Structure and Dynamics of eukaryotic H/ACA RNPs from *Saccharomyces cerevisiae*

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

vorgelegt beim Fachbereich Biochemie, Chemie und Pharmazie

der Johann Wolfgang Goethe-Universität

in Frankfurt am Main

von

Sven Trucks

aus Gelnhausen

Frankfurt am Main

2020

(D30)

Unter der Betreuung von Dr. Martin Hengesbach
durchgeführt am Institut für Organische Chemie und Chemische Biologie
im Fachbereich 14 für Biochemie, Chemie und Pharmazie
der Johann Wolfgang Goethe-Universität Frankfurt am Main
im Zeitraum von Oktober 2015 bis September 2020

vom Fachbereich Biochemie, Chemie und Pharmazie der
Johann Wolfgang Goethe-Universität angenommen.

Dekan: Prof. Dr. Clemens Glaubitz

Gutachter: Dr. Martin Hengesbach
Prof. Dr. Harald Schwalbe

Datum der Disputation:

Table of contents

Chapter 1

Introduction

1.1 Post-transcriptional RNA modifications

Post-transcriptional modification (and also a few examples of co-transcriptional modification) of non-coding as well as coding RNAs is highly abundant throughout all kingdoms of life, resulting in an incredible chemical diversity among RNA nucleosides. These modifications result in over 150 different modified nucleosides, reported for thousands of different RNA modification sites[1,2]. Post-transcriptional modifications are enzymatically introduced into the four basic RNA building blocks adenosine, uridine, cytidine and guanosine via various chemical reactions, including alkylations, hydroxylations, thiolations, selenations, esterifications, deaminations and peroxidations, with methylations being the most prominent (about ⅔ of all modifications)[3,4]. The modifications can take place at the ribose ring, for example via 2'-O-methylation or at the purine or pyrimidine backbone, resulting in the addition of single chemical groups or more complex side chains onto the base, with C5 of uridine being the most prominent target site for modifications. Figure 1 shows the most typical target sites in general[4–6].

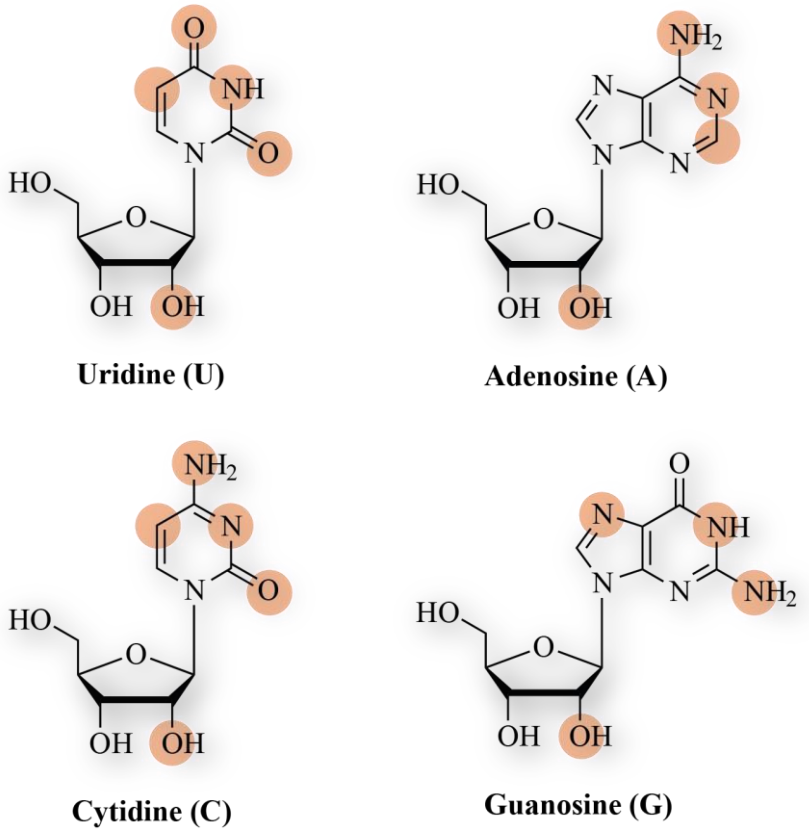


Figure 1: The four basic nucleosides uridine, adenosine, cytidine and guanosine with usual target sites for post-transcriptional modifications (orange)[6].

Some modifications also serve as the starting point for other, more complex modifications, resulting in different branching pathways during the modification reactions and modification cascades[3]. Several modifications can happen at a single nucleoside independently,

highlighting the orthogonality of the reactions and their respective enzymes. Some chemically identical nucleoside modifications in different organisms are even catalyzed by independently evolved enzymes, indicating convergent evolution[7,8]. While RNA modifications are occurring in all kingdoms of life, distribution between the three kingdoms is not uniform; some of them are specific to one kingdom or are specifically absent from one kingdom (Figure 2)[4,9].

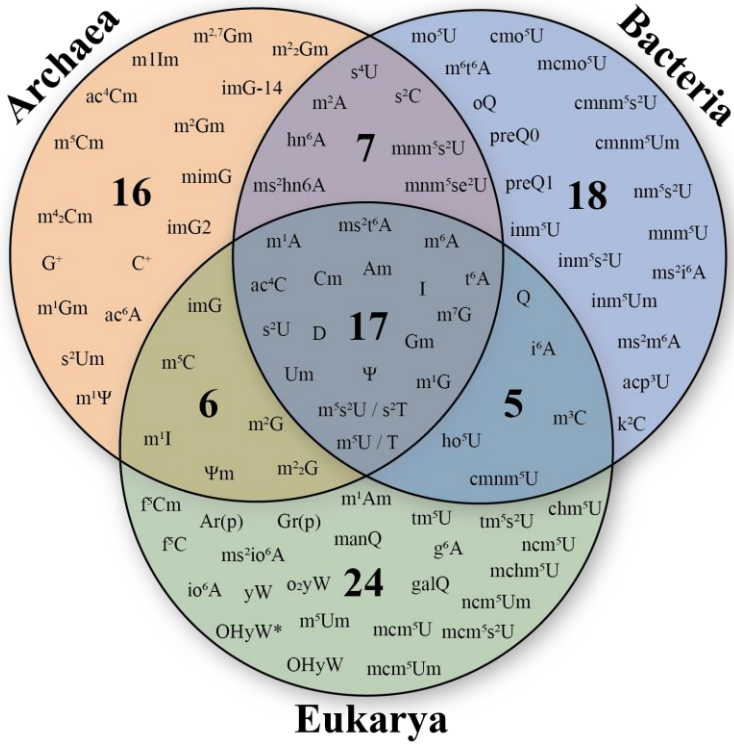


Figure 2: Post-transcriptional modifications in tRNA throughout the kingdoms of life. Capital letters denote the nucleobase, lowercase letters denote the type of modification, supersripted numbers denote the position of the modification in the nucleobase and subscripted numbers denote the number of modifications. Abbreviations as followed: ac: acetyl, acp: aminocarboxypropyl, chm: carboxyhydroxymethyl,cmo: oxyacetic acid, cmnm: carboxymethylaminomethyl, f: formyl, g: glycinyl, gal: galactosyl,hn: hydroxynorvalylcarbamoyl, ho: hydroxy, i: isopentenyl, inm: isopentenylaminomethyl, io: cis-hydroxyisopentenyl, m: methyl, man: mannosyl, mchm: carboxyhydroxymethyl methyl ester, mcm: methoxycarbonylmethyl, mcmo: oxyacetic acid methyl ester, mnm: methylaminomethyl, mo: methoxy, ncm: carbamoylmethyl, nm: aminomethyl, r(p): 5-O-phosphono-b-D-ribofuranosyl, s: thio, se: seleno, t: threonylcarbamoyl, tm: taurinomethyl (figure based on Lorenz *et al.*[9]).

For a long time, most RNA modifications were believed to be irrelevant and their impact on the function on the cell function minor, since the generation of phenotypes lacking single modifications led to negligible effects for the organism[3,10,11] and only the absence of multiple modifications led to obvious consequences[12–16]. However, it is more likely that the lack of a single modification can easily be compensated in part or completely by other modifications in the RNA, and that the actual function of most modifications is still not completely

understood. Moreover, a large number of defects in proteins that are involved in the post-transcriptional modification of RNAs are linked to human diseases[17,18].
To provide accurate quantification and correct positionally mapping of RNA modifications as well as the uncovering of still unknown RNA modifications, new methods are constantly being developed, like the utilization of liquid chromatography tandem mass spectrometry[19,20], high-throughput sequencing methods[21], signature chemical coupling with deep sequencing methods[22,23] and computational methods[24].
Post-transcriptional RNA modifications have been observed in rRNA, mRNA, snRNA, snoRNA, miRNA, viral RNA and most prominently in tRNA (approximately 14 modifications in each tRNA molecule on average)[4,25,26]. Especially positions 34 and 37 in the anticodon-loop of tRNA are hotspots for modification, as these affect the codon-anticodon function[3,6,27]. While modification of position 34 enhances base-pairing flexibility during wobble decoding, modification at position 37 ensures reading frame maintenance[28,29], having an impact on translation regulation[30–32]. Other tRNA modifications can have an impact on the structural stability of the RNA, exerting a rigidifying or loosening effect on the structure[33] and influence correct folding and function[34,35]. Many modified nucleotides have an effect on the local RNA structure in the Angstrom scale in form of stabilizing effects by restriction of conformational flexibility and stabilization of helical conformations[36–39]. By blocking hydrogen bond formation, RNA modifications can lead to rearrangements of Watson-Crick base pair interactions and thereby to rearrangements of the RNA structure in the nanometer scale, shifting the dynamic equilibrium in favor of a specific structure[14,40–42]. Furthermore, RNA modifications can lead to increased structural stabilization and thermal stability, which i.e. leads to better biochemical stability via increased resistance against degradation by bases and nucleases[43]. Especially in tRNA, the post-transcriptional modifications are believed to actively accompany the tRNA maturation step by step by shaping the unstructured tRNA precursor into a structured, functional tRNA[6,44–46].
In rRNA, post-transcriptional modifications occur on highly conserved areas, which are important for its function, indicating a strong selection process during evolution[10,25]. These modifications seem to be crucial for ribosome biogenesis and protein synthesis, being involved in intersubunit contacts with ribosomal proteins and tRNA binding[47–50].
In mRNA, post-transcriptional modifications are linked to mRNA maturation (including 5' capping, 3' polyadenylation and splicing) and transport[51], regulation of translation[52,53] and RNA decay pathways[51,54,55]. mRNA modifications seem to be highly regulated, showing dynamic in response to stress like heat shock and nutrient deprivation[25,56–59] as well as being reversible, for example by the regulation of methyltransferases and demethylases[18,25,60–62].

1.1.1 Pseudouridine and non-guided pseudouridylation

Pseudouridine is the most abundant post-transcriptional modification in RNA, as well as the first modified nucleotide that has been discovered in RNA, making it known as the "fifth nucleotide" with the abbreviation "Ψ"[63–65]. Pseudouridine has been identified in tRNA, rRNA, snRNA, snoRNA as well as mRNA, with an estimated frequency of 7-9% of uridines being modified into pseudouridine, and pseudouridine comprising around 4% of nucleotides

of yeast RNA[59,63,66–71]. In general, pseudouridine can be described as the C5-glycoside isomer of uridine, which is generated by a disruption of the N1-C1' bond between the uracil and the ribose, a rotation of the uracil base of 180° around the N3-C6 axis and the formation of a non-canonical C5-C1' glycosidic bond between the base and the ribose (Figure 3)[23,71].

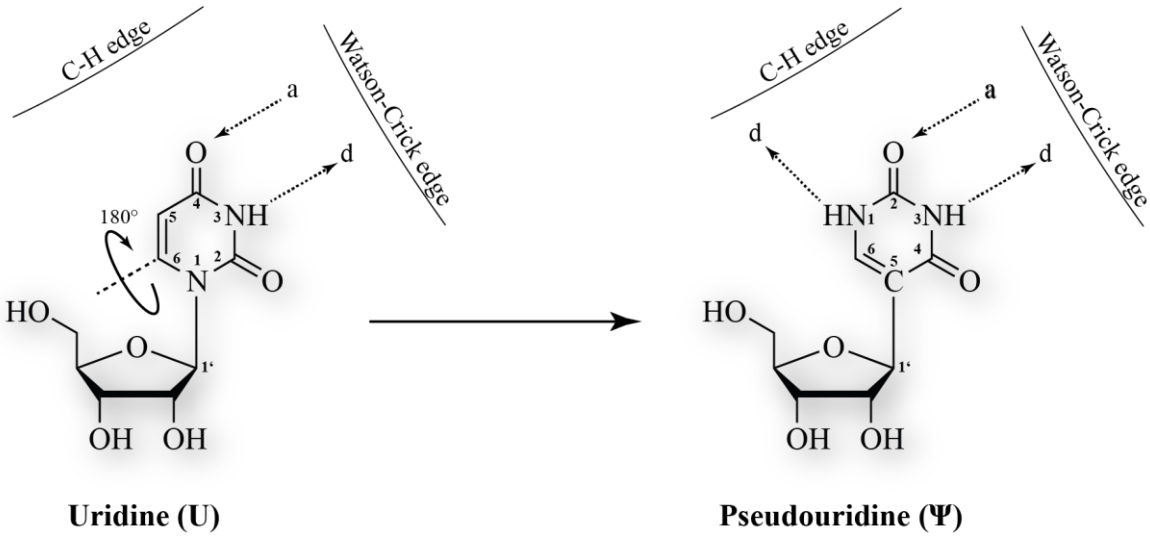


Figure 3: Uridine (U) and the generation of its isomer pseudouridine (Ψ) by disruption of the N1-C1' bond, rotation of 180° of the uracil base around the N3-C6 axis and formation of the C5-C1' bond. Both nucleosides contain an H-donor (d) and acceptor (a) at the Watson-Crick edge, pseudouridine contains an additional H-donor at the C-H edge.

While Ψ and U show some physiochemical similarities, like the UV spectrum and molecular mass[72], this isomerization gives some unique properties to pseudouridine[73,74]. While the Watson-Crick edge of pseudouridine in comparison to uridine is unchanged, which still allows Ψ-A base paring, pseudouridine contains an additional H-bond donor (in form of the NH1 imino proton) at the C-H edge of the molecule. Pseudouridine has been shown to coordinate a water molecule between NH1 and the 5'-phosphodiester moiety of both Ψ and the preceding base, restricting base conformation and mobility of the backbone, resulting in rigidity and therefore higher stability of the RNA structure (Figure 4)[75–78].

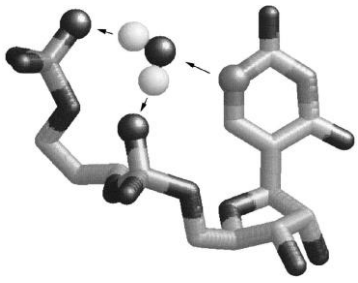


Figure 4: Coordination of a water molecule between NH1 of pseudouridine and the sugar-phosphate backbone of both pseudouridine and the preceding residue, forming a "water bridge" and stabilizing the RNA structure by induced rigidity. Hydrogen bonds are depicted as arrows.

(Figure taken from Charette *et al.*[71] with permission from John Wiley & Sons © 2008 John Wiley & Sons.)

Favoring the 3'-endo-conformation of the ribose leads to improved local base stacking in single stranded RNA as well as duplex regions, which can also involve the coordination of a water molecule and the stronger interaction with water of NH1 of Ψ in comparison to CH5 of U, resulting in stronger backbone rigidity[66,79,80]. This base stacking is believed to be the strongest contribution of Ψ towards structural stability of the RNA[80,81]. Pseudouridine has increased base pairing stability with adenosine in comparison to the U-A base pair[71,82] and also the base pairing capabilities with the other nucleotides is increased[78,83]. Additionally, pseudouracil is linked via a C-C bond instead of the N-C glycosyl bond to the sugar moiety, resulting in greater rotational freedom and therefore more conformational flexibility in comparison to uridine[71,84].
Interestingly enough, pseudouridine itself can become the target of further post-transcriptional modifications, resulting in hypermodified Ψ[85,86]. Observed Ψ modification sites are both nitrogen atoms in the pyrimidine ring and the 2'-O of the ribose subunit and include methylation and addition of a 3-amino-3-carboxypropyl-group (Figure 5).

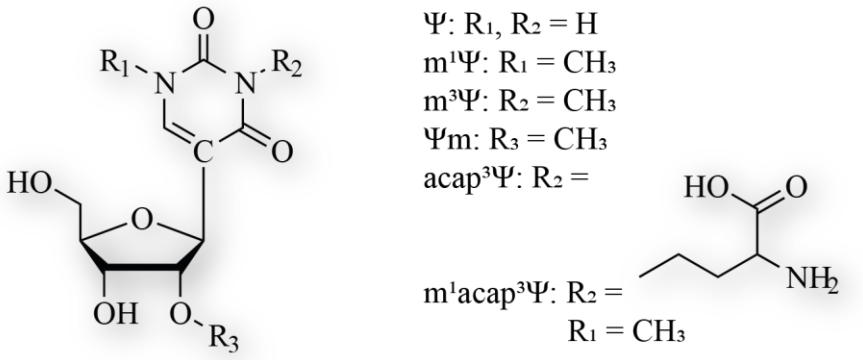


Figure 5: Observed hypermodifications of pseudouridine. Both nitrogen atoms as well as the 2'-O in the ribose moiety are prone to post-transcriptional modifications, with the addition of methyl- or 3-amino-3-carboxypropyl- (acap) group. Several modifications at once are possible[66].

In tRNAs, U to Ψ modifications have been identified in the D stem, the anticodon stem and loop and most prominently as the nearly universally conserved Ψ55 in the TΨC stem loop and contribute to the stabilization of these specific structural motifs[71,87]. Thereby, Ψ can influence the interactions of tRNA with mRNA and rRNA and stabilize the structure of the anticodon loop and the tRNA-ribosome interactions[79,88]. Pseudouridine directly in the anticodon loop has been linked to stronger codon-anticodon base pairings[79].
Pseudouridylation of mRNA may be involved in natural occurring nonsense codon suppression (see chapter 1.2.4.3), including the promotion of non-canonical purine-purine

base pairs in the anticodon region[89,90]. Additionally, pseudouridylated mRNA in combination with pseudouridylated non-coding RNAs may directly affect pre-mRNA splicing and translation fidelity[69]. Also regulatory roles of Ψ in mRNA are presumed, including alteration of the translation initiation efficiency and ribosome pausing[57].

In rRNA, pseudouridine clusters were found in functionally important domains, especially regions that are close to the A- and P-site of the ribosome where peptide bond formation takes place[91–93]. Ψ may therefore be involved in rRNA folding and ribosome assembly, enhancing translation efficiency[94]. An older, now more controversial hypothesis also suggests a catalytic role for Ψ in peptidyl transfer during the translation process[84].

In snRNA, Ψ is mostly located in functionally important regions which are linked to RNA-RNA and RNA-protein interactions during spliceosomal function[95].

The transformation of U into Ψ has also been shown to be inducible by stress. Positions 56 and 93 in *S. cerevisiae* U2 snRNA could be pseudouridylated during nutrient deprivation and specifically position 56 during heat shock[96]. The stress-induced target sites were similar, but not identical to the native target sites, that the respective enzymes (Pus7p targeted position 56, snR81 RNP targeted position 93) pseudouridylate under normal conditions.

Pseudouridylation can occur in an RNA guided manner (described in chapter 1.1.2.2) or is performed RNA-independent in a “protein only” fashion by a pseudouridylase. Those pseudouridylases can be found in all kingdoms of life and can be classified into six families, of which five are named after the respective *E. coli* enzyme responsible for pseudouridylation: RluA, RsuA, TruA, TruB and TruD[70] as well as Pus10[97]. Although sequence similarities between the enzymes belonging to different families are small, the structure of the pseudouridylases core is similar (Figure 6)[98–102]. This conserved core contains an eight-stranded mixed β sheet and several helices and loop structures, that are flanking the catalytic cleft, which bisects the sheet. One of the loops occupying this cleft carries a strictly conserved aspartate, that is found in all pseudouridylase families (and also in the catalytic core of RNA-guided pseudouridylases), which is crucial for the pseudouridylation reaction (see proposed mechanisms in Figure 7)[66,103]. Some psedouridylases carry additional structural elements in the form of N- or C-terminal extensions, like a PUA-domain, which can govern substrate specificity[70,104]. While RluA and RsuA family pseudouridylases are relatively closely related, even sharing conserved sequence motifs, TruA family enzymes are the most distantly related and they even seem to function as dimers[70,105]. Despite their structurally similar protein fold, pseudouridylases from all described families are able to recognize their respective substrate uridine in the RNA without any accessory factors, some being specific for a single RNA modification site, and some being able to modify multiple structurally similar target sites[70].

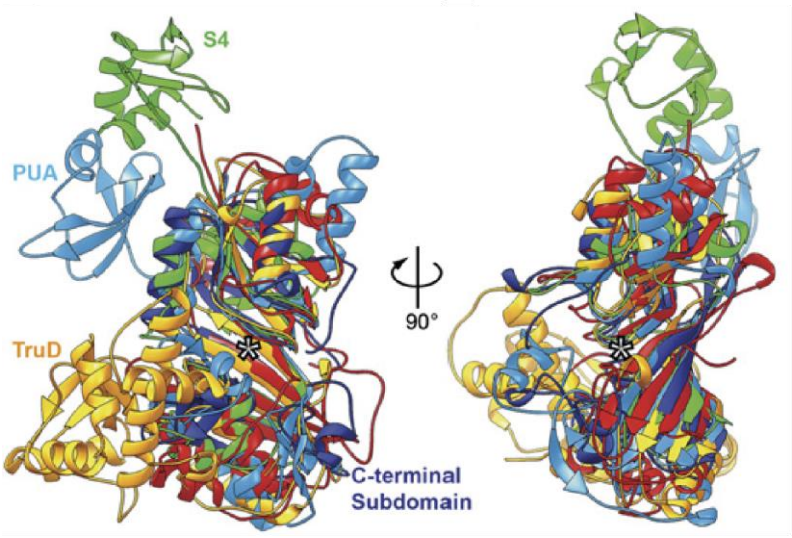


Figure 6: Superimposition of crystal structures from *E. coli* representatives of the five pseudouridylase families: TruD (yellow, pdb: 1SB7), TruA (red, pdb: IDJ0), TruB (cyan, pdb: 1K8W), RsuA (green, pdb: 1KSL) and RluA (navy, pdb: 2I82). (Figure taken from Hamma *et al.*[70] with permission from Elsevier, © 2006 Elsevier.)

The mechanisms of the uridine to pseudouridine conversion is still not fully understood and several mechanistic routes have been proposed[66]. For all described pseudouridylases (and also the snoRNPs) a conserved aspartate is known to be involved in the catalysis, by binding to a specific part of the target uridine[66,103]. Also, an arginine (or histidine in the case of TruB family members) is involved in the process of base-flipping[72,106,107]. The whole mechanism seems to take place in the four steps of initial RNA binding, induced fit (conformational change like unfolding/target placement at the catalytic site), catalysis and product release[108,109]. Kinetic studies on Ψ formation showed relatively slow multiple turnover rates, with K_m values in the nanomolar range[109]. While for Pus1, changes in the RNA conformation[110] and for RluA product release[109] have been proposed to be rate limiting steps, for TruA and TruB the actual chemical catalysis step (most likely the base rotation or C-C bond formation) has been shown to be the rate limiting factor of pseudouridylation[109,111]. The proposed reaction mechanisms are a Michael addition-like mechanism, where the catalytic aspartate would attack the Michael acceptor C6 of the base[106], or the interaction of the aspartate with C1' of the ribose in an acylal[112] or glycal[111,113] mechanism (Figure 7). The attack on the C1' of the ribose[114,115] and the Michael like attack on C6 of the base[116–118] have both been described for other modification enzymes. However, several co-crystal studies with the inhibitor-like uridine-analogue 5-fluorouridine (5FU, see Figure 109), which also undergoes rearrangement and hydration to 5S-6R-6-hydroxy-5-fluoro-pseudouridine in the presence of pseudouridylases, were not able to favor one mechanism over the other[100,106,119,120]. It is even unclear if U and 5FU undergo the same mechanism, due to specific side products identified via NMR in the reaction with 5FU but not with U[113,121].

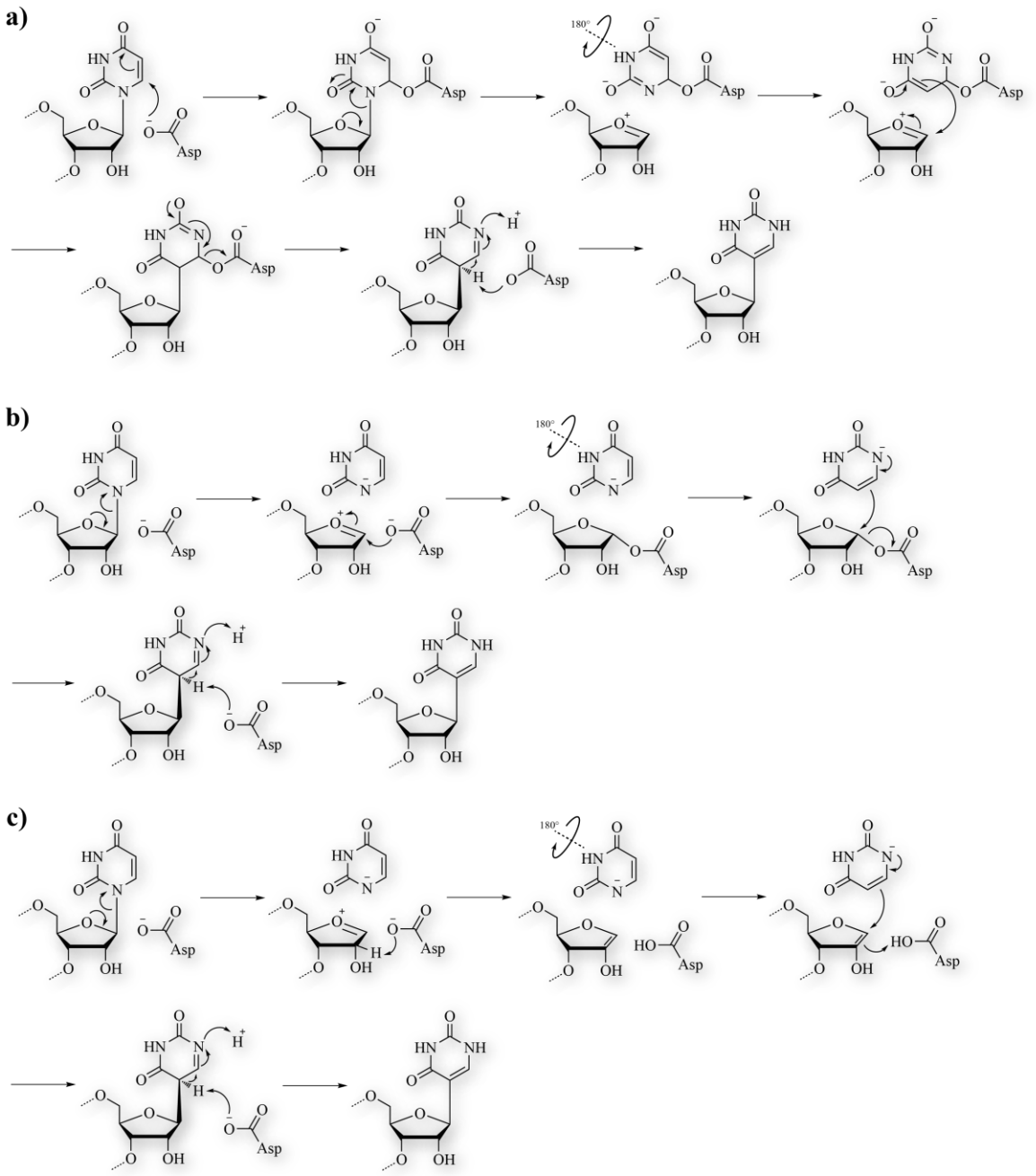


Figure 7: Proposed mechanisms for uridine to pseudouridine formation. a) Michael addition-like mechanism, b) acylal mechanism, c) glycal mechanism[66,106,112,113].

1.1.2 RNA-guided post-transcriptional RNA modifications

In eukaryotes and archaea, post-transcriptional RNA modification can occur in an RNA-guided manner. These guide RNAs act as a scaffold to coordinate the formation of a functional complex with several proteins, producing an active ribonucleoprotein particle (RNP)[122,123]. Furthermore, the guide snoRNA recruits the target RNA, which is modified, via Watson-Crick base pairing. snoRNPs exist in the form of box C/D RNPs, which are responsible for 2'-O-methylation and box H/ACA RNPs, which carry out pseudouridylation[123]. In eukaryotes, the guide RNAs involved in the modification of rRNAs

are called small nucleolar RNAs (snoRNAs) and are clustered in the nucleolus[122,124]. Modification of snRNAs in the small Cajal bodies is carried out by snoRNA-like guide RNAs (scaRNAs)[125]. In archaea, snoRNA-like RNAs (sRNAs) also carry out RNA modifications. In fact, most structural knowledge of RNPs stems from archaeal complexes, since it is oftentimes easier to purify the archaeal components in the required amount, and the lack of good *in vitro* reconstitution systems for non-archaeal RNPs, which until recently relied on *in vivo* preassembled protein complexes[126,127].

1.1.2.1 Box C/D RNPs

Box C/D RNPs are responsible for 2'-O-methylation of the ribose moiety at a specific target site in their substrate RNA. They contain a conserved C (RUGAUGA) and D (CUGA) motif at the termini of the RNA, forming the C/D box. This motif is part of a stem-internal loop-stem structure, forming a K-turn motif[128]. Additionally, the RNA contains an internal C'/D' motif with the same sequence, which adopts a stem-loop structure, forming a K-loop[129]. The RNA contains two identical target sites, between the C and D' and between the C' and D motives, which can recruit a substrate RNA via Watson-Crick base-pairing. The archaeal RNP is formed with the proteins L7Ae (Snu13 in yeast, 15.5K/NHPX in human), Nop5 (Nop56 in yeast, NOP56 in human) and fibrillarin (Nop1 in yeast), which is the catalytic subunit of the complex. A set of these four proteins binds to both the C/D and C'/D' motives, allowing to perform the methylation reaction on both target sites (Figure 8). L7Ae acts as the main RNA-binding protein and forms a pre-RNP with the RNA through binding to the K-motives[130–132]. This pre-RNP is recognized by Nop5 and fibrillarin, which binds to the complex and both Nop5 proteins homodimerize with their coiled-coils domains, providing a stable scaffold for the complex in addition to the RNA scaffold[131,133,134]. Fibrillarin, together with the N-terminal domain of Nop5, bind the first 6 nucleotides of the guide RNA-substrate RNA duplex and perform 2'-O-methylation at the fifth nucleotide[123,131]. However, it is unknown, whether RNA modification occurs at the same time at both modification sites in the complex[131]. Additionally, electron microscopy reveled a dimeric RNP, with two sets of guide RNA and four sets of proteins[135].

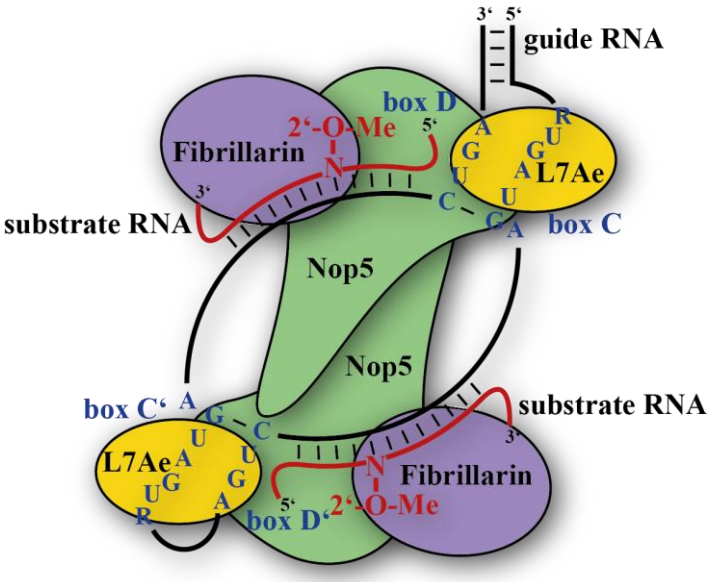


Figure 8: Schematic representation of a box C/D RNP. The guide RNA (black) contains the box C/D and C'/D' motifs, which act as a binding site for L7Ae (yellow) together with Nop5 (green) and fibrillarin (purple). 2'-O-methylation is performed on the fifth nucleotide of a substrate RNA (red) bound via Watson-Crick base pairing to the guide RNA. Both Nop5 proteins interact via their coiled-coils domains.

1.1.2.2 Box H/ACA RNPs

Box H/ACA RNPs carry out the pseudouridylation reaction in an RNA guided manner. The guide box H/ACA RNA adopts a secondary hairpin-hinge-hairpin-tail structure, with a conserved H- (ANANNA) and ACA-motif in the hinge and tail regions, respectively[122,124,136,137]. Both hairpins contain an internal loop structure ("pseudouridylation pocket"), which is the site of Watson-Crick base-pairing with the substrate RNA, that adopts a Ω-shaped structure[138–140]. After substrate recruiting, the target uridine is situated in the upper part of the loop, 13-16 nucleotides away from the H- or ACA-box. Both the target U and the adjacent 3' nucleotide R (A or G) are unpaired with the guide RNA[141]. Archaeal box H/ACA RNPs contain an additional K-turn motive, that is absent in eukaryotic H/ACA RNPs. While the typical H/ACA RNA in eukaryotes forms a bipartite structure with two hairpins, in archaea 1-3 hairpins have been observed an in humans an H/ACA RNA with four hairpins is possible[142,143]. To form a functional box H/ACA RNP, the guide RNA interacts with the four evolutionary conserved proteins Nop10, Cbf5 (Dyskerin in humans, NAP57 in rodents), Gar1 and Nhp2 (L7Ae in archaea), with a set of the four proteins on each hairpin (Figure 9)[144–148].

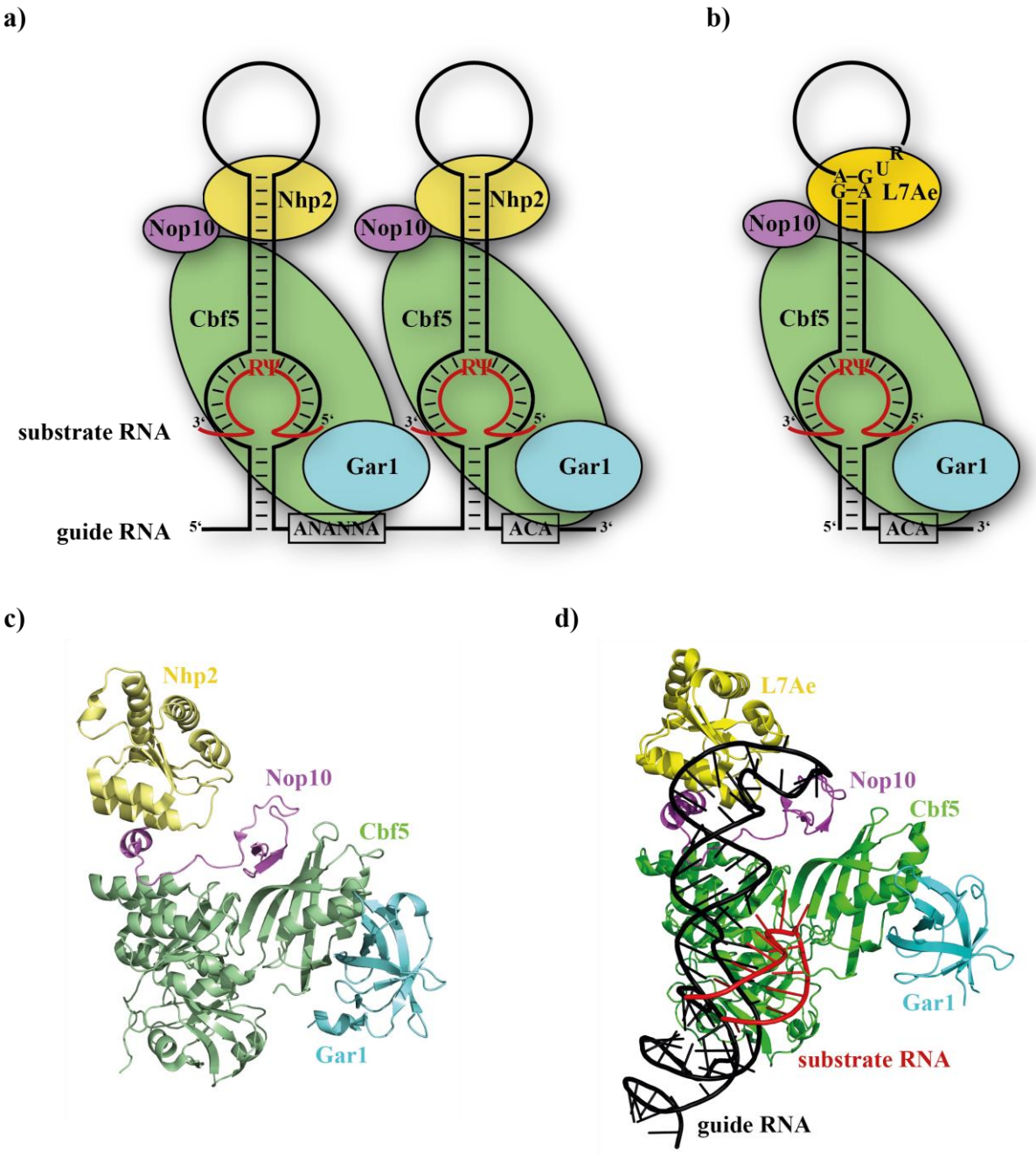


Figure 9: a) Schematic representation of a bipartite eukaryotic box H/ACA RNP, with the guide RNA (black) containing the H- and ACA-box motives and a set of four proteins (Nhp2 (yellow), Nop10 (pink), Cbf5 (green) and Gar1 (cyan)) situated at each hairpin. Two substrate RNAs (red) are bound via Watson-Crick base pairing in the pseudouridylation pocket and target nucleotide is marked as Ψ. b) Schematic representation of an archaeal H/ACA RNP, with the guide RNA containing the H-box motif. L7Ae (yellow) is bound via a kink-turn motif in the upper loop. c) Crystal structure of the trimeric Nop10 - Cbf5 (core domain) - Gar1 (core domain) complex from *Saccharomyces cerevisiae* (pdb: 3U28) as well as NMR structure of Nhp2 (S82W mutation) (pdb: 2LBW). Both structures are aligned after the archaeal RNP (d). d) Crystal structure of the full archaeal H/ACA RNP from *Pyrococcus furiosus* (pdb: 3HAY).

Cbf5 acts as the catalytic subunit and is structurally extremely similar to the *E. coli* TruB pseudouridine synthase, also carrying the N-terminal catalytic domain and the C-terminal

pseudouridine synthase and achaeosine transglycosylase (PUA) domain[101,149–151]. Cbf5 binds directly to the guide RNA via the uppers stem and the H- or ACA-motif, which interact with the PUA domain.
Nop10 binds to the upper stem of the guide RNA and to a conserved region of Cbf5, close to the catalytic site, in turn providing a binding interface for Nhp2. The binding region in Cbf5 for Nop10 is important for stability in other pseudouridine synthases, which suggests that Nop10 may also be involved in the structural organization and stabilization of the catalytic center in Cbf5[122]. Additionally, archaeal Nop10 contains a zinc ion binding site, which is absent in eukaryotic Nop10[152,153].
Nhp2 can interact with Cbf5 and Nop10 also in the absence of the guide RNA and additionally can bind unspecifically to RNA stem-loops[122,126,143,154–156]. In fact, the specific binding to H/ACA RNPs seems to arise from the trimeric Nhp2-Nop10-Cbf5 complex, that forms prior to assembly with the guide RNA. The archaeal homologue L7Ae on the other hand strictly requires the K-turn motif of the guide RNA to bind to the complex and can even bind to the substrate RNA in absence of the other proteins[126,143]. L7Ae is involved in the re-orientation of the substrate RNA bound to the guide RNA for catalysis[157].
Gar1 contains two glycine-arginine rich (GAR) domains at the N- and C-terminus of the protein, which flank the central core domain[158]. Gar1 does not interact with the guide RNA directly and is bound to Cbf5.
While the guide RNA in combination with Cbf5 and Nop10 are strictly required for basal pseudouridylation activity, Gar1 and L7Ae (in archaea) immensely increase the enzymatic activity of the complex, while Nhp2 (in eukaryotes) increases the activity, but seems to be dispensable in several RNPs[156,159,160]. Gar1 is involved in the substrate turnover after pseudouridine formation and is essential for multiple turnover activity of the complex[140,150,161,162]. Cbf5 contains a thumb-loop domain, which is responsible for the correct positioning of the target uridine at the catalytic active site and which contains the conserved aspartate, that is also the catalytic active site in non-guided pseudouridylation. Crystal structure analysis showed, that the thumb loop domain can either adapt a “closed” state, in which the substrate RNA is anchored at the catalytic site or an “open” conformation, in which the pseudouridylated substrate is released and the catalytic active site can be loaded with a new substrate RNA. In the closed state, the thumb loop forms nonspecific hydrogen bonds and electrostatic contacts with the sugar-phosphate backbone of the substrate RNA, which clarifies the non-specific protein interaction (in comparison to non-guided pseudouridylases) and the need for a guide RNA[138,140]. In the open state, the thumb loop is anchored at Gar1, which provides a binding surface for the opened loop[140,150,157,162].
The affinity of the four H/ACA proteins together towards the H/ACA guide RNA in eukaryotes and archaea is extremely strong. Archaeal complexes showed dissociation constants (K_d) in the nanomolar range, and eukaryotic complexes showed even tighter interactions with sub-nanomolar K_ds[159]. These experiments also revealed a very weak binding of Nhp2 alone towards the guide RNA, even though Nhp2 is an RNA-binding protein[155].
While it was possible to reconstitute eukaryotic RNPs *in vitro*[159,160], for the *in vivo* formation of RNPs the assembly factor Shq1 and the Gar1 homologue Naf1 are involved[163–165].

Additionally to the uridine to pseudouridine transformation, H/ACA RNPs have been associated with further functions, like the processing of pre-rRNA and ribosome formation[166,167] as well as telomere maintenance[168]. In fact, vertebrate telomerases contain an essential H/ACA RNP domain, and recently it was possible to obtain a cryo-EM structure of this H/ACA RNP domain as part of the human telomerase (Figure 10)[169].

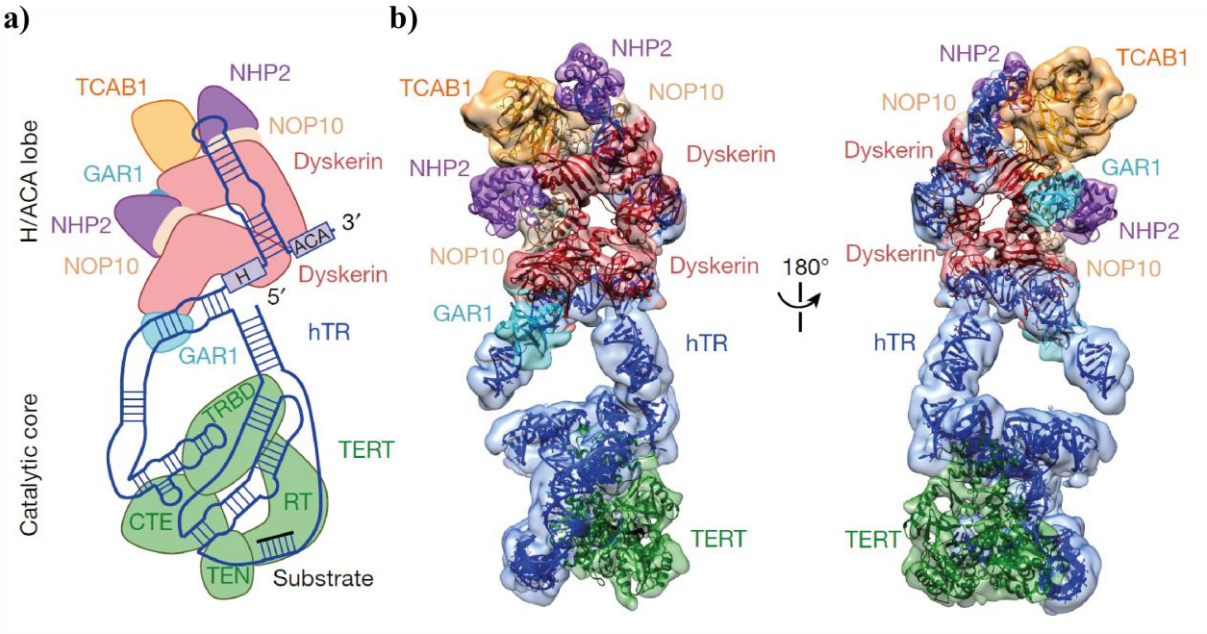


Figure 10: Structure of substrate-bound human telomerase enzyme, containing an H/ACA domain. a) Schematic representation of the subunits. b) Cryo-EM structures for the H/ACA domain at 8.2 Å and the catalytic core at 7.7 Å with fitted subunits (Figure taken from Nguyen *et al.*[169] with permission from the Springer Nature Customer Service Centre GmbH, © 2018 Springer Nature).

Mutations in the telomerase components hTR and hTERT, but also in the associated RNP proteins dyskerin, Nop10 and Nhp2 have been linked to Dykeratosis congenita (DC), a genetic disease including bone marrow failure, premature aging and increased susceptibility to cancer caused by abnormally short telomeres[170–173]. These DC mutations are also present in the conserved N- and C-terminal extensions of eukaryotic Cbf5. Furthermore, the X-linked form of DC is probably caused by mutations of the dyskerin PUA domain, weakening the interactions of dyskerin with the H/ACA domain of hTR and the assembly factor Shq1 and in general by impaired H/ACA RNP function[147,161,171,174].

1.2 Protein labeling by genetic code expansion

The occurrence and utilization of genetic code expansion was reviewed in "Genetic code expansion facilitates position-selective modification of nucleic acids and proteins" by Müller, Trucks, Schwalbe and Hengesbach[175].

1.2.1 General considerations for protein engineering

In general, proteins are composed of different amino acids, which are connected to one another by a peptide bound, defining the primary structure of the protein. The amino acids are defined by the characteristic side chain that they are carrying on their α-carbon atom. By the formation of different hydrogen bonds between amino- and carboxyl groups in the backbone of the peptide chain, which is influenced by the amino acids side chains, the secondary and furthermore the tertiary structure of a protein is defined[176,177]. For protein engineering, these characteristics of proteins give the opportunity to modify a protein either by the backbone of its peptide chain or by the various side chains of its amino acids. The backbone structure of a peptide chain can be changed by *in vivo* incorporation of proline analogues into the protein. The different preferences of proline analogues towards isomerization into a *cis* or *trans* conformation result in different structures of the peptide backbone, impacting the folding of the protein[178–180]. Another backbone modification technique utilizes the incorporation of a single[181,182] or several consecutive[183] β-amino acids – amino acids that carry two carbon atoms between the carboxyl and amine functional groups, with the amino acid side chain positioned at the β-carbon atom. These β-amino acids are present in cells, however they normally do not get incorporated by ribosomal protein synthesis[184], due to a strong selection of the ribosome against β-amino acids, manifesting as the disability of β-aminoacyl-tRNAs to effectively act as a substrate for the ribosome and participate in the peptide bond formation[185–187]. Incorporation *in vitro* can be achieved by chemical misacylation of suppressor tRNAs and utilization of modified ribosomes[181,188]. The backbone elongation, caused by the additional carbon atoms, and the adoption of a gauche conformation of the substituents at the α and β position of the β-amino acid, result in different secondary structures of the proteins. Proteins carrying β-amino acids are forming more stable helix structures in comparison to α-peptides[189] and have been utilized for β-peptide-based antibiotics, as a way to evade antibiotic resistances[190]. While backbone modifications on the peptide chain impact the core structure of the proteins, the modification of amino acid side chains leaves the peptide core structure intact. According to the "alanine world" hypothesis, todays canonical amino acids (except glycine, arginine and proline) originated from alanine during evolution[191]. This way, most of the canonical amino acids can be described as side chain modifications of alanine. Through side chain modifications of amino acids, a grand variety of functional groups can be added onto the amino acids, which then can be incorporated into proteins, while keeping the secondary protein structure (including protein folding and enzymatic activity) intact. The utilization of such "non-canonical" amino acids immensely broadens the field of protein engineering.

1.2.2 Non-canonical amino acids

All naturally occurring proteins are based on the basic building blocks of the 20 canonical amino acids[177]. The translational machinery incorporates these amino acids into peptide chains based on the genetic code. In all domains of life, the genetic code is universal, assigning the 64 possible triplet codons to a single amino acid or termination function (Figure 11). However, several deviations of this universal code during protein translation have been observed, resulting in abnormalities like unassigned codons, reassignment of codons to other canonic amino acids and furthermore the reassignment of codons to code for the proteinogenic amino acids selenocysteine and pyrrolysine (Figure 12a)[192–195]. In addition to this translational genetic code expansion, post-translational modifications of proteins have led to over 140 known non-proteinogenic amino acids in natural occurring proteins[196] and an additional 800 non-proteinogenic amino acids, which fulfill various functions, are known to exist in nature[197,198].

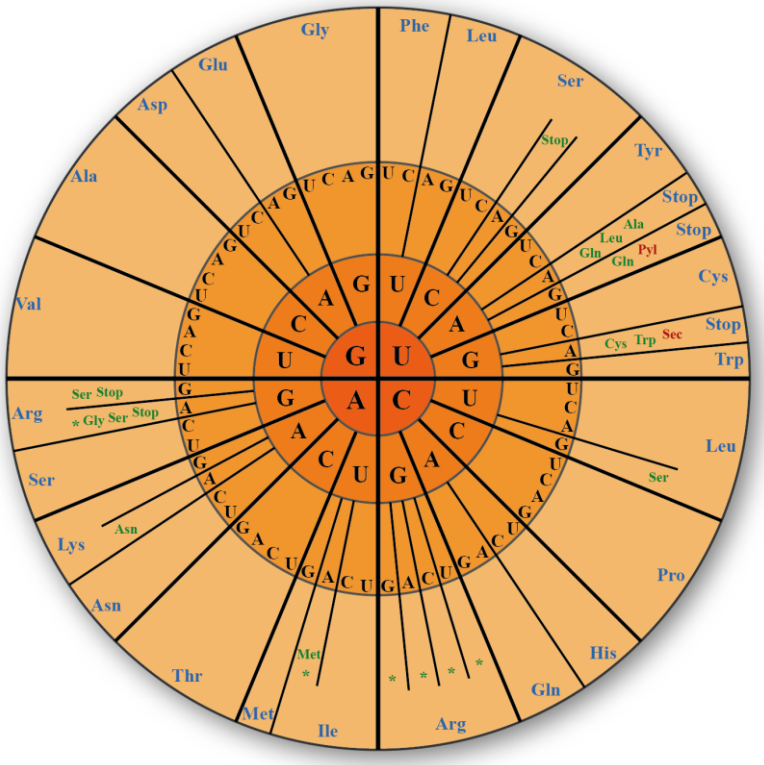


Figure 11: Codon wheel representing the standard genetic code with three letter codons (black) and the canonical amino acid/stop function they are assigned to (blue). In some organisms, several codons have been observed to have changed assignments to different canonical amino acids (green) or non-canonical amino acids (red). Some codons may be unassigned in some organisms or organelles (asterisk) (figure based on Ambrogelly *et al.*[192]).

In addition to these evolutionary genetic code expansions, which are sufficient for life, the field of protein modification gets extremely broadened by the addition of chemically

synthesized non-canonical amino acids, which add a completely new layer of complexity and potential to the possibilities of protein engineering. Like in nature, most non-canonical amino acids are synthesized by the modification of canonical amino acids[199–202]. Additionally, new synthesis strategies are constantly being developed, like the one pot reaction of aerobic oxidation and transamination of 1,2-diols into α-amino acids[203] or the chemoselective oxidation of α-hydroxy acids by a nitroxyl radical catalyst into α-keto acids[204]. Non-canonical amino acids can be designed in different ways to either directly act as spectroscopic probes or to be spectroscopic probe precursors with functional groups that can be modified post-translationally. Non-canonical amino acids that carry side chain modifications which are already defining them as functional spectroscopic probes, can be added translationally into the peptide chain and the protein can then be used in various spectroscopic techniques (Figure 12b). For *in vivo* incorporation of such spectroscopic probes, it is important that the functional groups, which are relevant for spectroscopic analysis, do not get destroyed in the environment of the cell. Non-canonical amino acids, that are spectroscopic probes, can be utilized for structural and dynamic investigations of a protein or protein complex. With the use of ^{13}C and/or ^{15}N isotopically labeled non-canonical amino acids, NMR spectroscopy investigations of large proteins and protein-ligand interactions can be performed[205,206]. Non-canonical amino acids with a paramagnetic probe, e.g. an attached nitroxide group, can be utilized for EPR spectroscopy, allowing distance measurements between two labeled proteins or inside a double labeled protein[207–211]. Combination of paramagnetic non-canonical amino acids and NMR spectroscopy allows the use of DNP techniques[212]. The utilization of UV-activatable groups in non-canonical amino acids can be utilized for photo-crosslinking between two proteins[213–215] or protein and RNA[216]. Furthermore, non-canonical amino acids bearing a photo-cleavable[209] or enzymatically cleavable[217] protection group can be used for deprotection and therefore activation of the spectroscopic function of the amino acid after incorporation. Lastly, the incorporation of fluorescent non-canonical amino acids can be used for fluorophore based techniques[218,219]. In addition to the incorporation of these non-canonical amino acids, that already act as spectroscopic probes, it is also possible to incorporate a non-canonical amino acid into the protein which is bearing a reactive linker group (Figure 12c). After the successful incorporation into the protein, this active group can be used to perform post-translational labeling of the protein, utilizing “click” chemistry (described in chapter 1.2.5).

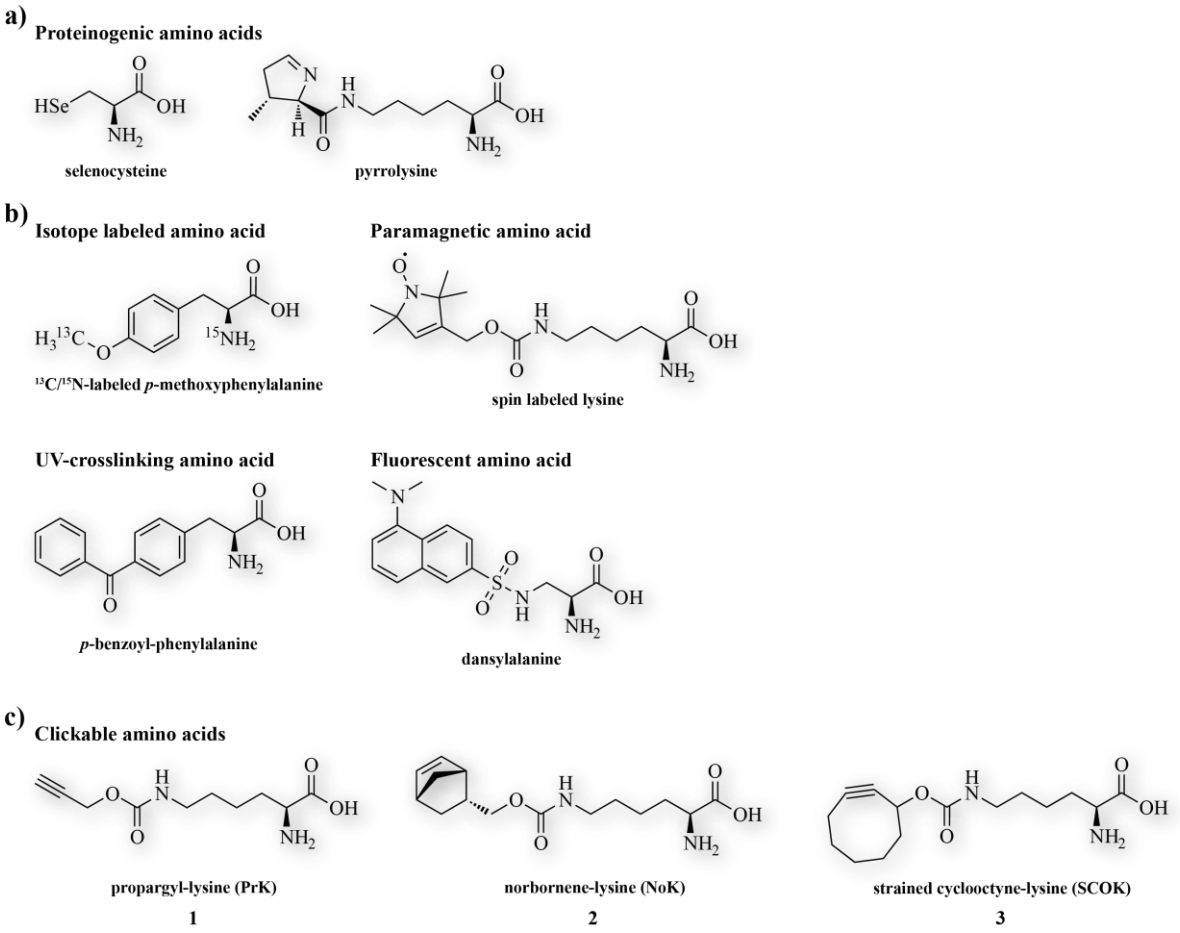


Figure 12: chemical structure of several exemplary non-canonical amino acids. a) proteinogenic amino acids: selenocysteine and pyrrolysine. b) non-canonical amino acids functioning as spectroscopic probes: $^{13}C/^{15}N$-labeled *p*-methoxyphenylalanine, spin labeled lysine, *p*-benzoyl-phenylalanine and dansylalanine. c) non-canonical amino acids for post-transcriptional labeling (which were utilized in this thesis): propargyl-lysine (1), norbornene-lysine (2) and strained cyclooctyne-lysine (3) [175].

1.2.3 Strategies for non-canonical amino acid incorporation

Several strategies have been developed for the incorporation of a non-canonical amino acid into a protein *in vitro* as well as *in vivo*. An *in vivo* approach for incorporation of non-canonical amino acids is the utilization of auxotrophic systems, in which one of the 20 canonical amino acids gets uniformly replaced by a non-canonical amino acid. The first utilization of this system was the replacement of methionine with selenomethionine[220]. The method relies on an auxotrophic *E. coli* strain, that cannot synthesize the canonical amino acid that is desired to be replaced. Additionally, the used medium also has be depleted of that amino acid. With the addition of the non-canonical amino acid, that is ideally a structure analogue of the canonical amino acid, into the medium, the translation machinery will incorporate the analogue non-canonical amino acid into the elongating peptide chain in place of the canonical amino acid. Utilizing this technique, it was possible to uniformly incorporate analogues of methionine[221–226], phenylalanine[227,228], proline[179] and tryptophan[229,230]. In

some cases, the aminoacyl tRNA synthetase had to be modified, to efficiently load the tRNA with the non-canonical amino acid.
Another approach, which can site-specifically incorporate a single non-canonical amino acid at a desired position, without the shortcomings of uniform amino acid replacement, is expressed protein ligation (EPL), which utilizes solid phase peptide synthesis to generate semi-synthetic proteins[231–234]. An artificial peptide is generated via solid phase synthesis, which gives the possibility to incorporate the non-canonical amino acid at a desired position. The synthesized peptide can then be ligated to an N-terminal cysteine of a protein generated by expression. A shortcoming of this method however is the limit of ~100 amino acids for peptides generated with peptide solid phase synthesis[235].

1.2.4 Amber suppression techniques

To overcome the shortcomings of the methods described in chapter 1.2.3, a technique to site-specifically incorporate non-canonical amino acids translationally *in vitro* as well as *in vivo* was developed. The development of this "amber suppression technique", that was also utilized in this thesis, will be described in this chapter.

1.2.4.1 Nonsense and frameshift codon utilization

The endogenous tRNAs in an organism are adhering to the universal code of triplet codon recognition (Figure 11). Three codon combinations code for the termination of protein translation, so called "stop codons" or "nonsense codons". For these codons, normally no tRNAs with the complementary anticodon are present in the cell. However, in some organisms, point mutations in the anticodon of tRNAs to match the nonsense codons, and utilization of the nonsense codons in the genetic code have been observed in nature, for example for the incorporation of the proteinogenic amino acids selenocysteine and pyrrolysine[192,236]. This method derived from nature has led to the development of *in vitro* mutation of the anticodon sequences of tRNAs derived from yeast[237] to match nonsense codons and it was possible to let tRNAs respond to the opal (UGA)[236], ochre (UAA)[238] and amber (UAG)[239,240] codons. The anticodon-bearing tRNAs are still in competition with the release factors, which would normally report to the nonsense codon to initiate termination, however they are able to suppress the stop function of the nonsense codon and incorporate the amino acid attached to them into the elongating protein. Of the three nonsense codons, the amber codon has mostly been used in this technique, because it is the least abundant termination codon in *E. coli*[241], leading to the term of "amber suppression technique". Another codon utilization method that has been derived from nature is the use of frameshift codons[242]. With the usage of tRNA, that responds to four-letter codons, the genetic code can theoretically be expanded from 64 to 256 possible codon assignments[243]. A benefit of this method is the fact that none of the already assigned triplet codons have to be repurposed and additionally, quadruplet codons are not in competition with release factors[244]. The suppression of frameshift codons has been utilized *in vitro*[244,245] as well as *in vivo*[246,247].

Utilizing site-directed mutagenesis, a nonsense or frameshift codon can be inserted into the plasmid at nearly any desired position, which highlights the site-specificity of this method in general.

1.2.4.2 Chemical misacylation of tRNAs with non-canonical amino acids

With the ability to recode tRNAs to respond to nonsense codons, it was also desired to use this technique to further manipulate the translation machinery and incorporate synthesized non-canonical amino acids by amber suppression. The basis for the rise of these methods was the fact that the pairing of mRNA and tRNA by the codon/anticodon recognition is independent of the amino acid that is attached to the tRNA, which is in line with Watson and Crick's adaptor hypothesis[248,249]. This made it possible to covalently attach non-canonical amino acids to tRNAs mediated by T4 RNA ligase and therefore "misacylate" canonical tRNAs with non-canonical amino acids[250]. At first, the amino acids attached to the tRNA had to be N-terminally protected, which made them unsuitable for application in translational systems[251]. However, with successful deprotection in good yield, it was possible to incorporate non-canonical amino acids via nonsense codon suppression using chemically misacylated suppressor tRNAs in an *in vitro* transcription/translation system[239,252]. To efficiently incorporate non-canonical amino acids, it was important that the chemically misacylated tRNAs did not have to compete with their endogenous counterparts, that were acylated by cognate aminoacyl tRNA synthetases (aaRS). For this reason, amber-mutated tRNAs, which were no longer recognized by their cognate aaRS and showed weak amber suppression efficiency *in vivo* were good candidates for misacylation with non-canonical amino acids[253,254]. With the injection of *in vitro* processed mRNA and chemically misacylated tRNA into *Xenopus* oocytes, first *in vivo* incorporation of non-canonical amino acids into proteins was made possible[255–257]. However, this approach still relied on the non-canonical amino acid being chemically attached to the tRNA beforehand.

