STRUCTURE-BASED H/ACA GUIDE RNA DESIGN AND TESTING EXPLAINS THE STRUCTURE-FUNCTION RELATIONSHIP OF H/ACA GUIDE RNA

DOMINIC PHILIP CZEKAY
Bachelor of Science, University of Lethbridge, 2016

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

MASTER OF SCIENCE

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

STRUCTURE-BASED H/ACA GUIDE RNA DESIGN AND TESTING EXPLAINS THE STRUCTURE-FUNCTION RELATIONSHIP OF H/ACA GUIDE RNA

DOMINIC PHILIP CZEKAY

Date of Defence: November 26,2018

Dr. Ute Kothe Thesis Supervisor	Professor	Ph.D.
Dr. Hans-Joachim Wieden Thesis Examination Committee Member	Professor	Ph.D.
Dr. Igor Kovalchuk Thesis Examination Committee Member	Professor	Ph.D.
Dr. Nils Walter External Examiner University of Michigan	Professor	Ph.D.
Dr. Peter Dibble Chair, Thesis Examination Committee	Associate Professor	Ph.D.

Abstract

H/ACA guide RNAs are a class of noncoding RNA that direct the pseudouridylation of many cellular RNA species. In most eukaryotes, H/ACA guide RNAs share a conserved hairpin-hinge-hairpin structure, where each hairpin can direct pseudouridylation when associated with evolutionarily conserved core proteins. Target selection occurs by base pairing between target RNA and single-stranded loops within each hairpin of the H/ACA guide RNA, called pseudouridylation pockets. Here, I have analyzed the structure-function relationship of H/ACA guide RNAs by applying a structure-focused approach to design H/ACA guide RNAs for pseudouridylation of novel substrates. Thereby, I designed and tested several artificial H/ACA guide RNAs that were both highly active and specific for their respective substrates *in vitro*. In addition, I generated multiple sub-optimal H/ACA guide RNA designs that reveal important information regarding H/ACA guide RNA features dictating productivity. My results open new avenues for evaluating, predicting/identifying, and designing cellular guide-substrate RNA combinations.

Acknowledgements

I would like to start by giving a huge thank you to my amazing supervisor – Dr. Ute Kothe. She has been an amazing mentor throughout so many years of my schooling and has never failed to be supportive. Most of all, it amazes me just how positive and calming she can be. Discussing research progress each week with her would always provide you with a positive outlook and direction for the following week, which was incredibly helpful in staying motivated. She always encouraged me to be my best and do my best work and nothing here would have been possible without her unmatched guidance.

I would also like to thank my committee members, Dr. Hans-Joachim Wieden and Dr. Igor Kovalchuk, for providing me with great ideas and suggestions during all our meetings.

I would like to thank Dr. Nils Walter, Professor of Chemistry at the University of Michigan, for accepting to act as an external examiner for my thesis examination. I look forward to meeting you, hearing about your research, and having you present during my examination.

Lastly, I would like to thank my family and friends, including all members of the Alberta RNA Research and Training Institute here at the University of Lethbridge, and of course to my wonderful and amazing girlfriend for the unfailing support and encouragement throughout my time here. This accomplishment would not have been possible without anyone mentioned here.

Thank you all.

Dominic

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Amp	ampicillin
bp	base pair
CMC	N-cyclohexyl-N′-β-(4-methylmorpholinium)ethylcarbodiimide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
GMP	guanosine monophosphate
iPPase	inorganic pyrophosphatase
IPTG	isopropyl β-D-1-thiogalactopyranoside
Kan	kanamycin
LB	lysogeny broth
MD	molecular dynamics
MFE	minimum free energy
mRNA	messenger RNA
ncRNA	noncoding RNA
nt	nucleotide
NTP	nucleotide triphosphate
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PMSF	phenylmethylsulfonyl fluoride
Pol	polymerase
PUA	Pseudouridine synthase and Archaeosine transglycosylase domain
RNA	ribonucleic acid
RNase	ribonuclease
RNP	ribonucleoprotein
rRNA	ribosomal RNA
SDS	sodium dodecyl sulfate
SEC	Size Exclusion Chromatography
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
TBE	Tris-Borate-EDTA buffer
TLC	thin layer chromatography
tRNA	transfer RNA
UTP	uridine triphosphate
UV	ultraviolet
β-ME	β-mercaptoethanol

Chapter 1: Introduction

1.1 RNA modifications

Ribonucleic acid (RNA) is a macromolecule composed of linked ribonucleotides and is involved in a large number of cellular processes. One major class of RNA, called messenger RNA (mRNA), encodes polypeptides; any RNA not translated into protein is classified as non-coding RNA (ncRNA). ncRNAs typically fold into complex secondary and tertiary structures giving them unique functions beyond acting as message carriers according to the central dogma of biology which postulates that information is transmitted from DNA to RNA and ultimately to proteins.

To promote RNA function and diversity, RNAs can contain one (or many) of the more than 160 currently described post-transcriptional modifications [1]. The majority of these modifications have been identified in ncRNAs in part due to their relative total abundance and higher stability when compared to mRNA [2]. RNA modifications can influence gene expression in many ways including (but not limited to) being required for pre-mRNA splicing, promoting ribonucleoprotein (RNP) biogenesis and function, and affecting RNA stability [3]. For decades, many methods to detect and quantify RNA modifications have been applied; however, recent advancements in next generation sequencing technologies has revolutionized this field by allowing for rapid, easy, and reasonably accurate detection of nucleotide modifications within RNA samples [4].

1.2 The diversity of pseudouridylation

Pseudouridine (ψ) is the most abundant post-transcriptional modification present in RNA (reviewed in [5]). This structural isomer of uridine is present in all major classes of RNA including, but not limited to, ribosomal RNA (rRNA), transfer RNA (tRNA), small nuclear RNA (snRNA), and mRNA [6-9]. Due in part to its high abundance, pseudouridine was the very first RNA modification to be identified experimentally and has also been sometimes referred to as "the fifth nucleotide" [10]. Since its discovery in yeast, pseudouridylation has been found to be present in all domains of life, and interestingly, the number of pseudouridines within an organism appears to increase with the species' complexity. For example, *Escherichia coli* rRNA contains 36 pseudouridines, while *Saccharomyces cerevisiae* and *Homo sapiens* rRNA have 46 and 95 pseudouridines, respectively [1].

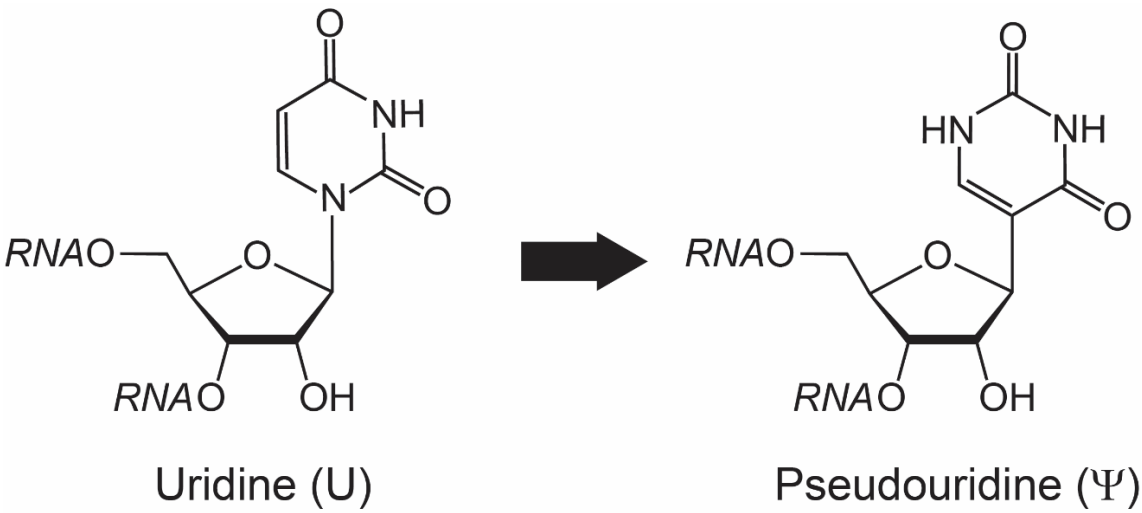


Figure 1. Comparison of a uridine nucleoside and a pseudouridine nucleoside.

Both uridine and pseudouridine have identical mass and produce the same ultraviolet (UV) spectra, but differ in their fragmentation patterns during mass

spectrometric analysis [11]. Additionally, uridine and pseudouridine differ slightly in their hydrophilicity allowing them to be distinguished from one another by thin layer chromatography (TLC), the method used in the initial discovery of pseudouridine [10]. Its unique features include the presence of an imino group at position 5 of the nucleobase (originally position 1) as a well as a C-C glycosidic bond replacing the original C-N bond (Figure 1).

For large scale mapping of pseudouridines, a pseudouridine-specific labelling approach is coupled with next-generation sequencing. Briefly, cellular RNA is labelled specifically at pseudouridines with N_3-[N-cyclohexyl-N′-β-(4-methylmorpholinium) ethylcarbodiimide]. The resulting bulky and stable ψ-CMC adduct terminates reverse transcription, which is followed by adaptor ligation, reverse transcription and Illumina® sequencing, allowing pseudouridylation sites to be mapped as regions of high read termination [12]. Pseudouridine has been identified to occur with the greatest frequency in rRNA when compared to other RNA species; most recently, pseudouridine sites in mRNA have been detected, but the functional roles of pseudouridylation in mRNAs remains unknown [9, 13, 14].

1.3 Pseudouridine synthases and mechanisms of pseudouridylation

Biosynthesis of pseudouridine occurs in the cell at the post-transcriptional level and requires enzymes known as pseudouridine synthases. Based on structure and sequence similarities, pseudouridine synthases are classified into six families: TruA, TruB, TruD, RsuA, RluA, and Pus10. Enzymes of the RsuA family are found only in bacteria, while those in the Pus10 family are only present in archaea and certain eukaryotes. Despite little sequence similarity between the families, all pseudouridine

synthases share a common core fold consisting of an eight-stranded mixed β sheet and a common active site cleft flanked by several helices and loops suggesting a common ancestor [15-19]. In eukaryotes, pseudouridine synthases localize are found in a variety of subcellular locations including the nucleus, nucleolus, Cajal bodies, cytoplasm, and mitochondria. In some cases, localization of a pseudouridine synthase has been been shown to be dynamic across different subcellular locations [14].

In bacteria, pseudouridylation is entirely achieved by stand-alone pseudouridine synthase enzymes, which both directly recognize and modify target RNAs (reviewed in [20]). Although eukaryotes have their respective versions of stand-alone pseudouridine synthases (reviewed in [21]), which introduce many of the pseudouridines known to be present in tRNA and mRNA, many pseudouridines in eukaryotes (particularly those in rRNA and snRNA) are introduced by a different cellular machinery known as H/ACA small ribonucleoproteins or H/ACA sRNPs (reviewed in [22]). Unlike pseudouridylation by stand-alone pseudouridines synthases, pseudouridylation by H/ACA sRNPs is RNA-dependent, requiring the presence of a guide RNA in association with core proteins to specify pseudouridylation sites.

The isomerization of uridine to pseudouridine requires the breakage of the uridine's C-N glycosidic bond, followed by a rotation of the nucleobase, and a reattachment of the nucleobase to the ribose sugar. Mechanistic studies have revealed that the pseudouridylation mechanism begins with a conserved catalytic aspartate residue (present in all pseudouridine synthase enzymes) attacking the C2 of the ribose sugar, breaking the C-N glycosidic bond, and forming a glycal intermediate [23, 24]. Besides the catalytic aspartate, molecular dynamics (MD) simulations have indicated the

presence of an electrostatic interaction network between multiple conserved residues at and near the catalytic core, a finding that was confirmed experimentally by showing that the charges of residues within the interaction network are important for catalysis [25]. There also exists a highly, but not universally conserved tyrosine residue in the active site cleft which is proposed to contribute to the active site structure and possibly acts as a general base during catalysis [26]. Kinetic investigations have demonstrated that pseudouridylation is slow with the rate-limiting step across multiple families of pseudouridine synthase enzymes being a uniformly slow catalysis rather than substrate binding[27]. Substrate recognition has been investigated in a number of stand-alone pseudouridine synthases, with most following a structure-specific mode of target recognition and specificity [28]. To date, only the eukaryotic stand-alone pseudouridine synthase Pus7 is confirmed to display some sequence-specific target selection, modifying the consensus sequence UGUAR (the modified uridine is underlined). Similarities in the pseudouridylation mechanism across all families of enzymes provide further support for the hypothesis that all pseudouridine synthase enzymes were derived from a common ancestor.

1.4 Roles of pseudouridine

The knockout of all stand-alone pseudouridine synthase enzymes (mostly in bacteria) typically does not impair cell growth under many conditions [29-31]. In all cases, except TruA, detrimental effects of the knockouts are seen only when grown in coculture with the wild type strain containing the pseudouridine synthase enzyme(s), where strains lacking pseudouridine synthase enzymes are outcompeted under most conditions. Similarly, eukaryotic stand-alone pseudouridine synthase deletions are also

viable and only small, if any, phenotypes are seen under stress conditions [32, 33]. Considering these results, it is surprising that pseudouridine synthases are not only conserved throughout evolution, but also that pseudouridines are clustered in functionally important regions of RNA [34]. Interestingly, in a few cases, cells harboring a catalytically inactive pseudouridine synthase gene can compete with a wild type strain, suggesting the possibility that the enzyme itself is serving an additional purpose whereas the pseudouridine itself is functionally unimportant [35, 36]. In the case of RNA-dependent pseudouridylation by H/ACA sRNPs, deletion of the core catalytic enzyme, Cbf5 (Dkc1 in humans), is lethal in all model species tested thus far, and mutations to this gene causes strong phenotypes including cold/heat sensitivity, reduced ribosome biogenesis, and in humans, a genetic disease called Dyskeratosis congenita causing bone marrow failure [37, 38].

Currently, only a small number of pseudouridines have partially characterized cellular roles (these instances are described later in this section). However, the majority of pseudouridines are not well enough understood to know the reason for their existence. One proposed role of pseudouridine stems from the fact that pseudouridine has additional hydrogen bonding capabilities compared to uridine due to the presence of an imino group on its Hoogsteen edge that is absent in the parent nucleotide. This imino group has the ability to coordinate a water molecule between the nucleobase and sugar-phosphate backbone [39]. Thereby, pseudouridines have a local stabilization effect on RNA structure and can also improve base stacking in RNAs [40].

Pseudouridines in rRNA have been shown to affect rRNA biogenesis and function, particularly in eukaryotes (reviewed in [41]). Interestingly, the removal of most

individual pseudouridines (through deletion of the responsible enzyme or guide RNA) seems to have little to no effect on cell growth or translation, but multiple deletions appear to cause synergistic effects that alter ribosome structure as well as reduce activity and fidelity of protein synthesis [42, 43]. Surprisingly, construction of an *E.coli* strain lacking all ribosomal pseudouridines by deleting every responsible pseudouridine synthase enzyme resulted only in minor effects on bacterial growth [44].

Like rRNA, snRNAs represent another class of functional RNA that are highly pseudouridylated, with 24 pseudouridines known to be present in the snRNA of the major human spliceosome alone (snRNA modification reviewed in [45]). Pseudouridines located within the 5′ end of U2 snRNA are critical for U2 snRNP maturation and pre-RNA splicing [46]. Specifically, Ψ34 in human U2 snRNA (ψ35 in yeast) induces a change in branch-site architecture that is required for recognition of the branch site adenosine within pre-mRNA, ultimately facilitating splicing [47].

tRNAs are another class of highly modified RNAs containing pseudouridines known to be important for correct function. In most tRNA isoforms, pseudouridines at positions 38 and 39 of the anticodon are required for efficient and accurate decoding [48]. tRNAs also have pseudouridines within the elbow region, but it appears that here the enzyme itself, but not necessarily the pseudouridine it introduces, is important; at least in one instance, *E. coli* TruB, the pseudouridine synthase enzyme acts as a tRNA chaperone with the pseudouridine introduced (Ψ55) being a byproduct of this process [30, 35]. Recently, pseudouridine has also been shown to be involved in the function of tRNA-derived fragments [49].

In mRNA, the role of pseudouridines has yet to be elucidated, but many studies indicate mRNA pseudouridylation to be a dynamic event (i.e. it responds to environmental stressors) [9, 14, 50]. This suggests the possibility that mRNA pseudouridylation may be altering gene expression, allowing cells to respond to changing conditions as needed. Remarkably, when present in the first position of a stop codon in mRNA, pseudouridines were determined experimentally to promote translational readthrough with high efficiency [51]. Although no evidence for stop codon pseudouridylation exists in nature, it is tempting to speculate that this could serve as a potential mechanism of gene expression regulation. When translated, ΨAA and ΨAG codons result in the incorporation of serine or threonine, while a ΨGA codon introduces tyrosine or phenylalanine [52]. Further studies revealed that readthrough is not caused by the pseudouridylated stop codon disrupting recognition by translational release factors, but rather that the pseudouridylated stop codon was recognized as a sense codon by non-cognate tRNAs [53]. For this to occur, a pair of purine-purine base pairs with unusual Watson-Crick/Hoogsten geometries is observed at positions 2 and 3 of codon-anticodon interactions in the crystal structure of a bacterial ribosome harboring the ψAG stop codon in the A site [54]. Further studies are needed to determine more examples of the functional contributions of pseudouridine in the cell.

1.5 H/ACA small ribonucleoproteins (sRNPs)

In archaea, box H/ACA sRNPs are responsible for introducing most pseudouridines in rRNA, as well as some in snRNA [55]. In yeast, RNA-dependent pseudouridylation is split between H/ACA snoRNPs, which pseudouridylate rRNA in nucleoli, and H/ACA scaRNPs which pseudouridylate snRNA in Cajal bodies;

collectively, the yeast H/ACA scaRNPs and H/ACA snoRNPs are referred to here as H/ACA sRNPs. A mature yeast H/ACA sRNP (Figure 2) consists of a box H/ACA guide RNA associated with four conserved proteins – Nhp2 (L7ae in archaea), Nop10, Gar1, and Cbf5 (Dkc1 in humans). In eukaryotes, H/ACA guide RNAs typically have two hairpins, and one set of core proteins can associate with each hairpin. Unlike stand-alone pseudouridine synthases, H/ACA sRNPs utilize a box H/ACA guide RNA to specify target RNAs for pseudouridylation in a sequence-specific manner by base pairing to available nucleotides within each internal loop [56].

To date, structure of the complete archaeal H/ACA sRNP complex as well as its eukaryotic counterpart in the form of human telomerase holoenzyme, whose RNA contains a 3′ H/ACA RNA-like structure bound by two heterotetramers of the four core H/ACA sRNP proteins, have been determined [57, 58]. The telomerase holoenzyme structure revealed a few interesting features of the H/ACA sRNP particle. First, the Dkc1 attached to one H/ACA guide hairpin can make extensive contacts with the second Dkc1 bound to the adjacent H/ACA RNA hairpin [58]. These contacts occur through the pseudouridine synthase and archaeosine transglycosylase (PUA) domain of Dkc1. Interestingly, many point mutations in residues involved in this interaction have been implicated in the rare human disease, X-linked Dyskeratosis Congenita [59]. The interaction between the two Dkc1 proteins also makes it tempting to speculate that a complex at one hairpin could potentially communicate with an adjacent complex as a means of modulating pseudouridylation activity. Second, the telomerase RNP structure reveals that the binding of core proteins to guide RNA is mediated almost entirely by Dkc1, indicating that Dkc1 alone is capable of anchoring the remaining core proteins to

the H/ACA guide. Notably, the proteins Nop10 and Nhp2 are essential for pseudouridylation activity [60-62], and it is suggested that their binding to the upper stem of H/ACA guide RNA is required for proper positioning of the guide RNA, in turn allowing the target uridine to be positioning correctly in the active site [60]. Besides the telomerase holoenzyme, the yeast Cbf5-Nop10-Gar1 crystal structure has been determined [63].

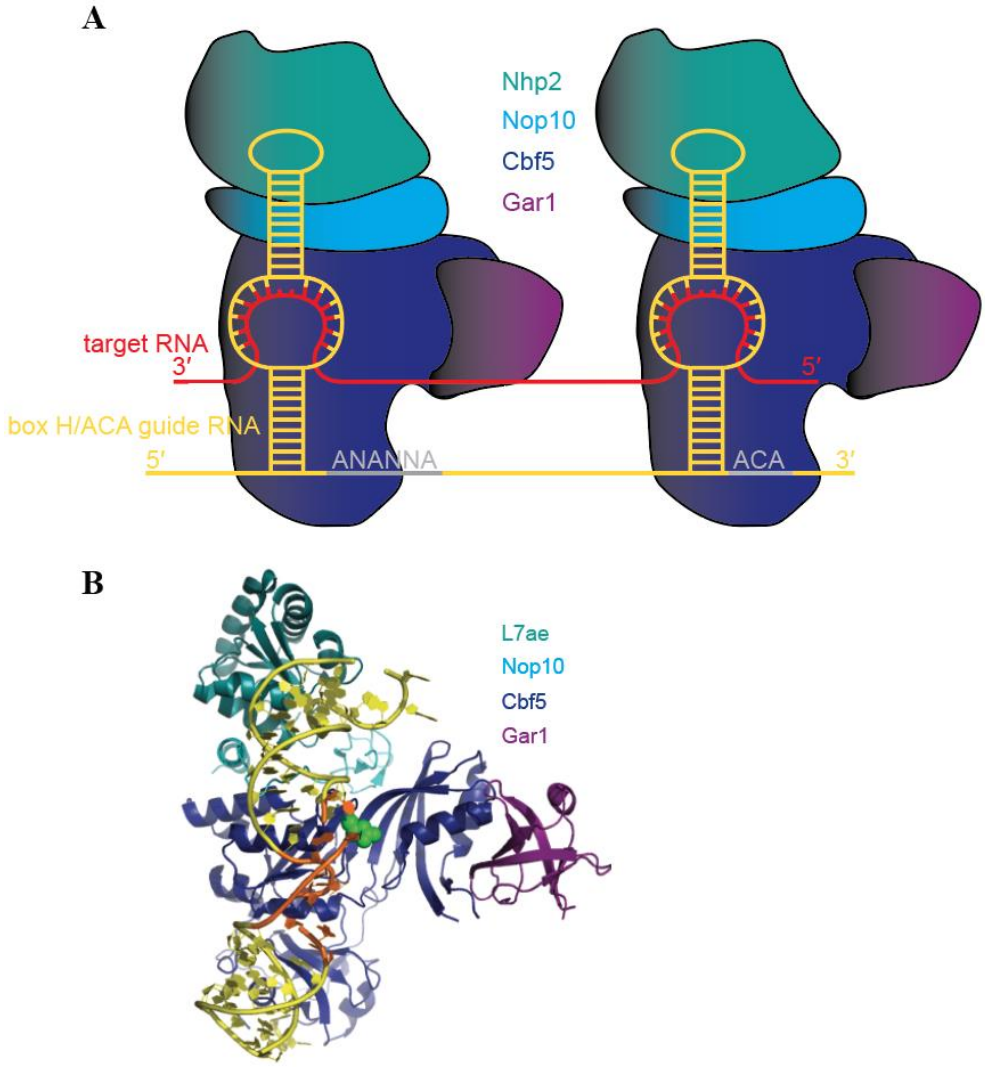


Figure 2. The structure of a mature H/ACA sRNP. A) Graphical representation of a mature eukaryotic H/ACA sRNP complex bound to a target RNA B) Crystal structure of the complete H/ACA sRNP particle from *Pyrococcus furiosus* (PDB=2HVY ; Li and Ye, *Nature* 2006). The single-hairpin guide RNA is depicted in yellow with the internal loop colored orange. The catalytic aspartate in the Cbf5 active site is shown as green spheres.

Cbf5 in yeast, and its homolog in other model eukaryota, is a 55-kDa essential protein that effects cellular fitness and cell cycle progression [64]. Years after its discovery, Cbf5 was shown to be important for rRNA pseudouridylation and processing, as well as for H/ACA snoRNA-specific accumulation [65]. Considering this, and its similarity to both the pseudouridine synthase TruB in bacteria and the pseudouridine synthase Pus4 in eukaryotes, Cbf5 was suggested and then confirmed to be the catalytic core component of a box H/ACA sRNP [65]. Cbf5 makes extensive contacts at the lower stem of each guide RNA hairpin with the conserved H and ACA box elements, and it mediates the recruitment of Nop10 and Gar1 to the H/ACA guide RNA through protein-protein interactions [66].

Gar1, another core component of H/ACA sRNPs, is an essential 21.5 kDa nucleolar protein with two glycine-arginine rich (GAR) regions [67]. Depletion of Gar1 causes abnormal rRNA processing which results in the accumulation of 35S pre-rRNA and an overall reduction in rRNA pseudouridylation [68]. Unlike Cbf5, Gar1 depletion has no effect on the nucleolar accumulation of H/ACA snoRNAs. Through its unique C-terminal domain, Gar1 associates with the Cbf5 thumb loop inducing an open conformation that is required for efficient substrate turnover [63].

