DISTINGUISHING FEATURES OF GUIDE AND SUBSTRATE RNA REGOGNITION BY H/ACA SNORNPS

ERIN KATELYN KELLY
Bachelor of Science, University of Lethbridge, 2017

A Thesis/Project
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfillment of the
Requirements of the Degree

MASTER OF SCIENCE

Department of Chemistry and Biochemistry
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

DISTINGUISHING FEATURES OF GUIDE AND SUBSTRATE RNA RECOGNITION BY H/ACA SNORNPS

ERIN KATELYN KELLY

Date of Defence: June 28, 2018

Dr. U. Kothe Thesis Supervisor	Professor	Ph.D.
Dr. T. Russell Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. T. Patel Thesis Examination Committee Member	Assistant Professor	Ph.D.
Dr. J. Jackman External Examiner The Ohio State University Columbus, Ohio, United States of America	Associate Professor	Ph.D.
Dr. M. Roussel Chair, Thesis Examination Committee	Professor	Ph.D.

Abstract

H/ACA small nucleolar ribonucleoproteins (snoRNPs) pseudouridylate RNA in eukaryotes and archaea. They consist of four proteins and a guide RNA. It is unclear which element of the guide RNA the H/ACA proteins recognize. By designing guide RNA variants with altered structural features, I determined that H/ACA proteins bind guide RNA extremely tightly, but non-specifically *in vitro*. Additionally, the base-pairing between guide and substrate RNA varies widely in nature. To elucidate substrate selection rules, I systematically altered the sequence of a substrate RNA and tested for snoRNP binding and activity. All substrate RNA variants tested are bound by the H/ACA snoRNP, but not all are effectively modified. Minimal continuous base pairing is required to position the target uridine in the active site for modification, and H/ACA snoRNPs can target structured RNAs. This information will improve the prediction of novel target sites of H/ACA snoRNPs, e.g. in messenger RNAs.

Acknowledgements

I would first like to thank my supervisor, mentor and friend – Dr. Ute Kothe. Under your instruction, I discovered my passion for science and learning. You taught me how to strive to be better and really learn about the type of person I want to be. Thank you for pushing me to be my best.

I would like to thank my committee members Drs. Tony Russell and Trushar Patel. I have learned so much from you both, whether it be in class, in a hallway discussion or in a committee meeting. Thank you for your input and support.

Thanks to Dr. Jane Jackman for taking the time to critique my thesis and attend my oral defense. I appreciate your input on my work.

Thank you to my lovely ARRTI lab mates, especially in the Kothe group. The love, support and ridiculous music will never be forgotten.

Thank you to all my friends and family, especially those who listened to my science rants and helped me make decent figures.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ATP	Adenosine triphosphate
BME	β-mercaptoethanol
Cbf5	Centromere binding factor 5
CD	Circular dichroism
CS	CHORD and Sgt1 domain
CTP	Cytidine triphosphate
DC	Dyskeratosis congenita
DCK1	Human dyskerin
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
Gar1	Glycine-Arginine rich protein 1
Gln	Glutamine
Glu	Glutamate
GMP	Guanosine monophosphate
GTP	Guanosine triphosphate
HH	Hoyeraal-Hreidarsson syndrome
HP	hairpin
iPPase	Inorganic pyrophosphatase
IPTG	isopropyl β-D-1-thiogalactopyranoside
IRES	Internal Ribosome Entry Site
K_D	Dissociation Constant
LB	Lysogeny Broth
mRNA	Messenger RNA
Naf1	Nuclear assembly factor 1
ncRNA	Non-coding RNA
NMR	Nuclear Magnetic Resonance
nt	nucleotides
NTP	Nucleoside triphosphate
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PMSF	phenylmethane sulfonyl fluoride
PUA	Pseudouridine synthase and Archaeosine transglycosylase
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
SDS	sodium dodecyl sulfate
Shq1	Small nucleolar RNAs of the box H/ACA family quantitative accumulation protein 1
SMN	Survival of motor neuron proteins
snoRNA	Small nucleolar RNA
snoRNP	Small nucleolar ribonucleoprotein
snRNA	Small nuclear RNA
snRNP	Small nuclear RNP
SSD	Shq1-Specific Domain
TERC	Telomerase RNA Component

TERT	Telomerase Reverse Transcriptase
TOR	Target of Rapamycin
tRF	tRNA Derived Fragments
tRNA	Transfer RNA
UTP	Uridine triphosphate
UV	Ultra-Violet
WT	Wild type
X-DC	X-linked Dyskeratosis congenita
Ψ	Pseudouridine

Chapter 1 - Introduction

1.1 - Pseudouridine

Chemical modifications conferred upon RNA molecules have a variety of effects, such as stabilization of the molecule, establishing a new molecular function or tagging the RNA to facilitate interaction with other cellular machines. The isomerization of uridine (U) to pseudouridine (Ψ) is the most abundant RNA modification across all domains of life. It was initially discovered in the 1950's (Cohn, 1959; Davis & Allen, 1957) and has since been identified in many types of RNA, including ribosomal RNA (rRNA), transfer RNA (tRNA), messenger RNA (mRNA) and various non-coding RNAs (ncRNA) (reviewed in (Li *et al.*, 2016)). Pseudouridine is the C-glycoside isomer of uridine, where the glycosidic bond between the ribose C1 and the uracil N1 has been broken, the uracil base rotated, and a new C1-C5 glycosidic bond formed between the base and the ribose (Figure 1) (Cohn, 1960). This seemingly subtle structural change has been shown to have significant effects on RNA stability. The presence of a new N_1H imino group gives the nucleotide potential to participate in another hydrogen bonding interaction. Various studies provide evidence that this proton is in fact involved in molecular interactions. A crystal structure of $tRNA^{Gln}$ complexed with a tRNA-synthetase revealed that this imino group could form a water bridge with the phosphate backbone of the RNA (Arnez & Steitz, 1994). Pseudouridines have been shown to form water bridges in the ribosome structures from *Escherichia coli*, *Saccharomyces cerevisiae* and humans, contributing to local rRNA stability (Ben-Shem *et al.*, 2011; Natchiar *et al.*, 2017; Noeske *et al.*, 2015). Additionally, it was shown through NMR, UV and CD spectroscopy that pseudouridine can increase the stability of RNA

molecules through increased base stacking interactions while maintaining an A-form helix, compared to RNA molecules that contain unmodified uridine (Davis, 1995).

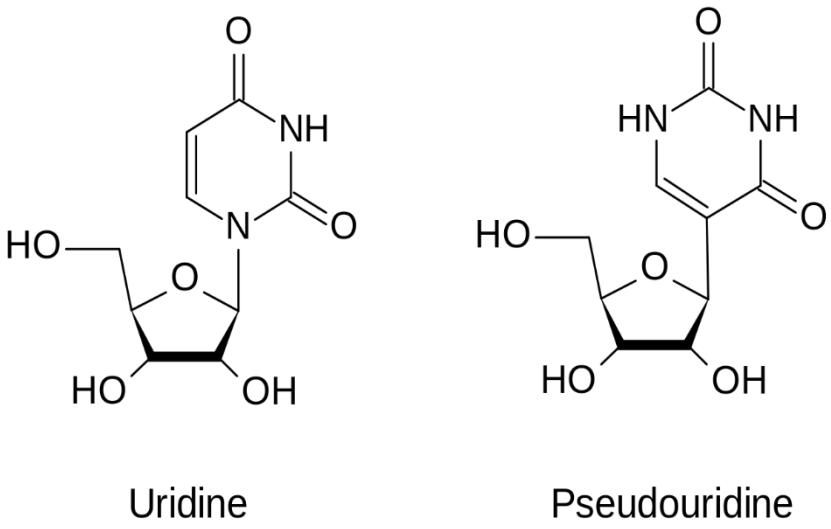

Figure 1. Chemical structures of uridine and pseudouridine. The N-C glycosidic bond between the uracil base and the ribose is replaced with a C-C glycosidic bond in pseudouridine, resulting in a free imino proton for hydrogen bonding.

1.2 - Biological Significance of Pseudouridylation

Pseudouridine has been shown to play a significant biological role in several systems. Most notably, pseudouridines are clustered in functionally important regions of the ribosome, such as the peptidyl transferase center, the peptide exit tunnel, the decoding centre and the ribosomal subunit interface (Decatur & Fournier, 2002). It has been shown in *S. cerevisiae* by deleting pseudouridine-guiding small nucleolar RNAs (snoRNAs) that while loss of individual pseudouridines in ribosomes does not appear to have any effect on function, the cumulative loss of pseudouridines in either the peptidyl transferase center or the decoding center causes impaired cell growth and translation rates, increased susceptibility to ribosome-targeting antibiotics and impaired polysome formation (King *et al.*, 2003; Liang *et al.*, 2009b). Yeast strains containing a catalytically inactive mutant of

the pseudouridine synthase Cbf5 (which targets rRNA) demonstrated that pseudouridine in rRNA is not essential for cell survival but does improve the cell's response to stress conditions. Furthermore, reduced pseudouridylation impaired the production of functional, cytoplasmic ribosomes (Zebarjadian *et al.*, 1999). Later, studies in yeast and human systems with inactive Cbf5 and its human homolog dyskerin demonstrated that pseudouridine-free ribosomes have a decreased affinity for tRNA at the A and P sites. Furthermore, the recruitment of ribosomes to mRNAs containing internal ribosomal entry sites (IRESs) is impaired (Jack *et al.*, 2011). Another study expanded on these effects using human cells in which the pseudouridine synthase DKC1 was depleted, resulting in pseudouridine-free ribosomes. It showed that while 5′-cap dependent translation initiation was not affected, IRES-dependent translation was reduced. Though translation could still occur, translational fidelity was lost as this group demonstrated an increased rate of stop codon read through (Penzo *et al.*, 2015).

Pseudouridines also play a functional role in other molecular machines such as the spliceosome. In vertebrates, the U2 snRNA has constitutive pseudouridylation in the branch site recognition region and at least one pseudouridine in this position is necessary for small nuclear ribonucleoprotein (snRNP) assembly and splicing activity (Zhao & Yu, 2004). These constitutive pseudouridines have been shown to stimulate the DEAD/H box ATPase Prp5 during spliceosome assembly, and pseudouridylated U2 is preferentially bound by Prp5 (Wu *et al.*, 2016a). Stress conditions have been shown to alter the pseudouridylation state of small nuclear RNAs (snRNAs). For example, Ψ93 of the U2 snRNA is induced under starvation conditions in yeast which is mediated by the Target of Rapamycin (TOR) signaling pathway. Similarly, Ψ56 in the U2 snRNA is induced by heat shock (Wu *et al.*,

2016b; Wu *et al.*, 2011). It was later confirmed that this stress-induced pseudouridylation changes the conformation of the U2 snRNA, but the effects of these induced modifications on splicing are still uncertain (van der Feltz *et al.*, 2018).

The development of pseudouridine sequencing first allowed the detection of widespread presence of pseudouridines in ncRNA and mRNA in both yeast and humans, and differential pseudouridylation of mRNAs under stress conditions suggested a possible role for pseudouridine in gene expression (Carlile *et al.*, 2014; Li *et al.*, 2015; Lovejoy *et al.*, 2014; Schwartz *et al.*, 2014). The exact biological purpose of pseudouridine in mRNA has not yet been determined. However, a regulatory role for a pseudouridine in tRNA-derived fragments (tRF) was recently characterized, showing the pseudouridylated tRF to have inhibitory effects on translation. This occurs through a pseudouridine-dependent interaction between tRFs and polyadenylate binding protein 1 (PABP1) and influences human embryonic stem cell differentiation (Guzzi *et al.*, 2018).

Finally, some pseudouridines, or at least the enzymes that produce them, have been implicated in several diseases. A well-known characteristic of cancer is increased ribosome biogenesis to accommodate the increased cellular proliferation. This is typically coupled with increased dyskerin expression to accommodate the modification of the new rRNA (Liu *et al.*, 2012; O'Brien *et al.*, 2016; Sieron *et al.*, 2009). Mutations in the genes encoding proteins of the H/ACA small nucleolar ribonucleoprotein (snoRNP) complex (which is responsible for directing pseudouridylation in rRNA) have been associated with X-linked Dyskeratosis congenita (X-DC) and its more severe form - Hoyeraal-Hreidarsson (HH) syndrome (Heiss *et al.*, 1998). These diseases are characterized as premature aging disorders in which patients experience abnormal skin pigmentation, nail dystrophy,

leukoplakia of the oral mucosa and an increased predisposition to cancer development. Though there is debate that these diseases are due to malfunction in the telomeres as the H/ACA proteins are part of both telomerase and the H/ACA snoRNP complex, it has been shown that mouse models for Dyskeratosis congenita indeed display reduced ribosomal RNA pseudouridylation, which may affect translation (Ruggero *et al.*, 2003). This finding appears to conflict with the paradigm of increased dyskerin expression and rRNA pseudouridylation in cancer. However, some tumor suppressor proteins are produced through IRES-dependent translation, and if this is impaired in hypomodified ribosomes found in Dyskeratosis congenita, lack of tumor suppressor expression may result in cancer development (Montanaro, 2010; Penzo *et al.*, 2017; Penzo *et al.*, 2015; Yoon *et al.*, 2006). Diseases other than Dyskeratosis congenita and cancer have also been associated with malfunctions in pseudouridylation and pseudouridine synthases. Mitochondrial tRNA mutations associated with maternally inherited diabetes and deafness have been shown to cause deficient pseudouridylation of position 55 in $tRNA^{Glu}$. It may be that lack of pseudouridylation at this position results in structural instability of the tRNA, and therefore may have implications in mitochondrial protein synthesis (Wang *et al.*, 2016). Mitochondrial myopathy, lactic acidosis and sideroblastic anaemia (MLASA) has also been associated with a mutation in pseudouridine synthase 1 (PUS1) (Bykhovskaya *et al.*, 2004). This mutation affects a highly conserved residue in the catalytic centre of PUS1, resulting in an inability to pseudouridylate its substrates, which include cytoplasmic tRNAs, U2 and U6 spliceosomal RNAs and mRNAs (reviewed in (Rintala-Dempsey & Kothe, 2017)), though the exact mechanism of disease is yet to be determined.

1.3 - Pseudouridine Synthases

The enzymes responsible for converting uridine to pseudouridine are known as pseudouridine synthases. There are two main classes of pseudouridine synthases – stand-alone and RNA-guided enzymes. The stand-alone pseudouridine synthases are protein-only enzymes that directly recognize the RNA and target site for pseudouridylation based on either sequence or secondary structure (Rintala-Dempsey & Kothe, 2017). All pseudouridine synthases share a conserved catalytic mechanism and contain a strictly conserved aspartate residue, which is responsible for catalysis. The catalytic mechanism for pseudouridine synthases was recently determined to proceed through a glycal intermediate (Veerareddygari *et al.*, 2016).

The RNA-guided pseudouridine synthase is the H/ACA snoRNP complex. This complex is found in both eukaryotic and archaeal organisms. It is composed of four core proteins, Cbf5, Nop10, Gar1 and Nhp2 (L7Ae in archaea), and these proteins form a complex with a guide RNA (Figure 2A). Cbf5 is the catalytic component in yeast and is homologous to human dyskerin, mouse Nap57, and bacterial TruB (though the latter is a stand-alone enzyme). Cbf5 contains two domains – the catalytic domain and the pseudouridylase archaeosine tRNA-guanine transglycosylase (PUA) domain. Mutations in the catalytic domain, particularly the conserved aspartate residue, have been shown to eliminate rRNA pseudouridylation *in vivo* (Zebarjadian *et al.*, 1999). The PUA domain is responsible for binding conserved sequence motifs in the guide RNA, and finally there is also a high degree of structural conservation of both the catalytic and PUA domains between homologous pseudouridine synthases in different species (Yu & Meier, 2014).

Nop10 binds to and stabilizes the catalytic domain of Cbf5 to form a soluble complex (Hamma *et al.*, 2005). It also extends the guide RNA-binding face of the complex as well as serves as a docking point for Nhp2 (Hamma *et al.*, 2005; Li & Ye, 2006) Furthermore, it has been shown to be required for substrate binding, despite having no direct contact with the substrate RNA (Charpentier *et al.*, 2005). Nop10 also contributes to pseudouridine formation by enhancing the catalytic ability of Cbf5 as the interaction between Cbf5 and Nop10 positions the conserved lysine of Motif I of Cbf5 for hydrogen bonding to the carbonyl oxygen of the catalytic aspartate , and increasing the affinity of Cbf5 for RNA (Hamma *et al.*, 2005; Kamalampeta & Kothe, 2012).

Gar1 increases the RNA binding ability and the catalytic activity of Cbf5 (Kamalampeta & Kothe, 2012). It has also been suggested that it may remodel the substrate if it is mis-docked to the complex (Liang *et al.*, 2008). Structural analysis, mutagenesis and enzyme kinetics studies have further shown that Gar1 has a role in substrate release through an interaction with the thumb loop of the Cbf5 catalytic domain, which has direct contacts with the substrate RNA (Duan *et al.*, 2009).

Nhp2 is proposed to interact with the complex on the upper stem of the H/ACA RNA similar to its archaeal homolog L7Ae, though it binds RNA non-specifically (Henras *et al.*, 2001). This contrasts with L7Ae, which specifically recognizes and binds kink-turn and kink-loop motifs in archaeal H/ACA guide RNAs and binds them with high affinity. Nhp2 does not specifically recognize kink-turns, nor do eukaryotic H/ACA RNAs contain any kink-turns. However, it was revealed in a solution structure that the guide RNA may adopt a similar fold to a kink-turn with a flipped out nucleotide, though this did not constitute a high affinity binding site for Nhp2 (Koo *et al.*, 2011). In the archaeal system,

it has been revealed in biochemical and structural studies that L7Ae contributes to positioning the substrate RNA in the active site of Cbf5 (Liang *et al.*, 2008; Liang *et al.*, 2007). This occurs through L7Ae anchoring the upper stem of the guide RNA, which induces a rotation of the guide RNA-substrate RNA helix that moves the target uridine into the active site (Liang *et al.*, 2009a). Compared to archaeal L7Ae, eukaryotic Nhp2 binds H/ACA guide RNA weakly and nonspecifically. Despite this contrast, the function of Nhp2 may be similar to L7Ae in positioning the substrate RNA appropriately in the active site of Cbf5. Though Nhp2 is not critical for binding of the guide RNA by the other H/ACA proteins, or for binding substrate RNA, it is required for efficient pseudouridylation of the substrate (Caton *et al.*, 2018).

The final component of the H/ACA snoRNP system is the H/ACA guide RNA. In most eukaryotes studied to date (in particular, yeast and vertebrates), this RNA forms two hairpin structures, each of which contains a single stranded region known as the pseudouridylation pocket (Ganot *et al.*, 1997a; Ni *et al.*, 1997). The single stranded regions are complementary to a sequence in a target RNA and provide sequence-specific substrate selection through base pairing a target RNA and guiding the target uridine to the active site of Cbf5. This guide RNA contains only two conserved sequences found immediately 3′ of the hairpins – the H box (ANANNA) and the ACA box (Figure 2B). These sequence elements have been shown to be vital for accumulation of snoRNPs in human cells (Ganot *et al.*, 1997b) and to be required for enzymatic activity of the complex (Bortolin *et al.*, 1999). They are specifically bound by the PUA domain of Cbf5, which has a pocket that accommodates the conserved A nucleotides of the ANANNA and ACA sequences (Zhou *et al.*, 2011). In the yeast system, it has been shown that mutation of these conserved

elements does not affect the affinity of guide RNA binding by the Cbf5-Nop10-Gar1 ternary complex, but the mutations do reduce pseudouridylation of substrate RNAs (Caton *et al.*, 2018). This information, coupled with the fact that there is a conserved distance of 14-16 nucleotides between the H or ACA boxes with the pseudouridylation pocket of the guide RNA (Ganot *et al.*, 1997a; Ni *et al.*, 1997) has led to the suggestion that the H/ACA boxes act as a molecular ruler to properly position Cbf5 against the guide RNA through the H/ACA box interaction with the PUA domain, such that the active site of Cbf5 aligns with the pseudouridylation pocket (Caton *et al.*, 2018; Ganot *et al.*, 1997a; Ni *et al.*, 1997).

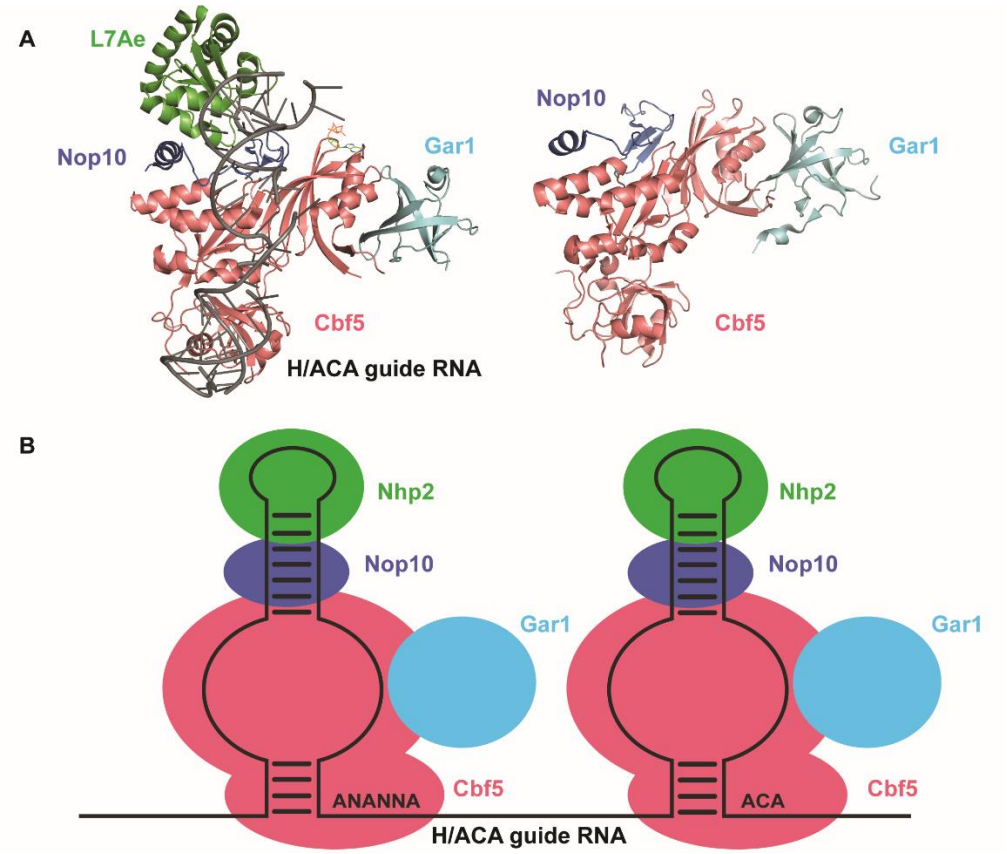

Figure 2. Structures of H/ACA snoRNPs. (A) Cartoon representations of the *Pyrococcus furiosus* H/ACA snoRNP (left - PDB ID 2HVY (Li & Ye, 2006)) and the *S. cerevisiae* Cbf5-Nop-10-Gar1 trimeric complex (right - PDB ID 3U38 (Li *et al.*, 2011b)). The catalytic protein Cbf5 (pink) as well as proteinsNop10 (blue), Gar1 (cyan) and L7Ae (green) are bound to an H/ACA guide RNA (black). Nhp2 replaces L7Ae in eukaryotic systems, and eukaryotic guide RNAs do not contain kink turns. (B) Schematic of the eukaryotic H/ACA snoRNP. Proteins are coloured according to (A). Nhp2, a homolog of L7Ae, binds the upper stem in eukaryotes (green). The guide RNA contains two conserved sequence elements, the H (ANANNA) and ACA boxes. The single stranded region in the guide RNA stem is the pseudouridylation pocket to which a substrate RNA can base pair. The PUA domains of the two Cbf5 copies bind the H and ACA boxes, respectively, to align the pseudouridylation pockets with the catalytic domains.