1.2.4.3 Bioorthogonal systems

The first amber suppression methods, that no longer had to rely on chemical acylation of tRNAs were utilizing aaRS/tRNA pairs from *Escherichia coli*. Orthogonality was introduced into these systems by mutating the tRNA in a way that its endogenous aaRS could no longer acylate the tRNA with its cognate amino acid, but still preserve its ability to partake in the translational process[258]. Also, the aaRS was altered by random point mutations and generation of mutant libraries, which were then tested with the intent to reenable recognition of the tRNA by the mutated aaRS, generating a completely orthogonal aaRS/tRNA pair[259]. A problem in this method however was the fact that the mutated aaRS was not only acylating the mutant tRNA but also its cognate wildtype tRNA, even after improving the discrimination of mutant aaRS towards mutant tRNA. Any misacylation of wildtype RNA with a non-canonical amino acid would result in uniform incorporation of this amino acid, undermining the site specificity purpose of the method. Furthermore, a cell-wide incorporation of non-canonical amino acids might be lethal for the cell.

Modern amber suppression techniques are based on bioorthogonal systems. The biorthogonality is ensured by utilizing aaRS/tRNA pairs from a different domain of life than the host expression organism. Heterologous aaRS/tRNA pairs from eukaryotes or archaea are used for amber anticodon mutation and transferred into *E. coli*, which allows their usage alongside the endogenous *E. coli* translational machinery[235,260]. Many such aaRS/tRNA pairs have been utilized, mainly from a variety of methane producing archaea like *Methanococcus jannaschii*[215,261], *Methanosarcina mazei*[199,262], *Methanosarcina barkeri*[263] or *Methanomethylophilus alvus*[264]. To use this technique to its full potential, orthogonality between the imported aaRS/tRNA and the host organisms endogenous aaRS/tRNA is strictly required. The imported aaRS must only be able to aminoacylate its cognate tRNA and none of the endogenous tRNAs. Likewise, none of the endogenous aaRS must be able to attach a canonical amino acid onto the imported tRNA. To generate an aaRS/tRNA pair, that completely complies to this bioorthogonality rule, first, a library of aaRS and tRNAs derived from the archaeal organisms is created, with different point mutations based on crystal structure data[215,235,261]. An important point is also to ensure the specificity of the aaRS towards the non-canonical amino acid, that is desired to be incorporated, since different non-canonical amino acids may require different mutations to the aaRS to ensure binding inside the substrate pocket[200,201]. The generated mutant libraries are subjected to several cycles of a positive/negative selection scheme[235]. The negative selection scheme utilizes the expression of the toxic barnase gene with an inserted amber codon. No non-canonical amino acid is added for this test, so if bioorthogonality is retained, no barnase is expressed. However, if endogenous amino acids get incorporated in response to the amber codon, cell death occurs. Likewise, the positive selection scheme utilizes the expression of the antibiotic resistance providing genes β-lactamase or chloramphenicol acetyltransferase which also contain an amber codon mutation. This time, the non-canonical amino acid is added and the expression is carried out in presence of ampicillin or chloramphenicol. If the non-canonical amino acid is incorporated and the enzymes are expressed, the cell is resistant to the antibiotic. If incorporation fails, cell death occurs.

To utilize the bioorthogonal amber suppression system, the genes coding for the bioorthogonal aaRS and tRNA have to be inserted into the host organism alongside the amber mutated plasmid, which is designed to express the amber mutated protein of interest. The amber suppression system is depicted in Figure 13. The bioorthogonal aaRS recognizes the medium-supplemented non-canonical amino acid and after binding, ATP-mediated activation of the non-canonical amino acid is performed. The bioorthogonal tRNA recognizes its cognate aaRS, binds to it and is acylated with the non-canonical amino acid. Afterwards, the loaded tRNA gets released and mediated by elongation factor Tu, the bioorthogonal tRNA with the non-canonical amino acid binds to the ribosome and the non-canonical amino acid is incorporated into the elongating peptide chain in response to an amber codon (instead of terminating the translation). After successful incorporation, the translation continues normally, until an opal or ochre codon is reached in the mRNA.

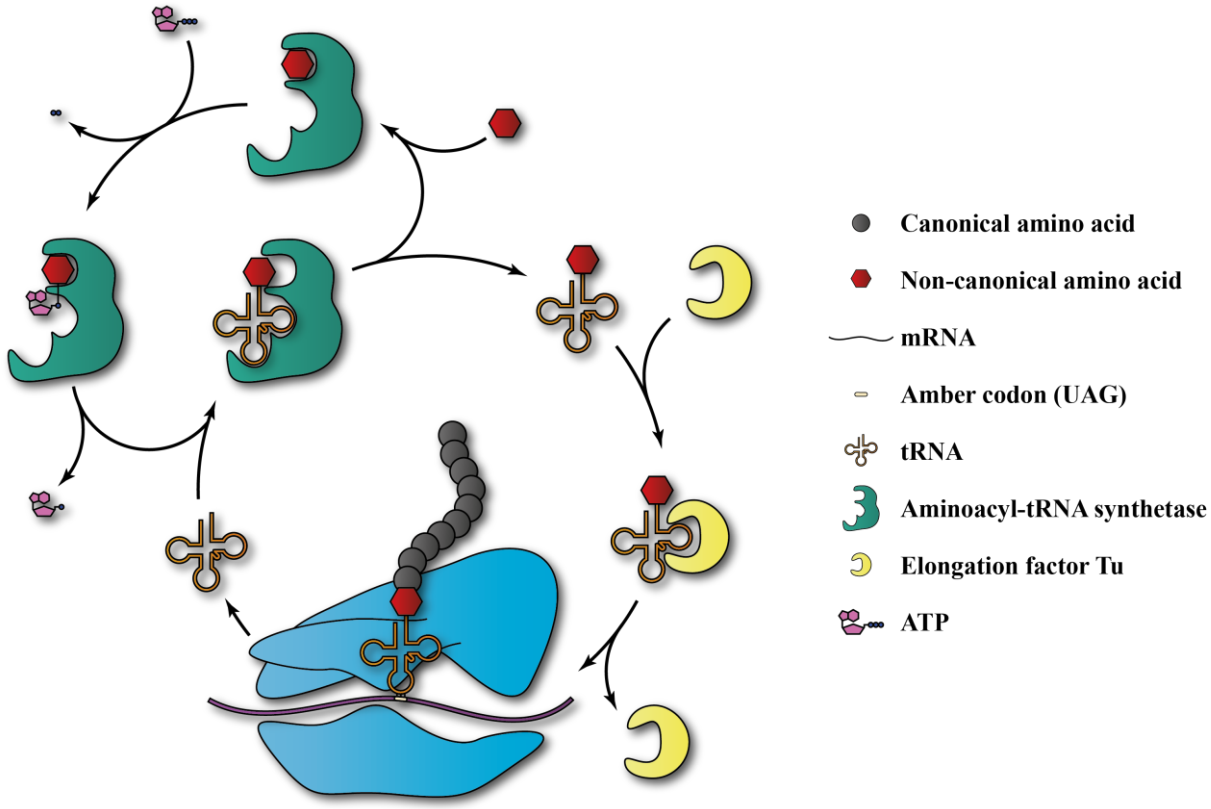


Figure 13: Schematic representation of non-canonical amino acid incorporation via the bioorthogonal amber suppression technique. The non-canonical amino acid (red) binds to a bioorthogonal aminoacyl-tRNA synthetase (green) and is activated by ATP (pink) under pyrophosphate release. The aaRS attaches the non-canonical amino acid onto a bioorthogonal tRNA (orange) and releases it. The tRNA is bound to the ribosome (blue), mediated by elongation factor Tu (yellow). In response to an amber codon, which was mutated into the mRNA (purple), the non-canonical amino acid gets incorporated into the elongating peptide chain[175].

The bioorthogonal amber suppression system has seen multiple improvements for incorporation yield of non-canonical amino acids over the course of time. The genes to produce aaRS and tRNA were added onto a single plasmid several times and put under the control of glnS' and proK promotors[265]. To ensure a constant production of aaRS, both constitutive and inducible promotors were added to the gene[266,267], resulting in the pEVOL plasmid, which was utilized in this thesis. To increase the expression yield of non-canonical amino acid bearing proteins even further, several modifications were applied during the last couple years. The affinity of the bioorthogonal tRNA towards elongation factor Tu, which is responsible for transporting the aminoacylated tRNA to the ribosome, was increased by several mutations to the tRNAs acceptor and T stem[268] and EF Tu itself was modified to increase its affinity towards a tRNA loaded with an non-canonical amino acid[269,270]. Another problem affecting protein production by amber codon suppression is the competition of bioorthogonal amber tRNA with release factors, that are still able to terminate the translation. To counter this problem, several strategies were employed recently[271]. The biorthogonality was expanded by utilizing bioorthogonal mRNA and ribosomes[272] and the release factor RF1 was completely deleted from the host system, after mutating every endogenous amber codon

into ochre codons, making RF1 non-essential and only retaining the single amber codon, which is desired for non-canonical amino acid incorporation[273]. Another interesting approach for nonsense suppression, utilizes the release factors specificity for the uridines in the three stop codons, by mutating uridine into pseudouridine, which resulted in a nonsense suppression an the incorporation of serine and threonine for ΨAA and ΨAG and tyrosine and phenylalanine for ΨGA[274–276].

1.2.5 Post-translational labeling strategies

If a non-canonical amino acid is incorporated that is not a spectroscopic probe on its own, but rather a linker to attach a spectroscopic probe, additional strategies have to be employed for post-translational labeling. Since this labeling has to be employed after the expression and purification of the natively folded protein, the labeling reaction has to fulfill several requirements. Mild reactions reaction conditions that are compatible with the folded protein have to be utilized to avoid denaturation, such as tolerable temperatures, aqueous environment and no inert gas. Additionally, the utilized reactions have to be broadly applicable and result in high labeling yields. Those reactions were first explored by Huisgen[277–279] and later summarized under the term “click reactions” by Kolb, Finn and Sharpless[280]. To comply with these requirements, these reactions have to be driven by a strong thermodynamic force, to readily attach large and sensitive labels to the protein. A commonly used reaction for this purpose between the functional group of an incorporated non-canonical amino acid and a spectroscopic probe is the 1,3-dipolar cycloaddition reactions. In this reaction, a 1,3-dipolar molecule (commonly an azide) is brought to reaction with a dipolarophile (commonly an alkyne), resulting in the formation of a 1,2,3-triazole[281,282]. To perform the specific reaction between an azide and alkine under the defined “click” conditions, either a copper(I) catalyst (usually reduced copper(II)-sulfate in combination with a ligand to prevent reoxidation) has to be utilized (copper(I)-catalyzed azide-alkyne cycloaddition, “CuAAC”)[283,284] or the reaction is performed in a copper-free way, gaining the activation energy from a strained ring system (typically cyclooctynes), which is released upon triazole formation (strain-promoted azide-alkyne cycloaddition, “SPAAC”) (Figure 14a,b)[200,285–288]. Another coupling approach is the reaction of strained dienophiles with tetrazines under N_2 release (strain-promoted inverse-electron-demand Diels-Alder cycloaddition, “SPIEDAC”) (Figure 14c)[201,289]. SPAAC and SPIEDAC reactions are especially important for *in vivo* labeling methods, since reactions performed in the environment of the cell have to be fully orthogonal to endogenous biochemical reactions and copper salt catalysts can act toxic towards the organism[290,291]. Also the utilized labels in *in vivo* reactions have to be inert towards biochemical processes in the cell, so they do not get metabolized or undergo reactions with endogenous functional groups.

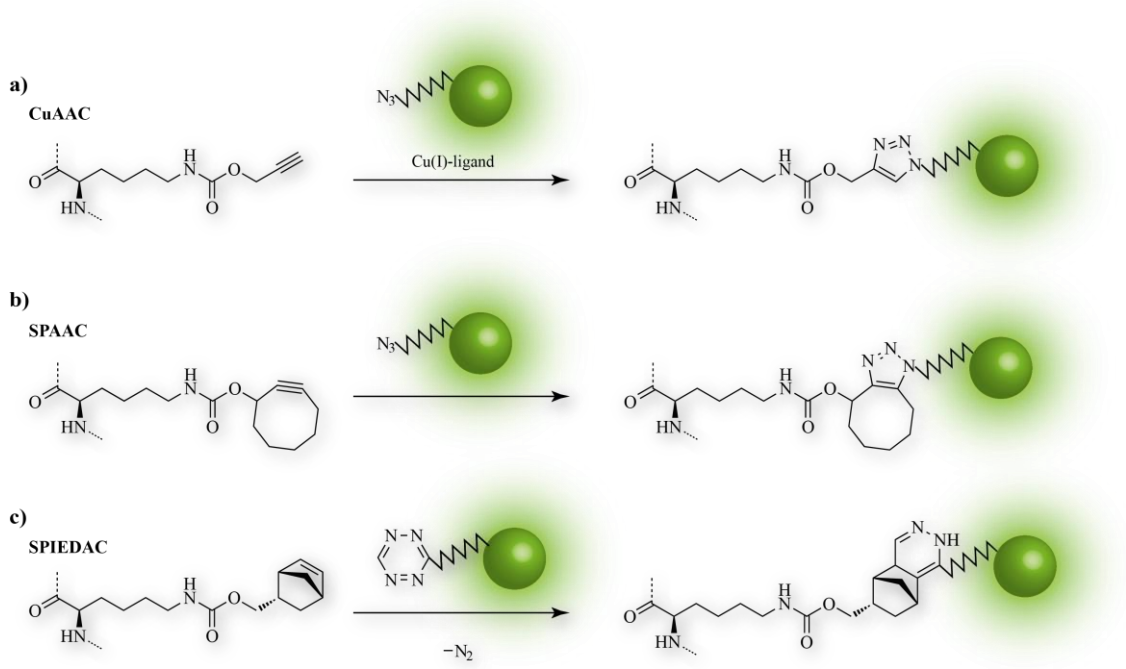


Figure 14: Click chemistry reactions illustrated for several non-canonical amino acids, after incorporation into a protein. a) Copper(I)-catalyzed azide-alkyne cycloaddition, "CuAAC" between an alkyne and an azide-label, typically performed with ligand stabilized Cu(I). b) Strain-promoted azide-alkyne cycloaddition, "SPAAC" between a strained alkyne and an azide-label. c) Strain-promoted inverse-electron-demand Diels-Alder cycloaddition, "SPIEDAC" between a norbornene and a tetrazine-label.

1.3 Principles and utilization of fluorescence single molecule spectroscopy

1.3.1 Theoretical background

1.3.1.1 Fluorescence and phosphorescence

Fluorescence describes the deactivation via radiation of a molecule from an excited state to the ground state[292–294]. The steps of fluorescence are described in Figure 15a. First, the molecule, which is in the electronic ground state (S_0), absorbs energy in the form of a photon and gets into an excited electronic state (S_1) via a vibronic transition. This transition between the electronic ground and excited state is much faster than any nuclear motions (10^{-15} s vs. 10^{-13} s), therefore it occurs without change of the nuclear coordinate (vertical transition). To ascertain compatibility of the excited electronic state with the ground state, simultaneously with the change of the electronic state, a change of the vibrational state occurs. According to the Franck-Codon principle, a vibronic transition is most likely to happen if the vibrational wave functions of both states are overlapping. Afterwards, the excited molecule emits energy via vibrational relaxation, until it reaches the vibrational ground state of the excited electronic state, in accordance with Kasha's rule[292–294].

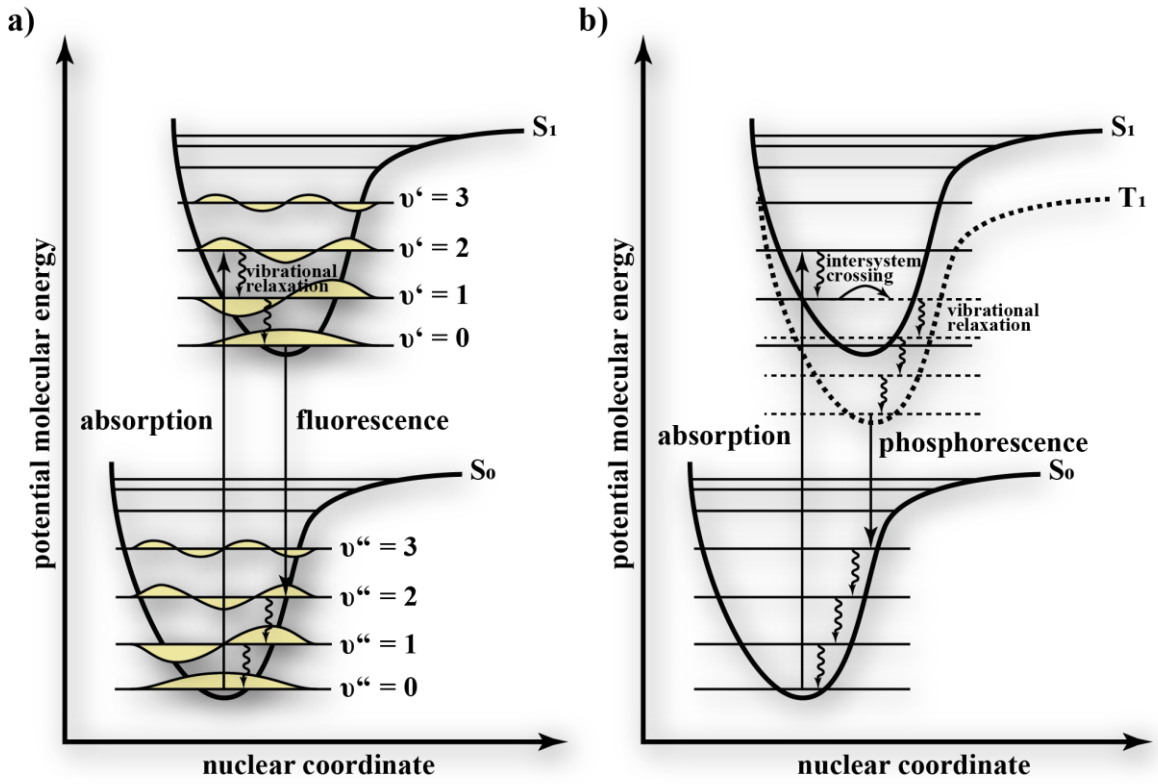


Figure 15: a) Schematic representation of the mechanism of fluorescence. A molecule in the electronic ground state (S_0) absorbs a photon, transitions into the excited state (S_1), reaches the vibrational ground state (v'=0) via vibrational relaxation and transitions back into the electronic ground state via emission of a photon. b) Simplified schematic representation of phosphorescence with an intersystem crossing into the excited triplet state (T_1), vibrational relaxation into the vibrational ground state of the triplet state and emission of a photon.

The energy gap between the vibrational ground state of the excited electronic state and a vibrational state of the electronic ground state (with an overlapping wave function) can be bridged by the spontaneous emission of a photon, which is called fluorescence. The emitted energy has a higher wavelength than the absorbed energy, due to the energy loss that occurrs through vibrational relaxation. In some cases, the energy gap between electronic excited and ground state can be transitioned by internal conversion, e.g. if the solvent surrounding the molecules is capable of absorbing the energy emitted by vibrational relaxation (fluorescence quenching). Additionally, in case of a strong spin-orbit interaction spin decoupling can occur, which can lead to an intersystem crossing between the excited electronic singlet (S_1) and triplet (T_1) state (Figure 15b). In the excited triplet state, the molecule also reaches the lowest vibrational level via vibrational relaxation. The transition back to the electronic singlet ground state is spin-forbidden and occurs only slowly, resulting in so called phosphorescence. Therefore, phosphorescence is dependent on heavy atoms with strong spin-orbit interaction and is mostly prominent in the solid state, where competition between internal conversion and phosphorescence is low. While fluorescence occurs nanoseconds after the excitation, phosphorescence occurs several seconds (or even hours) after the excitation.[292–294]

1.3.1.2 Förster Resonance Energy Transfer

Förster resonance energy transfer (FRET) describes a radiationless energy transfer between a donor and an acceptor molecule (Figure 16a). This energy transfer is caused by dipole-dipole interactions, if the donor and acceptor are in close proximity (nanometer ranges). According to the Förster theory, the energy transfer is most efficient, if the emission spectrum of the donor molecule and the absorption spectrum of the acceptor molecule significantly overlap (Figure 16b).

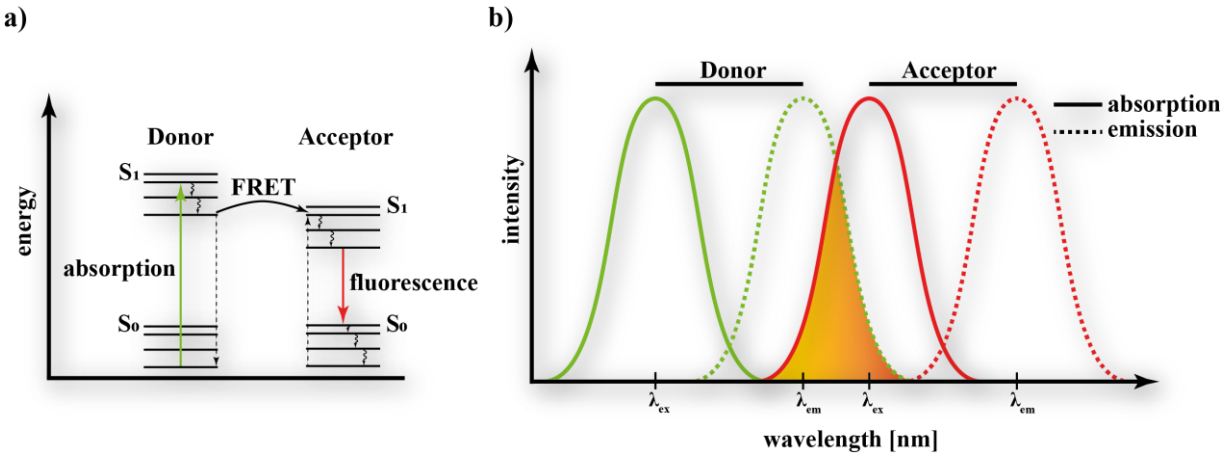


Figure 16: a) Simplified Jablonski diagram of a donor and acceptor molecule, describing the effect of FRET. b) Simplified schematic representation of absorption spectra (straight lines) and emission spectra (dashed lines) of a donor and acceptor molecule. Overlap of the donor emission and acceptor absorption spectra is marked in orange.

The FRET efficiency (E-FRET or E) is anti-correlated with the distance between the donor and acceptor molecule (Figure 17) and described through

$$E = \frac{R_0^6}{R_0^6 + R^6} = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6}$$

with dye distance R and Förster radius R_0.
The Förster radius R_0 describes a dye pair and system specific value, at which the dye distance R results in an E-FRET of 0.5 and is described through

$$R_0 = \frac{9 \cdot (\ln 10) \cdot (\kappa^2 Q_D J)}{128 \cdot \pi^5 \cdot N_A \cdot n^4}$$

with orientation factor between dye molecules κ, quantum yield of the donor molecule Q_D, spectral overlap J, Avogadro constant N_A, and refraction index of the surrounding medium n. FRET based spectroscopy methods cover a dye distance of approximately 2-10 nm and can be utilized as a molecular ruler[292–294].

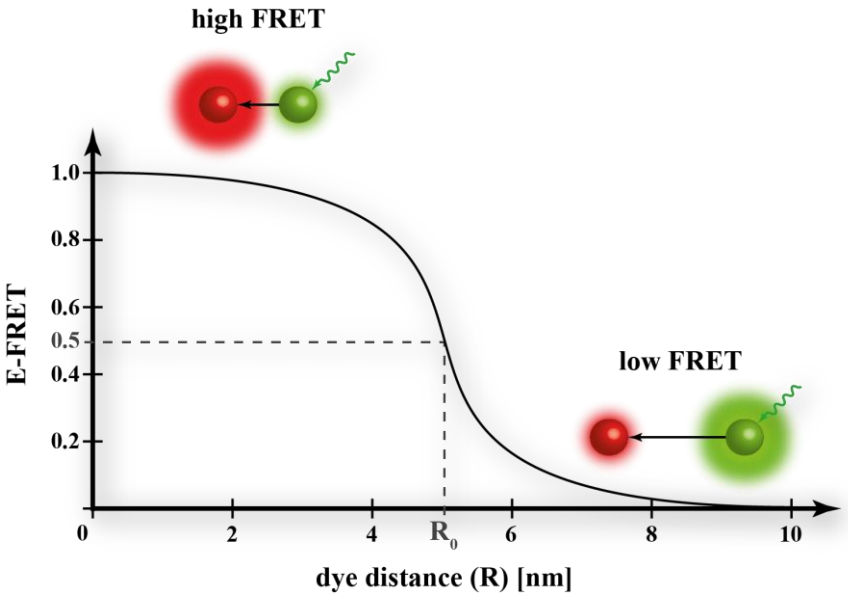


Figure 17: Schematic representation of E-FRET in correlation to dye distance (R). High and low FRET states can occur at small and great dye distances respectively. R_0 describes the dye distance at which E-FRET = 0.5.

1.3.2 Single-molecule spectroscopy

Many spectroscopic methods, like NMR, EPR, cryo-EM or crystallization are based on "ensemble" or "bulk" type experiments. The obtained data is integrated over a large number of molecules, to gain an average value out of the measured sample and minor populations can get lost by averaging over the ensemble of all populations. However, minor populations or states in biological, biochemical and biophysical systems can give crucial information about the dynamic of an observed system and may play an important role in the dissection and understanding of such systems. To investigate every population or state of a system of interest, approaches on a single-molecule level are important. Single-molecule data can be generated by techniques like detection of single ions by single-molecule mass spectrometry[295,296] or the tracing of polymer-chains by atomic force microscopy (AFM)[297,298]. While AFM was also utilized to study the folding pathways of RNA[299–301] and proteins[302] and to trace the polymerase[303] and the ribosome[304] during transcription/translation, many single-molecular approaches in the biochemical and biophysical field of research utilize the labeling of samples with fluorophores in combination with sensitive detectors, that are capable of detecting single photons. After the first observation of a nucleic acid on a single-molecular level[305], several developments in recent years made it possible to broadly establish single-molecule fluorescence techniques: the development of both the required hardware, like highly sensitive cameras (especially electron-multiplying charged couple device (EM-CCD) cameras[306]), microscopes (PTIR and OTIR setups, described in chapter 1.3.2.2), detectors and whole laser-setups[307] as well as "clickable" fluorophore dyes and dyes with increased brightness and chemical stability, like the now widely used Cyanine (Cy), ATTO and Alexa fluorophores[308,309].

The labeling of RNA for single molecule fluorescence techniques and the utilization of fluorophore single-molecule spectroscopy was reviewed in "Strategic Labelling Approaches for RNA Single-Molecule Spectroscopy" by Hanspach, Trucks and Hengesbach[310].

1.3.2.1 Super-resolution fluorescence microscopy

Fluorescence microscopy techniques are widely used to observe biological and biophysical systems. The fluorophore-based approach allows the specific labeling of molecules and in combination with microscopy, a sample can be observed in real time. One drawback from light microscopy however is the diffraction limit of light (Abbe-Limit), approximately 200-300 nm in the lateral and 500-700 nm in the axial direction, which does not allow the resolution of images below this limit[311,312]. However, many subcellular structures are below this limit and thus need non-conventional microscopy methods to be observable. To overcome this limit, several super-resolution microscopy methods have been established.
Reversible Saturable Optically Linear Fluorescence Transition (RESOLFT) microscopy utilizes fluorophores that can be reversibly photoswitched between a dark "off" state and a fluorescent "on" state. Stimulated Emission Depletion (STED) microscopy in particular utilizes fluorophores, where the dark state is the ground state of the fluorophore[312–315]. In these techniques, a sample is illuminated with a laser to activate the fluorophores and

simultaneously, a second laser (depletion laser) is used, that induces a photon with the same energy than the energy barrier between the excited and ground state of the fluorophore. This leads to a stimulated emission of the excited fluorophores, which are radiationlessly "switched off" and effectively leads to suppressed spontaneous fluorescence. The depletion laser is focused around the center of the excitation laser and has nearly zero intensity at the focal point. By utilizing an appropriately high laser power of the depletion laser, the region of depleted fluorophores can be expanded and only a small region at the focal point contains fluorophores which are emitting fluorescence. Through the overlap of both lasers in a circular area around the focal point, the area containing fluorescence emitting molecules can be decreased to a size below the Abbe-limit. This highly increases the resolution of STED compared to normal fluorescence microscopy, resulting in a spatial resolution of 20-70 nm and a temporal resolution of milliseconds to seconds[316]. Fluorophores often utilized in this method are fluorescent proteins and fluorophore-labeled antibodies[317,318].
Another set of methods combines fluorescence microscopy with single-molecule localization to generate super-resolution images. Stochastic Optical Reconstruction microscopy (STORM), Photoactivated Localization Microscopy (PALM) and Fluorescence Photoactivation Localization Microscopy (FPALM) utilize photoswitchable fluorescent dyes, that are excited at a different wavelength than the imaging light[311,319–321]. In this way, the number of molecules, which are emitting fluorescence at a given time can be controlled, which circumvents the Abbe-limit and results in individual fluorophores, which are optically resolvable from one another. The activation-imaging-deactivation process is done consecutively in a wide-field imaging approach, in parallelly mapping the coordinates of a large number of fluorophores and constructing a super-resolution image from the different snapshots. These methods can achieve a spatial resolution of 10-30 nm and a temporal resolution of seconds to minutes[316].

1.3.2.2 Single-molecule FRET spectroscopy

smFRET spectroscopy utilizes energy transfer between a donor and an acceptor molecule on a single-molecule level. Each individual (and observable) conformation that the donor/acceptor system can adapt results in a specific FRET-state (or rather a homogenous FRET-population as a result of shot noise). Due to the single-molecule character of the method, homogenous populations, each corresponding to a specific inter-dye distance, can be observed within a heterogenous population distribution. Experimental determination of a FRET-state (E-FRET) can be achieved by detection of donor and acceptor intensities via

$$E = \frac{I_{acceptor}}{I_{acceptor} + I_{donor}}$$

This allows to make conclusions about structural changes as well as to observe dynamics in the system of interest.
With the utilization of a confocal laser setup, molecules on a surface or in a solution can be observed. While a time resolution in the sub-milisecond range is possible with this method[322], the observation time for solution based approaches is limited by the time the

molecules pass through the excitation beam of the laser, which can be quite restricted and prevents the observation of processes and conformational changes that take longer than this short time window.

To overcome this shortcoming of confocal based methods, total internal reflection fluorescence (TIRF) microscopy is often utilized to allow longer observation times of the sample. In TIRF microscopy, the excitation beam of the laser meets the glass surface in a critical angle and is totally reflected, which creates an evanescent field, that extends only ~100-200 nm into the sample. The evanescent field only excites a small number of molecules and greatly reduces background fluorescence[307]. The total reflection can either be realized with a prism on top of the quartz slide surface of the measurement channels (prism type total internal reflection, PTIR) or on the surface of the objective below the cover slip of the measurement channels (objective type total internal reflection, OTIR) (Figure 18a,b)[323]. For the TIR based methods, the immobilization of the sample is required. This is realized by passivating the glass surface by adsorption of BSA/biotinBSA (for nucleic acids) or by aminosilanizing the surface and reaction with NHS-ester modified PEG/PEG-biotin (for proteins and nucleic acid/protein complexes)[307,324,325]. This procedure prevents non-specific binding of molecules to the glass surface and at the same time distributes biotin-linkers across the surface. By introducing a biotin-linker into the molecules of interest, the fluorophore containing samples can be specifically immobilized on the glass surface by the addition of streptavidin into the measurement channel prior to sample application via the formation of biotin-streptavidin-biotin bonds (Figure 18c). By the observation of immobilized molecules, basically the only time limiting factor is the photostability of the used fluorophores. This allows observation times in the seconds to minutes range up to over an hour, which gives the possibility for time-resolved analysis of the sample to observe the conversion between different states and identify reaction and folding pathways, intermediates and structural changes including complex dynamics[326–329]. Single-molecule FRET spectroscopy has also successfully been used in the study of many RNP complexes[330], including the spliceosome[331,332], a G-quadruplex helicase[333,334], the telomerase[335,336], the ribosome[337–340], argonaute proteins[341] as well as other complex structures like riboswitches[342,343] or the HIV envelope glycoprotein[344,345]. Also live imaging *in vivo* has been implemented[346,347] and furthermore, with the utilization of one donor and two acceptor dyes, three-color smFRET has successfully been utilized to study interactions between three different molecules[348–350].

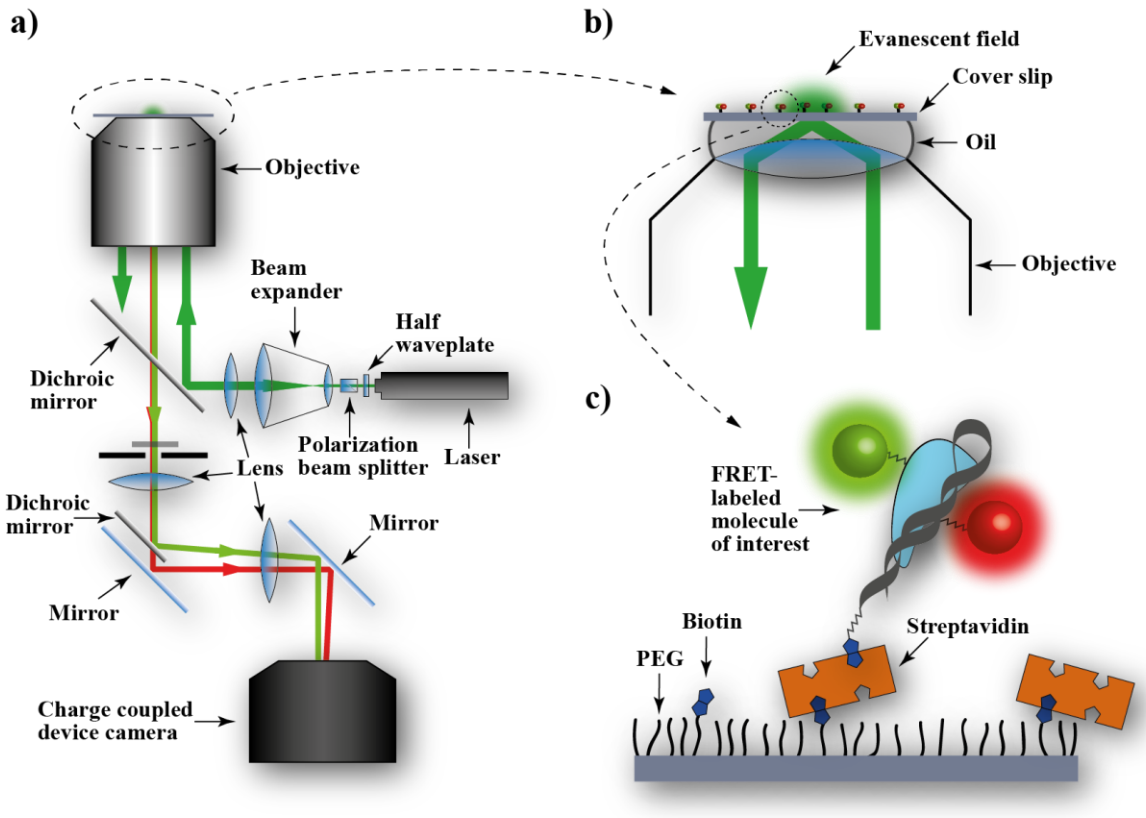


Figure 18: a) Typical setup of an OTIR-based microscope for smFRET spectroscopy. The laser beam is lead via a dichromatic mirror into the objective and totally reflected on the bottom of the cover slip, creating an evanescent field. The emitted fluorescence is split into the donor and acceptor channel by a dichroic mirror/mirror setup and the intensities are individually detected via a CCD camera. b) Magnification of the section enclosed by the dashed circle in (a). The evanescence field protrudes up to 200 nm into the sample, exciting only a small number of molecules. c) Magnification of the section enclosed by the dashed circle in (b). The surface of the cover slip is aminosilanized and PEGylated. Some PEG molecules contain biotin linkers, which allow the binding of a biotinylated fluorophore sample via biotin-streptavidin-biotin interactions (figure based on Roy *et al.*[307]).

For smFRET spectroscopy, a broad plethora of donor/acceptor fluorophore dyes is commercially available, which can be used for "click" based coupling methods to incorporate the dye into a nucleic acid or protein, preferably at a labeling site that can be chosen from available structural data[310]. Normally, these dyes also contain a linker between the clickable group and the fluorophore to prevent any negative effects of the dye towards the biomolecular structure of the system of interest. To prevent early photobleaching of the dyes, suppress dye blinking, and reduce the time that fluorophores are in the triplet dark state, a triplet state quencher in combination with an oxygen-scavenging system is usually used during spectroscopic measurements[351,352].

1.3.2.3 Two-colour coincidence detection

In Two-colour coincidence detection, two differently fluorophore-labeled molecules are utilized. Both dyes are simultaneously excited with two different lasers and corresponding events of fluorescence of both dyes, showing interaction between the two molecules, against a background of chance fluorescence events from unassociated molecules is detected on a single-molecule level[170,353,354]. These chance coincidence events need to be subtracted from the measured data. The fluorophores do not have to be placed in FRET-range, which makes this method useful for systems where limiting or insufficient structural data is available and the fluorophores can just be placed anywhere in the molecules. The method is highly sensitive, with minimal cross-talk between channels, and sub-femtomole quantities of labeled sample can be utilized even against a 1000-fold excess of a fluorescent background, allowing the analysis of less pure and more complex sample mixtures [353,355].

Chapter 2

Motivation and objective

The post-transcriptional modification of the canonical nucleoside uridine into its rotational isomer pseudouridine occurs in non-coding as well as coding RNA and is the most abundant post-transcriptional modification in all kingdoms of life[59,63–71]. While the occurrence of pseudouridine has been linked to the enhancement of stability and the codon-anticodon interaction in tRNAs, enhancement of the translation efficiency in rRNAs, regulatory functions in spliceosomal snRNA and nonsense codon suppression in mRNA (Figure 19), its exact role in many RNAs is still ambiguous and opens up a broad area of research[57,69,79,88–90,94].

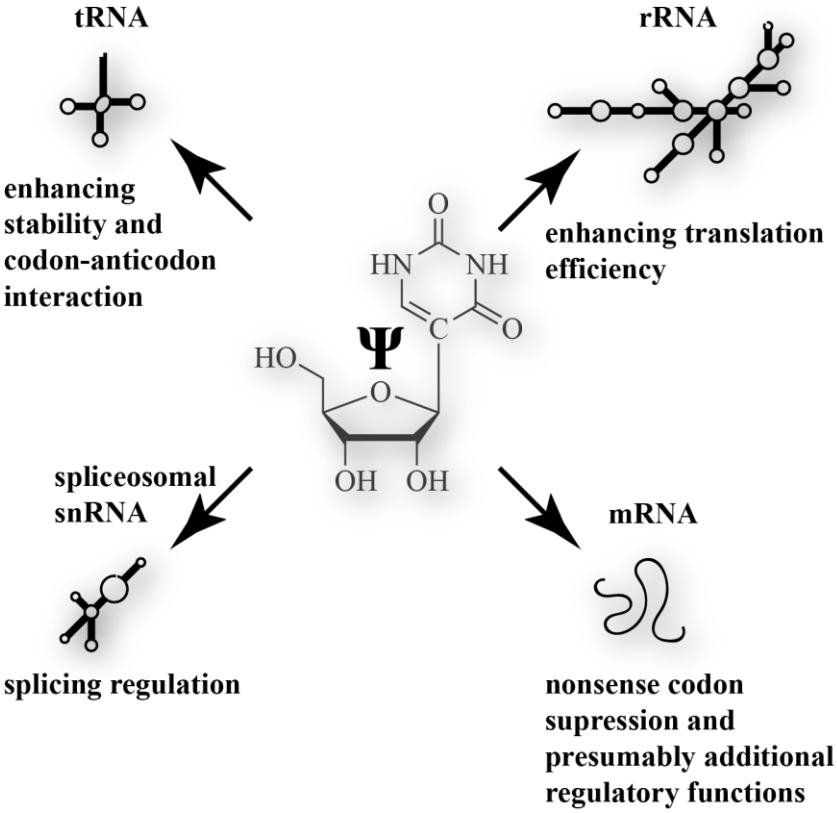


Figure 19: The role of pseudouridine in different types of coding and non-coding RNAs. Post-transcriptional modification of uridine into pseudouridine is linked to different enhancing and regulatory effects[69].

The uridine to pseudouridine isomerization can either be catalyzed by one of various standalone pseudouridylases or it can be performed in an RNA-guided manner by H/ACA ribonucleoproteins[70,122,124,136–140]. All these standalone enzymes as well as the catalytic subunits of H/ACA RNPs (Cbf5/dyskerin) share a common core structure with a conserved aspartate at the center, that is responsible for anchoring the uridine during catalysis[66,103,140]. However, while several mechanisms for the pseudouridylation reaction have been proposed, to this day it was not possible to identify the exact mechanism that occurs during isomerization[66].

Mutations in H/ACA RNP proteins have been linked to the human genetic disease Dyskeratosis congenita, which leads to shortened telomeres and thereby results in bone

marrow failure, premature aging and increased susceptibility to cancer[147,161,170–174]. While this makes investigation of human H/ACA RNPs especially interesting from a medicinal point of view, H/ACA RNPs from *Saccharomyces cerevisiae* are very similar to their human homologues and act as an excellent model system in the investigation of human-related diseases.

While archaeal H/ACA RNPs share many similarities with eukaryotic RNPs and act as good model system, there are also many differences between them. While archaeal H/ACA RNPs form single, double or triple hairpin motifs, eukaryotic RNPs always adapt a conversed double-hairpin conformation[122,124,136,137,143]. The two hairpins perform the pseudouridylation reaction on different target substrates, which can be on the same substrate RNA or on different RNAs altogether. While eukaryotic Cbf5, Nop10 and Gar1 seem to fulfill similar functions than their archaeal counterparts in maintaining catalytic activity, complex stability and substrate release, the role of Nhp2 is ambiguous[122,140,143,150,154–156,159–161]. Its archaeal counterpart L7Ae has been linked to the correct positioning of the substrate RNA in the complex and is strictly required for multiple turnover catalysis[157]. Nhp2 on the other hand, while exhibiting catalysis enhancing abilities, seems to be dispensable for the pseudouridylation reaction, and strong catalytic function has been observed in the absence of Nhp2[156,159,160]. Furthermore, both proteins exhibit completely different binding behaviors towards the RNP complex. L7Ae efficiently binds to the H/ACA RNA in absence of the other proteins via a K-turn motif and does not interact with the other proteins in the absence of H/ACA RNA[126]. Eukaryotic H/ACA RNA on the other hand does not contain a K-turn motif and Nhp2 unspecifically binds to a broad range of RNAs, however it binds with the proteins Cbf5 and Nop10 and forms a trimeric complex prior to RNP assembly, which seems to specify the interaction towards H/ACA RNAs[143,154–156].

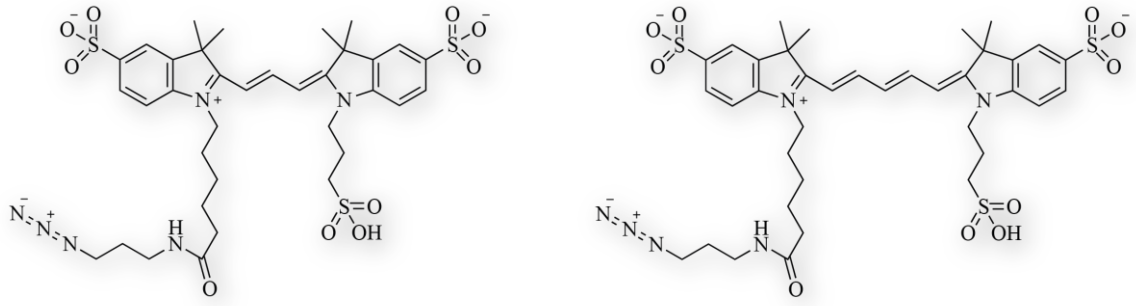
Up until now, most structural knowledge about H/ACA RNPs has been derived from archaeal complexes. This stems from the difficulties of efficient eukaryotic RNP preparation *in vitro*, which is hampered by small protein expression yields as well as proteins prone to degradation and oftentimes *in vivo* pre-assembled complexes have been used for eukaryotic RNP studies[126,127,140]. However, recent studies on two eukaryotic H/ACA RNPs – snR34 and snR5 - were performed with *in vitro* reconstituted complexes[159,160]. One of these studies led to the only available crystal structure of eukaryotic H/ACA proteins up until now: a structure of a trimeric complex of Nop10 and the core domains of Cbf5 and Gar1[160].

The main focus in this thesis lies on the investigation of the snR81 H/ACA snoRNP from *Saccharomyces cerevisiae*, an eukaryotic snoRNP that performs pseudouridylation on U42 of the U2 snRNA and U1051 of the 25S rRNA, which reside in the Cajal bodies/nucleoplasm and the nucleoli, respectively[96,356,357]. The goal is to gain information about the actual formation of the RNP, structural arrangement of the different complex components, interactions between the complex components during assembly and after the formation of an active RNP as well as dynamic information about the catalytic process during pseudouridine formation. The objectives of the thesis can be separated into two different main parts:

For the first part, it is necessary to efficiently reconstitute the snR81 H/ACA RNP *in vitro*. This includes the preparation and purification of the snoRNA as well as the associating proteins. For the RNA, the correct folding to a double hairpin structure is important, while for the proteins, expression and purification should lead to good purity as well as good yields.

While the synthesis of the individual building blocks one by one is possible, co-expression and co-purification strategies can be utilized. The correct assembly of the RNP should be verified by analysis of its catalytic pseudouridylation activity. Furthermore, different constructs of the complex, like standalone versions of both hairpins as well as shortened structures of both hairpins were to be utilized and kinetic analysis of the complex parts should be performed. Also, special focus should lie on two eukaryotic specific proteins/protein features and their influence on the catalytic capabilities of the snR81 H/ACA RNP: the conserved glycine-arginine rich "GAR" domains of protein Gar1, that are not present in archaeal Gar1[158,358], which leads to the assumption that they may play a special role in eukaryotic H/ACA RNPs, and the protein Nhp2, that shows unique characteristics in comparison to its archaeal counterpart L7Ae and has been shown to differently affect several other snoRNPs[159,160].

For the second part, the main goal is the labeling of specific RNP complex components with donor and acceptor dyes (Figure 20) and FRET analysis of the complex on a single-molecule level. Different RNA/protein interactions should be analyzed. Again, eukaryotic specific protein Nhp2 and its role in the complex assembly should be a focus. A large set of experimental work in this part of the thesis includes the efficient labeling of the proteins in sufficient quantity for smFRET analysis. For this purpose, the chemical synthesis of different non-canonical amino acids, their incorporation at suitable positions into the *Saccharomyces cerevisiae* proteins via amber suppression technique and the subsequent fluorophore labeling will play an important and major part and should be utilizing information on protein preparation gained in the first part of the thesis and also knowledge that was gained during master thesis as well as in other projects focusing on the archaeal H/ACA RNP from *Pyrococcus furiosus* in the Hengesbach group[359–361].



sulfo-Cy3-azide sulfo-Cy5-azide

Figure 20: The donor fluorophore dye sulfo-Cy3-azide and the corresponding acceptor fluorophore dye sulfo-Cy5-azide, which were to be utilized in this thesis for labeling of the snR81 RNP proteins.

Chapter 3

Reconstitution of snR81 H/ACA RNP and dissection of its enzymatic kinetics

3.1 Reconstitution of the eukaryotic snR81 snoRNP

The primary goal of the first project was the *in vitro* reconstitution of the snR81 snoRNP complex and the generation of kinetic data of the enzymatic pseudouridylation reaction of the complex. Up until now, it has only been managed to prepare an active eukaryotic snoRNP *in vitro* by two other groups, the snR34 RNP by the Kothe Lab[159] and the snR5 RNP by the Ye Lab[160].

The eukaryotic H/ACA snR81 RNP complex consists of multiple building blocks: the four *Saccharomyces cerevisiae* proteins Nhp2 (“W”), Cbf5 (“C”), Nop10 (“N”) and Gar1 (“G”) as well as the snR81 snoRNA, which contains two pseudouridylation sites. Additionally, for performance of enzymatic activity, the two substrate target RNAs U2 snRNA and 25S rRNA are necessary. After preparation of these individual complex components, an active complex had to be obtained by reconstitution.

3.2 Preparation of snoRNA constructs

The preparation of snR81 snoRNA was carried out via transcription. The snR81 snoRNA itself forms a double hairpin structure, containing an enzymatic pseudouridylation pocket in each hairpin (Figure 21). Furthermore, the RNA carries the conserved H- and ACA-box motifs. For studies of each individual hairpin of the RNP complex, a standalone construct of both hairpins was also be prepared. The 5’ and the 3’ hairpin standalone constructs are labeled as “H5” and “H3” respectively, while the complete snR81 RNA will be referred to as “FL”.

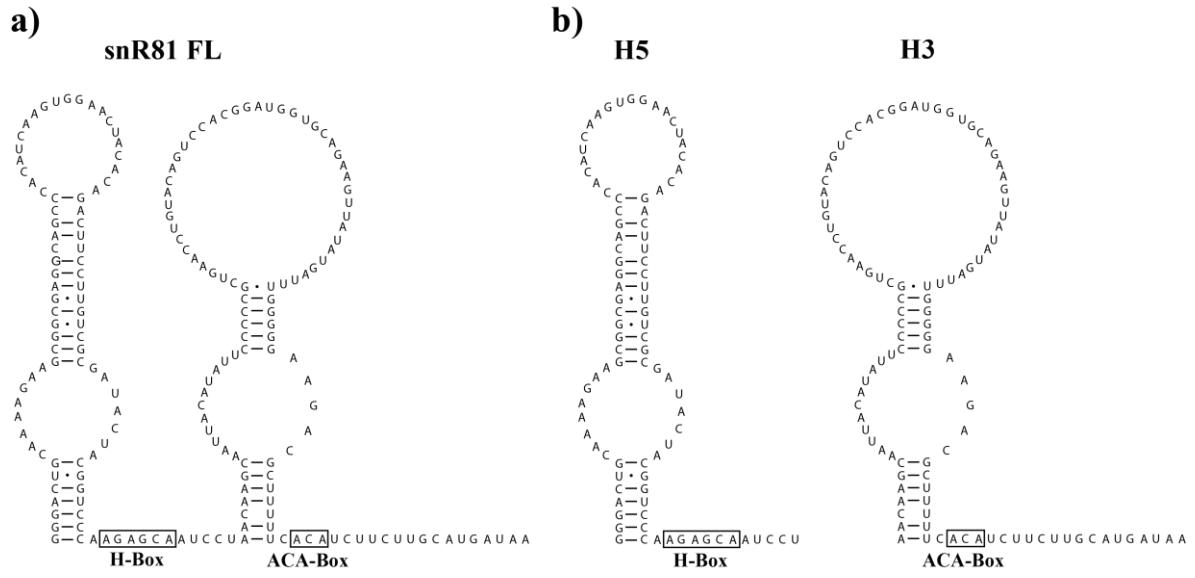


Figure 21: Schematic representation of snoRNA constructs. snR81 full length construct (a) and standalone constructs H5 and H3 (b).

To generate the transcription template, a plasmid containing the corresponding DNA with a T7 starting sequence at the 5' end of the full double-hairpin structure was used. The full DNA template was amplified with corresponding primers, as well as the two standalone constructs. For the 3' hairpin construct, a primer with T7 starting sequence overhang was used. The transcription templates were amplified via PCR and the transcription was carried out afterwards with T7 polymerase. Pyrophosphate was redissolved with EDTA, RNA was precipitated with ammonia acetate and ethanol and purified with preparative denaturing PAGE. A single RNA band for each construct could be identified by UV shadowing and excised from the gel. Gel elution in ammonia acetate was carried out overnight and the RNA pellet was redissolved in water after ethanol precipitation. From a 200 µl transcription reaction, 150.8 µg of H5, 185.0 µg of H3 and 80.4 µg of FL RNAs could be obtained. Analytical denaturing PAGE showed a single band for all 3 constructs, showing that transcription yielded one distinct RNA product for each construct (Figure 22a). RNAs were also denatured at 98 °C and put on ice immediately afterwards to promote the kinetic formation of the hairpin structure ("snap cooling"). Native PAGE revealed one single band for the H5 and FL constructs and one major band and a smaller secondary band above the major band for the H3 construct (Figure 22b).

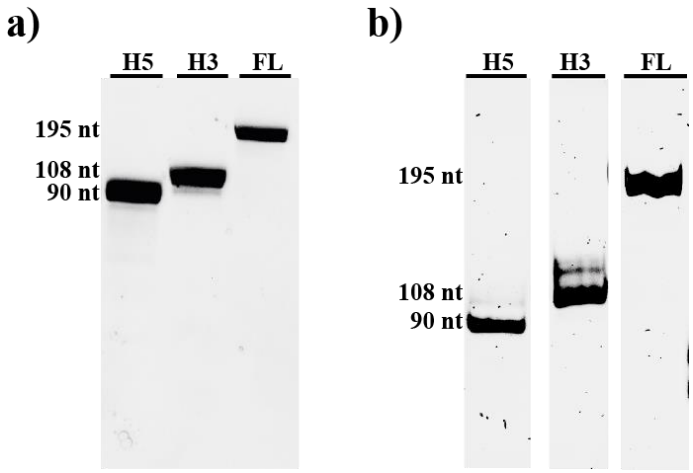


Figure 22: PAGE analysis of RNA constructs. Denaturing 12% PAGE (a) and native 8% PAGE (b).

As an addition to the constructs FL, H5 and H3, two additional constructs were prepared. These constructs H5Δ and H3Δ are both standalone constructs of the 5' and 3' hairpins, however, the upper stem, where the actual hairpin structure is formed, is cut out of both constructs and replaced by a tetraloop (Figure 23). Only part of the double stranded stem and the pseudouridylation pocket are remaining from the original structure. These constructs were prepared in order to analyze the function of the upper loop structure for the enzymatic activity of the complex as well as the influence on Nhp2 binding to the complex.

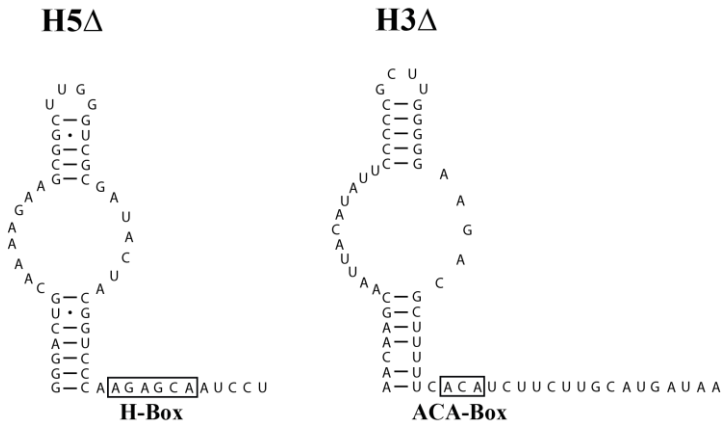


Figure 23: Schematic representation of shortened snoRNA constructs H5Δ and H3Δ.

3.3 Preparation of proteins

3.3.1 Generation of protein constructs

The four proteins WNCG had to be produced by recombinant expression in *E. coli* and subsequently purified by chromatography. As a first step for protein production, the respective genes were cloned into expression vectors: Cbf5 and Nop10 together into pET-Duet, Gar1 into pET-28b and Nhp2 into pET-24b. While Cbf5 contained six histidines (His-Tag) at the C-terminus, Gar1 contained a His-Tag at the N-terminus. For Nhp2, two constructs were prepared, one without a His-Tag and one with a His-Tag at the C-terminus. In addition to the genes of the four standard proteins, shortened versions of Cbf5 (Cbf5Δ) and Gar1 (Gar1Δ) were prepared in the same vectors as their full length counterparts. Also, the plasmid containing Cbf5/Nop10 and the plasmid containing Gar1 were co-transformed into BL21 (DE3) cell line to generate the NCG construct, which was used to co-express and co-purify Nop10, Cbf5 and Gar1 from a single expression culture. In the same manner, the constructs NCΔG, NCGΔ and NCΔGΔ were generated, allowing simultaneous expression and purification of this ternary protein complexes. Figure 24 shows the different constructs in regards of protein length and interaction with other proteins in the reconstituted RNP.

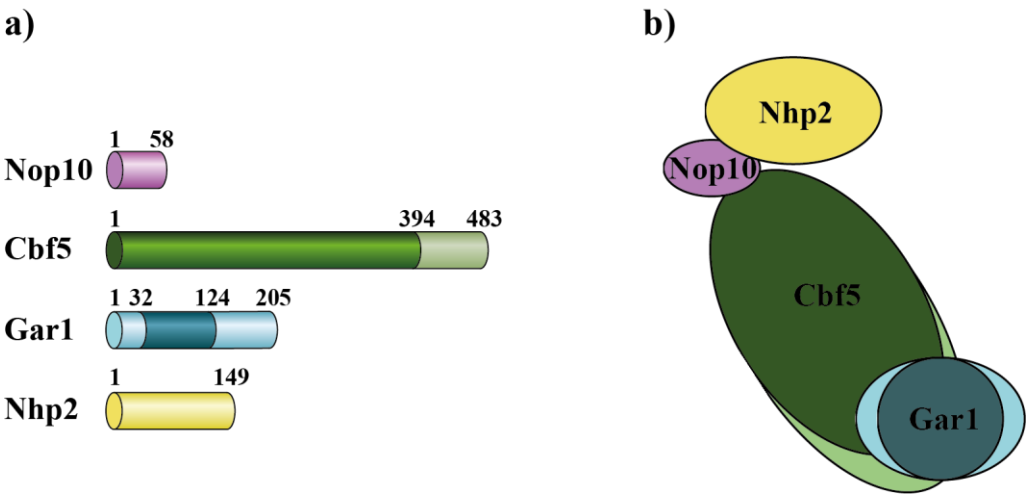


Figure 24: Schematic representation of the four *Saccharomyces cerevisiae* H/ACA proteins. Protein length (a) and arrangement in the quaternary protein complex (b). Core domains (Δ variants of proteins) are marked in darker color.

3.3.2 Initial recombinant expression tests of *S. cerevisiae* proteins in *E. coli*

After successful transformation of the plasmids into BL21 (DE3) *E. coli* cells, expression tests were performed to find suitable expression conditions for each protein and optimize protein yield. These expression tests were carried out at different expression temperatures (37 °C vs. 20 °C), with different media (LB vs. TB medium) and with different final concentrations of IPTG (0 mM vs. 0.1 mM vs. 1 mM). All cultures were incubated overnight and cell lysate of each test culture was analyzed by SDS PAGE (Figure 25 and appendix Figure 56).

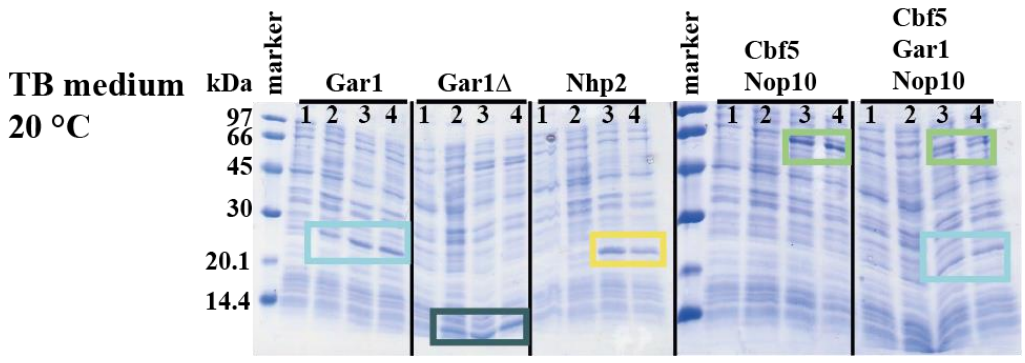


Figure 25: Coomassie stain - expression test of several *s.c.* proteins. Marker: low molecular weight marker (GE Healthcare). Samples: 1) before IPTG induction, 2) o.n. expression 0 mM IPTG, 3) o.n. expression 0.1 mM IPTG, 4) o.n. expression 1 mM IPTG. Expressed proteins are marked with boxes (note that Nop10 is not identifiable in cell lysate samples due to overlapping endogenous *E. coli* proteins and not ideal resolution of cell lysate on the SDS gel). Marker: low molecular weight marker (GE Healthcare).

Expression tests showed a very high basal expression (0 mM IPTG) at 37 °C. To prevent the formation of inclusion bodies and the likelihood of protein misfolding and precipitation,

20 °C was chosen as the more suitable expression temperature for the yeast proteins. LB medium and TB medium showed similar amounts of protein per OD_{600}. However, cells in TB medium showed a much higher OD_{600} in general, favoring TB medium over LB medium. Higher IPTG concentration did not result in higher protein yield, so 0.1 mM IPTG was chosen for further expressions.

3.3.3 Expression and purification of ternary Nop10-Cbf5-Gar1

After optimizing the expression conditions for the proteins, larger scale expressions were performed. NCG was co-expressed in a one-liter volume and cells were harvested after expression at 20 °C overnight. The buffer-resuspended cells were supplemented with DNase as well as RNase A to remove any potential endogenous nucleic acids binding to the proteins. Since resuspension of cells could be performed in a relatively small buffer volume, sonification was chosen as the method for cell lysis. The cell lysate was further supplemented with polyethyleneimine (PEI) for nucleic acid precipitation. Despite the DNase and RNase, there was still a massive precipitation of nucleic acids visible, so the PEI precipitation was done a second time. Purification of the ternary protein complex was done via Ni^{2+} affinity chromatography by binding the target proteins to the Ni-NTA matrix and eluting them with an imidazole gradient. SDS PAGE analysis of the eluted protein fractions showed that a large part of endogenous *E. coli* proteins could be removed by the affinity chromatography but the NCG complex was still contaminated with some endogenous *E. coli* proteins that had bound to the affinity column (Figure 26).

