Front. Biol. 2018, 13(1): 1–10
https://doi.org/10.1007/s11515-018-1480-8

REVIEW

RNA-dependent pseudouridylation catalyzed by box H/ACA RNPs

Meemanage D. De Zoysa, Yi-Tao Yu (✉)

University of Rochester Medical Center, Department of Biochemistry and Biophysics, Center for RNA Biology, 601 Elmwood Avenue, Rochester, NY14642, USA

BACKGROUND: Pseudouridine (Ψ) is the most abundant post-transcriptionally modified nucleotide found in RNA. Ψ is clustered in functionally important and evolutionary conserved regions of RNAs in all three domains of life. Pseudouridylation is catalyzed by two distinct mechanisms: an RNA-independent and an RNA-dependent mechanism. The former involves a group of stand-alone protein enzymes, and the latter involves a family of complex enzymes called box H/ACA RNPs, each of which consists of one RNA (box H/ACA RNA) and a set of four core proteins. Over the years, the mechanism of RNA-dependent pseudouridylation has been extensively studied. The crystal structures of partial and complete box H/ACA RNP have been solved. However, the detailed picture of RNA-dependent pseudouridylation is still not entirely clear.
OBJECTIVE: In this work, we review what is known about box H/ACA RNP and the mechanism by which box H/ACA RNP catalyzes RNA-dependent pseudouridylation. We also discuss some examples of the dual nature and redundancy of box H/ACA RNPs that deviate from the usual mechanism.
METHODS: A methodical literature search was performed using the Pubmed central search engine and International Digital Publishing Forum (EPUB) using the following keywords: "pseudouridylation," "pseudouridine," and "box H/ACA RNP." The necessary information was extracted and cited.
RESULTS: A detailed introduction is made including the discovery, mechanism and crystal structure of box H/ACA RNP. Three sequence/structural requirements for box H/ACA RNA-guided pseudouridylation are discussed and the exceptions to those rules are explored.
CONCLUSION: Over the years, box H/ACA RNP-catalyzed pseudouridylation has been extensively studied, generating fruitful results. However, a detailed picture regarding the mechanism of this reaction is still to be deciphered. More work is needed to fully understand box H/ACA RNP-catalyzed pseudouridylation.

Introduction

Ψ was first discovered in 1951 by Cohn and Volkin in the RNA hydrolysates of calf liver (Cohn, 1959). It is the 5-ribosyl isomer of uridine, resulting from a trans-glycosylation reaction of uridine (Fig. 1). Ψ Possesses a C-C bond rather than the usual N-C bond, which links the base to the ribose sugar. The enhanced rotational freedom in the C-C bond of Ψ gives it greater conformational flexibility. On the other hand, the free N1-H in Ψ can act as an additional hydrogen bond donor, providing extra rigidity to the phosphodiester backbone by participating in a water-mediated hydrogen bonding network (Cohn, 1959; Charette and Gray, 2000). Thus, the U-to-Ψ isomerization reaction endows the modified nucleotide (Ψ) with chemical properties that are distinct from those of uridine and all other known nucleotides.

Among the more than 140 post-transcriptionally modified nucleotides, identified in RNAs from all three domains of life, Ψ is the most abundant found in virtually all cellular RNAs. It is particularly concentrated in some major types of non-coding RNAs, including tRNA (tRNA) (Hopper and Phizicky, 2003; Grosjean, 2005; Sprinzl and Vassilenko, 2005), rRNA (rRNA) (Branlant et al., 1981; Ofengand and Fournier, 1998; Schattner et al., 2006) and spliceosomal small nuclear RNA (snRNA) (Reddy and Busch, 1988; Massenet et al., 1998; Wu et al., 2011; Yu et al., 2005, 2011) . Recently, Ψ has also been found in mRNA by high throughput sequencing methods (Carlile et al., 2014; Lovejoy et al., 2014; Schwartz et al., 2014;Li et al., 2015;). Ψ is not only abundant in cellular non-coding RNAs, it also almost always resides in regions that are highly conserved across species from yeast to humans. Due to its high abundance and conservation, Ψ is believed to be functionally important, and indeed, experi-

Received December 15, 2017; accepted January 15, 2018
Correspondence: Yi-Tao Yu
E-mail: yitao_yu@urmc.rochester.edu

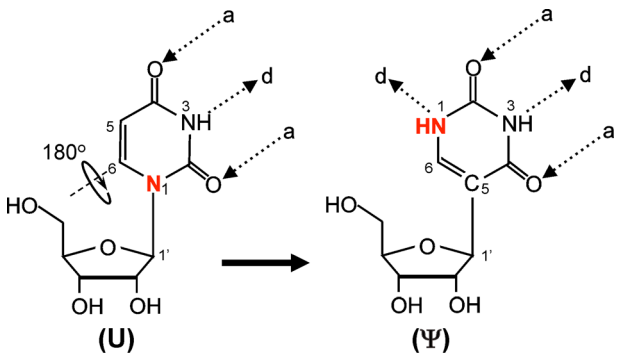

Figure 1 Pseudouridylation (U-to-Ψ Isomerization). The nucleotide pseudouridylation reaction is schematized. The N1-C1' bond in uridine is broken. The uracil base is lifted up and turned 180° around the C6-N3 axis. Upon reformation of the C5-C1' bond, Ψ is generated. "a" and "d" represent hydrogen bond acceptor and donor, respectively.

mental results indicate that Ψs in rRNAs and spliceosomal snRNAs play important roles in protein translation and pre-mRNA splicing, respectively.

There are around 44 Ψs discovered in *S. cerevisiae* rRNA and about 97 Ψs in vertebrate rRNA (Ofengand and Fournier, 1998; Schattner et al., 2004; Torchet et al., 2005; Schattner et al., 2006). It has been shown that deletion of the snoRNAs (small nucleolar RNA) responsible for Ψ formation (and hence the loss of Ψs) at some important sites of rRNA can have profound impact on rRNA processing, protein synthesis, and subsequently cell growth (King et al., 2003; Pieknа-Przybylska et al., 2008; Liang et al., 2009b). It has also been shown that hypo-pseudouridylated rRNA affects the ribosome-tRNA binding and translational fidelity (Jack et al., 2011).

There are a total of 24 Ψs in the five vertebrate snRNAs, including U1, U2, U4, U5, and U6, and 13 of them are clustered in U2 snRNA. In the *S. cerevisiae* spliceosomal snRNA, six Ψs have been identified, three in U2, two in U1 and one in U5 (Yu and Meier, 2014). Most (if not all) of these Ψs are in the evolutionary conserved regions and are shown to be functionally important as well. For example, it has been reported that the Ψs in the 5′ end region and in the branch site recognition region are important for pre-mRNA splicing (Yu et al., 1998; Dönmez, Hartmuth and LÜHrmann, 2004; Zhao and Yu, 2004; Yang et al., 2005). Recently, it has been demonstrated that yeast U2 lacking Ψ42 and Ψ44 in the branch site recognition region is deficient in interacting with Prp5 ATPase, thus resulting in inefficient spliceosome assembly and splicing (Wu et al., 2016). Pseudouridylation of U6 snRNA at position 28 has also been shown to be important for the filamentous growth in yeast, demonstrating the importance of Ψ in development (Basak and Query, 2014).

