MECHANISTIC CHARACTERIZATION OF PSEUDOURIDINE FORMATION BY EUKARYOTIC H/ACA SNORNPS

MEHAR GAYATRI DEVI NAMALA
Master of Science, Indian Institute of Science, 2018

A thesis submitted
in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOMOLECULAR SCIENCE

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

MECHANISTIC CHARACTERIZATION OF PSEUDOURIDINE FORMATION BY EUKARYOTIC H/ACA SNORNPS

MEHAR GAYATRI DEVI NAMALA

Date of Defence: November 30, 2022

Dr. U. Kothe Dr. T. Patel Thesis Co-Supervisors	Professor Associate Professor	Ph.D. Ph.D.
Dr. H-J. Wieden Thesis Examination Committee Member	Professor	Ph.D.
Dr. U.T. Meier External Examiner Department of Cell Biology Albert Einstein College of Medicine	Professor	Ph.D.
Dr. L. Barrett Internal External Examiner Department of Psychology Faculty of Arts & Science	Professor	Ph.D.
Dr. M. Gerken Chair, Thesis Examination Committee	Professor	Ph.D.

DEDICATION

In loving memory of my father

ఓం పూర్ణమదః పూర్ణమిదం పూర్ణాత్ పూర్ణముదచ్యతే |

పూర్ణస్య పూర్ణమాదాయ పూర్ణమేవావశిష్యతే ||

ఓం శాంతిః శాంతిః శాంతిః ||

"What is visible (Universe) is the infinite.
What is invisible (God) is also the infinite.
The infinite (Universe) proceeds from the infinite (God).
(Then) taking the infinitude of the infinite (universe),
It remains as the infinite (God) alone"
om santhi,santhi,santhi!!

ABSTRACT

The H/ACA snoRNP complex directs the formation of pseudouridines in ribosomal RNA (rRNA); however, the mechanisms by which the eukaryotic two-hairpin structured H/ACA snoRNP mediates pseudouridylation in rRNA are poorly understood. This study investigates the function of the conserved two-hairpin structure of eukaryotic H/ACA snoRNA while modifying large, structured rRNA. The biochemical analysis reveals that the two hairpins of the snoRNP complex can independently convert two uridines to two pseudouridines in long rRNA fragments. Moreover, the H/ACA snoRNP can likely unfold secondary structures in rRNA without the help of additional factors. Further, the GAR domains of Gar1 are required for pseudouridine formation, but the KKE/D extensions of Cbf5 play no role in pseudouridylation of rRNA. In summary, this study addresses the mechanism of the H/ACA snoRNP complex in modifying rRNA during ribosome biogenesis.

ACKNOWLEDGEMENTS

First, I would like to thank my wonderful supervisor, Prof. Ute Kothe. She is a great mentor; I am truly glad to have her as my Ph.D. supervisor. My special thanks to my co-supervisor, Dr. Trushar Patel, for his support in my Ph.D. studies. Moreover, I am also grateful for all our non-research conversations on gardening and cooking.

Thanks to my committee members, Dr. Tony Russell and Dr. Elizabeth Schultz, for their valuable suggestions during committee meetings. Thanks to Dr. H-J Wieden for agreeing to be a committee member at the last minute; I appreciate the gesture! Moreover, thank you, H-J and Ute, for hosting me at your house, I enjoyed our coffee conversations. My special thanks to U. Thomas Meier and Louise Barrett for their valuable time.

Thanks to the RNA innovation program and SGS for the great opportunity and funding. Thanks to Laura Keffer Wilkes for being a lovely RNA innovation program coordinator. Thanks to Susan and Gipsy for helping me with the documentation. Many thanks to Dr. Borries Demeler for his kindness. Thanks to Tim and Runhun for helping me with things in the new lab at the University of Manitoba.

Lastly, I thank my family, especially my parents, brother, and cousin Vasu, who listened to me without judgment and supported me through my difficult times. Love you all and thank you for your trust and respect!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ATP	Adenosine-5'-triphosphate
Amp	Ampicillin
Cbf5	Centromere binding factor 5
CAB	Cajal body-specific localizing box elements
C/D snoRNA	snoRD
DKC1	Dyskerin
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
Gar1	Glycine/arginine-rich domain containing protein 1
GTP	Guanosine-5'-triphosphate
His-tag	Histidine-tag
HEPES	*N*-2-Hydroxyethylpiperazine-*N*'2-ethanesulfonic acid
hTR	Human telomeric RNA
H/ACA snoRNA	snoRA
iPPase	Inorganic pyrophosphatase
IPTG	β-D-1-Thiogalactopyranoside
Kan	Kanamycin
LB	Lysogeny broth
mRNA	Messenger RNA
MM	Multiple myeloma cells
Nhp2	Non-histone protein 2
Nop10	Nucleolar protein 10
NMR	Nuclear magnetic resonance
NTP	Nucleotide-5'-triphosphate
OD600	Optical density at 600 nm
PCR	Polymerase chain reaction
PMSF	Phenylmethanesulfonylfluoride
PTC	Peptidyl transferase center
ψ	Pseudouridine
RNase	Ribonuclease
rRNA	Ribosomal ribonucleic acid
SDS	Sodium dodecyl sulfate
scaRA	Small Cajal body nucleolar ribonucleoprotein
Tris	Tris(hydroxymethyl)aminomethane
TERT	Telomerase reverse transcriptase
TOF	Tetralogy of Fallot
UTP	Uridine-5'-triphosphate
UV	Ultraviolet
WT	Wild-type
X-DC	X-linked Dyskeratosis Congentia

CHAPTER 1: INTRODUCTION

The chemical modification of nucleotides in RNA is a widespread phenomenon in the cell, and more than 170 distinct chemical modifications are observed in various RNA molecules [1, 2]. Among them, pseudouridine is one of the most common chemical modifications, which was discovered in 1957 by Davis and Allen, and Chon named the abbreviation of pseudouridine (ψ) in 1960 [3, 4]. This chemical modification has frequently been observed in coding and non-coding RNA molecules [5]. Pseudouridine is introduced by standalone pseudouridine synthases or by the RNA-dependent H/ACA snoRNP complex [6, 7]. My thesis focuses on the H/ACA snRNP complex and its role in ribosomal RNA (rRNA) modifications. Pseudouridine has an extra imino group, which stabilizes RNA by improving base-stacking interactions by forming an additional hydrogen bond with water. Moreover, pseudouridine has a unique C-C glycosidic bond, which is more stable than the regular C-N glycosidic bond **(Figure 1.1A)**. In humans and *Saccharomyces cerevisiae,* there are 95 and 47 sites in ribosomal RNA (rRNA) that are modified into pseudouridine, respectively [8, 9].

1.1. Small nucleolar RNAs

The following section provides a brief overview of the different small nucleolar RNAs (snoRNAs), their sequence characteristics, functions, and their occurrence in different organisms. snoRNAs are short non-coding RNAs enriched in the nucleolus. The snoRNAs are classified into two classes based on the characteristic nucleotide motifs present. The C/D snoRNAs (SNORD) fall under class-I, and the H/ACA-box snoRNAs (SNORA) are a class-II snoRNA [10]. C/D-box snoRNAs harbour the RUGAUGA sequence element, also called C-Box, where R is a purine base as well as a D-Box, which is a short sequence motif (CUGA); these motifs are brought in near proximity through the formation of a terminal stem structure which ultimately forms the kink-turn

structural element. Most C/D-box snoRNAs contain an additional pair of conserved C and D boxes, denoted as C′ and D′, respectively. The C/D-box snoRNA base pairs with target RNAs through a short 7–21 nucleotide antisense element (ASE) or a guide region located upstream of the D and/or D′ boxes. The box C/D snoRNPs are mainly involved in the 2′-O-methylation and processing of pre-rRNA [11].

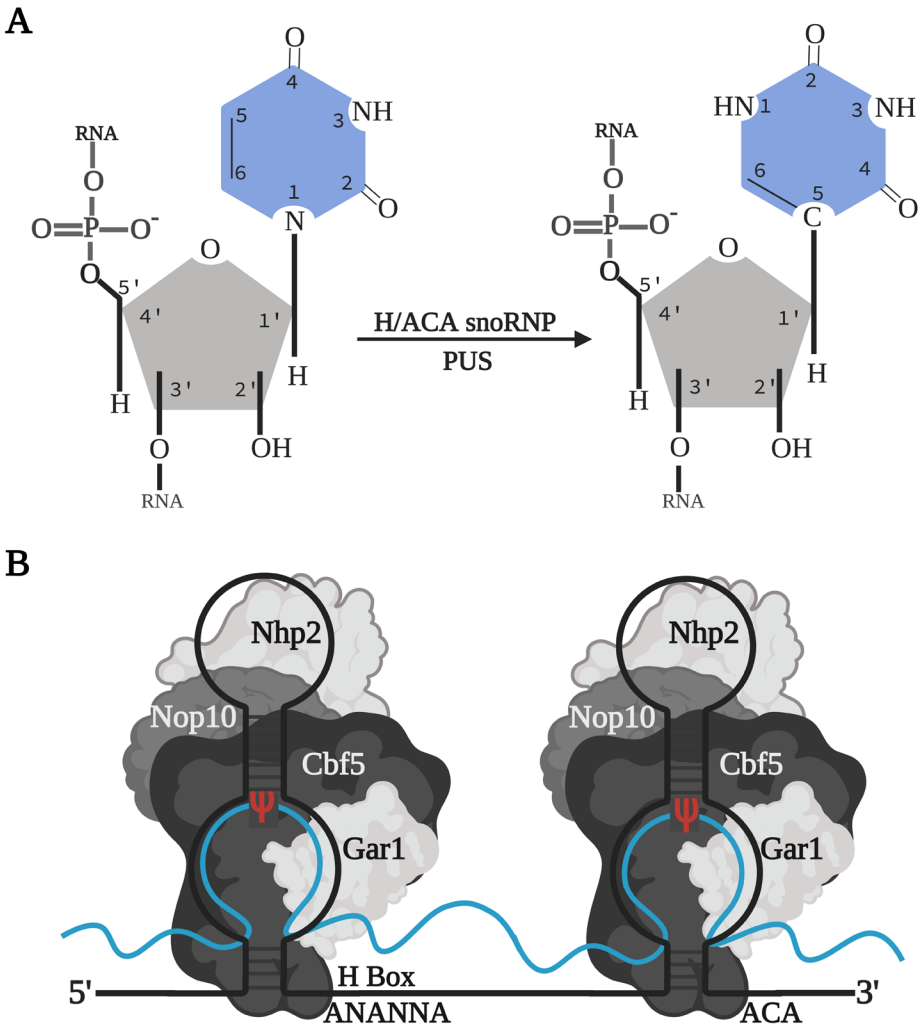

Figure 1.1. Pseudouridine and H/ACA snoRNP complex. A Conversion of uridine to pseudouridine by pseudouridine synthases (PUS) and H/ACA snoRNPs**. B** Structure of H/ACA snoRNP. In general, the eukaryotic H/ACA snoRNA harbours two hairpin structures connected by a hinge region. After the 5' hairpin, they contain a consensus sequence called H-box (ANANNA) and an ACA box located downstream of the second hairpin close to the 3' end. H/ACA snoRNAs are bound by the four H/ACA proteins (Cbf5, Gar1, Nop10 and Nhp2 in *S. cerevisiae*).

Eukaryotic H/ACA snoRNAs have two hairpins with internal loops connected by a hinge region harbouring the H-Box with a conserved ANANNA sequence element, where 'N' denotes any nucleotide and the ACA box near 3′ end. However, kinetoplastids have single hairpin structured H/ACA snoRNAs, which lack the H-Box but harbour the AGA box instead of an ACA box [12] (**Figure 1.1B)**. The H/ACA-box snoRNA antisense regions called pseudouridylation pockets are composed of two short nucleotide stretches with 9 -16 nt, located in the internal loops of the 5′ and 3′ hairpins [13]. Depending on their function, the H/ACA snoRNA molecules are further classified into guide and non-guide H/ACA snoRNAs. The guide H/ACA snoRNA is involved in pseudouridylation, and non-guide H/ACA snoRNA harbour non-conical functions, including the pre-rRNA processing [14, 15]. Additionally, there are orphan H/ACA snoRNAs discovered by the computational screening of the human genome (snoseeker and snoreport), whose functions are yet to be elucidated [16-19]. Moreover, it is important to note that some of the H/ACA snoRNA molecules have been reported to harbour additional sequence elements named CAB box or Alu elements and are therefore called scaRA and AluACA snoRNA, respectively [20]. The scaRNAs are known to accumulate in Cajal bodies, unlike other H/ACA snoRNAs [20, 21]. However, the Alu H/ACA snoRNA fail to form a stable H/ACA snoRNP complex due to their very long or very short 5′ hairpin; however, the 3′ hairpin of Alu H/ACA snoRNA harbour functional structural features. The function of Alu H/ACA snoRNAs has yet to be elucidated, and more than 348 Alu H/ACA snoRNAs have been reported in humans till now [22, 23]. *Saccharomyces cerevisiae* has 29 H/ACA snoRNAs, including 28 guide H/ACA snoRNA molecules involved in site-specific pseudouridylation and one H/ACA snoRNA molecule snR30 engaged in the processing of pre-rRNA. Moreover, one H/ACA snoRNA molecule snR10 is involved in both site-specific pseudouridylation and processing of the pre-rRNA **(Table 1.1)** [24].

Table 1.1. List of H/ACA snoRNA molecules in *S. cerevisiae.*
The complete list of *S. cerevisiae* snoRNA molecules involved in pseudouridylation of rRNA and rRNA processing.

Guide H/ACA snoRNA	Target rRNA	Position of the modification
snR3	25S rRNA	U2129 (25S rRNA), U2133 (25S rRNA), U2264 (25S rRNA)
snR5	25S rRNA	U1004 (25S rRNA), U1124 (25S rRNA)
snR8	25S rRNA	U960 (25S rRNA), U986 (25S rRNA)
snR9	25S rRNA	U2340 (25S rRNA)
snR10	25S rRNA	U2923 (25S rRNA)
snR11	25S rRNA	U2416 (25S rRNA)
snR31	18S rRNA	U999 (18S rRNA)
snR32	25S rRNA	U2191 (25S rRNA)
snR33	25S rRNA	U1042 (25S rRNA)
snR34	25S rRNA	U2826 (25S rRNA), U2880 (25S rRNA)
snR35	18S rRNA	U1191 (18S rRNA)
snR36	18S rRNA	U1187 (18S rRNA)
snR37	25S rRNA	U2944 (25S rRNA)
snR42	25S rRNA	U2975 (25S rRNA)
snR43	5.8S, 25S rRNA	U73 (5.8SrRNA), U966 (25S rRNA)
snR44	18S, 25S rRNA	U106 (18S rRNA), U1056 (25S rRNA)
snR46	25S rRNA	U2865 (25S rRNA)
snR49	18S, 25S rRNA	U120, U211, U302 (18S rRNA) U990 (25S rRNA)
snR80	18S, 25S rRNA	U759 (18S rRNA) U776 (25S rRNA)
snR81	25S rRNA, U2	U42 (U2 RNA) U1052 (25S rRNA)
snR82	25S rRNA	U1110, U2349, U2351 (25S rRNA)
snR83	18S rRNA	U1290 (18S rRNA), U1415 (18S rRNA)

snR84	25S rRNA	U2266 (25S rRNA)
snR85	18S rRNA	U1181 (18S rRNA)
snR86	25S rRNA	U2314 (25S rRNA)
snR161	18S rRNA	U632 (18S rRNA), U766 (18S rRNA)
snR189	18S, 25S rRNA	U466 (18S rRNA) U2735 (25S rRNA)
snR191	25S rRNA	U2258 (25S rRNA), U2260 (25S rRNA)
Non-Guide H/ACA snoRNA	**Target rRNA**	**Role in rRNA processing**
snR10	Pre-rRNA	35S rRNA processing
snR30	Pre-rRNA	35S rRNA processing

1.2. Structure of eukaryotic H/ACA snoRNP complex

This section describes the current knowledge on structural details of the eukaryotic H/ACA snoRNP complex and composition. The mature H/ACA snoRNP complex contains four core proteins, namely, Cbf5 (Dyskerin in humans), Gar1, N0p10 and Nhp2 (L7Ae in archaea) **(Figure 1.2A)**. All these four core proteins are highly basic; moreover, Cbf5 and Gar1 harbour N- and C-terminal unstructured tails named KKE/D and GAR domains, respectively. Eukaryotic Cbf5 has highly conserved TRUB and PUA domains, along with the KKE/D repeats. High-resolution structural information for many H/ACA snoRNP complex core proteins are available, ranging from archaea to humans, like the Cbf5•Nop10•Gar1 (CNG) ternary complex from *S. cerevisiae* or the cryo-electron microscopy (cryo-EM) structure of the substrate-bound human telomerase holoenzyme, which harbours a H/ACA snoRNP complex [25, 26]. The protein-protein interaction pattern of the Cbf5•Nop10•Gar1 (CNG) complex from *S. cerevisiae* is similar to that of the archaeal H/ACA snoRNP **(Figure 1.2B)** [26]. In general, Nop10 and Gar1 bind at two orthogonal faces of the catalytic domain of Cbf5 [26]. The eukaryotic H/ACA snoRNA has the ability to bind

with two sets of core proteins resulting in a bipartite structure, as it harbours two hairpins, as discussed in the previous section. There is no current detailed structural information on the eukaryotic H/ACA snoRNP complex alone available, except for human telomerase holoenzyme, which contains a H/ACA snoRNP lobe as part of its structure [27].

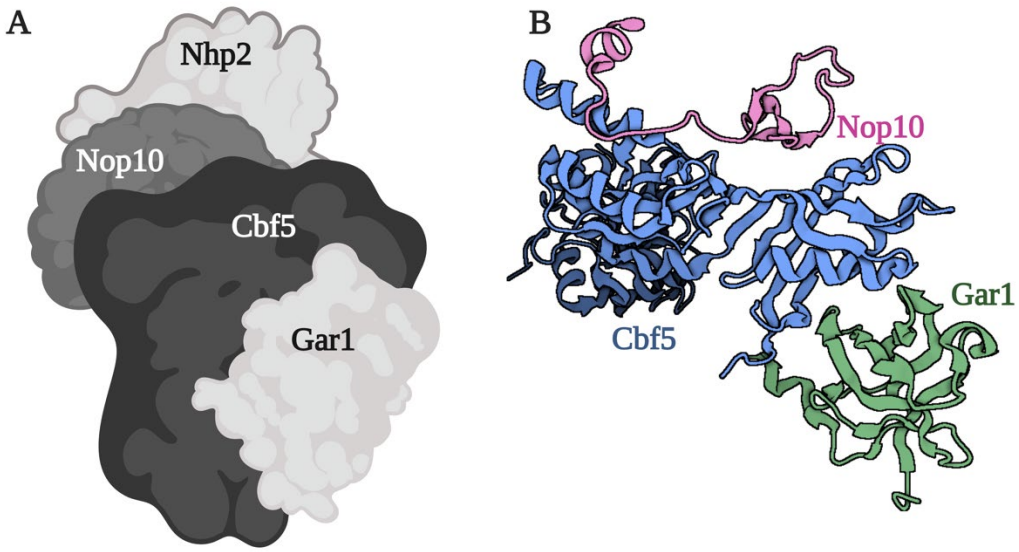

Figure 1.2. The core proteins of the H/ACA snoRNP complex. A The schematic representation of proteins involved in the formation of the H/ACA snoRNP complex. Cbf5, Nop10, Gar1, and Nhp2 represent accordingly. **B** X-ray Crystal Structure of *S. cerevisiae* Cbf5, Nop10, and Gar1 protein complex, Cbf5 in blue, Nop10, and Gar1are in pink and green, respectively (PDB ID:3U28).

In humans, the H/ACA snoRNP complex is also part of the telomerase enzyme [25]. The telomerase holoenzyme structures composition and organization were revealed using cryo-EM and homology modelling **(Figure 1.3)**. This telomerase holoenzyme structure provides insights into the architecture of the eukaryotic H/ACA snoRNP complex, as the human telomeric RNA (hTR) is a H/ACA snoRNA molecule (451nt in length) and has a H/ACA lobe, which binds to H/ACA snoRNP core proteins, and a catalytic lobe, which binds to telomeric reverse transcriptase (TERT) [27-29]. The hTR has eight stem loops with a pseudoknot (t/PK) and conserved regions 4 and 5 (CR4/5) **(Figure 1.4A)**. The H/ACA lobe of hTR harbours two sets of H/ACA heterotetrameric

proteins, bind to two RNA hairpins, and the 3′ hairpin loop interacts with the single holoenzyme TCAB1 subunit. Also, the two H/ACA protein heterotetramers interact with each other, and this interaction is mediated by the two Dyskerin molecules [25, 27]. While the first set of H/ACA proteins interacts with the 5′ hairpin P4 stem exclusively via Dyskerin, the second set interacts with the 3′ hairpin P7 stem and P8 stem loop. Dyskerin interacts with the P7 stem similarly to the 5′ hairpin P4 and continues to P8. The P8 stem-loop interacts with Nop10, Nhp2, and TCAB1. The binding of Dyskerin at the 5′ hairpin (P4) is sufficient to anchor the entire core complex **(Figure 1.4B)** [25, 27]. The single subunit of TCAB1 interacts with the 3′ hairpin CAB box (Cajal body-specific localizing box elements), Dyskerin in the second set of H/ACA proteins, and also Gar1. This structure provides significant insights into human Dyskeratosis Congenita and Hoyeraal–Hreidarsson disease mutations. It is clear from the structure that most of the Dyskerin mutations were mapped in the interface of two Dyskerin molecules, and this site is close to its RNA-binding surface [25, 27]. However, the Dyskerin R158 mutation is located near TCAB1, whereas R34 of Nop10 and V126 of Nhp2 are present near the P8 stem of hTR, respectively. This structure-based study revealed that telomerase deficiency mechanisms are associated with disrupted interactions between two Dyskerin/Cbf5 molecules [25, 27, 30-32].

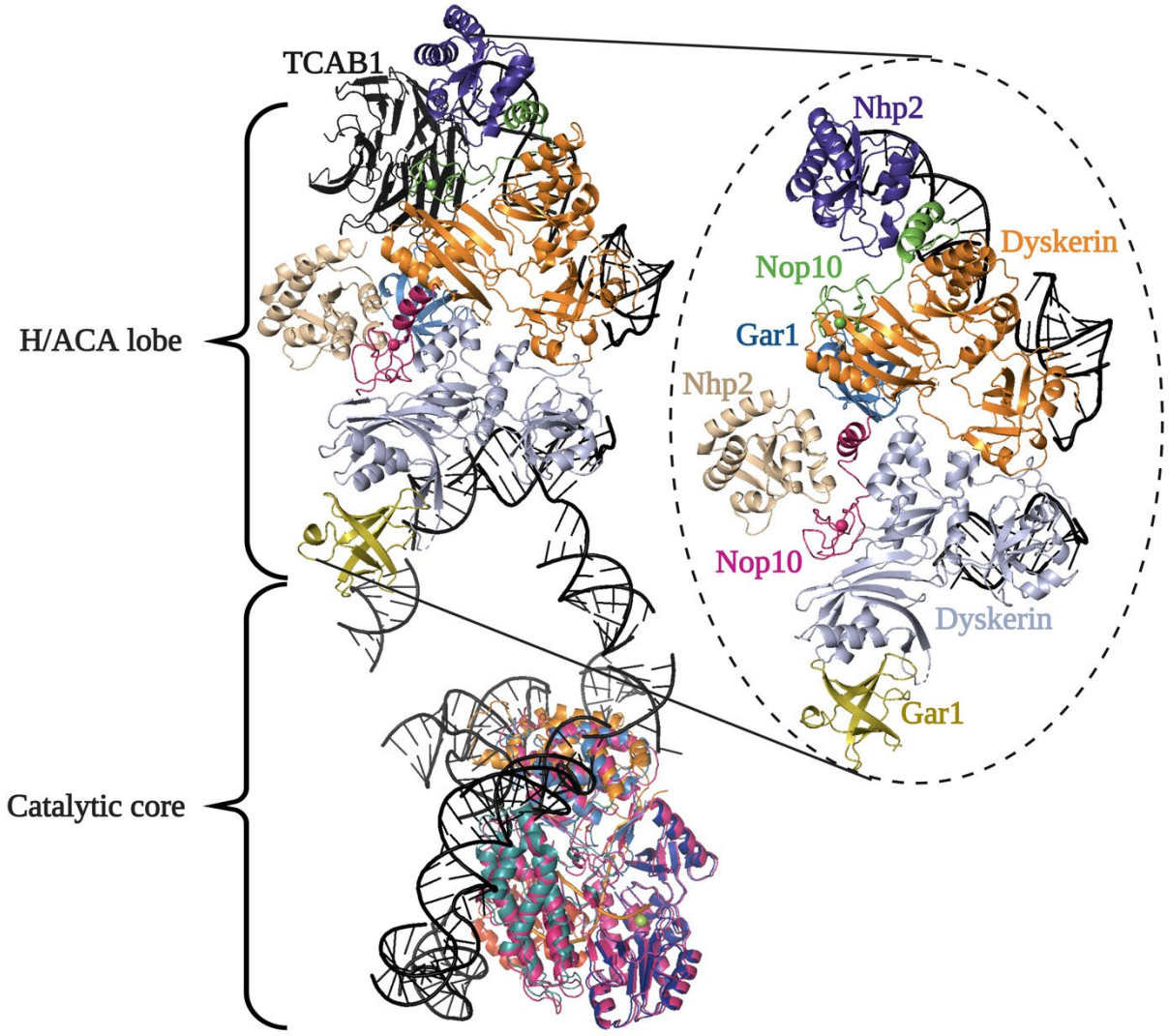

Figure 1.3. Cryo-EM structure of human telomerase holoenzyme. The Cryo-EM structure of the human telomerase holoenzyme complex harbours a H/ACA lobe and catalytic core. The hTR (human telomeric RNA) is represented in black. The H/ACA snoRNP complex is zoomed out and represented separately. The two sets of H/ACA snoRNP proteins are highlighted and labelled accordingly. Only the H/ACA snoRNP complex is shown in the zoom-in, and I excluded TCAB1 (Telomerase Cajal body protein 1) in the zoom-out illustration.

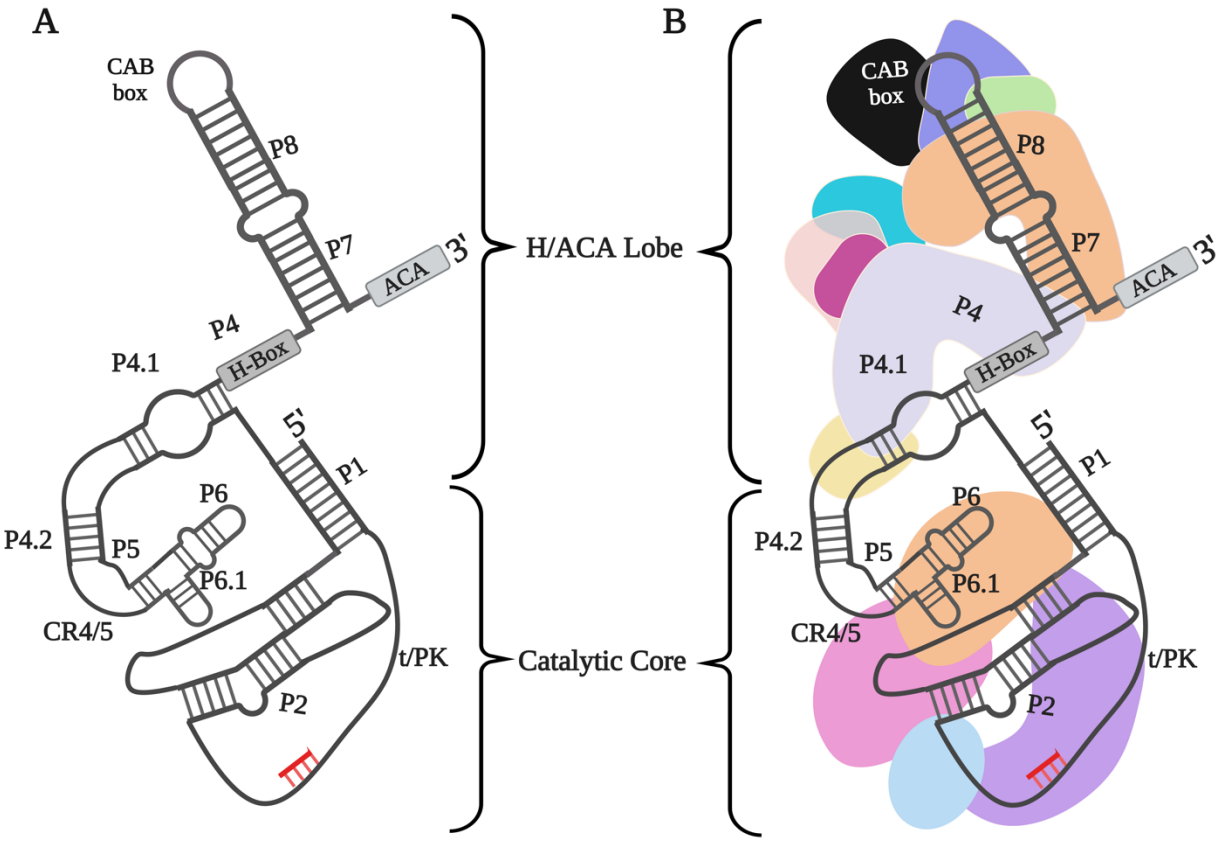

Figure 1.4. Illustration of human telomerase holoenzyme. A Structure of human telomeric RNA (hTR): The human telomeric RNA harbour H and ACA boxes along with the CAB box, which is specific for localization into Cajal bodies. hTR is composed of 8 stem-loop structures labelled P1-P8 accordingly. Further, the conserved regions 4 and 5 (CR4/5) and pseudoknot (t/PK) are labelled, and the substrate RNA is shown in red **B** Schematic representation of protein arrangement in the human telomerase enzyme: The first set of H/ACA snoRNP complex proteins in H/ACA lobe are Dyskerin (orange), Nop10 (green), Nhp2 (purple), Gar1 (blue). The second set represented in Dyskerin (gray), Nop10 (magenta), Nhp2 (wheat), Gar1 (yellow). The catalytic core binds the TERT enzyme, which is composed of TRBD (RNA-binding domain in orange), RT (reverse transcriptase domain in purple), TEN (the N-terminal domain in blue), and CTE (C-terminal extension in pink).

1.3. Importance of H/ACA snoRNP core proteins

The mature H/ACA snoRNP complex harbours four core proteins: Cbf5 (Dyskerin), Gar1, Nop10, and Nhp2. Previous studies revealed that all these proteins are essential for cell viability in *S. cerevisiae* except Gar1 [33]. However, Gar1 is required for stable association of the box H/ACA snoRNAs with the pre-rRNA and is also necessary for substrate turnover during the pseudouridine formation [26, 34]. Previously, it was reported that Gar1 binds to the Cbf5 catalytic domain, and the absence of Gar1 abolished the pseudouridine formation [26, 35]. Moreover, the archaeal Gar1 protein is reported to improve the catalysis of pseudouridine. It was predicted that Gar1 might be involved in the correct positioning of substrate RNA in the catalytic site of Cbf5 in the archaea [36]. According to iPTMnet source, in human Gar1 there are 25 sites that undergo post-translational modifications, and in yeast, there are 20 sites known to be modified post-translationally. Out of these 20 sites, two are ubiquitination sites located in the N -terminal of Gar1 at positions K59 and K77 [37]. Moreover, 18 sites of yeast Gar1 are known to be methylated, and all of these modification are sites are located in the N- and C-terminal GAR domains of the Gar1 protein. In yeast Cbf5, there are 14 known ubiquitination sites and 11 phosphorylation sites; human Dyskerin (DKC1) harbours 49 post-translational modifications, and out of these 49 modifications, 13 are acetylations and 19 phosphorylations. Interestingly, most of these phosphorylation sites are located in C -the terminal KKE/D domain in both yeast and human Cbf5/Dyskerin protein [38, 39]. In humans, post-translational modifications of Dyskerin were also reported near the C-terminal nuclear/nucleolar localization signal (N/NoLS), which has four SUMOylation sites. Moreover, the nucleolar localization of Dyskerin is mediated by SUMOylation at K467 located at C-terminal N/NoLS. And the K467 SUMOylation is also reported to mediate the interaction between Dyskerin and GAR1 [40]

Nhp2 and Nop10 depletion results in defective A_1 and A_2 site cleavage of pre-rRNA and accumulates 35S pre-rRNA and 23S pre-rRNA intermediates [41]. Furthermore, Nop10 is required to stabilize the H/ACA snoRNPs and depletion of Nop10 results in disruption of pseudouridine formation in 25S rRNA [41]. Unlike L7Ae (archaeal homologue of Nhp2), eukaryotic Nhp2 also interacts with Nop10 and the H/ACA snoRNA, whereas L7Ae solely binds to H/ACA snoRNA [7, 26, 42, 43].

The Cbf5 protein is an RNA-dependent pseudouridine synthase, and a single point mutation (D95N) in the catalytic site results in a complete loss of pseudouridine formation leading to severe growth defects in *S. cerevisiae* [43, 44]. The genetic depletion of Cbf5 inhibits both pre-rRNA processing and the formation of pseudouridines in the pre-rRNA and leads to the co-depletion of snR30 [45]. This depletion of snR30 leads to the loss of 18S rRNA synthesis since 18S rRNA maturation requires snR30 [46, 47]. Moreover, the depletion of Cbf5 is also associated with a substantial delay in the processing of 27S and 7S pre-rRNAs in *S. cerevisiae* cells and also leads to cell cycle arrest [44, 45, 48]. Moreover, a truncated version of Cbf5 lacking C-terminal KKE/D extensions leads to delayed cell division in *S. cerevisiae*; however, it was reported that the KKE/D extension of Cbf5 has no role in efficient pseudouridine formation [26, 49]. These results suggest that Cbf5/Dyskerin is a central core protein in the H/ACA snoRNP complex, which functions as pseudouridine synthase and might harbour other critical cellular functions yet to be uncovered.

1.4. Genomic organization of H/ACA snoRNA genes and assembly of mature H/ACA snoRNP complex

The RNA-Polymerase II (Pol II) is responsible for the transcription of snoRNA molecules, and these snoRNAs are predominately located in the nucleoli [50]. The genomic

organization of snoRNA genes varies between different species. In *S. cerevisiae*, most of the snoRNA molecules are produced from independent genes with few exceptions; all 76 *S. cerevisiae* snoRNA molecules are synthesized from 64 transcription units **(Figure 1.5A)** [50]. In plants and insects, the snoRNA molecules are polycistronic in origin, i.e., a single transcript can be processed into more than one snoRNA molecule **(Figure 1.5B)**. In primates (humans), most of the snoRNA molecules are derived from intronic regions of pre-mRNA, except for human telomeric RNA (hTR) **(Figure 1.5C)**. More recently, it was reported that the human genome encodes 2064 snoRNA molecules, of which 1391 are C/D snoRNAs, and 651 are H/ACA snoRNA molecules [51, 52].

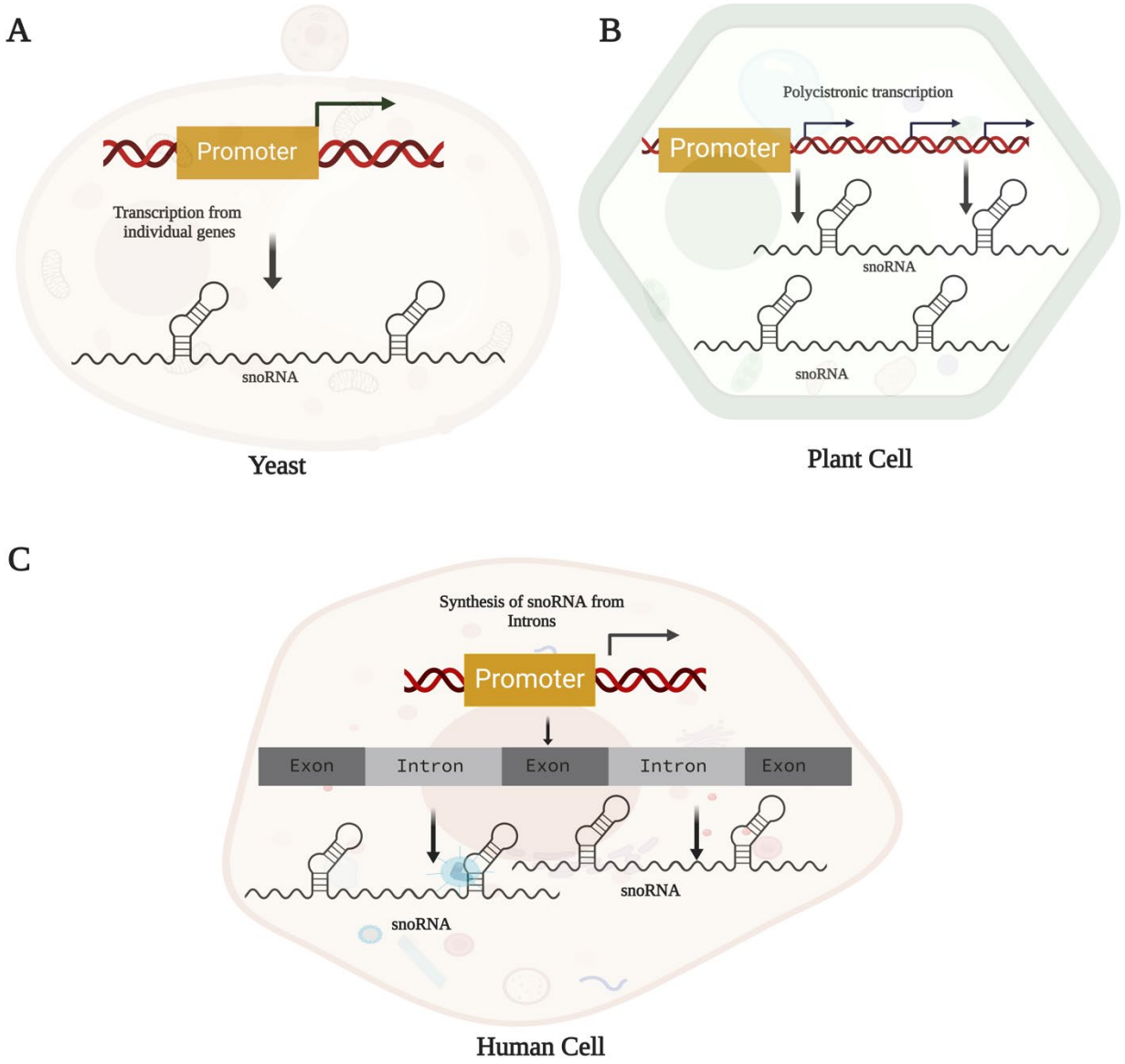

Figure 1.5. Genomic organization of snoRNA genes across the different organisms. A Transcription of snoRNA in *S. cerevisiae* occurs from independent genes. **B** snoRNAs in plants/insects are encoded as polycistronic genes. **C** snoRNAs in primates are processed from introns of pre-mRNA.

The assembly of a H/ACA snoRNP complex takes place co-transcriptionally and requires additional assembly factors and the assistance of chaperones and co-chaperon systems [53, 54]. In eukaryotes, the mature H/ACA snoRNA molecules interact with the proteins Cbf5 (Dyskerin in humans), Nop10, Gar1, and Nhp2 to form a stable H/ACA snoRNP complex [26, 55]. All four of these proteins have highly conserved orthologs ranging from archaea to humans, and the assembly of the H/ACA snoRNP complex in eukaryotes requires two additional proteins named Shq1 and Naf1 [56-58]. Shq1, a structural homolog of HSP90, acts as a chaperone to stabilize Cbf5 immediately after Cbf5 is translated in the cytoplasm since Cbf5 alone is unstable and capable of forming aggregates [58, 59]. Shq1 has two functional domains, a CHORD-Sgt1 domain (CS) and a C-terminal Shq1 specific domain (SSD); both domains of Shq1 are shown to interact with Dyskerin/Cbf5 by forming a tight clamp around Dyskerin /Cbf5 [60]. Furthermore, the SSD domain mimics the RNA molecule and interacts with the Cbf5 RNA binding domain in the initial stages of the H/ACA snoRNP complex assembly [61]. Given that Cbf5 can bind to any double-stranded RNA molecules in a non-specific manner, the interaction between Shq1 with Cbf5 might also prevent the unspecific binding of Cbf5 with other RNA molecules [55, 62]. Later, Naf1 interacts with Cbf5/Dyskerin, Nhp2, and Nop10 and with the phosphorylated C-terminal domain of RNA polymerase II (CTD). This CTD is predicted to help recruit the other H/ACA snoRNP core proteins to the site of the H/ACA snoRNA transcription [53, 63]. Naf1 harbours a core domain that shares similar structural features to the core domain of Gar1 [64]. The Gar1-like core domain of Naf1 interacts with Cbf5; however, the affinity is likely weaker than the affinity of Gar1 to Dyskerin/Cbf5. Later, the release of Shq1 from Shq1-dyskerin, Nop10, Nhp2, and Naf1 complex is mediated by the interaction of the R2TP complex [65]. In brief, the R2TP is a co-chaperon system that interacts with HSP90 and is composed of Ruvb1, Ruvb2, Tah1 and Pih1 proteins [66].

It was shown that the Ruvb1/2 proteins of the R2TP complex help release the Shq1 from Dyskerin/Cbf5 **(Figure 1.6)** [65, 67]. In later steps of H/ACA snoRNP complex maturation, the Naf1 is replaced with Gar1 in the Cajal body. It is shown that the Naf1 protein can dimerize, and this dimerized Naf1 has a lower affinity to Cbf5/Dyskerin than Gar1; this could be one of the mechanisms for how Naf1 is replaced with Gar1 in the final stages of H//ACA snoRNP assembly [26, 68]. In humans, it was also shown that the survival of motor neurons complex (SMN complex) along with Nopp140 plays a role in replacing Naf1 with Gar1 and is involved in the maturation of the H/ACA snoRNP complex [69-71] **(Figure 1.6)**.

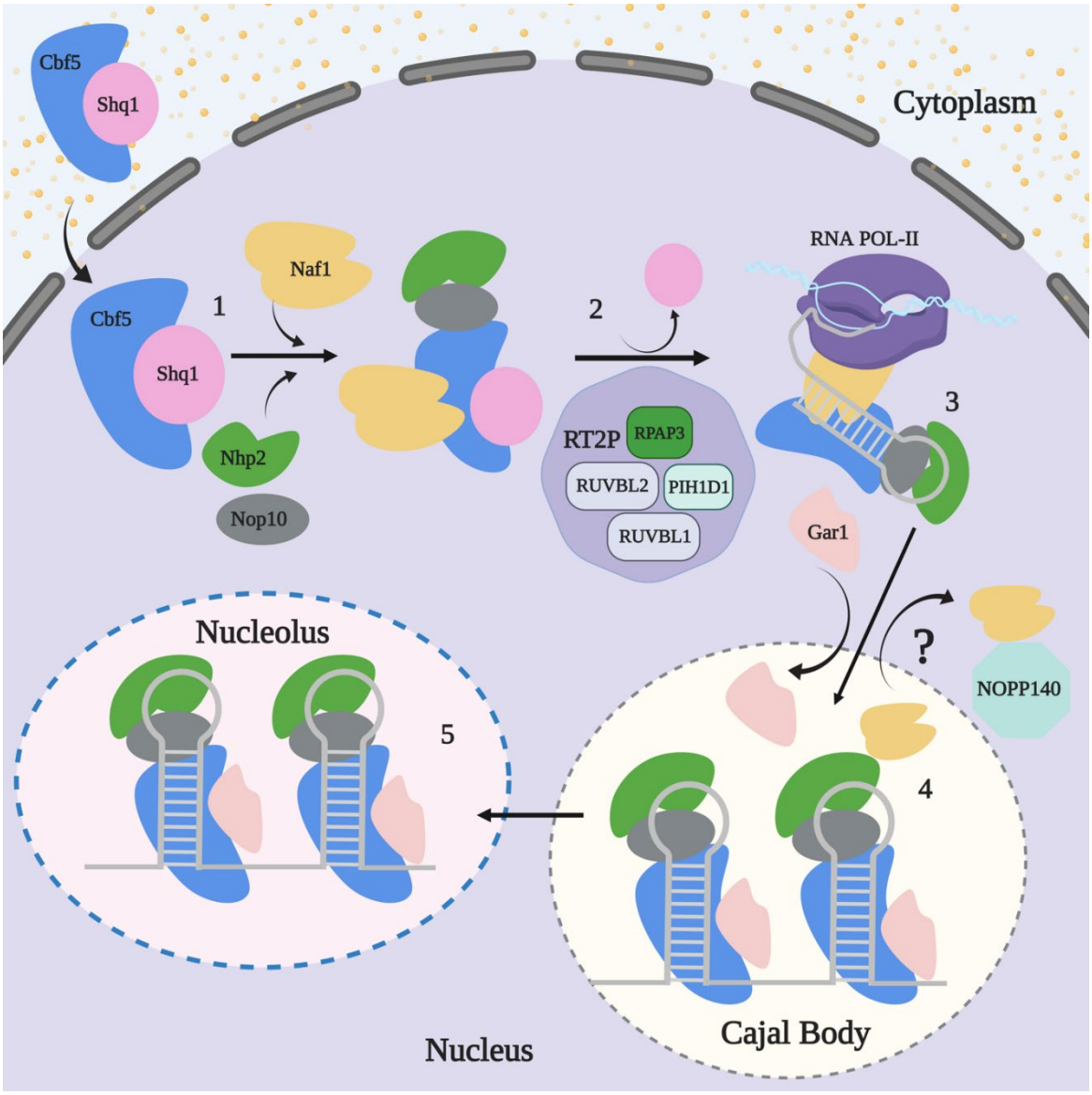

Figure 1.6. Assembly of the eukaryotic H/ACA snoRNP complex. Assembly of the eukaryotic H/ACA snoRNP is a coordinated pathway that requires additional assembly factors. The Cbf5/Dyskerin interacts with Shq1 in the cytoplasm, and this complex subsequently localizes into the nucleus. Upon the interaction of Cbf5/Dyskerin with Naf1-Nhp2-Nop10 proteins, the Shq1 protein is released from Cbf5/Dyskerin with the help of the R2TP complex composed of RUVBL1/2 (Pontin), Reptin, RPAP3, and P1H1D1 [65]. Later the Naf1-Cbf5-Nop10-Nhp2 complex interacts with phosphorylated RNA polymerase II carboxy-terminal domain (CTD) domain, and it is suggested that Naf1 recognizes the H/ACA snoRNA as it has RNA binding ability [53]. During maturation of the H/ACA snoRNP complex, the Naf1 protein is replaced with Gar1 protein either by forming self-dimers or with the help of multiple co-factors like SMN and Nopp140 proteins [67, 70]. This remodelling or replacement of Naf1 with Gar1 occurs in Cajal bodies in higher eukaryotes [68]. In the final step, the H/ACA snoRNP complex is transported into the nucleolus, which is involved in rRNA modifications and processing.

