THE BIOCHEMICAL CHARACTERIZATION OF *SACCHAROMYCES CEREVISIAE* H/ACA SMALL NUCLEOLAR RIBONUCLEOPROTEINS

JESSICA DURAND

B.Sc. BIOCHEMISTRY, UNIVERSITY OF LETHBRIDGE, 2008

A Thesis

Submitted to the School of Graduate Studies

of the University of Lethbridge

in Partial Fulfilment of the

Requirements for the Degree

MASTER OF SCIENCE

Department of Chemistry and Biochemistry

University of Lethbridge

LETHBRIDGE, ALBERTA, CANADA

Abstract

Ribosome biogenesis is a crucial yet poorly understood and complex process in all cells. To date, most studies on eukaryotic ribosome biogenesis have relied on yeast genetics and whole cell analysis of ribosomal RNA processing. An early and critical step in ribosome biogenesis is the post-transcriptional modification of rRNA. Pseudouridylation is the most frequently occurring modification. Pseudouridylation is catalyzed by H/ACA small nucleolar ribonucleoproteins (snoRNPs) which are one of the two major classes of snoRNPs found within eukaryotes and archaea. H/ACA snoRNPs consist of four conserved core proteins Cbf5, Gar1, Nop10, and Nhp2 (eukaryotes), and a substrate specific H/ACA snoRNA. Mutations causing the rare inherited disease Dyskeratosis congenita are found in *CBF5, NOP10,* and *NHP2*. Here I report the purification of H/ACA protein Cbf5 in the presence of detergents. Additionally, I report initial *in vitro* RNA binding studies using Nhp2 and the snoRNA snR34 as well as the effects of Dyskeratosis congenita substitutions within Nhp2 on this interaction.

Table of Contents

List of Tables

List of Figures

List of abreviations

ATP	adenosine-5'-triphosphate
BSA	bovine serum albumin
C-stem	canonical-stem
DC	Dyskeratosis congenita
ddH_2O	double distilled water
DNase	deoxyribonuclease
dNTP	deoxynucleoside triphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
ETS	external transcribed spacer
GMP	guanosine-5'-monophosphate
GST	glutathione *S*-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMG	high mobility group
HSP	heat shock protein
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
ITS	internal transcribed spacer
KOAc	potassium acetate
LB	Luria-Bertani Medium
MWCO	molecular weight cut off
MWM	molecular weight marker
NC-stem	non-canonical-stem
NTP	nucleoside triphosphate
OD	optical density
PAGE	polyacrylamide gel electrophoresis

PCR	polymerase chain reaction
PDB	protein data base
Pfu	*Pyrococcus furiosus*
PMSF	phenylmethylsulfonyl fluoride
PTC	peptidyltransferase center
PUA	pseudouridine synthase and archaeosine transglycosylase
RNase	ribonuclease
RNP	ribonucleoprotein
rRNA	ribosomal RNA
SAP	shrimp alkaline phosphatase
SDS	sodium dodecyl sulfate
sn	small nuclear
sno	small nucleolar
TERC	telomerase RNA
tRNA	transfer RNA
X-Gal	5-bromo-4-chloro-3-indolyl β-D-galactopyranoside

I Literature Review

I.I Ribosome Biogenesis

One of the most critical and energy consuming tasks encountered by any eukaryotic cell is the synthesis and assembly of ribosomes. Occurring successively within the nucleolus, nucleoplasm, and cytoplasm, ribosome biogenesis involves more than eighty ribosomal proteins, four ribosomal RNAs (rRNA), more than 150 transient accessory proteins and between sixty and two hundred small nucleolar RNAs (snoRNA). The eukaryotic 80S ribosome is composed of a small 40S subunit and a large 60S subunit. The 40S subunit consists of thirty to fifty ribosomal proteins and one rRNA, the 18S rRNA [1]. The 60S subunit is made up of forty to fifty ribosomal proteins and three rRNAs, the 5S, 5.8S, and 25S rRNAs [1].

Ribosome biogenesis begins in the nucleolus with the transcription of a single large pre-rRNA from tandemly arranged rRNA transcriptional units [2]. Contained within this pre-rRNA transcript are the sequences for the mature 18S, 5.8S and 25S rRNAs. However, the 5S rRNA is not transcribed from this pre-rRNA transcript, but is transcribed independently [2]. The mature rRNA sequences are separated by internal transcribed spacers (ITS1 and ITS2) and external transcribed spacers (5' ETS and 3' ETS), which are removed through a series of endonucleolytic and exonucleolytic cleavage events. In total, ten common cleavage events occur co-transcriptionally or rapidly after pre-rRNA transcription. The general yeast pre-rRNA cleavage pathway is outlined in Figure 1. The 3' ETS is the first to be removed [3]. Following the removal of the 3' ETS, the pre-rRNA is cleaved at the A0 and A1 sites removing the 5' ETS producing the 32S precursor transcript [4]. Cleavage at the A2 site within ITS1 separates the rRNA subunits forming the 20S and 27S precursor transcripts [4]. The mature 18S rRNA of the 40S ribosomal subunit is produced through the cleavage of the 20S precursor at the D site [5]. The mature 5.8S and 25S rRNAs of the 60S ribosomal subunit are formed by the endonucleolytic cleavage of ITS2 [4].

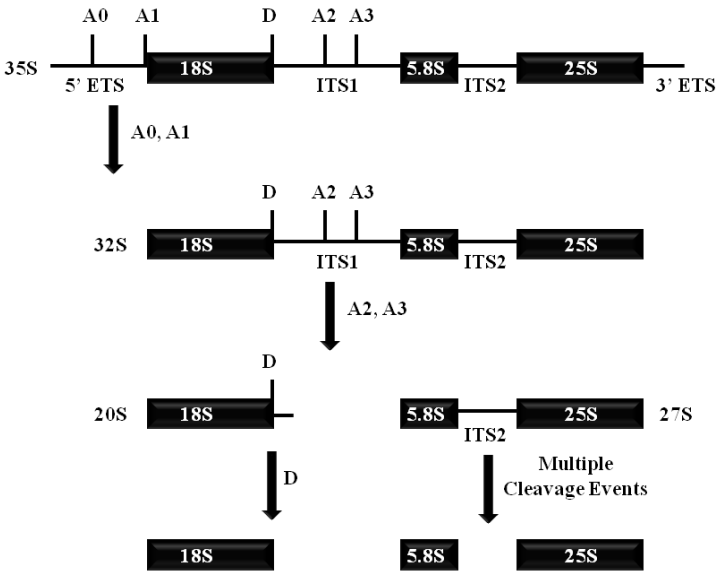


Figure 1: General pre-rRNA yeast processing pathway. The 18S, 5.8S, and 25S mature rRNA transcripts are processed from the 35S pre-rRNA transcript through the removal of internal and external transcribed spacer regions (adapted from [2-3, 5]).

During processing, pre-rRNA transcripts undergo extensive covalent chemical modifications of nucleotides within the mature rRNA sequences, occurring in functionally important regions of the ribosome. The most abundantly occurring chemical modifications are the isomerisation of uridines to pseudouridines and the 2'-O-methylation of riboses. Within mammalian rRNA, over two hundred nucleotides are modified producing approximately one hundred 2'-O-methylated riboses and approximately ninety five pseudouridines, while fifty five 2'-O-methylated riboses and twenty three pseudouridines can be found in *Saccharomyces cerevisiae* rRNA [6]. Two groups of small nucleolar ribonucleoproteins (snoRNPs) are responsible for these modifications in eukaryotes, C/D snoRNPs and H/ACA snoRNPs. C/D snoRNPs carry out the 2'-O-methylation of riboses while H/ACA snoRNPs are responsible for the isomerisation of uridines into pseudouridines [7].

Pseudouridines cluster in functionally important regions of the ribosome. Many pseudouridines have been found in the A and P site of the peptidyltransferase centre (PTC) on the

large ribosomal subunit whereas no clustering of pseudouridine residues occurs on the small ribosomal subunit [8].

I.II Pseudouridine

Pseudouridine is the most abundantly modified base found in most rRNAs, tRNAs, and small nuclear RNAs (snRNA). Post-transcriptional modifications of pre-RNA transcripts alter the chemical and structural properties of the RNA thus contributing to the function of the RNA. Pseudouridine is an isomer of uridine possessing a carbon-carbon rather than a nitrogen-carbon glycosidic bond, attaching the ribose to the uracil (Fig. 2).

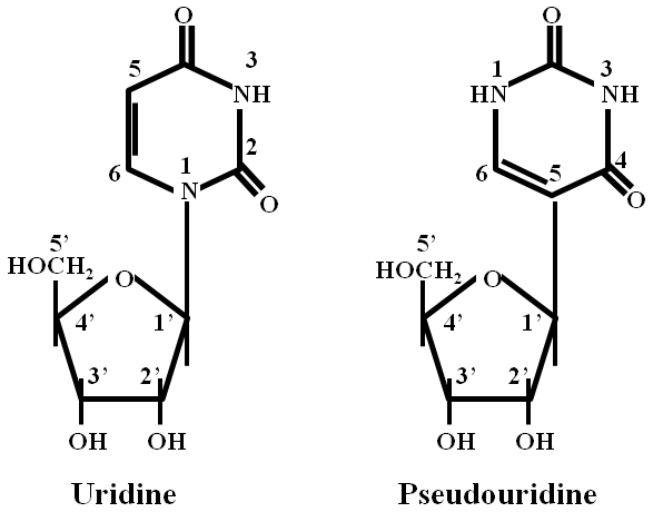


Figure 2: Structures of uridine and pseudouridine. Following the isomerisation of uridine to pseudouridine, uracil is attached to the ribose by a carbon-carbon glycosidic bond instead of a carbon-nitrogen bond.

The altered structure of pseudouridines compared to uridines introduces an extra free amino group which can act as an additional hydrogen bond donor [9]. As a free nucleotide, pseudouridine adopts a *syn* glycosyl conformation while an *anti* conformation is adopted by pseudouridines found within oligonucleotides and double helices [10-11]. In the *anti* conformation, pseudouridines confer increased rigidity [9-11]. The *anti* conformation of pseudouridines within RNA offers an appropriate geometry and distance for the formation of a

water bridge between the pseudouridine base and the backbone of the preceding residue, coordinating the water molecule between pseudouridines N1-H and the 5'-phosphates [12]. The presence of a water bridge in turn restricts the conformation and mobility of the backbone preceding the pseudouridines [10]. The formation of a water bridge may also enhance the local stacking of single-stranded and double-stranded regions of RNA by rigidifying the local phosphodiester backbone thus increasing the stacking of neighbouring nucleosides, in turn improving the stability of the RNA structure [10].

I.III Proposed Pseudouridylation Mechanisms

Pseudouridine synthases are responsible for the isomerisation of uridines to pseudouridines within archaea, bacteria, and eukaryotes. Pseudouridine synthases can be classified into five families: RluA, RsuA, TruA, TruB, and TruD. Structural analysis of the five pseudouridine synthase families has shown a common core domain and a highly conserved active-site cleft containing a universally conserved catalytic aspartate residue [13]. The *Escherichia coli* TruB protein is responsible for the pseudouridylation of uridine 55 within the T- loop of elongator tRNAs [14]. Structural analysis of TruB and the H/ACA small nucleolar ribonucleoprotein (snoRNP) complex component Cbf5 demonstrated that both share a similar structure, indicating that Cbf5 is a member of the TruB pseudouridine synthase family [15]. While Cbf5 possesses stand-alone enzymatic abilities, it requires the presence of a substrate specific guide RNA and three additional structural proteins to achieve optimal enzymatic activity.

The conversion of target uridines to pseudouridines involves three general steps i) the cleavage of the N-glycosidic bond between the uracil and ribose ii) the placement of carbon 5 of the cleaved uracil residue next to the carbon 1' of the ribose and iii) the formation of the C1'-C5 glycosidic bond. While the exact mechanism of pseudouridine formation remains under debate, two possible mechanisms have been proposed, the Michael mechanism (Fig. 3A) and the acylal mechanism (Fig. 3B).

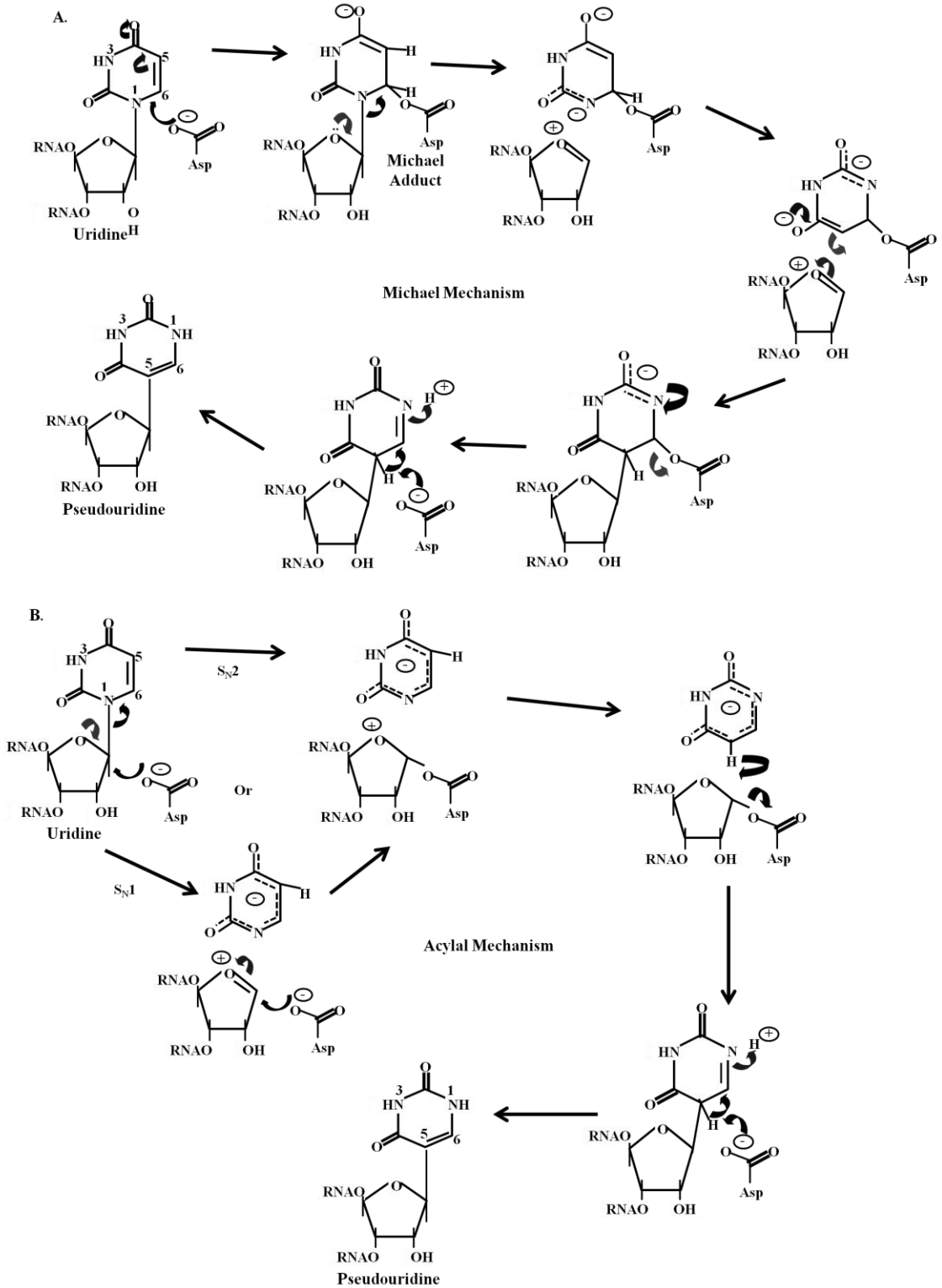

A.
Uridine
Michael Adduct
Michael Mechanism
Pseudouridine
B.
Uridine
Or
Acylal Mechanism
Pseudouridine

Figure 3: Proposed catalytic mechanisms of pseudouridine formation. A. The Michael mechanism, where the catalytic aspartate residue forms a Michael adduct with the uridine (adapted from [18-19]). **B.** The acylal mechanism, where an acylal intermediate is formed by the catalytic aspartate residue (adapted from [18-19]).

The Michael mechanism uses the catalytic aspartate as a nucleophile to form a Michael adduct with the target uracil [16-17]. The Michael addition to the uracil enables the N-C glycosidic bond to be cleaved allowing the uracil ring to rotate 180° around the new ester bond formed with the catalytic aspartate [16-17]. The rotation of the uracil repositions the carbon 5 atom of the uracil near the carbon 1' of the ribose allowing it to form a new C-C glycosidic bond. Subsequent removal of the catalytic aspartate and deprotonation of uracil's carbon 5 completes pseudouridine formation [16-17]. During the acylal mechanism, the catalytic aspartate residue forms an acylal intermediate through the nucleophilic attack of the ribose carbon 1' [20-21]. The resulting uracilate anion rotates to reposition the uracil carbon 5 and the ribose of carbon 1' next to each other thereby allowing subsequent formation of the C-C glycosidic bond and simultaneous removal of the catalytic aspartate [20-21]. Pseudouridine formation is completed with the deprotonation of uracil carbon 5.

I.IV H/ACA Small Nucleolar Ribonucleoproteins

Eukaryotes and archaea use H/ACA snoRNPs in addition to stand-alone pseudouridine synthases to catalyze the sequence-specific pseudouridylation of pre-rRNAs. Identified in the late 1990's, H/ACA snoRNPs are composed of four conserved proteins and a substrate-specific H/ACA snoRNA (Fig. 4) [22-23].

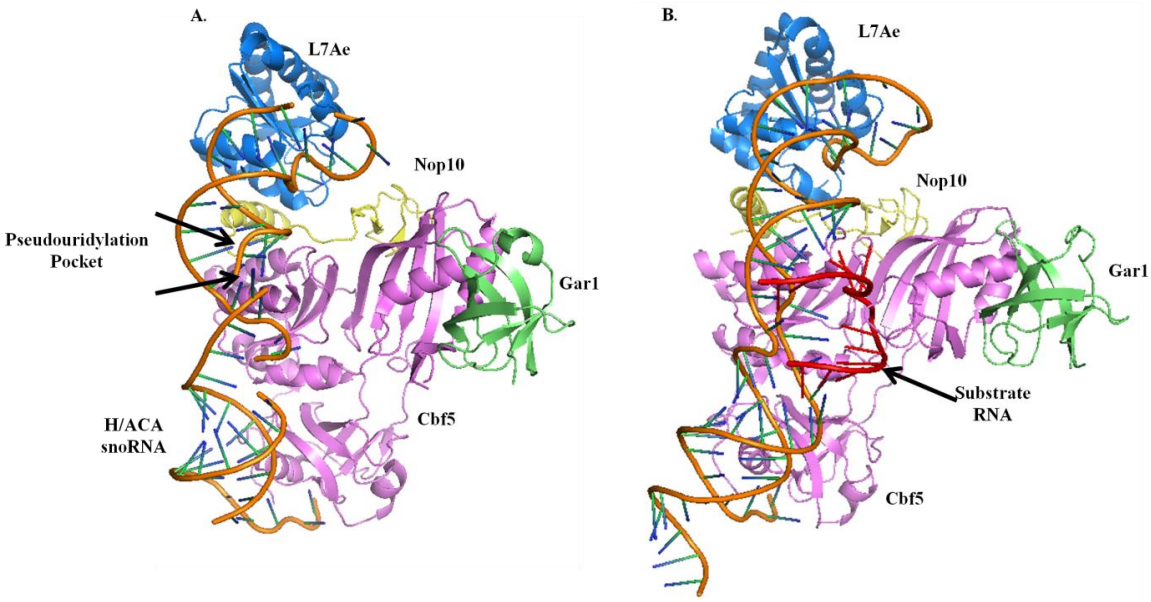


Figure 4: Structure of *Pyrococcus furiosus* H/ACA sRNP complex. **A.** No substrate RNA bound (PDB, 2HVY [24]). **B.** H/ACA sRNP with substrate RNA bound (PBD, 3HAY, [25]). The catalytic protein Cbf5 (purple) and structural protein L7Ae (blue, homolog of eukaryotic protein Nhp2) interact directly with H/ACA sRNA (orange). In addition, Cbf5 also interacts with substrate RNA (red) and associates independently with structural proteins Nop10 (yellow) and Gar1 (green).

H/ACA Small Nucleolar RNAs

The defining feature of all snoRNPs is their snoRNA. Eukaryotic H/ACA snoRNAs fold into a conserved 'hairpin-hinge-hairpin-tail' secondary structure [23]. The single-stranded hinge region and the 3' tail contain the essential Box H (ANANNA) and ACA sequences, respectively. Mutational analysis of the Box H and Box ACA sequence and their spatial arrangement have shown that both elements are required for snoRNA stability, localization, and function [23, 26]. The Box ACA makes critical RNA-protein contacts with the catalytic protein Cbf5's pseudouridine and archaeosine transglycosylase (PUA) domain [24].

Within each hairpin structure is a single-stranded motif, the pseudouridylation pocket, which is complementary to the sequence of the substrate rRNA around the site of pseudouridylation [26]. The specificity of an H/ACA snoRNP complex is determined by the pseudouridylation pocket as its single stranded sequence forms Watson and Crick base pairs with

three to ten nucleotides of the substrate rRNA. The Watson and Crick base pairing between the snoRNA and the substrate rRNA places the target uridine residue within the catalytic pocket of Cbf5 [26].

Found within the upper hairpin of all archaeal and some eukaryotic H/ACA snoRNAs is a kink-loop RNA structural motif, a variation of the kink-turn RNA secondary structure element. Kink-turns commonly consist of a helix – asymmetric internal bulge – helix structure, where the first helical stem, known as the canonical stem (C-stem), ends with two Watson-Crick base pairs at the internal bulge and the second helical stem, the non-canonical stem (NC-stem), follows the internal bulge with two tandem non-Watson-Crick base pairs (usually sheared A-G base pairs) [27]. The asymmetric loop typically contains three unpaired nucleotides on one strand and induces a 120^{o} bend in the phosphodiester backbone [27]. This kinking of the phosphodiester backbone causes the third unpaired nucleotide, commonly a uridine residue, to protrude into solution while A-minor interactions between the two helical stems stabilize the 120^{o} bend [27]. This secondary structure motif is recognized and bound by the archaeal H/ACA structural protein L7Ae. Unlike the archetypical kink-turn, kink-loop structural motifs contain a terminal stem loop with four to seven nucleotides which is incapable of forming both Watson-Crick base pairs and A-minor interactions, but which is flanked by two sheared A-G base pairs and can extrude a nucleotide in the loop similar to kink-turn structures [28].

H/ACA Protein Core

H/ACA snoRNAs associate with four eukaryotic highly conserved proteins: Cbf5 (dyskerin in humans, and NAP57 in mice), Gar1, Nhp2 (L7Ae in archaea), and Nop10 [14, 22-23, 29]. Depletion studies of all four proteins have shown that all H/ACA proteins with the exception of Gar1 are essential for complex viability and stability [29-32].

Cbf5 is the catalytic enzyme of the complex, sharing significant sequence similarity with the *E. coli* pseudouridine synthase TruB [33-34]. Depletion of Cbf5 results in the loss of pseudouridine formation and in the decrease of H/ACA snoRNP accumulation [14]. Cbf5 and TruB share a large N-terminal catalytic domain and a C-terminal PUA domain [35-36]. The N-terminal catalytic domain can be divided into two subdomains, D1 and D2, forming the catalytic pocket between them [24]. Within the catalytic pocket lies the universally conserved catalytic aspartate and two highly conserved polar residues [13, 21, 37]. Cbf5 interacts with two of the three H/ACA structural proteins (Nop10 and Gar1) in addition to the snoRNA. Cbf5's D1 and D2 subdomains make several contacts with structural protein Nop10 which presumably helps to stabilize Cbf5 in its catalytically active conformation, and which contributes to the structural stability of H/ACA snoRNPs [35]. Cbf5's D2 subdomain also makes critical contacts with structural protein Gar1 which has been proposed to assist in the loading and turnover of substrate rRNA [25].

Nop10 is a small two-domain protein consisting of an N-terminal β-domain and a C-terminal α-helix domain [35]. The two domains of Nop10 are connected by a highly conserved, unstructured linker region. On its own, Nop10 is a disordered protein, however Nop10 becomes highly structured upon binding to Cbf5's catalytic domain [15].

Eukaryotic H/ACA structural protein, Nhp2, has a similar structure to its archaeal homolog L7Ae and to its eukaryotic homolog snu13p/15.5K which are monomeric proteins folding into a single globular domain [38]. Accordingly, Nhp2 is a member of the L7Ae/L30 RNA-binding protein family. These proteins specifically recognize kink-turn RNA structural motifs. However, Nhp2 is the only member of this protein family which does not interact with kink-turn and kink-loop structures, binding double-stranded RNA instead [39-40]. In contrast, Nhp2's archaeal homolog L7Ae can interact with kink-turn and kink-loop structural motifs with high specificity. The Maxwell group demonstrated that replacement of L7Ae's signature residues,

responsible for kink-turn recognition, with those of Nhp2 resulted in the loss of L7Ae kink-loop binding and a reduction in L7Ae kink-turn binding [40].