RNA pseudouridylation (U-to-Ψ isomerization) is catalyzed by pseudouridine synthases (PUSs), which are a group of stand-alone protein enzymes, or by box H/ACA RNPs, which are a family of RNA-protein complexes each containing one unique box H/ACA RNA and four common core proteins (Cbf5, Nhp2, Nop10, and Gar1; Cbf5 is the pseudouridylation enzyme catalyzing the chemical reaction). While the former is an RNA-independent reaction in which the PUSs are capable of both recognizing/specifying the target uridine and catalyzing the chemical reaction, the latter is an RNA-dependent mechanism, where the RNA component (box H/ACA RNA) of the RNP serves as a guide that base-pairs with the substrate RNA and specifies the uridine to be subsequently pseudouridylated by Cbf5. In this review, we focus on the RNA-dependent mechanism.

The discovery of box H/ACA RNP components

In 1996, the Fournier group first reported the existence of two families of small RNAs in the nucleolus (snoRNAs). One is the C/D family and the other they named the "ACA family" because they identified a ACA triplet positioned 3 nucleotides upstream of the 3′ terminus of each member of this family (Balakin et al., 1996). Mutational analysis showed that the ACA triplet together with an adjacent upstream helix is required for snoRNA accumulation in yeast (Balakin et al., 1996). They also reported that all members of the ACA family are bound to Gar1, an essential, glycine/arginine-rich nucleolar protein (Girard et al., 1992). The ACA box might serve as a binding site for some specific proteins, and Gar1 probably serves as a key accessory protein.

The following year, the Kiss group characterized nine novel, intron–encoded human ACA snoRNAs with high-order nucleolar structures. Importantly, by cleaving snoRNA with single- and double-strand-specific nucleases under non-denaturing conditions, they deduced the secondary structure of U64. According to that model the ACA snoRNA consists of two major hairpin domains connected by a single-stranded hinge region, and the 3′ hairpin is followed by the consensus ACA motif (Fig. 2). For the first time with a systematic computer-aided modeling of all known ACA snoRNAs, they found that the majority of mammalian ACA snoRNAs share a common hairpin-hinge-hairpin-tail secondary structure (Fig. 2). Notably, there is an internal loop in each hairpin. Apart from the conserved ACA motif in the snoRNA, they identified another purine rich consensus sequence of ANANNA in the single-stranded hinge region, and because it is positioned in the hinge region, the ANANNA motif was named "Box H." Due to its conserved nature, box H was also considered to be a protein binding site, presumably playing an important role in processing, accumulation and function of ACA snoRNAs (Ganot et al., 1997).

Soon after, both the Fournier group and the Kiss group experimentally showed that the ACA family of snoRNAs was required for site-specific Ψ formation in rRNA. Specifically,

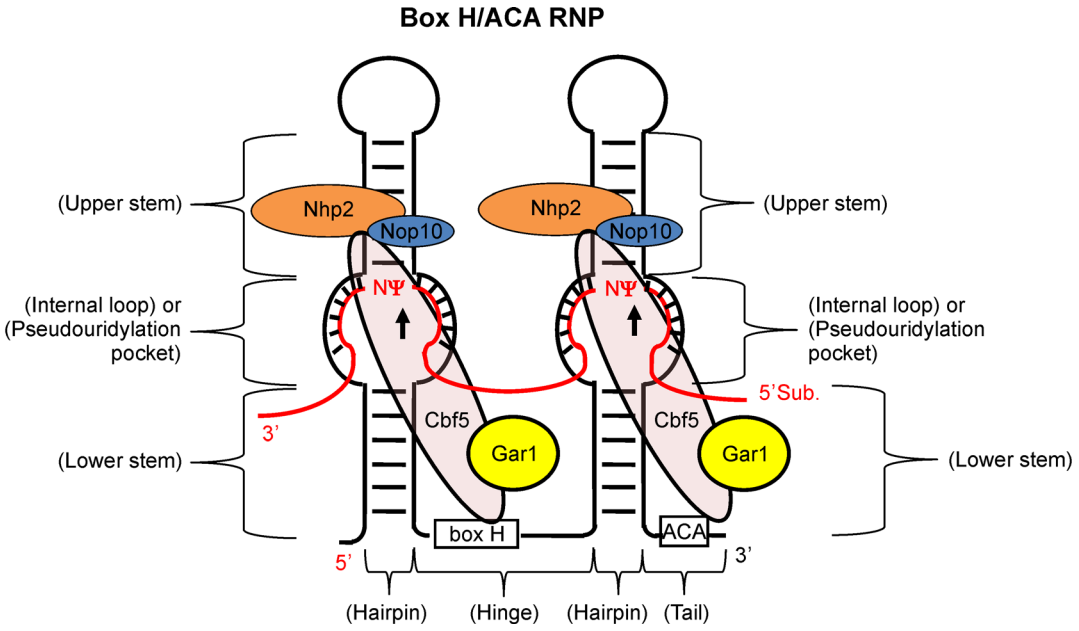

Figure 2 Box H/ACA RNP, a highly complex pseudouridylase that catalyzes RNA-dependent RNA pseudouridylation. The components of the RNP, including a box H/ACA RNA and four core proteins (Cbf5, Nhp2, Nop10 and Gar1), are shown. The two hairpins, the hinge region and the 3′ ACA tail of the box H/ACA RNA are indicated. The upper and lower stems, as well as the internal loop (or pseudouridylation pocket) of each hairpin are also indicated. Also depicted are the substrate RNA, which forms base-pairing interactions with the pseudouridylation pocket sequences, and the target uridines, which are to be pseudouridylated (Ψs, the arrows).

the Fournier laboratory discovered ten ACA snoRNAs that were required for site-specific pseudouridylation at 11 positions in rRNA. Using a computer folding algorithm, they also predicted that ACA snoRNAs can be folded into a common core structure that includes two stem-loop domains at the 5′ and 3′ ends and a single-stranded hinge region in the middle of the two stem-loops (Fig. 2). They made two major predictions: (1) Base-pairing is necessary between the snoRNA and the complementary rRNA that flanks the site of pseudouridylation. (2) A near constant distance is required between the target uridine and the ACA sequence element. The complementarity between snoRNA and rRNA was confirmed by mutational analysis. The distance between the target uridine and the ACA box was determined to be approximately 15 nucleotides, and this constant distance appears to be a key determinant of site selection (Ni et al., 1997) . The Kiss group, based on their structural mapping results, coined the name box H/ACA snoRNA and experimentally proved that these snoRNAs function as guide RNAs in yeast rRNA pseudouridylation. More specifically, they showed that the internal loop, located in each hairpin, serves as a guide that base-pairs with the rRNA sequences preceding and following a pseudouridylation site, specifically positioning the target uridine and its 3′ adjacent nucleotide at the base of the upper stem of the hairpin and leaving them unpaired (Fig. 2). Clearly, this base-pairing interaction between the internal loop of the snoRNA hairpin and rRNA determines the target (substrate) specificity. Because of its function, the internal loop is also referred to as pseudouridylation pocket (Ganot et al., 1997).