1.5. H/ACA snoRNP guided chemical modifications and processing of pre-rRNA

Functionally, the H/ACA snoRNPs are involved in rRNA processing and pseudouridine formation. However, in humans, the H/ACA snoRNA exerts various functions apart from its canonical role in rRNA modification and processing, such as serving as precursors for producing miRNA and maintaining telomeres [7]. The following sections will describe the role of the H/ACA snoRNP complex in the processing of rRNA and pseudouridine formation, along with the importance of pseudouridines for rRNA stability and function.

1.5.1 . Mechanism of H/ACA snoRNP in pseudouridine formation in rRNA

The pseudouridylation pocket of a H/ACA snoRNA possesses ~3-14 complementary nucleotides on either side of the target U, leaving one unpaired nucleotide next to the target U (NU) in a sequence-specific manner [8, 55, 72, 73]. A study conducted by Caton *et al.* provided evidence that the H box and ACA box elements of H/ACA snoRNAs are essential for proper pseudouridylation activity [55]. The deletion of the ACA box significantly reduces pseudouridine formation in both pockets. In contrast, the deletion of the H box only reduces the formation of pseudouridine on the substrate RNA recognized by the 5′ pseudouridylation pocket [55].

Separate studies conducted by the Kothe group and the Ye group showed that a single hairpin of a H/ACA snoRNA molecule has the ability of uridine isomerization on target RNA, albeit with reduced activity compared to a two-hairpin H/ACA snoRNP [55, 74]. Based on these studies, it is evident that in eukaryotes, the two-hairpin structure of H/ACA snoRNAs is an evolutionary advantage over the single-hairpin structure. In 2007, two independent groups, Moore and Feigon, showed that H/ACA snoRNA molecules can interact with substrate RNA without core proteins [75, 76]. They performed NMR experiments with H/ACA snoRNA U65 (snR34 in *S. cerevisiae*) alone and in the presence of substrate RNA to visualize the base-pairing interaction

between the guide RNA and target RNA in the absence of H/ACA snoRNP proteins [75, 76]. However, the affinity of guide RNA and substrate RNA in the absence of proteins is low with a high K_D (~200-300 μM), suggesting that the interaction between substrate RNA and guide RNA is strengthened by the core proteins [75]. A recent study from the Kothe group revealed that substitution or insertion of nucleotides in the substrate RNA near the target uridine significantly reduced the pseudouridylation of the target U [72]. The authors systematically substituted the nucleotides at both ends of short substrate RNAs guided by snR34 (U65 in humans), which is involved in the pseudouridylation at two sites in 25S rRNA (U2826 and U2880). They observed that the interaction between the guide RNA and the target RNA requires a minimum of 8 base pairs, and the base pairing needs to be a bipartite [55, 72]. Moreover, both cognate and near-cognate substrate RNA bind with similar nM affinity to H/ACA guide snoRNA. Interestingly, despite the similar affinity, the H/ACA guide RNA chose the cognate substrate RNA over the short near-cognate RNA substrate while modifying the target U [55, 72]. This finding indicates that the H/ACA snoRNA-directed pseudouridine formation is site-specific.

1.5.2. Mechanism of H/ACA snoRNA in pre-rRNA processing

The synthesis of rRNA and its assembly into ribosomes is a sophisticated and coordinated process that extends from the nucleolus to the cytosol **(Figures 1.7 & 1.8)**. The rRNAs are central molecules in all ribosomes and are essential for peptide bond formation and also for mRNA decoding [77]. In *S. cerevisiae*, the large ribosomal subunit (LSU) harbours 25S, 5.8S, and 5S rRNAs, and the small ribosomal subunits (SSU) harbours 18S rRNA, respectively [78, 79]. The pre-rRNA is synthesized as a 35S pre-rRNA in *S. cerevisiae* (47S pre-rRNA in humans) and undergoes a series of chemical modifications and stepwise processing events to remove the external transcribed spacer (ETS) and internal transcribed spacer (ITS) regions **(Figure 1.7)** [80].

It should be mentioned that the processing and chemical modifications of pre-rRNA are remarkably complex and coordinated, involving various proteins and different cellular compartments.

Non-guide H/ACA RNAs are required to process pre-rRNA into mature rRNA. One of the essential non-guide RNAs is snR30 (in humans U17), which is involved in pre-rRNA processing and cleavage of 18S rRNA from the 35S pre-rRNA [81, 82]. snR30 is one of the longest (608 nt) H/ACA snoRNA in *S. cerevisiae* and contains a non-canonical secondary structure. In early 1993, Morrissey *et al.* discovered that snR30 is involved in the 18S rRNA processing [46]. Morrissey *et al.* and Tollervey *et al.* further investigated the role of snR30 in pre-rRNA processing and showed that upon depletion of snR30, the levels of 18S rRNA are reduced while 23S rRNA accumulates. Moreover, snR30 was reported to be involved in the cleavage of A_0, A_1 and A_2 sites of pre-rRNA in both yeast and humans [46, 47, 81]. In subsequent years, the Kiss lab revealed that snR30 harbours two evolutionarily conserved sequence elements, m1 (AUAUUCCUA) and m2 (AAACCAU), located in the 3′ hairpin. These sequence elements base-pair with rm1/rm2 motifs in pre-18S rRNA in expansion segment 6 (ES6) and are involved in the maturation of the 18S rRNA [82, 83]. A recent study from the Kothe lab showed that this interaction between snR30 and rm1/rm2 of ES6 is stabilized by the interaction of Utp23 with snR30 [84]. In addition to snR30, the *S. cerevisiae* H/ACA snoRNA snR10 is involved in the 18S rRNA processing [85]. Interestingly, snR10 is also responsible for the pseudouridylation of 25S rRNA at position U2923. Moreover, snR10 deletion is not lethal in yeast; however, it leads to a cold sensitivity [8, 73, 86, 87]. In short, cleavage at the A_0, A_1, and A_2 sites are essential for the maturation of 18S rRNA, and U3, U14 (C/D snoRNAs) and snR30 (H/ACA snoRNAs) along with the proteins Utp23, Utp24, Rrp5 and Rcl1 are essential factors involved in this process [82, 88-91].

1.5.3. Brief overview of ribosome biogenesis in *S. cerevisiae*

In *S. cerevisiae,* the 35S pre-rRNA undergoes co-transcriptional cleavage by large RNP complexes and ribonucleases [92]. Formation of mature 18S rRNA from 35S pre-rRNA requires the removal of the 5′-ETS and processing of ITS1, an evolutionarily conserved 11 nucleotides stretch at 5'-ETS region requires U3 snRNA at 5'-ETS and the U3 snRNA plays a crucial role in the cleavage of 5'-ETS regions and acts as a chaperone for properly folding RNA molecules [89, 93, 94]. The 35S pre-rRNA is converted into 33S pre-rRNA by cleavage of the A_0 site, followed by the formation of 32S pre-rRNA by cleavage of the A_1 site. Utp24 is involved in processing A_0, A_1, and A_2 sites with the help of Rcl1 and Rrp5, which are required to cleave the A_2 site located at ITS1 and lead to the formation of the 20S and 27S pre-rRNAs [95, 96]. The 20S pre-rRNA is directly processed into 18S rRNA by an endonucleolytic cleavage at site-D by the Nob1 enzyme in the cytosol [97].

The 27S pre-rRNA can be processed differently by a series of exo- and endonucleases into 5.8S and 25S rRNA molecules. Cleaving the C2 site produces the 26S rRNA along with 7S rRNA, which is further processed into 25S and 5.8S rRNA by rapid exonuclease activity. In detail, the nuclear exosome involved in the processing of the 3' end of 7S rRNA with the help of Mtr4 (which is an RNA helicase) [98, 99], which leads to the production of 6S rRNA (5.8S rRNA with +30nt extension). Further, this will be processed by Rrp6 and produce a 6S intermediate with eight nucleotide extensions to 5.8S rRNA [100]. Finally, the Ngl2 process the 3' end of 6S rRNA and leads to the production of mature 5.8S rRNA in the cytoplasm [101].

On the other hand, Rat1 and Xrn1 are involved in processing 5' extension by 26S rRNA by their exonuclease activity and produce mature 25S rRNA [102]. Upon successful processing and chemical modifications, these rRNA molecules are packed into the SSU and LSU and are

transported into the cytoplasm. These Large Sub Unit (LSU) and Small Sub Unit (SSU) undergo final maturation before forming fully functional ribosomes **(Figure 1.8)**.

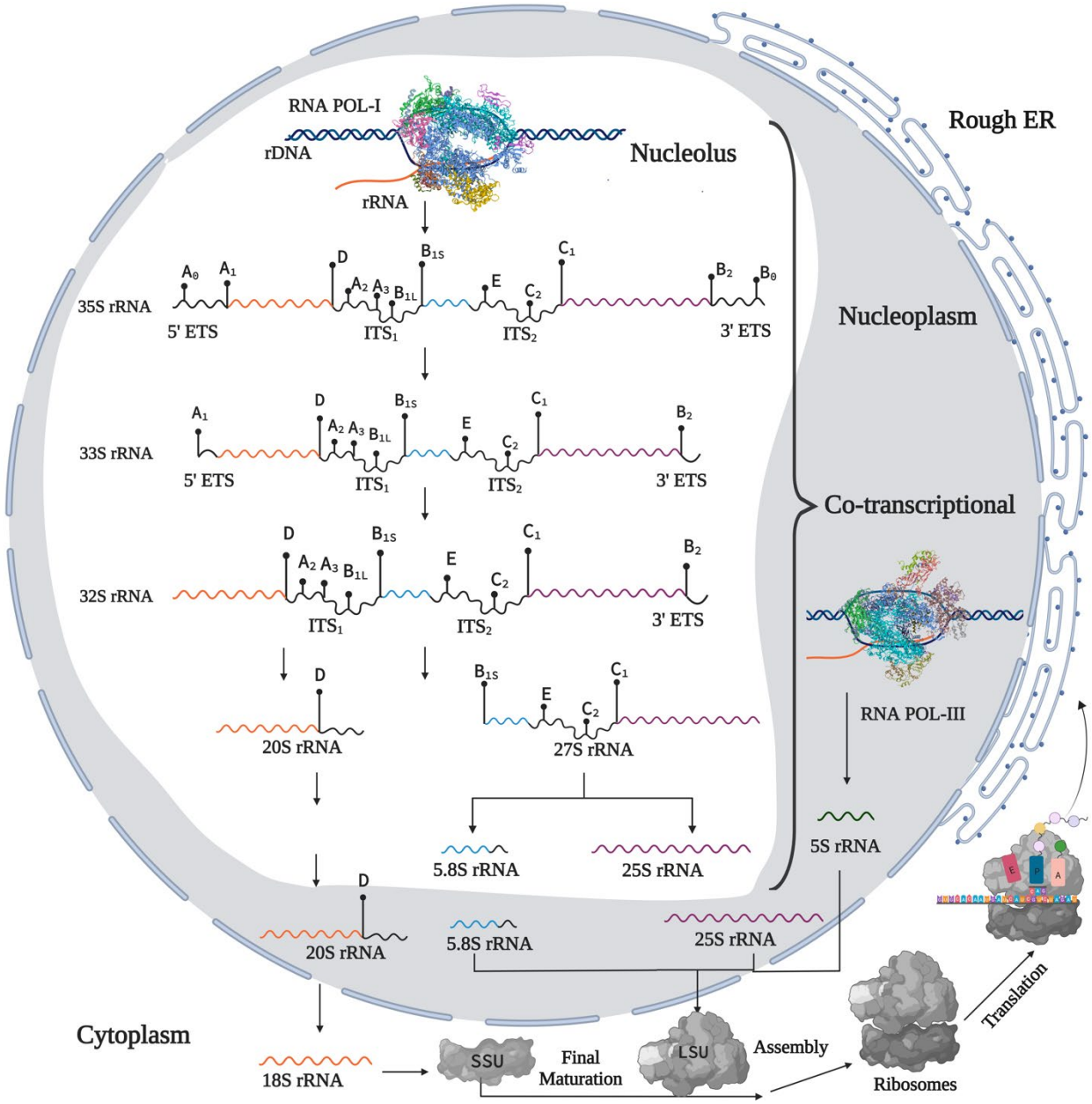

Figure 1.7. Pre-rRNA processing in *S. cerevisiae*. The pre-rRNA synthesized by RNA Pol-I (PDB ID:6HLS) in the nucleolus undergoes co-transcriptional exo/endonucleolytic cleavage by multiple enzymes and is processed into 18S (in orange), 25S (in purple), and 5.8S (in blue) rRNA. In eukaryotes, the small subunit rRNA sequences are located at the 5′ end, and the large subunit rRNA sequences are located at the 3′ end of the pre-rRNA. In *S. cerevisiae*, RNA polymerase I (Pol-I) transcribes a 35S rRNA precursor, which has external and internal transcribed spacers: the 5′ external transcribed spacer (5′ ETS), 3′ ETS, internal transcribed spacer 1 (ITS1) and ITS2. In the nucleoplasm (grey), Pol-III (PDB ID:5FJ9) separately transcribes the 5S rRNA (green), which is later integrated into the large subunit. In this figure, the depiction of 25S rRNA processing is incomplete since it is a simplified overview picture of pre-rRNA processing.

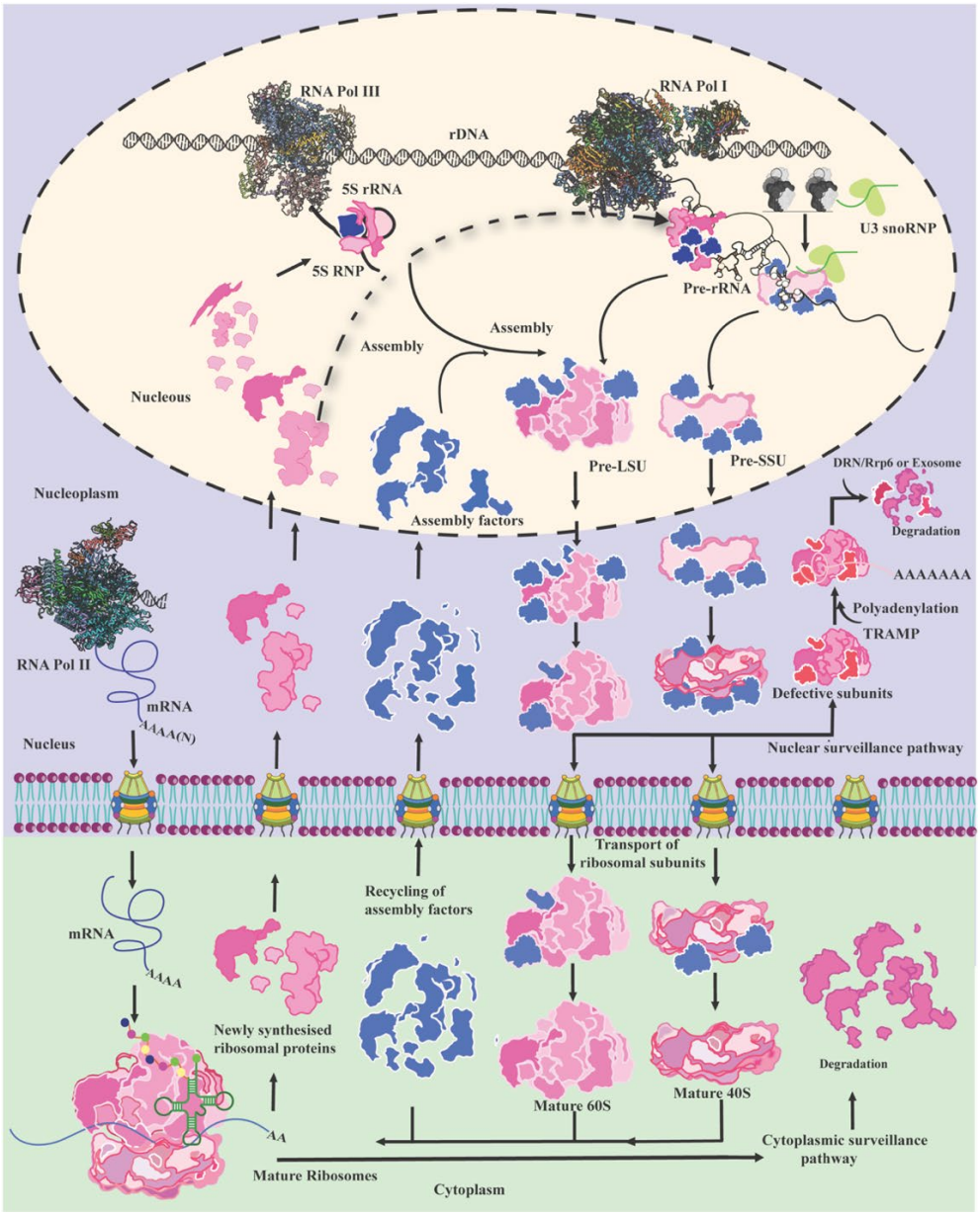

Figure 1.8. General description of ribosome biogenesis. In the nucleolus, Pol I and Pol III transcribe the pre-rRNA, 35S and 5S rRNAs, respectively. Ribosomal proteins and transacting elements are translated in the cytoplasm and imported into the nucleus. However, late ribosomal proteins assembly into the ribosome occurs in the cytoplasm in the later stages of ribosome biogenesis. During early ribosome biogenesis, the U3 snoRNP mediates pre-rRNA cleavage at the A_2 site to separate the small and large subunits. Moreover, it is yet to be elucidated when the incorporation of 5S RNP into the nascent large subunit occurs. Further, both the small and large subunits subsequently undergo maturation steps and export into cytosol independently of each other. The final maturation of ribosomes takes place in the cytosol, and the transacting assembly factors that migrate from the nucleus to the cytoplasm, along with the pre-ribosomal particles, are recycled to participate in the next round of ribosome biogenesis.

1.6. Importance of pseudouridine modification on ribosome function

Pseudouridylation is one of the most abundant chemical modifications observed in rRNA, and nearly 47 sites of rRNA undergo pseudouridylation in *S. cerevisiae* [103, 104]. The 25S rRNA has 31 pseudouridines, 14 pseudouridines are present in 18S rRNA, and each 5S and 5.8S have one pseudouridine within them [105]. This chemical modification has been known to have an impact on the decoding regions of rRNA and tRNA binding sites, suggesting that rRNA modifications play important roles in regulating ribosome function and are required for efficient mRNA translation [4]. The absence of pseudouridylation caused a delay in 18S rRNA processing in *S. cerevisiae*. For example, the deletion of snR35, which is involved in the formation of pseudouridine at hypermodified site 1191 (m^1 apc^3 ψ 1191) of 18S rRNA, reduced small subunit (SSU) biogenesis with cellular growth defects, further the lack of pseudouridine at 1191 (m^1 apc^3 ψ 1191 site) was shown to increase sensitivity to Neomycin [106, 107]. In humans, the lack of m^1 apc^3 ψ 1248 in 18S rRNA, introduced by snRA13, is reported to be a significant feature in various cancers [108].

Moreover, post-transcriptional chemical modifications have a unique role in maintaining ribosome fidelity [109]. In brief, the extended flexible helix-38, also known as the "A-site finger" (ASF) of *S. cerevisiae* rRNA, harbours seven pseudouridines, and these are known to give rigidity to the ASF region [109]. Moreover, it was reported that losing six of the seven pseudouridines in the ASF region leads to nonsense suppression of the UGA stop codon [109, 110].

The helix-69 (H69) region of the human 28S rRNA contains five pseudouridines (ψ) residues out of 19 total nucleotides; three of the pseudouridines are highly conserved from bacteria to humans. Pseudouridine at U3727 in human H69 rRNA is protecting the H69 region from metal complexes like $[Rh(DIP)3]^{3+}$, which can induce RNA strand scission independent of structure,

suggesting that these chemical modifications can increase the stability of rRNA [111]. Further, it was shown that the lack of pseudouridine modifications in H69 at positions U2264 and U2266 of *S. cerevisiae* results in an increased sensitivity toward the Neomycin [112]. And blocking the formation of pseudouridine in *E. coli* (K-12) H69 at positions U1911, U1915 and U1917 by inactivating the pseudouridine synthase RluD resulted in increased readthrough of stop codons [113, 114]. These results suggest that the differential role of pseudouridine at the various positions may lead to significant variations in the structure and loop dynamics of H69 and interactions with multiple protein factors in both bacterial and human ribosome [104].

RNA pseudouridylation has been reported to affect translation initiation on specific mRNAs containing internal ribosome entry sites (IRES) by altering the ribosome's affinity for these mRNA structures [115]. In *S. cerevisiae,* the Cbf5 D95A mutant causes a 90% reduction of translation in the IRES-mediated Cricket Paralysis Virus intergenic region compared to WT cells, suggesting that pseudouridylation is essential for cap-independent translation and recruitment of ribosomes to the IRES region [116]. Defective pseudouridylation in the Cbf5 D95A mutant *S. cerevisiae* cells also leads to defects in ribosome fidelity, which further results in multiple molecular impairments like poor reading frame maintenance and stop codon readthrough in both the Cbf5 D95A mutant *S. cerevisiae* and in Dyskeratosis Congenita patient human cells [44, 116]. Notably, pseudouridylation occurs in functionally important regions of ribosomes, and it might be necessary for the ribosome to function at their fullest potential [117]. It has been reported that decreased pseudouridylation due to decreased Dyskerin levels affects the translation of tumour suppressor genes p27 and p53 and increases the translation of VEGF (vascular endothelial growth factor) and HSP70 heat shock proteins [118]. VEGF is one of the vital angiogenesis factors involved in cancer progression. These results indicate that a lack of pseudouridylation affects

ribosomal function (translation fidelity in cells) and can contribute to cancer progression in various tissues [118-120]. Moreover, it has been reported that dysregulation of snoRNA molecules occurs in several neurodegenerative diseases, and mutations in H/ACA snoRNP core proteins were also reported in Dyskeratosis congenita [121]. However, the molecular mechanisms underlying these diseases have yet to be elucidated, and further studies are needed to understand the exact role of RNA modifications in maintaining cellular homeostasis.

1.7. Unsolved mysteries of H/ACA snoRNPs in human health and diseases

Initially, snoRNA function was believed to be confined to housekeeping and chemical modification of rRNA. However, further studies revealed the implications of snoRNAs in pathological conditions **(Figure 1.9)**. The box H/ACA RNAs are associated with tumour progression and genetically inherited diseases like X-linked Dyskeratosis Congenita (X-DC) [122, 123]. In X-DC, patient cells expressing the mutant *DKC1* (DKC1 c.-141 C>G) gene showed dysregulation of specific H/ACA snoRNA molecules, for example certain snoRNAs, like snoRA15,snoRA48 and snoRA24, showed a reduced expression in X-DC and also in the Acute Myeloblastic leukaemia [121]. scaRA U93 is a Cajal body-specific H/ACA snoRNA which harbours CAB box elements and is involved in the pseudouridylation of U2 and U5 snRNA at positions 53 and 54, respectively [20]. In X-DC patients, scaRA U93 is shown to be downregulated [124]. The snoRA42, also called ACA 42, is found to be significantly upregulated in lymphocytes expressing *DKC1* with N-terminal L37 deletion (*DKC1* ΔL37); Moreover, the dysregulation of snoRA48, snoRA22 and snoRA75 are observed in *DKC1* ΔL37 fibroblast cells [124]. A reduction in pseudouridylation in 18S rRNA was identified in fibroblasts with *DKC1* ΔL37 and lymphoblasts with *DKC1* mutation (T66A) [124]. Dysfunctions associated with Dyskerin increase cancer susceptibility. It is suspected that this is mainly due to defective chemical modifications in rRNA,

leading to altered translation and may affect the expression of tumour suppressor genes [123]. Further, mutations associated with Dyskerin lead to decreased activity of telomerase; this, in turn, leads to chromosome instability and cell cycle arrest in some cases, which can cause defective bone marrow function [123].

Further, Tetralogy of Fallot (TOF) is a genetic disorder characterized by abnormal fetal heart development during the first eight weeks of pregnancy. It was shown to be associated with box H/ACA guide RNAs [125]. Three box H/ACA snoRNAs, scaRA1, scaRA4, and scaRA8, are down-regulated in TOF [126].

1.7.1. H/ACA snoRNAs in cancer

It has been shown that dysregulation of box H/ACA snoRNAs is involved in the progression of several cancers or serves as biomarkers for cancer diagnosis [127]. However, the differential expression of H/ACA snoRNA in cancer is tissue-specific not well understood; for example, a snoRNA (h5sn2) was shown to have reduced expression levels in meningioma compared to the healthy brain suggesting the alerted expression of snoRNA in cancer tissues [128]. Overexpression of snoRA42 was observed in non-small cell lung cancer and colorectal cancer. The same study revealed that suppression of snoRA42 has adverse effects on cell proliferation, migration, invasion, and tumorigenicity in non-small cell lung cancer [129, 130]. This study suggests that snoRA42 can be a potential target in cancer therapeutics to cure non-small cell lung cancer since the downregulation of snoRA42 leads to the activation of p53-mediated apoptosis [129].

In human colorectal cancer, snoRA21 is reported to have oncogenic properties. Moreover, increased levels of snoRA21 are observed during metastasis cancer [131]. The upregulation of two other snoRNAs, snoRA47 and ACA11, is observed in hepatocellular

carcinoma tissues compared to healthy tissues. ACA11 is an orphan snoRNA believed to form a novel H/ACA snoRNP complex in multiple myeloma (MM) cells. Overexpression of ACA11 is reported to reduce oxidative stress and promote tumour progression in MM cells [132]. snoRA74B is involved in the pseudouridylation of 28S rRNA and is found to be upregulated in gallbladder cancer [133].

Moreover, the abnormal expression of box H/ACA snoRNAs has also been observed in acute myeloblastic and lymphoblastic leukemia, T-cell lymphoma, and multiple myelomas. However, how this dysregulation of snoRNA is involved in cancer progression is poorly understood, and further studies are required to understand the role of snoRNA molecules in tumour progression [125]. In the past decade, advanced techniques to study quantitative and qualitative RNA modifications have been developed, for example, chemical (CMC) and radiolabeling methods and sensitive mass spectroscopic techniques alongside SCARLET [134-136]. All these techniques open the door to studying the underexplored RNA epitranscriptomics field [137]. Until now, the precise functions of various site-specific rRNA modifications are not fully understood. It is essential to know how these site-specific rRNA modifications, like pseudouridine, are necessary for maintaining cellular homeostasis and how they are related to human diseases like cancer.

1.7.2. X-linked Dyskeratosis Congenita (X-DC)

Dyskerin is an enzyme that associates with H/ACA box snoRNAs and is involved in the catalysis of pseudouridylation at sites guided by these snoRNAs, and Dyskerin is also an essential component of mammalian telomerase holoenzyme complex. It has been shown that mutations in the Dyskerin gene (*DKC1*) lead to a pathological condition called X-linked Dyskeratosis Congenita (X-DC) [138]. X-DC is a congenital disorder characterized by bone marrow failure, skin abnormalities, hematopoietic malignancies, pulmonary fibrosis, reduced cell proliferation,

and increased cancer susceptibility [139]. A mutation in the *DKC1* gene leads to a reduction in telomerase activity. In some patients with dyskeratosis congenita, modifications are also found in telomerase Cajal body protein 1 (TCAB1), which is required for the localization of spliceosomal RNAs (scaRNAs) to Cajal bodies. TCAB1/WDR79 binds to the CAB box of hTR, and mutations in TCAB1 lead to the mislocalization of telomerase and Dyskerin into the nucleolus but not to Cajal bodies [139-141].

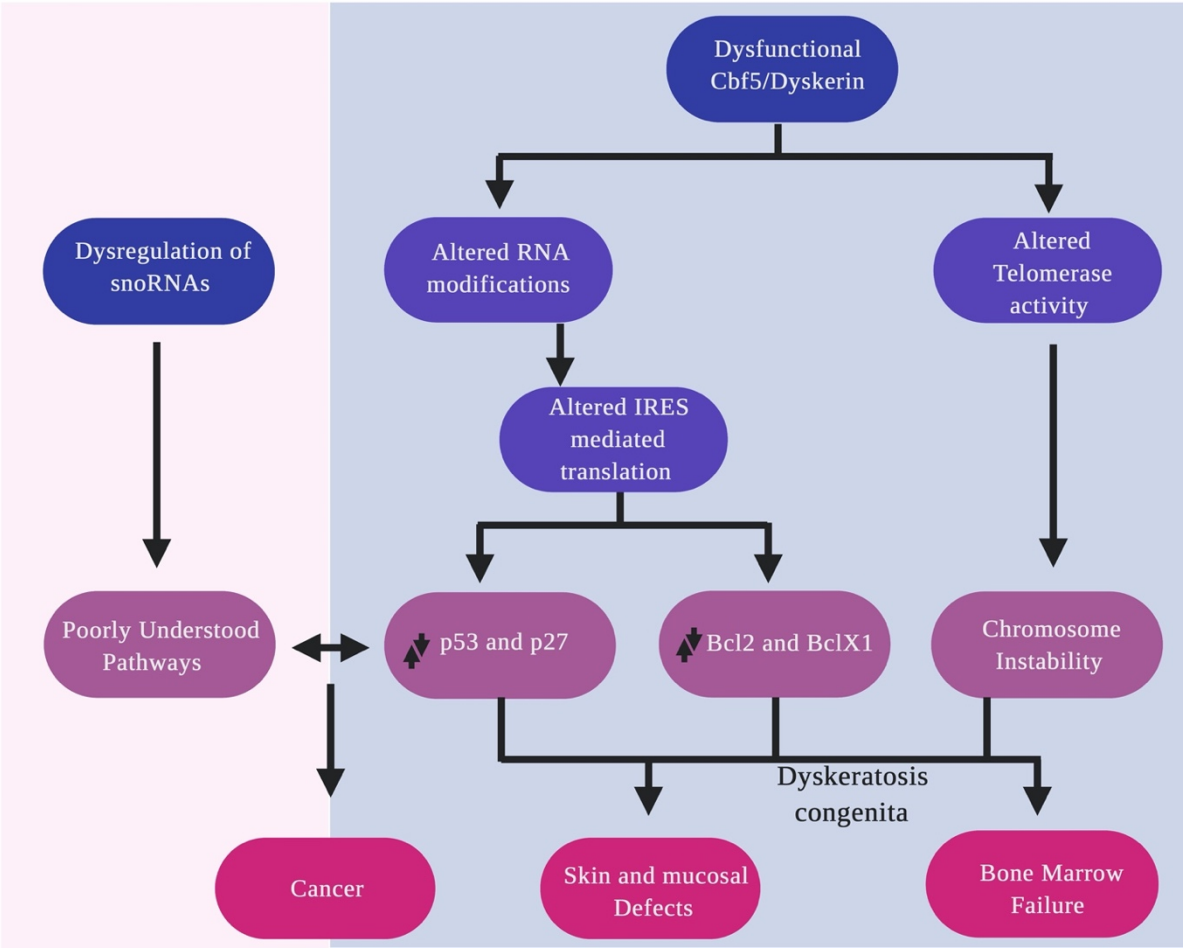

Figure 1.9. Schematic representation of H/ACA snoRNA and Dyskerin defects. This is a brief representation of indirect and direct effects caused by the dysregulation of H/ACA snoRNP complex molecules. We poorly understand the molecular mechanism underlying the dysregulation of snoRNA levels in cancer cells. On the other hand, mutations in Dyskerin lead to bone marrow failure and tumour progression in humans. The up and down arrows represent the dysregulation of p53, p27, Bcl2 and Bcl X1.

Even though the dysregulation of snoRNAs is involved in human diseases, the expression of snoRNA molecules linked to host gene transcription is not yet clear. Even though most of the snoRNA expression depends on host gene transcription, some snoRNAs can be expressed independently of their host gene [142]. In some cases, it was observed that deletion or mutation of host genes might alter the H/ACA snoRNA intronic expression. For example, two H/ACA type snoRNA genes, snoRA6 /ACA6 (U3616) and snoRA62/E2 (U3830), which are transcribed from

the host gene RPSA or Laminin receptor, are known to modify 28S rRNA [143, 144]. Mutations in these two genes can be coupled to Congenital asplenia, a functional disorder of the spleen [145]. In X-DC patients, a reduction in expression levels of snoRA66 is reported. Moreover, the *DKC1* gene was identified as a host gene for snoRA36 and snoRA56. Thus, *DKC1* mutations affect the expression of Dyskerin and also the expression of these two snoRNAs [146, 147]. In many cancers, including breast and lung cancer, the snoRNA molecules serve as biomarkers, and certain snoRNAs are highly upregulated in cancer cells [127].

A set of selective H/ACA snoRNA molecules have also been shown to act as precursors for miRNA and play a crucial role in regulating gene expression [148]. In 2008, the Gunter Meister lab found a miRNA derived from H/ACA snoRNA (ACA 45) and its role in regulating gene expression in a Dicer-dependent manner [148] **(Figure 1.10)**. The ACA 45 is processed into small 20-25 nt long miRNA (ACA45 sRNA) and associates stably with Argonaut (Ago) proteins. Using PicTar (miRNA target prediction tool) it is predicted that the target mRNA for ACA45 sRNA is Cyclin-dependent kinase 11a [138]. On the other hand, the miRNAs named mir-664, mir-1248, and mir-129 are derivatives of ACA36B, ACA34, and HBI-61 snoRNAs, respectively. Interestingly, these H/ACA snoRNA-derived miRNA molecules can still bind to Dyskerin [149].

1.7.3. The H/ACA snoRNA as regulators of cholesterol homeostasis in eukaryotes

The hypoxia-upregulated mitochondrial movement regulator (*Hummr*) is a mitochondrial adaptor protein implicated in mitochondrial motility. Recent studies revealed that the *Hummr* mRNA contains a region complementary to the highly conserved 3′ hairpin of snoRA73 (also U17 in humans, in *S. cerevisiae* snR30). This direct interaction between snoRA73 and *Hummr* mRNA was confirmed by pull-down assays [150]. The overexpression of *Hummr* results in a phenotype identical to snoRA73 knockdown, which leads to a decrease in cholesterol esterification. On the

other hand, *Hummr* downregulation by snoRA73 overexpression increased cholesterol esterification. Later it was shown that in the mouse ovaries, snoRA73 and *Hummr* show inverse expression profiles and are involved in regulating mitochondrial steroid hormone synthesis. These observations are consistent with the model in which *Hummr* is involved in the esterification of plasma membrane-derived cholesterol in ER by diverting cholesterol to mitochondria through enhancing contacts between the ER and mitochondrial, which leads to lipid exchange between ER and mitochondria [150].

1.8. snoRNAs and specialized ribosomes

Previously, it was believed that ribosomal RNA and ribosomal components are ubiquitous and well conserved; it is one of the main reasons our phylogenetic tree is based on 16/18S rRNA [151]. However, further studies proved that there is heterogeneity among rRNA and ribosomal proteins in various organisms [152, 153]; For example, the malarial parasite has two types of rDNA genes, one is S-type, which express in the sporozoite stage of *P. falciparum* in primary host mosquito, and another is A-type which expresses in gametocyte stage of *P. falciparum* in the secondary host humans [154]. *Drosophila melanogaster* harbours an additional 30nt length rRNA molecule called 2S rRNA, apart from 18S, 28S and 5.8S rRNA molecules [155, 156]. The rRNA is highly modified by snoRNA molecules, and tissue-specific snoRNA expression or heterogenicity in the expression of snoRNA has been reported in higher eukaryotes. The existence of specialized ribosomes has been hypothesized [157], and this heterogeneity or special functions for ribosomes could also arise from heterogenicity in rRNA modifications among the ribosomes [105, 158]. It is evident that different types of cancer cells exhibit differential expression of snoRNA molecules; however, how this tissue-specific or tumour-specific expression of snoRNA is involved in the production of specialized ribosomes is understudied. On the other hand,

the translation of tumour suppressor genes (p53 and p27) is IRES-mediated, and these IRES elements directly recruit ribosomes in a cap-independent manner. It is important to note that rRNA modifications, especially pseudouridine formation, can alter the expression of IRES containing mRNA [115]. The ribosomes lacking pseudouridine showed impairment in the IRES-mediated translation [159]. These results suggest that the presence of snoRNA might mediate the specialization of ribosomes in cancer cells **(Figure 1.10)**. However, further studies must be done to uncover the role of rRNA modifications in generating specialized ribosomes in cancer tissues [160, 161]. While multiple factors might contribute to the production of specialized ribosomes, here we have only discussed the potential role of rRNA modifications in specialized ribosome production, as this work focuses solely on the role of H/ACA snoRNA in rRNA modifications.

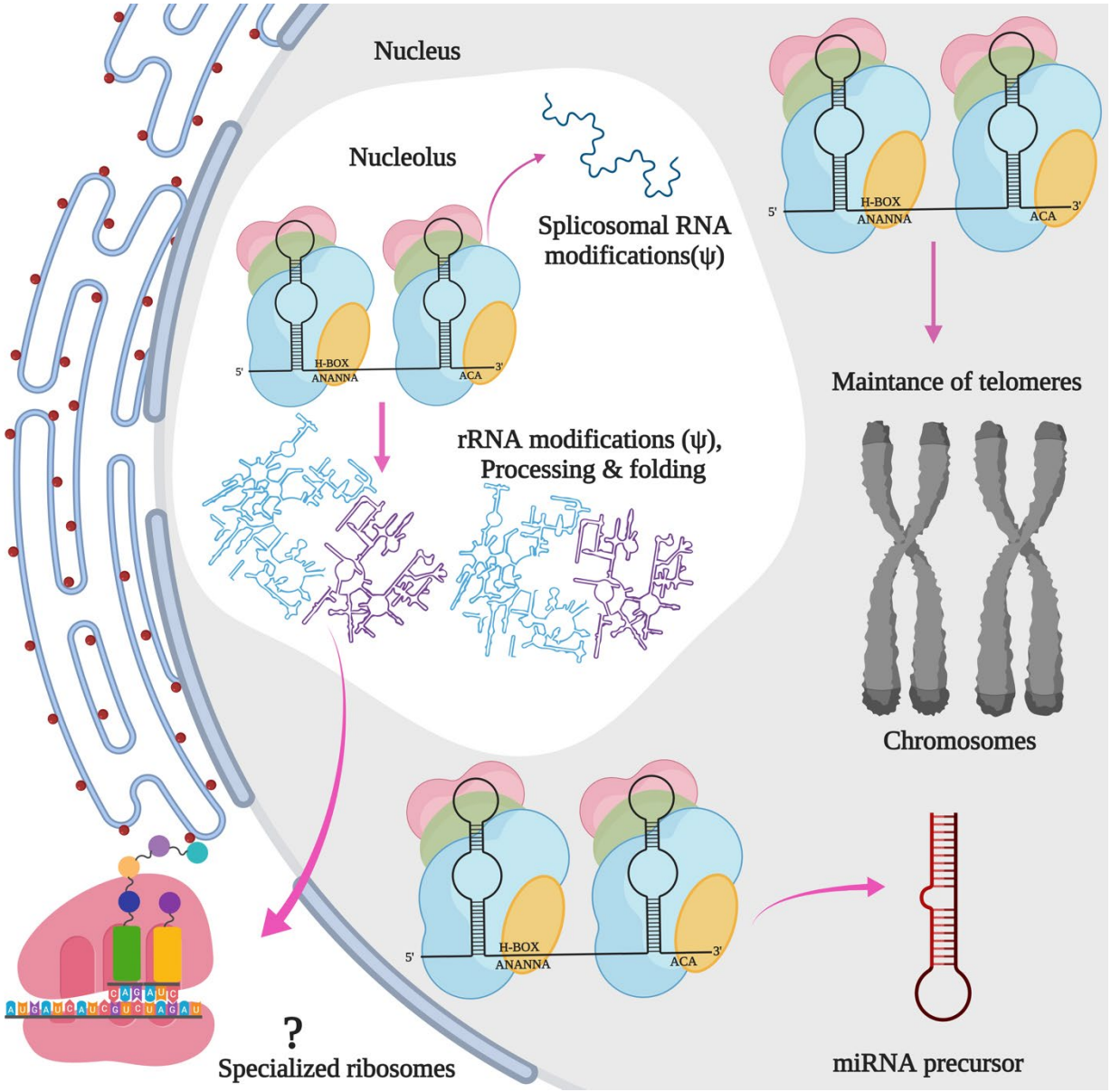

Figure 1.10. Functions of H/ACA snoRNP complexes. The snoRNP complex is mainly involved in rRNA modifications and processing. Moreover, they might be involved in producing heterogenic or specialized ribosomes by modifying rRNA differently. In general, rRNA is an invariable molecule of ribosomes. However, in some tissues, the composition of rRNA nucleotides and the composition of ribosomal proteins/associated factors are variable and involved in the production of heterogenic ribosomes, which can be engaged in the active translation of a specific set of proteins. The post-transcriptional modifications of rRNA and post-translational modifications of ribosomal proteins can give rise to special functions in the ribosome, such as IRES-mediated translation.

1.9. Conclusions and open questions

In both *S. cerevisiae* and humans, the snoRNA molecules function most often as housekeeping genes and act as a precursor for miRNA molecules, involved in maintaining genome stability (humans only), post-transcriptional chemical modifications, and pre-rRNA processing during ribosome biogenesis. The dysregulation/altered expression of snoRNA molecules has been shown in many pathological conditions [124]. However, the heterogeneity in the expression of snoRNA among the various cancer tissues is yet to be understood. Moreover, specific sets of snoRNA molecules are upregulated in multiple cancers, and how this upregulation of snoRNA molecules is involved in the tumour progression is poorly understood. Furthermore, it is not well-comprehended how the expression of snoRNA molecules is related to host gene expression, where the snoRNA is processed from their introns. The silencing of the Dyskerin gene also led to dysregulation of Rab5-Rab11 mediated vesicular trafficking and telomere shortening. However, how and why Dyskerin depletion affects vesicular trafficking independent of telomere instability is yet to be understood [162]. To better understand the biological significance of snoRNA/snoRNP complexes and their association with pathological conditions in humans, the mechanisms underlying pseudouridine formation by eukaryotic H/ACA snoRNAs on highly structured rRNA is yet to be elucidated.

Moreover, it is crucial to understand why eukaryotic H/ACA snoRNAs have two hairpin structure and how these two hairpins interact with each other while modifying the target uridine. Understanding the importance of the bi-parted snoRNP complex provides insights into the biochemical mechanisms. It might help us identify how snoRNA molecules are associated with efficient ribosomal RNA modifications and processing. It is essential to study the high-resolution structure of the eukaryotic H/ACA snoRNP complex for designing small molecule inhibitors or

drugs against certain snoRNAs, which can potentially be used in cancer therapeutics. Further, structure-based mutagenetic analysis, biochemical studies, and high-throughput nucleic acid sequencing techniques might reveal unknown functions of snoRNA molecules in the near future other than rRNA modification and rRNA processing in humans.

CHAPTER 2: OBJECTIVES

A deep understanding of the mechanistic role of H/ACA snoRNAs in pseudouridylation during ribosome biogenesis and how dysregulation of snoRNAs plays a role in promoting tumorigenesis will help us better understand the progression of human diseases like X-DC and cancer. Previous studies from the Kothe lab have shown that an H/ACA snoRNA with a two-hairpin structure is more efficient in pseudouridine formation than a truncated H/ACA snoRNA that harbours only one hairpin [6, 13]. Pseudouridylation is essential for forming functional ribosomes [163]; however, the molecular advantage of the conserved two-hairpin structure in eukaryotic H/ACA RNAs remains to be elucidated. Moreover, it is not well known how the H/ACA snoRNP complex recognizes its target substrate and how it might unfold the RNA while modifying the uridine to pseudouridine. In my Ph.D. thesis, the *in vitro* biochemical characterization of the H/ACA snoRNP complex was performed to provide insights into the importance of conserved two hairpin-structure of H/ACA snoRNAs and the role of two hairpins of the H/ACA snoRNAs in pseudouridine formation during ribosome biogenesis.