H/ACA structural protein Gar1 is a member of the reductase, isomerase and elongation factor fold superfamily, folding into the characteristic six-stranded β-barrel structure. Gar1 interacts with the catalytic domain of Cbf5 without having direct contacts to the H/ACA RNA. Gar1 has been suggested to play a role in the recruitment and release of substrate rRNA through its interaction with Cbf5's thumb loop [25].

I.V Assembly of H/ACA Small Nucleolar Ribonucleoproteins

Stable accumulation of mature eukaryotic H/ACA snoRNPs requires the co-transcriptional assembly of H/ACA snoRNAs with H/ACA snoRNP core proteins with the help of additional assembly factors [41]. Together the assembly factors and the core proteins prevent the degradation of the snoRNA and promote the stepwise assembly of the mature H/ACA snoRNP complex. Within eukaryotic organisms, two assembly factors, Shq1 and Naf1, appear to be required for the correct assembly and accumulation of mature H/ACA snoRNPs. Both Shq1 and Naf1 interact with Cbf5 during assembly.

Shq1 Assembly Factor

Shq1 is a modular protein consisting of two independent domains, an N-terminal HSP20-like domain and a C-terminal Shq1 domain. The heat shock protein (HSP) 20-like domain of Shq1 is also found in numerous co-chaperones for HSP90. Therefore, Shq1 may either act as an Hsp90-independent snoRNP-specific chaperone, or it may specifically recruit Cbf5 to the Hsp90 chaperone system [42-43]. Shq1 is involved in the initial steps of H/ACA biogenesis interacting only with free Cbf5 prior to Naf1 association [44]. Shq1 appears to function as a Cbf5 chaperone stabilizing free Cbf5 and potentially assisting in the folding of the protein while following Cbf5 from synthesis to its association with Nop10, Nhp2, and Naf1 [44].

Naf1 Assembly Factor

The structure of Naf1 most closely resembles that of the reductase, isomerase and elongation factor folds superfamily sharing the six-stranded β-barrel secondary structure common among members of this family [45]. H/ACA structural protein Gar1 is also a member of this superfamily, containing the same characteristic secondary and tertiary structure. Naf1 also shares a conserved Cbf5 binding surface with Gar1 [45]. The structural similarities found between Naf1 and Gar1 are the only commonalities shared between the two proteins as their functional roles are very different. While Gar1 is essential for H/ACA snoRNP function, it is the only H/ACA protein not required in the biogenesis of H/ACA snoRNPs [46]. Naf1 however is essential for the accumulation of H/ACA snoRNAs and the stable association of Cbf5 with the H/ACA snoRNA transcripts, but has no role in H/ACA snoRNP function [47-48]. Based on this information, it has been suggested that Naf1 and Gar1 each interact with Cbf5 in a mutually exclusive manner where Naf1 associates with the D2 subdomain of Cbf5 similarly to Gar1 [47]. Subsequently, the Naf1-Cbf5 complex is co-transcriptionally recruited to the H/ACA snoRNA transcript along with H/ACA structural proteins Nop10 and Nhp2 [47]. Finally, Naf1 is replaced by Gar1 forming a mature H/ACA snoRNP complex.

I.VI Substrate binding by H/ACA Small Nucleolar Ribonucleoproteins

Substrate RNA containing the target uridine for modification is recognized by complementary sequences found within the H/ACA snoRNA's internal pseudouridylation pocket, forming two substrate-guide helices [49-50]. The lower and upper stems of the H/ACA snoRNA are anchored by Cbf5's PUA domain and L7Ae/Nhp2 placing the substrate-guide RNA complex near the catalytic cleft of Cbf5 [51]. Correct placement of the substrate RNA within the catalytic cleft of Cbf5 is also highly dependent on the presence of H/ACA structural protein Nhp2 (archaeal L7Ae). Gar1 association with Cbf5 influences the conformation of Cbf5's thumb loop which is critical for substrate loading and release [25]. The thumb loop of Cbf5 pinches the

substrate RNA at the active site ensuring the correct placement of the target uridine within the catalytic cleft of Cbf5 [25]. Once in contact with Cbf5, the substrate RNA makes extensive interactions with Cbf5's polar catalytic cleft [25].

I.VII Functional Importance of H/ACA Small Nucleolar Ribonucleoproteins

Ribosome Biogenesis

During ribosome biogenesis, H/ACA snoRNAs have two primary functions, pre-rRNA modification through pseudouridylation and rRNA processing. As previously discussed, the isomerisation of specific uridine residues to pseudouridines has a profound effect on the chemical and structural properties of mature rRNA. While the majority of H/ACA snoRNAs function in the formation of pseudouridines, a few H/ACA snoRNAs function in rRNA processing through the nucleolytic cleavage of rRNA. To date, two H/ACA snoRNAs have been shown to be involved in rRNA processing, snR30 and snR10.

H/ACA snoRNA snR30 is essential for the correct processing and accumulation of 18S rRNA, and is required for correct cleavage of the 35S pre-rRNA transcript at the A0, A1 and A2 cleavage sites [52]. H/ACA snoRNA snR30 shares the common consensus hairpin-hinge-hairpin secondary structure with all H/ACA snoRNAs [53]. Interestingly all snR30 snoRNAs carry an additional internal hairpin upstream of the 3'-terminal hairpin [53]. Deletion analysis has demonstrated that this internal hairpin and the 5'-terminal hairpin are not essential for cell viability nor do they contain any functionally pertinent elements [53]. Similar analysis of the 3'-terminal hairpin revealed that this structural element of snR30 carries all of the functionally important elements needed for cell viability [53]. Contained within the lower region of the internal loop of the 3'-terminal hairpin are two highly conserved sequence motifs, m1 (AUAUUCCUA) and m2 (AAACCAU), which form transient Watson-Crick interactions with two short internal sequence elements within the 18S rRNA [53-54]. Hence, snR30s m1 and m2

structural motifs are 18S rRNA recognition elements interacting with two distantly located 18S rRNA internal sequence elements which are folded together by a stem-loop structure conserved in all 18S rRNAs [53-54]. This Watson-Crick interaction between snR30 and the 18S rRNA is essential for 18S rRNA production and cell viability. Alteration or complete deletion of either the m1 or m2 motif results in different rRNA processing pathways and in certain cases loss of cell viability [53]. The terminal stem-loop region of snR30's 3'-terminal hairpin functions as a snoRNP protein binding site which targets essential snR30-asscoiated processing factors to the pre-rRNA transcript [54].

H/ACA snoRNA snR10 also plays an important role in yeast rRNA processing and the accumulation of 18S rRNA. Unlike snR30, snR10 is not essential for cell viability as deletion of snR10 results in modest reductions in cellular growth rates and a slight cold-sensitive phenotype [55]. In addition to its role in rRNA processing, snR10 is also responsible for the pseudouridylation of uridine 2923 in the PTC region of the 25S rRNA [26]. Like all other H/ACA snoRNAs, snR10 folds into the hairpin-hinge-hairpin consensus secondary structure with the functional elements required for rRNA processing located exclusively in the 5'-terminal hairpin and the functional elements required for pseudouridylation mapped to the 3'-terminal hairpin [56]. The rRNA processing domain and pseudouridylation domain of snR10 function cooperatively in which a C-insertion mutation within the pseudouridylation pocket affects not only pseudouridylation but also impairs 18S rRNA processing [56]. The dual effect of such a mutation may be suggestive of snR10's role as a factor bringing together or stabilizing the precursor forms of the ribosomal subunits [56].

Pre-mRNA Splicing

The protein coding sequence of eukaryotic genes is disrupted by a series of interspersed non-coding sequences known as introns. Following transcription, the intervening sequences are

removed during mRNA splicing. Splicing is performed by spliceosomes, large RNP complexes consisting of several small nuclear RNP (snRNP) subunits. There are five snRNP subunits in total, each defined by a small nuclear RNA (snRNA) (U1, U2, U4, U5 and U6) in complex with several protein subunits [57]. The individual snRNP subunits assemble on a pre-mRNA transcript forming the spliceosome with portions of the spliceosome entering and leaving as the splicing reaction proceeds [57]. Throughout the splicing reaction important regions of the pre-mRNA are recognized by the snRNA through the formation of base-pairing between the snRNA and the pre-mRNA intron consensus sequences.

All five of the spliceosomal snRNAs are post-transcriptionally modified by pseudouridylation and 2'-O-methylation [58]. Again similar to rRNA, all modifications cluster in functionally important regions of the spliceosome and are involved in numerous intermolecular interactions. Of the five spliceosomal snRNAs, U2 is the most extensively modified snRNA containing ten 2'-O-methylated residues and thirteen pseudouridines [59]. The single stranded branch site recognition region of U2 snRNA contains six uridines which are all converted to pseudouridines [60]. One particular pseudouridine within the branch site recognition region is essential for the bulging out of the branch point adenosine nucleotide in the first step of RNA splicing [61-62]. The presence of this pseudouridine enhances the formation of the phosphodiester bonds during catalysis [63]. All of the pseudouridylation modifications of U2 snRNA are required for snRNP assembly and RNA splicing [64-65]. The presence of pseudouridines within the spliceosomal RNA appears to provide stabilization to the snRNP complex during the formation of the bulge at the branch point adenosine [64].

Telomere Maintenance

One vital step of the eukaryotic cell cycle is the duplication of chromosomes. The DNA polymerase however cannot completely replicate the chromosome ends resulting in incomplete

replication [57]. To prevent chromosome deterioration and the loss of critical genetic information during mitosis, regions of repetitive DNA repeats known as telomeres are found at the end of each chromosome [57]. Telomeres protect chromosomal ends by providing a type of disposable buffer which is deteriorated during cellular division [57]. To ensure that the telomere ends of chromosomes are never completely lost, telomeres of cells in the germ line and of some stem cells are maintained and replenished by telomerase reverse transcriptase [57].

Telomerase is a ribonucleoprotein particle consisting of telomerase RNA (TERC) and telomerase reverse transcriptase. The telomerase RNA contains the template for the synthesis of the telomeric repeats. While the size and sequence composition of telomerase RNA varies significantly among different species, telomerase RNAs share a common core secondary structure [66]. All telomerase RNAs share a 5' template boundary element and a large loop which contains the core pseudoknot domain, the template and a loop-closing helical region [66]. In addition to the common core secondary structure, the 3' end of vertebrate telomerase RNA folds into a characteristic H/ACA snoRNA structure [67]. This 3' H/ACA domain found within telomerase RNA contains both the H and ACA consensus sequences found in other H/ACA snoRNAs [67]. The H and ACA consensus sequences are important for the stability of telomerase RNA as alteration of either sequence results in reduced levels of telomerase RNA [67]. The telomerase RNA H/ACA domain also associates with H/ACA core proteins Cbf5, Nop10, and Nhp2 suggesting that H/ACA core proteins interact with the H/ACA domain of telomerase RNA in a similar manner as they assemble with H/ACA snoRNA [67-68]. A model explaining the role of H/ACA RNP proteins in 3'TERC formation and mature TERC accumulation was proposed by the Henry group: presumably the 3' H/ACA domain of telomerase RNA acts as a barrier for 3'→5' exonuculease, stopping its advancement during telomerase RNA maturation [30]. This in turn would prevent the turnover of telomerase RNA and allow for the development and maturation of the telomerase RNA 3'-end [30].

I.VIII Dyskeratosis Congenita

The rare inherited disease Dyskeratosis congenita (DC) is a clinically and genetically heterogeneous disease affecting multiple systems. DC patients typically share characteristic hyperpigmentation along the neck, shoulders, and face in addition to developing nail dystrophy and mucosal leukoplakia [69]. Bone marrow failure arises in the vast majority of DC patients frequently leading to death through increased susceptibility to infection and haemorrhaging [69-70]. DC has three known inherited forms: X-linked recessive, autosomal recessive and autosomal dominant [69].

The X-linked recessive form of DC is the most common and severe form of DC [69-71]. The X-linked recessive form of DC is primarily caused by missense mutations in the H/ACA catalytic protein dyskerin (of which Cbf5 is the yeast homolog) [72-73]. Dyskerin DC substitutions are primarily found within the PUA domain with additional substitutions localized to the N-terminal extension, resulting in a spatial cluster along one side of the PUA domain [15, 24, 71]. The autosomal recessive form of DC is the most genetically heterogeneous form, closely resembling X-linked recessive DC with the onset of symptoms occurring in the later stages of life [74]. Autosomal recessive DC has been linked to mutations in the genes of the H/ACA structural proteins Nhp2 and Nop10 [74-76]. Autosomal dominant DC is the result of point mutations within the RNA component of the telomerase. In general, autosomal dominant patients exhibit milder forms of DC, often possessing asymptomatic carriers [77-78]. Autosomal dominant DC is a prime example of disease anticipation, where the severity of the TERC mutations increases in later generations with the onset of symptoms and death occurring at earlier ages with each generation [78].

While research continues to bring forth increasing information on the molecular cause(s) and effects of DC, a comprehensive understanding of the disease is still lacking, leaving many unanswered questions open for discussion. One particular question that remains is whether DC is

a disease of ribosome biogenesis or telomere maintenance. DC patients are often diagnosed with critically short telomeres suggesting that DC is in fact a disease of telomerase [77]. A mouse model of DC based on the reduction of Dkc1 expression displayed no detectable changes in telomere length in early generations of the hypomorphic mice [79]. However, a marked reduction in pseudouridine formation within the 28S and 18S rRNA was observed in addition to an increased occurrence of immature pre-rRNA intermediates [79]. This would suggest that DC affects ribosome biogenesis. The fact that NAP57 interacts with both TERC and H/ACA snoRNA raises the possibility that DC may be a disease of ribosome biogenesis as well as telomere maintenance.

II

H/ACA snoRNP Structural Protein Nhp2

II.I Introduction

Nhp2, a Member of the L30 Protein Family

S. cerevisiae H/ACA snoRNP structural protein Nhp2 was first identified as a high mobility group (HMG)-like nuclear protein, sharing similar physical and chemical properties with other HMG proteins [80]. However, sequence analysis later showed that Nhp2 shared no sequence similarities to other known HMG proteins, but did in fact share significant sequence similarity with the RNA binding protein family L7Ae/L30 [81]. Members of the L7Ae/L30 protein family specifically recognize kink-turn structural RNA motifs.

Members of the L7Ae/L30 protein family contain a highly conserved RNA-binding region which directly interacts with the kink-turn motif. Primary sequence analysis of the L7Ae/L30 protein family revealed five signature amino acid residues which are variable and unique to each member, appearing to confer RNA recognition specificity exclusive to each member [40]. These five signature amino acid residues flank the RNA-binding region. Four of the five residues are found within the highly conserved C-terminal loop 9 region, positioning them in close vicinity to the RNA kink-turn motif [40]. The L7Ae/L30 protein family contains two members, archaeal protein L7Ae, and eukaryotic protein Nhp2, that have unique properties. L7Ae is the only protein which can interact with both kink-turn and kink-loop structures, while Nhp2 is the only member which has been reported not to interact with kink-turn and kink-loop structures [39-40]. Interestingly, both are structural proteins in H/ACA snoRNP complexes, with L7Ae replacing Nhp2 in the archaeal system. Replacement of L7Ae's signature RNA recognition residues flanking the N- and C-terminal sides of the protein's RNA recognition region with those of Nhp2 resulted in loss of L7Ae kink-loop binding and a reduction in L7Ae kink-turn binding, further supporting the idea that Nhp2 differs significantly from its homologues [40].

Nhp2 is a Eukaryotic H/ACA snoRNP Structural Protein

During early characterizations of H/ACA snoRNPs, Nhp2 was shown to be a component of eukaryotic H/ACA snoRNPs. Nhp2 has the ability to associate specifically with known H/ACA snoRNAs and was observed to be essential for H/ACA snoRNP function [29]. The Henry group demonstrated that purified *S. cerevisiae* Nhp2p could directly interact with *in vitro* transcribed H/ACA snoRNA, snR36, with a high affinity (K_D ~ 4 nM) using electrophoretic mobility shift assays (EMSA). Interestingly, multiple shifted bands were observed, suggesting that Nhp2 has more than one binding site [39]. This observation is in good agreement with the hypothesis that two sets of H/ACA core proteins associate with both hairpin structures of bipartite snoRNA in eukaryotes. However, most of what is currently known about eukaryotic H/ACA snoRNPs and Nhp2 comes from the work done on the archaeal H/ACA structural protein L7Ae (Fig. 5). Structural analysis and interaction studies involving L7Ae and Nhp2 strongly suggest that both proteins share a common function within H/ACA snoRNPs as similar protein-protein and protein-RNA interactions have been documented [24, 28, 39, 68, 82-83].

L7Ae is a monomeric protein which folds into a single globular domain [38]. Alternating α helices and β strands form an α-β-α sandwich structure, in which the central β-sheets contain three anti-parallel strands and one parallel strand while three α-helices pack along one side of the β-sheet and two other α-helices pack along the other side of the β-sheet [38]. Additionally, the recently determined structure of *C. parvum* Nhp2 shows the same folds (Fig. 5) [84].

L7Ae contains a significant number of positively charged residues. Electrostatic potential calculations have shown a positively charged region near the surface of the protein involving the solvent accessible regions of the α2 helix, the β1 strand, and the loops connecting α1-β1, α2-β2, α3-β3, and β4-α5 [38]. Structural analysis of L7Ae-RNA complexes show L7Ae interacting with

the RNA primarily through the α2 helix, β1 strand, and the β1-α2, β2-α3, and α4-β4 loops [28, 83].

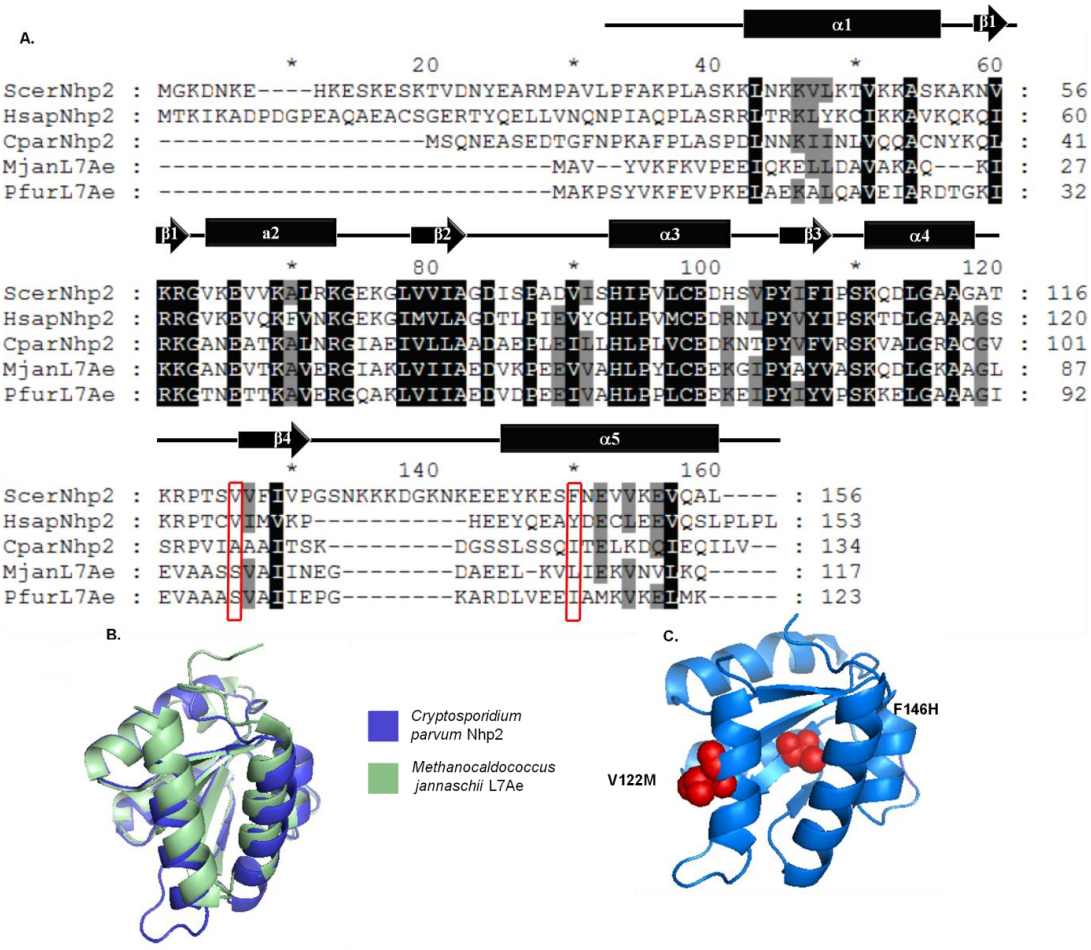


Figure 5: Sequence alignment and structural comparison of H/ACA structural proteins Nhp2 and L7Ae, and location of Dyskeratosis congenita substitutions within Nhp2. A. Multiple sequence alignment of eukaryotic Nhp2 and archaeal L7Ae. Of the five aligned sequences three are eukaryotic: *Saccharomyces cerevisiae* (ScerNhp2), *Homo sapiens* (HsapNhp2), and *Cryptosporidium parvum* (CparNhp2), and two are archaeal: *Methanocaldococcus jannaschii* (MjanL7Ae) and *Pyrococcus furiosus* (PfurL7Ae). The secondary structure elements are shown above the sequences, with α-helices depicted as rectangles and β-strands as arrows. The secondary structural elements were determined from the *Pyrococcus furiosus* H/ACA snoRNP structure [24]. Completely conserved residues among the aligned sequences are shaded in black, and residues with at least 60% conservation are shown in grey. Two positions in Nhp2 which are changed in Dyskeratosis congenita patients are outlined by red boxes. The alignment was prepared using ClustalW (EMBL-EBI) and Gene Doc (Dr. Nicholas Hugh). **B.** Superposition of *C. parvum* Nhp2 (blue, PDB 2AIF [84]) and *M. jannaschii* L7Ae (green, PDB 1XBI [85]) crystal structures. The structural alignment was made with Pymol Molecular Visualization Evaluation Product (Delano Scientific LLC). **C.** The crystal structure of *C. parvum* Nhp2 (blue, PDB 2AIF [84]) was used to model the *S. cerevisiae* DC substitution V122M (V126M in humans) and F146H (Y139H in humans) isolated in autosomal recessive DC patients.

The association of L7Ae with the kink-loop RNA motif in archaeal H/ACA snRNA anchors the upper stem of the guide RNA assisting in the correct placement of the pseudouridylation pocket within the catalytic domain of Cbf5 [24, 86]. The absence of L7Ae binding results in a dramatic shift in the location of the guide RNA relative to the catalytic domain of Cbf5, subsequently placing the target uridine approximately 11 Å away from the catalytic aspartate [51, 86].

L7Ae interacts with the characteristic protruding uridine residue and the tandem sheared G-A base pairs of the kink-loop within the upper hairpin of the H/ACA snoRNAs. Mutation of either the protruding uridine residue or of the G-A residues dramatically reduces or abolishes L7Ae binding [87]. L7Ae backbone atoms are predominantly responsible for the recognition of the protruding uridine residue, while side chain atoms are primarily responsible for the recognition of the tandem sheared G-A base pairs [28]. Interestingly, several basic residues making up the RNA binding surface of L7Ae are conserved in Nhp2 [24].

In addition to binding the H/ACA guide RNA, L7Ae makes several interactions with H/ACA structural protein Nop10, binding within the trough of Nop10 between the N-terminal zinc-ribbon and a C-terminal α helix [24]. The α3 helix of L7Ae contacts the α1 helix of Nop10 at a right angle through a salt bridge and three Nop10 aromatic rings stacked above a hydrophobic region of L7Ae [24]. Additionally, L7Ae's β2-α3 loop interacts with the linker region of Nop10 through van der Waals interactions and hydrogen bonding [24]. Residues within both L7Ae and Nop10 involved in this interface are conserved in eukaryotic Nhp2 and Nop10, strongly suggesting that Nhp2 associates with Nop10 in a similar manner as L7Ae [24]. Supporting this hypothesis is the observation that association of Nhp2 to the H/ACA snoRNP complex is dependent on the formation of a Nop10-Cbf5 complex [68, 82].

Dyskeratosis Congenita and Nhp2

As previously mentioned, DC has been linked to four known genes, including *NHP2*. Mutations within *NHP2* result in the autosomal recessive form of DC, the most genetically heterogeneous form of this inherited bone-marrow failure syndrome [76]. To date three Dyskeratosis congenita mutations have been indentified in Nhp2, V126M, Y139H, and X (stop codon) 154R [76].

Valine 126 is involved in the core packing of Nhp2 with Nop10, as it is positioned directly above the α helix which interacts with Nop10 [75]. It has been shown that mutation of this residue obstructs core packing, destabilizing Nhp2's interaction with Nop10. Tyrosine 139 is also located in the interior of the Nhp2 structure based on a model of the Nhp2 structure derived from the L7Ae structure suggesting an important role for Nhp2 protein structure and stability [75]. While tyrosine 139 does not appear to directly contact Nop10, mutation of this residue interferes with Nop10 association as Nhp2 stability may be compromised [75]. The Dragon group observed that Nhp2 containing V126M and Y139H mutations were unable to form the H/ACA protein tetramer complex, strongly suggesting that Nhp2's structural integrity/stability had been compromised [75]. However, the effects of these mutations on RNA binding by Nhp2 have not been studied.