Each box H/ACA guide RNA is complexed with a set of four common core proteins (Fig. 2). Besides Gar1, which was identified earlier by the Fournier group, three additional proteins were identified. Cbf5 (dyskerin in humans, NAP57 in rats, and Nop60B in *Drosophila*) is the key protein component that acts as the pseudouridylase and is the most well studied protein component in the box H/ACA RNP complex. It was first identified as low affinity centromere binding protein and was essential for cell viability (Jiang et al., 1993). The function of Cbf5 protein was elucidated by its homology to the *E. coli* pseudouridine synthase TruB (Nurse et al., 1995; Becker et al., 1997) and its association with box H/ACA RNA (Zebarjadian et al., 1999). Later, a detailed analysis of Cbf5 was conducted by various groups and they showed that Cbf5 contains a catalytic domain, which is common to all known pseudouridine synthases, as well as a PUA (pseudouridine synthase and archaeosine transglycosylase) domain, which is common to most of the TruB family PUSs. Importantly, there is a strictly conserved and essential aspartic acid residue positioned within the active site of Cbf5 (Zebarjadian et al., 1999; Hamma et al., 2005; Terns and Terns, 2006; Reichow et al., 2007).

Nhp2 (non-histone protein) and Nop10 are the other two protein components of box H/ACA RNP. Both Nhp2 and Nop10 are small basic proteins that were initially identified in 1998 by affinity chromatography with epitope-tagged Gar1

protein (Henras et al., 1998). The same group showed that cells lacking Nhp2 and Nop10 are impaired in global rRNA pseudouridylation (Henras et al., 1998; Watkins et al., 1998) . Detailed analysis indicated that all four core proteins play important roles in RNP biogenesis, stability and catalyzing pseudouridylation reaction (Li and Ye, 2006a; Li et al., 2011) .

Crystal structures and reconstitution of box H/ACA RNP

Several structures of partial and complete archaea and yeast box H/ACA RNP, with or without substrate RNA, have been solved (Hamma et al., 2005; Li and Ye, 2006b;Manival et al., 2006; Rashid et al., 2006; Liang et al., 2007; Ye, 2007; Liang et al., 2009a; Duan et al., 2009;Li et al., 2011). Li and Ye in 2006 reported the first crystal structure of an entire archaeal Pyrococcus furiosus H/ACA RNP, including the four core proteins, L7Ae (Nhp2 in eukaryotes), Cbf5, Nop10, and Gar1, as well as a 65-nucleotide long single-hairpin H/ACA RNA. In this structure (without substrate RNA), the pseudouridylation pocket is either disordered or stabilized by crystal packing interactions. Importantly, this structure also revealed that the PUA domain of Cbf5 recognizes the lower stem and the ACA box, placing the internal pseudouridylation pocket near the active site of Cbf5 (Li and Ye, 2006a) . Subsequently, the Li group solved the crystal structure of substrate-bound partial archaeal box H/ACA RNP complex (lacking protein L7Ae). They reported that, upon interaction with the substrate RNA, the pseudouridylation pocket of box H/ACA guide RNA became ordered by making specific contacts with the protein Cbf5 (Liang et al., 2007; Li, 2008) .

The crystal structure of substrate-bound complete archaeal box H/ACA RNP was reported in 2009 by the Ye group (Duan et al., 2009). Interestingly, the structure revealed some features that were absent in the previously solved structure of substrate-bound partial box H/ACA RNP lacking L7Ae (Liang et al., 2007). In particular, the structure of substrate-bound complete RNP revealed how the substrate RNA was recognized and how the target uridine was positioned at the active site of Cbf5 (such a detailed picture was not clearly revealed in the structure of substrate-bound partial box H/ACA RNP lacking L7Ae) (Duan et al., 2009).

Even though we can deduce the structure of eukaryotic box H/ACA RNP from the structural information of archaeal box H/ACA RNP, there are several differences between the two systems. Eukaryotic box H/ACA RNP almost exclusively consists of two hairpins and lacks the K-turn (kink turn) motif, a major structural feature of archaeal box H/ACA RNA that is recognized by L7Ae (Nhp2 in eukaryotes) (Rozhdestvensky et al., 2003). Nhp2 in eukaryotes associates with Cbf5-Nop10 protein complex (Meier, 2005) whereas L7Ae contacts only Nop10 in the archaeal complex (Li and Ye, 2006a). Another difference is that eukaryotic Nop10 lacks the cysteine residues that help coordinate a zinc ion in archaeal Nop10 (Hamma et al., 2005; Khanna et al., 2006). Therefore, it was important to solve the crystal structure of eukaryotic box H/ACA RNP complex.

In 2011, Ye and colleagues succeeded in reconstituting a functional yeast (*S. cerevisiae*) box H/ACA RNP using bacteria-expressed recombinant proteins, and solved the crystal structure of a partial complex consisting of Cbf5, Nop10, Gar1 and snR5, a two-hairpin H/ACA RNA (Li et al., 2011). Based on functional analysis, yeast Nhp2 did not appear to be as important as L7Ae in archaeal box H/ACA RNPs, as omission of Nhp2 virtually did not impact pseudouridylation activity. This result indicates that Nhp2 and L7Ae play a different role in facilitating box H/ACA RNP assembly and pseudouridylation in yeast and archaea, respectively (Li et al., 2011). On the other hand, even though Gar1 is important for substrate turnover, the omission of Gar1 from yeast RNP complex abolished most of the activity even under single turnover conditions. This is in contrast to the archaeal box H/ACA RNP, which is still functionally active in the absence of Gar1, suggesting a more important role of Gar1 in pseudouridylation in eukaryotes (Baker et al., 2005; Charpentier et al., 2005). It is also worth noting that the eukaryotic Cbf5 has both an N-terminal and C-terminal extensions compared to its archaeal counterpart, and the N-terminal extension appeared to form an additional important structural layer in the PUA domain (Li et al., 2011).