In short, my thesis was planned to address the following objectives,

- Identifying the importance of the two-hairpin structure for modifying two uridines in structured rRNA and how the modification of uridine in one pseudouridylation pocket affects the modification of the other uridine.
- Determining the importance of the hinge region length for pseudouridine formation
- Analyzing the importance of unstructured KKE/D and GAR extensions of the Cbf5 and Gar1 proteins for pseudouridine formation.

To analyze the function of the H/ACA snoRNP complex in unfolding structured rRNA fragments and converting the target uridine to pseudouridine without the help of additional factors like helicases, I performed *in vitro* biochemical studies using 25S rRNA fragments H89/H90-92/ H89-92 or H38/H39/ H38-39. These 25S rRNA fragments (H89/H90-92/ H89-92 or H38/H39/ H38-39) are predicted to have local secondary structures. These experiments help us understand how the H/ACA snoRNP complex act on the target site in structured rRNA fragments *in vitro*.

In this thesis the function of the conserved two-hairpin structure of the H/ACA snoRNA was analyzed with respect to pseudouridine formation to dissect whether the two hairpins work independently of each other or whether and how they cooperate with each other while modifying the target uridine. Towards this goal, a long 25S rRNA fragment comprising H89-92 was used, which harbours target sites U2826 and U2880 for snR34. To study the interactions of the two hairpins with rRNA while modifying the uridine, U-C mutations on sites U2826 and U2880 of H89-82 were introduced, respectively. These mutant H89-92 RNAs allow measuring pseudouridine formation on each site while the other site is binding with its cognate H/ACA snoRNA hairpin. Moreover, the function of two hairpins in competition conditions was also analyzed, as the nucleolus is a very dense and crowded organelle in the cell where many RNAs are present that may compete with binding to H/ACA snoRNPs. Thereby, we advance our understanding of the molecular mechanisms underlying the selection of target uridine.

Further, the minimum hinge region length required for efficient pseudouridine formation was determined using snR34 hinge truncations. Towards this goal, three variants of snR34, harbouring 5 nt, 10 nt, and 22 nt deletions at the hinge region were created and named after their

deletions as Δ5H snR34, Δ10H snR34, and Δ22H snR34, respectively. These hinge variants of snR34 reveal the lower limit for the hinge length required for efficient formation of pseudouridine.

Further, my thesis analyzed the importance of KKE/D and GAR domains of Cbf5 and Gar1, respectively, in isomerizing the target uridine to pseudouridine. These KKE/D and GAR domains are not present in the archaeal counterparts but only exist in eukaryotic Cbf5 and Gar1. Moreover, previously it was reported that the region of KKE/D harbours X-DC mutations, and truncation of this region was reported to delay the cell cycle in yeast [49]. However, it was also reported that the truncation of both Cbf5 and Gar1 unstructured KKE/D and GAR domains does not affect the formation of pseudouridines while using H/ACA snoRNPs containing snR5 [26]. In contrast, my analysis using the snR34 H/ACA snoRNP complex reveals that the GAR domains are essential for target uridine isomerization. My further investigation demonstrates that the GAR domains can bind to any RNA non-specifically with nanomolar affinity, and the GAR domains might not play a role in product release. Moreover, this data suggests that GAR domains function in stabilizing the base paring interactions of the pseudouridylation pocket with its substrate RNA.

Together, the work presented in this thesis advances our knowledge of the mechanism and function of the H/ACA snoRNP complex in rRNA modification during ribosome biogenesis. This research thereby contributes to our understanding of the mechanistic details of the two hairpins structure of the H/ACA snoRNP complex during pseudouridine formation in ribosomal RNA. Further, these studies help us to understand the importance of GAR domains in pseudouridine formation.

CHAPTER 3: TWO-HAIRPIN H/ACA SMALL NUCLEOLAR RIBONUCLEOPROTEINS CAN RAPIDLY UNFOLD AND PSEUDOURIDYLATE LARGE STRUCTURED RIBOSOMAL RNA*

3.1. Abstract

During ribosome biogenesis, ribosomal RNA is extensively modified by H/ACA small nucleolar Ribonucleoproteins (snoRNPs). Most eukaryotic H/ACA snoRNAs have two hairpins enabling them to identify two target uridines in ribosomal RNA (rRNA) for isomerization to pseudouridine. Using quantitative biochemical approaches and a defined in vitro reconstitution system for *Saccharomyces cerevisiae* H/ACA snoRNPs, I investigated how H/ACA snoRNPs recognize target sites in long and structured RNAs and why H/ACA snoRNAs possess two hairpins. H/ACA snoRNP can efficiently unwind secondary structure in rRNA, allowing rapid pseudouridylation of structured rRNA. The unfolding of rRNA by H/ACA snoRNPs may provide rRNA with a second chance at correct folding after pseudouridylation. Moreover, pseudouridylation directed by the two hairpins in H/ACA snoRNA occurs independently of each other without allosteric effects between the two active sites. The two hairpins enable H/ACA snoRNAs to correctly recognize two target uridines in the long rRNA, which is likely mediated by tethering the rRNA to the H/ACA snoRNP upon binding to one pseudouridylation pocket. This arrangement minimizes the risk of near-cognate RNA binding to a pseudouridylation pocket and competitive inhibition between two separate substrate RNAs. In conclusion, I uncover critical mechanistic principles allowing the efficient modification of long and structured rRNA.

**A version of this chapter has been submitted as a manuscript to Nucleic Acids Research (NAR). The manuscript has been prepared jointly by Mehar Gayatri D Namala and Ute Kothe.*

3.2. Introduction

H/ACA snoRNPs are important for ribosome biogenesis in both archaea and eukaryotes and telomere maintenance in the eukaryotes [7]. The H/ACA snoRNP complexes mainly function as pseudouridine synthases which chemically modify ribosomal RNA (rRNA), mRNAs, small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs) and other non-coding RNAs, including long-noncoding RNAs [73]. In addition to this function, a few H/ACA snoRNPs (snR30 and snR10 in *Saccharomyces cerevisiae)* contribute to the ribosomal RNA processing [46, 47, 84, 164].

Eukaryotic H/ACA snoRNAs have a highly conserved secondary structure and are organized as a 5'-hairpin-hinge-hairpin-tail-3' [165]. All eukaryotic H/ACA snoRNAs are about 60-300 nucleotides (nt) in length and characterized by two conserved motifs, the H-Box motif (ANANNA) and the ACA Box motif, which are located immediately downstream of the first and second hairpin within the guide RNA, respectively **(Figure 3.1)** [8, 73]. Notably, various dissimilarities are observed between archaeal and eukaryotic H/ACA RNPs [7]. Almost all eukaryotic H/ACA snoRNAs generally have a conserved structure with two hairpins with defined H and ACA Box elements. However, archaea also harbour unusual H/ACA RNAs containing one to three hairpins [7]. Each hairpin harbours an internal pseudouridylation pocket that base-pairs with target RNA and facilitates uridine modification. Previous studies of the *S. cerevisiae* H/ACA RNP complex revealed that both hairpins of the eukaryotic H/ACA snoRNA are required for an efficient pseudouridine formation [26, 55, 166]. Whereas a single hairpin can also perform pseudouridylation of target RNA, it has reduced activity compared to a two-hairpin H/ACA snoRNAs [26, 55]. To date, the functional importance and mechanistic role of the requirement for two hairpins in eukaryotic H/ACA RNAs remain unknown [26, 167].

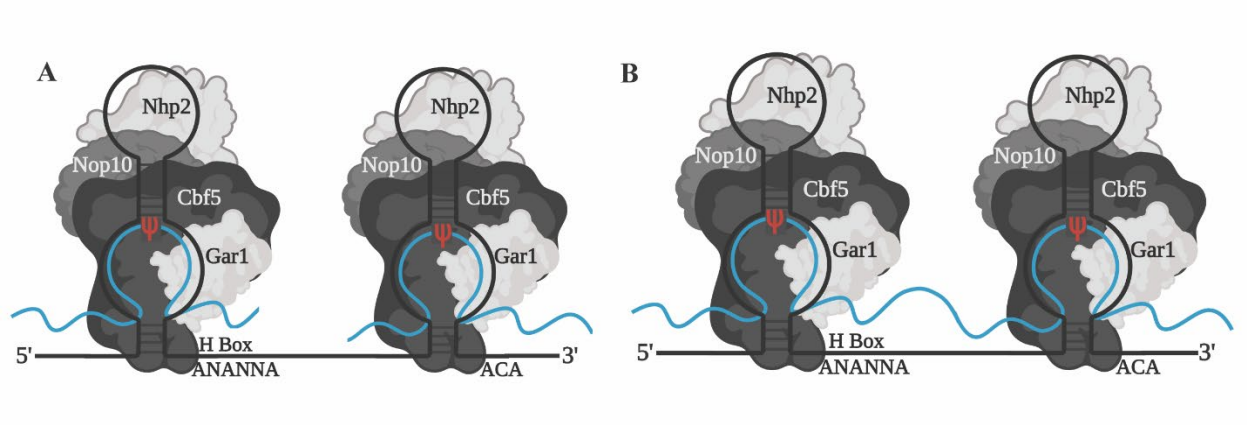

Figure 3.1. Schematic representation of the H/ACA snoRNP complex with bound substrate RNA. A H/ACA snoRNP complex with two individual substrate rRNA molecules bound to each hairpin. **B** Binding of long rRNA substrate, which base pairs to both the 5′ and the 3′ hairpin of the H/ACA snoRNP complex. In this study, individual substrate RNAs used bind separately to the two H/ACA snoRNA hairpins and long substrate RNAs binding to both hairpins.

The mature H/ACA snoRNP complex generally comprises the H/ACA RNA and the four proteins named Cbf5 in yeast/Dyskerin in humans, Nop10, Gar1 and Nhp2. Mutations in the human genes encoding H/ACA proteins Dyskerin, Nop10 and Nhp2 cause the inherited disease Dyskeratosis congenita, characterized by bone-marrow failure, premature ageing, increased tumorigenesis, and nail dystrophy [124, 142, 163]. Dyskeratosis Congenita has also been linked to impaired telomere maintenance [168]. Previous studies revealed that Gar1 is required for stable association of the box H/ACA snoRNAs with the pre-rRNA and is also necessary for substrate turnover during pseudouridine formation [26, 34]. Furthermore, Nop10 is required for the stability of the H/ACA snoRNPs, and knock out of the *NOP10* gene in *S. cerevisiae* leads to a null phenotype, indicating that Nop10 is essential for cell survival [41]. Cbf5 is an RNA-dependent pseudouridine synthase and a single point mutation (D95N) in the catalytic site results in a complete loss of pseudouridine formation leading to severe growth defects in *S. cerevisiae* [44]. Moreover, the depletion of Cbf5 in *S. cerevisiae* causes cell cycle arrest, and a truncated version of Cbf5 lacking the C-terminal domain (KKD/E) causes delayed cell division [49].

High-resolution structural information for various H/ACA snoRNP (sub-)complexes are available, ranging from archaebacteria to humans, like the Cbf5-Gar1-Nop10 ternary complex from *S. cerevisiae* or the cryo-electron microscopy (cryo-EM) structure of the human telomerase holoenzyme, which harbours a H/ACA snoRNP complex [25, 26]. In brief, Cbf5 binds to the lower stem, including the H or ACA box and the pseudouridylation pocket of the H/ACA snoRNA, whereas Nop10 and Nhp2 bind to the upper stem. Gar1 interacts with the catalytic domain of Cbf5 and forms no direct contact with the H/ACA snoRNA. Notably, the two hairpins in higher eukaryotes are associated with two sets of H/ACA proteins forming a bipartite structure [169]. The cryo-EM studies on human telomerase revealed that the conserved H box and ACA box elements are brought in close proximity by interactions between two Dyskerin proteins. Furthermore, nucleotides A377 of the H box and C447 of the ACA box are involved in intramolecular base stacking interactions, forming a cross-hairpin interaction between the 5′ and 3′ hairpins [27].

Given the implication of H/ACA snoRNPs in human diseases, it is essential to perform a detailed biochemical investigation of eukaryotic H/ACA snoRNPs to better understand their mechanism of action in pseudouridine formation during ribosome biogenesis. Previously, the Kothe group reported the reconstitution of *S. cerevisiae* H/ACA snoRNPs from purified components demonstrating that this complex is highly active in pseudouridine formation under physiological conditions [55, 72]. Here I dissect the interactions of the snR34 and snR5 H/ACA snoRNPs with long structured rRNA segments comprising single helices H89, H90-92 or H38, H39, and double helices 89-92 or helices 38-39 of 25S rRNA, respectively **(Figure 3.2 & 3.3)**. By comparing the binding and pseudouridylation of different substrate RNAs, this study shows that H/ACA snoRNPs can unfold structured substrate RNAs and modifies uridine to pseudouridine very efficiently on both target sites independently of each other. In addition, the minimal distance between two

hairpins of H/ACA snoRNA required for efficient pseudouridylation is identified. Interestingly, I observed that two individual substrate RNAs can compete for binding to a pseudouridylation pocket leading to inhibition which is overcome when a long substrate RNA comprising two target sites is modified. Based on these findings, I proposed a model why it is advantageous for eukaryotic H/ACA snoRNPs to use two-hairpin when pseudouridylating long rRNA at multiple sites.

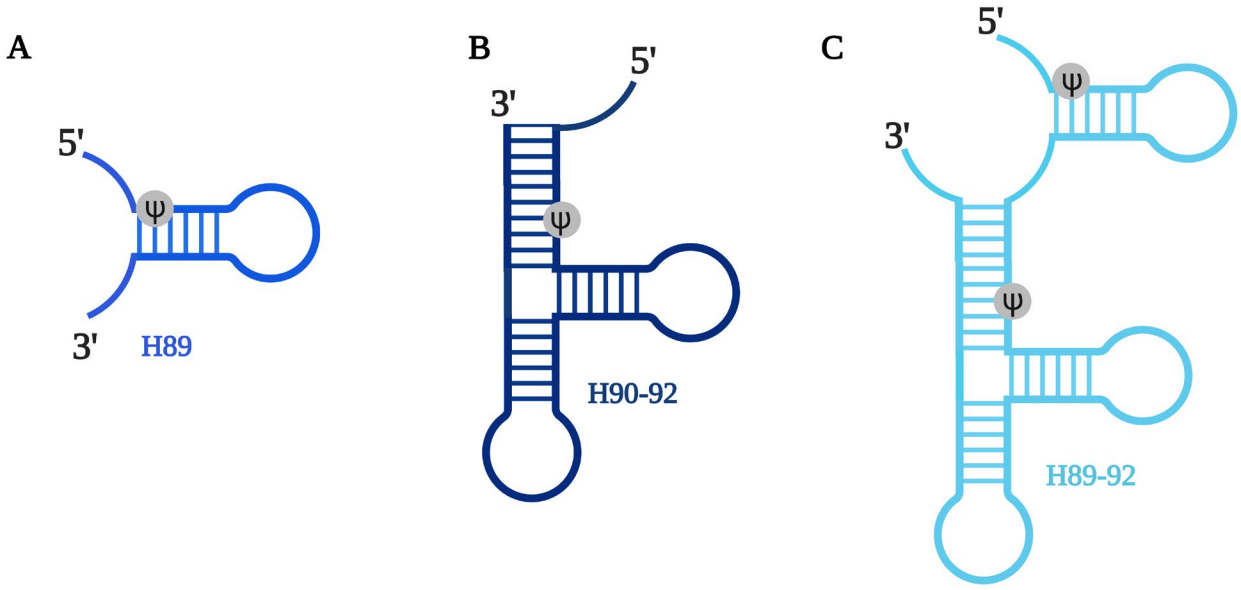

Figure 3.2. Schematic representation of snR34 structured substrates. A H89 RNA a fragment of 25S rRNA that is a substrate for the snR34 5′ hairpin. **B** H90-92 RNA of 25S rRNA, which binds to the 3′ hairpin of snR34. **C** The combined H89-92 rRNA substrate of snR34, which occupies both 5′ and 3′ pseudouridine pockets of the H/ACA snoRNP.

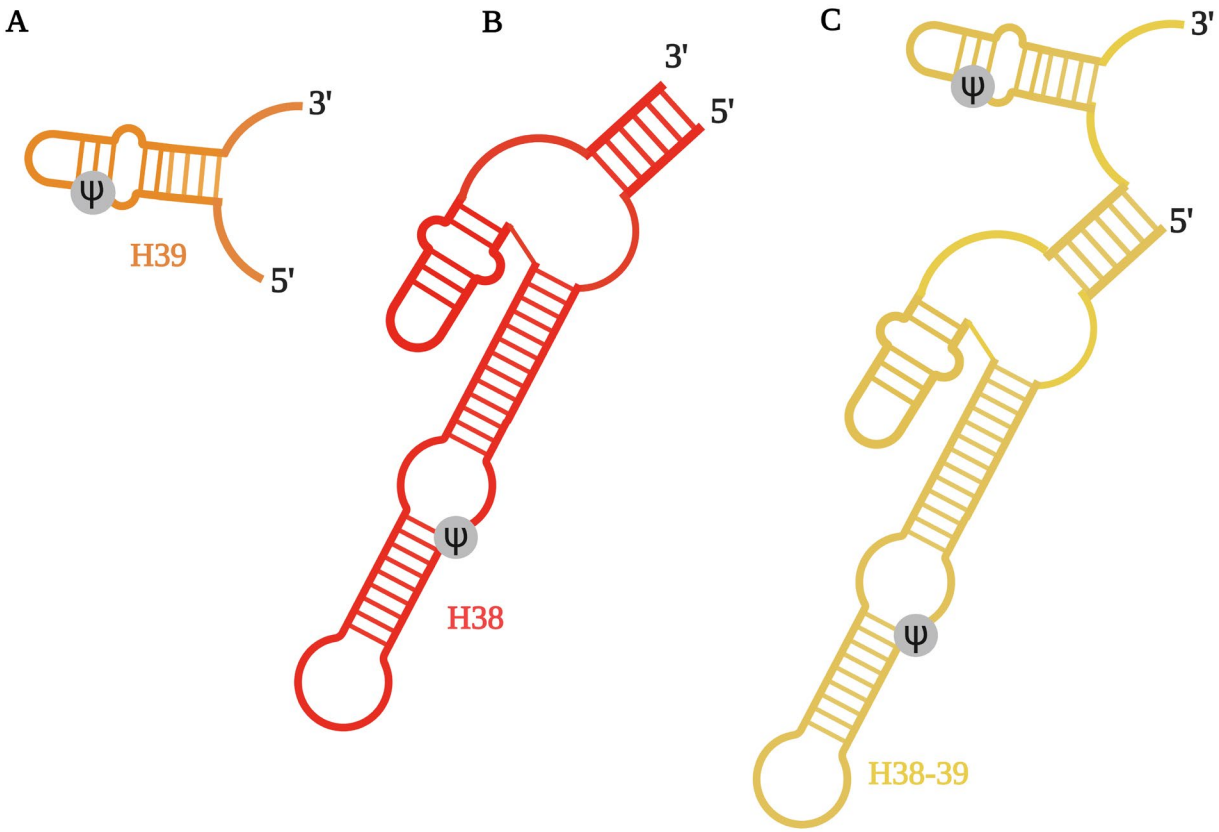

Figure 3.3. Schematic representation of snR5 structured substrates. A H39 rRNA is part of 25S rRNA and base-pair with the snR5 5′ hairpin. **B** H38 RNA, a helix within 25S rRNA that binds to the 3′ hairpin of snR5. **C** The H38-39 rRNA substrate of snR5 occupies both 3′ and 5′ pseudouridine pocket of the H/ACA snoRNP.

3.3. Materials and Methods

3.3.1. Materials

[^{3}H-C5]-UTP for generation of radiolabeled RNA was purchased from Moravek Biochemicals, oligonucleotides were procured from Integrated DNA Technologies (IDT), affinity chromatography materials from GE-Healthcare, and all other enzymes and chemicals were purchased from Bio-basic, Fisher Scientific and NEB.

3.3.2. Molecular cloning

Escherichia coli codon-optimized *CBF5* and *NOP10* genes from *S. cerevisiae* were synthesized (Genewiz), and the *GAR1* and *NHP2* genes of *S. cerevisiae* were cloned into expression vectors as described previously [55]. The snR34 and snR5 guide RNA were cloned into a pUC19 vector as described previously [55]. Site directed Quick change mutagenesis was used to generate the variants of snR34 with shortened hinge regions using gene specific primers listed in **Table 3.1.**

Table 3.1. List of primers used to synthesize 25SrRNA substrates and snR34/snR5 guide RNAs.

Name	Primer sequence
snR34-WT sense	*GCTAATACGACTCACTATAGGGGAATCAAAAA TTTATTTTTTACACGGAAACG*
snR34-WT antisense	*mUmAATGTAGACTTTCTTCGACATCCC*
snR34 Δ5H forward primer	*ATTTGGACAGGATAGGAAGTCCGATTTC*
snR34 Δ5H reverse primer	*ATTCTGGGAAACTCGTATTTCAAATCTTG*
snR34 Δ10H forward primer	*ATTTGGACAGGATAGGAAGTCCGATTTC*
snR34 Δ10H reverse primer	*GGGAAACTCGTATTTCAAATCTTGTTTG*
snR34 Δ22H forward primer	*TTATTTGGACAGGATAGGAAGTCCG*
snR34 Δ22H reverse primer	*TCAAATCTTGTTTGAAATACTGGCAATTAAC*
H89 (5′ sub) sense	*GCTAATACGACTCACTATAGGGATAACTGGCT TGTGGCAGTCAAGCGTTCATAGCGAC*
H89 (5′ sub) antisense	*CATCGAAGAATCAAAAAGCAATGTCGCTATGA ACGCTTGACTGCC*

H90-92 (3′ sub) sense	*GCTAATACGACTCACTATAGGGATGTCGGCT CTTCCTATCATACCGAAGCAGAATTCGGTAAG CG*
H90-92 (3′ sub) Anti sense	*CAGCTCACGTTCCCTATTAGTGGGTGAACAAT CCAACGCTTACCGAATTCTGCTTCGG*
H89-92 (5′-3′ sub) sense	*GCTAATACGACTCACTATAGGGATAACTGGCT TGTGGCAGTCAAGCGTTCATAGCGACATTGC TTTTTGATTCTTCGATGTCGGCTCTTC*
H89-92 (5′-3′ sub) anti sense	*CAGCTCACGTTCCCTATTAGTGGGTGAACAAT CCAACGCTTACCGAATTCTGCTTCGGTATGAT AGGAAGAGCCGACATCGAAGAATC*
H89 U2826C sense	*GCTAATACGACTCACTATAGGGATAACTGGC CTGTGGCAGTCAAGCGTTCATAGCGAC*
H89 U2826C antisense	*CATCGAAGAATCAAAAAGCAATGTCGCTATGA ACGCTTGACTGCC*
H90-92 U2880C sense	*GCTAATACGACTCACTATAGGGATGTCGGCC CTTCCTATCATACCGAAGCAGAATTCGGTAAG CG*
H90-92 U2880C antisense	*CAGCTCACGTTCCCTATTAGTGGGTGAACAAT CCAACGCTTACCGAATTCTGCTTCGG*
H89-92 U2826C sense	*GCTAATACGACTCACTATAGGGATAACTGGC CTGTGGCAGTCAAGCGTTCATAGCGACATTG CTTTTTGATTCTTCGATGTCGGCTCTTC*
H89-92 U2826C antisense	*CAGCTCACGTTCCCTATTAGTGGGTGAACAAT CCAACGCTTACCGAATTCTGCTTCGGTATGAT AGGAAGAGCCGACATCGAAGAATC*
H89-92 U2880C sense	*GCTAATACGACTCACTATAGGGATAACTGGCT TGTGGCAGTCAAGCGTTCATAGCGACATTGC TTTTTGATTCTTCGATGTCGGCCCTTC*
H89-92 U2880C antisense	*CAGCTCACGTTCCCTATTAGTGGGTGAACAAT CCAACGCTTACCGAATTCTGCTTCGGTATGAT AGGAAGGGCCGACATCGAAGAATC*
H89-92 U2826C- U2880C sense	*GCTAATACGACTCACTATAGGGATAACTGGC CTGTGGCAGTCAAGCGTTCATAGCGACATTG CTTTTTGATTCTTCGATGTCGGCCCTTC*
H89-92 U2826C-U2880C antisense	*CAGCTCACGTTCCCTATTAGTGGGTGAACAAT CCAACGCTTACCGAATTCTGCTTCGGTATGAT AGGAAGGGCCGACATCGAAGAATC*
T7 promoter sense	*GCTAATACGACTCACTATAGGG*
snR5 WT antisense	*mAmUATGTACACCTAGAGCGAACC*
snR34 short 3′ substrate sense	*GCTAATACGACTCACTATAGGGACGTCGGCT TTCCTATCATACC*
snR34 short 3′ substrate antisense	*GGTATGATAGGAAGAGCCGACGTCCCTATAG TGAGTCGTATTAGC*

3.3.3. Protein expression and purification

Cbf5 (wildtype), Cbf5 D95N and Nop10 were overexpressed in *E. coli* Lemo 21 DE3 cells (NEB) at 18 °C whereas GST-Gar1 expressing *E. coli* Rosetta cells were mixed with Cbf5- and Nop10-expressing cells for purification of Cbf5•Nop10•GST-Gar1 complex. Nhp2 was independently overexpressed in BL21 DE3 pLysS (NEB) at 37 °C as described in Caton. *et al.*[55] **(Figure A1 A, B)**.

3.3.4. *In vitro* transcription and purification of guide and substrate RNAs

The genes encoding H/ACA snoRNAs (snR5, snR34, Δ5H snR34, Δ10H snR34 and Δ22H snR34) were amplified by polymerase chain reaction (PCR) **(Table 3.2).** RNAs were *in vitro* transcribed as described and purified using size exclusion chromatography using a Superdex 200 column [55, 72]. All RNA concentrations were determined by spectroscopy (A_{260}) using extinction coefficients calculated by OligoAnalyzer 3.1 tool from IDT. Substrate RNAs for snR34 (H89, H90-92 and H89-H92) and snR5 (H38, H39 and H38-39) were generated by *in vitro* transcription in the presence of [C5–^{3}H]-UTP using annealed oligonucleotides as a template **(Table 3.2)**. All the substrate RNAs were purified using the crush and soak method, followed by RNA precipitation using 0.9 volumes of isopropanol with 0.1 volume of 3 M sodium acetate (pH 4.5). Subsequently washed three times with 70% ethanol. The RNA pellet was dissolved in Milli Q water **(Figure A2)**. The RNA concentration and specific activity of each RNA were determined by using A_{260} measurements and scintillation counting.

Table 3.2. List of the substrate and Guide RNA sequences used in this study.

Name	Sequence
H89	*GGGAUAACUGGCUUGUGGCAGUCAAGCGUUCAU AGCGACAUUGCUUUUUGAUUCUUCGAUG*
H90-92	*GAUGUCGGCUCUUCCUAUCAUACCGAAGCAGAA UUCGGUAAGCGUUGGAUUGUUCACCCACUAAUA GGGAACGUGAGCUG*

H89-92	*ACAGGGAUAACUGGCUUGUGGCAGUCAAGCGUU CAUAGCGACAUUGCUUUUUGAUUCUUCGAUGUC GGCUCUUCCUAUCAUACCGAAGCAGAAUUCGGU AAGCGUUGGAUUGUUCACCCACUAAUAGGGAAC GUGAGCUG*
H89 U2826C	*GGGAUAACUGGCCUGUGGCAGUCAAGCGUUCAU AGCGACAUUGCUUUUUGAUUCUUCGAUG*
H90-92 U2880C	*GAUGUCGGCCCUUCCUAUCAUACCGAAGCAGAA UUCGGUAAGCGUUGGAUUGUUCACCCACUAAUA GGGAACGUGAGCUG*
H89-92 U2826C	*ACAGGGAUAACUGGCCUGUGGCAGUCAAGCGUU CAUAGCGACAUUGCUUUUUGAUUCUUCGAUGUC GGCUCUUCCUAUCAUACCGAAGCAGAAUUCGGU AAGCGUUGGAUUGUUCACCCACUAAUAGGGAAC GUGAGCUG*
H89-92 U2880C	*ACAGGGAUAACUGGCUUGUGGCAGUCAAGCGUU CAUAGCGACAUUGCUUUUUGAUUCUUCGAUGUC GGCCCUUCCUAUCAUACCGAAGCAGAAUUCGGU AAGCGUUGGAUUGUUCACCCACUAAUAGGGAAC GUGAGCUG*
H89-92 2U826C & U2880C	*ACAGGGAUAACUGGCCUGUGGCAGUCAAGCGUU CAUAGCGACAUUGCUUUUUGAUUCUUCGAUGUC GGCCCUUCCUAUCAUACCGAAGCAGAAUUCGGU AAGCGUUGGAUUGUUCACCCACUAAUAGGGAAC GUGAGCUG*
H38	*GGGAAGCUCGUAUCAGUUUUAUGAGGUAAAGCG AAUGAUUAGAGGUUCCGGGGUCGAAAUGACCUU GACCUAUUCUCAAACUUUAAAUAUGUAAGA*
H39	*GGGAGCUUUUAGUGGGCCAUUUUUGGUAAGCA GAACUGGCGAUGCGGGAUGAACCGAACGUAG*
H38-39	*AAGCUCGUAUCAGUUUUAUGAGGUAAAGCGAAU GAUUAGAGGUUCCGGGGUCGAAAUGACCUUGAC CUAUUCUCAAACUUUAAAUAUGUAAGAAGUCCU UGUUACUUAAUUGAACGUGGACAUUUGAAUGAA GAGCUUUUAGUGGGCCAUUUUUGGUAAGCAGAA CUGGCGAUG*
snR34 WT	*GAAUCAAAAAUUUAUUUUUUACACGGAAACGAUG CCACAGUUGACUGAACCUGUCUUCUAACAGUAG UUAAUUGCCAGUAUUUCAAACAAGAUUUGAAAUA CGAGUUUCCCAGAAUAAUUUAUUUGGACAGGAU AGGAAGUCCGAUUUCUGUGUUGUCUCAAACGAG GCGAUAGAAUUGGGAUGUCGAAGAAAGUCUACA UUA*
Δ5H snR34	*GAATCAAAAATTTATTTTTTACACGGAAACGATGCC ACAGTTGACTGAACCTGTCTTCTAACAGTAGTTAAT TGCCAGTATTTCAAACAAGATTTGAAATACGAGTTT*

	CCCAGAATATTTGGACAGGATAGGAAGTCCGATTT CTGTGTTGTCTCAAACGAGGCGATAGAATTGGGAT GTCGAAGAAAGTCTACATTA
Δ10H snR34	*GAATCAAAAATTTATTTTTTACACGGAAACGATGCC ACAGTTGACTGAACCTGTCTTCTAACAGTAGTTAAT TGCCAGTATTTCAAACAAGATTTGAAATACGAGTTT CCCATTTGGACAGGATAGGAAGTCCGATTTCTGTG TTGTCTCAAACGAGGCGATAGAATTGGGATGTCGA AGAAAGTCTACATTA*
Δ22H snR34	*GAATCAAAAATTTATTTTTTACACGGAAACGATGCC ACAGTTGACTGAACCTGTCTTCTAACAGTAGTTAAT TGCCAGTATTTCAAACAAGATTTGATTATTTGGACA GGATAGGAAGTCCGATTTCTGTGTTGTCTCAAACGA GGCGATAGAATTGGGATGTCGAAGAAAGTCTACAT TA.*
snR5 WT	*AUCAUUCAAUAAACUGAUCUUCCGGAUUACCAU GCUUAAGACAUCACGCCUCCAUAUGUCUAUAUAA AGCGCAAAUGGCUGGAAGUAGACCAAUUUUUUU UGUUCCUAGCUUUUCAUUAUUGAAAUCUAAUCC AGUUUUAAUGGUUUUUCUUAAUUAAGAAAACAA AUUAUCAUUGGUUCGCUCUAGGUGUACAUAU.*

3.3.5. H/ACA snoRNP reconstitution

Guide RNAs (snR34/snR5) were unfolded at 65°C for 5 min and refolded by slowly cooling from 65°C to room temperature for 30 min in reaction buffer (20 mM HEPES KOH (pH 7.4), 150 mM NaCl, 0.1 mM EDTA, 1.5 mM $MgCl_2$, 10% (v/v) glycerol, 0.75 mM dithiothreitol (DTT)) and incubated with H/ACA proteins in reaction buffer (20 mM HEPES-KOH (pH 7.4), 150 mM NaCl, 0.1 mM EDTA, 1.5 mM $MgCl_2$, 10% (v/v) glycerol, 0.75 mM dithiothreitol (DTT)) in a 0.5:1 ratio of guide RNA: protein for two-hairpin guide RNAs, respectively. Proteins and RNA were incubated for 10 min at 30°C to allow for complex formation.

3.3.6. Tritium release assay

Pseudouridine formation by H/ACA RNP complex was measured by a tritium release assay as previously described [55]. In brief all substrate RNAs are heated at 65°C for 5 minutes and either snap-cooled on ice for 5 minutes or slow-cooled at room temperature for 30 minutes in

reaction buffer (20 mM HEPES KOH (pH 7.4), 150 mM NaCl, 0.1 mM EDTA, 1.5 mM $MgCl_2$, 10% (v/v) glycerol, 0.75 mM dithiothreitol (DTT)). Subsequently, multiple turnover assays were performed in triplicates or duplicates at 30 °C using 50 nM H/ACA RNP and 500 nM substrate.

3.3.7. Analysis of H/ACA snoRNP complex proteins

The *in vitro* reconstituted H/ACA snoRNP complexes with snR34 WT and Δ22H snR34 (50 nM) were subjected to ultrafiltration using a 100 kDa molecular weight cutoff membrane to remove unbound proteins by stepwise dilution with buffer A containing 20 mM HEPES KOH (pH 7.4), 150 mM NaCl, 0.1 mM EDTA, 1.5 mM $MgCl_2$, 10% (v/v) glycerol, 0.75 mM dithiothreitol (DTT)). To remove unbound proteins from the *in vitro* reconstituted 50 nM H/ACA snoRNP complex, it was diluted 5-fold with buffer A to final conc. 10 nM and concentrated to 50 µl corresponding to about 100 nM. This step was repeated twice using 100 kDa Amicon ultrafiltration devices to get final volume of 30 µl with predicted conc. around 45 nM. If the complexes dissociate during this process, the dissociated proteins will be removed during Amicon filtration and can be detected in the flow-through. As negative control, a similar reaction was carried out with only Cbf5•Nop10•GST-Gar1•Nhp2 proteins without guide RNA. In all cases, the flow-through from the ultrafiltration was subjected to trichloroacetic acid (TCA) precipitation by adding 4 volumes of 100% TCA and incubating overnight at 4°C, followed by three washes with ice-cold acetone. The resultant total pellet was dissolved in 30 µl of 50 mM Tris pH 8, and all 30 µl sample was loaded on 12% tris-tricine gels followed by silver staining.

3.3.8. Nitrocellulose filtration assay

Binding assays used tritium-labelled substrate RNAs heated at 65°C for 5 minutes and slow-cooled at room temperature for 30 minutes. Binding affinities were measured by incubating increasing concentrations of the substrate RNA for 3 min at 30°C in the presence of 5 nM H/ACA

snoRNP reconstituted with Cbf5 wild type. The complete reaction was filtered through a nitrocellulose membrane, followed by washing the nitrocellulose membrane with 1 mL cold reaction buffer. The nitrocellulose membrane was dissolved in a 10 mL EcoLite scintillation cocktail [EcoLite (+), MP Biomedical], followed by scintillation counting to determine the amount of substrate RNA bound to the H/ ACA snoRNP. Dissociation constants (K_D) were determined by fitting the binding curves to the following hyperbolic function using GraphPad Prism.

$Y = B_{max} \times [S] / (K_D + [S])$ (Equation 1)

with [S] being the substrate concentration and B_{max} the maximum binding. The substrate RNA: enzyme ratio was calculated by dividing the picomoles of substrate RNA retained on the nitrocellulose membrane by the picomoles of H/ACA RNP in the reaction.

3.3.9. Nitrocellulose filtration competition assay

Competition binding assays were performed using both tritium-labelled and non-radioactive substrate RNAs. All RNAs are heated at 65°C for 5 min and slow cooled at room temperature for 30 min in reaction buffer (20 mM HEPES KOH (pH 7.4), 150 mM NaCl, 0.1 mM EDTA, 1.5 mM $MgCl_2$, 10% (v/v) glycerol, 0.75 mM dithiothreitol (DTT)). Binding affinities are measured by incubating increasing concentrations of the non-radioactive substrate RNA and fixed concentration of radioactive RNA (20 nM or 30 nM as indicated) in the presence of 5 nM H/ACA snoRNP for 3 min at 30°C. Dissociation constants (K_D) for the non-radioactive substrate RNA were determined by fitting the binding curves to the hyperbolic function in GraphPad Prism one-site non-specific binding function and set the non-specific binding to zero non-specific bindings.

$Y = B_0 - Amp \times [S] / (K_D + [S])$ (Equation 2)

where B_0 is the initial binding, Amp is the amplitude of the decrease in binding, and [S] is the concentration of the non-radioactive RNA being titrated.

3.4. Results

3.4.1. H/ACA snoRNPs can form pseudouridines rapidly in structured 25S rRNA fragments

Previously, the Kothe group demonstrated that the H/ACA snoRNP complex is capable of pseudouridine formation in short (23 nt), unstructured rRNA fragments representing segments of the 25S rRNA helices H89 or H90-92 [55, 72]. To elucidate the biochemical mechanism of the *S. cerevisiae* H/ACA RNPs acting on longer structured rRNA fragments, I reconstituted the H/ACA snoRNP complex from purified proteins with *in vitro* transcribed snR34 or snR5 guide RNA. Moreover, structured fragments of 25S rRNA H89, H90-92 and H89-92 or H38, H39 and H38-39 corresponding to both target sites of snR34 and snR5 were transcribed using [C5^{-3}H] UTP to generate radioactively labelled substrate RNA allowing us to measure pseudouridine formation using a tritium release assay. *S. cerevisiae* H/ACA RNPs were formed by incubating all four H/ACA core proteins with snR34 or snR5 guide RNA, and time courses of pseudouridine formation were recorded for all three structured substrate RNAs of snR34. The H89 and H39 RNA are recognized by the 5′ hairpin and the H90-92 and H38 RNA by the 3′ hairpin of snR34 and snR5, respectively. The H89-92 and H38-39 RNA will be recognized by both 5′ and 3′ hairpins of snR34 and snR5, respectively. Using these rRNA fragments containing secondary structure, time courses of pseudouridine formation were recorded. Here, I observed an initial velocity of 32.0 ± 7.0 nM min^{-1} for modification of the structured H89 by the 5′ hairpin of snR34 H/ACA snoRNP. Similarly, the activity of the 3′ hairpin of snR34 in modifying H90-92 was recorded as 42.0 ± 2.0 nM min^{-1} **(Figure 3.4A-B, Table 3.3)**. Moreover, the initial velocity of 65.0 ± 3.0 nM min^{-1} was observed for the H89-92 structured substrate RNA, which occupies both 5′ and 3′ hairpins of snR34 **(Figure 3.4C, Table 3.3).** Next, I performed similar experiments with snR5 and its structured substrate RNAs H38, H39 and H38-39 to confirm that the formation of pseudouridine

on structured RNA molecules is similar for all H/ACA snoRNPs. I observed an initial velocity of 71.0 ± 9.0 nM min^{-1}, 59.0 ± 9.0 nM min^{-1}, and 85.0 ± 15.0 nM min^{-1}, respectively **(Figure 3.4D-F, Table 3.3)**. These data demonstrate that the H/ACA snoRNP complex is able to pseudouridylate long folded 25S rRNA [55, 72]. Moreover, this data suggests that the H/ACA snoRNP complex is capable of efficiently unfolding local secondary interactions while chemically modifying the target uridine to pseudouridine on structured substrate RNAs.

Furthermore, I conducted similar tritium release assays to monitor if the H/ACA snoRNP complex is involved in non-specific pseudouridine formation in regions of 25S rRNA other than H89 U2826 or H90-92 U2880. To answer this question, I used mutated substrate RNA, where all target uridines are converted into cytosine. The H89 U2826C RNA, the H90-92 U2880C RNA, and the H89-92 U2826C and U2880C RNA are not pseudouridylated as evident by comparison to the negative control without H/ACA snoRNP **(Figure 3.4A-C).** In conclusion, this data proves that the snR34 harbouring H/ACA snoRNP complex does not have any off-targets in this region of 25S rRNA and that it only modifies the U2826 and U2880 sites on H89-92 25S rRNA.

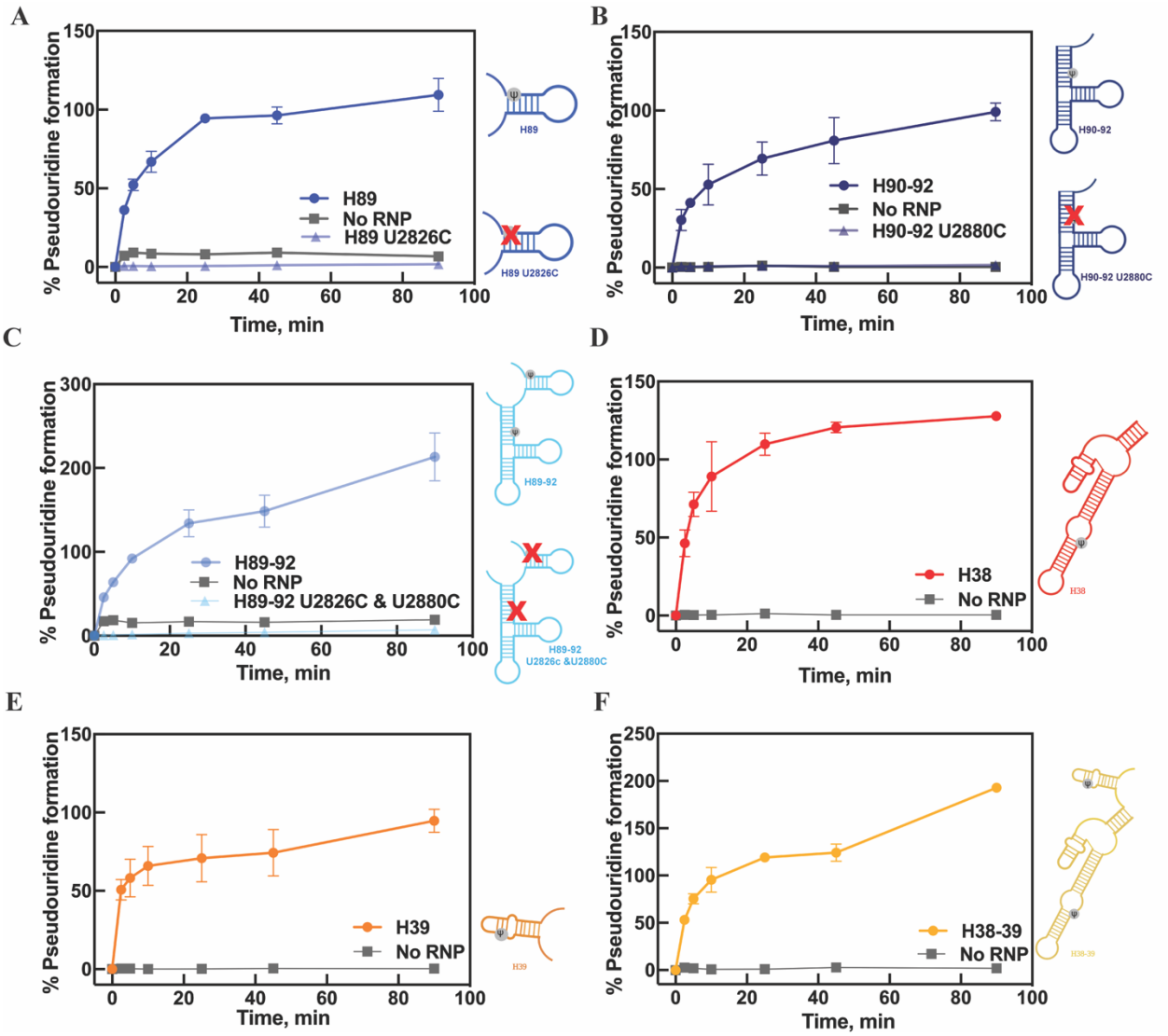

Figure 3.4. Pseudouridine formation by the snR34 and snR5 H/ACA RNP in structured substrate RNAs. A Pseudouridine formation was detected by tritium release using H89 as substrate for the 5′ hairpin of snR34, which has the target U2826. As negative controls, H89 with the target U substitute by C was used (H89 U2826C), and the H89 substrate RNA was incubated in the absence of H/ACA RNP. **B** Pseudouridine formation in H90-92, a substrate for 3′ hairpin of snR34, which has the target U2880. **C** Pseudouridine formation in a long substrate RNA for snR34 combining H89 and H90-92. This 25S rRNA fragment has two target uridines and can base-pair with both hairpins of snR34. As negative control, both target Us were substituted with C. **D** Pseudouridylation by an H/ACA RNP reconstituted with snR5 and using H38 as a substrate for the 3′ hairpin of snR5, which has the target U1004. **E** Pseudouridine formation in H39 as a substrate for 5′ hairpin of snR5, which has the target U1124. **F** Pseudouridylation time courses for long substrate RNAs (H38-39) with two target uridines and the ability to base-pair to both hairpins of snR5.