The third mutation X154R, results in a 51 amino acid extension of Nhp2's C-terminus, as the wild-type stop codon of Nhp2 is mutated to an arginine residue [76]. *In vitro* studies have shown that H/ACA protein tetramer formation is not impaired by the X154R mutation, however the possibility of such a mutation affecting RNP association cannot be ruled out as it might affect RNA binding [75].

II.II Objective

The protein Nhp2 is part of the eukaryotic H/ACA small nucleolar ribonucleoprotein complex which introduces pseudouridine modifications in many cellular RNAs. Eukaryotic H/ACA

structural protein Nhp2 is a member of the L7Ae/L30 RNA-binding protein family. Members of the L7Ae/L30 protein family specifically recognize kink-turn RNA structural motifs. Nhp2 is however the only member which does not interact with kink-turn structures. Nhp2's archaeal homolog L7Ae can interact with kink-turn structural motifs with high specificity. Mutations within three of the four H/ACA core proteins, Cbf5, Nop10 and Nhp2 have been linked to the rare inherited disease DC. To date three DC substitutions have been identified in Nhp2: V126M, Y139H, and X154R. These findings raise the questions how Nhp2 recognizes and interacts with H/ACA snoRNAs and what effects the DC substitutions in Nhp2 have on RNA recognition.

To answer these questions, it was my objective to purify *S. cerevisiae* Nhp2 and to *in vitro* transcribe an H/ACA snoRNA, snR34. Furthermore, it was my aim to create and purify two Nhp2 variants, V122M and F146H, corresponding to mutations in human Nhp2 found in DC patients. By using [^{32}P]-labelled RNA in EMSA analysis, we can detect and quantify the interaction of Nhp2 with H/ACA snoRNA. A initial truncation of snR34 was transcribe to begin the identification of the minimal binding site of Nhp2 on snR34. In addition, the initial characterization of the binding affinity, stoichiometry and thermodynamics of the Nhp2-RNA interaction was analyzed by isothermal titration calorimetry (ITC).

II.III Materials and Methods

Recombinant expression of Nhp2

Nhp2 was recombinantly expressed in *E. coli* BL21 (DE3) Rosetta cells (VWR). 100 mL of LB medium (1% (w/v) tryptone, 0.5% (w/v) yeast extract, and 1% (w/v) NaCl, pH 7.0) with 0.1 mg/mL kanamycin was inoculated with BL21 (DE3) Rosetta cells containing *pET28a-ScNhp2* and grown overnight at 37 ^{o}C. The cells were pelleted and resuspended in 50 mL of fresh kanamycin-containing LB medium. 2 L of LB-kanamycin medium was inoculated to an optical density of approximately 0.1 (OD_{600}). Cultures were incubated at 37 ^{o}C while shaking with cell

growth monitored at 600 nm. Recombinant protein expression was induced at an optical density of 0.6 OD_{600} by adding IPTG to a final concentration of 1 mM. The cell density was recorded for three additional hours with cell samples of 1 OD_{600} being taken at times 0 hr, 0.5 hr, 1.0 hr, 2.0 hr, and 3.0 hr after induction. After 3 hr cell cultures were pelleted by centrifugation at 5000 xg and shock frozen with liquid nitrogen and stored at -80 ^{o}C. Cell samples were suspended in 80 µL of cell lysis buffer (0.1 M Tris-HCl pH 8.5 containing 8 M urea) and boiled for 5 minutes at 95 ^{o}C to open cells. Cell samples were analyzed by SDS-PAGE and visualized using Coomassie staining.

Site directed mutagenesis of *Nhp2*

DC point mutation V122M and F146H were introduced into *pET28a-ScNhp2* wild-type plasmid DNA by site directed quick change mutagenesis using the following primer pairs:

Nhp2 V122M F: GACCTACCTCAATGGTCTTTATCGTCCCGGGTAGCAATAAG

R: CTTATTGCTACCCGGGACGATAAAGACCATTGAGGTAGGTC

Nhp2 F146H F: GAAGAATACAAGGAATCTCACAACGAAGTTGTCAAAGAAGTTCAAGC

R: GCTTGAACTTCTTTGACAACTTCGTTGTGAGATTCCTTGTATTCTTC

Plasmid DNA was purified with an EZ-10 Spin Column PCR Products Purification Kit (Bio Basic Inc.), following the manufacturer's instructions. Mutagenesis reactions were carried out in 25 µL volumes with final component concentrations as follows: 1x Pfu buffer with $MgSO_4$, 0.4 mM dNTPs, 0.4 µM primers, 0.12 U/µL Pfu DNA polymerase (Fermentas), and 40 ng/µL *pET28a-ScNhp2* wild-type plasmid DNA. Amplification parameters included an initial denaturation at 95 ^{o}C for 5 min followed by 18 cycles of denaturation at 95 ^{o}C for 45 s, annealing at 45 ^{o}C for 60 s, extension at 68 ^{o}C for 900 s, and a final extension at 68 ^{o}C for 15 min. The mutagenesis product was subsequently incubated with 1µL of DpnI (Fermentas) overnight at 37 ^{o}C. DpnI was inactivated at 65 ^{o}C for 10 min. 5 µL samples of the mutagenesis reaction were taken prior to amplification, following the first amplification cycle, after mutagenesis, and after

DpnI treatment for analysis by agarose gel electrophoresis. Mutagenesis products were analyzed on a 1% agarose gel at 110 V for 1hr and visualized by ethidium bromide staining [88].

pET28a-ScNhp2 V122M and *pET28a-ScNhp2 F146H* were transformed into high-efficiency competent *E. coli* DH5α cells (NEB). Transformation mixtures were plated on LB agar containing 0.1 mg/mL kanamycin. Plates were incubated overnight at 37 °C. *E. coli* colonies were transferred into 5 mL of LB containing 0.1 mg/mL kanamycin and grown overnight at 37 ^{o}C. Plasmid DNA from the resulting cultures was isolated using an EZ-10 Spin Column Plasmid DNA Miniprep Kit (Bio Basic Inc.), following the manufacturer's instructions. *pET28a-SNhp2 V122M* plasmid DNA was isolated and screened for positive mutagenesis through restriction digestion. 20 µL restriction digestions of *pET28a-SNhp2 V122M* (*Sma*I) was incubated at 30 ^{o}C for 2 hr with final component concentrations of: 1x Tango Buffer (Fermentas), ~2 µg *pET28a-SNhp2 V122M* plasmid DNA, and 2 U/µL *Sma*I restriction enzyme (Fermentas). Restricted colonies were analyzed on a 1% agarose gel at 110 V for 1 hr and visualized with ethidium bromide staining [88]. *pET28a-SNhp2 F146H* were sent directly for sequencing for mutagenesis confirmation. *pET28a-SNhp2 V122M* clones appearing to be positive were then sent to Macrogen Inc. for sequencing.

Purification of H/ACA protein Nhp2

Recombinantly expressed, histidine (His)-tagged Nhp2 was initially purified by nickel-sepharose affinity chromatography. *E. coli* BL21 (DE3) Rosetta cells containing His-tagged Nhp2 were resuspended in cell lysis buffer (1 mL/g of cells) containing: 50 mM Tris-HCl pH 8.0, 400 mM KCl, 0.1 mM phenylmethylsulfonyl fluoride (PMSF), 5 mM β-mercaptoethanol, 5 % glycerol, 30 mM imidazole. 1 mg/ml lysozyme was added to the resuspended cells followed by incubation for 30 min. Next 12.5 mg/g of cells sodium deoxycholate was added to the resuspended cells, and the suspension was incubated further for 20 min. Cells were opened by sonication for 1 min with 20 s

breaks, repeated five times using an intensity level of 6, and duty cycle of 60%. Cell debris was pelleted at 30 000 x g for 30 min at 4 ^{o}C. The cleared cell lysate was loaded onto a 5 mL Ni-Sepharose column (GE Healthcare) at a flow rate of 0.5 ml/min using a BioLogic LP chromatography system (BioRad) while the presence of protein was observed through absorbance measurements at 280 nm. The column was washed at a flow rate of 1.5 mL/min with Buffer A, comprised of: 50 mM Tris-HCl pH 8.0, 400 mM KCl, 0.1 mM phenylmethylsulfonyl fluoride (PMSF), 5 mM β-mercaptoethanol, 5 % glycerol, and 30 mM imidazole. His-tagged Nhp2 was eluted using a gradient in which the concentration of imidazole was increased at a rate of 6 mM/mL to a final concentration of 500 mM at a flow rate of 1 mL/min. Fractions of 1 mL/min were collected throughout the purification. Fractions were analyzed by 15% SDS-PAGE and visualized using Coomassie staining [88]. Fractions containing His-tagged Nhp2 were concentrated by ultrafiltration in Vivaspin MWCO 10000 (VWR) and re-buffered 1:100 in cation exchange low salt buffer (20 mM HEPES-KOH pH 8.0, 150 mM KCl, 0.5 mM EDTA, 1.5 mM $MgCl_2$, 2% glycerol, 5 mM β-mercaptoethanol).

His-tagged Nhp2 was further purified by cation exchange chromatography. His-tagged Nhp2 was loaded onto a 100 mL SP-sepharose cation exchange column (GE Healthcare) at a flow rate of 0.5 mL/min in the presence of low salt buffer using a BioLogic DuoFlow chromatography system (BioRad) while the protein absorbance at 280 nm was recorded. The column was washed at a flow rate of 4.0 mL/min with 60 mL of low salt buffer. His-tagged Nhp2 was eluted using a gradient in which the concentration of salt was increased at a rate of 20 mM/mL to a final concentration of 1000 mM at a flow rate of 4.0 mL/min. Fractions of 4 mL/min were collected throughout the purification. Fractions were analyzed by SDS-PAGE and visualized using Coomassie staining [88]. Fractions containing His-tagged Nhp2 were concentrated by ultrafiltration using a Vivaspin MWCO 10000 (VWR) and were subsequently shock frozen in liquid nitrogen and stored at -80 °C. The protein concentration was determined using absorbance

readings at 210 nm and an extinction coefficient of 20.5 mL mg^{-1} cm^{-1} as well as by quantification of SDS-PAGE with the software Image J 1.41 (Apache Software Foundation).

PCR amplifications of DNA template for H/ACA snoRNA

The template DNA for wild-type and truncated H/ACA snoRNA snR34 was amplified from *pTZ19r-snR34* plasmid DNA by polymerase chain reaction (PCR) using the following primer pairs:

snR34 wt F: AGCTATGACCATGATTACGCC
snR34 wt R: mUmAATGTAGACTTTCTTCGACATCCC
snR34 T F: AGCTATGACCATGATTACGCC
snR34 T R: mUmCTTGTTTGAAATACTGGCAATTAACTACTG

PCR reactions were carried out in 50 µL volumes with final concentrations as follows: 1x *Pfu* buffer with $MgSO_4$, 200 µM dNTPs, 0.5 µM of each primer, 0.02 U/µL *Pfu* DNA polymerase, and 40 ng/µL *pTZ19r-snR34* plasmid DNA. PCR cycle parameters included an initial denaturation at 95 ^oC for 5 min followed by 6 cycles of denaturation at 95 ^oC for 30 s, touch-up annealing starting for 30 s at 45 ^oC with a 1 °C increase in each cycle, and extension at 72 °C for 20 s, with an additional 30 cycles of denaturation at 95 ^oC for 30 s, annealing at 50 ^oC for 30 s, and extension at 72 ^oC for 20 s followed by a final extension at 72 ^oC for 11 min. PCR products were analyzed on a 2% agarose gel and visualized by ethidium bromide staining [88].

***In vitro* transcription of H/ACA snoRNA**

Wild-type and truncated H/ACA snoRNA, snR34, was transcribed *in vitro* from the PCR amplified DNA template. Transcription reactions were carried out in 50 µL or 1.0 mL reaction volumes with final concentrations as follows: 1x transcription buffer (40 mM Tris-HCl pH 7.5, 15 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl), 10 mM DTT, 3 mM of each NTP, 5 mM GMP, 0.01 U/µL inorganic pyrophosphatase (Sigma), 0.3 µM T7-RNA polymerase, 0.12 U/µL

RiboLock RNase Inhibitor (Fermentas) and 10% (v/v) template DNA. Transcriptions were performed at 37 °C for 4 hr. The DNA template was degraded through the addition of one crystal of DNase (Sigma) and further incubation for 1 hr at 37 °C. RNA products were analyzed by 11% urea-PAGE at 300 V for 25 min and visualized by ethidium bromide staining [88].

Purification of H/ACA snoRNA

In vitro transcribed wild-type and truncated H/ACA snoRNA snR34 (1 mL) were purified with anion exchange chromatography using a Nucleobond AX500 gravity flow column (Machery-Nagel). The column was equilibrated with 6 mL of equilibration buffer containing 100 mM Tris-acetate pH 6.3, 10 mM $MgCl_2$, and 15% (v/v) ethanol. The RNA (1 mL) was bound to the column after dilution with 13 mL of binding buffer composed of 200 mM KCl, 100 mM Tris-acetate pH 6.3, 10 mM $MgCl_2$, and 15% ethanol. The column was washed with 32 mL of washing buffer containing 100 mM Tris-acetate pH 6.3, 10 mM $MgCl_2$, 300 mM KCl, and 15% (v/v) ethanol and eluted in 12 mL high salt elution buffer comprised of 100 mM Tris-acetate pH 6.3, 10 mM $MgCl_2$, 1150 mM KCl, and 15% (v/v) ethanol. The RNA was subsequently precipitated with isopropanol for three days at -20 ^{o}C and re-suspended in 500 µL of ddH_20. The purified RNA was analyzed by 11% urea-PAGE and visualized by ethidium bromide staining.

Radiolabeling of H/ACA snoRNA

Wild-type and truncated H/ACA snoRNA snR34 were radiolabeled with ^{32}P. The H/ACA snoRNA was first dephosphorylated in a reaction volume of 20 µL containing: 0.2 µM RNA, 1x SAP buffer (Fermentas), and 1 U/µL SAP (Fermentas). The dephosphorylation reaction was performed at 37 ^{o}C for 30 min and stopped by heating at 65 ^{o}C for 15 min. H/ACA snoRNA was rephosphorylated in a reaction volume of 30 µL containing: 0.13 µM RNA, 1x SAP buffer (Fermentas), 5 mM DTT, 0.1 mM spermidine, 0.1 mM EDTA, 0.67 µM of [γ-^{32}P]-ATP (20 µCi), 10 U/µL T4 polynucleotide kinase (Fermentas). The phosphorylation reaction was carried out at

37 ^{o}C for 30 min and quenched by the addition 0.25 mM EDTA followed by heating at 75 ^{o}C for 10 min. The labeled snoRNA was purified using Sephadex G-25 filtration spin column. 400 µL of Sephadex G-25 (Pharmacia) was added to a SigmaPrep spin column, the column was washed with 400 µL ddH_2O following which labeled snoRNA was loaded and eluted from the column by centrifugation. The activity of labeled snoRNA was determined by liquid scintillation counting.

Electrophoretic mobility shift assays

Electrophoretic mobility shift assays (EMSA) were performed to analyze the interaction between Nhp2 and [^{32}P]-radiolabeled H/ACA snoRNA. snR34 H/ACA snoRNA was incubated at 85 ^{o}C for 2 min and permitted to slowly cool to 30 ^{o}C to promote correct folding. RNA (2 nM) was titrated with increasing concentrations of Nhp2 (0 to 50 µM) in 15 µL reaction volumes in 1x binding buffer (30 mM Tris-HCl pH 7.4, 150 mM KCl, 2 mM $MgCl_2$, 0.1% (v/v) Triton X-100, 1 mM DTT, 20% (v/v) glycerol, 100 ng/µL BSA). Binding was allowed to occur for 30 min at 30 ^{o}C. Samples were analyzed on an 8% Tris-glycine native gel (8% Acrylamide/Bisacrylamide 19:1 solution (Biobasic), 0.5x Tris-glycine buffer) in 0.5x Tris-glycine buffer (5 mM Tris pH 8.5 and 5 mM glycine) for 18 hr at 100 V at 4 ^{o}C and visualized by autoradiography. By plotting the concentration of Nhp2 was versus the intensity of the shifted band (determined by ImageJ) in Prism GraphPad3, a K_D for the interaction was extrapolated using the equation: $y = B_{max}*x/(K_D+x)$, where B_{max} is the maximal binding and K_D is the concentration of protein required to reach half-maximal binding

tRNA competition experiments

tRNA competition experiments were performed to analyze the binding specificity of Nhp2 to a [^{32}P]-radiolabeled H/ACA snoRNA using previously described EMSA conditions. 2 nM of snR34 H/ACA snoRNA was incubated at 85 ^{o}C for 2 min and permitted to slowly cool to 30 ^{o}C to promote correct folding. The RNA was then incubated with Nhp2 wild-type (50 nM), Nhp2

F146H (30 µM), and Nhp2 V122M (30 µM) in the presence and absence of 10 µM unlabeled tRNA in 15 µL reaction volumes in 1x binding buffer. Binding was allowed to occur for 30 min at 30 ^{o}C. Samples were analyzed on an 8% Tris-glycine native gel in 0.5x Tris-glycine buffer for 18 hr at 100 V at 4 ^{o}C and visualized by autoradiography.

Isothermal titration calorimetry

All isothermal titration calorimetry (ITC) experiments were performed at 30 ^{o}C on a VP ITC Microcalorimeter (Microcal). The cell volume was 1.7 mL, and the syringe volume was 300 µL. All solutions and samples were degassed prior to each experiment. Nhp2 was dialyzed in 2 L of ITC buffer (20mM HEPES-KOH pH 7.5, 150 mM KCl, and 15 mM $MgCl_2$) for 48 hr. After dialysis, the ITC buffer was saved for the actual ITC experiments. snoRNA was diluted in ITC buffer to a concentration of 5 µM (snR34 wt) and 2.8 µM (snR34 truncated), then heated to 85 ^{o}C for 2 min and permitted to slowly cool to 30 ^{o}C to promote correct folding. The experimental parameters used in each ITC analysis can be found in Table 1. snR34 concentrations used were determined by the equation, $c = K \times M_{tot} \times n$, in which K is the association constant, n is the stoichiometry parameter, M_{tot} is the total biomolecule concentration within the cell, and c is a unitless constant which determines the shape of the resulting thermogram. For ideal K measurements, c should be between 5 – 500. By solving the equation for M_{tot}, a minimal concentration of RNA is estimated for the cell. To determine the concentration of protein needed for the syringe, the equation $X_{tot} \times \Delta V/V = n \times M_{tot} \times 1.5$, was used where X_{tot} is the total concentration of protein in the syringe, ΔV is the total volume of protein to be injected, and V is the total volume of the cell. The reference power determines the value to which the baseline will settle once equilibrium has been reached within the cell. As it was initially hypothesized that the system would be largely exothermic, a large reference power was used (30 µCal/sec).

Table 1: Experimental parameters used in ITC analysis.

	Water – Water	**Methanol – Water**	**Buffer – Buffer**	**Protein – Buffer**	**Buffer – RNA**	**Protein – RNA**	
						snR34 wt	**snR34 Truncated**
Total No. of Injections:	30					28	30
Cell Temperature (^{o}C):	30						
Reference Power (µCal/s):	15					30	15
Initial Delay (s):	60	60	60	600	600	120	600
Nhp2 Syringe Concentration (µM):	0	0	0	100	0	100	40
snoRNA Cell Concentration (µM):	0	0	0	0	5	5	2.8
Stirring Speed (RPM):	450						
Feedback Mode/Gain:	Hi						
Injection Parameters							
Volume (mL):	10	10	10	10	10	10	10
Duration (s):	20	20	20	20	20	20	20
Spacing (s):	300	300	300	300	300	300	300
Filter (s):	2	2	2	2	2	2	2
ITC Equilibration Options	Fast Equilibration, Auto						

After analysis of the wild-type system, a smaller more conservative reference power (15 µCal/sec) was selected for analysis of the truncated RNA. As controls, water-water, methanol-water, buffer-buffer, protein-buffer, and buffer-RNA controls were performed. Data was fit in the Origin 7 software (MicroCal™ Software, Inc.) using a two-binding-site model.

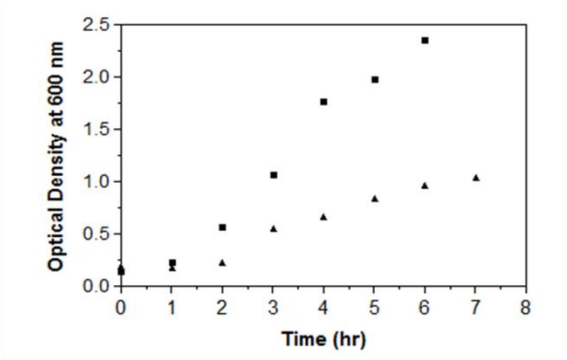
II.IV Results

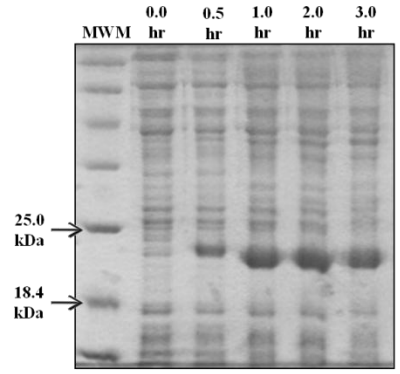
Recombinant Expression and Purification of S. cerevisiae H/ACA protein Nhp2

H/ACA structural protein Nhp2 was recombinantly expressed in *E. coli* BL21 (DE3) *Rosetta* cells at 37°C. Optical density readings at 600 nm of the *E. coli* BL21 (DE3) *Rosetta* cells containing *pET28a-Nhp2* were plotted in comparison to *E. coli* BL21 (DE3) *Rosetta* cells without plasmid to determine the cell growth efficiency of both strains (Fig. 6 A). In parallel, cell samples were collected upon induction to confirm the expression of Nhp2 by SDS-PAGE. A band of ~ 21 kDa is observed with increasing intensity over a three hour time period, indicating that the observed band is His-tagged Nhp2 (Fig. 6A).

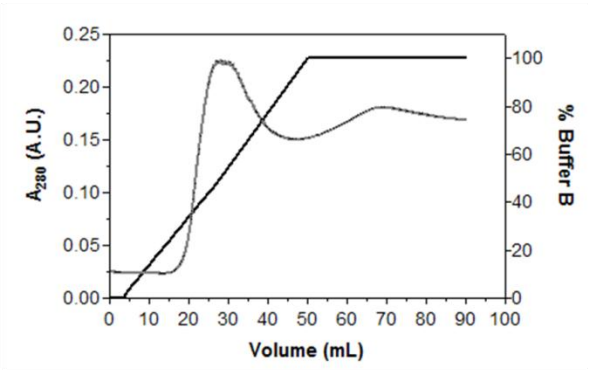
His-tagged Nhp2 was purified using a two-step purification procedure including Ni^{2+}-sepharose affinity chromatography and SP Sepharose Fast Flow cation exchange chromatography. *E. coli BL21 (DE3) Rosetta* cells containing His-Nhp2 were opened in a cell lysis buffer, and the resulting cellular lysate was loaded onto a Ni-Sepharose affinity chromatography column. Following washing with cell lysis buffer, His-Nhp2 was subsequently eluted from the column using a gradient with an increasing imidazole concentration. A large peak in absorbance at 280 nm is observed in the chromatogram at 250 mM imidazole (Fig. 6B). Fractions from this peak were analyzed by SDS-PAGE. As expected, a band of approximately 21 kDa corresponding to His-Nhp2 is observed in fractions corresponding to the peak in the chromatogram (Fig. 6B). In addition to the prominent 21 kDa band, several larger and smaller molecular weight bands are visible indicating that further purification of Nhp2 is required.

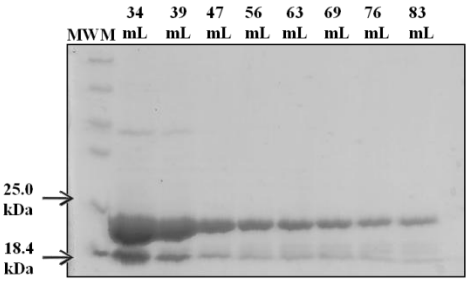
A. Over expression of Nhp2



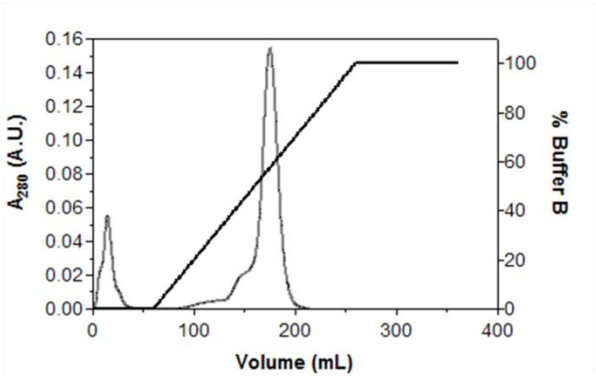


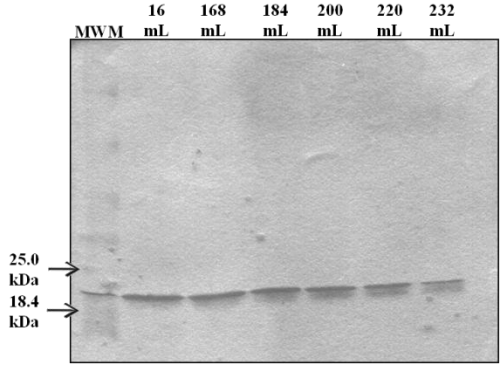
B. Ni sepharose purification of Nhp2





C. Cation exchange purification of Nhp2





D.