Nop10 and Nhp2 are the remaining two protein components of a box H/ACA sRNP, and similar to Gar1, both are essential proteins that are required for formation of 18S rRNA [69]. In a mature H/ACA sRNP, Nop10 facilitates an indirect interaction between Cbf5 and Nhp2 in addition to binding the upper stem of an associated H/ACA guide RNA. This interaction, along with Nhp2-guide RNA association at the upper stem, helps to anchor the protein complex on a guide RNA allowing for proper docking of the

target uridine in the substrate RNA into the Cbf5 active site upon association of substrate RNA with guide RNA [57, 60]. Free Nop10 has a largely unstructured C-terminal region, but it is suggested to become structured upon association with Cbf5 forming many contacts across the entire surface of Cbf5 [61]. Nhp2 is recruited to the H/ACA sRNP through its interactions with Nop10, which is different from its archaeal counterpart L7Ae, as L7Ae recognizes and tightly binds a specific motif in the upper stem of each H/ACA RNA hairpin known as a k-turn, which is absent from eukaryotic H/ACA snoRNAs [70].

1.6 H/ACA guide RNA organization, processing, and sRNP assembly

The human genome has 181 known locations containing H/ACA guide RNA genes (corresponding to 108 unique H/ACA guides) largely found within the introns of housekeeping genes [71]. In yeast on the other hand, there is just one genomic occurrence for each of its 29 box H/ACA guide RNAs, with almost all being organized as individual genes under the control of their own promoter [72]. In both organisms, the vast majority of box H/ACA guides are transcribed by RNA Pol II [73]. Guide RNA expression appears mostly constitutive with only some cases of variable H/ACA guide RNA expression, usually during development [74, 75].

Following transcription and splicing of intron-encoded H/ACA guide RNAs, the lariat intron containing pre-snoRNA is debranched by the lariat debranching enzyme Dbr1 [76]. Removal of excess upstream and downstream RNA is performed by the exonuclease Rnt1 and the RNA exosome, respectively [77, 78]. The binding of core H/ACA sRNP proteins to newly transcribed H/ACA snoRNA protects it from further processing resulting in defined 5′ and 3′ ends. H/ACA sRNP biogenesis is a complex

process that is still not entirely understood; however, in general biogenesis involves the coupling of H/ACA snoRNA maturation and core protein assembly (reviewed in [79]).

In humans, assembly of core proteins begins with the binding of the early assembly protein, Shq1, to nascent dyskerin (Dkc1) in the cytoplasm which stabilizes Dkc1, preventing its aggregation and degradation [80]. Shq1 forms a tight clamp around Dkc1 with the C-terminal domain of Shq1 acting as an RNA mimic that is tightly bound by Dkc1 [81]. Thereby, Shq1 is believed to prevent the binding of unwanted cellular RNAs by Dkc1. Following nuclear localization to the site of H/ACA snoRNA transcription, the core proteins Nop10 and Nhp2 bind to the Shq1-Dkc1 complex and Shq1 is removed by the AAA+ ATPases pontin and reptin, allowing for Dkc1 to tightly bind pre-snoRNA, facilitating snoRNA processing [82]. During this time, the assembly factor Naf1 binds to Dkc1 through a domain which mimics the Gar1 core domain [83, 84]. Following nucleolar or Cajal body localization, Naf1 is replaced by Gar1, which is highly abundant in these subcellular compartments, and the production of a mature H/ACA sRNP is complete [85].

1.7 Current understandings of H/ACA snoRNA structure and function

Almost all eukaryotic box H/ACA guide RNAs in model organisms studied so far share a conserved secondary structure consisting of two hairpins (one to three hairpins in archaea), each of which contains an internal loop known as the pseudouridylation pocket. Additional features common to all H/ACA guide RNAs include the presence of two conserved sequences known as the H and ACA boxes, which are always located between the hairpins and three nucleotides from the 3′ end, respectively [86, 87]. The H and ACA boxes are required for snoRNA/snoRNP accumulation and activity *in* vivo, and for

pseudouridylation activity in sRNPs reconstituted *in vitro* [60, 88]. H/ACA RNAs can be further divided into two subclasses, H/ACA snoRNAs and H/ACA scaRNAs (reviewed in [89]), which guide the modification of rRNA in nucleoli and snRNA in Cajal bodies, respectively. Both have identical overall structure and consensus sequences; however, box H/ACA scaRNAs contain an additional localization sequence at the top of each hairpin known as a CAB box which specifically localizes them to Cajal bodies [90].

Unlike H/ACA guide RNAs in eukaryotes, archaeal guide RNAs contain secondary structures known as K-turns located within the upper stem of the hairpins which are specifically recognized and tightly bound by the core protein L7ae [91]. Besides this example, the binding of other H/ACA sRNP constituents to an H/ACA guide are mediated mostly through Cbf5 which appears to recognize structured RNA in general with nanomolar affinities [60]. H/ACA sRNPs have also been implicated in the stabilization of the human telomerase RNA component (hTR) as the H/ACA proteins bind and protect an H/ACA-like element at the 3′ end of hTR [92].

A guide RNA in a box H/ACA sRNP specifies where pseudouridine formation occurs in a sequence-specific manner by forming base pairs between nucleotides flanking a target uridine and nucleotides in the internal loop of the H/ACA guide RNA. Typically, the target uridine along with one additional substrate ribonucleotide immediately 3′ of the target uridine remain unpaired (Figure 3, green for reference). Analysis of all putative wild-type yeast H/ACA guide:substrate RNA-RNA interactions indicates considerable variability within these interactions. First, there is a large variation of the number of base pairs formed between guide RNA and substrate RNA ranging from as little as 8 bp to as many as 16 bp and with anywhere from 3 bp to 10 bp on either side of the target uridine.

Second, some guide:substrate interactions contain one or more noncanonical base pairs and have guide RNAs with various upper and lower stem lengths and stabilities. Such variability creates problems both in the prediction of productive guide:target RNA-RNA interactions and in the design of artificial H/ACA guide RNAs for efficient targeted pseudouridylation.

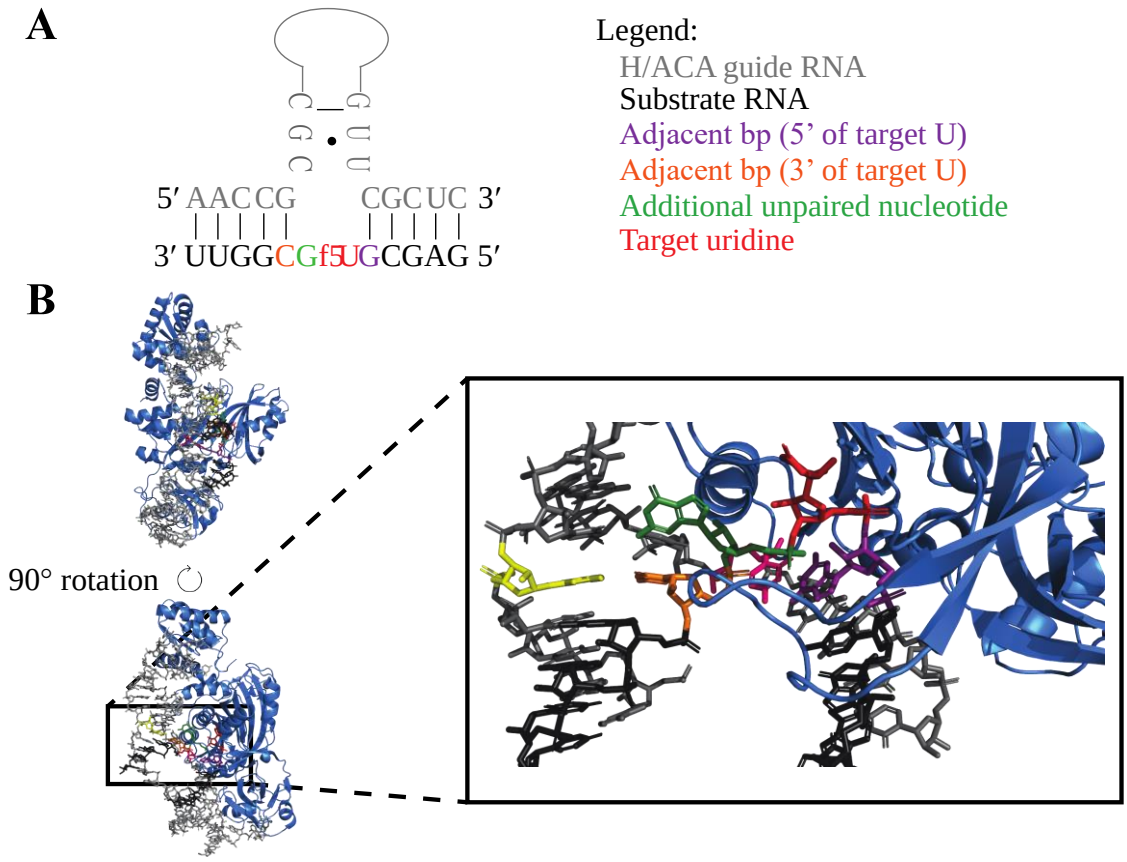


Figure 3. Depiction of H/ACA guide RNA:target RNA interactions within the active site of a functional H/ACA sRNP. A) Base pairing between an H/ACA guide RNA and target RNA with specific nucleotides colored for clarity. B) Enhancement of the substrate RNA:guide RNA interactions at the active site of a functional *P. furiosus* H/ACA sRNP (PDB:3HJY, reference [107]).

In archaea, an analysis of all putative guide:substrate pairings revealed a few interesting features, some of which may be common to guides of other organisms [93]. First, the length of the upper stem is strictly conserved, with 10 bp being present between the top of the internal loop and the G-A shear pair in the K-turn motif. This distance of

conserved across all archaeal H/ACA guide RNAs such that L7ae can specifically bind the K-turn to correctly position the guide RNA on the core proteins and allow for substrate catalysis. There also exists a strict 9 bp lower stem and 13-16 nt between the target uridine and consensus sequences; in eukaryotes, this corresponds to a conserved 14-15 nt between the target uridine and H or ACA box elements [94]. These findings suggest a model in which the conserved box H and box ACA elements are specifically recognized by Cbf5 and act as molecular rulers that determine correct guide positioning [60]. Lastly, the energetic contributions of the guide RNA:target RNA duplex on the 5′ side of the pseudouridylation pocket appear to be more important than the duplex on the 3′ side of the pseudouridylation pocket, a finding which may be generalizable to yeast H/ACA guides which, like archaea, have on average more base pairs on the 5′ side of the pseudouridylation pocket compared to the 3′ side. Further prediction of guide:substrate interactions is hampered by the absence of algorithms capable of predicting the energy of box H/ACA guide:target RNA-RNA interactions due to the fact that these interactions do not form standard A-form helices because of the unusual bending of the substrate RNA into an omega (Ω) shape [95].

Recently, a systematic investigation into *in vivo* box H/ACA guide:target RNA pairing was published that provides new insights into yeast H/ACA snoRNA target selection [96]. First, the minimum number of base pairs capable of directing pseudouridylation was confirmed to be 8 bp. However, these base pairs have to be distributed in such a way that base-pairing remained bipartite; if only two base-pairs were present on one side of the pseudouridylation pocket then activity was not observed, likely due to the inability of these base pairs to anchor one side of the substrate RNA

long enough for pseudouridylation. Next, the base pairs formed immediately adjacent to the target uridine (Figure 3, purple/orange) appear to play a more important role than other base pairs within the guide:substrate RNA-RNA interaction, especially when relatively few base-pairs are present on the same side of the pseudouridylation pocket. When these base pairs were disrupted, it was more likely that activity was abolished than if a base pair further away from the target uridine was disrupted, especially when approaching the 8 bp limit. Lastly, it was suggested that up to 4 unpaired nucleotides could be accommodated in the pseudouridylation pocket in addition to the target uridine, although the vast majority of wild-type guide RNA:substrate RNA interactions appear to have only 1 (at most 2) unpaired nucleotides adjacent to the target uridine (Figure 3, green).

A similar study was undertaken recently by our group which focused on establishing rules governing substrate:guide RNA pairing *in vitro* [97]. Interestingly, the number and type of base pairs between a guide RNA and a substrate RNA barely influenced the affinity of the H/ACA sRNP complex for substrate RNA; when nucleotides of the target RNA are changed to introduce a few mismatches to the H/ACA guide RNA, the affinity for substrate RNA remains similar to a wild type interaction. Although additional nucleotides next to the target uridine were suggested to be accommodated in the previous study, this was not the case in the *in vitro* study conducted by our group; the insertion of one additional unpaired nucleotide 3′ of the target uridine disrupted activity entirely. This difference implies that the ability to accommodate additional free nucleotides in the pseudouridylation pocket may vary from guide RNA to guide RNA. Lastly, in the guide:substrate RNA-RNA interaction examined by our group,

the 5′ and 3′ base pairs immediately adjacent the target uridine (Figure 3, purple/orange) were essential for activity, even in cases where the minimum number of 8 bp was well exceeded. Again, this result differs from the published study which determined these flanking base-pairs to be important mostly in cases where the 8 bp limit was nearing.

Although some concrete rules regarding guide RNA function have been established, much of the finer details determining guide RNA function remain elusive. Aspects such as preferred nucleotide composition and overall guide RNA stability may affect box H/ACA guide function. Additionally, as evidenced by recent studies, the findings observed at one H/ACA guide RNA hairpin may not be generalizable for all H/ACA guide RNAs [96, 97]. Thus, it is clear that a greater understanding of H/ACA guide RNA function and guide:substrate RNA pairing is needed, particularly in defining any rules that are generalizable for all H/ACA guide RNAs (if such rules exist).

1.8 Objective and Hypothesis

The objective of this study is to expand the investigations aimed at establishing a comprehensive understanding of box H/ACA guide RNA function and specificity and to further develop artificial guide RNA design strategies for targeted pseudouridylation. As described earlier, besides a few concrete rules that form the foundation of our understanding of H/ACA guide RNAs, current information available in this area appears to be guide RNA specific. Furthermore, published methods of artificial guide RNA design for targeted pseudouridylation of mRNA substrates tend to produce inefficient H/ACA guide RNAs [52]; guide RNAs produced following these protocols, in most instances, achieve pseudouridylation activities of no more than 15% when tested both *in vitro* and *in vivo*.

I hypothesize that there is a number of features of the H/ACA guide RNA and the guide:substrate RNA-RNA interaction that have yet to be entirely explored/understood, and that a comprehensive understanding of these will allow for efficient development of artificial H/ACA guide RNAs. First and foremost, I suggest that H/ACA guide RNA fold needs to be a major consideration when designing artificial H/ACA guide RNAs. Rather than simply changing nucleotides in the pseudouridylation pocket to target a sequence of interest, changes that must maintain the overall hairpin-hinge-hairpin structure as well as an open internal loop to allow for substrate RNA recognition. Furthermore, I hypothesize that by designing guide:substrate RNA-RNA interactions to more closely resemble their respective wild-type interaction (i.e. maintaining number and type of base pairs) higher activity can be achieved.

In this study, I aimed to design a number of artificial H/ACA guide RNAs using these rational principles and to test their activity against novel RNA substrates *in vitro*. Analyzing the activity of a number of these artificial guide RNAs will provide valuable information that can be applied as guiding principles for future H/ACA guide RNA design. Working towards a comprehensive understanding of H/ACA guide RNA function will aid in the identification of novel guide:target RNA possibilities, the identification of novel guides for orphan pseudouridines, and the design of efficient guide RNAs for the application of targeted pseudouridylation.

Chapter 2 – Materials and Methods

2.1 Overexpression and purification of *S. cerevisiae* Cbf5-Nop10-Gar1

pETDuet-6xHis-Cbf5-Nop10 and pET28a-Nop10 [60] were co-transformed intro BL21(DE3) *E. coli* cells (NEB) for expression and plated on selection media (LB + Ampicillin + Kanamycin). Positive transformants were screened for expression using 50 mL cultures of LB + Ampicillin + Kanamycin. Briefly, cultures were started at an initial OD_{600} of 0.1, then grown at 37°C with shaking to an OD_{600} of 0.6-0.8, at which point protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Protein expression was monitored by analyzing equivalent cell samples on a 12% SDS-PAGE and visualizing proteins with Coomassie Brilliant Blue staining. The best-expressing colony was used to inoculate cultures for large scale protein expression (6 L culture volume) performed similarly to test expressions described above. 6xHis-Cbf5-Nop10 was expressed overnight (14 - 16 hours) at 18°C, at which point cells were harvested by centrifugation (5000xg, 15 min, 4°C), shock frozen in liquid nitrogen to preserve protein integrity, and stored at -80°C until purification. GST-Gar1 was expressed independently for 3 hours at 37°C from pGEX-5X-3-Gar1 [60] in Rosetta 2(DE3) *E. coli* (NEB).

The *S. cerevisiae* Cbf5-Nop10-Gar1 complex was purified using a tandem glutathione-sepharose chromatography and nickel-sepharose chromatography approach. First, 6xHis-Cbf5-Nop10 expressing cells and GST-Gar1 expressing cells were combined in cell opening buffer: 50 mM Tris (pH 8.0), 1 M NaCl, 5% (v/v) glycerol, 0.5 mM phenylmethylsulfonyl fluoride (PMSF), and 1 mM β-mercaptoethanol (β-ME). Cell opening by sonication (level 6 intensity, 60% duty cycle, Branson Sonifier) was followed

by centrifugation at 30,000 x g for 45 min at 4°C. Cleared lysate containing all soluble components was combined with 5 mL of previously equilibrated Glutathione Sepharose™ 4 Fast Flow resin (GE Healthcare) for 1 hour on ice to allow for protein complex binding to the resin. The resin was subsequently washed with 150 mL of cell opening buffer lacking PMSF to remove undesired proteins, then Cbf5-Nop10-Gar1 was eluted from the resin with cell opening buffer supplemented with 10 mM reduced L-glutathione. Eluates were pooled (45 mL total) and combined with 3 mL of equilibrated Ni Sepharose™ 6 Fast Flow resin (GE Healthcare) for 1 hour on ice. The resin was washed 40 mL of Ni^{2+}-wash buffer (50 mM Tris (pH 8.0), 20% (v/v) glycerol, 1 mM β-ME, 20 mM imidazole), then eluted successively in 1 mL fractions with identical buffer containing 300 mM imidazole. Elutions were shock frozen in liquid nitrogen and stored at -80°C.

Yeast Cbf5-Nop10-Gar1 concentrations were determined independently for each elution. Briefly, purified Cbf5-Nop10-Gar1 was analyzed on a 10% SDS-PAGE gel along with known amounts of pseudouridine synthase TruB as a standard. The concentration of TruB was accurately determined previously using A_{280} measurements [27]. Following electrophoresis and visualization with Coomassie Brilliant Blue, the band intensity of Cbf5 in lanes containing Cbf5-Nop10-Gar1 complex was compared with the band intensity of the TruB standard using ImageJ software (accessed at https://imagej.nih.gov/ij/index.html). The concentration of Cbf5-Nop10-Gar1 complex was determined from the band intensity of Cbf5 in each elution relative to the protein standard. Cbf5-Nop10-Gar1 concentrations were additionally confirmed by testing activity with a wild-type snR34 guide RNA for its natural 25S rRNA substrate, an

interaction well characterized in our lab [60]. Briefly, guide RNA of known concentration was increasingly titrated with H/ACA proteins until no further increase in pseudouridylation activity was observed, indicating saturation of the protein complex on the guide RNA. Saturation suggests complete complex assembly at both hairpins of the guide RNA, and thus, any further additions of protein would not result in further increases of activity.

2.2 *In vitro* transcription and purification of H/ACA guide RNAs

Artificial H/ACA guide RNA gene sequences were synthesized as double-stranded gBlocks® Gene Fragments from Integrated DNA Technologies (IDT), and dissolved in MilliQ H_2O according to the suppliers' recommendations. Each gBlock® was blunt-end cloned into a SmaI site in pFL45 [98]. To generate DNA template for run-off *in vitro* transcription, the guide RNA coding sequence was amplified from the plasmid in a polymerase chain reaction (PCR) with primers specifying the artificial guide RNA (Table 2). For snR5_sub 3.v1, template DNA through PCR was unobtainable using the plasmid template; therefore, the gBlock® itself was used as template in the PCR. A T7 RNA polymerase promoter was introduced by incorporating its sequence in the forward primer. *In vitro* transcription reactions were performed for 3 – 4 hours at 37°C using the following reaction components: 40 mM Tris (pH 7.5), 15 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 10 mM dithiotreitol (DTT), 3 mM nucleotide triphosphates (NTPs), 5 mM guanosine monophosphate (GMP), 0.01 U/µL inorganic pyrophosphatase (iPPase), 0.3 µM T7 RNA polymerase, 1.0 U/µL RiboLock Ribonuclease (RNase) inhibitor, and 1 ng/µL template DNA (purified PCR product). Following transcription, 0.02 U/µL of deoxyribonuclease (DNaseI) was added to the reaction, and each reaction

was incubated for an additional 30 min at 37°C to remove DNA template. The reaction was stopped by extracting the RNA with phenol/chloroform. Briefly, RNA was extracted with a 1:1 phenol(pH 4.3):chloroform mixture, followed by two extractions with 1 volume each of chloroform. Subsequently, RNA was precipitated overnight with isopropanol at 4°C.

To generate H/ACA guide RNA containing only half of the artificial pseudouridylation pocket (set v2 guide RNAs), the initial PCR step to generate *in vitro* transcription template was optimized to use a primer with mismatches to the template on one half of the pocket specifying it as wild-type rather than artificial (Table 2). For example, for a 3′ hairpin set v2 guide RNA, the 3′ side of the 3′ pseudouridylation pocket was reverted to wild type. Similarly, for 5′ hairpin mutants of this type, the 5′ side of the pseudouridylation pocket was kept wild type. All artificial guide RNAs designed to target substrates 2 and 3 (Table 1 for substrate sequences) were produced as described above. To produce the snR34 variant targeting substrate RNA 1, multiple rounds of site directed mutagenesis (primers snR34_sub1_M1-M4 – Table 2) were performed on a wild type snR34 gene in pUC19. This plasmid was used as template for PCR to generate *in vitro* transcription template, with the remaining synthesis staying identical to what was described previously.

We also produced a pair of snR34 guide RNA variants with additional sequence added onto the 5′ end of the guide RNA that contained a short segment of rRNA sequence normally targeted for pseudouridylation by snR34; one of these chimeras (snR34_rRNA sub 1-chimera) had a 5′ extension containing the rRNA sequence normally targeted by the snR34 5′ hairpin, and the other (snR34_rRNA sub 2-chimera)

had the corresponding 3′ hairpin rRNA target sequence. Additional linker nucleotides were strategically chosen to avoid any unwanted effects on RNA secondary structure and to provide sufficient distance for the attached substrate RNA to reach its respective pseudouridylation pocket in 3D space (Figure 5A – design schematic). H/ACA guide:substrate RNA chimeras were generated by adding sequence to the 5′ end of the snR34 gene using insertional mutagenesis. Briefly, primers designed to extend in opposite directions (snR34_25S rRNA sub 1/2 primers – Table 2) were created to anneal at the 5′ terminus of the snR34 gene within pUC19. One primer had a large 5′ overhang with the sequence to be inserted. Following PCR, T4 DNA ligase was used to circularize the plasmid, followed by the transformation into *E. coli* DH5α, DNA miniprep and confirmation by sequencing. Template generation, *in vitro* transcription, and purification were identical to other H/ACA guide RNAs described here.

All H/ACA guide RNAs and their respective variants were purified by size exclusion chromatography on a Superdex 200 10/300 GL Column (GE Life Sciences) at a flow rate of 0.4 mL/min in 50 mM Tris (pH 7.5), 70 mM NH_4Cl, 30 mM KCl, 1 mM EDTA, 4 mM $MgCl_2$ buffer. Fractions corresponding to guide RNA with A_{254} values above 0.250 were pooled and precipitated with isopropanol overnight at 4°C. Guide RNAs were suspended in MilliQ H_2O, and their concentration was determined in triplicate using A_{260} measurements and the calculated molar extinction coefficient for each guide (OligoAnalyzer® from IDT).

2.3 *In vitro* transcription and purification of substrate RNAs

Template for run-off *in vitro* transcriptions of substrate RNAs was generated by annealing two oligonucleotides encoding a T7 RNA polymerase promoter upstream of

each 15 – 25 nt substrate RNA sequence (Table 2). Transcription of substrate RNA is identical to guide RNA transcription (detailed earlier) except for a few differences described here. First, final template concentration in the reaction is 0.05 µM of the annealed primers. Also, since substrate RNA must contain C5-tritiated uridine for pseudouridylation to be detectable using our activity assay, NTPs were added to the reaction individually; importantly, uridine triphosphate was added to a lower final concentration (0.1 mM) compared to the other NTPs, and consisted of a mixture of both non-radioactive UTP and [^{3}H]-UTP supplied by Moravek Biochemicals. Following transcription and template removal, substrate RNA was precipitated using isopropanol overnight at 4°C.