Table 3.3 Binding and modification of structured substrate RNAs by H/ACA snoRNPs. The corresponding modification time courses are shown in **Figure 3.4 and Figure 3.5**, and the binding curves for determining dissociation constants are displayed in **Figure 3.6 (snR34) and Figure 3.7 (snR5).**

Guide RNA	Substrate RNA	v_0 (nM min^{-1})	K_D (nM)
snR34 WT	H89	32.0 ± 7.0	28.6 ± 2.5
	H89 U2826C	0.08 ± 0.03	11.4 ± 4.3
	H90-92	42.0 ± 2.0	18.0 ± 3.9
	H90-92 U2880C	0.07 ± 0.02	15.0 ± 3.5
	H89-92	65.0 ± 3.0	9.0 ± 2.6
	H89-92 U2826C U2880C	0.36 ± 0.02	7.6 ± 2.7
	H89-92 U2826C	30.0 ± 4.0	7.0 ± 2.0
	H89-92 U2880C	36.0 ± 2.0	12.5 ± 3.0
snR5 WT	H38	71.0 ± 9.0	14.0 ± 3.0
	H39	59.0 ± 9.0	18.5 ± 4.0
	H38-39	85.0 ± 15.0	7.0 ± 2.0
	H89	-	36.1 ± 13.0

3.4.2. The two hairpins of the snR34 H/ACA snoRNP complex work independently

Next, I focused on investigating how two hairpins of H/ACA snoRNA function while modifying its target U on structured rRNA fragments when both hairpins of H/ACA snoRNA are bound to its target substrate. Therefore, substrates with uridine to cytosine mutations at positions 2826 and 2880 of the H89-92 long-structured substrate RNA were designed, as the H89-92 predicted to have the ability to occupy both hairpins of snR34 at a given time. In the H89-92 U2826C and H89-92 U2880C RNAs, only one target U was changed to C, and in H89-92 U2826C-U2880C RNA, both target uridines at 2826 and 2880 were mutated to C. I performed tritium release assays to monitor the formation of pseudouridine at one target uridine while the other H/ACA snoRNA hairpin was bound by the substrate RNA region containing the U-to-C mutations. Further I performed the tritium release assay with both snap-cooled and slow-cooled H89-92 U2880C, and recorded the initial velocity of formation of pseudouridine on the U2826 position as 36.0 ± 2.0 nM min^{-1}, which is the same as the initial velocity of 32.0 ± 7.0 nM min^{-1} for H89 RNA alone harbouring U2826 **(Figure 3.5A, Figure A4, Table 3.3)**. Further, I performed a similar tritium release assay with H89-92 U2826C mutant substrate RNA where I monitored the formation of pseudouridine at U2880, which was modified by 3′ pseudouridylation pocket of snR34. My data showed a slightly reduced initial velocity 30.0 ± 4.0 nM min^{-1} for modification of U2880 in H89-92 U2826C RNA when compared with the initial velocity of 42.0 ± 2.0 nM min^{-1} for pseudouridylation of U2880 in H90-92 RNA **(Figure 3.5B, Table 3.3)**. Thus, this data indicates that both the 5′ and the 3′ pseudouridylation pockets in the two hairpins work equally fast while modifying the target U on both target sites with the initial velocity of 36.0 ± 2.0 nM min^{-1} modifying U2826 and 30.0 ± 4.0 nM min^{-1} modifying U2880 in long H89-92 25S rRNA fragment.

Moreover, these findings show that the activity in one pseudouridylation pocket is not influenced by binding of RNA to the other pseudouridylation pocket of a two-hairpin H/ACA snoRNA.

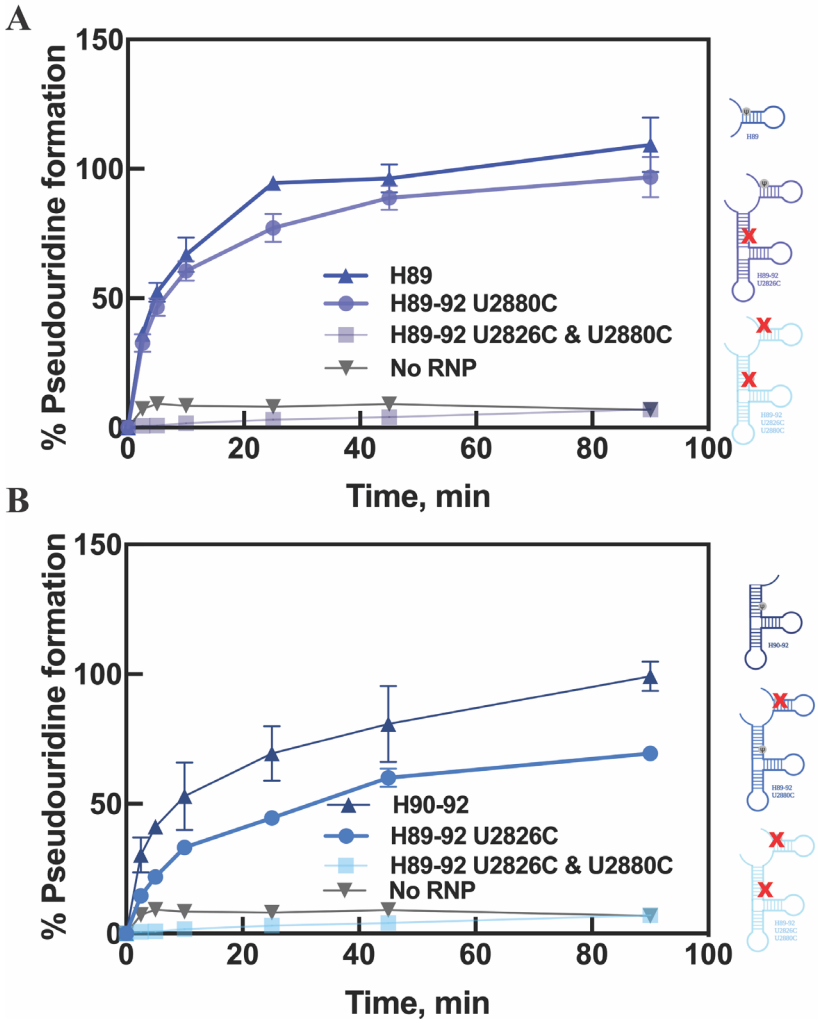

Figure 3.5. Pseudouridine formation directed by one hairpin is independent of substrate RNA binding to the other hairpin of H/ACA snoRNA. A Modification of U2826 by the 5′ hairpin of snR34: pseudouridylation time courses were conducted with the H89-92 U2880C substrate RNA binding to both hairpins in comparison to H89 substrate RNA containing U2826 and binding only to the 5′ hairpin of snR34. **B** Pseudouridine formation at position U2880 by the 3′ hairpin of snR34: modification was compared in H90-92 binding only to the 3′ hairpin and H89-92 U2880C substrate binding to both hairpins of snR34. H89-92 U2826C & U2880C substrate RNA and no RNP samples were used as negative controls for off-target pseudouridine measurements.

3.4.3. H/ACA snoRNPs bind to structured rRNA substrate with high affinity

Further, the binding of substrate RNAs to H/ACA RNPs was evaluated by analyzing the affinity of the H/ACA snoRNP complex with radioactively labelled structured substrate RNA. In brief, a nitrocellulose filter binding assay was performed to determine the K_D using a 5 nM constant concentration of reconstituted H/ACA RNP and titrating with high concentrations of substrate RNA up to 500 nM. The substrate RNA titrations without H/ACA snoRNP serve as a negative control, which confirmed a minimal background signal from substrate RNA alone binding to the nitrocellulose membrane. 5 nM active H/ACA snoRNPs were incubated with excess substrate RNA for 3 min. In this short time, only partial pseudouridine formation can occur, and the substrate RNA should contain mostly uridines rather than pseudouridine. My data show that the single helices harbouring H89 or H90-92 and double helices harbouring H89-92 25S rRNA substrates bind to H/ACA snoRNP with a high affinity of 28.6 ± 2.5 nM, 18.0 ± 3.9 nM, and 9.0 ± 2.6 nM respectively **(Figure 3.6A-B, Table 3.3)**. These results suggest that the binding affinity of the H/ACA snoRNP is increased for the long H89-92 rRNA fragment compared to its two halves, H89 and H90-92. To validate this finding with a different H/ACA snoRNA and its substrates RNAs, a similar filter binding assay was performed to analyze the binding affinity of structured H38, H39 and H38-39 rRNA fragments with snR5 harbouring snoRNP. My data showed that the H38, H39 and H38-39 RNAs bind to snR5 H/ACA snoRNP complex with affinities of 14.0 ± 3.0 nM, 18.5 ± 4.0 nM and 7.0 ± 2.0 nM, respectively which further confirms that all H/ACA snoRNP complexes to bind tightly to long structured rRNA fragments. **(Figure 3.7A-C, Table 3.3)**.

Next, I conducted similar filter binding experiments to determine the affinity of H/ACA snoRNPs for the U-C mutant H89 (U2826C), H90-92 (U2880C) and H89-92 (U2826C, U2880C) substrate RNA molecules that cannot be pseudouridylated. The only difference between the wild-

type and mutant substrate RNA was a single nucleotide change. I observed that the H89 U2826C, H90-92 U2880C, H89-92 U2826C, H89-92 U2880C and H89-92 U2826C-U2880C RNAs bind with 11.4 ± 4.3 nM, 15.0 ± 3.5 nM, 7.0 ± 2.0 nM, 12.5 ± 3.0 nM and 7.6 ± 2.7 nM affinity, respectively **(Figure 3.6C-H, Table 3.3)**. Hence, the substrate RNAs carrying a C at the modification position bind slightly tighter to the wild-type snR34 snoRNP than the RNAs containing target uridine that may be (partially) modified during the pseudouridylation assay.

To further confirm how the catalytic inactive H/ACA snoRNP complex binds uridine-containing rRNA, I performed a filter binding analysis with an snR34 H/ACA snoRNP complex harbouring catalytic inactive Cbf5 D95N. I observed that the structured substrate RNAs H89, H90-92 and H89-92 bind with affinities of 19.0 ± 4.2 nM, 23.3 ± 4.0 nM and 7.0 ± 1.9 nM, respectively **(Figure 3.8, Table 3.4)**. Therefore, these substrate RNAs are binding with equal or higher affinity to the inactive H/ACA snoRNP than to wild-type H/ACA snoRNP, again suggesting that unmodified substrate RNAs are bound tighter than (partially) modified pseudouridine containing RNAs.

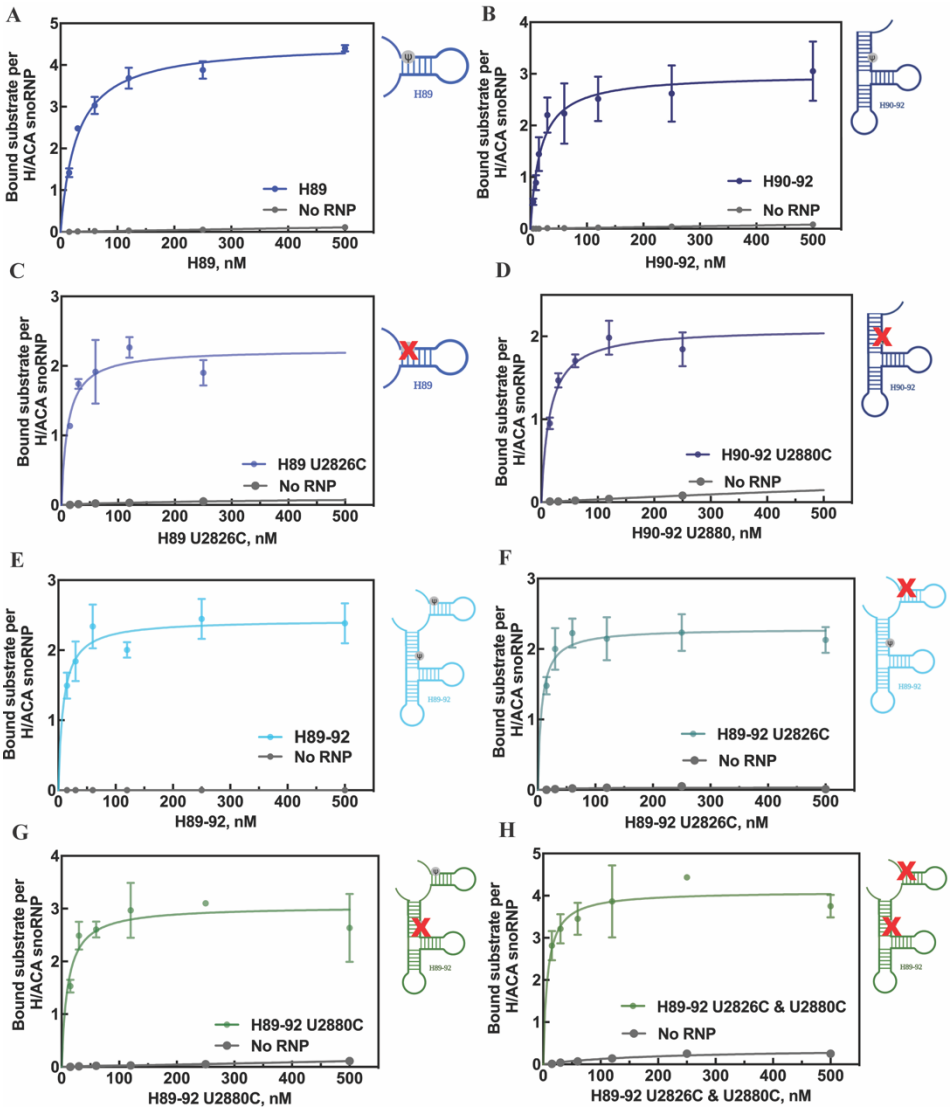

Figure 3.6. Binding of structured substrate RNA to the snR34 H/ACA RNP. To monitor binding, increasing concentrations of different radiolabeled substrate RNAs were incubated with a constant concentration of snR34 H/ACA RNP, and the amount of substrate RNA binding was determined by nitrocellulose filtration. **A** H89 substrate RNA binding to the 5′ hairpin snR34. **B** H90-92 substrate RNA binding to the 3′ hairpin snR34. **C** H89 U2826C mutant RNA binding to the 5′ hairpin snR34. **D** H90-92 U2880C mutant RNA binding to the 3′ hairpin of snR34. **E** H89-92 substrate RNA binding to both the 5′ and 3′ hairpin of snR34. **F** H89-92 U2826C mutant RNA binding to both hairpins of snR34 **G** H89-92 U2880C mutant RNA binding to both hairpins of snR34. **H** H89-92 U2826C & U2880C mutant RNA binding to both hairpins of snR34 **(Table 3.3)**. The dissociation constants (K_D) were determined by fitting the data with a hyperbolic equation (see Materials and Methods).

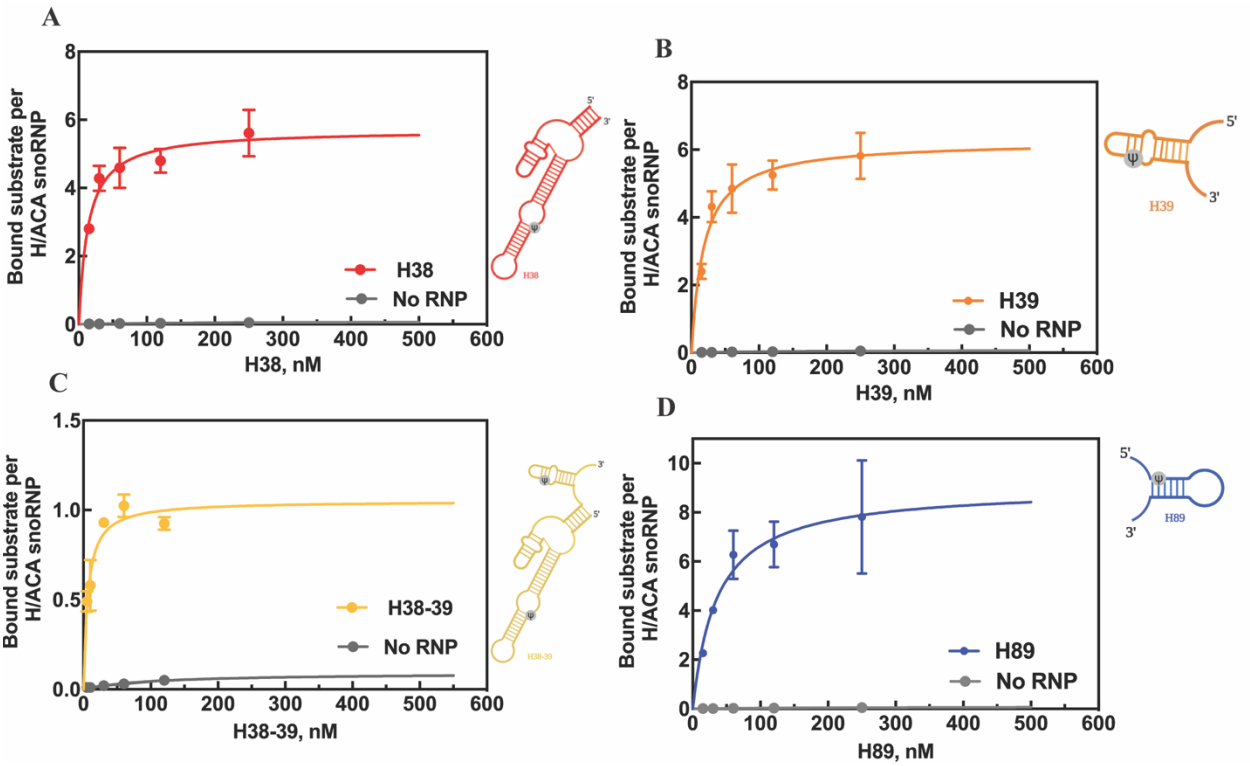

Figure 3.7. Binding of structured substrate RNA to the snR5 H/ACA RNP. The interaction of different substrate RNAs with the snR5 H/ACA RNP was detected by nitrocellulose filtration. **A** H38 substrate RNA binding to the 3′ hairpin of snR5. **B** H39 substrate RNA base-pairing with the 5′ hairpin of snR5. **C** H38-39 substrate RNA binding to both hairpins of snR5. **D** Binding of the snR5 H/ACA RNP to H89 RNA which is a substrate for snR34 **(Table 3.3)**. The dissociation constant (K_D) was determined by fitting with a hyperbolic equation (see Materials and Methods).

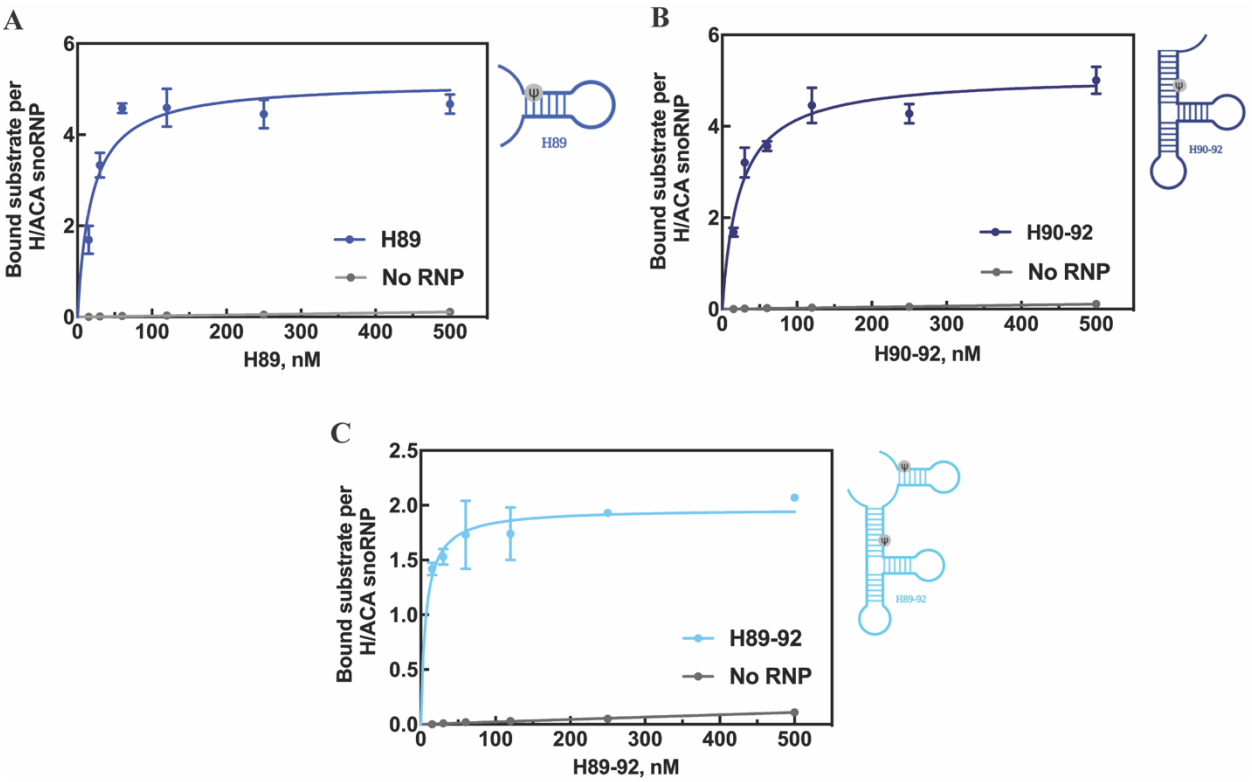

Figure 3.8. Binding of structured substrate RNA to snR34 H/ACA RNP harbouring catalytically inactive Cbf5 D95N. Nitrocellulose filtration assays were used to monitor the binding of increasing concentrations of substrate RNA to 50 nM of H/ACA snoRNP reconstituted with snR34 H/ACA RNA. **A** Binding of H89 RNA of to the 5′ hairpin of snR34. **B** Binding of H90-92 RNA to the 3′ hairpin snR34. **C** Binding of H89-92 RNA to both hairpins of snR34. Hyperbolic fitting of the data yielded the dissociation constants, which are summarized in **Table 3.4.**

Table 3.4. Substrate RNA binding by H/ACA snoRNPs containing a Cbf5 D95N variant. The corresponding modification time courses and the binding curves are shown in **Figure 3.8**.

Guide RNA	Substrate RNA	K_D (nM)
snR34 WT	H89	19.0 ± 4.2
	H90-92	23.3 ± 4.0
	H89-92	7.0 ± 1.9

3.4.4. Shortening the hinge region of H/ACA snoRNA impairs pseudouridylation

Two previous studies found that even though a single hairpin of H/ACA snoRNA is capable of pseudouridine formation, its efficiency is reduced compared to full length H/ACA snoRNA [26, 55]. Therefore, I further analysed the importance of the hinge region of snR34 for the co-ordination of the two hairpins. I systematically deleted a stretch of 5nt, 10nt and 22nt sequences from snR34 hinge region called Δ5H, Δ10H and Δ22H snR34, respectively. While designing the deletions of these sequences from the hinge region, I used m-fold RNA structure prediction tool to verify that the pseudouridylation pockets of snR34 guide RNA are not disturbed **(Figure 3.9)**. To analyze the activity of the H/ACA snoRNP complex harbouring the snR34 hinge variants Δ5H, Δ10H and Δ22H in pseudouridine formation, tritium releases assays were further carried out. Initially, the activity of the 5′ hairpin of snR34 variants with H89 of 25S rRNA was monitored in tritium release assays. Interestingly, it was observed that the deletion of 5nt in the hinge region of snR34 (Δ5H) has no effect, and the formation of pseudouridine in H89 is proceeding with an initial velocity of 31.0 ± 7.0 nM min^{-1} similar to snR34 type guide RNA **(Figure 3.10A, Table 3.3)**. The initial velocity is the same for snap-cooled and slow-cooled substrate RNA H89 (**Figure A4**). Furthermore, the formation of pseudouridine by the 3′ hairpin on H90-92 was monitored with an initial velocity of 43.0 ± 6.0 nM $min^{-1,}$ which is equal to the initial velocity of snR34 wild-type on H90-92 **(Figure 3.10B, Table 3.3)**. However, the deletion of 10 and 22nt in the snR34 hinge region significantly affects the activity of the 5′ pseudouridine pocket of snR34 when compared with wild type snR34. The initial velocities of pseudouridine formation in H89 by the 5′ pseudouridine pocket of Δ10H and Δ22H are reduced almost three-fold compared to wild-type to 13.0 ± 1.0 nM min^{-1} and 11.0 ± 2.0 nM min^{-1}, respectively **(Table 3.5)**. Moreover, I observe slightly reduced pseudouridine activity in H90-92 by the 3′

pseudouridine pocket of Δ10H and Δ22H snR34 variants with initial velocities of 23.0 ± 2.0 nM min^{-1} and 25.0 ± 1.0 nM min^{-1}, respectively **(Table 3.5)**. Additional analysis of the potential hinge length in all yeast H/ACA snoRNAs was carried out using the m-fold tool **(Table A1)**. This analysis suggests that most of the yeast H/ACA snoRNA have hinge lengths in the range of 10-25nt **(Figure A5)**, with exceptions like snR34, snR42, snR43, snR46 and snR84, which have longer hinge regions than usual.

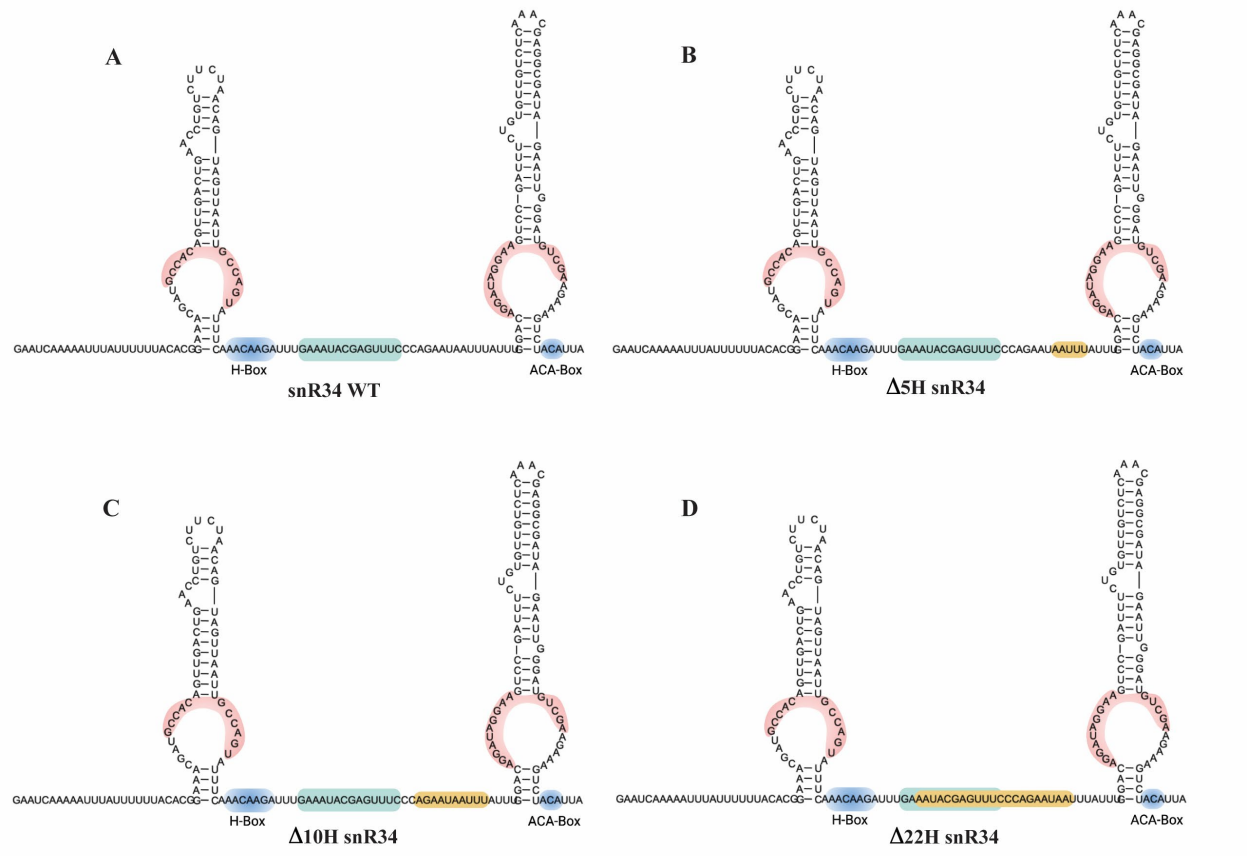

Figure 3.9. Successive deletions within the hinge region of snR34 H/ACA snoRNA. The H-box and ACA box are highlighted in blue color. The pseudouridylation pockets are highlighted in pink color, the cyan color represents a possible internal helix within the hinge region and the yellow color indicates the deleted sequences. **A** snR34 WT, **B** Δ5H snR34, **C** Δ10H snR34, and **D** Δ22H snR34.

Table 3.5. Binding and modification of substrate RNAs by snR34 H/ACA snoRNPs with shortened hinge regions. The corresponding modification time courses are shown in **Figure 3.10**, and the binding curves for determining dissociation constants are displayed in **Figure 3.11**. For reference, the data for full-length snR34 summarized in **Table 3.3** are listed here as well.

Guide RNA	Substrate RNA	v_0 (nM min^{-1})	K_D (nM)
snR34 WT	H89	32.0 ± 7.0	28.6 ± 2.5
snR34 Δ5H		31.0 ± 7.0	37.9 ± 5.1
snR34 Δ10H		13.0 ± 1.0	31.7 ± 2.9
snR34 Δ22H		11.0 ± 2.0	40.1 ± 9.6
snR34 WT	H90-92	42.0 ± 2.0	18.0 ± 3.9
snR34 Δ5H		43.0 ± 6.0	14.3 ± 3.6
snR34 Δ10H		23.0 ± 2.0	15.3 ± 4.3
snR34 Δ22H		25.0 ± 1.0	13.6 ± 3.5
snR34 WT	H89-92	65.0 ± 3.0	9.0 ± 2.6
snR34 Δ5H		67.0 ± 5.0	19.2 ± 6.0
snR34 Δ10H		15.0 ± 4.0	19.1 ± 5.0
snR34 Δ22H		12.0 ± 0.0	18.4 ± 6.0

To further characterize the importance of the snR34 hinge length for the formation of pseudouridine by both hairpins, I performed a tritium release assay with long H89-92 substrate RNA, which has two targets U at positions 2826 and 2880 and occupies both the 5′ and 3′ pseudouridylation pocket. I observed that the rate of formation of pseudouridine with Δ5H snR34 and wild-type snR34 are equal with 67.0 ± 5.0 nM min^{-1} and 65.0 ± 3.0 nM min^{-1}, respectively **(Figure 3.10C)**. On the other hand, I observed that the initial velocities for modifying H89-92 long RNA are only 15.0 ± 4.0 nM min^{-1} with Δ10H snR34 and 12.0 ± 0.0 nM min^{-1} with Δ22H snR34 **(Figure 3.10C)**. These reduced initial velocities in pseudouridine formation again suggest that the length of the hinge region of snR34 is important for efficient pseudouridylation in both pseudouridine pockets **(Table 3.5)**.

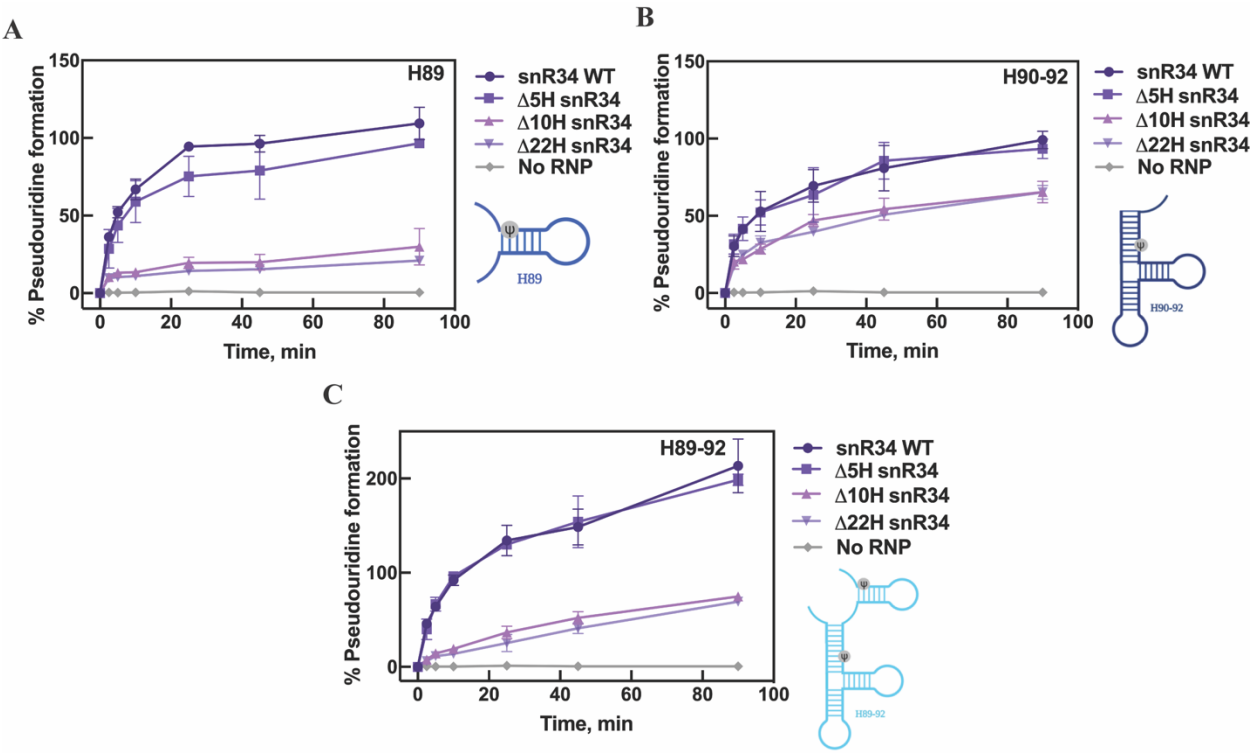

Figure 3.10. Pseudouridine formation by snR34 H/ACA RNP with different hinge regions between the hairpins. The hinge region of snR34 was systematically shortened by deleting 5, 10 or 22nt between the two hairpins (Δ5H, Δ10H, Δ22H snR34, respectively), and pseudouridine formation was observed by a tritium release assay. **A** Formation of pseudouridine in H89 which base-pairs to the 5′ hairpin. **B** Formation of pseudouridine in H90-92, a substrate for the 3′ hairpin. **C** Formation of pseudouridine in H89-92 RNA which can bind to both hairpins.

Based on this data, I hypothesized that the reduced pseudouridine formation might be the result of weaker interactions of H/ACA snoRNP core proteins with Δ10H and Δ22H snR34 guide RNA. To validate this hypothesis, the protein composition of H/ACA snoRNP assembled with snR34 WT and Δ22H snR34 was analyzed **(Figure A6)**. As observed in Figure A6, all four proteins Cbf5, Nop10, GST-Gar1 and Nhp2, are present in both samples of H/ACA snoRNP complexes with snR34WT and Δ22H snR34, respectively, which suggests that the core proteins might interact similarly with both snR34WT and Δ22H snR34.

3.4.5. Substrate RNA binding by H/ACA snoRNPs with shortened hinge regions

Since I observed a reduced initial velocity of pseudouridine formation for Δ10H and Δ22H snR34 variants when compared with wild type snR34 and Δ5H snR34 **(Table 3.5)**, I analyzed the interaction of substrate rRNA with H/ACA snoRNA hinge variants. This helps us understand how the hinge variant snR34 H/ACA snoRNA interacts with its substrate RNA molecules. Towards this goal, I measured the substrate RNA binding affinity to all three snR34 hinge variants by performing nitrocellulose filtration. The affinities of 5′ substrate H89 to Δ5H, Δ10H and Δ22H snR34 H/ACA snoRNP complex were 37.9 ± 5.1 nM, 31.7 ± 2.9 nM and 40.1 ± 9.6 nM, respectively **(Figure 3.115A-C)**. The H90-92 RNA binds to Δ5H, Δ10H and Δ22H snR34 H/ACA snoRNP with 14.3 ± 3.6 nM, 15.3 ± 4.3 nM and 13.6 ± 3.5 nM affinity, respectively **(Figure 3.11D-F)**. These results suggest that all snR34 variants harbouring H/ACA snoRNP complexes bind to its substrate RNA H89, H90-92 and H89-92 tightly and with similar affinity to wild-type. Moreover, the filter binding results confirmed that the long H89-92 5′-3′ substrate binds to the hinge variant Δ5H, Δ10H and Δ22H snR34 H/ACA snoRNP complex with the affinity of 19.2 ± 6.0 nM, 19.1 ± 5.0 nM and 18.4 ± 6.0 nM, respectively. Thus, in comparison to the affinity of wild type snR34 H/ACA snoRNA, all snR34 hinge variants bind to their substrates with similar affinity **(Table 3.5)**. These results suggest that the deletion of hinge regions has no effect on substrate RNA binding to the H/ACA snoRNP complex. Moreover, binding of the substrate RNA and pseudouridine formation are not correlated when the hinge region is shortened in snR34 H/ACA snoRNA.

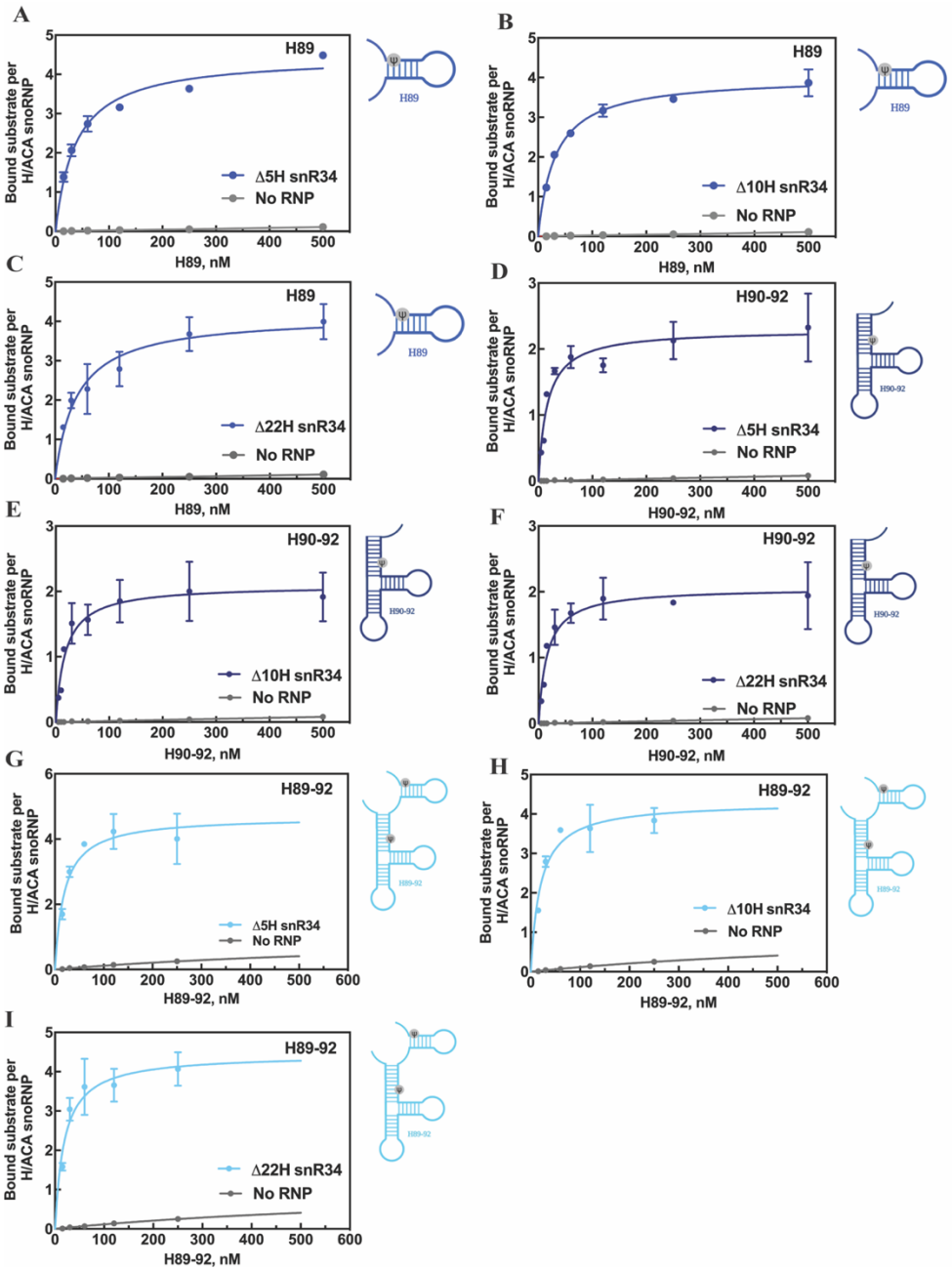

Figure 3.11. Substrate RNA binding affinity to H/ACA snoRNPs harbouring snR34 variants with different hinge lengths. The interaction of different substrate RNAs with the hinge variant snR34 H/ACA RNP was detected by nitrocellulose filtration **A, B,** and **C** Binding of H89 RNA to snR34 H/ACA snoRNPs with hinge variants Δ5H, Δ10H and Δ22H snR34, respectively. **D, E,** and **F** H90-92 RNA associating with snR34 H/ACA snoRNPs with hinge variants Δ5H, Δ10H and Δ22H snR34, respectively. **G, H** and **I** H89-92 RNA binding to snR34 H/ACA snoRNPs with hinge variants Δ5H, Δ10H and Δ22H snR34, respectively. The dissociation constant (K_D) was determined by fitting the data with a hyperbolic equation (see Materials and Methods).

3.4.6. Competitive inhibition between substrates for the two hairpins of snR34 H/ACA snoRNA

It was previously reported that H/ACA snoRNPs bind to short target RNAs (cognate RNA) as well as to near-cognate RNAs with similar sequences [55, 72]. To further understand the advantage of two-hairpin H/ACA snoRNAs for pseudouridine formation compared to single-hairpin RNAs, the formation of pseudouridine was assessed in the presence of independent substrate RNAs for each hairpin. Thereby I wanted to assess whether the presence of other structured RNA affects the formation of pseudouridine in the target rRNA substrate. We performed tritium release assays using combinations of radioactive target RNA (500 nM) and other non-radioactive RNA (500 nM) and monitored the formation of pseudouridine on radioactive RNAs. First, I analyzed the formation of pseudouridine in 5′ radioactive H89 structured substrate RNA in the presence of nonradioactive H90-92. The initial velocity of pseudouridine formation was observed to be 19 ± 4 nM min^{-1} **(Table 3.6)**, which indicates that the formation of pseudouridine is significantly decreased when compared with the initial velocity of modifying H89 alone (**Figure 3.12A**). Similarly, I performed the competitive tritium release experiments with the radioactive H90-92 substrate RNA and non-radioactive H89, where I measured the formation of pseudouridine directed by the 3′ hairpin of snR34. Here, the initial velocity is again significantly reduced to 12 ± 2 nM min^{-1} **(Table 3.6)**, in comparison to pseudouridylation of H90-92 alone which is modified with an initial velocity of 42.0 ± 2.0 nM min^{-1} (**Figure 3.12B**). In summary, these results indicate that the rate of formation of pseudouridine was affected in the presence of the other substrate RNA for snR34 H/ACA snoRNP complex. To further confirm these findings, I performed a similar tritium release assay in the presence of both radioactive substrate RNAs, H89 and H90-92, and again we observed that the initial velocity of pseudouridine formation is reduced to 16.0 ± 4.0 nM

min^{-1} **(Figure A3, Table 3.6).** This result is highly surprising as the corresponding long H89-92 substrate RNA is modified with a much higher initial velocity (65.0 ± 3.0 nM min^{-1}, **Table 3.3**).

Previously, from the Kothe lab, it was reported that the formation of pseudouridine on unstructured, short 23 nt substrate RNA molecule is not affected in competition conditions with short near-cognate RNA [72]. Hence, I hypothesized that the reduced pseudouridine formation observed here might be specific for competition between long-structured substrate RNAs but not present in the case of short competitor RNA. To confirm this hypothesis, I performed tritium release experiments with radioactive H89 RNA and a short unstructured RNA comprising U2880, which is a 23 nt segment of H90-92. The initial velocity for modifying H89 in the presence of this short RNA is 35.0 ± 7.0 nM min^{-1}, which confirms that the short unstructured substrate RNA is not competing with H89 for the same hairpin (**Figure 3.12A, Table 3.6)**.

Next, I asked whether any structured RNA could compete with cognate rRNA for binding the same hairpin and hinder the formation of pseudouridine. To answer this question, I performed a tritium release assay with radioactive H89 or H90-92 substrate RNA in the presence of non-radioactive *E. coli* $tRNA^{Phe}$ as a competitor. The initial velocity of the formation of pseudouridine in H89 in the presence of 500nM $tRNA^{Phe}$ is 38.0 ± 4.0 nM min^{-1} and 47.0 ± 5.0 nM min^{-1} for the H90-92 substrate RNA which is very similar to the initial velocities in the absence of $tRNA^{Phe}$. Thus, not all structured RNAs are competing for binding to the same hairpin of H/ACA snoRNA **(Figure 3.12A-B)**.