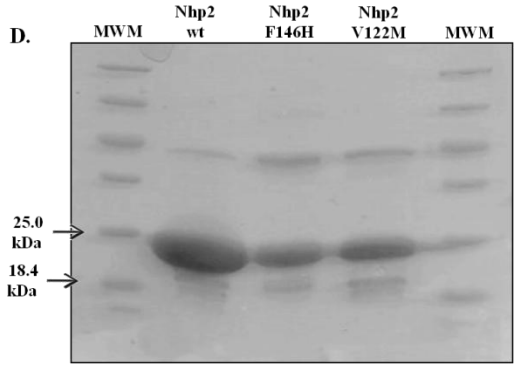


Figure 6: Recombinant expression and purification of *S. cerevisiae* H/ACA protein Nhp2. A. Recombinant expression of His-Nhp2 wild-type in *E. coli* BL21 (DE3) Rosetta cells. Cells were grown at 37 ^{o}C with cell samples taken for three hours after induction. Cell samples were analyzed by SDS–PAGE, and cellular growth (triangles) was plotted in comparison to *E. coli BL21* (*DE3*) Rosetta cells (squares) standard growth. His-tagged Nhp2 has a molecular weight of ~21 kDa. **B.** Nickel sepharose affinity chromatography purification of His-Nhp2 wild-type. Collected fractions were analyzed by SDS–PAGE. **C.** Cation exchange chromatography purification of His-Nhp2 wild-type. Collected fractions were analyzed by SDS–PAGE. **D.** His-tagged Nhp2 F146H and V122M variants were expressed and purified in the same manner as Nhp2 wild-type and were analyzed by SDS-PAGE.

As Nhp2 has a pI of 9.1 (as computed from its amino acid composition using the tools section on the UniProt database [Swiss Institute of Bioinformatics]), it is anticipated that it will interact with a cation exchange chromatography column with a strong affinity at low salt and a pH 8.0. Therefore, His-tagged Nhp2 was further purified by SP Sepharose Fast Flow cation exchange chromatography. Nhp2 was eluted from the column using a gradient from low to high salt. The resulting chromatogram displays a large peak during Nhp2 elution at 600 mM KCl (Fig. 6C). Samples taken from the corresponding fractions display a prominent band of approximately 21 kDa strongly indicating that His-tagged Nhp2 was successfully further purified with the additional higher and lower molecular weight contaminants significantly reduced (Fig. 6C). The presence of a single band corresponding to His-Nhp2 indicates that Nhp2 has been successfully purified.

As mentioned above, the rare inherited disease DC autosomal form has been linked to three substitutions found within the human H/ACA structural protein Nhp2: V126M, Y139H, and X154R [76]. To determine the possible effects such DC substitutions in Nhp2 may have on RNA recognition, two *S. cerevisiae* Nhp2 variants, V122M and F146H, corresponding to mutations in human Nhp2 found in DC patients, were created through site-directed mutagenesis and purified. *S. cerevisiae* Nhp2 DC variants F146H and V122M were purified using the same procedure as for wild-type Nhp2. Samples taken from purified and concentrated Nhp2 wild-type, F146H, and V122M were analyzed by SDS-PAGE (Fig. 6D). Analysis by Image J 1.41 (Apache Software Foundation) indicated that all protein preparations are ~95% pure.

In vitro Synthesis of *S. cerevisiae* H/ACA snoRNA snR34

Eukaryotic H/ACA snoRNAs are composed of two irregular hairpin structures linked by a hinge sequence [23]. A secondary structure model of *S. cerevisiae* H/ACA snoRNA snR34 is shown in Figure 7A outlining essential features of all H/ACA snoRNAs. The hinge region (Box H) following the 5' hairpin consists of an ANANNA sequence. Located downstream of the 3' hairpin structure is the Box ACA consisting of an ACA sequence [89]. The first step towards studying the protein-RNA and RNA-RNA interactions between H/ACA snoRNA snR34 and the core proteins and substrate RNA is to *in vitro* synthesize snR34. To this end, the DNA template for H/ACA snoRNA snR34 wild-type was amplified from *pTZ19r-snR34* plasmid DNA by PCR using appropriate primer pairs in which the reverse primer was 2'-O-methylated at the 5' end. The 2'-O-methylation prevents the addition of non-specific nucleotides to the 3' end of the snoRNA during the *in vitro* transcription. The amplified DNA template was analyzed by agarose gel electrophoresis in order to control for the correct template size and to estimate its concentration. A band of ~250 bp is observed corresponding to the correct size of *snR34* wild-type gene plus a T7 promoter region, indicating that the DNA template for snR34 wild-type was correctly amplified (Fig. 7B). snR34 wild-type was then produced in a four hour *in vitro* transcription reaction. Subsequently, the DNA template was degraded by DNase I. The resulting RNA product was purified by an AX500 Nucleobond column (Machery-Nagel). RNA samples were collected after the *in vitro* transcription and during purification and were analyzed by urea–PAGE. As shown in Figure 7B, significant amounts of RNA product are visible after both the *in vitro* transcription and the purification leading to the conclusion that snR34 wild-type was successfully synthesized and purified.

As a first step to determine the specific binding site of H/ACA protein Nhp2 on H/ACA snoRNA snR34, a truncated version of snR34 was synthesized. The truncated construct removes the 3'-hairpin structure leaving the hinge sequence and the 5' hairpin only as outlined by a black

box in Figure 7A. The snR34 truncation was synthesized using the same experimental set up used for snR34 wild-type. The amplified DNA template has the expected band size of ~200 bp corresponding to the correct size of truncated construct with a T7 promoter region (Fig. 7C). Urea-PAGE analysis of the resulting RNA samples collected from the subsequent *in vitro* transcription and purification shows significant amounts of RNA product suggesting that truncated snR34 was successfully synthesized (Fig. 7C). Analysis of the RNA products by urea-PAGE shows that both RNA preparations consist predominantly of a single RNA species thus indicating that little degradation of the RNA product took place during the *in vitro* transcription and purification (Fig. 7B-C).

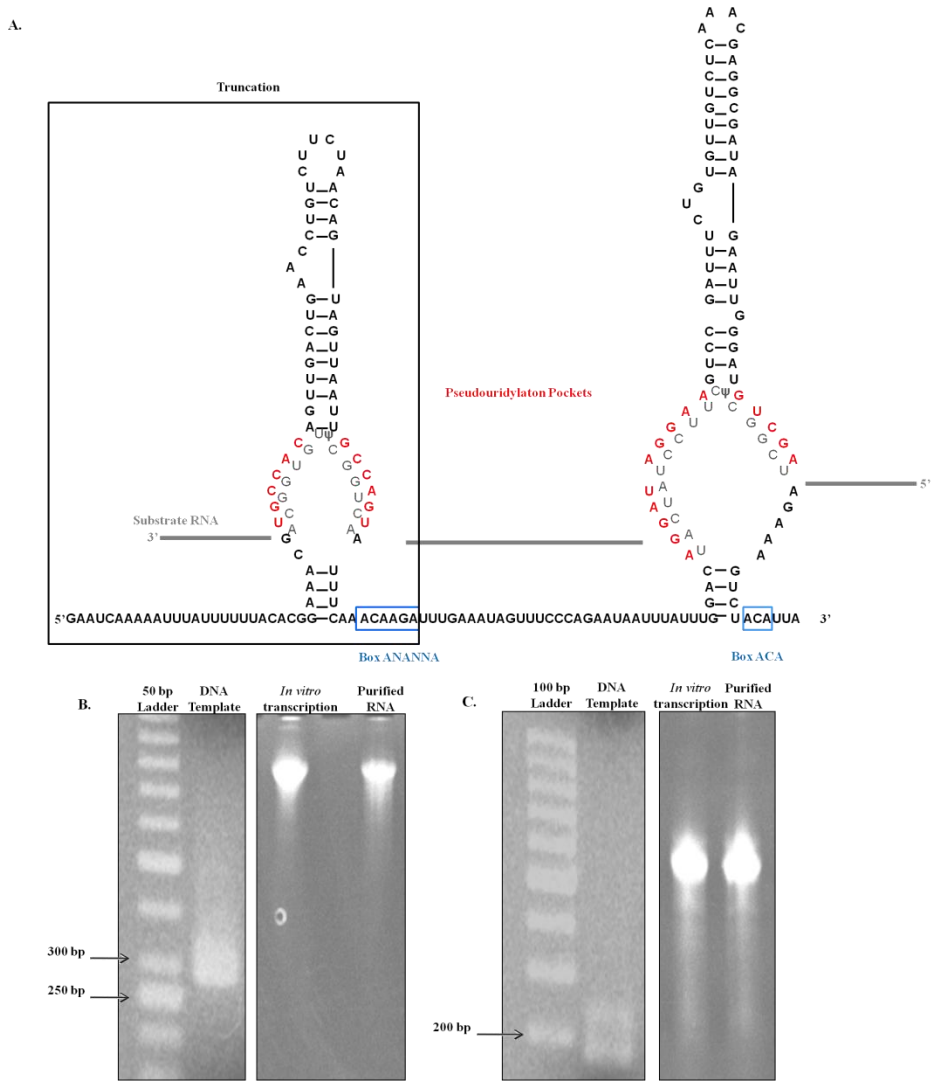


Figure 7: Synthesis of *S. cerevisiae* H/ACA snoRNA snR34 wild-type and truncated. A. Secondary structure of H/ACA snoRNA snR34. The pseudouridylation pockets are indicated in red and the substrate RNA is indicated in grey. snR34 5'-hairpin targets U2826, and snR34 3'-hairpin targets U2880 of the 25S rRNA. The conserved box ANANNA and ACA are highlighted in blue. The black box indicates a shortened RNA construct which is used to identify the binding site of Nhp2. **B.** Synthesis of snR34 wild-type DNA template and the *in vitro* transcription and purification of snR34 wild-type H/ACA snoRNA. **C.** snR34 truncation DNA template and *in vitro* transcription and purification of truncated snR34 H/ACA snoRNA.

Towards the identification of the minimal H/ACA snoRNA binding site for Nhp2 by EMSA

By using [^{32}P]-labeled snR34 in EMSA, I was able to detect and quantify the interaction of Nhp2 with H/ACA snoRNA. 2 nM of radiolabeled snR34 wild-type was incubated with increasing concentrations of Nhp2 wild-type, and the snR34-Nhp2 interaction was analyzed by native-PAGE. An upwards shift of the RNA band was observed upon addition of Nhp2 (Fig. 8A) which is indicative of complex formation between snR34 and Nhp2. The intensity of the shifted band increases with increasing Nhp2 concentration over the whole concentration range. The fact that a band corresponding to free RNA remains visible at the highest tested Nhp2 concentration (72 nM) suggests that complete binding has not been reached yet. By plotting the concentration of Nhp2 versus the quantified intensity of the shifted band (Image J 1.41 [Apache Software Foundation] pixel quantification), an approximate value for the K_D of the interaction between Nhp2 and snR34 wild-type can be extrapolated to be 36 +/- 9 nM. While Nhp2's archaeal homolog L7Ae can interact with kink-turn structural motifs with high specificity, Nhp2 however does not have this specificity. This puzzling trait of Nhp2 puts forth several questions including how Nhp2 specifically recognizes and binds H/ACA snoRNAs. To answer this question, the minimal binding site of Nhp2 on the H/ACA snoRNA must be identified by successively shortening the snoRNA and studying the resulting interaction between Nhp2 and the shortened construct. Towards this aim, qualitative EMSA was used to analyze the interaction between Nhp2 wild-type and truncated snR34 (outlined in Fig. 7A). Similarly to the Nhp2-snR34 wild-type system, 2 nM of radiolabeled snR34 truncation was incubated with increasing concentrations of Nhp2 wild-type, and the RNA-Nhp2 interaction was analyzed by native-PAGE. An observable

shift in band location is visible with increasing concentrations of Nhp2 (Fig. 8B), suggesting an interaction between truncated snR34 and Nhp2 wild-type.

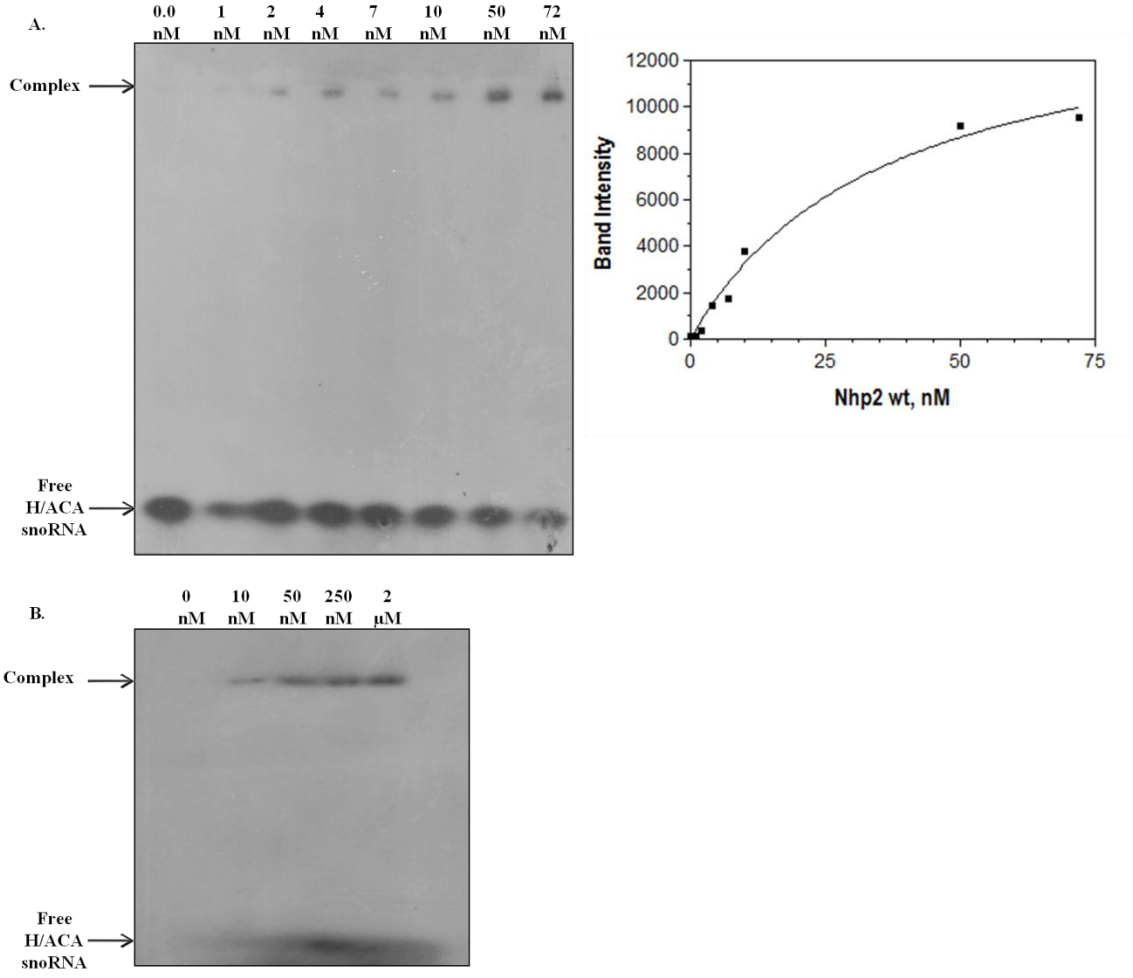


Figure 8: Analysis of *S. cerevisiae* Nhp2 wild-type interaction with snR34 wild-type and truncated snR34 by EMSA. H/ACA snoRNA, snR34, was radiolabeled at its 5' end with [^{32}P]. snR34 concentrations remained constant at 2 nM while increasing concentrations of Nhp2 were titrated. All Nhp2-snR34 interactions were analyzed on an 8% native-PAGE and visualized using autoradiography. **A.** EMSA analysis of the interaction between Nhp2 wt and snR34 wild-type. The intensities of the shifted band were estimated by Image J 1.41 (Apache Software Foundation) and plotted against the Nhp2 concentration, in order to estimate the K_D (36 +/- 9 nM). **B.** EMSA analysis of the interactions between Nhp2 wt and snR34 truncation.

Analysis of the interaction between S. cerevisiae Nhp2 and H/ACA snoRNA snR34 by isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a highly sensitive, powerful technique used to study the thermodynamics of the binding of the two interacting molecules [90]. As the name implies, ITC is based on the gradual titration of a ligand into a solution of its binding partner. By monitoring the amount of energy over time required to maintain a constant temperature within the sample cell, ITC is able to provide essential information on the binding affinity, stoichiometry and thermodynamics of an interaction [90].

Here, ITC was used to further examine the binding affinity, stoichiometry and thermodynamic properties of the *S. cerevisiae* Nhp2 and H/ACA snoRNA snR34 wild-type and truncation interactions. 100 µM Nhp2 was titrated into 5.0 µM snR34 wild-type, while 40 µM Nhp2 was titrated into 2.8 µM truncated snR34. In both experimental setups, injection of Nhp2 produced an upward inflection observed in the raw ITC data indicating that an increase in energy over time was required to maintain the system at a constant temperature (Fig. 9A, B), thus indicating that an endothermic reaction was occurring. The gradual decrease in the magnitude of each inflection with the increase of Nhp2 concentration corresponds to increasing binding of Nhp2 to the RNA (Fig. 9A, B). The small inflections at the end of each titration can be attributed to energy released from the heat of dilution and mechanical energy as protein saturation has occurred. Controls to account for the heat of dilution and mechanical energy released during this experiment were performed and found to produce minimal to no effect on the system energy (data not shown). By integrating the area under each inflection point, a thermogram of the injection heats for the specific amount of Nhp2 added at a specific injection point was generated. Quantification of these injection heats produced from Nhp2 interactions with both snR34 wild-type and truncation revealed a biphasic behavior in which the first Nhp2 binding event is

enthalpically driven with enthalpies of binding of ΔH_{1wt} = -3000 +/- 9000 cal/mole and $\Delta H_{1truncated}$ = -1600 +/- 400 cal/mole respectively.

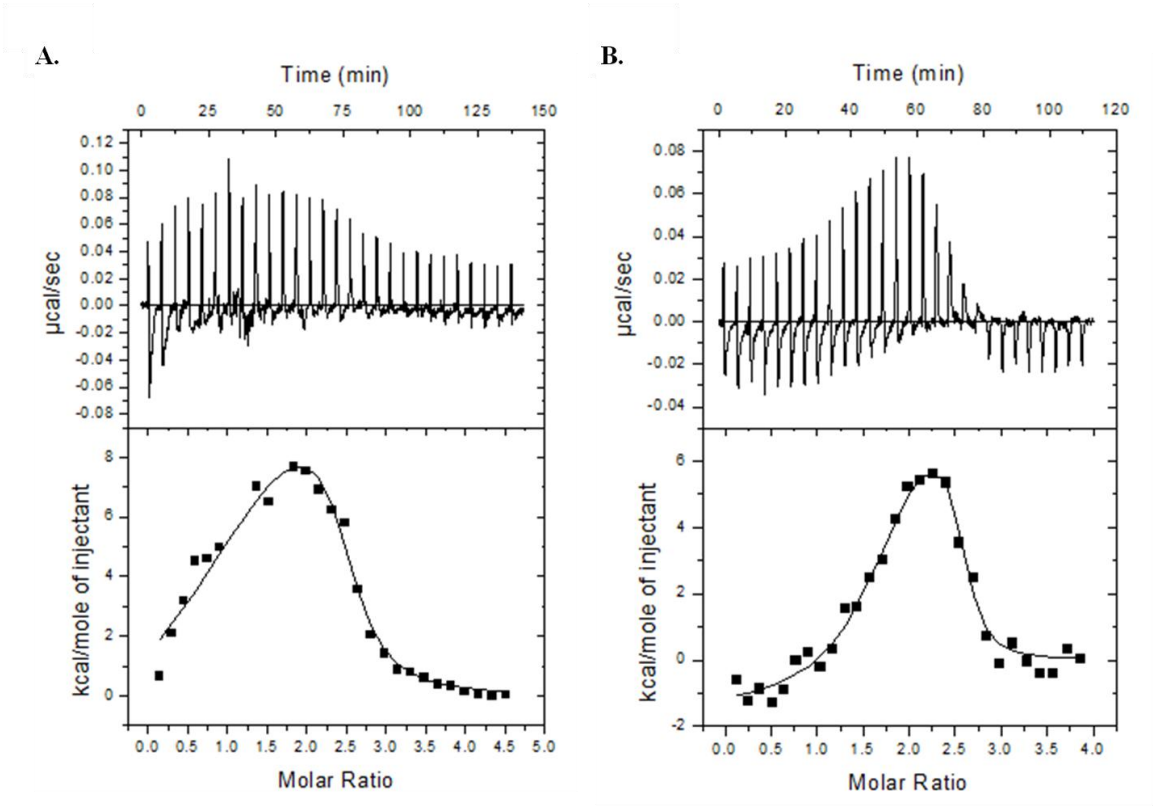


Figure 9: Analysis of *S. cerevisiae* Nhp2 wild-type interaction with snR34 wild-type and truncated snR34 by ITC. All ITC analysis was performed at 30 ^{o}C on a VP ITC Microcalorimeter. Raw data and ΔH graph generated from the interactions between Nhp2 wt and **A.** snR34 wt. and **B.** snR34 truncation.

The enthalpies of binding for the second Nhp2 binding event were determined to be ΔH_{2wt} = 14,000 +/- 11,000 cal/mole and $\Delta H_{2truncated}$ = 8500 +/- 2000 cal/mole, indicating an entropically driven reaction is occurring. The binding entropy was calculated from ΔH and K using $\Delta G^{o} = -RT \ln K = \Delta H^{o} - T\Delta S^{o}$. Comparison of the K_D value for the first binding event in the Nhp2-snR34 wild-type interaction (12 +/- 8 nM) determined by ITC with that estimated from the EMSA analysis (36 +/- 9.0 nM) shows that both values are in the same order of magnitude (low nanomolar range), similar to the previously determine K_D value for the interaction between Nhp2 and H/ACA snoRNA snR36 (4 nM, [39]). A complete comparison of thermodynamic parameters

determined by ITC for the interactions between Nhp2 and snR34 wild-type and truncation can be found in Table 2.

Table 2: Comparison of parameters for snR34 wild-type and truncated snR34 determined by ITC where n, K_D, ΔH, and ΔS are stoichiometry, dissociation constant, enthalpy of binding and entropy of binding.

ITC Determined Parameter	Nhp2-snR34 wt Interaction	Nhp2-snR34 truncation Interaction
n_1	1.1 +/- 0.4	1.6 +/- 0.1
K_{DI}	12 +/- 8 nM	2.4 +/- 1.6 nM
ΔH_1	-3,000 +/- 9,000 cal·mole^{-1}	-1,600 +/- 400 cal·mole^{-1}
ΔS_1	26.8 cal·mole^{-1}K^{-1}	34.3 cal·mole^{-1}K^{-1}
n_2	1.4 +/- 0.5	0.9 +/- 0.1
K_{D2}	40 +/- 80 nM	29 +/- 13 nM
ΔH_2	14,000 +/- 11,000 cal·mole^{-1}	8,500 +/- 2,000 cal·mole^{-1}
ΔS_2	79.6 cal·mole^{-1}K^{-1}	62.6 cal·mole^{-1}K^{-1}

Analysis of the effects of Dyskeratosis congenita substitutions in S. cerevisiae Nhp2 on the interaction with H/ACA snoRNA snR34

The autosomal form of the multisystem disease DC is associated with three substitutions found within the human Nhp2: V126M, Y139H, and X154R [76]. The generation of two DC substitutions in *S. cerevisiae* Nhp2, V122M and F146H, which correspond to the mutations found in human Nhp2 DC patients allows for the *in vitro* study of the effects such substitutions in Nhp2 may have on RNA recognition. To examine the effects these substitutions have on Nhp2's

interaction with H/ACA snoRNA, Nhp2 variants, F146H and V122M, were used in EMSA analysis with snR34 wild-type.