Substrate RNAs were purified using PAGE purification. Briefly, the entire RNA sample was subject to electrophoresis through a 15% polyacrylamide 8M Urea gel in 1X TBE buffer for 30 minutes at 300 V. Substrate RNA was visualized by UV shadowing over a thin layer chromatography (TLC) plate, excised, crushed into a fine powder, and soaked overnight at room temperature in 1X TBE buffer in the presence of 1.0 U/µL RNase inhibitor. The next day, RNA in the supernatant was extracted with phenol/chloroform (exactly as described for guide RNAs), then precipitated with ethanol overnight, and resuspended in MilliQ H_2O. Substrate RNA concentration and specific activity (dpm/pmol·U) was determined (at least) in triplicate for each substrate RNA using A_{260} measurements and scintillation counting (PerkinElmer Tri-Carb 2810 TR). The specific activities of the substrate RNAs ranged from 1500 – 1800 dpm/pmol·U.

2.4 Tritium release assay for *in vitro* pseudouridylation activity determination

H/ACA guide RNA was diluted to a concentration of 0.5 µM in activity buffer (20 mM HEPES (pH 7.4), 150 mM NaCl, 0.1 mM EDTA, 1.5 mM $MgCl_2$, 10% (v/v) glycerol, 0.75 mM DTT) and refolded by heating to 65°C for 5 min and then cooling slowly to room temperature. Refolded guide RNA and Cbf5-Nop10-Gar1 as well as Nhp2 proteins were combined in activity buffer in a 0.45:1 ratio (guide RNA:proteins), ensuring complete complex formation on both hairpins. Complex formation was allowed to occur for 10 min at 30°C prior to addition of substrate RNA. Excess substrate RNA (500 nM) was added to 45 nM reconstituted H/ACA sRNP complexes (45 nM H/ACA guide RNA + 100 nM H/ACA proteins) and then incubated at 30°C assaying activity at multiple points during the time course. Samples removed from reactions were quenched immediately with 5% (w/v) Norit A in 0.1 M HCl. Following centrifugation, the supernatant was again combined with 5% (w/v) Norit A in 0.1 M HCl, centrifuged, and filtered through a glass wool plug. Prior to scintillation counting, the filtrate was subject to centrifugation to ensure the absence of charcoal in the counted samples. Disintegrations per minute (dpm) were determined for each reaction at every time point, corresponding to the amount of tritium released into the supernatant, and finally used to determine the percentage of pseudouridines formed.

Table 1. Name and sequence of each substrate RNA generated.

Substrate RNA	Sequence (5′ – 3′)
SUB 1	GGGAAUUAACUUCUAAAAUGAAGGUCAUG
SUB 2	GGGCACGUUGGGUAAAGCGAAUUACCA

SUB 3	GGGCACUGGGUUCUAAGAAUUAACC
25S rRNA SUB 1	GGGGACAACUGGCUUGUGGCAGCCA
25S rRNA SUB 2	GGGACGUCGGCUCUUCCUAUCAUACC

Table 2. Oligonucleotide sequences use in the study (PCR, template generation, mutagenesis, etc.)

Primer Name	Sequence (5′ – 3′)
25S rRNA sub 1+T7_Fwd	AGGCAGCCACAAGCCAGTTGTCCCCTATAGTGAGTCGTATTAGC
25S rRNA sub 1+T7_Rev	GCTAATACGACTCACTATAGGGGACAACTGGCTTGTGGCTGCCT
25S rRNA sub 2+T7_Fwd	GGTATGATAGGAAGAGCCGACGTCCCTATAGTGAGTCGTATTAGC
25S rRNA sub 2+T7_Rev	GCTAATACGACTCACTATAGGGACGTCGGCTCTTCCTATCATACC
Substrate 1+T7_Fwd	GCTAATACGACTCACTATAGGGAATTAACTTCTAAAATGAAGGTCATG
Substrate 1+T7_Rev	CATGACCTTCATTTTAGAAGTTAATTCCCTATAGTGAGTCGTATTAGC
Substrate 2+T7_Fwd	GCTAATACGACTCACTATAGGGCACGTTGGGTAAAGCGAATTACCA
Substrate 2+T7_Rev	TGGTAATTCGCTTTACCCAACGTGCCCTATAGTGAGTCGTATTAGC
Substrate 3+T7_Fwd	GCTAATACGACTCACTATAGGGCACTGGGTTCTAAGAATTAACC
Substrate 3+T7_Rev	GGTTAATTCTTAGAACCCAGTGCCCTATAGTGAGTCGTATTAGC
snR34+T7_Fwd	GCTAATACGACTCACTATAGGGGAATCAAAAATTTATTTTTTACACGGAAACG
snR34_Rev	mUmAATGTAGACTTTCAACTTCATCCC
snR5+T7_Fwd	GCTAATACGACTCACTATAGGGATCATTCAATAAACTGATCTTCCGG
snR5_Rev	mAmUATGTACACCTAGAGCGAACC
snR81+T7_Fwd	GCTAATACGACTCACTATAGGGAGGTTAGGGACTGCAAAAGAAGCG
snR81_Rev	mAmGATGTGAAAAAGCGTCTTCCCCC
snR34_25S rRNA sub 1_Fwd	Phos-TGTGGTTGTTGGTTGTGGGTTTGTGGTTGTCCCTATAGTGAGTCGTATTAGC

snR34_25S rRNA sub 2_Fwd	Phos-ACGTCGGCTCTTCCTATCATACCACAACCACAAACCCACACAACC
snR34_25S rRNA sub_Rev	Phos-GACAACTGGCTTGTGGCAGCC
snR34_sub 1_M1	TCATTTAGTTGACTGAACCTGTCTTCTAACAG
snR34_sub 1_M2	GTCAGTCAACTAAATGATCGTTTCCGTG
snR34_sub 1_M3	GAAGTTATTTCAAACAAGATTTGAAATACGAGTTTCCCAG
snR34_sub 1_M4	CTTGTTTGAAATAACTTCAATTAACTACTGTTAGAAGACAGG

Table 3. gBlock® sequences containing artificial (set v1) H/ACA guide RNAs.

gBlock® name	Sequence (5′ – 3′)
snR5_sub2.v1_gB	TTAATAGGAACTCATGGTGTAATTTTGCTTCACTTTTCTGAATTCAAATGGAAATAGAAGTAAAAATGGGCAGAAAAGTTACCTCATGCTTTAAATATACTCAACTTTTCTCTACCGACAGTAGGAGCCCTAAGAGATAAAGCGAAAAATTTTCACAGCGCGCTCCTTGGTGCGGCGAGGGTATAAAAGGCGATTTCATTTCAATTCTTTGAACTTACTTCATTTGCAGGATCCTTCAGGATAAGAAAACCATCACTCGAGATCATTCAATAAACTGATCTTCCGGATTACCATGCTTAAGACATCACGCCTCCATATGTCTATATAAAGCGCAAATGGCTGGAAGTAGACCAATTTTTTTTGTTCCTAGATTTTCATTATTGAAATCGCTTCCAGTTTTAATGGTTTTTCTTAATTAAGAAAACAAATTATCATTGGCCCAACCTAGGTGTACATATGCTAGCCTTCACTTCATTACTCTCTTGTTTACATTGTATACATCGTACATATACTTCAATTAT
snR5_sub3.v1_gB	TTAATAGGAACTCATGGTGTAATTTTGCTTCACTTTTCTGAATTCAAATGGAAATAGAAGTAAAAATGGGCAGAAAAGTTACCTCATGCTTTAAATATACTCAACTTTTCTCTACCGACAGTAGGAGCCCTAAGAGATAAAGCGAAAAATTTTCACAGCGCGCTCCTTGGTGCGGCGAGGGTATAAAAGGCGATTTCATTTCAATTCTTTGAACTTACTTCATTTGCAGGATCCTTCAGGATAAGAAAACCATCACTCGAGATCATTCAATAAACTGATCTTCCGGATTACCATGCTTAAGACATCACGCCTCCATATGTCTATATAAAGCGCAAATGGCTGGAAGTAGACCAATTTTTTTTGTTCCTAGCTTTTCATTATTGAAACCCAAGCCAGTTTTAATGGTTTTTCTTAATTAAGAAAACAAATTATCATTGGTCTTAACTAGGTGTACATATGCTAGCCTTCACTTCATTACTCTCTTGTTTACATTGTATACATCGTACATATACTTCAATTAT
	TTAATAGGAACTCATGGTGTAATTTTGCTTCACTTTTCTGAATTCAAATGGAAATAGAAGTAAAAATGGGCAGAAAAGTTACCTCATGCTTTAAATATACTCAACTTTTCTCTAC

snR34_sub2.v1_gB	CGACAGTAGGAGCCCTAAGAGATAAAGCGAAAAATTT TCACAGCGCGCTCCTTGGTGCGGCGAGGGTATAAAAG GCGATTTCATTTCAATTCTTTGAACTTACTTCATTTGCA GGATCCTTCAGGATAAGAAAACCATCACTCGAGGAAT CAAAAATTTATTTTTTACACGGAAACGATGCCACAGTT GACTGAACCTGTCTTCTAACAGTAGTTAATTGCCAGTA TTTCAAACAAGATTTGAAATACGAGTTTCCCAGAATAA TTTATTTGGACTGATTCGCTTGTCCGATTTCTGTGTTGT CTCAAACGAGGCGATAGAATTGGGATCTCAAAGAAAG TCTACATTAGCTAGCCTTCACTTCATTACTCTCTTGTTT ACATTGTATACATCGTACATATACTTCAATTAT
snR34_sub3.v1_gB	TTAATAGGAACTCATGGTGTAATTTTGCTTCACTTTTCT GAATTCAAATGGAAATAGAAGTAAAAATGGGCAGAAA AGTTACCTCATGCTTTAAATATACTCAACTTTTCTCTAC CGACAGTAGGAGCCCTAAGAGATAAAGCGAAAAATTT TCACAGCGCGCTCCTTGGTGCGGCGAGGGTATAAAAG GCGATTTCATTTCAATTCTTTGAACTTACTTCATTTGCA GGATCCTTCAGGATAAGAAAACCATCACTCGAGGAAT CAAAAATTTATTTTTTACACGGAAACGAAATTCTAGTT GACTGAACCTGTCTTCTAACAGTAGTTAATTGAACCCA TTTCAAACAAGATTTGAAATACGAGTTTCCCAGAATAA TTTATTTGGACAGGATAGGAAGTCCGATTTCTGTGTTG TCTCAAACGAGGCGATAGAATTGGGATGTCGAAGAAA GTCTACATTAGCTAGCCTTCACTTCATTACTCTCTTGTT TACATTGTATACATCGTACATATACTTCAATTAT
snR81_sub2.v1_gB	TTAATAGGAACTCATGGTGTAATTTTGCTTCACTTTTCT GAATTCAAATGGAAATAGAAGTAAAAATGGGCAGAAA AGTTACCTCATGCTTTAAATATACTCAACTTTTCTCTAC CGACAGTAGGAGCCCTAAGAGATAAAGCGAAAAATTT TCACAGCGCGCTCCTTGGTGCGGCGAGGGTATAAAAG GCGATTTCATTTCAATTCTTTGAACTTACTTCATTTGCA GGATCCTTCAGGATAAGAAAACCATCACTCGAGAGGT TAGGGACTGCAAAAGAAGCGGCGAGGCAGCCCACATC AAGTGGAACTACACAGACTTCCTTGTCGCGATACTACG GTCCCAAGAGCAATCCTAACAAGCAAAATTCGCTTCCC CCGCTGAACCTGTACAGTCCACGGATGGTGCAGAAGTT ATATGATTTGGGGGCCCACGCTTGTTCACATCTGCTAG CCTTCACTTCATTACTCTCTTGTTTACATTGTATACATC GTACATATACTTCAATTAT
snR81_sub3.v1_gB	TTAATAGGAACTCATGGTGTAATTTTGCTTCACTTTTCT GAATTCAAATGGAAATAGAAGTAAAAATGGGCAGAAA AGTTACCTCATGCTTTAAATATACTCAACTTTTCTCTAC CGACAGTAGGAGCCCTAAGAGATAAAGCGAAAAATTT TCACAGCGCGCTCCTTGGTGCGGCGAGGGTATAAAAG GCGATTTCATTTCAATTCTTTGAACTTACTTCATTTGCA GGATCCTTCAGGATAAGAAAACCATCACTCGAGAGGT

	TAGGGACTGCTAATTCTGCGGCGAGGCAGCCCACATCA AGTGGAACTACACAGACTTCCTTGTCGCGAACCCACGG TCCCAAGAGCAATCCTAACAAGCAATTACATATTCCCC CGCTGAACCTGTACAGTCCACGGATGGTGCAGAAGTTA TATGATTTGGGGGAAGACGCTTTTTCACATCTGCTAGC CTTCACTTCATTACTCTCTTGTTTACATTGTATACATCG TACATATACTTCAATTAT

Chapter 3 – Results

3.1 Purification of H/ACA guide RNAs

To reliably analyze pseudouridylation activity of H/ACA guide RNAs *in vitro*, I purified a number of different guide (wild-type, artificial, and chimeric guide RNAs – introduced later), in order to use them in activity assays. For comparative purposes, it was also important that all H/ACA guide RNAs were of similar purity. Following *in vitro* transcription, all H/ACA guide RNAs were purified by size exclusion chromatography (SEC) to remove abortive transcripts, unincorporated NTPs, and proteins (i.e. RNA polymerase, RNase inhibitor, and iPPase). Figure 4A shows a representative chromatogram obtained during SEC purification of snR5_sub 2.v1 (version 1 of artificial snR5 guide RNA designed to target substrate 2).

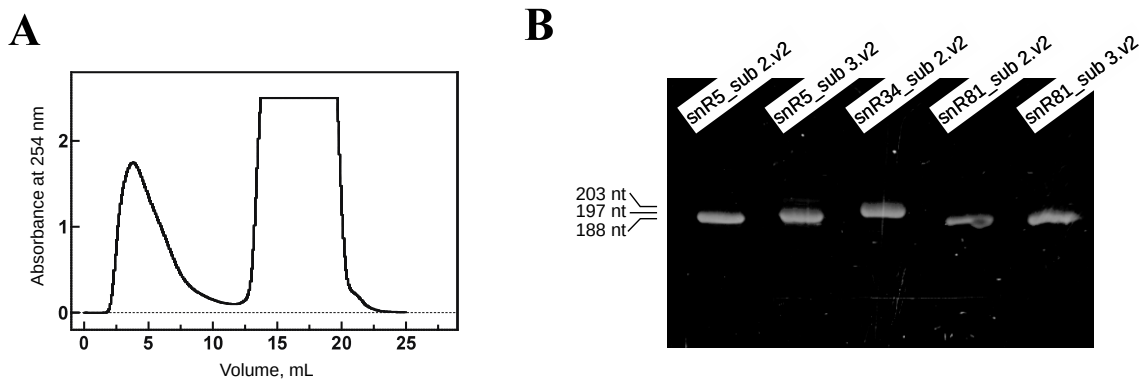


Figure 4. Size Exclusion Chromatography purification of artificial H/ACA guide RNAs. All H/ACA guide RNAs were purified using size exclusion chromatography using a Superdex 200 10/300 GL column (GE Life Sciences). A) Representative chromatogram measuring the absorbance at 254 nm of fractions eluted off the size exclusion column during the purification of the artificial guide RNA - snR5_sub 2.v1. B) 8M urea 12% PAGE of set v1 artificial guide RNAs following SEC. 10 pmol of each guide RNA was loaded. The sizes of guide RNA variants are as follows: snR5 = 197 nt, snR34 = 203 nt, snR81 = 188 nt. Ethidium bromide staining followed by UV illumination was used for visualization.

Two peaks were obtained during the chromatography, one small and sharp peak around 4 mL elution volume followed by a much larger and broader peak eluting between 14 – 20 mL, which is characteristic of free NTPs. The early peak fractions were analyzed by 8M Urea 12% PAGE, and those fractions containing RNA were pooled and precipitated. After purification, the concentration of each H/ACA guide RNA was determined using UV absorbance, and 10 pmol of each guide RNA was subject to Urea PAGE to confirm relative size, concentration, and purity (Figure 4B). Each purified guide RNA is clearly visible with similar band intensities, providing support for consistent concentration determination between the H/ACA guide RNA samples. Relative to one another, the sizes of each artificial guide RNA are all correct as well; an snR34 artificial guide RNA (203 nt) should be slightly larger than an snR5 artificial guide RNA (197 nt), and both should be larger than an artificial snR81 guide RNA (188 nt). In RNA preparations with multiple bands (ex. the faint larger band in snR5_sub 3.v2 – Figure 4B), the larger bands are a result of T7 RNA polymerase having a tendency to add additional nucleotides to the 3′ end of a transcript with high self-complementarity [99]. Following purification, H/ACA guide RNAs were diluted to 0.5 µM in 1X activity buffer (see methods) and stored at -20°C until used.

3.2 A cis-substrate RNA element competes for H/ACA guide pseudouridylation pockets

To investigate how H/ACA guide RNA activity was influenced by the addition of a 5′ RNA element and to investigate the possibility of pseudouridylation in cis, I generated a pair of snR34 guide RNA:substrate RNA chimeras (Figure 5A). I hypothesize that if the attached element does not interfere with the overall structure of

the H/ACA guide RNA, it should be able to be bound and potentially be modified in cis. Furthermore, I would expect that if the above were true, using an H/ACA guide:substrate chimera to target free substrate RNA would result in a competition effect where the free substrate RNA and the substrate RNA element attached in cis would compete for access to the pseudouridylation pocket. This would result in a reduction in pseudouridylation activity for the free substrate RNA in trans, as it must outcompete the RNA bound to the pseudouridylation pocket in cis.

Following reconstitution with the four core proteins, three H/ACA sRNP complexes (snR34 wild-type, snR34_rRNA sub 1-chimera, and snR34_rRNA sub 2-chimera) were tested in biological triplicate for activity against both 25S rRNA sub 1 and 25S rRNA sub 2 (Figure 5B). Pseudouridylation activity was determined using the tritium release assay, which measures the activity of a pseudouridine synthase for a substrate RNA containing [^{3}H-C5]-uridine, which is generated by *in vitro* transcription using [^{3}H-C5]-UTP. Pseudouridylation of a labeled uridine results in a detectable release of tritium into the supernatant due to bond cleavage of the C5-^{3}H bond in the target uridine when the new C1′-C5 glycosidic bond in pseudouridine is formed. In the assays performed with chimeric H/ACA guide RNAs, only modification of free radioactive 25S rRNA substrate (targeted in trans) is detectable, since the 25S rRNA substrate elements attached in cis to the 5′ extension of snR34 are not radioactively labeled. Thus, the modification of these cis RNA elements will not produce any detectable activity.

A

B

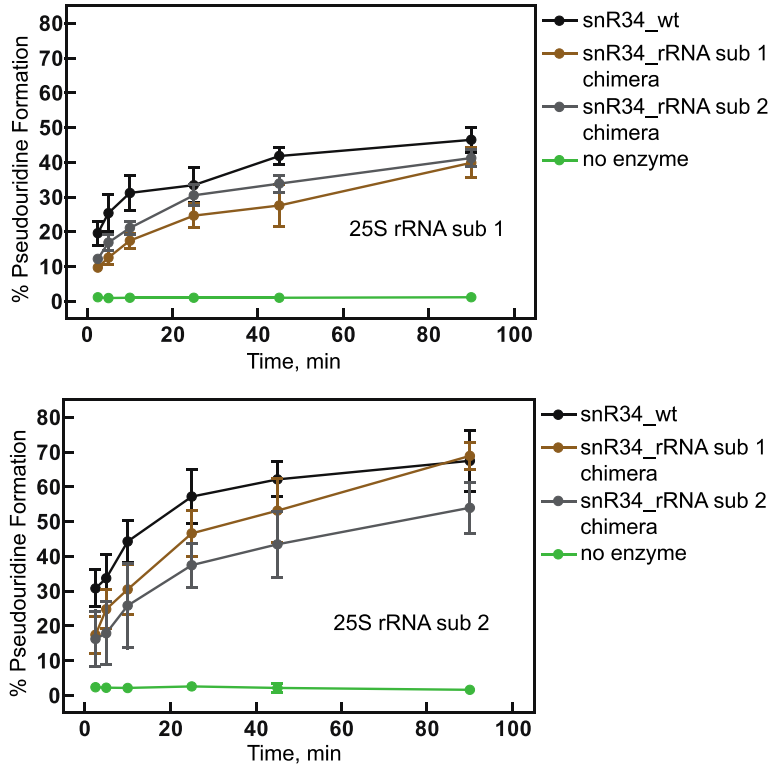


Figure 5. Pseudouridylation activity of H/ACA guide RNA:substrate chimera sRNPs for substrate RNA in trans. A) Schematic representation of a H/ACA guide:substrate chimeric RNA design. B) Tritium release assay examining the pseudouridylation activity of *in vitro* reconstituted snR34_wt (black), snR34_rRNA sub 1-chimera (brown), and snR34_rRNA sub 2-chimera (gray) box H/ACA sRNPs for 25S rRNA sub 1 (top) and 25S rRNA sub 2 (bottom). Reactions were performed under multiple turnover conditions (500 nM substrate RNA; 100 nM proteins (Cbf5-Nop10-Gar1-Nhp2); 45 nM guide RNA).

Each guide:substrate chimera provided pseudouridylation activity for both 25S rRNA substrate 1 and 25S rRNA substrate 2 which are normally targeted by the 5′ and 3′ pseudouridylation pockets of snR34, respectively (Figure 5B). Although the highest level of pseudouridine formation was provided by the wild type snR34 H/ACA sRNP, both

chimeric H/ACA sRNPs maintained relatively high levels of pseudouridine formation. The difference between wild-type and chimeric H/ACA sRNPs was significant, particularly at time points earlier than 30 minutes in the reaction. The slight decrease in pseudouridylation activity of the chimeric guide RNAs confirms that the 5′ additions made to snR34 minorly influenced H/ACA-guided pseudouridylation activity.

Interestingly, when targeting the same 25S rRNA substrate in trans as the 25S rRNA sequence attached in cis, a small, but significant decrease in activity was observed when compared to the other chimeric H/ACA sRNP for the same substrate. In other words, the snR34_rRNA sub 1-chimera was more active for 25S rRNA sub 2 than the snR34_rRNA sub 2-chimera, and vice versa for 25S rRNA sub 1 (Figure 5B). These results can be explained by the competition effect described earlier, stating that reduced activity in trans could be explained by the binding the attached substrate RNA element in cis. Binding of attached substrate RNA in cis and binding of free substrate RNA in trans cannot be simultaneous if they are both bound by the same pseudouridylation pocket. Thus, if an assembled chimeric snR34 guide is binding the rRNA substrate attached to its 5′ end at the 5′ hairpin, then it would not be able to modify a free 5′ hairpin substrate until the attached substrate is released, allowing binding of the free substrate RNA to the pseudouridylation pocket. The relative pseudouridine formation of the chimeric H/ACA sRNPs to one another are consistent at each time point during the reactions, providing further evidence that these complexes are significantly different throughout the reaction. These results generally agree with the hypothesis made earlier that chimeric guides may in fact have the capacity to target pseudouridylation in cis. Although pseudouridylation of the 25S rRNA element in cis cannot be confirmed directly using our system, one may

speculate that since the chimeric H/ACA guide RNAs can bind an RNA element in cis, activity could possibly be achieved, albeit with unknown (potentially low) efficiency.

3.3 Basic artificial box H/ACA guide design strategies are not always effective

To begin my investigation into optimizing artificial guide RNA design strategies, I wanted to establish the efficacy of the most basic design principles used to engineer artificial H/ACA guide RNAs for targeted pseudouridylation. This strategy has just one consideration – maximizing base pairing between a guide RNA pseudouridylation pocket and a target RNA by changing pseudouridylation pocket nucleotides in an H/ACA guide RNA to be complementary to a new substrate RNA sequence. Along these lines, I performed mutagenesis of the 3′ hairpin pseudouridylation pocket of snR34 to provide as many base pairs as possible between the artificial snR34 guide RNA and an artificial substrate RNA (sub 1); compared to the wild-type snR34 guide:substrate interaction, this approach increased the number of base pairs with substrate RNA from 15 to 18 (Figure 6B). However, this strategy does not take into consideration how the changes made at the pseudouridylation pocket affect the overall fold of the RNA, and thus relies on the the assumption that the changes introduced do not negatively influence H/ACA guide RNA fold and also that all nucleotides needed for substrate RNA pairing are available to do so. This guide RNA variant, designated as snR34_sub 1 (since it was designed to target artificial substrate RNA 1), was reconstituted *in vitro* with purified core proteins (Cbf5 Gar1, Nop10, and Nhp2), then tested for pseudouridylation activity using a tritium release assay. The activity of a wild type snR34 H/ACA sRNP for its respective 3′ hairpin substrate RNA, a short fragment of *S. cerevisiae* 25S rRNA (25S rRNA sub 2), was used as an activity reference [60].