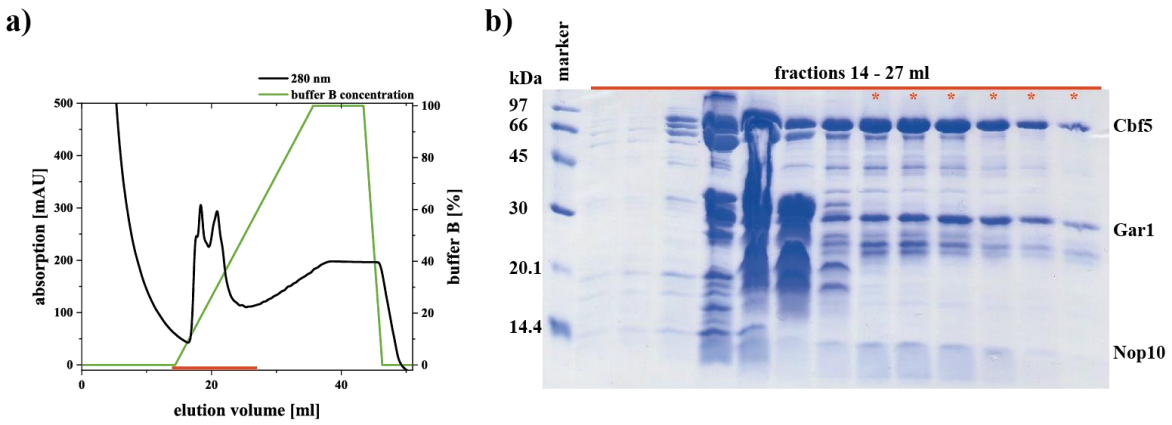


Figure 26: Affinity chromatography purification of ternary NCG complex. a) Ni-NTA elution chromatogram using an imidazole gradient. Fractions used for analysis are marked with a red bar. b) PAGE analysis of fractions from Ni-NTA. Fractions used for further purification are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

As a second chromatographic purification step, the proteins were loaded onto a size exclusion column made of cross-linked agarose and dextran and size exclusion chromatography was performed. It was possible to remove any contaminations. Also, spare Cbf5 and Nop10 not bound to the other proteins were removed and the NCG containing fractions show, that the

proteins were purified together as a ternary protein complex (Figure 27). 280/260 nm ratio in the chromatograms as well as during concentration determination via UV-vis spectroscopy showed, that the protein complex was purified RNA free and could be obtained in good yield of ~5 mg protein from 1 liter expression culture. However, Gar1 was split up into multiple bands on the PAGE gel. This degradation of Gar1 is described in literature[148,160].

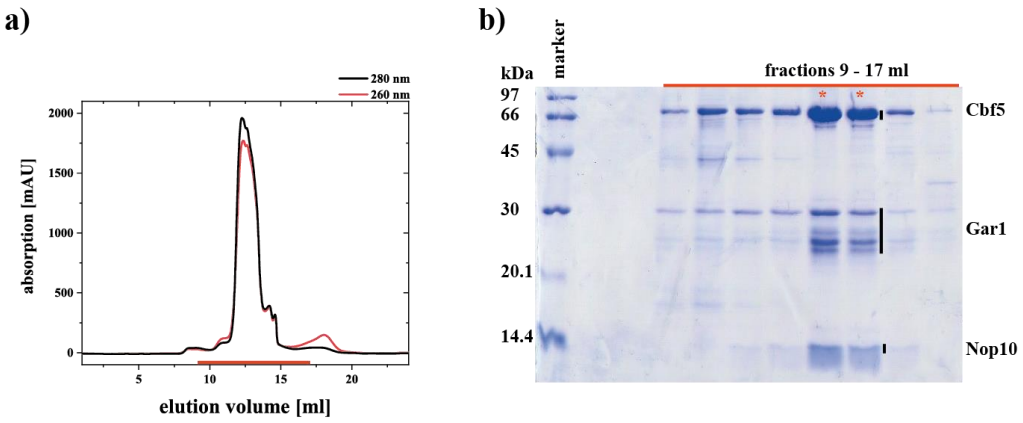


Figure 27: Size exclusion chromatography purification of ternary NCG complex. a) SEC elution chromatogram using a Superdex 200 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) PAGE analysis of fractions from SEC. Fractions used for further experiments are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare). For calibration of column with standard proteins for size comparison with the complex see appendix Figure 57.

This purification procedure was also done for the NCΔG, NCGΔ and NCΔGΔ complexes (Figure 28, Figure 29 and appendix Figure 58, Figure 59). During purification of these protein complex, it became apparent that especially protein complexes with the shortened Gar1 construct GΔ were extremely prone to precipitation upon temperature changes, if the proteins were brought to a high concentration. For further purifications involving these protein constructs, all buffers were pre-cooled and protein was always stored on ice during purification steps where the protein had to be brought to a higher concentration. The shortened constructs could be obtained in yields of 200 µg - 1 mg from one liter expression culture.

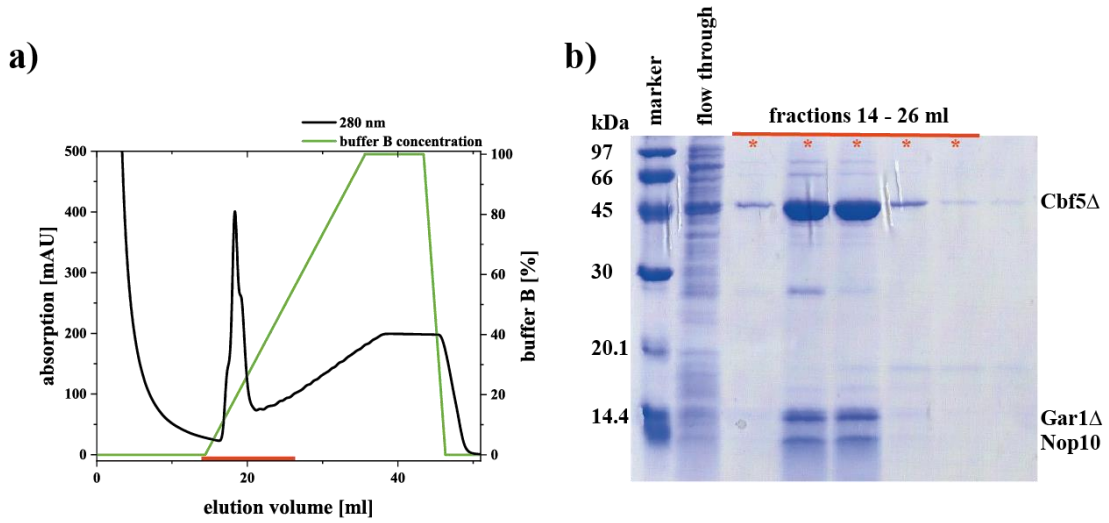


Figure 28. Affinity chromatography purification of ternary NCΔGΔ complex. a) Ni-NTA elution chromatogram using an imidazole gradient. Fractions used for analysis are marked with a red bar. b) PAGE analysis of fractions from Ni-NTA. Fractions used for further purification are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

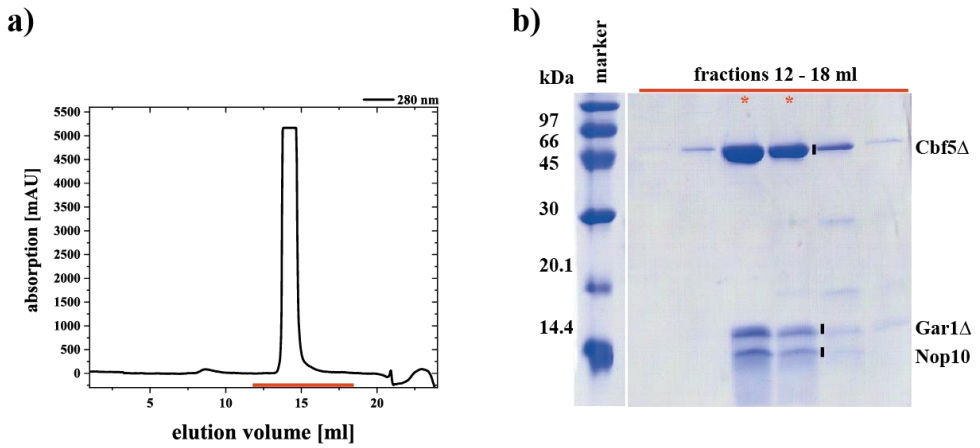


Figure 29: size exclusion chromatography purification of ternary NCΔGΔ complex. a) SEC elution chromatogram using a Superdex 200 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) PAGE analysis of fractions from SEC. Fractions used for further experiments are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare). For calibration of column with standard proteins for size comparison with the complex see appendix Figure 57.

3.3.4 Expression and purification of Nhp2

For Nhp2, the mutation of serine at position 82 to tryptophan was performed via site directed mutagenesis, creating the Nhp2 S82W construct ("W"). This mutation favors a *cis* conformation of the adjacent proline at position 83[362]. Normally, the *cis*/*trans* ratio of P83 is 40:60, but with this mutation it is 60:40. In this way, the "proline spine" which is formed throughout Cbf5, Nop10 and Nhp2 is more stable, which in turn stabilizes the whole complex

and promotes a stronger association of Nhp2 towards the complex[362]. This higher stability of the system was deemed favorable for reconstitution experiments as well as later reconstitution of fluorophore labeled samples and smFRET experiments. Additionally, the *cis* conformation renders Nhp2 structurally more similar to its archaeal counterpart L7Ae, a former protein of interest used for research in the Hengesbach group[359–361]. As a side effect, the addition of the tryptophan residue made the protein more easily identifiable in absorption chromatograms. In initial complex reconstitutions and initial activity tests, the Nhp2 WT construct without the S82W mutation was used. However, an activity test of both the WT and S82W constructs was performed, which showed that the mutated construct did not change the activity behavior (see appendix, Figure 72). For all multiple turnover pseudouridylation assays, the S82W construct was used. Also, for all smFRET experiments, including the generation of fluorophore labeled Nhp2, the S82W construct was used, and is henceforth referred to as Nhp2.

Nhp2 was expressed under the conditions optimized in the expression tests. The same treatment with DNase, RNase and PEI, which was used to remove bound endogenous RNAs from NCG was used for Nhp2. This was especially necessary for Nhp2, since it is known to unspecifically bind a broad range of RNAs[122,126,143,154–156]. Additionally, Nhp2 binding to the Ni^{2+} affinity column seemed to be diminished by higher salt concentrations of 1 M NaCl (see appendix Figure 60, Figure 61), so the NaCl concentration in all buffers was lowered to 500 mM. After affinity and size exclusion chromatography, Nhp2 could be obtained in high yield (Figure 30, Figure 31). However, the 260/280 nm absorption ratio still showed a massive RNA contamination, so additional measures had to be taken. The lower salt concentration in the buffer also benefited the increase in RNA amount bound to the protein.

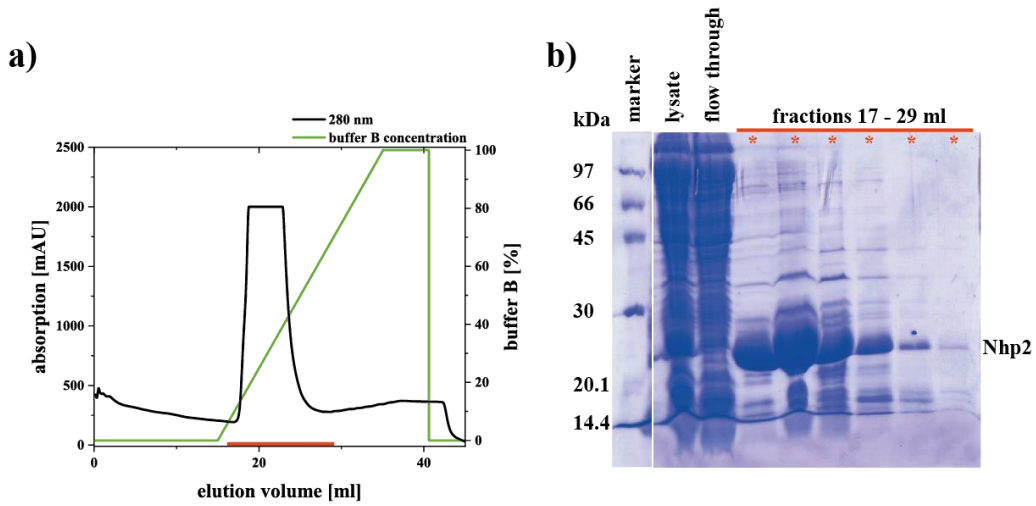


Figure 30: Affinity chromatography purification of Nhp2 before RNase A treatment. a) Ni-NTA elution chromatogram using an imidazole gradient. Fractions used for analysis are marked with a red bar. b) PAGE analysis of fractions from Ni-NTA. Fractions used for further purification are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

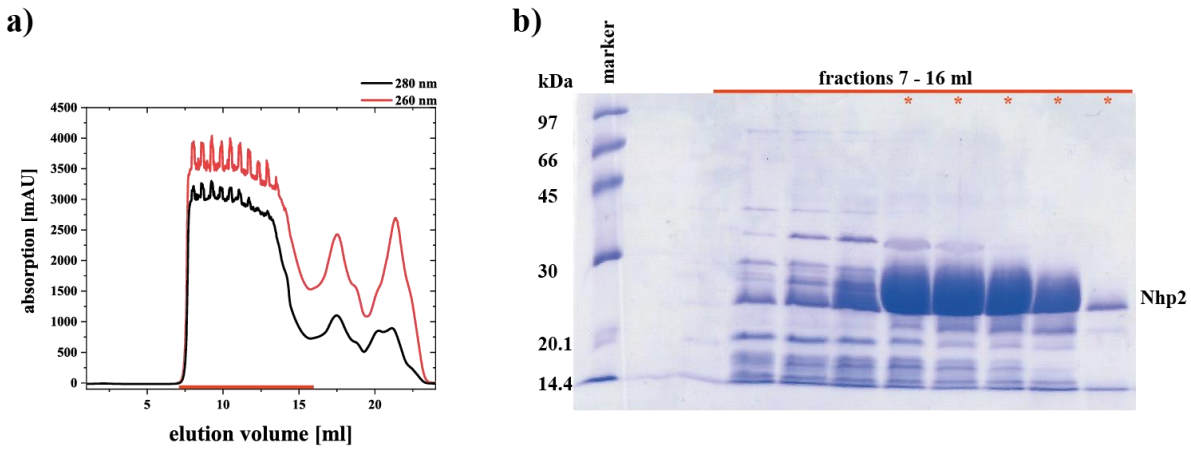


Figure 31: Size exclusion chromatography purification of Nhp2 before RNase A treatment. a) SEC elution chromatogram using a Superdex 75 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) PAGE analysis of fractions from SEC. Fractions used for further experiments are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

To remove the RNA, the pooled fractions of Nhp2 were supplemented with an additional 20 mg of RNase A. The incubation with RNase A was done overnight at room temperature, since Nhp2 seemed to be much more stable at higher temperatures than the NCG proteins. Afterwards, the protein could be loaded onto a Ni^{2+} affinity column again and purified by affinity and size exclusion chromatography to remove any RNase A (Figure 32, Figure 33 and appendix Figure 62, Figure 63). A small fraction of protein/RNA contamination (absorption at 260 nm higher than at 280 nm) was also removed via SEC (see Figure 33, peak at ~7.5 ml). In further preparations of Nhp2, the RNase A treatment was done directly after the first affinity chromatography, the pooled fractions were diluted with imidazole-free buffer to an imidazole concentration of 20 mM and a second affinity chromatography was performed to remove RNase A. Afterwards, the SEC was done to remove any remaining impurities. RNA-free Nhp2 could be obtained in good yield of ~5 mg from one liter expression culture.

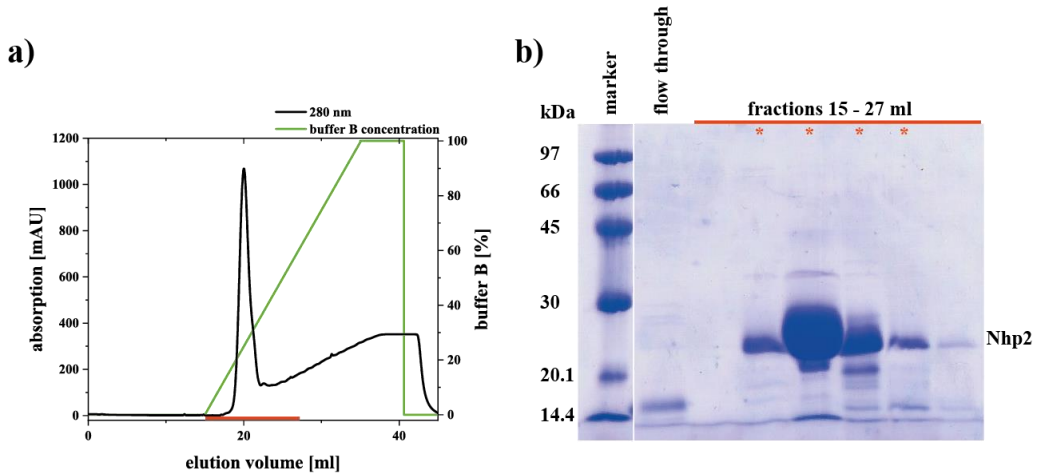


Figure 32: Affinity chromatography purification of Nhp2 after RNase A treatment. a) Ni-NTA elution chromatogram using an imidazole gradient. Fractions used for analysis are marked with a red bar. b) PAGE analysis of fractions from Ni-NTA. Fractions used for further purification are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

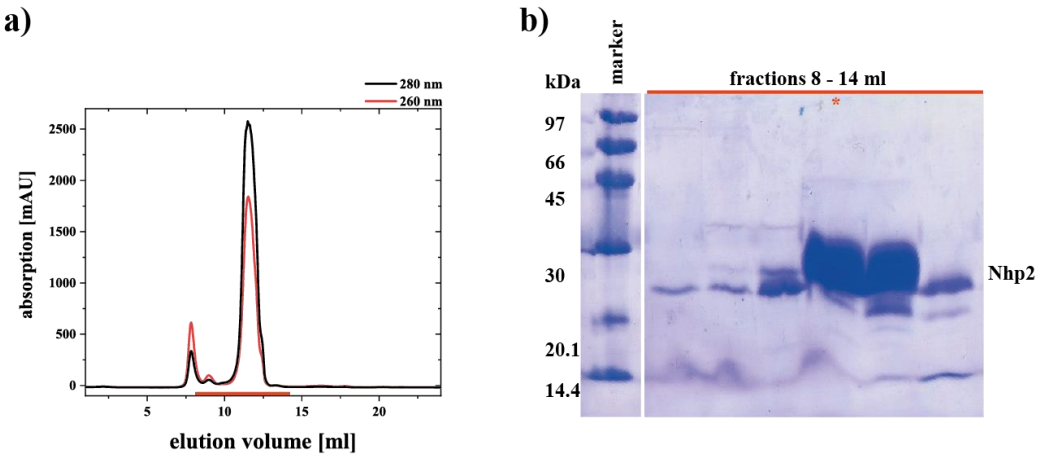


Figure 33: Size exclusion chromatography purification of Nhp2 after RNase A treatment. a) SEC elution chromatogram using a Superdex 75 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) PAGE analysis of fractions from SEC. Fractions used for further experiments are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

3.3.5 Simultaneous purification of NCG and Nhp2

Since co-purification of Cbf5 and Nop10 by protein-protein interaction was successful, the same was tested for Nhp2 (Nhp2 co-purification with Cbf5/Nop10 was previously shown[160]). Since Nhp2 is bound to the RNP via RNA-protein interactions as well as protein-protein interactions – most likely between Nop10 and Nhp2, it was tested if all proteins could be purified simultaneously. For this approach, Nhp2 without a His-Tag was used to verify if during affinity chromatography, the binding of Nhp2 to the rest of the proteins is sufficient for a co-purification. Nhp2 (without His-Tag) as well as NCG were individually expressed as

previously described. Before cell lysis, the protein containing cell pellets of NCG and Nhp2 were mixed together. Afterwards, lysis and further purification steps were carried out as described before. Affinity chromatography and SEC were performed and protein containing fractions were analyzed with SDS PAGE. However, no band with the migration behavior of Nhp2 could be identified in the gel (see appendix Figure 64). Additionally, more endogenous protein impurities were co-purified with the mixing of protein pellets and the overall smaller expression volumes.

3.3.6 Expression and purification of Nop10-Cbf5 and Cbf5

For reconstitution and activity experiments in the absence of Gar1, Cbf5 together with Nop10 were also expressed and purified without the Gar1 plasmid. Expression and purification were done accordingly to the NCG expression and Cbf5/Nop10 could be obtained in good yield (see appendix Figure 65, Figure 66). To test the influence of Nop10 towards enzymatic activity of the reconstituted RNP complex as well as the behavior of reconstituted samples for smFRET spectroscopy without Nop10, standalone expression of Cbf5 and Cbf5Δ should also be performed. Both gene sequences of Cbf5 and Cbf5Δ were gained from restriction digestion of the pET-Duet vector plasmid with NcoI and EcoRI (see appendix Figure 67a). The gene sequences were separated from the vector gene sequence by preparative 1% agarose gel electrophoresis. New pET-Duet vector was prepared by Midiprep and the vector was also cut with the two restriction enzymes and purified. Ligation of the genes of Cbf5 and Cbf5Δ into the vector was performed with T4 DNA ligase. The correct insertion of the gene sequence was tested via colony PCR and sequencing (see appendix Figure 67b). Expression and purification of Cbf5 and Cbf5Δ was carried out in a 500 ml culture volume and the proteins were afterwards purified via affinity chromatography and size exclusion chromatography. Cbf5 could be obtained in good yield of 2.9 mg of protein. For Cbf5Δ however, expressions were not successful. For both Cbf5/Nop10 and Cbf5 endogenous *E. coli* proteins with a similar length than Gar1 were co-expressed and could not be removed by affinity chromatography. However, removal via SEC was successful and only a minute amount of contaminations was co-purified with the *S. cerevisiae* proteins.

3.4 Complex reconstitution

3.4.1 Electrophoretic mobility shift assays

A first step for complex reconstitution was performed in the form of electrophoretic mobility shift assays (EMSAs). In these assays, a defined amount of snoRNA (1-2 µM) was incubated with equivalents of the NCG ternary protein complex (and its shortened Δ versions). Binding of NCG to snoRNA has been shown to have dissociation constants in the low nanomolar range.[159] The snoRNA was ‘snap cooled’ beforehand. RNA was heated to 98 °C and then immediately put into ice water. During the heating, the RNA was denatured and with the rapid cooling, the correct folding should be kinetically promoted, leading to the correct formation of the hairpin or double-hairpin structure of the snoRNA. Varying amounts of proteins were then added to the RNA in SEC buffer and samples were incubated at 30 °C for 30 minutes,

simulating native conditions for yeast proteins. Afterwards, native PAGE was performed with the samples. Only small electric current was utilized and the PAGE was performed at 4 °C to prevent heating up of the samples. As a negative control, RNA only samples were also incubated at 30 °C and added onto the gel. The gels were analyzed by both GelRed (Figure 34, Figure 35) and Coomassie blue (Figure 68, Figure 69) staining.

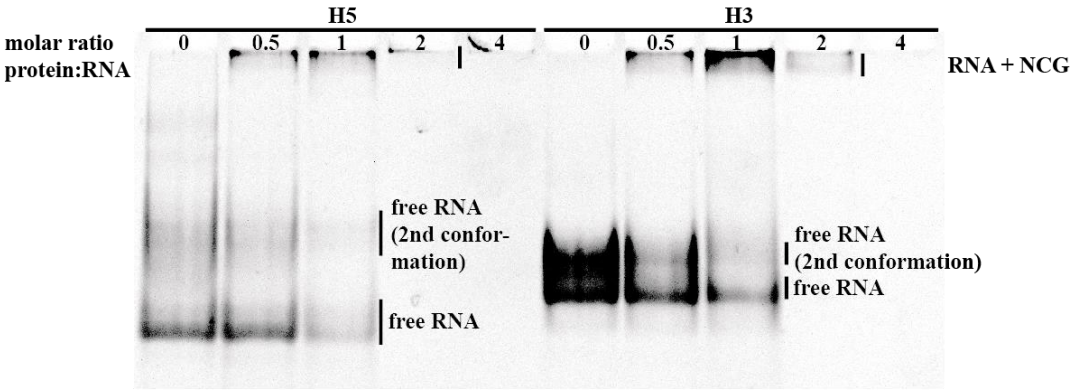


Figure 34: EMSA of snoRNAs H5 and H3 with NCG on a 12% native PAA gel. RNA concentration is 2 µM, protein concentration varies between 0 and 8 µM.

For H5 and H3, one major band and one minor band can be identified on the gel. The minor band most likely represents a minor secondary conformation of the snoRNA. This second conformation was already observed for H3 during the analysis on a native gel after the RNA purification. The incubation at 30 °C seems to increase the amount of this second conformation for H3 as well as for H5. With the addition of NCG proteins, the major as well as the minor band disappear with increasing protein concentration and a new RNA band is formed, which migrates into the gel slower. For some constructs, this new RNA band could not migrate into the gel at all and instead precipitated in the pocket of the gel. Coomassie staining showed that the new RNA band arising from higher protein concentrations indeed contained protein (see appendix Figure 68, Figure 69) and represents a band of formed RNP complex. The inability to migrate into the gel of some constructs – mainly the constructs with full length proteins Cbf5 and Gar1 – could arise from protein and RNA clumping together at too high protein concentrations. A complete binding of NCG towards the RNA constructs H5, H3 and FL could be observed at the addition of two molar equivalents of proteins to the RNA.

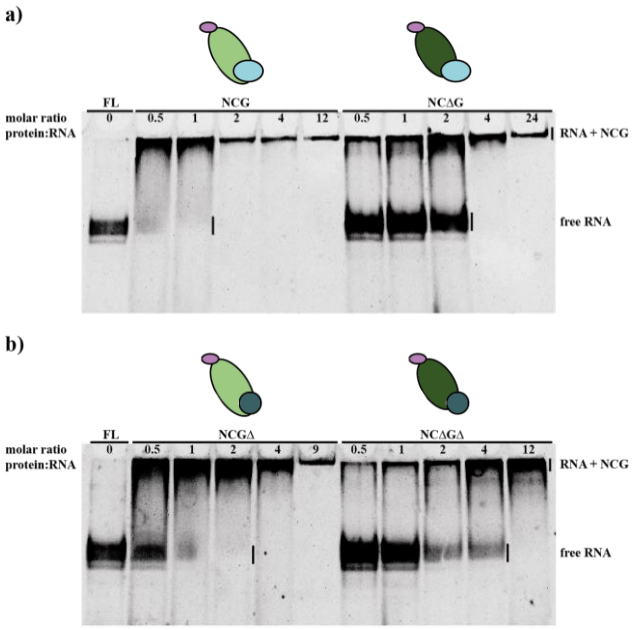


Figure 35: EMSA of snoRNA FL with NCG, NCΔG (a), NCGΔ and NCΔGΔ (b) on a 12% native PAA gel. RNA concentration is 1 µM, protein concentration varies between 0 and 24 µM.

The comparison of the different NCG full length and shortened construct show, that the removal of the *s.c.* specific domains of Cbf5 and Gar1 decrease the binding capacity of the proteins towards the RNA. While for the full length NCG, 2 µM of protein are sufficient to completely bind 1 µM of RNA, for both NCΔG as well as NCGΔ, full binding is achieved at 4 µM of protein for 1 µM of RNA and for the double shortened construct NCΔGΔ even a greater amount of protein is necessary.

3.4.2 Size exclusion purification after complex assembly

Complex assembly was done with an excess of protein in comparison to snoRNA in general. This approach was sufficient for activity assays as well as smFRET experiments. For complex activity, it was verified in negative tests, that proteins without snoRNA did not show any detectable enzymatic activity, so a protein excess was deemed to be unproblematic. In smFRET experiments, only samples containing the snoRNA with the biotin-linker were able to bind via biotin-streptavidin interactions to the glass surface of the microscopy slides, so a protein excess was also not hindering these experiments. However for analytical approaches like cryo-EM, an excess of protein in the sample would not be feasible. This approach preconditioned a reconstituted complex with equimolar amounts of snoRNA and proteins with proteins only present in the sample that are bound to the snoRNA hairpins. As a way to remove unbound components from the complex, a way to utilize size exclusion chromatography was established. Complex reconstitution was done with 1 µM snoRNA and 2 µM protein per hairpin. The reconstitution was done in a 50 µl reaction volume. Afterwards, the sample was loaded onto a size exclusion column with 2.4 ml bed volume, allowing the separation of such a small sample. To determine the migration behavior of RNA and proteins alone on the column, samples of only the RNA constructs H5, H3 and FL were loaded onto

the column and absorption at 260 nm was detected. The RNAs could be identified and the migration behavior of all three constructs was as expected, with FL migrating faster than H3 and H3 migrating faster than H5 (Figure 36a). Afterwards, each of the RNA constructs was reconstituted with NCG and size exclusion chromatography of the sample was performed. The addition of the NCG trimer resulted in a shift of the RNA peaks in the chromatogram in comparison to standalone RNA. The addition of WNCG for full complex reconstitution resulted in a bigger shift compared to standalone RNA and RNA+NCG. For both H5 and H3, this shift resulted in earlier elution of the RNA from the column, presumably because of bound protein resulting in reconstituted complex (Figure 36b, c). For the FL construct, also a shift could be observed with the addition of proteins, however this shift was not as clear as for the standalone hairpin constructs (Figure 36d). Unbound proteins were eluted at a later time at the same elution volume as standalone proteins. This led to a removal of unbound protein, making this method a promising technique for generation of reconstituted RNP without excess protein.

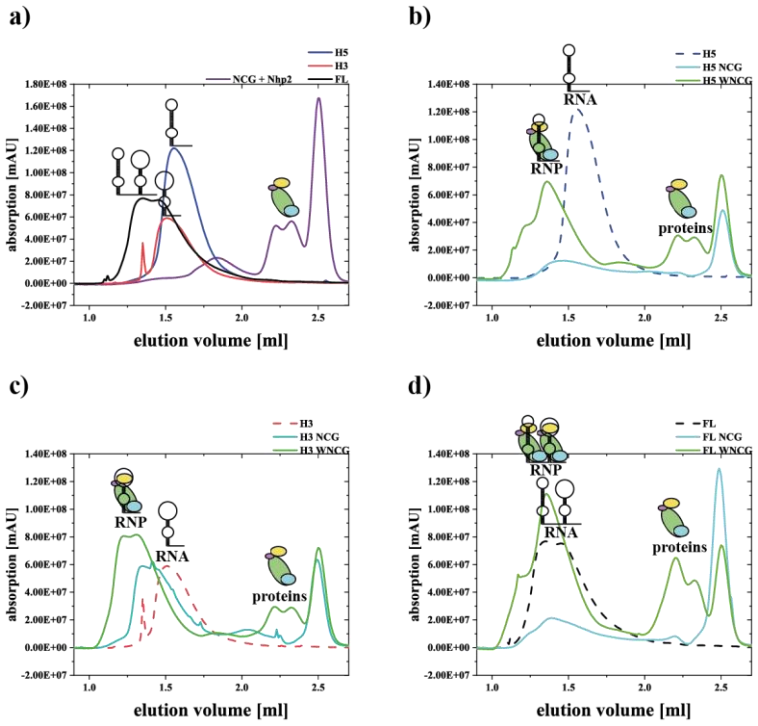


Figure 36: Size exclusion chromatography of snoRNAs and RNPs using a Superdex 200 3.2/300 GL increase column. Elution profiles of snoRNAs H5, H3, FL and proteins NCG + Nhp2 (a), H5, H5 NCG and H5 WNCG (b), H3, H3 NCG and H3 WNCG (c), FL, FL NCG and FL WNCG (d). e) Thin layer chromatography of multiple turnover pseudouridylation activity tests performed with site specific ^{32}P-labeled 5' substrate RNA (1 µM) for 2 h. H5-RNP eluted after size exclusion chromatography shows the same activity as the positive control.

Additionally, RNP with Cy3-labeled Nhp2 was reconstituted and purified via SEC (see appendix, Figure 70a). The PAGE analysis showed an RNP band on the gel via UV as well as fluorescence scan, confirming the presence of RNA and Nhp2, as well as Cbf5 and Nop10, which are necessary for Nhp2 binding to the complex (see appendix, Figure 70c,d). The

eluted RNP fractions were also tested in a multiple turnover pseudouridylation activity test and showed the expected enzymatic activity, which also confirms the presence of a fully reconstituted RNP (including Gar1) (see appendix, Figure 70b).

3.4.3 Influence of *Saccharomyces cerevisiae* specific domains of Gar1 towards complex reconstitution

The protein Gar1 contains several glycine-arginine rich (GAR) domains at its N- and C-terminus, which are flanking the core domain of Gar1 (Gar1Δ) (Figure 37). Those GAR domains consist of an arginine-glycine-glycine (RGG) motif, which is repeated several times in these domains. Both domains make up ~40% of the full length Gar1 protein. The GAR domains are believed to have an accessory role in RNA binding[358,363–368]. However, they seem to be dispensable *in vivo*[160,358]. Moreover, in the fully assembled complex, the main interaction of Gar1 is believed to be with Cbf5 rather than the RNA itself; Gar1 is believed to directly interact with the thumb loop domain of Cbf5 and is responsible for substrate release[140,157,158,161]. However, the core domain Gar1Δ also contains an RNA binding motif that was described to be more efficient in RNA binding than the GAR domains[358]. Still, the GAR domains seem to affect catalytic behavior of the complex, since activity assays performed with the snR5 snoRNA revealed Gar1Δ is sufficient for catalytic activity, but GAR domains still had an effect on the pseudouridylation, since activity was overall decreased for both hairpins of the RNP in absence of GAR domains[160]. Interestingly the core domain Gar1Δ is structurally much more similar to archaeal Gar1, which does not possess any GAR domains. Moreover so, the crystal structure available of eukaryotic Gar1 was obtained with a construct lacking the GAR domains[160].

Figure 37: Schematic representation of the different domains in protein Gar1. RGG motifs inside the N- and C-terminal domains are marked in royal blue.

To test the binding capabilities of Gar1 and specifically the GAR domains of Gar1, several constructs were prepared. Besides the full length Gar1 and the core domain Gar1Δ, two other constructs with only the GAR domains present at the N-terminus (Gar1N) and at the C-terminus (Gar1C) were designed. For Gar1N, a stop codon was inserted behind the core domain of Gar1 via site directed mutagenesis. For Gar1C, the N-terminal Gar1 domain was removed with primer based whole plasmid PCR and the plasmid was religated afterwards. Both constructs were expressed from TB medium cultures and expression conditions were carried out by standard protocol. Purification of protein constructs was carried out by affinity and size exclusion chromatography and all constructs could be obtained in good yield (Figure 38). Since Gar1 on its own shows a very weak absorption, concentrations of the protein could

not be obtained via UV-vis spectroscopy. Instead, the protein concentrations of the different Gar1 constructs were determined via Bradford assay.

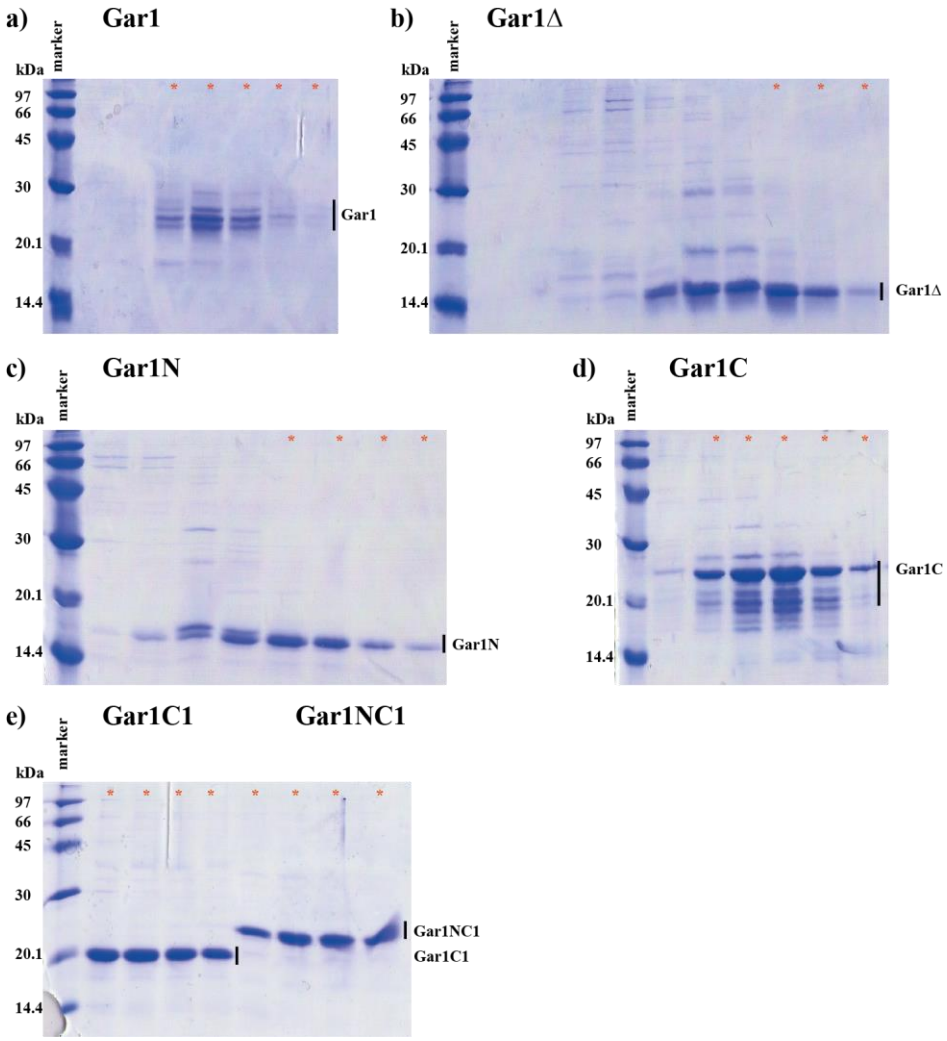


Figure 38: SDS PAGE analysis of fractions from SEC (Superdex 200 10/300 GL increase column) of the different Gar1 constructs. Fractions used for further experiments are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

For the purified proteins Gar1N and Gar1C, as well as for Gar1 and Gar1Δ, EMSAs with different protein to RNA molar ratios from 0.5:1 to 10:1 were performed. The EMSAs were performed with H5 and H3 snoRNA to determine whether the presence or absence of the GAR domains had an impact on the binding capabilities of the proteins towards the snoRNA. The snoRNA constructs were snap cooled, proteins were added in Ψ buffer and afterwards samples were incubated at 30 °C for 30 minutes. As a negative control, an RNA-only sample was also incubated at 30 °C and loaded onto the EMSA gel. As with other EMSAs before, the incubation at 30 °C led to a minor population of snoRNA besides the main population, most likely due to formation of a second conformation during the incubation. If protein was bound to the RNA, the RNA did not migrate into the gel anymore, indicating a saturation of RNA-protein complex. For H5 and Gar1, this point was reached at an equimolar concentration of

1 µM for both RNA and protein (Figure 39a). For both Gar1C and Gar1N, H5 was completely bound at a molar concentration of protein to RNA of 2 µM : 1 µM (Figure 40a). Gar1Δ showed the weakest binding to H5 with complete complexation at a concentration of 5 µM : 1 µM (Figure 39a). This shows indeed that the GAR domains facilitate RNA-binding of the protein, with the weakest protein-RNA interaction with the absence of both GAR domains. For the H3 RNA construct, the complete association of Gar1 towards the snoRNA was reached at a concentration of protein to RNA of 2 µM : 1 µM (Figure 39b). For Gar1C and Gar1N, the bands disappeared at a molar ratio of 5:1 (Figure 40b). For Gar1Δ, complete association was also reached at a concentration of 5 µM : 1 µM (Figure 39b). This shows the same trend that was observed for H5 snoRNA. However, H3 seems to be even more dependent on the GAR domains. On the other hand, H3 shows a weaker binding to the different protein constructs in general.

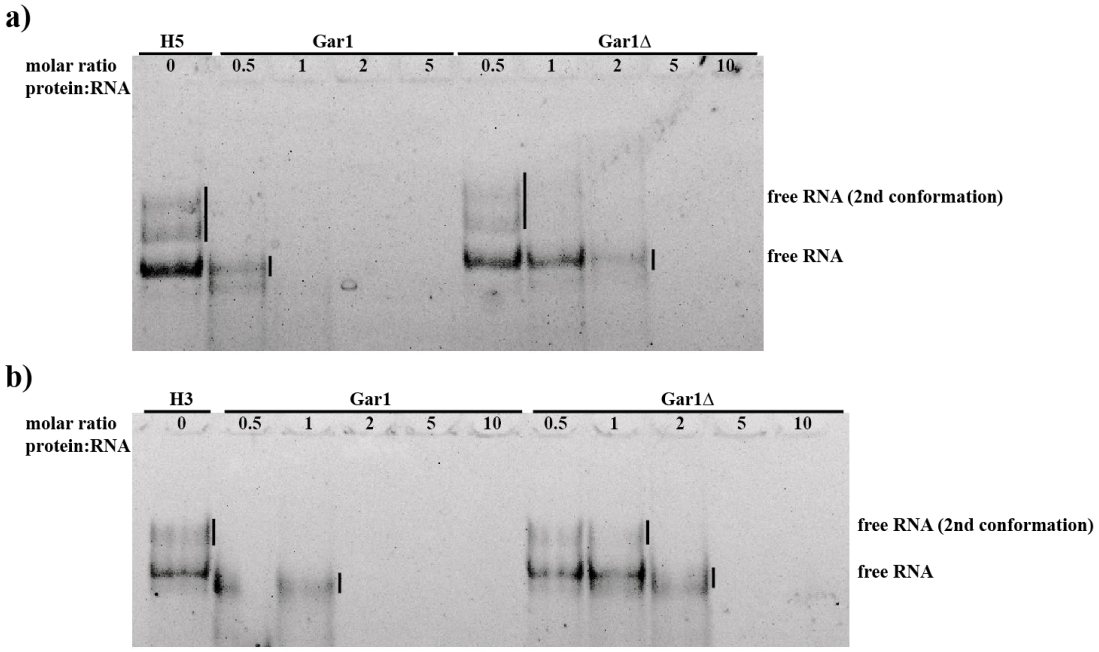


Figure 39: EMSA of snoRNAs H5 (a) and H3 (b) with Gar1 and Gar1Δ on a 12% native PAA gel. RNA concentration is 1 µM, protein concentration varies between 0 and 10 µM.

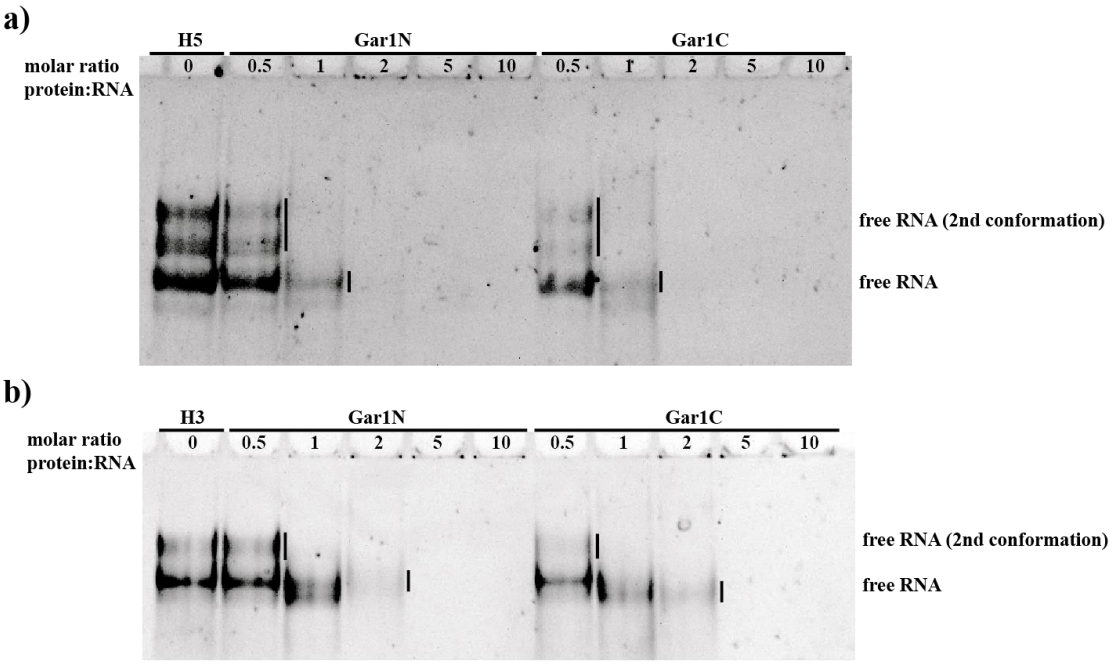


Figure 40: EMSA of snoRNAs H5 (a) and H3 (b) with Gar1N and Gar1C on a 12% native PAA gel. RNA concentration is 1 µM, protein concentration varies between 0 and 10 µM.

Interestingly, the protein degradation that takes place during the purification of Gar1 full length protein, can also be observed for the Gar1C construct, but not for the Gar1Δ and the Gar1N constructs. This highly suggests that the degradation that is visible by SDS PAGE gel analysis takes place in the C-terminal domain of the protein. To further analyze the binding capabilities of Gar1 to the snoRNAs as well as the role of the C-terminal GAR domain, two additional constructs were generated. The C-terminal domain contains 14 RGG motifs in total and can be split up into two individual GAR domains, with 7 RGG motifs in the first half and 6 RGG motifs in the second half. Two additional Gar1-constructs were generated: Gar1C1, which consist of the core domain and the first half of the C-terminal GAR domain and Gar1NC1, which additionally consist of the N-terminal GAR domain. Both these constructs did not show any degradation during protein purification (Figure 38e). This again highly suggests, that the degradation of Gar1 takes place in the second half of the C-terminal GAR domain. EMSA experiments with both constructs were carried out with H5 as well as H3 (Figure 41).

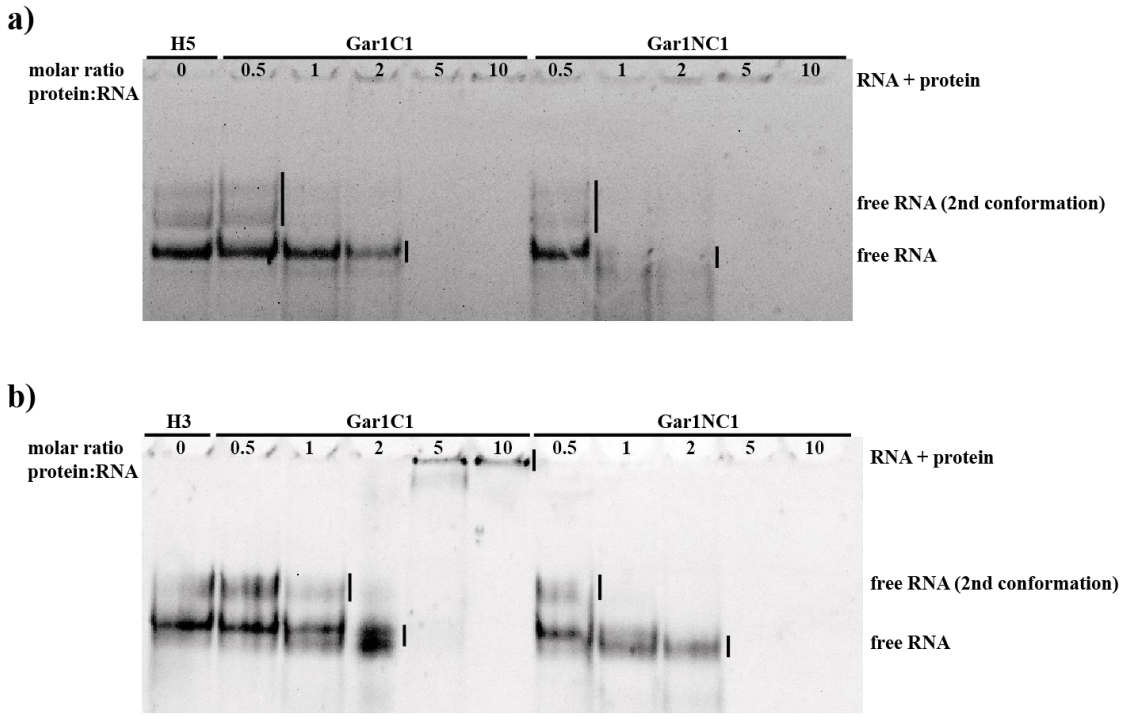


Figure 41: EMSA of snoRNAs H5 (a) and H3 (b) with Gar1C1 and Gar1NC1 on a 12% native PAA gel. RNA concentration is 1 µM, protein concentration varies between 0 and 10 µM.

For Gar1C1, 5 µM protein were needed for complete binding of protein to 1 µM H5 snoRNA, while for Gar1C, 2 µM protein were sufficient. Also, for Gar1NC1, 2 µM protein were needed, while for Gar1, 1 µM protein was sufficient. For H3, constructs Gar1NC1 completely bound the RNA at 5 µM protein, while for Gar1, 2 µM protein were sufficient. For both Gar1C1 and Gar1C 2 molar equivalents of protein were necessary for complete binding of the RNA. This shows, that for both snoRNA constructs, the second half of the C-terminal GAR domain has an impact of RNA binding capabilities to the protein. The binding capabilities towards the H5 hairpin again seem to be higher than for the H3 hairpin construct. Table 1 summarizes the concentrations of protein needed for complete binding to snoRNA.

Table 1: molar ratio of different Gar1 constructs needed to fully bind 1 µM snoRNAs H5 and H3.

	H5	**H3**
Gar1	1 µM	2 µM
Gar1Δ	5 µM	5 µM
Gar1N	2 µM	5 µM
Gar1C	2 µM	5 µM
Gar1C1	5 µM	5 µM
Gar1NC1	2 µM	5 µM

Additionally to the already expressed and purified constructs of the Gar1 protein, it was also tried to express the C and the N terminal GAR domains on their own, not including the core domain of Gar1. For the C-terminal domain, expression and purification was possible, however, the C-terminal domain on its own also showed degradation behavior on the SDS PAGE gel. It was not possible to express a standalone construct of the N-terminal domain of Gar1. It is most likely, that this domain is just too small to be expressed by normal expression protocol and that the use of a fusion protein is necessary to generate a sufficient amount from protein expression.

The degradation taking place in Gar1 seemed to not affect the enzymatic activity of the protein complex. However, for methods like cryo-EM, degradation products would interfere with the structure and a complex preparation without these side products would be favorable. With the knowledge, that protein degradation takes place in the second half of the C-terminal GAR domain, Gar1 gene was cloned into a pET 24b vector with a C-terminal His-Tag. This would ensure, that with the degradation of the C-terminal domain, the His-Tag would get destroyed and therefore the degradation products would not carry an affinity tag anymore and would not be co-purified with the intact protein. Expression of Gar1 from pET 24b vector confirmed this, as in the SDS PAGE no degradation products are visible anymore in comparison to expression from pET 28b vector with an N-terminal His-Tag (Figure 42, Figure 43). However, over the course of a day between affinity and size exclusion chromatography, new degradation bands are forming, so the degradation of Gar1 does not only take place during cell lysis but seems to be an ongoing process. This problem could be to some extent be reduced by consecutively performing all purification steps without any delay or by utilizing one of the shortened Gar1 constructs, that did not show degradation.

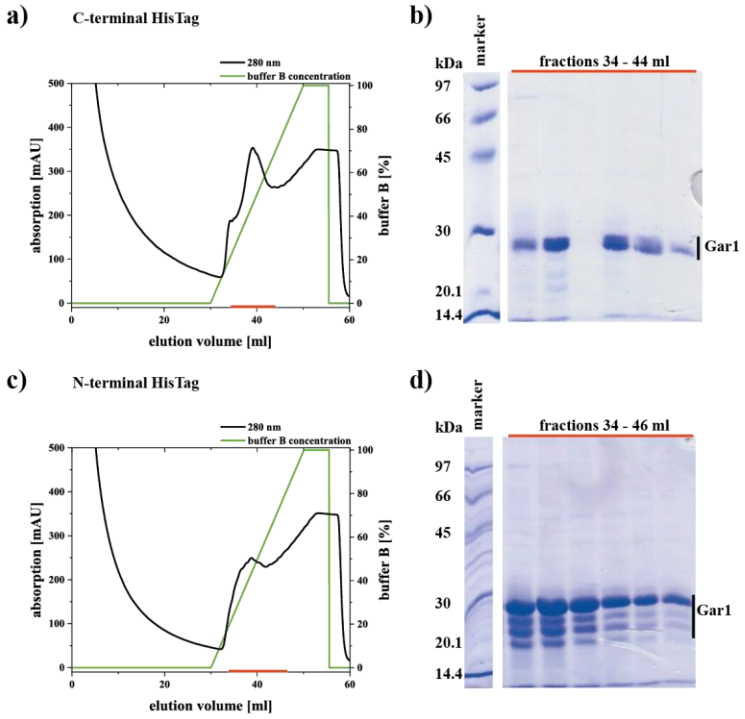


Figure 42: Affinity chromatography purification of Gar1 from pET 28b (a+b) and pET 24b (c+d). a+c) Ni-NTA elution chromatogram using an imidazole gradient. Fractions used for

analysis are marked with a red bar. b+d) PAGE analysis of fractions from Ni-NTA. Marker: low molecular weight marker (GE Healthcare).

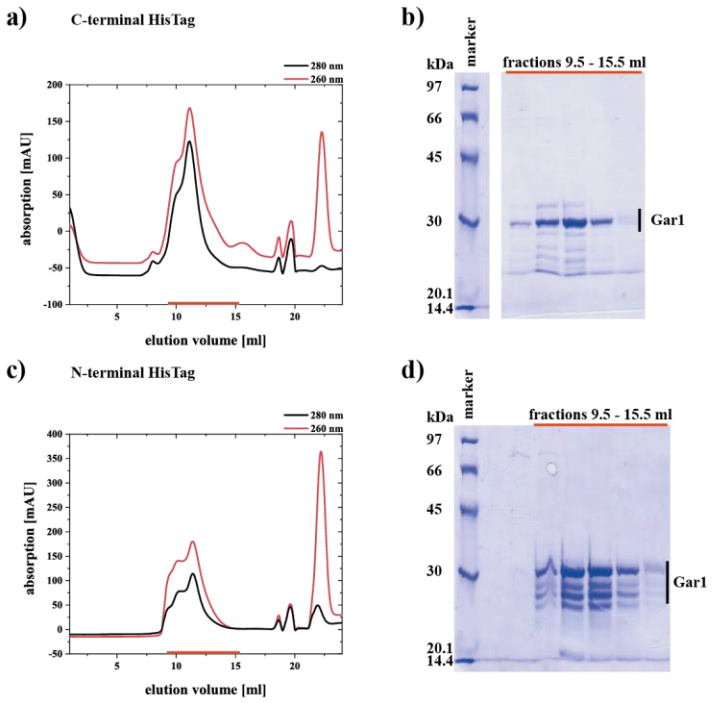


Figure 43: Size exclusion chromatography purification of Gar1 from pET 28b (a+b) and pET 24b (c+d). a+c) SEC elution chromatogram using a Superdex 200 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b+d) PAGE analysis of fractions from SEC. Marker: low molecular weight marker (GE Healthcare).

3.5 Dissection of catalytic activity of snR81 RNP

3.5.1 Preparation of substrate RNA via transcription

In vivo, the actual target substrates of the snR81 RNP are U42 of the U2 snRNA and U1051 of the 25S rRNA for the 5' and the 3' hairpins respectively[96,356,357]. For testing of the enzymatic activity of the complex, target RNAs corresponding to the actual target RNA sequences were used, however only the part of the substrate RNA binding inside the pseudouridylation pocket with a short overhang on both sides were used. Substrate RNA for initial tests was generated via transcription with [α-^{32}P] UTP. By this method, all uridines (including the target uridine) are radioactively labeled and uridine to pseudouridine formation can be quantified via autoradiography. However, to first test and optimize the transcription reaction conditions in a radioactive free environment, transcription tests were carried out with unlabeled UTP. DNA template for transcription was generated by PCR reaction with a template and corresponding primers. The amplified DNA template showed two bands on an analytic agarose gel for the 5' substrate and a single band for the 3' substrate. Test transcriptions were performed, however for both constructs two band were visible on a

denaturing analytic PAGE, pointing towards two RNA products of different length for both constructs rather than two different folding conformations. First, several improving steps were undertaken, to optimize PCR conditions towards a single band template. Different annealing temperatures were used as well as shortened versions of primers. A "hot start-like PCR" was performed, in which the Phusion polymerase was added after heating up the samples to 95 °C. Also, addition of DMSO to the PCR reaction was tried. However, it was not possible to generate a single DNA template product. As a second way of template generation, the template and its base-pairing counterpart were annealed in high concentrations and purified by filter column. This approach was more promising, generating a single transcription template without side products (see appendix Figure 71a). Test transcriptions showed that it was possible to mainly generate the desired main product. However, bands with lower molecular weight were still visible on the analytical PAGE gel (see appendix Figure 71b). It was assumed that these smaller side products would not interfere with the pseudouridylation reaction. Since the uniformly labeled substrate RNA from the transcription approach was supposed to be used as an initial test only, to see if the reconstituted RNP performs catalytic activity at all, it was decided to use the transcription conditions from the test transcriptions for generation of the radioactively labeled substrate RNA. Transcription was performed with [α-^{32}P] UTP instead of UTP. This generated substrate RNAs in which all uridines carried the radioactive label. Incorporation rate of [α-^{32}P] UTP was determined by measuring the radioactive counts per minute of the transcription reaction mixture before and after filtration and with the number of uridines in the construct (10 for 5' substrate, 6 for 3' substrate) the concentration of samples could be determined as 8.12 µM for 5' substrate (incorporation rate 0.081) and 2.82 µM for 3' substrate (incorporation rate 0.017).

3.5.2 Initial pseudouridylation activity tests

RNP complexes of H5, H3 and FL were reconstituted in Ψ buffer. A concentration of 1 µM of snoRNA and 2 equivalents of protein per hairpin (2 µM for H5 and H3, 4 µM for FL) were used. The snoRNA was snap cooled before adding buffer and proteins. For single turnover conditions, 0.5 µM of uniformly [α-^{32}P] UTP substrate RNA was added to their respective constructs. For FL, tests with both 5' and 3' substrate were performed. Single turnover experiments were performed at 30 °C for 1-2 h. For verification of catalytic activity and pseudouridine formation, the proteins were removed from the samples by phenol extraction and the RNA was ethanol precipitated. Afterwards, the RNA was resuspended in P1 buffer and a digestion with P1 endonuclease was performed for several hours. A small amount of sample was then spotted onto a thin layer chromatography cellulose plate and thin layer chromatography was performed with an iso-propanol based buffer to separate the nucleotides from each other. The radioactively labeled nucleotides could then be visualized with a storage phosphor screen. For all constructs, 2 spots could be identified on the TLC plate. Those spots could be assigned to known literature values[369] for 1D-TLC uridine and pseudouridine (Figure 44b). With 10 and 6 uridines in the 5' and 3' substrates respectively and the assumption of only one specific uridine being transformed into pseudouridine, the yield of the pseudouridylation reaction could be determined. Since the radioactive samples smeared a bit on the TLC plate, and for some samples the identification of two separate spots was difficult,

additional 2D-TLC was performed. First, the samples were separated on the TLC plate with an iso-butyric acid-based buffer, afterwards the TLC plates were dried and rotated by 90° and a second TLC with the iso-propanol based buffer was performed. This more clearly separated the U and Ψ spots from one another, which allowed a better assignment to literature values[369] (Figure 44b). Overall, most of the constructs showed catalytic activity. The activity for H5 was higher than for H3, and also in FL, pseudouridine formation in 5' substrate was higher than in 3' substrate (Figure 44a, lane 1 and 3). Also H5 was not able to catalyze the pseudouridine formation on the 3' substrate and H3 not on the 5' substrate (Figure 44c). Furthermore, it was tested (with 1D TLC), how pseudouridylation reaction would perform in the FL construct, if both hairpins would be loaded with substrate RNA simultaneously. For this purpose, equimolar amounts of [α-^{32}P] UTP labeled 5' substrate and unlabeled 3' substrate were added as well as [α-^{32}P] UTP labeled 3' substrate and unlabeled 5' substrate. It was revealed that the pseudouridylation formation on either of the radioactively labeled substrates were neither inhibited nor promoted by the presence of unlabeled substrate RNA in the other pseudouridylation pocket (Figure 44a, lane 2 and 4).

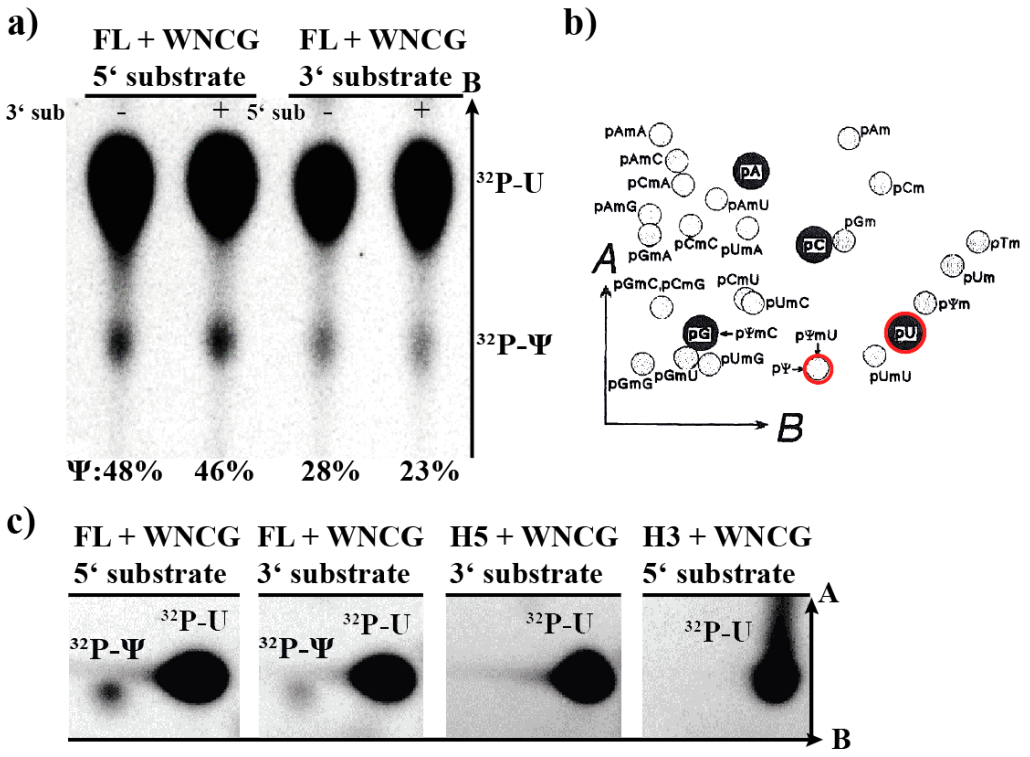


Figure 44: Thin layer chromatography of single turnover pseudouridylation activity tests performed with uniformly ^{32}P-labeled 5' and 3' substrate RNA and FL snoRNA for 2 h. a) Simultaneous pseudouridylation reaction on both fully reconstituted haripins does not interfere with pseudouridylation yield. b) Literature values for uridine (U) and pseudouridine (Ψ) using TLC buffers A and B (see material and methods). U and Ψ are marked with red circles. (figure taken from Keith *et al.*[369] with permission from Elsevier, © 1995 Elsevier.) c) 2D TLC of pseudouridylation reaction of 5' and 3' substrates with fully reconstituted FL snoRNP, 3' substrate with H5 and 5' substrate with H3. H5 and H3 do not pseudouridylate the target RNAs of the other hairpin, respectively.