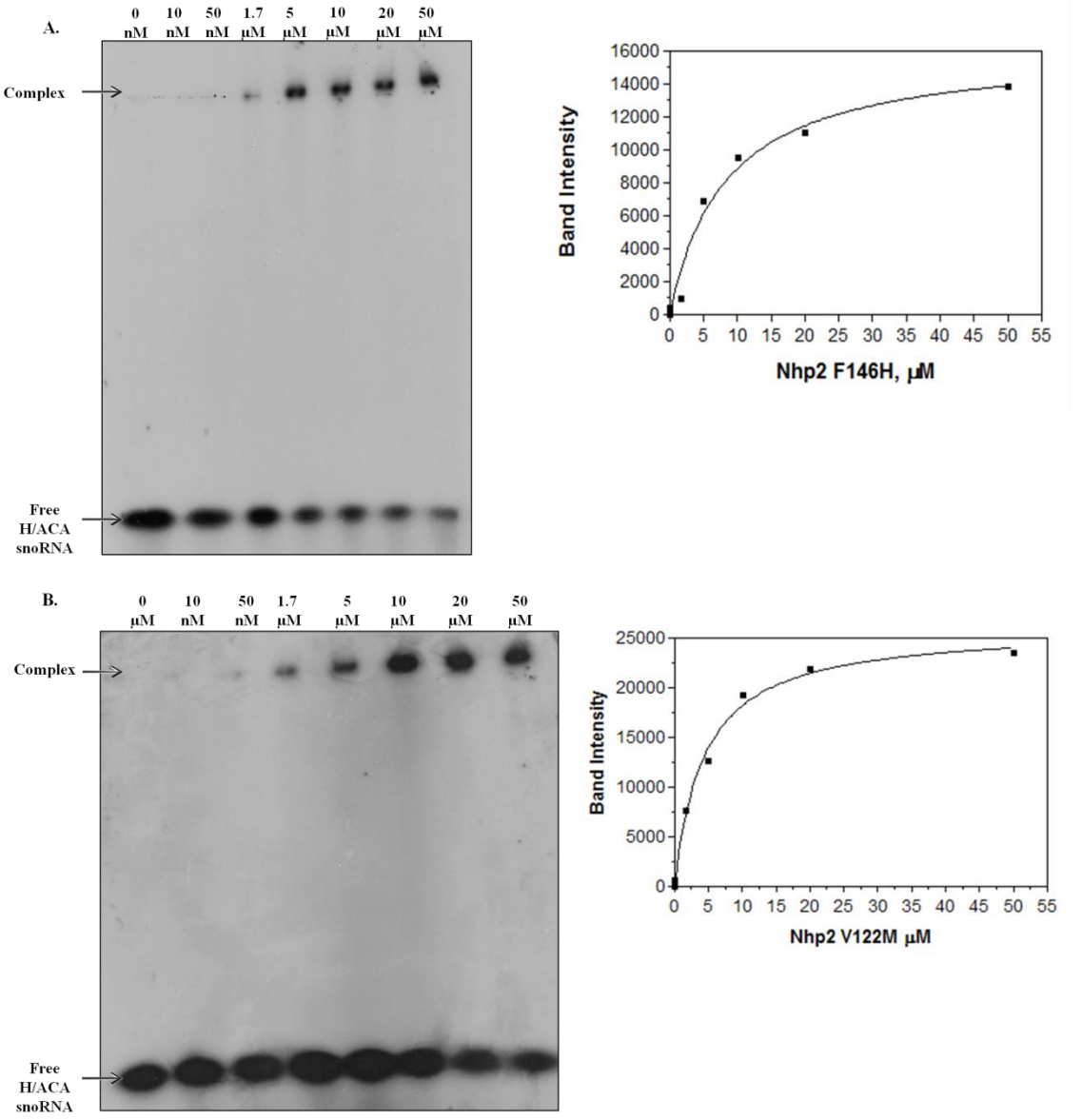


Figure 10: Analysis of *S. cerevisiae* Nhp2 Dyskeratosis congenita variants F146H and V122M interaction with snR34 wild-type by EMSA. snR34 wild-type concentration remained constant at 2 nM while increasing concentrations of Nhp2 were titrated in. All Nhp2-snR34 interactions were analyzed by 8% native-PAGE. **A.** EMSA analysis of the interaction between Nhp2 F146H and snR34 wt with a K_D estimated of 8.0 +/- 2.0 µM. **B.** Analysis of the interactions between Nhp2 V122M and snR34 wt using EMSA, where a K_D of 4.4 +/- 0.6 µM was estimated.

2 nM of radiolabeled snR34 wild-type was incubated with increasing concentrations of Nhp2 F146H or V122M respectively, and their interactions were analyzed by native PAGE. The

titration of both Nhp2 F146H and V122M to snR34 wild-type produces a shifted band of higher molecular weight visible in Figure 10A-B, suggesting an interaction between snR34 wild-type and Nhp2 F146H and V122M is occurring. Complex formation is also supported by the decrease in band intensity corresponding to free RNA (Fig. 10A- B). K_Ds of 8.0 +/- 2.0 µM and 4.4 +/- 0.6 µM were estimated for Nhp2 F146H and V122M, respectively, from the dependence of the shifted band intensity on the Nhp2 concentration. When compared with the estimated K_D (36 +/- 9 nM) calculated from the snR34-Nhp2 wild-type interaction, a notable decrease in RNA affinity is observed for Nhp2 F146H and V122M indicating that both substitutions of Nhp2 impair RNA binding.

Analysis of Nhp2 wild-type and Dyskeratosis congenita variants specificity for H/ACA snoRNA snR34

Nhp2 has been reported to bind non-specifically to RNA molecules when alone [39, 68]. Taking this into consideration we must question whether the interactions we see within our binding assays between Nhp2 and snR34 are the result of specific or non-specific binding. To test Nhp2's binding specificity for H/ACA snoRNA snR34, a tRNA competition experiment was performed with Nhp2 wild-type, and Nhp2's two DC variants V122M and F146H. 2 nM of radiolabelled snR34 wild-type and 10 µM unlabeled tRNA were used in all three chase experiments. A concentration of 50 nM of Nhp2 wild-type was added to snR34 alone and to snR34 in the presence of tRNA. Similarly, 30 µM Nhp2 DC variants V122M and F146H were added to snR34 alone and to snR34 plus tRNA. Analysis of the shifted band intensities in the presence and absence of tRNA was performed by Image J 1.41 (Apache Software Foundation) and is summarized in Table 3. The presence of tRNA appears to have an effect on Nhp2 binding to snR34, as a decrease in the shifted band intensities in the presence of tRNA is visible for Nhp2 wild-type (Fig. 11A), F146H (Fig. 11B lanes 2-3), and V122M (Fig. 11B lanes 4-5).

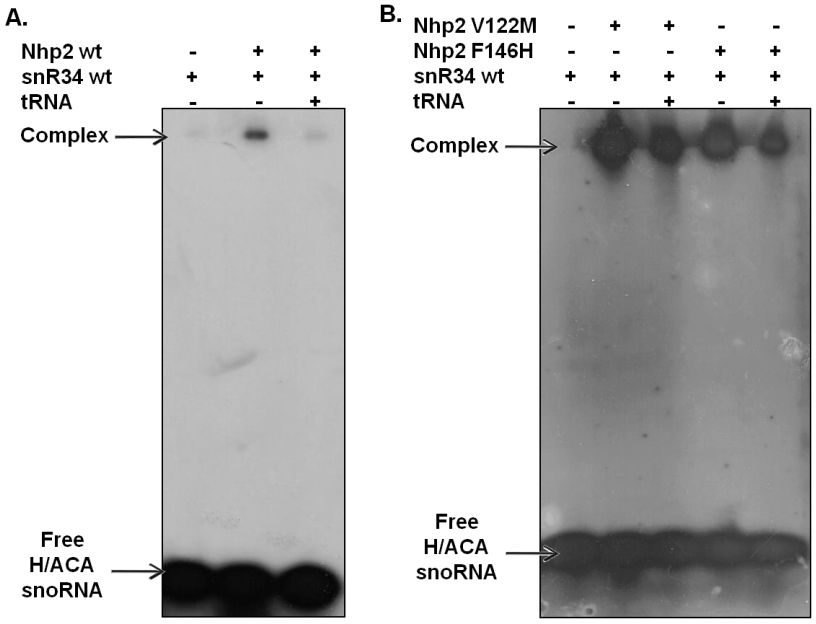


Figure 11: Analysis of *S. cerevisiae* Nhp2 specificity for H/ACA snoRNA snR34. Nhp2 wt, V122M, and F146H were incubated with radiolabeled snR34 wild-type in the presence and absence of unlabeled tRNA. snR34 concentration remained constant at 2 nM while tRNA concentration remained constant at 10 µM. tRNA chase experiments were analyzed by 8% native-PAGE. **A.** Nhp2 wt. (50 nM). **B.** Nhp2 V122M (30 µM, lanes 2-3) and F146H (30 µM, lanes 4-5).

Table 3: Comparison of band intensities of Nhp2-snR34 complexes in the presence and absence of tRNA.

Nhp2 wt			**Nhp2 V122M**			**Nhp2 F146H**		
snR34 alone	snR34 + tRNA	Percent Change in Binding	snR34 alone	snR34 + tRNA	Percent Change in Binding	snR34 alone	snR34 + tRNA	Percent Change in Binding
11,200	2,400	78 %	45,000	26,000	42 %	47,000	36,000	23 %

II.V Discussion

Binding affinity of Nhp2 wild-type and Nhp2 DC variants to H/ACA snoRNA snR34

The small RNA-binding protein Nhp2 is an important structural member of the eukaryotic H/ACA snoRNP complex which is responsible for the introduction of pseudouridine modifications in many cellular RNAs. Mutations in Nhp2 (V126M, Y139H, and X154R) have been directly linked to the autosomal recessive form of the rare inherited disease DC. The determination of Nhp2's binding affinity to H/ACA snoRNAs allows for the study of the effects such DC substitutions have on Nhp2 interaction with H/ACA snoRNAs. Here approximate dissociation constant values were extrapolated from the EMSA analysis of the interactions between Nhp2 wild-type, V122M, and F146H with snR34 wild-type to be K_D(wt) = 36 +/- 9.0 nM, K_D(F146H) = 8.0 +/- 2.0 µM, and K_D(V122M) = 4.4 +/- 0.6 µM. While these K_D values are only estimations derived from the quantification of the shifted band intensities, it is clear that the binding affinity of Nhp2 to H/ACA snoRNA snR34 is dramatically affected by approximately a 100 fold decrease when DC substitutions are introduced into Nhp2.

Valine 126 (valine 122 in yeast) and tyrosine 139 (phenylalanine 146 yeast) are highly conserved residues within the human H/ACA structural protein Nhp2 [75, 84]. Both residues are also directly involved in packing of the hydrophobic core of the protein, indicating an important role in Nhp2's structural integrity and stability [75, 84]. Valine 126 is positioned directly above the α helix which interacts with Nop10; replacement of the valine residue with a bulkier methionine residue destabilizes the interaction between Nhp2 and Nop10 presumably by altering the core packing of Nhp2 [75]. Tyrosine 139 has an important role in Nhp2's protein structure and stability as tyrosine 139 is also located within the interior of the Nhp2 [75]. Analysis of Nhp2's primary sequence shows that tyrosine 139 is conserved as a hydrophobic residue in all species. The introduction of a histidine residue dramatically affects the overall core packing of

Nhp2 and thus the stability of the protein [75]. Similarly to the effects on Nop10 interactions, it can be envisioned that destabilization of the Nhp2 hydrophobic core also affects RNA binding as shown here.

In the cell, Nhp2's association with H/ACA snoRNPs and thus with H/ACA snoRNAs is dependent on its interaction with the Nop10-Cbf5 binary complex [68, 82]. As shown here, destabilization of the packing of Nhp2's hydrophobic core dramatically affects Nhp2's affinity to H/ACA snoRNAs thus impairing the assembly of the pre-H/ACA RNP complex. Interaction studies of archaeal H/ACA snoRNP complex have shown that the absence of L7ae (the archaeal homolog of Nhp2) dramatically affects placement of substrate RNA within the catalytic site. As H/ACA snoRNPs are involved not only in RNA modifications but also in telomerase RNA maturation, impaired pre-RNP assembly would also result in the degradation of the telomerase RNA. Without the telomerase RNA template, replenishment of the telomere repeating sequence would not occur resulting in the occurrence of critically short telomere sequences often observed in DC patients.

S. cerevisiae Nhp2 recognition of H/ACA snoRNAs

Using [^{32}P]-labeled snR34 in EMSA analysis, we have been able to detect and quantify the interaction between Nhp2 and H/ACA snoRNA snR34 wild-type with an estimated K_D of 36 +/- 9.0 nM. This K_D value within the low nM range is in the same order of magnitude as in the previous reported Nhp2-H/ACA snoRNA interaction studies performed by the Henry lab where the interaction between Nhp2 and H/ACA snoRNA snR36 was analyzed, and a K_D of 4 nM was determined using EMSA [39]. In an effort to illuminate the details of Nhp2's interaction with H/ACA snoRNAs, I have begun to determine of the minimal binding site of Nhp2 on H/ACA snoRNAs through the successive shortening of the H/ACA snoRNA snR34. Analysis of the first Nhp2-snR34 truncation interaction by EMSA demonstrates that removal of the conserved box

ACA motif and the 3'-hairpin have no effect on Nhp2's interaction with the guide RNA as binding is observed in a similar Nhp2 concentration range. Nhp2's ability to interact with the guide RNA in the absence of both the box ACA motif and the 3'-hairpin may partially be explained through Nhp2's archaeal homolog L7Ae. Like Nhp2, L7Ae interacts with H/ACA snoRNAs at the tip of each irregular-hairpin, and thus Nhp2 might still be able to bind to the first hairpin when we removed the second [28, 39, 87]. Similar findings have been reported with another guide RNA [28, 39, 87]. While Nhp2 lacks the ability to interact with kink-loop structural motifs, several basic residues within the conserved RNA binding region of L7Ae are found in Nhp2 suggesting a common mechanism for interacting with H/ACA snoRNAs [24]. The difference observed between Nhp2's interaction with H/ACA snoRNAs and that of L7Ae may also be attributed to five signature amino acids which are variable and unique to each member of the L7Ae/L30 protein family [40]. These five signature amino acids flank the RNA-binding region and appear to confer RNA recognition specificity which is distinct to each member of the L7Ae/L30 protein family [40]. Replacement of L7Ae's signature RNA recognition residues with those of Nhp2 results in loss of L7Ae's kink-loop binding capabilities and a reduction in L7Ae's kink-turn binding abilities [40].

EMSA analysis is a useful tool in determining the binding affinity of Nhp2 to H/ACA snoRNAs. EMSA does however have several limitations which influence the accuracy of the experimental data collected from such analyses. EMSA is a non-equilibrium method, and hence there is the possibility of the protein-RNA complex being disrupted during electrophoresis. The disruption of the complex would thus result in the underestimation of protein binding. Therefore, the determined K_D's are estimates only, as the relative percentage of the RNA-protein complex might be underestimated. Additionally, EMSA analysis of the interaction between Nhp2 and snR34 can only give a clear indication that the two recombinant components are interacting, it

does not provide information on where the components are interacting or how, providing a rather shallow view of the protein-RNA interaction occurring.

To address the limitation presented in EMSA analysis, a second more precise quantitative technique, ITC, was used to examine the binding affinity, stoichiometry and thermodynamic properties of the Nhp2 and H/ACA snoRNA snR34 interaction. As ITC is an equilibrium-based technique, it is capable of providing more quantitative information on protein binding. ITC is an extremely sensitive and direct technique which measures the amount of energy over time needed to maintain a constant temperature within the sample cell as complex formation occurs [90]. Using L7Ae's interaction with archaeal H/ACA snoRNAs as a model, it was believed that Nhp2 would interact with eukaryotic bipartite H/ACA snoRNAs in a similar manner to L7Ae's interactions in which one Nhp2 protein would associate with each hairpin structure. Upon analysis of the Nhp2-snR34 wild-type and Nhp2-snR34 truncation interactions, it became apparent that we were working with a two-site binding model, where the first Nhp2 binding event is enthalpically driven, while the second Nhp2 binding event is entropically driven. As high RNA and protein concentrations were used, Nhp2 presumably binds to both sites simultaneously, explaining why all inflections are upward although the first binding event is enthalpically driven. Interestingly, examination of the Nhp2-snR34 truncated results revealed no significant alteration in the stoichiometry possibly suggesting that both binding events occur on the 5'-hairpin. This raises the question whether Nhp2 can bind to the 3' hairpin as predicted and where the two binding sites are located on the 5' hairpin.

Within the raw ITC data a small negative inflection is visible directly after the observed positive inflections. This pattern of positive and negative inflections is suggestive of a binding event followed by a conformational change within the newly formed complex as protein binding can be endothermic, withdrawing energy from the system, while a conformational change typically is exothermic, releasing energy from the system. These features of Nhp2's interactions

with snR34 wild-type and truncation are similar to those observed in L7Ae's binding of archaeal H/ACA snoRNAs. Binding of L7Ae to an archaeal H/ACA snoRNA induces a conformational change within the guide RNA forming the notable 120^{o} bend in the phosphodiester backbone of the snoRNA characteristic to all kink-loop and kink-turn structural motifs [28]. As Nhp2 shares many conserved residues in the RNA-recognition region with L7Ae, it is reasonable to hypothesize that the mechanism of interaction between the two homologues will be conserved as well, suggesting that binding of Nhp2 to eukaryotic H/ACA snoRNA may also induce a conformational change within the snoRNA. To determine whether Nhp2's interaction mechanism is similar to that of L7Ae, the minimal binding site of Nhp2 on the H/ACA snoRNA must be established. Therein, we can examine the binding specificity of Nhp2 by systematically changing nucleotides within the minimal binding site of H/ACA snoRNAs. This would allow in turn for the determination of a more precise binding affinity, stoichiometry, and thermodynamic properties for the interaction between Nhp2 and H/ACA snoRNAs. Additionally, to observe whether Nhp2 induces a similar kinking effect within the H/ACA snoRNA as L7Ae, the crystal structure of Nhp2 bound to the H/ACA snoRNA would need to be determined, as well as the crystal structure of the H/ACA snoRNA in the absence of Nhp2.

While the advancements in ITC equipment have made it an increasingly popular technique for obtaining and analyzing the quantitative thermodynamic proprieties of biological systems, it has several limitations as an analytical tool. ITC experiments require large amounts of sample, which may be difficult to produce when working with recombinant components [90]. Additionally, the data gathered from ITC analysis only provide information on how the respective biological components are interacting, but not where they are interacting [90]. While the stoichiometry, binding affinity, and thermodynamic properties of a system provide invaluable information, without knowing the interaction sites, such values often raise more questions than answers [90].

Nhp2's specificity for H/ACA snoRNAs

The binding specificity of a protein is determined by several factors including the affinity (K_D) of the protein for different ligands and the concentration of these different ligands within the cell. If a protein interacts with several different ligands in the cell, this binding is characterized as being unspecific. Binding is specific when a protein interacts only with a single ligand in the cell either because the affinity for this ligand is much higher than for other similar ligands or because this ligand is present in a much higher concentration in the cell than other ligands.

Nhp2 has previously been reported to bind non-specifically to RNA molecules, bringing into question whether the interaction observed in our binding assays is specific or unspecific [39, 68]. To test for Nhp2's RNA specificity, a tRNA competition experiment was performed with Nhp2 wild-type, and Nhp2's two DC variants V122M and F146H. Analysis of the competition experiments showed a strong effect on Nhp2's interaction with H/ACA snoRNA snR34 wild-type. Such a result suggests that Nhp2's affinity for tRNA is within a range from low nano-molar to 5 µM (half of the added tRNA concentration), as addition of tRNA clearly affects Nhp2-snR34 wild-type binding. A low affinity of 5 µM to tRNA would be sufficient to explain the observation as in this case Nhp2 would quantitatively bind to tRNA which is present at high concentrations (10 µM). Thereby, the concentration of free Nhp2 which can bind to snR34 would be highly reduced. Another explanation for the results observed in the tRNA competition assay would be that Nhp2's affinity for tRNA was equal to or stronger than its affinity for snR34. However, under *in vivo* cellular conditions this would seem highly unlikely as Nhp2 has been document through *in vivo* pull down assays to interact specifically with H/ACA snoRNAs and protein components [39, 68, 91]. The effect observed in the snR34-Nhp2 mutant tRNA chase experiment are the result of approximately 10 µM of Nhp2 interacting with the tRNA and 20 µM of Nhp2 interacting with the H/ACA snoRNA.

A review of the literature shows comparable results when Nhp2 is exposed to RNA structures containing irregular hairpins similar to those found in H/ACA snoRNAs [39]. It has also been reported that while Nhp2 does interact with RNA molecules *in vitro*, it does not interact with all RNA molecules equally, displaying extremely poor interactions with single-strand poly(A) and poly(U) RNAs, and little affinity for perfect double stranded RNA molecules [39]. As tRNA's clover leaf structure contains several hairpin structures, it is not unreasonable to hypothesize that Nhp2 may recognize secondary RNA structures rather than specific residues within the RNA's primary sequence. If Nhp2 specificity arises from its ability to recognize and interact with certain secondary structures of RNA (irregular hairpins), it is not unreasonable to hypothesize that Nhp2 would have a similar affinity to tRNA as it does to H/ACA snoRNA as both contain irregular hairpin secondary structures. To test this hypothesis, Nhp2's affinity for tRNA would need to be analysed in a comparable manner as Nhp2's affinity for H/ACA snoRNA, potentially using EMSA and ITC to determine the thermodynamic properties of the interaction. It has also been well documented that Nhp2's association with the H/ACA snoRNP complex is dependent on the formation of a Nop10-Cbf5 binary complex [68, 82]. The presence of this Nop10 –Cbf5 binary complex may provide Nhp2 with the RNA binding specificity observed in other L7Ae/L30 protein family members. While both hypotheses of Nhp2's interaction specificity have merit, further research is required to ultimately determine how Nhp2 interacts with H/ACA snoRNAs [68, 82].

III

H/ACA snoRNP Catalytic Protein Cbf5

III.I Introduction

Cbf5, a Member of the TruB Pseudouridine Synthase Family

S. cerevisiae H/ACA catalytic protein Cbf5 was first identified as a low-affinity centromere-binding protein, sharing similar charge distribution and a sequence repeat with known microtubule-associated proteins [92]. Later sequence analysis revealed significant sequence similarity with other bacterial and eukaryotic proteins, including the *E. coli* pseudouridine synthase TruB [33].

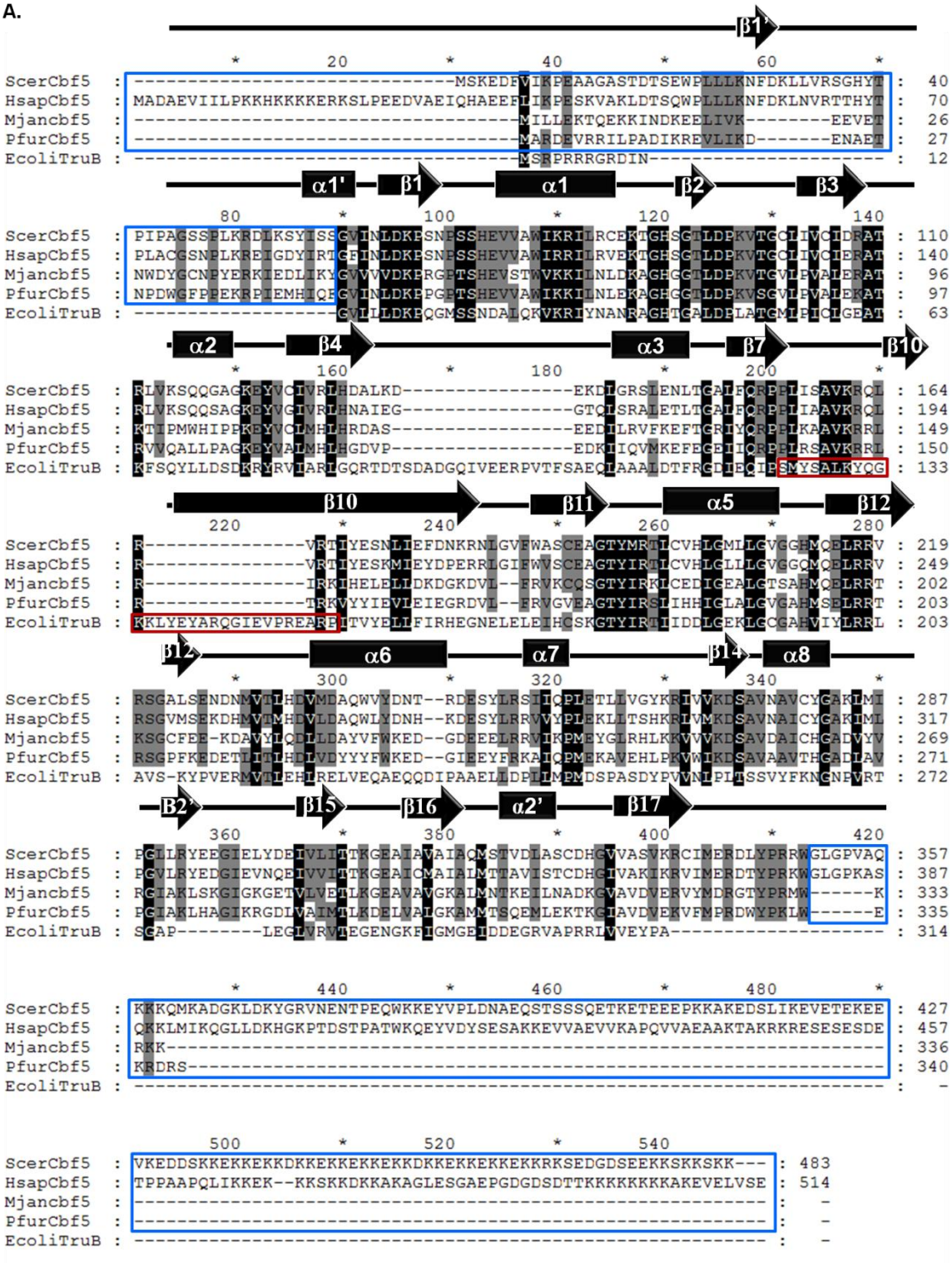
To date five pseudouridine synthase families within archaea, bacteria, and eukaryotes are known: RluA, RsuA, TruA, TruB, and TruD. Structural analysis of the five pseudouridine synthase families has shown a common core domain and a highly conserved active-site cleft [13]. The pseudouridine synthase core domain is composed of an eight stranded mixed β-sheet bisected by a highly conserved hydrophobic active-site cleft, flanked with numerous helices and loops [13]. Within the active site loop is a universally conserved aspartate residue. The aspartate found within the active site loop has been shown to be essential for catalysis in all five pseudouridine synthase families as substitution of this residue abolishes *in vivo* pseudouridine formation [21, 37]. In addition to the absolutely conserved aspartic acid, two other highly conserved polar residues are found within the active site cleft, a lysine or an arginine residue and an invariant tyrosine residue (except in TruD) [13]. The side chain of the lysine or arginine residue interacts with the invariant tyrosine residue and the catalytic aspartate residue to form a buried salt bridge [13]. Despite the shared core domain, the five pseudouridine synthase families share little sequence similarity and contain a variety of independent N- and C-terminal extensions.