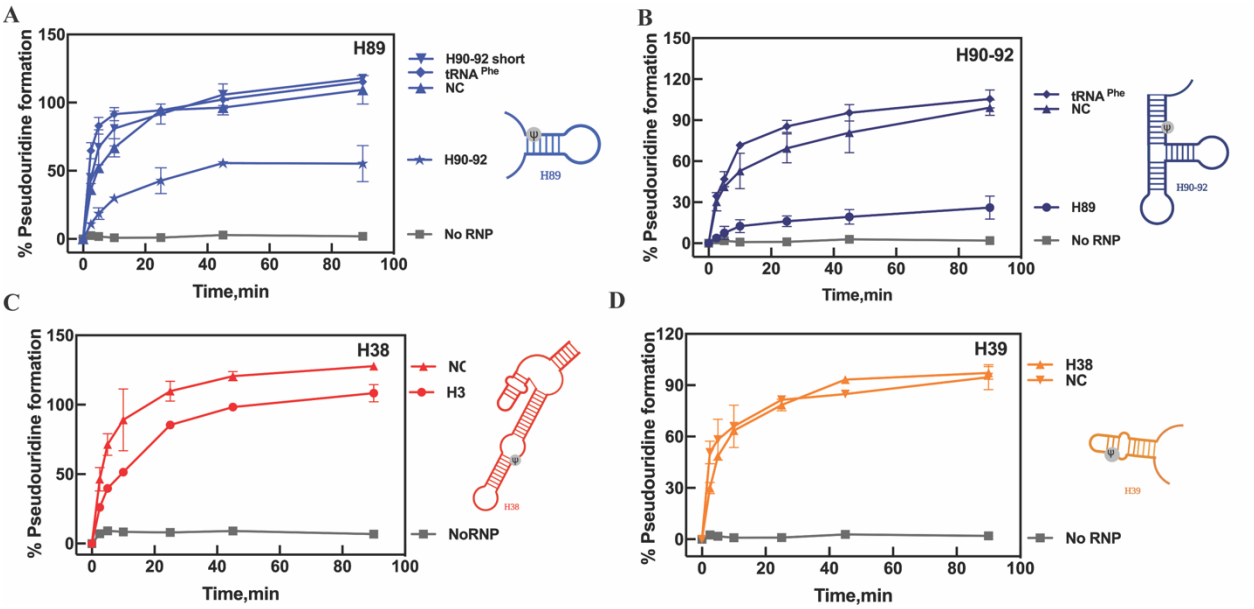

Figure 3.12. Pseudouridine formation by the snR34 and snR5 H/ACA RNP in the presence of other RNAs. Pseudouridine formation was analyzed by incubating 50 nM *in vitro* reconstituted H/ACA snoRNP complex containing either snR34 (**A** and **B**) or snR5 (**C** and **D**) with 500 nM of both [^{3}H-C5] uridine-labeled substrate RNA and non-radiolabeled RNA. Pseudouridylation was observed without the presence of competing non-radiolabeled RNA indicated as no competition (NC). At selected time points, the percentage of pseudouridine formation was determined by the tritium release assay. **A** Pseudouridine formation in [^{3}H-C5] uridine-labeled H89 substrate RNA directed by the 5′ hairpin of snR34 was measured in the absence or presence of non-radiolabeled RNA (tRNA^{Phe}, H90-92, and a short unstructured substrate H90-92 for the 3′ hairpin of snR34). **B** Analyzing the formation of pseudouridine in [^{3}H-C5] uridine labeled H90-92 substrate RNA base-pairing to the 3′ hairpin of snR34 was measured in the absence and presence of non-radiolabeled RNA (H89, tRNA^{Phe}). **C** Pseudouridylation in [^{3}H-C5] uridine-labeled H38 substrate RNA recognized by the 5′ hairpin of snR5 was measured in the absence or presence of non-radiolabeled H39 RNA. **D** Formation of pseudouridine on [^{3}H-C5] uridine-labeled H39 substrate RNA of snR5 was measured in the absence or presence of non-radiolabeled RNA H38, and a short unstructured substrate H90-92.

Table 3.6. Modification of substrate RNAs by H/ACA snoRNPs in the presence of competing RNA substrates. The corresponding modification time courses are shown in **Figure 3.12**, and the binding curves for determining dissociation constants are displayed in **Figures 3.6, 3.7 & 3.13**. For reference, the data for modification in the absence of competing RNA summarized in **Table 3.3** are listed here as well. **NC: no competitor RNA**.

Guide RNA	Substrate RNA (radioactive)	Competitor RNA (nonradioactive)	v_0 (nM min^{-1})
snR34	H89	NC	32.0 ± 7.0
	H90-92	NC	42.0 ± 2.0
	H89-92	NC	65.0 ± 3.0
	H89	H90-92	19.0 ± 4.0
	H90-92	H89	12.0 ± 2.0
	H89 & H90-92	-	16.0 ± 4.0
	H89	short H90-92	35.0 ± 11.0
	H89	$tRNA^{Phe}$	38.0 ± 4.0
	H90-92	$tRNA^{Phe}$	47.0 ± 5.0
snR5	H38	NC	71.0 ± 9.0
	H39	NC	59.0 ± 9.0
	H38-39	NC	85.0 ± 15.0
	H38	H39	53.0 ± 3.0
	H39	H38	60.0 ± 5.0
	H39	short H90-92	54.0 ± 8.0

3.4.7. Substrate RNAs can compete for binding to the same hairpin of H/ACA snoRNA

To confirm the hypothesis that the long rRNA competes with cognate rRNA for binding to the same hairpin, I performed competitive filter binding experiment with 5 nM snR34 H/ACA snoRNP in the presence of H89 or H90-92 radioactive RNA with increasing concentrations of non-radioactive H90-92 or H89, respectively. A fixed concentration of H89 (30nM), and H90-92 (20nM) was chosen based on the previous filter binding data to achieve approximately 50% binding as the H89 binds to snR34 H/ACA snoRNP complex with 28.6 ± 2.5 nM affinity and H90-92 bind with 18.0 ± 3.9 nM affinity **(Table 3.3)**. I observed a hyperbolic decrease in the binding of the radioactive RNA upon increasing the concentration of the other, non-radioactive substrate RNA, suggesting direct competition for binding of the two RNAs to the snR34 H/ACA snoRNP.

Hyperbolic fitting allowed us to determine that the H89 and H90-92 RNAs bind to its non-cognate hairpin with affinities of 79.0 ± 24.3 nM and 43.1 ± 16.2 nM, respectively. Thus, at the higher concentrations used in the tritium release assays (500 nM), both H89 and H90-92 RNA compete with each other for binding to the same hairpins of snR34 **(Figure 3.13D-E, Table 3.7)**.

3.4.8. Some, but not all RNA substrates compete for binding to H/ACA snoRNAs

To further confirm whether two long rRNA substrates can compete for the same hairpin also for other H/ACA snoRNAs than snR34, I performed tritium release assay with snR5 H/ACA snoRNP and H38 and H39 substrate RNAs in non-radioactive and radioactive states. The initial velocity in tritium release assay for radioactive H38 in the presence of non-radioactive H39 is 53.0 ± 3.0 nM min^{-1}, and it is 60.0 ± 5.0 nM min^{-1} for modifying radioactive H39 in the presence of non-radioactive H38 **(Figure 3.12C-D, Table 3.6)**. The comparison with initial velocities for modifying H38 (71.0 ± 9.0 nM min^{-1}) and H39 alone (59.0 ± 9.0 nM min^{-1}) demonstrates that there is no competition between the substrate RNA for snR5, unlike for snR34. To further confirm that the short unstructured rRNA molecules do not compete with the target substrate RNA, similar experiments were performed with radioactive H39 in the presence of short non-radioactive unstructured H90-92 and recorded the initial velocity of 54.0 ± 8.0 nM min^{-1}. When comparing this initial velocity with the initial velocity for H39 alone (59.0 ± 9.0 nM min^{-1}), it is evident that there is no competition between the short unstructured H90-92 substrate RNA and H39 **(Fig 3.12D, Table 3.6)**

To further understand these results, I asked the question whether these two rRNA fragments, H38 and H39, are competing for binding to same hairpin or not. To answer this question, I performed a competition filter binding experiments in the presence of a constant concentration of H38 (20 nM) and H39 (30 nM); again, I chose these concentrations based on the

affinity of H38 and H39 for the snR5 H/ACA snoRNP complex **(Table 3.3)**. Our data showed a hyperbolic decrease in the binding of radioactive RNA with increasing concentrations of the other substrate RNA, suggesting that the H38 and H39 compete with each other and that the H38 and H39 bind to its non-cognate hairpin with an affinity of 36.0 ± 17.0 nM and 106.0 ± 40.0 nM, respectively (**Figure 3.13A-B, Table 3.7)**.

To further verify that rRNA fragments can compete with the pseudouridylation pocket of other target substrates for binding, I performed a competition filter binding assay between snR5 substrate (non-radioactive H38) and snR34 substrate (radioactive H89) in the presence of H/ACA snoRNP complex harbouring snR34 as guide RNA (**Figure 3.13F)**. Surprisingly, I observed that H38 competes with H89 in binding to snR34 and binds with an affinity of 38.0 ± 12.0 nM to the snR34 H/ACA snoRNP complex. Further, a similar experiment with non-radioactive H90-92 and radioactive H39 with H/ACA snoRNP complex harbouring snR5 was performed, and the non-radioactive H90-92 competes with the binding of radioactive H39 **(Figure 3.13C, Table 3.7)** suggesting that the H/ACA snoRNP complex binds to many rRNA non-specifically with affinities in the nanomolar range. In other words, certain structured rRNA fragments can compete with cognate RNAs when binding to H/ACA snoRNPs.

Since this data suggests that the H/ACA snoRNP complex binds other substrate RNA molecules, I performed a filter binding assay with H89, which is an 5′ substrate for snR34, to analyze how it interacts with snR5 harboring H/ACA snoRNP complex to determine if the H/ACA snoRNP complex can bind un-specifically to 25S rRNA in search of its target substrate. Interestingly I found that the H89 binds to snR5-harboring H/ACA snoRNP complex with 36.1 ± 13.0 nM affinity which is comparable to the binding affinity 28.6 ± 2.5 nM of H89 substrate with

snR34 harbouring H/ACA snoRNP complex which confirms that the H/ACA snoRNP complex binds to other areas of rRNA during ribosome biogenesis (**Figure 3.7D, Table 3.3)**.

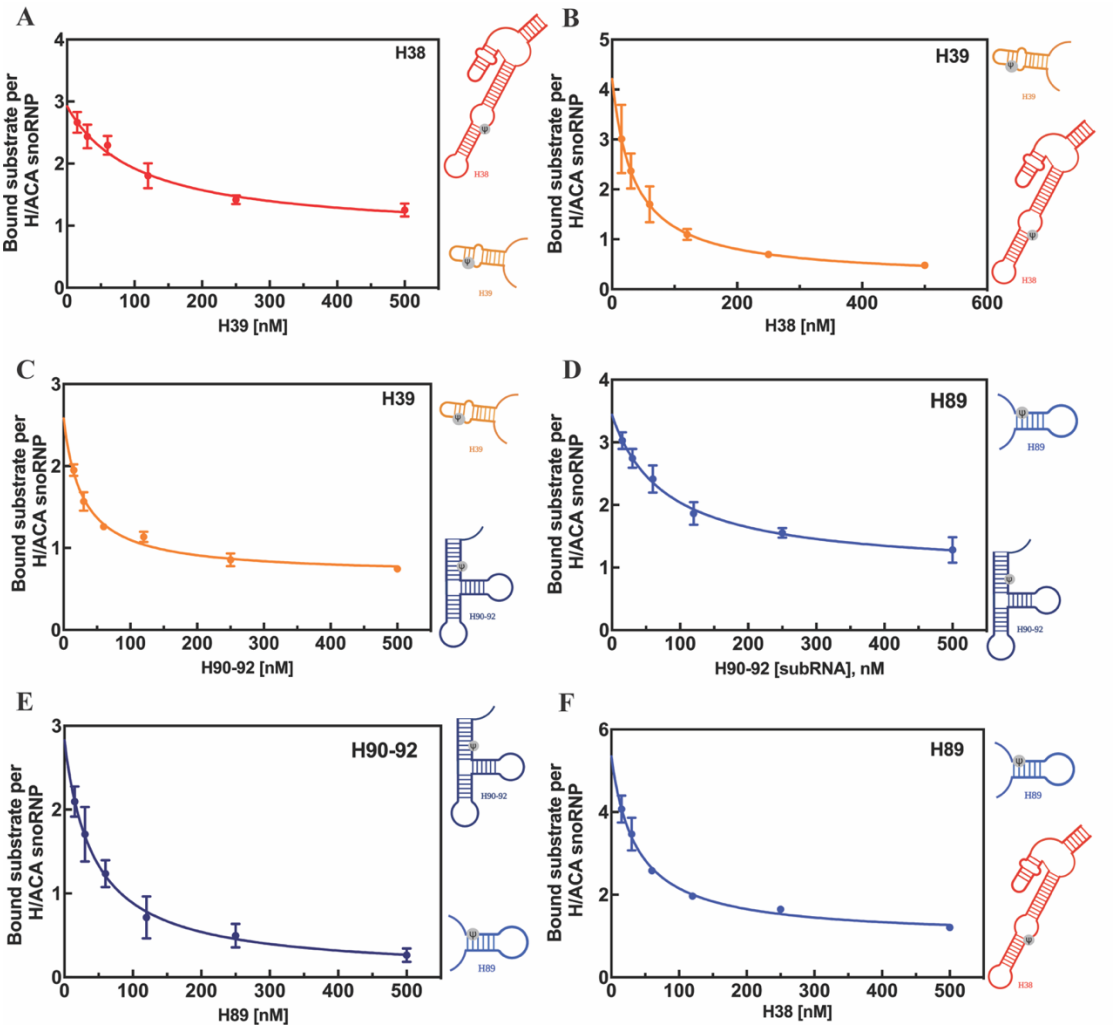

Figure 3.13. Competition of non-target RNA with the binding substrate RNAs to the snR34 and snR5 H/ACA snoRNP. The interaction of different substrate RNAs with the snR5 / snR34 H/ACA RNP was detected by nitrocellulose filtration with 5 nM snR5 H/ACA snoRNP in the presence of a constant concentration of H38 (20 nM) and H39 (30 nM) with increasing concentrations of non-radioactive H39 or H38 as competitor RNA, respectively. A similar competitive filter binding experiment with 5 nM snR34 H/ACA snoRNP in the presence of 30 nM H89 or 20 nM H90-92 with increasing concentrations of non-radioactive H90-92 or H89, respectively, was performed. **A** Binding of [^{3}H-C5] uridine-labelled H38 substrate RNA of snR5 under the competition of increasing concentration of non-radioactive H39 RNA. **B** Binding of [^{3}H-C5] uridine-labelled H39 substrate RNA of snR5 in the presence of increasing concentrations of non-radiolabeled H38 RNA. **C** Binding of [^{3}H-C5] uridine-labelled H39 substrate RNA of snR5 in the presence of non-radiolabeled H90-92 substrate RNA. **D** Binding of [^{3}H-C5] uridine-labelled H89 substrate RNA of snR34 in the presence of non-radioactive H90-92 RNA. **E** Binding of [^{3}H-C5] uridine-labelled H90-92 substrate RNA of snR34 in the presence of non-radiolabeled H89 RNA. **F** Binding of [^{3}H-C5] uridine-labelled H89 substrate of snR34 in the presence of non-radiolabeled H38 RNA. Dissociation constants (K_D) for the non-radioactive substrate RNA were determined by fitting the binding curves to the hyperbolic function (see Materials and Methods).

Table 3.7. Binding of substrate RNAs by H/ACA snoRNPs in the presence of competing RNA substrates. The corresponding nitrocellulose filtration experiments are shown in **Figure 3.13**.

Guide RNA	Substrate RNA (radioactive)	Competitor RNA (nonradioactive)	K_D (nM) of competitor RNA
snR34	H89	H90-92	79.0 ± 24.3
	H90-92	H89	43.1 ± 16.1
	H89	H38	38.0 ± 12.0
snR5	H39	H90-92	38.0 ± 12.0
	H38	H39	106.1 ± 40.0
	H39	H38	36.0 ± 17.0

3.4.9. Base-pairing potential between near-cognate substrates and H/ACA snoRNAs

Next, I investigated why certain RNA substrates lead to competitive inhibition of pseudouridine formation, whereas other RNA substrates can compete for binding without inhibiting RNA modification. Therefore, I systematically screened for potential base-pairing between the 5′ and 3′ pseudouridylation pockets of snR34 and the entire sequence of the respective other RNAs, H90-92 and H89 (**Figure 3.14**). Notably, this analysis uncovered several opportunities for H89 to base-pair in its 3′ region with the 3′ pseudouridylation pocket of snR34. One of these interactions is comprised of seven and three base pairs on either side of the pseudouridylation pocket, and another potential interaction includes seven base pairs on one site of the pseudouridylation pocket. Based on the similarities in base-pairing with the pseudouridylation pocket, these RNA sequences are called near-cognate RNAs. The extensive base-pairing explains that interaction with high affinity occurs between the near-cognate H89 RNA and the 3′ pseudouridylation pocket of snR34, which competes with binding of the cognate substrate region in H90-92 and inhibits pseudouridylation (**Figure 3.12B**). Notably, one of the predicted interactions even includes U2866 positioned at the 5′ site of the two unpaired nucleotides, which is usually pseudouridylated. However, since the H89 U2826C substrate cannot be pseudouridylated by snR34 (**Figure 3.4A**), U2866 is not modified by the 3′ pseudouridylation

pocket of snR34. Presumably, the three base pairs on the 5′ site of the pseudouridylation pocket do not provide sufficient stability for an efficient pseudouridylation [72]. Interestingly, the adjacent U2865 is pseudouridylated with the guidance of snR46 H/ACA snoRNA in *S. cerevisiae,* suggesting that snR34 and snR46 can compete in binding to this region of 25S rRNA.

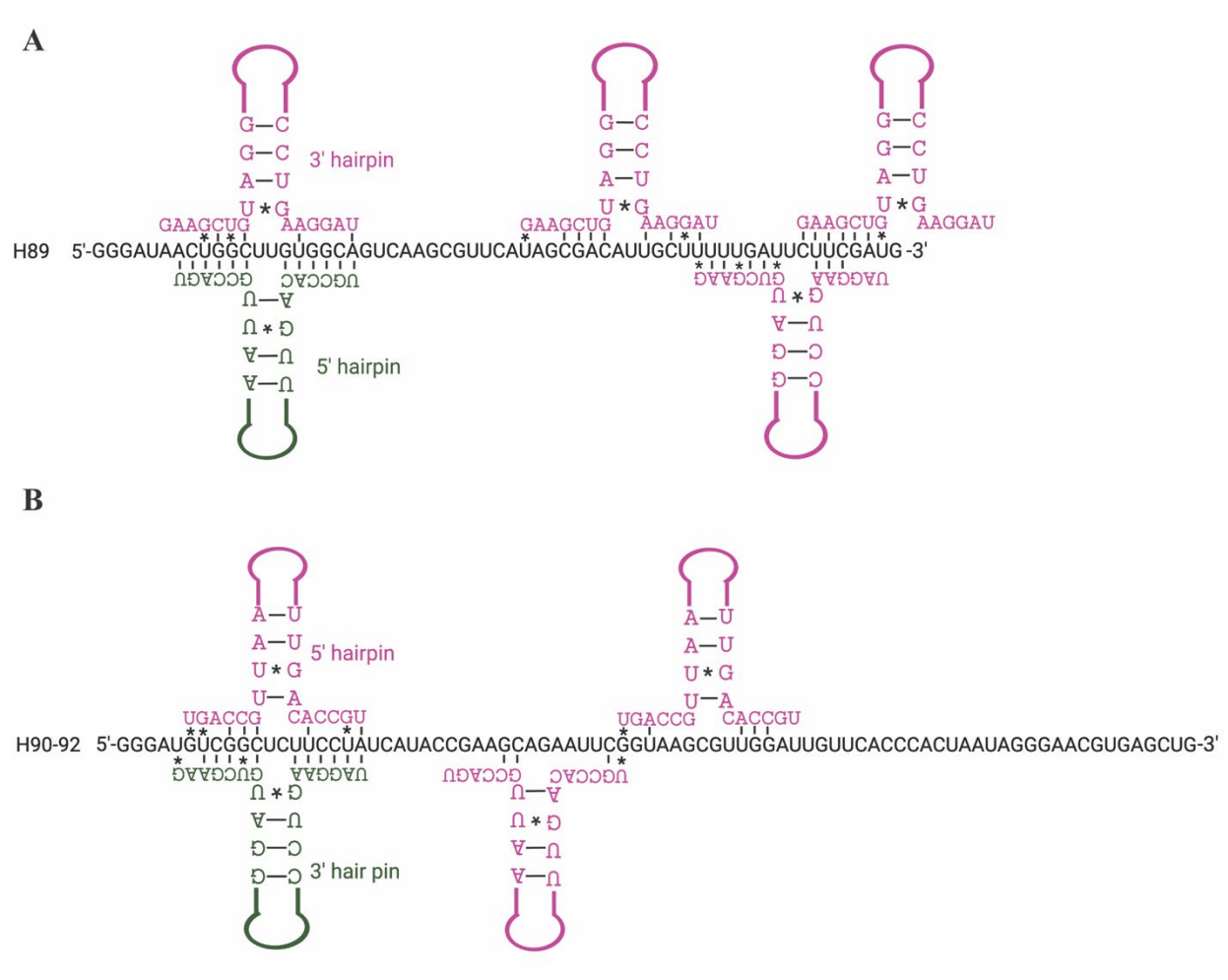

Figure 3.14. Potential base paring between H89 and H90-92 with snR34 pseudouridylation pockets. A Base pairing of H89 RNA (black color) with the 5′ hairpin pseudouridylation pocket of snR34 (green, cognate interaction), and potential base paring of H89 RNA with the 3′ pseudouridylation pocket (pink, near-cognate interactions). **B** Base pairing of H90-92 RNA (black color) with the 3′ hairpin pseudouridylation pocket of snR34 (green, cognate interaction), and potential base paring of H90-92 RNA with the 5′ pseudouridylation pocket (pink, near-cognate interactions).

Similarly, I identified potential near-cognate base-pairing interactions between H90-92 and the 5′ pseudouridylation pocket of snR34 (**Figure 3.14B**). Here, at most, five plus three base-pairs can be formed between the two sites of the pseudouridylation pocket. This explains the inhibition pseudouridylation of H89 by snR34 in the presence of H90-92 (**Figure 3.12A**), albeit to a lesser extent than in the previously described case where H89 strongly binds to the 3′ pseudouridylation pocket and inhibits modification of H90-92.

Surprisingly, there is no significant difference in pseudouridine formation in H38 and H39 when the other rRNA fragment is present (**Fig 3.12**), even though the structured rRNA fragments H38 and H39 can bind to the non-cognate hairpin (**Figure 3.13**). To explain this, I again analyzed the base-pairing potential between H38 and H39 with their respective non-cognate hairpins in snR5 (**Figure 3.15**). Whereas some base-pairing potential was identified, no more than four consecutive base-pairs are possible for the non-cognate interaction of H38 and for the non-cognate interaction of H39. While these numbers of base-pairing may allow for binding to occur, the interaction is likely weak and dissociates rapidly such that pseudouridylation is not inhibited, thereby explaining our experimental observations. Therefore, these interactions can be classified as non-cognate.

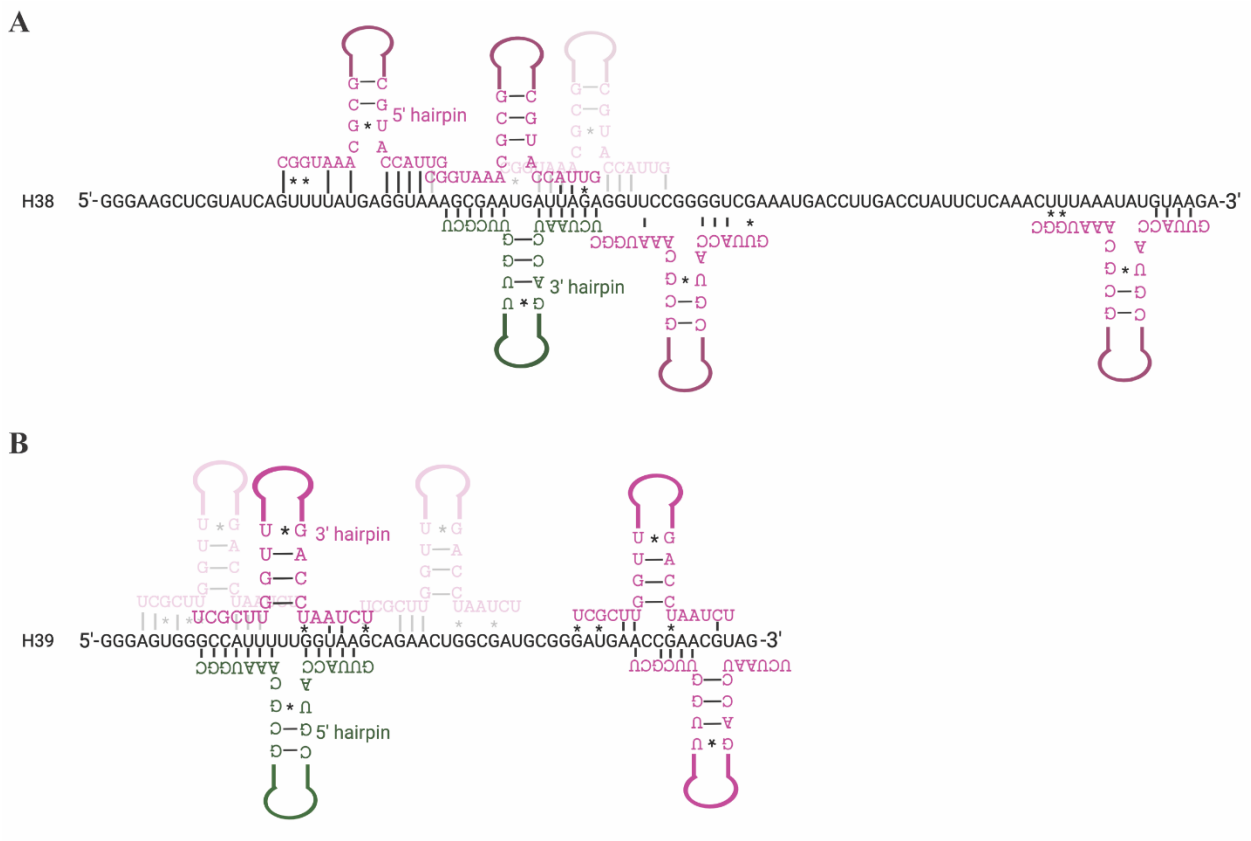

Figure 3.15. Potential base paring between H38 and H39 with snR5 pseudouridylation pockets. A Base pairing of H38 RNA (black color) with the 3′ hairpin pseudouridylation pocket (green, cognate interaction) and potential base paring of H38 RNA with the 5′ hairpin pseudouridylation pocket (pink, non-cognate interactions). **B** Base paring of H39 RNA (black color) with the 5′ hairpin pseudouridylation pocket (green, cognate interaction) and potential base paring of non-cognate H39 RNA with the 3′ hairpin pseudouridylation pocket (pink, non-cognate interactions).

3.5. Discussion

In this study, I have addressed the question of how eukaryotic H/ACA snoRNPs comprised of two hairpins with two pseudouridylation pockets efficiently modify the long rRNA. First, it was demonstrated that the H/ACA snoRNPs tightly bind and rapidly modify folded substrate RNA, indicating that they can unwind secondary structure in rRNA without impact on the catalytic activity **(Figure 3.4, 3.6, and 3.7).** Second, a complementary set of multiple data was collected to understand why most eukaryotic H/ACA snoRNAs have two hairpins allowing them to target two sites for pseudouridylation. This data showed that pseudouridylation in one active site is independent of binding substrate RNA to the other active site suggesting that there is no allosteric interaction between the two active sites (**Figure 3.5**). Moreover, this study revealed that a minimum distance between the two hairpins in H/ACA snoRNA is required to facilitate efficient pseudouridine formation, which is likely due to steric constraints by the H/ACA proteins, in particular Cbf5. Notably, this proves that H/ACA snoRNPs rapidly modify long substrate RNA at two distinct pseudouridylation sites **(Figure 3.4)**. However, unexpectedly I observed competitive binding and inhibition for some combinations of two separate substrate RNAs due to binding of the wrong RNA to the pseudouridylation pocket, thereby blocking its activity **(Figures. 3.12 & 3.13)**.

3.5.1. H/ACA snoRNP efficiently binds and might unwind structured rRNA

H/ACA snoRNPs function in the dense structure of the nucleolus where the local RNA concentration is very high. Presumably, H/ACA snoRNPs act very early during ribosome biogenesis while the pre-rRNA is being transcribed by RNA polymerase-I. Most likely, the pre-rRNA will begin folding immediately such that H/ACA snoRNPs will encounter pre-rRNA that may contain structure resembling the fold of the mature ribosome or pre-rRNA that is misfolded

adopting other secondary structures. This raises the question how H/ACA snoRNPs identify their correct target sites among the high concentration of many large RNAs and how they access their target sites which are likely to be part of a secondary structure. As evident in our experiment, the presence of secondary structure is no hindrance for H/ACA snoRNPs as they modify structured RNAs as quickly as they modify short, unstructured RNA fragments **(Figure 3.4)** [72]. Here, I have thereby demonstrated the ability of H/ACA snoRNPs to unwind secondary structure for two snoRNAs (snR5 and snR34) and four pre-rRNA fragments (H38, H39, H89, H90-92). I performed the experiments using both slow-cooled (likely folded) and snap-cooled (potentially unfolded) RNA (**Figure A4)**, and these findings showed that H/ACA snoRNPs does not rely on the action of helicases or other external factors to unwind pre-rRNA prior to modification or to be efficiently released from pre-rRNA. Thus, recent reports about the function of helicases during ribosome synthesis are likely specific for certain snoRNAs and the assembly of specific ribosomal elements [170, 171]. But how do H/ACA snoRNPs gain access to their target site? In many cases, such as in the pre-rRNA fragments used here, the pseudouridylation site is located within an RNA helix, but adjacent single-stranded regions also contribute to base-pairing to the H/ACA snoRNA. Therefore, it has been hypothesized that H/ACA snoRNPs base-pair first to the adjacent single-stranded regions in pre-rRNA and then melt the secondary structure in a strand-displacement mechanism where each melted base-pair in pre-rRNA is replaced by a base-pair of pre-rRNA to H/ACA snoRNA. This mechanism will likely apply to both native rRNA structures as well as misfolded pre-rRNA structures which are likely to also possess single-stranded regions in the vicinity to the target uridine. Thus, H/ACA snoRNPs are most likely capable of unwinding also misfolded pre-rRNA structures. Upon release of the pre-rRNA from H/ACA snoRNPs, the pre-rRNA will spontaneously fold again. Thus, the H/ACA snoRNPs may provide pre-rRNA with a

second chance at correct folding thereby acting indirectly as pre-rRNA chaperones similar to stand-alone RNA modifying enzymes [172-174].

The RNA unwinding activity of H/ACA snoRNPs thus depends on their ability to bind the substrate RNA through base-pairing to the pseudouridylation pocket. Our quantitative experiments showed that substrate RNA is bound with low nanomolar affinity for all four pseudouridylation pockets tested **(Figure 3.6 & 3.7).** As previously reported for other pseudouridine synthases, higher affinity of H/ACA snoRNPs for target RNAs harbouring a C at the modification sites (such that they cannot be modified) was observed compared to U-containing RNAs that will be pseudouridylated during the nitrocellulose filtration assays **(Figure 3.5 & Table 3.3)**. Thus, the formation of pseudouridine seems to result in a decrease in affinity and facilitate the product release. When comparing the binding of single-target RNAs to one pseudouridylation pocket with the binding of longer, dual-target RNA to two pseudouridylation pockets, the affinity is increased up to two-fold for long substrates **(Table 3.3)**. Hence, the presence of two hairpins in H/ACA snoRNA strengthens the interaction of H/ACA snoRNPs with pre-rRNA.

3.5.2. The two H/ACA snoRNA hairpins operate independently and require a minimum distance

In order to assess how efficient our *in vitro* reconstituted H/ACA snoRNPs form pseudouridines, it is interesting to compare our reported initial velocities (**Table 3.3**, 30 – 85 nM min^{-1}) to previously reported activities of archaeal and yeast H/ACA snoRNPs. Unlike most eukaryotes, the archaeal H/ACA sRNP possesses one-two-three hairpins in the H/ACA sRNA. When reconstituting the archaeal H/ACA sRNP, initially only single-turnover experiments were conducted which do not allow determination of an initial velocity [42, 175]. However, the Keyiong lab reported two studies with multiple turnover activity of *Pyrococcus furiosus* (Pf) H/ACA sRNP.

In Duan et al. the Keyiong lab reported that the *Pyrococcus furiosus* (Pf) H/ACA sRNPs (1000 nM) were able to convert 61% (12,200 nM) of 20 μM substrate RNA into the pseudouridine-containing product in one hour at 37 °C in multi-turnover conditions, so, the approximate initial velocity of *Pyrococcus furiosus* (Pf) H/ACA RNPs is ~204 nM min^{-1}[176]. To compare this initial velocity measured at high enzyme concentration (1000 nM) to our data, the initial velocity is divided by the enzyme concentration to estimate an "initial velocity per enzyme". According to the Duan et al. the initial velocity per *Pyrococcus furiosus* (Pf) H/ACA sRNPs is 0.20 min^{-1} In another study from Keyiong's lab, Li et al., reported that single hairpin *Pyrococcus furiosus* (Pf) H/ACA sRNPs (1000 nM) were able to convert only 21% (~4000 nM) of 20 μM (20000 nM) substrate RNA into the pseudouridine-containing product in one hour at 37 °C in multi-turnover conditions, such that the approximate initial velocity of *Pyrococcus furiosus* (Pf) H/ACA RNPs is ~70 nM min^{-1} [26]. This corresponds to an initial velocities per Pf H/ACA RNP of 0.07 min^{-1}. Comparing these initial velocities per enzyme to our data (v0 of 30 – 80 nM min^{-1} with 50 nM enzyme, i.e., about 1 min^{-1}) suggests that our reconstituted yeast H/ACA snoRNPs are much more active. However, it should be noted that the activity of *P. furiousus* H/ACA snoRNPs was measured at 37 °C, although the optimal growth temperature of this thermophile is 98 to 100 °C.

In addition, Li et al. also determined the pseudouridylation efficiency of reconstituted yeast H/ACA snoRNPs. They reported that the snR5 H/ACA snRNP complex (200 nM) has an initial velocities of ~166 nM min^{-1} and ~116 nM min^{-1} for 3' and 5' hairpin substrates, respectively. Thus, we can calculate an initial velocity per enzyme of 0.83 min^{-1} and 0.58 min^{-1}. Interestingly, we observed that our snR5 H/ACA snoRNP complex (50 nM) was able to convert target uridine to pseudouridine with initial velocities of 59 ± 9 nM min^{-1},71 ± 9 nM $min^{-1}$for 5' and 3' hairpins, respectively, corresponding to initial velocities per enzyme of 1.18 min^{-1} and 1.42 min^{-1},

respectively, demonstrating that our H/ACA snoRNPs are significantly more active when using the same guide RNA, snR5. The Hengesbach lab has also reconstituted and characterized *S. cerevisiae* H/ACA snoRNPs. According to Trucks et.al, the initial velocity of pseudouridine formation by yeast snR81 H/ACA snoRNP in the 5' substrate is 54 nM min^{-1}, and for the 3' substrate, it is reported as 17 nM min^{-1}. However, their reaction contains 100 nM H/ACA snoRNA and 2 μM H/ACA proteins; as the snoRNA is the limiting factor, we assume an enzyme concentration of 100 nM [166]. Accordingly, the initial velocity per enzyme reported by Trucks et al. is 0.54 min^{-1} and 0.17 min^{-1} for the 5' and 3' substrate of snR81, respectively. Again, the comparison to the data reported in this thesis suggests that our H/ACA snoRNP preparations are more active than the ones reported in the literature by Li et al. and Trucks et al. supporting our assessment that H/ACA snoRNPs are efficient in forming pseudouridines in our hands.

A fundamental question regarding snoRNAs is why eukaryotic snoRNAs typically target two sites for modification, in the case of H/ACA snoRNAs, by having two hairpins with two pseudouridylation pockets. As single-hairpin H/ACA snoRNA has a much lower modification activity than two-hairpin H/ACA snoRNA [26, 55], it has been hypothesized that the two active sites could act co-operatively where one site enhances the activity of the other. However, a series of experiments using a combination of RNAs binding to one or both hairpins of H/ACA snoRNAs reveals that pseudouridine formation occurs equally fast whether or not the RNA is bound to one or both hairpins (**Figure 3.4 & 3.5**). Therefore, the two active sites are functioning independently from each other ruling out allosteric effects, and other reasons must have led to the wide conservation of two hairpins in H/ACA snoRNAs.

Next, through a series of deletions in the hinge region of the H/ACA snoRNA snR34, it has been revealed that a minimum length of this region is required for activity in both

pseudouridylation pockets of this complex **(Figure 3.8)** whereas the binding of RNA to these snR34 variants is not significantly affected **(Table 3.5)**. Only 5 nt can be removed from the 41-nt hinge region without loss of pseudouridylation activity by the snR34 H/ACA snoRNP. This is quite surprising since snR34 has a much longer hinge region than other *S. cerevisiae* H/ACA snoRNA, e.g., snR3 has a 15 nt hinge region, and snR36 has a 10 nt hinge region. However, yeast H/ACA snoRNAs often have additional helices referred to as inserted hairpin elements positioned, and RNA structure prediction reveals that at least one small hairpin comprised of 14 nt can form within the snR34 hinge region, which effectively reduces the length of the hinge [165]. Further, shortening the hinge may reduce H/ACA snoRNP activity due to steric hindrance between the H/ACA proteins. In particular, Cbf5 is bound to the base of each H/ACA snoRNA hairpin and interacts with the adjacent H box and ACA box through its PUA domain [27]. Thus, Cbf5 presumably occupies parts of the hinge region, and shortening the hinge region could result in a steric clash of the two Cbf5 proteins. It was reported previously that correct positioning of the Cbf5 protein relative to the H/ACA snoRNA and the pseudouridylation pocket is essential for activity [55], and therefore, the steric clash between the Cbf5 proteins may result in their displacement and loss of catalytic activity. Interestingly, the pseudouridylation effects of shortening the hinge regions are more pronounced for modification of a substrate binding to the 5′ hairpin **(Figure 3.8A)** than for a substrate binding to the 3′ hairpin **(Figure 3.8B)**. Hence, the Cbf5 protein binding to the 5′ hairpin and the H box in the hinge region may be most impaired in its pseudouridylation activity when the hinge region is shortened, and steric clashes arise. On the other hand, the pseudouridylation pockets remain unperturbed upon deletions within the hinge regions explaining why binding of substrate RNA to the H/ACA snoRNA is not affected. However, the bound substrate RNA will likely not be positioned in the active site of Cbf5.

3.5.3. Two-hairpin H/ACA snoRNAs facilitate RNA target selection by minimizing competitive inhibition by near-cognate RNA sequences

Interestingly, it was observed that a long substrate RNA is rapidly modified at both target uridines; however, when this long substrate is split into two separate RNAs, pseudouridylation by snR34 H/ACA snoRNPs (but not by snR5 H/ACA snoRNPs) is significantly impaired **(Figure 3.9)**. This observation is at first very surprising as it suggests an inhibitory effect of one RNA on the pseudouridylation of the other RNA. Indeed, these results are able to show that the two RNAs directly compete with each other for binding to the snR34 H/CA snoRNP **(Figure 3.10)**. This competing binding of both RNAs to the snR34 H/ACA snoRNPs can be explained by their base-pairing potential **(Figure 3.14)**. Upon close inspection, it is confirmed that the H89 RNA substrate can form several near-cognate base-pairing interactions with the 3′ hairpin of snR34 although it is pseudouridylated when interacting with the snR34 5′ hairpin. Likewise, the H90-92 RNA substrate can base-pair to both its cognate 3′ hairpin in snR34 as well as to the 5′hairpin. These near-cognate base-pairing interactions are not leading to pseudouridine formation; however, they are stable enough and comprise sufficient base-pairs to lead to the competition of the RNAs binding to the snR34. In contrast, no inhibitory competitions are observed when one long substrate RNA, H89-92, is modified by the snR34 H/ACA snoRNP **(Fig 3.4C)**.

Our findings that one long substrate RNA (**Figure 3.4**) is modified more efficiently than two shorter RNAs with the same overall sequence (**Figure 3.13, Figure A3**) can explain the wide-spread conservation of H/ACA snoRNAs containing two hairpins and two pseudouridylation pockets. As further outlined below, the presence of two hairpins allows H/ACA snoRNA to modify long substrate RNAs at multiple sites in an efficient manner (**Figure 3.16**). When a substrate RNA region base-pairs to its cognate pseudouridylation pocket in a two-hairpin H/CA snoRNA, then

the substrate RNA is effectively tethered to the H/ACA snoRNPs. This tethering will greatly facilitate the cognate interaction of the other target site with the second pseudouridylation pocket. Therefore, upon correct binding of a long substrate in one pseudouridylation pocket, the likelihood is very high that correct substrate RNA interaction with the other pseudouridylation pocket will occur. Thus, the main advantage of two-hairpin H/ACA snoRNAs is likely the efficiency of binding the second target site in a correct manner. In contrast, single-hairpin H/ACA snoRNAs may be prone to inhibition by binding to near-cognate RNAs similar to the situation described here where separate target RNAs cause competitive inhibition. When a single-hairpin H/ACA snoRNAs selects its cognate RNA from a pool of RNAs, then near-cognate RNAs with similar sequences to the cognate RNA may have an equal chance of binding to the H/ACA snoRNA. In other words, the advantage of pre-tethering substrate RNA in the correctly bound manner to the H/ACA snoRNP complex is lost for single-hairpin H/ACA snoRNAs.

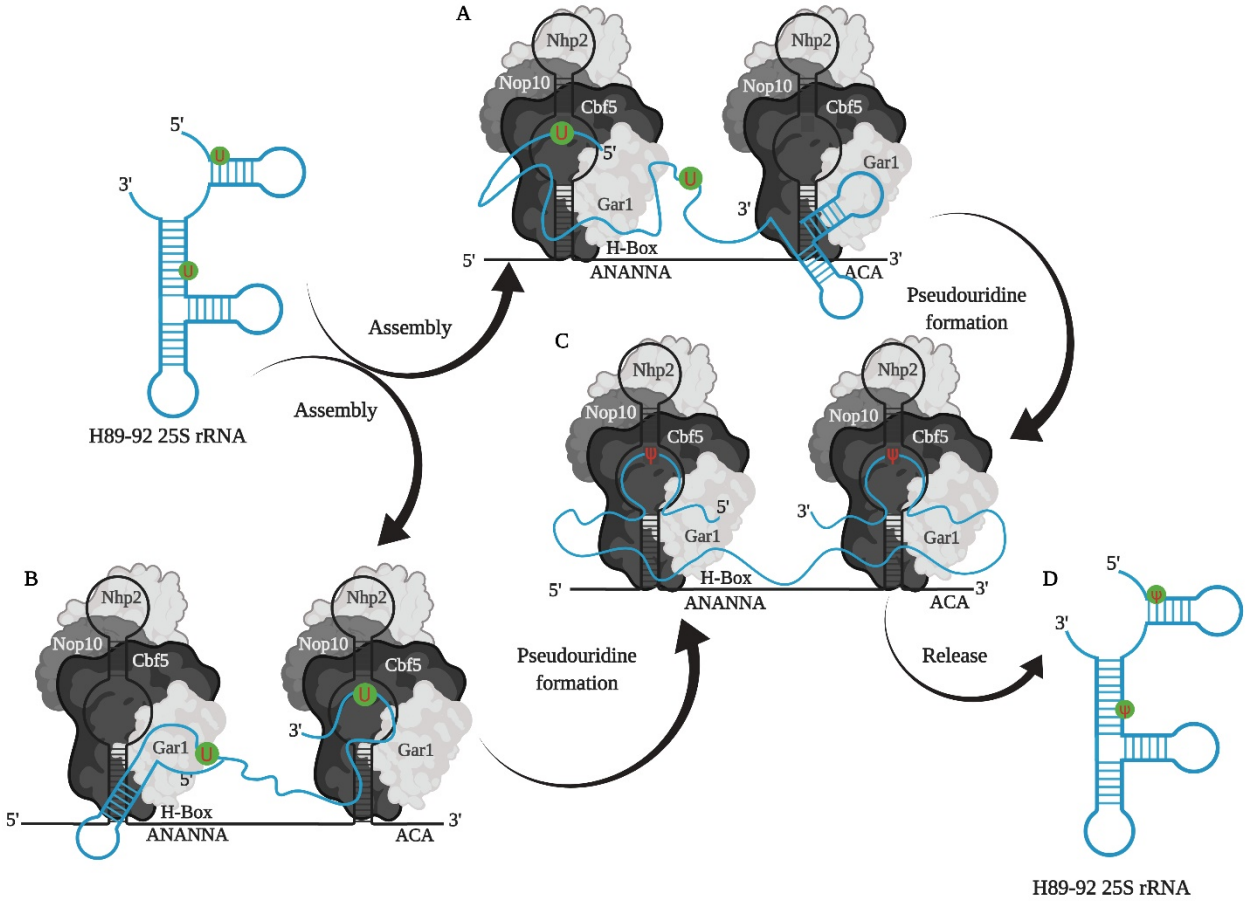

Figure 3.16. Model for efficient substrate RNA recruitment by two-hairpin H/ACA snoRNPs. Long substrate RNAs with multiple pseudouridylation sites can first bind to one of the pseudouridylation pockets of a two-hairpin H/ACA snoRNP. Thereby, the substrate RNA is tethered to the H/ACA snoRNP, and the correct binding of the second target uridine to the second pseudouridylation pocket is greatly facilitated. By increasing the chance of both target sites to interact with the correct pseudouridylation pocket, two-hairpin H/ACA snoRNPs can efficiently modify both sites while minimizing the competition between these sites for base-pairing to the pseudouridylation pockets. Hence, two-hairpin H/ACA snoRNAs are advantageous for the modification of substrate RNA at multiple sites.