3.5.3 Generation of radioactively labeled substrate by splinted ligation

After the initial tests for pseudouridylation activity of the RNP complex were successful, more precise experiments were designed to obtain more precise values for the yield determination of pseudouridine formed as well as a more detailed insight into the time course of the pseudouridine formation. For this purpose, it was necessary to generate site-specifically ^{32}P-labeled substrate RNA via splinted ligation. For both substrate RNAs, the same nucleotide sequence as for the substrates generated via transcription was used. Two oligonucleotides were utilized per substrate RNA, with the target uridine at the 5' end of the 3' oligonucleotide (Figure 45a). In a first step, this target uridine on both constructs was radioactively labeled. This was performed via phosphorylation with [γ-^{32}P] ATP and T4 polynucleotide kinase. After the phosphorylation with the radioactive ATP, additional non-radioactive ATP was added to perform phosphorylation quantitively. Unincorporated ATP was removed with a G25 size exclusion spin column. Afterwards, the ^{32}P-labeled 3' oligonucleotide and its corresponding 5' oligonucleotide were first annealed with a DNA splint and slow cooling from 75 °C to room temperature and eventually to 4 °C. The ligation of the two RNA oligonucleotides was performed with T4 ligase overnight.

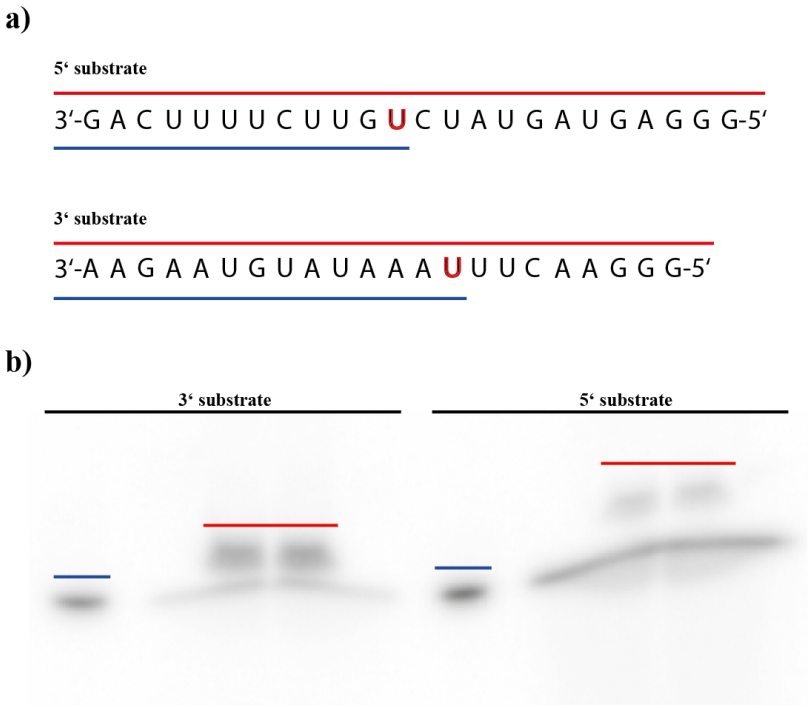


Figure 45: a) Substrate RNA constructs used for pseudouridylation activity tests. The ^{32}P-labeled target uridine is marked in red. The 3' oligonucleotides used for [γ-^{32}P] ATP phosphorylation are marked with a blue line. The full length substrates are marked with a red line. b) Autoradiography scan of preparative 15% denaturing PAGE of splint ligated constructs (note: the non-radioactively labeled 3' oligonucleotide is not visible in this scan).

The ligated product was separated from unligated oligonucleotides with denaturing PAGE. The radioactive samples on the gel could be visualized with a storage phosphor screen and

could be excised from the gel (Figure 45b). A second scan of the gel was done to verify if the excising of the gel was done correctly. The RNA was eluted from the gel with 0.5 M ammonia acetate or 0.3 M sodium acetate and ethanol precipitated. Concentrations for both 5' and 3' substrate in the 10 – 50 µM range (resuspended in 50 µl H_2O) could be achieved, which is 10-50% of the theoretical yield of 50 µM (in 100 µl reaction volume), which was more than sufficient for activity tests.

3.5.4 Activity tests with single turnover conditions

To get more insight into the pseudouridylation reaction on the different snoRNA constructs, pseudouridylation activity tests with the ligated substrate RNAs were performed. RNP complexes were reconstituted with conditions previously found favorable during the initial activity tests with transcripted substrate RNA. 1 µM snoRNA and 2 µM of each protein per hairpin of the construct were used for RNP formation. To the reconstituted complexes, 0.5 µM guide RNA was added and the pseudouridylation reaction was carried out for 2 h at 30 °C. Like before, proteins were removed with phenol extraction, RNA ethanol precipitated, digested with P1 endonuclease, NTPs separated via thin layer chromatography and visualized with a storage phosphor screen (Figure 46). This time, the yield of the pseudouridylation reaction could easily be determined by the uridine/pseudouridine intensity ratio, since only the target uridine was radioactively labeled. Under standard conditions, both constructs H5 and FL showed nearly quantitative pseudouridine formation of the 5' substrate with 93% and 91% yield respectively. Also the 3' substrate was pseudouridylated by FL however the yield was lower than for the 5' constructs with 70% of uridine transformed to pseudouridine for FL and 60% for H3. Omission of Nhp2 lead to reduced activity for both hairpins (see appendix, Figure 72) Additionally, the pseudouridylation reaction was tested with several subcomplex compositions. In negative control experiments, it was revealed that without the snoRNA or without NCG, the complex was not able to perform catalytic activity at all. With the omission of Nhp2 however, pseudouridine formation could still be observed, however with reduced yield. Single turnover assays were also carried out with shortened protein variants of Cbf5 and Gar1 for comparison of the influence of the eukaryotic specific domains of the proteins on catalytic activity. Namely the constructs NCG, NCΔG, NCGΔ and NCΔGΔ together with Nhp2 with both H5 and H3 snoRNAs were tested for comparison. For H5, the amount of formed pseudouridine was relatively similar between full length and shortened protein constructs, resulting in nearly quantitative pseudouridylation of around 90% for all 4 protein constructs. The yield for the NCG construct was still slightly higher than for the shortened constructs. For H3 the NCG construct also showed the highest activity of around 61%. Also, the difference between the full length and the shortened constructs was much higher, with pseudouridine yield being around 15-30% smaller for the shortened proteins.

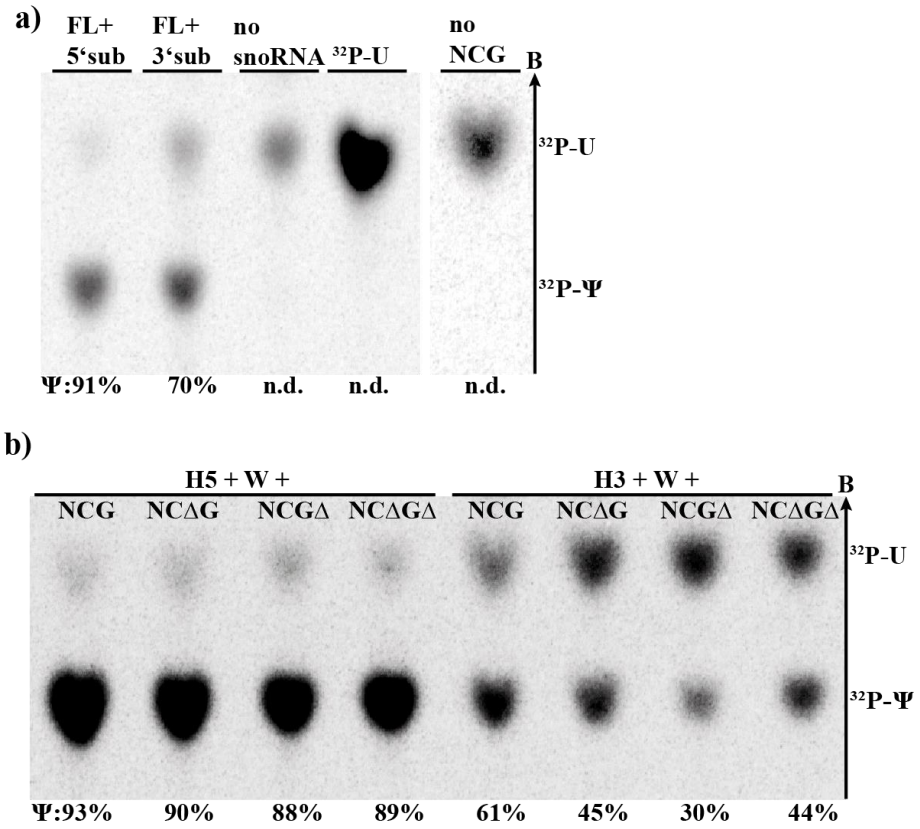


Figure 46: Thin layer chromatography of single turnover pseudouridylation activity tests performed with site specific ^{32}P-labeled 5' and 3' substrate RNA for 2 h. a) Pseudouridylation with FL snoRNA as well as negative controls. b) Pseudouridylation with H5, H3 and full length as well as shortened NCG constructs.

3.5.5 Activity tests with multiple turnover conditions

To gain a more detailed insight into the reaction kinetics of the pseudourdylation reaction, assays under multiple turnover conditions were performed. The key observations from single turnover assays (different catalytic behavior of 5' and 3' hairpin, different catalytic behavior between full length and standalone hairpin constructs) were investigated in more detail during the multiple turnover assays. The 5' hairpin showed a higher enzymatic activity than the 3' hairpin. Also differences between the full length snR81 construct and the standalone constructs H5 and H3 were observed. Multiple turnover assays were performed with 100 nM snoRNA and a tenfold excess of proteins per hairpin. In order to increase the substrate concentration, unlabeled (non-radioactive) substrate (3.9 µM) was used in presence of 100 nM radioactively labeled substrate ("Spiking"), resulting in a 40-fold substrate excess overall. In general, the multiple turnover assays were performed in a 125 µl reaction volume with the reconstitution of the RNP in Ψ-buffer and the time course of the reaction was started with the addition of the substrate RNA. Samples of 20 µl were taken after 5 min, 15 min, 1 h, 2 h, 4 h and 22 h. The samples taken were immediately mixed with pre-cooled aqua-phenol to precipitate the proteins and inhibit the enzymatic activity of the complex. The RNA was ethanol precipitated, digested with P1 endonuclease, spotted on a cellulose plate, U and Ψ separated via TLC and the amount of U and Ψ determined via autoradiography (see appendix

Table 6). The data was fitted with a Michaelis-Menden like equation[370] $\Psi = \frac{\Psi_{max} \cdot t}{t_{\Psi_{1/2}} + t}$. Ψ_{max} describes the maximum amount of pseudouridine that is produced by the snoRNP at the end of the reaction. In theory, this amount should always be 100% after an indefinite amount of time. However, due to protein degradation and general decomposition of the complex at a longer time at 30 °C, the reaction curve reaches a plateau at a defined Ψ_{max}. In the first 4 hours, the most activity behavior could be observed, so the sample taken after 22 h was used as the reaction endpoint. $t_{\Psi_{1/2}}$ marks the point in time when half of the maximum substrate turnover has been reached. The complex reconstitution was first performed with the different snoRNA constructs FL, H5 and H3 as well as a full set of proteins WNCG (Figure 47). For both FL as well as H5 with the 5' substrate RNA, very fast pseudouridine formation could be observed with 80% pseudouridine yield in the first hour and the plateauing of pseudouridine at around 90% during the first four hours (Figure 47 blue and cyan curve). The difference of 5' substrate turnover between H5 and FL was minute. Pseudoudidylation of 3' substrate was significantly slower compared to the 5' substrate. 3' substrate turnover of FL reached 45% pseudouridine yield during the first four hours and plateaued at around 55% after 22 hours (Figure 47 red curve). H3 showed an even lower activity than FL with the 3' substrate, with lower than 10% substrate turnover during the first four hours and around 25% after 22 h (Figure 47 orange curve).

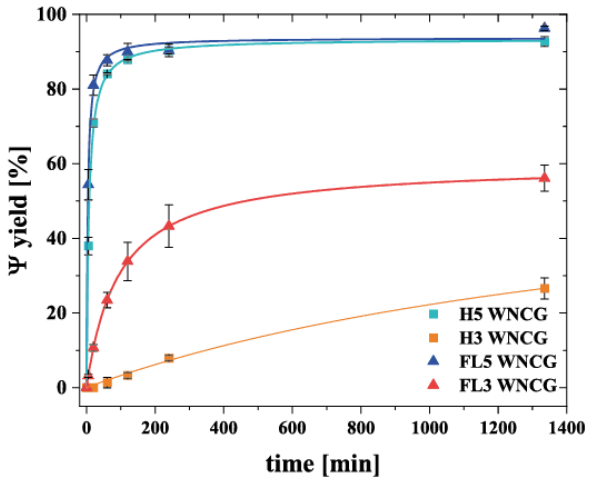


Figure 47: time course of pseudouridylation reaction under multiple turnover conditions with H5 (cyan), H3 (orange) and FL with 5' substrate (blue) and 3' substrate(red). Errors of each data point were generated by standard deviation of two individual measurements.

To further investigate the importance of different building blocks of the RNP towards the kinetic of the reaction, RNP complexes with the absence of several proteins were reconstituted. An activity assay performed by Gerd Hanspach (Goethe University) showed that in the absence of Gar1 (WNC) no pseudouridine formation could be observed[371]. Also the absence of Nop10 (WCG) inhibited the complex, resulting in no detectable catalytic

activity (see appendix, Figure 73a). However, in the absence of Nhp2 (NCG), the RNP was still able to perform catalytic activity. For 5' substrate turnover, the NCG construct was still able to maintain activity for both FL as well as H5. However, the reaction kinetics were slowed down significantly. In the first four hours, FL reached around 65% substrate turnover (Figure 48a, black curve) and H5 reached a smaller amount of only 45% (Figure 48b, black curve). FL started to reach a plateau of around 80% uridine turned to pseudouridine at 22 h (Figure 48a, black curve), while H5 only reached an amount of around 70% (Figure 48b, black curve).

In addition to the absence of Nhp2, the influence of the eukaryotic specific GAR domains within the protein Gar1 towards reaction kinetic were also investigated by assembling a complex with WNCGΔ. Pseudouridine formation was only slightly hindered by the absence of the GAR domains, resulting in a slower start of the reaction, but still an amount of 80-85% pseudouridine was reached during the first four hours for FL (Figure 48a, green curve) as well as H5 (Figure 48b, green curve) with the 5' substrate and pseudouridine yield plateaued at 85-90% for both constructs, almost plateauing at the same amount as with full length Gar1.

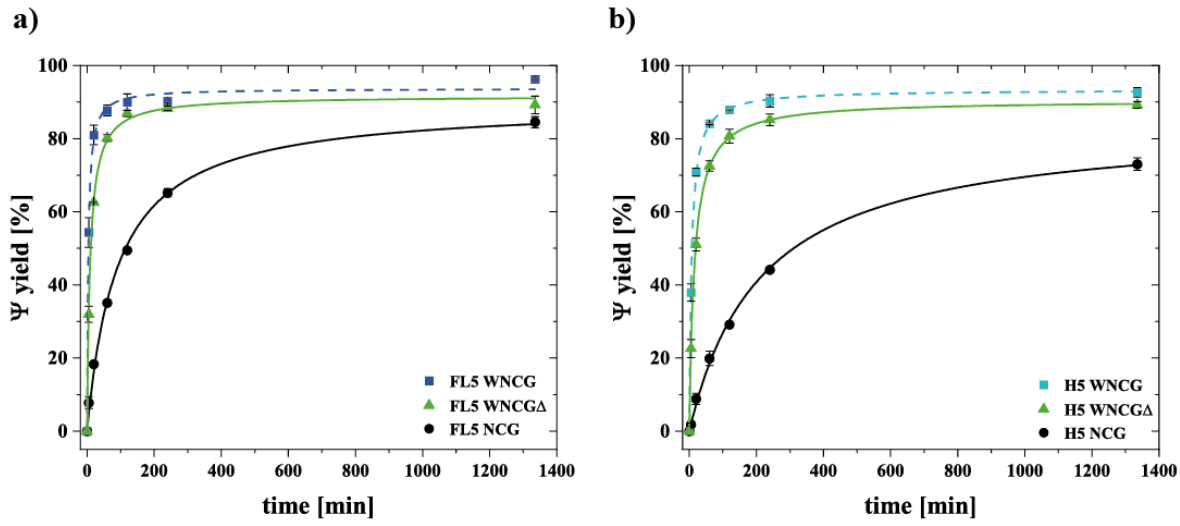


Figure 48: Time course of pseudouridylation reaction under multiple turnover conditions with 5' substrate RNA and snR81 FL snoRNA (a) and H5 (b). Protein constructs WNCG (blue/cyan), WNCGΔ (green) and NCG (black) were utilized. Errors of each data point were generated by standard deviation of two individual measurements.

For 3' substrate turnover, the same aspects of Nhp2 omission and influence of GAR domains was tested with FL and H3. The absence of Nhp2 had a major impact on the kinetic of the pseudouridylation reaction. For FL, during the first four hours of the reaction, no pseudouridine formation could be observed and only a small amount of <10% turnover could be identified after 22 h (Figure 49a, black curve). For H3, the impact of Nhp2 omission for the enzymatic reaction was even stronger, with no pseudouridine observable at all after 22 h (Figure 49b, black curve). The omission of the GAR domains of Gar1 on the other hand resulted in a faster reaction kinetic on the 3' substrate for both FL as well as H3 under multiple turnover conditions. For FL, in the first four hours of the reaction, 55%

pseudouridine yield could be observed (compared to 45% for full length Gar1) and after 22 h the amount of pseudouridine plateaued at 70% (compared to 55%) (Figure 49a, green curve). For H3, the reaction kinetic was slower compared to FL, but the omission of GAR domains also had a positive effect on the course of the reaction with 20% pseudouridine after four hours (compared to 10% for full length Gar1) and an amount of 40% (compared to 25%) after 22 h (Figure 49b, green curve).

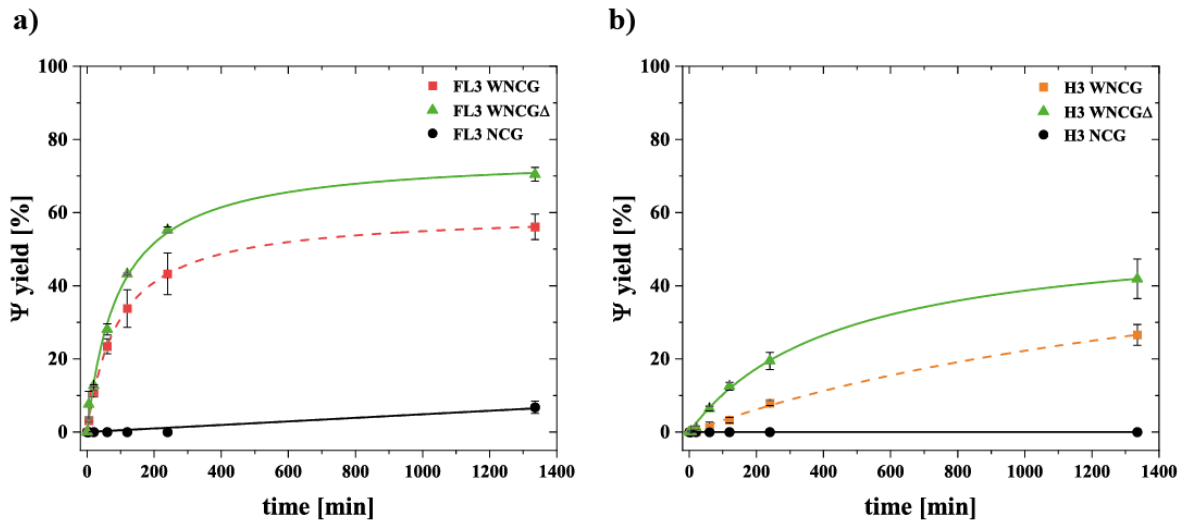


Figure 49: Time course of pseudouridylation reaction under multiple turnover conditions with 3' substrate RNA and snR81 FL snoRNA (a) and H3 (b). Protein constructs WNCG (red/orange), WNCGΔ (green) and NCG (black) were utilized. Errors of each data point were generated by standard deviation of two individual measurements.

Multiple turnover assays were also performed for the shortened snoRNA constructs H5Δ and H3Δ. The assays were performed with the protein constructs WNCG, NCG and WNCGΔ to analyze how the upper stem loop structure of the two snoRNA hairpins affects catalytic activity. For the full RNP, almost no difference in activity was visible. Both H5 as well as H5Δ in addition with WNCG reached around 90% pseudouridine yield in the first 4 hours of the reaction (Figure 50, cyan curves). The omission of Nhp2 lead to a decrease in activity, however this decrease for H5Δ was not as strong as for H5. The shortened RNA construct still reached a pseudouridine yield of 70% after the first four hours, while the H5 construct only pseudouridylated around 45% of the substrate (Figure 50, black curves). On the contrary, the omission of GAR domains with the WNCGΔ construct had a much stronger effect on the activity of H5Δ. While the activity loss for H5 with GAR domain omission was only around 5% in comparison to the full RNP, for H5Δ the amount of pseudouridine after the first four hours was 70% (compared to 90% for H5Δ WNCG) (Figure 50, green curves).

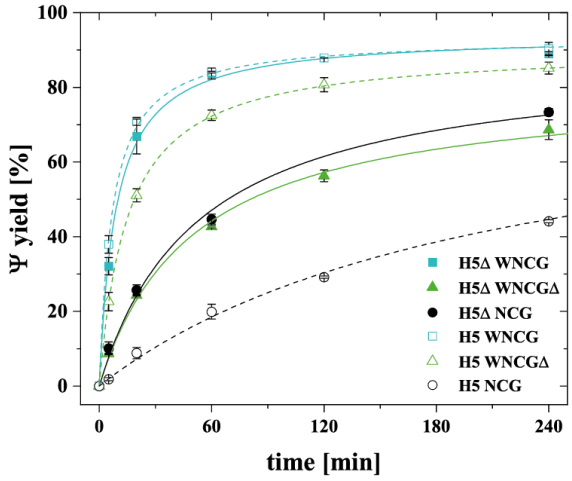


Figure 50: Time course of pseudouridylation reaction under multiple turnover conditions with H5Δ (solid lines) in comparison to H5 (dashed lines). Protein constructs WNCG (cyan), WNCGΔ (green) and NCG (black) were utilized. Errors of each data point were generated by standard deviation of two individual measurements.

For H3Δ, the same pseudouridylation assays were performed. However, for none of the used protein constructs, any pseudouridine could be observed (see appendix Figure 73b). Even after 22 h, only uridine could be identified in the sample. H3Δ seems to be completely inactive.

For all multiple turnover assays of the various protein and RNA constructs, the initial reaction velocity (starting turnover rate) was determined by linear regression (Table 2). In the first few minutes of the pseudouridylation reaction, the reaction kinetic can be seen as nearly linear. The fitted pseudouridine yield at the time point 5 min was determined, and the starting turnover rate of the enzymatic reaction v_{start} was determined by $v_{start} = \frac{\Psi_{fit(5\ min)}}{t_{fit\ (5\ min)}} \cdot 40$. Since the multiple turnover reactions were performed with 100 nM snoRNA and 4 µM substrate RNA (100 nM labeled, 3.9 µM unlabeled), the factor 40 describes the excess of substrate RNA versus snoRNA. Additionally, Table 3 lists the substrate half-life time for all snoRNP constructs, determined by the Michaelis-Menden fit of each reaction kinetic.

Table 2: Starting turnover rates v_{start} for all snoRNP constructs

v_{start} [min^{-1}]	**FL5**	**H5**	**H5Δ**	**FL3**	**H3**	**H3Δ**
WNCG	4.21	3.04	2.67	0.24	0.01	n.a.
WNCGΔ	2.53	1.76	0.58	0.31	0.05	n.a.
NCG	0.37	0.15	0.61	n.a.	n.a.	n.a.

Table 3: Substrate half-life time $t_{\Psi 1/2}$ for all snoRNP constructs. Errors result from two individual measurements and fitting of the data sets.

$t_{\Psi 1/2}$ [min]	**FL5**	**H5**	**H5Δ**	**FL3**	**H3**	**H3Δ**
WNCG	3.5±0.3	6.9±0.3	8.8±0.8	93.3±0.3	1883±426	n.a.
WNCGΔ	9.1±0.4	15±0.2	50±3	94±9	445±32	n.a.
NCG	91±7	214±12	53±7	n.a.	n.a.	n.a.

3.6 Discussion of chapter 3

3.6.1 Preparation of snoRNA

Several snoRNA constructs of snR81 were generated via *in vitro* transcription. It was possible to obtain the full length construct, the standalone hairpin constructs H5 and H3 as well as the shortened standalone hairpin constructs H5Δ and H3Δ. Since the transcription reaction with standard parameters generated all constructs in good yield and denaturing gel analysis showed a single band, no extra steps for transcription optimizations were undergone. The generated amount of snoRNA was sufficient for complex reconstitution and catalytic activity assays as well as for possible PELDOR samples (in the milligram range). The folding of the RNA constructs was tested with native PAGE and for all constructs, a single major band was visible, showing the formation of the kinetic hairpin fold of the RNA constructs. Construct H3 showed a minor band above the main product. Since a second RNA product from the transcription can be excluded from denaturing PAGE analysis, this secondary band likely represents a minor conformation of misfolded RNA. However, this slight amount of misfolded RNA might even be a feature of the H3 standalone hairpin, resulting from the absence of the 5' hairpin, which will be discussed later.
The generated shortened constructs H5Δ and H3Δ represent each hairpin in which the characteristic stem-loop structure was removed and replaced with a stable tetraloop five nucleotides above the pseudouridylation pocket. This structural element is not involved in the recruitment of the substrate RNA via Watson-Crick base-pairing and seems to function as a binding site for proteins – Nop10 and Nhp2 specifically. However, while archaeal L7Ae strictly needs the upper part of the hairpin structure (with its characteristic K-turn motif in archaeal H/ACA RNAs)[126], the eukaryotic homologue Nhp2 forms a stable tertiary complex with Cbf5 and Nop10[122,126,155]. This seems odd, since Nhp2 appears to not require the conserved upper stem-loop structure and should be able to bind to the full RNP only via protein-protein interactions (a hypothesis that will be analyzed and discussed in chapter 4). However, due to the extremely similar structure of L7Ae and Nhp2, binding of Nhp2 with Nop10 but also with the upper stem loop of the snoRNA are presumed (Figure 51)[362].

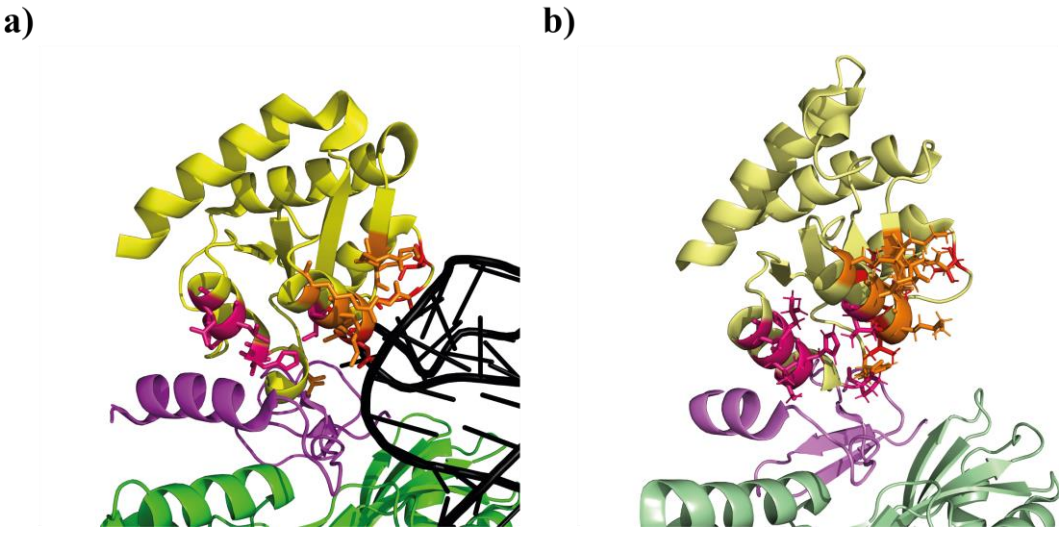


Figure 51: a) Crystal structure of an archaeal H/ACA RNP from *Pyrococcus furiosus* (pdb: 3HAY) with sRNA (black), L7Ae (yellow), Nop10 (magenta) and Cbf5 (green) illustrating the binding site of L7Ae to the complex. Marked amino acids in L7Ae are interacting with the H/ACA RNA (R34, K35, T37, E39, K42, R46, D59: orange, with Van der Waals contacts N38, I93, V95, A96, A98: red) and Nop10 (T41, A65, H66, P69, L70, E73: pink). b) NMR structure of Nhp2 (yellow) (pdb: 2LBW) and crystal structure of Nop10 (magenta) and Cbf5Δ (green) (pdb: 3U28), aligned with the archaeal crystal structure from a). Marked amino acids in Nhp2 are corresponding to archaeal amino acids pointed out above (K57, R58, K61, E62, K65, K69, W82: orange, V60, T116, T120, S121, V123: red, V64, P83, S88, H89, V92, L93, D96: pink)[362].

3.6.2 Preparation of proteins

The formation of the eukaryotic snR81 H/ACA snoRNP from *Saccharomyces cerevisiae* requires the four proteins Nhp2, Nop10, Cbf5 and Gar1, which were generated via recombinant expression from *E. coli*. For all proteins, expression tests were performed at 37 °C (temperature at which the expression of homologous archaeal proteins in a previous work[359] were performed) and 20 °C (which is more suitable for yeast proteins). The expression at 37 °C showed very strong basal expression of proteins and misfolding of the proteins and formation of inclusion bodies was expected, which would lead to a precipitation of the protein. In fact, one large scale expression of NCG at 30 °C was carried out for test purposes, and while protein expression seemed fine in the PAGE analysis of cell lysate, during the analytic purification, the protein nearly quantitatively precipitated together with cell debris after cell lysis. Also, the lower expression yields at 20 °C could be countered by longer expression times. While for archaeal proteins, the expression at 37 °C for 1-3 h was usually sufficient[359], the expression at 20 °C for the eukaryotic proteins was carried out overnight. The use of TB medium also made a strong impact on protein yield, since a much higher cell density can be reached compared to standard LB medium as evidenced by test expressions.

From literature, it is known that the proteins Nop10, Cbf5 and Gar1 are able to form a trimeric complex, of which also a crystal structure is available[160,372]. These protein-protein interactions could be utilized in a co-expression and co-purification strategy, to generate the trimeric NCG complex in a stoichiometric amount. The protein purifications were carried out

with affinity chromatography (His-Trap) and size exclusion chromatography (Superdex). During the affinity chromatography step, most of the endogenous *E. coli* proteins could be removed from the sample. A relatively large amount of endogenous proteins, that run at a similar height in PAGE gels than Gar1 did bind to the affinity chromatography column, but eluted at a lower imidazole concentration than the actual NCG proteins and could therefore be removed. Another large protein band at approximately 45 kDa could not be removed during affinity chromatography, however during size exclusion chromatography this protein could be removed from the sample, which led to a pure NCG protein complex (see Figure 26, Figure 27).

The protein preparation of shortened versions of the NCG subcomplex, which contained either or both of only the core domains of Cbf5 and Gar1 was performed in a similar fashion than the NCG protein expression and purification. The shortened constructs were more prone to precipitation at higher temperatures, especially constructs containing only the Gar1 core domain (GΔ). For these constructs, handling on ice was crucial to prevent spontaneous precipitation. The N- and C-terminal extensions of the eukaryotic proteins might help in stabilizing the protein structure and prevent precipitation of the full length construct proteins. The core domain-only constructs of Cbf5 (CΔ) and Gar1 (GΔ) are actually more similar to their archaeal counterparts, which are also lacking the N-/C-terminal domains. Even with the lack of those structural elements, the archaeal proteins from *Pyrococcus furiosus* are extremely thermostable, even at temperatures above 90 °C (native conditions for *P. furiosus* archaea), which is in stark contrast to their eukaryotic counterparts. This shows, that even despite a very similar structure of the archaeal and eukaryotic proteins, thermal stability might be governed through additional factors, likely through primary sequence elements or other inter-complex interactions of the proteins. In this way, archaeal proteins might be adapted to the higher temperatures and eukaryotic proteins are more reliant on additional conserved domains for proper structural stability. Also, the short Cbf5Δ construct has been shown to be more prone towards precipitation, even if expressed together with Nop10. Nop10 might have some stabilizing abilities towards Cbf5 when forming the bipartite protein complex and might be necessary altogether to express and purify the Cbf5Δ core domain. *In vivo*, the assembly factors Naf1 and Shq1 are binding to Cbf5 during the maturation of the RNP[122,163–165], which might provide stabilization for Cbf5 and in the *in vitro* expression of Cbf5, the immediate assembly with Nop10 might help to substitute for the missing assembly factors. *In vivo*, the KKD/E repeat containing C-terminal domain of Cbf5 (which is missing in Cbf5Δ) is necessary for a timely progression through the cell cycle and was also shown to be needed for the binding of microtubules *in vitro*.[373]

In literature, the co-purification of all four proteins at the same time is described[160]. Since Nhp2 interacts with Nop10 and is believed to form a stable Nhp2-Nop10-Gar1 complex *in vivo* prior to complex assembly with the H/ACA snoRNA[122,126,155], one might assume that the co-purification of all four proteins is feasible. While a co-expression of all four proteins would require the transformation of three different plasmids simultaneously into *E. coli*, this procedure would also put a lot of strain onto the expression system and was not tried. For a co-purification strategy, expression of His-tagged trimeric NCG and untagged Nhp2 was performed individually and the proteins were mixed together prior to cell lysis. However, after affinity chromatography, only NCG could be identified in the sample with Nhp2 in the

flow through. Presumably, the interaction of Nhp2 towards Nop10/Cbf5 was not strong enough for a co-purification strategy, at least at the buffer conditions utilized in this thesis (note: a co-purification of tagged Cbf5 and untagged Nop10/Gar1 was possible).
Therefore, Nhp2 expression was carried out individually with His-tagged Nhp2. The expression was done in the same way as for the NCG constructs. However, during Nhp2 purification, one major problem was the strong binding and co-purification of endogenous *E. coli* RNA towards Nhp2, which is not surprising, since Nhp2 is known to unspecifically bind to RNA[126,143,154,155]. While nucleic acid-binding could also be observed in part during NCG purification, several steps were undertaken to counter this problem: The addition of DNase and RNase A prior to cell lysis as well as the use of polyethyleneimine (PEI), which precipitates nucleic acids, which could then be removed from the sample via centrifugation. While this strategy was sufficient for NCG, Nhp2 still contained vast amounts of RNA even after those purification steps. To counter this problem, after affinity chromatography, the Nhp2 fractions were pooled and RNase A incubation was carried out overnight at room temperature (Nhp2 was much more stable at ambient temperatures than NCG). To quantitatively remove the RNase A from the sample, affinity chromatography and size exclusion chromatography were performed and RNase A removal was tested by incubation of snoRNA with Nhp2, which did not lead to RNA degradation. A 260/280 ratio of 0.58 verified the removal of endogenous RNA, and RNA-free Nhp2 could be obtained in good yield. In PAGE analysis, a small protein band below the actual Nhp2 band is visible (see Figure 33). However, Nhp2 fractions contained very high concentrations of Nhp2 and on PAGE analysis with “normal” amounts of Nhp2, the secondary band could no longer be identified (Figure 53), showing that this second protein was only contained in a very minor (and negligible) amount.
For Nhp2, the mutation S82W was introduced into the protein. This mutation induces a *cis* conformation into the adjacent proline 83 and shifts the *cis*/*trans* conformation equilibrium form a 40/60 ratio to a 60/40 ratio (Figure 52)[362]. In several archaeal L7Ae variants, a corresponding proline is locked in the cis conformation, which forms a “proline spine” with other prolines in Nop10 and Cbf5, which is considered to have a stabilizing effect on the overall structure of the H/ACA RNP[374]. With P83 adapting primarily a *cis* conformation, like L7Ae, it is also possible that it interacts with P33 of Nop10 and becomes part of a similar proline spine in eukaryotic H/ACA RNPs[362]. Furthermore, the *cis* conformation of P83 seems to point the backbone amide and carbonyl group of the adjacent amino acid at position 82 (S or W) towards the snoRNA, which may facilitate Nhp2-RNA interactions. Interestingly, a large number of snoRNAs contain a conserved U ten or eleven base pairs above the pseudouridylation pocket[362], which is similar to the conserved archaeal U in the K-turn motif, that is strictly required for L7Ae binding[126]. It was hypothesized, that also in eukarytotic H/ACA RNPs, this U can interact with the backbone of amino acid 82 of Nhp2 in a L7Ae-like way[362], although specific Nhp2-RNA interactions in absence of Nop10 are not possible, which will also be demonstrated in chapter 4.

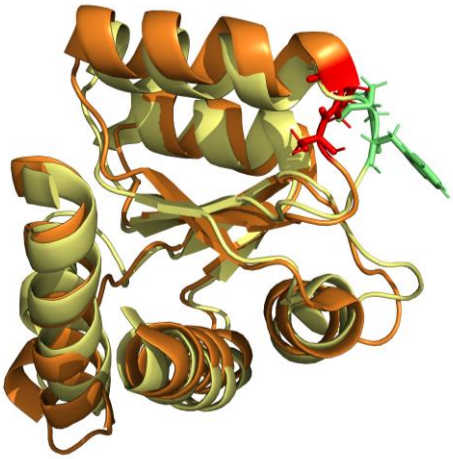


Figure 52: Overlap of NMR structures of Nhp2 wildtype (orange) (pdb: 2LBX) with amino acids S82 and P83 (red), preferring a *cis* conformation of P83 and Nhp2 S82W (yellow) (pdb: 2LBW) with amino acids W82 and P83 (green), preferring a *trans* conformation of P83[362].

Figure 53 shows the "main" purified RNP proteins. For Nop10-Cbf5, a small impurity, which runs slightly higher than Gar1 can be observed. This impurity can be attributed to a small amount of endogenous *E. coli* proteins that were co-purified (discussed above) and could not fully be removed by affinity chromatography and size exclusion chromatography. For NCG, the degradation of Gar1 into several discernible polypeptides can be observed. This degradation is known from literature[160]. While the Gar1 used here still showed strong catalytic activity during multiple substrate turnover experiments (discussed in chapter 3.6.6), the preparation of non-degraded Gar1 would be much desirable for structural analysis like crystallization and cryo-EM. While the Gar1 seen in Figure 53 contains a N-terminal His-Tag, a new construct with an C-terminal His-tag was generated and expression and purification was tested. Since the degradation seems to take place in the C-terminal domain (discussed in chapter 3.6.4), the utilization of an N-terminal His-Tag inevitably leads to the purification of all degradation products. With a C-terminal His-Tag however, this can be circumvented. During degradation, a C-terminal His-Tag is also removed from the main protein, and affinity and size exclusion chromatography would lead to more non-degraded protein. This was tested and eventually, a single band for Gar1 could be detected without any side products (see Figure 42). However, with longer storage of the affinity chromatography purified Gar1, degradation products would become visible again, while cryo storage of the samples did not lead to enhanced amount of degraded Gar1 (a test performed by Christin Fuks). With affinity chromatography and immediate subsequent size exclusion chromatography and cryo storage, the formation of vast amounts of degradation products of Gar1 should be preventable.

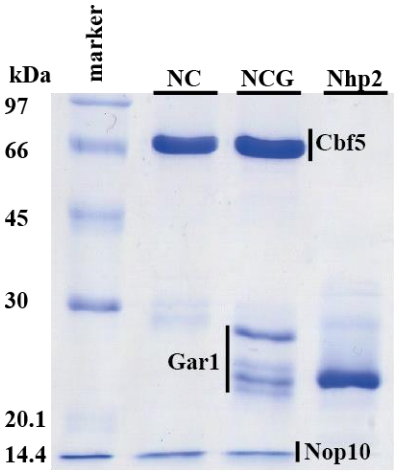


Figure 53: Purified protein constructs Nop10-Cbf5, Nop10-Cbf5-Gar1 and Nhp2.

3.6.3 Complex reconstitution

Since no crystal structure of the whole eukaryotic H/ACA RNP is available, but only a substructure of Nop10 and the conserved core domains of Cbf5 and Gar1[160], the exact interactions of the proteins with the snoRNA are not known. Based on crystal structure data from an archaeal H/ACA RNP[140], it is presumed that in the full eukaryotic H/ACA RNP complex, the main interaction with the snoRNA is performed by the protein Cbf5[101,149–151]. Nop10 interacts with Cbf5, the snoRNA and Nhp2, and Nhp2 interacts with Nop10 and the snoRNA. In archaeal complexes, Gar1 only interacts with Cbf5, but not with the snoRNA[140,158].

For complex reconstitution *in vivo*, the assembly factor Shq1 and the Gar1 analogue Naf1 are involved (Figure 54b)[122,163–165]. Shq1 only interacts with free Cbf5 and is believed to immediately bind to newly synthesized Cbf5 and provide stability for the protein. Afterwards, Cbf5 interacts with Nop10 and Nhp2 and forms a stable trimeric complex. The interactions of Nhp2 with the proteins Nop10 and Cbf5 are believed to facilitate the specificity of the RNA-binding protein Nhp2 towards H/ACA RNAs and prior assembly of this trimeric complex might be crucial for correct complex assembly *in vivo*. Naf1 binds to the trimeric complex in a way analogous to Gar1[126,163]. Since Naf1 also interacts with RNA polymerase Pol II and the Nhp2-Nop10-Gar1-Naf1 complex accumulates at actively transcribed H/ACA RNA genes, Naf1 is believed to target the complex towards the H/ACA RNA and promote RNP maturation[122,164,375]. The last step in the *in vivo* RNP assembly is the replacement of Naf1 with Gar1, transitioning the complex from an inactive to an active state[163].

Complex assembly *in vitro* is not in need of the assembly factors and is done in a different fashion (Figure 54a), but still results in an active H/ACA RNP (as discussed in chapters 3.6.5 and 3.6.6). The co-expression and co-purification of the Nop10-Cbf5-Gar1 trimeric subcomplex leads to an early Gar1 interaction with Cbf5. As discussed before, the immediate interaction of Nop10 with Cbf5 may provide stability to the otherwise not so stable Cbf5 and compensate for the missing Shq1. The trimeric protein complex was incubated with snoRNA at 30 °C (native yeast conditions) and the protein-RNA complex assembly was tested with electrophoretic mobility shift assays. The interaction of both biomolecules with an increasing

amount of protein *versus* a constant amount of RNA could be verified by the decreasing intensity of the gel-band of free RNA. The RNA-protein complex did only poorly migrate into the PAGE gel and mostly got stuck in the pockets of the gel (see Figure 34, Figure 35, Figure 68, Figure 69). Still, in the amount of complex that migrated into the gel, protein and RNA could be identified by both Coomassie and GelRed staining. The assembly was performed with the standalone hairpins H5 and H3 and with full length snR81 and was also tested with the core domain-only constructs of Cbf5 and Gar1. As expected, the core domain constructs showed weaker binding to the RNA than the full length NCG construct. This shows that the eukaryotic specific domains of Cbf5 as well as Gar1 (which is believed to not interact with the RNA) both play a role in correct complex assembly. During the heating at 30 °C for 30 minutes, a secondary RNA structure could be observed for RNA constructs H5 and H3 in the absence of proteins. The prolonged incubation at higher temperatures seem to induce a misfolding of the standalone RNA. Interestingly, in an EMSA with Nhp2 (shown in chapter 4, Figure 82), this second conformation disappeared first in the presence of Nhp2, showing the (unspecific) interaction with RNA in the absence of Nop10.

It is interesting to point out that for the full length construct NCG, full binding of the standalone RNA constructs H5 and H3 as well as the FL construct is achieved at around 2 molar equivalents of proteins. One might assume that the double amount of protein would be necessary to fully bind the FL construct in comparison to the standalone constructs, if always both hairpins bind to one equivalent of NCG respectively. This could point toward a preference of NCG towards one of the two hairpins, instead of equally binding to the 5' and 3' hairpins at the same time. However, this might also be attributed to the secondary conformations that H5 and H3 undergo during heating at 30 °C.

Nhp2, which was expressed and purified alone, was added separately to the snoRNA and NCG. In another study, for snoRNAs snR34 and snR5, dissociation constants for the different eukaryotic proteins were determined by tritium release assay[159]. For both snoRNAs, a nanomolar K_d was observed for the full complex with Nhp2-Nop10-Cbf5-Gar1, and the absence of Nhp2 did not influence the K_ds. The K_d for snoRNA and only Nhp2, without the other proteins, was greater by three orders of magnitude, showing the strong reliance of Nhp2 on the other proteins to efficiently bind to the complex. The co-purification of Nhp2 together with the other proteins however did not work (as discussed in chapter 3.6.2) since Nhp2-Nop10 interactions in absence of the snoRNA were apparently too weak *in vitro* compared to the *in vivo* reconstitution to efficiently co-purify Nhp2 with the other proteins. This may strongly be influenced by endogenous *E. coli* RNA that binds to Nhp2. Since the Nhp2-Nop10-Cbf5 formation *in vivo* is believed to be responsible for specific Nhp2 binding to the RNP complex, the binding of endogenous snoRNA during the *in vitro* purification may hinder the assembly of the Nhp2-Nop10-Cbf5 complex and only after the complete removal of the endogenous RNA, Nhp2 is able to efficiently bind to Nop10 and also the snoRNA. Since complete removal of endogenous RNA was only possible after the first affinity chromatography with most of endogenous *E. coli* proteins/RNA removed, Nhp2 could not form a stable complex with the other proteins in the cell lysate.

Interestingly, some single turnover assays showed the same complex activity, if Gar1 was expressed and purified without the other proteins and then added during complex assembly prior to the assays, while multiple turnover assays performed by Gerd Hanspach showed

weaker complex activity. This might suggest a very slow formation of the Gar1-Cbf5 interaction. Compared to the *in vivo* assembly, Gar1 is the last protein to bind to the complex, but also substitutes the already present assembly factor Naf1, which is not present in *in vitro* reconstitution.

In Figure 54, a simplified scheme shows the differences in complex assembly of the eukaryotic H/ACA for the *in vivo* route as well as for the *in vitro* reconstitution utilized in this thesis.

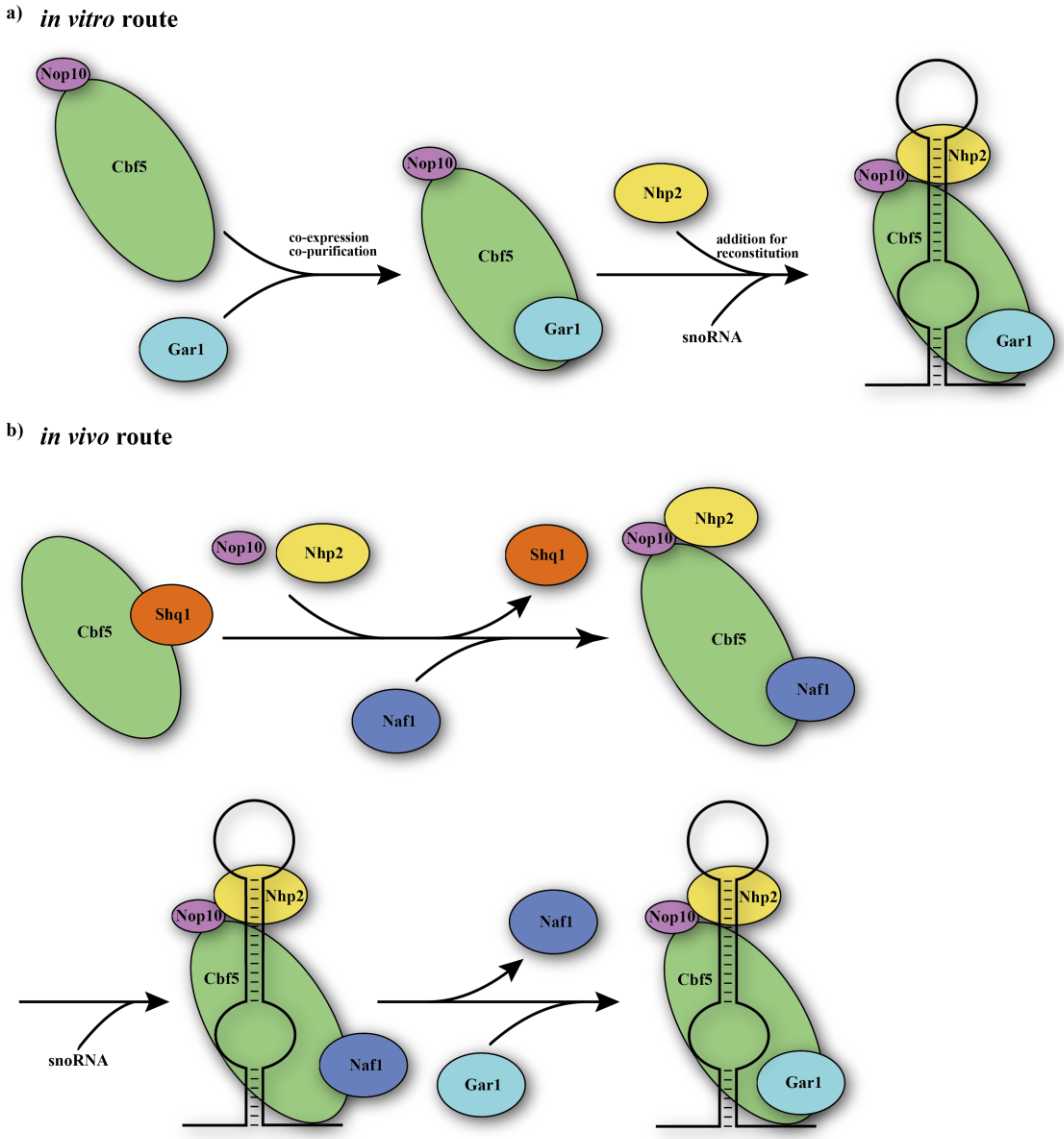


Figure 54: a) *In vitro* RNP complex assembly. Cbf5/Nop10 and Gar1 are co-expressed and co-purified as a stable trimeric complex. For complete reconstitution, Nhp2 and snoRNA are added. b) *In vivo* complex assembly. Cbf5 is stabilized by assembly factor Shq1. After formation of a stable trimeric complex with Nop10 and Nhp2, assembly factor Naf1 occupies the Gar1 binding site. Together with the snoRNA, a pre-snoRNP is formed, which becomes an active snoRNP after the substitution of Naf1 with Gar1[122].

The snoRNA is the limiting step in the complex assembly (as no activity could be observed for the proteins in absence of snoRNA). For this reason, to ascertain complete complex assembly, the proteins were added in an excess to the snoRNA for activity assays as well as smFRET experiments. In activity assays, the excess of proteins is not problematic, as they display no activity on their own, if not bound to the snoRNA. In smFRET experiments, the excess proteins, which were not bound to a biotinylated snoRNA, did not contribute to observed RNPs (discussed in chapter 4).
To provide a method that resulted in quantitative RNP complexes with no excess proteins, the complex assembly was furthermore tested with size exclusion chromatography of the full complex. The migration behavior on the SEC column of the RNA constructs FL, H5 and H3 and also of the proteins was tested and then an assembled complex was loaded onto the column. A shift of the main peak indicated protein bound to the complex and excess protein was separated from the RNP peak. Whether both NCG as well as Nhp2 are correctly bound to the snoRNA can also not be verified by this method. However, the presence of Nhp2 in the RNP fraction could be verified by fluorophore labeled Nhp2, and bound Nhp2 at least indicates the presence of Cbf5 and Nop10 as well. Furthermore, an activity test of the main fraction, that was believed to contain the fully reassembled RNP showed activity similar to a positive control, which also confirms the presence of Gar1.
For the full length snR81 construct, no apparent shift was visible in the elution chromatogram of the size exclusion chromatography between the RNA and the RNP. This behavior could be explained by the fact that the migration behavior of big molecules, especially RNA and RNA-protein complexes, is not only determined by the molecular weight of the biomolecules but also by the hydrodynamic radius of the molecules or complexes. Especially with the FL snoRNA with the double hairpin structure, it might be possible that the migration behavior of this construct differs from the others because of its hydrodynamic radius.

3.6.4 Eukaryotic specific domains in Gar1

In archaeal H/ACA RNPs, Gar1 only interacts with Cbf5, and is the only protein to not interact with the snoRNA[140,158]. It acts as mediator between the open and closed state of the thumb-loop domain of Cbf5, which is part of the catalytic center of uridine to pseudouridine formation. In eukaryotic RNPs, a similar role for Gar1 is proposed[140,157,158,161]. However, in contrast to archaeal Gar1, eukaryotic Gar1 contains arginine-glycine-glycine rich domains (“RGG” or “GAR” domains) at its N- and C-terminus[158,358]. Those GAR domains have been observed in other proteins and play an accessory role in RNA binding by enhancing the function of primary RNA binding motifs[358,363–368]. Gar1 has been shown to specifically bind to the snoRNAs snR10 and snR30 via an internal RNA binding motif in the core domain, and binding is enhanced by the presence of the GAR domains[358]. For H/ACA RNPs however, Gar1 is believed to only bind to Cbf5 and not the snoRNA, based on crystal structure data[122,140]. In this thesis, several Gar1 constructs with presence or absence of the specific GAR domains were expressed and purified and the interactions of the snoRNA constructs H5 and H3 were probed via EMSA for binding capabilities of the Gar1 constructs. Interestingly, an increasing concentration of Gar1 lead to protein-RNA binding which was also enhanced by the GAR domains. As for the EMSAs performed with NCG, no clear protein-RNA band was

visible and the protein-RNA complex was not able to enter the gel, which led to precipitation in the gel pocket. However, with the "vanishing" of the free RNA band and the formation of an RNA-protein band in the gel pocket, still a qualitative statement about the RNA-binding capabilities of the protein can be made. While for the full length Gar1, 1-2 molar equivalents of Gar1 lead to full binding of the RNA, for the core domain 5 molar equivalents were necessary. While for construct H3, a considerable binding of RNA by Gar1 could only be observed for the full length Gar1, for construct H5 more gradual effects were visible. Both C- and N-terminal GAR domains seemed to have the same effect on RNA binding, which was tested with constructs that only contained one of the two domains. Furthermore, a shortening of the C-terminal GAR domain could be used to pinpoint the area of the described degradation of Gar1 to the end of the C-terminal GAR domain. However, with these experiments it cannot be discerned if the Gar1 binding to the snoRNA is of specific or unspecific character.
One thing to point out for these EMSAs is the fact, that some Gar1 constructs showed a stronger binding towards the snoRNAs than the full ternary NCG constructs. This might have two reasons. For once, Gar1 on its own may have indeed a stronger binding towards snoRNAs as it has in combination with Cbf5 and Nop10. In the ternary construct, the protein may be aligned differently and parts of possible binding sites may be obstructed or used for the binding towards Cbf5. A second explanation is the concentration determination of the proteins. While it was possible to measure the concentration of NCG constructs with UV-vis spectroscopy, for Gar1 standalone constructs a Bradford assay was utilized. This method may contain an error, also since the standard curve was obtained by a different protein, which may show stronger intercalation of dye then Gar1. Since the Bradford method was used for all standalone Gar1 constructs, the Gar1 EMSAs are comparable to one another but might not be comparable to the NCG EMSAs. Furthermore, quantification via Bradford assay is not the most precise of methods and the RGG domains may be able to influence the binding of Coomassie dye to the different constructs.

3.6.5 Single turnover pseudouridylation assays

To verify the catalytic activity of *in vitro* reconstituted eukaryotic H/ACA RNPs, pseudouridylation assays under single turnover conditions with an excess of complex over substrate were performed. First, to ascertain the general ability of the full length as well as standalone hairpin complexes to perform uridine to pseudouridine formation, uniformly ^{32}P-labeled substrate RNA was generated via transcription. Pseudouridylation reactions were performed at 30 °C, the same temperature that was used for complex annealing, to mimic native yeast conditions. After a reaction time of 1-2 h, the substrate RNA was digested and pseudouridine was separated from uridine via thin layer chromatography. Furthermore, to more precisely identify the pseudoruridine spots, 2D TLC was performed. The uridine and pseudouridine spots migrated on the TLC according to literature values[369], verifying active RNP complexes. To more reliably identify the Ψ to U ratio, site specifically ^{32}P-labeled substrate RNA was generated by splinted ligation. Both 5' and 3' substrate RNAs showed a single radioactive product band on the PAGE gel, which could be excised and substrate RNA could be obtained in good yield. Single turnover reactions with reconstituted complex,

containing the snoRNA FL or H5/H3 constructs, NCG and Nhp2 all showed pseudouridine formation. For the 5' hairpin in the full length construct as well as for H5, higher Ψ formation could be observed than for the 3' hairpin and H3.
Activity was also tested for the core domain versions of Cbf5 and Gar1. As expected, the absence of the eukaryotic specific domains resulted in activity loss of the RNP, which is in line with the observation from the EMSA experiments, which revealed a weaker binding of the proteins to the RNA in case of eukaryotic domain omission. While for construct H5, the reduced activity was only around 5% for the shortened proteins, the effect of the domain omission was more drastic for construct H3 (see Figure 46). Interestingly, while NCG resulted in 61% substrate turnover after 2 hours, the activity for both NCΔG and NCΔGΔ was 45% and 44% respectively, while for NCGΔ it dropped to 30%. One might assume that the construct NCΔGΔ would result in the lowest activity, which however was not the case. Maybe the omission of the eukaryotic specific C-terminal Cbf5 domain is not as drastic for the complex activity than the omission of both C- and N-terminal Gar1 domains. Furthermore, the Cbf5 core domain may more easily interact with the Gar1 core domain, than the full length Cbf5 can with the Gar1 core domain. This might explain, why the combination of full length Cbf5 and core domain Gar1 resulted in the lowest activity of the different full length/core domain protein combinations. For H5, the same effect could be observed, with also NCGΔ having the lowest turnover rate with 88%. However, the difference to the other shortened protein constructs was only 1-2% and therefore within the margin of error. Interestingly, the full length Gar1 shows higher activity, even though several degradation products are visible during PAGE analysis (discussed above). The decrease of complex activity for eukaryotic specific domain omission is in accordance with literature, where the full length H/ACA proteins were tested vs. the core domain proteins on snR5 snoRNA[160].
Additionally, negative control experiments showed no catalytic activity if either the H/ACA snoRNA or NCG proteins were omitted from the reaction, showing the strong reliance of the protein-RNA interactions for H/ACA RNPs in comparison to pseudouridine formation by standalone pseudouridylases. Also, the reaction of H5 RNP with 3' substrate and H3 RNP with 5' substrate did not result in any pseudouridine formation in the respective hairpins. This shows, that the recruitment of the correct substrate is highly specific and can only be achieved by the correct Watson-Crick base-paring in the pseudouridylation pocket.
First tests with Nhp2 omission showed, that the complexes still exhibit basal activity in the absence of Nhp2. The Nhp2 omission was impacting the 3' hairpin more heavily than the 5' hairpin. However, no statement about the impact on multiple substrate turnover ability can be made from these tests and will be discussed in the next chapter.

3.6.6 Multiple turnover pseudouridylation assays

For a more detailed insight into the reaction kinetics of the pseudouridine formation and the influence of different snR81 snoRNA constructs and protein constructs, multiple turnover reactions were performed. Only 2.5% of the used substrate RNA was ^{32}P-labeled and the reaction was spiked with unlabeled substrate, to gain a 40fold substrate excess over RNP. The pseudouridylation reactions were carried out overnight and samples were taken at several time points during the first four hours and at 22 h after reaction start, at which point the catalytic

activity of the complex was declining and most constructs had reached a plateau of pseudouridylated substrate. With the amount of ^{32}P-labeled pseudouridine vs. uridine at different time points, the reaction kinetic could be fitted after a Michaelis-Menten like enzyme kinetic equation, which determined the amount of pseudouridine at the end of the reaction and the substrate half-life, which marks the time point at which half the substrate was pseudouridylated in comparison to the reactions endpoint (Table 3). With the substrate half-life, a statement about the average velocity of the catalytic process for each individual construct in comparison to the other construct can be made. Additionally, the starting turnover rate of the catalysis for each construct was determined via linear regression of the first five minutes, during this time (the first time point at which a sample was taken was 5 minutes), the course of the catalytic reaction was assumed to be linear (Table 2, starting turnover rates for all constructs are shown in Figure 55).

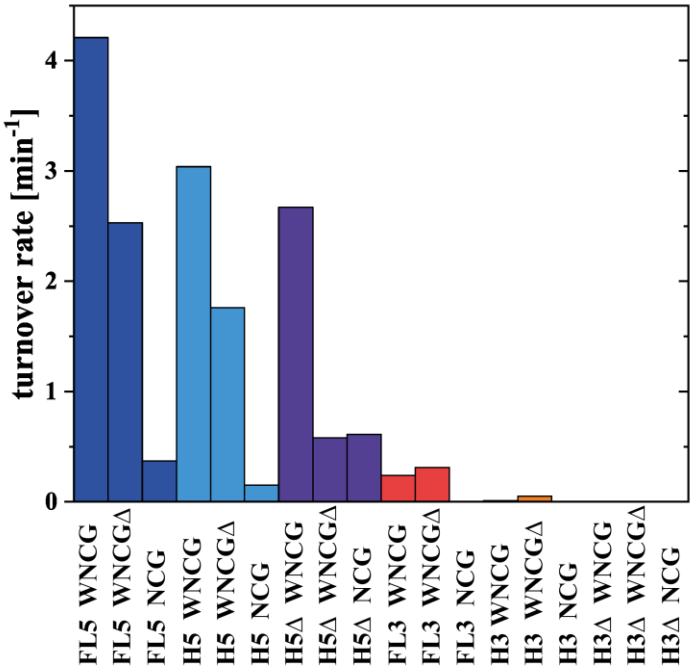


Figure 55: Comparison of starting turnover rates for all investigated RNPs.

First, the kinetics of fully reconstituted H/ACA RNP for full length snR81 was tested with both 5' and 3' substrate. For the 5' substrate, the RNP showed a starting turnover rate of 4.2 min^{-1}, which means one RNP molecule can pseudouridylate ~4 substrate RNAs in the course of one minute, taking approximately 15 seconds for the binding of the RNA, the substrate turnover and the product release. For the 3' substrate, the turnover rate was substantially reduced with 0.24 min^{-1}, being 18 times slower than for the 3' hairpin. The standalone hairpins H5 and H3 also showed a reduced activity for the 3' hairpin, with 3.04 min^{-1} for H5 and 0.01 min^{-1} for H3. In general, the standalone constructs H5 and H3 showed lower activity than their full length counterparts FL5 and FL3. However, while for construct H5, the activity was only slightly lowered in comparison to FL5, the difference between H3 and FL3 was already in the order of one magnitude. It seems, that the 5' hairpin

(with the H-motif) is not reliant on the presence on the 3' hairpin (with the ACA motif) and can perform the pseudouridylation reaction also in full scale on its own. The 3' hairpin however seems to be somewhat reliant on the presence of the 5' hairpin, only being able to effectively perform catalytic activity in the presence of the 5' hairpin. This may have to do with correct RNA folding, since on the native gels one minor second conformation for H3 was visible, while H5 and FL only showed a single conformation. On the other hand, this second conformation for H3 was only minimal and would not account for the drastic activity loss (assuming that the major conformation is the active conformation). Maybe, the correct folding of the H3 hairpin and therefore the association with the proteins takes longer, which results in such a slow starting turnover rate. But also the substrate half-life determined for H3 was extremely high with 31 hours, which lies well outside the measurement time of the experiment, since after 22 hours, H3 still had not reached a plateauing pseudouridine concentration, while the full length FL3 showed a substrate half-life time of "only" 1.5 hours.