E. coli TruB is responsible for the isomerisation of uridine 55 to pseudouridine within the T- loop of elongator tRNAs [14]. The TruB pseudouridine synthase family contains a C-terminal pseudouridine and archaeosine transglycosylase (PUA) domain and a thumb-like loop protrusion

flanking the active-site cleft, both of which are essential for pseudouridylation [36, 93]. Structural analysis of TruB and Cbf5 demonstrated that both share a similar structure, indicating that Cbf5 is a member of the TruB pseudouridine synthase family (Fig. 12) [15]. Like TruB, Cbf5 contains a large N-terminal catalytic domain and a small C-terminal PUA domain. The N-terminal catalytic domain of Cbf5 can be divided into two subdomains, D1 and D2, forming a catalytic cleft between them [24]. The active site lies towards the bottom of the catalytic cleft [35, 94]. The C-terminal PUA domain contains a mix of α-helix and β-strand secondary structures [35, 94].

While the overall structure of Cbf5 is similar to TruB, there are three fundamental structural differences between the two proteins. The N-terminus of yeast Cbf5 holds an additional extension compared to TruB allowing the N-terminus to wrap around the C-terminal PUA domain forming an extra β-strand (Fig. 12) [94]. The catalytic domain of TruB contains a well defined thumb loop which is responsible for the placement of the substrate RNA within the catalytic site of TruB (Fig. 12) [35, 94]. Cbf5 contains a thumb loop structure which is considerably smaller than TruB's [35, 94]. To compensate for this loss, H/ACA protein Gar1 interacts with Cbf5 within the thumb loop region allowing for proper placement of substrate RNA within the catalytic site [25]. The C-terminal PUA domain of Cbf5 is significantly larger than TruB's, containing a phylogenetically conserved extension elongating the PUA domain (Fig. 12) [35, 94]. The structural differences observed between Cbf5 and TruB might explain Cbf5's need of a complex RNP system for optimal catalytic activity.

A.


ScerCbf5
HsapCbf5
Mjancbf5
PfurCbf5
EcoliTruB
β1
α1'
α1
β2
β3
α2
β4
α3
β7
β10
β11
α5
β12
α6
α7
β14
α8
β2'
β15
β16
α2'
β17

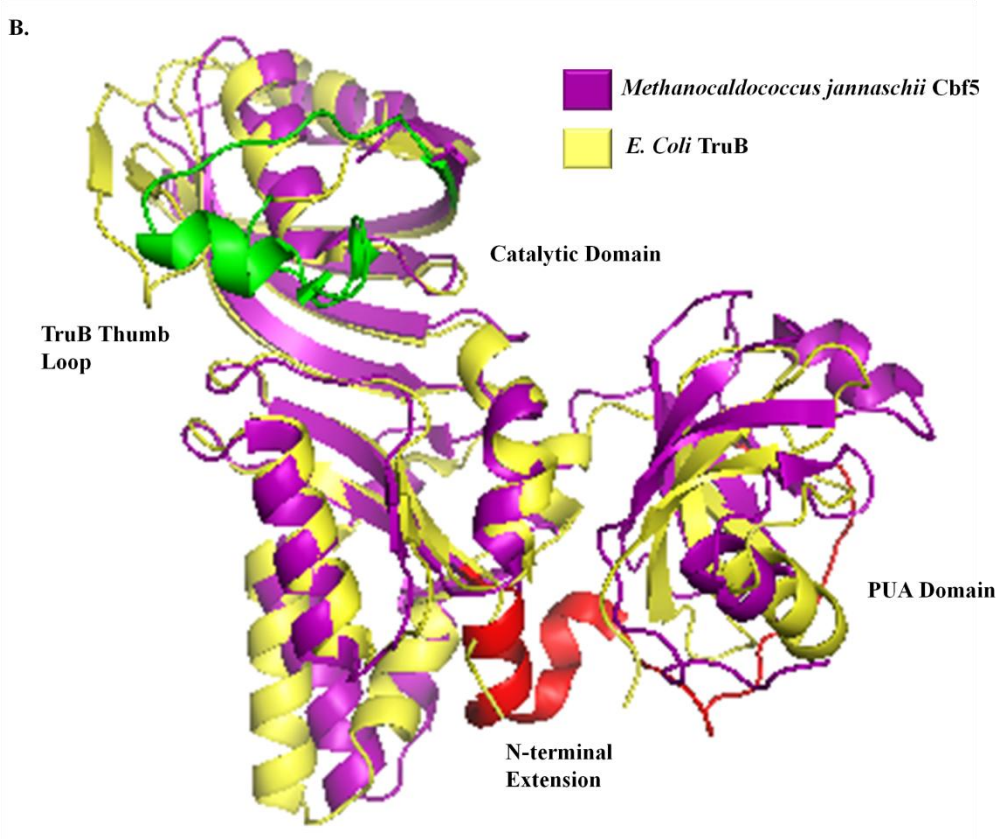


Figure 12: Sequence alignment and structure of H/ACA catalytic protein Cbf5. A. Multiple sequence alignment of Cbf5. Of the five aligned sequences two are eukaryotic: *Saccharomyces cerevisiae* (ScerCbf5) and *Homo sapiens* (HsapCbf5), two are archaeal: *Methanocaldococcus jannaschii* (MjanCbf5) and *Pyrococcus furiosus* (PfurCbf5), and one is bacterial: *Escherichia coli* (EcolTruB). The secondary structure elements are shown above the sequences, with α-helices depicted as rectangles and β-strands as arrows. The secondary structural elements indicated above the sequence align are determined by Li *et al.* [24]. Completely conserved residues among the sequences aligned are shaded in black, and residues with at least 60% conservation are shown in grey. N-terminal and C-terminal extensions are outline by blue boxes and the *E. coli* TruB thumb loop is outlined in red. The alignment was prepared using ClustalW (EMBL-EBI) and Gene Doc (Dr. Nicholas Hugh). **B.** Superposition of *M. jannaschii* Cbf5 (purple, PDB 2APO [35]) and *E. coli* TruB (yellow, PDB 1K8W [36]) crystal structures. The N-terminal Cbf5 extension is indicated in red; the C-terminal extension is not resolved in crystal structure. The TruB thumb loop is indicated in green. The structural alignment was made with Pymol Molecular Visualization Evaluation Product (Delano Scientific LLC).

Cbf5 is the Catalytic Protein of Archaeal and Eukaryotic H/ACA snoRNPs

Early characterization of Cbf5 strongly suggested that Cbf5 was the H/ACA snoRNP catalytic enzyme as Cbf5 shares significant sequence similarity with the *E. coli* pseudouridine synthase TruB and to the rat nucleolus protein NAP57 [33, 95-96]. In 1998 the Tollervey group confirmed

Cbf5 as the catalytic component of H/ACA snoRNPs, demonstrating that depletion of Cbf5 results in the loss of pseudouridine formation and the reduction of H/ACA structural protein accumulation [14]. In addition, the Carbon group in 1999 further demonstrated that Cbf5 was indeed the catalytic residue of H/ACA snoRNPs using mutational studies in which the proposed catalytic aspartate residue was mutated [97]. Biochemical disruption studies and immunoprecipitation assays of eukaryotic H/ACA snoRNPs indicated that Cbf5 forms specific interactions with H/ACA snoRNAs and the H/ACA structural proteins Gar1, Nhp2, and Nop10, in which each set of H/ACA proteins may potentially interact with both H/ACA snoRNA hairpins (Fig. 13) [14, 31, 91]. Biochemical studies and crystallographic analysis indicate that Cbf5 interacts independently with both Nop10 and Gar1, while the recruitment of structural protein Nhp2 is contingent on the formation of a Cbf5-Nop10 complex [24, 46, 91].

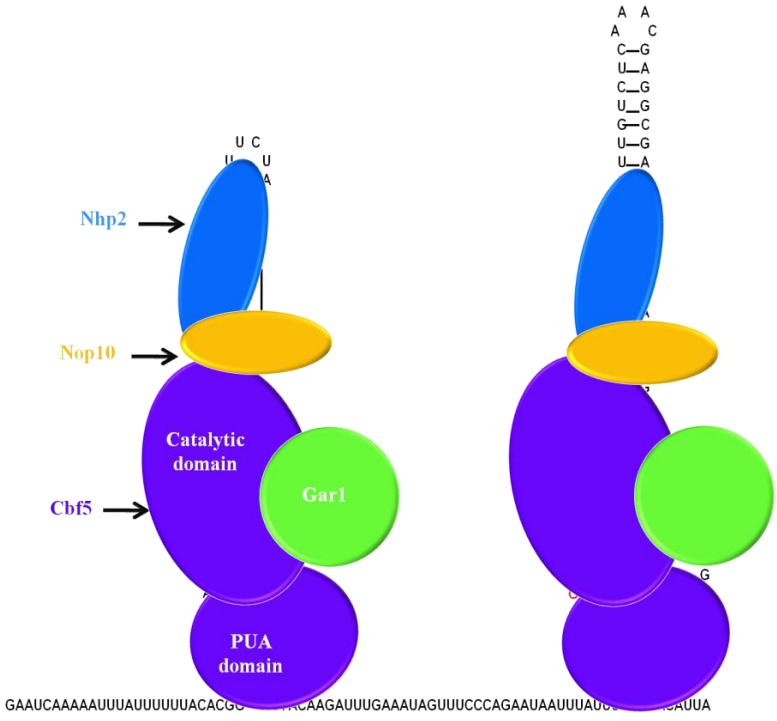


Figure 13: Proposed model of a eukaryotic H/ACA snoRNP complex. Studies of the archaeal H/ACA snoRNP system suggest that the H/ACA core components will interact in a similar manner in eukaryotes as they do in archaea, with a copy of the core H/ACA proteins assembling at each hairpin motif of the snoRNA.

H/ACA Structural Protein Nop10 Interacts with the Catalytic Domain of Cbf5

H/ACA structural protein Nop10 is a small two-domain protein consisting of an N-terminal β-domain with two closely packed β-hairpins and a C-terminal α-helix domain [15, 35]. The two domains are connected by a highly conserved, unstructured linker region forming a random coil. On its own, Nop10 is a disordered protein; however, upon binding to Cbf5, Nop10 becomes highly structured, interacting with both Cbf5's catalytic D1 and D2 subdomains [15, 24, 35]. Many highly conserved residues within the linker region of Nop10 associate with residues in Cbf5 which are also highly conserved [35]. The C-terminal α-helix of Nop10 interacts with Cbf5's D1 α6 helix forming a coiled-coil, while the N-terminal β-domain of Nop10 associates with Cbf5's β10-β12 central β-sheet [15, 24]. The interactions formed between Nop10's C-terminal domain and the D1 subdomain of Cbf5 within eukaryotes are required for both the association of H/ACA structural protein Nhp2 and stable binding of the H/ACA snoRNA and substrate RNA [15]. The association of Nop10 to Cbf5 enables a universally conserved lysine residue of Cbf5 to form hydrogen bonds with Cbf5's catalytically aspartate residue forming a catalytic active Cbf5 enzyme. This suggests that Nop10 could act as a molecular switch between catalytically active and inactive forms of Cbf5, thereby modulating the enzymatic dynamics of the active site of Cbf5 [35].

H/ACA Structural Protein Gar1 Interacts with the Catalytic Domain of Cbf5

H/ACA structural protein Gar1 is a member of the reductase, isomerase and elongation factor fold superfamily, containing a characteristic six-stranded β-barrel structure. Gar1 interacts with Cbf5's catalytic subdomain D2. Gar1's β1/β2 loop and β4/β5 loop associate with Cbf5's β7/β10 hairpin loop [15]. The β7/β10 hairpin loop of Cbf5 (also called the thumb loop) contains a highly conserved arginine rich region [25]. The thumb loop of Cbf5 plays an essential role in substrate recruitment and release, undergoing a conformational change between a substrate-free Gar1-

bound open state to a substrate-bound loading state [25]. Mutational studies of Gar1 have shown that depletion of Gar1 not only prohibits product release after catalysis, but also inhibits substrate loading and further catalysis, supporting the idea that Gar1 plays a critical role in substrate loading and turnover [25].

Cbf5 Interacts with H/ACA snoRNAs

As previously mentioned, H/ACA snoRNAs fold into a characteristic secondary structure: eukaryotic H/ACA snoRNAs contain two irregular hairpin structures, linked by a hinge sequence (Box H) followed by a short 3' tail (Box ACA) [23, 98]. Box H, following the 5' hairpin, consists of an ANANNA sequence; while Box ACA, located three nucleotides upstream of the 3' end, consists of an ACA sequence [89]. Each hairpin structure contains a large internal loop, the pseudouridylation pocket, which is complementary to the substrate RNA [23, 26]. H/ACA snoRNAs make several critical contacts with the H/ACA protein complex, specifically with Cbf5's catalytic domain and PUA domain. The snoRNA binds along the front face of the protein complex, where the upper hairpin structure interacts with the composite surface of Cbf5-Nop10-L7Ae (Nhp2 in eukaryotes) while the lower hairpin structure and the ACA sequence is bound by Cbf5's PUA domain [24]. The binding of the snoRNA on either end anchors the snoRNA on the protein complex, positioning the pseudouridylation pocket in close proximity to Cbf5's active site cleft [24].

Dyskeratosis Congenita and Cbf5

H/ACA snoRNPs have been linked to the rare inherited disease DC. DC is a clinically and genetically heterogeneous disease and is characterized by premature aging, nail dystrophy, mucosal leukoplakia, an increased chance of malignancies, and often fatal bone marrow failure [99]. To date three forms of Dyskeratosis congenita have been characterized: X-linked, autosomal dominant and autosomal recessive. X-linked, the most common and severe form, is attributed to

point mutations within the human homolog of Cbf5, dyskerin [70]. The majority of these DC mutations have been mapped to the PUA domain of Cbf5 with additional mutations being localized to the catalytic domain [15, 24, 35]. The DC substitutions within the PUA domain of Cbf5 form a spatial mutational cluster [15, 24, 35]. While the effect of such a mutational hot spot is not fully understood, it is believed that these substitutions either directly disrupt Cbf5's interaction with H/ACA snoRNAs or compromise the structural stability of Cbf5 [15, 24]. A recent publication by the Meier group showed that DC substitutions within NAP57 (the rat homolog of Cbf5) affect the interaction between NAP57 and the H/ACA snoRNP accessory protein SHQ1 [100]. DC substitutions in NAP57 causes an imbalance in the NAP57-SHQ1 interaction which may result in the sequestering, improper folding or degradation of newly synthesized NAP57 [100]. This would directly lead to the impairment of H/ACA snoRNP biogenesis which in turn could be also connected to a decrease in H/ACA snoRNP accumulation [100].

III.II Objective

H/ACA small ribonucleoproteins (RNPs) catalyze the formation of pseudouridines in many different RNAs in archaea and eukaryotes. The H/ACA RNP complex is comprised of four highly conserved core proteins, Cbf5, Gar1, Nop10, and L7ae (archaea), as well as one of several different H/ACA snoRNAs. The H/ACA snoRNA guides substrate RNA modification by site-directed base-pairing while Cbf5 catalyzes pseudouridylation. Mutations in Cbf5 have been shown to cause the rare X-linked disease Dyskeratosis congenita. In eukaryotes, pseudouridine formation plays an essential role in pre-ribosomal RNA processing. The modification of rRNA is the first important step after transcription in the overall process of ribosome biogenesis. Despite its importance, the detailed mechanism of pseudouridine formation by H/ACA RNPs is not fully understood. To understand the mechanism of H/ACA RNPs, it is necessary to gain detailed knowledge of how the components interact with one another.

Much of what is currently known and hypothesized about eukaryotic H/ACA snoRNPs has been obtained from the *in vitro* study of the homologous archaeal H/ACA sRNP complex. While this information is invaluable, fundamental questions regarding stoichiometry, association, and catalytic efficiency of H/ACA snoRNPs have yet to be answered within a eukaryotic system. The reconstitution of a functionally active *S. cerevisia*e H/ACA snoRNP complex will allow us to address these essential, unanswered questions using powerful *in vitro* studies. As Cbf5 is the central catalytic protein within H/ACA snoRNPs, it is critical that we begin to understand how Cbf5 interacts with the other H/ACA snoRNP components. To study such interactions, a recombinant *S. cerevisiae* Cbf5 protein must first be successfully expressed and purified.

III.III Materials and Methods

Molecular cloning of H/ACA catalytic protein Cbf5

The gene encoding H/ACA catalytic protein Cbf5 was amplified from *S. cerevisiae* genomic DNA and a 5'-*EcoRI* and 3'-*HindIII* restriction sites were introduced by PCR using the following primer pairs:

Cbf5 F: GAATTCGTCAAAGGAGGATTTCGTTATTA

Cbf5 R: AAGCTTTCATTTCTTAGATTTCTTAGATTTCC

PCR reactions were carried out in 50 µL volumes with final concentrations as follows: 1x Pfu buffer with $MgSO_4$, 200 µM dNTPs, 0.5 µM of each primer, 0.02 U/µL *Pfu* DNA polymerase, and 2.0 µL *S. cerevisiae* genomic DNA. PCR cycle parameters included an initial denaturation at 98 ^{o}C for 5 min followed by 5 cycles of denaturation at 98 ^{o}C for 30 s, touch-up annealing starting for 30 s at 41.5 ^{o}C with a 1 °C increase in each cycle, and extension at 72 °C for 1 min, with an additional 30 cycles of denaturation at 98 ^{o}C for 30 s, annealing at 46.5 ^{o}C for 30 s, and extension at 72 ^{o}C for 1 min followed by a final extension at 72 ^{o}C for 10 min. PCR products were analyzed on a 1% agarose gel and visualized by ethidium bromide staining. The PCR

product was purified with an EZ-10 Spin Column PCR Products Purification Kit (Bio Basic Inc.), following the manufacturer's instructions and blunt end ligated into the *Sma*I cleaved cloning vector *pUC19* [88]. Ligated *pUC19-ScCbf5* was transformed into competent high-efficiency *E. coli* DH5α cells (NEB). Transformation mixtures were plated on X-Gal-treated LB agar plates containing 0.1 mg/mL ampicillin [88]. Plates were incubated overnight at 37 °C. White *E. coli* colonies (positive for gene insertion) were transferred into 5 mL of LB containing 0.1 mg/mL ampicillin and grown overnight at 37 ^{o}C. Plasmid DNA from the resulting cell cultures was isolated using an EZ-10 Spin Column Plasmid DNA Minipreps Kit (Bio Basic Inc.), following the manufacturer's instructions. Isolated plasmid DNA was tested for positive gene insertion through restriction digestion with *EcoRI*. 20 µL restriction digestion reactions with *pUC19-ScCbf5* (*EcoRI*) were incubated at 37 ^{o}C for 2 hr with final component concentrations of: 1x Tango Buffer (Fermentas), 2 µg *pUC19-ScCbf5* plasmid DNA, and 2 U/µL *EcoRI* restriction enzymes (Fermentas). Restricted plasmids were analyzed on a 1% agarose gel at 110 V for 1hr and visualized with ethidium bromide staining [88]. Clones appearing to be positive were sent to Macrogen Inc. for sequencing. Positively sequenced clones were then maxiprepped and used for further molecular cloning of *CBF5*.

pUC19-Cbf5 plasmid DNA was purified with EZ-10 Spin Column PCR Purification Kit (Bio Basic Inc.). *CBF5* was then PCR amplified with a primer pair introducing a 5'- *BamHI* and 3'- *Eco*RI restriction site (underlined in primer sequences) at the 5' and 3' end of the Cbf5 gene:

Cbf5F: CGTGGGATCCCCTCAAAGGAGGATTTCGTTATTA

Cbf5 R: CCGGGAATTCTCATTTCTTAGATTTCTTAGATTTC

PCR reactions were carried out in 50 µL volumes with final concentrations as specified above. PCR cycle parameters included an initial denaturation at 98 ^{o}C for 5 min followed by 6 cycles of denaturation at 98 ^{o}C for 30 s, touch-up annealing starting for 30 s at 48 ^{o}C with a 1 °C increase

in each cycle, and extension at 72 °C for 1 min, with an additional 30 cycles of denaturation at 95 ^{o}C for 30 s, annealing at 53 ^{o}C for 30 s, and extension at 72 ^{o}C for 1 min followed by a final extension at 72 ^{o}C for 10 min. PCR products were once again analyzed on a 1% agarose gel and visualized by ethidium bromide staining and purified using an EZ-10 Spin Column PCR Products Purification Kit (Bio Basic Inc.) [88]. Purified *CBF5* was restricted with *Eco*RI and *BamHI* restriction enzymes (Fermentas) in a 100 µL reaction volume for 4 hours at 37 ^{o}C with a final components concentration of: 2X Tango Buffer (Fermentas), 2 µg *CBF5*, 2 U/µL *Eco*RI restriction enzyme (Fermentas), 4 U/µL *BamHI* restriction enzyme (Fermentas). *pGEX-5x-3* plasmid was restricted in the same manner and was subsequently gel extracted using a Silica Bead DNA Gel Extraction kit (Fermentas) following the manufacturer's instructions. Purified *Cbf5* was sticky end ligated into purified *Eco*RI and *BamHI* restricted expression vector *pGEX-5X-3* [88]. Ligated *pGEX-5X-3-Cbf5* was subsequently transformed into high-efficiency competent *E. coli* DH5α cells (NEB) and plated on LB agar containing 0.1 mg/mL ampicillin. Plates were incubated overnight at 37 ^{o}C from which resulting colonies were selected and grown overnight once again at 37 ^{o}C in 5 mL LB. Plasmid DNA from the selected colonies was isolated by miniprep as described previously and subjected to *Eco*RI restriction digestion to confirm for positive clones. 20 µL *Eco*RI restriction volumes were incubated at 37 ^{o}C for 1 hr and analyzed by 1% agarose gel electrophoresis (both as previously discussed). Plasmids positive for *CBF5* insertion were sent to Macrogen Inc. for sequencing.

Optimization of Cbf5 recombinant expression

E. coli BL21 (DE3) Rosetta cells containing the *pGEX-5x-3-ScCbf5* plasmid were grown overnight at 37^{o}C in 100 mL of LB medium (1% (w/v) tryptone, 0.5% (w/v) yeast extract, and 1% (w/v) NaCl pH 7.0) containing 0.1 mg/mL of ampicillin. Cells were pelleted and resuspended in 50 mL of fresh LB medium containing 0.1 mg/mL of ampicillin. Subsequently, 2 L of LB medium containing 0.1 mg/mL of ampicillin were inoculated to a cellular density of

approximately 0.1 OD_{600}. Cell cultures were incubated at 37°C and the OD_{600} was recorded every hour. Once a cellular density of 0.4 OD_{600} was reached, 1 mM of IPTG was added to induce expression of GST-Cbf5 fusion protein. The cell density was recorded for three additional hours with cell samples of 1 OD_{600} being taken at times 0 hr, 0.5 hr, 1.0 hr, 2.0 hr, and 3.0 hr after induction. After 3 hr of growth, cell cultures were pelleted by centrifugation, shock frozen with liquid nitrogen and stored at -80 °C. Cell samples were suspended in 80 µL of cell lysis buffer (0.1 M Tris-HCl pH 8.5 containing 8 M urea) and boiled for 5 minutes at 95 °C to open cells. A cellular sample containing GST-tagged Cbf5 was also taken and subjected to solubility testing. All samples were analyzed by 12 % SDS-PAGE and visualized using Coomassie stain.