Sequence and structural elements of box H/ACA RNA required for pseudouridylation

To date many laboratories have predicted and experimentally tested the ability of various box H/ACA RNAs (wild-type and mutants) to direct RNA pseudouridylation, yet, despite the large amount of data and information generated over the years, the picture regarding box H/ACA RNA sequence and structural requirements for pseudouridylation remains fragmentary and incomplete. Sometimes it is still difficult to determine the correct target pseudouridylation site based on prediction algorithms currently available. Even though all box H/ACA RNAs fold into a conserved secondary structure (hairpin-hinge-hairpin-tail) and bound to the same set of core proteins, not all of them have assigned substrates (Hüttenhofer et al., 2001; Vitali et al., 2003). Therefore, substantial efforts have been made to understand how a box H/ACA guide RNA recognizes and specifies a particular uridine to be modified.

Based on computational alignments and experimental results, three sequence/structural elements within a box H/ACA RNA were identified to be important for guiding pseudouridylation (Ge and Yu, 2013b; Xiao et al., 2009). The first element is the distance between the ACA or H box and substrate uridine (Bortolin et al., 1999). It was first

demonstrated that there was a constant distance of 15 nucleotides between the ACA box and substrate uridine (Fig. 2). Insertion or deletion of nucleotides between the ACA box and the target uridine led to either alteration in site specificity or reduction in pseudouridylation at the original site (Ni et al., 1997). By examining the increased number of experimentally verified box H/ACA RNAs, Xiao et al. were able to confirm a consensus optimal distance of 14-16 nucleotides between the substrate uridine and the ACA or H box (Xiao et al., 2009). It turns out that this spatial distance is important for the binding of the Cbf5 PUA domain to the ACA/H box and positioning of the substrate/target uridine in the catalytic center of Cbf5 (Duan et al., 2009; Xiao et al., 2009; Ge and Yu, 2013a).

The second element is the stability of the hairpin structure harboring the pseudouridylation pocket (internal loop) (Fig. 2). The formation of the hairpin depends on the stability of the upper and lower stems flanking the pseudouridylation pocket. In this regard, it was experimentally verified when an imperfect upper stem (with mismatches) of a non-functional box H/ACA RNA was converted to a perfect upper stem by point mutations (Watson-Crick base-pairing was re-established), the box H/ACA RNA became functional (Xiao et al., 2009).

The third element is the sufficient and proper Watson-Crick base pairing interactions between the guide sequence of box H/ACA RNA and substrate RNA (Fig. 2). In 1996, upon checking 10 ACA snoRNAs, the Fournier group identified complementarities between the snoRNAs and the rRNA sequences at the pseudouridylation sites, eventually leading to the current model of box H/ACA RNA-guided RNA pseudouridylation (Ni et al., 1997). In 2009, the Yu group tested this third element experimentally and systematically by making mutations in the pseudouridylation pocket to alter the base pairing interactions with the substrate RNA. By increasing the complementarity, thereby strengthening the interaction of box H/ACA RNA with substrate RNA, they again demonstrated, both in vitro and in yeast cells, that sufficient base-pairing between the guide sequence and the substrate sequence is required for box H/ACA RNA-guided RNA pseudouridylation (Xiao et al., 2009).

Inducible pseudouridylation catalyzed by box H/ACA RNP

In a study of dynamic pseudouridylation in yeast cells, the Yu group found that a box H/ACA RNP was also able to catalyze inducible RNA pseudouridylation under stress conditions (Wu et al., 2010). Specifically, upon starvation or nutrient deprivation, the yeast box H/ACA RNP snR81 became capable of pseudouridylating yeast U2 snRNA at position 93, which is not modified under normal conditions. Interestingly, while the 5′ pseudouridylation pocket of snR81 remained specific for the uridine at position 42 of U2, the 3′ pocket, which is normally specific for the uridine at position 1051 of 25S rRNA, loosened its specificity to also include U93 of U2. Sequence alignment revealed two mismatches in the base-pairing interactions between the guide sequence and the U2 sequence surrounding U93 (Fig. 3). When one (or both) of the mismatches was converted to perfect Watson-Crick base-pairing, U93-to-Ψ93 conversion became constitutive. Detailed analysis indicated that the two mismatches are necessary but not sufficient for induced pseudouridylation mediated by snR81. This work suggests the multifaceted role of box H/ACA RNP in pseudouridylation (Wu et al., 2010).

Redundancy and dual functionality of box H/ACA RNA

With the advancement of knowledge in box H/ACA RNA, several groups identified, using computational and experimental (RNomics) methodologies, a large number of putative box H/ACA RNAs in eukaryotic cells, although only a few were experimentally verified (Hüttenhofer et al., 2001; Schattner et al., 2004; Torchet et al., 2005;Schattner et al., 2006). For example, in yeast, a complete set of 28 snoRNAs were identified that guide the U-to-Ψ conversion at 44 sites in 18S and 28S rRNA (Torchet et al., 2005). In humans, experimental and computational approaches have identified 82 putative box H/ACA RNAs for the 97 pseudouridylation sites in rRNA (Schattner et al., 2004; Schattner et al., 2006). Even though in most cases, a single box H/ACA RNA has been assigned to a particular pseudouridylation site, it is quite possible that some other box H/ACA RNAs exhibit similar (or identical) guide sequences to target the same site (Schattner et al., 2004; Schattner et al., 2006). Conversely, it has also been proposed that a single box H/ACA RNA can fold into more than one slightly different structures and can thus adjust their guide pocket sequences to target more than one substrates, thereby directing U-to-Ψ conversion at multiple sites (Schattner et al., 2004).

To gain a better understanding of these phenomena, the Yu group experimentally tested a number of predicted box H/ACA RNAs using both *in vitro* and *in vivo* systems. While some of these predicted guide RNAs shared similar guide sequences and therefore could target the same sites, the others were each predicted to fold into alternate structures exhibiting different guide sequences that could target different sites (Xiao et al., 2009). For instance, the 5′ guide pocket of human ACA19 box H/ACA RNA was predicted to fold into three alternate pockets with three guide sequences directing the pseudouridylation of 18S rRNA at positions U863 and U866 and 28S rRNA at position U3709, respectively (Fig. 4). The same sites (U863 and U866) were also predicted to be targeted by the 5′ pocket of ACA24 box H/ACA RNA and the 3′ pocket of ACA28 box H/ACA RNA, respectively. Target redundancies were also predicted for ACA67 box H/ACA RNA and ACA42 box H/ACA RNA in targeting U109 and

