, et al. Dyskeratosis congenita: telomerase, telomeres and anticipation. *Curr Opin Genet Dev*. 2005; 15:249–257. [PubMed: 15917199]

21. Heiss NS, et al. X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat Genet.* 1998; 19:32–38. [PubMed: 9590285]
22. Armanios M, et al. Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita. *Proc Natl Acad Sci USA.* 2005; 102:15960–15964. [PubMed: 16247010]
23. Vulliamy T, et al. The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature.* 2001; 413:432–435. [PubMed: 11574891]
24. Ruggero D, et al. Dyskeratosis congenita and cancer in mice deficient in ribosomal RNA modification. *Science.* 2003; 299:259–262. [PubMed: 12522253]
25. Mochizuki Y, et al. Mouse dyskerin mutations affect accumulation of telomerase RNA and small nucleolar RNA, telomerase activity, and ribosomal RNA processing. *Proc Natl Acad Sci USA.* 2004; 101:10756–10761. [PubMed: 15240872]

Box 1**Dyskeratosis congenita (DC)****Phenotype**

DC is a rare bone marrow failure syndrome that is inherited in three modes (listed in descending order of frequency and severity) X-linked, autosomal recessive, and autosomal dominant. DC patients are mostly identified in their first decade of life by the triad of nail dystrophy, abnormal skin pigmentation, and mucosal leucoplakia, but often die in their third decade due to bone marrow failure. Patients are also predisposed to malignancies of rapidly dividing tissues and other complications [20].

Molecular pathogenesis

X-linked DC is caused by mutations in the H/ACA RNP core protein and pseudouridylase NAP57 [21]. The autosomal dominant form is caused by mutations in telomerase RNA and reverse transcriptase [22, 23], whereas the gene(s) responsible for the autosomal recessive form remain(s) to be identified. Impaired telomere maintenance could explain both X-linked and autosomal dominant DC because telomerase RNA also forms an H/ACA RNP. Such a mechanism is supported by shortened telomeres in DC patients. Do mutations in NAP57, therefore, specifically affect its interaction with telomerase RNA but not with the other one hundred or so H/ACA RNAs? The answer appears to be no as the DC phenotype can be reproduced by mutations in NAP57 that, in the absence of telomere defects, impair ribosome biogenesis (and possibly pre-mRNA splicing) through reduced RNA pseudouridylation [24, 25].

Figure 1. Archaeal Gar1-Cbf5-Nop10 complex and mammalian classes of H/ACA RNAs
 Crystal structure of the H/ACA protein complex of Gar1 (blue), Nop10 (red), and Cbf5 with its catalytic domain (green), PUA domain (cyan), and amino terminus (yellow). The alpha carbon backbones in ribbon diagram are shown through a surface rendering based on the atomic coordinates for the *P. furiosus* proteins deposited in the Protein Data Bank (accession code 2ey4) [4] and generated using PyMOL (<http://pymol.sourceforge.net>). The amino (N) and carboxy (C) termini of the proteins and the catalytic aspartate (arrow, black ball and stick model) are indicated. A top view, with the RNA binding surface on top, and a side view are presented. The likely area and surface for binding of one hairpin of any of the depicted H/ACA RNAs are outlined by a grey dashed oval and a grey dashed line, respectively. In the RNA schematics, the location of antisense elements in the pseudouridylation pockets and the template region of telomerase RNA are highlighted (thick black lines). Note, relative to the protein complex, the H/ACA RNAs are not drawn to scale and are randomly arranged around the protein structures. The location of the DC mutation cluster in the Cbf5 PUA domain and amino terminus (based on a model of human NAP57) is pointed out by arrowheads.