Multicomponent Machines in RNA Modification: H/ACA Ribonucleoproteins

Petar Grozdanov and U. Thomas Meier*

Abstract
Pseudouridylation, the isomerization of uridine to pseudouridine, is the most frequent post-transcriptional modification of RNA, such that pseudouridine has even been termed the fifth nucleotide. Whereas eubacteria employ single protein enzymes to identify and modify target uridines, archaeabacteria and eukaryotes additionally evolved more complex modification machines, H/ACA ribonucleoproteins (RNPs). Each H/ACA RNP consists of a short RNA and the same four core proteins, one of which is the pseudouridine synthase related to the bacterial single protein enzymes. In this chapter, we will give an overview of these multicomponent machines with emphasis on the eukaryal systems that have acquired additional functions and that are the subject of the inherited bone marrow failure syndrome dyskeratosis congenita.

Introduction
Nuclei of metazoans harbor several hundred individual small nucleolar ribonucleoproteins (snoRNPs) that predominantly function in RNA modification. They are divided into two major classes according to their function-defining snoRNAs, box H/ACA and box C/D snoRNPs, which pseudouridylate and 2′-O-methylate their target RNAs, respectively. SnoRNAs guide the modification by site-specific base pairing while an enzyme (which is one of four core proteins of each RNP) catalyzes the reaction. Collectively, the snoRNAs account for one of the largest families of noncoding RNAs. In this overview, we will focus on the H/ACA class of RNPs (see chapter by Gagnon et al for C/D RNPs).

H/ACA RNAs
H/ACA RNAs are generally 60-150 ribonucleotides in length, noncoding, trans-acting molecules, for reviews see. 1, 2, 3, 4, 5, 6 Defining features of H/ACA RNAs are two hairpins separated by a short single stranded sequence (hinge), which includes an ANANNA consensus hexanucleotide, and an ACA triplet exactly three nucleotides from their 3′-end (Fig. 1A). 7, 8 Although the number of hairpins can vary, H/ACA RNAs are conserved from archaea to mammals. The hairpins contain internal bulges and can differ in size and organization of stems and loops (Fig. 1A). The vast majority of H/ACA RNAs contain in their bulges two 3-10 ribonucleotide long stretches (3′ and 5′ of the upper stem) that are complementary to the sequences flanking their target uridines (Fig. 1A, arrows). 9, 10 Hence, these internal loops are also known as pseudouridylation pockets. So
far, targets of these so-called guide RNAs are ribosomal RNAs (rRNAs) and spliceosomal small nuclear RNAs (snRNAs). 12–15 Although originally and together with C/D RNAs identified in nucleoli as snoRNAs, H/ACA RNAs are now subdivided into guide and nonguide RNAs (that function in pseudouridylation or not). The former are further categorized into snoRNAs (located in nucleoli and functioning in the pseudouridylation of rRNA) and small Cajal body RNAs (scaRNAs, located in Cajal bodies and functioning in the pseudouridylation of snRNAs). Cajal bodies are approximately one micron sized structures numbering one to five in most nuclei and serving as locale of snRNA modification.16–18 ScaRNAs contain a Cajal body-localizing element, the CAB box (5′-ugAG-3′), in the terminal loop of one or both hairpins (see chapter by Yu et al).19 Two types of scaRNAs are unique, one, in combining features of H/ACA and C/D RNAs yielding hybrids and, two, in forming a twin H/ACA RNA with four hairpins.7,15,16,20 In mammals, H/ACA RNAs target the modification of ∼100 uridines in rRNAs and 27 in snRNAs. However, it should be noted that not all pseudouridines are specified by H/ACA RNPs. For example, the pseudouridines of eukaryotic tRNAs and yeast 5S rRNA21 are generated by protein-only enzymes that recognize the uridine and catalyze its modification and yeast U2 snRNA is the target of both H/ACA RNPs and stand-alone pseudouridylases (see chapter by Karijolich et al).

H/ACA Core Proteins

All H/ACA RNAs associate with four conserved core proteins that are responsible for the metabolic stability of the RNAs and catalyze the isomerization of uridine to pseudouridine. These proteins are the mammalian pseudouridine synthase NAP57 (aka dyskerin or in yeast Cbf5p and in archaea Cbf5), NOP10, NHP2 (L7Ae in archaea) and GAR1 (Fig. 1B).