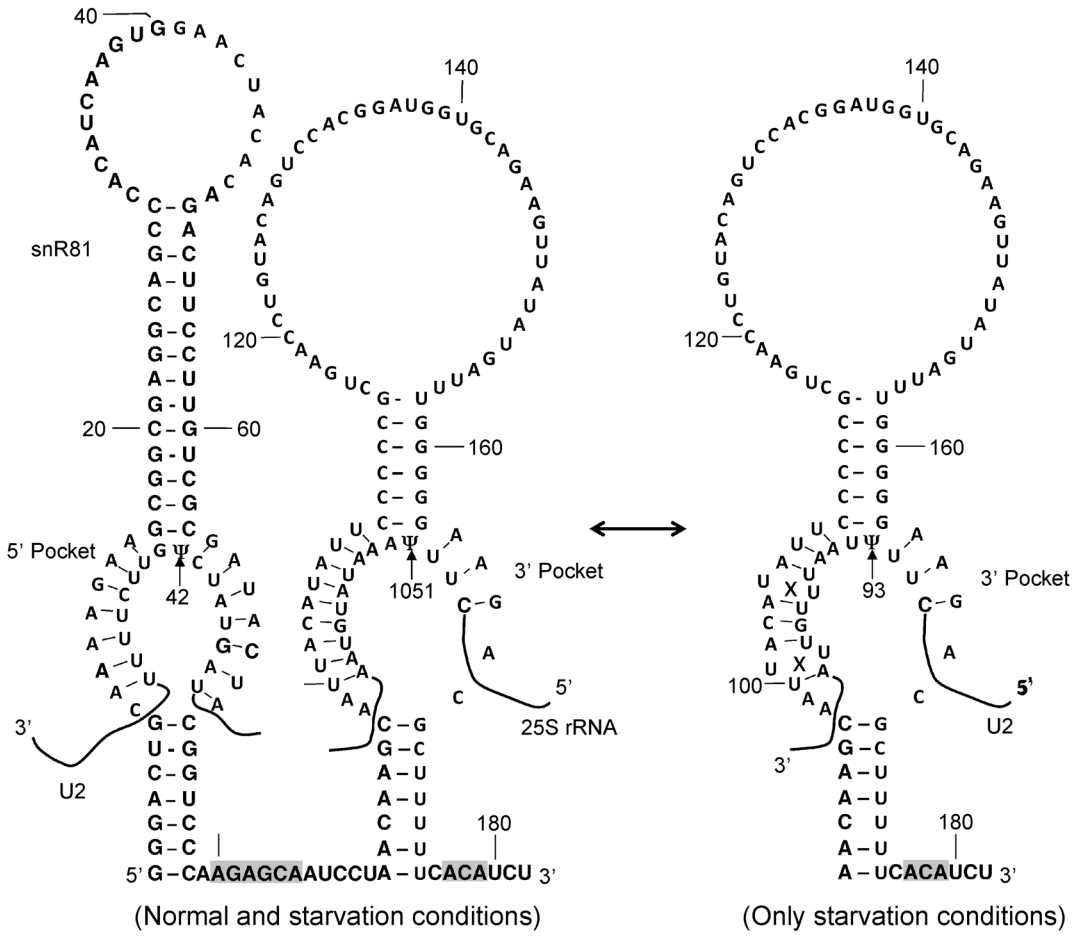

Figure 3 Box H/ACA RNA-guided RNA pseudouridylation under normal and stress conditions. Shown is the *S. cerevisiae* snR81 box H/ACA RNA, which guides the pseudouridylation of U2 snRNA at position 42 (the 5′ pocket) and 25 S rRNA at position 1051 (the 3′ pocket) under normal and starvation conditions. Under starvation conditions, the 3′ pocket of snR81 loosens its specificity to also guide U2 pseudouridylation at a novel site, position 93. The substrate RNAs (U2 snRNA and 25S rRNA), which form base-pairing interactions with the sequences of the pseudouridylation pockets, and the target nucleotides to be pseudouridylated (Ψs, the arrows) are also shown. The H and ACA boxes of snR81 are indicated (gray boxes) as well.

U572 in 18S rRNA, and for ACA50 box H/ACA RNA and ACA62 box H/ACA RNA in targeting U34 and U105 in 18S rRNA. While many of the predictions tested were correct, some were not. Remarkably, for those that were incorrect, minor changes made to box H/ACA RNAs (only a few nucleotide changes in the three elements discussed above) can turn the incorrect predictions to correct ones. For example, the 5′ pocket of ACA42, which did not modify U572 in 18S rRNA, became perfectly functional when the imperfect upper stem of the hairpin was changed to a perfect stem, and the complementarity between the guide sequence and the substrate RNA was enhanced. Among the three predicted alternate structures of the ACA19 5′ guide pocket, only one was correct that guides 28S rRNA pseudouridylation at position U3709. By improving the base-pairing interactions between the guide pocket and the substrate RNA, they were able to turn the ACA19 5′ guide pocket to direct 18S rRNA pseudouridylation at position U863. Along this line, it has been reported that several *Drosophila* box H/ACA RNAs can probably also fold into alternative structures with distinct pseudouridylation pockets targeting different sites (Deryusheva and Gall, 2013). In addition, the Gall group also showed that the 5′ pocket of U2-34/44 (a box H/ACA RNA, also known as U92) was responsible for U2 pseudouridylation at position 34, whereas the 3′ pocket was able to guide U2 pseudouridylation at two adjacent positions 43 and 44. From these data, it is clear that one guide RNA pocket can act on different substrates depending on the structure of the guide RNA and the binding affinity between the guide RNA and its substrates (Xiao et al., 2009). On the other hand, there can be multiple guide RNAs targeting a single site as long as they share similar (or same) guide sequences (Xiao et al., 2009).

Several lines of evidence also suggest that there can co-exist, in cells, an RNA-dependent and RNA-independent pseudouridylation activities targeting the same site of an RNA for modification (Fig. 5). For instance, the Yu group observed that when *Xenopus* pugU2-34/44, (a box H/ACA RNA responsible for U2 pseudouridylation at positions 34