To optimize GST-Cbf5 protein solubility at the expression level, the temperature of cell growth was reduced to 20 °C and 2% ethanol was added prior to induction. *E. coli* BL21 (DE3) Rosetta cells containing the *pGEX-5x-3-ScCbf5* plasmid were grown and prepared for recombinant expression in the same manner as discussed above. Cultures were placed into a refrigerated cell growth chamber 1 hr prior to IPTG induction and equilibrated to a temperature of 20 °C. Immediately preceding induction, cell cultures were treated with 2% ethanol to cause cellular stress. Upon induction with IPTG, cells were permitted to grow for an addition 17 hr at 20 °C, with cell density reading and cell samples taken at times 1 hr and at 17 hr after induction. A cellular sample containing GST-tagged Cbf5 was taken and subjected to solubility testing. All samples were again prepared and analyzed as previous described.

Solubility Testing of GST-Cbf5

1 g of *E. coli* BL21 (DE3) Rosetta cells containing GST-tagged Cbf5 was resuspended in 3 mL of cell lysis buffer (50 mM Tris-Cl (pH 8.0), 1 mM EDTA, 10 mM NaCl). 0.13 mM PMSF and 0.27 mg/ml lysozyme was added to resuspended cells and incubated for 20 min at 4 °C. 4 mg of sodium deoxycholate was then added followed by incubation for an additional 10 min at 4 °C.

Once the cell lysate has become viscous, 0.0067 mg/ml DNase I was added followed by incubation until the cell lysate was no longer viscous (~30 min). The cell lysate was centrifuged at 14,000 rpm for 15 min at 4°C. The isolated supernatant was saved for SDS-PAGE analysis as the soluble protein fraction. The pelleted cell debris was washed with 1 mL H_20 and resuspended in 1 ml of 0.1 M Tris-Cl (pH 8.5) containing 8 M urea. Solubilized protein was separated from remaining insoluble protein through centrifugation, where the resulting supernatant was saved for SDS-PAGE analysis as the urea-dissolved fraction. The remaining insoluble fraction was resuspended in 1 mL H_20 and saved for SDS-PAGE analysis as the insoluble inclusion body fraction. All saved samples were analyzed by 12% SDS-PAGE for the presence of GST-Cbf5.

Optimization of Cbf5 Purification

Recombinantly expressed, GST-tagged Cbf5 was initially purified by glutathione-sepharose affinity chromatography. *E. coli* BL21 (DE3) Rosetta cells containing GST-tagged Cbf5 were resuspended in cell lysis buffer (1 mL/g of cells) containing: 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10 mM $MgCl_2$, 0.5 mM EDTA, 0.1 mM PMSF, 1 mM β-Mercaptoethanol and 5% glycerol. 1 mg/ml lysozyme and 12.5 mg/g (of cells) of sodium deoxycholate were then added to the resuspended cells. Cells were opened by sonication using an intensity level of 6, and duty cycle of 60% for 1 min repeated 5 times. The cell debris was pelleted at 30 000 xg for 30 min at 4 ^oC. The cleared cell lysate was incubated with 5 mL of glutathione-sepharose (GE Healthcare) for 1 hr at room temperature. All unbound protein was removed by washing of the glutathione sepharose with 3 times the column volume of cell lysis buffer, through centrifugation at 500 xg for 5 min at 4 ^oC. The washing step was repeated 5 times. GST-Cbf5 was eluted from the glutathione sepharose with ½ the column volume of elution buffer (50 mM Tris-HCl, pH 8.0, 10 mM reduced glutathione, and 10% glycerol) incubated for 10 min at room temperature. Eluted GST-Cbf5 was isolated from the glutathione sepharose through centrifugation at 500 xg for 5 min

at 4 ^{o}C. GST-Cbf5 elution was repeated 6 times. 50 µL samples were collected throughout the purification procedure and were analyzed as previous described.

To optimize the solubility of GST-Cbf5 during purification, sarkosyl, CHAPS and triton X-100 were added to the purification buffers. Cell lysis and opening was performed as previously described. The cellular debris pellet containing insoluble GST-Cbf5 was resuspended in 15 mL of cell lysis buffer containing 10% (w/v) sarkosyl and incubated for 15 min at room temperature. Solubilized GST-Cbf5 was separated from remaining insoluble cell debris through centrifugation at 30 000 xg for 30 min at 4 ^{o}C. The resulting supernatant was then diluted 1:10 with cell lysis buffer, thereby diluting the sarkosyl concentration to 1% (w/v). 2% (v/v) Triton X-100 and 20 mM CHAPs was subsequently added to the diluted supernatant. The diluted supernatant was incubated with 5 mL of glutathione-sepharose (GE Healthcare) for 1 hr at room temperature. All unbound protein was removed through the washing of the glutathione sepharose with 3 times the column volume of cell lysis buffer containing 1% (w/v) sarkosyl, 2% (v/v) Triton X-100 and 2 mM CHAPS for the first 2 washes, while cell lysis buffer containing 1% (v/v) Triton X-100 and 10% (v/v) glycerol was used for the last 3 washes. Wash solution and glutathione sepharose resin were separated through centrifugation at 500 xg for 5 min at 4 ^{o}C. GST-Cbf5 was eluted from the glutathione sepharose with ½ the column volume of elution buffer (50 mM Tris-HCl, pH 8.0, 10 mM reduced glutathione, 10% glycerol, and 0.5% (v/v) Triton X-100) incubated for 10 min at room temperature. Eluted GST-Cbf5 was isolated from the glutathione sepharose through centrifugation at 500 xg for 5 min at 4 ^{o}C. The elution of GST-Cbf5 was repeated 6 times. 50 µL samples were collected throughout the purification procedure and were prepared and analyzed as previous described.

III.IV Results

Optimization of the recombinant expression of *S. cerevisiae* H/ACA protein Cbf5

In an effort to optimize the recombinant expression efficiency and protein solubility of *S. cerevisiae* H/ACA catalytic protein Cbf5, multiple protein expression conditions were tested. For comparison, the growth of *E. coli* BL21 (DE3) Rosetta cells containing the plasmid *pGEX-5x-3-ScCbf5* was monitored at 37 °C for seven hours, at 20 °C for seven hours and at 20 ^{o}C in the presence of two percent ethanol for seventeen hours (Fig. 14D). A protein solubility test was subsequently performed to determine the effect of the different cell growth conditions of *E. coli BL21 (DE3)* Rosetta cells on GST-Cbf5 protein solubility. The solubility test separates the soluble, urea-soluble and insoluble protein fractions. Samples were collected during each protein expression step and throughout the protein solubility test and were then analyzed by SDS-PAGE.

First, expression and solubility levels of GST-Cbf5 were tested at 37 °C, i.e. at standard expression conditions in *E. coli*. As shown in Figure 14A, a band of ~80 kDa increases in intensity over a three hour time period, indicating that recombinant expression of GST-Cbf5 is successful at 37 °C. However, samples analyzed from the subsequent solubility test (Fig. 14A) indicated that GST-Cbf5 is predominantly in the insoluble pellet and urea dissolved fraction. In an effort to increase the solubility of GST-Cbf5, two cellular growth conditions were varied: the temperature at which the cells were grown upon induction was lowered, and the addition of ethanol upon induction was tested as well.

Recombinant expression at lower temperatures was first tried as it has been well documented that lower cell growth temperatures reduces cellular growth (observed in Fig. 14D) resulting in the slowing of all cellular processes including protein folding. While the use of a lower temperature appears to increase GST-Cbf5 solubility, some protein was still observed in the insoluble fraction (Fig. 14B). The introduction of ethanol during recombinant expression has been reported to induce cellular stress which is thought to enhance the expression of chaperones to

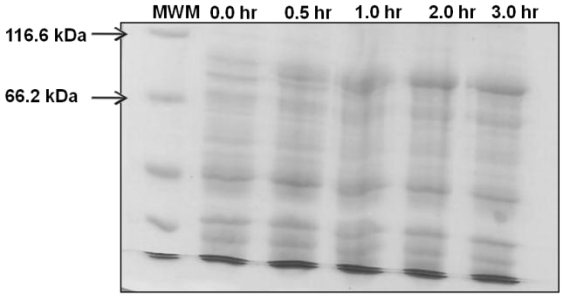
assist in protein folding [101-102]. The utilization of both lower cell growth temperatures and the introduction of ethanol during protein expression were believed to further aid in the folding of GST-Cbf5 and thus improve the solubility of the protein. As both reduced growth temperatures and the addition of a foreign stress appear to affect the rate of cellular growth (Fig. 14D), cells were permitted to grow for seventeen hours after induction. Analysis of the cellular samples collected during protein expression and during the solubility test indicate that prolonged cellular growth increases GST-Cbf5 expression levels and that the combination of a reduced growth temperature and the addition of two percent ethanol increases protein solubility levels (Fig. 14C).

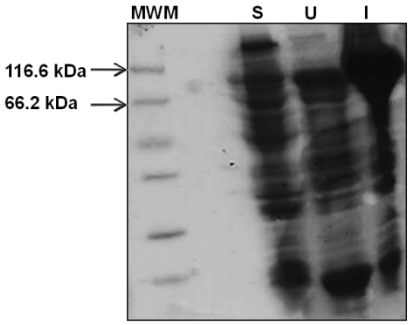
Optimization of the purification of *S. cerevisiae* H/ACA protein Cbf5

As previously discussed, GST-Cbf5 is predominantly found within the insoluble cell pellet when expressed at standard 37 ^{o}C conditions, presenting a significant problem in the purification of the fusion protein. To address this problem at the protein expression level, all cells were grown for seventeen hours at 20 ^{o}C and in the presence of two percent ethanol. GST-tagged Cbf5 was subsequently purified using glutathione sepharose affinity chromatography.

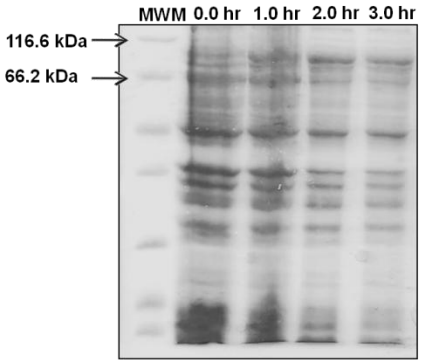
E. coli BL21 (DE3) *Rosetta* cells containing GST-Cbf5 were opened in a cell lysis buffer, and the resulting cellular lysate was incubated with glutathione-sepharose affinity chromatography matrix. Following washing with cell lysis buffer, GST-Cbf5 was subsequently eluted from the glutathione sepharose with elution buffer containing reduced glutathione. Samples were taken at every step during the purification procedure and were analyzed by SDS-PAGE. A band of approximately 80 kDa corresponding to GST-Cbf5 was expected to be observed throughout the purification; however, a band of 80 kDa was only observed within the insoluble cellular pellet (Fig. 15A). This indicates that insufficient amounts of GST-Cbf5 are soluble for protein purification.

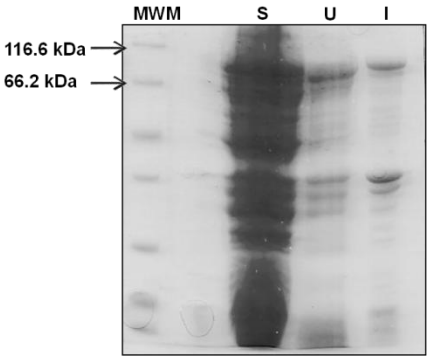
A. Over expression at 37 °C



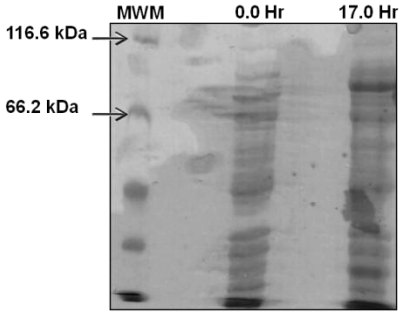


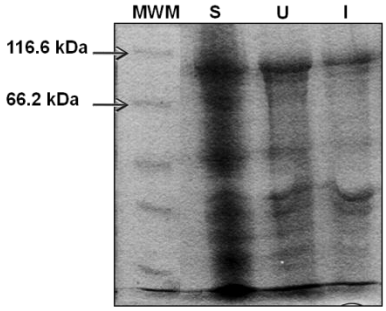
B. Over expression at 20 °C





C. Over expression at 20 °C with 2% ethanol





D. Cellular growth curves for GST-Cbf5 recombinant expression using the three discussed growth conditions

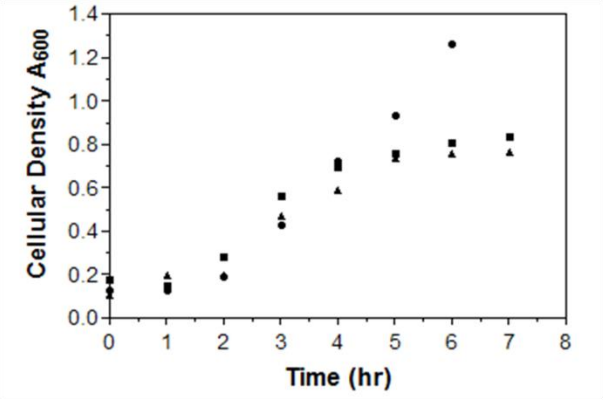


Figure 14: Optimization of the recombinant expression of *S. cerevisiae* Cbf5 in *E. coli*. A. Recombinant expression at 37 ^{o}C and solubility test of GST-Cbf5 wild-type in *E. coli* BL21 (DE3) Rosetta cells. Cell samples were taken for three hours after induction and analyzed by 12 % SDS – PAGE. GST-Cbf5 fusion protein has a molecular weight of ~80 kDa. Soluble cell lysate (S), the urea-dissolved fraction of the cell pellet (U) and the insoluble fraction (I) were analyzed by 12% SDS-PAGE. MWM: molecular weight marker. **B.** Recombinant expression at 20 ^{o}C and solubility test of GST-Cbf5 wild-type. Cell samples were taken at induction and for three hours after induction and analyzed by 12 % SDS – PAGE. The solubility of expressed GST-Cbf5 was analyzed as before. **C.** GST-Cbf5 wild-type recombinant expression at 20 ^{o}C with the addition of 2% ethanol. Cell samples were taken upon induction and seventeen hours after induction and analyzed by 12 % SDS – PAGE. The solubility of expressed GST-Cbf5 was analyzed as before. **D.** Cellular growth rates of *E. coli BL21* (*DE3*) Rosetta cells under the three growth condition were plotted to compare cellular growth rates. Cell growth at 37 ^{o}C (circles), at 20 ^{o}C (triangles), and at 20 ^{o}C with 2% ethanol (squares).

While the GST-Cbf5 protein solubility was significantly improved by expression at decreased temperatures and the addition of two percent ethanol, the solubility was still not sufficient to obtain detectable amounts of GST-Cbf5 after purification by glutathione sepharose affinity chromatography. In an effort to optimize GST-Cbf5 purification, a combination of detergents, sarkosyl, CHAPS and Triton X-100, were employed as this detergent combination had been previously reported to aid in protein solubility during purification of GST-fusion proteins [103]. The cell pellet was first treated with sarkosyl in order to solubilize GST-Cbf5 (Fig. 15B, lane Spt. AS). CHAPS and Triton X-100 were subsequently added to the newly solubilized fusion protein enabling GST-Cbf5 to quantitatively bind to the glutathione sepharose chromatography column as confirmed by the lack of a GST-Cbf5 band in the supernatant after column binding (Fig. 15B, lane Spt. AB). The glutathione sepharose was washed with two different wash buffers in which the first contained all three detergents and the second contained only Triton X-100 and 10% glycerol (Fig. 15B, lanes W1-5). Two wash buffers differing slightly in their composition were used in hope of maintaining GST-Cbf5 solubility while removing all unbound proteins. GST-Cbf5 was then eluted with an elution buffer containing reduced glutathione; however, SDS-PAGE analysis of samples taken from the elution step did not reveal any GST-Cbf5 (Fig. 15B, lanes E1-4).

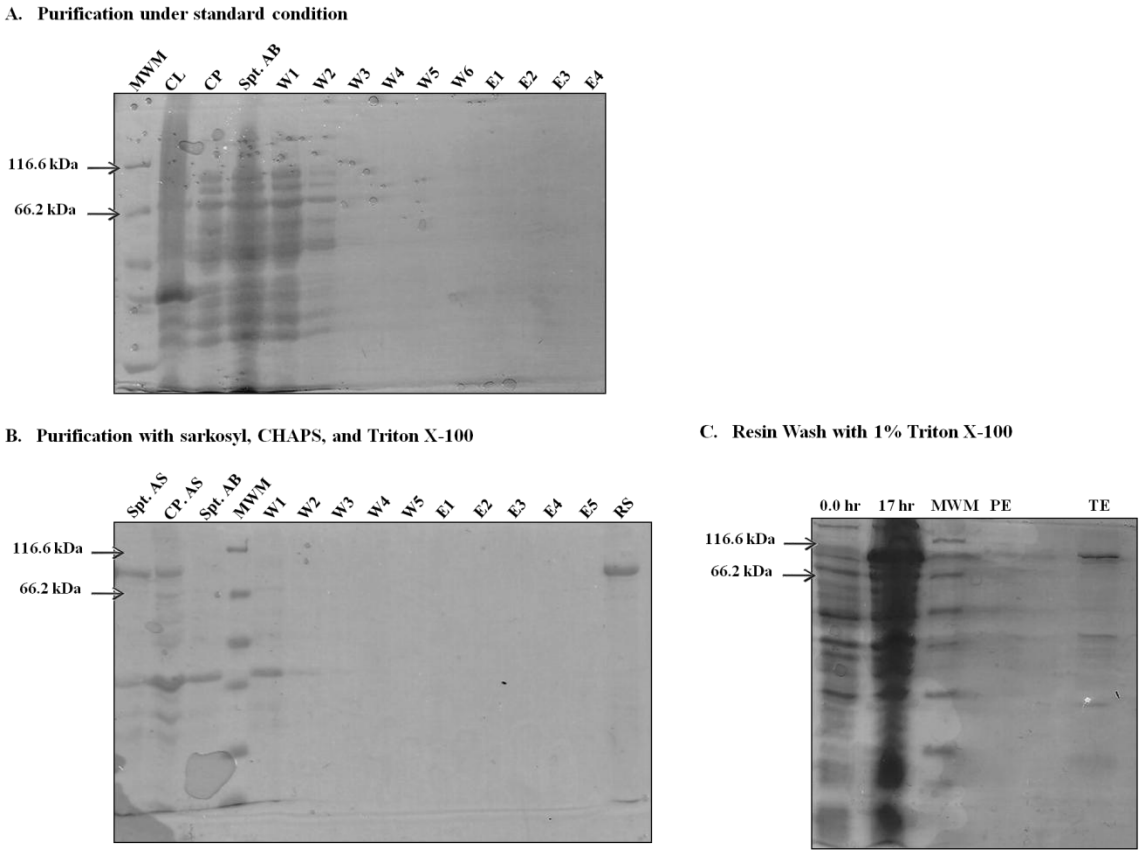


Figure 15: Optimization of the purification of *S. cerevisiae* Cbf5. A. Glutathione sepharose affinity chromatography purification of GST-Cbf5 fusion protein using standard purification conditions (described in Materials and Methods). During the purification, samples were collected from the cell lysate (CL), cell pellet (CP), supernatant after resin binding (Spt. AB), washes (W), and elutions (E) and analyzed by 12 % SDS-PAGE. **B.** Glutathione sepharose affinity chromatography purification of GST-Cbf5 fusion protein using the detergent sarkosyl to extract GST-Cbf5 from the cell pellet, and the detergents CHAPs and Triton-X-100 to increase GST-Cbf5's solubility in the binding and washing buffer (for details see Materials and Methods). Samples were taken throughout the purification and analyzed: supernatant after sarkosyl treatment (Spt. AS), cell pellet after sarkosyl treatment (CP. AS), and a resin sample taken after GST-Cbf5 elutions (RS). **C.** The glutathione sepharose column after the purification shown in B. was treated with 1% Triton X-100 to remove all unspecifically bound proteins. A sample was taken from supernatant after Triton X-100 treatment (TE) and compared with samples taken during GST-Cbf5 over-expression and with a sample taken from the pool GST-Cbf5 elutions (PE).

In contrast, analysis of a sample of the glutathione sepharose resin showed that a significant amount of GST-Cbf5 remained on the resin despite treatment with reduced glutathione (Fig. 15B, lane RS), suggesting that GST-Cbf5 was non-specifically aggregating on the glutathione sepharose column. This hypothesis could be confirmed by treatment of the glutathione sepharose resin with 1% Triton X-100 which denatured and solubilised GST-Cbf5

(Fig. 15C). In summary, this experiment shows that GST-Cbf5 can be successfully solubilized from the cell pellet by sarkosyl and that it binds to the glutathione sepharose column under the described conditions. The elution step remains to be optimized as GST-Cbf5 seems to precipitate on the glutathione column during the washing steps which gradually remove the detergents (sarkosyl and CHAPS).

III.V Discussion

Optimization of recombinant expression of GST-Cbf5 fusion protein

Recombinant expression of proteins in *E. coli* is dependent on numerous features throughout the process of gene expression including the transcription of a stable mRNA transcript containing a correct translation initiation start site and the proper translation and folding of polypeptide chains into functional proteins structures [104-105]. However, high levels of expression in *E. coli* often lead to the aggregation of recombinant proteins as insoluble inclusion bodies as a result of improper protein folding. Like many other eukaryotic proteins, recombinantly expressed in *E. coli*, GST-Cbf5 experienced significant issues with protein solubility levels as the majority of GST-Cbf5 was found within insoluble inclusion bodies. GST-Cbf5 was expressed in *E. coli* BL21 (DE3) Rosetta cells. BL21 (DE3) Rosetta cells contain the genes for tRNA argU, argW, ileX, glyT, leuW, proL, metT, thrT, tyrU, and thrU and provide supplementation for the rare eukaryotic codons AGG, AGA, AUA, CUA, CCC, and GGA. Expression of GST-Cbf5 in *E. coli* BL21 (DE3*)* Rosetta cells at standard 37 ^{o}C expression conditions was not sufficient to produce soluble protein, as the majority of GST-Cbf5 remained in the insoluble inclusion body fraction.

It has been previously reported that the addition of approximately two percent ethanol to *E. coli* cells during recombinant protein expression can increase the production of soluble recombinant proteins [101]. It is believed that the addition of ethanol increases cellular stress levels within the *E. coli* cells resulting in the increased production of *E. coli* chaperonins GroEL

and GroES [101]. High concentrations of chaperones within the cell would increase the potential for the correct folding of newly synthesized proteins, thus increasing the production of soluble recombinant proteins. It has also been reported that the reduction of cellular growth temperatures during protein expression can assist in the synthesis of soluble recombinant proteins. The reduction of cellular growth temperatures slows replication, transcription and translation in the cell while increasing the activity and expression levels of many *E. coli* chaperones [102]. The slowing of cellular processes and the increase in chaperone availability at reduced growth temperature provides a more stable environment for protein expression with a greater potential for efficient protein folding [102]. Accordingly, the expression of GST-Cbf5 under reduced cellular growth temperatures and with the addition of two percent ethanol significantly improved the solubility of the recombinant fusion protein. However, as a result of the lower growth temperature, the duration of recombinant protein expression had to be extended from three hours to seventeen hours to allow an adequate amount of GST-Cbf5 to be produced.

Another possible approach to increasing the synthesis of soluble GST-Cbf5 fusion protein is to co-express GST-Cbf5 with one of the H/ACA snoRNP early accessory proteins which interact only with free Cbf5. H/ACA accessory proteins Shq1 and Naf1 have both been reported to interact exclusively with Cbf5 stabilizing the free Cbf5 and potentially assisting in the proper folding of the protein [43-45, 48, 49]. Co-expression of either Shq1 or Naf1 with GST-Cbf5 may increase GST-Cbf5 protein solubility by associating with newly synthesized GST-Cbf5. Neither Shq1 nor Naf1 are part of the mature H/ACA snoRNP complex as both accessory proteins are removed from Cbf5 either upon snoRNA binding (Shq1) or upon Gar1 association (Naf1) [43-45, 48, 49]. This would potentially allow for the successful expression and purification of a stable GST-Cbf5-Shq1 or GST-Cbf5-Naf1 complex, which can be used for both interaction studies and H/ACA snoRNP complex recombination as both accessory proteins could

potentially be selectively removed by addition of snoRNA or Gar1, respectively. The GST domain could be removed by site-specific protease cleavage.