NAP57 was identified in the immunoprecipitate of the highly phosphorylated nucleolar protein Nopp140 and termed Nopp140 associated protein with a relative molecular mass of 57 kD.22 NAP57 localizes to nucleoli and Cajal bodies and is 70% identical to yeast Cbf5p, which was previously identified as a low-affinity centromeric DNA binding protein.23 The central part
of NAP57 (later identified as the catalytic domain) showed 34% identity to a bacterial protein that was subsequently purified based on its pseudouridylase activity.\textsuperscript{24}

Analysis of the primary amino acid sequence of NAP57 reveals several distinct domains. One lysine-rich motif at the amino and three at the carboxyl terminus are separated by the catalytic and the pseudouridine and archaeosine transglycosylase (PUA) domains (see chapter by Mueller and Ferre-D’Amare). The catalytic domain contains a conserved aspartate that is important for catalysis (see chapter by Mueller and Ferre-D’Amare).\textsuperscript{25-28} The PUA domain is an RNA binding motif\textsuperscript{29-32} and the lysine-rich stretches can function as nuclear localization signals.\textsuperscript{33,34}

NOP10 is the smallest polypeptide of the RNP with only 64 amino acids in mammals and a molecular mass of 7.7 kD.\textsuperscript{35} In the complex, it lines the catalytic domain of NAP57 stabilizing it and providing a docking site for NHP2.\textsuperscript{36,37}

NHP2 was discovered as a nonhistone protein with molecular mass of 17 kDa.\textsuperscript{38} It is homologous to the ribosomal protein L30 and to 15.5K/NHP2L1/NHPX (Snu13p in yeast), which is part of C/D RNPs and the snrRNPs U4.\textsuperscript{39-42} The archaeal ortholog L7Ae is part of both archaeal H/ACA and C/D RNPs.\textsuperscript{43} L7Ae (and 15.5K) binds specifically to a kink-turn motif in RNA, whereas NHP2 binds RNA secondary structures in an unspecific manner (see chapter by Gagnon et al for more details).\textsuperscript{35,37,44,45}

GAR1 is a protein with a molecular mass of 22 kDa and consists of a central domain flanked by glycine-arginine rich (GAR) domains.\textsuperscript{46} GAR1 is an integral part of the active RNP complex and binds directly to NAP57.\textsuperscript{37,47-50}

According to the crystal structure of an archaeal H/ACA RNP and to cryoelectron microscopic studies of purified H/ACA particles, each of the normally two hairpins of H/ACA RNAs associates with its own set of four core proteins placing the catalytic core at the pseudouridylation pocket.\textsuperscript{40,48,51} Therefore H/ACA RNPs consist of one RNA and two each of the four core proteins.

Beyond Formation of Pseudouridines

Although most H/ACA RNAs guide the modification of RNA, their most prominent members do not. They are the only essential H/ACA RNA, U17/E1 (snR30 in yeast), required for ribosomal RNA processing and the mammalian telomerase RNA, required for telomere maintenance.\textsuperscript{52,53} Of additional interest are tissue-specific and orphan H/ACA RNAs (without complementarity to any stable RNAs).

Ribosomal RNA Processing

The H/ACA RNA U17/E1 is required for a processing event in the formation of 18S rRNA.\textsuperscript{54} Thus, U17/E1 is essential for ribosome biogenesis and cell viability. Specifically, short stretches of highly conserved nucleotides in the bulge of the 3’ hairpin are engaged in the early cleavage steps of 35S pre-rRNA in yeast.\textsuperscript{55} The importance of these sequences is illustrated by their high degree of evolutionary conservation in budding and fission yeasts and in all vertebrates.\textsuperscript{53,56} In addition to the H/ACA core proteins, U17/E1 associates with the DEAD box helicase Has1p, which is required for snRNP release from pre-rRNA.\textsuperscript{57} Additional interacting but as of yet uncharacterized proteins have been identified.\textsuperscript{51,58} These may be testimony of the specialized function of U17/E1.