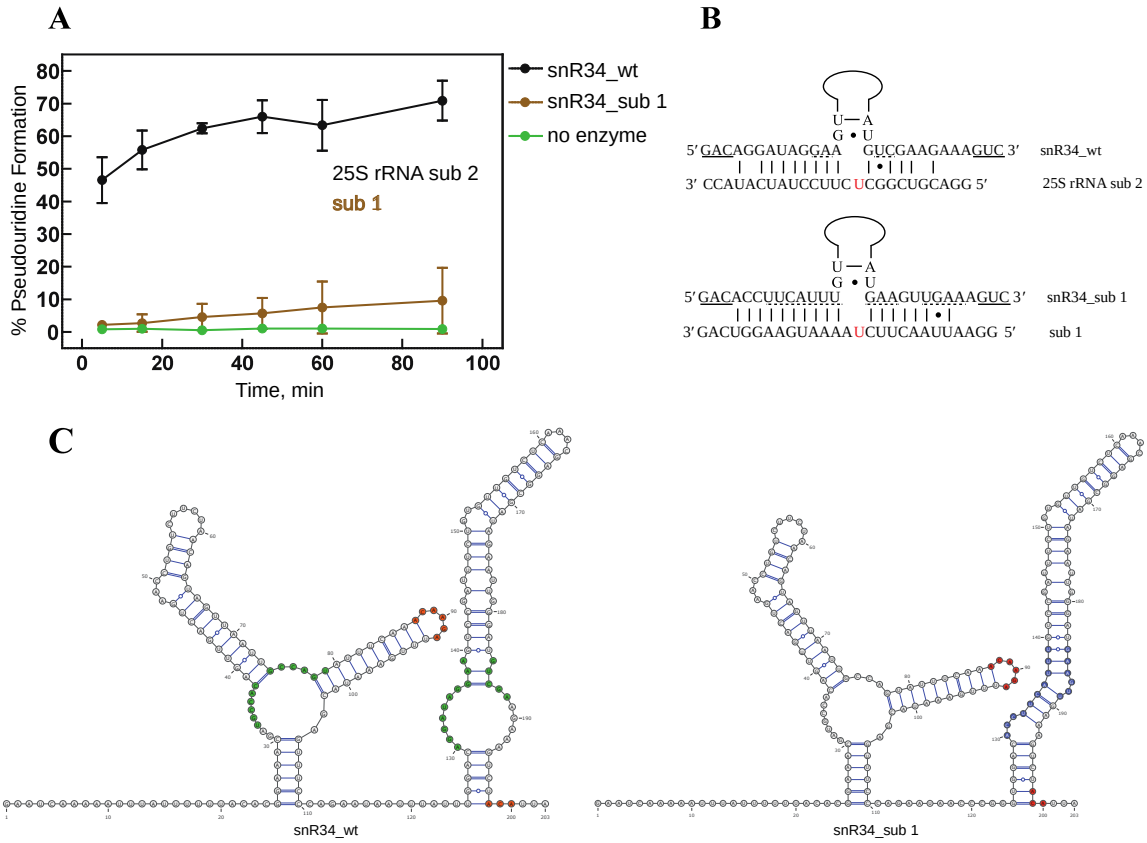


Figure 6. Pseudouridylation activity of snR34_sub 1 H/ACA sRNP and snR34_wt. A) Tritium release assay of *in vitro* reconstituted snR34_wt (black) and snR34_sub 1 (brown) H/ACA sRNPs for their respective tritiated substrate RNAs, performed under multiple turnover conditions (500 nM substrate RNA; 100 nM proteins (Cbf5-Nop10-Gar1-Nhp2); 45 nM guide RNA). B) The base pairing possibilities between pseudouridylation pockets on snR34_wt (top) and snR34_sub 1(bottom) for 25S rRNA sub 2 and artificial RNA sub 1, respectively. Underlined nucleotides in the guide RNA are nucleotides which are part of a stem/base pair according to structure prediction while dashed underlines indicate nucleotides whose pairing needed to be disrupted to achieve the suggested outcome. The red uridine is believe to be isomerized if active. C) The Boltzmann-weighted centroid structure prediction of snR34_wt (left) and snR34_sub 1 (right). Nucleotides in the wild-type snR34 pseudouridylation pockets needed for target recognition are colored green while nucleotides in the 3′ pseudouridylation pocket of snR34_sub 1 required for targeting substrate RNA 1 are highlighted in blue. The conserved box H and ACA elements are colored orange.

Pseudouridylation guided by the artificially designed H/ACA guide RNA was significantly lower when compared to pseudouridylation by wild-type snR34 (Figure

6A). Precisely, pseudouridine formation with snR34_sub 1 remains barely above baseline, reaching only 10% modification after 90 min. By comparison, snR34 H/ACA sRNP activity is very high, with almost 50% of 25S rRNA sub 2 being modified after just 2.5 minutes.

The observed activity of each guide RNA can be rationalized by their predicted secondary structures, obtained using an open-source folding algorithm that determines the Boltzmann-weighted ensemble centroid structure (webserver can be found at https://rna2dmut.bb.iastate.edu/ - [100]). Ensemble centroid structure prediction algorithms make fewer prediction errors compared to free energy minimization by determining and characterizing both the minimum free energy structure in addition to a set of suboptimal structures [101]. Amongst this ensemble of structures, the centroid structure is that which best represents the ensemble (i.e. the structure which has the smallest base-pair distance to all structures in the ensemble). The structure of the snR34_sub 1 indicates that the changes introduced through mutagenesis resulted in the formation of many base pairs within the internal loop essentially eliminating the pseudouridylation pocket (Figure 6C, right). In comparison, the much more active wild-type snR34 guide RNA is predicted to fold into a structure where nucleotides needed for guiding modification are available to pair with substrate RNA (Figure 6C, left).

3.4 Rational artificial H/ACA guide RNA design overview

It was my objective to create an effective and consistent strategy for the design of artificial H/ACA guide RNAs for targeted pseudouridylation. To accomplish this, I developed a more rational approach based on current understandings of H/ACA guide RNA, in particular, the structure-function relationship of H/ACA guide RNAs. Rather

than designing artificial H/ACA guide RNAs by simply changing pseudouridylation pocket nucleotides to form as many base pairs as possible with a new target RNA sequence, I applied the following design principles. First, changes to guide RNA sequence were only made at positions in the guide RNA that normally base pair to target RNA in the wild-type scenario (i.e. maintaining "wt-like" H/ACA guide RNA:substrate RNA pairing). This allows for the maintenance of key features observed in the wild-type guide:substrate RNA-RNA interaction (i.e. keeping the same number and position of Watson-Crick and non-Watson-Crick base pairs whilst still promoting pairing with a new substrate sequence). Keeping the base-pairing similar to what is seen in the wild-type guide RNA ensures that any putative requirements, such as number/strength of base pairing on one or both sides of the pseudouridylation pocket, are maintained.

Second, additional consideration was taken to ensure that the fold of each box H/ACA guide RNA hairpin, predicted via a Boltzmann-weighted centroid structure folding algorithm [102], was maintained after any changes to guide RNA sequence were introduced. To evaluate the predicted H/ACA guide RNA structure, the following set of criteria was applied: 1. Maintenance of an overall hairpin-hinge-hairpin structure and/or maintenance of the overall fold observed in the parent wild-type guide RNA. 2. Presence of conserved box elements as single stranded regions in the expected location (i.e. H box in the hinge between the two hairpins and ACA box after the 3′ hairpin). 3. Existence of a (mostly) single-stranded pseudouridylation pocket to allow for pairing with target RNA. If a fold was obtained that did not conform to the listed criteria (predicted to be "ideal"), further optimizations to H/ACA guide design were carried out, usually by introducing mutations in regions outside the pseudouridylation pocket, but avoiding

regions known to be important for guide RNA function (i.e. H and ACA box elements). The goal of further optimization was to disrupt the formation of undesirable structures and promote the formation of an artificial H/ACA guide RNA fold that aligns with the proposed criteria.

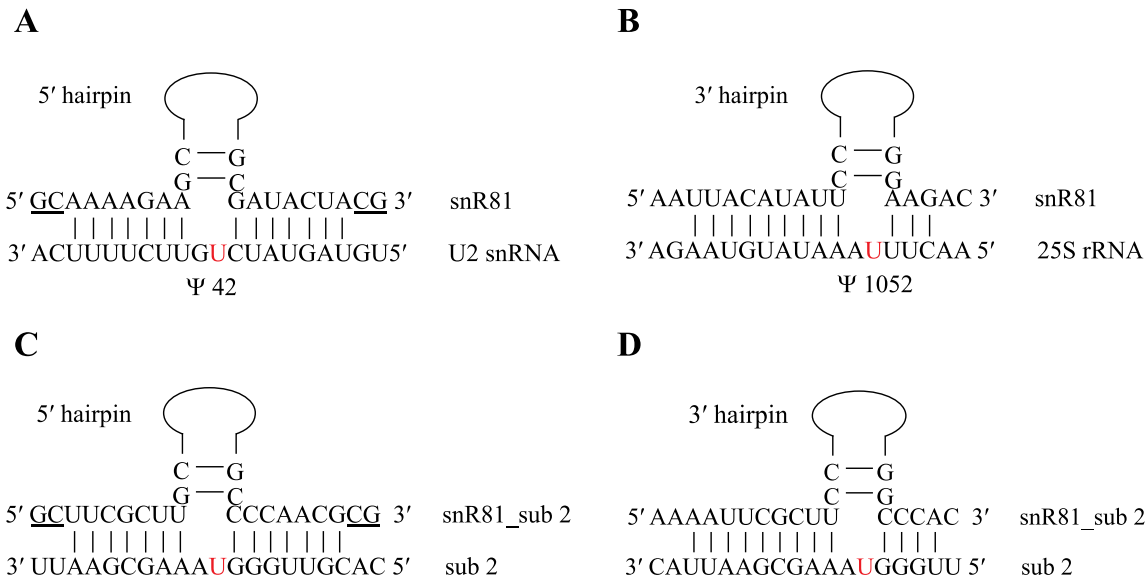


Figure 7. Guide:substrate RNA pairing of wild-type and artificial snR81 H/ACA guide RNA variants. A) snR81_wt 5′ hairpin pseudouridylation pocket interacting with its natural U2 snRNA target sequence. B) snR81_wt 3′ hairpin pseudouridylation pocket interacting with its natural 25S rRNA target sequence. C) 5′ hairpin variant of snR81_sub 2 base-pairing to artificial substrate RNA 2. D) 3′ variant of snR81_sub 2 base-pairing to artificial substrate RNA 2. Underlined nucleotides are those which are part of a stem according to the Boltzmann-weighted centroid structure of the H/ACA guide RNA. The red uridine is targeted for isomerization to pseudouridine.

Ultimately, while the majority of guide RNAs I designed and chose to produce/test met all the criteria, due to the nature of the internal loop sequence requiring complementarity to artificial target RNA, not every single criterion outlined here could be met in every case. In particular, an H/ACA guide RNA design can be severely compromised if the pseudouridylation pocket nucleotides on one side of the pseudouridylation pocket form base-pairs with the nucleotides on the other side, which

was the case for the snR34_sub 1 artificial guide RNA (Figure 6B, right). Thus, the design of an effective H/ACA guide RNA is at least to some extent limited by the nature of the substrate RNA sequence being targeted.

As an example of the design process, the complete design of an snR81 guide RNA engineered to target artificial substrate RNA 2 (snR81_sub2.v1) is described in detail here. To begin, the wild-type snR81 guide:target RNA-RNA pairing was examined at both the 3′ hairpin and 5′ hairpin pseudouridylation pockets, which are responsible for directing the pseudouridylation of U2 sRNA position 42 and 25S rRNA position 1052, respectively [103]. At the 5′ hairpin, the guide:target RNA-RNA pairing includes the formation of 7 base pairs on each side of the target uridine (Figure 7A). At the 3′ hairpin, the guide:target RNA-RNA pairing includes the formation of 9 base pairs downstream of the target uridine, and 3 base pairs upstream of the target uridine (Figure 7B). Next, at either the 5′ or 3′ hairpin, every nucleotide within snR81 that pairs to its natural target RNA was changed to be complementary to artificial substrate 2 (Table 1 for substrate RNA sequence). The pairing with substrate RNA 2 (sub 2) can be seen in Figures 7C and 7D for the 5′ and 3′ hairpin of the snR81_sub 2 variants, respectively.

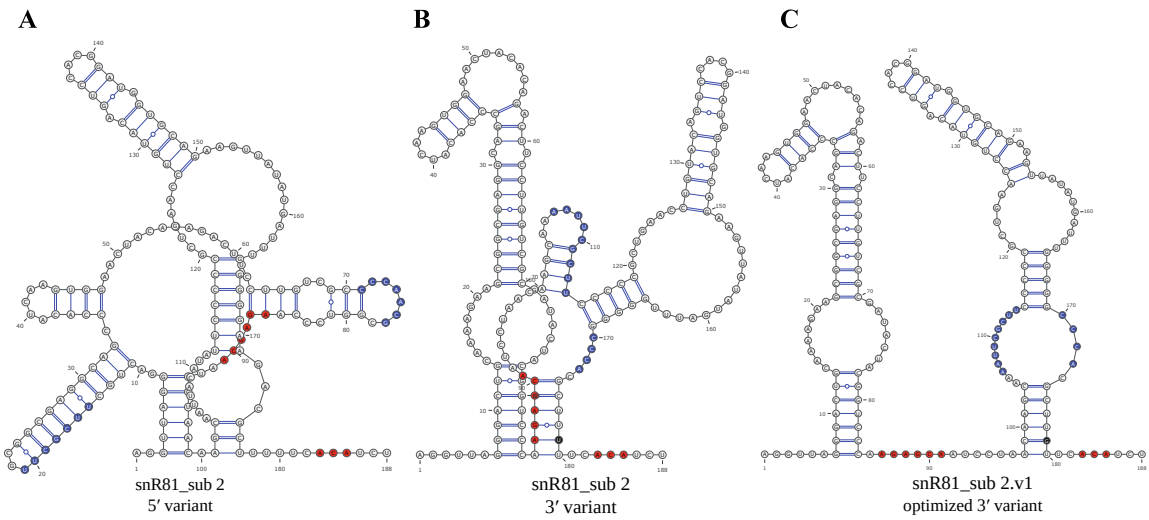


Figure 8. The Boltzmann-weighted centroid structure prediction of H/ACA guide RNAs during the design process of snR81_sub 2.v1. A) Predicted secondary structure of the 5′ variant of snR81_sub 2 pre-optimization. B) Predicted secondary structure of 3′ variant of snR34_sub 2 before optimization. C) Predicted secondary structure after optimization of the 3′ variant of snR34_sub 2, designated now as snR81_sub 2.v1. Nucleotides needed for artificial substrate RNA 2 recognition are colored blue. Conserved H box and ACA box sequences are colored orange. The U179G point-mutation needed to optimize the structure of the 3′ variant of snR81_sub 2 is colored black.

After changing the pseudouridylation pocket nucleotides, the fold of each snR81_sub 2 variant was predicted using a Boltzmann-weighted centroid structure folding algorithm. The predicted structure of the 5′ variant of snR81_sub 2 (Figure 8A) indicates that the nucleotide changes made at the 5′ pseudouridylation pocket to target substrate RNA 2 resulted in a large deviation from a canonical hairpin-hinge-hairpin structure. The formation of this structure is largely influenced by a long stretch of base pairs formed between the nucleotides of the (now artificial) 5′ pseudouridylation pocket and nucleotides which normally form the upper stem.

On the other hand, the predicted structure of the un-optimized 3′ variant of snR81_sub 2 (Figure 8B), while still imperfect, due to the conserved H box element and

3′ pseudouridylation pocket forming base pairs with other nucleotides in the structure, shares more semblance to an expected canonical H/ACA guide RNA structure and more closely resembles the predicted structure of snR81_wt, particularly at the 5′ hairpin (Figure 9A). At this step, further optimizations (described further) to the 3′ variant snR81_sub 2 structure were performed, since it already was closer to a structure that aligned with my criteria. The largest issue with the un-optimized 3′ variant of snR81_sub 2 was the formation of a small stem-loop that ties up nucleotides needed for base-pairing with the substrate RNA downstream of the target uridine (Figure 8B). To remedy this, a strategically chosen point mutation was made at a nucleotide on the right side of the 3′ lower stem (U179G – black in Figure 8). This single-nucleotide substitution destabilized the lower stem (seen in Figure 8B) and created an alternative lower stem (seen in Figure 8C), requiring the disruption of the undesirable stem that formed with the nucleotides needed for substrate RNA 2 targeting. Following this optimization, the finalized version was designated as snR81_sub 2.v1 as it aligns very well with the criteria I created to assess H/ACA guide RNA fold.

This process was followed to design twelve artificial H/ACA guide RNAs, six to target substrate RNA 1 and six to target substrate RNA 2. For each substrate, these six included two designs based on each of three different yeast guide RNAs (snR5, snR34, and snR81), and with each guide RNA, a design changing either the 5′ pseudouridylation pocket or a design changing the 3′ pseudouridylation pocket. From these twelve designs, five high-quality candidates (one of each guide RNA/substrate RNA combination, minus an artificial snR5 guide RNA targeting artificial substrate RNA 3) were selected based on how well the guide RNAs met the design criteria. This set of artificial H/ACA guide

RNAs included: snR5_sub 2.v1, sn34_sub 2.v1, snR34_sub 3.v1, snR81_sub 2.v1, snR81_sub 3.v1. To clarify naming, snR5_sub 2.v1 indicates that I changed one hairpin of snR5 to target artificial substrate RNA 2 and that this design was the first iteration (v1 = version 1; version 2 will be described below). A brief design process and the rationale behind choosing artificial guide RNAs to produce is described below.

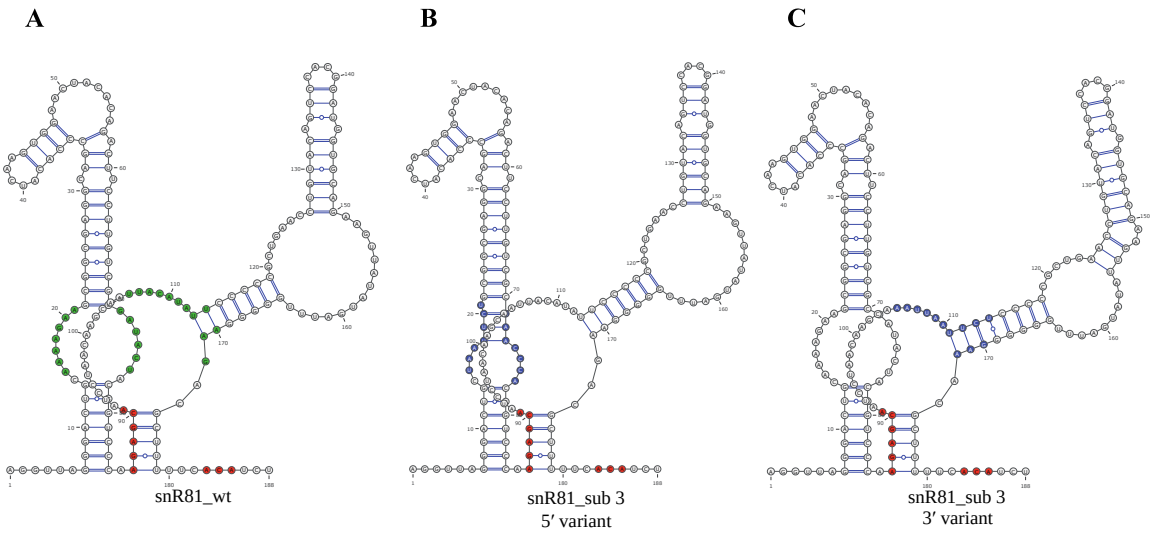


Figure 9. The Boltzmann-weighted centroid structure prediction of H/ACA guide RNAs during the design process of snR81_sub 3.v1. A) Predicted secondary structure of the snR81_wt. Nucleotides needed for wild-type target RNA pairing are colored green. B) Predicted secondary structure of 5′ variant of snR81_sub 3. C) Predicted secondary structure of the 3′ variant of snR81_sub 3. Nucleotides needed for artificial substrate RNA 3 pairing are colored blue. Conserved H box and ACA box sequences are colored orange.

For the design of a substrate 3-targeting snR81 guide RNA, two variants (one changed at each hairpin) were designed and their predicted centroid structures were determined (Figure 9B-C). Due to the complementarity of the sequence surrounding the target uridine, both of these snR81_sub 3 variants are predicted to form three base pairs that close off the top of the respective pseudouridylation pocket. Overall, both of the snR81_sub 3 variants share a similar structure with the predicted fold of snR81_wt

(Figure 9A). Since the structure of the 5′ variant (Figure 9B) resembled snR81_wt so well, and since it had more nucleotides needed for targeting substrate RNA 2 unpaired in the predicted structure than the 3′ variant (Figure 9C), the 5′ variant of snR81_sub 3 was designated as snR81_sub 3.v1 with no further optimizations.

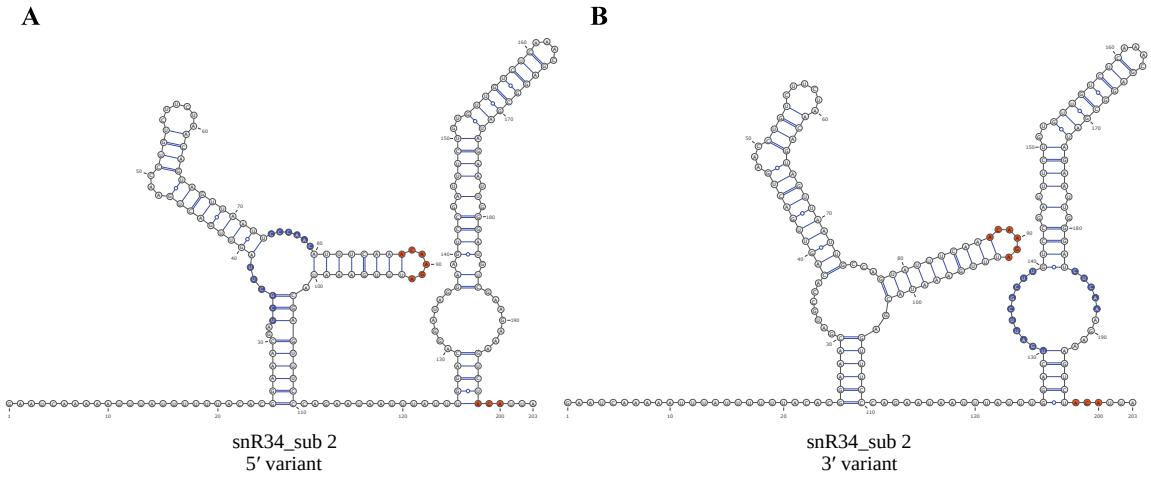


Figure 10. The Boltzmann-weighted centroid structure prediction of H/ACA guide RNAs during the design process of snR34_sub 2.v1. A) Predicted secondary structure of 5′ variant of snR34_sub 2. C) Predicted secondary structure of the 3′ variant of snR34_sub 2. Nucleotides needed for artificial substrate RNA 2 pairing are colored blue. Conserved H box and ACA box sequences are colored orange.

For artificial snR34 guide RNAs targeting substrate RNA 2, the predicted centroid structures for the two designs, the 5′ variant and the 3′ variant, can be seen in Figure 10. Between these two designs, the 3′ hairpin variant of snR34_sub 2 shares a virtually identical structure to snR34_wt (Figure 6C, left) and has every nucleotide required for substrate RNA 2 targeting unpaired within the pseudouridylation pocket. Thus, the 3′ variant of snR34_sub 2 was chosen instead of the 5′ variant and was designated as snR34_sub 2.v1 with no further optimizations.