1.4 - Assembly of the H/ACA ribonucleoprotein complex

H/ACA snoRNPs function primarily in the nucleolus where they modify ribosomal RNAs, and some function in the Cajal bodies to modify snRNAs (Darzacq *et al.*, 2002; Jády *et al.*, 2003; Narayanan *et al.*, 1999). The new evidence that H/ACA snoRNPs may also target mRNAs brings their localization into question (Carlile *et al.*, 2014; Li *et al.*,

2015; Lovejoy *et al.*, 2014; Schwartz *et al.*, 2014). Though many details of H/ACA snoRNP assembly and localization are still unknown, the major steps of assembly have been described.

Cbf5 is translated in the cytoplasm, where it is bound to another protein known as Shq1. Shq1 is a chaperone protein that contains a CHORD and Sgt1 (CS) domain, similar to that of many Hsp90 co-chaperones (Godin *et al.*, 2009). Shq1 also has a Shq1-specific domain (SSD) that has been described as an RNA mimic, as it binds to the PUA domain of Cbf5 by mirroring contacts created between the guide RNA and Cbf5 (Machado-Pinilla *et al.*, 2011). The CS domain also binds to the PUA domain, though on the opposite face bound by the SSD, creating a vice around the Cbf5 RNA binding domain (Li *et al.*, 2011a; Machado-Pinilla *et al.*, 2012; Singh *et al.*, 2015). Next, Cbf5 associates with Nop10, Nhp2 and another assembly factor known as Naf1. It is not explicitly known whether or not Shq1 must be removed before Cbf5 can interact with these other proteins, as a structure with all four proteins assembled together has been published, but other *in vitro* experiments have shown that Shq1 may exclude Nop10, Nhp2 and Naf1 binding to Cbf5 (Grozdanov *et al.*, 2009; Li *et al.*, 2011a). Shq1 removal from Cbf5 is dependent on the R2TP complex, which contains the AAA+ ATPases pontin and reptin. It has been shown that these ATPases target the CS domain on Shq1 and remove Shq1 from Cbf5 in an ATP-independent manner (Machado-Pinilla *et al.*, 2012).

Naf1 associates with Cbf5 via a domain structurally similar to Gar1 and interacts with the C-terminal domain of RNA polymerase II. This provides a bridge to recruit Cbf5-Nop10-Nhp2 to the nascent H/ACA RNA as it is transcribed. Additional interesting features of Naf1 are that it only binds RNA polymerase II when the C-terminal domain of

RNA polymerase II is phosphorylated, and Naf1 can also form a homodimer with itself via its Cbf5 binding face, suggesting it may be playing a regulatory role in H/ACA snoRNP biogenesis (Fatica *et al.*, 2002). Naf1 binding to Cbf5 is also exclusive of Gar1 binding, given that they are structurally similar. This may attenuate the activity of immature H/ACA snoRNPs as it prevents Gar1 functioning in substrate placement and turnover (Leulliot *et al.*, 2007). Naf1 is exchanged for Gar1 in the Cajal bodies through an unknown mechanism, though it may be mediated through survival of motor neuron (SMN) proteins (Pellizzoni *et al.*, 2001). From the Cajal bodies, the now mature H/ACA snoRNPs will be shuttled to the nucleoli where they will modify rRNAs.

In some organisms, the H/ACA guide RNAs undergo several processing steps as well. In yeast, most H/ACA RNAs are encoded as individual transcription units and undergo 3′-end digestion by exonucleases. The RNA is protected by binding of the H/ACA proteins, and the binding of the Cbf5 PUA domain on the ACA box prevents nucleases from cleaving the RNA farther up than three nucleotides 3′ of the ACA box. In contrast, some yeast H/ACA RNAs and most vertebrate H/ACA RNAs are encoded within introns. Following lariat debranching, exonucleolytic processing occurs. In mammals, it was found the process also involved subsequent addition of a poly-A tail by PAPD5, followed by trimming by poly-A specific ribonuclease (PARN). This may serve to stabilize the RNAs against shortening (Berndt *et al.*, 2012).

1.5 – Cellular Function of H/ACA snoRNPs

H/ACA snoRNPs and RNA modification play significant roles in ribosome biogenesis. In yeast, ribosome biogenesis begins with the RNA polymerase I – dependent transcription of the 35S pre-rRNA. This pre-rRNA undergoes cleavage steps to eventually

form the 18S, 5.8S and 25S rRNAs. The 5S rRNA is transcribed independently by RNA polymerase III (Henras *et al.*, 2015; Woolford & Baserga, 2013). Cleavage of the 35S pre-rRNA can occur either co-transcriptionally or post-transcriptionally (Kos & Tollervey, 2010; Osheim *et al.*, 2004; Turowski & Tollervey, 2015). Each pre-rRNA interacts with several snoRNP complexes, both H/ACA snoRNPs and C/D snoRNPs. The U3, U8, U14 and snR30/U17 snoRNPs contribute to rRNA cleavage. Other snoRNP species are responsible for either 2′O-methylation (C/D snoRNPs) or pseudouridylation (H/ACA snoRNPs) of the rRNA. In one case involving the H/ACA snoRNA snR10, rRNA processing and modification events directed by this snoRNA are coupled (Liang *et al.*, 2010). In yeast, approximately 2% of rRNA nucleotides are covalently modified, the majority of which are 2′O-methylations or pseudouridylations (Boccaletto *et al.*, 2018). These modifications have been implicated in roles involving translation and pre-rRNA processing, as depletion of snoRNAs results in accumulation of pre-rRNA (Liang *et al.*, 2009b). It has been shown that 2′O-methylation can occur co-transcriptionally and post-transcriptionally, however the timing of pseudouridylation has not been confirmed. It is reasonable to speculate that it can also occur co-transcriptionally (Kos & Tollervey, 2010; Sloan *et al.*, 2017). Modification of rRNA at different stages of ribosome assembly contributes to appropriate rRNA folding and may promote the recruitment of ribosomal proteins, and the specific binding of snoRNPs may be remodeling the rRNA (Turowski & Tollervey, 2015). Lastly, it has been suggested that ribosome heterogeneity can be defined on the basis of different rRNA modification levels and locations, and that modification patterns may differ between cell types in multicellular organisms (Natchiar *et al.*, 2017).

Telomeres are repetitive DNA sequence elements that cap the ends of linear eukaryotic chromosomes. With each round of genome replication and cell division, the telomeres are shortened, eventually leading to genome instability and cell death (Levy *et al.*, 1992). Telomerase is a ribonucleoprotein enzyme responsible for extending the length of telomeres. It is composed of a telomerase reverse transcriptase (TERT) and uses the telomerase RNA component (TERC) as a template (Blackburn & Collins, 2011). Mammalian TERC has the conserved H and ACA boxes found in H/ACA snoRNAs, giving it the ability to bind the H/ACA proteins (Mitchell *et al.*, 1999a). The TERC forms a complex with two subunits each of the four H/ACA proteins – dyskerin, Nop10, Gar1 and Nhp2 (Mitchell & Collins, 2000; Nguyen *et al.*, 2018). The binding of the H/ACA proteins stabilizes the TERC, and interruption of this interaction reduces the accumulation of the TERC (Mitchell *et al.*, 1999b; Shukla *et al.*, 2016). The two subunits of dyskerin form a dimer within the telomerase complex, and mutations associated with the disease Dyskeratosis congenita are found at this interface (Nguyen *et al.*, 2018).

Dyskeratosis congenita (DC) is characterized by abnormal skin pigmentation, mucosal leukoplakia and nail dystrophy. The most common cause of death in DC patients is bone marrow failure (Dokal, 2000). Mutations in the components of both H/ACA snoRNPs and telomerase have been associated with DC. These mutations have been found in genes encoding dyskerin, Nop10, Nhp2 and both the telomerase reverse transcriptase and the telomerase RNA component (Mason & Bessler, 2011). Mutations affecting dyskerin are typically point mutations that cluster in the RNA-binding PUA domain and are sometimes also found in the catalytic domain (Heiss *et al.*, 1998; Li & Ye, 2006; Li *et al.*, 2011a; Mason & Bessler, 2011). This has the potential to affect the function of both

H/ACA snoRNPs and telomerase. It was very recently shown that DC mutations causing amino acid substitutions in dyskerin are found at both RNA binding site and the dyskerin-dyskerin interface in the telomerase holoenzyme, potentially affecting the function of telomerase (Nguyen *et al.*, 2018). Disruption of the dyskerin-dyskerin interaction may also affect H/ACA snoRNPs as snoRNAs are complexed with two sets of protein heterotetramers, but no structure of the dual hairpin system has been described so far. DC mutations have also been described in the genes encoding Nop10 and Nhp2 (Vulliamy *et al.*, 2008; Walne *et al.*, 2007). It was determined that mutations in the Nhp2 gene impairs the ability of Nhp2 to assemble in the H/ACA snoRNP complex, but Nop10 mutations did not affect assembly (Trahan *et al.*, 2010). As components of both telomerase and the H/ACA snoRNP are affected by DC mutations, there has been much debate concerning which function is primarily affected, or if the disease is a result of combined effects of reduced telomere maintenance or ribosome biogenesis and modification (Montanaro, 2010; Narla & Ebert, 2010; Ruggero & Shimamura, 2014).

1.6 - Interactions with the Substrate RNA

The binding of a target RNA in the pseudouridylation pocket and correct positioning of the target uridine into the active site of Cbf5 is essential for pseudouridylation. The binding and positioning of substrate RNA can be affected by a variety of factors. Firstly, each of the H/ACA proteins has a unique role in substrate binding and placement. Cbf5 is the only protein to have direct contact with the substrate RNA, and this interaction is largely nonspecific, consisting of hydrophobic and electrostatic contact between Cbf5 residues and the sugar-phosphate backbone of the substrate RNA (Liang *et al.*, 2009a). The thumb loop of Cbf5 also undergoes an open/closed conformational change upon binding of the substrate, locking it into place (Duan *et al.*, 2009). Gar1 interacts

directly with the Cbf5 thumb loop and may remodel the substrate RNA if it is mis-docked and upon pseudouridylation will help change the thumb loop conformation from closed to open to release the product (Liang *et al.*, 2008; Wang *et al.*, 2015). The interaction between Nop10 and the guide RNA improves the docking of the target uridine in the Cbf5 active site (Liang *et al.*, 2008). The role of archaeal L7Ae in substrate accommodation has been studied thoroughly, and it is likely that eukaryotic Nhp2 functions similarly (Caton *et al.*, 2018). A structure of an archaeal H/ACA snoRNP that does not contain L7Ae shows that the substrate RNA is bound by the complex, but the target uridine is not docked in the active site (Liang *et al.*, 2007). Upon inclusion of L7Ae, the target uridine moves into the active site through an induced conformational change of the upper stem of the H/ACA guide RNA (Liang *et al.*, 2008; Liang *et al.*, 2007; Liang *et al.*, 2009a).

The structure and stability of the H/ACA guide RNA is also critical for substrate RNA binding and pseudouridylation. Previously, it was determined that not all predicted guide RNA-target RNA pairs resulted in successful pseudouridylation. The modification ability is strongly dependent on three factors, namely the stability of the guide RNA hairpin, sufficient base pairing between the guide and the target in the pseudouridylation pocket, and a conserved distance between the target uridine and the H or ACA boxes (Xiao *et al.*, 2009). The stability of the guide RNA hairpins contribute to substrate binding through coaxial stacking interactions that form between both the upper and lower stems of the guide and the helices formed by the binding of the substrate RNA to the single stranded pseudouridylation pocket, and these interactions likely confer the substrate remodeling effects of L7Ae (Duan *et al.*, 2009; Liang *et al.*, 2009a). The conserved distance of 14-16 nucleotides between the target uridine and the H or ACA boxes is important for properly aligning the substrate to the catalytic domain of Cbf5, such that the target uridine can be

appropriately docked into the active site (Caton *et al.*, 2018; Duan *et al.*, 2009; Wu & Feigon, 2007).

The interaction between the H/ACA guide RNA and the substrate RNA has been characterized as an Ω-structure, forming a three-way junction between the two helices formed between the substrate and both single stranded regions of the pseudouridylation pocket and the upper stem of the guide RNA (Figure 3) (Jin *et al.*, 2007; Wu & Feigon, 2007). The substrate RNA only interacts with the guide RNA on one side, rather than being threaded through the pseudouridylation pocket. The H/ACA proteins assemble on the opposite side of the guide RNA leaving a free side to engage the substrates (Jin *et al.*, 2007; Wu & Feigon, 2007). The nature of the base pairing between the guide RNA and the substrate RNA can vary quite dramatically between guide-target pairs. There is no consistent length of base pairing region on either the 5′ or 3′ side of the pseudouridylation pocket, and an inconsistent number of non-canonical pairs and mismatches in the pseudouridylation pocket can be accommodated. For example, the snR191 3′ hairpin makes only 8 base pairs with its substrate in the 25S rRNA, with 4 base pairs on either side of the target uridine. The longest known interaction in nature is 17 base pairs and it occurs between the 3′ hairpin of snR82 and the 25S rRNA (8 base pairs in the 5′ side of the pocket and 9 base pairs in the 3′ side of the pocket). The fewest number of base pairs made on one side of the target uridine is 3 base pairs (e.g. in snR3 and snR81) and the maximum number of pairs on one side is 10 base pairs. Typically, the duplex between the guide and substrate RNA is shorter in the 3′ side of the pseudouridylation pocket, though there are some exceptions. Finally, a maximum 2 mismatches and 3 non-canonical base pairs occur in known guide-substrate interactions (Piekna-Przybylska *et al.*, 2008; Torchet *et al.*, 2005). It is also not clear if or how the H/ACA snoRNP can accommodate structured RNA

substrates, given that a drastically different structure must form between the substrate and the guide in the pseudouridylation pocket. The possibility of pseudouridylation to occur co-transcriptionally may alleviate structural barriers of the substrate, but post-transcriptional modification may be retarded if the target uridine is buried deep in secondary structure (Kos & Tollervey, 2010). Though helicases are required to remove H/ACA snoRNPs from their substrates, it is not known if helicases are required for substrate loading as well (Liang & Fournier, 2006).

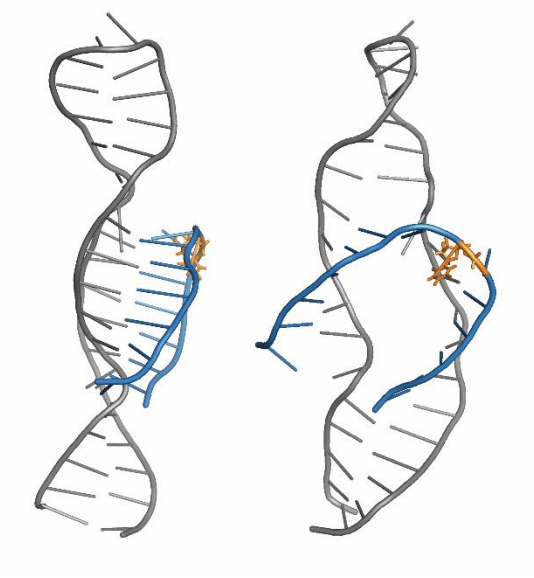

Figure 3. The solution NMR structure of a substrate RNA bound to an H/ACA guide RNA. A cartoon representation of a substrate RNA (blue) fragment bound to an H/ACA guide RNA pseudouridylation pocket (grey) in the absence of proteins (side and front view). The target uridine is show in orange. The complex forms an Ω-structure and a three-way junction between the two helices between the substrate and either side of the pseudouridylation pocket and the upper stem of the guide RNA (PDB ID 2PCW (Jin *et al.*, 2007)).

The C/D box snoRNP class performs analogous RNA-guided modification to H/ACA snoRNPs but directs 2′O methylation rather than pseudouridylation which has mostly been studied in archaeal organisms. he C/D snoRNP is composed of the methyltransferase fibrillarin (Nop1), L7Ae, Nop56/58 and a guide RNA. The guide RNA contains two sets of conserved sequence elements – the C box (RUGAUGA; R is a purine) and D box (CUGA) and the C′ and D′ boxes (Reichow *et al.*, 2007). These sequence elements form kink-turn and kink-loop structures to which L7Ae binds. Two sets of the proteins can bind a single guide RNA, giving the snoRNP a bipartite structure. The substrate RNA base pairs with a single stranded region upstream of the D/D′ boxes. The nucleotide paired 5 nucleotides upstream of the D/D′ box is the one that is methylated (N+5 rule). The base pairing between the C/D guide RNA and the substrate RNA has been well described. The base pair between the target nucleotide and the guide RNA must be a

Watson-Crick pair, and base pairing adjacent to the target nucleotide must be continuous (Appel & Maxwell, 2007). In order to obtain the optimal conformation for modification, the single stranded region of the guide RNA cannot be less than twelve nucleotides, though the substrate RNA does not form more than 10 base pairs with the guide sequence (Yang *et al.*, 2016). The bipartite structure of the C/D snoRNP is also necessary to confer substrate specificity. Methylation can also occur with a "hemi-complex" (containing only one set of C/D associated proteins); however non-target nucleotides can be methylated. It was determined that both the C/D and C′/D′ boxes are required to confer nucleotide specificity of methylation (Hardin & Batey, 2006). Finally, a model of the C/D snoRNP that describes the complex as a di-RNP has been proposed consisting of four sets of proteins and two C/D guide RNAs. In this model, the face of fibrillarin containing the active site is away from the substrate RNA bound to the nearest single stranded guide sequence. There is some evidence that supports the idea that fibrillarin may be modifying a substrate RNA bound to a different C/D guide RNA within this di-RNP complex (Xue *et al.*, 2010).

1.7 – Objectives and Significance

In this thesis, I will investigate the features that distinguish H/ACA guide RNAs and the substrates they target. First, as the features of eukaryotic guide RNA that facilitate recognition by the H/ACA proteins are not well understood, and binding is not dependent on the conserved H and ACA boxes, guide RNA variants will be used to clarify the mechanism through which guide RNAs are recognized (Caton *et al.*, 2018). Each variant will target a structural feature of the H/ACA guide RNA, such as the 5′ tail or the single stranded pseudouridylation pocket. Secondly, as the details of base-pairing between the guide and substrate RNAs in the pseudouridylation pocket vary widely between different

guide RNAs in all organisms studied to date, I will elucidate the rules that define an RNA as a substrate for pseudouridylation by a yeast H/ACA snoRNP. This may serve as a model for substrate specificity in other eukaryotic organisms. To this end, a variety of substrate RNAs targeted by the yeast H/ACA guide RNA snR34 will be used to systematically introduce mismatches between the substrate RNA and the guide RNA in the pseudouridylation pocket. Finally, longer fragments of the 25S rRNA containing the sequence targeted by snR34 will be used to determine how the H/ACA snoRNP accommodates structured versus unstructured substrates, and how structure may affect the snoRNPs ability to modify an RNA.

The determination of the mechanism of guide RNA recognition and binding by the H/ACA enzymes will provide insight into the important steps of H/ACA snoRNP assembly. Additionally, the role of H/ACA RNA identification and misidentification may be considered in disease. Finally, knowledge of the key elements of H/ACA RNA recognition will define constraints for synthetic biologists to design an produce custom H/ACA guide RNAs to direct pseudouridylation to specific sites.

The variation of guide-substrate interactions in the pseudouridylation pocket can make the prediction of novel modification sites difficult. Defining the limitations of the guide-substrate interaction using the short substrate variants will ultimately enable the prediction of previously unknown modification sites in different types of RNAs as well as identify guide RNAs responsible for modification of known sites. Furthermore, understanding how H/ACA snoRNPs can target structured RNA substrates may provide some evidence into the possible timing of pseudouridylation, whether it be co-transcriptional or post-transcriptional. This knowledge will significantly increase our

understanding of global pseudouridine formation in the cell, which is predicted to be an epitranscriptomic regulator of gene expression (Carlile *et al.*, 2014; Chen *et al.*, 2016; Li *et al.*, 2015; Lovejoy *et al.*, 2014; Schwartz *et al.*, 2014). Moreover, this research will allow us to design artificial H/ACA sRNAs to target particular uridines with two important applications. First, artificial guide RNAs will be powerful tools in synthetic biology to regulate gene expression and to alter the function of stop codons (Fernández *et al.*, 2013; Karijolich & Yu, 2011) Second, such guide RNAs can serve as promising therapeutics for genetic diseases resulting from premature stop codons, such as cystic fibrosis, Duchenne muscular dystrophy and many cancers, because pseudouridylated stop codons allow translational read-through (Bordeira-Carrico *et al.*, 2012; Karijolich & Yu, 2011).

Chapter 2 - Materials and Methods

2.1 - Reagents

[5-^{3}H] uridine triphosphate (UTP) for *in vitro* transcriptions was purchased from Moravek Biochemicals, and [γ-^{32}P] adenosine triphosphate (ATP) for guide RNA 5′ end labelling was obtained from Perkin Elmer. Nickel Sepharose 6 Fast Flow and Glutathione Sepharose 4 Fast Flow resins were purchased from GE Healthcare. DNA oligonucleotides were obtained from Integrated DNA Technologies (IDT). Polymerase chain reaction (PCR) reagents and *Pfu* DNA polymerase were purchased from Truin Science. All other chemicals were purchased from Fisher Scientific.

2.2 - Overexpression of *S. cerevisiae* Cbf5-Nop10 and Gar1

pETDuet-ScCbf5-ScNop10 and pET28a-ScNop10 were co-transformed into *E. coli* BL21(DE3) (New England Biolabs) (Caton *et al.*, 2018). Test expression cultures were grown in 50 ml LB media with ampicillin and kanamycin at 37 °C to an OD_{600} of ~0.6 - 0.8, at which point 100 µL of the culture was re-plated on LB-ampicillin-kanamycin plates. The remaining culture was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and grown at 18 °C for 16 h. One OD_{600} of cells were harvested by centrifugation. The cell pellet was resuspended in 80 µL of 100 mM Tris-HCl pH 8.5, 5 M urea and analyzed by 12% sodium dodecyl sulfate (SDS)-PAGE. The test culture demonstrating the highest expression of Cbf5 was used to inoculate large-scale overexpression cultures. Large-scale overexpression cultures were grown in 5 L of LB media with ampicillin and kanamycin at 37 °C to an OD_{600} of ~ 0.8, induced with 1 mM IPTG and grown at 18 °C for 16 h. Cells were harvested by centrifugation at 5000 x g for 15 min at 4 °C, flash frozen and stored at -80 °C.

pGEX-5X-3-ScGar1 was transformed into *E. coli* Rosetta2(DE3) (New England Biolabs) (Caton *et al.*, 2018). Large scale overexpression cultures were grown in 2 L LB media with ampicillin at 37 °C to an OD_{600} of ~0.6 - 0.8 and were induced with 1 mM IPTG. The culture was grown at 37 °C for 3 h. Samples for SDS-PAGE analysis were taken as described above. Cells were harvested by centrifugation at 5000 x g for 15 min at 4 °C, flash frozen and stored at -80 °C.

2.3 - Purification of Cbf5-Nop10-Gar1 complex and Nhp2

Cell pellets containing Cbf5, Nop10 and Gar1 were combined and resuspended in 10 mL/g Buffer A (50 mM Tris-HCl pH 8.0, 1 M NaCl, 5% glycerol, 0.5 mM phenylmethane sulfonyl fluoride (PMSF), 1 mM β-mercaptoethanol (BME)). Cells were stirred on ice for 30 min and were lysed by sonication at intensity level 6, duty cycle 60% (Branson Sonifier 450) in eight 1-min intervals with 1 min rest between pulses. The lysate was then centrifuged at 30,000 x g for 45 min at 4 °C. The supernatant was applied to 10 mL glutathione sepharose resin and incubated for 1 h on ice with gentle shaking. The slurry was centrifuged at 500 x g for 5 min at 4 °C. The supernatant was removed, and the resin was washed 3 times with Buffer B (50 mM Tris-HCl pH 8.0, 1 M NaCl, 5% glycerol, 1 mM BME). Proteins were eluted with 90 % column volume of Buffer C (50 mM Tris-HCl pH 8.0, 1 M NaCl, 5 % glycerol, 0.5 mM PMSF, 1 mM BME, 10 mM glutathione) and incubated on ice for 5 min before centrifugation. The glutathione sepharose elutions were pooled and applied to 700 µL of nickel sepharose resin and incubated on ice for 1 h with gentle shaking. The slurry was centrifuged at 500 x g for 5 min at 4 °C, and the supernatant was removed. The resin was washed 3 times with 14 mL Buffer D (50 mM Tris-HCl pH 8.0, 1 M NaCl, 20 % glycerol, 1 mM BME) followed by 3 washes with Buffer E (50 mM

Tris-HCl pH 8.0, 1 M NaCl, 20 % glycerol, 1 mM BME, 20 mM imidazole). Proteins were eluted with 600 µL Buffer F (50 mM Tris-HCl pH 8.0, 1 M NaCl, 20 % glycerol, 1 mM BME, 300 mM imidazole) and incubated on ice for 5 min before centrifugation. Elutions were aliquoted, shock frozen and stored at -80 °C. The purification was evaluated using 12% SDS PAGE, and protein concentration was determined by 12% SDS PAGE and analyzed using ImageJ. Nhp2 was expressed and purified as described previously (Caton *et al.*, 2018).