Telomerase

Maintenance of chromosome ends (telomeres), which plays a crucial role in cellular senescence and cancer, is mediated by telomerase, an H/ACA RNP.\textsuperscript{52} Specifically, human telomerase consists of a 451 nucleotide long RNA (hTR) whose 3’ end is an H/ACA domain.\textsuperscript{59} Like all H/ACA RNAs, hTR associates with all four core proteins that are important for its accumulation and stability.\textsuperscript{59,60} Activity of telomerase is dependent on the template region in the 5’ half of hTR and on the reverse transcriptase TERT. Although hTR (and its H/ACA core proteins) is (are) expressed in all cells, TERT (and telomerase activity) is (are) mostly restricted to stem and cancer cells. Not only is hTR an H/ACA RNA but it is also a scaRNA with a CAB box that localizes telomerase to Cajal bodies in a cell cycle and TERT dependent manner.\textsuperscript{61-65}
Additional Functions

New H/ACA RNAs are still being identified using a combination of biochemical and in silico approaches. These approaches unearthed novel H/ACA RNAs that lack complementarity to any of the stable RNAs. These so-called orphan H/ACA RNAs appear either to guide the pseudouridylation of yet to be identified RNAs (e.g., mRNAs) or to exhibit separate functions (like U17/E1 and hTR).

One of these orphan H/ACA RNAs, HBI-36, is of particular interest because, unlike all other H/ACA RNAs, it is expressed in a tissue specific manner. Specifically, HBI-36 is expressed from an intron of the serotonin C2 receptor gene only in the choroid plexus of the brain suggesting a developmentally regulated function.

In another case, scaRNA U100 possesses complementarity to a target RNA, however, the uridine that it specifies in U6 snRNA is apparently not modified. Therefore, even apparent guide RNAs may serve different purposes.

Architecture of H/ACA RNPS

Overview

Recent years have produced a detailed view of H/ACA RNPs. Biochemical analyses revealed intra RNP protein-RNA interactions of eukaryal particles and X-ray crystallographic studies provided the details of partially and fully reconstituted archael RNPs. The major difference between archael and eukaryal H/ACA RNPs is between the homologous proteins L7Ae and NHP2, respectively. Whereas L7Ae recognizes and binds archael H/ACA RNAs independently, NHP2 does so only when complexed with NAP57 via NOP10.37,43-45

H/ACA RNPs appear unique among RNA-protein complexes. In place of the usual intertwined structures of proteins and RNA, e.g., in the cases of the U1 snRNP and C/D RNPs (see chapter by Gagnon et al), the four H/ACA core proteins form a planar, coherent surface accommodating individual H/ACA RNAs and their targets like a slice of bread being buttered. This arrangement may allow the accommodation of the 150 or so different H/ACA RNAs by the same protein complex.79,80

Intra-RNP Interactions

In eukaryotes the four core proteins can form an independent complex (archael L7Ae is held in place by the RNA) that resembles an equilateral triangle (Fig. 1B). Its corners are formed by GAR1, NHP2 and the C-terminal PUA domain of NAP57 (which also associates with the N-terminus). The body consists of the catalytic domain of NAP57, which is lined by NOP10 (that in turn binds NHP2) and which binds GAR1. One hairpin of an H/ACA RNA stretches across the NAP57-NOP10-NHP2 axis. The PUA domain of NAP57 anchors the ACA triplet on one end and NHP2 the terminal loop of a hairpin on the other thereby placing the pseudouridylation pocket over the catalytic domain of NAP57. The confinement of the ACA triplet to the PUA pocket explains the constraint of 14 nucleotides between the ACA and the top of the pseudouridylation pocket (where the target uridine will be situated) for placement of the latter near the active site of NAP57. GAR1 is not required for RNA binding and the three proteins NAP57, NOP10 and NHP2 form an independent complex (the core trimer) that provides the specificity for H/ACA RNA recognition. Despite this separation of GAR1 from the core trimer, UV-crosslinking experiments suggest that all eukaryal core proteins contact the H/ACA RNA in some fashion, whereas only NAP57 and GAR1 crosslink to the target uridine.87,73