This study further demonstrated the advantage of two-hairpins for modification of long substrate RNA as a proof-of-concept for snR34 H/ACA snoRNA and its interaction with helices H89 to H92, a 140-nt fragment of 25S rRNA. This effect was not observed for the interaction of snR5 with H38 and H39, a 174-nt fragment of 25S rRNA where efficient base-pairing is only observed between the cognate target site and the respective pseudouridylation pocket **(Figure**

3.15). In future, it would be interesting to systematically analyze the potential of H/ACA snoRNAs to bind to near-cognate sites which may lead to competition and inhibition. It would also be insightful to test whether the advantage of tethering substrate RNA to the H/ACA snoRNPs for efficient binding of the second target site holds true for H/ACA snoRNAs that modify two uridines that located in very different regions of pre-rRNA. Notably, similar mechanistic principles may apply to C/D snoRNPs mediating the 2`-O-methylation of rRNA at two different sites. However, this model may not be directly applicable for H/ACA snoRNAs like snR9, snR10, snR11, snR32, snR33, snR35, snR36, snR37, snR46, snR84 and snR85 as only one single hairpin of these H/ACA snoRNA is involved in pseudouridylation of rRNA. Therefore, these snoRNAs might have a different mechanism in target U binding. The H/ACA snoRNAs have previously been suggested to have RNA chaperone function and aid in rRNA folding; accordingly it is possible that the second hairpin of these H/ACA snoRNA might possess stabilizing function and might help the rRNA folding. U3 C/D snoRNP and the U17/snR30 H/ACA snoRNP have also been hypothesized to act as chaperons in pre-rRNA folding [82, 177]. However, how many other H/ACA snoRNAs possess this function has yet to be elucidated [178]. Moreover, recently it has been reported by the Kiss lab that human H/ACA snoRNAs snoRA53, snoRA57 showed differential base pairing ("Acrobatics of H/ACA snoRNA") with target substrate RNA while modifying target uridines which are located in close proximity. [179]. They proposed a novel mechanism called "one-for-two" where a single guide hairpin of H/ACA snoRNA can target two uridines adjacent to each other or separated by 1 nt. The authors performed extensive mutational analysis and confirmed that the base pairings of the guide RNA with the substrate RNA determine the dual guide activity [180]. The authors also found a similar H/ACA snoRNA (snR82) in yeast, which is predicted to follow a "one-for-two" mechanism. However, the molecular mechanism

underlying guide RNA acrobatics in this one-for-two mechanism is yet to be understood. Our model described in this study might be one of the many mechanisms adopted by two hairpin H/ACA snoRNA molecules while converting the target uridine to pseudouridine. It is highly possible that our model will be applicable to human H/ACA snoRNAs, as humans have more than 100 snoRNAs unlike yeast which only possess 29 H/ACA snoRNAs.

3.6. Conclusion

In summary, our study provides mechanistic insights into the function and the evolutionary advantage of H/ACA snoRNAs with two hairpins for modifying long pre-rRNA during ribosome biogenesis. Through systematically dissecting the interactions of H/ACA snoRNPs with long, structured rRNA fragments, two key features of two-hairpin H/ACA snoRNPs are disclosed or discovered. First, H/ACA snoRNPs can efficiently unwind secondary structures in their target RNA, which may be facilitated by a strand-displacement mechanism. Second, the two-hairpin arrangement enables H/ACA snoRNPs to correctly bind to both target sites, presumably because binding of the first target uridine tethers the substrate RNA to the H/ACA snoRNPs and facilitates correct binding of the second target uridine **(Figure 3.11)**. Thereby, two-hairpin H/ACA snoRNAs are very effective at modifying long substrate RNAs by minimizing near-cognate, and possibly inhibitory interactions with other regions in the substrate RNA.

CHAPTER 4: GAR DOMAINS PLAY A PIVOTAL ROLE IN EFFICIENT PSEUDOURIDINE FORMATION

4.1. Introduction

The mature H/ACA snoRNP complex harbours four core proteins, including Cbf5/Dyskerin (C), Nop10 (N), Gar1 (G) and Nhp2 (P). Cbf5/Dyskerin is an RNA-dependent pseudouridine synthase with N- and C- terminal lysine-rich regions called KKE/D domains, and these KKE/D domains were previously reported to interact with microtubules [181]. Earlier studies on the H/ACA snoRNP complex revealed that the core domain of Cbf5 1–394 ($Cbf5^{\Delta}$) **(Figure 4.1)**, lacking the C-terminal KKE/D domain, is capable of pseudouridine formation *in vitro* in rRNA fragments [26].

Mutations in the Dyskerin gene (*DKC1*) lead to a pathological condition called X-linked Dyskeratosis congenita (X-DC) [182]. In brief, X-DC is a congenital disorder characterized by bone marrow failure, hematopoietic malignancies, reduced cell proliferation, skin abnormalities, pulmonary fibrosis, and increased tumorigenesis [139]. Various Dyskeratosis congenita mutations were reported within or near the KKE/D domains of Cbf5, which suggests that they might have a vital role in the function of Cbf5/Dyskerin other than catalysis of pseudouridine formation as these regions are absent in the archaeal homolog of the Cbf5 [26]. The structure of human telomerase holoenzyme reveals that most of the X-DC mutations are found in the interface of two Dyskerin molecules and near the RNA-binding domain of the Dyskerin protein [27]. In addition, a few X-DC substitutions (e.g. of R158) are located near the telomerase-specific protein TCAB1 [25, 30-32, 139-141].

The Gar1 protein was previously reported to be involved in substrate turnover, and it has N- and C-terminal glycine arginine-rich domains called GAR domains [26]. A previous study

reported that the truncation of GAR domains from Gar1 protein (Gar1$^{\Delta}$ (32-124)) **(Figure 4.1)** does not influence pseudouridine formation [26]. Interestingly, these GAR domains are absent in archaeal homologs suggesting that the eukaryotic Gar1 protein might have additional abilities **(Figure A7)**.

The current chapter discusses how these unstructured C-terminal KKE/D tails of Cbf5/Dyskerin and N- and C-terminal GAR domains of Gar1 play a role in t pseudouridine formation. Towards this, a series of ternary complexes, such as Cbf5•Nop10•GST-Gar1 (CNG), Cbf5$^{\Delta}$•Nop10•GST-Gar1 (C$^{\Delta}$NG), Cbf5$^{\Delta}$•Nop10•GST-Gar1$^{\Delta}$ (C$^{\Delta}$NG$^{\Delta}$) and Cbf5•Nop10•GST-Gar1$^{\Delta}$ (CNG$^{\Delta}$) are purified using bacterial co-expression systems and used to understand the importance of N- and C-terminal GAR domains of Gar1 and the C-terminal KKE/D domain of Cbf5 for the function of H/ACA snoRNP complex concerning pseudouridine formation in *in vitro* biochemical assays. In this study, yeast snR34 (corresponding to U64 in humans) is used as H/ACA guide RNA, which has two target uridines, U2826 and U2880, located on the PTC domain of the 25S rRNA in H89 and H90-92 helices, respectively **(Figure 4.2)**. The results discussed in this chapter reveal that the unstructured regions of Gar1 are essential for efficient pseudouridine formation but not the C-terminal KKE/D domain of Cbf5. Moreover, this study shows that the KKE/D and GAR domain truncations do not affect the formation of a stable H/ACA snoRNP complex and do not alter its ability to bind substrate RNA. Further, this chapter discloses that truncation of GAR domains abolishes the RNA-binding ability of the Gar1 protein. This study uncovers that GAR extensions help Gar1 to bind RNA non-specifically. Moreover, the GAR domains might be helpful in the efficient formation of pseudouridine by placing the substrate RNA accessible to the catalytic site of Cbf5.

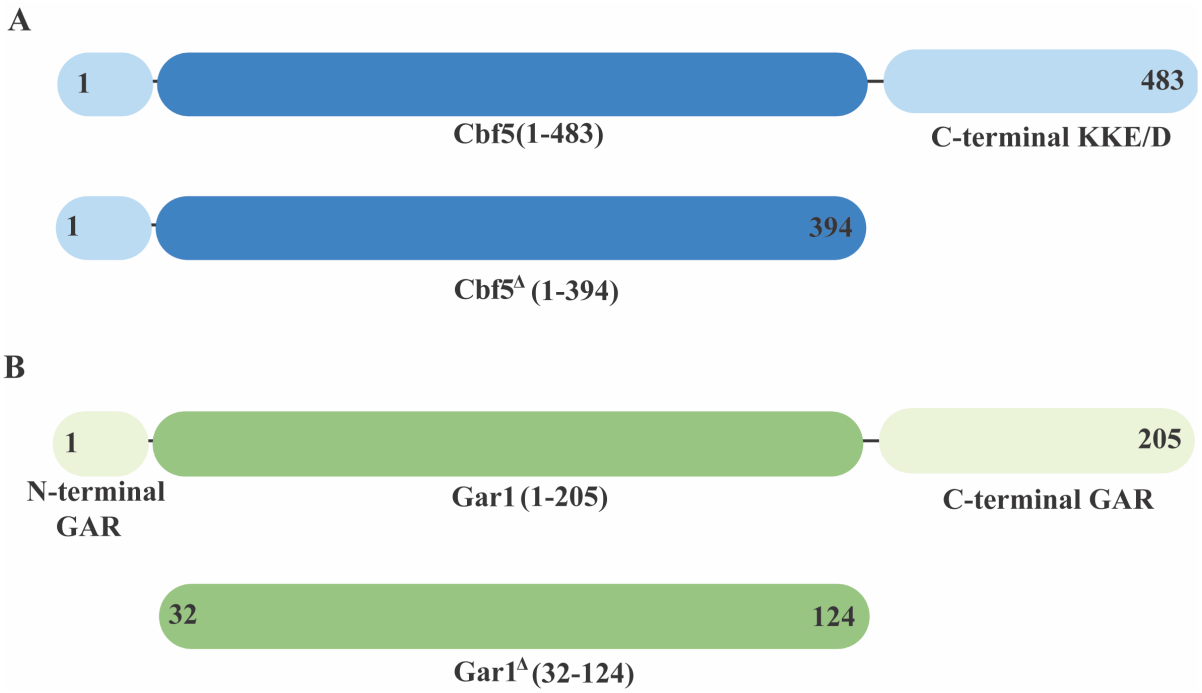

Figure 4.1. Schematic representation of Cbf5 and Gar1 truncations. A. Cbf5 core domain is depicted in dark blue, and N- and C-terminal KKE/D domains are in pale blue. **B** The Gar1 core domain is shown in green, with N- and C-terminal GAR domains in pale green.

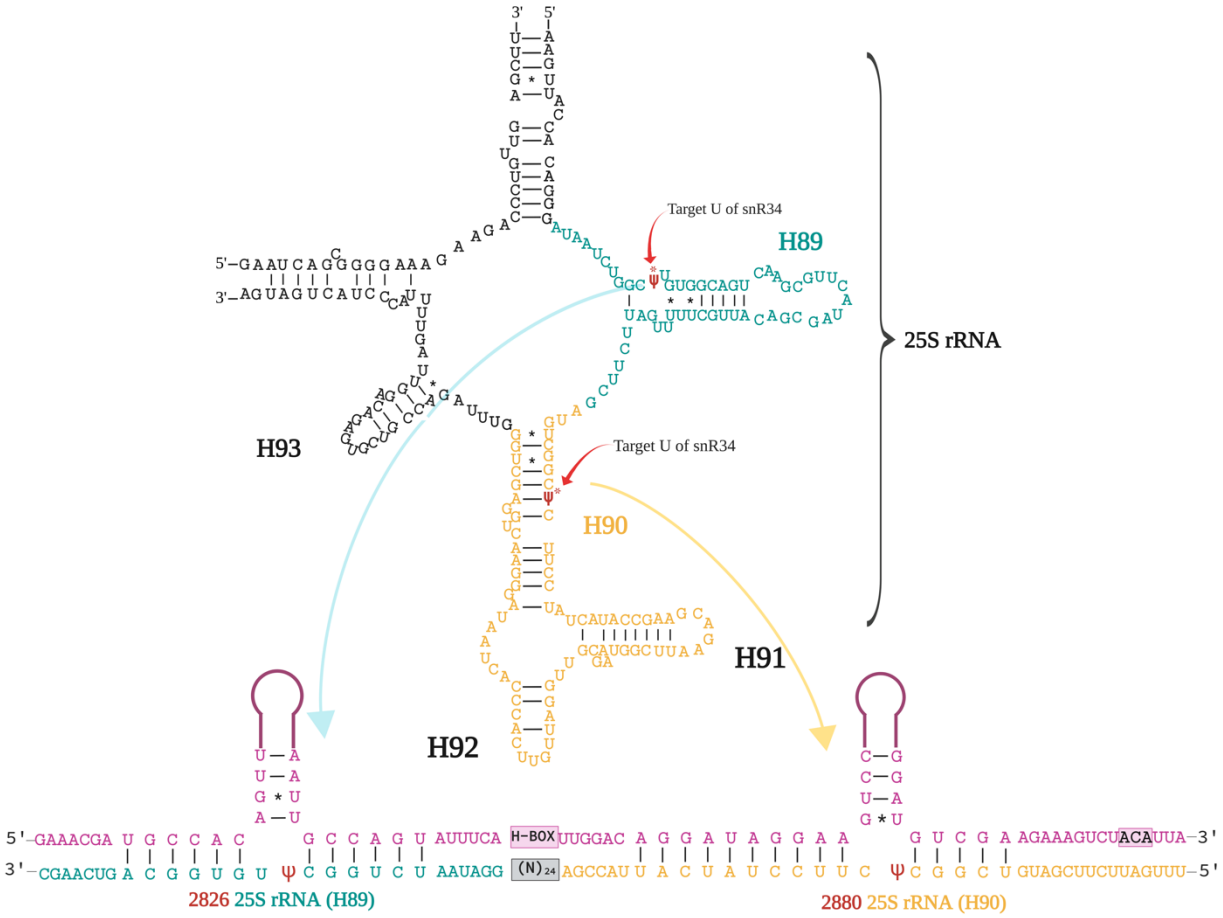

Figure 4.2. Representation of snR34 and 25S rRNA binding. The PTC domain of 25S rRNA harbours helices H89 and H90-92, which are pseudouridylated by snR34 H/ACA snoRNA at positions U2826 and U2880, respectively. The cyan colour represents H89 and yellow stands for H90-92. The pink colour represents snR34 H/ACA snoRNA.

4.2. Materials and Methods

4.2.1. Materials

[^{3}H-C5]-UTP was purchased from Moravek Biochemicals, chromatography resin was obtained from GE-Healthcare, oligonucleotides were purchased from Integrated DNA Technologies (IDT), scintillation cocktail was purchased from [EcoLite (+), MP Biomedical], and all other enzymes and chemicals were procured from Biobasic, Sigma, Fisher Scientific and NEB.

4.2.2. Deletion Mutagenesis

Bacterial (*E. coli*) codon-optimized *CBF5* and *NOP10* genes from *S. cerevisiae* were purchased from Azenta (Genewiz), and the *GAR1* and *NHP2* genes were cloned into expression vectors, as reported previously [11]. In brief, the C-terminal truncated Cbf5$^{\Delta}$ (1–394) was cloned by performing deletion PCR of pETDuet-1 Cbf5•Nop10 plasmid using the following forward and reverse primers /5Phos/CACCACCATCATCATCACTAAGAATTCGAGCTCGGCG and CTCGGCGTTGTCCAGCGGAACGTACTCCTTCTTCC, respectively. The N- and C- terminal GAR domain truncations harbouring Gar1$^{\Delta}$ (32-124) were synthesized by performing deletion PCR of pGEX-5x-3 Gar1 plasmid using forward primer /5Phos/CCACCAGACACTGTTCTAG AAATGGGAGC and reverse primer TCCTTGGAAGTACAGGTTTTCCAGATCCGATTTTG GAGGATGG (these primers also introduce the TEV protease site between GST and Gar1). Followed by site-directed mutagenesis to introduce a stop codon to generate C-terminal truncated Gar1 using the following forward and reverse primers CTATTGCCTATTGAAAGATTTTTACC TAAGTAACCTAAGGTGGTAGGC and GGTTTTGGTGGGCCTACCACCTTAGGTTACTT AGGTAAAAATCTTTC respectively. All clones were confirmed by Sanger sequencing by Azenta (Genewiz). Upon successful generation of truncated Cbf5 and Gar1 harbouring plasmids,

Cbf5 contains a core region 1–394 named Cbf5$^{\Delta}$ hereafter, and Gar1 with a central core domain 32–124 named Gar1$^{\Delta}$ hereafter.

4.2.3. Protein expression and purification

The Cbf5•Nop10 or Cbf5$^{\Delta}$•Nop10 were overexpressed in *E. coli* Lemo 21 DE3 cells (NEB) at 18 °C for 20 hours in the presence of 0.75 mM IPTG final concentration in LB ampicillin, kanamycin containing media. GST-Gar1/ GST-Gar1$^{\Delta}$ was over-expressed independently in Bl21 DE3 in ampicillin-containing LB media supplemented with 1 mM IPTG at 30 °C for 6-7 hours. For the purification of Cbf5•Nop10•GST-Gar1 or Cbf5$^{\Delta}$•Nop10•GST-Gar1$^{\Delta}$ or Cbf5$^{\Delta}$•Nop10•GSTGar1; the GST-Gar1/GST-Gar1$^{\Delta}$ expressing cells were mixed with Cbf5•Nop10 or Cbf5$^{\Delta}$•Nop10 overexpressing cells as previously described in Caton *et al*. [55]. Nhp2 was independently overexpressed in *E. coli* BL21 DE3 pLysS (NEB) at 18°C in the presence of 1 mM IPTG final concentration as described previously [55].

Moreover, the Cbf5•Nop10•GST-Gar1$^{\Delta}$ were co-expressed using pRSF-Duet-Cbf5-Nop10 and pGEX-5x-3-Gar1$^{\Delta}$ in Lemo 21 DE3 cells at 18 °C for 20 hours in LB media supplemented with ampicillin, kanamycin and 0.75 mM IPTG. All proteins were purified using buffers described previously from the Kothe lab [26, 55] **(Figure A1 C-I)**

4.2.4. *In vitro* transcription and purification of guide and substrate RNAs

The genes encoding H/ACA snoRNAs (snR34, snR81) were amplified by polymerase chain reaction (PCR), including a T7 promoter. The H/ACA snoRNAs were *in vitro* transcribed, purified, and concentrations were determined as described before [9, 26]. Substrate RNAs were generated using annealed oligonucleotides as a template using [C5–^{3}H]-UTP and purified as described previously [55, 72]. The final RNA pellet was resuspended in nuclease-free water. The specific activity of each RNA molecule was calculated as described before [72].

4.2.5. Reconstitution of H/ACA RNP complex

Guide RNAs were folded by cooling from 65 °C to room temperature for 30 min in 20 mM HEPES KOH (pH 7.4), 150 mM NaCl, 0.1 mM EDTA, 1.5 mM $MgCl_2$, 10% (v/v) glycerol, 0.75 mM dithiothreitol (DTT). The folding of guide RNAs was followed by incubating guide RNA: protein (H/ACA core proteins) in a 1:2 ratio for 10 min at 30 °C to allow for the H/ACA snoRNP formation.

4.2.6. Tritium release assay

Tritium release assays were performed to analyze pseudouridine formation by the H/ACA RNP complex [9]. All substrate RNAs used in this study are heated at 65°C for 5 minutes and snap-cooled on ice for 5 minutes in reaction buffer (20 mM HEPES KOH (pH 7.4), 150 mM NaCl, 0.1 mM EDTA, 1.5 mM $MgCl_2$, 10% (v/v) glycerol, 0.75 mM dithiothreitol (DTT)). In brief, *in vitro* reconstituted H/ACA RNPs were used for multiple turnovers, as reported previously [55, 72]. The single turnover assays were performed using 100 nM snR34 H/ACA snoRNP with 50 nM of substrate RNA. The time points were collected starting from 30 sec up to 2 min, as indicated in the results section. The results were analyzed using GraphPad Prism-9 using one phase association function.

$Y=Y_0 + (Plateau-Y_0) *(1^{-exp}(-k_{app}*x))$

Y0 is the pseudouridine level when X (time) is zero. Plateau is the maximum pseudouridine level at infinite times, and k_{app} is the rate.

4.2.7. Nitrocellulose filtration assay

The affinity of guide and substrate RNA for the H/ACA proteins and the H/ACA snoRNP, respectively, were determined using a filter binding assay as described before in chapter 3 [55, 72]. The affinity of H/ACA snoRNA and substrate RNA with H/ACA proteins were

determined using, tritium-labelled H/ACA snoRNA (0.4 nM final concentration) heated at 65°C for 5 min and slow-cooled on the bench for 30 min in reaction buffer A (20 mM HEPES KOH (pH 7.4), 150 mM NaCl, 0.1 mM EDTA, 1.5 mM $MgCl_2$, 10% (v/v) glycerol, 0.75 mM dithiothreitol (DTT)) and substrate RNA heated at 65 °C for 5 min and slow-cooled at room temperature for 30 min in reaction buffer A, were incubated with increasing concentrations of the H/ACA proteins (0-15 nM) or GST-Gar1/ GST-Gar1$^{\Delta}$ (0-10 nM) for 3 min at 30 °C. Following nitrocellulose filtration, the membrane was washed with 1 mL cold reaction buffer containing 20 mM HEPES KOH (pH 7.4), 150 mM NaCl, 0.1 mM EDTA, 1.5 mM $MgCl_2$, 10% (v/v) glycerol, 0.75 mM dithiothreitol (DTT). To analyze substrate RNA binding (substrate RNAs heated at 65 °C for 5 min and slow cooled at room temperature for 30 minutes in reaction buffer A,), the tritium-labelled substrate RNAs were incubated in increasing concentrations (0-500 nM) for 3 min at 30 °C in the presence of 5 nM *in vitro* reconstituted H/ACA snoRNP. Following nitrocellulose filtration, the membrane was washed with 1 mL cold reaction buffer as described above. Upon washing, the nitrocellulose membrane was dispensed in a 10 mL EcoLite scintillation cocktail overnight, followed by scintillation counting. Dissociation constants (K_D) were determined using the hyperbolic function in GraphPad Prism-9.0.

$Y = B_{max} \times [A] / (K_D + [A])$,

[A] is the concentration of H/ACA core proteins when measuring binding to snR34, whereas [A] is the concentration of substrate RNA when the binding of substrate RNA to the reconstituted H/ACA snoRNP is analyzed. B_{max} represents the maximum binding.

4.3. Results

4.3.1 Analyzing the role of KKE/D and GAR extensions in efficient pseudouridine formation

Previously, it was reported that the KKE/D domains of Cbf5 interact with the microtubules and deletion of the C-terminal KKE/D domain of Cbf5 was shown to delay the cell cycle in yeast [181]. Gar1 is an essential snoRNP protein required for the pre-rRNA processing [183], and it was shown that the absence of GAR domains in Gar1 protein leads to a reduced growth rate at 30°C and 19°C [184]. Further, it was reported that the deletion of GAR domains from Gar1 protein resulted in the delay of both 18S and 25S rRNA synthesis [184]. Suggesting that GAR domains might play a significant role in the ribosome biogenesis [184].To further understand the importance of the C-terminal KKE/D domain of Cbf5 and both N- and C-terminal GAR domains of Gar1 in pseudouridine formation, I performed multi-turnover tritium release experiments using 500 nM of [C5^{-3}H] UTP labelled structured fragments of 25S rRNA H89, H90-92 and 50 nM snR34 H/ACA RNPs reconstituted from purified *S. cerevisiae* proteins and *in vitro* transcribed snR34 guide RNA as reported previously [55, 72]. This experiment demonstrates that the formation of pseudouridine on both H89 and H90-92 by Cbf5$^{\Delta}$•Nop10•GST-Gar1$^{\Delta}$•Nhp2 (C$^{\Delta}$NG$^{\Delta}$P) harbouring snR34 H/ACA snoRNP is significantly decreased when compared to pseudouridine formation by snR34 H/ACA snoRNP complex having Cbf5•Nop10•GST-Gar1•Nhp2 (CNGP) **(Figure 4.3)**.

To further analyze the phenomenon of reduced pseudouridine formation while using snR34 H/ACA snoRNP harbouring C$^{\Delta}$NG$^{\Delta}$P, I hypothesize that the reduced pseudouridine formation might be due to the truncation of the C-terminal KKE/D domain of Cbf5, as Cbf5 is the main pseudouridine synthase in H/ACA snoRNP complex, and mutations at the C-terminal region of Cbf5 were previously linked to the progression of X-linked Dyskeratosis congenita [182]. To

answer this question, I performed a multi-turnover tritium release assay using snR34 H/ACA snoRNP complex harbouring Cbf5$^{\Delta}$•Nop10•GST-Gar1•Nhp2 (C$^{\Delta}$NGP), where the Cbf5 is truncated but not Gar1, and H89 and H90-92 25S rRNA fragments were used as substrate RNAs. Interestingly, this experiment uncovers that the snR34 H/ACA snoRNP complex harbouring C$^{\Delta}$NGP is capable of pseudouridine formation similar to the snR34 H/ACA snoRNP complex with Cbf5•Nop10•GST-Gar1•Nhp2 (CNGP) **(Figure 4.3)**. These results suggest that Cbf5 unstructured KKE/D tails may not have an impact on pseudouridine formation.

To further confirm these results, the Cbf5•Nop10•GST-Gar1$^{\Delta}$ (CNG$^{\Delta}$) complex was purified, and similar multi-turnover tritium release experiments were performed using the snR34 H/ACA snoRNP complex harbouring Cbf5•Nop10•GST-Gar1$^{\Delta}$•Nhp2 (CNG$^{\Delta}$P) as core proteins. This experiment results in reduced pseudouridine formation in both substrate RNAs (H89 and H90-92) when compared with pseudouridine formation by snR34 H/ACA snoRNP harbouring CNGP **(Figure 4.3)**. Therefore, this biochemical data indicate that GAR domains might have a role in pseudouridine formation. Even though Gar1 is not an RNA-dependent pseudouridine synthase, like Cbf5, Gar1 was previously reported to have a role in the product release [26].

To analyze whether the GAR domains are specifically involved in product release, single turnover tritium release experiments were performed using 0.1 μM (100 nM) of snR34 H/ACA snoRNP complex and 0.05 μM (50 nM) of the H89 and H90-92 substrate RNAs, and pseudouridine formation was monitored up to 2 min. Unexpectedly, the formation of pseudouridine in both H89 and H90-92 is reduced with snR34 H/ACA snoRNP complex harbouring C$^{\Delta}$NG$^{\Delta}$P/CNG$^{\Delta}$P as core proteins; however, efficient formation of pseudouridine was reported while using CNGP/C$^{\Delta}$NGP proteins having snR34 H/ACA snoRNP in the reaction **(Figure 4.4)**. These results indicate that N-

and C-terminal GAR regions of Gar1 are important for a step prior to product release as the GAR domains are required for efficient pseudouridine formation in multi- and single-turnover conditions **(Figures 4.3 & 4.4)**.

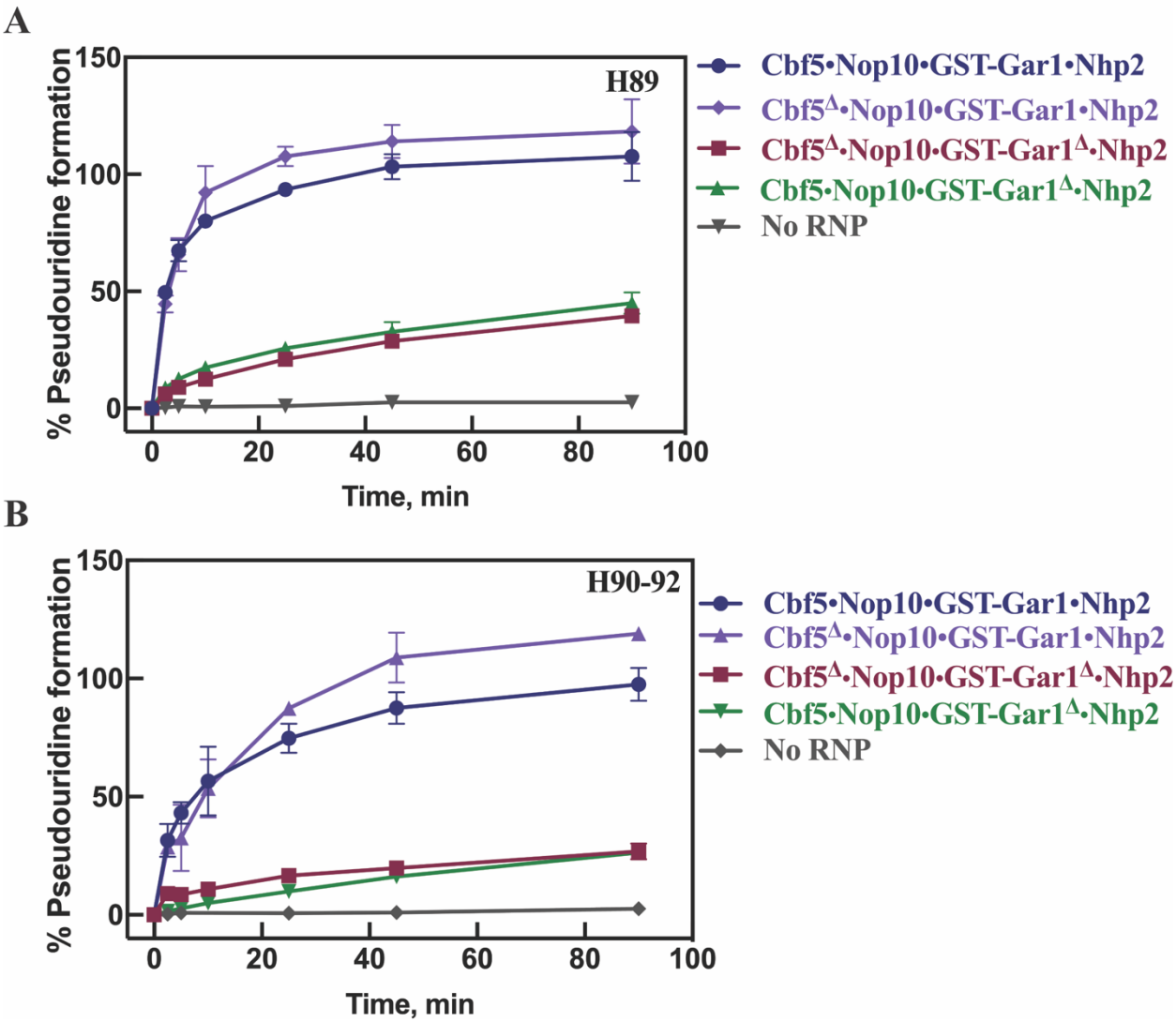

Figure 4.3. Determination of pseudouridine formation in multi-turnover conditions. Tritium release experiments were performed to measure pseudouridine formation in 25S rRNA fragments H89 and H90-92 in multi-turnover reaction conditions as mentioned in materials and methods. **A.** Pseudouridine formation is measured for H89 as substrate using CNGP, C$^{\Delta}$NGP, C$^{\Delta}$NG$^{\Delta}$P and CNG$^{\Delta}$P. **B** Pseudouridine formation is recorded in H90-92 in the presence of snR34 H/ACA snoRNP complex harbouring either CNGP, C$^{\Delta}$NGP, C$^{\Delta}$NG$^{\Delta}$P or CNG$^{\Delta}$P, respectively.

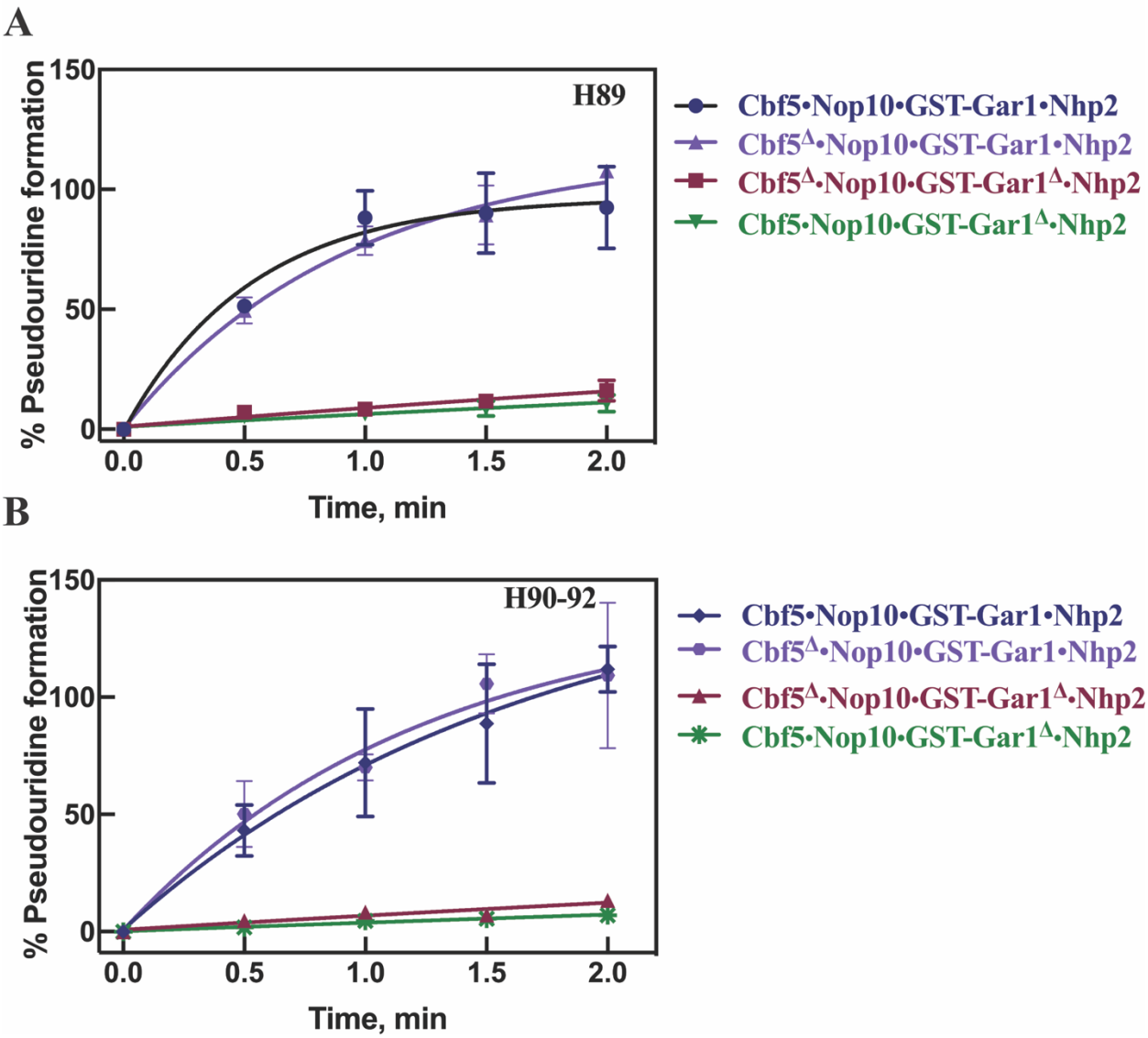

Figure 4.4. Analysis of pseudouridine formation in single turnover conditions. Tritium release experiments are used to analyze the formation of pseudouridine in a single-turnover reaction, where an excess of H/ACA snoRNP (100 nM) is incubated with low amounts of H89 or H90-92 (50 nM). **A.** Pseudouridine formation is monitored for H89 using CNGP, C$^{\Delta}$NGP, C$^{\Delta}$NG$^{\Delta}$P and CNG$^{\Delta}$P proteins harbouring snR34 H/ACA snoRNP. **B.** Formation of pseudouridine in H90-92 was analyzed using snR34 H/ACA snoRNP containing CNGP, C$^{\Delta}$NGP, C$^{\Delta}$NG$^{\Delta}$P and CNG$^{\Delta}$P proteins for up to 2 min. The data is analyzed by using a one-phase association function in GraphPad Prism-9.0 (see Materials and Methods)

4.3.2. The Cbf5$^{\Delta}$/Gar1$^{\Delta}$ proteins bind with H/ACA snoRNA with high affinity

The importance of KKE/D and GAR domains for the interaction of H/ACA proteins with H/ACA snoRNA was analyzed by performing a nitrocellulose filter-binding assay using truncated Cbf5 (Cbf5$^{\Delta}$) and Gar1 (Gar1$^{\Delta}$) proteins. In brief, increasing concentrations of the H/ACA proteins (0-15 nM) are incubated with 0.4 nM of guide RNA snR34/snR81. As mentioned in the Materials and Methods section, the H/ACA snoRNA bound to proteins was quantified using scintillation counting. The dissociation constant (K_D) for the interaction of snR34 with C$^{\Delta}$NG$^{\Delta}$P is recorded as 0.15 ± 0.055 nM similar to previously reported snR34 binding affinity with CNGP which is 0.5 ± 0.2 nM [55].

Next, the binding of CNG$^{\Delta}$P and C$^{\Delta}$NGP proteins with snR34 was analyzed, and the observed K_D values are reported as 0.23 ± 0.085 nM and 0.38 ± 0.054 nM, respectively. These results suggest that both KKE/D and GAR domains have no significant role in binding H/ACA snoRNA with its core proteins Cbf5 and Gar1. To further confirm these results, the binding of snR81 with C$^{\Delta}$NG$^{\Delta}$P was analyzed to assess the importance of the unstructured regions of Cbf5 and Gar1 in binding with different H/ACA snoRNAs. Our analysis suggests that the affinity of snR81 to C$^{\Delta}$NG$^{\Delta}$P (2.2 ± 0.7 nM) is similar to the association of snR81 to the CNGP core proteins (3.6 ± 0.7 nM). These results further confirm that the truncations of the C-terminal KKE/D domain of Cbf5 or the N- and C-terminal GAR domains of Gar1 do not affect the binding of H/ACA snoRNA to the H/ACA core proteins **(Figure 4.5, Table 4.1)**.

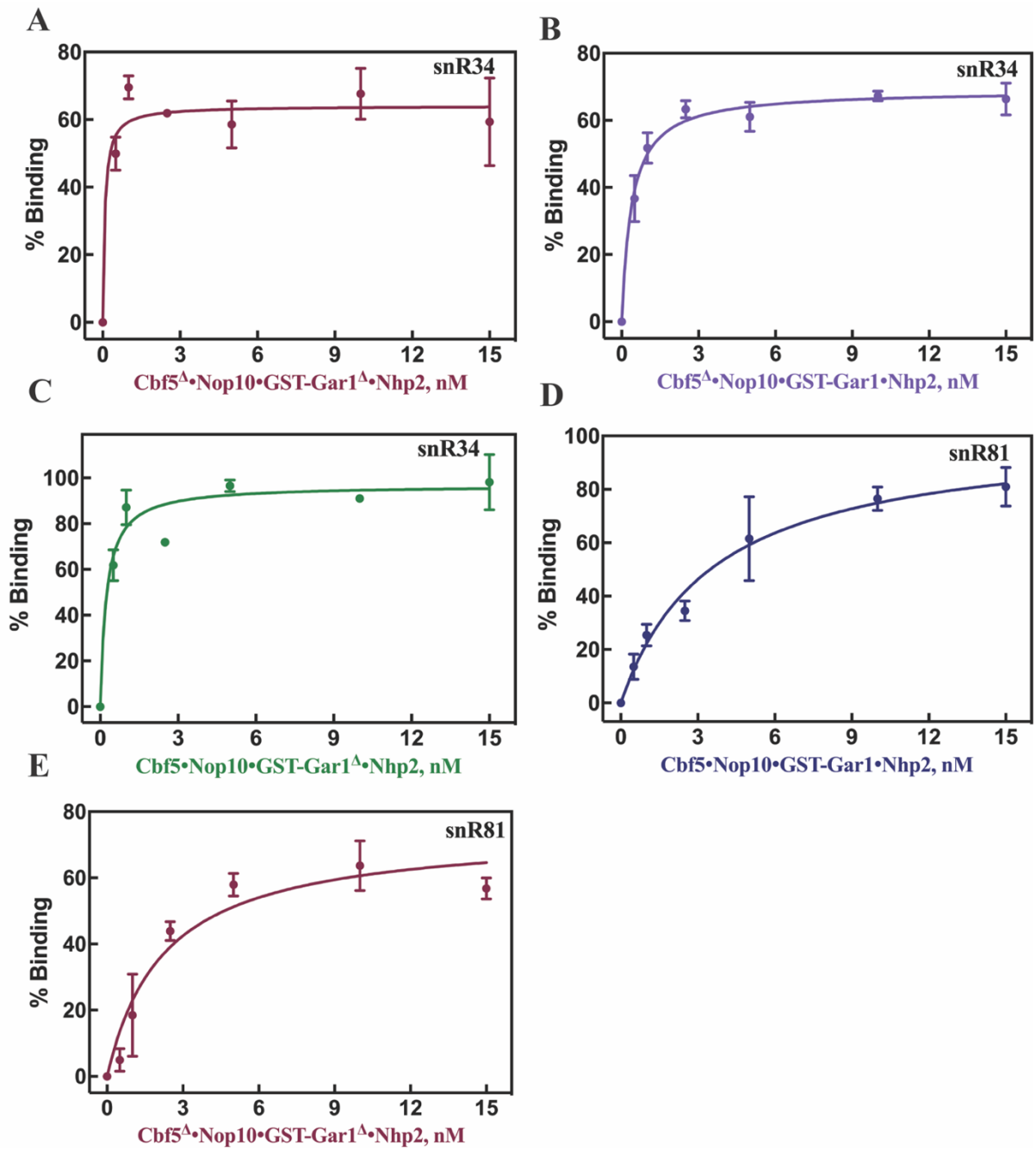

Figure 4.5. Binding of H/ACA snoRNA to CNGP, CNG$^{\Delta}$P, C$^{\Delta}$NGP and C$^{\Delta}$NG$^{\Delta}$P core proteins. The interaction of H/ACA snoRNA with CNGP/ C$^{\Delta}$NG$^{\Delta}$P was detected by nitrocellulose filtration. **A**, **B**, and **C** represent snR34 binding with C$^{\Delta}$NG$^{\Delta}$P, C$^{\Delta}$NGP and CNG$^{\Delta}$P, respectively. **D and E** represent the binding of snR81 H/ACA snoRNA to CNGP and C$^{\Delta}$NG$^{\Delta}$P, respectively. The dissociation constants (K_D) were determined by fitting them with a hyperbolic equation (see Materials and Methods).

Table 4.1. Binding of H/ACA snoRNA with H/ACA proteins harbouring KKE/D or GAR truncations. The corresponding binding curves for determining the dissociation constants for each combination of core protein truncations are shown in **Figure 4.5**

Guide RNA	Proteins	K_D (nM)
snR81	CNGP	2.2 ± 0.7
	$C^{\Delta}NG^{\Delta}P$	3.6 ± 0.7
snR34	$C^{\Delta}NG^{\Delta}P$	0.15 ± 0.055
	$C^{\Delta}NGP$	0.38 ± 0.054
	$CNG^{\Delta}P$	0.23 ± 0.085

4.3.3. The snR34 H/ACA snoRNP complex harbouring Cbf5$^{\Delta}$ and Gar1$^{\Delta}$ has a high affinity for substrate RNAs

Previously, I showed that the truncation of KKE/D and GAR domains does not affect the binding of the H/ACA snoRNA to the H/ACA proteins. However, reduced pseudouridine formation was observed in both multi- and single-turnover experiments while using H/ACA snoRNP harbouring either $C^{\Delta}NG^{\Delta}P$ or $CNG^{\Delta}P$ **(Figures 4.3 & 4.4)**. Based on these results, it has been hypothesized that the truncation of KKE/D or GAR domains might affect substrate RNA binding to the H/ACA snoRNP complex. Therefore, nitrocellulose filter binding experiments were performed to analyze the interaction of substrate rRNAs H89 and H90-92 with H/ACA snoRNP harbouring $C^{\Delta}NG^{\Delta}P$, $CNG^{\Delta}P$ and $C^{\Delta}NGP$, respectively. These experiments will show how the truncations of Cbf5 and Gar1 influence the binding of substrate RNA to the H/ACA snoRNP complex. Towards this goal, the substrate RNA binding affinity was measured by using 5 nM reconstituted H/ACA snoRNP and increasing concentrations of radiolabeled substrate RNA H89 or H90-92 up to 500 nM. The affinities of H89 to snR34 H/ACA snoRNP complex harbouring CNGP, $C^{\Delta}NG^{\Delta}P$, $CNG^{\Delta}P$ and $C^{\Delta}NGP$ are 29.0 ± 2.5 nM, 23.0 ± 6.1 nM, 23.0 ± 7.5 nM and 14.0 ± 5.0 nM, respectively, suggesting that the binding of 5' substrate H89 with snR34 H/ACA snoRNP

complex harbouring $C^{\Delta}NG^{\Delta}P$, $CNG^{\Delta}P$ and $C^{\Delta}NGP$ is not hindered, and it is similar to the binding of H89 with snR34 snoRNP complex having CNGP.