Optimization of GST-Cb5 protein purification

While the expression of GST-Cbf5 at reduced temperature and in the presence of two percent ethanol aided in the synthesis of soluble GST-Cbf5 fusion protein, the solubility effect noticeable during expression was not significant enough to prevent GST-Cbf5 from aggregating during glutathione sepharose affinity purification. A recent publication by the Massiah group reported the use of a combination of three detergents, sarkosyl, Triton X-100, and CHAPS, during protein purification of insoluble GST-fusion proteins to solubilize proteins aggregated in the insoluble cell pellet and thus to successfully purify soluble protein [103]. The Massiah group hypothesized that the addition of sarkosyl to the insoluble pellet disrupts protein aggregates by encasing the insoluble protein [103]. The addition of Triton X-100 and CHAPS reduces the concentration of sarkosyl around the protein by absorbing the sarkosyl molecules into large micelle structures allowing for protein folding [103]. The Massiah group speculated that the combination of the three detergents transfers the aggregated protein into a molten globule state [103]. This in turn would allow the protein a second chance at correct folding. Modification of the GST-Cbf5 purification strategy to incorporate these detergents allowed for the successful solubilisation and purification of GST-Cbf5. However, upon removal of these detergents during washing and elution, GST-Cbf5 protein solubility was lost resulting in the precipitation of the protein onto the glutathione sepharose column. To circumvent this problem, the next logical step would be to incorporate sarkosyl, Triton X-100, and CHAPS in both the washing buffer and elution buffer. However, the presence of such detergents may affect future protein interaction studies. Therefore, extensive controls and trials will have to be performed to investigate if the detergents could be removed subsequent to elution or upon addition of other H/ACA components.

Future Directions

The successful purification of *S. cerevisiae* Cbf5 will allow for the investigation of how eukaryotic H/ACA snoRNP components interact with one another, and for the determination of their individual functional roles in establishing an active H/ACA snoRNP structure. H/ACA snoRNPs are intricately involved in critical cellular processes through post-transcriptional modifications, cleavage events, and RNA maturation. The detailed functional mechanism of H/ACA snoRNPs is however unknown as a functionally active eukaryotic H/ACA snoRNP complex has yet to be successfully produced *in vitro*. The ability to successfully reconstitute a functionally active H/ACA snoRNP complex in yeast would provide a model system in which the full extent of H/ACA snoRNP involvement in ribosome biogenesis, mRNA splicing and telomere maintenance could be studied.

The establishment of a yeast H/ACA snoRNP system would also allow for the direct analysis of the effect DC mutations within H/ACA snoRNP components have on ribosome biogenesis. The amino acid positions at which DC substitutions are found in the human dyskerin protein are conserved in yeast allowing us to study the effects such mutations have at a diploid and haploid level *in vivo* and hopefully in future also *in vitro*. As DC is often associated with men carrying a DC mutation on their single X-chromosome, the use of haploid *S. cerevisiae* cells should display similar symptoms as male patients with a DC mutation while it would be expected that diploid *S. cerevisiae* cells would exhibit a similar asymptomatic behaviour as heterozygous females [99].

IV References

1. Planta RJ, Mager WH., *The list of cytoplasmic ribosomal proteins of Saccharomyces cerevisiae.* Yeast, 1998. **14**(5): p. 471–477.
2. Cheutin T, O'Donohue MF, Beorchia A, Vandelaer M, Kaplan H, Deféver B, Ploton D, Thiry M., *Three-dimensional organization of active rRNA genes within the nucleolus* Journal of Cellular Science 2002. **115**(16): p. 3297-3307
3. Spasov K, Perdomo LI, Evakine E, Nazar RN., *RAC protein directs the complete removal of the 3 ' external transcribed spacer by the Pac1 nuclease.* Molecular Cell 2002. **9**(2): p. 433-437
4. MOJ., Olson, *The Nucleolus*. Molecular Biology Intellegence Unit, ed. O. MOJ. 2004, George Town, Texas: Kluwer Academic/Plenum Publishers. 1-343.
5. Vanrobays E, Gleizes PE, Bousquet-Antonelli C, Noaillac-Depeyre J, Caizergues-Ferrer M, Gélugne JP., *Processing of 20S pre-rRNA to 18S ribosomal RNA in yeast requires Rrp10p, an essential non-ribosomal cytoplasmic protein.* The EMBO Journal, 2001. **20**(15): p. 4204-4213
6. JR., Warner, *The economics of ribosome biosynthesis in yeast.* Trends in Biochemical Science, 1999. **24**(11): p. 437-440
7. UT., Meier, *How a single protein complex accommodates many different H/ACA RNAs.* Trends in Biochemical Science, 2006. **31**(6): p. 311-315.
8. Brimacombe R, Mitchell P, Osswald M, Stade K, Bochkariov D. , *Clustering of modified nucleotides at the functional center of bacterial ribosomal RNA* FASEB Journal, 1993. **7**(1): p. 161-167
9. Lane BG, Ofengand J, Gray MW., *Pseudouridine and O2'-methylated nucleosides. Significance of their selective occurrence in rRNA domains that function in ribosome-catalyzed synthesis of the peptide bonds in proteins.* Biochemie, 1995. **77**(1-2): p. 7-15.
10. DR., Davis, *Stabilization of RNA stacking by pseudouridine.* Nucleic Acid Research, 1995. **23**(24): p. 5020-5026.
11. Neumann JM, Bernassau JM, Gueron M, Tran-Dinh S *Comparative conformations of uridine and pseudouridine and their derivatives.* European Journal of Biochemistry 1980. **108**(2): p. 457-463
12. Arnez JG, Steitz TA., *Crystal structure of unmodified tRNA(Gln) complexed with glutaminyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure.* Biochemstry, 1994. **33**(24): p. 7560-7567.
13. Hamma T, Ferré-D'Amaré AR., *Pseudouridine synthases.* Chemistry and Biology, 2006. **13**(11): p. 1125-1135.
14. Lafontaine DL, Bousquet-Antonelli C, Henry Y, Caizergues-Ferrer M, Tollervey D., *The box H+ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase.* Genes and Development, 1998. **12**(4): p. 527-537.
15. Rashid R, Liang B, Baker DL, Youssef OA, He Y, Phipps K, Terns RM, Terns MP, Li H., *Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita.* Molecular Cell, 2006. **21**(2): p. 249-260.
16. Kammen HO, Marvel CC, Hardy L, Penhoet EE., *Purification, Structure and Properties of Escherichia coli tRNA pseudouridine synthase I.* The Journal of Biological Chemistry, 1988. **263**(5): p. 2255-2263.
17. Gu X, Liu Y, Santi DV., *The mechanism of pseudouridine synthase I as deduced from its interaction with 5-fluorouracil-tRNA.* PNAS, 1999. **96**(25): p. 14270-14275.
18. Spedaliere CJ, Ginter JM, Johnston MV, Mueller EG., *The pseudouridine synthases: revisiting a mechanism that seemed settled.* Journal of American Chemical Society, 2004. **126**(40): p. 12758-12759.

19. Hamilton CS, Greco TM, Vizthum CA, Ginter JM, Johnston MV, Mueller EG., *Mechanistic investigations of the pseudouridine synthase RluA using RNA containing 5-fluorouridine.* Biochemistry, 2006. **45**(39): p. 12029-12038.
20. EG., Mueller, *Chips off the old block.* Nature Structural Biology, 2002. **9**(5): p. 320-322.
21. Huang L, Pookanjanatavip M, Gu X, Santi DV., *A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst.* Biochemistry, 1998. **37**(1): p. 344-351.
22. Balakin AG, Smith L, Fournier MJ., *The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions.* Cell, 1996. **86**(5): p. 823-834.
23. Ganot P, Caizergues-Ferrer M, Kiss T., *The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation.* Genes & Development, 1997. **11**(7): p. 941-956.
24. Li L, Ye K., *Crystal structure of an H/ACA box ribonucleoprotein particle.* Nature, 2006. **443**(7109): p. 302–307.
25. Duan J, Li L, Lu J, Wang W, Ye K., *Structural mechanism of substrate RNA recruitment in H/ACA RNA-guided pseudouridine synthase.* Molecular Cell, 2009. **34**(4): p. 427–439.
26. Ni J, Tien AL, Fournier MJ., *Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA.* Cell, 1997. **89**(4): p. 565-573.
27. Klein DJ, Schmeing TM, Moore PB, Steitz TA., *The kink-turn: a new RNA secondary structure motif.* The EMBO Journal, 2001. **20**(15): p. 4214-4221.
28. Hamma T, Ferré-D'Amaré AR, *Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 A resolution.* Structure, 2004. **12**(5): p. 893–903.
29. Henras A, Henry Y, Bousquet-Antonelli C, Noaillac-Depeyre J, Gélugne JP, Caizergues-Ferrer M., *Nhp2 and Nop10 are essential for the function of the H/ACA snoRNPs.* The EMBO Journal, 1998. **17**(23): p. 7078-7090.
30. Dez C, Henras A, Faucon B, Lafontaine D, Caizergues-Ferrer M, Henry Y., *Stable expression in yeast of the mature form of human telomerase RNA depends on its assiociation with the box H/ACA small nucleolar RNP proteins Cbf5p, Nhp2p and Nop10p.* Nucleic Acid Research, 2001. **29**(3): p. 598-603.
31. Watkins NJ, Gottschalk A, Neubauer G, Kastner B, Fabrizio P, Mann M, Lührmann R., *Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure.* RNA, 1998. **4**(12): p. 1549–1568.
32. Bousquet-Antonelli C, Henry Y, G'elugne JP, Caizergues-Ferrer M, Kiss T., *A small nucleolar RNP protein is required for pseudouridylation of eukaryotic ribosomal RNAs.* The EMBO Journal, 1997. **16**(15): p. 4770-4776.
33. Meier UT, Blobel G., *NAP57, a mammalian nucleolar protein with a putative homolog in yeast and bacteria.* The Journal of Cell Biology, 1994. **127**(6): p. 1505–1514.
34. Lafontaine DL, Tollervey D., *Birth of the snoRNPs: the evolution of the modification-guide snoRNAs.* Trends in Biochemical Science, 1998. **23**(10): p. 383-388.
35. Hamma T, Reichow SL, Varani G, Ferré-D'Amaré AR., *The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs.* Nature Structural and Molecular Biology, 2005. **12**(12): p. 1101-1107.
36. Hoang C, Ferré-D'Amaré AR., *Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme.* Cell, 2001. **107**(7): p. 929-939.
37. Ramamurthy V, Swann SL, Paulson JL, Spedaliere CJ, Mueller EG., *Critical aspartic acid residues in pseudouridine synthases.* Journal of Biology and Chemistry, 1999. **274**(32): p. 22225-22230.

38. Charron C, Manival X, Cléry A, Senty-Ségault V, Charpentier B, Marmier-Gourrier N, Branlant C, Aubry A., *The archaeal sRNA binding protein L7Ae has a 3D structure very similar to that of its eukaryal counterpart while having a broader RNA-binding specificity.* Journal of Molecular Biology, 2004. **342**(3): p. 757–773.
39. Henras A, Dez C, Noaillac-Depeyre J, Henry Y, Caizergues-Ferrer M., *Accumulation of H/ACA snoRNPs depends on the integrity of the conserved central domain of the RNA-binding protein Nhp2p.* Nucleic Acids Research, 2001. **29**(13): p. 2733-2746.
40. Gagnon KT, Zhang X, Qu G, Biswas S, Suryadi J, Brown BA 2nd, Maxwell ES., *Signature amino acids enable the archaeal L7Ae box C/D RNP core protein to recognize and bind the K-loop RNA motif.* RNA, 2010. **16**(1): p. 79-90.
41. Darzacq X, Kittur N, Roy S, Shav-Tal Y, Singer RH, Meier UT., *Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells.* The Journal of Cells Biology, 2006. **173**(2): p. 207-218.
42. Godin KS, Walbott H, Leulliot N, van Tilbeurgh H, Varani G., *The box H/ACA snoRNP assembly factor Shq1p is a chaperone protein homologous to Hsp90 cochaperones that binds to the Cbf5p enzyme.* Journal of Molecular Biology, 2009. **390**(2): p. 231-244.
43. Singh M, Gonzales FA, Cascio D, Heckmann N, Chanfreau G, Feigon J., *Structure and functional studies of the CS domain of the essential H/ACA ribonucleoparticle assmble protein SHQ1.* Journal of Biologal Chemistry, 2009. **284**(3): p. 1906-1916.
44. Grozdanov PN, Roy S, Kittur N, Meier UT., *SHQ1 is required prior to NAF1 for assembly of H/ACA small nucleolar and telomerase RNPs.* RNA, 2009. **15**(6): p. 1188–1197.
45. Leulliot N, Godin KS, Hoareau-Aveilla C, Quevillon-Cheruel S, Varani G, Henry Y, Van Tilbeurgh H., *The box H/ACA RNP assembly factor Naf1p contains a domain homologous to Gar1p mediating its interaction with Cbf5p.* Journal of Molecular Biology, 2007. **371**(5): p. 1338–1353.
46. Baker DL, Youssef OA, Chastkofsky MI, Dy DA, Terns RM, Terns MP., *RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP.* Genes & Development, 2005. **19**(10): p. 1238-1248.
47. Darzacq, X., Kittur, N., Roy, S., Shav-Tal, Y., Singer, R.H., and Meier, U.T., *Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells.* Journal of Cellular Biology, 2005. **173**: p. 207–218.
48. Yang PK, Hoareau C, Froment C, Monsarrat B, Henry Y, Chanfreau G., *Cotranscriptional recruitment of the pseudouridylsynthetase Cbf5p and of the RNA binding protein Naf1p during H/ACA snoRNP assembly.* Molecular and Cellular Biology, 2005. **25**(8): p. 3295–3304.
49. Jin H, Loria JP, Moore PB., *Solution structure of an rRNA substrate bound to the pseudouridylation pocket of a box H/ACA snoRNA.* Molecular Cell, 2007. **26**(2): p. 205-215.
50. Wu H, Feigon J., *H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification.* PNAS, 2007. **104**(16): p. 6655-6660.
51. Liang B, Zhou J, Kahen E, Terns RM, Terns MP, Li H., *Structure of a functional ribonucleoprotein pseudouridine synthase bound to a substrate RNA.* Nature Structural and Molecular Biology, 2009. **16**(7): p. 740-746.
52. Morrissey JP, Tollervey D., *Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis.* Molecular and Cellular Biology, 1993. **13**(4): p. 2469-2477
53. Atzorn V, Fragapane P, Kiss T., *U17/snR30 is a ubiquitous snoRNA with two conserved sequence motifs essential for 18S rRNA production.* Molecular and Cellular Biology, 2004. **24**(4): p. 1769–1778.

54. Fayet-Lebaron E, Atzorn V, Henry Y, Kiss T., *18S rRNA processing requires base pairings of snR30 H/ACA snoRNA to eukaryote-specific 18S sequences.* The EMBO Journal, 2009. **28**(9): p. 1260-1270.
55. D., Tollervey, *A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA.* The EMBO Journal, 1987. **6**(13): p. 4169-4175.
56. Liang XH, Liu Q, Liu Q, King TH, Fournier MJ., *Strong dependence between functional domains in a dual-function snoRNA infers coupling of rRNA processing and modification events.* Nucleic Acid Research, 2010. **38**(10): p. 3376–3387.
57. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walters P., *Molecular biology of the cell*. fourth ed, ed. D.B. Anderson MS. 2002, New York, NY: Garland Science.
58. Massenet S, Mougin A, Branlant C., *Posttranscriptional modifications in the U small nuclear RNAs*. Modification and editing of RNA, ed. B.R. Grosjean H. 1998, Washingtion, DC: ASM Press.
59. H, Reddy R and Busch, *Structure and function of major and minor small nuclear ribonucleoprotein particles*, in *Small nuclear RNAs: RNA sequences, structure and modifications*, M.L. Birnstiel, Editor. 1988, Springer-Verlag Press: Berlin, Germany. p. 1-37.
60. Parker R, Siliciano PG, Guthrie C., *Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA.* Cell, 1987. **49**(2): p. 229-239.
61. Newby MI, Greenbaum NL., *A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture.* RNA, 2001. **7**(6): p. 833-845.
62. Newby MI, Greenbaum NL., *Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine.* Nature Structural and Molecular Biology, 2002. **9**(12): p. 958-965.
63. Valadkhan S, Manley JL., *Characterization of the catalytic activity of U2 and U6 snRNAs.* RNA, 2003. **9**(7): p. 892-904.
64. Zhao X, Yu YT., *Pseudouridines in and near the branch site recognition region of U2 snRNA are required for snRNP biogenesis and pre-mRNA splicing in Xenopus oocytes.* RNA, 2004. **10**(4): p. 681-690.
65. Yu YT, Shu MD, Steitz JA., *Modification of U2 snRNA are required for snRNP assembly and pre-mRNA splicing.* The EMBO Journal, 1998. **17**(19): p. 5783-5795.
66. Chen JL, Greider CW., *An emerging consensus for telomerase RNA structure.* PNAS, 2004. **101**(41): p. 14683–14684.
67. Mitchell JR, Cheng J, Collins K., *A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end.* Molecular and Cellular Biology, 1999. **19**(1): p. 567–576.
68. Wang C, Meier UT., *Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins.* The EMBO Journal, 2004. **23**(8): p. 1857–1867.
69. A., Gupta V and Kumar, *Dyskeratosis Congenita*, in *Diseases of DNA repair*, A. SI., Editor. 2010, Landes Bioscience and Springer Science: New York. p. 215-219.
70. Vulliamy TJ, Dokal I., *Dyskeratosis congenita: the diverse clinical presentation of mutations in the telomerase complex.* Biochemie, 2008. **90**: p. 122-130.
71. Vulliamy TJ, Marrone A, Knight SW, Walne A, Mason PJ, Dokal I., *Mutations in dyskeratosis congenita: their impact on telomere length and the diversity of clinical presentation.* Blood, 2006. **107**(7): p. 2680-2685.
72. Marrone A, Walne A, Dokal I., *Dyskeratosis Congenita: telomerase, telomeres and anticipation.* Current Opinion in Gene Development, 2005. **15**(3): p. 249-257.
73. Heiss NS, Knight SW, Vulliamy TJ, Klauck SM, Wiemann S, Mason PJ, Poustka A, Dokal I., *X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions* Nature Genetics, 1998. **19**(1): p. 32-38.

74. Walne AJ, Vulliamy T, Marrone A, Beswick R, Kirwan M, Masunari Y, Al-Qurashi FH, Aljurf M, Dokal I., *Genetic heterogeneity in autosomal recessive dyskeratosis congenita with one subtype due to mutations in the telomerase-associated protein NOP10.* Human Molecular Genetics, 2007. **16**(13): p. 1619-1629.
75. Trahan C, Martel C, Dragon F., *Effects of dyskeratosis congenita mutations in dyskerin, NHP2 and NOP10 on assembly of H/ACA pre-RNPs.* Human Molecular Genetics, 2010. **19**(5): p. 825–836.
76. Vulliamy T, Beswick R, Kirwan M, Marrone A, Digweed M, Walne A, Dokal I., *Mutations in the telomerase component NHP2 cause the premature ageing syndrome dyskeratosis congenita.* PNAS, 2008. **105**(23): p. 8073–8078.
77. Vulliamy TJ, Knight SW, Mason PJ, Dokal I., *Very short telomeres in the peripheral blood of patients with X-linked and autosomal dyskeratosis congenita.* Blood Cells Molecules and Diseases, 2001. **27**(2): p. 353-357.
78. Vulliamy T, Marrone A, Szydlo R, Walne A, Mason PJ, Dokal I., *Disease anticipation is associated with progressive telomere shortening in families with dyskeratosis congenita due to mutations in TERC.* Nature Genetics, 2004. **36**(5): p. 447-449.
79. Ruggero D, Grisendi S, Piazza F, Rego E, Mari F, Rao PH, Cordon-Cardo C, Pandolfi PP., *Dyskeratosis congenita and cancer in mice deficient in ribosomal RNA modification.* Science, 2003. **299**: p. 259-262.
80. Kolodrubetz D, Haggren W, Burgum A., *Amino-terminal sequence of a Saccharomyces cerevisiae nuclear protein, NHP6, shows significant identity to bovine HMG1.* FEBS Letters, 1988. **238**(1): p. 175-179.
81. Kolodrubetz D, Burgum A., *Sequence and genetic analysis of NHP2: A moderately abundant high mobility group-like nuclear protein with an essential function in Saccharomyces cerevisiae.* Yeast, 1991. **7**(2): p. 79-90.
82. Henras AK, Dez C, Henry Y., *RNA structure and function in C/D and H/ACA s(no)RNPs.* Current Opinion in Structural Biology, 2004. **14**(3): p. 335-343.
83. Moore T, Zhang Y, Fenley MO, Li H., *Molecular basis of box C/D RNA–protein interactions: cocrystal structure of archaeal L7Ae and a box C/D RNA.* Structure, 2004. **12**(5): p. 807–818.
84. Vedadi M, Lew J, Artz J, Amani M, Zhao Y, Dong A, Wasney GA, Gao M, Hills T, Brokx S, Qiu W, Sharma S, Diassiti A, Alam Z, Melone M, Mulichak A, Wernimont A, Bray J, Loppnau P, Plotnikova O, Newberry K, Sundararajan E, Houston S, Walker J, Tempel W, Bochkarev A, Kozieradzki I, Edwards A, Arrowsmith C, Roos D, Kain K, Hui R., *Genome-scale protein expression and structural biology of Plasmodium falciparum and related Apicomplexan organisms.* Molecular and Biochemical Parasitology, 2007. **151**(1): p. 100-110.
85. Suryadi J, Tran EJ, Maxwell ES, Brown BA 2nd., *The crystal structure of the Methanocaldococcus jannaschii multifunctional L7Ae RNA-binding protein reveals an induced-fit interaction with the box C/D RNAs.* Biochemistry, 2005. **44**(28): p. 9657-9672.
86. Liang B, Xue S, Terns RM, Terns MP, Li H., *Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex.* Nature Structural and Molecular Biology, 2007. **14**(12): p. 1189-1195.
87. Kuhn JF, Tran EJ, Maxwell ES., *Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein.* Nucleic Acids Research, 2002. **30**(4): p. 931–941.
88. Russel, Joseph Sambrook and David W., *Molecular Cloning a Laboratory Manual.* 3 ed, ed. J. Argentine. Vol. 1. 2001, New York: Cold Spring Harbor Laboratory Press.

89. Bortolin ML, Ganot P, Kiss T., *Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs.* The EMBO Journal, 1999. **18**(2): p. 457-469.
90. Holde KE, Johnson C, Ho PS *Principles of Physical Biochemistry* 2ed. 2006, Upper Saddle River, New Jersey: Pearson/Prentic Hall.
91. Henras AK, Capeyrou R, Henry Y, Caizergues-Ferrer M., *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. 1704–1712.
92. Jiang W, Middleton K, Yoon HJ, Fouquet C, Carbon J., *An essential yeast protein, CBF5p, binds in vitro to centromeres and microtubules.* Molecular and Cellular Biology, 1993. **13**(8): p. 4884-4893.
93. Pan H, Agarwalla S, Moustakas DT, Finer-Moore J, Stroud RM., *Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit.* PNAS, 2003. **100**(22): p. 12648-12653.
94. Manival X, Charron C, Fourmann JB, Godard F, Charpentier B, Branlant C., *Crystal structure determination and site-directed mutagenesis of the Pyrococcus abyssi aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity.* Nucleic Acids Research, 2006. **34**(3): p. 826-839.
95. EV., Koonin, *Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases.* Nucleic Acids Research, 1996. **24**(12): p. 2411–2415.
96. Cadwell C, Yoon HJ, Zebarjadian Y, Carbon J., *The yeast nucleolar protein Cbf5p is involved in rRNA biosynthesis and interacts genetically with the RNA polymerase I transcription factor RRN3.* Molecular and Cellular Biology 1997. **17**(10): p. 6175–6183.
97. Zebarjadian Y, King T, Fournier MJ, Clarke L, Carbon J., *Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA.* Molecular and Cellular Biology, 1999. **19**(11): p. 7461–7472
98. Ganot P, Bortolin ML, Kiss T., *Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs.* Cell, 1997. **89**(5): p. 799-809.
99. Handley TP, McCaul JA, Ogden GR., *Dyskeratosis congenita.* Oral Oncology, 2006. **42**(4): p. 331-336.
100. Grozdanov PN, Fernandez-Fuentes N, Fiser A, Meier UT., *Pathogenic NAP57 mutations decrease ribonucleoprotein assembly in dyskeratosis congenita.* Human Molecular Genetics, 2009. **18**(23): p. 4546-4551.
101. Thomas JG, Baneyx F., *Protein misfolding and inclusion body formation in recombinant Escherichia coli cells overexpressing Heat-shock proteins.* The Journal of Biological Chemistry, 1996. **271**(19): p. 11141–11147.
102. Sørensen HP, Mortensen KK., *Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli.* Microbial Cell Factories 2005. **4**(1): p. 1.
103. Tao H, Liu W, Simmons BN, Harris HK, Cox TC, Massiah MA., *Purifying natively folded proteins from inclusion bodies using sarkosyl, Triton X-100, and CHAPS.* Biotechniques, 2010. **48**(1): p. 61-64.
104. F., Baneyx, *Recombinant protein expression in Escherichia coli.* Current Opinion in Biotechnology, 1999. **10**(5): p. 411-421.
105. Nygaard FB, Harlow KW., *Heterologous expression of soluble, active proteins in Escherichia coli: The human estrogen receptor hormone-binding domain as paradigm.* Protein Expression and Purification, 2001. **21**(3): p. 500-509.