In different studies, the pseudouridylation rate of the *in vitro* reconstituted snoRNPs for snR5 and snR34 where determined[159,160]. While different multiple turnover conditions were used and the results are not entirely comparable, in those studies also the reaction rate of the 5' hairpin vs. the 3' hairpin was tested with both full length and standalone-hairpin constructs and can be compared with the observations in this thesis. For snR5, the reaction was performed with a 100-fold substrate excess and only carried out for 2.5 h. For FL5, a starting turnover rate of 0.33 min^{-1} and for FL3, a starting turnover rate of 1.15 min^{-1} was determined, while H5 showed a starting rate of 0.10 min^{-1} and H3 of 0.63 min^{-1}. These starting turnover rates are approximately in the same area of magnitude than the starting reaction rates for snR81 determined in this thesis. However, for snR5 the 3' hairpin shows a higher activity than the 5' hairpin. Furthermore, both hairpins seem to be dependent on the presence of the other hairpin, since for both a relatively strong reduction of catalytic activity is visible, when the other hairpin is omitted from the construct. Interestingly, FL5 of snR81 completes pseudouridylation almost four times faster than FL3 of snR5. Here, also a reduced activity for the standalone hairpin constructs in comparison to the full length constructs was observed[160].

For snR34, turnover conditions were analyzed with a 10-fold substrate excess and the reaction kinetic was analyzed for 1.5 h. No exact starting reaction turnover rates were determined in this study. However, in the kinetic plot it is shown, that the 3' hairpin is more active than the 5' hairpin[159].

One might assume, that the multiple turnover reaction rate of a specific hairpin is tied to the Watson-Crick base-pairing capabilities of the pseudouridylation pocket. If more base-pairs are formed, the substrate RNA is bound more tightly, hence the release of the product takes longer and the reaction rate is slowed down. Table 4 shows the number of base-pairs involved in substrate recruitment and the amount of G-C and A-U base pairs.

Table 4: Number of base-pairs forming in the 5' and 3' pseudouridylation pockets of several snoRNAs for substrate recruitment. ΔG for each construct was calculated with IDT OligoAnalyzer[376]. For comparability reasons, for the unpaired target U and the unpaired adjacent base, the corresponding base-pairs (G-C, A-U) were used in the calculation.

snoRNA	5' hairpin base-pairs				3' hairpin base-pairs			
	total	G-C	A-U	ΔG [kcal/mole]	total	G-C	A-U	ΔG [kcal/mole]
snR5	12	5	7	-29.75	12	4	8	-24.43
snR34	12	8	4	-28.06	13	6	7	-16.97
snR81	14	3	11	-23.76	12	2	10	-20.45

While the formed base-pairs are approximately the same for the constructs (12-14), the content of G-C and A-U base-pairs differs greatly. Especially for construct snR81, the amount of G-C base-pairs is extremely low. Exemplarily, the ΔG values for the base pairing in the pseudouridylation pocket were calculated (Table 4). Even though this approach calculates helical interactions and is therefore of limited comparability to the facial interaction found in H/ACA RNPs, it may allow for a qualitative comparison of the interaction strengths between different guide RNA/substrate RNA pairs. Interestingly, ΔG for the 3' hairpin of snR34 is reduced due to two mismatches in the substrates base-pairing (G-A and U-G). Possibly, the mismatches and the reduced ΔG facilitate product release and are the reason for the higher enzymatic activity of the 3' hairpin in snR34 in comparison to the 5' hairpin. But then again, the 3' hairpin of snR81 shows the second lowest ΔG of the constructs compared, while at the same time showing decreased activity in comparison to the 5' hairpin.

For snR5 and snR34, also the dissociation constants of the substrate RNAs were determined and showed 53 nM for snR34-H5, 100 nM for snR34-H3, 160 nM for snR5-H5 and 330 nM for snR5-H3[159]. The amount of G-C base-pairs correlates with the dissociation constants in a way that more G-C base-pairs result in lower K_ds.

For both snR5 and snR34, the 3' hairpin forms fewer G-C base-pairs, which results in higher activity for this hairpin in comparison to the 5' hairpin. However, for snR81, this trend holds not true. The 5' hairpin forms one more G-C base-pair and two more base-pairs in total than the 3' hairpin, which should result in lower activity for the 5' hairpin than for the 3' hairpin. However, the results in this thesis show otherwise, with extremely strong activity for the 5' hairpin and reduced activity for the 3' hairpin. Maybe the two G-C base-pairs for snR81-H3 are already too low for efficient substrate binding, and while substrate release is faster than for the 5' hairpin, the substrate recruitment is a more rate limiting step.

Furthermore, while for the 5' hairpin of snR81 the base-pairings of the substrate is distributed between both sides of the pseudouridylation pocket (7 base-pairs one either side), the base-pairing in the 3' pseudouridylation pocket is heavily one-sided, with 3 base-pairs on the 3' side of the pseudouridylation pocket and 9 base-pairs on the 5' side of the pseudouridylation pocket. For comparison, the longest known base-pair length in nature for the whole pseudouridylation pocket is 17 base-pairs (in snR82), while the shortest is 8 base-pairs (in snR191), with 8 base-pairs seeming to be the minimum number of required base-pairs between substrate and guide RNA for pseudouridylation to occur[377,378]. For a single side of the pseudouridylation pocket, the longest interaction counts 10 base-pairs (in snR82), while the shortest is actually 3 (in snR3 and snR81). The 3' hairpin of snR34 is somewhat similar to

the 3' hairpin of snR81, with more base-pairs forming on the 5' side of the pseudouridylation pocket between guide and substrate RNA than on the 3' side. Reducing the base-pairs on the 3' side of the pseudouridylation pocket of snR34 via mutations had a major impact on pseudouridylation activity, while the effect by mutations on the 5' side was not as drastic. This shows, that already the disruption of one of the three base-pairs on the 3' side of the 3' hairpin of snR81 can have a major impact on the enzymatic activity. Interestingly, it was shown that even with the disruption of several base-pairs through mutations, while the activity was decreased, the actual binding of the substrate RNAs to the snoRNA was not reduced, showing that the reduced activity seems to occur due to incorrect binding and the target uridine not being correctly placed at the catalytic center of Cbf5[377].

Also, the 3' hairpin of snR81 is somewhat unique, since under stress conditions (nutrient deprivation), it is able to induce pseudouridylation not only on its native target (U1051 of 25S rRNA), but also on U93 of U2 sRNA[96]. This is especially interesting, since U42 of U2 sRNA is the native target of the 5' hairpin of snR81, which results in both hairpins performing pseudouridylation on the same target RNA, and even on two target uridines, that are relatively close to one another. While the native 25S rRNA target forms 12 base-pairs with the snoRNA, for the stress-induced target site in U2 sRNA, two mismatches are present, resulting in 10 base-pairs. Further analysis revealed that the presence of two mismatches is the requirement for inducible pseudouridylation at position U93. However, the actual position of these mismatches seemed to be unimportant, since efficient pseudouridylation was also observable for mutant constructs, with the native mismatches mutated to base-pairs and the introduction of two new mismatches, either via modification of the target U2 sRNA or the 3' pocket of the snR81 snoRNA[96].

Furthermore, it is also likely, that other factors like the incorrect folding and assembly with the proteins somehow restrict the 3' hairpin in efficient catalytic activity. Also, snR81 is somewhat unique in the behavior, that H3 is strongly reliant on the presence of H5, but H5 not on H3, which was not observed for snR5 and snR34[159,160].

As a next step, the influence of several RNP proteins on the pseudouridylation reaction was tested for FL as well as H5 and H3. Cbf5, containing the catalytic subunit as well as Nop10 are strictly required for complex activity. Also, Gar1 is needed for multiple turnover of substrate. Therefore, the requirement for Nhp2 for efficient substrate turnover was tested by Nhp2 omission from the complex. For FL5 as well as H5, the starting turnover rate was significantly reduced, to 0.37 min^{-1} for FL5 and 0.15 min^{-1} for H5, which represents a reduction by a factor of ~10 and 20, respectively. However, after 22 h the amount of pseuodouridine at which the reaction plateaued was still relatively high (~80% compared to ~95% with Nhp2). For the 3' hairpin however, omission of Nhp2 led to a complete shutdown of complex activity. For H3, no pseudouridine at all could be detected, while for FL3 only after 22 h trace amounts of pseudouridine could be identified. Since activity for H3 could be detected in single turnover experiments, although it was decreased, the omission of Nhp2 mostly seems to affect the multiple turnover abilities of the 3' hairpin. Without Nhp2, the uridine to pseudouridine formation seemed just to be slowed down at the 5' hairpin, pointing towards a catalyzing factor for Nhp2. For the 3' hairpin however, Nhp2 seems to be strictly required for basic complex activity under multiple turnover conditions.

For snR5, also the omission of Nhp2 was tested in a different study[160]. Here, under multiple turnover conditions, Nhp2 also reduced the amount of pseudouridine, however only one time point was measured. With a 20-fold substrate excess and a reaction time of 1 h at 30 °C, the 5' hairpin of snR5 produced 52% Ψ if all proteins were present and 26% Ψ with Nhp2 omission. For the 3' hairpin, the effect was not as strong, with 98% Ψ with all proteins and 94% Ψ with omission of Nhp2. Even so the trend in hairpin activity is switched for snR5, the omission of Nhp2 seems to affect the less active hairpin more heavily than the higher active hairpin, showing the same trend as for snR81 in this thesis.
For snR34, also the effect of missing Nhp2 on complex activity was tested[159]. For this snoRNA, both hairpins seem majorly affected by the omission of Nhp2 (however still not as strong as the 3' hairpin in snR81). Interestingly, the more active 3' hairpin for snR34 is stronger affected by Nhp2 omission than the less active 5' hairpin, which is in contrast to snR5 and snR34.
In archaeal H/ACA RNPs, the Nhp2 analogue L7Ae is involved of the re-orientation and correct positioning of the substrate RNA and is strictly required for efficient substrate turnover[156,157]. While Nhp2 may have a similar function in eukaryotic H/ACA RNPs, it is also likely that Nhp2 has additional stabilizing functions, of which the requirement is different for each hairpin. While for all snoRNA hairpins, Nhp2 seems to perform a catalyzing function, that results in faster substrate turnover rates, for some hairpins Nhp2 seems to be strictly required for basic catalytic activity.

Table 5: Different snoRNPs show varying requirements of Nhp2 for effective multiple turnover enzymatic activity.

snoRNP and hairpin	Effect of Nhp2 omission on multiple turnover rates
snR5 5' hairpin	**activity loss**
snR5 3' hairpin	**minor activity loss**
snR34 5' hairpin	**major activity loss**
snR34 3' hairpin	**complex mostly inactive**
snR81 5' hairpin	**activity loss**
snR81 3' hairpin	**complex mostly inactive**

The eukaryotic specific domains of Gar1 showed increased RNA binding capabilities (of both NCG as well as standalone Gar1) and omission of these domains lead to decreased activity in single turnover assays. Since Gar1 is mainly responsible for efficient substrate release, the performance of pseudourydilation in a multiple turnover manner was also tested in the absence of the GAR domains. For FL5 and H5, starting turnover rates of 2.53 min^{-1} and 1.76 min^{-1} respectively were determined, which is approximately a reduction by factor 2 in comparison to full length Gar1. The GAR domains may provide a stabilizing effect on Gar1, provide stronger binding to Cbf5 and may also influence the binding of the Cbf5 thumb loop domain to Gar1 during substrate turnover. Surprisingly, the omission of the GAR domains had an activity increasing effect for both FL3 and H3 under multiple turnover conditons, with

initial turnover rates of 3.06 min^{-1} and 0.50 min^{-1} respectively, resulting in an activity increase in comparison to full length Gar1 by factor 1.3 and 3.6 respectively. For some reason, multiple turnover pseudouridylation works better on the 3' hairpin if the GAR domains are absent, contrary to single turnover conditions, where they showed an activity inhibiting effect. The GAR domains of Gar1 may play different roles for the two hairpins in the substrate turnover process, resulting in a promoting effect for the 5' hairpin, but slowing down substrate turnover for the 3' hairpin. Even though Gar1Δ showed decreased activity under single turnover conditions, some positive secondary effect for multiple substrate turnover seems to compensate for the decelerating effect. Maybe the RNA-binding promoting capabilities of the GAR domain play a specific role in this, interacting with the snoRNA in a way that stabilizes one hairpin, but at the same time destabilizes the other hairpin – possibly due to unspecific binding of the large unstructured region of the 3' hairpin loop. It is also possible that interactions with the substrate RNA during correct placement of the substrate or during the product release occur. Maybe this effect is also different for both hairpins due to the different binding of substrate RNA – evenly distributed between both strands of the pseudouridylation pocket for the 5' hairpin and asymmetrical for the 3' pseudouridylation pocket.

While the presence of one hairpin showed an effect in catalytic activity on the respective other hairpin, the effects induced by either Nhp2 and the GAR domains of Gar1 were in the similar range on both the standalone and the full length snoRNA. This suggest, that interactions between the two hairpins, that have an effect on the enzymatic activity, are not mediated by Nhp2/Nhp2 or GAR domain/GAR domain interactions between the two hairpins and are also not caused by interactions of Nhp2 or the GAR domains bound on one hairpin with components of the other hairpin. In a recent study, the H/ACA domain of the human telomerase RNA hTR, which also forms a double-hairpin structure like eukaryotic snoRNAs was analyzed by cryo-EM[169]. The fit of the structure shows no interaction of Gar1 with one of the hairpins, but interestingly may indicate a direct interaction of Gar1 with the upper area of the pseudouridylation pocket and the connected stem structure of the other hairpin (see Figure 10), which may point to an interaction with either snoRNA or substrate RNA. However, this interaction may well be caused by the core domain of Gar1, independent from the GAR domains. Furthermore, hTR has not been established as a guide RNA for pseudouridylation and therefore conclusions derived from the structure of the H/ACA domain of hTR may be of limited value for H/ACA snoRNPs.

The multiple turnover activity assays were also performed for the shortened snoRNA constructs H5Δ and H3Δ, which have their upper stem-loop structure replaced by a stable tetraloop, five base-pairs above the pseudouridylation pocket. For H5Δ, the omission of the upper snoRNA structure led to a minor decrease in activity for the full RNP with an initial turnover rate of 2.7 min^{-1}. This is interesting, since the upper stem loop structure together with Nop10 provides a binding interface for Nhp2. However, Nhp2 still seemed to be able to efficiently bind to the RNP. Additionally, the omission of Nhp2 led to an initial turnover rate of 0.6 min^{-1}, which is four times higher than for H5 without Nhp2. Not only was Nhp2 still able to bind to the complex without the upper hairpin structure, also its omission did not affect the complex as drastically as it did when the upper hairpin structure was present. This may point to a more sophisticated Nhp2-snoRNA interaction than just simply providing a binding

site for the protein. Maybe Nhp2, additionally to other functions, stabilizes the upper stem loop, which leads to increased catalytic activity of the complex, and destabilization and decreased function in the absence of Nhp2. If the upper stem loop is not present in H5Δ, the stabilizing interactions of Nhp2 are not needed anymore, which leads to an increased activity in the absence of Nhp2 in comparison to H5. However, Nhp2 must have additional functions, otherwise the activity of H5Δ both with and without Nhp2 would be nearly identical, which is not the case.

The omission of the GAR-domains led to a decrease in activity by a factor of ~4.5 and a starting turnover rate of 0.6 min^{-1}. For the H5Δ construct, the GAR-domain omission had a stronger impact than on H5, where it decreased activity only by factor 2. Possibly, the GAR domains interact in some way with the upper-stem loop of the snoRNA, and provide increased stability for the complex, again indicating that Gar1 does not only interact with Cbf5, but also with the snoRNA.

Surprisingly, the construct H3Δ did not show any catalytic activity at all for all tested protein constructs. For some reason, the 3' hairpin seems to be reliant on the upper stem hairpin to perform catalytic activity at all, which is in stark contrast to the 5' hairpin. This observation also seems connected with the reliance on Nhp2 for the 3' hairpin. When comparing the two hairpins in regards of paired and unpaired nucleotides, the 5' hairpin forms a double stranded stem of 28 paired nucleotides and a loop of 22 unpaired nucleotides in the upper stem region, while the 3' hairpin only forms a stem of 12 paired nucleotides and a loop of 43 unpaired nucleotides. Maybe, the 3' hairpin requires the upper stem-loop structure for correct folding and complex assembly, but at the same time is reliant on Nhp2 to bind to the upper loop and stabilize it, since the high amount of unpaired nucleotides in the 3' loop otherwise lead to misfolding and incorrect complex reconstitution.

3.7 Appendix of chapter 3

3.7.1 Expression tests

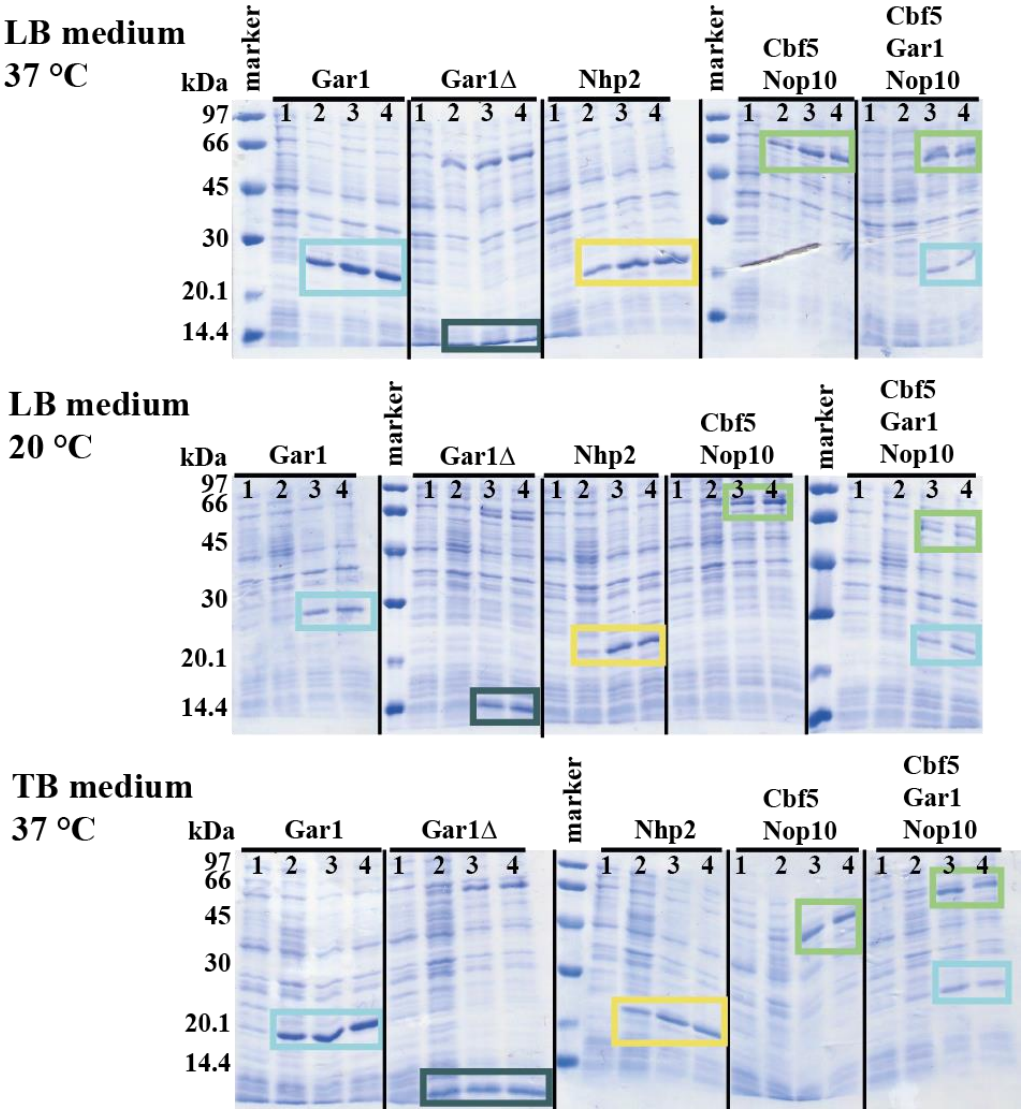


Figure 56: Coomassie stain - expression test of several *s.c.* proteins. Marker: low molecular weight marker (GE Healthcare). Samples: 1) before IPTG induction, 2) o.n. expression 0 mM IPTG, 3) o.n. expression 0.1 mM IPTG, 4) o.n. expression 1 mM IPTG. Expressed proteins are marked with boxes. Marker: low molecular weight marker (GE Healthcare).

3.7.2 Calibration of Superdex 200 10/300 GL increase column with standard proteins

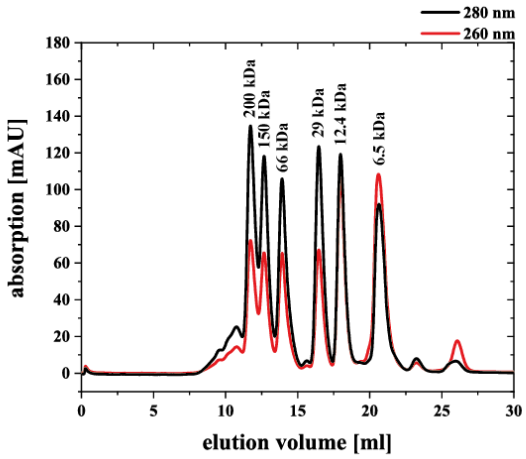


Figure 57: Size exclusion chromatography purification of standard proteins (β-amylase (200 kDa), alcohol dehydrogenase (150 kDa), bovine serum albumin (66 kDa), carbonic anhydrase (29 kDa), cytochrome C (12.4 kDa), aprotinin (6.5 kDa)) for calibration with the Superdex 200 10/300 GL increase column.

3.7.3 Expression of NCGΔ

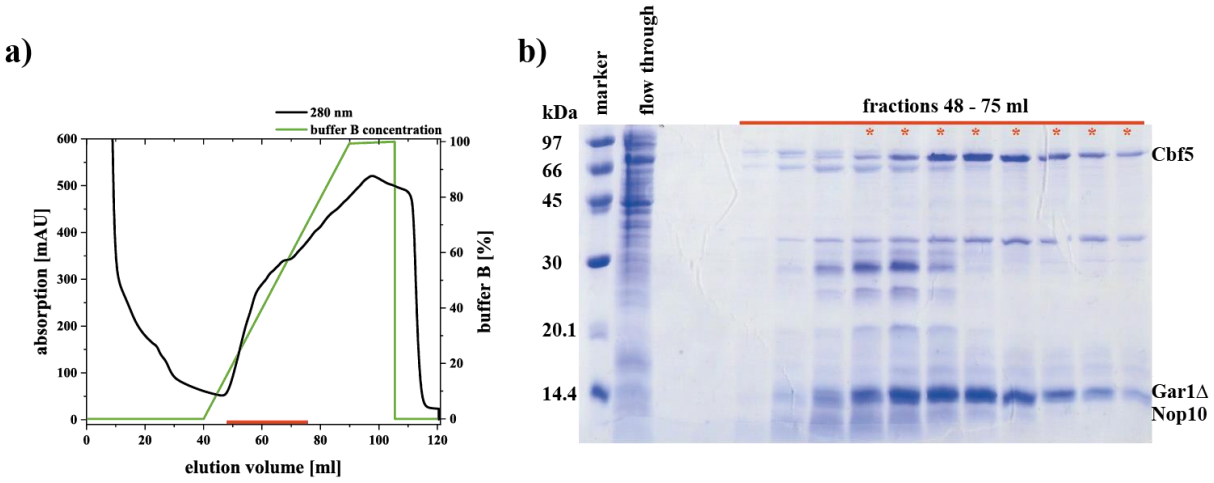


Figure 58: Affinity chromatography purification of ternary NCGΔ complex. a) Ni-NTA elution chromatogram using an imidazole gradient. Fractions used for analysis are marked with a red bar. b) PAGE analysis of fractions from Ni-NTA. Fractions used for further purification are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

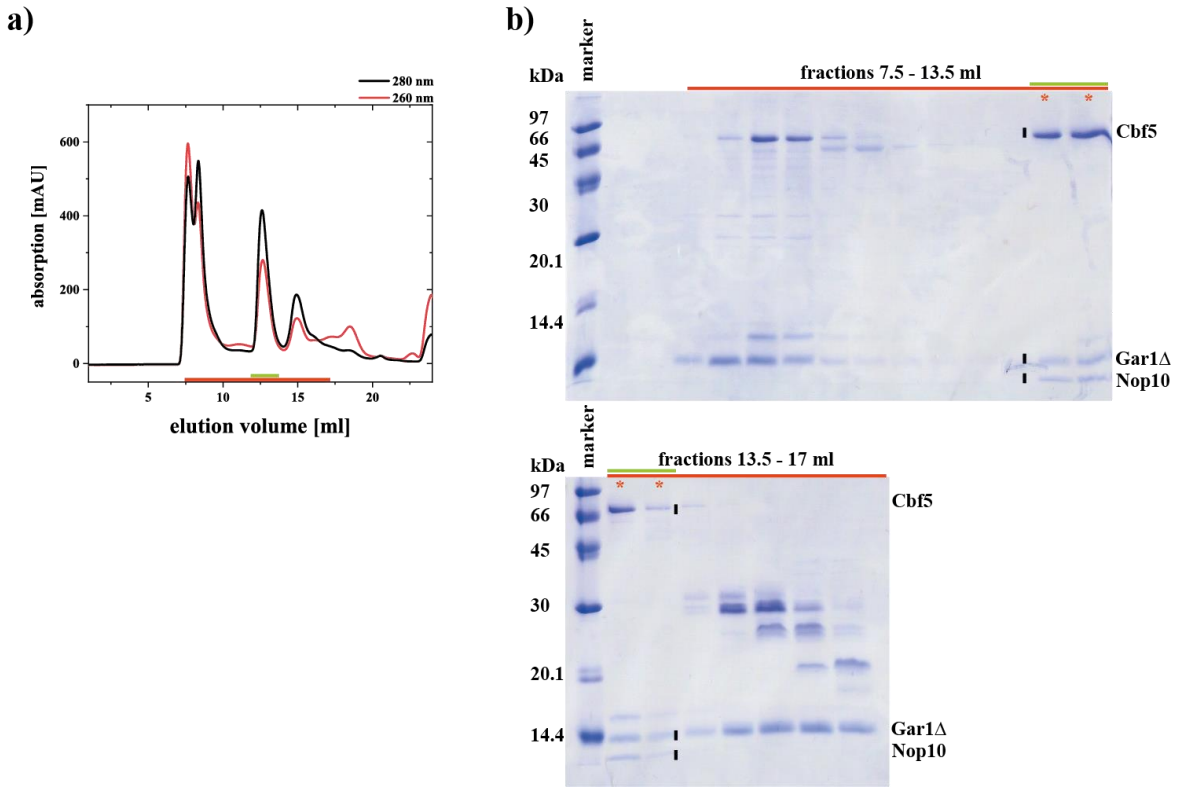


Figure 59: Size exclusion chromatography purification of ternary NCGΔ complex. a) SEC elution chromatogram using a Superdex 200 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. The relevant peak is marked with a green bar. b) PAGE analysis of fractions from SEC. Fractions used for further experiments are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare). For calibration of column with standard proteins for size comparison with the complex see appendix Figure 57.

3.7.4 Expression of Nhp2 in 500 mM NaCl buffer

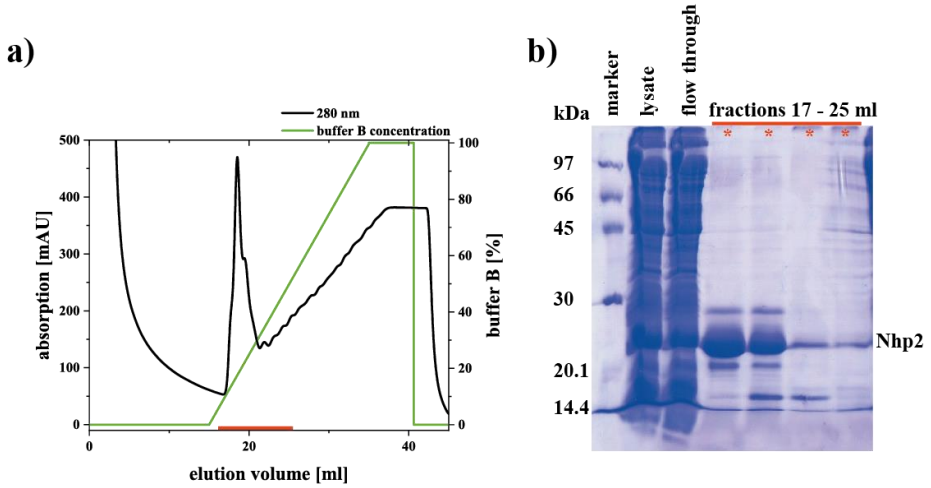


Figure 60: Affinity chromatography purification of Nhp2 before RNase A treatment. a) Ni-NTA elution chromatogram using an imidazole gradient. Fractions used for analysis are marked with a red bar. b) PAGE analysis of fractions from Ni-NTA. Fractions used for further purification are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

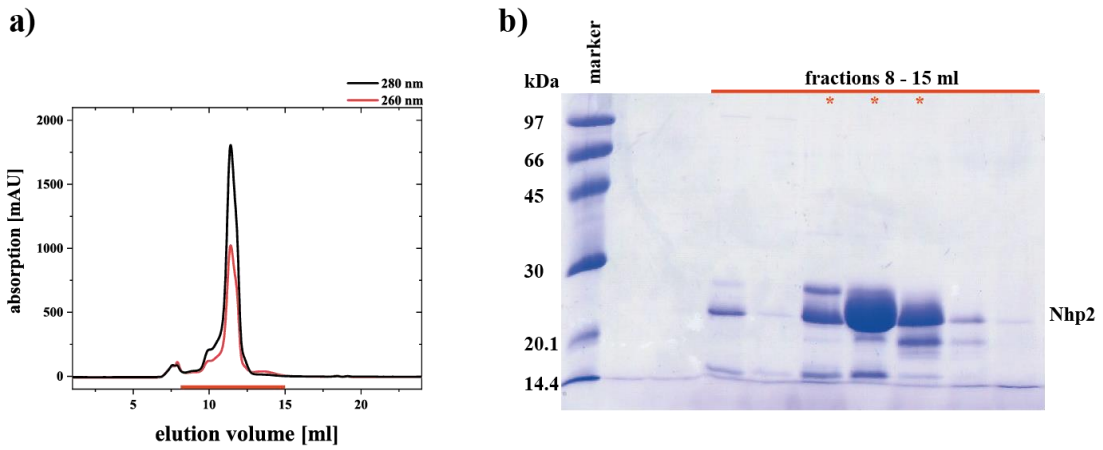


Figure 61: Size exclusion chromatography purification of Nhp2 before RNase A treatment. a) SEC elution chromatogram using a Superdex 75 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) PAGE analysis of fractions from SEC. Fractions used for further experiments are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

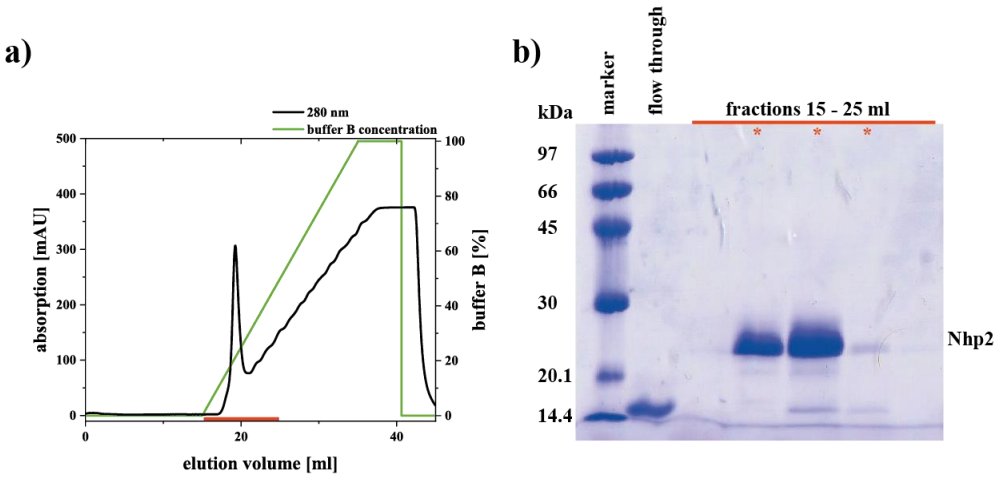


Figure 62: Affinity chromatography purification of Nhp2 after RNase A treatment. a) Ni-NTA elution chromatogram using an imidazole gradient. Fractions used for analysis are marked with a red bar. b) PAGE analysis of fractions from Ni-NTA. Fractions used for further purification are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

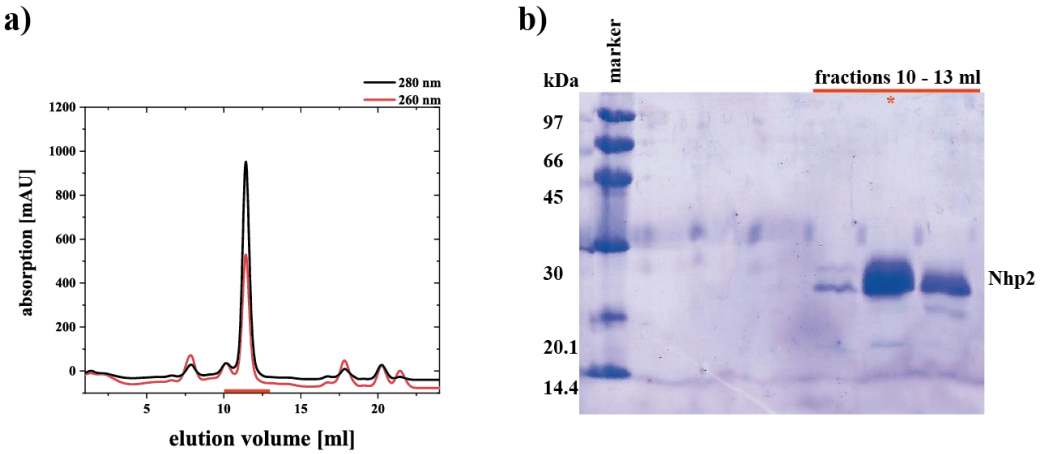


Figure 63: Size exclusion chromatography purification of Nhp2 after RNase A treatment. a) SEC elution chromatogram using a Superdex 75 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) PAGE analysis of fractions from SEC. Fractions used for further experiments are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

3.7.5 Simultaneous purification of NCG and Nhp2

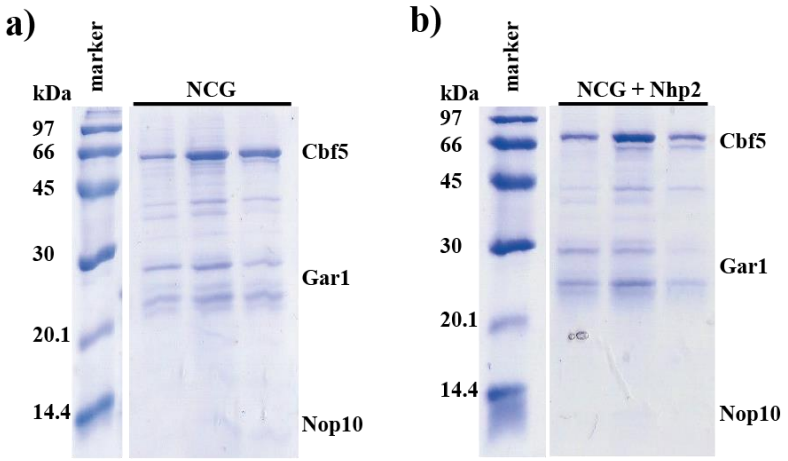


Figure 64: Size exclusion chromatography purification of NCG without (a) and with Nhp2 (b). In the PAGE analysis of fractions from SEC no Nhp2 band can be detected. Marker: low molecular weight marker (GE Healthcare).

3.7.6 Preparation of NC

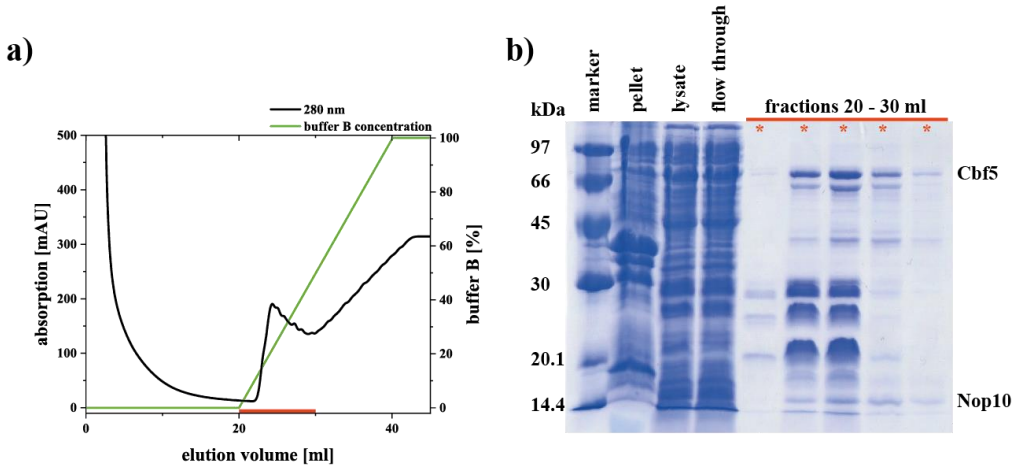


Figure 65: Affinity chromatography purification of Nop10/Cbf5. a) Ni-NTA elution chromatogram using an imidazole gradient. Fractions used for analysis are marked with a red bar. b) PAGE analysis of fractions from Ni-NTA. Fractions used for further purification are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

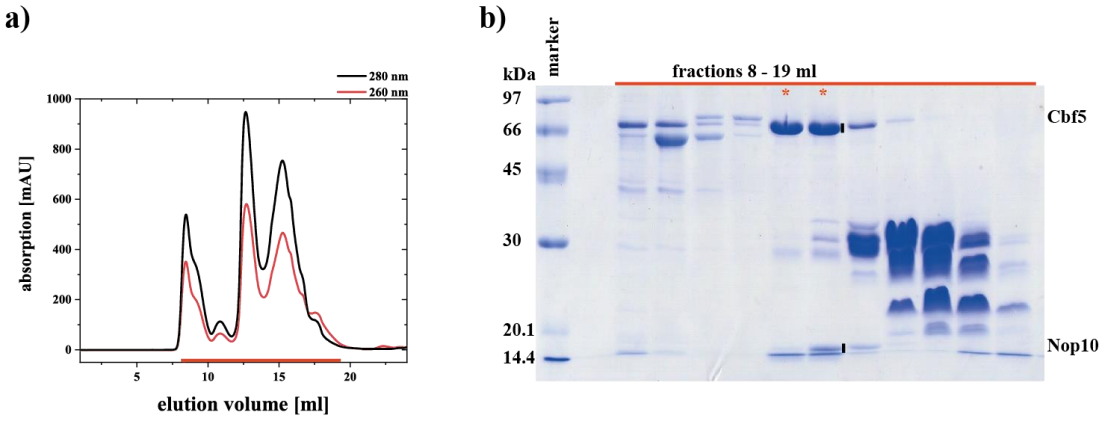


Figure 66: Size exclusion chromatography purification of Nop10/Cbf5. a) SEC elution chromatogram using a Superdex 200 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) PAGE analysis of fractions from SEC. Fractions used for further experiments are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

3.7.7 Preparation of plasmid for standalone Cbf5

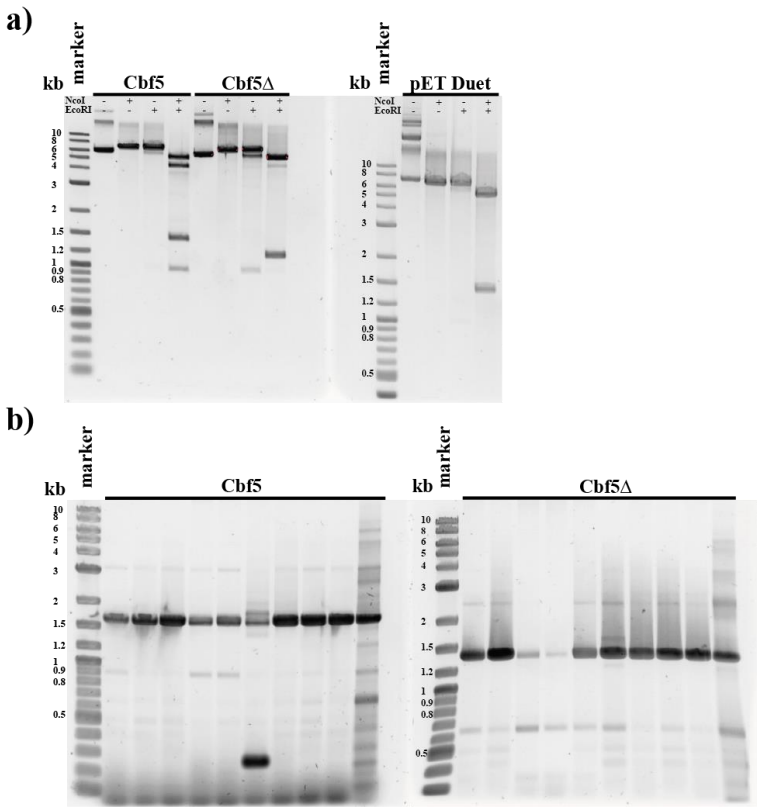


Figure 67: a) Restriction digest of Cbf5 and Cbf5Δ genes and pET Duet vector. b) colony PCR of Cbf5/Cbf5Δ standalone plasmid after cloning into pET Duet vector. Marker: 2-Log DNA Ladder (NEB).

3.7.8 EMSAs

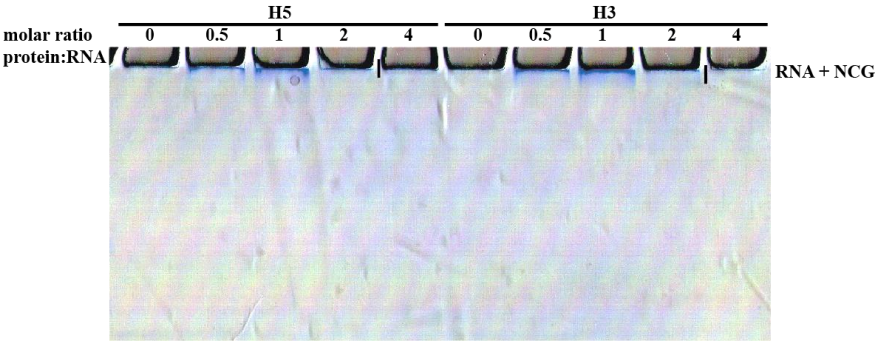


Figure 68: EMSA of snoRNAs H5 and H3 with NCG on a 12% native PAA gel (Coomassie stain).

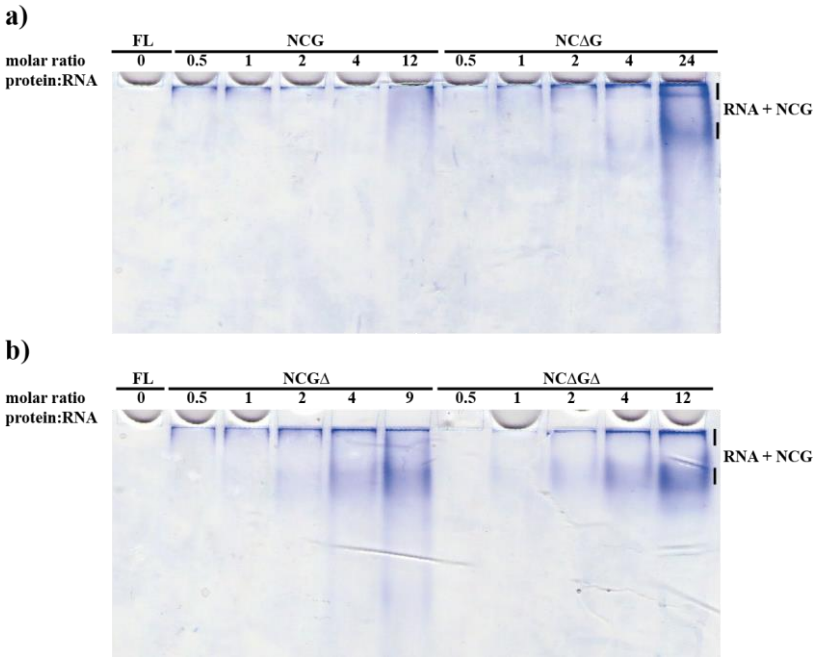


Figure 69: EMSA of snoRNA FL with NCG, NCΔG (a), NCGΔ and NCΔGΔ (b) on a 12% native PAA gel (Coomassie stain).

3.7.9 Size exclusion purification after complex assembly with fluorophore

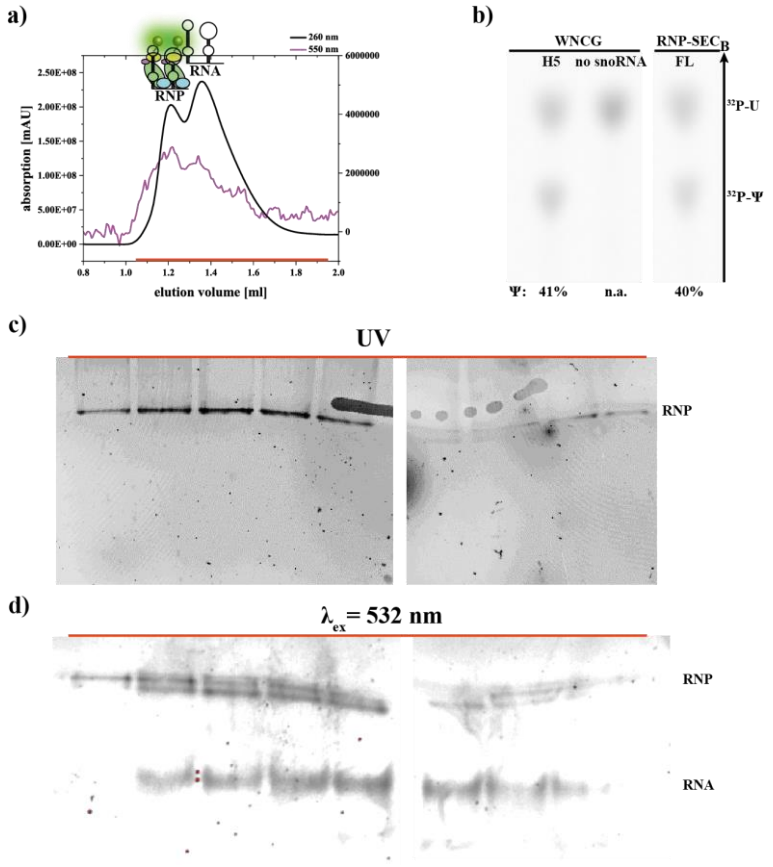


Figure 70: Size exclusion chromatography of RNP using a Superdex 200 3.2/300 GL increase column. a) Elution profile of snoRNA FL + NCG + Nhp2-K48-Cy3. b) Thin layer

chromatography of multiple turnover pseudouridylation activity tests performed with site specific ^{32}P-labeled 5' substrate RNA (1 µM) for 2 h. FL-RNP eluted after size exclusion chromatography shows a similar activity as the positive control. c+d) Eluted fractions analyzed on a 12% native PAA gel. UV scan (c) and fluorescence scan (d).

3.7.10 Preparation of substrate RNA via transcription

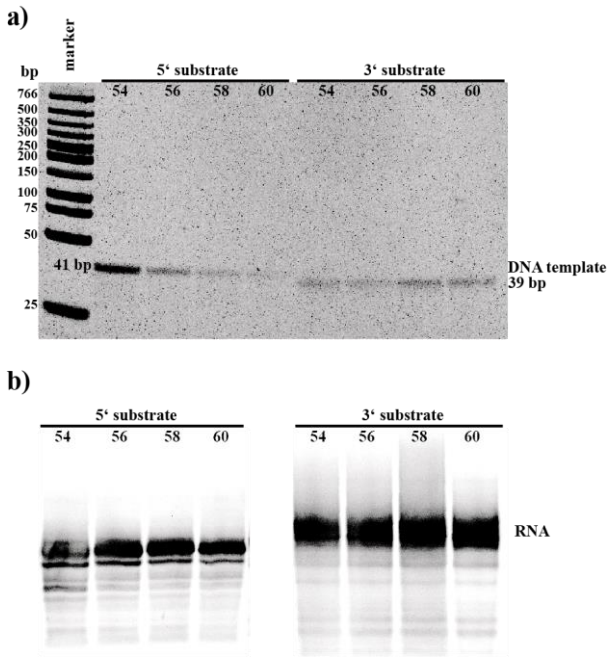


Figure 71: a) DNA template generation for substrate RNAs via annealing at different temperatures on a 2. Marker: Fast DNA ladder (NEB). b) transcription products of substrate RNAs on a 20% denaturing PAA gel.

3.7.11 Activity tests for Nhp2 WT/S82W and without Nhp2

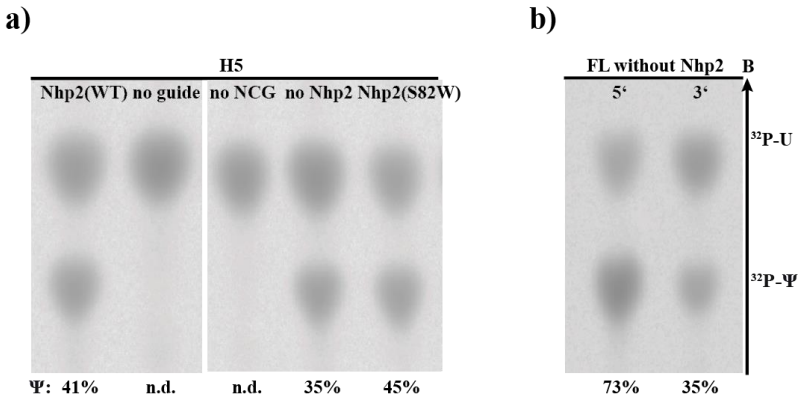


Figure 72: a) Thin layer chromatography of multiple turnover pseudouridylation activity tests performed with site specific ^{32}P-labeled 5' substrate RNA for 2 h (standard multiple turnover conditions). Nhp2(WT) and Nhp2(S82W) both show a similar activity promoting effect. b) Thin layer chromatography of single turnover pseudouridylation activity tests performed with FL snoRNA and site specific ^{32}P-labeled 5' and 3' substrate RNA (standard single turnover conditions). Omission of Nhp2 leads to reduced activity.

3.7.12 Multiple turnover assay data

Table 6: Reaction kinetics raw data

H5 WNCG				
time [min]	assay 1	assay 2	mean	Δ
5	36.28	39.62	37.95	2.36
20	70.12	71.64	70.88	1.08
60	83.93	84.27	84.10	0.24
120	87.85	87.87	87.86	0.01
240	91.54	89.13	90.34	1.70
1335	91.79	93.64	92.72	1.31

FL5 WNCG				
time [min]	assay 1	assay 2	mean	Δ
5	51.45	57.22	54.34	4.08
20	79.09	82.92	81.00	2.71
60	86.62	88.74	87.68	1.50
120	88.38	91.55	89.96	2.24
240	89.66	90.83	90.24	0.82
1335	95.81	96.67	96.24	0.60

H5 WNCGΔ				
time [min]	assay 1	assay 2	mean	Δ
5	24.36	20.87	22.62	2.47
20	52.29	49.82	51.05	1.74
60	71.50	73.48	72.49	1.40
120	79.34	82.01	80.68	1.89
240	83.99	86.29	85.14	1.62
1335	88.57	89.85	89.21	0.91

FL5 WNCGΔ				
time [min]	assay 1	assay 2	mean	Δ
5	33.54	30.40	31.97	2.22
20	62.62	62.53	62.57	0.06
60	79.46	80.81	80.14	0.96
120	86.29	87.58	86.93	0.91
240	87.87	88.75	88.31	0.62
1335	87.53	90.94	89.23	2.41

H5 NCG				
time [min]	assay 1	assay 2	mean	Δ
5	2.11	1.55	1.83	0.40
20	9.90	7.75	8.83	1.52
60	21.33	18.50	19.92	2.00
120	28.94	29.36	29.15	0.29
240	44.41	43.82	44.12	0.42
1335	71.84	74.25	73.05	1.71

FL5 NCG				
time [min]	assay 1	assay 2	mean	Δ
5	8.95	6.65	7.80	1.63
20	18.16	18.50	18.33	0.24
60	35.16	35.05	35.10	0.07
120	49.48	49.44	49.46	0.03
240	64.44	65.97	65.20	1.08
1335	85.57	83.42	84.50	1.52

H3 WNCG					
time [min]	assay 1	assay 2	assay 3	mean	Δ
5	n.d.	n.d.	n.d.	n.d.	n.d.
20	n.d.	n.d.	n.d.	n.d.	n.d.
60	2.77	n.d.	1.25	1.34	1.39
120	4.05	2.60	3.10	3.25	0.73
240	8.00	7.23	8.54	7.92	0.66
1335	24.80	24.99	29.85	26.55	2.86

FL3 WNCG				
time [min]	assay 1	assay 2	mean	Δ
5	2.83	3.44	3.13	0.43
20	10.01	11.25	10.63	0.88
60	21.98	24.86	23.42	2.04
120	30.14	37.42	33.78	5.15
240	39.23	47.27	43.25	5.68
1335	53.64	58.56	56.10	3.48

H3 WNCGΔ				
time [min]	assay 1	assay 2	mean	Δ
5	0.28	0.00	0.14	0.20
20	0.72	1.50	1.11	0.55
60	6.90	5.97	6.43	0.66
120	13.29	11.79	12.54	1.06
240	21.11	17.81	19.46	2.33
1335	38.05	45.73	41.89	5.43

FL3 WNCGΔ				
time [min]	assay 1	assay 2	mean	Δ
5	10.07	5.18	7.63	3.45
20	12.88	12.50	12.69	0.27
60	27.00	29.15	28.08	1.52
120	43.36	42.99	43.18	0.26
240	54.69	55.85	55.27	0.82
1335	69.15	71.82	70.49	1.89

H3 NCG				
time [min]	assay 1	assay 2	mean	Δ
5	n.d.	n.d.	n.d.	n.d.
20	n.d.	n.d.	n.d.	n.d.
60	n.d.	n.d.	n.d.	n.d.
120	n.d.	n.d.	n.d.	n.d.
240	n.d.	n.d.	n.d.	n.d.
1335	n.d.	n.d.	n.d.	n.d.

FL3 NCG				
time [min]	assay 1	assay 2	mean	Δ
5	n.d.	n.d.	n.d.	n.d.
20	n.d.	n.d.	n.d.	n.d.
60	n.d.	n.d.	n.d.	n.d.
120	n.d.	n.d.	n.d.	n.d.
240	n.d.	n.d.	n.d.	n.d.
1335	7.92	5.57	6.74	1.66

H5Δ WNCG				
time [min]	assay 1	assay 2	mean	Δ
5	33.70	30.46	32.08	2.29
20	70.13	63.52	66.83	4.68
60	84.71	82.64	83.68	1.47
120	--	--	--	--
240	89.37	88.62	88.99	0.53
1335	--	--	--	--

H5Δ WNCGΔ				
time [min]	assay 1	assay 2	mean	Δ
5	8.60	8.74	8.67	0.10
20	24.15	24.47	24.31	0.23
60	42.93	42.43	42.68	0.35
120	55.19	57.39	56.29	1.55
240	66.78	70.50	68.64	2.63
1335	76.39	79.29	77.84	2.05

H5Δ NCG				
time [min]	assay 1	assay 2	mean	Δ
5	11.33	8.87	10.10	1.74
20	26.69	24.71	25.70	1.40
60	45.62	43.74	44.68	1.34
120	--	--	--	--
240	74.07	72.75	73.41	0.93
1335	--	--	--	--

3.7.13 Activity tests for H5 in absence of Gar1 or Nop10 and for H3Δ

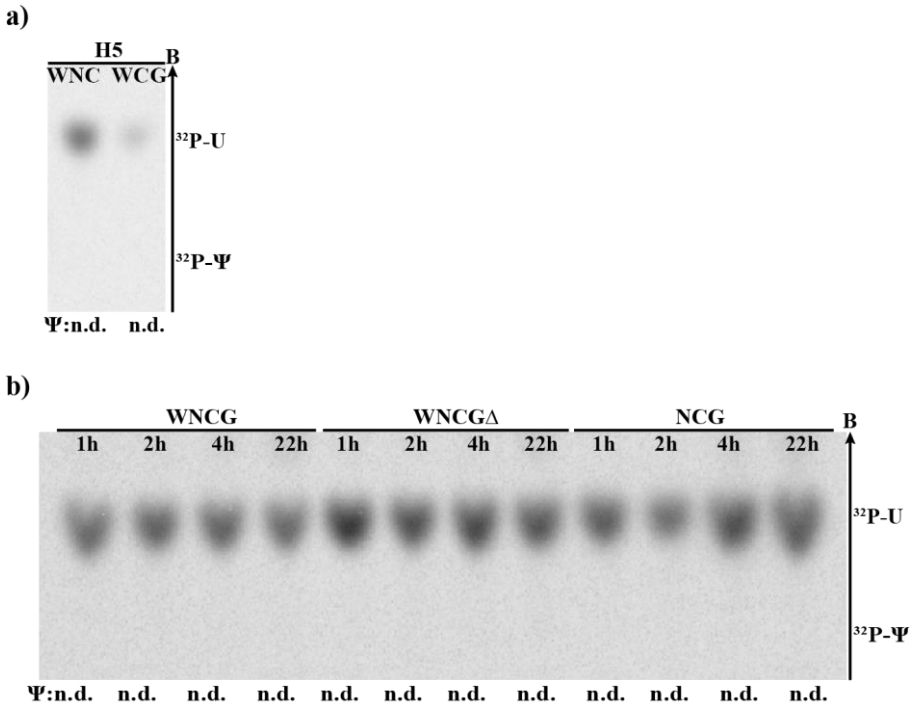


Figure 73: a) Thin layer chromatography of single turnover pseudouridylation activity tests performed with site specific ^{32}P-labeled 5' substrate RNA for 2 h (standard single turnover conditions). No activity can be observed in the absence of either Gar1 or Nop10. b) Thin layer chromatography of multiple turnover pseudouridylation activity tests performed with snoRNA H3Δ and site specific ^{32}P-labeled 3' substrate RNA (standard multiple turnover conditions). No activity can be observed for this construct.

Chapter 4

Functionalization of proteins and spectroscopic analysis of snR81 H/ACA RNP

4.1 Synthesis of non-canonical amino acids

For the generation of fluorophore- or spin-labeled H/ACA RNP, it was necessary to generate an effective labeling scheme to site-specifically label the proteins of the complex. For this purpose, different non-canonical amino acids were synthesized, with the goal to incorporate them into the proteins to act as a reactive functional group. Utilizing click chemistry, the non-canonical amino acids can be used to attach spectroscopic labels onto the proteins to generate labeled RNP complexes.
While the basic building blocks of the genetic code includes the 20 canonic amino acids as well as the 2 proteinogenic amino acids selenocysteine and pyrrolysine, over 800 non-proteinogenic amino acids have been found in nature, fulfilling various functions[176,177,192–194,197,198]. Of these non-proteinogenic amino acids, over 140 have been found in natural occurring proteins[196]. However, these non-proteinogenic amino acids result from post-translational modifications rather than translational incorporation. With the functionalization of an amber codon via orthogonal tRNA/aminoacyl-tRNA synthetase pairs, it is possible to incorporate a non-canonical amino acid into the peptide chain of a protein translationally[175]. These non-canonical amino acids mostly contain a particular functional group at their side chain. While the core structure of the protein is most often kept intact, the functionalization of the side chain can be used to directly act as a spectroscopic probe or for post-translational modifications.
Most of the non-canonical amino acids are prepared by the modification of natural occurring amino acids. For the amino acids that were utilized in this thesis, the canonical amino acid L-lysine acted as the precursor. The chemical functionalization was added on the side chain amino group of lysine, while the amino group at the β-carbon atom was protected by a tert-butoxycarbonyl (BOC) group.

4.1.1 Synthesis of propargyl-lysine (PrK)

Propargyl-lysine (**1**) was synthesized by the reaction of BOC protected L-lysine (**4**) with propargyl chloride (**5**) in a 2-step reaction according to literature[199]. The high electronegativity of the chlorine atom results in a negative inductive effect on the propargyl chloride, which allows a nucleophilic substitution of the amino group on the ε-carbon atom of the lysine side chain, with Cl^- acting as a potent leaving group. The released HCl from the reaction is neutralized by the basic reaction milieu to prevent an early BOC deprotection reaction. This way, it is ensured that during workup of the reaction mixture, the intermediate product (**6**) remains in the organic phase. After workup, the deprotection reaction to release the BOC protecting group was performed with concentrated HCl and the product was obtained as the HCl salt of propargyl-lysine (**1**) after lyophilization in an overall yield of 97% (Figure 74).