2.4 - *In vitro* transcription, and purification of H/ACA guide RNA variants and substrate RNA variants

The DNA sequence encoding the H/ACA guide RNA snR34 was PCR-amplified to include a T7 promoter as previously described (Caton *et al.*, 2018). The template DNA for snR34 5′ hairpin variants was assembled by PCR extension of two partially complementary oligonucleotides (Table 1). RNAs were *in vitro* transcribed as previously described (Wright *et al.*, 2011). Briefly, all *in vitro* transcription reactions were performed for 4 h at 37 °C in transcription buffer (40 mM Tris-HCl pH 7.5, 15 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl) with 10 mM dithiothreitol (DTT), 3 mM nucleoside triphosphates (NTPs) (adenosine triphosphate - ATP, cytidine triphosphate - CTP, guanosine triphosphate - GTP and uridine triphosphate - UTP), 5 mM guanosine monophosphate (GMP), 0.01 U/µL inorganic pyrophosphatase (iPPase), 0.3 µM T7 RNA polymerase and 0.12 U/µL RiboLock RNase inhibitor (Thermo Fisher Scientific). Subsequently, DNaseI was added to a final concentration of 0.002 U/µL, and the reaction was incubated for 1 h at 37 °C. Reactions were quenched with 1/10 volume of 3 M NaOAc.

Template DNA for substrate RNA was generated from PCR extension of two partially overlapping oligonucleotides ((Milligan *et al.*, 1987), Table 2). Radioactive substrate RNAs were generated by *in vitro* transcriptions including 3 mM ATP, CTP and GTP, and 0.1 mM [5-^{3}H] UTP (16.2 Ci/mmol). RNA concentration was determined by A_{260} using extinction coefficients calculated by OligoAnalyzer 3.1 (IDT), and specific activity was determined by scintillation counting.

RNAs were purified by crush and soak gel extraction. RNAs were separated by 15% urea-polyacrylamide gel electrophoresis (PAGE), and the appropriate band was identified by UV shadowing. The band was excised, crushed, and soaked in 1x TBE for at least 6 h. The gel was separated from the buffer by centrifugation at 4000 x g for 10 min, and the supernatant was removed. The RNA was purified from the supernatant using phenol-chloroform extraction. RNA was precipitated by adding 1/10 volume 3 M NaOAc and ethanol to a final concentration of 70% to the aqueous phase and incubated overnight at -20 °C. The pellet was resuspended in deionized water and stored at -20 °C. The CrPV IRES was kindly provided by Luc Roberts from the Wieden Lab.

2.5 - Reconstitution of H/ACA snoRNPs

Full length snR34 or the 5′ hairpin of the snR34 guide RNA was refolded by heating to 75°C for 5 min and cooling slowly to room temperature. Guide RNA was combined with Cbf5, Nop10, Gar1 and Nhp2 in a 0.45:1 guide RNA: protein ratio for the full length snR34 complex, or a 0.9:1 guide RNA:protein ratio for the 5′ hairpin snR34 complex in Reaction Buffer (20 mM HEPES-KOH pH 7.4, 150 mM NaCl, 0.1 mM ethylenediaminetetraacetic acid (EDTA), 1.5 mM $MgCl_2$, 10 % (v/v) glycerol, 0.75 mM DTT). The mixture was incubated for 10 min at 30 °C to allow complex formation.

2.6 - Tritium Release Assay

Multiple turnover assays were performed with 50 nM reconstituted H/ACA snoRNP and 500 nM substrate RNA. For experiments using long substrate RNAs, the substrates were heated to 75°C for 5 min and cooled slowly to room temperature to promote folding or snap cooled on ice to keep the RNA unfolded. The modification reaction was performed at 30 °C. Samples containing 7.5 – 25 pmol of RNA (depending on the specific activity) were taken over a time course. Samples were quenched and incubated for 10 minutes in 1 mL 5 % (w/v) activated charcoal (Norit A) in 0.1 M HCl. Samples were centrifuged at 10,000 x g for 2 min, and 850 μL of the supernatant was mixed with 300 μL 5% Norit A (w/v) in 0.1 M HCl and centrifuged again. The supernatant was filtered through glass wool and 800 μL of the filtrate was subjected to scintillation counting to determine the amount of pseudouridine formed. Competitive tritium release assays contained 50 nM reconstituted H/ACA snoRNP, 250 nM non-radioactive unmodifiable substrate (G7 insert) and 250 nM tritium-labelled 3′ substrate wild-type. The G7 insert substrate RNA was pre-bound to the H/ACA complex for 3 min before the 3′ substrate wild-type was added to the reaction. Samples were taken and analyzed as described above. Data was analyzed using Microsoft Excel and GraphPad Prism (La Jolla, California, USA). Initial velocities were estimated by linear regression of the initial region of the tritium release assay time course (<70% of the measured end-level) and forcing the fitted line through zero. The slope of the line was taken as the initial velocity.

2.7 - Nitrocellulose Filtration Assay

Guide RNAs and selected substrate RNAs were ^{32}P-labelled. RNA was dephosphorylated with Calf Intestinal Alkaline Phosphatase (New England Biolabs, 0.1 U/μL) and rephosphorylated with T4 polynucleotide kinase (0.5 U/μL) in the presence of [γ-^{32}P] ATP (5–10 μCi). Excess [γ-^{32}P] ATP was removed by using a SigmaPrep spin column (Sigma-Aldrich) containing Sephadex G-25 resin. To measure H/ACA guide RNA binding to the Cbf5-Nop10-Gar1-Nhp2 complex, refolded guide RNA (0.05 nM) was incubated with increasing concentrations of Cbf5-Nop10-Gar1-Nhp2 in Reaction Buffer for 10 min at 30 °C. Binding assays using tritium-labelled substrate RNAs were performed by incubating increasing concentrations of substrate in the presence of 5 nM H/ACA snoRNPs for 3 min at 30 °C. The complete 200 μL reaction was filtered through a nitrocellulose membrane, followed by washing of the nitrocellulose membrane with 1 mL cold Reaction Buffer. The nitrocellulose membrane was dissolved in 10 mL EcoLite scintillation cocktail (EcoLite (+), MP Biomedical) followed by scintillation counting.

The dissociation constant for guide RNA binding was determined as described in Caton *et al.,* 2018 by fitting to the quadratic equation with [RNA] = 0.05 nM:

$$P_{bound} = Amp \times \left[\frac{(K_D + [RNA] + [protein])}{2} - \sqrt{\left\{ \frac{(K_D + [RNA] + [protein])^2}{4} - [protein] \times [RNA] \right\}} \right] \quad (1)$$

where, P_{bound} is the percentage of bound RNA and Amp is the amplitude or final level of bound RNA.

The amount of substrate RNA bound to the H/ACA snoRNP was determined by scintillation counting. Dissociation constants (K_D) were determined by fitting the binding curves to the hyperbolic function in GraphPad Prism

$$Y = B_{max} \times [S] / (K_D + [S]) \quad (2)$$

where [S] is the substrate concentration and B_{max} is the maximum binding. The substrate RNA: enzyme ratio was calculated by dividing the picomoles of substrate RNA retained on the nitrocellulose membrane by the picomoles of enzyme in the reaction.

Table 1. Oligonucleotides for *in vitro* transcription template generation for snR34 5′ hairpin guide RNA variants. Underlined sequences indicate overlapping regions for PCR extension. All oligos are shown in the 5′ to 3′ direction.

Primer Name	**Sequence**	**Paired with:**
snR34 5′HP sense	GCGCTAATACGACTCACTATAGGGAAT CAAAAATTTATTTTTTACACGGAAACGA TGCCACAGTTGACTGAACCTGTCTTCTA ACAG	Various
snR34 5′HP no 5′ tail sense	GCGCTAATACGACTCACTATAGGGAAA CGATGCCACAGTTGACTGAACCTGTCTT CTAACAG	snR34 5′HP antisense
snR34 5′HP antisense	TCTTGTTTGAAATACTGGCAATTAACTA CTGTTAGAAGACAGGTTCAGTCAAC	snR34 5′HP sense
snR34 5′HP no pocket antisense	TCTTGTTTGAAACGATGCCACAATTAAC TACTGTTAGAAGACAGGTTCAGTCAAC	snR34 5′HP sense
snR34 5′HP small pocket antisense	TCTTGTTTGAAACGAACTGGCAATTAAC TACTGTTAGAAGACAGGTTCAGTCAAC	snR34 5′HP sense
snR34 5′HP no lower stem antisense	TCTTGTTTCTTTTACTGGCAATTAACTAC TGTTAGAAGACAGGTTCAGTCAAC	snR34 5′HP sense snR34 5′HP no 5′ tail sense
snR34 5′HP extended lower stem antisense	TCTTGTTTACACGGAAATACTGGCAATT AACTACTGTTAGAAGACAGGTTCAGTC AAC	snR34 5′HP sense snR34 5′HP no 5′ tail sense

Table 2. Oligonucleotides for *in vitro* transcription template generation for snR34 5′ and 3′ substrates. All oligos are shown in the 5′ to 3′ direction.

Primer Name	**Sequence**
T7 promoter extended sense	CGTACAAGCCTTGACGATCGGATGCGCTAATACGACTCACTATAGGG
snR34 3′ sub wt extended antisense	GGTATGATAGGAAGAGCCGACGTCCCTATAGTGAGTCGTATTAGCGCATCCGATCGTCAAGGCTTGTACG
snR34 3′sub Δ1-2 extended antisense	GGTATGATAGGAAGAGCCCTCGTCCCTATAGTGAGTCGTATTAGCGCATCCGATCGTCAAGGCTTGTACG
snR34 3′sub Δ 1-4 extended antisense	GGTATGATAGGAAGAGGGCTCGTCCCTATAGTGAGTCGTATTAGCGCATCCGATCGTCAAGGCTTGTACG
snR34 3′sub Δ16-17 extended antisense	GGTTCGATAGGAAGAGCCGACGTCCCTATAGTGAGTCGTATTAGCGCATCCGATCGTCAAGGCTTGTACG
snR34 3′Δ 14-17 extended antisense	GGTTCCTTAGGAAGAGCCGACGTCCCTATAGTGAGTCGTATTAGCGCATCCGATCGTCAAGGCTTGTACG
snR34 3′sub Δ12-17 extended antisense	GGTTCCTCTGGAAGAGCCGACGTCCCTATAGTGAGTCGTATTAGCGCATCCGATCGTCAAGGCTTGTACG
snR34 3′sub Δ1,12-17 extended antisense	GGTTCCTCTGGAAGAGCCGTCGTCCCTATAGTGAGTCGTATTAGCGCATCCGATCGTCAAGGCTTGTACG
snR34 3′sub Δ1-2,12-17 extended antisense	GGTTCCTCTGGAAGAGCCCTCGTCCCTATAGTGAGTCGTATTAGCGCATCCGATCGTCAAGGCTTGTACG
snR34 3′sub 10CC-GG extended antisense	GGTATGATACCAAGAGCCGACGTCCCTATAGTGAGTCGTATTAGCGCATCCGATCGTCAAGGCTTGTACG
snR34 3′ sub 7CU-GA extended antisense	GGTATGATAGGATCAGCCGACGTCCCTATAGTGAGTCGTATTAGCGCATCCGATCGTCAAGGCTTGTACG
snR34 3′sub G7 insert extended antisense	GGTATGATAGGAAGCAGCCGACGTCCCTATAGTGAGTCGTATTAGCGCATCCGATCGTCAAGGCTTGTACG
snR34 5′ sub wt extended antisense	AGGCAGCCACAAGCCAGTTGTCCCCTATAGTGAGTCGTATTAGCGCATCCGATCGTCAAGGCTTGTACG

snR34 5′sub YRA1 extended antisense	GATGTTAGCCATACCAGTAGATTGGCCCCTATAGTGAGTCGTATTAGCGCATCCGATCGTCAAGGCTTGTACG
Sc 25S rRNA H89 U2826 (5′ sub) sense	GCTAATACGACTCACTATAGGGATAACTGGCTTGTGGCAGTCAAGCGTTCATAGCGAC
Sc 25S rRNA H89 U2826 (5′ sub) antisense	CATCGAAGAATCAAAAAGCAATGTCGCTATGAACGCTTGACTGCC
Sc 25S rRNA H90-92 U2880 (3′ sub) sense	GCTAATACGACTCACTATAGGGATGTCGGCTCTTCCTATCATACCGAAGCAGAATTCGGTAAGCG
Sc 25S rRNA H90-92 U2880 (3′ sub) antisense	CAGCTCACGTTCCCTATTAGTGGGTGAACAATCCAACGCTTACCGAATTCTGCTTCGG

Chapter 3 – Results

3.1 - Expression and Purification of Cbf5-Nop10 and Gar1

To reconstitute an H/ACA snoRNP complex *in vitro* from purified components, the proteins and the guide RNA are first generated separately. Cbf5 and Nop10 are co-expressed in *E. coli* to ensure that Cbf5 remains stable and soluble. Freshly transformed test expression cultures of Cbf5-Nop10 were performed using *E. coli* BL21(DE3) (Caton *et al.*, 2018). Individual colonies were used to inoculate separate cultures. Each culture was induced with ITPG when it reached mid-log phase (OD_{600} ~0.6-0.8) and the temperature was reduced to 18 °C. A culture sample of 1 OD_{600} was taken at 0 h and at 16 h and was analyzed by SDS-PAGE (Figure 4A). Cbf5 is a 65 kDa protein, and a band corresponding to Cbf5 appears in the 16 h samples of all cultures. However, the intensity of this band varies considerably between cultures; the most intense band appears in sample 4 and the Cbf5 band in sample 1 is very faint. This differential expression may be the result of escape mutants that appeared after induction generated stress on the cells. Nop10 cannot be observed by standard SDS-PAGE as it is too small (7 kDa); however, it has been shown to be expressed using this system (Caton *et al.*, 2018). Culture 4 was used to inoculate a large-scale expression culture. An intense band corresponding to Cbf5 appears after 1 h and increased slightly after 16 h (Figure 4B). GST-tagged Gar1 (45 kDa) was also overexpressed, with intense bands appearing 1 h after induction with IPTG (Figure 4B). Cells harvested from these two over expressions were combined and lysed, allowing the trimeric Cbf5-Nop10-Gar1 complex to form. The complex was purified by sequential affinity purification, first with glutathione sepharose to pull down GST-tagged Gar1 and

then nickel sepharose to capture the hexahistidine-tagged Cbf5 (Figure 4C). Glutathione sepharose elutions (E) contained both Cbf5 and GST-Gar1.

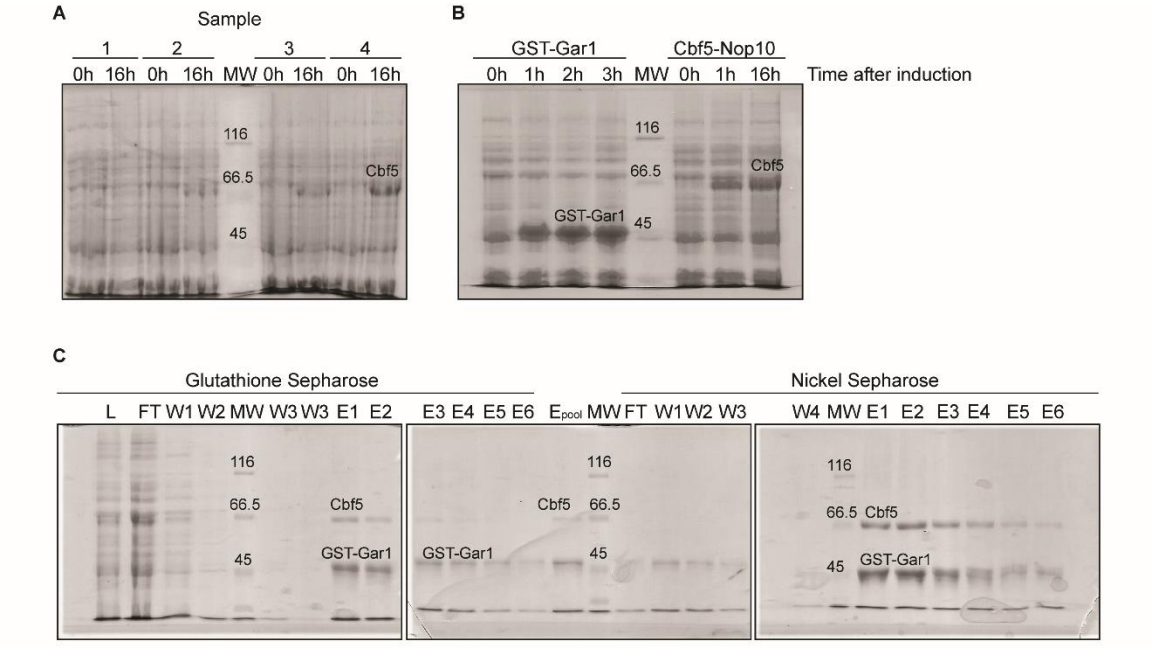

Figure 4. Expression and purification of recombinant H/ACA proteins. (A) 12% SDS-PAGE of a Cbf5-Nop10 test expression from four different colonies. Strong Cbf5 expression can be seen at 16 h in sample 4. (B) 12% SDS-PAGE of a large-scale overexpression of GST-tagged Gar1 (45 kDa) and of co-expressed histidine-tagged Cbf5-Nop10 (Cbf5: 65 kDa, Nop10: 7 kDa). Nop10 was expressed from two plasmids (see Materials & Methods). (C) 12% SDS-PAGEs of the affinity purification of the Cbf5-Nop10-Gar1 ternary complex. Bands corresponding to Cbf5 and GST-Gar1 are clearly visible. Sizes of proteins in the molecular weight marker (MW) are indicated. Cell lysate (L), flow through (FT), wash (W), elution (E), pooled elutions (E_{pool}). Nop10 is not visible on 12% SDS-PAGEs. All gels are stained with Coomassie brilliant blue.

These elutions were pooled before they were loaded onto nickel sepharose resin. The nickel sepharose elutions also contained both Cbf5 and GST-Gar1. Equal volumes of each elution sample were loaded onto the SDS-PAGE, and the increasing intensity of both the Cbf5 and GST-Gar1 bands compared to the glutathione sepharose elutions indicates that the nickel sepharose elutions are more concentrated. The GST-Gar1 band is more intense than the Cbf5 band, demonstrating that 1:1 stoichiometry was not obtained. This is likely due to the ability of GST to dimerize such that not all GST-Gar1, that is not complexed with Cbf5,

was removed in the glutathione sepharose or nickel sepharose washes (Terpe, 2003). The purification yielded 2470 pmol Cbf5-Nop10-Gar1 complex from 25 g Cbf5-Nop10 expression cells and 5 g GST-Gar1 expression cells (based of Cbf5 quantification using band intensity analysis in ImageJ).

3.2 – Determinants of Guide RNA Recognition (Adapted from (Caton *et al.*, 2018)[1])

Though the two-hairpin structure of the H/ACA snoRNA is common in eukaryotes, the sequence of the RNA varies widely, leaving the only conserved sequence elements to be the H (ANANNA) and ACA boxes (Piekna-Przybylska *et al.*, 2008). Additionally, it was shown that mutation of the H and ACA boxes in snR34 did not affect the affinity of the Cbf5-Nop10-Gar1-Nhp2 complex for the guide RNA (Caton *et al.*, 2018). To test if the Cbf5-Nop10-Gar1-Nhp2 complex recognized another feature of the guide RNA, snR34 guide RNA 5′ hairpin variants were designed based on the 5′ hairpin sequence and structure of the H/ACA RNA snR34 (Figure 5). Various structural features of the guide RNA were altered. The 5′ single stranded region (5′ extension) was removed, the pseudouridylation pocket was reduced or removed through the introduction of complementary nucleotides, the lower stem was abolished through the removal of complementary nucleotides, and the lower stem was extended by introducing nucleotides complementary to the 5′ single-stranded region. The internal ribosome entry site (IRES) of the Cricket Paralysis Virus (CrPV) was used as a structured RNA control that did not contain features of an H/ACA RNA. These RNAs were used in nitrocellulose filtration assays to determine the affinity of the Cbf5-Nop10-Gar1-Nhp2 complex for the guide RNA variant. These guide RNAs are

[1] In (Caton *et al.*, 2018), I performed the experiments described in this thesis. Evan A. Caton and Rajashekhar Kamalampeta performed all other experiments in the study.

also used in tritium release assays with short [^{3}H-C5] uridine-labeled substrate RNA to determine if any variations to the guide RNA affect the ability of the snoRNP to pseudouridylate a substrate.

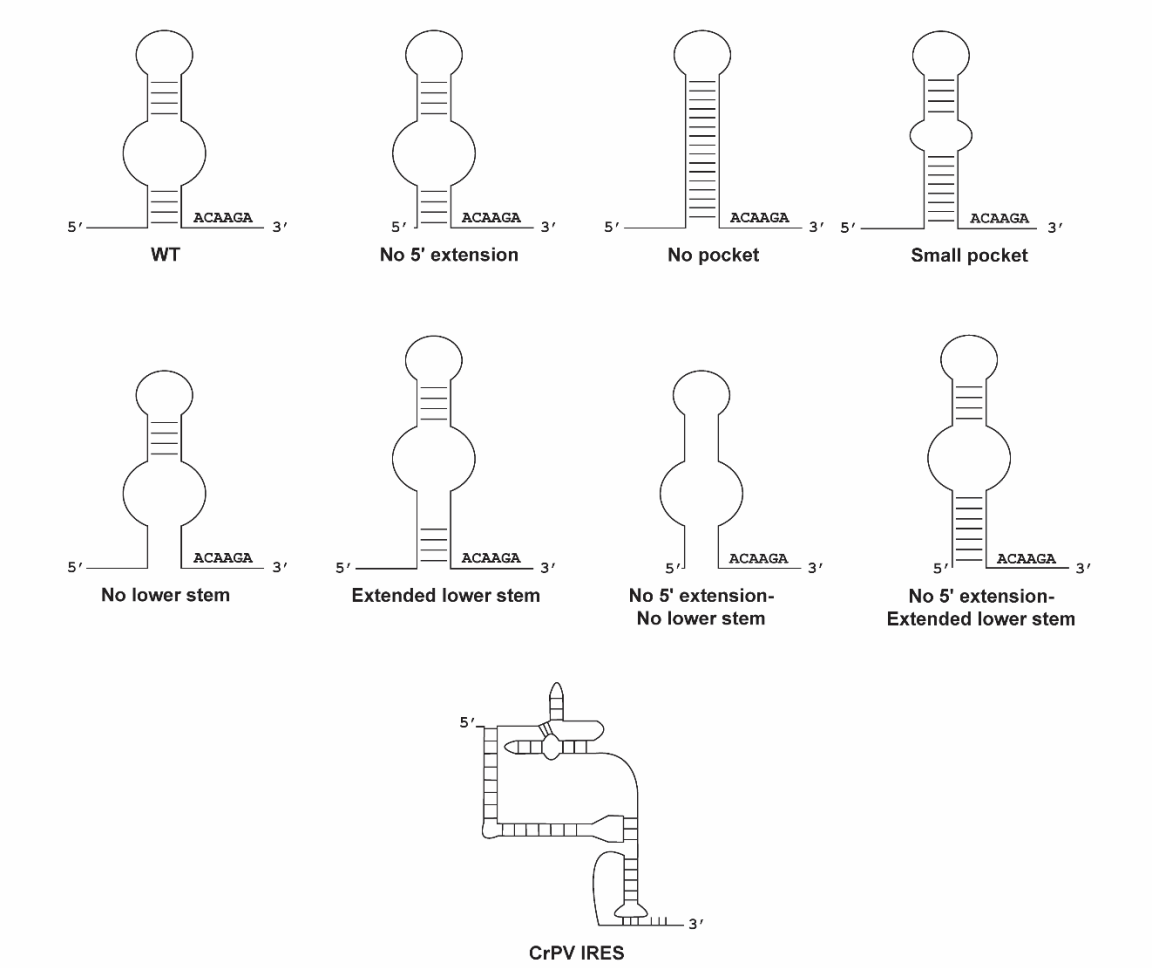

Figure 5. Variants of the 5′ hairpin of the H/ACA guide RNA snR34. The 5′ hairpin was mutated to delete the 5′ extension, eliminate or reduce the single stranded pseudouridylation pocket, and remove or extend the lower stem of the hairpin. The Cricket Paralysis Virus Internal Ribosome Entry Site (CrPV IRES) was used as a non-H/ACA structured RNA comparison (Jan & Sarnow, 2002).