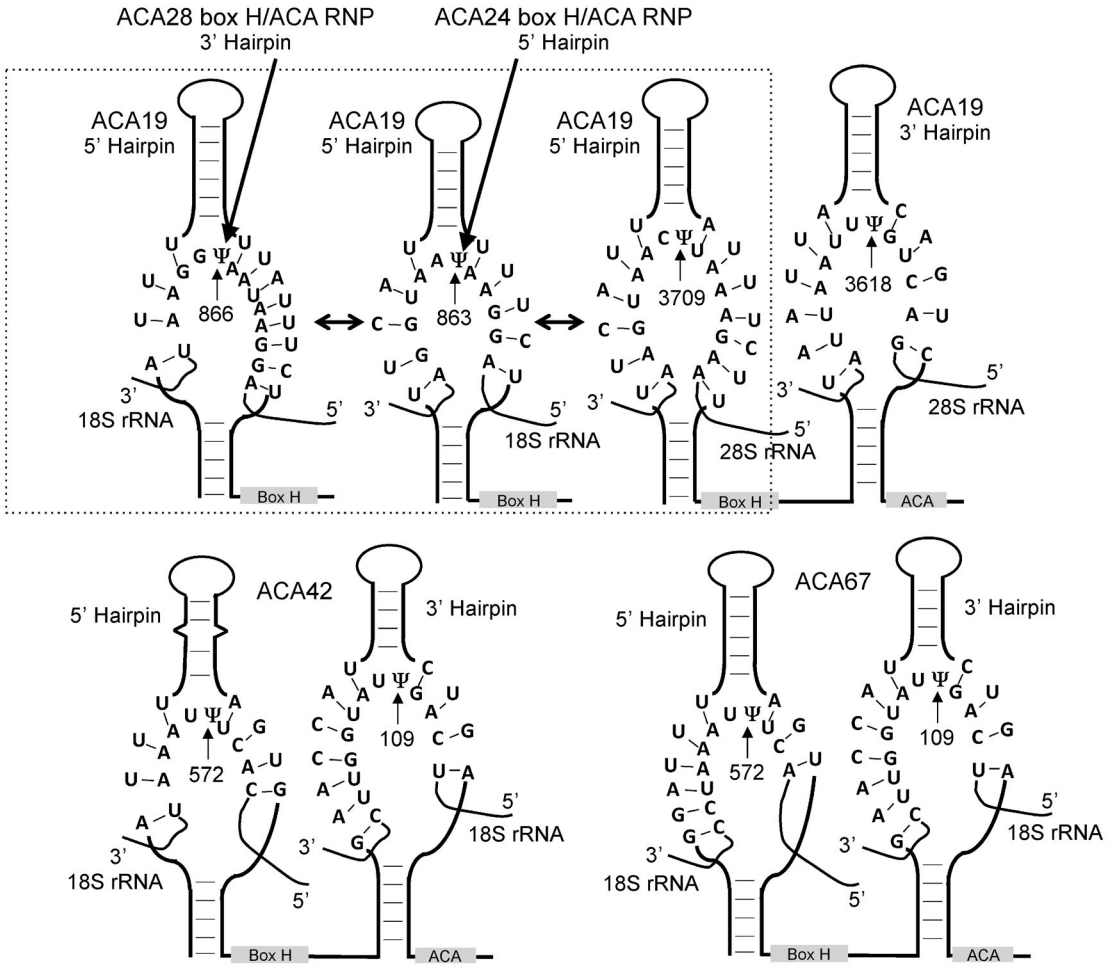

Figure 4 Intricate box H/ACA RNA-substrate network. Several mammalian box H/ACA RNAs, including ACA19, ACA42 and ACA67, are schematized. According to computational predictions, the 5′ hairpin of ACA19 folds into three alternative structures, forming three distinct pseudouridylation pockets that target three different sites, U866 and U863 of 18S rRNA and U3709 of 28S rRNA. However, experimental results indicate that the 5′ pocket of ACA19 only targets U3709 of 28S rRNA (Xiao et al., 2009). U866 and U863 of 18S rRNA are the targets of ACA28 (its 3′ pocket) and ACA24 (its 5′ pocket), respectively (indicated) (Xiao et al., 2009). Based on computational predictions, ACA42 and ACA67 possess almost identical pseudouridylation pockets, targeting the same sites: the 5′ pocket of ACA42 and the 5′ pocket of ACA67 both target U572 of 18S rRNA, and the ACA42 3′ pocket and the ACA67 3′ pocket both target U109 of 18S rRNA. But, in reality, the 5′ pocket of ACA42 fails to guide U572-to-Ψ572 conversion (Xiao et al., 2009). Interestingly, small nucleotide changes can turn a non-functional pseudouridylation pocket into a functional pocket.

and 44), was depleted, pseudouridylation was initially abolished at positions 34 and 44, but was soon recovered at position 34 suggesting that there is another (or more than one) independent pseudouridylation activity that targets position 34 (Zhao et al., 2002). Indeed, it was later reported that an artholog of yeast Pus7, which pseudouridylates *S. cerevisiae* U2 at position 35 (equivalent to position 34 of vertebrate U2), was also identified in *Xenopus laevis* (Ma et al., 2003) and unpublished work). Likewise, the Gall group showed that depletion of *Drosophila* pugU2-35/45, an ortholog of *Xenopus* pugU2-34/44, did not block U2 pseudouridylation at position 35 in *Drosophila* cells (positions 35 and 45 of *Drosophila* U2 are equivalent to positions 34 and 44, respectively, of vertebrate U2). They suggested that Ψ35 formation in the absence of pugU2-35/45 could be due to an endogenous pseudouridine synthase enzyme (an artholog of yeast Pus7) that also pseudouridylates U2 at position 35 (Deryusheva and Gall, 2013). The Gall group also showed that pugU2-34/44 (also known as SCARNA8 or U92 scaRNA) and Pus1 (stand-alone protein enzyme) were redundant in targeting U2 pseudouridylation at position 43 in vertebrates. While Pus1 specifically catalyzes U2 pseudouridylation at position 43, pugU2-34/44 (SCARNA8) targets three different positions, 34, 43 and 44. In *pombe*, U2 pseudouridylation at position 43 is also targeted by two different enzymatic activities, Pus1, a stand-alone protein enzyme, and *sp*pugU2-43/44, a box H/ACA RNP (Deryusheva and Gall, 2017) (Fig. 5).

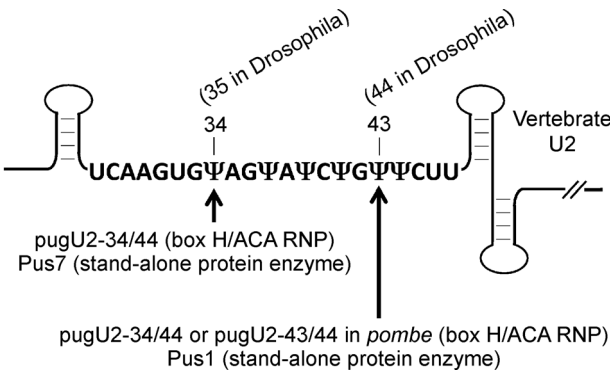

Figure 5 Redundancy between stand-alone pseudouridine synthases and box H/ACA RNP pseudouridylases. Vertebrate U2 RNA is schematized and its branch site recognition region sequence is shown. Pus7 (stand-alone protein enzyme) and the 5′ pocket of pugU2-34/44 (a box H/ACA RNP also known as SCARNA8 or U92 scaRNP) both target U34 (equivalent to U35 in Drosophila) of U2. Pus1 (stand-alone protein enzyme) and pugU2-34/44 (or pugU2-43/44 in *pombe*) both target U43 (equivalent to U44 in Drosophila) of U2 (Deryusheva and Gall, 2017).