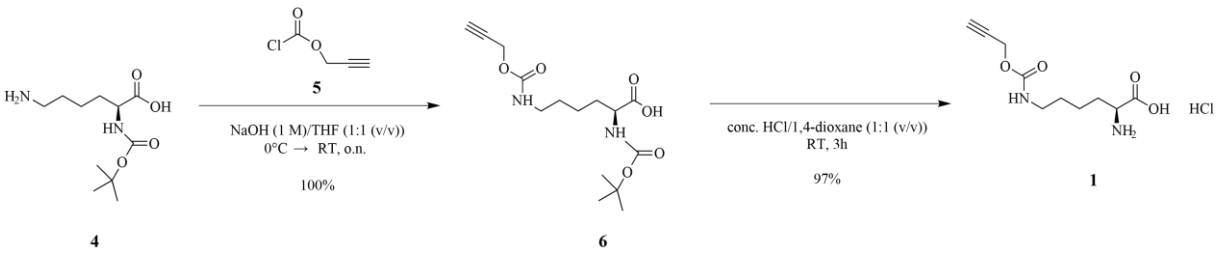


Figure 74: Synthesis scheme of propargyl-lysine

4.1.2 Synthesis of norbornene-lysine (NoK)

Norbornene-lysine (**2**) was synthesized in a 2-step reaction according to literature[201]. A carbonyl group from triphosgene (**8**) was used as a linker between the lysine and the norbornene moiety. The high reactivity of triphosgene towards moisture made it necessary to perform the reactions under argon atmosphere and to use dry solvents. The usage of triphosgene prevents the need to directly use phosgene as an educt, however still phosgene is produced during the reaction, so special care to prevent leakage of phosgene gas had to be taken. First, 5-norbornene-2-methanol (**7**) was added to a solution of triphosgene in THF. In this way, the hydroxy group of norbornene-2-methanol could perform a nucleophilic substitution on the carbonyl-group of triphosgene. Afterwards, the product from this reaction was added to a solution of BOC protected L-lysine (**4**), to perform a second nucleophilic substitution of the amino group on the ε-carbon atom of the lysine side chain with the carbonyl-group of triphosgene. The reaction was performed in a basic milieu to prevent an early deprotection reaction of the BOC group and was carried out overnight. When the pH was checked at the next day, the reaction milieu had turned acidic. The reaction mixture was worked up nonetheless and column chromatography was performed to purify the BOC protected norbornene-lysine (**9**), however only a yield of 17% for this first reaction step could be achieved. One probable explanation is that because the reaction solution turned acidic overnight, the BOC group was removed early and a considerable amount of the product was lost in the aqueous phase during the workup of the reaction mixture. An explanation for the pH change was probably that some moisture had entered the reaction mixture and caused a reaction of excess triphosgene with water to HCl. Triphosgene can theoretically react with three molar equivalents of the 5-norbornene-2-methanol, since phosgene is produced *in situ* during the reaction. The reaction conditions were changed in two ways to achieve a higher yield of product. In the beginning, only 0.7 molar equivalents of triphosgene was used for the start of the reaction and the rest was added later. Also, after the addition of the BOC protected L-lysine to the intermediate product, additional NaOH was added to keep a basic milieu overnight. These improvements led to an increase in yield for the first reaction step to 79%. After workup and column chromatography of the reaction mixture, the deprotection reaction was performed with formic acid in chloroform. After removal of all volatile components, the product was taken up in a small amount of 50 mM HCl and lyophilization was performed to obtain the HCl salt of the norbornene-lysine (**2**) in an overall yield of 59% (Figure 75).

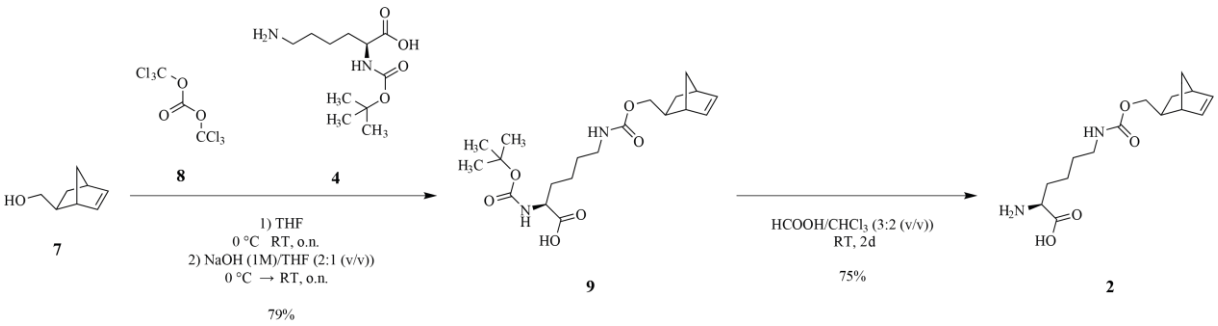


Figure 75: Synthesis scheme of norbornene-lysine

4.1.3 Synthesis of strained cyclooctyne-lysine (SCOK)

Strained cyclooctyne-lysine (**3**) was synthesized in a 6-step reaction which was described in detail during my master thesis[359]. In brief, the construction of the cyclooctyne ring moiety was performed with cycloheptene (**10**) as a starting material. KO^{t}Bu activated bromoform was reacted with cycloheptene to form the halocarbon-adduct 8,8-dibromobicyclo[5.1.0]octane (**11**). Through silverperchlorate mediated hydrolysis, the eight-ring was formed with 2-bromo-*trans*-cyclooct-2-enol (**12**) as the product. This was brought to reaction with DBU to eliminate HBr and form 2-cyclooctyne-1-ol (**13**). The carbonyl linking group was inserted by a reaction of 2-cyclooctyne-1-ol with 4-nitrophenyl-chloroformate (**14**) and as with the other non-canonical amino acid synthesis, BOC protected L-lysine (**4**) was added to the carbonyl linking group of the intermediate product (**15**) via nucleophilic substitution to obtain the BOC protected product (**16**). Deprotection reaction was performed with formic acid in chloroform. After removal of all volatile components and lyophilization, the product was obtained as the formate salt of the strained cyclooctyne-lysine (**3**) in an overall yield of 7% (Figure 76).

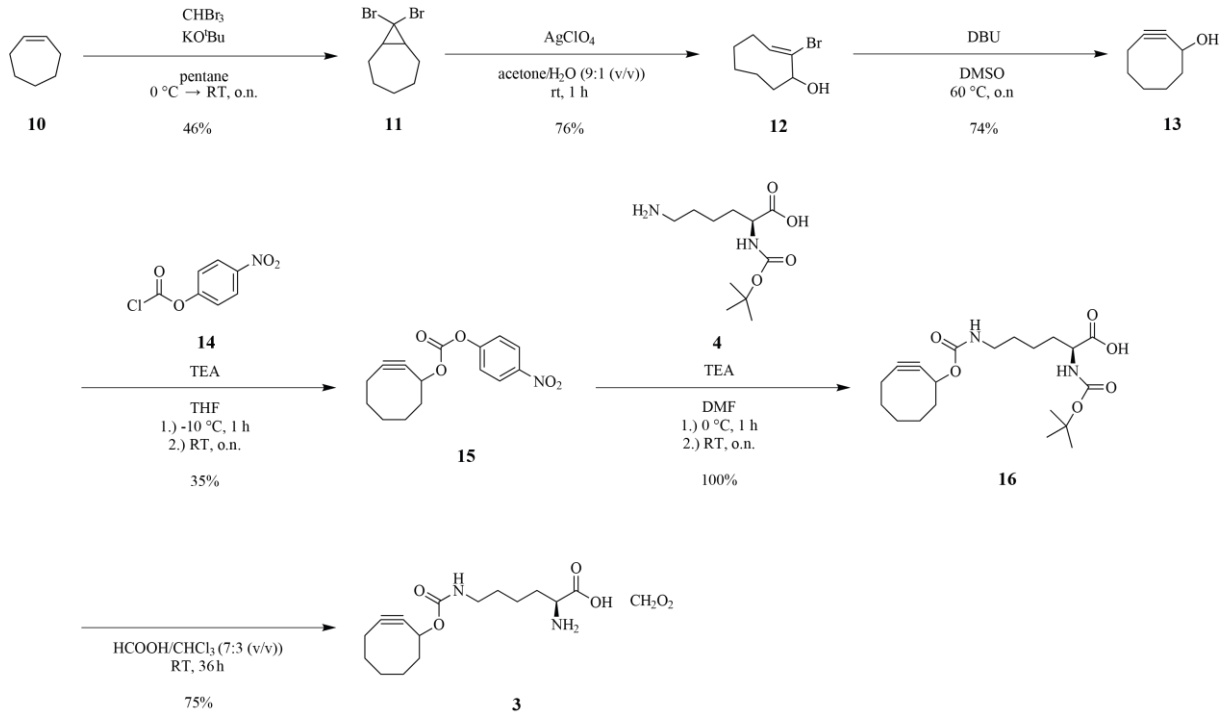


Figure 76: Synthesis scheme of strained cyclooctyne-lysine

4.2 Preparation of fluorophore labeled Nhp2 via copper catalyzed click chemistry

4.2.1 Incorporation of PrK into Nhp2

For a closer dissection of the snR81 RNP complex, it was desired to attach spectroscopic probes to the different RNP building blocks. To gain a more detailed insight into protein-protein or protein-RNA interaction, labeling of the proteins with fluorophores was performed to gain constructs for single molecule FRET spectroscopy. While label attachment is also possible via the utilization of natural occurring amino acids, e.g. cysteine-labeling, this method also has some shortcomings. For once, for site specific single labeling, only a single cysteine has to be present in the protein. If the protein contains other native cysteines, they have to be removed beforehand via site directed mutagenesis. This procedure can easily disrupt the protein structure and native folding, preventing any further analysis. With the site-specific incorporation of a non-canonical amino acid, only a single position in the peptide chain has to be modified. Furthermore, with this method an amino acid with varying reactive groups at the amino acid side chain can be incorporated at a desired position, which widens the scope of possible spectroscopic probes that can be linked to the amino acid. First non-canonical amino acid incorporation tests were done with the non-canonical amino acid propargyl-lysine (PrK) into the protein Nhp2. This protein was chosen because it seemed to be the most stable protein during purification procedures and it was readily overexpressed, resulting in very high yield. Additionally, this protein could easily be generated without co-expression of other proteins. To incorporate the non-canonical amino acids, an amber codon suppression system was utilized. An amber codon (TAG) was inserted into the plasmid at the desired position with complementary primer-based site directed mutagenesis. For Nhp2, the

TAG mutation was done at positions K37, K48, K54, G70, K107 and E152 (Figure 77). Those positions were chosen both on the NMR structure of Nhp2 (pdb: 2LBW) in comparison to its supposed binding sites towards Nop10 and the snoRNA[362] as well as previous successful labeling of the structurally highly similar L7Ae (archaeal counterpart of Nhp2) in the Hengesbach group[359–361].

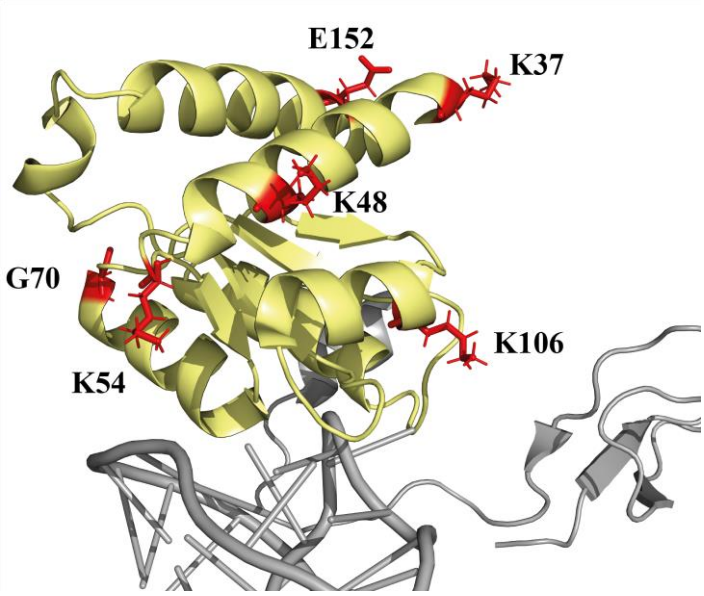


Figure 77: NMR structure of Nhp2 (pdb: 2LBW). Mutation sites K37, K48, K54, G70, K107, and E152 are marked in red. *S. cerevisiae* Nop10 (pdb:3U28) and *P. furiosus* snoRNA (pdb: 3HAY) are shown in gray to demonstrate potential binding sites towards Nhp2.

The arabinose-inducible plasmid pEVOL (WT) was co-transformed with the Nhp2 containing pET24b plasmid into E. coli BL21 (DE3) cells. The pEVOL plasmid is responsible for the generation of amber codon recognizing tRNA as well as aminoacyl tRNA synthetase (aaRS) from *Methanosarcina mazei*, responsible for loading the PrK onto the amber codon recognizing tRNA[199]. Cell colonies containing both plasmids were selected against kanamycin and chloramphenicol. Test expressions were done in 3 ml volume and cells were incubated at 37 °C in TB medium. After one hour, 5 mM of PrK (in DMSO) was added to the cells and cell growth was continued for an additional hour. At an OD_{600} of 0.5-0.9, arabinose was added to the cultures. Production of tRNA and aaRS was undergone for an additional five hours to ensure an overabundance of those amber codon recognizing components. Even with the amber suppression in place, it is still possible for the ribosomal machinery to occasionally recognize the amber codon as an actual stop codon and terminate the translation, which could result in a fraction of the peptide being truncated. Afterwards, expression cultures were cooled on ice, 0.1 mM IPTG was added to start protein expression and the cultures were shaked for several days at 20 °C. Every day, a sample was taken from the expression test cultures, the cells in this sample were pelleted and dissolved in an amount of 2xSDS loading buffer according to the OD_{600} of the expression culture. Samples were then separated by SDS PAGE and analyzed with Western Blotting. With Western Blotting, the amount of protein per

samples could be visualized by anti-HisTag based staining. And since truncated protein would not contain the C-terminal HisTag, this way only HisTag containing protein was visualized. While constructs G70, K107 and E152 only showed small amount of protein produced after several days, the constructs K37, K48 and K54 showed promising amounts of protein compared to the WT protein, which was also added onto the Western Blot for comparison (Figure 78). The WB band intensities were plotted and showed, that after approximately 3-4 days, the highest yield of protein was reached, and that additional incubation/expression time only resulted in a very small increase in yield. Compared to WT protein, approximately 60 - 80% protein expression was achieved with the non-canonical amino acid incorporation method.

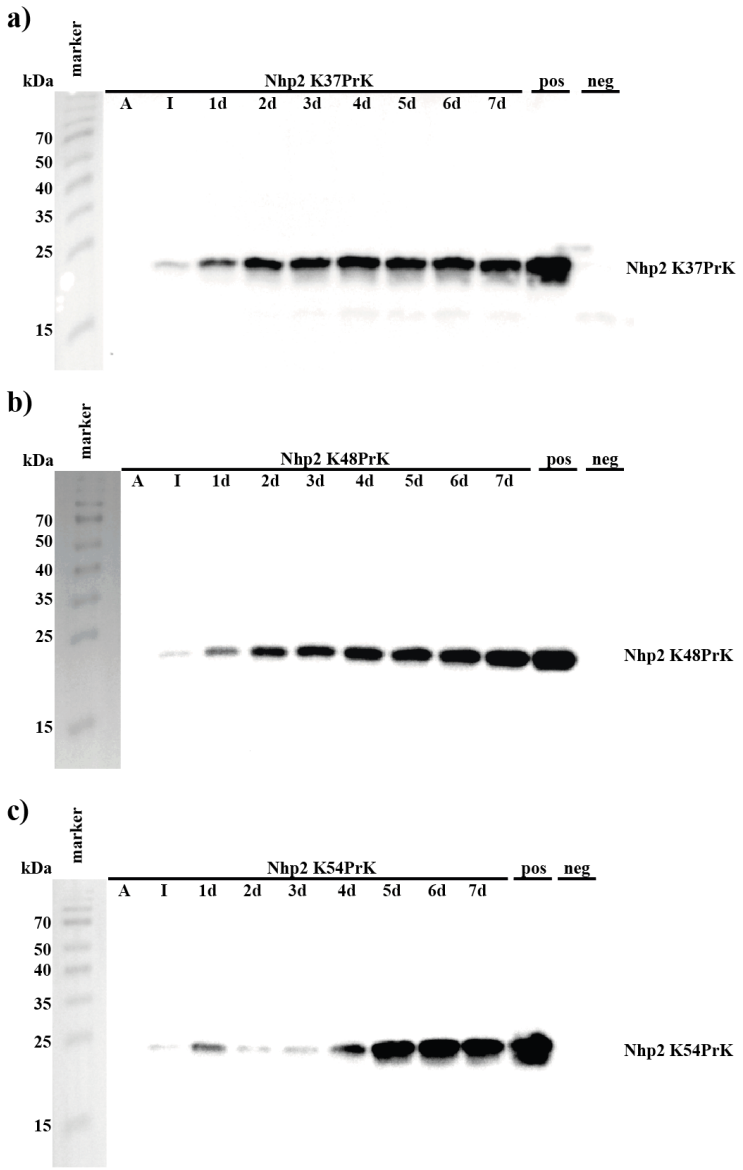


Figure 78: Expression test of PrK incorporation into the Nhp2 constructs K37 (a), K48 (b) and K54 (c). Marker: Spectra Multicolor Broad Range Protein Ladder (Thermo Scientific). Positive control: Nhp2. Negative control: no addition of PrK into expression medium.

After the successful expression tests, expression of the constructs K37, K48 and K54 was done in a larger scale to potentially purify these constructs for further labeling experiments. Expression was carried out in 250 ml TB medium. The PrK concentration was lowered to 2.5 mM, based on the assumption that this amount would be sufficient for an adequate yield[361] and to save some PrK, considering the high synthesis cost of the compound. Expression was carried out with the parameters used during expression tests. After the initial seven hours of cell growth, before the addition of IPTG, cultures in TB medium reached an OD_{600} of 7-8. After the cultures were supplemented with 0.1 mM IPTG, the expression of protein constructs was carried out over the weekend for 3 days at 20 °C. Afterwards, the cells were harvested, shock frosted in liquid nitrogen and stored at -80 °C. Cell lysis was carried out as described before for wildtype Nhp2. Purification was performed with affinity chromatography and fractions analyzed with SDS PAGE (Figure 79). The protein containing fractions were pooled and supplemented with RNase A. After incubation overnight at room temperature, RNase A was removed with a second affinity chromatography (Figure 80) and size exclusion chromatography was performed to further purify the protein constructs (Figure 81). For all the constructs, the two fractions with the highest purity were pooled and concentrated to a volume of 100 µl. All constructs could be obtained in good yield with 697 µg for K37, 628 µg for K48 and 741 µg for K54 out of the 250 ml expression culture. All samples showed an 260/280 ratio of 0.55-0.58, indicating the quantitative removal of all endogenous RNA in the samples. PrK incorporation was verified by mass spectrometry (see appendix, Figure 132).

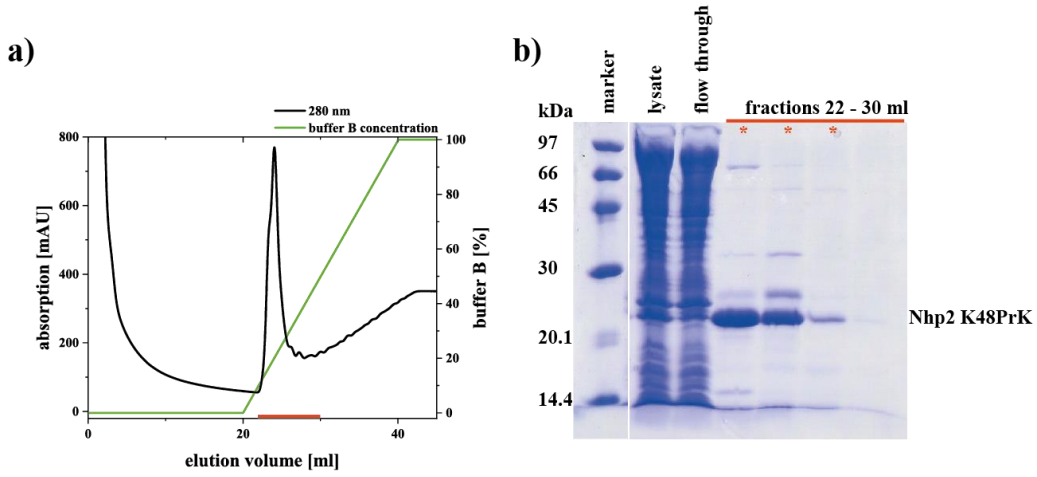


Figure 79: Affinity chromatography purification of Nhp2K48PrK before RNase A treatment. a) Ni-NTA elution chromatogram using an imidazole gradient. Fractions used for analysis are marked with a red bar. b) PAGE analysis of fractions from Ni-NTA. Fractions used for further purification are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

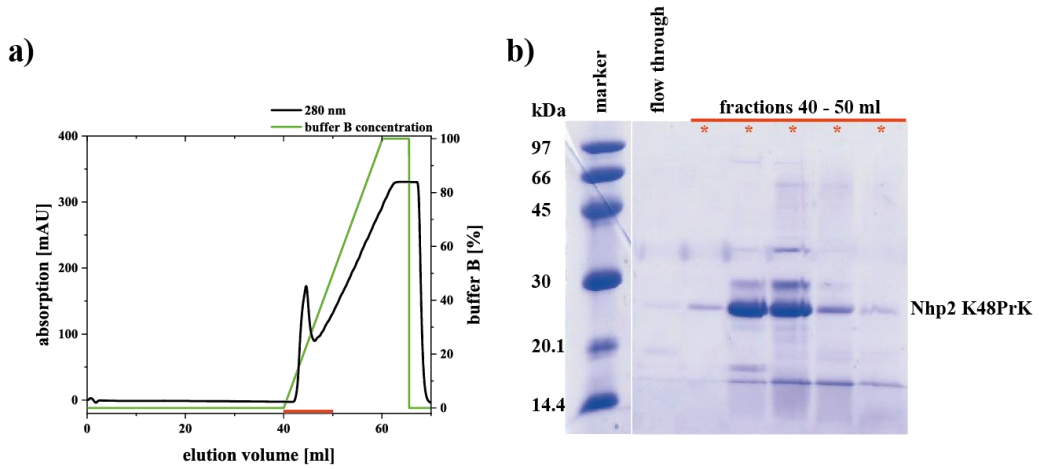


Figure 80: Affinity chromatography purification of Nhp2 K48PrK after RNase A treatment. a) Ni-NTA elution chromatogram using an imidazole gradient. Fractions used for analysis are marked with a red bar. b) PAGE analysis of fractions from Ni-NTA. Fractions used for further purification are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

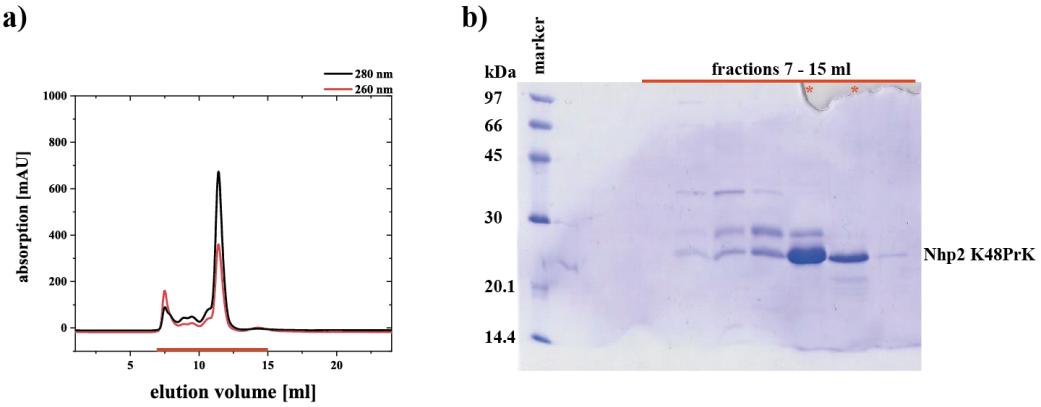


Figure 81: Size exclusion chromatography purification of Nhp2 K48PrK after RNase A treatment. a) SEC elution chromatogram using a Superdex 75 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) PAGE analysis of fractions from SEC. Fractions used for further experiments are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

An EMSA was performed to check binding capacity of the constructs K37 and K48 in comparison to Nhp2 with the H5 RNA (Figure 82). Protein was added to the RNA in increasing amounts and samples were incubated at 30 °C for 30 min before the EMSA was performed. One sample contained only RNA as a control. The incubation at 30 °C shifted the RNA to a second folding conformation, as it was observed before. Interestingly, with the addition of an increasing amount of all Nhp2 constructs, the RNA is shifted back to its major conformation. Nhp2 either seems to not directly bind to the RNA on its own in equimolar concentrations but rather shifts its conformation back into the correct hairpin conformation, or Nhp2 binds to the RNA (unspecifically), which leads to the RNA adapting a conformation

similar to the conformation of the free RNA. Possibly, the RNA contains a recognition motif for Nhp2 which lets the RNA adapt the correct folding in the presence of Nhp2. The affinity of Nhp2 seemed to be a bit higher to H5 than for the K37 and K48 constructs. However, at an excess of protein to RNA of 2:1, the samples looked the same. At a high excess of Nhp2 constructs to RNA of 6:1, the protein clumped together with the RNA and the complex was not able to migrate into the gel anymore, the same behavior that could be observed for the NCG constructs. No degradation bands are visible, which shows no RNAse contamination of the protein.

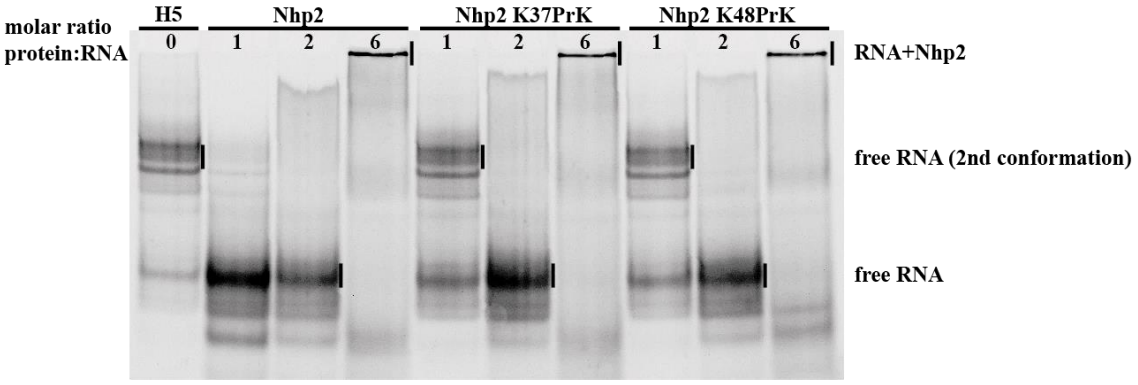


Figure 82: EMSA of H5 snoRNA with Nhp2, Nhp2 K37PrK and Nhp2K48PrK.

In an effort to optimize the yield of protein containing a non-canonical amino acid compared to the actual amount of non-canonical amino acid used for the expression, a test was performed for construct K37. The expression was carried out in 500 ml TB medium, which was supplemented with only 1.25 mM PrK. The expression and purification was done according to protocol and afterwards, a yield of 1.11 mg of protein could be obtained, so basically 60% more yield for the same amount of PrK used. This shows, that a PrK concentration of 1.25 mM is sufficient for complete incorporation of the non-canonical amino acid during expression, and it was decided to use these expression conditions for further non-canonical amino acid incorporation expressions.

4.2.2 Fluorophore labeling of Nhp2 with copper catalyzed click chemistry

The incorporation of propargyl-lysine into Nhp2 added a reactive alkyne group to the protein, which made it possible to use azide-containing spectroscopic probes to perform azide-alkyne cycloaddition reactions. These “click” reactions were performed with the use of a copper catalyst. Copper(II)sulfate was used as a starting material and reduced with ascorbic acid into its copper(I) state. THPTA was used as a ligand to chelate and stabilize the oxidation state of copper(I) to prevent reoxidation. As spectroscopic probe, sulfo-Cy3-azide was used to attach a donor fluorophore label onto Nhp2. Additionally, aminoguanidinium was added to the reaction mixture to prevent reactions with protein side chains[283]. Initial test click reactions were carried out in small scale with construct K37 in a molar ratio of 3:1 of fluorophore dye to protein. Reactions were carried out in the dark at 37 °C since Nhp2 seemed to be relatively

stable at higher temperatures compared to the other proteins and a reaction temperature of 37 °C for the click reaction resulted in good yield for L7Ae[361]. Samples were taken after 3 and 24 hours. SDS PAGE and a subsequent fluorophore scan showed that the click reaction was successful (Figure 83). However, after longer reaction times, some di- and multimers of Nhp2 were visible on the denaturing gel. Coomassie staining revealed that only a small portion of protein formed these side products. Additionally, a negative control experiment with Nhp2 was performed to reveal that the dye attachment was indeed specific and due to click reaction and not some other side reaction resulting in attachment of the fluorophore to the protein.

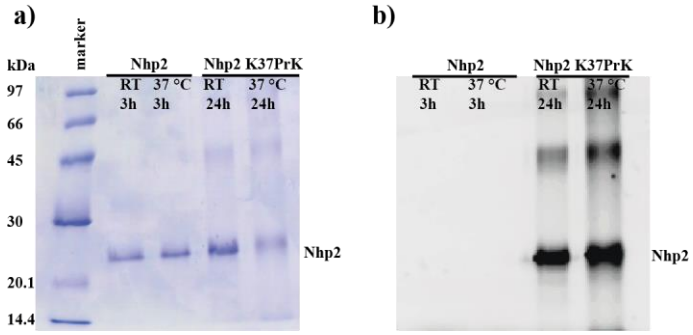


Figure 83: Test click reaction of Nhp2 (negative control) and Nhp2 K37PrK with sulfo-Cy3-azide. Coomassie stain (a) and fluorescence scan (b).

To test if the formation of Nhp2 multimers was caused by prolonged incubation at 37 °C, the reaction times were lowered to 1-3 h and test click reactions were performed for K37 and K48 constructs. For these shorter reaction times, the formation of multimers was minute in comparison to the main product (Figure 84).

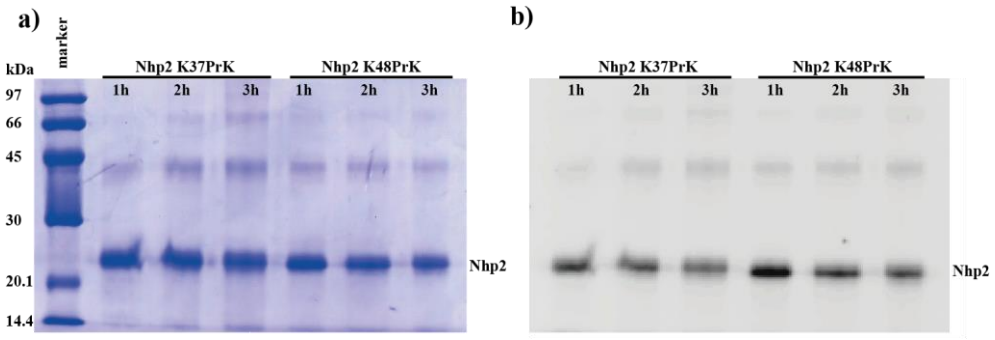


Figure 84: Test click reaction of Nhp2 K37PrK and Nhp2 K48PrK with sulfo-Cy3-azide. Coomassie stain (a) and fluorescence scan (b).

The click reaction with sulfo-Cy3-azide was carried out in a larger scale for constructs K37 and K48. Reaction volume was increased to 110 µl with a fluorophore to protein ratio of 2.5:1. The reaction was carried out at 37 °C for 3 h and subsequently the reaction mixture was

loaded onto a SEC column. Size exclusion was performed to separate the protein from unreacted dye (Figure 85). Notably, fluorophore labeled Nhp2 was eluted earlier from the SEC column (in the exclusion volume) than unlabeled Nhp2, which most probably is caused by the influence of the dye molecule on the hydrodynamic radius of the protein. Both constructs could be obtained in good yield of 100-250 µg for single molecule experiments. Labeling efficiency was determined via UV-vis spectroscopy and could be determined to be 70-80%.

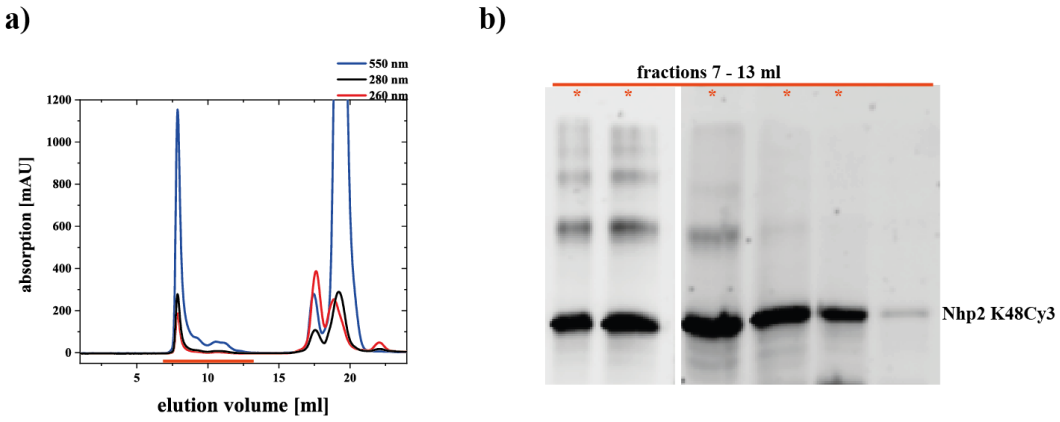


Figure 85: Size exclusion chromatography purification of Nhp2 K48Cy3. a) SEC elution chromatogram using a Superdex 75 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) Fluorescence scan of fractions from SEC. Fractions used for further experiments are marked with a red asterisk.

For a potential investigation of protein/protein interactions or for a reversed RNA/protein labeling scheme, an Nhp2 construct with an acceptor fluorophore label was required. For construct K48 a click reaction with sulfo-Cy5-azide was therefore performed. The construct was purified by affinity and size exclusion chromatography (Figure 86). As observed for the Cy3-labeled constructs, Cy5-labeled Nhp2 was also eluted in the exclusion volume. It was possible to obtain 100 µg of protein with a labeling efficiency of 60%.

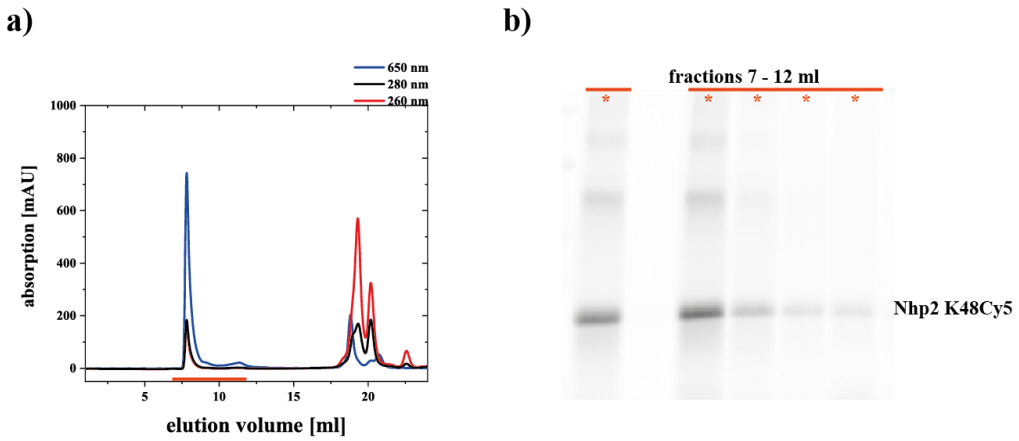


Figure 86: Size exclusion chromatography purification of Nhp2 K48Cy5. a) SEC elution chromatogram using a Superdex 75 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) Fluorescence scan of fractions from SEC. Fractions used for further experiments are marked with a red asterisk.

Folding and RNA binding may be altered by modification and fluorophore labeling. To assess this effect, an activity test of the fluorophore labeled constructs was performed to check if the attachment of the fluorophore interferes with enzymatic activity of the complex. Cy3- as well as Cy5-labeled Nhp2 showed an activity promoting effect similar to non-labeled Nhp2 in comparison to complex without Nhp2, showing that the fluorophores do not reduce the effect of Nhp2 on the pseudouridylation reaction. (appendix, Figure 133).

4.3 Preparation of fluorophore labeled Nhp2 via copper free click chemistry

4.3.1 Incorporation of SCOK into Nhp2

The synthesis of the non-canonical amino acid strained cyclooctyne-lysine (SCOK) was optimized and first incorporation tests on archaeal proteins were performed during a preciding master thesis[359]. The amino acid SCOK bears an alkyne group similar to PrK, making it accessible for click reactions. However, the alkyne group is situated inside a cyclooctane ring, resulting in a strained ring system. The energy from this ring strain is released during the coupling with an azide moiety and serves as energy for the click reaction. Therefore, no copper catalyst is needed for the reaction[200,285–288]. This approach has several benefits compared to the copper catalyzed reaction. The copper reagent can act toxic towards biomolecules, making the copper catalyzed reaction less suitable for *in vivo* experiments. Furthermore, a reducing agent like ascorbic acid, which is required during the copper catalyzed reaction may be fine for spectroscopic probes like fluorophores, but on the other hand reacts with and thus destroys other probes like nitroxide spin labels. The incorporation of the amino acid SCOK was tested with the protein constructs Nhp2W K37 and K48. Since SCOK is sterically very bulky, the synthetase provided by the pEVOL plasmid is not able to provide a binding surface for this non-canonical amino acid. For this reason, the mutant pEVOL^{AF} was used for incorporation of SCOK. In this mutant, the amino acids of the aaRS Y306 and Y384 were replaced by alanine and phenylalanine, respectively, widening the

binding pocket for amino acids on the synthetase and allowing larger amino acids like SCOK to bind to the synthetase and be attached to the tRNA[200]. The expression tests were performed according to protocol, with an SCOK amount of 1 mM supplemented one hour after the start of cell growth at 37 °C, arabinose induction (0.2%) after an additional hour and IPTG (0.1 mM) induction after additional five hours. Protein expression was carried out at 20 °C. Expression tests were analyzed by western blot and showed the successful incorporation of SCOK into both protein constructs (Figure 87).

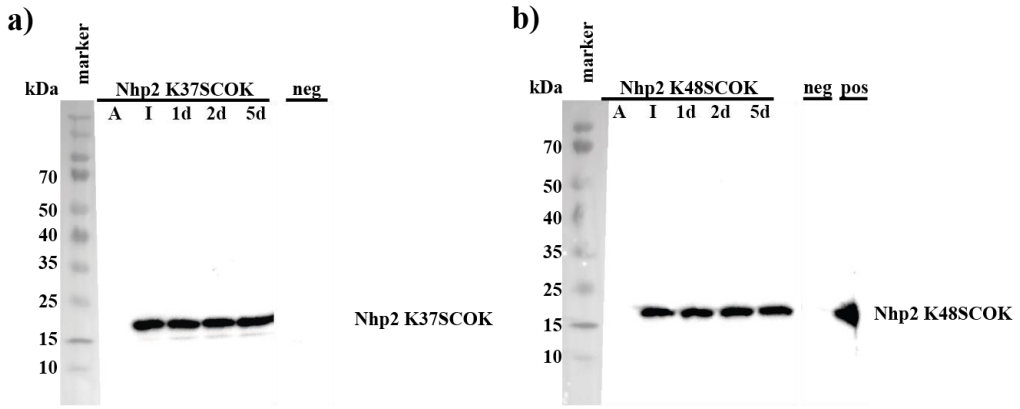


Figure 87: Expression test of SCOK incorporation into the Nhp2 constructs K37 (a) and K48 (b). Marker: Spectra Multicolor Broad Range Protein Ladder (Thermo Scientific). Positive control: Nhp2. Negative control: no addition of SCOK into expression medium.

Expression for SCOK incorporation into construct K48 was scaled up and performed with a 250 ml expression culture. Expression was done with the same parameters as in the expression tests, and cells were harvested after 4 days of expression. Protein was obtained by cell lysis and subsequent affinity chromatography. RNase A treatment was performed as described before and a second affinity chromatography and size exclusion chromatography were performed (Figure 88). Nhp2W K48SCOK could be obtained with a relatively high yield of 3.2 mg of protein in good purity.

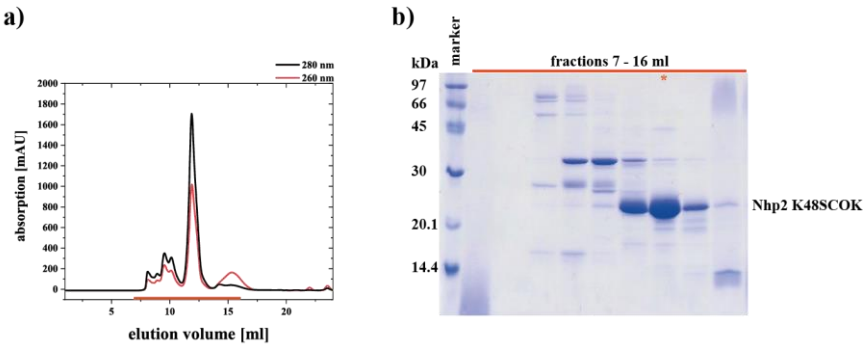


Figure 88: Size exclusion chromatography purification of Nhp2 K48SCOK. a) SEC elution chromatogram using a Superdex 75 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) Coomassie stain of fractions from SEC. Fractions used

for further experiments are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

4.3.2 Fluorophore labeling of Nhp2 with copper free click chemistry

With the purified protein, click test reactions were performed with sulfo-Cy3-azide. Reactions were carried out at room temperature and 37 °C, and samples were collected after several points in time up to 48 h. SDS PAGE analysis of the samples and fluorescence scan of the gel showed, that click reactions were successful (Figure 89). The highest yield was achieved after 48 h, showing that the strain promoted version of the click reaction is slower than the copper catalyzed variant. Overall yield was higher for reactions carried out at 37 °C.

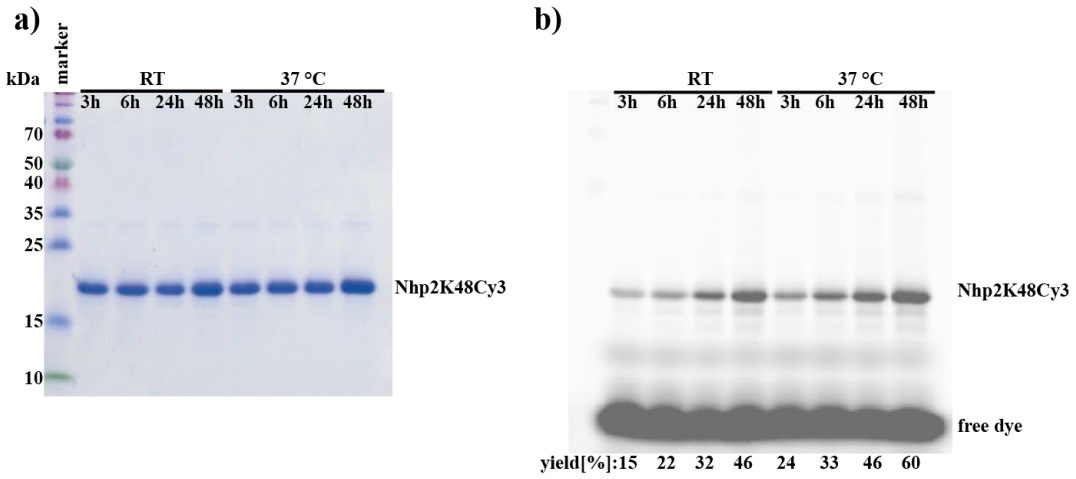


Figure 89: Test click reaction of Nhp2 K48SCOK with sulfo-Cy3-azide. Coomassie stain (a) and fluorescence scan (b). Labeling yields were quantified via ImageJ by the fluorescence intensities of protein and free dye bands.

4.4 Preparation of fluorophore labeled Cbf5

4.4.1 Incorporation of PrK into Cbf5

Cbf5 acts as the catalytic core of the RNP complex, where the amino acid D95 binds the target uridine, that gets transformed into pseudouridine. The protein also contains a “thumb” loop domain, which is believed to position the target uridine close to D95 and thereby facilitate catalytic activity. For archaeal RNPs, this thumb loop could be observed in different states with crystallography and its function was also dissected via smFRET spectroscopy in the Hengesbach group[360]. For eukaryotic Cbf5 however, it was not possible to resolve a crystal structure of the thumb loop domain, possibly due to high flexibility of this part of the protein[160]. Several sites for incorporation of non-canonical amino acids were chosen inside or in close proximity of the thumb loop domain (Figure 90). Namely, those mutations sites were P154, L156, I157, K161 and V166. All mutations sites were chosen to replace amino acids that have already been shown to not be essential for enzymatic activity of Cbf5[379].

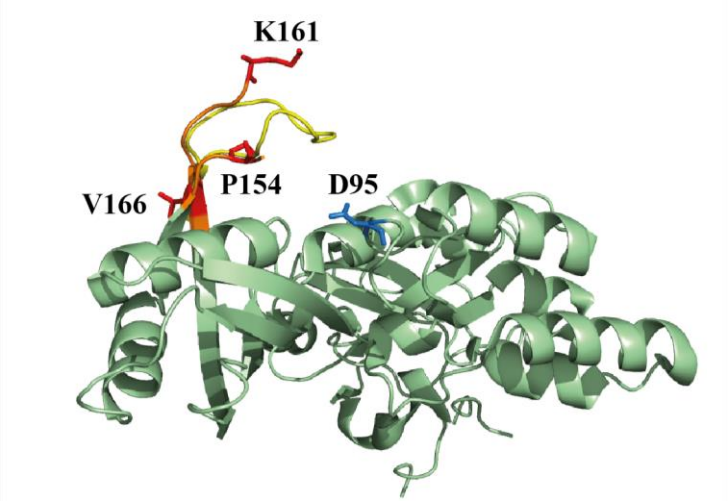


Figure 90: Crystal structure of Cbf5Δ (pdb: 3U28). The thumb loop domain, which is only partly resolved in the crystal structure, is marked in orange and overlayed with the structurally similar thumb loop domain from the crystal structure of archaeal Cbf5 from *pyrococcus furiosus* (pdb: 3HAY). Labeling sites P154, K161 and V166 are marked in red (Labeling sites L156 and I157 are contained in the unresolved part of the structure). The catalytic core residue D95 is marked in blue.

Several constructs of Cbf5 as well as Cbf5Δ were generated via site directed mutagenesis. All constructs also contained the Nop10 gene on the plasmid. Expression tests were carried out in small scale with the expression conditions that had already been established for non-canonical amino acid incorporation into Nhp2 (1.25 mM PrK, 0.2% arabinose, 0.1 mM IPTG). Samples were taken after 24 h, 48 h and finally after one week of expression at 20 °C. Expression tests were analyzed by SDS PAGE and western blot (Figure 91). The expression with PrK incorporation worked for all of the constructs tested except K161.

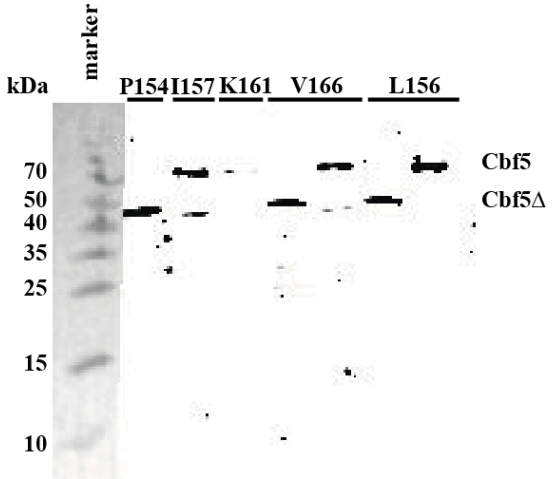


Figure 91: Expression test of PrK incorporation into Cbf5 and Cbf5Δ constructs. Comparison of samples taken after expression of one week. Marker: Spectra Multicolor Broad Range Protein Ladder (Thermo Scientific).

Since some constructs showed secondary bands in the western blot, the construct L156 was chosen to be the most promising construct for further expression in larger scale. Both Cbf5 L156 and Cbf5Δ L156 were expressed in a larger scale of one liter expression culture volume. Gar1 was expressed separately and before cell lysis, the pellets of both the NC and the G expression cultures were mixed together for co-purification of the ternary NCG complex. For both Cbf5 constructs expression was successful and the constructs were purified by affinity and subsequent size exclusion chromatography. SDS PAGE analysis showed that it was possible to obtain both NCG constructs, however both constructs still were contaminated with other proteins. The sample contained about 10% w/w contaminating RNA as revealed by a 260/280 ratio of 1.56 for Cbf5 L156/Nop10/Gar1 and 1.04 for Cbf5Δ L156/Nop10/Gar1. A total yield of 2.7 mg for Cbf5 L156/Nop10/Gar1 and 1.7 mg for Cbf5Δ L156/Nop10/Gar1 could be obtained from the 1 l expression. However, these values do not account for the impurities. The construct Cbf5Δ L156 was expressed again from a two liter culture. Before cell lysis, a pellet from a Gar1 expression was added. The co-purification of the ternary complex was performed as before. However, after the first affinity chromatography, the protein containing fractions were pooled and RNase A was added. The fractions were incubated at 4 °C for 18 h and RNase A was removed with a second affinity chromatography. After size exclusion chromatography, the protein containing fractions were analyzed via SDS PAGE Coomassie blue staining (Figure 92). This time, protein could be obtained in better purity, however still one endogenous protein band could not be removed via SEC. Endogenous RNA could be removed however, as revealed by a 260/280 ratio of 0.66. A total yield of 262 µg of purified protein could be obtained from the two liter expression.

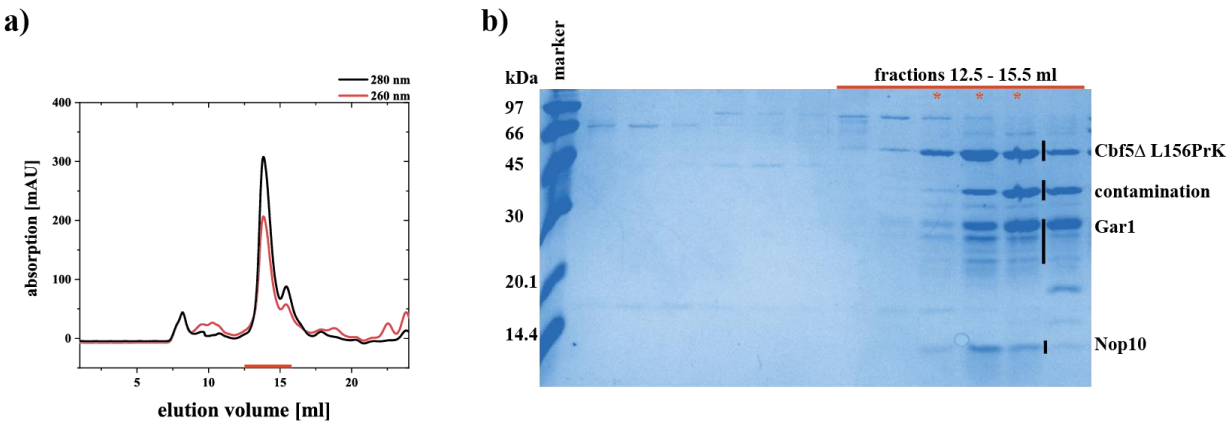


Figure 92: Size exclusion chromatography purification of Cbf5Δ L156PrK together with Nop10 and Gar1. a) SEC elution chromatogram using a Superdex 200 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) PAGE analysis of fractions from SEC. Fractions used for further experiments are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

4.4.2 Fluorophore labeling of Cbf5 with copper catalyzed click chemistry

Click test reactions were performed with the Cbf5 L156PrK and Cbf5Δ L156PrK samples which contained impurities, in order to verify dye labeling, while not using up the small

available amount of pure Cbf5Δ L156PrK. Despite the proteins being not entirely pure, fluorophore labeling conditions could still be optimized with these samples. Click reactions were carried out a 4 °C, 25 °C and 37 °C and samples were taken after 2 h, 4 h, 6 h and 24 h. Samples were analyzed by SDS PAGE Coomassie stain and fluorescence scan (appendix, Figure 135). A reaction temperature of 37 °C was not suitable for the proteins, since degradation was observed and only a smaller amount of labeled protein is visible on the fluorescence scan. Also, a reaction time of 24 h seemed not suitable, since the apparent labeling yield seemed to be actually lower. A reaction temperature of 4 °C with a reaction time of 6 h seemed to be the best reaction conditions for click reaction of Cbf5 constructs. With the purified Cbf5Δ L156 a click reaction with sulfo-Cy3-azide was performed with a dye to protein ratio of 2.5:1. The click reaction was incubated at 4 °C for 4 h. Afterwards, unreacted fluorophore was removed with size exclusion chromatography and protein containing fractions were analyzed by SDS PAGE fluorescence scan (Figure 93). Labeling reaction was successful for both constructs, resulting in 56 µg of labeled and purified Cbf5Δ L156/Nop10/Gar1.

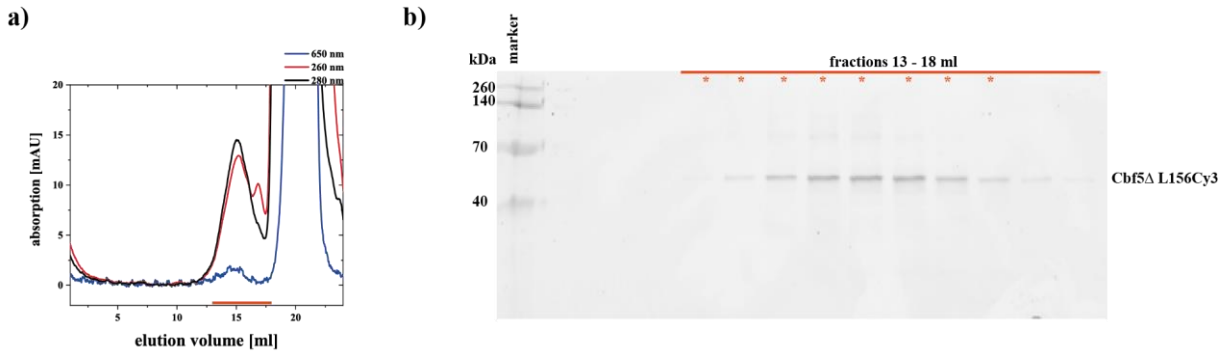


Figure 93: Size exclusion chromatography purification of Cbf5Δ L156Cy3 with Nop10 and Gar1. a) SEC elution chromatogram using a Superdex 200 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) Fluorescence scan of fractions from SEC. Fractions used for further experiments are marked with a red asterisk.

4.5 Preparation of fluorophore labeled Gar1

4.5.1 Incorporation of PrK into Gar1

To examine interactions of the protein Gar1 with Cbf5, Nhp2 or the snoRNA, it was also desired to generate fluorophore labeled constructs of this protein. Especially the interactions of Gar1 with the Cbf5 thumb loop would be an interesting target for investigations, as Gar1 has been shown to act as a mediator between the closed and open state of the thumb loop domain[140,150,157,360]. For fluorophore probe attachment sites, several amino acids on the proteins surface were chosen as suitable targets to be replaced with PrK. Those amino acids namely were N71, D96, G97, K108 and K115 (Figure 94).

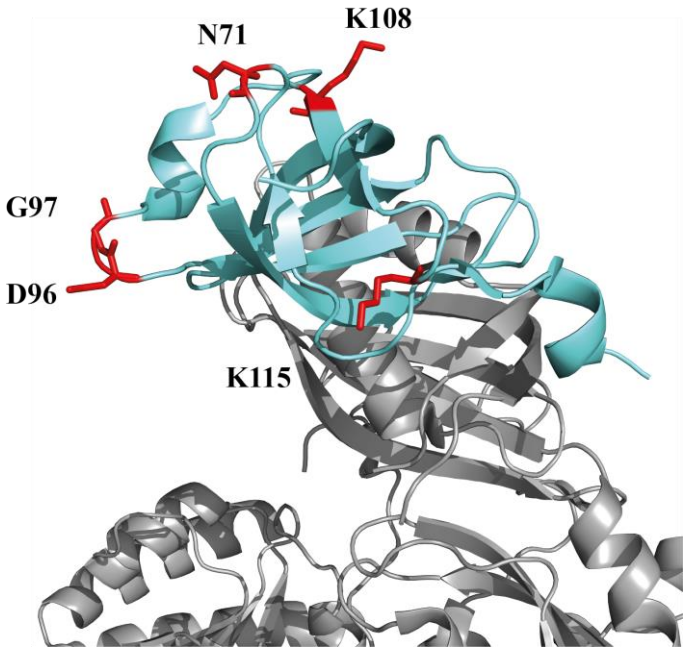


Figure 94: Crystal structure of Gar1Δ (pdb: 3U28). Mutation sites are marked in red. The other RNP proteins are shown in grey.

As before, amber stop codons were inserted via site directed mutagenesis into the gene sequence of Gar1 in pET24b vector. In this vector, Gar1 contains the His-tag at the C-terminus, which makes it easier to validate the incorporation of the non-canonical amino acid. If no amino acid is incorporated at the amber location, protein expression is stopped and the C-terminal His-tag is not included in the truncated protein, resulting in no western blot signals as well as no binding to His-Trap columns during purification. Furthermore, as was shown during the expression of shortened Gar1 constructs (chapter 3.4.3), the degradation during Gar1 purification seems to take place in the C-terminal GAR domain of the protein. A C-terminal His-tag ensures that no degradation products are purified together with the full length protein. For constructs N71 and K108 expression tests were performed under standard conditions. Cells were incubated at 37 °C, PrK (1.25 mM) was added after 1 h and arabinose (0.2%) was added after an additional 1 h. After additional 5 h, expression of target protein was initiated with IPTG (0.1 mM) at 20 °C for several days. Samples of cell lysate were taken after 24 h and after 4 days. SDS PAGE and western blot analysis showed successful protein expression. However, protein was already strongly expressed before the addition of IPTG, showing a significant basal expression of the constructs (appendix, Figure 136a). It was tried to minimize the basal expression by two different approaches. One approach was to incubate the expression cultures at 37 °C until the addition of arabinose. Afterwards, the incubation was continued at 20 °C until the addition of IPTG and expression was carried out at 20 °C. Alternatively, glucose (2%) was added to the expression cultures. While the glucose inhibited inducible expression of protein, the lower temperatures helped to prevent basal expression (appendix, Figure 136b). Constructs N71 and K108 were then expressed in larger scale with 500 ml culture volume. After affinity chromatography and SDS PAGE gel analysis of the fractions, it became apparent that protein yield was extremely low. After RNase A treatment and the second affinity chromatography and size exclusion chromatography, only construct

N71 could be obtained in a very small amount. SDS PAGE analysis of the cell debris after cell lysis showed, that most of the expressed protein was contained in the cell pellet rather than in the lysate. The incorporation of the non-canonical amino acid into Gar1 seems to result in some sort of misfolding of the protein, which results in precipitation of most of the protein after cell lysis, opposite to wildtype Gar1. To test if changes in the lysis buffer could somehow solve this problem, Gar1 N71PrK was expressed in a small scale expression with the standard parameters and afterwards, cells were harvested in 200 µl aliquots. Several buffers were prepared, with different pH-values and various additives like lysozyme, EDTA and urea. Each pellet was dissolved in 100 µl of a different buffer and the lysis in this small scale volume was performed by several freeze-thaw cycles, which were performed by freezing the cell solution with dry ice and afterwards thawing it up to 42 °C several times. The lysate was washed with acetone and lysate as well as pellet were mixed with SDS loading buffer and SDS PAGE was performed and analyzed with western blot. Only the addition of urea led to a visible protein band in the lysate, suggesting some sort of misfolding of the Gar1 protein upon non-canonical amino acid incorporation (appendix, Figure 137).

4.6 Generation of fluorophore labeled snoRNAs

Fluorophore labeled RNA constructs were provided by Gerd Hanspach (Goethe University) [371]. Attachement of Cy5 label was performed by the introduction of an aminoallyl group at positions U61 for H5 and U26 for H5Δ (Figure 95a). For H3 and H3Δ, the native C21 was replaced by U21, since insertion of functional groups works more reliable for uridine[327]. This point mutation leading to a non-canonical U-G pair instead of the canonical C-G pair was deemed to be non-disruptive to the secondary RNA structure, since it is flanked by two C-G pairs on both sides. An aminohexylacrylamino group was introduced at position U21 for both H3 and H3Δ (Figure 95b). The RNA was ordered in separate oligonucleotides, the labeling was performed by coupling of Cy5-NHS-ester with the primary amine groups of the modified oligonucleotides, the labeled oligonucleotides were separated from unreacted dye via HPLC purification and the dye labeled oligonucleotides and the unmodified oligonucleotides were splint ligated with T4 DNA ligase or T4 RNA ligase 2. Purification of the ligated products was performed by preparative PAGE, the product bands were excised and eluted from the gel and precipitated afterwards[371]. The Cy5-labeled RNA constructs were snap cooled prior to use in smFRET experiments (see appendix, Figure 138).

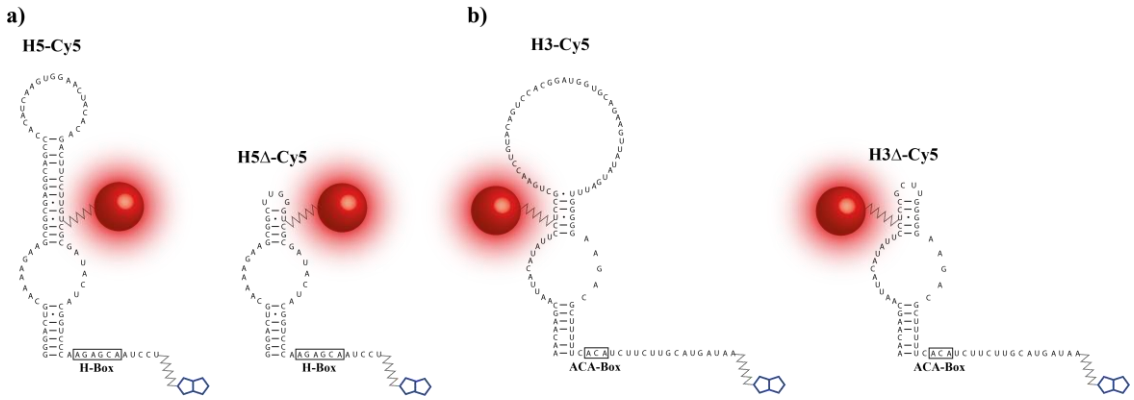


Figure 95: schematic representation of fluorophore labeled snoRNA constructs used for smFRET experiments. a) H5-Cy5 (U61) and H5Δ-Cy5 (U26) constructs. a) H3-Cy5 (U21) and H3Δ-Cy5 (U21) constructs. Cy5-label are marked in red. Biotin is marked in blue. Linkers are marked in grey.