Guide RNAs were *in vitro* transcribed and purified by gel extraction. Purified RNAs were analyzed by 10% Urea-PAGE for relative size confirmation (Figure 6, A, B). The snR34 5′ hairpin wild type guide RNA is 92 nt long. Deleting the 5′ extension changes the

size of the RNA to 68 nt, resulting in faster migration through the Urea-PAGE. The other guide RNAs have the following expected sizes: no pocket – 94 nt, small pocket – 94 nt, no lower stem – 92 nt, extended lower stem – 97 nt. These guide RNAs migrate similarly through the gel to 5′ hairpin wild type indicating that they are the appropriate size. The guide RNAs shown in Figure 6A were used in nitrocellulose filtration and tritium release assays (vide infra). Additionally, two guide RNA variants have been produced, but have only been preliminarily tested (Figure 5 and 6B). These RNAs are a combination of the no 5′ extension and either the no lower stem or extended lower stem variants. The expected sized of these guide RNAs are 68 nt (no 5′ extension-no lower stem) and 73 nt (no 5′ extension- extended lower stem). These RNAs migrated through the gel at a relatively correct distance for their sized compared to each other. Some lanes reveal a secondary IVT product of a smaller size than the main product (Figure 6A, Extended lower stem). These additional, smaller products may be the result of incomplete transcription.

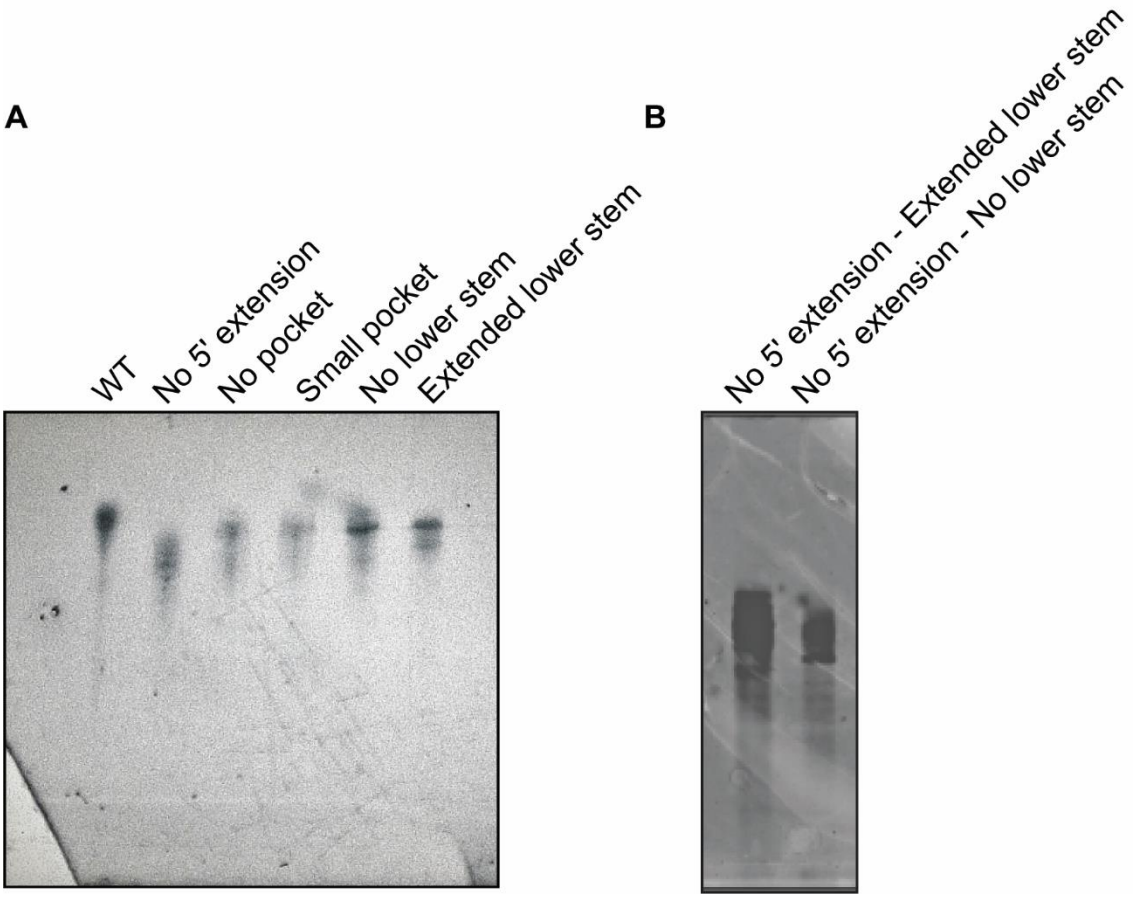

Figure 6. Urea PAGE analysis of *in vitro* transcribed and purified guide RNA variants. (A) 10% Urea-PAGE of the snR34 5′ hairpin guide RNA variants stained with ethidium bromide and imaged by UV transillumination (10 pmol each). (B) 10% Urea-PAGE of additional 5′ hairpin guide RNA variants with combines features removed, stained with SYBR Green and imaged with a Typhoon imager (GE Healthcare).

For all guide RNA variants, the affinity to the Cbf5-Nop10-Gar1-Nhp2 protein complex and the ability to facilitate pseudouridine formation was assessed. Nitrocellulose filter binding assays were performed to determine the dissociation constant (K_D) of the Cbf5-Nop10-Gar1-Nhp2 complex for each guide RNA variant (Figure 7 A-G, Table 3). One set of proteins should bind the 5′ hairpin. Each nitrocellulose filtration assay was performed in triplicate. The K_D of Cbf5-Nop10-Gar1-Nhp2 for snR34 5′ hairpin wild-type is 1.5 ± 0.8 nM, indicating that the binding interaction is very strong. snR34 has an

unstructured extension on its 5′ end and deletion of the 5′ extension did not exhibit any effect on the ability of Cbf5-Nop10-Gar1-Nhp2 to bind the guide RNA (K_D = 1.0 ± 0.3 nM). Abolishing or shortening the single stranded pseudouridylation pocket also did not affect Cbf5-Nop10-Gar1-Nhp2 binding. Mutating the lower stem of the guide RNA to prevent base pairing or extending the stem did not affect binding either. The binding curve for the Extended Lower Stem guide RNA did not reach the same end levels as the other guide RNAs in the nitrocellulose filtration assays. This is likely because this guide was not as efficiently labelled with ^{32}P compared to the other guides, and there was more γ-^{32}P-ATP remaining in the sample after purification. Excess γ-^{32}P-ATP would increase the apparent specific activity of the sample, but would be washed way during nitrocellulose filtration such that 100% of the radioactivity cannot be retained on the filter. Alternative explanations may include that some RNA was incorrectly folded and therefore unable to bind to the protein complex. However, the K_Ds calculated for the Extended Lower Stem guide RNA were consistent between trials. Guide RNA variants combining the features of the 5′ extension deletion and the "no lower stem" and the "extended lower stem" have been produced but have not yet been tested for Cbf5-Nop10-Gar1-Nhp2 binding ability. Together, these data show that Cbf5 binds H/ACA guide RNA irrespective of key H/ACA RNA structural features and independent of the presence of single-stranded or double-stranded regions in specific positions. Next, we asked whether Cbf5–Nop10-Gar1-Nhp2 can bind to an unrelated RNA with complex structure utilizing a cricket paralysis virus internal ribosome entry site (IRES) RNA (Jan & Sarnow, 2002). Surprisingly, even this viral RNA was bound very tightly by the H/ACA protein complex with a sub-nanomolar affinity (Figure 7G, Table 3). In conclusion, the Cbf5–Nop10–Gar1–Nhp2 complex binds all guide RNA variants analyzed so far as tightly to the H/ACA protein complex as the

wild-type 5′ hairpin of snR34, indicating that the H/ACA protein complex binds RNA non-specifically.

The guide RNAs were also tested for their ability to pseudouridylate a short 5′ substrate RNA using a tritium release assay (Figure 7H). As expected, elimination of the pseudouridylation pocket abolishes the activity to modify the substrate RNA. Interestingly, the No 5′ Extension and Small Pocket guide RNAs confer slightly faster rates of pseudouridylation than the wild type 5′ hairpin. Even more interestingly, the No Lower Stem and Extended Lower Stem guides were very efficient at pseudouridylation of the short substrate RNA and demonstrated rates even faster than the No 5′ Extension and Small Pocket. It was not necessarily unexpected that these guides would be active in the H/ACA snoRNP in guiding pseudouridylation; however, it was not expected that they were much more active than the wild type guide 5′ hairpin. The No 5′ extension-No lower stem and No 5′ extension-Extended lower stem combination guide RNAs were also analyzed by tritium release assay (Figure 8). Tritium release by snoRNPs using these guide RNA variants is much more comparable to the snR34 5′ hairpin wild-type. In these assays, the snR34 5′ hairpin wild-type snoRNP pseudouridylated the substrate to a final level of approximately 55% at 150 min, the No 5′ extension-No lower stem snoRNP reached a final level of 33% and the No 5′ extension-Extended lower stem snoRNP reached a final level of 51% pseudouridine formation.

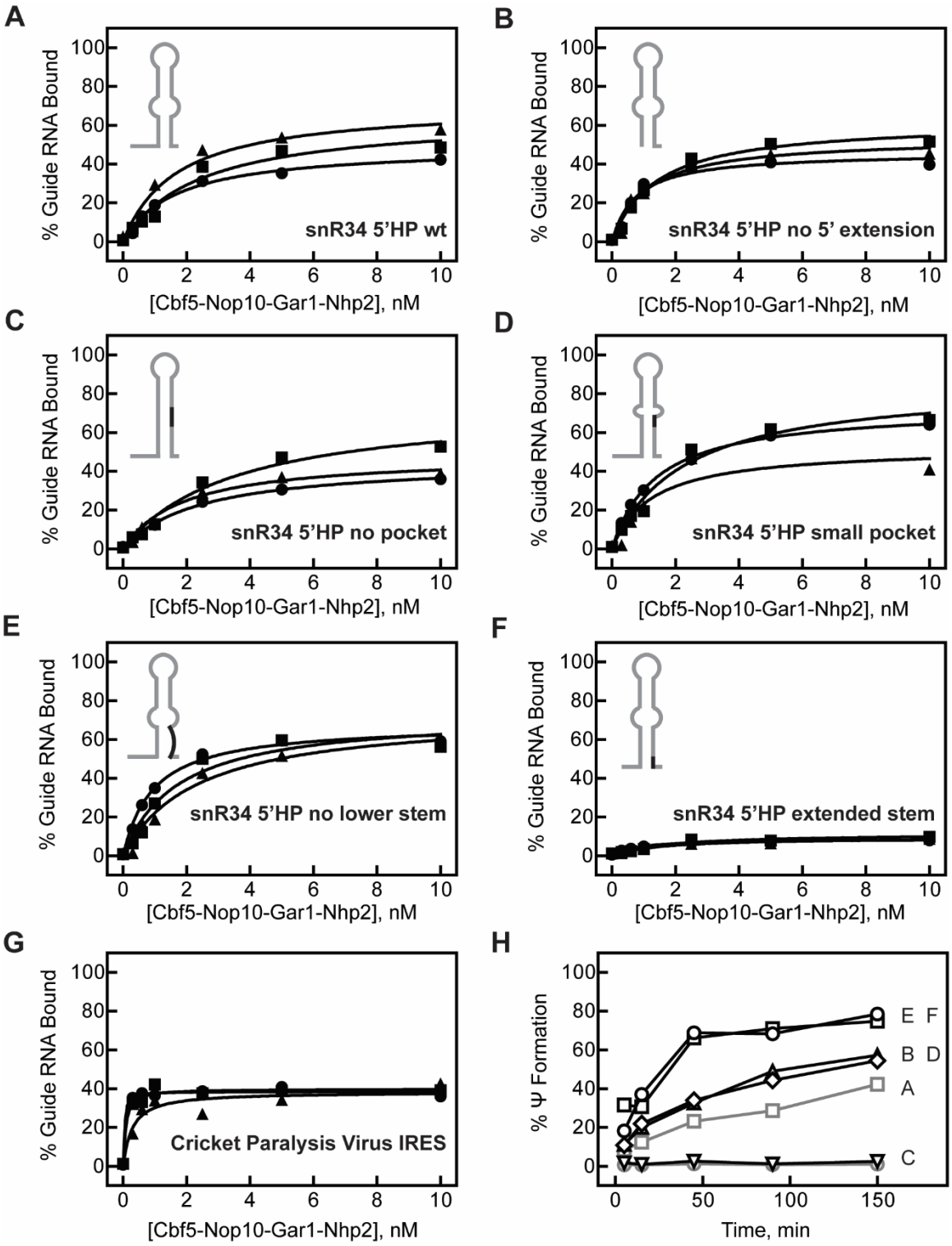

Figure 7. Specificity of H/ACA proteins for guide RNA. A series of different variants of the 5′ hairpin of snR34 were generated, and the affinity of Cbf5-Nop10-Gar1-Nhp2 for these guide RNAs was determined by nitrocellulose filtration. Each panel shows three replicates of the same experiment generated on different days. For each experiment the

dissociation constant, K_D, was obtained by fitting to a quadratic equation (smooth lines, see Materials and Methods). The average K_D values are given in Table 1. (A) snR34 5′ hairpin wild-type. (B) snR34 5′ hairpin without 5′ extension, i.e. the first 24 nt of snR34 were deleted that do not form base pairs with the rest of the 5′ hairpin. (C) snR34 5′ hairpin without pseudouridylation pocket. The single-stranded pseudouridylation pocket was converted into a double-stranded region by changing the 3′ site of the pocket rendering it complementary to the 5′ site (indicated in black in the schematic representation). (D) snR34 5′ hairpin with a small pseudouridylation pocket. Three additional base pairs were introduced at the bottom of the pseudouridylation pocket by changing the sequence on the 3′ site (black). The nucleotides that interact with substrate RNA remain unpaired in this structure. (E) snR34 5′ hairpin without lower stem underneath the pseudouridylation pocket. The base-pairing in the lower stem was abolished by mutating the residues on the 3′ site (black). (F) snR34 5′ hairpin with an extended lower stem. The lower stem was elongated by inserting 5 nucleotides on the 3′ site between the lower stem and the Box H element (black) which will form 5 additional base pairs with the 5′ extension at the bottom of the lower stem. (G) Cricket Paralysis Virus IRES. This is a 216 nt long, viral RNA with extensive secondary and tertiary structure that generates an internal ribosome entry site (IRES) to facilitate cap-independent mRNA translation (Jan & Sarnow, 2002). (H) Pseudouridylation assays using the different snR34 5′ hairpin variants (100 nM) together with 500 nM 5′ substrate RNA. Letters on the right indicate the guide RNA used as shown in panels A to F. Specifically, grey symbols and lines show the control experiments without enzyme (circles) and snR34 5′ hairpin wild-type (squares). The snR34 5′ hairpin variants are displayed in black as follows: snR34 5′ hairpin without 5′ extension (triangles), snR34 5′ hairpin without pseudouridylation pocket (inverted triangles), snR34 5′ hairpin with a small pseudouridylation pocket (diamonds), snR34 5′ hairpin without lower stem (squares), and snR34 5′ hairpin with an extended lower stem (circles). Reprinted from (Caton *et al.*, 2018).

Table 3. Binding of guide RNA to H/ACA proteins forming an H/ACA RNP complex. Reprinted from (Caton *et al.*, 2018). Dissociation constants are stated with standard deviation and were determined from nitrocellulose filter binding shown in Figure 7.

Guide RNA	**K_D (nM)**
snR34 5′ hairpin WT	1.5 ± 0.8
snR34 5′ hairpin no 5′ extension	1.0 ± 0.3
snR34 5′ hairpin no pocket	2.6 ± 0.8
snR34 5′ hairpin small pocket	1.6 ± 0.5
snR34 5′ hairpin no lower stem	1.6 ± 0.7
snR34 5′ hairpin extended stem	1.3 ± 0.5
CrPV IRES	0.1 ± 0.1

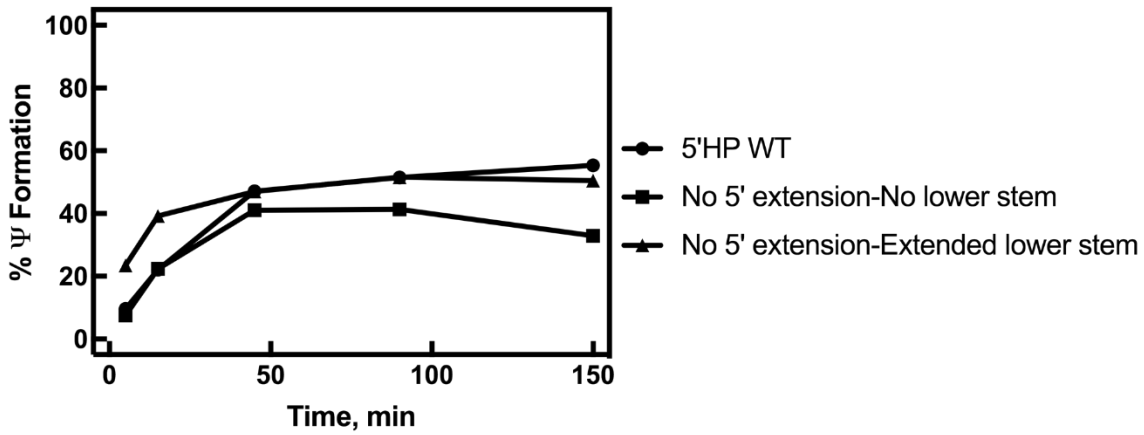

Figure 8. Pseudouridylation of 5′ substrate RNA using No 5′ extension-No lower stem and No 5′ extension-Extended lower stem variants. Pseudouridylation assays were conducted using the different snR34 5′ hairpin variants (100 nM) together with 500 nM 5′ substrate RNA. One replicate of each assay was performed.

3.3 - Substrate RNA Variants Require a Minimal Continuous Base Pairing Interaction with the Pseudouridylation Pocket to be Modified

To investigate the requirements of the H/ACA guide-substrate interaction in the pseudouridylation pocket, short substrate RNAs were designed based on the sequence of the 25S rRNA in yeast complementary to the pseudouridylation pocket of the 5′ and 3′ hairpins of snR34 (designated 5′ substrate and 3′ substrate) (Figure 9). Mismatches were introduced into the 3′ substrate by substituting nucleotides by the following rules: G to C, C to G, U to A and A to G. Adenine nucleotides were not changed to uridines to avoid the introduction of novel uridines. A short substrate RNA was also designed for the yeast mRNA YRA1, which had been predicted to be pseudouridylated by the 5′ hairpin of snR34 (Figure 9B) (Schwartz *et al.*, 2014). Flanking sequences were added to substrate RNAs to increase their length to improve *in vitro* transcription and purification (6 nt on the 5′ end and 3 nt on the 3′ end of the 3′ hairpin substrates, 7 nt on the 5′ end and 3 nt on the 3′ end of the 5′ hairpin substrate, 10 nt on the 5′ end and 7 nt on the 3′ end of the YRA1 mRNA substrate – Figure 9A, B). The substrate RNAs are named according to the location within

the base pairing region of the substrate RNA (5′ to 3′) in which mismatches are introduced. For example, Δ1-2, 12-17 indicates that from the 5′ side of the base pairing region of the substrate RNA, mismatches occur at nucleotides 1-2 and 12-17. 10CC-GG indicates that the two CG base pairs beginning at position 10 in the substrate RNA are mutated to GG mismatches. Short substrate RNA base pairing is described in Figure 9.

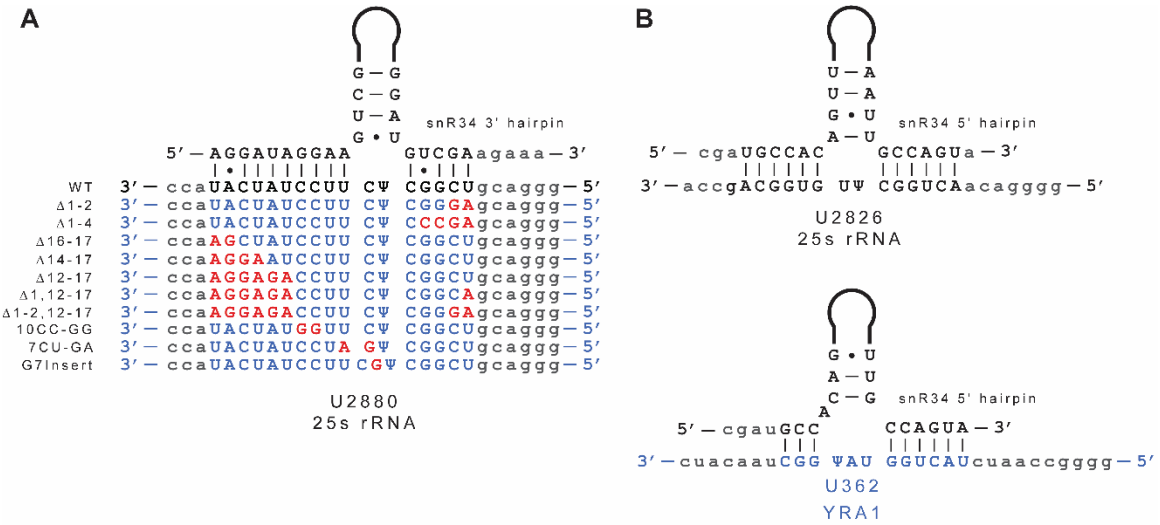

Figure 9. Short substrate RNA variants in the pseudouridylation pockets of snR34. (A) The base pairing region of the pseudouridylation pocket and a segment of the upper stem of snR34 3′ hairpin is displayed 5′ to 3′. The wild type sequence of the region of 25S rRNA that base-pairs in the pseudouridylation pocket of snR34 3′ hairpin is displayed 3′ to 5′ (black sequence – WT). Sequences written in blue are the substrate variants designed to test the limits of substrate binding and modification. Red nucleotides indicate the changed nucleotides from the wild type sequence. The names of the substrate RNAs indicate which nucleotides have been changed, referring to their position in the substrate RNA from the 5′ end of the base pairing region. In the G7 insert substrate, an extra unpaired nucleotide was inserted adjacent to the target uridine. Grey lower-case sequences are regions in the pseudouridylation pocket and substrate RNAs that do not base pair. (B) Short substrate RNAs were designed for the 5′ hairpin pseudouridylation pocket of snR34. The wild type 25S rRNA sequence is displayed bound in the pseudouridylation pocket in black. The predicted mRNA substrate YRA1 is shown bound to the pseudouridylation pocket in blue. Grey lower-case sequences are regions in the pseudouridylation pocket and substrate RNAs that do not base pair.

Substrate RNAs were also *in vitro* transcribed and purified by gel extraction (Figure 10). Each RNA migrates similarly to the 3′ hairpin wild type substrate as they are designed to be the same size. The 5′ hairpin substrate YRA1 migrates slightly more slowly through the gel compared to the 3′ hairpin wild type substrate as it is 29 nt, rather than 25 nt like the rest of the substrates. The 10CC-GG substrate RNA was purified previously (data not shown). Again, the urea PAGE reveals that some substrate RNA samples contain a second *in vitro* transcription product, as indicated by multiple bands in some samples (Ex. Figure 10 3′sub Δ1-4). This may be due to incomplete *in vitro* transcription.