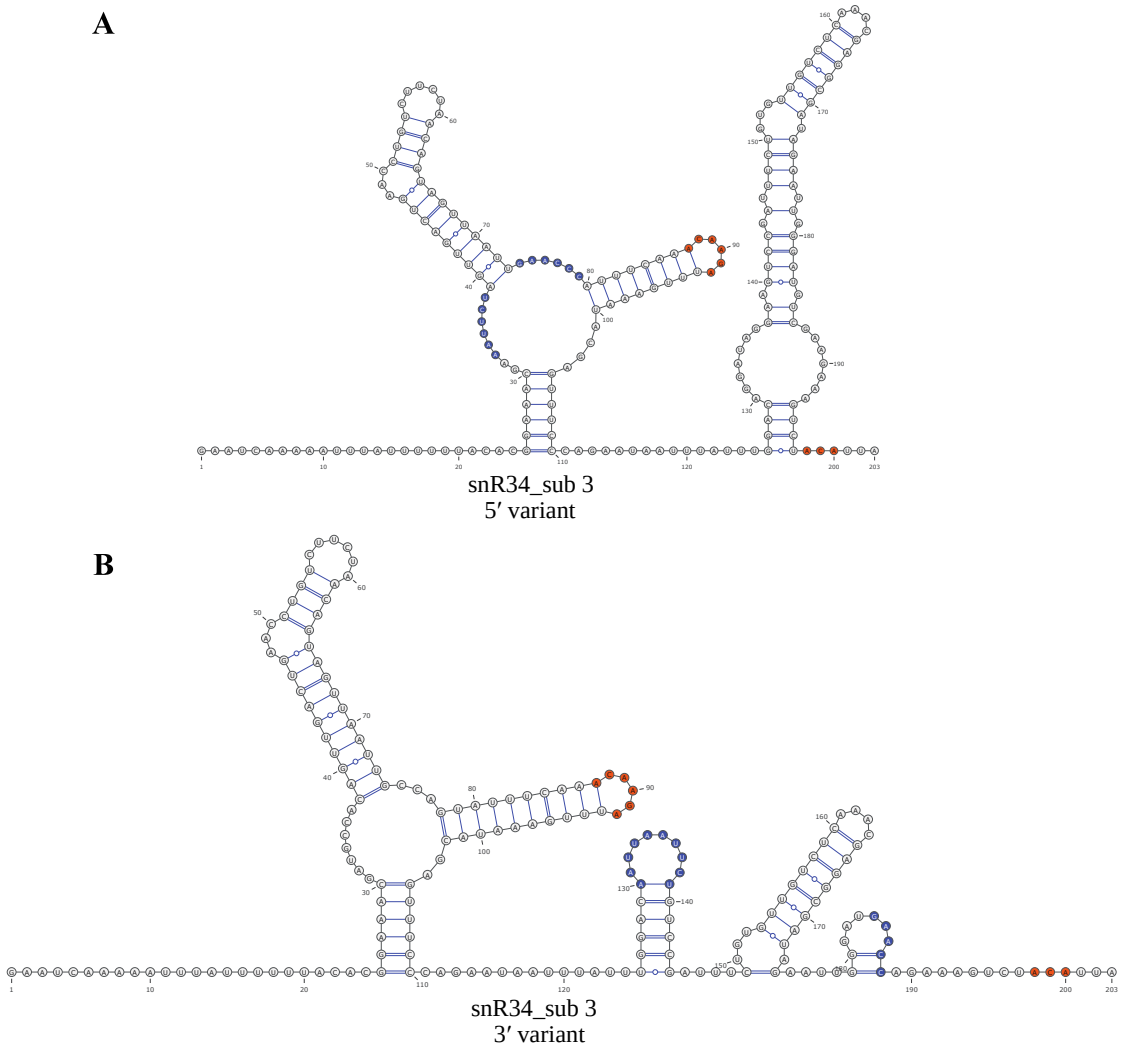


Figure 11. The Boltzmann-weighted centroid structure prediction of H/ACA guide RNAs during the design process of snR34_sub 3.v1. A) Predicted secondary structure of 5′ variant of snR34_sub 3. C) Predicted secondary structure of the 3′ variant of snR34_sub 3. Nucleotides needed for artificial substrate RNA 3 pairing are colored blue. Conserved H box and ACA box sequences are colored orange.

When designing an snR34 guide RNA to target substrate RNA 3, the 3′ variant structure prediction indicated a complete loss of the 3′ hairpin (Figure 11B), so the clear choice in this case was to produce the 5′ hairpin variant of snR34_sub 3. This variant agrees well with the structure of snR34_wt (Figure 6C, left) and so without any further optimizations, it was produced and designated as snR34_sub 3.v1.

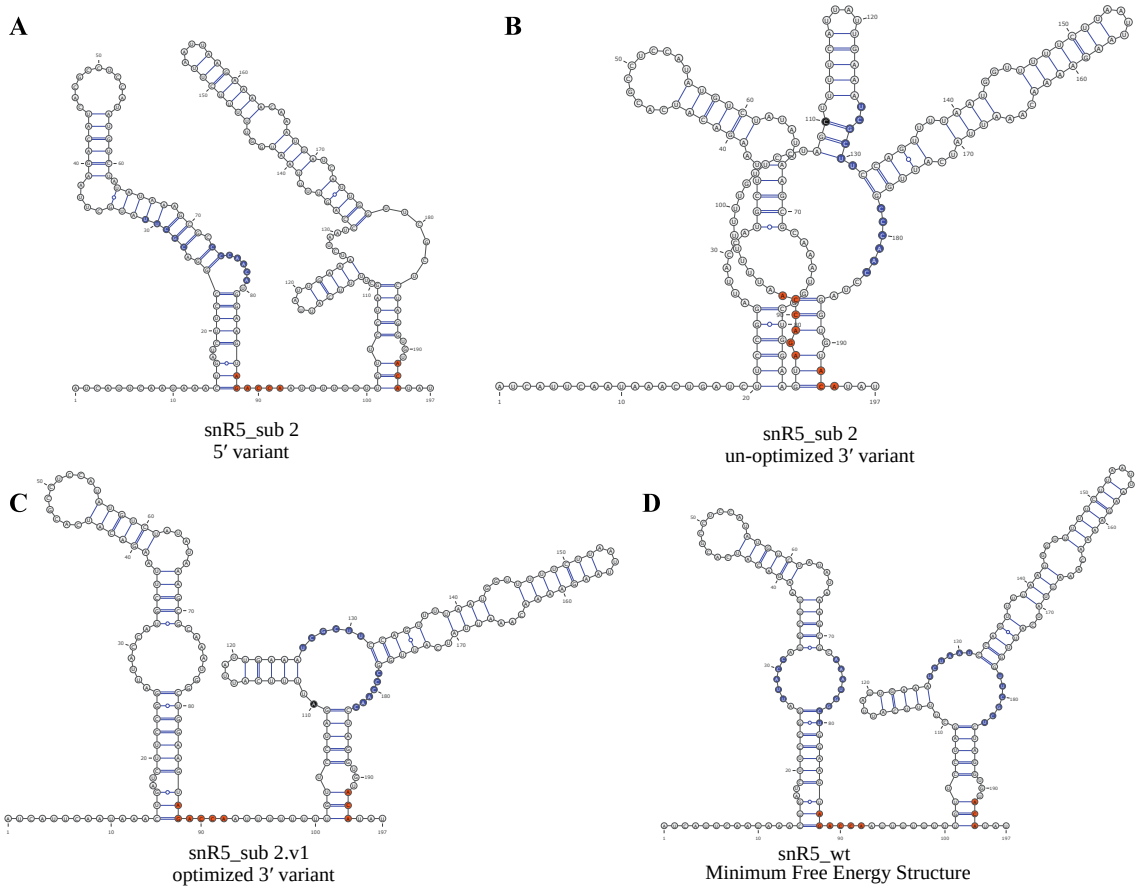


Figure 12. The minimum free energy (MFE) structure predictions of H/ACA guide RNAs during the design process of snR5_sub 2.v1. A) Predicted secondary structure of the 5′ variant of snR5_sub 2. B) Predicted secondary structure of un-optimized 3′ variant of snR5_sub 2. C) Predicted secondary structure of optimized 3′ variant of snR5_sub 2, designated as snR5_sub 2.v1. D) Predicted secondary structure of snR5_wt. Nucleotides needed for substrate RNA pairing are colored blue. Conserved H box and ACA box sequences are colored orange. The single-nucleotide substitution used to optimize the 3′ variant is colored black.

Finally, snR5 was used as a parent RNA to design a set of artificial H/ACA guide RNAs to target either artificial substrate RNA 2 or artificial substrate RNA 3. Unlike all wild-type guide RNAs examined here, the snR5 wild-type guide RNA structure is not well predicted by the Boltzmann-weighted centroid structure folding algorithm (Figure A2) in that it does not show any semblance to an expected H/ACA guide RNA structure as the second hairpin is missing. When designed at either hairpin to target substrate

RNAs 2 and 3, the structure predictions using this algorithm remained poor, so instead, the structure of artificial snR5 designs were predicted using using a minimum free energy (MFE) structure prediction (mFold webserver found at http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form - [104]), which predicted structures that included two hairpins as expected for H/ACA guide RNAs.

Using an MFE algorithm, the structure predictions for the 5′ and 3′ variant of snR5_sub 2 can be seen in Figure 12A and 11B, respectively. Attempts were first made to optimize the 5′ hairpin variant by introducing point mutations *in silico* with the goal of unwinding the long 5′ stem to free the nucleotides required for substrate RNA 2 recognition. However, these attempts were largely unsuccessful in achieving this goal. Thus, optimizations were performed on the 3′ variant of snR5_sub 2 (Figure 12B), and one point mutation (C110A – colored black in Figure 12) allowed for the formation of a structure which aligned well with my criteria. The optimized 3′ hairpin variant (Figure 12C) was produced and designated as snR5_sub 2.v1 in this study.

In order to probe the plasticity of guide:substrate RNA base-pairing, a second set of guide RNAs based on the first set was produced and designated as artificial guide RNA set v2. The set v2 guide RNAs differ from set v1 in that they only contain changes to nucleotides on one side of the original pseudouridylation pocket as opposed to both. The set v2 artificial H/ACA guide RNAs produced were: snR5_sub 2.v2, snR81_sub 2.v2, snR34_sub 3.v2, and snR81_sub 3.v2. While the structures of each set v2 guide RNA were predicted (shown and discussed later), no attempts to optimize the structures of these guide RNAs were made, even if they did not align with my criteria.

Shown below (Figure 13) is an overview of the design strategy utilized in this work to generate effective artificial H/ACA guide RNAs. With an input consisting of a wild-type H/ACA guide RNA and pseudouridylation pocket, following the process depicted, I have created a number of artificial H/ACA guide RNAs based on either snR5, snR34, and snR81 to target two different artificial substrate sequences. The process is formatted to resemble what could be a future guide RNA design algorithm.

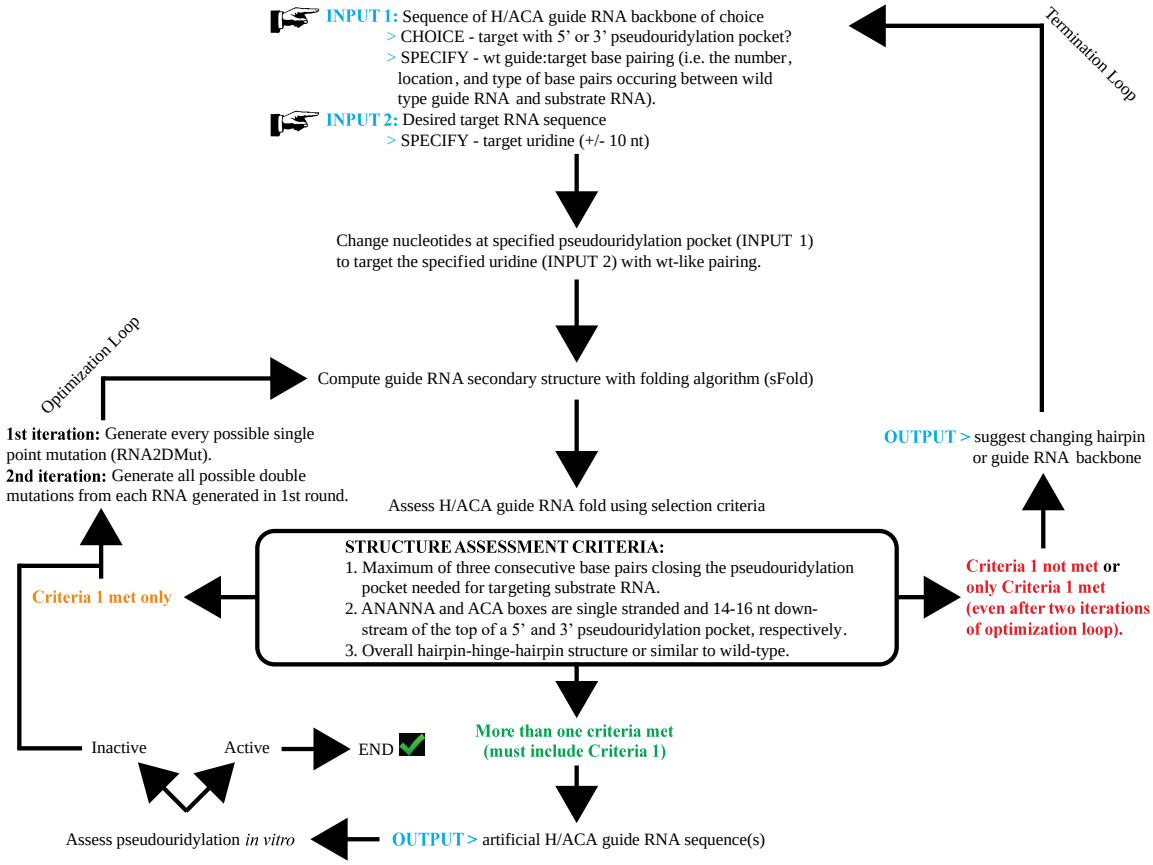


Figure 13. Algorithm for the design of artificial H/ACA guide RNAs for pseudouridylation of new substrates. In brief, based on the input of a desired target RNA and a naturally occurring H/ACA snoRNA, an artificial H/ACA guide RNA is designed based on sequence complementarity and structural criteria. Promising H/ACA snoRNA designs are then assessed for efficient pseudouridine formation *in vitro*.

3.5 Experimental evaluation of artificial H/ACA guide RNA activity for substrate 2

I wanted to determine the effectiveness of the rational H/ACA guide RNA design approach in creating artificial H/ACA guide RNAs capable of forming pseudouridine at desired substrate RNA sequences. Beginning with the set v1 guide RNAs designed to target substrate RNA 2 (snR5_sub 2.v1, snR34_sub 2.v1, and snR81_sub 2.v1), these guide RNAs were reconstituted *in* vitro with core H/ACA sRNP components and tested for activity against [^{3}H-C5] uridine-containing substrate RNA 2 using a tritium release assay (see methods for details). Additionally, each guide RNA's v2 counterpart (snR5_sub 2.v2, snR34_sub 2.v2, and snR81_sub 2.v2) was also tested for activity and evaluated alongside set v1 guide RNAs to evaluate how sub-optimal folding and base pairing with substrate RNA could affect pseudouridylation. All reactions with set v1 guide RNAs were performed in biological triplicate (duplicate for set v2 guide RNAs) at 30°C under multiple turnover conditions (500 nM substrate RNA 2; 100 nM proteins (Cbf5-Nop10-Gar1-Nhp2); 45 nM H/ACA guide RNA), and pseudouridine formation was determined after 2.5, 5, 10, 25, 45, and 90 minutes.

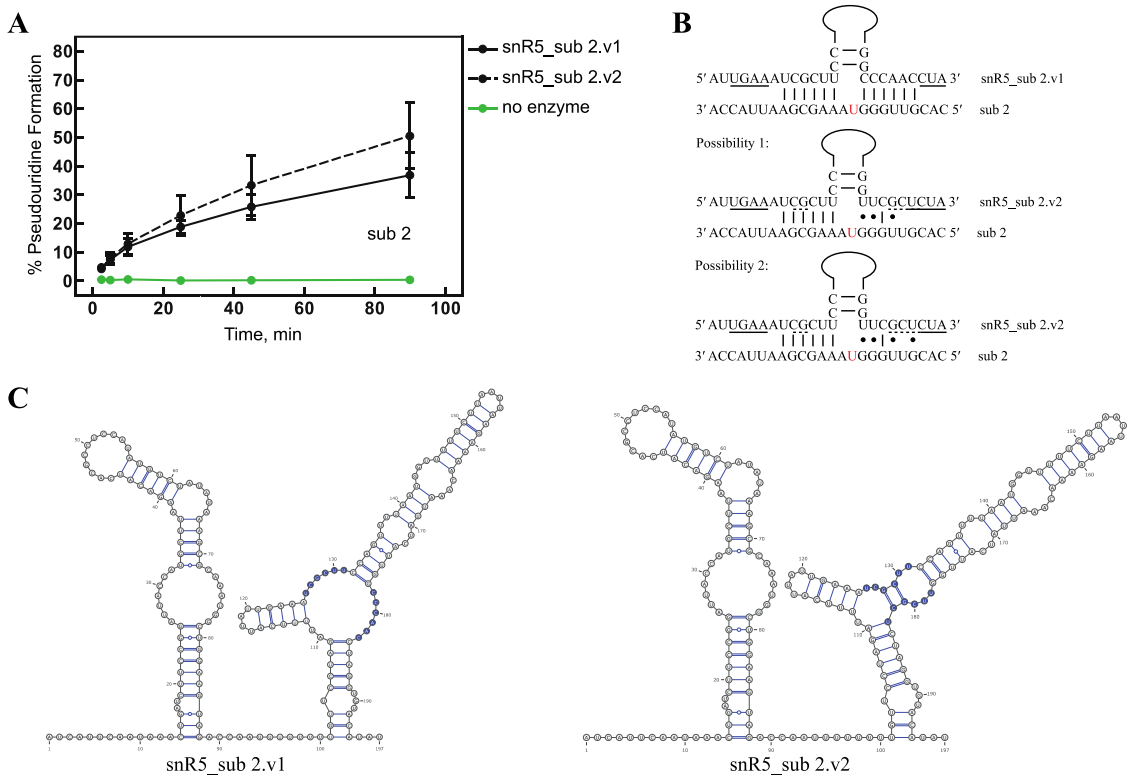


Figure 14. Pseudouridylation activity of artificial snR5_sub 2 H/ACA sRNPs for substrate RNA 2. A) Tritium release assay determining the activity of reconstituted snR5_sub 2.v1 (solid line) and snR5_sub 2.v2 (broken line) H/ACA sRNPs for artificial substrate RNA 2. Reactions were performed under multiple turnover conditions (500 nM substrate RNA 2; 45 nM H/ACA sRNP). B) The putative base pairing possibilities of snR5_sub 2.v1 and snR5_sub 2.v2 guide RNA pseudouridylation pockets for artificial substrate RNA 2. Two possibilities for snR5_sub 2.v2 can be envisioned which are both shown. Underlined nucleotides in the guide RNA are nucleotides which are part of a stem/base pair according to structure prediction while dashed underlines indicate nucleotides whose pairing needed to be disrupted to achieve the suggested outcome. The red uridine is isomerized. C) The Boltzmann-weighted centroid structural prediction of snR5_sub 2.v1 (left) and snR5_sub 2.v2 (right). Nucleotides in the 3′ pseudouridylation pocket needed for targeting substrate RNA 2 are highlighted in blue.

Relative to a no enzyme (negative) control, which should only detect baseline tritium release due to random proton exchange of tritium with surrounding water molecules, tritium release from substrate RNA 2 in the presence of an snR5_sub 2.v1 H/ACA sRNP was significant after just 2.5 minutes (Figure 14A, solid lines). After 90 minutes, the artificial set v1 H/ACA sRNP reached greater than 30% pseudouridylation,

while the no enzyme control remained at baseline throughout the reaction. To investigate how sub-optimal base pairing with substrate RNA will affect pseudouridylation, snR5_sub 2.v2 sRNP pseudouridylation activity was determined and compared to the snR5_sub 2.v1 H/ACA sRNP. Interestingly, snR5_sub 2.v2 also directed pseudouridylation of substrate RNA 2, again with significantly higher pseudouridine formation observed versus the no enzyme control after just 2.5 minutes (Figure 14A, broken line). Although the fold of set v2 guide RNAs was not strictly controlled like the set v1 guide RNAs, structure prediction of snR5_sub 2.v2 suggests a nearly identical overall structure to its counterpart – snR5_sub 2.v1 (Figure 14C). Compared to the structure of snR5_sub 2.v1 (Figure 14C, left), the structure of snR5_sub 2.v2 (Figure 14C, right) indicates that the 3′ hairpin pseudouridylation pocket is closed by two G-C base pairs which would need to be broken to target substrate RNA 2. Unexpectedly, snR5_sub 2.v2 pseudouridine formation (Figure 14B, broken line) was higher than its v1 counterpart, particularly after 10 minutes of reaction time. This was a very striking result considering that the set v2 guide RNA was produced with the expectation of suboptimal pairing to substrate (assuming wild-type snR5_wt:substrate RNA pairing as optimal).

Taking into account the structure predictions of both snR5_sub 2 guide RNAs, a number of base pairing possibilities between each guide RNA and substrate RNA 2 were manually predicted (Figure 14B). snR5_sub 2.v1 was designed to maintain the same number and location of base-pairing to substrate RNA 2 as observed in wild-type snR5 3′ hairpin pseudouridylation pocket and thus is predicted to form a total of 12 base pairs with substrate RNA, 6 on each side of the pseudouridylation pocket (Figure 14B, top). The set v2 guide RNA differs in sequence from its v1 counterpart only on the right side

of the 3′ hairpin pseudouridylation pocket, and thus, we expect no differences between base pairing on the left of the pseudouridylation pocket in either of the predicted pairings (1 and 2). However, despite a very different sequence on the right side of the pseudouridylation pocket, many G-U wobble pairs appear to replace Watson-Crick base pairs seen in the snR5_sub 2.v1:substrate 2 interaction. The “best case scenario” (Figure 14B, possibility 2) requires the breakage of a single base pair at the top of the lower stem to allow a uridine in the guide RNA upper stem to be available to pair with a guanine in substrate RNA 2. In this case, a net loss of only one base pair occurs compared to snR5_sub 2.v1:substrate RNA 2 pairing.

Next, the *in vitro* pseudouridylation activity of the snR34_sub 2 artificial H/ACA guide RNAs for substrate RNA 2 was examined. The snR34_sub 2.v1 sRNP showed significant activity for substrate RNA 2 compared to the no enzyme control (Figure 15A, solid lines). However, in this case the substrate 2-targeting v2 guide RNA, snR34_sub 2.v2, was unable to guide modification of substrate RNA 2 (Figure 15A, broken line). From the predicted centroid structure, the snR34_sub 2.v2 guide RNA appears to form multiple base pairs within the 3′ pseudouridylation pocket causing a partial closure of the pocket not seen in the snR34_sub 2.v1 counterpart (Figure 15C). These base pairs mimic what was observed earlier in the predicted structure of snR34_sub 1 (Figure 6C for reference).

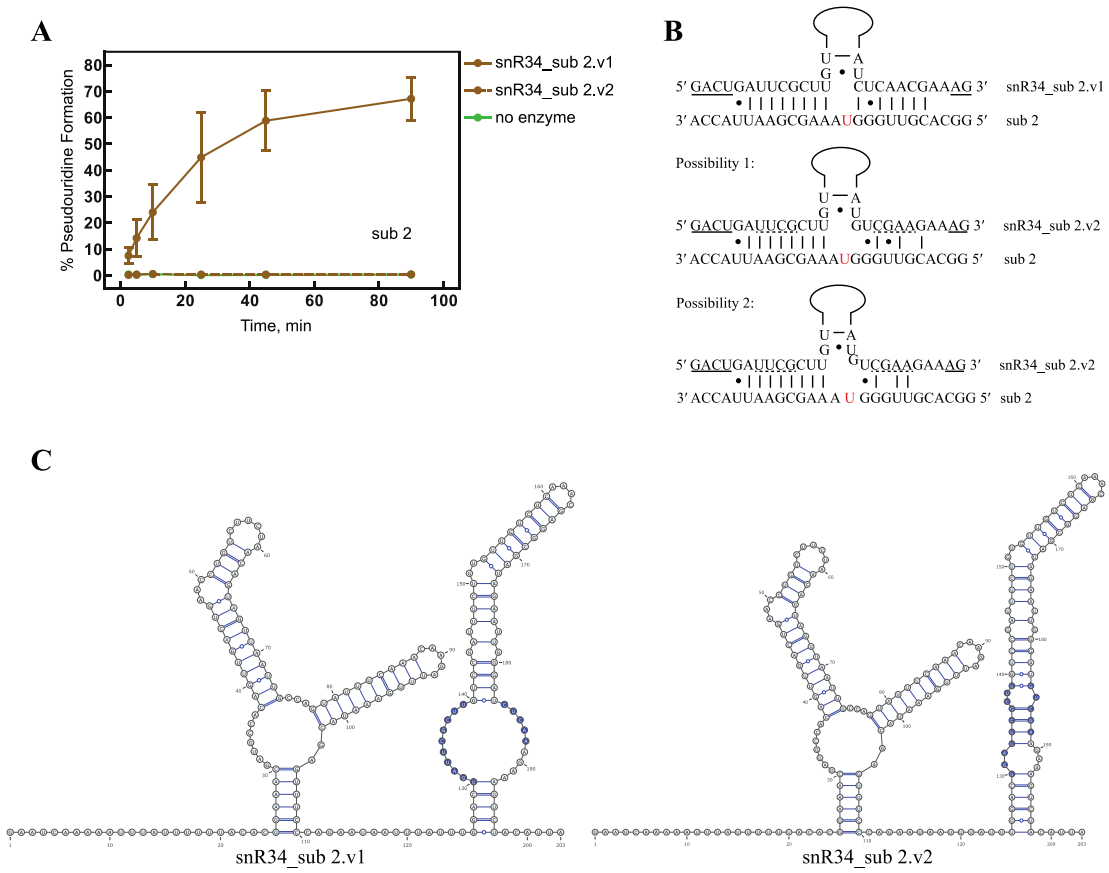


Figure 15. Pseudouridylation activity of artificial snR34_sub 2 H/ACA sRNPs for substrate RNA 2. A) Tritium release assay determining the activity of reconstituted snR34_sub 2.v1 (solid line) and snR34_sub 2.v2 (broken line) H/ACA sRNPs for artificial substrate RNA 2. Reactions were performed under multiple turnover conditions (500 nM substrate RNA 2; 45 nM H/ACA sRNP). B) The putative base pairing possibilities of snR34_sub 2.v1 and snR34_sub 2.v2 guide RNA pseudouridylation pockets for artificial substrate RNA 2. Underlined nucleotides in the guide RNA are nucleotides which are part of a stem/base pair according to structure prediction while dashed underlines indicate nucleotides whose pairing needed to be disrupted to achieve the suggested outcome. The red uridine is isomerized. C) The Boltzmann-weighted centroid structural prediction of snR34_sub 2.v1 (left) and snR34_sub 2.v2 (right). Nucleotides in the 3′ pseudouridylation pocket needed for targeting substrate RNA 2 are highlighted in blue.