4.7 Investigation of snR81 RNP via smFRET spectroscopy

4.7.1 Sample preparation and processing of data

Samples for the single molecule FRET analysis were prepared with 600 nM snoRNA and a tenfold excess of proteins. The snoRNA constructs used for the smFRET experiments contained a biotin-linker at the 3' end. The passivated objective slides, on which the measurement channels for smFRET experiments were prepared, were also biotinylated. Prior to the addition of the samples into the measurement channels, streptavidin was added to allow the effective pull-down of snoRNA via biotin-streptavidin-biotin interaction and therefore the immobilization of samples on the glass surface. After reconstitution of the samples in Ψ-buffer, samples were highly diluted to a concentration of 0.25 - 1 nM. To see when an ideal particle density of fluorophore labeled samples was reached and the channel was not overloaded with fluorophore labeled sample, the loaded slides were checked on the microscope and loading of sample was adjusted accordingly. Prior to the smFRET measurements, an oxygen scavenging system was added to the measurement channel and the smFRET measurements were performed at room temperature. Activity tests of unlabeled snR81 RNP at room temperature were performed by Gerd Hanspach (Goethe University) and showed only a minor reduction of activity in comparison to activity tests at 30 °C. Activity tests with dye labeled proteins (appendix, Figure 133) as well as dye labeled RNAs[371] were performed. Fluorophore labeled proteins did not result in decreased activity and fluorophore labeled RNAs showed a small decrease (<5%) during activity assays[371]. RNPs used for the measurements usually contained a protein carrying a donor fluorophore (Cy3) and a snoRNA carrying an acceptor fluorophore (Cy5), resulting in the detection of only completely reconstituted complexes. For the generation of smFRET histograms, movies with a length of 7 seconds were recorded and 40 individual measurements for each sample were performed. For each movie, during the first 2.5 seconds, only the green laser was turned on and the FRET value averaged during the first 2 seconds was used for the histograms. During the last 4.5 seconds, the red laser was also turned on and the averaged acceptor value of the last 3 seconds

of the movie was used as a filter value during data analysis to exclude incompletely assembled RNPs. If the intensity of the donor channel reached a sufficiently high value and the difference between the donor and acceptor channel therefore exceeded a certain threshold, the trace was accepted (Figure 96a). In this scenario, the acceptor intensity was the product from FRET between a donor-acceptor pair as well as the red laser irradiation. If the difference between donor and acceptor channel did not reach a certain threshold, the intensity in the acceptor channel was only caused by the red laser irradiation and no FRET pair was present, therefore the trace was discarded (Figure 96b). This procedure was implemented, since during the first measurements a significant donor-only peak could be observed. Normally, Nhp2 should only be able to bind to the snoRNA, that was immobilized by biotin-streptavidin interaction on the glass surface. However, it seemed that either a high number of early bleached molecules were accumulating in the measurement channel or that even after passivation of the glass surface, some Nhp2 was still able to nonspecifically bind inside the measurement channels. With the above described filtering method, these unspecifically bound molecules could be detected and filtered, which removed the donor only peak in an unbiased experimental way rather than utilizing a gaussian fit. As an alternative approach, a reverse labeling scheme was tested with Cy3-labeled snoRNA and Cy5-labeled Nhp2. This however, was also not ideal, since not all snoRNA that were immobilized in the measurement channel did result in a fully reconstituted RNP, which also resulted in a donor only peak and furthermore an overloading of the channel with donor marked molecules (appendix, Figure 139).

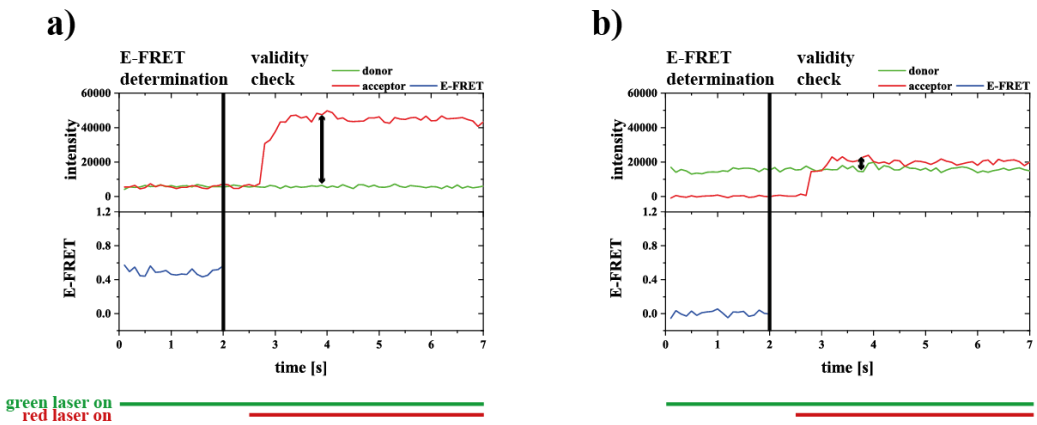


Figure 96: Generation of valid smFRET traces for histograms. a) Difference between donor and acceptor channel exceeds threshold and trace is accepted. b) Difference between donor and acceptor channel does not exceed threshold and trace is denied.

Furthermore, 120 seconds long movies were recorded and analyzed. The analysis of these movies lead to the determination of the different FRET states, that are represented in the histograms. Additionally, with these movies it is possible to check for transitions between the different states to observe possible FRET dynamics. With E-FRET states determined, state populations could be fitted into the histograms with a gaussian function.

4.7.2 smFRET investigation of the 5’ and 3’ hairpin of snR81 with labeled Nhp2

First smFRET investigations were done with the construct H5-Cy5 and Nhp2 K48Cy3 to gain a more detailed insight into the formation of the RNP and the role of Nhp2 during complex formation. The fluorophore on the snoRNA was attached above the pseudouridylation pocket in proximity to the supposed binding surface of Nhp2 atop the hairpin. The position of the fluorophore on Nhp2 was inside of helix α1, which is supposed to not be a part of the binding surface of Nhp2 towards the rest of the RNP, therefore not interfering with the RNP formation, but still in proximity to the other proteins (Figure 97).

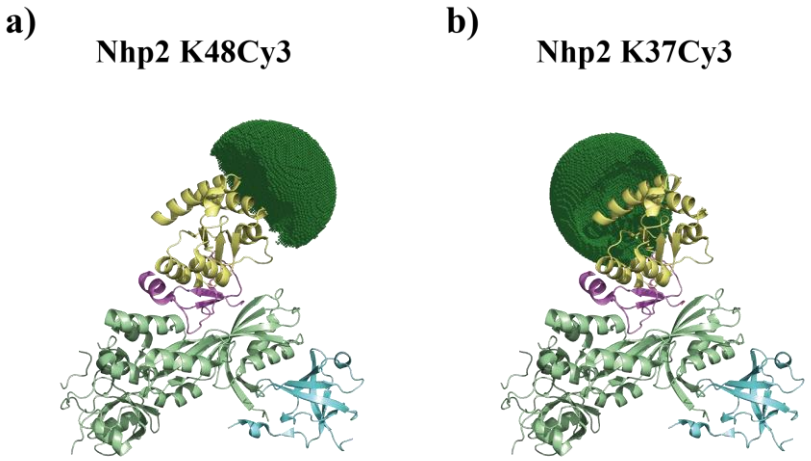


Figure 97: Cartoon representation of accessible volume clouds of sulfo-Cy3 (dark gren) in relation to the RNP proteins Nop10 (magenta), Cbf5 (green), Gar1 (cyan) for labeling sites Nhp2 (yellow) K48 (a) and K37 (b) (pdb: 3U28 and 2LBW). AV clouds were generated using FPS software[380] using the NMR structure of Nhp2 (2LBW) and the corresponding parameters for sulfo-Cy3 including linker length.

As a negative control, only H5 and Nhp2 were incubated together. Nhp2 is known to unspecifically bind to RNA stem-loop structures, so it was necessary to rule out any interactions of Nhp2 with the RNA in absence of the other proteins. The FRET experiment showed that in absence of NCG, Nhp2 was not able to bind to the snoRNA, neither in a specific nor unspecific way (appendix, Figure 140a). To see whether Nhp2 was able to bind to the complex in the absence of Nop10, but presence of Cbf5, the complex H5 WCG was reconstituted. In the presence of only Cbf5 and Gar1, again no FRET population could be identified, showing that the presence of Cbf5 and the snoRNA is not sufficient for Nhp2 binding to the complex and in fact Nop10 is required for Nhp2 to specifically bind to the RNP (appendix, Figure 140b).

As a next step, the full RNP was reconstituted. smFRET measurements revealed a single binding mode for Nhp2 K48Cy3 to the complex with a FRET efficiency of 0.6 (Figure 98). The experiments were also performed with Nhp2 K37Cy3, to verify if the observed data was not only labeling site specific. The utilization of the second labeling site also led to the observation of a single binding mode for Nhp2 to H5 with an E-FRET of 0.5.

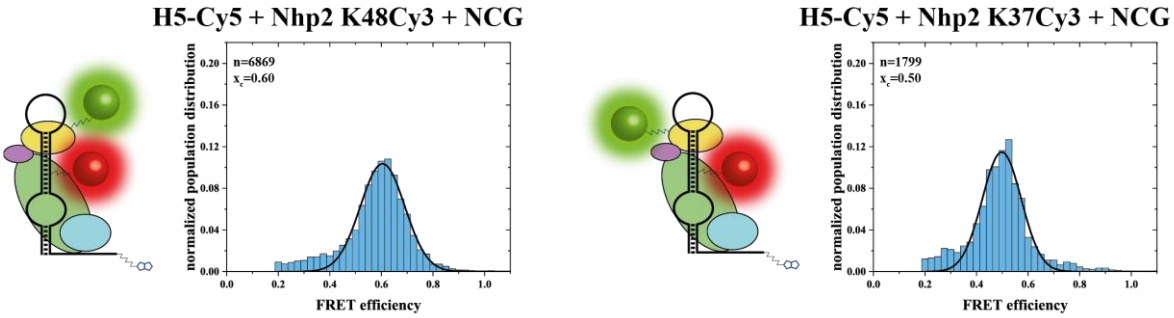


Figure 98: Histograms of fully reconstituted 5' hairpin complexes H5-Cy5 + Nhp2 K48Cy3 + NCG (Ieft) and H5-Cy5 + Nhp2 K37Cy3 + NCG (right). n represents the number of molecules observed for E-FRET ≥ 0.2. The states x_c were determined by analyzing 120 second movies. Gaussian fitting was used to determine and plot the Gaussian distribution in the histogram (error for x_c<0.01).

Analysis of recorded movies, each representing the time course of donor and acceptor intensities of a single FRET pair on one RNP particle over the time course of 120 seconds showed typical anticorrelated donor-acceptor intensities for FRET (Figure 99). Both histograms in Figure 99 show the bleaching of the acceptor dye and as a consequence thereof the increase in donor intensity. The corresponding E-FRET results from $\frac{intensity\ (acceptor)}{intensity\ (acceptor)+intensity\ (donor)}$.

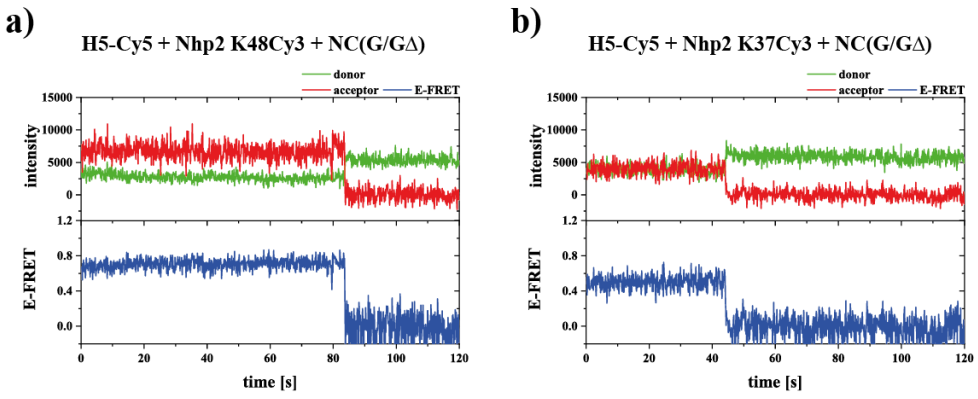


Figure 99: Exemplary time resolved FRET traces, representing donor and acceptor intensity as well as corresponding E-FRET. Traces for H5-Cy5 with Nhp2 K48Cy3 (a) and Nhp2 K37Cy3 (b).

The experiments were repeated with the snoRNA H3-Cy5 to analyze the binding of Nhp2 to the 3' hairpin of the RNP complex. For H3, not a single, but several binding states could be observed. Analysis of the 120 second movies revealed 3 different states for the H3 construct, for both Nhp2 K48Cy3 (Figure 101a) and Nhp2 K48Cy3 (Figure 101b). In a few rare cases, transitions between the three states could be observed, but most of the molecules analyzed did not change their FRET state during the 120 seconds of observations or the FRET pairs bleached out, before any transition could be observed. The three observed states x_{c1-3} including the error, which was determined by standard deviation between all performed smFRET experiments with each construct, could be fitted into the histograms with three gaussian functions (Figure 100). The three FRET states represent one major population at a high E-FRET for both Nhp2 constructs which are populated by 40-50% of the molecules and two minor states at lower FRET values (Table 7).

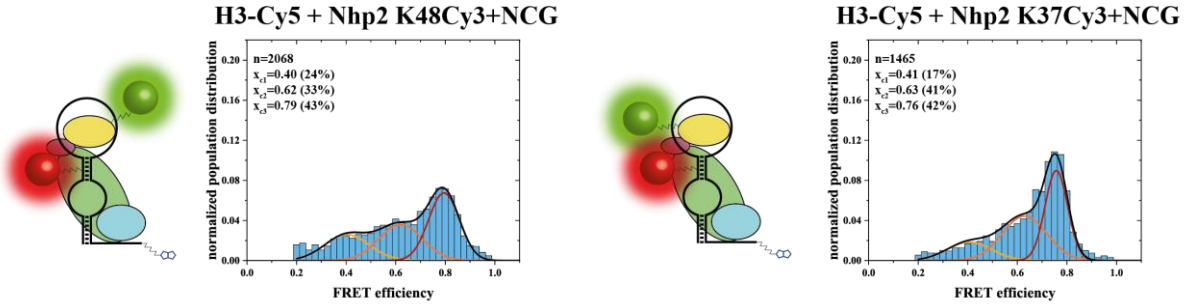


Figure 100: Histograms of fully reconstituted 3' hairpin complexes H3-Cy5 + Nhp2 K48Cy3 + NCG (Ieft) and H3-Cy5 + Nhp2 K37Cy3 + NCG (right). n represents the number of molecules observed for E-FRET ≥ 0.2. The three states x_{c1-3} including error were determined by analyzing 120 second movies (see Table 7) and Gaussian fitting within the error of x_c was used to determine and plot three Gaussian distributions in the histogram with a given peak width between 0.05 and 0.18 based on the single state constructs. Values in % are calculated by determining the area under the curve using the Gaussian distribution and represent the amount of molecules in one of the three states.

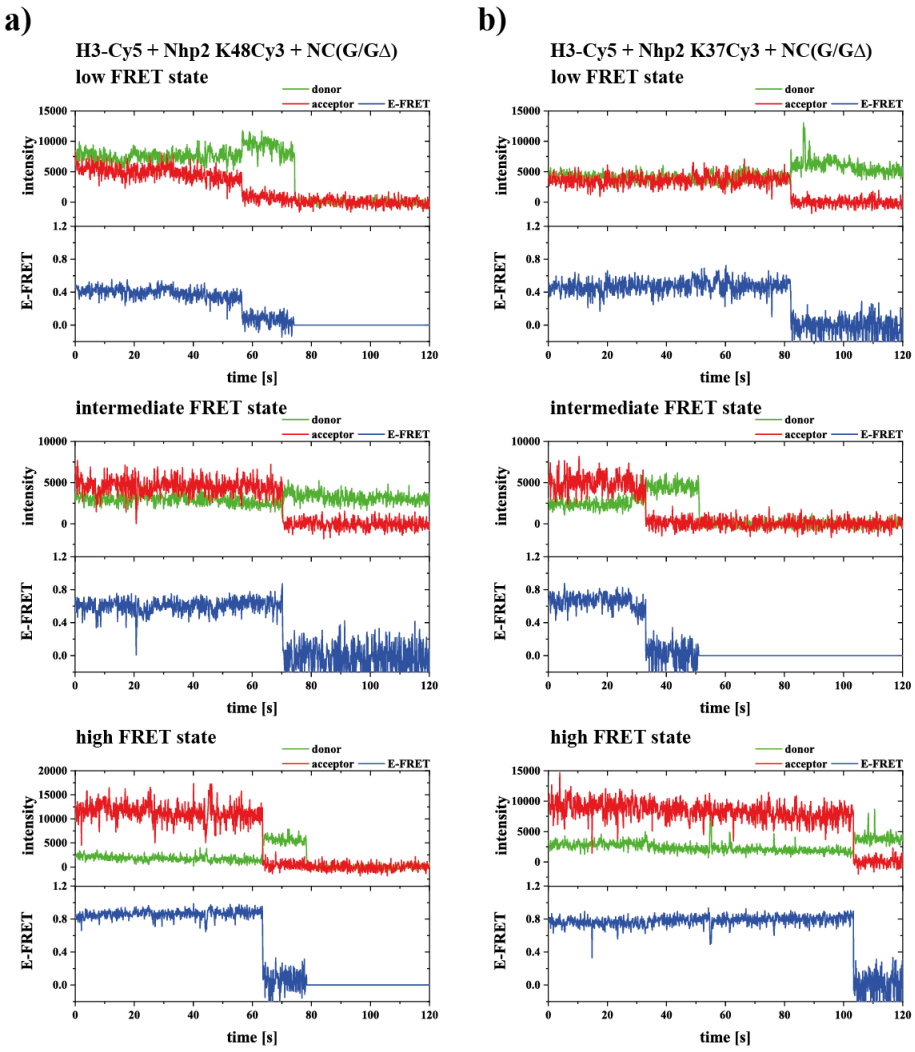


Figure 101: Exemplary time resolved FRET traces, representing donor and acceptor intensity as well as corresponding E-FRET. Traces for the three observed states of H3-Cy5 with Nhp2 K48Cy3 (a) and Nhp2 K37Cy3 (b).

Table 7: States observed for constructs H3 and H3Δ. FRET states were determined by analyzing 120 second movies and state identification by HaMMy[381]. Errors were calculated from standard deviations between 24 individual measurements for each construct.

	low FRET	intermediate FRET	high FRET
Nhp2 K48Cy3	0.43 ± 0.03	0.65 ± 0.03	0.82 ± 0.03
Nhp2 K37Cy3	0.43 ± 0.05	0.60 ± 0.03	0.77 ± 0.04

4.7.3 smFRET investigation of the influence of Gar1 and eukaryotic specific GAR domains

To gain more insight into the influence of the GAR-domains, that already showed different effects on pseudouridylation activity during activity test, the smFRET experiments were also performed with the omission of the eukaryotic specific GAR domains and the use of Gar1Δ as well as the complete omission of Gar1. The experiments were performed with the snoRNA constructs H5-Cy5 and H3-Cy5 and both Nhp2 constructs Nhp2 K48Cy3 and K37Cy3. The complete omission of Gar1 did not change the E-FRET value of the observed state, nor did other states arise. Nhp2 was still bound towards the RNP in a single binding mode, which was observed for both labeling sites (Figure 102a). However, the omission of Gar1 led to slightly sharper states observed in the histograms. Interestingly, the omission of only the eukaryotic specific domains of Gar1 with the use of the construct Gar1Δ, led to a major increase in peak sharpening. Still, no secondary state could be observed, nor did E-FRET value change for the single observable state (Figure 102b). Both Nhp2 K37Cy3 and Nhp2 K47Cy3 showed similar results, the only difference between the two constructs was the fact that the peak sharpening was not as strong for the NC construct for K48Cy3 than it could be observed for K37Cy3. For both NCG and NCGΔ, the states looked similar for both constructs.

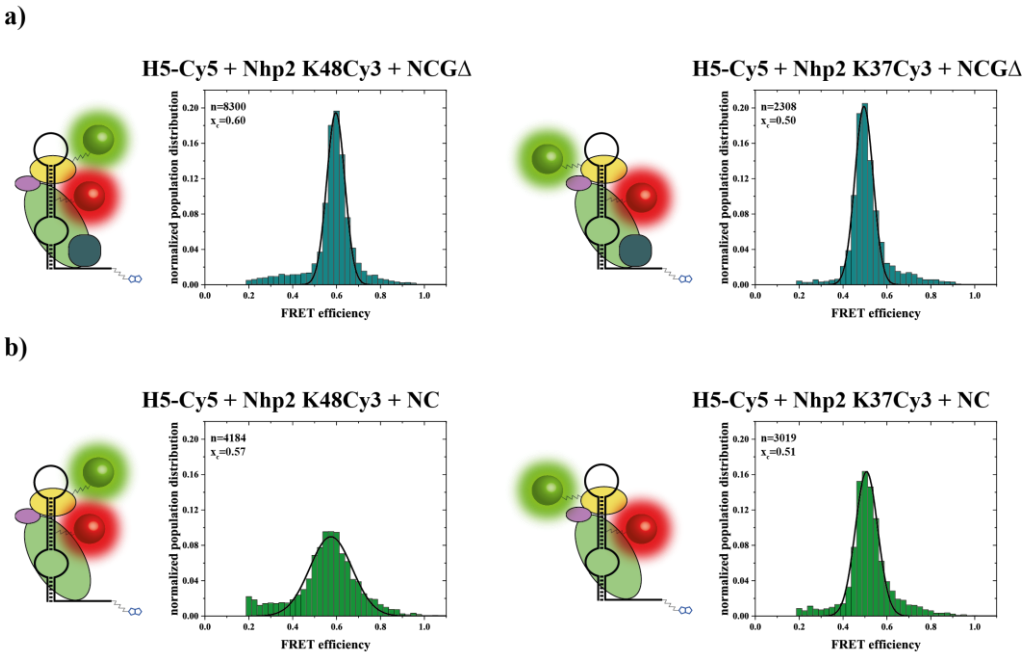


Figure 102: Histograms of fully reconstituted 3' hairpin complexes H3-Cy5 + Nhp2 K48Cy3 + NCG (Ieft) and H3-Cy5 + Nhp2 K37Cy3 + NCG (right). n represents the number of molecules observed for E-FRET ≥ 0.2. The states x_c were determined by analyzing 120 second movies. Gaussian fitting was used to determine and plot the Gaussian distribution in the histogram (error for x_c<0.01).

The described experiments were also performed for construct H3-Cy5. As before observed for the full complex reconstitution, several states could be seen in the histograms, which could be pointed down to three individual states by analyzing 120 second movies. The three states represented the same states as before (Table 7). The high FRET state at around E-FRET ~ 0.8 again was the most populated state (Figure 103a). However, omission of the GAR domains led to an increase of this population. For construct Nhp2 K48Cy3, the amount of molecules increased from 43% to 51%. For K37Cy3, the increase in population was from 42% to 60%. While the intermediate FRET state stayed roughly the same (around 30-40% of the molecules), the low FRET state decreased for both constructs. With the complete omission of Gar1, the same observations could be made as for GAR domain omission. While the low FRET state declined, the high FRET state increase in population (Figure 103b). For both constructs, the complete omission of Gar1 led to roughly the same population of states as only the GAR domain omission did.

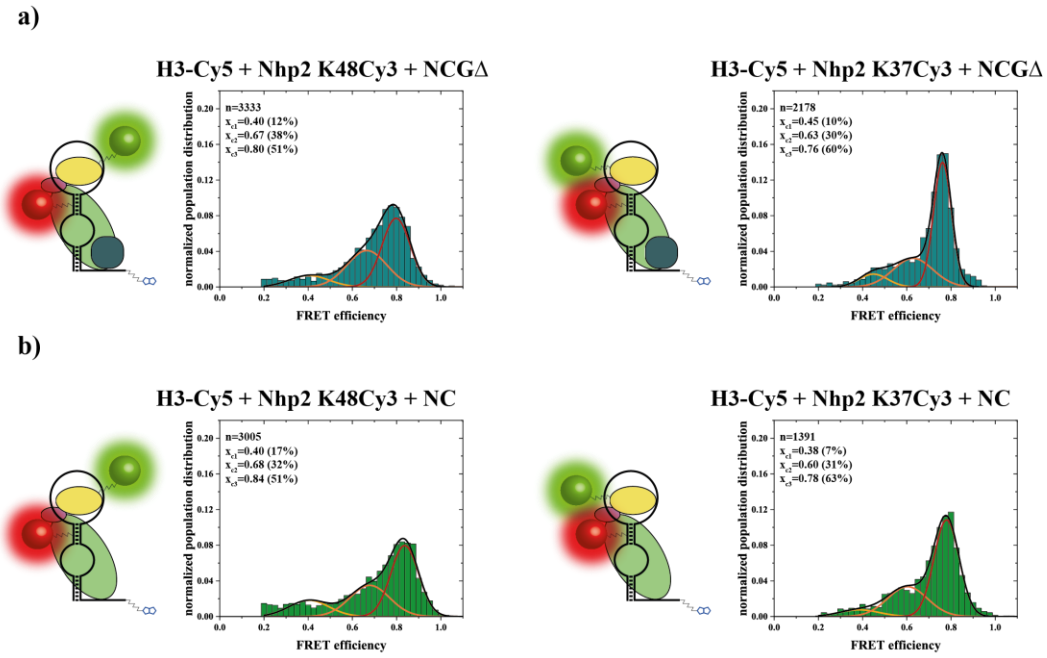


Figure 103: a) Histograms of GAR domain omission for the 3' hairpin complexes H3-Cy5 + Nhp2 K48Cy3 + NCGΔ (Ieft) and H3-Cy5 + Nhp2 K37Cy3 + NCGΔ (right). b) Histograms of Gar1 omission for the 3' hairpin complexes H3-Cy5 + Nhp2 K48Cy3 + NC (Ieft) and H3-Cy5 + Nhp2 K37Cy3 + NC (right). n represents the number of molecules observed for E-FRET ≥ 0.2. The three states x_{c1-3} including error were determined by analyzing 120 second movies (see Table 7) and Gaussian fitting within the error of x_c was used to determine and plot three Gaussian distributions in the histogram with a given peak width between 0.05 and 0.18 based on the single state constructs. Values in % are calculated by determining the area under the curve using the Gaussian distribution and represent the amount of molecules in one of the three states.

4.7.4 smFRET investigation of shortened RNA constructs

After the smFRET experiments with the H5 and H3 hairpins respectively, the shortened hairpin constructs H5Δ and H3Δ were also measured for comparison. Both constructs H5Δ-Cy5 and H3Δ-Cy5 were tested with both Nhp2 constructs K48Cy3 and K37Cy3. Since both of the hairpins are missing the upper part of their stem-loop structure, which got replaced by a tetraloop, the actual binding site for Nhp2 is missing. Since Nop10 was deemed to be necessary for the binding of Nhp2 as well as the performance of catalytic activity at all – even in the absence of Nhp2 – it was interesting to see in which way Nhp2 would interact with the shortened snoRNA constructs. Activity test had already shown that even without the snoRNA binding site of H5, still the absence or presence of Nhp2 had an impact on the catalytic activity of the RNP, so even without the upper stem-loop structure, in some way Nhp2 should still be able to specifically bind to the RNP complex. First, the full RNP of H5Δ was reconstituted with NCG and the labeled Nhp2 constructs. Interestingly, smFRET investigations did not show a single binding mode, as was the case for H5, but several states could be observed in the histograms. Again, 120 second movies of the construct were analyzed to obtain information about the actual states visible in the histogram and, as was the case for construct H3, three distinct states could be pinpointed (Table 8). For both constructs Nhp2 K48Cy3 and K37Cy3, one intermediate FRET state and two high FRET states could be observed (Figure 105). Using Gaussian fitting, these three states could be placed into the histograms and population of each state could be determined (Figure 104). For both constructs, the high FRET state around 0.78 / 0.77 was the most populated state with 54% / 60% respectively.

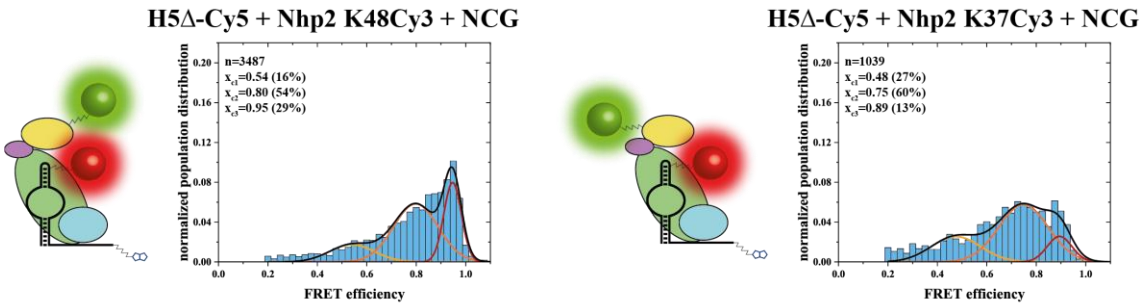


Figure 104: Histograms of fully reconstituted 3' hairpin complexes H5Δ-Cy5 + Nhp2 K48Cy3 + NCG (Ieft) and H5Δ-Cy5 + Nhp2 K37Cy3 + NCG (right). n represents the number of molecules observed for E-FRET ≥ 0.2. The three states x_{c1-3} including error were determined by analyzing 120 second movies (see Table 8) and Gaussian fitting within the error of x_c was used to determine and plot three Gaussian distributions in the histogram with a given peak width between 0.05 and 0.18 based on the single state constructs. Values in % are calculated by determining the area under the curve using the Gaussian distribution and represent the amount of molecules in one of the three states.

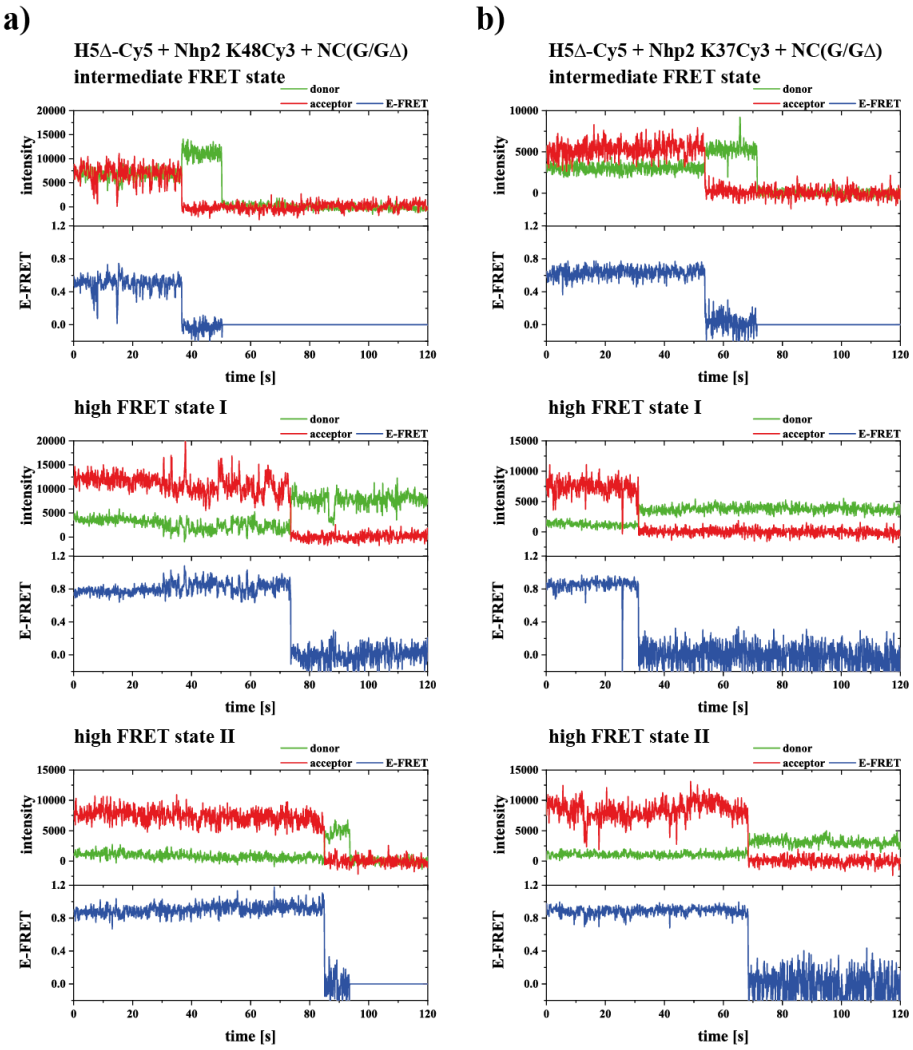


Figure 105: Exemplary time resolved FRET traces, representing donor and acceptor intensity as well as corresponding E-FRET. Traces for the three observed states of H5Δ-Cy5 with Nhp2 K48Cy3 (a) and Nhp2 K37Cy3 (b).

Table 8: FRET states observed for construct H5Δ. FRET states were determined by analyzing 120 second movies and state determination by HaMMy[381]. Errors were calculated from standard deviations between 12 individual measurements for each construct.

	intermediate FRET	high FRET I	high FRET II
Nhp2 K48Cy3	0.57 ± 0.04	0.78 ± 0.02	0.96 ± 0.01
Nhp2 K37Cy3	0.49 ± 0.03	0.77 ± 0.02	0.89 ± 0.02

Like for the other constructs, also the smFRET experiments with GAR domain and Gar1 omission were performed. Like before, the same three E-FRET states could be identified without any significant change of the states (Figure 106). Upon GAR domain as well as Gar1 omission, high FRET state II increased in population (around 20%), while high FRET state I decreased (around 10%). While high FRET state I did not show any distinctive characteristics, high FRET state II showed the same behavior that was previously observed for construct H5. For the GAR domain omission, the high FRET became much sharper, while the complete omission of Gar1 only led to a small sharpening of the state. This effect was more prominent upon Gar1 omission for the Nhp2 K37Cy3 construct than for the Nhp2 K48Cy3 construct.

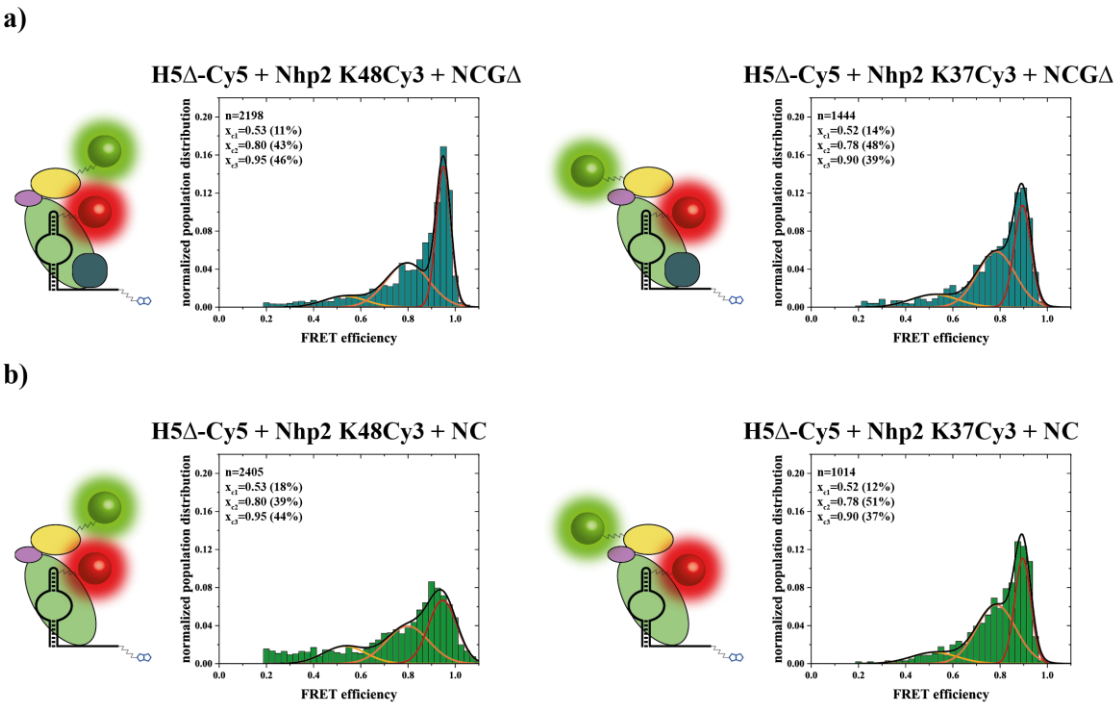


Figure 106: a) Histograms of GAR domain omission for the 5' hairpin complexes H5Δ-Cy5 + Nhp2 K48Cy3 + NCGΔ (Ieft) and H5Δ-Cy5 + Nhp2 K37Cy3 + NCGΔ (right). b) Histograms of Gar1 omission for the 5' hairpin complexes H5Δ-Cy5 + Nhp2 K48Cy3 + NC (Ieft) and H5Δ-Cy5 + Nhp2 K37Cy3 + NC (right). n represents the number of molecules observed for E-FRET ≥ 0.2. The three states x_{c1-3} including error were determined by analyzing 120 second movies (see Table 8) and Gaussian fitting within the error of x_c was used to determine and plot three Gaussian distributions in the histogram with a given peak width between 0.05 and 0.18 based on the single state constructs. Values in % are calculated by determining the area under the curve using the Gaussian distribution and represent the amount of molecules in one of the three states.

While H5Δ still showed catatalytic activity, even in the absence of the upper stem-loop, no catalytic activity at all could be observed for H3Δ. smFRET experiments with a fully reconstituted WNCG complex on H3Δ-Cy5 again showed several states instead of one clear state. The collected 120 second movies for H3Δ-Cy5 were analyzed and revealed three distinct states for construct H3Δ-Cy5. Furthermore, these three states were the same states, that could be observed for construct H3-Cy5, with Nhp2 K37Cy3 as well as Nhp2 K48Cy3 (appendix, Figure 141). Interestingly, while the states observed for H5 and H5Δ were completely different (additional to the fact that for H5 a single state and for H5Δ three different states could be observed), the states that could be observed for H3 and H3Δ appeared identical (Table 7). However, the population observed for these states was majorly different. While for H3, the high FRET state was the most distinct and most populated state, for H3Δ this was not the case. For Nhp2 K48Cy3, it was even the least populated state of the three states. For both Nhp2 K48Cy3 and Nhp2 K37Cy3, the intermediate FRET state was populated with the highest number of molecules (Figure 107).

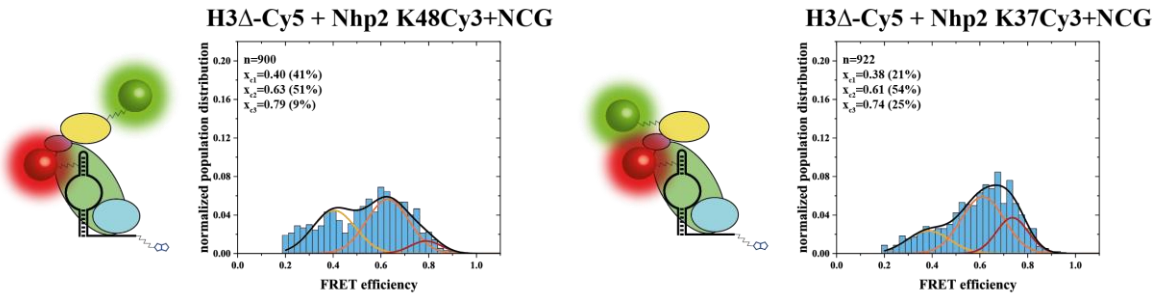


Figure 107: Histograms of fully reconstituted 3' hairpin complexes H3Δ-Cy5 + Nhp2 K48Cy3 + NCG (Ieft) and H3Δ-Cy5 + Nhp2 K37Cy3 + NCG (right). n represents the number of molecules observed for E-FRET ≥ 0.2. The three states x_{c1-3} including error were determined by analyzing 120 second movies (see Table 7) and Gaussian fitting within the error of x_c was used to determine and plot three Gaussian distributions in the histogram with a given peak width between 0.05 and 0.18 based on the single state constructs. Values in % are calculated by determining the area under the curve using the Gaussian distribution and represent the amount of molecules in one of the three states.

To test whether the GAR domain and Gar1 omission would have an impact on the states, smFRET experiments were performed with these constructs as well. However, no clear pattern could be identified from these two constructs. While in general, the intermediate FRET state was mostly populated, the difference between the three states was not huge, and no characteristic peak (like it was the case for the high FRET state of H3 and the high FRET state II of H5Δ) could be identified in the histograms (Figure 108). The GAR domain omission and the complete omission of Gar1 did in fact lead to a higher population of the high FRET state of H3Δ. However, this observation was not as characteristic as it was for construct H3.

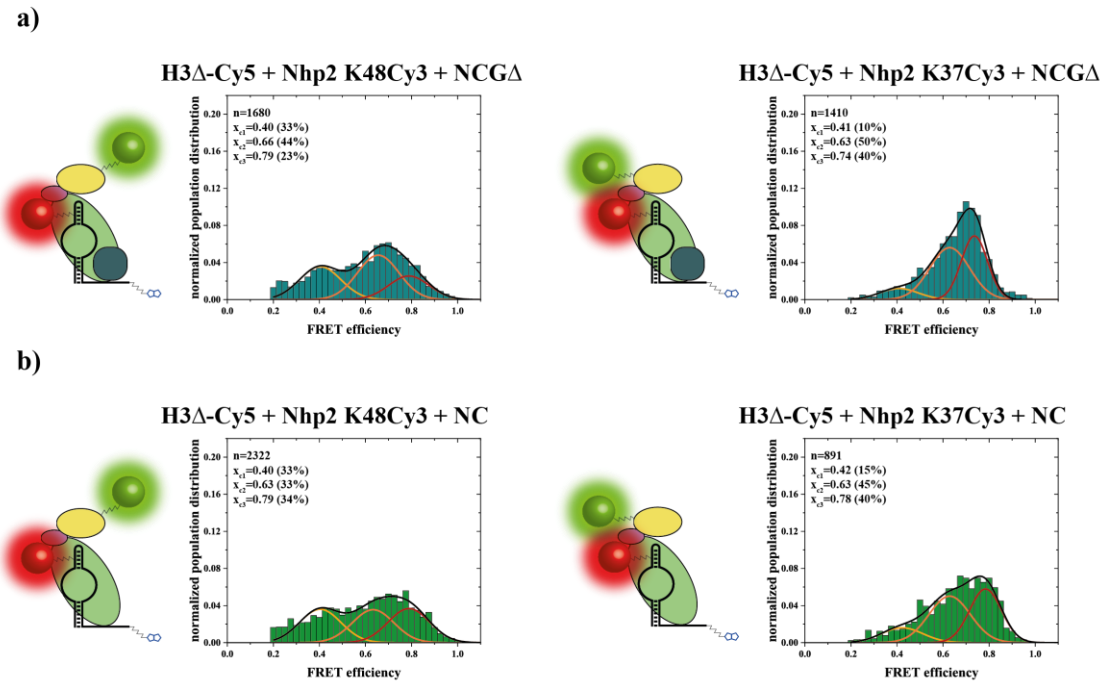


Figure 108: a) Histograms of GAR domain omission for the 3' hairpin complexes H3Δ-Cy5 + Nhp2 K48Cy3 + NCGΔ (Ieft) and H3Δ-Cy5 + Nhp2 K37Cy3 + NCGΔ (right). b) Histograms of Gar1 omission for the 3' hairpin complexes H3Δ-Cy5 + Nhp2 K48Cy3 + NC (Ieft) and H3Δ-Cy5 + Nhp2 K37Cy3 + NC (right). n represents the number of molecules observed for E-FRET ≥ 0.2. The three states x_{c1-3} including error were determined by analyzing 120 second movies (see Table 7) and Gaussian fitting within the error of x_c was used to determine and plot three Gaussian distributions in the histogram with a given peak width between 0.05 and 0.18 based on the single state constructs. Values in % are calculated by determining the area under the curve using the Gaussian distribution and represent the amount of molecules in one of the three states.

4.7.5 Effect of substrate RNA on the binding state of Nhp2

The smFRET measurements to this point were made with a fully or partially reconstituted snoRNP, however no substrate RNA was present during the experiments. To check whether the presence of target RNA had an influence on the snoRNA/Nhp2 interaction, an excess of target RNA was added into the imaging buffer (the buffer containing the oxygen scavenging system) before the buffer was added onto the measurement channel. All smFRET experiments with all constructs that were described previously, were also done in the presence of substrate RNA (appendix, H5: Figure 142, H3: Figure 143, H5 Δ: Figure 144, H3Δ: Figure 145). With the addition of substrate RNA, no apparent changes in the states could be observed. Also, the three states that could be detected for H5Δ and for H3/H3Δ were not affected and stayed the same. The population of the states also did not change drastically. Most of the time, a difference of a few percent could be detected, which is well inside the assumed error between populations. The only bigger differences in state population that were observed upon substrate RNA addition were for H3-Cy5 and Nhp2K37 with NCG and for H3Δ-Cy5 and Nhp2K48 with NCG. For both experiments, a higher population of the high FRET state could be observed (58% vs. 42% for H3 and 22% vs. 9% for H3Δ). However, for H5 constructs, a change in the peak width (full width at half maximum, FWHM) could be noted. For a more in depth comparison, H5-Cy5 and Nhp2K48-Cy3 were tested with two different target RNAs. The natural 5’ substrate RNA segment was tested as well as a variant of the 5’ substrate where the target uridine was replaced by 5-flurouridine (5FU, Figure 109a). This target has been shown to be able to bind to the catalytic core of the H/ACA RNP, however pseudouridylation is inhibited or slowed down by several orders of magnitude[112,113,121,382,383]. This leads to a longer dwell time of the 5FU-substrate on the pseudouridylation pocket of the complex. The presence of any of the target RNAs did not lead to a change of the E-FRET states. Still, for H5 a single binding state could be observed. However, a sharpening of the E-FRET state could be observed for the presence of 5FU-substrate RNA and an even stronger sharpening for the 5’ substrate (Figure 109b). The FWHM for the E-FRET state in a consecutive experiment decreased from 0.17 to 0.16 with the addition of 5FU substrate and to 0.14 with the addition of 5’ substrate. In an experiment performed without Gar1, the FWHM of the E-FRET state decreased from 0.19 to 0.14 upon 5FU substrate as well as 5’ substrate addition.

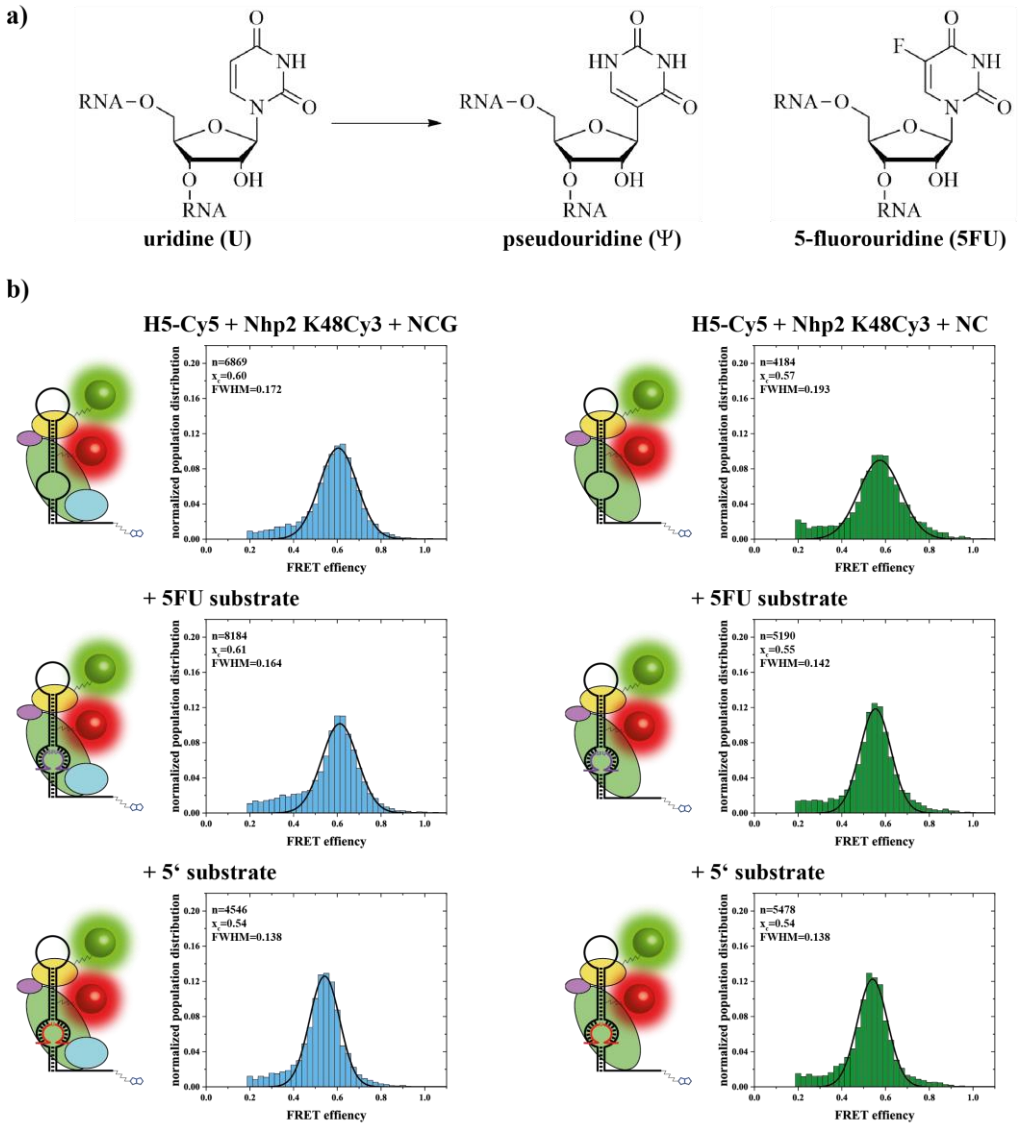


Figure 109: a) Structure of normal substrate uridine and its transformation into pseudouridine as well as the inhibitor 5-fluorouridine. b) Histograms of fully reconstituted 5’ hairpin complexes H5-Cy5 + Nhp2 K48Cy3 + NCG (left) and complex without Gar1 H3-Cy5 + Nhp2 K37Cy3 + NC (left) without substrate, with 5-FU substrate and with normal substrate (top to bottom). n represents the number of molecules observed for E-FRET ≥ 0.2. The states x_c were determined by analyzing 120 second movies. Gaussian fitting was used to determine and plot the Gaussian distribution in the histogram (error for x_c<0.01).

4.7.6 smFRET investigation on H5 and H3 with labeled Cbf5Δ

Investigations on Nhp2 revealed a single binding mode for H5 and three different binding states for H3. This raises the question whether the different binding modes for H3 are actually caused by Nhp2 binding in different ways to the RNP complex, or if Nhp2 always binds in the same way (like for H5), but the RNA stem folds in different ways or misfolds, causing several binding states of Nhp2 for H3 (and possibly also causes the lower activity observed for H3). To follow up on this topic, smFRET measurements with H5-Cy5 and H3-Cy5 and Cbf5Δ L156Cy3 were performed. The shortened version of Cbf5 was used, because the Cy3-labeled

full length Cbf5 showed some degradation during the labeling, very low yield and contained some contaminations after purification. The shorter construct Cbf5Δ could be purified in an acceptable yield with better purity. In activity tests between H5 and H3, Cbf5Δ showed a lower activity in general, but like with the full length Cbf5, H5 was more active than H3. To confirm if labeling interfered with enzymatic activity of the complex, an activity assay was performed, which showed an active fluorophore labeled complex (appendix, Figure 146). With the donor label on the protein Cbf5Δ and the acceptor label still on the same positions in the H5 and H3 hairpin stems, the smFRET measurements were performed with a fully reconstituted complex, containing Nhp2, Nop10 and Gar1. Also, substrate RNA was added to the samples. Since the donor labeled was attached close to the highly dynamic thumb loop domain, also 120 second movies were recorded and analyzed for possible dynamics. The histograms of both constructs showed a single binding mode for Cbf5Δ, with an E-FRET of 0.88 for H5 and 0.89 for H3 (Figure 110). The FRET traces showed typical anticorrelated behavior, however no dynamics of the thumb loop domain could be observed in the movies (Figure 111). Also in a second measurement no dynamics could be observed, however one interesting thing to point out is the fact, that for H5, always four to five times as many FRET pairs could be observed than for H3, indicating a stronger binding of Cbf5Δ to H5 than to H3.

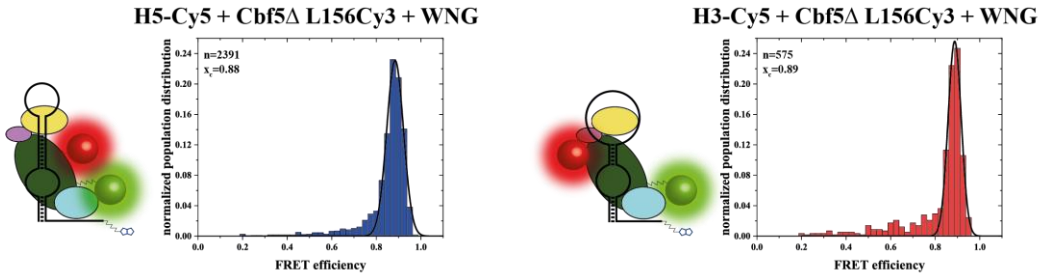


Figure 110: Histograms of fully reconstituted 3' hairpin complexes H3-Cy5 + Nhp2 K48Cy3 + NCG (Ieft) and H3-Cy5 + Nhp2 K37Cy3 + NCG (right). n represents the number of molecules observed for E-FRET ≥ 0.2. The states x_c were determined by analyzing 120 second movies. Gaussian fitting was used to determine and plot the Gaussian distribution in the histogram (error for x_c<0.01).

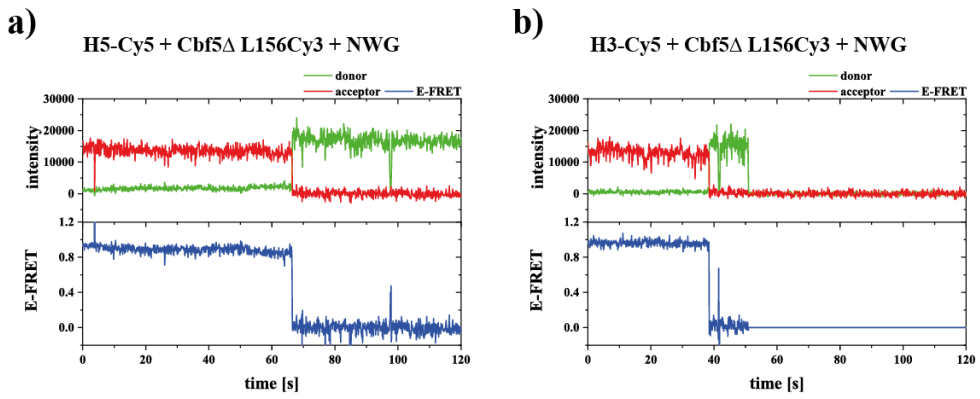


Figure 111: Exemplary time resolved FRET traces, representing donor and acceptor intensity as well as corresponding E-FRET. Traces for H5-Cy5 (a) and H3-Cy5 (b) with Cbf5Δ L156Cy3.

4.8 Investigation of the full length snR81 RNP

Since the previously described smFRET experiments were carried out on the individual standalone hairpins H5 and H3 of snR81, an investigation of the full snoRNP could provide more insight into effects between the two individual hairpins. It was already shown in catalytic activity assays, that there seems to be some effect between the two hairpins that affects their ability to pseudouridylate substrate RNAs. Enzymatic activity on a single standalone hairpin was always slower than on the full length snR81, an effect that was only minute for the 5' hairpin and had major influence on the 3' hairpin. Since very little structural data is available for full length snR81 RNP, it was however not clear whether the two hairpins were in a favorable distance for spectroscopic analysis of one another. Two separate labeling strategies were implemented for the full length snR81 RNP (Figure 112). The double hairpin complex could be labeled with fluorophores and investigated by smFRET spectroscopy and spin labels should be attached and PELDOR distance measurement were to be performed.

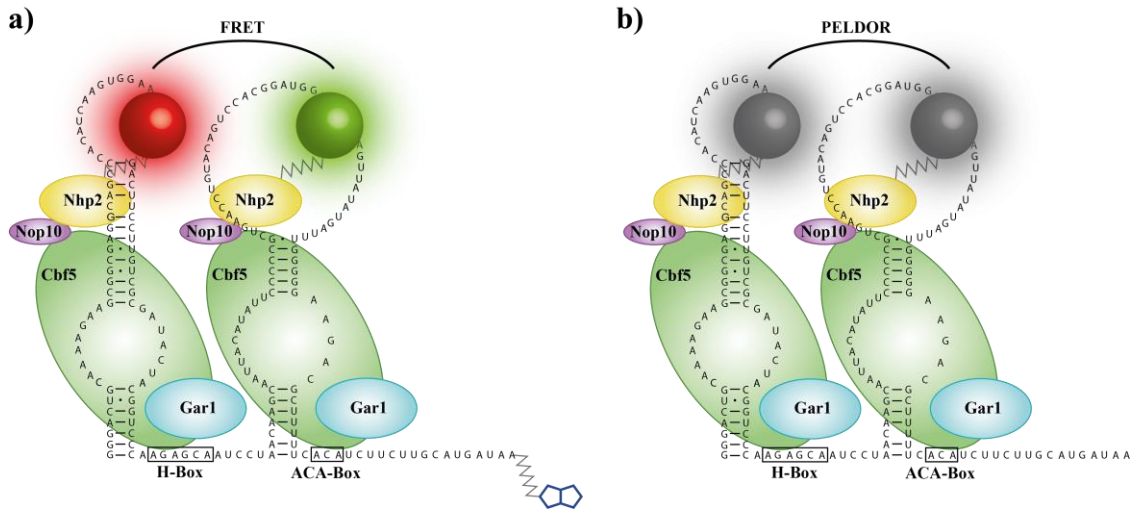


Figure 112: Labeling scheme for the full length snR81 hairpin. a) Labeling with fluorophores. Approximately half of the sample were to be reconstituted with a donor-acceptor pair of Cy3 and Cy5 on the two hairpins. Shown is one of the two possible outcomes with a Cy5-labeled Nhp2 on the 5' hairpin and a Cy3-labeled Nhp2 on the 3' hairpin. b) Labeling with spin labels. Attachment of a spin label to Nhp2 and complex reconstitution could lead to a double spin labeled complex.

4.8.1 Investigation by smFRET

To gain more insight into this question, samples with 600 nM full length snR81 were prepared and incubated with 6 µM NCG as well as 3 µM Nhp2 K48Cy5 and 3 µM Nhp2K48Cy3. The idea behind this approach was, that an equimolar amount of both acceptor and donor labeled Nhp2 was utilized, which in theory should lead to 25% of the sample with two donor molecules and 25% of the sample with two acceptor molecules, but still 50% of the sample would have either an acceptor at the 5' hairpin and a donor at the 3' hairpin or vice versa. 50% of the molecules being eligible to give a FRET signal was more than enough to test out

the interaction between the two hairpins. However, no smFRET population could be detected in the histograms and also analysis of 120 second movies did not show any typical FRET behavior.

4.8.2 Investigation by PELDOR

4.8.2.1 Preparation of spin labeled Nhp2 and investigation of snR81 RNP via EPR spectroscopy

As a parallel approach to investigate the relation between the two hairpins and the attached proteins on each hairpin of snR81, efforts were undertaken to prepare a fully reconstituted RNP with two spin labels, one on each hairpin. Since the main focus up until this point lay upon the labeling of proteins via non-canonical amino acids, an approach with spin labeled Nhp2 was made. In theory, each hairpin is able to hold a single set of RNP proteins, including one Nhp2 per hairpin. This allowed the spin labeling of Nhp2 at a single position inside the protein with one spin label. With the reconstitution of the snR81 RNP, it should be possible to gain double labeled RNPs in this way, which then could be used in PELDOR distance measurements between the two labels.

4.8.2.2 Incorporation of NoK into Nhp2

For spin label attachment, the non-canonical amino acid norbornene-lysine (NoK) was incorporated into several constructs of Nhp2 that were already available. Incorporation of the amino acid was achieved via amber suppression technique, utilizing the pEVOL^{AF} mutant, which was already utilized for SCOK incorporation. Small scale test expressions revealed a successful incorporation of the non-canonical amino acid into Nhp2, however expression of protein seemed to be slower for this amino acid and expression time was prolonged to seven days of cell expression (Figure 113).

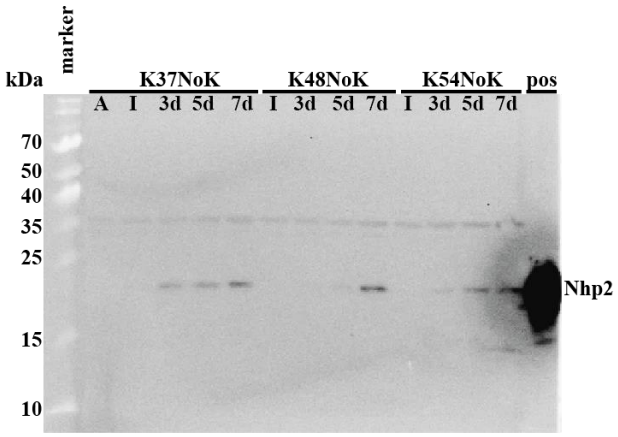


Figure 113: Expression test of NoK incorporation into the Nhp2 constructs K37, K48 and K54. marker: Spectra Multicolor Broad Range Protein Ladder (Thermo Scientific). positive control: Nhp2. A: before arabinose induction. I: before IPTG induction.

Expression seemed to work most reliably with construct Nhp2 K37, so with this construct a larger scale expression of 2 x 2 liters in TB medium was set up. Each pellet from a 2 liter expression was lysed separately and purified via affinity chromatography. Afterwards, the protein containing fractions of both samples were pooled and RNase A treatment was done overnight. RNase A was removed with a second affinity chromatography. Since protein yield was really low compared to the large expression volume of 4 liters, it was decided to pool all Nhp2 containing fractions after the affinity chromatography to not waste any of the protein, even the ones with impurities (appendix, Figure 147). Nhp2 K37NoK was further purified via size exclusion chromatography, and unfortunately, some of the impurities overlapped with Nhp2 containing fractions, so in the end, some of the Nhp2 containing fractions had to be discarded for higher protein purity (Figure 114).

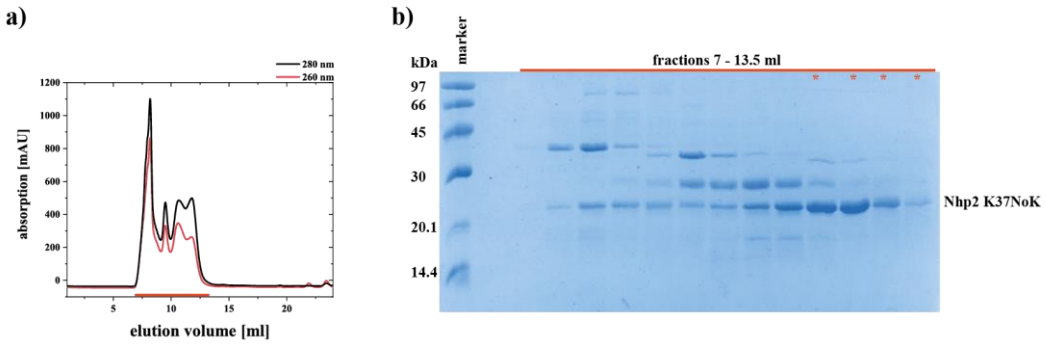


Figure 114: Size exclusion chromatography purification of Nhp2 K37NoK. a) SEC elution chromatogram using a Superdex 75 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) PAGE analysis of fractions from SEC. Fractions used for further experiments are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

Nhp2 K37NoK could be obtained in a yield of 1.3 mg out of four liter expression medium.