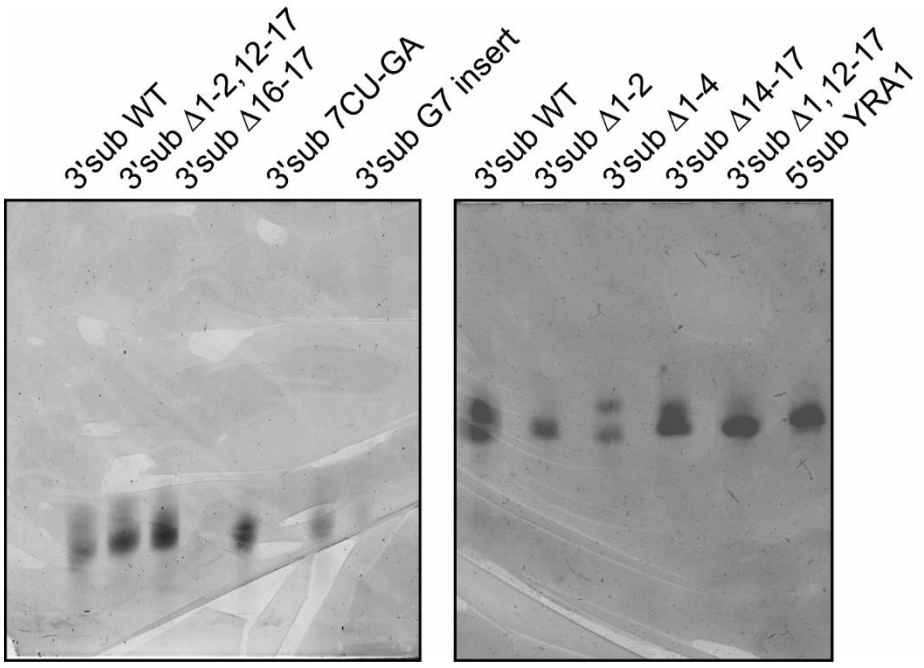

Figure 10. Urea PAGE analysis of *in vitro* transcribed and purified substrate RNA variants. 15% urea PAGEs of tritium-labelled, gel-extracted short substrate RNA variants, stained with SYBR Green and imaged with a Typhoon imager (GE Healthcare).

Substrate RNA variants were analyzed for pseudouridine formation by the snR34 H/ACA snoRNP. Substrate RNA variants that made fewer base pairs with the 3′ side of the pseudouridylation pocket (Δ1-2, and Δ1-4) revealed slower pseudouridine formation compared to the wild type substrate RNA (Figure 11A, Table 4). The wild type substrate

forms 5 base pairs with the 3′ side of the pseudouridylation pocket, whereas the Δ1-2 and Δ1-4 substrates form only 3 and 1 base pair, respectively. The estimated initial velocity of the wild type substrate is 26 ± 5 nM min^{-1} (Table 4). However, the estimated initial velocity of Δ1-2 and Δ1-4 are 2.0 ± 0.1 nM min^{-1} and 0.20 ± 0.02 nM min^{-1}, respectively (Table 4). These substrates are modified 10- and 100-fold more slowly than the wild type substrate, respectively, suggesting that reduced base pairing in the 3′ side of the pseudouridylation pocket strongly affects the rate at which pseudouridines can be formed in substrate RNAs. In these cases, only 3 or 1 base-pair are remaining on the 3′ side of the pseudouridylation pocket indicating that these are not sufficient to allow for efficient pseudouridine formation.

RNA variants with reduced base pairing in the 5′ side of the pseudouridylation pocket (Δ12-17, Δ14-17 and Δ16-17) were also tested for activity (Figure 11B). The Δ16-17 substitutions were farthest from the target uridine of any of the substrate variants. The Δ16-17 substrate was pseudouridylated at a rate similar to the wild type substrate. These substitutions do not influence the rate at which this substrate is modified compared to the wild type sequence (Figure 11B, Table 4). The Δ14-17 substrate is ~50% pseudouridylated after 45 min and does not increase substantially after 360 min (Figure 11B). The estimated initial velocity of Δ14-17 is 12.3 ± 0.5 nM min^{-1}. The estimated initial velocity of Δ12-17 is 6.5 ± 0.4 nM min^{-1} (Table 4). The initial rate of the Δ14-17 reaction is faster than the Δ12-17 initial rate, though it appears to reach a lower end level near 50% pseudouridylation. This may be indicative of a problem with the substrate RNA sample where only half of the RNA may be able to be modified. Importantly, the Δ14-17 substrate RNA that can form more base pairs with the pocket (6 base pairs with the 5′ side of the pocket, 5 base pairs with the 3′ side of the pocket) is initially modified faster than the Δ12-

17 substrate that forms fewer base pairs with the pseudouridylation pocket (4 base pairs with the 5′ side of the pocket and 5 base pairs with the 3′ side of the pocket).

Substrate RNAs with substitutions that result in fewer base pairs formed with both sides of the pseudouridylation pocket were also analyzed. The Δ1,12-17 was pseudouridylated more slowly than the wild type substrate with an estimated initial velocity of 4.2 ± 0.5 nM min^{-1} (Figure 11C, Table 4). The Δ1-2,12-17 substrate was modified extremely slowly with an estimated initial velocity of 0.43 ± 0.01 nM min^{-1} (Figure 11C, Table 4). The gradual reduction of base pairing in the 3′ side of the pseudouridylation pocket further diminishes the rate of pseudouridylation compared to substrates already exhibiting fewer base pairs with the 5′ side of the pseudouridylation pocket. The slow modification of the Δ1-2, 12-17 substrate RNA appears to be a combined effect of the diminished modification rates of both the Δ1-2 and the Δ12-17 substrate RNAs.

Finally, substrate RNAs that would introduce a bulge in the helix through mismatches with the substrate RNA (10CC-GG and 7CU-GA), and an RNA containing an extra unpaired nucleotide adjacent to the target uridine (G7 insert) were analyzed with a tritium release assay. The 10CC-GG, 7CU-GA and G7 insert substrate RNAs all displayed drastically reduced initial rates compared to the wild-type substrate reaching less than 15% pseudouridine formation after 150 min, with all obtaining estimated initial velocities less than 1 nM min^{-1} (Figure 11D, Table 4).

Modification of substrates targeted by the 5′ pseudouridylation pocket of snR34 was also investigated. Schwartz *et al* (2014) predicted that the snR34 5′ pseudouridylation pocket could modify the YRA1 mRNA at position 362 (Figure 12). However, accommodation of this mRNA into the pseudouridylation pocket would require

remodelling of the pseudouridylation pocket and the upper stem of the H/ACA RNA. Tritium release assays of the *in vitro* transcribed 5′ wild type substrate yielded an estimated initial velocity of 11 ± 0.5 nM min^{-1}. Analysis of the YRA1 fragment substrate revealed that this substrate was modified at an extremely slow rate compared to the 5′ wild type substrate preventing the determination of an initial velocity (Figure 12).

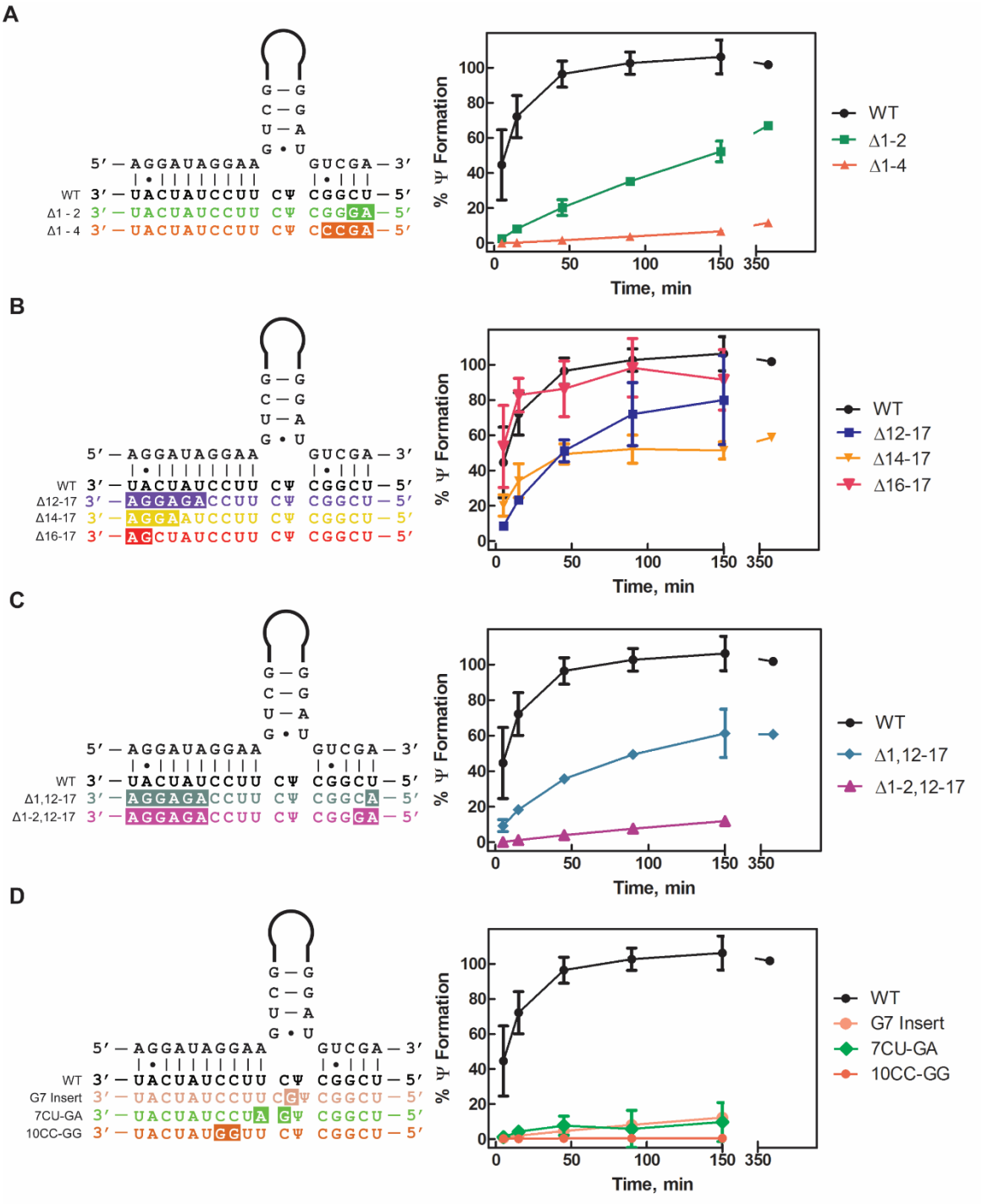

Figure 11. ***In vitro*** **pseudouridylation of 3′ hairpin short substrate variants by the snR34 H/ACA snoRNP.** An excess (500 nM) of each [^{3}H-C5] uridine-labeled 3′ substrate RNA was incubated with 50 nM reconstituted snR34 H/ACA snoRNP. At the given time points, pseudouridylation was quantified by tritium release assay. Location and sequence of substitutions are highlighted in the substrate RNA sequences on the left side. (A) Modification of substrate RNA variants with mismatches in the 3′ side of the

pseudouridylation pocket. The wild-type substrate (black sequence, circles), Δ1-2 substrate (green sequence, squares) and the Δ1-4 substrate (orange sequence, triangles) are compared. (B) Modification of substrate RNA variants with mismatches in the 5′ side of the pseudouridylation pocket. The wild type substrate (black sequence, circles), Δ12-17 substrate (indigo sequence, squares), Δ14-17 substrate (yellow sequence, inverted triangles) and Δ16-17 substrate (red sequence, bold inverted triangles) are compared. (C) Modification of substrate variants with mismatches in both sides of the pseudouridylation pocket. The wild type substrate (black sequence, circles), Δ1,12-17 substrate (teal sequence, diamonds) and the Δ1-2,12-17 substrate (magenta sequence, triangles) are compared. (D) Modification of substrates with mismatches at internal sites in the pseudouridylation pocket. The wild type substrate (black sequence, circles), 10CC_GG substrate (orange sequence, circles) 7CU-GA substrate (green sequence, diamonds), and G7 insert substrate (peach sequence, circles) are compared. Mean and standard deviation of three replicates are shown.

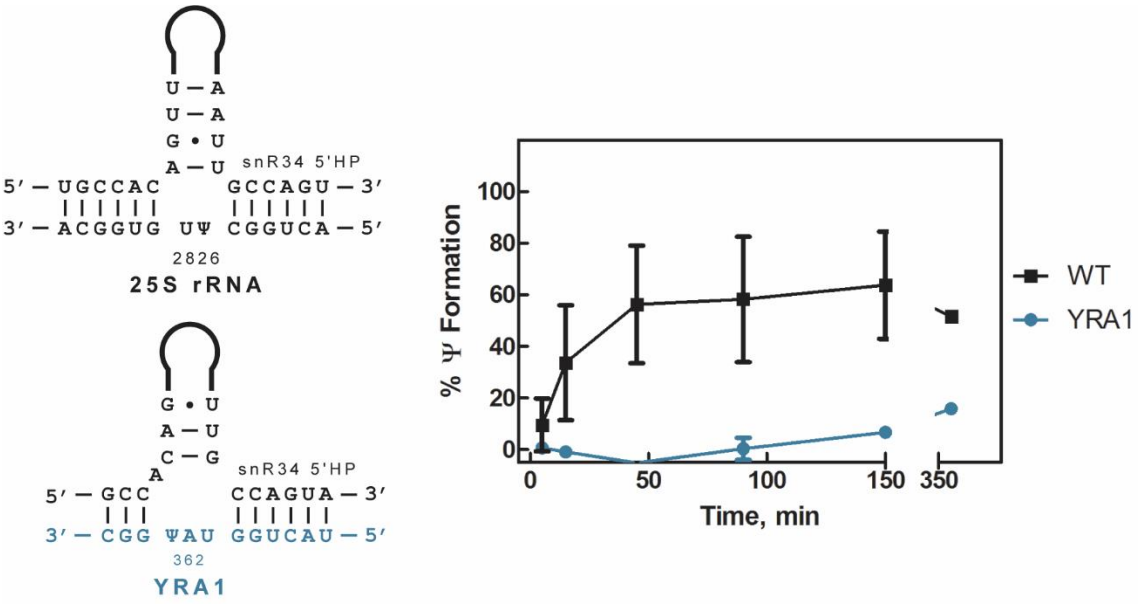

Figure 12. *In vitro* pseudouridylation of 5′ hairpin short substrate variant by the snR34 H/ACA snoRNP. An excess (500 nM) of each [^{3}H-C5] uridine-labeled 5′ substrate RNA was incubated with 50 nM reconstituted snR34 H/ACA snoRNP. At the given time points, pseudouridylation was quantified by tritium release assay. The mRNA YRA1 was predicted to be pseudouridylated by the 5′ hairpin of snR34 through the indicated base pairing by (Schwartz *et al.*, 2014). The 5′ wild type substrate (black sequence, squares) and the YRA1 substrate fragment (blue sequence, circles) are compared. Mean and standard deviation of three replicates are shown.

Next, nitrocellulose filtration assays were used to determine the affinity of the snR34 H/ACA snoRNP for the different substrate RNA variants to investigate if any of the substitutions cause a change in the ability of the H/ACA snoRNP to bind the substrate

RNAs. Thereby, it can be assessed whether the lack of modification is due to a change in the binding ability. The concentration of Cbf5-Nop10-Gar1 does not allow titration to high concentrations; therefore, 5 nM reconstituted snR34 H/ACA snoRNP was used in the filtration experiments as this complex is stable at such low concentration (Caton *et al.*, 2018). Radiolabelled substrate RNA was titrated against the H/ACA snoRNP, and control background (no H/ACA snoRNP) experiments were performed in tandem confirmed a minimal background signal from substrate RNA alone (Figure 13, Figure A1, Table 4). Theoretically, the bound substrate per H/ACA snoRNP ratio should be 1, as the substrate RNA should only bind to one pseudouridylation pocket of the snoRNP. However, we have observed that this ratio does change depending on the age of the substrate RNA sample and may increase over time. This does not change the K_D determined by the fitting of Equation 2, as the K_Ds measured between replicates are within the same order of magnitude. The H/ACA snoRNP generally binds all substrate RNAs tightly, with K_Ds in the nanomolar range.

Comparing the dissociation constants with the activity of the H/ACA snoRNP on the substrate RNA suggests that binding and pseudouridylation are not correlated (Table 4). Of the substrates in which mismatches are introduced from the periphery of the base pairing region, the only substrate with an increased K_D is Δ1-2,12-17. Of the substrates introducing substitutions at internal sites in the helices between the substrate and pseudouridylation pocket, only 10CC-GG displays an increased K_D. Otherwise, some of

the substrates that present suboptimal or minimal activity are very tightly bound by the H/ACA snoRNP, such as Δ1-4, 7CU-GA and G7 insert.

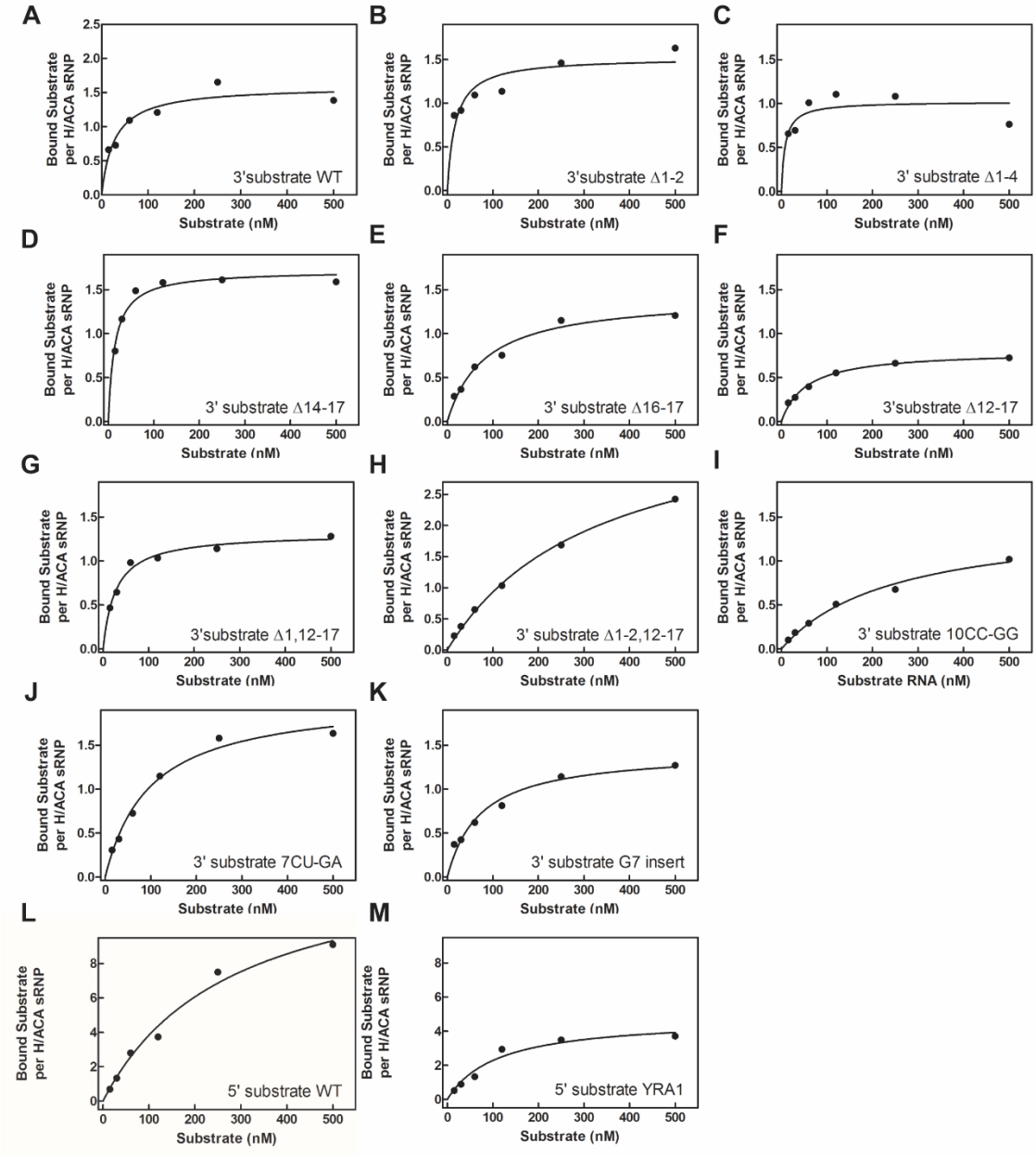

Figure 13. Affinity of the snR34 H/ACA snoRNP for short substrate RNA variants. Affinity of the snoRNP was determined for each substrate RNA variant using nitrocellulose filtration. Each panel shows a representative curve of either duplicate or triplicate trials. Full data sets are shown in the Appendix (Figure A1). The dissociation constant (K_D) was determined by fitting the data to Equation 2 (smooth lines – see Materials and Methods). Dissociation constants are listed in Table 4. (A) 3′ hairpin substrate wild type (n=1, previously determined by (Caton *et al.*, 2018)). (B) 3′ hairpin substrate Δ1-2 (n=3). (C) 3′ hairpin substrate Δ1-4 (n=3). (D) 3′ hairpin substrate Δ14-17 (n=3). (E) 3′ hairpin substrate

Δ16-17 (n=3). (F) 3′ hairpin substrate Δ12-17 (n=3). (G) 3′ hairpin substrate Δ1,12-17 (n=3). (H) 3′ hairpin substrate Δ1-2,12-17 (n=3). (I) 3′ hairpin substrate 10CC-GG (n=3). (J) 3′ hairpin substrate 7CU-GA (n=3). (K) 3′ hairpin substrate G7 insert (n=3). (L) 5′ hairpin substrate wild type (n=3). (M) 5′ hairpin substrate YRA1 (n=3).

Table 4. Activity and affinity of snR34 H/ACA snoRNP for short substrate variants. Activity was determined by tritium release assay (Figure 11, Figure 12) and the initial velocity was estimated by linear regression (Materials and Methods). Dissociation constants were determined by nitrocellulose filtration (Figure 13).

Substrate RNA	**Estimated Initial Velocity (nM min^{-1})**	**K_D (nM)**
3′ hairpin substrate WT	26 ± 5	27 ± 9 (this work), 100 ± 30 (Caton *et al.*, 2018)
3′ hairpin substrate Δ1-2	2.0 ± 0.1	26 ± 16
3′ hairpin substrate Δ1-4	0.20 ± 0.02	18 ± 15
3′ hairpin substrate Δ12-17	6.5 ± 0.4	48 ± 11
3′ hairpin substrate Δ14-17	12.3 ± 0.5	22 ± 17
3′ hairpin substrate Δ16-17	30 ± 10	63 ±28
3′ hairpin substrate Δ1,12-17	4.2 ± 0.5	24 ± 10
3′ hairpin substrate Δ1-2,12-17	0.43 ± 0.01	167 ±140
3′ hairpin substrate 10CC-GG	0.055 ± 0.03	266 ± 26
3′ hairpin substrate 7CU-GA	0.90 ± 0.2	76 ±43
3′ hairpin substrate G7 insert	0.47 ± 0.03	51 ± 34
5′ hairpin substrate WT	11 ± 0.5	222 ± 81
5′ hairpin substrate YRA1	ND	132 ±28

3.4 – The H/ACA snoRNP can Dissociate Rapidly from Non-Modifiable Sequences

Since the H/ACA snoRNP could bind certain substrate RNA sequences with nanomolar affinities but not modify these RNAs, we wondered if the H/ACA snoRNP could select a modifiable sequence while in competition with a non-modifiable sequence. To answer this question, a competitive tritium release assay between the non-modifiable 3′ substrate G7 insert sequence and the 3′ substrate wild-type sequence was performed. Non-radiolabelled G7 insert was pre-bound to the H/ACA snoRNP before addition of a tritiated 3′ substrate wild-type. There is no difference in the rate of pseudouridylation of the 3′ substrate WT in the presence or absence of the competitive G7 insert sequence (Figure 14).