Although snR34_sub 2.v2 was inactive when tested for pseudouridylation activity for substrate RNA 2, the base pairing possibilities between snR34_sub 2.v2 and substrate RNA 2 remained interesting to predict (Figure 15B) so that one may speculate as to why activity may not have been observed. If the four internal base pairs closing the artificial

3′ pseudouridylation pocket of snR34_sub 2.v2 were disrupted, several base pairs between guide and substrate RNA could be formed upstream of the target uridine, with a net loss of just 2-3 base pairs compared to the snR34_sub 2.v1:sub 2 interaction predicted in possibility 1 (Figure 15B, top). It is worth noting that although possibility 1 suggests more base pairs formed between snR34_sub 2.v2 and sub 2, possibility 2 maintains a base pair immediately adjacent to the target uridine through a G-U wobble pair. I cannot distinguish whether the lack of activity of the snR5_sub 2.v2 sRNP for substrate RNA 2 is due to the inability to create a single-stranded pseudouridylation pocket, the loss of base pairing, the absence of the base pair adjacent to the target uridine, or some combination of the three.

Finally, the artificial snR81 guide RNAs designed to target substrate RNA 2 were tested for pseudouridylation activity with substrate RNA 2. Again, the artificial v1 guide RNA which I designed displayed pseudouridylation activity for the artificial substrate RNA (Figure 16A, solid line). On the other hand, the snR81_sub 2.v2 sRNP showed no activity for substrate RNA 2 (Figure 16A, broken line). Similar to snR34_sub 2.v2 and snR34_ sub 1, snR81_sub 2.v2 is predicted to fold into a structure with a partially closed pseudouridylation pocket (Figure 16C, right), although the effect is not as drastic (3 closing base pairs instead of 4). However, even if these base pairs were broken, there is limited options for snR81_sub 2.v2 to base pair with substrate RNA 2 upstream of the target uridine (Figure 16B, bottom). The base pairing possibility shown is envisioned to be the “best-case scenario” since it forms the greatest number of base pairs upstream of the target uridine. Despite attempting to predict pairings at each uridine within substrate

RNA 2, there was no base pairing possibility which retained the base pair adjacent to the target uridine.

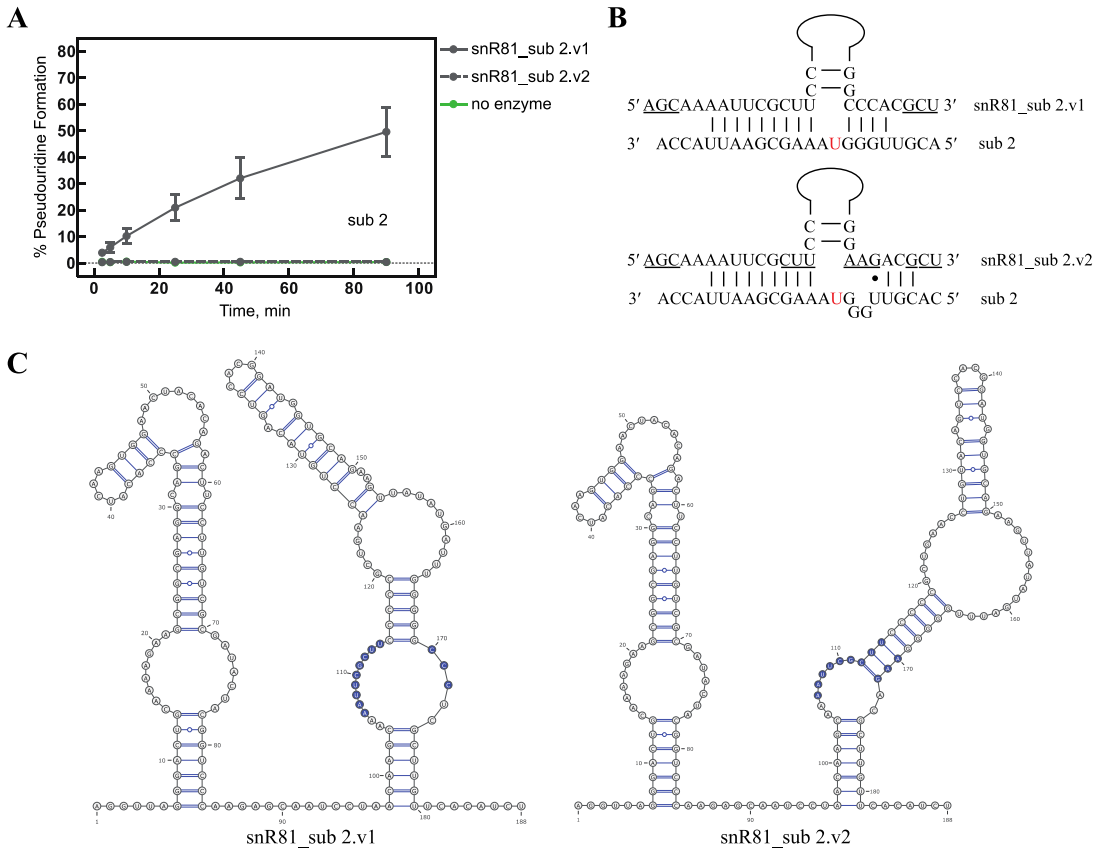


Figure 16. Pseudouridylation activity of artificial snR81_sub 2 H/ACA sRNPs for substrate RNA 2. A) Tritium release assay determining the activity of reconstituted snR81_sub 2.v1 (solid line) and snR81_sub 2.v2 (broken line) H/ACA sRNPs for artificial substrate RNA 2. Reactions were performed under multiple turnover conditions (500 nM substrate RNA 2; 45 nM H/ACA sRNP). B) The putative base pairing possibilities of snR81_sub 2.v1 and snR81_sub 2.v2 guide RNA pseudouridylation pockets for artificial substrate RNA 2. Underlined nucleotides in the guide RNA are nucleotides which are part of a stem/base pair according to structure prediction while dashed underlines indicate nucleotides whose pairing needed to be disrupted to achieve the suggested outcome. The red uridine is isomerized. C) The Boltzmann-weighted centroid structural prediction of snR81_sub 2.v1 (left) and snR81_sub 2.v2 (right). Nucleotides in the 3′ pseudouridylation pocket needed for targeting substrate RNA 2 are highlighted in blue.

3.6. Experimental evaluation of artificial H/ACA guide RNA activity for substrate 3

I also wanted to establish how effective the rationally designed substrate 3-targeting guide RNAs (snR34_sub 3.v1 and snR81_sub 3.v1) were for their respective substrate RNA (substrate 3 – see Table 1). Additionally, snR81_sub 3.v1 activity was compared to the activity of its v2 counterpart – snR81_sub 3.v2; snR34_sub 3.v2 was not produced, and thus is not tested. The substrate 3-targeting guide RNAs were tested for pseudouridylation activity against substrate RNA 3 under identical conditions as described earlier for the substrate 2-targeting guide RNAs.

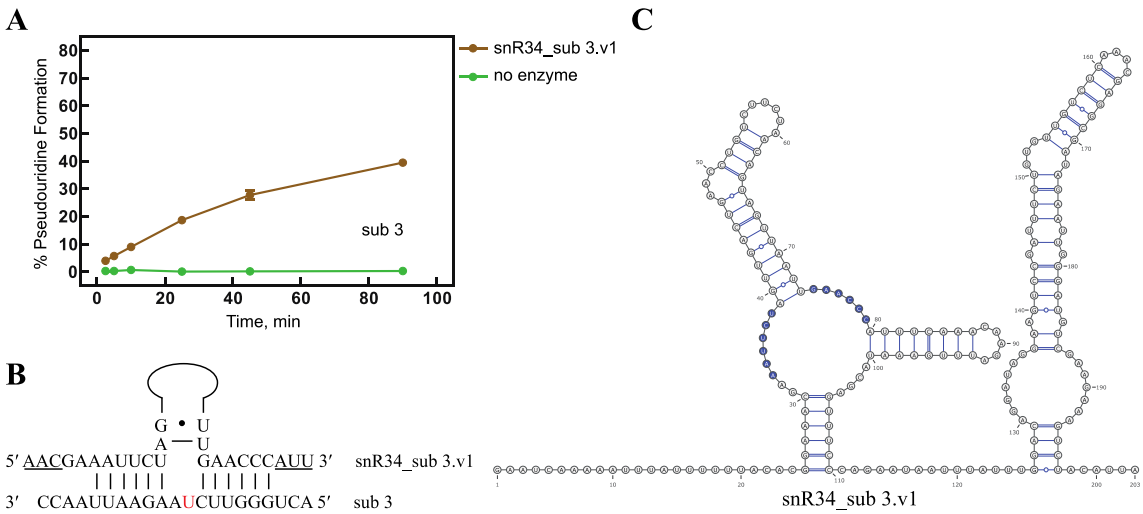


Figure 17. Pseudouridylation activity of artificial snR34_sub 3.v1 H/ACA sRNP for substrate RNA 3. A) Tritium release assay determining the activity of a reconstituted snR5_sub 3.v1 (brown) H/ACA sRNP for artificial substrate RNA 3. Reactions were performed under multiple turnover conditions (500 nM substrate RNA 3; 45 nM H/ACA sRNP). B) The putative base pairing of the snR34_sub 3.v1 guide RNA 5′ pseudouridylation pocket for artificial substrate RNA 3. Underlined nucleotides in the guide RNA are nucleotides which are part of a stem/base pair according to structure prediction. The red uridine is isomerized. C) The Boltzmann-weighted centroid structural prediction of snR34_sub 3.v1. Nucleotides in the 5′ pseudouridylation pocket needed for targeting substrate RNA 3 are highlighted in blue.

Similar to all three of the substrate 2-targeting v1 guide RNAs, snR34_sub 3.v1 had significantly more pseudouridine formation after just 2.5 mins of reaction time with substrate RNA 3 versus the no enzyme control, reaching about 40% pseudouridine formation after 90 min (Figure 17A, solid line). This result agrees with both the predicted structure of snR34_sub 3.v1 (Figure 17C) which was designed to align to my design criteria and also with the predicted base pairing between guide RNA and substrate RNA (Figure 17B) which too, was designed to be "optimal" such that it resembled the pairing seen between wild-type snR34 and its target at the 5′ pseudouridylation pocket.

It is interesting to consider that although their structures are virtually identical, snR34_sub 2.v1 shows greater activity for its respective substrate RNA (Figure 15A) than snR34_sub 3.v1 for substrate RNA 3. While this could indicate a potential issue with either the artificial substrate RNA 3 or with snR34_sub 3.v1, this could also be explained by the fact that these guide RNAs are using opposite hairpins to target their respective substrate RNAs. In our group, it has been observed that for snR34, the 3′ hairpin consistently outperforms the 5′ hairpin in tritium release assays [60].

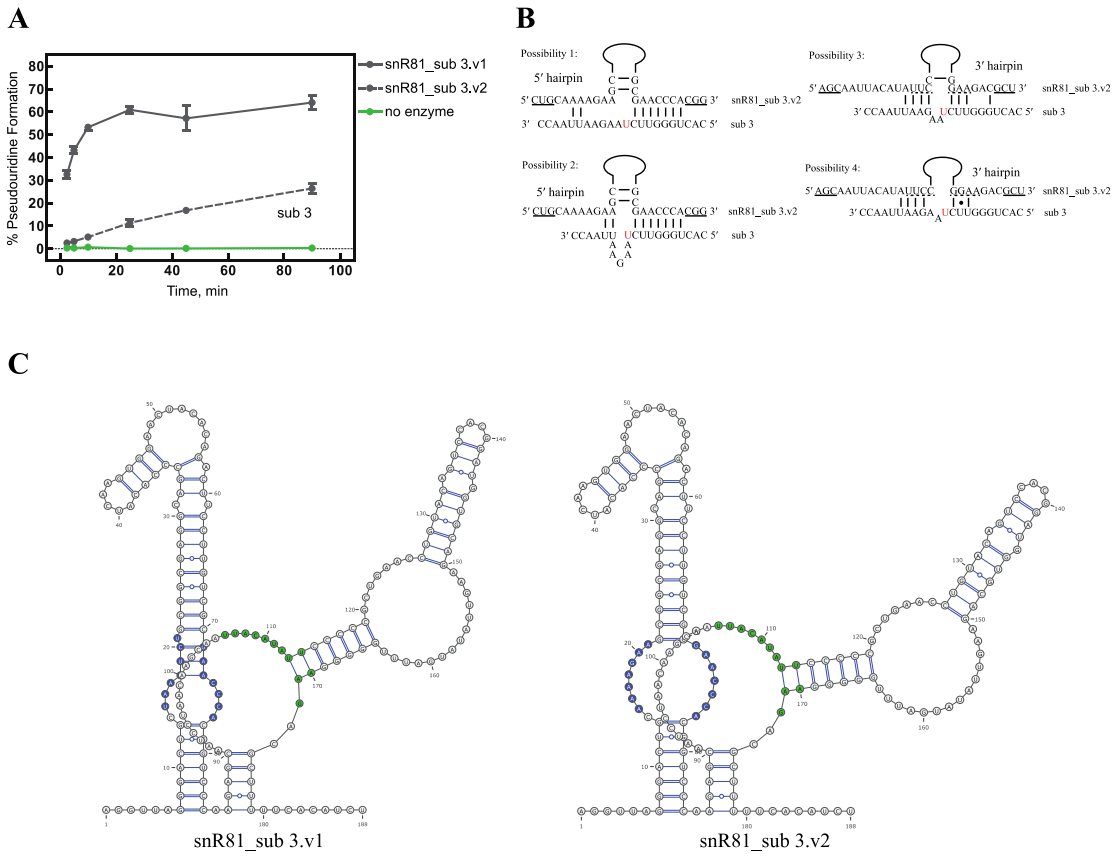


Figure 18. Pseudouridylation activity of artificial snR81_sub 3 H/ACA sRNPs for substrate RNA 3. A) Tritium release assay determining activity of reconstituted snR81_sub 3.v1 (solid line) and snR81_sub 3.v2 (broken line) H/ACA sRNPs for artificial substrate RNA 3. Reactions were performed under multiple turnover conditions (500 nM substrate RNA 3; 45 nM H/ACA sRNP). B) The putative base pairing possibilities of snR81_sub 3.v1 and snR81_sub 3.v2 guide RNA pseudouridylation pockets for artificial substrate RNA 3. Multiple possibilities for snR81_sub 3.v2 pairing with substrate 3 can be envisioned which are all shown; possibilities 1 and 2 suggest target recognition is through the 5′ hairpin while possibilities 3 and 4 suggest target recognition is though the 3′ hairpin. Underlined nucleotides in the guide RNA are nucleotides which are part of a stem/base pair according to structure prediction while dashed underlines indicate nucleotides whose pairing needed to be disrupted to achieve the suggested outcome. The red uridine is isomerized. C) The Boltzmann-weighted centroid structural prediction of snR81_sub 3.v1 (left) and snR81_sub 3.v2 (right). Nucleotides in the 5′ pseudouridylation pocket needed for targeting artificial substrate RNA 3 are highlighted in blue and nucleotides in the 3′ pseudouridylation pocket needed for targeted the natural 3′ pseudouridylation pocket 25S rRNA substrate are highlighted in green.

Comparing snR81_sub 3.v1 to its v2 counterpart (snR81_sub 3.v2), pseudouridylation of substrate RNA 3 is able to be directed by both guide RNAs (Figure 18A). These two guide RNAs differ from one another in sequence only at the 5′ pseudouridylation pocket, where the left side in snR81_sub 3.v1 has a wild-type sequence. Although the v1 counterpart was designed to be structurally as ideal as possible (based on my design criteria), the different sequence on the left side of the 5′ pseudouridylation pocket results in snR81_sub 3.v2 forming a more optimal fold when assessed using my criteria (Figure 18C). Specifically, the artificial 5′ pseudouridylation pocket of snR81_sub 3.v2 is entirely single-stranded, whereas and in the v1 guide RNA (Figure 18C, right) this pocket contains three base pairs that extend the upper stem (Figure 18C, left). Manually predicting the pairing possibilities between the artificial 5′ hairpin of snR81_sub 3.v2 guide RNA and substrate RNA 3 (Figure 18B – possibility 1 and 2) reveals a major loss of base pairing downstream of the target uridine with only 2 base pairs being formed in any predicted scenario. Across the entire pseudouridylation pocket, 5 base pairs are lost when compared to snR81_sub 3.v1:sub 3 pairing. Not only is there very minimal pairing downstream of the target uridine in these predictions, but there are also many additional unpaired nucleotides adjacent to the target uridine (Figure 3, green), which could have severe implications on the productivity of the interaction.

Considering that neither predicted guide:substrate pairing appears convincing, the activity of the snR81_sub 3.v2 guide for substrate 3 was puzzling and prompted an investigation into the possibility that modification may be occurring through the 3′ pseudouridylation pocket, which remained wild-type in both snR81_sub 3 variants. Unexpectedly, a number of base pairing scenarios (Figure 18B – possibilities 3 and 4)

can be envisioned if as few as 3 base pairs at the base of the 3′ hairpin upper stem are broken. In fact, in order for snR81_wt to guide pseudouridylation of position 1052 in 25S rRNA substrate with the 3′ hairpin, two of these 3-4 base pairs (two A-U pairs at the top of the pseudouridylation pocket) are normally broken to allow for pairing, suggesting that the upper stem of the snR81 3′ hairpin may be prone to remodeling/unwinding. Although base pairing is far less extensive on the right side of the pseudouridylation pocket when comparing possibilities 1 and 2 versus 3 and 4 (7 bp vs 3-4 bp), possibilities 1 and 2 have less extensive pairing on the left side of the pseudouridylation pocket (2 bp vs 4 bp). Possibilities 3 and 4 also have an advantage over possibilities 1 and 2 in that they reduce the number of unpaired nucleotides downstream of the target uridine from five to just two. Amongst all wild-type yeast guide:substrate RNA pairings, there exists none with more than 2 unpaired nucleotides downstream of the target uridine.

3.7. Evaluation of artificial H/ACA guide RNA activity for a different substrate

Lastly, I wanted to examine the specificity of the v1 artificial H/ACA guide RNAs and was also interested in establishing how a few changes to one side of the pseudouridylation pocket (i.e. set v2 guide RNAs) impacted specificity. To investigate these questions, each set v1 and set v2 H/ACA guide RNA used in this study was reconstituted *in* vitro with core H/ACA sRNP components and tested for activity (in duplicate) against the [^{3}H-C5] uridine-containing substrate RNA for which it was not designed for (e.g. sub 2.v1 guide RNA for sub 3).

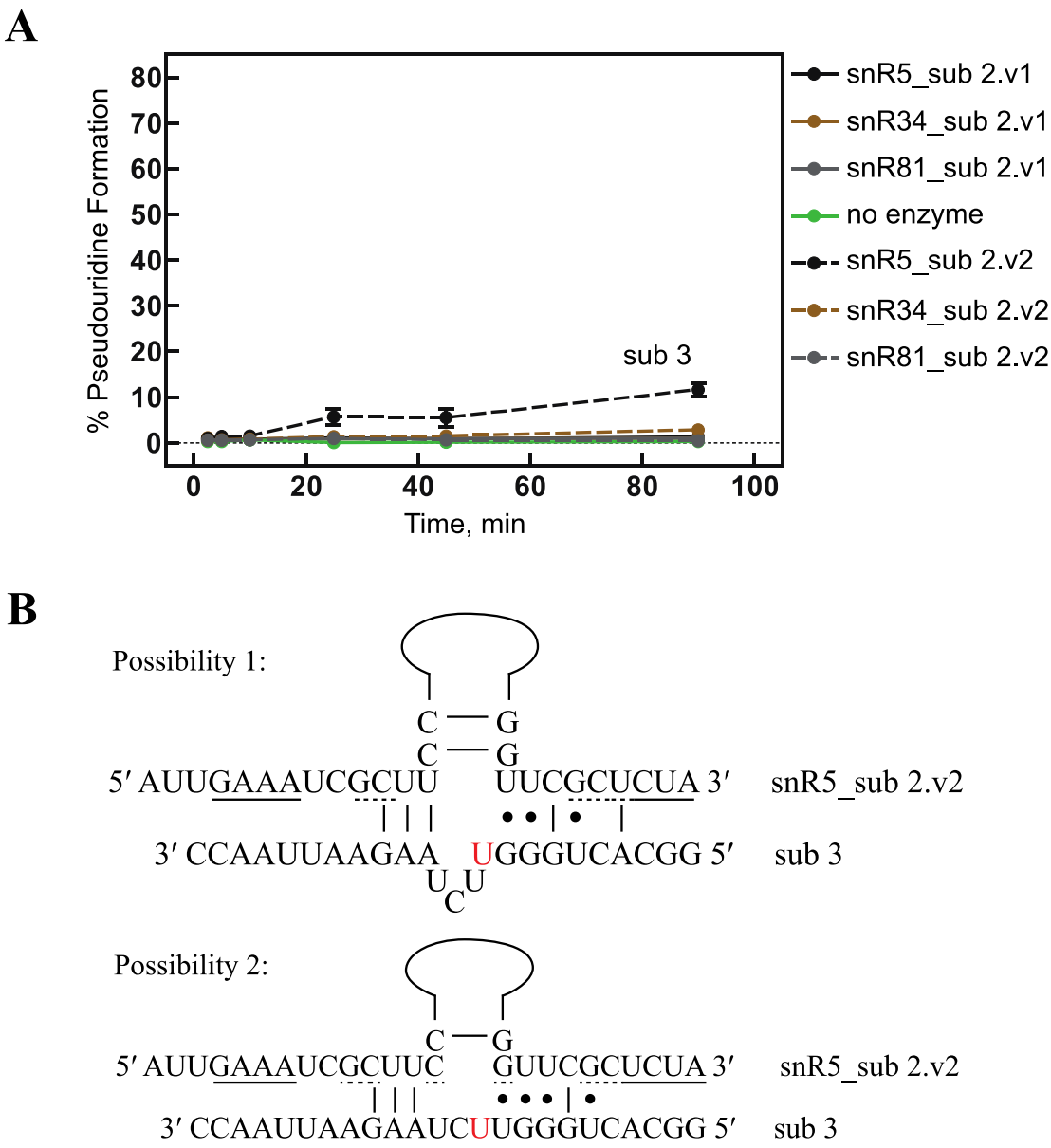


Figure 19. Pseudouridylation activity in substrate RNA 3 by H/ACA sRNPs harboring artificial guide RNAs designed for substrate RNA 2. A) Tritium release assay using substrate RNA 3 determining the activity of reconstituted H/ACA sRNPs assembled with each of the six artificial H/ACA guide RNAs designed to target artificial substrate RNA 2. Reactions were performed under multiple turnover conditions (500 nM substrate 3; 100 nM proteins (Cbf5-Gar1-Nop10-Nhp2); 45 nM H/ACA guide RNA). Solid lines correspond to set v1 guide RNAs while broken lines correspond to set v2 guide RNAs. B) The putative base pairing possibilities of snR5_sub 2.v2 guide RNA pseudouridylation pocket for artificial substrate RNA 3. Underlined nucleotides in the guide RNA are nucleotides which are part of a stem/base pair according to structure prediction while dashed underlines indicate nucleotides whose pairing needed to be disrupted to achieve the suggested outcome. The red uridine is isomerized.