Further, the binding of H90-92 was analyzed with snR34 H/ACA snoRNP complex having CNGP, $C^{\Delta}NG^{\Delta}P$, $CNG^{\Delta}P$ and $C^{\Delta}NGP$ which resulted in the K_D values 18.0 ± 3.9 nM, 43.0 ± 8.8 nM, 60.0 ± 22.0 nM, and 27.0 ± 10.0 nM, respectively **(Figure 4.6, Table 4.2)**. These results suggest that the binding of 3' substrate H90-92 is affected by the truncation of GAR domains as the K_D is approximately two-fold increased for complexes with GST-$Gar1^{\Delta}$.

To further confirm the role of GAR domains in substrate RNA binding in the absence of the other H/ACA snoRNP components, a nitrocellulose filter binding assay was performed with both increasing concentrations of GST-Gar1 and GST-$Gar1^{\Delta}$ (0-10 nM) and 0.4 nM H89, H90-92, H39 and bacterial $tRNA^{Phe}$ respectively, which will allow us to analyze the RNA binding nature of GAR domains **(Figure 4.7)**. These experiments confirm that the full-length GST-Gar1 protein bearing GAR domains can bind RNA in low nanomolar affinity. However, GST-$Gar1^{\Delta}$, which lacks GAR domains, fails to interact with RNA **(Table 4.3).**

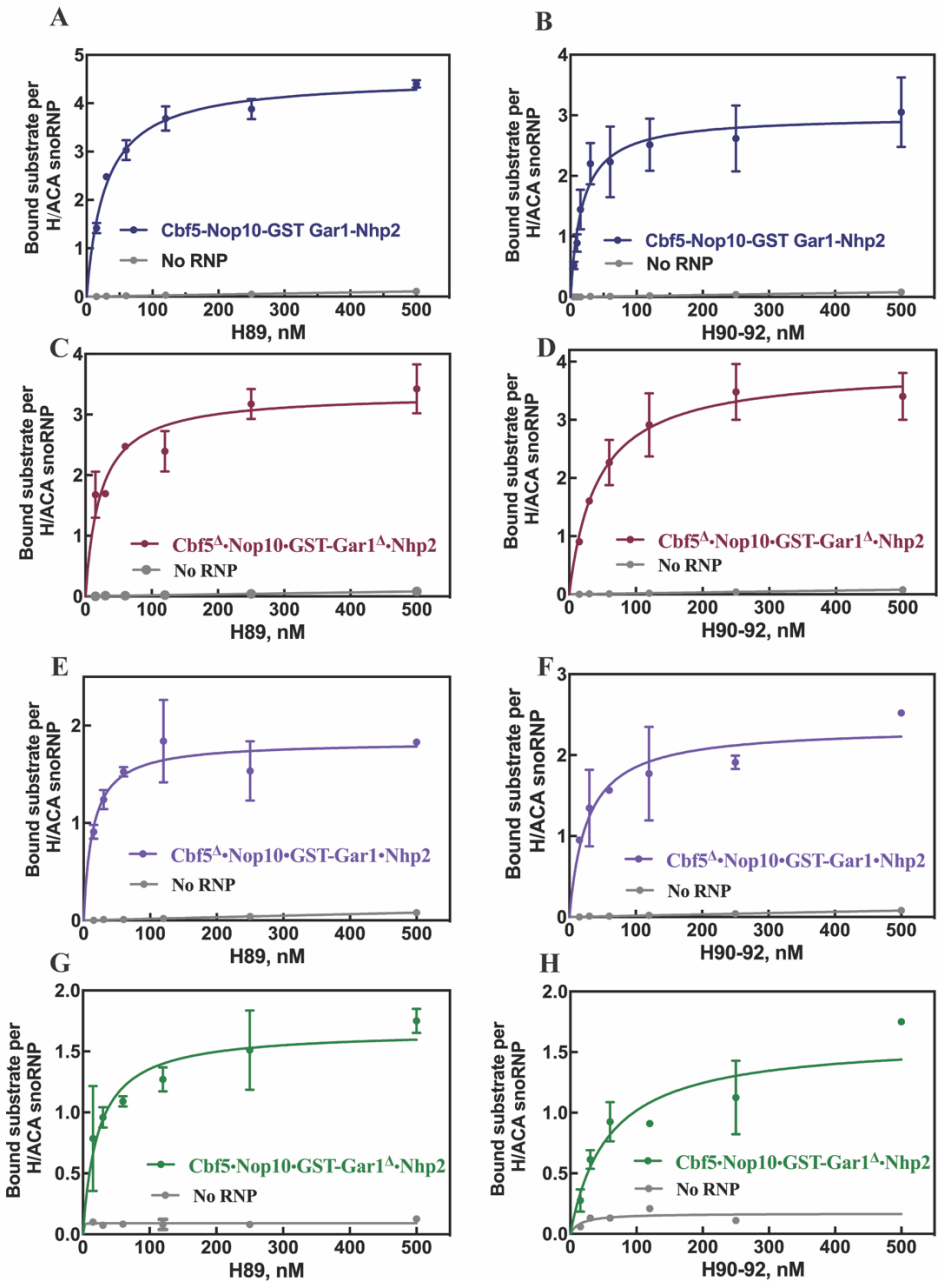

Figure 4.6. Determination of binding affinities of H89 and H90-92. Nitrocellulose filtration assay was carried out to determine the affinity of substrate RNA binding, where increasing concentrations of radiolabeled H89 or H90-92 substrate RNAs were incubated with a specific concentration of snR34 H/ACA RNP bearing truncated or full-length core proteins **A** and **B** represent the binding of H89, H90-92 substrate RNA to the snR34 H/ACA RNP harbouring CNGP **C** and **D** indicates the H89 and H90-92 substrate RNAs binding to the snR34 H/ACA RNP harbouring C^{Δ}NGP. **E** and **D** show the binding data of H89 and H90-92 to the snR34 H/ACA RNP harbouring C^{Δ}NGP. Further, **G** and **H** represent snR34 H/ACA snoRNP harbouring CNG$^{\Delta}$P binding with H89 and H90-92, respectively. The dissociation constants (K_D) were determined by using a hyperbolic equation (see Materials and Methods).

Table 4.2. The affinity of H/ACA snoRNP complex harbouring truncated Cbf5 or truncated Gar1 to substrate RNA. The corresponding binding curves for determining the dissociation constants for each substrate RNA with snR34 H/ACA snoRNP complex with truncated proteins are displayed in **Figure 4.6**

Proteins	Substrate RNA	K_D (nM)
CNGP	H89	29.0 ± 2.5
$C^{\Delta}NG^{\Delta}P$		23.0 ± 6.1
$C^{\Delta}NGP$		14.0 ± 4.8
$CNG^{\Delta}P$		23.0 ± 7.5
CNGP	H90-92	18.0 ± 3.9
$C^{\Delta}NG^{\Delta}P$		43.0 ± 8.8
$C^{\Delta}NGP$		27.0 ± 10.0
$CNG^{\Delta}P$		60.0 ± 22.0

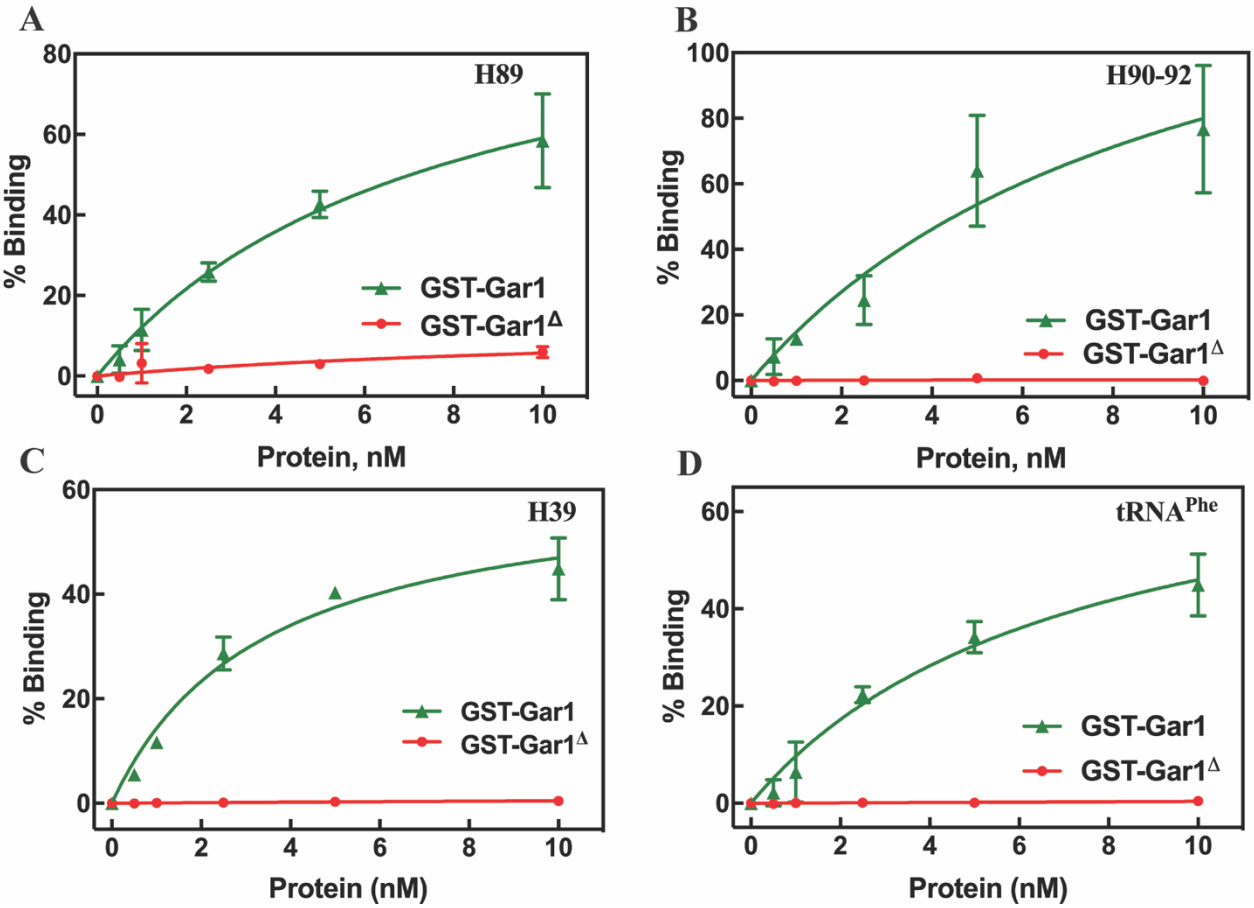

Figure 4.7. Analysis of RNA binding to GST-Gar1 and GST-Gar1$^{\Delta}$. Nitrocellulose filtration assay was performed to determine the affinity of GST-Gar1 and GST-Gar1$^{\Delta}$ for RNA. Increasing concentrations of Gar1/Gar1$^{\Delta}$ were incubated with either H89/H90-92/H39/tRNA^{Phe} for 3 min at 30 °C **A, B, C** and **D** represents H89, H90-92, H39 and tRNA^{Phe} binding to Gar1(green) and Gar1$^{\Delta}$ (red) as indicated. The dissociation constants (K_D) were determined by using a hyperbolic function in GraphPad Prism 9.0 (see Materials and Methods).

Table 4.3. The affinity of GST-Gar1 for RNA. The corresponding data for determining the dissociation constants for H89, H90-92, H39, and tRNA^{Phe} are shown in **Figure 4.7**

RNA	K_D (nM)
H89	8 ± 2.4
H90-92	10 ± 5.8
H39	3.4 ± 0.7
tRNA^{Phe}	7 ± 2.4

4.4. Discussion

This study analyzes the role of unstructured regions of Cbf5 and Gar1 in pseudouridine formation. Towards this, the importance of the C-terminal lysin-rich KKE/D domain of Cbf5 and the N- and C-terminal GAR domains of Gar1 in pseudouridine formation is elucidated by monitoring the isomerization of U2826 in H89 and U2880 in H90-92 25S rRNA using snR34 as H/ACA guide RNA. Our results suggest that H/ACA snoRNP complex harbouring the truncated core proteins Gar1$^{\Delta}$ show reduced pseudouridine formation in H89 and H90-92 compared with snR34 H/ACA snoRNP complex carrying full-length Gar1. This indicates that the unstructured domains of Gar1 might have a role in proper pseudouridylation. In contrast, further analysis of reduced pseudouridine formation concludes that the C-terminal KKE/D domain of Cbf5 has no role in efficient pseudouridine formation **(Figure 4.3 & 4.4)**. These results support the previously reported data, where the truncation of the C-terminal KKE/D domain of Cbf5 does not affect pseudouridine formation *in vitro* when using the snR5 H/ACA snoRNP complex [26].

4.4.1. Truncation of the KKE/D or GAR domain does not affect the interaction of H/ACA snoRNA with core proteins

Further investigation on the role of both KKE/D domains and GAR domains in stable H/ACA snoRNP complex formation with H/ACA snoRNA using filter binding assays reveals that the CNG$^{\Delta}$P/ C$^{\Delta}$NG$^{\Delta}$P/ C$^{\Delta}$NGP bind to snR34 with nM affinity similar to CNGP **(Figure 4.5, Table 4.1**). These results are further confirmed by analyzing the binding of snR81 to C$^{\Delta}$NG$^{\Delta}$P and CNGP, which reveals that the truncated proteins C$^{\Delta}$NG$^{\Delta}$P can bind with all H/ACA snoRNA with nM affinity similar to the full-length core proteins CNGP. Further, to better understand the role of KKE/D and GAR truncations in the binding of substrate RNA with the snR34 H/ACA snoRNP complex, the binding affinities of the H89 and H90-92 to snR34 H/ACA snoRNP complex are

analyzed. The measured K_D values **(Figure 4.6, Table 4.2)** reveal that both H89 and H90-92 bind to H/ACA snoRNP complex with similar affinity as wild-type when the KKE/D domain of Cbf5 is deleted. However, the affinity of H90-92 to H/ACA snoRNPs harboring GAR truncations is two-fold reduced when compared to binding affinities with snR34 H/ACA snoRNP complex harbouring CNGP. This suggests that GAR domains might play a role in binding substrate RNA to the H/ACA snoRNP complex.

4.4.2. GAR domains are important for efficient pseudouridine formation

The Gar1 protein in eukaryotes has glycine and arginine-rich N- and C- terminal GAR domains; however, these domains are absent in archaea **(Figure A7)**. This study reports that these GAR domains are important for pseudouridine formation in multi- and single-turnover conditions. Pseudouridine formation is compared between snR34 H/ACA snoRNP complex harbouring CNGP, $CNG^{\Delta}P$, $C^{\Delta}NGP$ and $C^{\Delta}NG^{\Delta}P$ using H89 and H90-92 25S rRNA fragments as substrate RNAs **(Figure 4.3 & 4.4)**. My results indicate that the KKE/D domain of Cbf5 is not essential for the isomerization of uridine. Furthermore, my results confirm that the GAR domains of Gar1 are important for efficient pseudouridine formation. This observation might be possible only if GAR domains are involved in positioning the target site accessible to the Cbf5 catalytic domain, which might help efficient pseudouridine formation by Cbf5.

4.4.3. GAR domains might be involved in positioning the substrate RNA near the pseudouridine pocket of H/ACA snoRNA or the catalytic site of Cbf5

Previously, it was reported that the Gar1 protein is required in the archaea for positioning the substrate RNA accessible to the Cbf5 catalytic site [36]. This phenomenon might also be true for eukaryotic Gar1 protein, especially the GAR domains might be involved in this process in eukaryotes. Previous reports suggest that Gar1 is required for substrate turnover in eukaryotes [26].

Our experiments reveal that the GAR domains of Gar1 might not have a role in product release **(Figure 4.4)**; instead, they might be essential for placing the substrate RNA close to the catalytic site of Cbf5. **(Figure 4.8)**. Moreover, recent studies using snR81 as snoRNA along with Gar1$^{\Delta}$ reveal that pseudouridylation is affected under single turnover conditions [166]. This further explains that GAR domains might have no role in product release but might be involved in enhancing the interactions between the Cbf5 catalytic site and substrate RNA.

My filter binding analysis of RNA with Gar1/Gar1$^{\Delta}$ shows that the full-length Gar1 protein with GAR domains can bind any RNA with high affinity **(Table 4.1)**. However, truncated Gar1 (Gar1$^{\Delta}$) has lost the ability to bind RNA **(Figure 4.5)**. This observation might explain snR34 H/ACA snoRNP complex with Gar1$^{\Delta}$ is less efficient in pseudouridine formation compared with full-length snR34 H/ACA snoRNP complex harbouring Gar1 protein. The RNA-binding function of the GAR domains might help the Cbf5 catalytic site to quickly access the target uridine and facilitate the efficient formation of pseudouridine. This might be one of the potential reasons that the Naf1 protein is replaced by Gar1 in the final stages of H/ACA snoRNP complex assembly, as the significant difference between Naf1 and Gar1 is the absence and presence of GAR domains respectively. Overall, this study indicates that the GAR domains might have a vital role in the function of H/ACA snoRNP complex in efficient pseudouridine formation [69]. Moreover, our Gar1$^{\Delta}$ protein has a GST tag, and deletion of N- and C-terminal GAR domains might bring the GST and Gar1$^{\Delta}$ in close vicinity; this could potentially result in steric hindrance for the H/ACA snoRNP which might be a potential reason why we observed reduced pseudouridine formation **(Figure 4.3** and **4.4)**. To better understand this, we have to perform similar experiments with Gar1$^{\Delta}$ after removing the GST tag by proteolytic cleavage. Further characterization of GAR domains can be done by studying the interaction of RNA with GST protein containing N- and C-terminal GAR

domains. This experiment will reveal whether the GAR domains are not only required, but also sufficient for RNA binding. Furthermore, GST protein alone without GAR domains will be the control for this experiment to confirm that the GAR domains solely contributing to RNA binding function of Gar1.

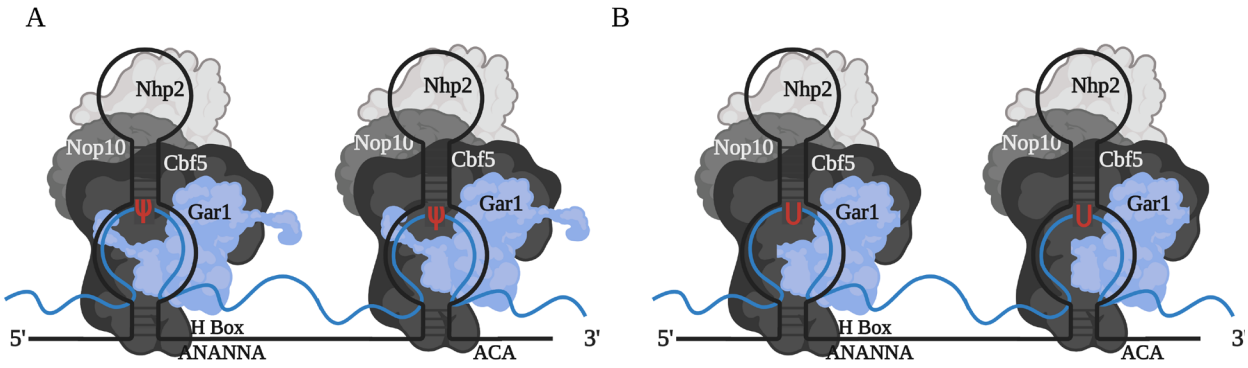

Figure 4.8. Schematic representation of disorientation of Cbf5 catalytic domain due to lack of GAR domain of Gar1. The GAR domains might be essential for placing substrate RNA accessible to the pseudouridine pockets, and the catalytic site of Cbf5 **A** represents the proper orientation of the Cbf5 catalytic site near the pseudouridylation pocket where GAR domains might stabilize. **B** represents the dislocated Cbf5 catalytic domain due to Gar1 truncation ($Gar1^{\Delta}$).

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS

The H/ACA snoRNP complex is an RNA-dependent pseudouridine synthase, mainly involved in nucleotide modifications of rRNA. Moreover, H/ACA snoRNP complexes are also involved in rRNA processing and, in turn, play a vital role in the ribosome biogenesis [7, 47]. In eukaryotes, a two-hairpin structure is conserved for almost all H/ACA snoRNAs, whereas archaea harbour one-, two-, and three-hairpin H/ACA snoRNAs [7]. Previously, two independent studies reported that a single, eukaryotic H/ACA snoRNA hairpin is an independent functional unit, albeit with less efficiency in the pseudouridine formation than two-hairpin H/ACA snoRNAs [26, 55]. This thesis has answered a long-lasting question in the H/ACA snoRNP field: why eukaryotes have a conserved two-hairpin structured H/ACA snoRNA, and what is the importance of this conserved two-hairpin structure H/ACA snoRNA for efficient rRNA modifications.

Further, I have elucidated the importance of the hinge region length in two-hairpin H/ACA snoRNAs for proper pseudouridine formation. In addition, in this dissertation, I have characterized the importance of the unstructured KKE/D and GAR domains of Cbf5 and Gar1, respectively, for pseudouridine formation. Taken together, a detailed analysis of the role of the H/ACA snoRNP complex in pseudouridine formation is provided.

As demonstrated in chapter 3, the H/ACA snoRNP complex is actively involved in pseudouridine formation in structured H89, H90-92 and H89-92 or H38, H39 and H38-39 of 25S rRNA fragments. This indicates that under *in vitro* conditions, the H/ACA snoRNP complex may unwind the local secondary structures of rRNA. Therefore, these results suggest that the H/ACA snoRNP complex might not depend on additional factors such as helicases and that it also potentially aids the proper folding of rRNA *in vivo* as the H/ACA snoRNP complex functions during the initial stages of rRNA biogenesis. Thus, the H/ACA snoRNP complex might be leading

to the functional folding of rRNA during ribosome biogenesis. It is known that RNA folds from a single-stranded primary form to secondary, tertiary, and quaternary structures and unfolds in reverse order [185]. *In vivo,* several proteins*,* including DEAD-Box helicases and ribosomal proteins, play a role in unwinding the rRNA before it is packaged into mature ribosomes [185, 186]. With this current knowledge, the rRNA unfolding function of the H/ACA snoRNP complex can be studied in detail using RNA foot-printing and Fluorescence Resonance Energy Transfer (FRET) analysis. These methods will provide insights into mechanistic details of the unfolding and folding process of rRNA by the H/ACA snoRNP complex. Foot-printing is an important tool for understanding RNA structure . In time-resolved foot-printing, we will be able to follow the changes in the structure of substrate RNA molecules by the binding to H/ACA snoRNP [187]. The foot-printing brings the advantage that it can provide single nucleotide resolution, which will give a very accurate determination of protein binding sites and how that binding affects the RNA folding time-resolved foot-printing methods [188]. The RNA or RNP is treated with a low concentration of DMS to produce methylated RNA. This methylated RNA can be deproteinized and isolated, followed by a reverse transcription reaction using a radiolabeled primer, and the positions of modification are determined by denaturing PAGE. The sites of modification and the extent of modification are then compared with the reference which is typically the RNA in the absence of H/ACA snoRNP and a non-DMS control. These controls are necessary to understand the termination of reverse transcription reaction, which can be caused by DMS methylation or by DMS-independent source in the reaction, such as truncated RNA[189]. Towards FRET, the long substrate RNA H89-92 or H38-39 can be labelled with two donor and acceptor dyes; this will allow us to monitor the folding and unfolding states of each helix of H89-92/H38-39 individually. Further, to analyze the folding state of H90-92 in long RNA H89-92, labelled at

3' ends with an acceptor dye and the middle region of H89 and H90-92 with donor dye can be used. For the analysis of H89 folding states in H89-92 long RNA, labelling at the 5' end with donor dye and the region intermediate to the H89 and H90-92 with acceptor dye can be performed using a splint ligation method [190, 191] **(Figure 5.1)**. This experiment will uncover RNA folding and unfolding states of substrate RNA when it is bound to the H/ACA snoRNP complex. A high FRET signal will be observed upon folding of substrate RNA, and low FRET in the case of the unfolded state of substrate RNA or, in other words, a low FRET signal will be observed when the rRNA is bound to the snR34 H/ACA snoRNP complex. This method will allow us to directly test the hypothesis that the H/ACA snoRNP is involved in the unfolding of 25S rRNA upon binding. Moreover, FRET allows us to monitor the folding states of the 25S rRNA upon dissociation from the snR34/snR5 H/ACA snoRNP complex. In turn, this will enable us to monitor the kinetics of unfolding and folding of rRNA in the presence of the H/ACA snoRNP complex. Overall, these experiments are crucial to further characterize the function of the H/ACA snoRNP complex during ribosome biogenesis, other than pseudouridine formation.

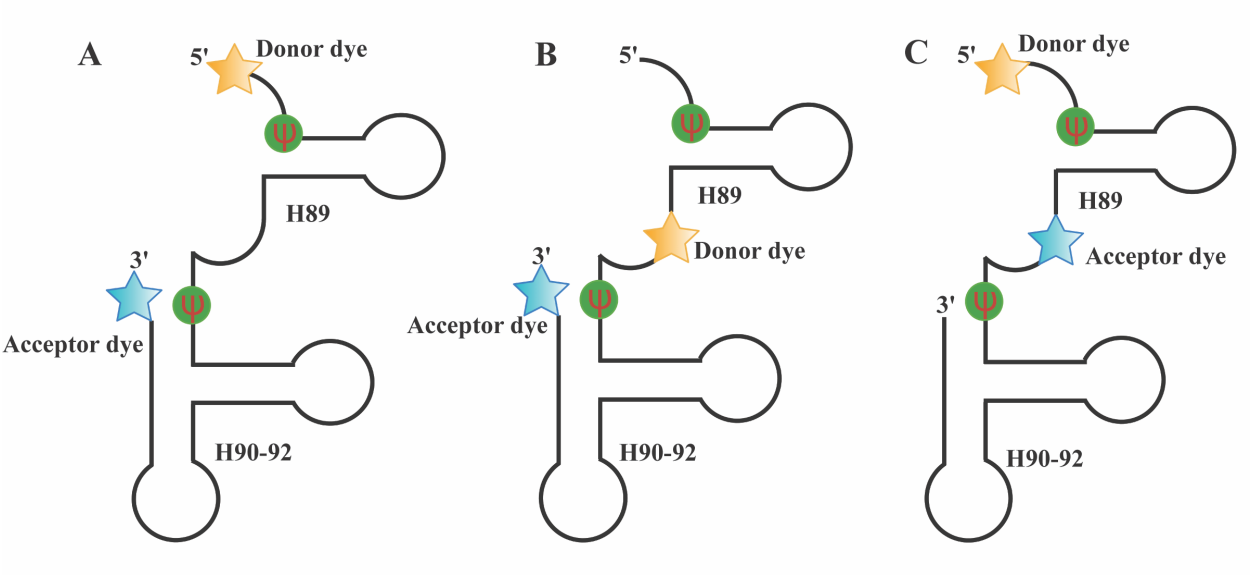

Figure 5.1. Labelling of H89-92 25S rRNA for FRET analysis. A The H89-92 long substrate RNA will be labelled with donor and acceptor dye on 5' and 3'ends, respectively. **B** The unstructured region between helices H89 and H90-92 is labelled with donor dye and acceptor dye is placed at 3'end of H89-92. This will allow us to monitor the folding and unfolding of H90-92. **C** The 5' end is marked with the donor, and the unstructured region between helices H89 and H90-92 is labelled with acceptor dye to monitor H89 folding and unfolding states.

Chapter 3 describes the importance of the hinge region length for proper pseudouridine formation. Here, the hinge region of snR34 is truncated, starting from 5 nt, 10 nt, and up to 22 nt, named Δ5H, Δ10H and Δ22H snR34, respectively. Truncation of 5 nt from the snR34 hinge regions does not alter the formation of pseudouridine; however, truncation of 10 and 22 nt alters the formation of pseudouridine drastically on the 5' hairpin. In the case of the 3' hairpin, the formation of pseudouridine is reduced but is not affected as drastically as for the 5' hairpin, indicating that the minimum length of the hinge region is essential for efficient pseudouridine formation.

As described in chapter 3, near-cognate rRNA fragments compete with the cognate substrate for the same pseudouridine pocket, and near-cognate substrates hinder efficient pseudouridine formation *in vitro*. Based on these results, we propose a model to explain the importance of the conserved two-hairpin structure of the H/ACA snoRNA in efficient

pseudouridine formation. According to this model, binding of substrate RNA at one pseudouridylation pocket allows the tethering of the substrate RNA to the H/ACA snoRNP complex, which in turn enables the binding of proper substrate RNA to the second pocket by minimizing the competitive inhibition with near cognate rRNA. Further analysis using long H89-92 U2826C and H89-92 U2880C reveals that the formation of pseudouridine is not affected by the RNA-bound state of the other hairpin, and there is no reported allosteric interaction between two hairpins according to our *in vitro* biochemical analysis using snR34/snR5. This data confirms that two hairpins are independent functional units while modifying the target uridine.

In chapter 4, various combinations of the truncated and total length of core proteins are purified, such as Cbf5•Nop10•GST-Gar1•Nhp2 (CNGP), Cbf5$^{\Delta}$•Nop10•GST-Gar1$^{\Delta}$•Nhp2 (C$^{\Delta}$NG$^{\Delta}$P), Cbf5•Nop10•GST-Gar1$^{\Delta}$•Nhp2 (CNG$^{\Delta}$P) and Cbf5$^{\Delta}$•Nop10•GST-Gar1•Nhp2 (C$^{\Delta}$NGP), to characterize the unstructured domains of Cbf5 and Gar1. This chapter discusses that the KKE/D domains of Cbf5 have no role in pseudouridine formation, as previously reported [26]. However, the GAR domains of the Gar1 protein are essential for proper pseudouridine formation in multiple turnover and single-turnover conditions while using snR34 as H/ACA snoRNP complex. Previous studies reported that Gar1 is critical for substrate turnover during pseudouridine formation with snR5 H/ACA snoRNP complex [26]; however, as per our knowledge, this is the first time it has been shown that GAR domains of Gar1 are necessary for the function of the H/ACA snoRNP complex harbouring snR34. Moreover, it is shown that Gar1 full-length protein carrying GAR domains can bind to any RNA in a non-specific manner and truncation of GAR domains abolish this RNA-binding capacity of Gar1 protein.

Purification of various H/ACA snoRNP core proteins, including Cbf5•Nop10•GST-Gar1 •Nhp2 (CNGP) or Cbf5$^{\Delta}$•Nop10•GST-Gar1$^{\Delta}$•Nhp2 (C$^{\Delta}$NG$^{\Delta}$P) or Cbf5$^{\Delta}$•Nop10•GST-Gar1•Nhp2

(C$^{\Delta}$NGP) or Cbf5•Nop10•GST-Gar1$^{\Delta}$•Nhp2 (CNG$^{\Delta}$P), are successfully achieved in this research. This will allow us to elucidate the cryo-EM structure of the H/ACA snoRNP complex in different combinations of full-length or truncated core proteins (Cbf5/Gar1). The structural analysis with CNGP or C$^{\Delta}$NGP, harbouring H/ACA snoRNP complex, will help us further understand how these GAR domains interact with the substrate RNA for efficient pseudouridine formation. Moreover, analyzing the structural features of C$^{\Delta}$NGP, harbouring H/ACA snoRNP complex, will help us understand why the deletion of KKE/D domains has no effect on pseudouridine formation.

Further, the structure of the H/ACA snoRNP complex can be elucidated in its apo and substrate-bound states, which allows us to understand substrate binding with H/ACA snoRNP complex in its mature form. This can be done using a longer substrate RNA (H89-92/H38-39), which covers both 5'-3' hairpins of the H/ACA snoRNP complex (snR34/snR5). To maintain the stable binding of substrate RNA with H/ACA snoRNA without converting it into pseudouridine, the H89-92 U2826C & U2880C can be used, or uridine can be replaced with 5-fluorouracil (5FU) at the target uridine position using splint end ligation method. The presence of 5-fluorouracil is predicted to prevent the dissociation of the product after the catalysis [192]. Moreover, the structural studies can be performed using *in vitro* reconstituted H/ACA snoRNP complex containing catalytic mutant D95N Cbf5, which has been shown to bind with H89-82 with an affinity of 7.0 ± 1.9 nM; this affinity is similar to the binding affinity of H89-92 (9.0 ± 2.6 nM) with H/ACA snoRNP complex harbouring WT Cbf5. Suggesting that the H/ACA snoRNP complex containing D95N Cbf5 will form a stable complex with H89-92 substrate RNA without being involved in pseudouridine formation.

Furthermore, to stabilize the complex by restricting the flexibility of the two halves of the snR34 H/ACA snoRNP, the snR34 variant with a shortened hinge region Δ5H snR34 can be

used, as it still displays high activity similar to snR34 WT as described in chapter 3. Upon successful elucidation of substrate-bound and apo structures of eukaryotic H/ACA snoRNP complex harbouring CNGP and C$^{\Delta}$NGP, respectively, structure-based mutagenesis can be performed to better understand the role of H/ACA snoRNPs in ribosome biogenesis and in the progression of X-DC. The cryo-EM analysis will further reveal the fine molecular architecture of the eukaryotic H/ACA snoRNP complex in the presence of substrate RNA for the first time, which will allow us to visualize the molecular interactions within the H/ACA snoRNP complex and its interaction with the rRNA. Overall, this study would be the first to explore the structural details of the eukaryotic H/ACA snoRNP complex in the substrate-bound state. More importantly, this structure-based mutagenesis and biochemical characterization of the H/ACA snoRNP complex will also uncover strategies for designing artificial H/ACA snoRNPs, which can be used to target novel sites of rRNA or mRNA to produce specialized ribosomes involved in the incorporation of rare amino acids and to proceed the stop codon read-through in mRNA [193, 194]. These novel artificial H/ACA snoRNPs can be explored as potential RNA therapeutics for different diseases like cancer [194]. Moreover, inhibition of the H/ACA snoRNP complex, like snR30, can be explored as cancer therapeutics as it impedes ribosome synthesis [127, 195]. Overall, this structural study will help us understand the mechanistic role of the H/ACA snoRNP complexes in ribosome biogenesis and, in turn, in tumour progression, as it was previously reported that various cancer cells show differential regulation of H/ACA snoRNA molecules, as discussed in chapter 1.

CHAPTER 6: REFERENCES

1. Sloan, K.E., et al., *Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function.* RNA Biol. **14**(9): p. 1138-1152.

2. Boccaletto, P., et al., *MODOMICS: a database of RNA modification pathways. 2021 update.* Nucleic Acids Res, 2022. **50**(D1): p. D231-D235.

3. Davis, F.F. and F.W. Allen, *Ribonucleic acids from yeast which contain a fifth nucleotide.* J Biol Chem, 1957. **227**(2): p. 907-15.

4. Sloan, K.E., et al., *Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function.* RNA Biol, 2017. **14**(9): p. 1138-1152.

5. Sun, W.J., et al., *RMBase: a resource for decoding the landscape of RNA modifications from high-throughput sequencing data.* Nucleic Acids Res. **44**(D1): p. D259-65.

6. 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.

7. Czekay, D.P. and U. Kothe, *H/ACA Small Ribonucleoproteins: Structural and Functional Comparison Between Archaea and Eukaryotes.* Front Microbiol, 2021. **12**: p. 654370.

8. 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.

9. Decatur, W.A. and M.N. Schnare, *Different mechanisms for pseudouridine formation in yeast 5S and 5.8S rRNAs.* Mol Cell Biol, 2008. **28**(10): p. 3089-100.

10. Liang, J., et al., *Small Nucleolar RNAs: Insight Into Their Function in Cancer.* Front Oncol, 2019. **9**: p. 587.

11. Falaleeva, M., et al., *C/D-box snoRNAs form methylating and non-methylating ribonucleoprotein complexes: Old dogs show new tricks.* Bioessays, 2017. **39**(6).

12. Russell, A.G., M.N. Schnare, and M.W. Gray, *Pseudouridine-guide RNAs and other Cbf5p-associated RNAs in Euglena gracilis.* RNA, 2004. **10**(7): p. 1034-46.

13. Cohn, W.E., *Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: isolation, structure, and chemical characteristics.* J Biol Chem, 1960. **235**: p. 1488-98.

14. 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. 823-34.

15. 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. 305-12.

16. Hertel, J., I.L. Hofacker, and P.F. Stadler, *SnoReport: computational identification of snoRNAs with unknown targets.* Bioinformatics, 2008. **24**(2): p. 158-64.

17. de Araujo Oliveira, J.V., et al., *SnoReport 2.0: new features and a refined Support Vector Machine to improve snoRNA identification.* BMC Bioinformatics, 2016. **17**(Suppl 18): p. 464.

18. Jorjani, H., et al., *An updated human snoRNAome.* Nucleic Acids Res, 2016. **44**(11): p. 5068-82.

19. Yang, J.H., et al., *snoSeeker: an advanced computational package for screening of guide and orphan snoRNA genes in the human genome.* Nucleic Acids Res, 2006. **34**(18): p. 5112-23.

20. Kiss, A.M., et al., *A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains.* Nucleic Acids Res, 2002. **30**(21): p. 4643-9.

21. Darzacq, X., et al., *Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs.* EMBO J, 2002. **21**(11): p. 2746-56.

22. Jady, B.E., A. Ketele, and T. Kiss, *Human intron-encoded Alu RNAs are processed and packaged into Wdr79-associated nucleoplasmic box H/ACA RNPs.* Genes Dev, 2012. **26**(17): p. 1897-910.

23. Ketele, A., T. Kiss, and B.E. Jady, *Human intron-encoded AluACA RNAs and telomerase RNA share a common element promoting RNA accumulation.* RNA Biol, 2016. **13**(12): p. 1274-1285.

24. Samarsky, D.A. and M.J. Fournier, *A comprehensive database for the small nucleolar RNAs from Saccharomyces cerevisiae.* Nucleic Acids Res, 1999. **27**(1): p. 161-4.

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

26. Li, S., et al., *Reconstitution and structural analysis of the yeast box H/ACA RNA-guided pseudouridine synthase.* Genes Dev, 2011. **25**(22): p. 2409-21.

27. Ghanim, G.E., et al., *Structure of human telomerase holoenzyme with bound telomeric DNA.* Nature, 2021. **593**(7859): p. 449-453.

28. Chan, H., Y. Wang, and J. Feigon, *Progress in Human and Tetrahymena Telomerase Structure Determination.* Annu Rev Biophys, 2017. **46**: p. 199-225.

29. Mitchell, J.R. and K. Collins, *Human telomerase activation requires two independent interactions between telomerase RNA and telomerase reverse transcriptase.* Mol Cell, 2000. **6**(2): p. 361-71.

30. Fu, D. and K. Collins, *Distinct biogenesis pathways for human telomerase RNA and H/ACA small nucleolar RNAs.* Mol Cell, 2003. **11**(5): p. 1361-72.

31. Hamma, T. and A.R. Ferre-D'Amare, *The box H/ACA ribonucleoprotein complex: interplay of RNA and protein structures in post-transcriptional RNA modification.* J Biol Chem. **285**(2): p. 805-9.

32. Li, L. and K. Ye, *Crystal structure of an H/ACA box ribonucleoprotein particle.* Nature, 2006. **443**(7109): p. 302-7.

33. Tollervey, D., et al., *Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly.* Cell, 1993. **72**(3): p. 443-57.

34. Wang, P., et al., *Accurate placement of substrate RNA by Gar1 in H/ACA RNA-guided pseudouridylation.* Nucleic Acids Res, 2015. **43**(15): p. 7207-16.

35. Wang, C. and U.T. Meier, *Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins.* EMBO J, 2004. **23**(8): p. 1857-67.

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

37. Swaney, D.L., et al., *Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation.* Nat Methods, 2013. **10**(7): p. 676-82.

38. Huang, H., et al., *iPTMnet: an integrated resource for protein post-translational modification network discovery.* Nucleic Acids Res, 2018. **46**(D1): p. D542-D550.

39. Rigbolt, K.T., et al., *System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation.* Sci Signal, 2011. **4**(164): p. rs3.

40. MacNeil, D.E., et al., *SUMOylation- and GAR1-Dependent Regulation of Dyskerin Nuclear and Subnuclear Localization.* Mol Cell Biol, 2021. **41**(4).

41. Henras, A., et al., *Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs.* EMBO J, 1998. **17**(23): p. 7078-90.

42. Charpentier, B., S. Muller, and C. Branlant, *Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation.* Nucleic Acids Res, 2005. **33**(10): p. 3133-44.

43. Henras, A.K., et al., *Cbf5p, the putative pseudouridine synthase of H/ACA-type snoRNPs, can form a complex with Gar1p and Nop10p in absence of Nhp2p and box H/ACA snoRNAs.* RNA, 2004. **10**(11): p. 1704-12.

44. Zebarjadian, Y., et al., *Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA.* Mol Cell Biol, 1999. **19**(11): p. 7461-72.

45. Lafontaine, D.L., et al., *The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase.* Genes Dev, 1998. **12**(4): p. 527-37.

46. Morrissey, J.P. and D. Tollervey, *Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis.* Mol Cell Biol, 1993. **13**(4): p. 2469-77.

47. Vos, T.J. and U. Kothe, *snR30/U17 Small Nucleolar Ribonucleoprotein: A Critical Player during Ribosome Biogenesis.* Cells, 2020. **9**(10).

48. Andrew, C., A.K. Hopper, and B.D. Hall, *A yeast mutant defective in the processing of 27S r-RNA precursor.* Mol Gen Genet, 1976. **144**(1): p. 29-37.

49. Jiang, W., et al., *An essential yeast protein, CBF5p, binds in vitro to centromeres and microtubules.* Mol Cell Biol, 1993. **13**(8): p. 4884-93.

50. Dieci, G., M. Preti, and B. Montanini, *Eukaryotic snoRNAs: a paradigm for gene expression flexibility.* Genomics, 2009. **94**(2): p. 83-8.

51. Bouchard-Bourelle, P., et al., *snoDB: an interactive database of human snoRNA sequences, abundance and interactions.* Nucleic Acids Res, 2020. **48**(D1): p. D220-D225.

52. 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. D158-62.

53. Ballarino, M., et al., *The cotranscriptional assembly of snoRNPs controls the biosynthesis of H/ACA snoRNAs in Saccharomyces cerevisiae.* Mol Cell Biol, 2005. **25**(13): p. 5396-403.

54. Kiss, T., E. Fayet-Lebaron, and B.E. Jady, *Box H/ACA small ribonucleoproteins.* Mol Cell, 2010. **37**(5): p. 597-606.

55. 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.

56. Yang, P.K., et al., *The Shq1p.Naf1p complex is required for box H/ACA small nucleolar ribonucleoprotein particle biogenesis.* J Biol Chem, 2002. **277**(47): p. 45235-42.

57. 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. 7053-65.

58. Grozdanov, P.N., et al., *SHQ1 is required prior to NAF1 for assembly of H/ACA small nucleolar and telomerase RNPs.* RNA, 2009. **15**(6): p. 1188-97.

59. 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. 231-44.

60. Singh, M., et al., *Structure and functional studies of the CS domain of the essential H/ACA ribonucleoparticle assembly protein SHQ1.* J Biol Chem, 2009. **284**(3): p. 1906-16.

61. Walbott, H., et al., *The H/ACA RNP assembly factor SHQ1 functions as an RNA mimic.* Genes Dev, 2011. **25**(22): p. 2398-408.

62. Li, S., et al., *Structure of the Shq1-Cbf5-Nop10-Gar1 complex and implications for H/ACA RNP biogenesis and dyskeratosis congenita.* EMBO J, 2011. **30**(24): p. 5010-20.

63. Fatica, A., M. Dlakic, and D. Tollervey, *Naf1 p is a box H/ACA snoRNP assembly factor.* RNA, 2002. **8**(12): p. 1502-14.

64. 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. 1338-53.

65. 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. 1833-45.

66. Kakihara, Y. and M. Saeki, *The R2TP chaperone complex: its involvement in snoRNP assembly and tumorigenesis.* Biomol Concepts, 2014. **5**(6): p. 513-20.

67. 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.

68. Darzacq, X., et al., *Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells.* J Cell Biol, 2006. **173**(2): p. 207-18.

69. Pellizzoni, L., et al., *The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1.* Curr Biol, 2001. **11**(14): p. 1079-88.

70. Kolb, S.J., D.J. Battle, and G. Dreyfuss, *Molecular functions of the SMN complex.* J Child Neurol, 2007. **22**(8): p. 990-4.

71. Yang, Y., et al., *Conserved composition of mammalian box H/ACA and box C/D small nucleolar ribonucleoprotein particles and their interaction with the common factor Nopp140.* Mol Biol Cell, 2000. **11**(2): p. 567-77.

72. Kelly, E.K., D.P. Czekay, and U. Kothe, *Base-pairing interactions between substrate RNA and H/ACA guide RNA modulate the kinetics of pseudouridylation, but not the affinity of substrate binding by H/ACA small nucleolar ribonucleoproteins.* RNA, 2019. **25**(10): p. 1393-1404.

73. 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. 565-73.

74. Li, Y., et al., *A polysaccharide from Pinellia ternata inhibits cell proliferation and metastasis in human cholangiocarcinoma cells by targeting of Cdc42 and 67kDa Laminin Receptor (LR).* Int J Biol Macromol, 2016. **93**(Pt A): p. 520-525.

75. 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. 6655-60.

76. Jin, H., J.P. Loria, and P.B. Moore, *Solution structure of an rRNA substrate bound to the pseudouridylation pocket of a box H/ACA snoRNA.* Mol Cell, 2007. **26**(2): p. 205-15.

77. Tschochner, H. and E. Hurt, *Pre-ribosomes on the road from the nucleolus to the cytoplasm.* Trends Cell Biol, 2003. **13**(5): p. 255-63.