This indicates that binding of the G7 insert sequence does not competitively inhibit the binding and modification of the wild-type substrate.

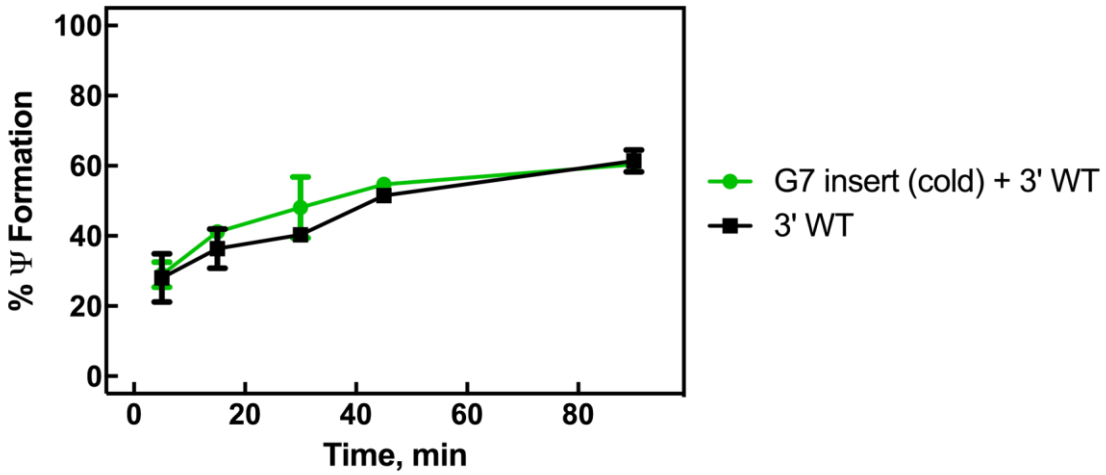

Figure 14. Competitive *in vitro* pseudouridylation of 3′ substrate wild-type. An excess (250 nM) of non-radioactive 3′ substrate G7 insert was incubated with 50 nM reconstituted snR34 H/ACA snoRNP for 3 min before 250 nM [^{3}H-C5] uridine-labeled 3′ substrate wild-type was introduced to the reaction (green circles). As control, an excess (250 nM) of 250 nM [^{3}H-C5] uridine-labeled 3′ substrate wild-type was incubated with 50 nM reconstituted snR34 H/ACA snoRNP in absence of 3'substrate G7 insert (black squares). Mean and standard deviation of duplicate reactions are shown.

3.5 – Modification of structured RNAs by H/ACA snoRNPs

As the substrates used in this study so far have been short oligomers that are unlikely to form secondary structure, longer substrate RNAs comprising RNA helices were designed based on the sequence and secondary structure of the 25S rRNA (Figure 15A). These long substrates include Helix 89 (H89) of the 25S rRNA where U2826 is targeted by the 5′ hairpin of snR34, and Helices 90-92 of the 25S rRNA where U2880 is targeted by the 3′ hairpin of snR34. The long substrate RNAs H89 and H90-92 were *in vitro* transcribed and purified by gel extraction. H89 and H90-92 are expected to be 61 and 79 nt, respectively (Figure 15B). These RNAs also migrate according to their relative sizes.

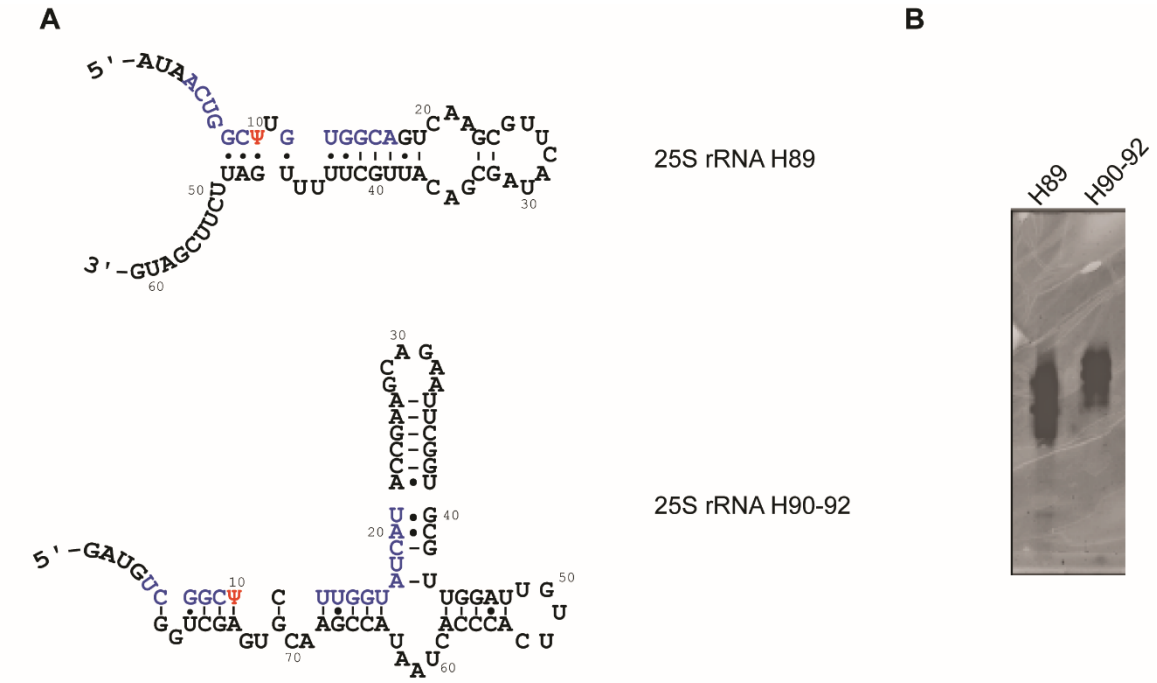

Figure 15. Schematic representation of the sequence and structure of the 25S rRNA fragments used as long substrates. (A) H89 is targeted by the 5′ hairpin of snR34 and H90-92 is targeted by the 3′ hairpin of snR34. Locations of pseudouridines are indicated in red. The region that base pairs with the pseudouridylation pockets of snR34 are indicated in blue. RNA sequence and structure was obtained from the 3D Ribosomal Modification Maps Database (Piekna-Przybylska *et al.*, 2008). (B) 10% Urea PAGE analysis of the tritium labelled long substrate RNAs, 25S rRNA fragments H89 and H90-92, stained with SYBR Green and imaged with a Typhoon imager (GE Healthcare).

Radiolabelled substrates were used in tritium release assays to determine if structure around the target uridine affects the ability of the H/ACA snoRNP to modify the substrate. The long substrates were either allowed to refold by slowly cooling from 80°C (folded) or they were heated and then snap cooled on ice to prevent refolding before they were introduced into the tritium release reaction. Uridine 2826 in H89 of the 25S rRNA is modified by the 5′ pseudouridylation pocket of snR34. The pseudouridylation time courses show that it is modified in both the folded and unfolded state (Figure 16 A & B). Uridine 2880 in H90-92 of the 25S rRNA is targeted by the 3′ pseudouridylation pocket of snR34. It is modified in the folded and unfolded state (Figure 16 C & D). For both long substrate

RNAs, the folded and unfolded states display similar pseudouridylation activity. Compared to the short 5′ substrate, both the folded and unfolded H89 substrate RNAs display increased pseudouridylation activity. In contrast, the H90-92 substrate RNA shows slightly reduced pseudouridylation activity compared to the short 3′ substrate. Similarly, the H89 substrate which is modified by the 5′ hairpin of snR34 appears to be modified at a higher rate than the H90-92 substrate which is targeted by the 3′ hairpin. The unfolded substrates likely also fold during the reaction process over several hours because the reactions are performed at 30°C.

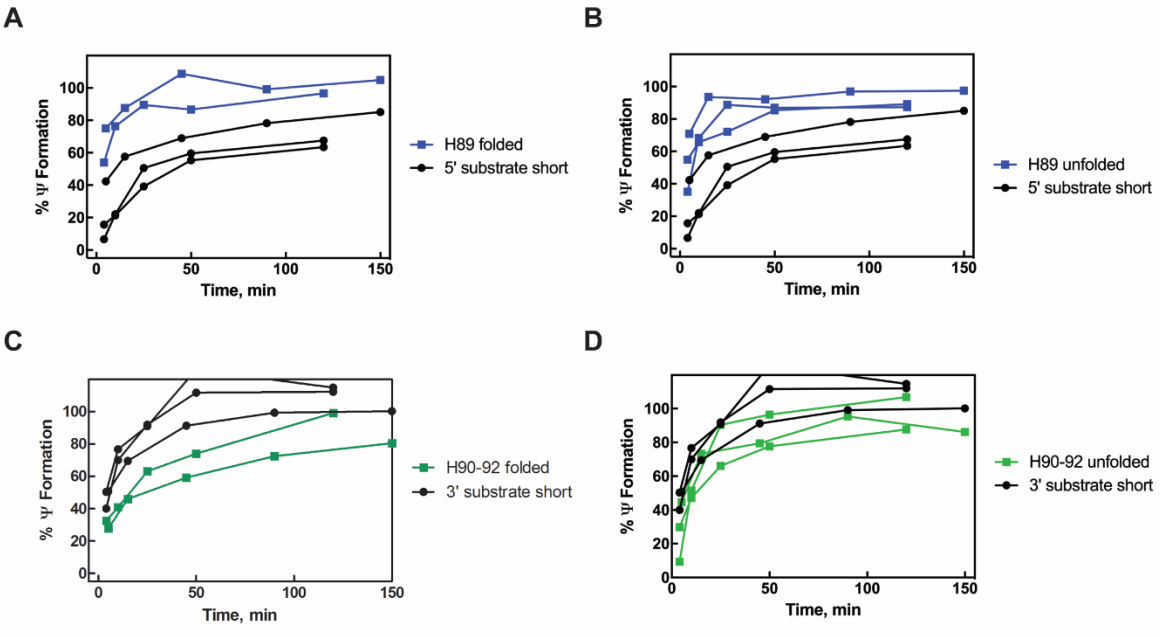

Figure 16. *In vitro* pseudouridylation of long substrate RNAs based on H89 and H90-92 of the 25S rRNA by the snR34 H/ACA snoRNP. An excess (500 nM) of each [^{3}H-C5] uridine-labeled substrate RNA was incubated with 50 nM reconstituted snR34 H/ACA snoRNP. At the given time points, pseudouridylation was quantified by tritium release assay. Three replicates of each substrate were performed, though one replicate was measured at different time points (each shown as a separate connecting line). Each long substrate RNA was either refolded (folded) or heated and snap cooled to prevent refolding (unfolded) before each assay. (A) Pseudouridylation of folded H89 (blue squares) by the 5′ pseudouridylation pocket of the snR34 H/ACA snoRNP, compared to pseudouridylation of the short 5′ hairpin substrate (black circles). (B) Pseudouridylation of unfolded H89 (blue squares) by the 5′ pseudouridylation pocket of the snR34 H/ACA snoRNP, compared to

pseudouridylation of the short 5′ hairpin substrate. (C) Pseudouridylation of folded H90-92 (green squares) by the 3′ pseudouridylation pocket of the snR34 H/ACA snoRNP, compared to pseudouridylation of the short 3′ hairpin substrate (black circles). (D) Pseudouridylation of unfolded H90-92 (green squares) by the 3′ pseudouridylation pocket of the snR34 H/ACA snoRNP, compared to pseudouridylation of the short 3′ hairpin substrate (black circles).

Chapter 4 – Discussion

Characterizing the guide RNA and substrate RNA specificity of the H/ACA snoRNP provides critical information towards developing a more comprehensive understanding of the biological roles of H/ACA snoRNPs. I have shown that the Cbf5-Nop10-Gar1-Nhp2 complex binds guide RNAs non-specifically, though with exceptionally high, sub-nanomolar affinity. Next, I showed that substitutions in the substrate-guide base pairing interaction can be tolerated to allow substrate binding; however, not all substrate variants that are bound are also pseudouridylated. Following this, I demonstrated that structured substrates can be pseudouridylated, despite having the target uridine encompassed by secondary structure.

4.1 - Yeast H/ACA Proteins Bind Guide RNA with High Affinity, but are Non-Specific

The two interesting pieces of information obtained from the guide RNA variant study are the high affinity of the H/ACA proteins for guide RNA, and their lack of specificity for RNA binding. No change in protein affinity for RNA was observed when features of the 5′ hairpin of snR34 were altered, indicating that no single structural feature of an H/ACA RNA tested is important for recognition of a guide RNA. The H/ACA boxes are also not specifically required for guide RNA binding, even though they are required for pseudouridylation activity (Caton *et al.*, 2018). Tight binding of the guide RNA by the H/ACA proteins means these complexes are highly stable *in vivo* and explains why H/ACA RNAs are not exchanged between RNPs, requiring *de novo* synthesis for every RNP (Wang & Meier, 2004). Binding of the Cbf5-Nop10-Gar1-Nhp2 complex to the CrPV IRES revealed that these proteins have a high affinity for structured, non-H/ACA RNA. It is therefore possible that free Cbf5 could bind any structured cellular RNA. Structural studies

of H/ACA snoRNPs have revealed that the contacts formed between Cbf5 and the H/ACA RNA are mostly non-specific. Interactions typically occur between amino acids and the sugar-phosphate backbone of the RNA (Liang *et al.*, 2009a). Additionally, some Cbf5 residues interact with the backbone of the substrate RNA involved in the helix of the 3′ side of the pseudouridylation pocket. This explains why guide RNA selection is not sequence specific and that Cbf5 could interact with any RNA backbone. Finally, the RNA binding face of Cbf5 is extended by Nop10 and Nhp2, increasing the collective ability of the H/ACA proteins to bind tightly to the guide RNA (Hamma *et al.*, 2005; Li & Ye, 2006; Li *et al.*, 2011b; Liang *et al.*, 2009a). It would be interesting to investigate the affinity of the H/ACA proteins for a single stranded RNA. Short RNA substrates could be used in an assay with the H/ACA proteins in the absence of guide RNA, as these small RNAs should not form strong secondary structure. We could then determine if the H/ACA proteins could bind unstructured RNA, or if a double helix is required for the high affinity.

Furthermore, the tight binding of H/ACA guide RNA highlights the importance of the role of Shq1 in the assembly of H/ACA snoRNPs. In the cytoplasm, Shq1 functions as an RNA mimic and shields the Cbf5 RNA binding surface before it is exchanged for an H/ACA RNA in the nucleus, preventing the non-specific binding of other RNAs (Godin *et al.*, 2009; Li *et al.*, 2011a; Machado-Pinilla *et al.*, 2011; Singh *et al.*, 2015). At the site of H/ACA RNA transcription, AAA+ ATPases assist in the dissociation of Shq1 and its replacement with the H/ACA RNA (Machado-Pinilla *et al.*, 2012). Accordingly, my data explain why these assembly mechanisms are required in the biogenesis of H/ACA snoRNPs. To explore this idea further, mutations in the H/ACA snoRNP assembly factors can be considered. For example, mutations in the chaperone Shq1 occur in the regions

responsible for binding Cbf5, which could reduce the efficacy of shielding this site from binding other RNAs (Bizarro & Meier, 2017; Li *et al.*, 2011a; Singh *et al.*, 2015). These mutations were identified in a patient presenting with intrauterine growth retardation and neurological symptoms reminiscent of the severe variant of Dyskeratosis congenita – Hoyeraal-Hreidarsson syndrome (Bizarro & Meier, 2017). In addition to improper assembly of the H/ACA snoRNP required for pseudouridylation in ribosome biogenesis or stabilizing the telomerase RNA component, perhaps effects of Cbf5 indiscriminately binding non-H/ACA RNAs can also contribute to pathology in patients with SHQ1 mutations.

The pseudouridylation activity assays performed with the guide RNA variants were generally consistent with the current understanding of the molecular mechanism of H/ACA snoRNPs as explained in the following. Removal of the 5′ extension did not affect pseudouridylation, which is expected because structural studies have demonstrated that the interaction between the H/ACA proteins and guide RNA occur primarily with the stems of the guide RNA and the H/ACA boxes on the 3′ sides of the stems (Li & Ye, 2006; Li *et al.*, 2011b; Zhou *et al.*, 2011). Removing the pseudouridylation pocket resulted in no pseudouridylation activity which can be explained by the lack of substrate binding site. In the small pseudouridylation pocket guide RNA, only the region known to base pair with a substrate RNA remained single-stranded, and this guide retained pseudouridylation activity. This finding indicates that extra single-stranded RNA in the pseudouridylation pocket is not required for substrate binding and modification. However, this extra single stranded region is still available for base pairing in the wild type guide and may be used for targeting different pseudouridylation sites that are yet to be confirmed.

It was unexpected that the guide RNAs without lower stem and with an extended lower stem displayed enhanced pseudouridylation activity in the H/ACA snoRNP. These guide RNAs were designed to investigate if the H/ACA proteins had any specificity for the lower stem of the H/ACA RNA. The conserved distance between the H/ACA box and the target uridine of 14-16 nt should be disrupted in these guide RNAs which was expected to influence pseudouridylation, if not guide RNA binding. However, upon reinvestigation of how these RNAs may form secondary structure using the folding program mfold (Zuker, 2003), we noticed that the no-lower-stem guide RNA may in fact be forming an alternative lower stem. The mutations introduced into the original lower stem, intended to disrupt the expected base pairing, can also base pair with a region in the 5′ extension, resulting in a longer lower stem and an alternative conformation of the pseudouridylation pocket (Figure 17B). In the case of the extended-lower-stem guide RNA variant, the increased distance to the H box in guide RNA should have misaligned the substrate RNA with Cbf5 such that it would not be positioned in the pseudouridylation pocket. However, there is a second sequence element near the 3′ end of this guide variants that could serve as an alternative H box if unpaired, and in doing so restore the conserved distance from the H box to the target uridine. A longer lower stem can be formed in this guide RNA but might be partially melted (Figure 17C). These lower stem structures may in fact be stabilizing to the H/ACA snoRNP and promote activity. A previous study showed that the stability of the upper stem of the H/ACA RNA directly affects the pseudouridylation ability of the H/ACA RNP (Xiao *et al.*, 2009). While the previous study focused on the upper stem, the lower stem may exhibit similar characteristics. It has also been demonstrated that a stable lower stem in the H/ACA RNA contributed to accumulation of the RNAs and contributed to pseudouridylation activity (Balakin *et al.*, 1996; Bortolin *et al.*, 1999). To see effects from guide RNA variants

that are completely missing the lower stem or contain an extended lower stem without alternative H box, guide variants combining the features of the No 5′ extension and the Now lower stem and extended lower stem RNAs were produced (No 5′ extension-No lower stem and No 5′ extension-Extended lower stem, Figure 5 and Figure 8). These RNA variants should eliminate the alternative structures formed by the No lower stem and Extended lower stem variants. Indeed, the pseudouridylation activity of these guide RNAs is much more comparable to the 5′ hairpin wild-type guide RNA (Figure 8). In the case of the No 5′ extension-No lower stem variant, it does not appear that loss of formation of the lower stem hinders pseudouridylation compared to the wild-type guide RNA. Similarly, extending the lower stem beyond the conserved 14-16 nt distance from the H box does not eliminate pseudouridylation. This may be explained by a certain degree of flexibility between the Cbf5 catalytic and PUA domains that could maybe accommodate a longer distance between the H box and the target uridine to align the pseudouridylation pocket properly against the catalytic domain (Hamma *et al.*, 2005).

The lack of specificity of the H/ACA proteins for guide RNA may provide an advantage in site selection for modification. There is diversity in the sequences and structures of H/ACA RNAs with the only conserved features being the H/ACA boxes and the distance of 14-16 nt between the H/ACA boxes and the target uridine. The proteins' non-specific binding allows the snoRNP to be modular and accommodate various guide

RNA sequences and structures. This feature provides the H/ACA snoRNP the flexibility to bind and modify many target sites in various cellular RNAs.

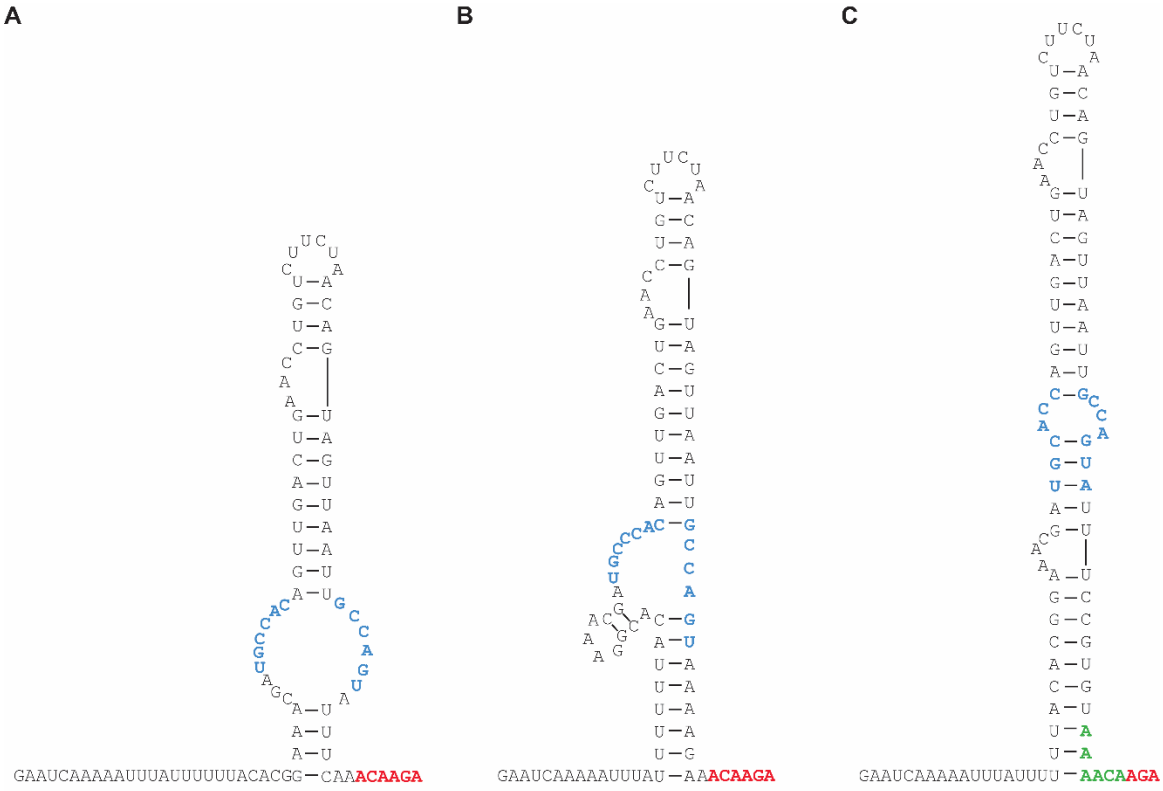

Figure 17. Secondary structure predictions of the 5′ hairpin of snR34. (A) snR34 5′ hairpin wild-type**.** (B) snR34 5′ hairpin no-lower-stem variant. (C) snR34 5′ hairpin extended-lower-stem variant. Blue letters indicate the portion of the pseudouridylation pocket that base pairs to a substrate RNA. Red letters indicate the H box (ACAAGA). Green letters in the extended-lower-stem variant indicates the alternative H box (AAAACA). Structures B and C were predicted by mfold (Zuker, 2003).