RNP-Substrate Interactions

How an H/ACA guide RNA accommodates its target RNA has been visualized in solution and in the context of three core proteins. The pseudouridylation pocket of the guide RNA (Fig. 1C, in black) forms a more or less straight opening that base pairs on one side with the 5’ half of the target RNA (gray) (extending the bottom helix of the hairpin) and on the other with the 3’ half (extending the top helix of the hairpin) (arrows). This unique conformation forces the
substrate RNA into a tight turn at the two unpaired nucleotides flipping out the target uridine (Fig. 1C, gray U and arrowhead), which becomes accessible to the active site of NAP57. Additionally, this arrangement of the H/ACA guide-target RNA complex obviates the necessity of a helicase for loading and release of target RNAs.81,82

**RNP Stability**

Each of the proteins of the core trimer, but not GAR1, is essential for cell viability and for metabolic stability of all H/ACA RNAs and of each other.35,40,60,83,84 Consistent with these observations in yeast, mammalian RNP complexes of the core trimer and an H/ACA RNA, once assembled do not exchange their RNA.37 In particular, NAP57 remains stably associated with its H/ACA RNA in cell extracts, whereas NOP10 and NHP2 exchange to some extent and GAR1 more readily.85 In conclusion, H/ACA RNPs are stable complexes and formation of new particles requires de novo synthesis and assembly of its individual components.

**Biogenesis of H/ACA RNPs**

Despite the simple five-component composition of H/ACA RNPs, eukaryal particles rely on accessory factors for their assembly. In particular, two factors, Naf1p and Shq1p have been identified in yeast to be essential for the stable accumulation of H/ACA RNPs.86-88 Both proteins have homologs in mammals, NAF1 and SHQ1. NAF1 is recruited cotranscriptionally to the site of H/ACA RNA transcription and is also required for the assembly of human H/ACA RNPs including telomerase.89-92 NAF1 binds NAP57 at the same site as GAR1 indicating a sequential assembly.37,87,93 Although less is known about Shq1p, it also binds Cbf5p (the yeast NAP57) without being part of mature H/ACA RNPs.88 Consistent with these findings, both proteins are excluded from nucleoli and Cajal bodies, the sites of mature particles and localize to the nucleoplasm. In contrast to eukaryotes, archaea lack recognizable homologs of these assembly factors and their H/ACA RNPs can be functionally reconstituted with just the five core components alone.49,50

Two additional proteins, Nopp140 and SMN, have been implicated in H/ACA RNP biogenesis and/or function due to their ability to interact with them. In fact, NAP57 was identified in immunoprecipitates of the highly phosphorylated nucleolar protein Nopp140,94 whereas the survival of motor neuron protein (SMN) that is affected in spinal muscular atrophy binds GAR1.95-97 Although SMN is clearly involved in the assembly of spliceosomal snRNPs, evidence for a similar function in H/ACA RNP biogenesis is lacking. Therefore, NAF1 and SHQ1 are to date the only bona fide H/ACA RNP assembly factors.

Finally, factors that may be involved in the biogenesis of both H/ACA and C/D RNPs have been identified. These include AAA+ helicases and chaperone proteins, e.g., the helicases Rvb1 (Tih1, TIP48, pontin, etc.) and Rvb2 (Tih2, TIP49, reptin, etc.) and the heat shock protein HSP90.98-102 These factors may be more generally required for RNP biogenesis and, like that of the other assembly factors, their precise mechanism of action remains to be determined.

**Dyskeratosis Congenita**

**Overview**

H/ACA RNPs have gained significant attention due to their association with the bone marrow failure syndrome dyskeratosis congenita (DC). DC is a rare but often fatal inherited disease leading to stem cell loss particularly in rapidly proliferating tissues such as the bone marrow, skin and intestine.103,104 It is mainly characterized by bone marrow failure and the mucocutaneous triad of abnormal skin pigmentation, nail dystrophy and mucosal leukoplakia, but also causes a predisposition to malignant tumor formation.105 DC is inherited in three patterns, X-linked recessive (accounting for ~45% of cases), autosomal recessive (~50%) and autosomal dominant (~5%). The X-linked and autosomal recessive forms usually are most severe with extreme cases of intrauterine growth retardation, whereas the autosomal dominant form is milder and can go unnoticed until the fourth or fifth decade of life. The X-linked form is caused exclusively by mutations in NAP57, which is hence also referred to as dyskerin.106,107 The autosomal recessive
form is genetically heterogeneous. Although families with mutations in NOP10, NHP2 and the telomeric factor TIN2 have been identified, the affected gene(s) of most families remain to be discovered.\textsuperscript{108-111} The autosomal dominant form is due to mutations in the telomerase RNA and reverse transcriptase genes.\textsuperscript{112,113}