Concluding remarks

Since their discovery in 1996, box H/ACA RNPs have been extensively studied. Particular attention has been paid to the molecular mechanism by which box H/ACA RNPs recognize and convert specific uridines into Ψs within RNAs. While a great deal of progress has been made, there are still a number of questions that remain unanswered. For instance, we now know that there are a large number of box H/ACA RNPs (either computer predicted or experimentally identified) in the cell that target various types of RNA for pseudouridylation. However, only a small fraction of these RNPs have been experimentally tested and their function verified. With regards to the details of box H/ACA RNA-guided pseudouridylation, our knowledge of the rules of this modification remains incomplete and fragmentary, despite the fact that the structures of the RNP complexes have been solved, and that the three key elements (a constant distance between the target uridine and box H or between the target uridine and box ACA, a stable hairpin that harbors the pseudouridylation pocket, and a stable interaction of the pseudouridylation pocket with its substrate) have been experimentally identified. In addition, it remains puzzling that in order to target multiple sites, some box H/ACAs fold into alternative hairpin structures, some of which are apparently unstable. Even less is understood as to the roles that the protein components of box H/ACA RNPs play during pseudouridylation. What is known, however, is the fact that some diseases are associated with mutations in box H/ACA RNP proteins. One example is the disease X-linked Dyskeratosis Congenita (X-DC). X-DC is a rare inherited syndrome caused by mutations in the DKC1 gene encoding the box H/ACA RNP protein dyskerin (NAP57 in rats and Cbf5 in yeast) (Mitchell et al., 1999), thus linking a human disease to box H/ACA RNPs and pseudouridylation.

It is clear that the experimental data accumulated thus far only scratch the surface of the complexity of box H/ACA RNA-guided RNA pseudouridylation. More attention is needed to understand this complex modification.

Acknowledgments

We thank the members of the Yu laboratory for helpful discussions on the manuscript. The work carried out in the Yu laboratory was supported by grant GM104077 from NIH (to Y-T Yu).

Compliance with ethics guidelines

Yi-Tao Yu and Meemanage D. De Zoysa declare that they have no conflict of interest. This manuscript is a review article and does not involve a research protocol requiring approval by the relevant institutional review board or ethics committee.

References

Baker D L, Youssef O A, Chastkofsky M I R, Dy D A, Terns R M, Terns M P (2005). RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. Genes Dev, 19(10): 1238–1248

Balakin A G, Smith L, 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(5): 823–834

Basak A, Query C C (2014). A pseudouridine residue in the spliceosome core is part of the filamentous growth program in yeast. Cell Reports, 8(4): 966–973

Becker H F, Motorin Y, Planta R J, Grosjean H (1997). The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of psi55 in both mitochondrial and cytoplasmic tRNAs. Nucleic Acids Res, 25(22): 4493–4499

Bortolin M L, Ganot P, Kiss T (1999). Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. EMBO J, 18(2): 457–469

Branlant C, Krol A, Machatt M A, Pouyet J, Ebel J P, Edwards K, Kössel H (1981). Primary and secondary structures of *Escherichia coli* MRE 600 23S ribosomal RNA. Comparison with models of secondary structure for maize chloroplast 23S rRNA and for large portions of mouse and human 16S mitochondrial rRNAs. Nucleic Acids Res, 9 (17): 4303–4324

Carlile T M, Rojas-Duran M F, Zinshteyn B, Shin H, Bartoli K M, Gilbert W V (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature, 515 (7525): 143–146

Charette M, Gray M W (2000). Pseudouridine in RNA: what, where, how, and why. IUBMB Life, 49(5): 341–351

Charpentier B, Muller S, Branlant C (2005). Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridyla-

tion. Nucleic Acids Res, 33(10): 3133–3144

Cohn W E (1959). 5-Ribosyl uracil, a carbon-carbon ribofuranosyl nucleoside in ribonucleic acids. Biochim Biophys Acta, 32: 569–571

Deryusheva S, Gall J G (2013). Novel small Cajal-body-specific RNAs identified in Drosophila: probing guide RNA function. RNA, 19(12): 1802–1814

Deryusheva S, Gall J G (2017). Dual nature of pseudouridylation in U2 snRNA: Pus1p-dependent and Pus1p-independent activities in yeasts and higher eukaryotes. RNA, 23(7): 1060–1067

Dönmez G, Hartmuth K, Lührmann R (2004). Modified nucleotides at the 5′ end of human U2 snRNA are required for spliceosomal E-complex formation. RNA, 10(12): 1925–1933

Duan J, Li L, Lu J, Wang W, Ye K (2009). Structural mechanism of substrate RNA recruitment in H/ACA RNA-guided pseudouridine synthase. Mol Cell, 34(4): 427–439

Ganot P, Bortolin M L, Kiss T (1997a). Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. Cell, 89(5): 799–809

Ganot P, Caizergues-Ferrer M, 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(7): 941–956

Ge J, Yu Y T (2013). RNA pseudouridylation: new insights into an old modification. Trends Biochem Sci, 38(4): 210–218

Girard J P, Lehtonen H, Caizergues-Ferrer M, Amalric F, Tollervey D, Lapeyre B (1992). GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast. EMBO J, 11(2): 673–682

Grosjean H (2005). Modification and editing of RNA: historical overview and important facts to remember. In: H. Grosjean, ed. Fine-Tuning of RNA Functions by Modification and Editing. Berlin: Springer Berlin Heidelberg, pp. 1–22

Hamma T, Reichow S L, Varani G, Ferré-D'Amaré A R (2005). The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. Nat Struct Mol Biol, 12(12): 1101–1107

Henras A, Henry Y, Bousquet-Antonelli C, Noaillac-Depeyre J, Gélugne J P, Caizergues-Ferrer M (1998). Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. EMBO J, 17(23): 7078–7090

Hopper A K, Phizicky E M (2003). tRNA transfers to the limelight. Genes Dev, 17(2): 162–180

Hüttenhofer A, Kiefmann M, Meier-Ewert S, O'Brien J, Lehrach H, Bachellerie J P, Brosius J (2001). RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. EMBO J, 20(11): 2943–2953

Jack K, Bellodi C, Landry D M, Niederer R O, Meskauskas A, Musalgaonkar S, Kopmar N, Krasnykh O, Dean A M, Thompson S R, Ruggero D, Dinman J D (2011). rRNA pseudouridylation defects affect ribosomal ligand binding and translational fidelity from yeast to human cells. Mol Cell, 44(4): 660–666

Jiang W, Middleton K, Yoon H J, Fouquet C, Carbon J (1993). An essential yeast protein, CBF5p, binds in vitro to centromeres and microtubules. Mol Cell Biol, 13(8): 4884–4893

Khanna M, Wu H, Johansson C, Caizergues-Ferrer M, Feigon J (2006). Structural study of the H/ACA snoRNP components Nop10p and the 3′ hairpin of U65 snoRNA. RNA, 12(1): 40–52

King T H, Liu B, McCully R R, 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(2): 425–435

Li H (2008). Unveiling substrate RNA binding to H/ACA RNPs: one side fits all. Curr Opin Struct Biol, 18(1): 78–85

Li L, Ye K (2006a). Crystal structure of an H/ACA box ribonucleoprotein particle. Nature, 443(7109): 302–307

Li L, Ye K (2006b). Crystal structure of an H/ACA box ribonucleoprotein particle. Nature, 443(7109): 302–307