4.2 - Positioning of the Target Uridine in the Cbf5 Active Site Requires Sufficient Base Pairing on Both Sides of the Target Uridine

Substitutions in the substrate RNA that result in fewer base pairs in the 3′ side of the pseudouridylation pocket of snR34 (Δ1-2 and Δ1-4 substrates, Figure 11A) are detrimental to pseudouridylation activity. The Δ1-2 substrate makes 3 base pairs in the 3′ side of the pseudouridylation pocket and the Δ1-4 substrate makes one base pair with the 3′ side, compared to the 5 base pairs made by the wild type substrate. Each of these substrates make 10 base pairs (one G-A pair) with the 5′ side of the pseudouridylation pocket. These substitutions appear to be more detrimental to pseudouridylation of the substrate than substitutions that reduce base pairs in the 5′ side of the pseudouridylation pocket of snR34 (Δ16-17, Δ14-17 and Δ12-17 substrates, Figure 11B, Table 4). Each of the substrates described in Figure 11B makes 5 base pairs with the 3′ side of pseudouridylation pocket. In the 5′ side of the pseudouridylation pocket of snR34, the Δ16-17 substrate makes 8 base pairs, the Δ14-17 substrate makes 6 base pairs and the Δ12-17 substrate makes 4 base pairs. Though the Δ14-17 substrate reached a lower end level than Δ1-2 substrate, the rate at which the end level was reached is faster than for the Δ1-2 substrate. The Δ14-17 substrate RNA sample may contain a population of RNA that cannot be modified for unknown reasons, resulting in an end level of ~50% pseudouridylation. The next substrate set analyzed had substitutions that resulted in reduced base pairing in both sides of the pseudouridylation pocket of snR34 (Figure 11C, Table 4). The Δ1,12-17 substrate makes 4 base pairs with both the 5′ and 3′ sides of the pseudouridylation pocket (8 base pairs total). The Δ1-2,12-17 makes 4 base pairs with the 5′ side and 3 base pairs with the 3′ side of the pseudouridylation pocket (7 base pairs total). The reduced base

pairing of substrates with substitutions in both the 5′ and 3′ sides of the pseudouridylation pocket causes a combined effect on the rate of pseudouridylation, influenced by substitutions affecting pairing in both sides of the pocket. Since both the Δ1-2 and the Δ12-17 substrate RNAs were modified slowly compared to the wild type substrate, it makes sense that the rate of the Δ1-2,12-17 substrate modification is even further depressed compared to the substrates with substitutions on individual sides of the pseudouridylation pocket. A similar effect occurs with the Δ1,12-17 substrate, though not to the same degree. Finally, substrates causing mismatches in the guide-substrate helix were analyzed (Figure 11D, Table 4). The 10CC-GG substrate makes 5 base pairs with the 3′ side of the pseudouridylation pocket of snR34 and 8 base pairs with the 5′ side, but these are not consecutive base pairs. The 10CC-GG substitution introduces a bulge in the helix between the substrate RNA and the pseudouridylation pocket, which may disrupt the positioning of the target uridine in the active site of Cbf5. It is also possible that the two A-U pairs directly adjacent to these substitutions are not forming, resulting in an inability of the target uridine to be placed in the Cbf5 active site. The 7CU-GA substrate makes 5 base pairs with the 3′ side of the pseudouridylation pocket and 9 base pair with the 5′ side of the pocket and changes the identity of the nucleotides directly 3′ of the target uridine. Though one of these nucleotides is unpaired, the introduction of a purine-purine mismatch in position 8 may cause displacement of the target uridine and inhibit pseudouridylation. Lastly, introduction of an additional unpaired nucleotide adjacent to the target uridine (G7 insert, Figure 11D, Table 4) may also displace the uridine from the Cbf5 active site, impeding modification despite complete wild type base pairing with the pseudouridylation pocket.

Somewhat surprisingly, the lack of base pairing between most substrate RNAs and the snR34 guide RNA did not influence the affinity of the reconstituted H/ACA snoRNP for the substrates. All substrates were bound with nanomolar affinities. The Δ1-2,12-17 and 10CC-GG substrates may represent affinity thresholds where the base pairing between the guide and substrate RNAs may be sufficiently reduced as to influence the affinity of the H/ACA snoRNP for the substrate, given that the K_Ds for these substrates are high in comparison to the other 3′ hairpin substrates. The Δ1-2,12-17 substrate makes the fewest base pairs with the pseudouridylation pocket of all the substrate variants tested, engaging in only 7 base pairs. The shortest guide-substrate interaction known in yeast is between the 3′ hairpin of snR191 and its substrate in the 25S rRNA, modifying U2260 (Badis *et al.*, 2003). There are 8 base pairs formed between the snR191 3′ hairpin and the sequence surrounding U2260, with 4 base pairs on either side of the target uridine. The Δ1,12-17 substrate represents an example of sequence that can form 8 base pairs with the pseudouridylation pocket, mimicking the features of the 3′ hairpin of snR191. The data presented here indicate that an 8-base pair match is required to bind and position the substrate RNA in such a way that it can be pseudouridylated with reasonable efficiency. It would be interesting to test the snR191 3′ hairpin and substrate pair in this reconstituted system to compare results. In the case of the 10CC-GG substrate, the location of the mismatches in the helix and the fact that two purines are replacing two pyrimidines may be sterically destabilizing enough to increase the K_D.

The fact that all the tested substrate RNAs were bound to the H/ACA snoRNP, but not all were effectively modified, leads us to ask the question of how the helices formed between the guide and substrate RNA influence the exact position of the target uridine

relative to the active site of Cbf5. When the substrate RNA base pairs with the pseudouridylation pocket, two new helices composed of guide and substrate RNA strands form. The substrate RNA adopts a severely bent Ω-structure to accommodate these helices (Figure 3) (Jin *et al.*, 2007). Additional stabilization of the complex is derived from coaxial stacking that occurs between the upper stem of the guide RNA and the helix formed between the 5′ side of the pseudouridylation pocket and the 3′ side of the substrate RNA, and the coaxial stack between the lower stem of the guide RNA and the helix formed between the 3′ side of the pocket and the 5′ side of the substrate RNA (Wu & Feigon, 2007). These stacking interactions and the Ω conformation help preserve the distance from the H and ACA boxes to the target uridine and therefore directly contribute to positioning of the target uridine in the Cbf5 active site. Disruption of these stacking interactions through reduced base pairing, such as in the case of some substrate variants, could result in improper target uridine placement and inhibit modification. It appears that the mismatches in the 3′ side of the pseudouridylation pocket have a greater effect on activity than mismatches in the 5′ side. This may be due to the difference in length of the base pairing regions on either side of the pseudouridylation pocket. The shorter base pairing region on the 3′ side of the pseudouridylation pocket makes it easier to disrupt with substitutions in the substrate RNA. General inspection of known H/ACA guide – substrate interactions reveals that the helix in the 3′ side of the pseudouridylation pocket is typically shorter than that on the 5′ side of the pocket, though there are some exceptions (Piekna-Przybylska *et al.*, 2008). It may be interesting to also test an example H/ACA RNA and substrate pair that has a shorter helix in the 5′ side of the pseudouridylation pocket than the 3′ side, such as the 3′ hairpin of snR189 which target U2735 in the 25S rRNA (Piekna-Przybylska *et al.*, 2008; Torchet *et al.*, 2005). This guide-substrate pair makes 4 base pairs on the 5′ side of the pocket and 7

base pairs on the 3′ side of the pocket. Examination of this system will differentiate if the length of the helices on either side of the target uridine have similar contributions to the positioning of the target uridine, or if the helix in the 3′ side of the pocket is more important, especially given its proximity to the H/ACA boxes.

It is also interesting to note that the guide and substrate RNA pairs used in the NMR structure studies did not leave unpaired nucleotides in the 3′ side of the pseudouridylation pocket, though this often occurs between guide and substrate RNAs found in nature (Jin *et al.*, 2007; Piekna-Przybylska *et al.*, 2008; Wu & Feigon, 2007). In the situation where no unpaired nucleotide of the pseudouridylation pocket is left in the 3′ side of the pocket, the helix formed in the 3′ side of the pocket can directly stack on the lower stem of the guide RNA. In the case of the 3′ hairpin of snR34 and the 25S rRNA substrate it targets, there are 5 unpaired nucleotides in the single stranded region of the pseudouridylation pocket that do not base pair with the substrate. This may prevent stacking of the 3′ pseudouridylation pocket-substrate helix with the lower stem of the guide RNA. However, in most natural cases, the 5′ side of the pseudouridylation pocket is paired completely with a substrate sequence in the region closest to the upper stem of the guide RNA, thus allowing stacking between the upper stem and the helix in the 5′ side of the pseudouridylation pocket. Therefore, the reason for the reduction of pseudouridylation activity in the cases of substrates with substitutions in the 3′ side of the pseudouridylation pocket may have less to do with base stacking. Another structural study revealed that interactions between Cbf5 and the pseudouridylation pocket occur at the 3′ side of the pseudouridylation pocket, and that amino acids contact both the guide and substrate RNA backbones (Duan *et al.*, 2009; Liang *et al.*, 2009a). It is possible that destabilizing the helix in the 3′ side of the pseudouridylation

pocket disrupts these contacts with Cbf5 and ultimately results in improper placement of the target uridine in the Cbf5 active site.

Recently, another study investigating the base-pairing requirement between H/ACA snoRNPs and their substrates was published. This study focused on the 5′ hairpin of snR81 which targets position 42 of the U2 snRNA in yeast. By mutating the sequence of the guide RNA to introduce mismatches with the substrate within the pseudouridylation pocket and detecting pseudouridylation by CMC-labelling of RNA extracted from yeast cells, the study showed that a minimum of 8 base pairs between the guide and substrate RNA was required to generate detectable levels of pseudouridylation (De Zoysa *et al.*, 2018). This phenomenon was confirmed in two other pseudouridylation pockets and is consistent with the results of this thesis. However, they methods used by Zoysa *et al* (2018) do not provide information on the rate of pseudouridylation since CMC-labelling followed by primer extension only reveals relative final levels of pseudouridylation. The data generated by tritium release assays in this thesis describe how mismatches within the pseudouridylation pocket affect the rate of modification, which may have physiological relevance. Notably, the Δ1,12-17 substrate RNA (that forms 8 base pairs with snR34) is modified 8-times more slowly than the wild type substrate RNA.

De Zoysa *et al* (2018) also made substrate RNAs that have more than one unpaired nucleotide adjacent to the target uridine. The Yu group specifically added more unpaired uridines and showed that more unpaired nucleotides could be accommodated by the snoRNP, and additionally, the location of the pseudouridine shifted to the 5′-most unpaired uridine in the pocket. This is inconsistent with the results obtained from tritium release assays using a substrate containing an extra unpaired nucleotide in this thesis (G7 insert,

Figure 11D). In this case, the introduction of an unpaired guanosine abolished pseudouridylation. The discrepancies seen here may stem from differences in the guide RNAs themselves, and different pseudouridylation pockets may be able to tolerate extra unpaired nucleotides or not.

Additionally, De Zoysa *et al* (2018) performed assays in which mismatches were introduced either immediately 5′ or 3′ to the unpaired dinucleotide. In the case where the remainder of the naturally occurring base pairs were left intact, these pairs appeared to play no significant role in pseudouridylation. They only became essential when the minimum number of base pairs between the substrate and the guide RNA was reached when the additional removal of base pairs immediately 5′ or 3′ of the unpaired dinucleotides resulted in abolished pseudouridylation. The tritium release assay of the snR34 3′ substrate 7CU-GA contradicts this finding (Figure 11D). In the presence of all other naturally occurring base pairs, disrupting the base pair immediately 3′ of the unpaired dinucleotide renders the substrate unmodifiable. This may indicate that there is a difference in mismatch tolerance between guide RNAs.

Finally, De Zoysa *et al* (2018) commented on an attempt to calculate a ΔG of binding based on the base-pairing between the substrate and pseudouridylation pocket. However, this was deemed incalculable as most standard algorithms are only suitable for standard Watson-Crick base pairing and cannot accommodate the base pairing patterns and torsion of the substrate RNA in the case of H/ACA snoRNP-substrate binding in the Ω conformation. However, by determining the dissociation constants of binding between the snoRNP and substrate obtained in this thesis, it may be possible to calculate ΔGs of binding for the substrates tested in this thesis using the following equation:

$$\Delta G = -RTln(K_D) \quad (3)$$

Therefore, the filter binding data described here may serve as a useful starting point for developing an appropriate algorithm which may then be used in pipelines to predict novel pseudouridylation target sites.

Analysis of the putative YRA1 mRNA substrate of the 5′ hairpin of snR34 revealed that it is modified very slowly compared the 5′ wild type substrate sequence (Figure 12). This interaction was predicted after a pseudouridine sequencing study revealed that pseudouridylation of the YRA1 mRNA was reduced in cells upon knock down of Cbf5 (Schwartz *et al.*, 2014). It is possible that sites in this mRNA are modified by H/ACA snoRNPs. In the predicted interaction, the 5′ hairpin of snR34 would have to undergo structural rearrangements to accommodate base pairing with the YRA1 mRNA to situate the target uridine (U362) in the active site of Cbf5. The required rearrangement of the 5′ hairpin may limit the modification of the YRA1 mRNA substrate; however, a previous study has demonstrated that alternative guide RNA structures may exist to accommodate modification of different sites (Xiao *et al.*, 2009). Despite the slow modification of the YRA1 mRNA, it is still bound by the H/ACA snoRNP at a similar K_D to the wild type 5′ hairpin substrate, which supports the notion that the 5′ hairpin can bind the substrate either in its canonical structure or after rearrangement to the alternative conformation. In the canonical conformation, base-pairing interactions close to the target uridine would be missing which would likely prevent placement of the target U in the active site. However, even in the alternative conformation, the location of the target uridine in the pseudouridylation pocket next to two unpaired nucleotides 5′ to the target uridine probably also impedes proper placement of the uridine in the Cbf5 active site. Similarly, in the case

of the G7 insert substrate in the 3′ hairpin of snR34, an extra unpaired nucleotide adjacent to the target uridine severely decreases the rate of pseudouridylation (Figure 11D). In the case of the YRA1 mRNA, the 5′-most unpaired uridine in the pseudouridylation pocket (U360) may be slowly pseudouridylated, as this uridine also contains a tritium label and therefore would be detected in our assay. To confirm which uridine is modified by a tritium release assay, a mutant YRA1 mRNA substrate would have to be produced where the U360 is substituted with another nucleotide.

The ability of H/ACA snoRNPs to bind substrate sequences, but not modify them is likely of significance *in vivo*. We speculate that H/ACA snoRNPs also transiently bind non-modifiable sites *in vivo*. This could implicate the complex in a possible regulatory role – either blocking access of certain sites from other enzymes or regulatory factors or assisting in RNA folding or processing. We can also then ask to what extent other factors, such as helicases like Rok1 and Has1, would be required to moderate these kinds of interactions (Liang & Fournier, 2006; Martin *et al.*, 2014). It has been confirmed that *in vitro* substrate turnover is possible, as demonstrated by the multiple turnover tritium release assays. This is mediated by the detection of the conversion of uridine to pseudouridine by Gar1 and its subsequent influence on the thumb loop of Cbf5 (Wang *et al.*, 2015). It would be interesting to determine how slow or fast the complex could dissociate from an unmodifiable target. To investigate this, a competitive tritium release assay was performed (Figure 14). In this assay, an unlabeled unmodifiable substrate RNA was pre-bound to a reconstituted H/ACA snoRNP. Subsequently, tritium-labelled wild type substrate was introduced. Tritium release was detected at similar rates in the presence and absence of unmodifiable substrate RNA. This indicates that the snoRNP can release unmodified

substrate quickly, and subsequently bind and modify a new substrate. This finding further reveals that sequences can be released by the snoRNP without dependence on pseudouridylation of that sequence, and that is it possible to dissociate non-modifiable sequences without the assistance of helicases. It would be interesting to repeat a competitive tritium release assay in the presence of helicases to see if the rate of pseudouridylation of a wild type substrate could be enhanced. Finally, to complement the competitive tritium release assay, a filter binding chase experiment could also be performed to determine the rate at which substrate turnover can occur with an unmodifiable variant. In this case, the unmodifiable substrate would be radioactively labelled, and unlabeled wild type substrate would be introduced in excess to compete for binding in the pseudouridylation pocket.

4.3 - H/ACA snoRNPs can modify uridines encompassed by structure

The tritium release assays using the long 25S rRNA fragments as substrates revealed that secondary structure surrounding a target uridine does not impede the H/ACA snoRNP from modifying the target site (Figure 16). The H89 substrate RNA is modified at a faster rate than the H90-92 substrate, targeted by the 5′ hairpin and 3′ hairpin of snR34, respectively. Examination of the secondary structures can provide an explanation for this phenomenon (Figure 15). The target uridine in H89 is near the base of a helix and a single-stranded region of which 4 nucleotides base pair with the 5′ side of the pseudouridylation pocket. These single-stranded nucleotides can provide a nucleation point for the 5′ side of the snR34 5′ hairpin pseudouridylation pocket to base pair and initiate a strand displacement event to unwind the H89 helix and allow base pairing of the rRNA fragment to the rest of the pseudouridylation pocket. Conversely, the H90-92 rRNA fragment only

has one single-stranded nucleotide that would base pair with the 5′ side of the pseudouridylation pocket in the snR34 3′ hairpin. One nucleotide would be less effective at initiating strand displacement in this case, which may explain the slower rate of modification of H90-92.

The timing of pseudouridylation during ribosome biogenesis has been difficult to elucidate, and partial modification (<85%) of some sites has also been observed (Sloan *et al.*, 2017). 2′O-methylation by C/D snoRNAs has been shown to occur in both early and late stages of ribosome biogenesis, and these modifications can be directed co-transcriptionally (Kos & Tollervey, 2010; Sloan *et al.*, 2017). Though similar behaviour has not been confirmed for H/ACA snoRNPs, the ability of these complexes to pseudouridylate a structured substrate *in vitro* implies that H/ACA snoRNPs could modify pre-rRNA at various stages of biogenesis. If the complex can overcome structural constraints to modify uridines encompassed by secondary structure, then it is possible that these complexes have an additional function to pseudouridylation. The disruption of the structure of the substrate RNA suggests that the H/ACA snoRNP can remodel the structure of the pre-rRNA. This ability might even be enhanced and controlled by the bipartite nature of the H/ACA snoRNP, and the sequence targeted by each hairpin could have some significance in determining the location and final folding state of that sequence in the mature rRNA.

Nitrocellulose filtration experiments can also be performed with the long substrate RNAs to determine the effect structure has on the affinity of the H/ACA snoRNP for the substrate sequence. Given that these substrates can be modified, it is already implied that they are bound by the snoRNP. It will be interesting to compare the affinity of long

substrates to those of the short substrates, to see the influence of structure as this may be more comparable to substrate binding *in vivo*. It may also be interesting to investigate product release of long substrate RNAs. Helicases are required for the release of H/ACA snoRNPs from pre-rRNA *in vivo* and are involved in the remodelling of pre-rRNA (Liang & Fournier, 2006; Martin *et al.*, 2014). Perhaps the addition of helicases to *in vitro* experiments will increase modification rates by enabling more efficient product release.

4.4 - A Comparison of Systems: Substrate binding by H/ACA snoRNPs and C/D snoRNPs

C/D snoRNPs are responsible for the RNA-guided 2′O-methylation of nucleotides and have been primarily studied in archaeal systems. Like H/ACA snoRNPs, C/D snoRNPs engage in a base-pairing interaction with a substrate RNA sequence to select a target nucleotide for modification. Whereas a minimum of 8 base pairs is required between yeast H/ACA guide RNA and a substrate sequence to attain a reasonable rate of modification, it has been shown that C/D guide RNA from *Solfolobus solfataricus* forms a maximum 10-base pair duplex with a substrate sequence (Yang *et al.*, 2016). While shorter substrates are tolerated by C/D snoRNPs, longer substrates are not accommodated. Like the H/ACA snoRNP, the nature of substrate binding is optimized for spatial positioning of the target nucleotide. In the C/D snoRNP, a substrate, that forms more than 10 base pairs, does not have the target nucleotide properly positioned for methylation by fibrillarin. It is interesting that the archaeal C/D guide RNA spacer sequence that binds to substrate RNAs is typically 12 nucleotides long, but the system can only effectively methylate substrates engaging in 10 base pairs (though there are some exceptions). Interestingly, human and yeast C/D snoRNAs do not exhibit a constrained spacer length (Tran *et al.*, 2005). The

pseudouridylation pocket of the H/ACA snoRNP also typically contains more nucleotides than base pairs formed with a substrate. It would be interesting to see if substrates that can form a longer duplex with the pseudouridylation pocket also affect the rate and degree of pseudouridylation of a substrate.

Unlike the uridine targeted by H/ACA snoRNPs, the nucleotide targeted by C/D snoRNPs is also base paired to the guide RNA. However, the systems are similar in that they require continuous base pairing in close proximity to the target nucleotides to achieve successful modification. A study of substrate binding by the C/D snoRNP demonstrated that only a mismatch with the nucleotide immediately 5′ of the D box was tolerated in an *in vitro* methylation assay, and mismatches at any other location drastically reduced methylation of the substrate (Appel & Maxwell, 2007). Additionally, a similar activity assay was performed in this report using long, structured substrate RNAs, and it was found that C/D snoRNPs can also overcome secondary structure to methylate a target nucleotide. This observation strengthens the idea that both CD/snoRNPs and H/ACA snoRNPs can overcome secondary structure *in vitro* and likely also *in vivo*. This publication also argued that decreased methylation was not necessarily a result of decreased base pairing strength, but as an effect of improper positioning of the target nucleotide. This dependency on conformation is further supported by the substitution of deoxynucleotides in the substrate RNA of the C/D snoRNP, which would disrupt the A-form helix of the guide-substrate duplex. Finally, the relevance of the bipartite nature of the C/D snoRNP to substrate recognition has also been investigated (Hardin & Batey, 2006). It was revealed that the full C/D guide RNA confers specificity of the site of methylation, and that a hemi-RNP (composed of one set of proteins and only one C/D box set) can direct methylation, but at

non-specific sites. In eukaryotes, the H/ACA snoRNP also has a bipartite structure and perhaps the presence of the second hairpin and set of proteins also contributes to substrate binding, positioning and pseudouridylation. More detailed biochemical studied will be required to confirm this.

4.5 - Conclusion

The data presented in this thesis provide insight into the mechanisms by which guide RNAs are recognized and bound by H/ACA proteins, and the mechanisms by which substrate RNAs are recognized and modified or not. The Cbf5-Nop10-Gar1-Nhp2 complex binds H/ACA guide RNAs extremely tightly, but lacks specificity for features of the H/ACA RNA and can bind other structured, non-H/ACA RNAs, such as the Cricket Paralysis Virus IRES. This finding highlights the importance of the proper assembly of H/ACA snoRNPs and the utility of the modular nature of the H/ACA snoRNPs given their ability to form a complex with many different guide RNAs to target multiple sites for pseudouridylation. The substrate RNA studies revealed that positioning of the target uridine in the Cbf5 active site requires sufficient base pairing on both sides of the target uridine for fast and efficient pseudouridine formation. The base pairing between the substrate RNA and the pseudouridylation pocket can contribute to appropriate coaxial stacking between the 5′ pseudouridylation pocket helix and the upper stem of the H/ACA RNA, and the 3′ pseudouridylation pocket helix and the lower stem of the H/ACA RNA. Furthermore, proper formation of a helix in the 3′ side of the pseudouridylation pocket ensures appropriate contacts between Cbf5 amino acids and the guide and substrate backbones. Lastly, I showed that structured RNA substrates can be modified by H/ACA snoRNPs. This indicates that it may be possible for pseudouridylation to occur post-transcriptionally, as it

is possible for H/ACA snoRNPs to initiate unwinding of helices to access the target uridine through strand displacement. This may also allow H/ACA snoRNPs to play a role in remodeling of rRNA.

References

Appel, C. D., & Maxwell, E. S. (2007). Structural features of the guide:target RNA duplex required for archaeal box C/D sRNA-guided nucleotide 2'-O-methylation. *RNA, 13*(6), 899-911.

Arnez, J. G., & 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*(24), 7560-7567.

Badis, G., Fromont-Racine, M., & Jacquier, A. (2003). A snoRNA that guides the two most conserved pseudouridine modifications within rRNA confers a growth advantage in yeast. *RNA, 9*(7), 771-779.

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.

Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., & Yusupov, M. (2011). The structure of the eukaryotic ribosome at 3.0 A resolution. *Science, 334*(6062), 1524-1529.

Berndt, H., Harnisch, C., Rammelt, C., *et al.* (2012). Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. *RNA, 18*(5), 958-972.

Bizarro, J., & Meier, U. T. (2017). Inherited SHQ1 mutations impair interaction with NAP57/dyskerin, a major target in dyskeratosis congenita. *Molecular Genetics & Genomic Medicine, 5*(6), 805-808.

Blackburn, E. H., & Collins, K. (2011). Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harbor Perspectives in Biology, 3*(5).

Boccaletto, P., Machnicka, M. A., Purta, E., *et al.* (2018). MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Research, 46*(Database issue), D303-D307.