78. Klinge, S., et al., *Atomic structures of the eukaryotic ribosome.* Trends Biochem Sci. **37**(5): p. 189-98.

79. Klinge, S. and J.L. Woolford, Jr., *Ribosome assembly coming into focus.* Nat Rev Mol Cell Biol. **20**(2): p. 116-131.

80. Warner, J.R., *The economics of ribosome biosynthesis in yeast.* Trends Biochem Sci, 1999. **24**(11): p. 437-40.

81. Bally, M., J. Hughes, and G. Cesareni, *SnR30: a new, essential small nuclear RNA from Saccharomyces cerevisiae.* Nucleic Acids Res, 1988. **16**(12): p. 5291-303.

82. Fayet-Lebaron, E., et al., *18S rRNA processing requires base pairings of snR30 H/ACA snoRNA to eukaryote-specific 18S sequences.* EMBO J, 2009. **28**(9): p. 1260-70.

83. Atzorn, V., P. Fragapane, and T. Kiss, *U17/snR30 is a ubiquitous snoRNA with two conserved sequence motifs essential for 18S rRNA production.* Mol Cell Biol, 2004. **24**(4): p. 1769-78.

84. Vos, T.J. and U. Kothe, *Synergistic interaction network between the snR30 RNP, Utp23, and ribosomal RNA during ribosome synthesis.* RNA Biol, 2022. **19**(1): p. 764-773.

85. Liang, X.H., et al., *Strong dependence between functional domains in a dual-function snoRNA infers coupling of rRNA processing and modification events.* Nucleic Acids Res, 2010. **38**(10): p. 3376-87.

86. Bohnsack, M.T., M. Kos, and D. Tollervey, *Quantitative analysis of snoRNA association with pre-ribosomes and release of snR30 by Rok1 helicase.* EMBO Rep, 2008. **9**(12): p. 1230-6.

87. Tollervey, D. and C. Guthrie, *Deletion of a yeast small nuclear RNA gene impairs growth.* EMBO J, 1985. **4**(13B): p. 3873-8.

88. Liang, W.Q. and M.J. Fournier, *U14 base-pairs with 18S rRNA: a novel snoRNA interaction required for rRNA processing.* Genes Dev, 1995. **9**(19): p. 2433-43.

89. Sharma, K. and D. Tollervey, *Base pairing between U3 small nucleolar RNA and the 5' end of 18S rRNA is required for pre-rRNA processing.* Mol Cell Biol, 1999. **19**(9): p. 6012-9.

90. Henras, A.K., et al., *The post-transcriptional steps of eukaryotic ribosome biogenesis.* Cell Mol Life Sci, 2008. **65**(15): p. 2334-59.

91. Strunk, B.S. and K. Karbstein, *Powering through ribosome assembly.* RNA, 2009. **15**(12): p. 2083-104.

92. Woolford, J.L., Jr. and S.J. Baserga, *Ribosome biogenesis in the yeast Saccharomyces cerevisiae.* Genetics, 2013. **195**(3): p. 643-81.

93. Dutca, L.M., J.E. Gallagher, and S.J. Baserga, *The initial U3 snoRNA:pre-rRNA base pairing interaction required for pre-18S rRNA folding revealed by in vivo chemical probing.* Nucleic Acids Res, 2011. **39**(12): p. 5164-80.

94. Gerczei, T., et al., *RNA chaperones stimulate formation and yield of the U3 snoRNA-Pre-rRNA duplexes needed for eukaryotic ribosome biogenesis.* J Mol Biol, 2009. **390**(5): p. 991-1006.

95. Billy, E., et al., *Rcl1p, the yeast protein similar to the RNA 3'-phosphate cyclase, associates with U3 snoRNP and is required for 18S rRNA biogenesis.* EMBO J, 2000. **19**(9): p. 2115-26.

96. Horn, D.M., S.L. Mason, and K. Karbstein, *Rcl1 protein, a novel nuclease for 18 S ribosomal RNA production.* J Biol Chem, 2011. **286**(39): p. 34082-7.

97. Pertschy, B., et al., *RNA helicase Prp43 and its co-factor Pfa1 promote 20 to 18 S rRNA processing catalyzed by the endonuclease Nob1.* J Biol Chem, 2009. **284**(50): p. 35079-91.

98. de la Cruz, J., et al., *Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3' end formation of 5.8S rRNA in Saccharomyces cerevisiae.* EMBO J, 1998. **17**(4): p. 1128-40.

99. Allmang, C., et al., *Functions of the exosome in rRNA, snoRNA and snRNA synthesis.* EMBO J, 1999. **18**(19): p. 5399-410.

100. van Hoof, A., P. Lennertz, and R. Parker, *Three conserved members of the RNase D family have unique and overlapping functions in the processing of 5S, 5.8S, U4, U5, RNase MRP and RNase P RNAs in yeast.* EMBO J, 2000. **19**(6): p. 1357-65.

101. Thomson, E. and D. Tollervey, *The final step in 5.8S rRNA processing is cytoplasmic in Saccharomyces cerevisiae.* Mol Cell Biol, 2010. **30**(4): p. 976-84.

102. Geerlings, T.H., J.C. Vos, and H.A. Raue, *The final step in the formation of 25S rRNA in Saccharomyces cerevisiae is performed by 5'-->3' exonucleases.* RNA, 2000. **6**(12): p. 1698-703.

103. Torchet, C., et al., *The complete set of H/ACA snoRNAs that guide rRNA pseudouridylations in Saccharomyces cerevisiae.* RNA, 2005. **11**(6): p. 928-38.

104. Decatur, W.A. and M.J. Fournier, *rRNA modifications and ribosome function.* Trends Biochem Sci, 2002. **27**(7): p. 344-51.

105. Henras, A.K., et al., *Synthesis, Function, and Heterogeneity of snoRNA-Guided Posttranscriptional Nucleoside Modifications in Eukaryotic Ribosomal RNAs.* Enzymes. **41**: p. 169-213.

106. Meyer, B., et al., *Ribosome biogenesis factor Tsr3 is the aminocarboxypropyl transferase responsible for 18S rRNA hypermodification in yeast and humans.* Nucleic Acids Res, 2016. **44**(9): p. 4304-16.

107. Liang, X.H., Q. Liu, and M.J. Fournier, *Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing.* RNA, 2009. **15**(9): p. 1716-28.

108. Babaian, A., et al., *Loss of m(1)acp(3)Psi Ribosomal RNA Modification Is a Major Feature of Cancer.* Cell Rep, 2020. **31**(5): p. 107611.

109. Baudin-Baillieu, A., et al., *Nucleotide modifications in three functionally important regions of the Saccharomyces cerevisiae ribosome affect translation accuracy.* Nucleic Acids Res, 2009. **37**(22): p. 7665-77.

110. Piekna-Przybylska, D., et al., *Ribosome performance is enhanced by a rich cluster of pseudouridines in the A-site finger region of the large subunit.* J Biol Chem, 2008. **283**(38): p. 26026-36.

111. Sumita, M., et al., *Comparison of solution conformations and stabilities of modified helix 69 rRNA analogs from bacteria and human.* Biopolymers, 2012. **97**(2): p. 94-106.

112. Liang, X.H., Q. Liu, and M.J. Fournier, *rRNA modifications in an intersubunit bridge of the ribosome strongly affect both ribosome biogenesis and activity.* Mol Cell, 2007. **28**(6): p. 965-77.

113. O'Connor, M. and S.T. Gregory, *Inactivation of the RluD pseudouridine synthase has minimal effects on growth and ribosome function in wild-type Escherichia coli and Salmonella enterica.* J Bacteriol, 2011. **193**(1): p. 154-62.

114. Meroueh, M., et al., *Unique structural and stabilizing roles for the individual pseudouridine residues in the 1920 region of Escherichia coli 23S rRNA.* Nucleic Acids Res, 2000. **28**(10): p. 2075-83.

115. Yoon, A., et al., *Impaired control of IRES-mediated translation in X-linked dyskeratosis congenita.* Science, 2006. **312**(5775): p. 902-6.

116. Jack, K., et al., *rRNA pseudouridylation defects affect ribosomal ligand binding and translational fidelity from yeast to human cells.* Mol Cell, 2011. **44**(4): p. 660-6.

117. 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. 457-69.

118. Rocchi, L., et al., *Dyskerin depletion increases VEGF mRNA internal ribosome entry site-mediated translation.* Nucleic Acids Res, 2013. **41**(17): p. 8308-18.

119. Bellodi, C., N. Kopmar, and D. Ruggero, *Deregulation of oncogene-induced senescence and p53 translational control in X-linked dyskeratosis congenita.* EMBO J, 2010. **29**(11): p. 1865-76.

120. Bellodi, C., et al., *Loss of function of the tumor suppressor DKC1 perturbs p27 translation control and contributes to pituitary tumorigenesis.* Cancer Res, 2010. **70**(14): p. 6026-35.

121. McMahon, M., A. Contreras, and D. Ruggero, *Small RNAs with big implications: new insights into H/ACA snoRNA function and their role in human disease.* Wiley Interdiscip Rev RNA, 2015. **6**(2): p. 173-89.

122. Marciniak, R.A., F.B. Johnson, and L. Guarente, *Dyskeratosis congenita, telomeres and human ageing.* Trends Genet, 2000. **16**(5): p. 193-5.

123. Montanaro, L., *Dyskerin and cancer: more than telomerase. The defect in mRNA translation helps in explaining how a proliferative defect leads to cancer.* J Pathol. **222**(4): p. 345-9.

124. Bellodi, C., et al., *H/ACA small RNA dysfunctions in disease reveal key roles for noncoding RNA modifications in hematopoietic stem cell differentiation.* Cell Rep, 2013. **3**(5): p. 1493-502.

125. McMahon, M., A. Contreras, and D. Ruggero, *Small RNAs with big implications: new insights into H/ACA snoRNA function and their role in human disease.* Wiley Interdiscip Rev RNA. **6**(2): p. 173-89.

126. Nagasawa, C., et al., *The Role of scaRNAs in Adjusting Alternative mRNA Splicing in Heart Development.* J Cardiovasc Dev Dis, 2018. **5**(2).

127. Langhendries, J.L., et al., *The human box C/D snoRNAs U3 and U8 are required for pre-rRNA processing and tumorigenesis.* Oncotarget, 2016. **7**(37): p. 59519-59534.

128. Chang, L.S., et al., *Differential expression of human 5S snoRNA genes.* Biochem Biophys Res Commun, 2002. **299**(2): p. 196-200.

129. Mei, Y.P., et al., *Small nucleolar RNA 42 acts as an oncogene in lung tumorigenesis.* Oncogene, 2012. **31**(22): p. 2794-804.

130. Okugawa, Y., et al., *Clinical significance of SNORA42 as an oncogene and a prognostic biomarker in colorectal cancer.* Gut, 2017. **66**(1): p. 107-117.

131. Yoshida, K., et al., *SNORA21 - An Oncogenic Small Nucleolar RNA, with a Prognostic Biomarker Potential in Human Colorectal Cancer.* EBioMedicine, 2017. **22**: p. 68-77.

132. Chu, L., et al., *Multiple myeloma-associated chromosomal translocation activates orphan snoRNA ACA11 to suppress oxidative stress.* J Clin Invest, 2012. **122**(8): p. 2793-806.

133. Qin, Y., et al., *SNORA74B gene silencing inhibits gallbladder cancer cells by inducing PHLPP and suppressing Akt/mTOR signaling.* Oncotarget, 2017. **8**(12): p. 19980-19996.

134. Liu, N., et al., *Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA.* RNA, 2013. **19**(12): p. 1848-56.

135. Wetzel, C. and P.A. Limbach, *Mass spectrometry of modified RNAs: recent developments.* Analyst, 2016. **141**(1): p. 16-23.

136. Heiss, M. and S. Kellner, *Detection of nucleic acid modifications by chemical reagents.* RNA Biol, 2017. **14**(9): p. 1166-1174.

137. Muller, G., *Epitranscriptomics: RNA Biology Moves into the Focus of Medicinal Chemistry.* Chimia (Aarau), 2020. **74**(3): p. 194-195.

138. Callea, M., et al., *Multisystemic Manifestations in Rare Diseases: The Experience of Dyskeratosis Congenita.* Genes (Basel), 2022. **13**(3).

139. Parry, E.M., et al., *Decreased dyskerin levels as a mechanism of telomere shortening in X-linked dyskeratosis congenita.* J Med Genet, 2011. **48**(5): p. 327-33.

140. Tycowski, K.T., et al., *A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles.* Mol Cell, 2009. **34**(1): p. 47-57.

141. Batista, L.F., et al., *Telomere shortening and loss of self-renewal in dyskeratosis congenita induced pluripotent stem cells.* Nature, 2011. **474**(7351): p. 399-402.

142. Williams, G.T. and F. Farzaneh, *Are snoRNAs and snoRNA host genes new players in cancer?* Nat Rev Cancer, 2012. **12**(2): p. 84-8.

143. Deogharia, M. and M. Majumder, *Guide snoRNAs: Drivers or Passengers in Human Disease?* Biology (Basel), 2018. **8**(1).

144. 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. 246-68.

145. DiGiacomo, V. and D. Meruelo, *Looking into laminin receptor: critical discussion regarding the non-integrin 37/67-kDa laminin receptor/RPSA protein.* Biol Rev Camb Philos Soc, 2016. **91**(2): p. 288-310.

146. Garcon, L., et al., *Ribosomal and hematopoietic defects in induced pluripotent stem cells derived from Diamond Blackfan anemia patients.* Blood, 2013. **122**(6): p. 912-21.

147. Trahan, C. and F. Dragon, *Dyskeratosis congenita mutations in the H/ACA domain of human telomerase RNA affect its assembly into a pre-RNP.* RNA, 2009. **15**(2): p. 235-43.

148. Ender, C., et al., *A human snoRNA with microRNA-like functions.* Mol Cell, 2008. **32**(4): p. 519-28.

149. Scott, M.S., et al., *Human miRNA precursors with box H/ACA snoRNA features.* PLoS Comput Biol, 2009. **5**(9): p. e1000507.

150. Jinn, S., et al., *snoRNA U17 regulates cellular cholesterol trafficking.* Cell Metab, 2015. **21**(6): p. 855-67.

151. Woese, C.R., O. Kandler, and M.L. Wheelis, *Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.* Proc Natl Acad Sci U S A, 1990. **87**(12): p. 4576-9.

152. Ramagopal, S., *Are eukaryotic ribosomes heterogeneous? Affirmations on the horizon.* Biochem Cell Biol, 1992. **70**(5): p. 269-72.

153. Shi, Z., et al., *Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide.* Mol Cell, 2017. **67**(1): p. 71-83 e7.

154. Velichutina, I.V., et al., *Chimeric rRNAs containing the GTPase centers of the developmentally regulated ribosomal rRNAs of Plasmodium falciparum are functionally distinct.* RNA, 1998. **4**(5): p. 594-602.

155. Pavlakis, G.N., et al., *Sequence and secondary structure of Drosophila melanogaster 5.8S and 2S rRNAs and of the processing site between them.* Nucleic Acids Res, 1979. **7**(8): p. 2213-38.

156. Pellegrini, M., J. Manning, and N. Davidson, *Sequence arrangement of the rDNA of Drosophila melanogaster.* Cell, 1977. **10**(2): p. 213-4.

157. Segev, N. and J.E. Gerst, *Specialized ribosomes and specific ribosomal protein paralogs control translation of mitochondrial proteins.* J Cell Biol, 2018. **217**(1): p. 117-126.

158. Krogh, N., et al., *Profiling of 2'-O-Me in human rRNA reveals a subset of fractionally modified positions and provides evidence for ribosome heterogeneity.* Nucleic Acids Res, 2016. **44**(16): p. 7884-95.

159. Montanaro, L., et al., *Novel dyskerin-mediated mechanism of p53 inactivation through defective mRNA translation.* Cancer Res, 2010. **70**(11): p. 4767-77.

160. Xue, S. and M. Barna, *Specialized ribosomes: a new frontier in gene regulation and organismal biology.* Nat Rev Mol Cell Biol, 2012. **13**(6): p. 355-69.

161. Haag, E.S. and J.D. Dinman, *Still Searching for Specialized Ribosomes.* Dev Cell, 2019. **48**(6): p. 744-746.

162. Di Maio, N., et al., *A new role for human dyskerin in vesicular trafficking.* FEBS Open Bio, 2017. **7**(10): p. 1453-1468.

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

164. Tollervey, D., *A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA.* EMBO J, 1987. **6**(13): p. 4169-75.

165. 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. 941-56.

166. Trucks, S., G. Hanspach, and M. Hengesbach, *Eukaryote specific RNA and protein features facilitate assembly and catalysis of H/ACA snoRNPs.* Nucleic Acids Res, 2021. **49**(8): p. 4629-4642.

167. Rashid, R., et al., *Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita.* Mol Cell, 2006. **21**(2): p. 249-60.

168. Trahan, C., C. Martel, and F. Dragon, *Effects of dyskeratosis congenita mutations in dyskerin, NHP2 and NOP10 on assembly of H/ACA pre-RNPs.* Hum Mol Genet, 2010. **19**(5): p. 825-36.

169. Tang, T.H., et al., *Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus.* Proc Natl Acad Sci U S A, 2002. **99**(11): p. 7536-41.

170. Aquino, G.R.R., et al., *RNA helicase-mediated regulation of snoRNP dynamics on pre-ribosomes and rRNA 2'-O-methylation.* Nucleic Acids Res, 2021. **49**(7): p. 4066-4084.

171. Jaafar, M., et al., *Association of snR190 snoRNA chaperone with early pre-60S particles is regulated by the RNA helicase Dbp7 in yeast.* Nat Commun, 2021. **12**(1): p. 6153.

172. 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.

173. Keffer-Wilkes, L.C., E.F. Soon, and U. Kothe, *The methyltransferase TrmA facilitates tRNA folding through interaction with its RNA-binding domain.* Nucleic Acids Res, 2020. **48**(14): p. 7981-7990.

174. Porat, J., U. Kothe, and M.A. Bayfield, *Revisiting tRNA chaperones: New players in an ancient game.* RNA, 2021.

175. Baker, D.L., et al., *RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP.* Genes Dev, 2005. **19**(10): p. 1238-48.

176. Duan, J., et al., *Structural mechanism of substrate RNA recruitment in H/ACA RNA-guided pseudouridine synthase.* Mol Cell, 2009. **34**(4): p. 427-39.

177. Dragon, F., et al., *A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis.* Nature, 2002. **417**(6892): p. 967-70.

178. Meier, U.T., *The daunting task of modifying ribosomal RNA.* RNA, 2022. **28**(12): p. 1555-1557.

179. Jady, B.E., et al., *Guide RNA acrobatics: positioning consecutive uridines for pseudouridylation by H/ACA pseudouridylation loops with dual guide capacity.* Genes Dev, 2022. **36**(1-2): p. 70-83.

180. Meier, U.T., *Guide RNA acrobatics: the one-for-two shuffle.* Genes Dev, 2022. **36**(1-2): p. 1-3.

181. Winkler, A.A., et al., *The lysine-rich C-terminal repeats of the centromere-binding factor 5 (Cbf5) of Kluyveromyces lactis are not essential for function.* Yeast, 1998. **14**(1): p. 37-48.

182. 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. 32-8.

183. 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. 673-82.

184. Bousquet-Antonelli, C., et al., *Rrp8p is a yeast nucleolar protein functionally linked to Gar1p and involved in pre-rRNA cleavage at site A2.* RNA, 2000. **6**(6): p. 826-43.

185. Li, P.T., J. Vieregg, and I. Tinoco, Jr., *How RNA unfolds and refolds.* Annu Rev Biochem, 2008. **77**: p. 77-100.

186. Mitterer, V. and B. Pertschy, *RNA folding and functions of RNA helicases in ribosome biogenesis.* RNA Biol, 2022. **19**(1): p. 781-810.

187. Hampel, K.J. and J.M. Burke, *Time-resolved hydroxyl-radical footprinting of RNA using Fe(II)-EDTA.* Methods, 2001. **23**(3): p. 233-9.

188. Peng, Y., T.J. Soper, and S.A. Woodson, *RNase footprinting of protein binding sites on an mRNA target of small RNAs.* Methods Mol Biol, 2012. **905**: p. 213-24.

189. Tijerina, P., S. Mohr, and R. Russell, *DMS footprinting of structured RNAs and RNA-protein complexes.* Nat Protoc, 2007. **2**(10): p. 2608-23.

190. Depmeier, H., et al., *Strategies for Covalent Labeling of Long RNAs.* Chembiochem, 2021. **22**(19): p. 2826-2847.

191. Kershaw, C.J. and R.T. O'Keefe, *Splint ligation of RNA with T4 DNA ligase.* Methods Mol Biol, 2012. **941**: p. 257-69.

192. Hoang, C. and A.R. Ferre-D'Amare, *Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme.* Cell, 2001. **107**(7): p. 929-39.

193. Eyler, D.E., et al., *Pseudouridinylation of mRNA coding sequences alters translation.* Proc Natl Acad Sci U S A, 2019. **116**(46): p. 23068-23074.

194. Pelletier, J., G. Thomas, and S. Volarevic, *Ribosome biogenesis in cancer: new players and therapeutic avenues.* Nat Rev Cancer, 2018. **18**(1): p. 51-63.

195. Gaviraghi, M., C. Vivori, and G. Tonon, *How Cancer Exploits Ribosomal RNA Biogenesis: A Journey beyond the Boundaries of rRNA Transcription.* Cells, 2019. **8**(9).

APPENDIX 1: SUPPLEMENTARY INFORMATION FOR CHAPTERS 3 & 4

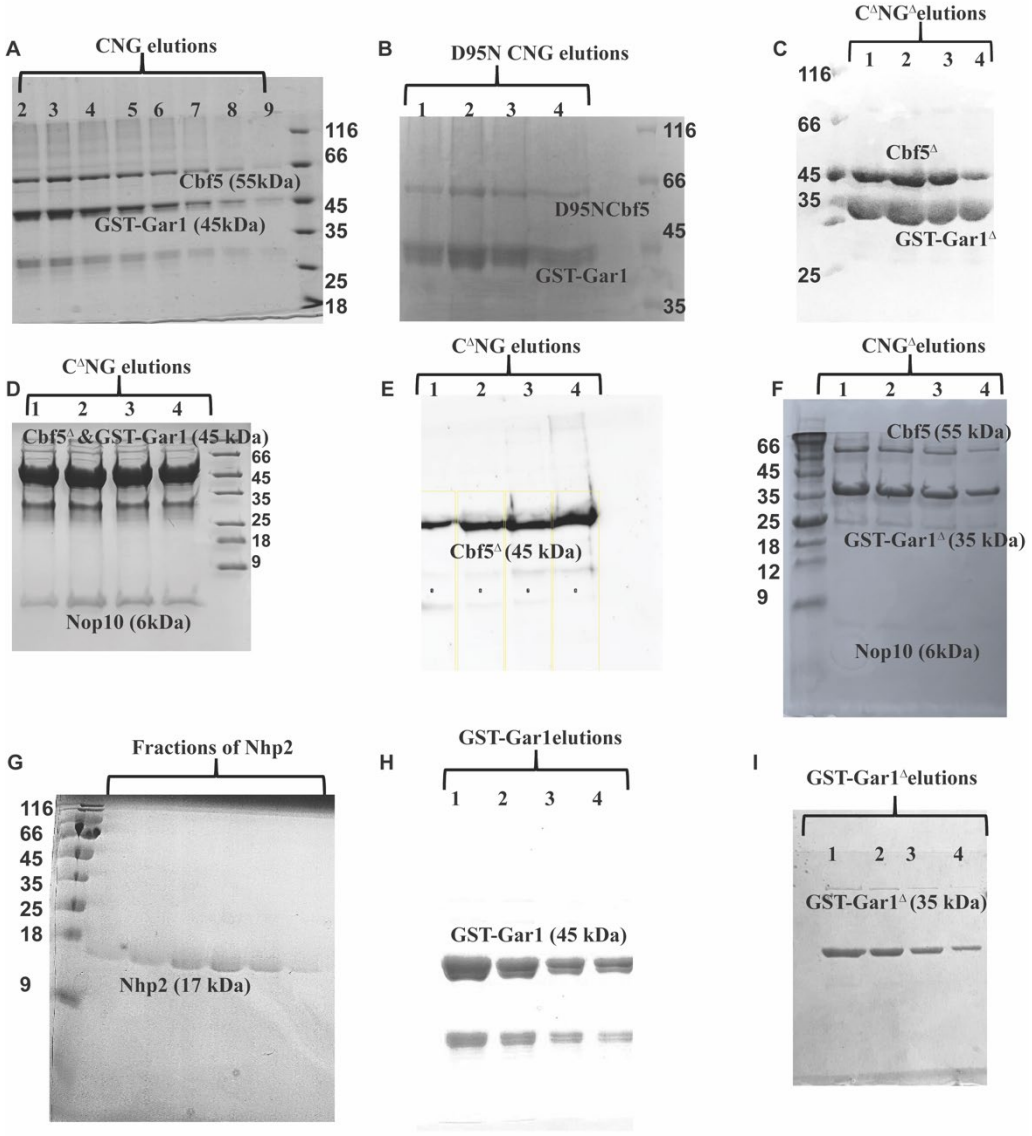

Figure A1. Protein purifications: A WT CNG (Cbf5•Nop10•GST-Gar1) elutions 2- 9 from Ni-Sepharose; Cbf5 is 55 kDa, and migrates at 65 kDa in the SDS-PAGE due to its high basic charge. **B** D95N CNG (D95N Cbf5•Nop10•GST-Gar1) elutions 1-4 from Ni-Sepharose **C** is Cbf5$^{\Delta}$•Nop10•GST-Gar1$^{\Delta}$•Nhp2 (C$^{\Delta}$NG$^{\Delta}$P) elutions 1-4 from Ni-Sepharose; where Cbf5$^{\Delta}$ is 45 kDa and GST-Gar1$^{\Delta}$ is 35 kDa, **D** and **E** Cbf5$^{\Delta}$•Nop10•GST-Gar1•Nhp2 (C$^{\Delta}$NGP) elutions 1-4 from Ni-Sepharose; where Cbf5$^{\Delta}$ is 45 kDa, GST-Gar1 is 45 kDa and Nop10 is 6 kDa, **E** represents confirmation Cbf5$^{\Delta}$ by western blotting using anti His tag antibodies, **F** Cbf5•Nop10•GST-Gar1$^{\Delta}$•Nhp2 (CNG$^{\Delta}$P), **G** Nhp2 (17 kDa) purification, the gel represents fractions from cation exchange chromatography column (SP Sepharose Fast Flow), **H** and **I** are elutions 1-4 representing GST-Gar1 and GST-Gar1$^{\Delta}$ purifications, respectively. All gels are stained using Coomassie brilliant blue.

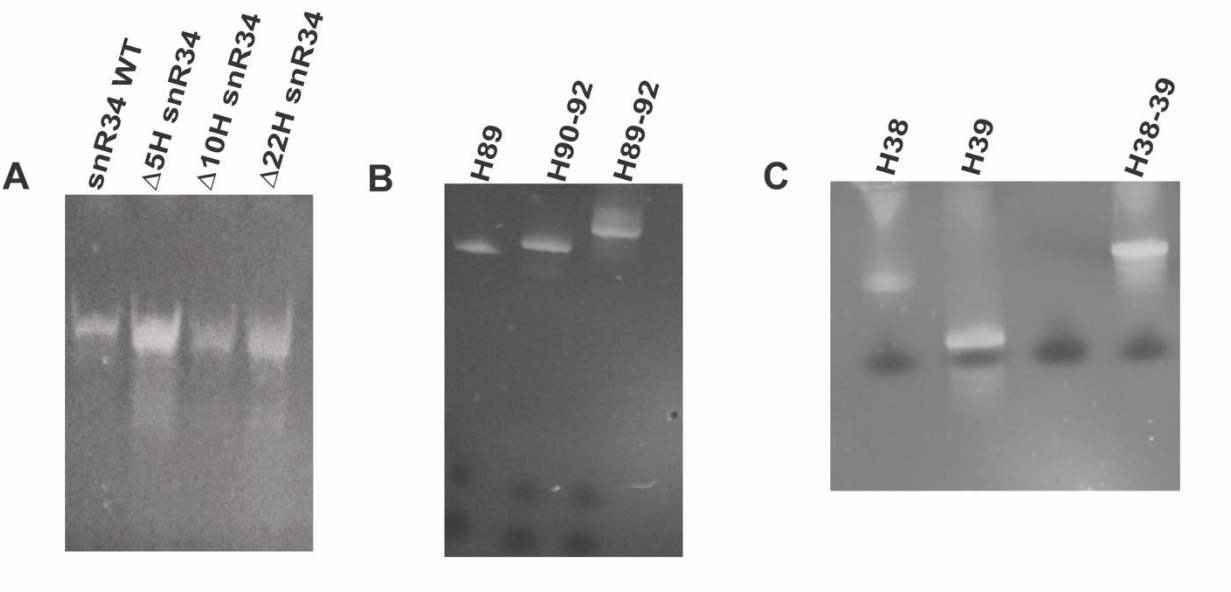

Figure A2. RNA purifications visualized by Urea-PAGE and stained with syber gold. A WT snR34 WT (203 nt), Δ5H snR34 (198 nt), Δ10H snR34 (193 nt) and Δ22H snR34 (181 nt) respectively **B** H89 (61 nt), H90-92 (80 nt) and H89-92 (140 nt) respectively **C** H38 (93 nt), H39 (56 nt) and H38-39 (159 nt) respectively.

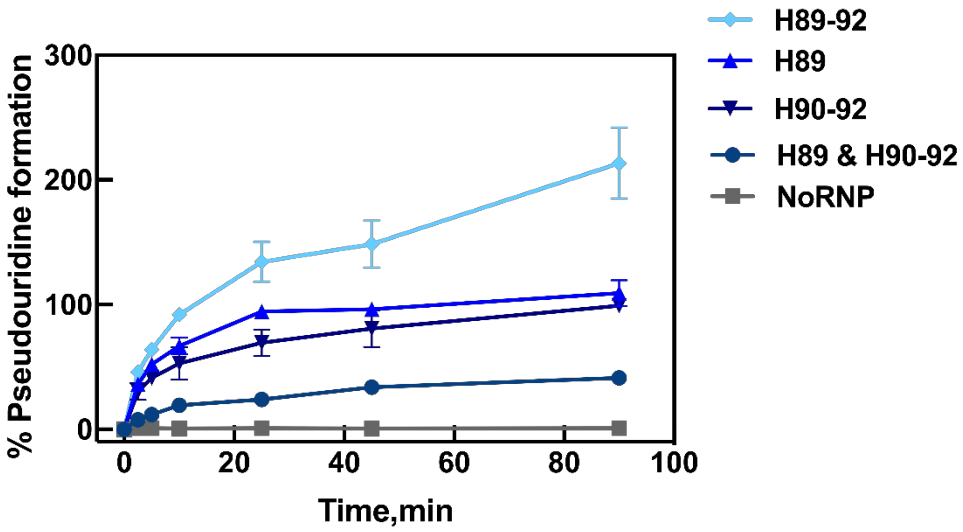

Figure A3. Pseudouridine formation by snR34 H/ACA RNP in the presence and absence of radioactive competing RNA. Pseudouridine formation was analyzed by incubating 50 nM *in vitro* reconstituted H/ACA snoRNP complex containing snR34 with 500 nM [^{3}H-C5] uridine-labelled substrate RNA. Pseudouridine formation in H89-92 (directed by both the 5′ and 3′ hairpin), H89 (directed by the 5′ hairpin) and H90-92 (directed by the 3′ hairpin) were measured in the absence of competing RNA. Furthermore, I analyzed the simultaneous pseudouridylation of both [^{3}H-C5] uridine-labelled H90-92 and H89 substrate RNAs. A negative control reaction was conducted without H/ACA snoRNP (labelled No RNP).

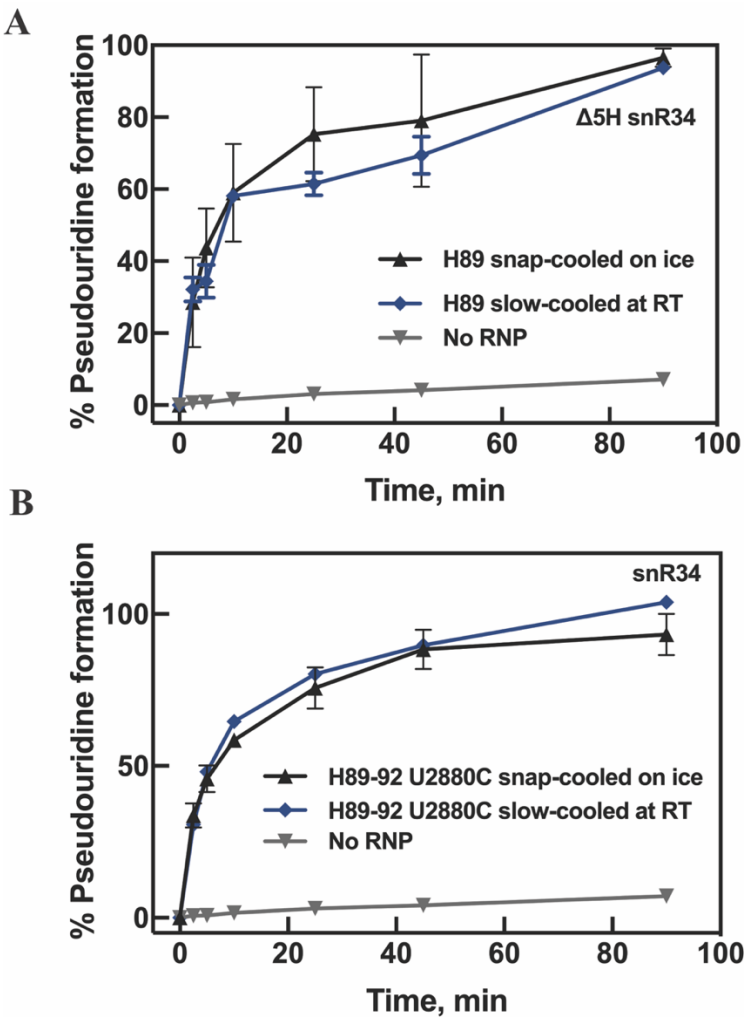

Figure A4. Pseudouridine formation comparing slow-cooled and snap-cooled substrate RNA. Pseudouridine formation was analyzed by incubating 50 nM *in vitro* reconstituted H/ACA snoRNP complex with 500 nM [^{3}H-C5] uridine-labelled substrate RNA. The dark blue color represents the substrate RNA unfolded at 65 ºC for 5 min, and slow cooled at room temp for 30 min. The black color represents the snap-cooled RNA incubated on ice for 5 min after the heating step at 65 ºC. No RNP is a control for the experiments. **A** Pseudouridine formation in H89 by 50 nM *in vitro* reconstituted H/ACA snoRNP complex containing Δ5H snR34. **B** Pseudouridine formation in H89 and H89-92 U2880C by 50 nM *in vitro* reconstituted H/ACA snoRNP complex containing snR34 wild-type.

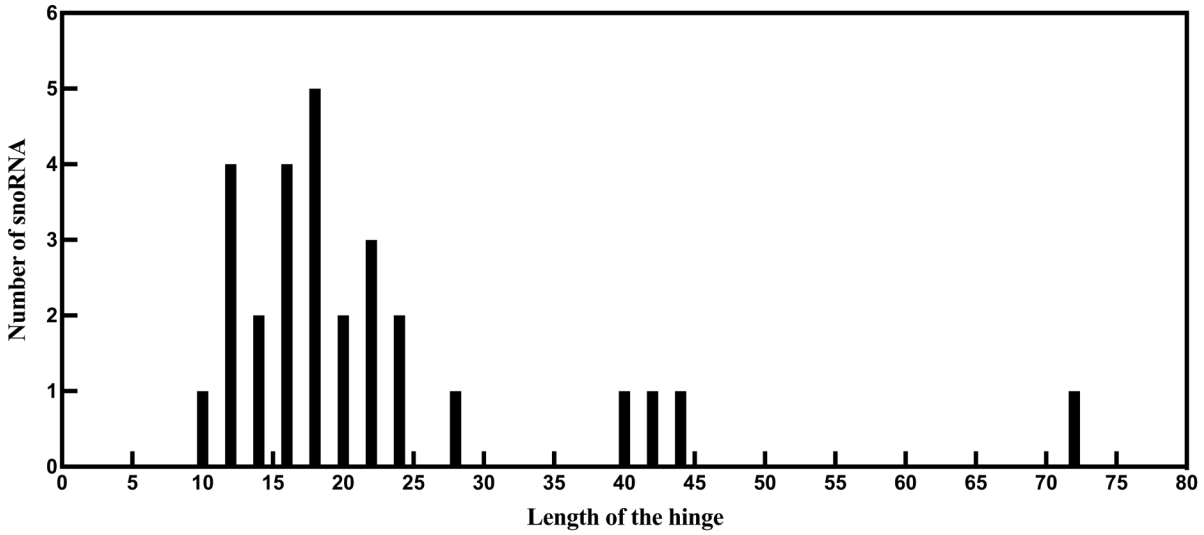

Figure A5. Frequency distribution map of yeast H/ACA snoRNAs with respect to their hinge region length. The 29 yeast H/ACA snoRNA molecules are analyzed using m-fold to predict the hinge length **(Table A1)**. Here, the hinge length including the 6-nucleotide (nt) H box is reported. However, for snR86 the H box is 7 nt in length instead of 6 nt.

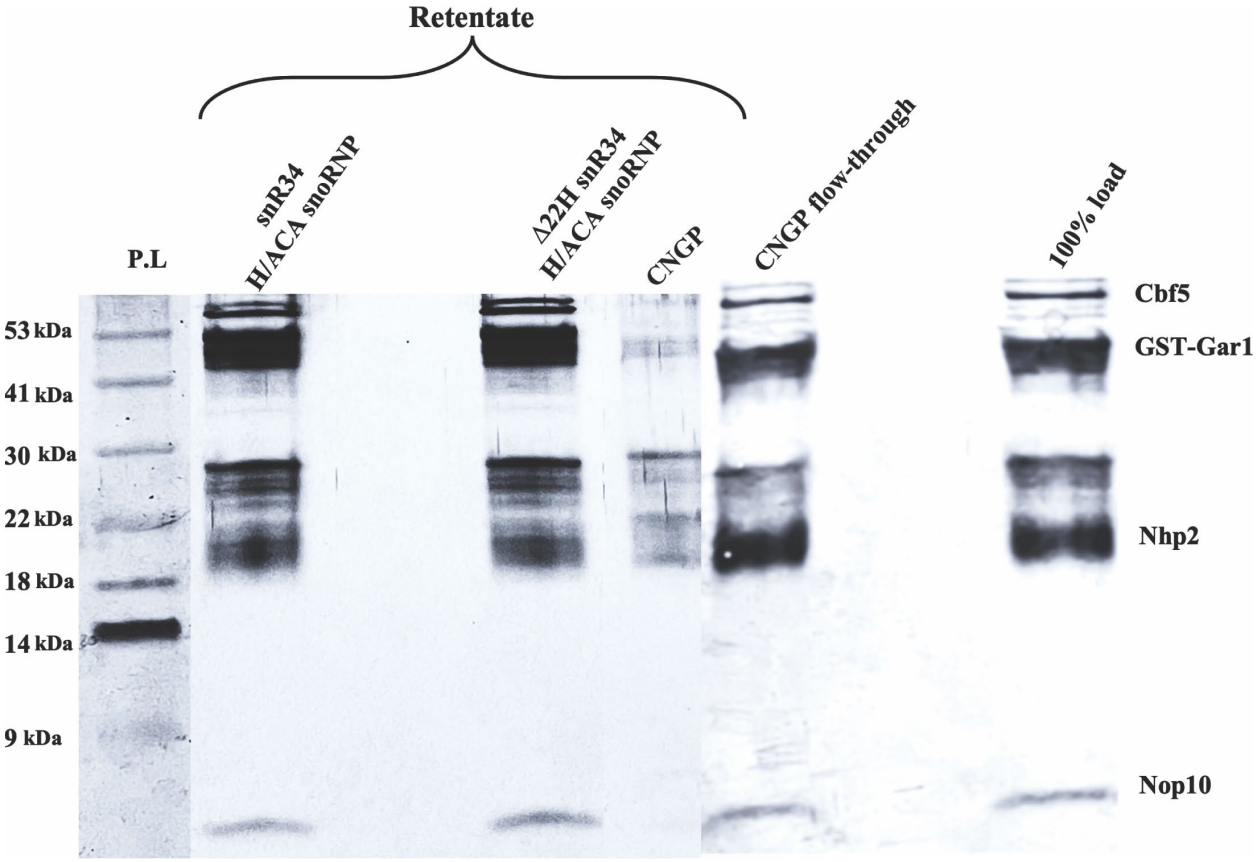

Figure A6. Protein composition analysis of H/ACA snoRNPs reconstituted with full-length snR34 or Δ22H snR34 with a shortened hinge region. After reconstituting the snR34 H/ACA snoRNP complex, it was subjected to ultrafiltration using 100kDa cutoff filters to remove unbound protein. The protein bound to snR34 and retained above the filter is analyzed by 12% Tris-Tricine gel. The protein ladder (P.L), followed by snR34 and Δ22H snR34 *in vitro* reconstituted samples respectively. The Cbf5•Nop10•GST-Gar1•Nhp2 (CNGP) alone sample (negative control), where we are not expecting any proteins presence except the background. For the negative control, the Cbf5•Nop10•GST-Gar1•Nhp2 flow through (CNGP FT) was also TCA precipitated and loaded in lane 6. Purified Cbf5•Nop10•GST-Gar1•Nhp2 proteins are loaded in lane 8 as a loading control for the experiment. The band around 28 kDa might be GST contamination, which is also present in our original purifications **(Figure A1 A** and **H).** Moreover, the snR34 and Δ22H snR34 *in vitro* reconstituted samples has snR34/Δ22H snR34 RNA present along with CNGP, which might cause proteins to migrate differently; as a result we are observing multiple close bands on the gel which are not present in CNGP and loading control sample. Moreover, the samples are analyzed using silver staining method which also detects RNA.

Figure A7. Sequence alignment of Gar1 protein. I compared the Gar1 protein sequence ranging from archaeal to higher eukaryotes. Only higher eukaryotes have GAR domains at both the N- and C- terminus of Gar1 protein, and these domains are absent in archaeal Gar1, which indicates that these domains might have a unique role for the function of H/ACA snoRNPs in eukaryotes. The truncation of yeast Gar1 used in this study starts and ends with amino acids highlighted in red color boxes.

Appendix 1

Table A1. The predicted hinge length of yeast H/ACA snoRNAs. All 29 H/ACA snoRNAs are divided into 3 classes depending on the range of hinge region length without including the H-Box region. The class I contains the H/ACA snoRNAs with 4-7 nt hinge length. Class II contains H/ACA snoRNA with an 8-20 nt range hinge length. Class III has a 21-66 nt range of hinge length.

Class-I is 4-7 nt as a hinge

Guide H/ACA snoRNA	Target rRNA	Potential Hinge length (nt) without H box	Number of potential internal helices near hinge	Functional hairpin	Position of the modification
snR10	25S rRNA	7	NA	3'	U2923
snR31	18S rRNA	6		3'	U999
snR32	25S rRNA	6		3'	U2191
snR36	18S rRNA	4	Long helices in 5’ hairpin	3'	U1187
snR80	18S, 25S rRNA	7		Both hairpins	U759, U776
snR81	25S rRNA, U2	5		Both hairpins	U1052, U42
snR86	25S rRNA	4	Has a long extremely folded 5’ hairpin	3'	U2314

Class-II with 8-20 nt as hinge range

Guide H/ACA snoRNA	Target rRNA	Potential Hinge length (nt) without H-box	Number of potential internal helices near hinge	Functional hairpin	Position of the modification
snR3	25S rRNA	11	NA	5' and 3'	U2129, U2133, U2264
snR5	25S rRNA	12	NA	5' and 3'	U1004, U1124
snR8	25S rRNA	12	NA	5' and 3'	U960, U986
snR9	25S rRNA	18	1 (long helices is possible)	5'	U2340
snR11	25S rRNA	11	NA	5'	U2416
snR35	18S rRNA	18		5'	U1191

snR33	25S rRNA	15		3'	U1042
snR37	25S rRNA	10		5'	U2944
snR44	18S, 25S rRNA	10		5' and 3'	U106, U1056
snR49	18S, 25S rRNA	16	1	5' and 3'	U120, U211, U302 U990
snR82	25S rRNA	10		5' and 3'	U1110, U2349, U2351
snR83	18S rRNA	14		5' and 3'	U1290, U1415
snR161	18S rRNA	13		5'and 3'	U632, U766
snR85	18S rRNA	16		5'	U1181
snR189	18S, 25S rRNA	11		5' and 3'	U466 U2735
snR191	25S rRNA	10	With extra hairpin or two internal helices	5' and 3'	U2258, U2260

Class-III with 21-66 nt as hinge region length

Guide H/ACA snoRNA	Target rRNA	Potential Hinge length (nt) without H-box	Number of potential internal helices near hinge	Functional hairpin	Position of the modification
snR34	25S rRNA	33		5' and 3'	U2826, U2880
snR42	25S rRNA	38	1 (long helices in hinge)	3'	U2975
snR43	5.8S, 25S rRNA	66	Has an extra hairpin	5'	U73, U966

snR46	25S rRNA	35	1	5'	U2865
snR84	25S rRNA	21		5'	U2266