**Pathogenesis**

Although DC patients of all inheritance patterns exhibit shortened telomeres in peripheral blood, the degree to which the other functions of H/ACA RNPs are contributing to the pathogenesis and if and how certain classes of H/ACA RNPs are preferentially impaired in the recessive forms remains to be established. The autosomal dominant form is due to haploinsufficiency of telomerase and shows disease anticipation, i.e., shorter telomeres and earlier onset in subsequent generations.\textsuperscript{112,114} The recessive forms are more complex and mouse models point to a mixture of affected H/ACA RNP functions with telomerase featured prominently.\textsuperscript{107,115-118} The level of understanding or lack thereof is perhaps best illustrated by the absence of an explanation for the molecular impact of the many mutations in NAP57 in X-linked DC.

**NAP57 Mutations**

The forty or so DC mutations identified in NAP57 cluster to its PUA domain including the C-terminus and to its N-terminus mostly avoiding the catalytic domain.\textsuperscript{119} In a model of the 3D structure (based on those from archaea) most of these mutations come together on one solvent accessible surface (at the bottom of the molecule in Fig. 1B).\textsuperscript{47,48} Despite their location in the PUA domain, the mutations apparently fail to impact the binding of the ACA triplet of the H/ACA RNAs. Moreover, except for potential allosteric effects, the DC mutations do not impact intra-RNP protein-protein interactions. Therefore, the mutation cluster may impair the interaction of the RNP with (a) yet to be identified factor(s). Such a factor could be RNP-specific and thus explain a preferential impact on, e.g., telomerase.

**Conclusions and Anticipated Developments**

The main function of H/ACA RNPs is the modification of target RNAs and based on genetic, biochemical and more recently structural studies we have gained detailed insight into their structure and function. Some specialized aspects, such as their catalytic mechanism (see chapter by Mueller and Ferre-D’Amare) and their action on spliceosomal snRNPs (see chapter by Karijolich et al) are discussed in separate chapters of this book. In particular, two aspects have boosted research into H/ACA RNPs, first, their involvement in an inherited disease (DC) and, second, their forming part of mammalian telomerase. Despite the wealth of information accumulated on these five component particles, many questions remain.

Although it is clear that overall and partial pseudouridylation of ribosomal RNA is important for ribosome biogenesis and function,\textsuperscript{120-122} we are far from understanding the importance of individual modifications, e.g., is it really the modification that matters or is it the action (hybridization) of the respective H/ACA RNP on (to) the target site? In the future, the targets and functions of orphan H/ACA RNAs will undoubtedly be unraveled potentially opening entire new areas of H/ACA RNP research.

The differences between archaeal and eukaryal H/ACA RNPs have hampered extending findings from one to the other. Although archaeal RNPs can be functionally reconstituted from recombinant components and crystallized, mammalian RNPs require assembly factors. Moreover, the structures of mammalian RNPs can be modeled based on those of the archaeal ones, but about one third of their entire RNP structure is still missing due to N- and C-terminal extensions of the individual proteins. In the future, mammalian H/ACA RNPs will need to be functionally reconstituted and crystallized from recombinant components and the action of their assembly factors determined in more detail.\textsuperscript{123} Eventually, the analysis of RNPs reconstituted from proteins with and without DC mutations and their impact on individual particles will provide insight into the molecular mechanism underlying DC.
Acknowledgements

We thank Sujayita Roy for critical reading of the manuscript. The work in the authors’ laboratory is supported by grant HL079566 (to U.T.M.) from the National Institute of Health.

References

57. Liang XH, Fournier MJ. The helicase Has1p is required for snoRNA release from prerRNA. Mol Cell Biol 2006; 26:7437-7450.


74. Henras AK, Capeyrou R, Henry Y 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.