Li S, Duan J, Li D, Yang B, Dong M, Ye K (2011). Reconstitution and structural analysis of the yeast box H/ACA RNA-guided pseudouridine synthase. Genes Dev, 25(22): 2409–2421

Li X, Zhu P, Ma S, Song J, Bai J, Sun F, Yi C (2015). Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. Nat Chem Biol, 11(8): 592–597

Liang B, Xue S, Terns R M, Terns M P, Li H (2007). Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex. Nat Struct Mol Biol, 14(12): 1189–1195

Liang B, Zhou J, Kahen E, Terns R M, Terns M P, Li H (2009a). Structure of a functional ribonucleoprotein pseudouridine synthase bound to a substrate RNA. Nat Struct Mol Biol, 16(7): 740–746

Liang X H, Liu Q, Fournier M J (2009b). Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. RNA, 15(9): 1716–1728

Lovejoy A F, Riordan D P, Brown P O (2014). Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. PLoS One, 9(10): e110799

Ma X, Zhao X, Yu Y T (2003). Pseudouridylation (Ψ) of U2 snRNA in S. cerevisiae is catalyzed by an RNA-independent mechanism. EMBO J, 22(8): 1889–1897

Manival X, Charron C, Fourmann J B, Godard F, Charpentier B, Branlant C (2006). 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, 34(3): 826–839

Massenet S, Mougin A, Branlant C (1998). Posttranscriptional Modifications in the U Small Nuclear RNAs. In: Grosjean H, Benne R, eds. Modification and Editing of RNA. Washington D. C.: ASM Press

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

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

Ni J, Tien A L, Fournier M J (1997). Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. Cell, 89(4): 565–573

Nurse K, Wrzesinski J, Bakin A, Lane B G, Ofengand J (1995). Purification, cloning, and properties of the tRNA psi 55 synthase from Escherichia coli. RNA, 1(1): 102–112

Ofengand J, Fournier M J (1998). The Pseudouridine Residues of rRNA: Number, Location, Biosynthesis, and Function. In: Grosjean H, Benne R., eds. Modification and Editing of RNA. Washington D. C.: ASM Press

Piekna-Przybylska D, Przybylski P, Baudin-Baillieu A, Rousset J P,

Fournier M J (2008). Ribosome performance is enhanced by a rich cluster of pseudouridines in the A-site finger region of the large subunit. J Biol Chem, 283(38): 26026–26036

Rashid R, Liang B, Baker D L, Youssef O A, He Y, Phipps K, Terns R M, Terns M P, Li H (2006). Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. Mol Cell, 21(2): 249–260

Reddy R, Busch H (1988). Small Nuclear RNAs: RNA Sequences, Structure, and Modifications. In: Birnstiel M L, ed. Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles. Berlin: Springer Berlin Heidelber, pp. 1–37

Reichow S L, Hamma T, Ferré-D'Amaré A R, Varani G (2007). The structure and function of small nucleolar ribonucleoproteins. Nucleic Acids Res, 35(5): 1452–1464

Rozhdestvensky T S, Tang T H, Tchirkova I V, Brosius J, Bachellerie J P, Hüttenhofer 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(3): 869–877

Schattner P, Barberan-Soler S, Lowe T M (2006). A computational screen for mammalian pseudouridylation guide H/ACA RNAs. RNA, 12(1): 15–25

Schattner P, Decatur W A, Davis C A, Ares M Jr, Fournier M J, Lowe T M (2004). Genome-wide searching for pseudouridylation guide snoRNAs: analysis of the Saccharomyces cerevisiae genome. Nucleic Acids Res, 32(14): 4281–4296

Schwartz S, Bernstein D A, Mumbach M R, Jovanovic M, Herbst R H, León-Ricardo B X, Engreitz J M, Guttman M, Satija R, Lander E S, Fink G, Regev A (2014). Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell, 159(1): 148–162

Sprinzl M, Vassilenko K S (2005). Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res, 33(Database issue): D139–D140

Terns M, Terns R (2006). Noncoding RNAs of the H/ACA family. Cold Spring Harb Symp Quant Biol, 71(0): 395–405

Torchet C, Badis G, Devaux F, Costanzo G, Werner M, Jacquier A (2005). The complete set of H/ACA snoRNAs that guide rRNA pseudouridylations in Saccharomyces cerevisiae. RNA, 11(6): 928–938

Vitali P, Royo H, Seitz H, Bachellerie J P, Hüttenhofer A, Cavaillé J (2003). Identification of 13 novel human modification guide RNAs. Nucleic Acids Res, 31(22): 6543–6551

Watkins N J, Gottschalk A, Neubauer G, Kastner B, Fabrizio P, Mann M, Lührmann 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(12): 1549–1568

Wu G, Adachi H, Ge J, Stephenson D, Query C C, Yu Y T (2016). Pseudouridines in U2 snRNA stimulate the ATPase activity of Prp5 during spliceosome assembly. EMBO J, 35(6): 654–667

Wu G, Xiao M, Yang C, Yu Y T (2011). U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. EMBO J, 30(1): 79–89

Wu G, Yu A T, Kantartzis A, Yu Y T (2011). Functions and mechanisms of spliceosomal small nuclear RNA pseudouridylation. Wiley Interdiscip Rev RNA, 2(4): 571–581

Xiao M, Yang C, Schattner P, Yu Y T (2009). Functionality and substrate specificity of human box H/ACA guide RNAs. RNA, 15(1): 176–186

Yang C, McPheeters D S, Yu Y T (2005). y35 in the branch site recognition region of U2 small nuclear RNA is important for pre-mRNA splicing in Saccharomyces cerevisiae. J Biol Chem, 280(8): 6655–6662

Ye K (2007). H/ACA guide RNAs, proteins and complexes. Curr Opin Struct Biol, 17(3): 287–292

Yu A T, Ge J, Yu Y T (2011). Pseudouridines in spliceosomal snRNAs. Protein Cell, 2(9): 712–725

Yu Y T, Meier U T (2014). RNA-guided isomerization of uridine to pseudouridine–pseudouridylation. RNA Biol, 11(12): 1483–1494

Yu Y T, Shu M D, Steitz J A (1998). Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. EMBO J, 17 (19): 5783–5795

Yu Y T, Terns R M, Terns M P (2005). Mechanisms and functions of RNA-guided RNA modification. In: Grosjean H. ed. Fine-Tuning of RNA Functions by Modification and Editing. Berlin: Springer Berlin Heidelberg, pp. 223–262

Zebarjadian Y, King T, Fournier M J, Clarke L, Carbon J (1999). Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. Mol Cell Biol, 19(11): 7461–7472

Zhao X, Li Z H, Terns R M, Terns M P, Yu Y T (2002). An H/ACA guide RNA directs U2 pseudouridylation at two different sites in the branchpoint recognition region in Xenopus oocytes. RNA, 8(12): 1515–1525

Zhao X, 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(4): 681–690