Bordeira-Carrico, R., Pego, A. P., Santos, M., & Oliveira, C. (2012). Cancer syndromes and therapy by stop-codon readthrough. *Trends in Molecular Medicine, 18*(11), 667-678.

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. *The EMBO Journal, 18*(2), 457-469.

Bykhovskaya, Y., Casas, K., Mengesha, E., Inbal, A., & Fischel-Ghodsian, N. (2004). Missense Mutation in Pseudouridine Synthase 1 (PUS1) Causes Mitochondrial Myopathy and Sideroblastic Anemia (MLASA). *American Journal of Human Genetics, 74*(6), 1303-1308.

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.

Caton, E. A., Kelly, E. K., Kamalampeta, R., & Kothe, U. (2018). Efficient RNA pseudouridylation by eukaryotic H/ACA ribonucleoproteins requires high affinity binding and correct positioning of guide RNA. *Nucleic Acids Research, 46*(2), 905-916.

Charpentier, B., Muller, S., & Branlant, C. (2005). Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Research, 33*(10), 3133-3144.

Chen, K., Zhao, B. S., & He, C. (2016). Nucleic acid modifications in regulation of gene expression. *Cell Chemical Biology, 23*(1), 74-85.

Cohn, W. E. (1959). 5-Ribosyl uracil, a carbon-carbon ribofuranosyl nucleoside in ribonucleic acids. *Biochimica et Biophysica Acta, 32*, 569-571.

Cohn, W. E. (1960). Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: isolation, structure, and chemical characteristics. *Journal of Biological Chemistry, 235*, 1488-1498.

Darzacq, X., Jady, B. E., Verheggen, C., Kiss, A. M., Bertrand, E., & Kiss, T. (2002). Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *The EMBO Journal, 21*(11), 2746-2756.

Davis, D. R. (1995). Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Research, 23*(24), 5020-5026.

Davis, F. F., & Allen, F. W. (1957). Ribonucleic acids from yeast which contain a fifth nucleotide. *Journal of Biological Chemistry, 227*(2), 907-915.

De Zoysa, M. D., Wu, G., Katz, R., & Yu, Y. T. (2018). Guide-substrate base-pairing requirement for box H/ACA RNA-guided RNA pseudouridylation. *RNA, 24*(8), 1106-1117.

Decatur, W. A., & Fournier, M. J. (2002). rRNA modifications and ribosome function. *TRENDS in Biochemical Sciences, 27*(7), 344-351.

Dokal, I. (2000). Dyskeratosis congenita in all its forms. *British Journal of Haematology, 110*(4), 768-779.

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

Fatica, A., Dlakic, M., & Tollervey, D. (2002). Naf1p is a box H/ACA snoRNP assembly factor. *RNA, 8*(12), 1502-1514.

Fernández, I. S., Ng, C. L., Kelley, A. C., Wu, G., Yu, Y. T., & Ramakrishnan, V. (2013). Unusual base pairing during the decoding of a stop codon by the ribosome. *Nature, 500*(7460), 107-110.

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 & Development, 11*(7), 941-956.

Godin, K. S., Walbott, H., Leulliot, N., van Tilbeurgh, H., & Varani, G. (2009). The box H/ACA snoRNP assembly factor Shq1p is a chaperone protein homologous to Hsp90 cochaperones that binds to the Cbf5p enzyme. *Journal of Molecular Biology, 390*(2), 231-244.

Grozdanov, P. N., Roy, S., Kittur, N., & Meier, U. T. (2009). SHQ1 is required prior to NAF1 for assembly of H/ACA small nucleolar and telomerase RNPs. *RNA, 15*(6), 1188-1197.

Guzzi, N., Ciesla, M., Ngoc, P. C. T., *et al.* (2018). Pseudouridylation of tRNA-Derived Fragments Steers Translational Control in Stem Cells. *Cell, 173*(5), 1204-1216.

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. *Nature Structural & Molecular Biology, 12*(12), 1101-1107.

Hardin, J. W., & Batey, R. T. (2006). The bipartite architecture of the sRNA in an archaeal box C/D complex is a primary determinant of specificity. *Nucleic Acids Research, 34*(18), 5039-5051.

Heiss, N. S., Knight, S. W., Vulliamy, T. J., Klauck, S. M., Wiemann, S., Mason, P. J., Poustka, A., & Dokal, I. (1998). X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nature Genetics, 19*(1), 32-38.

Henras, A., Dez, C., Noaillac-Depeyre, J., Henry, Y., & 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 Research, 29*(13), 2733-2746.

Henras, A. K., Plisson-Chastang, C., O'Donohue, M. F., Chakraborty, A., & Gleizes, P. E. (2015). An overview of pre-ribosomal RNA processing in eukaryotes. *WIREs RNA, 6*(2), 225-242.

Jack, K., Bellodi, C., Landry, D. M., *et al.* (2011). rRNA pseudouridylation defects affect ribosomal ligand binding and translational fidelity from yeast to human cells. *Molecular Cell, 44*(4), 660-666.

Jády, B. E., Darzacq, X., Tucker, K. E., Matera, A. G., Bertrand, E., & Kiss, T. (2003). Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm. *The EMBO Journal, 22*(8), 1878-1888.

Jan, E., & Sarnow, P. (2002). Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. *Journal of Molecular Biology, 324*(5), 889-902.

Jin, H., Loria, J. P., & Moore, P. B. (2007). Solution Structure of an rRNA Substrate Bound to the Pseudouridylation Pocket of a Box H/ACA snoRNA. *Molecular Cell, 26*(2), 205-215.

Kamalampeta, R., & Kothe, U. (2012). Archaeal proteins Nop10 and Gar1 increase the catalytic activity of Cbf5 in pseudouridylating tRNA. *Scientific Reports, 2*(663), 1.

Karijolich, J., & Yu, Y. T. (2011). Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature, 474*(7351), 395-398.

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. *Molecular Cell, 11*(2), 425-435.

Koo, B. K., Park, C. J., Fernandez, C. F., Chim, N., Ding, Y., Chanfreau, G., & Feigon, J. (2011). Structure of H/ACA RNP protein Nhp2p reveals cis/trans isomerization of a conserved proline at the RNA and Nop10 binding interface. *Journal of Molecular Biology, 411*(5), 927-942.

Kos, M., & Tollervey, D. (2010). Yeast pre-rRNA processing and modification occur cotranscriptionally. *Molecular Cell, 37*(6), 809-820.

Leulliot, N., Godin, K. S., Hoareau-Aveilla, C., Quevillon-Cheruel, S., Varani, G., Henry, Y., & Van Tilbeurgh, H. (2007). The box H/ACA RNP assembly factor Naf1p

contains a domain homologous to Gar1p mediating its interaction with Cbf5p. *Journal of Molecular Biology, 371*(5), 1338-1353.

Levy, M. Z., Allsopp, R. C., Futcher, A. B., Greider, C. W., & Harley, C. B. (1992). Telomere end-replication problem and cell aging. *Journal of Molecular Biology, 225*(4), 951-960.

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

Li, S., Duan, J., Li, D., Ma, S., & Ye, K. (2011a). Structure of the Shq1–Cbf5–Nop10–Gar1 complex and implications for H/ACA RNP biogenesis and dyskeratosis congenita. *The EMBO Journal, 30*(24), 5010-5020.

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

Li, X., Ma, S., & Yi, C. (2016). Pseudouridine: the fifth RNA nucleotide with renewed interests. *Current Opinion in Chemical Biology, 33*, 108-116.

Li, X., Zhu, P., Ma, S., Song, J., Bai, J., Sun, F., & Yi, C. (2015). Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nature Chemical Biology, 11*(8), 592-597.

Liang, B., Kahen, E. J., Calvin, K., Zhou, J., Blanco, M., & Li, H. (2008). Long-distance placement of substrate RNA by H/ACA proteins. *RNA, 14*(10), 2086-2094.

Liang, B., Xue, S., Terns, R. M., Terns, M. P., & Li, H. (2007). Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex. *Nature Structural & Molecular Biology, 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. *Nature Structural & Molecular Biology, 16*(7), 740-746.

Liang, X. H., & Fournier, M. J. (2006). The helicase Has1p is required for snoRNA release from pre-rRNA. *Molecular and Cellular Biology, 26*(20), 7437-7450.

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.

Liang, X. H., Liu, Q., Liu, Q., King, T. H., & Fournier, M. J. (2010). Strong dependence between functional domains in a dual-function snoRNA infers coupling of rRNA processing and modification events. *Nucleic Acids Research, 38*(10), 3376-3387.

Liu, B., Zhang, J., Huang, C., & Liu, H. (2012). Dyskerin overexpression in human hepatocellular carcinoma is associated with advanced clinical stage and poor patient prognosis. *PLoS ONE, 7*(8), e43147.

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.

Machado-Pinilla, R., Liger, D., Blaud, M., *et al.* (2011). The H/ACA RNP assembly factor SHQ1 functions as an RNA mimic. *Genes & Development, 25*(22), 2398-2408.

Machado-Pinilla, R., Liger, D., Leulliot, N., & Meier, U. T. (2012). Mechanism of the AAA+ ATPases pontin and reptin in the biogenesis of H/ACA RNPs. *RNA, 18*(10), 1833-1845.

Martin, R., Hackert, P., Ruprecht, M., *et al.* (2014). A pre-ribosomal RNA interaction network involving snoRNAs and the Rok1 helicase. *RNA, 20*(8), 1173-1182.

Mason, P. J., & Bessler, M. (2011). The genetics of dyskeratosis congenita. *Cancer Genetics, 204*(12), 635-645.

Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987). Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Research, 15*(21), 8783-8798.

Mitchell, J. R., Cheng, J., & Collins, K. (1999a). A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. *Molecular and Cellular Biology, 19*(1), 567-576.

Mitchell, J. R., & Collins, K. (2000). Human telomerase activation requires two independent interactions between telomerase RNA and telomerase reverse transcriptase. *Molecular Cell, 6*(2), 361-371.

Mitchell, J. R., Wood, E., & Collins, K. (1999b). A telomerase component is defective in the human disease dyskeratosis congenita. *Nature, 402*(6761), 551-555.

Montanaro, L. (2010). Dyskerin and cancer: more than telomerase. The defect in mRNA translation helps in explaining how a proliferative defect leads to cancer. *Journal of Pathology, 222*(4), 345-349.

Narayanan, A., Lukowiak, A., Jady, B. E., Dragon, F., Kiss, T., Terns, R. M., & Terns, M. P. (1999). Nucleolar localization signals of box H/ACA small nucleolar RNAs. *The EMBO Journal, 18*(18), 5120-5130.

Narla, A., & Ebert, B. L. (2010). Ribosomopathies: human disorders of ribosome dysfunction. *Blood, 115*(16), 3196-3205.

Natchiar, S. K., Myasnikov, A. G., Kratzat, H., Hazemann, I., & Klaholz, B. P. (2017). Visualization of chemical modifications in the human 80S ribosome structure. *Nature, 551*(7681), 472-477.

Nguyen, T. H. D., Tam, J., Wu, R. A., Greber, B. J., Toso, D., Nogales, E., & Collins, K. (2018). Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature, 557*(7704), 190–195.

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.

Noeske, J., Wasserman, M. R., Terry, D. S., Altman, R. B., Blanchard, S. C., & Cate, J. H. D. (2015). High-resolution structure of the *Escherichia coli* ribosome. *Nature Structural & Molecular Biology, 22*(4), 336-341.

O'Brien, R., Tran, S. L., Maritz, M. F., *et al.* (2016). MYC-Driven Neuroblastomas Are Addicted to a Telomerase-Independent Function of Dyskerin. *Cancer Research, 76*(12), 3604-3617.

Osheim, Y. N., French, S. L., Keck, K. M., Champion, E. A., Spasov, K., Dragon, F., Baserga, S. J., & Beyer, A. L. (2004). Pre-18S ribosomal RNA is structurally compacted into the SSU processome prior to being cleaved from nascent transcripts in *Saccharomyces cerevisiae*. *Molecular Cell, 16*(6), 943-954.

Pellizzoni, L., Baccon, J., Charroux, B., & Dreyfuss, G. (2001). The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1. *Current Biology, 11*(14), 1079-1088.

Penzo, M., Guerrieri, A. N., Zacchini, F., Trere, D., & Montanaro, L. (2017). RNA Pseudouridylation in Physiology and Medicine: For Better and for Worse. *Genes (Basel), 8*(11), 301-315.

Penzo, M., Rocchi, L., Brugiere, S., Carnicelli, D., Onofrillo, C., Coute, Y., Brigotti, M., & Montanaro, L. (2015). Human ribosomes from cells with reduced dyskerin levels are intrinsically altered in translation. *FASEB Journal, 29*(8), 3472-3482.

Piekna-Przybylska, D., Decatur, W. A., & Fournier, M. J. (2008). The 3D rRNA modification maps database: with interactive tools for ribosome analysis. *Nucleic Acids Research, 36*(Database issue), D178-183.

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

Rintala-Dempsey, A. C., & Kothe, U. (2017). Eukaryotic stand-alone pseudouridine synthases - RNA modifying enzymes and emerging regulators of gene expression? *RNA Biology, 14*(9), 1185-1196.

Ruggero, D., Grisendi, S., Piazza, F., Rego, E., Mari, F., Rao, P. H., Cordon-Cardo, C., & Pandolfi, P. P. (2003). Dyskeratosis Congenita and Cancer in Mice Deficient in Ribosomal RNA Modification. *Science, 299*(5604), 259-262.

Ruggero, D., & Shimamura, A. (2014). Marrow failure: a window into ribosome biology. *Blood, 124*(18), 2784-2792.

Schwartz, S., Bernstein, D. A., Mumbach, M. R., *et al.* (2014). Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell, 159*(1), 148-162.

Shukla, S., Schmidt, J. C., Goldfarb, K. C., Cech, T. R., & Parker, R. (2016). Inhibition of telomerase RNA decay rescues telomerase deficiency caused by dyskerin or PARN defects. *Nature Structural & Molecular Biology, 23*(4), 286-292.

Sieron, P., Hader, C., Hatina, J., Engers, R., Wlazlinski, A., Muller, M., & Schulz, W. A. (2009). DKC1 overexpression associated with prostate cancer progression. *British Journal of Cancer, 101*(8), 1410-1416.

Singh, M., Wang, Z., Cascio, D., & Feigon, J. (2015). Structure and interactions of the CS domain of human H/ACA RNP assembly protein Shq1. *Journal of Molecular Biology, 427*(4), 807-823.

Sloan, K. E., Warda, A. S., Sharma, S., Entian, K.-D., Lafontaine, D. L. J., & Bohnsack, M. T. (2017). Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biology, 14*(9), 1138-1152.

Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology, 60*(5), 523-533.

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.

Trahan, C., Martel, C., & Dragon, F. (2010). Effects of dyskeratosis congenita mutations in dyskerin, NHP2 and NOP10 on assembly of H/ACA pre-RNPs. *Human Molecular Genetics, 19*(5), 825-836.

Tran, E., Zhang, X., Lackey, L., & Maxwell, E. S. (2005). Conserved spacing between the box C/D and C′/D′ RNPs of the archaeal box C/D sRNP complex is required for efficient 2′-O-methylation of target RNAs. *RNA, 11*(3), 285-293.

Turowski, T. W., & Tollervey, D. (2015). Cotranscriptional events in eukaryotic ribosome synthesis. *Wiley Interdisciplinary Reviews: RNA*, *6*(1), 129-139.

van der Feltz, C., DeHaven, A. C., & Hoskins, A. A. (2018). Stress-induced Pseudouridylation Alters the Structural Equilibrium of Yeast U2 snRNA Stem II. *Journal of Molecular Biology*, *430*(4), 524-536.

Veerareddygari, G. R., Singh, S. K., & Mueller, E. G. (2016). The Pseudouridine Synthases Proceed through a Glycal Intermediate. *Journal of the American Chemical Society, 138*(25), 7852-7855.

Vulliamy, T., Beswick, R., Kirwan, M., Marrone, A., Digweed, M., Walne, A., & Dokal, I. (2008). Mutations in the telomerase component NHP2 cause the premature ageing syndrome dyskeratosis congenita. *PNAS, 105*(23), 8073-8078.

Walne, A. J., Vulliamy, T., Marrone, A., Beswick, R., Kirwan, M., Masunari, Y., Al-Qurashi, F.-h., Aljurf, M., & Dokal, I. (2007). Genetic heterogeneity in autosomal recessive dyskeratosis congenita with one subtype due to mutations in the telomerase-associated protein NOP10. *Human Molecular Genetics, 16*(13), 1619-1629.

Wang, C., & Meier, U. T. (2004). Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *The EMBO Journal, 23*(8), 1857-1867.

Wang, M., Liu, H., Zheng, J., *et al.* (2016). A Deafness- and Diabetes-associated tRNA Mutation Causes Deficient Pseudouridinylation at Position 55 in tRNAGlu and Mitochondrial Dysfunction. *Journal of Biological Chemistry, 291*(40), 21029-21041.

Wang, P., Yang, L., Gao, Y. Q., & Zhao, X. S. (2015). Accurate placement of substrate RNA by Gar1 in H/ACA RNA-guided pseudouridylation. *Nucleic Acids Research, 43*(15), 7207-7216.

Woolford, J. L., & Baserga, S. J. (2013). Ribosome Biogenesis in the Yeast *Saccharomyces cerevisiae*. *Genetics, 195*(3), 643-681.

Wright, J. R., Keffer-Wilkes, L. C., Dobing, S. R., & Kothe, U. (2011). Pre-steady-state kinetic analysis of the three *Escherichia coli* pseudouridine synthases TruB, TruA, and RluA reveals uniformly slow catalysis. *RNA, 17*(12), 2074-2084.

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

Wu, G., Radwan, M. K., Xiao, M., Adachi, H., Fan, J., & Yu, Y.-T. (2016b). The TOR signaling pathway regulates starvation-induced pseudouridylation of yeast U2 snRNA. *RNA, 22*(8), 1146-1152.

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

Wu, H., & Feigon, J. (2007). H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification. *PNAS, 104*(16), 6655-6660.

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.

Xue, S., Wang, R., Yang, F., Terns, R. M., Terns, M. P., Zhang, X., Maxwell, E. S., & Li, H. (2010). Structural basis for substrate placement by an archaeal box C/D ribonucleoprotein particle. *Molecular Cell, 39*(6), 939-949.

Yang, Z., Lin, J., & Ye, K. (2016). Box C/D guide RNAs recognize a maximum of 10 nt of substrates. *PNAS, 113*(39), 10878-10883.

Yoon, A., Peng, G., Brandenburger, Y., Zollo, O., Xu, W., Rego, E., & Ruggero, D. (2006). Impaired control of IRES-mediated translation in X-linked dyskeratosis congenita. *Science, 312*(5775), 902-906.

Yu, Y. T., & Meier, U. T. (2014). RNA-guided isomerization of uridine to pseudouridine--pseudouridylation. *RNA Biology, 11*(12), 1483-1494.

Zebarjadian, Y., King, T., Fournier, M. J., Clarke, L., & Carbon, J. (1999). Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. *Molecular and Cellular Biology, 19*(11), 7461-7472.

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.

Zhou, J., Liang, B., & Li, H. (2011). Structural and functional evidence of high specificity of Cbf5 for ACA trinucleotide. *RNA, 17*(2), 244-250.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research, 31*(13), 3406-3415.

Appendix

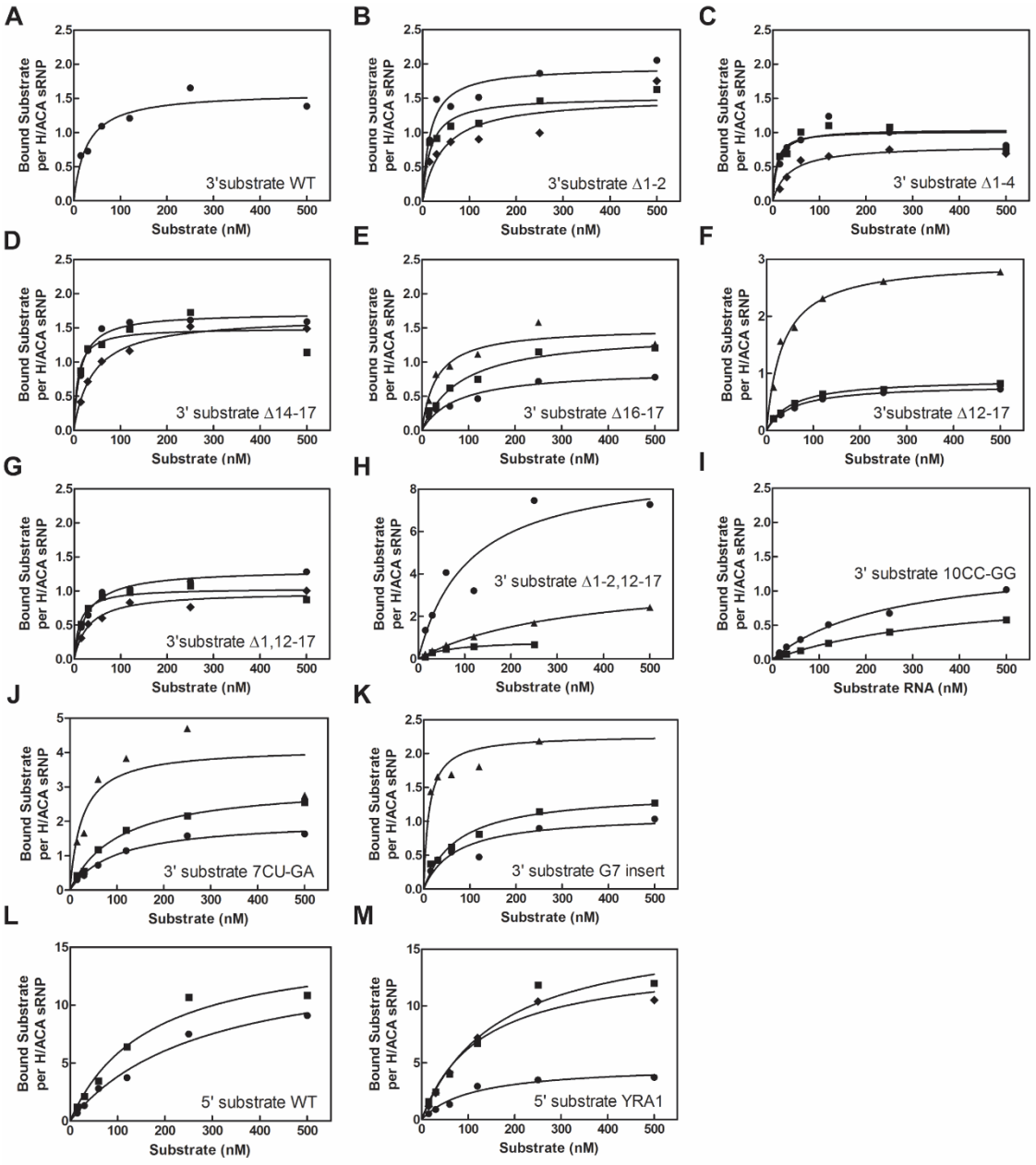

Figure A1. Affinity of the snR34 H/ACA snoRNP for short substrate RNA variants. Affinity of the snoRNP was determined for each substrate RNA variant using nitrocellulose filtration. The dissociation constant (K_D) was determined by fitting the data to Equation 2 (smooth lines – see Materials and Methods). Dissociation constants are listed in Table 4. (A) 3′ hairpin substrate wild type (n=1, previously determined by (Caton *et al.*, 2018)). (B) 3′ hairpin substrate Δ1-2 (n=3). (C) 3′ hairpin substrate Δ1-4 (n=3). (D) 3′ hairpin substrate Δ14-17 (n=3). (E) 3′ hairpin substrate Δ16-17 (n=3). (F) 3′ hairpin substrate Δ12-17 (n=3). (G) 3′ hairpin substrate Δ1,12-17 (n=3). (H) 3′ hairpin substrate Δ1-2,12-17 (n=3). (I) 3′ hairpin substrate 10CC-GG (n=3). (J) 3′ hairpin substrate 7CU-GA (n=3). (K) 3′ hairpin

substrate G7 insert (n=3). (L) 5′ hairpin substrate wild type (n=3). (M) 5′ hairpin substrate YRA1 (n=3).