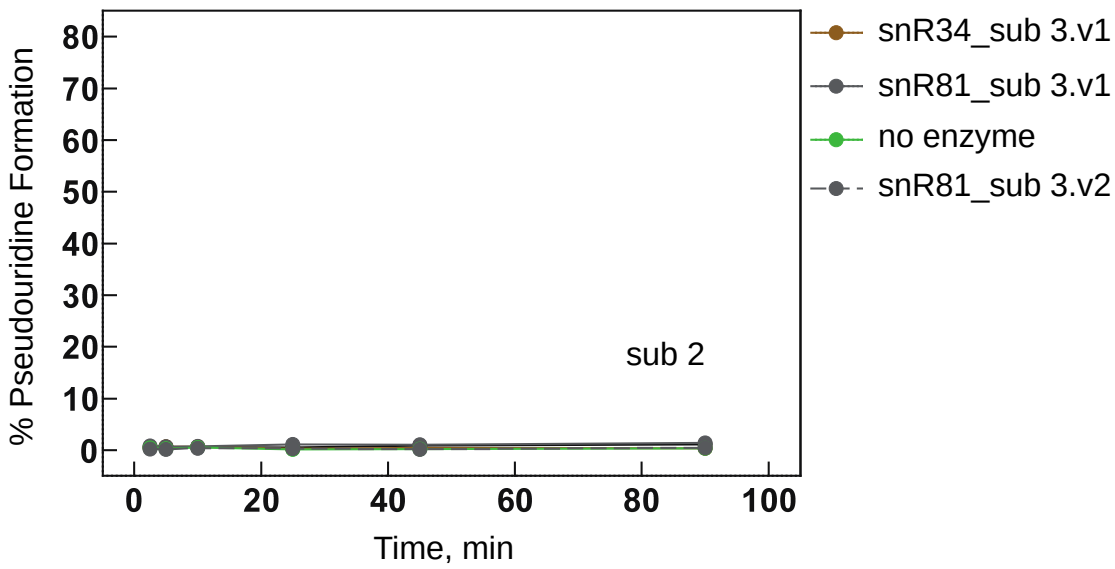


Figure 20. Pseudouridylation activity in substrate RNA 2 by H/ACA sRNPs harboring artificial guide RNAs designed for substrate RNA 3. Tritium release assay using substrate RNA 2 determining the activity of reconstituted H/ACA sRNPs assembled with each of the three artificial H/ACA guide RNAs designed to target artificial substrate RNA 3. Reactions were performed under multiple turnover conditions (500 nM substrate 2; 100 nM proteins (Cbf5-Gar1-Nop10-Nhp2); 45nM H/ACA guide RNA). Solid lines correspond to set v1 guide RNAs while broken lines correspond to set v2 guide RNAs.

Considering that the artificial substrate RNA sequences are different from one another, I expected that if H/ACA-guided pseudouridylation is highly sequence specific, then pseudouridine formation should not be observed when tested with the substrate RNA which the guide RNA was not designed for. Indeed, all artificial set v1 guide RNAs were unable to direct the pseudouridylation of the substrate RNA they were not designed for (Figure 19A and 18, solid lines). However, when the set v2 artificial guide RNAs were tested for activity against the substrate their v1 counterpart was not designed for, activity was achieved in one combination – snR5_sub 2.v2 for substrate RNA 3 (Figure 19A, broken black line). Although very inefficient in comparison to all other productive

guide RNA:substrate pairings in this study, the aforementioned combination results in a low, but significant, level of pseudouridine formation, reaching about 10% pseudouridylation after 90 minutes. With this guide RNA, a clear increase from baseline begins to emerge between 10 and 25 minutes. It is interesting to consider that snR5_sub 2.v2 was also the most active of the set v2 guide RNAs when tested for the substrate its v1 counterpart (snR5_sub 2.v1) was designed for. This result demonstrates a situation where one H/ACA guide RNA pseudouridylation pocket is active for more than one substrate RNA sequence, indicating at least some level of H/ACA guide RNA targeting flexibility. However, this is not a novel observation. In fact, several yeast H/ACA guide RNAs are suggested to target two slightly different RNA sequences with a single pseudouridylation pocket and, in one case (snR49) up to 3 sequences with one pseudouridylation pocket; in this case, the target sequences are quite similar to one another [72].

To explain the activity observed with snR5_sub 2.v2, I predicted base pairing possibilities between snR5_sub 2.v2 and substrate RNA 3, depicted in Figure 19B. Notably, the pairing predictions I envisioned result in a different uridine in the substrate RNA being targeted; pseudouridylation activity using the tritium release assay is unspecific since the [^{3}H]-UTP can be incorporated at any positions in the substrate during *in vitro* transcription. Due to the predicted snR5_sub 2.v2 secondary structure having two G-C base pairs closing the internal loop (Figure 14B, right), in both pairing possibilities these base pairs were assumed to be broken to allow for substrate RNA access to pseudouridylation pocket nucleotides. Possibility 1 retains the G-C base pair at the bottom of the upper stem, which is broken in possibility 2 to allow for a different

pairing combination. Both pairing possibilities maintain the 8 minimum base pairs needed for activity. While possibility 1 predicts that there are 3 unpaired nucleotides aside from the target uridine in the pseudouridylation pocket, possibility 2 suggests only 2 are present, which agrees more with typical natural guide RNA:substrate RNA interactions. Furthermore, pairing possibility 2 has a longer stretch of continuous base pairs which would result in increased base stacking on the right side of the pseudouridylation pocket; discontinuous base pairings have previously been shown by our group to negatively influence H/ACA guide RNA:target RNA interactions likely as a result of the loss of important base-stacking interactions [97]. I cannot say conclusively which pairing possibility is responsible for providing the pseudouridylation activity for substrate RNA 3, and realistically, it could be the case that a combination of both possibilities contribute, at least to some extent.

Chapter 4 - Discussion

4.1 – Overview of the major contributions towards project objectives

The goal of this study was to develop a deeper understanding of H/ACA guide RNA features, such as structure, function, and specificity, and how all these factors together influence H/ACA sRNP guided pseudouridylation. Recently, there have been two studies investigating specific H/ACA guide RNA features and their implications for H/ACA sRNP guided pseudouridylation [60, 62]. These studies examined *in vivo* guide:substrate RNA-RNA interaction limitations in addition to how H/ACA guide RNA features (stem lengths, box elements, etc.) affect pseudouridylation activity and assembly *in vitro*. My thesis provides a unique perspective by having greater emphasis on the relationship between H/ACA guide RNA fold, pseudouridylation pocket architecture, and the functional implications for pseudouridylation activity. Here, I have designed and produced several artificial H/ACA guide RNAs and guide RNA variants and evaluated their fold and pseudouridylation activities across different artificial substrate RNAs. In combination with previous knowledge in the field, my results help in establishing a thorough understanding of H/ACA sRNP-guided pseudouridylation.

4.2 – Optimal structure is an important aspect of H/ACA guide RNA design

Several artificial H/ACA guide RNAs derived from snR5, snR34, and snR81 were designed and tested for pseudouridylation activity within an *in vitro* reconstituted H/ACA sRNP (Figures 6A, 14A-19A, and 20). In addition, for each guide RNA:substrate RNA combination, I manually determined possible guide:substrate RNA pairing possibilities (Figures 6B, 14B-19B), and the secondary structure of each H/ACA guide

RNA was predicted using a Boltzmann-weighted folding algorithm (Figures 6C, 14C-19C). Together, activity and structural data were evaluated to reveal the presence of features common to either functional or nonfunctional H/ACA guide RNAs.

One such feature that appears critical for H/ACA guide RNA function in this study was the number of internal base pairs predicted to form within the pseudouridylation pocket of an H/ACA guide RNA in the absence of proteins and substrate RNA. Within my data set, every active H/ACA guide RNA was predicted to form less than three base pairs (and in fact, usually none) within their active pseudouridylation pockets. In order of appearance in the results, the active guide RNAs evaluated in this study included: snR34_wt, snR5_sub 2.v1, snR5_sub 2.v2, snR34_sub 2.v1, snR81_sub 2.v1, snR34_sub 3.v1, snR81_sub 3.v1, and snR81_sub 3.v2. Although the structural fold predictions are not shown, the "less than three base pair in the pseudouridylation pocket" trend also applies to every pseudouridylation pocket of each wild-type H/ACA guide RNA in this study which, by nature of being wild-type and having a cellular target, must be active (Figure 6A and 8A). Notably, when a pseudouridylation pocket was predicted to contain more than three continuous base pairs within the pseudouridylation pocket, such as in snR34_sub 1 (4 bp; Figure 6C, right) and snR34_sub 2.v2 (4 bp; Figure 15C, right), guide RNAs consistently showed no activity for their respective substrates at said pseudouridylation pocket (Figures 6A and 15A, respectively). Interestingly, the guide RNA:substrate RNA pairing possibilities which I predicted for both of these guide RNAs appear very reasonable and typical of a guide RNA:target RNA interaction, provided the base pairs within the pseudouridylation pocket can be broken to accommodate substrate RNA. Altogether, this suggests that the

base pairs formed within these pseudouridylation pockets could be inhibiting the activity of the assembled H/ACA sRNP, likely because the nucleotides are not available to base pair to and stabilize substrate RNA in the active site of Cbf5.

Structural studies of human H/ACA U65 (homolog of yeast snR34) support the fact that, in the absence of proteins, H/ACA guide RNAs can have structured pseudouridylation pockets, resulting in a "closed" pseudouridylation pocket conformation [105, 108]. Internal base pairs within a pseudouridylation pocket must be broken to allow target RNA access to the nucleotides of the pseudouridylation pocket, and in turn, allow pseudouridylation to occur. The authors of one study speculate that the binding of proteins to the H/ACA guide RNA disrupts base pairing within the pseudouridylation pocket although this has never been experimentally confirmed [105].

Given my results and the structural information available, I speculate that the formation of base pairs between the two sides of the pseudouridylation pockets is a tolerable, and somewhat common feature of H/ACA guide RNAs. However, my results indicate that too many base pairs within the pseudouridylation pocket can be disruptive to pseudouridylation. In this study, the tolerable limit for continuous internal base pairs within the pseudouridylation pocket appears to be three base pairs although further testing is needed, particularly with other guide RNAs, to confirm or adjust this value and make it generalizable. I hypothesize that internal base pairing within pseudouridylation pockets is disrupted by either protein binding or through strand displacement as a result of target RNA binding possibly in combination with protein binding to the guide RNA. Both scenarios agree with the fact that in most wild-type cases where pairing is seen within the pseudouridylation pocket, the pairing often closes off only the top or bottom

of the pseudouridylation pocket and rarely the entire pocket. This allows for at least some nucleotides within the pseudouridylation pocket to be available to initiate binding of substrate RNA, which can then displace any remaining base pairs formed.

It would be interesting to perform a systematic investigation to confirm precisely the maximum number of base pairs within the pseudouridylation pocket which are tolerable, provided this limit exists. The presence of such a limit has important implications for the design of artificial H/ACA guide RNAs. First, when designing an artificial H/ACA guide RNA for targeted pseudouridylation, designing the pseudouridylation pocket sequence rationally to avoid internal base pairs would have to be a priority. In fact, this study takes into account this exact principle in the design process of each set v1 guide RNA, and the activity of each guide RNA designed this way (Figures 14A - 18A) speaks towards the value of rational H/ACA guide RNA design, a novel approach described in detail here in my research.

Due to substrate RNA recognition requiring complementarity between target RNA and the H/ACA guide RNA pseudouridylation pocket, there is inherent limitations to a target sequence for which an artificial H/ACA guide RNA can be designed. An example of such a target would be a uridine within a small loop at the top of an RNA hairpin. The nucleotides flanking such a uridine would be, by nature, complementary to another and as a result, the guide RNA necessary to target this sequence would have complementarity between the two sides of its pseudouridylation pocket as well. If too extensive, this would not allow for the formation of an open internal loop within the H/ACA guide RNA hairpin. This may partially explain why such a small percentage

(4%) of pseudouridines in yeast rRNA are located within a small loop (< 5 nt) at the end of a hairpin (Ψ960 and Ψ2923 in 25S rRNA – [72]).

4.3 – Investigations into H/ACA guide:substrate RNA-RNA pairing

In addition to developing and testing an effective method for the generation of artificial H/ACA guide RNAs, I performed further investigations into the base pairing requirements of the H/ACA guide RNA:substrate RNA interaction. Therefore, I created and tested the activity of 9 artificial H/ACA guide RNAs for two unique artificial substrate RNAs (see Table 1 for list of substrates). Within these artificial guide RNAs, a subset (set v1) was designed rationally to provide base pairing with the artificial substrate RNA similar to that of the wild-type interaction and have an ideal overall structure, while the remaining guide RNAs (set v2) were designed using set v1 as a template and reverting half the pseudouridylation pocket to wild-type. It was expected prior to testing that the set v1 guide RNAs would be more active than set v2 guide RNAs considering they were rationally designed to form all the required base pairs to one of two artificial substrate RNAs.

When assessed for pseudouridylation activity, each set v1 guide RNA showed activity for its respective substrate RNA and was also highly specific, since no activity was observed when tested against the opposite substrate RNA (Figures 14A – 19A and Figure 20). These results clearly demonstrate the power of the guide RNA design principles that I applied. Amongst the set v2 guide RNAs, which contain an artificial, "half wild-type" pseudouridylation pocket, two of the four designs were active for the substrate which their corresponding v1 H/ACA guide RNA was designed for. This included snR5_sub 2.v2 which surprisingly caused comparable, if not more,

pseudouridylation of substrate RNA 2 than its corresponding v1 guide RNA, and also snR81_sub 3.v2 which experienced decreased, but significant, pseudouridine formation when compared to its parent v1 guide RNA (Figures 14A and 18A, respectively). Interestingly, when all guide RNAs were tested for the substrates for which they were not designed for, one set v2 guide RNA (snR5_sub2.v2) showed activity for substrate RNA 3, albeit with only 10% pseudouridylation achieved after 90 min of reaction (Figure 19A) suggesting the existence of sequence flexibility between a guide RNA and target RNA sequence.

For every active guide RNA:substrate RNA combination, the sequence of the pseudouridylation pocket expected to target the substrate RNA was compared against the substrate RNA sequence, and then manually analyzed to determine the pairing possibilities between the two RNAs (Figures 14B-19B); if more than one reasonable pairing possibility was predicted, each is shown. Across the entire set of active guide RNA:substrate RNA pairings, the number of base pairs between guide RNA and substrate RNA ranged from as little as 8 bp (Figure 18B, possibility 3) to as many as 16 bp (Figure 15B, top), which fall within the range seen both *in vivo* in natural yeast guide RNA:substrate pairings as well as what was determined in a systematic study assessing H/ACA guide:substrate RNA pairing [72, 96, 97].

The importance of the base pairs nearest the target uridine (Figure 3, purple and orange) within a pseudouridylation pocket is not entirely understood. Most research has suggested that these base pairs, which are close to the site of pseudouridylation, are more important in determining activity than those further away, particularly when the base pairing is nearing the determined minimum of 8 base pairs [96, 97]. Focusing on the base

pair 5′ of the target uridine first, we can see that amongst all the guide RNA:substrate RNA pairings within our data (Figures 14B-19B), there is no active interaction in which the base pair immediately 5′ of the target uridine in the substrate is missing (Figures 14A-19A). For clarity, the purple base pair in Figure 3 shows an example of the base pair described here. In fact, even if extensive base pairing between guide RNA and substrate RNA occurs on both sides of the pseudouridylation pocket, as is the case of snR34_sub 2.v2 targeting substrate RNA 2 (Figure 15B, possibilities 1 and 2), the absence of the base pair 5′ of the target uridine seems to be a determining factor for the lack of pseudouridylation activity observed (Figure 15A, dotted line).

However, it cannot be said conclusively that the absence of this adjacent base pair alone is entirely responsible for the loss of snR34_sub 2.v2-guided pseudouridylation activity since snR34_sub 2.v2 is also predicted to form four consecutive base pairs within the pseudouridylation pocket needed to target substrate RNA 2 (Figure 15C, right). Explained earlier in detail, extensive base pairs within the pseudouridylation pocket appear to be a contributing factor of guide RNA inactivity due to the formation of a “closed” pseudouridylation pocket. If possible, it would be potentially worthwhile to generate an optimized version of snR34_sub 2.v2 which still lacks the base pair 5′ of the target uridine when paired with substrate RNA 2, but forms fewer internal base pairs between nucleotides on each side of its pseudouridylation pocket. Evaluating the activity of this mutant should allow for a more confident determination of the contribution of the single base pair in question alone within this pseudouridylation pocket. Alternatively, one could also measure whether a snR34_sub 2.v2 sRNP is capable of binding substrate RNA 2 *in vitro*; the inability to bind substrate RNA 2 would be suggestive of a closed

pseudouridylation pocket whereas the absence of one base pair within the interaction should only have a small effect on the affinity.

The predicted base-pairing between snR81_sub 2.v2 and substrate RNA 2 (Figure 16B, bottom) is another example where the absence of adjacent base pairs upstream of the target uridine result in a complete loss of pseudouridylation activity even though there is extensive base pairing on both sides of the pseudouridylation pocket (14A, dotted line). Although not entirely conclusive due to the added effects of base pairs within the pseudouridylation pocket, my results appear to agree with what has been observed in the literature that the base pair immediately adjacent the target uridine plays an important role in determining the productivity of the guide RNA:substrate RNA interaction [96, 97].

Similarly, I evaluated the results with a focus on identifying the importance of the nearest base pair formed downstream, i.e. 3′ of the target uridine. Again, for clarity, this base pair is depicted in orange in Figure 3. The majority of H/ACA guide:substrate RNA-RNA interactions in yeast (> 95%) are predicted to only have one nucleotide (and at most two) next to the target uridine remaining unpaired within the pseudouridylation pocket of the H/ACA guide RNA [72]. The unpaired nucleotide is always immediately 3′ of the target uridine (green in Figure 3) and the following nucleotide then forms the nearest base pair with H/ACA guide RNA on the 5′ side of the pseudouridylation pocket (orange in Figure 3).

Besides the set v1 guide RNAs, which were rationally designed to have just one unpaired substrate RNA nucleotide (as seen in the majority of wild-type interactions), the active set v2 H/ACA guide RNAs in our study often have 2 (rarely more) unpaired

nucleotides predicted to be present in the pseudouridylation pocket. This can be seen in the predicted base pairings for both snR81_sub 3.v2 and snR5_sub 2.v2 for substrate RNA 3 (Figures 18B and 19B, respectively). When reconstituted *in vitro*, both H/ACA sRNPs display activity for substrate RNA 3 (Figures 18A and 19A, dotted lines). If targeted by the 5′ hairpin pseudouridylation pocket of snR81_sub 3.v2, the best case scenario pairings (pairing possibilities 1 and 2 in Figure 18B) indicate the need to accommodate 6 unpaired nucleotides within the pseudouridylation pocket, with only 2 base pairs anchoring the entire 3′ end of the substrate RNA to the H/ACA guide RNA. Although published data suggests the ability of the pseudouridylation pocket to accommodate up to 6 unpaired nucleotides ([96]), this disagrees with unpublished data obtained in our group ([97]) and therefore the predicted pairings between the 5′ hairpin of snR81_sub 3.v2 and substrate RNA 3 seemed unlikely. Equally questionable is the fact that there are only 2 base pairs anchoring an entire side of a pseudouridylation pocket, a feature not seen in any natural H/ACA guide RNA:target RNA pairing. In fact, in some cases, having only three base pairs on one side of the pseudouridylation pocket rendered a guide RNA inactive, even when more than 8 base pairs are maintained across the entire pseudouridylation pocket [96].

However, at the wild-type snR81 3′ pseudouridylation pocket, the base pairing possibilities with substrate RNA 3 (possibilities 3 and 4 – Figure 18B) appear more reasonable. Both possibilities have a continuous stretch of base pairs on each side of the pseudouridylation pocket and also only 2 unpaired nucleotides downstream of the target uridine. Recent investigations into the limitations of H/ACA guide RNA targeting *in vivo* has suggested that, at the very least, 8 base pairs must be formed between the guide RNA

and substrate RNA for pseudouridylation [96]. If three base pairs at the top of the 3′ snR81_sub 3.v2 pseudouridylation pocket are disrupted (Figure 18B, possibility 3), the interaction agrees with all the putative pseudouridylation activity criteria (i.e. 8 base pairs overall, a stretch of base pairs on each side with no more than 2 unpaired nucleotides downstream of target uridine, and no base pairs closing off the pseudouridylation pocket in the guide RNA).

It would be interesting to test whether or not pseudouridylation of substrate RNA 3 is indeed being directed by the wild-type 3′ hairpin of snR81_sub 3.v2 (and potentially snR81_sub 3.v1). Through site-directed mutagenesis, it would be possible to mutate the 3′ hairpin of both snR81_sub 3.v1 and snR81_sub 3.v2 and see its effect on pseudouridylation of substrate RNA 3, provided the mutations don't disrupt the overall structure of the H/ACA guide RNA. If pseudouridylation of substrate RNA 3 is occurring through the 3′ hairpin in these two mutants, this provides us with the ideal scenario to confirm if 8 base pairs is indeed the minimum required for directing pseudouridylation since the pairing possibilities at this hairpin suggest a possibility where only 7 base pairs are formed between guide and substrate RNA. From the predicted pairings, it would appear more likely that pseudouridylation activity at this hairpin would follow pairing possibility 3 (Figure 18B) since it both requires the least number of base pairs within the pseudouridylation pocket needing to be broken and contains 8 base pairs with substrate RNA 3. However, testing pseudouridylation activity after disrupting the furthest lone G-C base pair between guide RNA and substrate upstream of the target uridine would confirm if pseudouridylation activity at this hairpin requires a minimum of 8 base pairs in the guide:substrate RNA-RNA interaction.

When examining the activity of artificial H/ACA guide RNAs for substrate RNA 2, it was very surprising that snR5_sub 2.v2 outperformed its corresponding v1 guide RNA (snR5_sub 2.v1) in modification of substrate RNA 2 (Figure 14A). Comparing the base pairing with substrate RNA 2 (Figure 14B), it is clear that the set v1 variant displays a longer consecutive sequence of base pairs upstream of the target uridine. As stated before, the set v1 guide RNA was designed to maintain the number and location of base pairs as seen in the wild-type interaction at that hairpin, which in the case of snR5, corresponds to six Watson-Crick base pairs on each side of the pseudouridylation pocket. Interestingly, the difference between these two guide RNAs suggests the possibility that having a too stable guide RNA:substrate RNA duplex upstream of the target uridine is inhibitory. This agrees strongly with a comprehensive analysis of archaeal H/ACA guide RNAs which typically have shorter stretches of base pairing upstream of the target uridine [93].

Considering also the fact that many wild-type putative guide RNA:substrate RNA pairings have maintained mismatches throughout evolution suggests that a more complex model for predicting guide RNA:substrate RNA productivity is needed than simply "presence of base pairs with substrate RNA on both sides of the pseudouridylation pocket". It is worth noting the possible existence of an optimal "pairing energy" parameter across the entire pseudouridylation pocket, or on each side of the pocket individually, that may influence pseudouridylation by modulating substrate RNA binding and substrate turnover. Although very interesting, progress in dissecting such parameters is currently halted by the fact that current algorithms are unable to accurately predict the

stability of H/ACA guide RNA:substrate RNA pairing due to the nature of this interaction forming an atypical A-form helix rather than a standard B-form helix [106].

4.4 – Investigating the 5′ extension of snR34 and pseudouridylation in cis

A recent investigation by our group determined that the unstructured 5′ extension of the H/ACA guide RNA snR34, corresponding to the first 24 nt which do not yet participate in the lower stem of the 5′ hairpin (Figure 6C,left for reference), is not required for protein binding or activity; surprisingly, removal of the snR34 5′ extension actually increased pseudouridylation formation *in vitro* [60, 97]. Considering the 5′ extension is not necessary for assembly or activity *in vitro*, it appears to be a good candidate for engineering purposes since its alteration would not be expected to negatively influence H/ACA sRNP function. Since only deletion of the 5′ extension was tested, I wanted to instead perform investigations lengthening the 5′ extension. The ability to add nucleotides to the 5′ end of H/ACA guide RNAs without detriment creates many engineering possibilities, such as the addition of protein recognition sequences or localization signals.

At the same time, I was also curious about the potential for H/ACA guide RNAs to direct pseudouridylation in cis (i.e. directing the pseudouridylation of a uridine within its own sequence). Two snR34 guide RNA:substrate RNA chimeras were created each containing 83 nt of additional sequence attached to the normal 5′ end. In each variant, this sequence contained a portion of 25S rRNA normally targeted by either the 5′ or 3′ snR34 pseudouridylation pockets as well as additional flanking/linker sequence (Figure 5A). The sequences flanking the 25S rRNA target sequences were strategically designed to be unstructured to limit unwanted effects on overall H/ACA guide RNA fold.

Each guide RNA was tested under multiple turnover conditions for pseudouridylation activity against free [^{3}H]UTP-labeled 25S rRNA substrates 1 and 2, targeted by the 5′ and 3′ pseudouridylation pockets of snR34, respectively. Due to technical limitations of the tritium release assay, pseudouridylation activity cannot be tested for directly in cis. Both chimeric guide RNAs showed activity for both rRNA substrates, maintaining some percentage of wild-type snR34 pseudouridylation throughout the reaction (Figure 5B). Considering pseudouridylation activity was observed our results suggest that lengthening the 5′ extension of snR34 likely did not entirely disrupt the H/ACA guide RNA fold or the assembly of a mature and functional H/ACA sRNP.

The results discussed here indicate that increasing the length of the snR34 5′ extension reduced pseudouridylation formation slightly in free substrate RNA and previous results in our lab indicate that removal of the 5′ extension increased pseudouridine formation to levels greater than that of wild-type snR34 [97]. Considered together, it appears as though there exists an inverse relationship between the length of the 5′ extension of an H/ACA guide RNA and ability to direct pseudouridylation *in vitro*. *In vivo*, the presence of the 5′ extension of H/ACA guide RNAs is believed to be a by-product of pre-snoRNA exonucleolytic processing, due to supposed protection of these regions from core protein association [77]. However, if this was the sole determinant, we would expect the 5′ terminus of all mature H/ACA guide RNAs in the cell to be a similar distance away from the base of each guide RNA hairpin, which is not the case. This suggests that the length of the H/ACA guide RNA 5′ extension may be of some importance in each H/ACA guide RNA. Considering my results alongside the previous

results obtained in our group, I speculate that the length of the 5′ extension of H/ACA guide RNAs varies as a mechanism to modulate the activity of each H/ACA guide RNA in the cell. The majority of H/ACA guide RNAs are expressed constitutively to relatively equal levels such a mechanism would contribute to a pseudouridine modification landscape in which some pseudouridines are present at higher frequency in RNA, due to being introduced by a guide RNA with a short(er) 5′ extension, and some pseudouridines at lower frequency, due to a guide RNA with a relatively long(er) 5′ extension. To elucidate this relationship, further systematic testing is needed, particularly with several H/ACA guide RNAs, to confidently draw a conclusion which could be generalized across all H/ACA guide RNAs.

Excitingly, pseudouridylation assays with chimeric H/ACA sRNPs (Figure 5B) indicated that when targeting the same 25S rRNA substrate in trans as the substrate attached to the 5′ end of the chimeric guide RNA, pseudouridylation formation was reduced more than if the same chimeric guide RNA was targeting the other 25S rRNA substrate. In other words, while snR34_rRNA sub 1-chimera directs greater pseudouridine formation than snR34_rRNA sub 2-chimera when targeting 25S rRNA substrate 2 (Figure 5B, bottom), the opposite is true if targeting 25S rRNA substrate 1 (Figure 5B, top). This result could be caused by a competition effect observed when targeting the same substrate RNA in trans as the one attached to the chimeric H/ACA guide RNA. Binding the attached target rRNA element in cis with a pseudouridylation pocket would mean that to bind and modify free radiolabeled substrate RNA (in trans), the pseudouridylation pocket would first have to modify and release the substrate bound in cis so that the free radiolabeled substrate RNA would have access to the

pseudouridylation pocket. The activity differences used to draw this conclusion are rather small, partially since our pseudouridylation assays have a much greater concentration of free radiolabeled substrate RNA than H/ACA sRNP (10-fold greater). Because of the drastic concentration differences, free substrate RNA would very effectively outcompete any pseudouridylation pocket binding in cis. It may be valuable to repeat similar assays in the future with a reduced concentration of free substrate RNA, potentially under single turnover conditions where substrate RNA and H/ACA sRNP are at equal concentrations.

4.5 Conclusions

The major contributions of this thesis stem largely from the establishment and utilization of a structure-directed guide RNA design, testing, and analysis approach. In this thesis, I describe and exhibit the ability of this approach to consistently design artificial H/ACA guide RNAs with pseudouridylation activity *in vitro* for a pair of artificial substrate RNAs. The development of artificial H/ACA guide RNAs for targeted pseudouridylation has several applications in the cell. With progressively more examples of pseudouridylation in mRNA emerging, and the fact that mRNA pseudouridylation has been shown to be a dynamic event, targeted pseudouridylation could be used as a potential tool to regulation of gene expression. Excitingly, it has been shown that pseudouridylation of termination codons causes translational read-through, a phenomenon with very wide-reaching applicability. Although I establish the efficacy of the structure-based guide RNA design approach *in vitro*, further work needs to be done to corroborate success *in vivo*.

In addition, my data expand on current understandings of H/ACA guide RNA activity and the structure-function relationship of H/ACA guide RNAs. Results obtained

here demonstrate that extensive base-pairing within the pseudouridylation pocket of an H/ACA guide RNA renders the pseudouridylation pocket unable to guide modification. Based on structure predictions of the H/ACA guide RNAs designed and tested in this study, pairing between the two sides of the pseudouridylation pocket is a common feature of H/ACA guide RNAs, but tolerable only if 3 or less continuous base pairs are formed. This advances H/ACA guide RNA design strategies by highlighting the need to incorporate a strategy that is cognizant of the presence of internal pseudouridylation pocket base pairing. It also indicates the likelihood of an intrinsic limitation to which uridines are capable to be targeted for pseudouridylation by H/ACA guide RNAs, limited to those whose surrounding sequence does not have high complementarity. Further investigations using a systematic approach would be very helpful in confirming the veracity of my results.

My results also contribute towards establishing a comprehensive understanding of H/ACA guide:substrate RNA-RNA interaction features that influence productivity. Particularly, in every active guide:substrate combination, the base pair immediately 5′ of the target uridine (Figure 3, purple for reference) was always maintained. When this base pair was lost, even in cases where pairing appeared otherwise sufficient, activity was not observed. Previous studies have reported contradicting requirement of this base pair in directing pseudouridylation [96, 97], so it is possible that this base pair while essential at some guide RNA pseudouridylation pockets, may not be essential in others. On the 3′ side of the target uridine, most active guide:substrate combinations I tested appear to always have one or two unpaired nucleotides, which is consistent with what is seen in nature and in published studies (Figure 3, green for reference). However, while it has

been reported that H/ACA guide RNAs can tolerate up to 5 unpaired nucleotides downstream of the target uridine [96], this is not well reflected in my data since no pairings with more than 2 unpaired nucleotides displayed activity. It is possible that there exists alternative pairing possibilities between guide RNA and substrate RNA in the published cases, which reduce the number of unpaired nucleotides downstream of the target uridine to two or less. More work in this area is needed to elucidate further details of the H/ACA guide RNA:substrate RNA interaction. A comprehensive understanding of H/ACA guide RNA targeting will allow for better evaluation of the productivity of guide:substrate pairing combinations.

Finally, I suggest a model of snR34 activity regulation in which the length of the 5′ extension of snR34 is indirectly proportional to the ability of an H/ACA guide RNA to direct pseudouridylation. Here, I show that this is applicable to at least one guide RNA – snR34; however, it would be interesting to know whether this is a generalizable feature of all cellular H/ACA guide RNAs. This is worthwhile to determine and could fill an otherwise limited understanding of the mechanisms involved in the heterogeneity of pseudouridylation frequency.

Bibliography

1. Boccaletto, P., et al., *MODOMICS: a database of RNA modification pathways. 2017 update.* Nucleic Acids Res, 2018. **46**(D1): p. D303-D307.

2. Palazzo, A.F. and E.S. Lee, *Non-coding RNA: what is functional and what is junk?* Front Genet, 2015. **6**: p. 2.

3. Karijolich, J., A. Kantartzis, and Y.T. Yu, *RNA modifications: a mechanism that modulates gene expression.* Methods Mol Biol, 2010. **629**: p. 1-19.

4. Ayub, M., et al., *Nanopore-based identification of individual nucleotides for direct RNA sequencing.* Nano Lett, 2013. **13**(12): p. 6144-6150.

5. Spenkuch, F., Y. Motorin, and M. Helm, *Pseudouridine: still mysterious, but never a fake (uridine)!* RNA Biol, 2014. **11**(12): p. 1540-1554.

6. Brand, R.C., et al., *Pseudouridylation of yeast ribosomal precursor RNA.* Nucleic Acids Res, 1979. **7**(1): p. 121-134.

7. Madison, J.T., G.A. Everett, and H. Kung, *Nucleotide sequence of a yeast tyrosine transfer RNA.* Science, 1966. **153**(3735): p. 531-534.

8. Kato, N. and F. Harada, *Nucleotide sequence of nuclear 5.4 S RNA of mouse cells.* Biochim Biophys Acta, 1984. **782**(2): p. 127-131.

9. Carlile, T.M., et al., *Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells.* Nature, 2014. **515**(7525): p. 143-146.

10. Davis, F.F. and F.W. Allen, *Ribonucleic acids from yeast which contain a fifth nucleotide.* J Biol Chem, 1957. **227**(2): p. 907-915.

11. Pomerantz, S.C. and J.A. McCloskey, *Detection of the common RNA nucleoside pseudouridine in mixtures of oligonucleotides by mass spectrometry.* Anal Chem, 2005. **77**(15): p. 4687-4697.

12. Ofengand, J. and A. Bakin, *Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaebacteria, mitochondria and chloroplasts.* J Mol Biol, 1997. **266**(2): p. 246-268.

13. Lovejoy, A.F., D.P. Riordan, and P.O. Brown, *Transcriptome-Wide Mapping of Pseudouridines: Pseudouridine Synthases Modify Specific mRNAs in S. cerevisiae.* PLoS ONE, 2014. **9**(10): p. 1-15.

14. Schwartz, S., et al., *Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA.* Cell, 2014. **159**(1): p. 148-162.

15. Foster, P.G., et al., *The structural basis for tRNA recognition and pseudouridine formation by pseudouridine synthase I.* Nat Struct Biol, 2000. **7**(1): p. 23-27.

16. Hoang, C. and A.R. Ferré-D'Amaré, *Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme.* Cell, 2001. **107**(7): p. 929-939.

17. Kaya, Y., et al., *Crystal structure of TruD, a novel pseudouridine synthase with a new protein fold.* J Biol Chem, 2004. **279**(18): p. 18107-1810.

18. Sivaraman, J., et al., *Structure of the 16S rRNA pseudouridine synthase RsuA bound to uracil and UMP.* Nat Struct Biol, 2002. **9**(5): p. 353-358.

19. Hoang, C., et al., *Crystal structure of pseudouridine synthase RluA: indirect sequence readout through protein-induced RNA structure.* Mol Cell, 2006. **24**(4): p. 535-545.

20. Hamma, T. and A.R. Ferré-D'Amaré, *Pseudouridine synthases.* Chem Biol, 2006. **13**(11): p. 1125-1135.

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

22. Watkins, N.J. and M.T. Bohnsack, *The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA.* Wiley Interdiscip Rev RNA, 2012. **3**(3): p. 397-414.

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

24. Huang, L., et al., *A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst.* Biochemistry, 1998. **37**(1): p. 344-351.

25. Friedt, J., et al., *An arginine-aspartate network in the active site of bacterial TruB is critical for catalyzing pseudouridine formation.* Nucleic Acids Res, 2014. **42**(6): p. 3857-3870.

26. Phannachet, K., Y. Elias, and R.H. Huang, *Dissecting the roles of a strictly conserved tyrosine in substrate recognition and catalysis by pseudouridine 55 synthase.* Biochemistry, 2005. **44**(47): p. 15488-15494.

27. Wright, J.R., et al., *Pre-steady-state kinetic analysis of the three Escherichia coli pseudouridine synthases TruB, TruA, and RluA reveals uniformly slow catalysis.* RNA, 2011. **17**(12): p. 2074-2084.

28. Gu, X., et al., *Molecular recognition of tRNA by tRNA pseudouridine 55 synthase.* Biochemistry, 1998. **37**(1): p. 339-343.

29. Raychaudhuri, S., et al., *Functional effect of deletion and mutation of the Escherichia coli ribosomal RNA and tRNA pseudouridine synthase RluA.* J Biol Chem, 1999. **274**(27): p. 18880-18886.

30. Gutgsell, N., et al., *Deletion of the Escherichia coli pseudouridine synthase gene truB blocks formation of pseudouridine 55 in tRNA in vivo, does not affect exponential growth, but confers a strong selective disadvantage in competition with wild-type cells.* RNA, 2000. **6**(12): p. 1870-1881.

31. Conrad, J., et al., *16S ribosomal RNA pseudouridine synthase RsuA of Escherichia coli: deletion, mutation of the conserved Asp102 residue, and sequence comparison among all other pseudouridine synthases.* RNA, 1999. **5**(6): p. 751-763.

32. Carbone, M.L., et al., *A gene tightly linked to CEN6 is important for growth of Saccharomyces cerevisiae.* Curr Genet, 1991. **19**(1): p. 1-8.

33. Ansmant, I., et al., *Identification of the Saccharomyces cerevisiae RNA:pseudouridine synthase responsible for formation of psi(2819) in 21S mitochondrial ribosomal RNA.* Nucleic Acids Res, 2000. **28**(9): p. 1941-1946.

34. Bakin, A., B.G. Lane, and J. Ofengand, *Clustering of pseudouridine residues around the peptidyltransferase center of yeast cytoplasmic and mitochondrial ribosomes.* Biochemistry, 1994. **33**(45): p. 13475-13483.

35. Keffer-Wilkes, L.C., G.R. Veerareddygari, and U. Kothe, *RNA modification enzyme TruB is a tRNA chaperone.* Proc Natl Acad Sci U S A, 2016. **113**(50): p. 14306-14311.

36. Gutgsell, N.S., et al., *A second function for pseudouridine synthases: A point mutant of RluD unable to form pseudouridines 1911, 1915, and 1917 in Escherichia coli 23S ribosomal RNA restores normal growth to an RluD-minus strain.* RNA, 2001. **7**(7): p. 990-998.

37. Zebarjadian, Y., et al., *Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA.* Mol Cell Biol, 1999. **19**(11): p. 7461-7472.

38. Heiss, N.S., et al., *X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions.* Nat Genet, 1998. **19**(1): p. 32-38.

39. Arnez, J.G. and T.A. Steitz, *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, 1994. **33**(24): p. 7560-7567.

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

41. Penzo, M. and L. Montanaro, *Turning Uridines around: Role of rRNA Pseudouridylation in Ribosome Biogenesis and Ribosomal Function.* Biomolecules, 2018. **8**(2).

42. Ofengand, J., *Ribosomal RNA pseudouridines and pseudouridine synthases.* FEBS Lett, 2002. **514**(1): p. 17-25.

43. King, T.H., et al., *Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center.* Mol Cell, 2003. **11**(2): p. 425-435.

44. O'Connor, M., M. Leppik, and J. Remme, *Pseudouridine-Free Escherichia coli Ribosomes.* J Bacteriol, 2017. **200**(4): p. 1-10.

45. Bohnsack, M.T. and K.E. Sloan, *Modifications in small nuclear RNAs and their roles in spliceosome assembly and function.* Biol Chem, 2018. **399**(11): p. 1265-1276.

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

47. Newby, M.I. and N.L. Greenbaum, *A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture.* RNA, 2001. **7**(6): p. 833-845.

48. Lecointe, F., et al., *Lack of pseudouridine 38/39 in the anticodon arm of yeast cytoplasmic tRNA decreases in vivo recoding efficiency.* J Biol Chem, 2002. **277**(34): p. 30445-30453.

49. Guzzi, N., et al., *Pseudouridylation of tRNA-Derived Fragments Steers Translational Control in Stem Cells.* Cell, 2018. **173**(5): p. 1204-1216.

50. Li, X., et al., *Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome.* Nat Chem Biol, 2015. **11**(8): p. 592-597.

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

52. Huang, C., G. Wu, and Y.T. Yu, *Inducing nonsense suppression by targeted pseudouridylation.* Nat Protoc, 2012. **7**(4): p. 789-800.

53. Svidritskiy, E., R. Madireddy, and A.A. Korostelev, *Structural Basis for Translation Termination on a Pseudouridylated Stop Codon.* J Mol Biol, 2016. **428**(10 Pt B): p. 2228-2236.

54. Fernandez, I.S., et al., *Unusual base pairing during the decoding of a stop codon by the ribosome.* Nature, 2013. **500**(7460): p. 107-110.

55. Kiss, T., *Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs.* EMBO J, 2001. **20**(14): p. 3617-3622.

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

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

58. Nguyen, T.H.D., et al., *Cryo-EM structure of substrate-bound human telomerase holoenzyme.* Nature, 2018. **557**(7704): p. 190-195.

59. Ashbridge, B., et al., *Single-molecule analysis of the human telomerase RNA.dyskerin interaction and the effect of dyskeratosis congenita mutations.* Biochemistry, 2009. **48**(46): p. 10858-10865.

60. Caton, E.A., et al., *Efficient RNA pseudouridylation by eukaryotic H/ACA ribonucleoproteins requires high affinity binding and correct positioning of guide RNA.* Nucleic Acids Res, 2018. **46**(2): p. 905-916.

61. Hamma, T., et al., *The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs.* Nat Struct Mol Biol, 2005. **12**(12): p. 1101-1107.

62. Kamalampeta, R. and U. Kothe, *Archaeal proteins Nop10 and Gar1 increase the catalytic activity of Cbf5 in pseudouridylating tRNA.* Sci Rep, 2012. **2**: 663.

63. Li, S., et al., *Reconstitution and structural analysis of the yeast box H/ACA RNA-guided pseudouridine synthase.* Genes Dev, 2011. **25**(22): p. 2409-2421.

64. Jiang, W., et al., *An essential yeast protein, CBF5p, binds in vitro to centromeres and microtubules.* Mol Cell Biol, 1993. **13**(8): p. 4884-4893.

65. Lafontaine, D.L., et al., *The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase.* Genes Dev, 1998. **12**(4): p. 527-537.

66. Baker, D.L., et al., *RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP.* Genes Dev, 2005. **19**(10): p. 1238-1248.

67. Girard, J.P., et al., *GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast.* EMBO J, 1992. **11**(2): p. 673-682.

68. Bousquet-Antonelli, C., et al., *A small nucleolar RNP protein is required for pseudouridylation of eukaryotic ribosomal RNAs.* EMBO J, 1997. **16**(15): p. 4770-4776.

69. Henras, A., et al., *Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs.* EMBO J, 1998. **17**(23): p. 7078-7090.

70. Hamma, T. and A.R. Ferré-D'Amaré, *Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 A resolution.* Structure, 2004. **12**(5): p. 893-903.

71. Lestrade, L. and M.J. Weber, *snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs.* Nucleic Acids Res, 2006. **34**(Database issue): p. D158-D162.

72. Piekna-Przybylska, D., W.A. Decatur, and M.J. Fournier, *New bioinformatic tools for analysis of nucleotide modifications in eukaryotic rRNA.* RNA, 2007. **13**(3): p. 305-312.

73. Steinmetz, E.J., et al., *Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase.* Mol Cell, 2006. **24**(5): p. 735-746.

74. Deng, W., et al., *Organization of the Caenorhabditis elegans small non-coding transcriptome: genomic features, biogenesis, and expression.* Genome Res, 2006. **16**(1): p. 20-29.

75. Angrisani, A., et al., *Developmentally regulated expression and expression strategies of Drosophila snoRNAs.* Insect Biochem Mol Biol, 2015. **61**: p. 69-78.

76. Ooi, S.L., et al., *Intronic snoRNA biosynthesis in Saccharomyces cerevisiae depends on the lariat-debranching enzyme: intron length effects and activity of a precursor snoRNA.* RNA, 1998. **4**(9): p. 1096-1110.

77. Petfalski, E., et al., *Processing of the precursors to small nucleolar RNAs and rRNAs requires common components.* Mol Cell Biol, 1998. **18**(3): p. 1181-1189.

78. Allmang, C., et al., *Functions of the exosome in rRNA, snoRNA and snRNA synthesis.* EMBO J, 1999. **18**(19): p. 5399-5410.

79. Massenet, S., E. Bertrand, and C. Verheggen, *Assembly and trafficking of box C/D and H/ACA snoRNPs.* RNA Biol, 2017. **14**(6): p. 680-692.

80. Godin, K.S., et al., *The box H/ACA snoRNP assembly factor Shq1p is a chaperone protein homologous to Hsp90 cochaperones that binds to the Cbf5p enzyme.* J Mol Biol, 2009. **390**(2): p. 231-244.

81. Walbott, H., et al., *The H/ACA RNP assembly factor SHQ1 functions as an RNA mimic.* Genes Dev, 2011. **25**(22): p. 2398-2408.

82. Machado-Pinilla, R., et al., *Mechanism of the AAA+ ATPases pontin and reptin in the biogenesis of H/ACA RNPs.* RNA, 2012. **18**(10): p. 1833-1845.

83. Dez, C., et al., *Naf1p, an essential nucleoplasmic factor specifically required for accumulation of box H/ACA small nucleolar RNPs.* Mol Cell Biol, 2002. **22**(20): p. 7053-7065.

84. Leulliot, N., et al., *The box H/ACA RNP assembly factor Naf1p contains a domain homologous to Gar1p mediating its interaction with Cbf5p.* J Mol Biol, 2007. **371**(5): p. 1338-1353.

85. Girard, J.P., et al., *Identification of a segment of the small nucleolar ribonucleoprotein-associated protein GAR1 that is sufficient for nucleolar accumulation.* J Biol Chem, 1994. **269**(28): p. 18499-18506.

86. Balakin, A.G., L. Smith, and M.J. Fournier, *The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions.* Cell, 1996. **86**(5): p. 823-834.

87. Ganot, P., M. Caizergues-Ferrer, and T. Kiss, *The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation.* Genes Dev, 1997. **11**(7): p. 941-956.

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

89. Meier, U.T., *RNA modification in Cajal bodies.* RNA Biol, 2017. **14**(6): p. 693-700.

90. Richard, P., et al., *A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs.* EMBO J, 2003. **22**(16): p. 4283-4293.

91. Rozhdestvensky, T.S., et al., *Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea.* Nucleic Acids Res, 2003. **31**(3): p. 869-877.

92. Mitchell, J.R., J. Cheng, and K. Collins, *A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3′ end.* Mol Cell Biol, 1999. **19**(1): p. 567-576.

93. Toffano-Nioche, C., D. Gautheret, and F. Leclerc, *Revisiting the structure/function relationships of H/ACA(-like) RNAs: a unified model for Euryarchaea and Crenarchaea.* Nucleic Acids Res, 2015. **43**(16): p. 7744-7761.

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

95. Wu, H. and J. Feigon, *H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification.* Proc Natl Acad Sci U S A, 2007. **104**(16): p. 6655-6660.

96. De Zoysa, M.D., et al., *Guide-substrate base-pairing requirement for box H/ACA RNA-guided RNA pseudouridylation.* RNA, 2018. **24**(8): p. 1106-1117.

97. Kelly, E.K., *Distinguishing features of guide RNA recognition by H/ACA snoRNPs.* MSc Thesis. University of Lethbridge, 2018.

98. Bonneaud, N., et al., *A family of low and high copy replicative, integrative and single-stranded S. cerevisiae/E. coli shuttle vectors.* Yeast, 1991. **7**(6): p. 609-615.

99. Cazenave, C. and O.C. Uhlenbeck, *RNA template-directed RNA synthesis by T7 RNA polymerase.* Proc Natl Acad Sci U S A, 1994. **91**(15): p. 6972-6976.

100. Moss, W.N., *RNA2DMut: a web tool for the design and analysis of RNA structure mutations.* RNA, 2018. **24**(3): p. 273-286.

101. Ding, Y., C.Y. Chan, and C.E. Lawrence, *RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble.* RNA, 2005. **11**(8): p. 1157-1166.

102. Moss, W.N., *RNA2DMut: A web tool for the design and analysis of RNA structure mutations.* RNA, 2017. **24**(3): p. 273-286.

103. Schattner, P., et al., *Genome-wide searching for pseudouridylation guide snoRNAs: analysis of the Saccharomyces cerevisiae genome.* Nucleic Acids Res, 2004. **32**(14): p. 4281-4296.

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

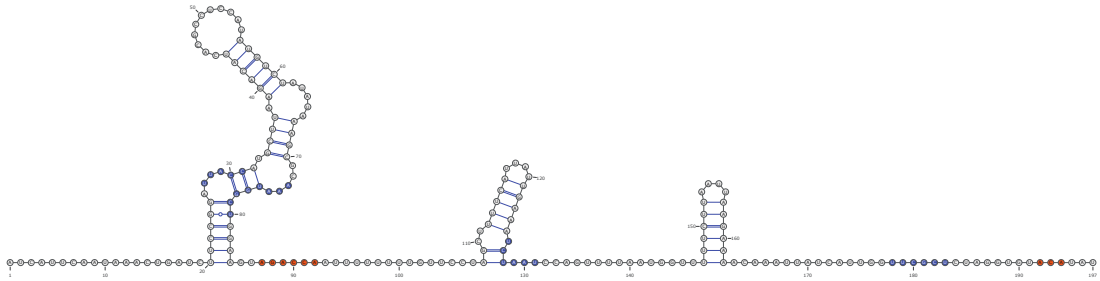
105. Khanna, M., et al., *Structural study of the H/ACA snoRNP components Nop10p and the 3′ hairpin of U65 snoRNA.* RNA, 2006. **12**(1): p. 40-52.

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

107. Liang, B., et al., *Structure of a functional ribonucleoprotein pseudouridine synthase bound to a substrate RNA.* Nat Struct Mol Biol, 2009. **16**(7): p. 740-746.

108. Jin H., et al., *Solution struture of an rRNA substrate bound to the pseudouridylation pocket of a box H/ACA snoRNA.* Mol Cell, 2007. **26**(2): p. 205-215.

Appendix



snR5_wt
Boltzmann-weighted centroid structure

Figure A1. The Boltzmann-weighted centroid ensemble structure of snR5_wt. Nucleotides used for base pairing to target RNA at each pseudouridylation pocket are colored blue. The conserved H and ACA box element are colored orange.