4.8.2.3 Spin labeling of Nhp2NoK constructs

Nhp2 K37NoK (40 µM) was incubated with a ten-fold excess of the tetrazine spinlabel (see appendix, Figure 148) for one hour, according to a similar literature approach for tetrazine labels[209]. Afterwards, the sample was diluted to a protein concentration of 5 µM and the tetrazine spin label concentration was increased to a 40-fold excess (in line with the standard MTSSL labeling procedure) and the reaction was carried out for an additional hour. The unreacted spin label was removed by filter column and the protein solution was brought to a high concentration via Vivaspin concentrator. EPR spin counting showed the successful attachment of the spin label, however only a labeling efficiency of 25% could be reached with this method, which would be not sufficient for PELDOR measurements.

4.8.2.4 Spin labeling of Nhp2 with MTSSL

As an alternate strategy for spin labeling of the protein Nhp2, an additional strategy was pursued. In this labeling strategy, a disulfide bond labeling scheme was to be utilized to couple the spin label MTSSL with the side chain of a cysteine. Conveniently, Nhp2 only contained one native cysteine at position 94. As a first test, this native cysteine was labeled with MTSSL. 400 µg of the protein was brought to a concentration of 5 µM and incubated with a 40-fold excess of MTSSL. The unreacted MTSSL was removed by filter column and the protein solution was brought to a high concentration via Vivaspin concentrator. Successful labeling of Nhp2 was confirmed by EPR spin counting (see appendix, Figure 149). By comparison of label concentration and the protein concentration determined via UV-vis spectroscopy, a labeling efficiency of ~90% could be determined. However, labeling site 94 was not optimal for PELDOR measurements. For once, the labeling site was very rigid, on the other hand the labeling site was contained relatively close to one of the binding sites of Nhp2 towards the RNP complex. To optimize the labeling scheme, several labeling sites were determined based on the available NMR structure of Nhp2 and a calculation by program MMM[384]. As favorable labeling sites, positions K37, K54, G70, K131, K136 and K151 were chosen (Figure 115).

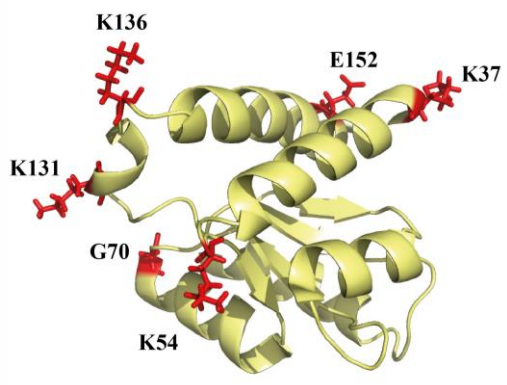


Figure 115: NMR structure of Nhp2W (pdb: 2LBW). Mutation sites K37, K54, G70, K131, K136 and K151 are marked in red.

In order to perform the labeling reaction with MTSSL in a site-specific way, it was necessary to first alter the cysteine at position 94. Two Nhp2 constructs were prepared with the native cysteine replaced by another amino acid. In construct C94A, cysteine was replaced by alanine, which provides a rather neutral and non-disturbing replacement of the cysteine. The second construct C94S contained its hydrophilic character by replacing the cysteine with a serine, switching out the sulfur with oxygen. Both constructs C94A and C94S were then mutated a second time to introduce a cysteine at the different labeling positions. All twelve constructs were used in a test expression to verify that protein was produced. Expression was performed in TB medium with standard expression conditions for Nhp2. All constructs were able to overexpress protein (Figure 116), however, for some constructs the protein band was running at an abnormal height (K151 and to a lesser extent G70 and K136). To produce

cysteine mutants in a larger scale for spin labeling and potential PELDOR measurements, the two constructs Nhp2 C94A/K37C and Nhp2 C94S/K54C were chosen by the fact, that these labeling sites were also used for the fluorophore labeling.

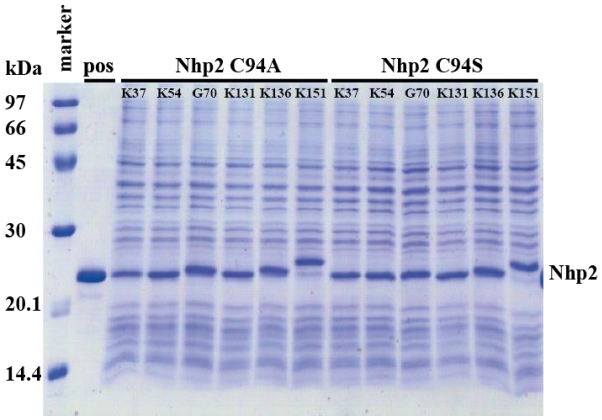


Figure 116: Expression test of Nhp2 mutants with C94 mutated to either alanine or serine and K37, K54, G70, K131, K136 or K151 mutated to cysteine. Marker: low molecular weight marker (GE Healthcare). Positive control: Nhp2.

Large scale expression of both mutants were performed in TB medium and cells harvested. Afterwards, standard purification procedure was performed with affinity chromatography, incubation with RNase A overnight and a second affinity chromatography and SEC afterwards (Figure 117). For both proteins, an unexpectedly high amount of 10 mg per construct could be purified from one liter medium, indicated also by the overloading of the protein gel and the SEC chromatogram. An 260/280 nm absorption ratio of 0.60 showed, that RNA free proteins could be obtained.

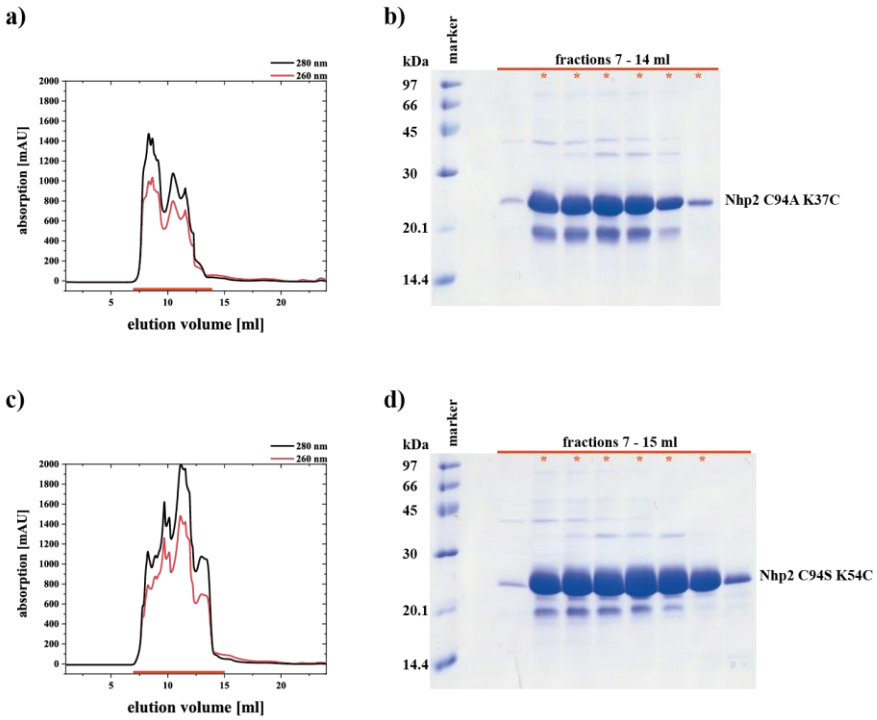


Figure 117: Size exclusion chromatography purification of Nhp2 cysteine mutants. a) SEC elution chromatogram of Nhp2 C94A K37C using a Superdex 75 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) PAGE analysis of fractions from SEC. c) SEC elution chromatogram of Nhp2 C94S K48C using a Superdex 75 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. d) PAGE analysis of fractions from SEC. Fractions used for further experiments are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

With both proteins, labeling reactions were performed with a 40-fold excess of MTSSL and purification of the samples were done with a filter column to remove excess MTSSL. The samples were brought to a high concentration with vivaspin concentrators and EPR spin counting measurements were performed (Figure 118). Like for the wildtype Nhp2, a labeling efficiency of ~90% could be obtained for both constructs.

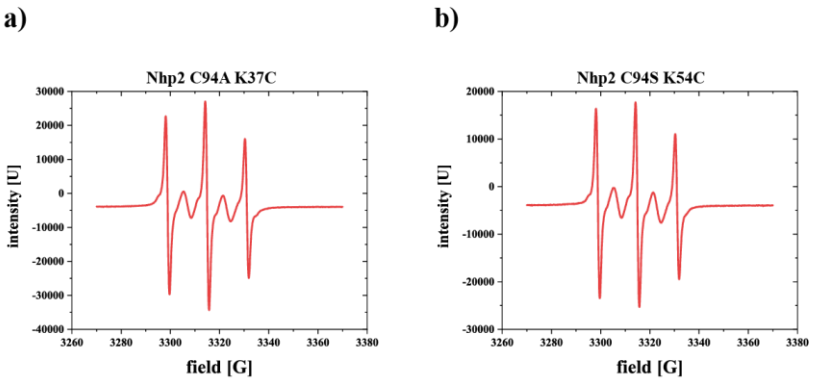


Figure 118: Spin counting of MTSSL labeled Nhp2 constructs. a) Nhp2 C94A K37C and b) Nhp2 C94S K54C.

Activity test performed with the MTSSL labeled wildtype Nhp2 as well as the constructs C94A K37C and C94S K54C in comparison to samples with unlabeled Nhp2 and without Nhp2 showed, that label attachment did not interfere with enzymatic activity (appendix, Figure 150).

4.9 Discussion of chapter 4

4.9.1 Synthesis of non-canonical amino acids

In this thesis, the three non-canonical amino acids propargyl-lysine (PrK), strained cyclooctyne-lysine (SCOK) and norbornene-lysine (NoK) were utilized for incorporation into *Saccharomyces cerevisiae* proteins (Figure 119). All three synthesis routes are based on the functionalization of Boc-protected L-lysine with the addition of a functional group, that can be used in CuAAC, SPAAC and SPIEDAC "click" chemistry respectively. While PrK and NoK were synthesized during this thesis, SCOK was synthesized during my master thesis[359]. Altogether, synthesis was done according to literature[200,201,288,385–387] and analysis of synthesized compounds was done with NMR spectroscopy. Utilizing NMR techniques like ^{1}H-^{1}H-2D-COSY and ^{1}H-^{13}C-2D-HSQC, it was possible to provide a correct assignment of signals, which was partly lacking in literature. Special care had to be taken during NoK synthesis to prevent hydrolysis of sensitive compounds and the possibility of leaking phosgene gas. Synthesis yields were comparable to literature values, however during NoK synthesis a problem with acidification of the reaction mixture overnight was observed. This was attributed to moisture entering the reaction and reacting with the excess triphosgene, producing HCl in the process. The acidic solution was believed to prematurely deprotect the Boc-protected amino group, which resulted in a loss of compound during column chromatography (the deprotected amino acid does not elute under used column chromatography conditions). With the addition of an excess of NaOH, to counter possible HCl production overnight, this problem could be fixed and in subsequent synthesis attempts NoK could also be obtained in good yield.

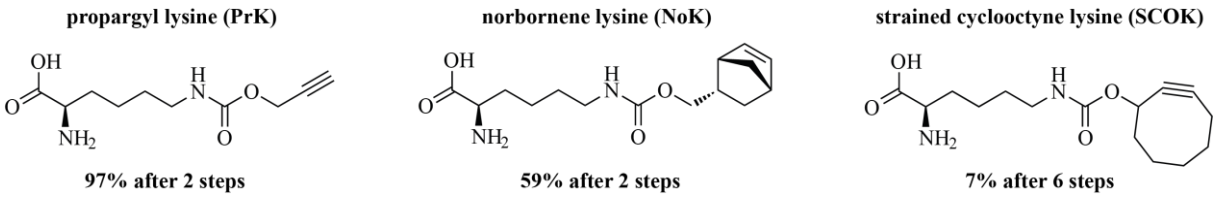


Figure 119: Overall yield of synthesized non-canonical amino acids utilized in this thesis.

4.9.2 Incorporation of non-canonical amino acids into proteins and fluorophore labeling

The incorporation of non-canonical amino acids into proteins by amber suppression provides a site-specific approach for selective labeling with a spectroscopic probe. A spectroscopic probe in the form of a fluorophore or spin label can then be attached to the protein post-transcriptionally for the application of smFRET or PELDOR spectroscopy on the snoRNP. Furthermore, this method only replaces a single amino acid in the peptide chain, which most likely does not interrupt the structural integrity and enzymatic activity of the protein and is not reliant on the replacement of several amino acids, like it would be necessary i.e. for site-

specific cysteine-labeling approaches. The non-canonical amino acid incorporation was performed at several labeling sites in Nhp2, Cbf5 and Gar1. Since no crystal structure of the full eukaryotic H/ACA RNP of snR81 (or any eukaryotic RNP at all) is available, the labeling sites had to be chosen based on the available crystal structure of the ternary Nop10-Cbf5Δ-Gar1Δ complex (pdb: 3U28) as well as the NMR structure of Nhp2 (pdb: 2LBW).

For Nhp2, the labeling sites were chosen on several criteria. Since Nhp2 and L7Ae show significant structural similarity, the labeling sites were chosen based on previous non-canonical amino acid incorporation into L7Ae[360,361]. Furthermore, the labeling sites had to be apart from the supposed binding sites of Nhp2 towards Nop10 and the snoRNA[362]. The expression times needed for efficient incorporation were analyzed and expression for 2-3 days seemed to work best. One important role played the sufficient supply of the expression system with the bioorthogonal aminoacyl-tRNA synthetase and tRNA which was produced by the arabinose inducible pEVOL plasmid. A delayed IPTG induction for the protein expression five hours after the induction with arabinose worked fine for efficient production of protein with incorporated non-canonical amino acid. The incorporation could be verified by Western blot analysis. Amber codon suppression still does compete with release factors and can result in truncated protein. The truncated protein in this case however does not contain the C-terminal His-Tag, which is only attached, if the amber codon at the incorporation sites gets suppressed and non-canonical amino acid gets incorporated. For this reason, all non-canonical amino acid incorporations were performed with C-tagged proteins. Likewise, the biorthogonality of the system could be verified, since without the supply of non-canonical amino acid during expression, no protein was visible on the Western blot analysis, showing that no canonical amino acid gets incorporated either by the host's or the biorthogonal aaRS/tRNA. It was however possible to vary the amount of non-canonical amino acid used for the expression to optimize the non-canonical amino acid amount required for a specific amount of medium, which led to more protein yield per amount of used amino acid. For Nhp2, six different labeling sites were tested, of which three (K37, K48 and K54) did result in good protein yield, while the other three (G70, K107, E152) only showed small amounts of protein. The exact reason for this discrepancy was not analyzed further, maybe the latter labeling sites had a negative effect on the protein structure of Nhp2 or amber suppression in the C-terminal region of the protein did not work as efficiently as for the N-terminal region. For Nhp2, the incorporation of all three non-canonical amino acids was performed successfully and utilized for several click-based labeling approaches. For SCOK and NoK incorporation, the mutant strain pEVOL^{AF} was used, which provides a bigger binding pocket in the aminoacyl-tRNA synthetase for the larger amino acids[200]. While for PrK- and SCOK-incorporated Nhp2, good yield could be obtained, for NoK-incorporated constructs the yield was only mediocre. Possible explanations would be that the pEVOL^{AF} mutant is not sufficient for efficient NoK incorporation and the aaRS requires additional modifications or NoK has a negative effect on the structural integrity of the protein.

With the PrK-incorporated constructs, copper(I) catalyzed click reactions were performed with fluorophore dyes sulfo-Cy3-azide and sulfo-Cy5-azide. The reactions were performed with 2.5 – 3 molar equivalents of fluorophores in comparison to protein concentration and excess fluorophore was removed via size exclusion chromatography. The copper-catalyzed approach seemed to somehow produce dimers of Nhp2 if the reaction was carried out for too

long. Since Nhp2 was relatively stable even at higher temperatures, the click reactions could be performed at 37 °C and the reaction time could be reduced to < 3 hours, resulting only in a negligible amount of the Nhp2 dimer.

To test out the opportunities of a copper free approach, the SCOK-incorporated constructs were reacted with sulfo-Cy3-azide, which resulted in labeled protein. However, labeling times had to be increased to 48 hours for efficient incorporation. While this seemed unproblematic for the stable Nhp2, it may be problematic for other eukaryotic proteins, which may not be able to handle the prolonged reaction time (which would have to be even more increased since reaction can not be carried out at 37 °C). The main reason for the testing of the copper free approach was the idea to incorporate an azide spin-label into the protein via this route, which would allow the application of EPR spectroscopy in addition to smFRET spectroscopy. With a double-labeled complex, generation of distance information in the form of PELDOR measurements would be possible. Since for the copper catalyzed approach, the use of a reducing agent is applied, this approach is not suitable for a nitroxide, since it would be reduced. After a nitroxide spin label with a tetrazine linker was provided by the Sigurdsson group, a second copper free approach was attempted with the use of the NoK incorporated Nhp2.

For Cbf5, incorporation of PrK was performed at the thumb-loop domain. This highly flexible domain can adapt an “open” state, in which it is bound to Gar1 and a target uridine in a substrate RNA can “enter” the catalytic center of Cbf5 and a “closed” state in which it pushed the target uridine to the catalytic aspartate in Cbf5[140,150,157,360]. Several labeling sites in the thumb-loop domain were chosen based on amino acids, which are not essential for the catalytic activity of the complex[379]. Labeling on full length and core domain-only Cbf5 (Cbf5Δ) was performed, and expression tests showed the best results for labeling site L156 for both Cbf5 constructs. The expression of these constructs was carried out together with Nop10 and the purification was done together with Gar1, since previous experiments suggested a weak interaction of Gar1 with the complex when just added to the rest of the proteins during complex assembly. While the purification of Cbf5 L156 suffered from very low yield and impurities that could not be removed, Cbf5Δ L156 could be obtained in better yield, however also with one protein impurity that could not be removed (see Figure 92), and was never observed during the expression and purification of wildtype Cbf5. One problem might lie in the prolonged expression for the non-canonical amino acid incorporation, which had to be extended to five days for efficient incorporation into Cbf5.

PrK incorporation was also performed for Gar1, for which several labeling sites were chosen. Gar1 with a C-terminal His-Tag was prepared for this purpose, so that during purification, the degradation products of Gar1 would not be co-purified along the full length Gar1. Since Gar1Δ was problematic during protein expression and purification, it was decided that the prolonged expression conditions of the PrK incorporation would not be suitable for Gar1Δ constructs. Expression tests of several Gar1 construct were performed and the successful incorporation at several labeling sites could be verified by Western blot analysis. However, during purification, only traces of Gar1 could be identified after the affinity chromatography step, which were not sufficient for a labeling approach with fluorophores. Further analysis showed, that Gar1 had already precipitated during cell lysis together with the cell debris and was no longer in solution, pointing towards protein aggregation and formation of inclusion

bodies. A small scale protein lysis in combination with changes of the buffer and addition of several additives showed that the addition of urea retained Gar1 in the cell lysate and not the precipitated pellet. This indicates some kind of folding problem for Gar1, causing protein precipitation. Since it was possible to readily express standalone wildtype Gar1, the incorporation of the non-canonical amino acid seems to result in the disruption of the Gar1 structure, causing instability and misfolding. The prolonged expression time for non-canonical amino acid incorporation may also contribute to this problem.
In recent studies, several new approaches for improved amber suppression, which can lead to higher yields of non-canonical amino acid incorporated proteins, have been reported. One major problem is the competition of release factors during the protein translation with the bioorthogonal amber suppression aaRS/tRNA. Bioorthogonal ribosomes and mRNA have been utilized and furthermore all amber codons in the host *E. coli* strain have been replaced by ochre codons, with only the amber codon utilized for amber suppression remaining[271–273]. This allowed the complete deletion of amber codon specific release factor RF1, preventing premature termination of the protein translation. Also, the interaction between elongation factor EF-Tu and tRNA loaded with a non-canonical amino acid were improved, i.a. by mutations to the tRNAs acceptor and T stem[268–270]. These improvements can result in a better incorporation efficiency, higher yield and make it also possible to incorporate several non-canonical amino acids at different incorporation sites at the same time[238,247,266,388]. Additionally, the utilization of quadruplet codons can completely negate the competition with release factors and can be combined with bioorthogonal ribosomes to efficiently incorporate several non-canonical amino acids simultaneously[243,244,247,272].

4.9.3 smFRET analysis of snR81 single hairpin constructs with fluorophore labeled proteins

The smFRET analysis of various constructs was performed with acceptor fluorophore labeled snoRNA and donor fluorophore labeled Nhp2. Constructs K37 and K48 of Nhp2 were used for the measurements. Based on the NMR structure of Nhp2, the crystal structure of Nop10 and a comparison with the crystal structure of the archaeal H/ACA RNP from *P. furiosus*, both labeling sites are expected to not interfere with Nop10 or snoRNA binding sites and therefore complex formation. Two different constructs, with labeling sites close to each other, on both ends of a helix structure, were used to see whether the collected data is reproducible and whether observed effects may be labeling site specific. For the snoRNA, both constructs were labeled with an acceptor fluorophore slightly above the pseudouridylation pocket. It has to be noted, that for H5 and H3, different “sides” opposite of each other of the Watson-Crick base-pairing stem were labeled, which should however not be an issue due to the large accessible volume of the fluorophore dye in comparison to the RNA, which should result in similar average dye positions in regard to the Nhp2 labeling sites.
For the measurements, aminosilanized and PEGylated glass slides were used to inhibit unspecific binding of proteins to the surface and at the same time allow the specific binding of snoRNP binding via biotin-streptavidin-biotin linking of biotinylated PEG and snoRNA. However, during measurements, still an unexpectedly large zero-FRET peak (E-FRET = 0) could be observed. Since the snoRNA carrying the acceptor dye was used as an

immobilization handle and the donor dye was attached to Nhp2 binding to the RNA, data points with low apparent FRET could arise from either unspecifically bound Nhp2, very early bleaching events, or from molecules exhibiting very low E-FRET. To solve this problem, a "molecule filter" was added into the FRET measurement. After the collection of the two second trace, that was used for E-FRET determination, the red laser was switched on additionally, which allowed the distinction between molecules, that were emitting fluorescence due to FRET and molecules that were emitting fluorescence only because of the excitation by the red laser. This effectively eliminated the zero-FRET peak from all measurements, showing that this population did not correspond to any FRET effect. Additionally, E-FRET below 0.2 was omitted from the data evaluation, since a dye distance for the applied system of more than 7.5 nm (see below for distance determination) was extremely unlikely, based on the assumption that a single hairpin adopts a similar structure than it was shown for archaeal hairpins by crystal structure analysis[140]. Besides, after the elimination of the zero-FRET peak by the applied filter, below an E-FRET of 0.2 only background noise was detectable, that could however interfere with the correct population distribution analysis.

For Cy3-Cy5 dye pairs, usually a Förster radius of R_0 = 6 nm is assumed[307,389]. The distance between donor and acceptor dye R can be determined by

$$R = \sqrt[6]{\frac{1}{E} - 1} \cdot R_0$$

However, distances calculated with this Förster radius can only be seen as a rough estimate, since the exact Förster radius is strongly dependent on the electrochemical environment of the fluorophores and for exact distance calculations, the Förster radius has to be experimentally determined by measurement of the quantum yield of the donor fluorophore and the overlap of the emission and excitation spectrum of both dyes for each individual construct[390,391].

For H5, the RNP reconstitution with all four *S. cerevisiae* proteins resulted in a single intermediate FRET state for both constructs. For construct K48, an E-FRET of 0.60 and for construct K37 an E-FRET of 0.50 was determined. This results in a dye distance of 5.61 nm for construct K48 and 6.00 nm for construct K37. If compared to the crystal structure of the archaeal H/ACA snoRNP from *P. furiosus* and an overlap of L7Ae with the NMR structure of Nhp2, than a shorter dye distance for K48 is actually to be expected, showing that the arrangement of Nhp2 towards the snoRNA may be similar to L7Ae in the 5' hairpin of the eukaryotic H/ACA RNP. The effects of GAR domain removal and complete removal of Gar1 from the RNP were determined and only had a negligible effect on the E-FRET value. However, especially the GAR domain removal led to a significant sharpening of the FRET peak, which got weaker again for the complete removal of Gar1. In absence of observable single molecule dynamics, a sharper FRET peak signifies a more rigid complex, since the dye molecules are more restricted in their accessible volume, while a broad peak suggests a more dynamic system. Somehow, the presence of only the core domain of Gar1 in comparison to full length Gar1 has a rigidifying effect on the complex structure of the RNP and may also be another indicator that Gar1 and its GAR domains directly interact with the snoRNA in some

way. The presence of substrate RNA in the measurement channel did not lead to any new populations. However, for construct H5 with the addition of substrate RNA, a sharpening of the FRET state to a narrower peak could be observed for both Nhp2 fluorophore constructs, which represents a more rigid complex (similar to GAR domain omission). The substitution of uridine to 5-fluoro-uridine, an inhibitor like substrate in the target RNA, also led to more narrow peaks, but not as strong as for the normal substrate.

The shortened H5Δ-construct was also analyzed in the same way as H5. Interestingly, this construct did not show a single binding mode for Nhp2 but rather three different binding modes, that could be identified with the analysis of two minute videos of the system and plotting of the different donor/acceptor intensities observed (Figure 120). Each observed molecule adapted one of the three conformations, and in nearly all the cases, during the course of the movie, the molecules remained in one FRET state and did not switch between them. This shows that the different binding states that Nhp2 adopts are either distinct binding modes of Nhp2, which are locked upon complex formation and do not transition into one another, or they represent a dynamic change in the complex structure, with transitions between states either taking much longer then the observed time window (minutes or even hours) or being much faster then the time resolution of the experiment. For the three states, for construct K48 an average E-FRET of 0.57 for the intermediate FRET state, 0.78 for the high FRET state I and 0.96 for the high FRET state II could be observed, which represent dye distances of 5.72 nm, 4.86 nm and 3.54 nm respectively. For construct K37 an average E-FRET of 0.49 for the intermediate FRET state, 0.77 for the high FRET state I and 0.89 for the high FRET state II could be observed, which represent dye distances of 6.04 nm, 4.91 nm and 4.23 nm respectively. Like for H5, the dye distance is shorter for construct K48 than for construct K37, which is to be expected, if the arrangement of Nhp2 in the 3’ hairpin is similar to archaeal L7Ae. With the removal of GAR domains as well as complete Gar1 omission from the constructs, the same peak sharpening that was observed for H5 could be observed for the high FRET state II of H5Δ. Furthermore, an increase of high FRET state II and decrease of high FRET state I could be observed upon removal of GAR domains / Gar1. While for the full RNP, high FRET state II represented a minor population, upon Gar domain / Gar1 removal, both high FRET states I and II were more evenly populated. High FRET state II may represent a similar binding mode of Nhp2 than in construct H5, however the dye distance is significantly decreased, for what most likely the removal of the upper stem loop structure is responsible. If Nhp2 is responsible for anchoring the upper stem loop structure in the 5’ hairpin, the removal of this structure seems to abolish the requirement of Nhp2 to firmly bind in one conformation and stabilize the complex, which leads to the different binding states. While in the presence of the upper-stem loop, a single binding state is heavily preferred, with the removal of the stem-loop structure, this preference is also abolished. The presence of substrate RNA did not lead to any significant changes in the distribution of the molecules between the three states. This may have different explanations. For once, it might be possible, that an actual pseudouridylation of substrate is not possible for the immobilized complexes on the glass surface. However, at least binding of the substrate RNA to the complex has been shown with immobilized RNPs[360,371], so at least the Watson-Crick base pairing with the snoRNA is not inhibited by surface immobilization. More likely, the actual position of Nhp2 in reference to the snoRNA might not change during the actual pseudouridylation reaction,

and Nhp2 might not be directly involved in the catalytic process, but more in a stabilizing process of the complex, as discussed before. So Nhp2 might influence and "catalyze" the pseudouridylation reaction by providing more stable and "active" complexes, which then can more effectively perform the enzymatic reaction.

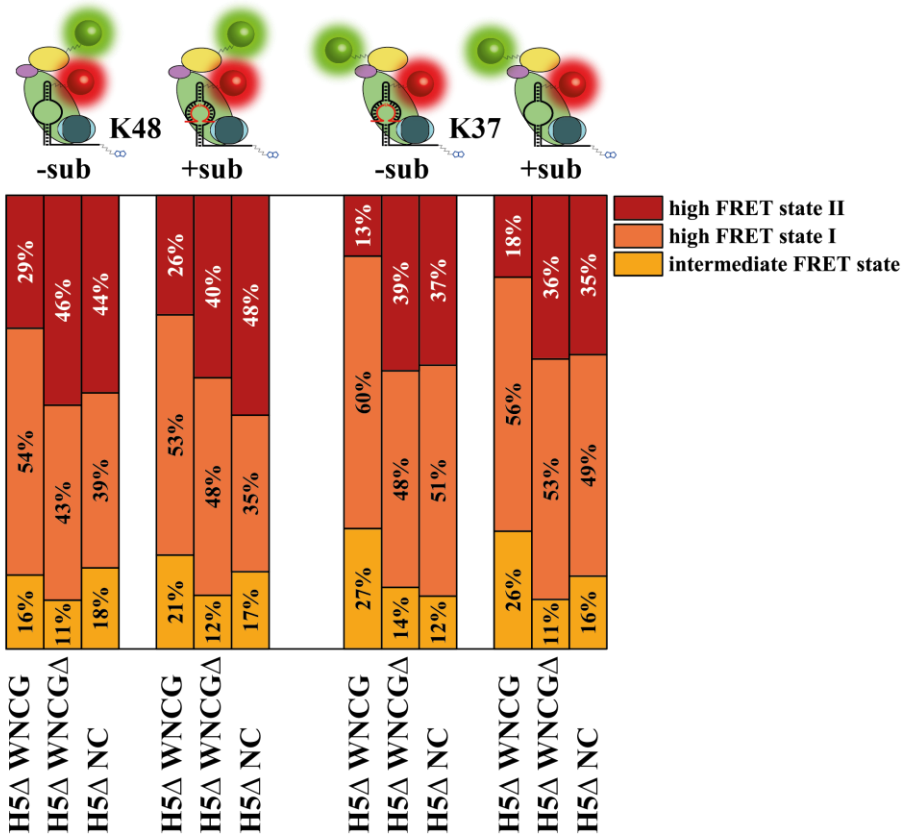


Figure 120: Population distribution of FRET states for snoRNP construct H5Δ-Cy5 + Nhp2K48Cy3 / Nhp2K37Cy3 and WNCG / WNCGΔ / WNC without substrate (-sub) and in the presence of substrate (+sub). The overall comparable distribution between states for both labeling schemes indicates validity of labeling sites and limited structural disturbance by labels.

Similar to the 5' hairpin constructs, Cy5-labeled H3 was analyzed under the same conditions. Interestingly, even for the full H3 construct (including the upper stem structure), several different populations could be observed, as was the case for construct H5Δ (Figure 121). With the analysis of the collected two minute movies, three states could be pinpointed. For construct K48 an average E-FRET of 0.43 for the low FRET state, 0.65 for the intermediate FRET state and 0.82 for the high FRET state could be observed, which represent dye distances of 6.29 nm, 5.41 nm and 4.66 nm respectively. For construct K37 an average E-FRET of 0.43 for the low FRET state, 0.60 for the intermediate FRET state and 0.77 for the high FRET state could be observed, which represent dye distances of 6.29 nm, 5.61 nm and 4.91 nm respectively. Again, the dyes distances for construct K37 are slightly higher than for construct K48, suggesting a similar orientation of Nhp2 towards the pseudouridylation pocket than it was the case for the 5' hairpin constructs.

Additionally, the dye distances of the three states are higher than what was observed for construct H5Δ, suggesting that the different binding modes of Nhp2 are further away from the pseudouridylation pocket. Again, no transitions between the different states could be observed, pointing to an extremely slow complex rearrangement or a locking of the binding modes of Nhp2 after the complex assembly with no structural rearrangement afterwards. Both the intermediate and high FRET states were the dominant states for both constructs, indicating that one of these two states might represent an "active" state of Nhp2, while the other state represents an "inactive" or "misbound" state. This observation is in line with the activity data from chapter 3, where it was shown that H3 shows a drastically decreased activity in comparison to H5. This may be caused by an inactive state of Nhp2, that does not contribute to complex activity and may be responsible for the lower activity in comparison to H5, where Nhp2 only binds in a single conformation. Since H3 is drastically dependent on the effects of Nhp2 (H3 in the full length snR81 is nearly completely inactive without Nhp2), the misbound conformation of Nhp2 might also contribute to the lower complex activity, even in the presence of Nhp2. With the removal of the GAR domains from Gar1, the population of the high FRET state was slighly increased. If the high FRET state is assumed to be the "active" state of Nhp2, the increase of this state upon GAR domain removal could be brought into correlation with the higher enzymatic activity upon GAR domain removal.

Building on this hypothesis, the GAR domains of Gar1 may have an influence on the snoRNA, that influences the "correct" binding of Nhp2. With GAR domains present, the correct positioning of Nhp2 to the 3' hairpin is disturbed and Nhp2 more often binds in the "inactive" state, albeit reducing the complex activity. This observation may actually be *in vitro* specific, since during *in vitro* complex reconstitution, the trimeric complex of Nop10-Cbf5-Gar1 is formed first, while in the *in vivo* complex assembly, the trimeric complex Nhp2-Nop10-Cbf5 is formed first, with Gar1 being the last protein binding to the complex and Naf1 (which is replaced by Gar1 at the end) not containing any GAR domains (also discussed in chapter 3.6). With the removal of the GAR domains, the negative influence of Gar1 towards the H3 snoRNA is weakened, which allows Nhp2 to bind in the correct state, increasing the "active" FRET state population as well as the activity of the construct. However, the change of state populations was not as drastic as the change in enzymatic activity, so the higher enzymatic activity induced by GAR domain removal and the change of FRET populations for Nhp2 might be two completely independent effects and Gar1 might not have an influence on Nhp2 binding at all. Summed up, the data shows that H3 seems to not only be in need of the presence of Nhp2 for effective pseudouridylation, but furthermore suffers from wrongly bound Nhp2, which may be influenced by Gar1 GAR domains.

Similar to the 5' hairpin, the presence of substrate RNA did not lead to any significant changes in the distribution of the molecules between the three states, leading to the same assumptions made for the 5' hairpin for the substrate RNA effect on Nhp2 binding.

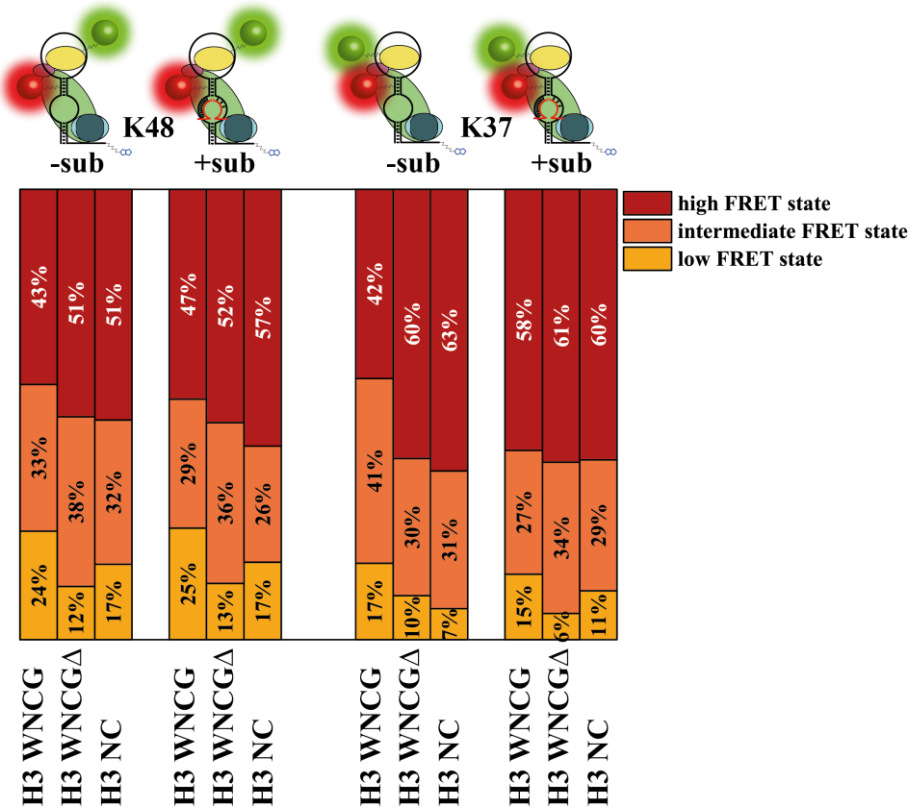


Figure 121: Population distribution of FRET states for snoRNP construct H3-Cy5 + Nhp2K48Cy3 / Nhp2K37Cy3 and WNCG / WNCGΔ / WNC without substrate (-sub) and in the presence of substrate (+sub). The overall comparable distribution between states for both labeling schemes indicates validity of labeling sites and limited structural disturbance by labels.

Construct H3Δ did also show several FRET states upon analysis (Figure 122). Interestingly, these states could be identified as states with the same E-FRET that were observed for H3 before via analysis of single-molecule fluorescence time traces. It appears that even in the absence of the upper stem loop structure, Nhp2 binds in a similar conformation than it does with the full H3 construct. Contrary to construct H5, the removal of the upper stem loop structure seems to not affect the binding states of Nhp2. This is even more surprising, since for construct H5, the removal of the upper stem loop resulted in different FRET states, while only mildly affecting the activity of the complex, while for construct H3, the removal of the stem loop structure did not change the FRET states, but at the same time resulted in complete activity loss for the complex. This observation again points to different requirements and functions of Nhp2 for the 5' and 3' hairpin and may also be attributed to different formations of the RNP on each hairpin.

While for construct H3, a major state could be identified, which might be brought into correlation with the activity data, for construct H3Δ the molecules were more broadly distributed between all three states. Upon GAR domain/Gar1 removal, the high FRET state was also more strongly populated, similar to construct H3. Also the low FRET state was more strongly populated than for the other constructs. This broader distribution may represent a more unspecific binding of Nhp2 between the three states, which is induced by the absence of the upper stem loop structure. While the presence of this stem loop structure, of which the

presence is critical for complex activity, induces the correct positioning of Nhp2, its absence leads to the distribution between “active” and “inactive” states of Nhp2, which however does not matter, since the construct H3 is inactive anyway without the full hairpin structure present.
Stronger fluctuations between the measurements with and without substrate RNA could be detected, which might also be attributed to the more “unspecific” behavior of Nhp2 for the construct H3Δ. Furthermore, the inability of H3Δ to perform any catalytic activity at all may be represented by this behavior, possibly due to substrate RNA binding to the complex via Watson-Crick base-pairing, while the rearrangement towards the catalytic core and/or the U to Ψ reaction are not happening.

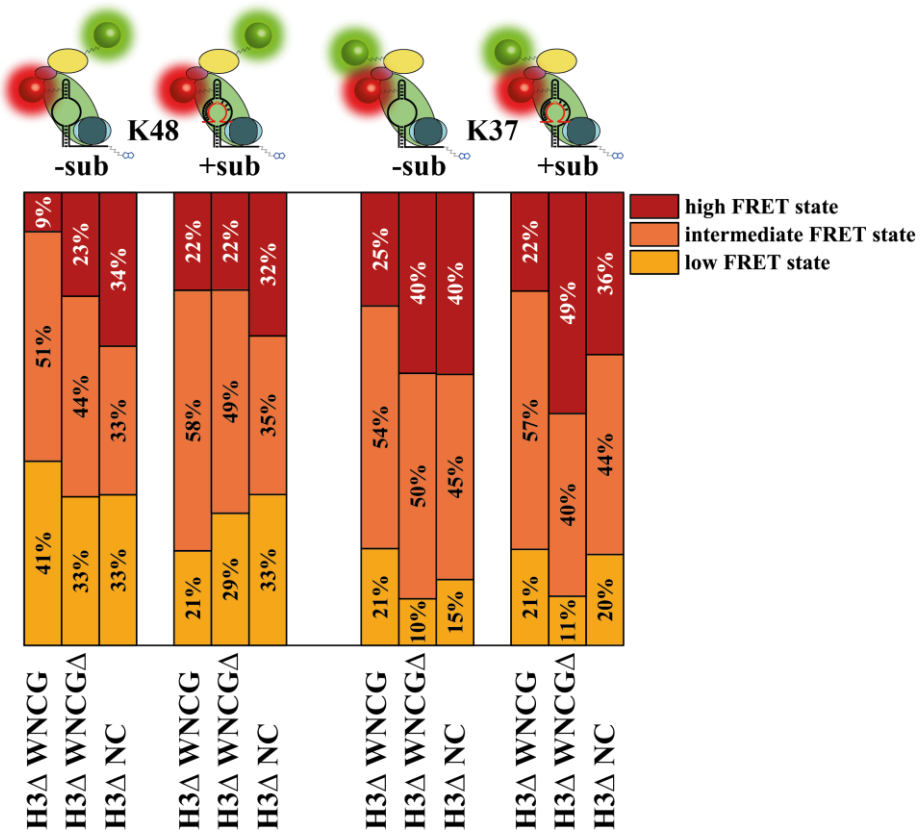


Figure 122: Population distribution of FRET states for snoRNP construct H3Δ-Cy5 + Nhp2K48Cy3 / Nhp2K37Cy3 and WNCG / WNCGΔ / WNC without substrate (-sub) and in the presence of substrate (+sub). The overall comparable distribution between states for both labeling schemes indicates validity of labeling sites and limited structural disturbance by labels.

In addition to Nhp2 labeled constructs, the Cy5-labeled constructs H5 and H3 were also analyzed with Cy3-labeled Cbf5Δ. More specifically, the label was attached to the thumb-loop domain of Cbf5, a highly flexible domain, which adopts an “open” and “closed” state in the archaeal H/ACA sRNP from *Pyrococcus furiosus* during catalysis[140,150,157,162]. In an earlier work, these different states could be identified with a fluorophore labeled thumb-loop domain and snoRNA[360]. However, since *P. furiosus* is highly thermophilic and the archaeal sRNP is catalytic active at ~90 °C and mostly inactive at room temperature, it was not possible to perform real time measurements of the thumb loop dynamic (since the microscope

can not be heated up to a temperature where the sRNP would be active). Instead, the complex was trapped during different states of catalysis by heating to catalytic temperatures and direct storing on ice afterwards with different targets RNAs (substrate, product, inhibitor etc.) and afterwards a "snapshot" of the complex at room temperature was made. While this provided interesting data into the different states of the thumb loop domain during catalysis, it was not possible to generate "real time" dynamic traces of the complex, with an opening and closing of the thumb loop domain observable during the measurement.

The smFRET measurements with labeled H5/H3 and the labeled thumb loop domain of the *S. cerevisiae* snoRNP showed a very sharp single E-FRET state of 0.88 and 0.89, which represent a dye distance of 4.30 nm and 4.23 nm respectively. The analysis of two minute movies also showed this state and typical anticorrelated FRET behavior was observable. However, no population for a second FRET state – and hence a conformational change of the thumb loop – could be observed. For once, it might be possible that similar to the stable FRET states during the addition of substrate to the Nhp2-labeled complexes, the immobilized RNP particles are simply not able to perform catalytic activity. The binding via linker to the glass surface may in some way disrupt the catalytic activity or the low concentration, in which the complex was added onto the glass surface may in some way affect activity (concentration in activity assays was 100-1000 fold higher), although eukaryotic RNPs have been shown to have extremely low dissociation constants[159].

Another explanation might be, that the observed state represents just the "open" state of the complex, while the states with a "closed" thumb loop are only adapted for a very short time window (<<< 1 ms), which would not be observable by smFRET. Analysis of the H/ACA RNP from *P. furiosus* showed, that during the initial association of the substrate, a kinetic intermediate is formed which is largely devoid of thumb interaction (ES1) (Figure 123)[162]. Afterwards, the thumb loop adopts a "closed" state, which results in a more stable complex (ES2). After the reaction of U to Ψ (EP2), the thumb loop has no energetic contribution to the binding of the product anymore and again adopts an "open" state (EP1). The observed FRET states might represent the "open" loop state, which is adapted if either no substrate is bound (E+S, E+P), after the substrate RNA gets recruited by Watson-Crick base pairing, but before the correct placement at the catalytic site (ES1) and directly before product release (EP1) (Figure 123). The fastest observed turnover rate for all constructs utilized in this thesis was approximately 15 seconds for one turnover cycle, while for *P. furiosus* (at room temperature) 250 seconds for one cycle[162] and for TruA, TruB and RluA approximately 2 seconds for one cycle[109] were observed. While the substrate loading[162], substrate positioning[110] or the product release[109] have been proposed to be rate-limiting steps for some pseudouridylases, in TruA and TruB the actual catalytic step was shown to be the rate-limiting factor during pseudouridylation[109,111]. Since the actual uridine-to-pseudouridine reaction is assumed to be similar in protein-only pseudouridylases and H/ACA RNPs[66], it is rather unlikely that the catalytic reaction is "too fast", and the states ES2 and EP2 would be adapted for a longer time frame, making the "closed" state of the thumb loop observable by smFRET. Also, *in vitro*, H/ACA RNPs are suspected to bind and dissociate relatively fast from cognate as well as near-cognate RNAs, in order to search for the correct target among a large range of RNA sequences[377].

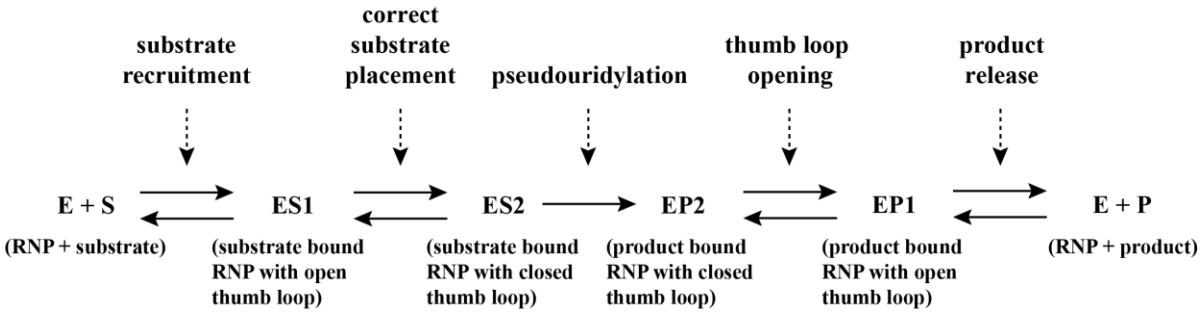


Figure 123: Schematic representation of the assumed reaction kinetic for the H/ACA RNP pseudouridylation reaction. In a first reversible step, the substrate is recruited via Watson-Crick base-pairing. Afterwards, the target uridine is placed at the catalytic center via thumb loop interactions. In an irreversible step the uridine to pseudouridine isomerization takes place. The thumb loop has no interaction with the pseudouridine and opens, which leads to product release[162].

A third explanation lies in a different structural alignment of the eukaryotic thumb loop domain compared to archaeal Cbf5 altogether. While the thumb loop domain seems to be partially disordered and was not resolved as part of the crystal structure of eukaryotic Cbf5 in the Nop10-Cbf5Δ-Gar1Δ subcomplex[160], still some assumptions can be made by the alignment and comparison of the P. *furiosus* and *S. cerevisiae* crystal structures. The "open" conformation of the eukaryotic thumb loop actually seems to be more similar to the "closed" conformation of the archaeal thumb loop[160]. This results from several Gar1 amino acids physically blocking the docking site, that the eukaryotic thumb loop would adapt in an archaeal like "open" conformation and instead the eukaryotic thumb loop interacts with a hydrophobic pocket of Gar1, placing the loop in a position similar to the archaeal "closed" state while it is in the "open" state[160]. This may result in only very subtle conformational changes upon the adaption of a "closed" conformation during the substrate loading and the adaption of the "open" conformation during substrate release. However, smFRET spectroscopy might still be able to detect such small changes.

4.9.4 Analysis of the full length snR81 bipartite complex

After the successful analysis of the standalone hairpin constructs H5 and H3, it was also tried to gain insights into the full snR81 snoRNA construct, containing both 5' and 3' hairpins. For this approach, the full length snoRNA was incubated with an excess of the *S. cerevisiae* proteins. Cy3- as well as Cy5-labeled Nhp2 was used. Since it is not possible to specifically label one hairpin in the complex in the presence of the other hairpin (using a protein-labeling approach), the strategy was to statistically distribute the donor and acceptor fluorophores between the two hairpins, so that at least 50% of reconstituted complexes would contain both a donor and an acceptor labeled Nhp2. However, no smFRET signal could be detected with the double labeled complex. In a different approach, Nhp2 with an attached nitroxide spin label was utilized. In theory, one molecule of Nhp2 should interact with each hairpin, resulting in a double spin labeled complex, which should be able to provide a PELDOR signal. But also in PELDOR measurements, no signal could be obtained for the complex. An important point to notice is the fact that both smFRET as well as PELDOR spectroscopy are

applicable for a label distance of up to 10 nm[292,392]. If for some reason, both labeling sites and hence both Nhp2 molecules are further apart than 10 nm, both methods would not be able to provide a signal.

In a recent study, a cryo-EM structure of substrate bound human telomerase was solved[169]. The human telomerase RNA hTR contains an H/ACA motif in the 3' region, which also adopts a double-hairpin structure and both hairpins associate with as set of the four H/ACA proteins, forming a bipartite complex. The provided structure does not suggest a parallel arrangement of the two RNA hairpins but rather a kink of 90° between both hairpins. While this comparison is of limited value, since hTR has not been established as a pseudouridylation guide RNA, and one of the hairpins contains an extension that is required for telomerase activity, the idea of an angled hairpin structure could be applied to the snR81 H/ACA RNP.

Based on the crystal structure of the *P. furiosus* H/ACA sRNP, the RNA interacts in a way with the proteins that from top to bottom, the complex approximately spans a range of 10 nm[140]. If snR81 indeed forms a similar kink-like structure and additionally the proteins are arranged in a similar fashion then in the archaeal complex, then the distance between both Nhp2 molecules would be greater than 10 nm. If this holds true, the absence of both detectable smFRET signals with fluorophore labeled Nhp2 as well as no measurable PELDOR distance between spin labeled Nhp2 might provide a hint that indeed the bipartite structure of the complex may be arranged in a way that the hairpins are arranged in a considerable angle to one another (Figure 124).

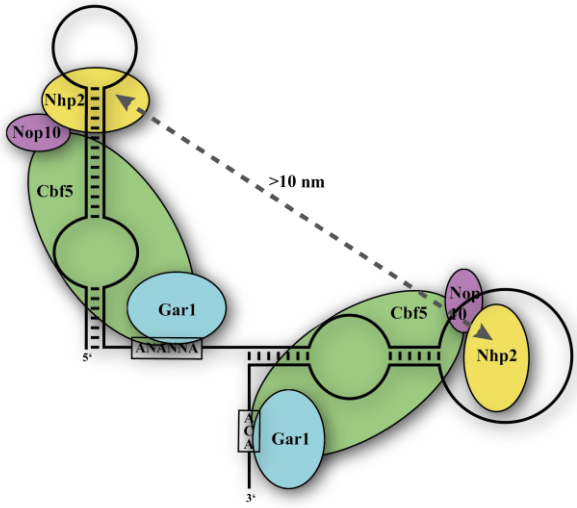


Figure 124: Schematic illustration of the bipartite complex structure with angled hairpins. This arrangement would result in the upper parts of the two hairpins being further apart then 10 nm, which could explain the absence of both FRET and PELDOR signals.

4.10 Appendix of chapter 4

4.10.1 Preparation of Nhp2 K37PrK

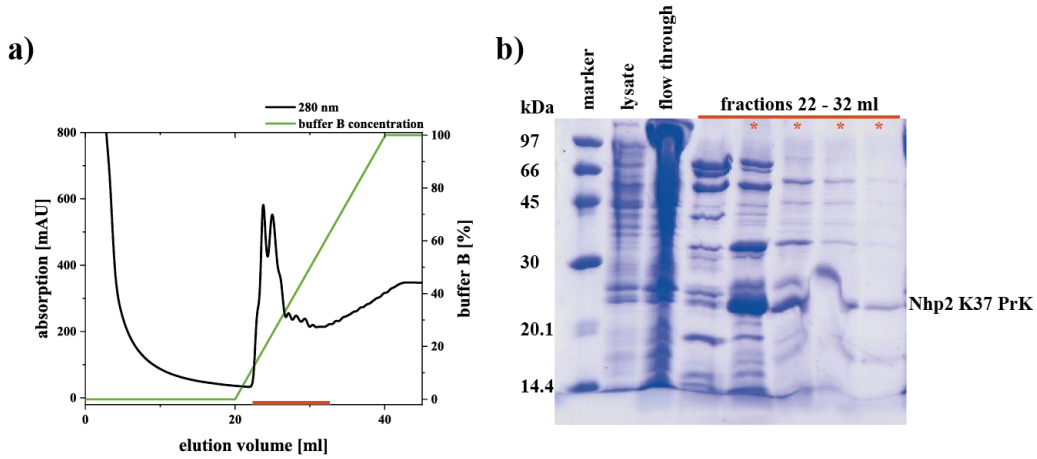


Figure 125: Affinity chromatography purification of Nhp2K37PrK before RNase A treatment. a) Ni-NTA elution chromatogram using an imidazole gradient. Fractions used for analysis are marked with a red bar. b) PAGE analysis of fractions from Ni-NTA. Fractions used for further purification are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

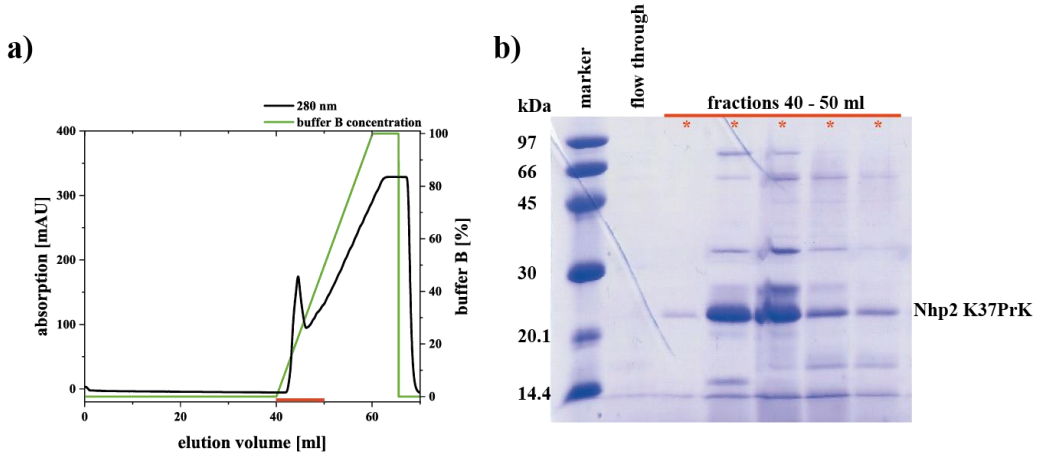


Figure 126: Affinity chromatography purification of Nhp2 K37PrK after RNase A treatment. a) Ni-NTA elution chromatogram using an imidazole gradient. Fractions used for analysis are marked with a red bar. b) PAGE analysis of fractions from Ni-NTA. Fractions used for further purification are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

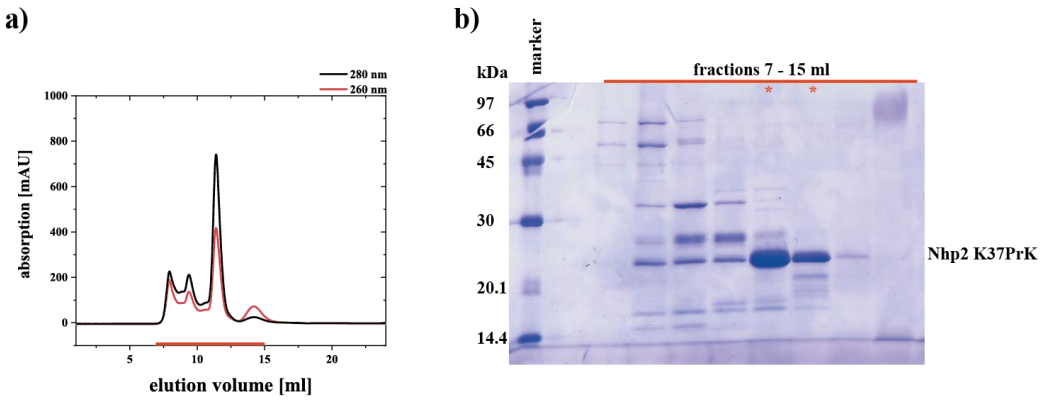


Figure 127: Size exclusion chromatography purification of Nhp2 K37PrK. a) SEC elution chromatogram using a Superdex 75 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) PAGE analysis of fractions from SEC. Fractions used for further experiments are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

4.10.2 Preparation of Nhp2 K54PrK

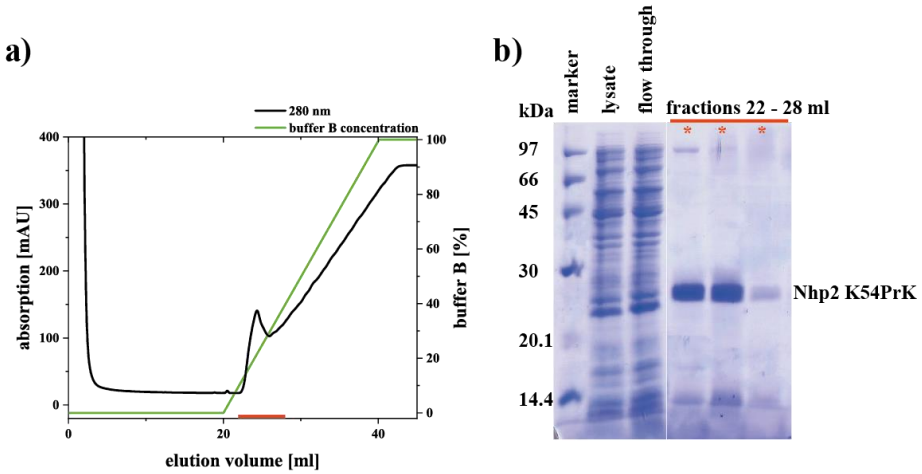


Figure 128: Affinity chromatography purification of Nhp2K54PrK before RNase A treatment. a) Ni-NTA elution chromatogram using an imidazole gradient. Fractions used for analysis are marked with a red bar. b) PAGE analysis of fractions from Ni-NTA. Fractions used for further purification are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

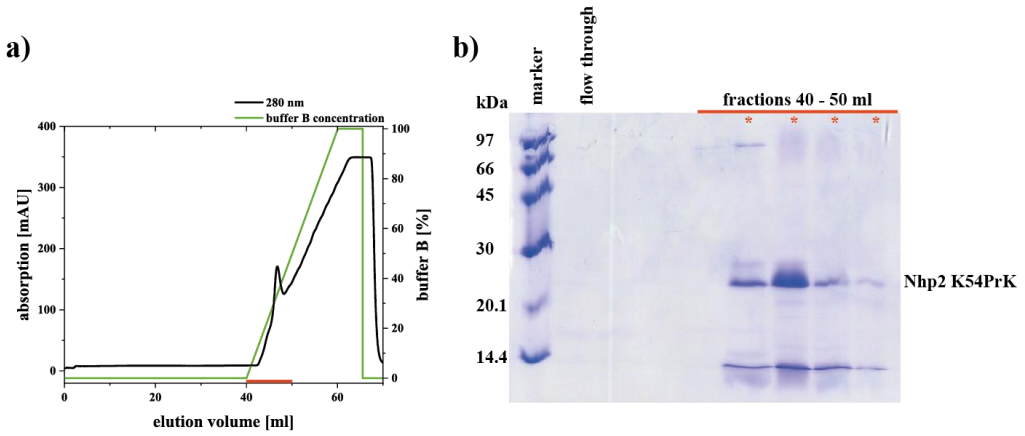


Figure 129: Affinity chromatography purification of Nhp2 K54PrK after RNase A treatment. a) Ni-NTA elution chromatogram using an imidazole gradient. Fractions used for analysis are marked with a red bar. b) PAGE analysis of fractions from Ni-NTA. Fractions used for further purification are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

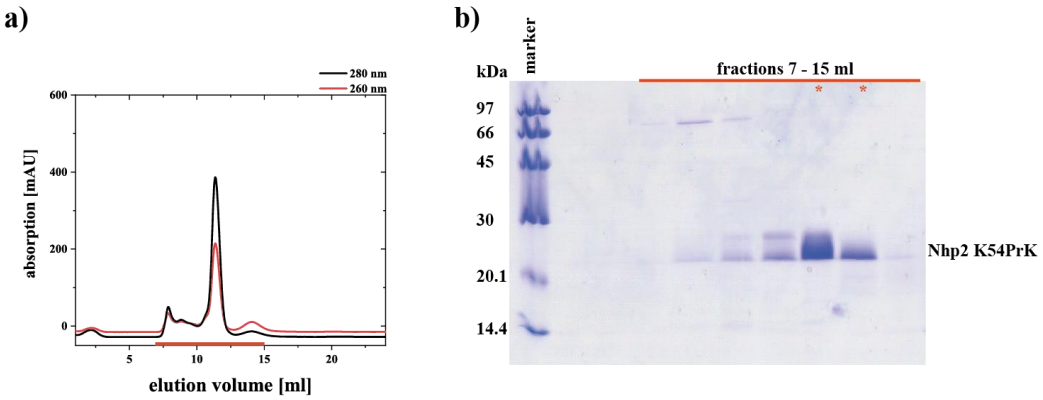


Figure 130: Size exclusion chromatography purification of Nhp2 K54PrK. a) SEC elution chromatogram using a Superdex 75 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) PAGE analysis of fractions from SEC. Fractions used for further experiments are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

4.10.3 Preparation of Nhp2 K37Cy3

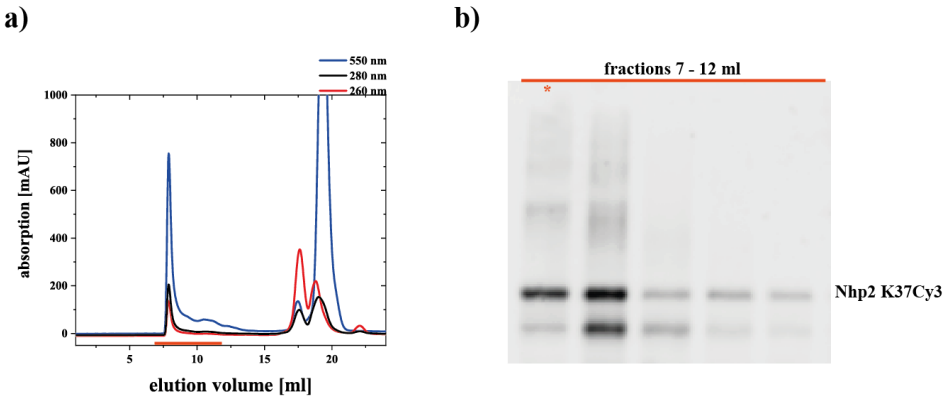


Figure 131: Size exclusion chromatography purification of Nhp2 K37Cy3. a) SEC elution chromatogram using a Superdex 75 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) PAGE analysis of fractions from SEC. Fractions used for further experiments are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

4.10.4 MALDI of Nhp2 constructs

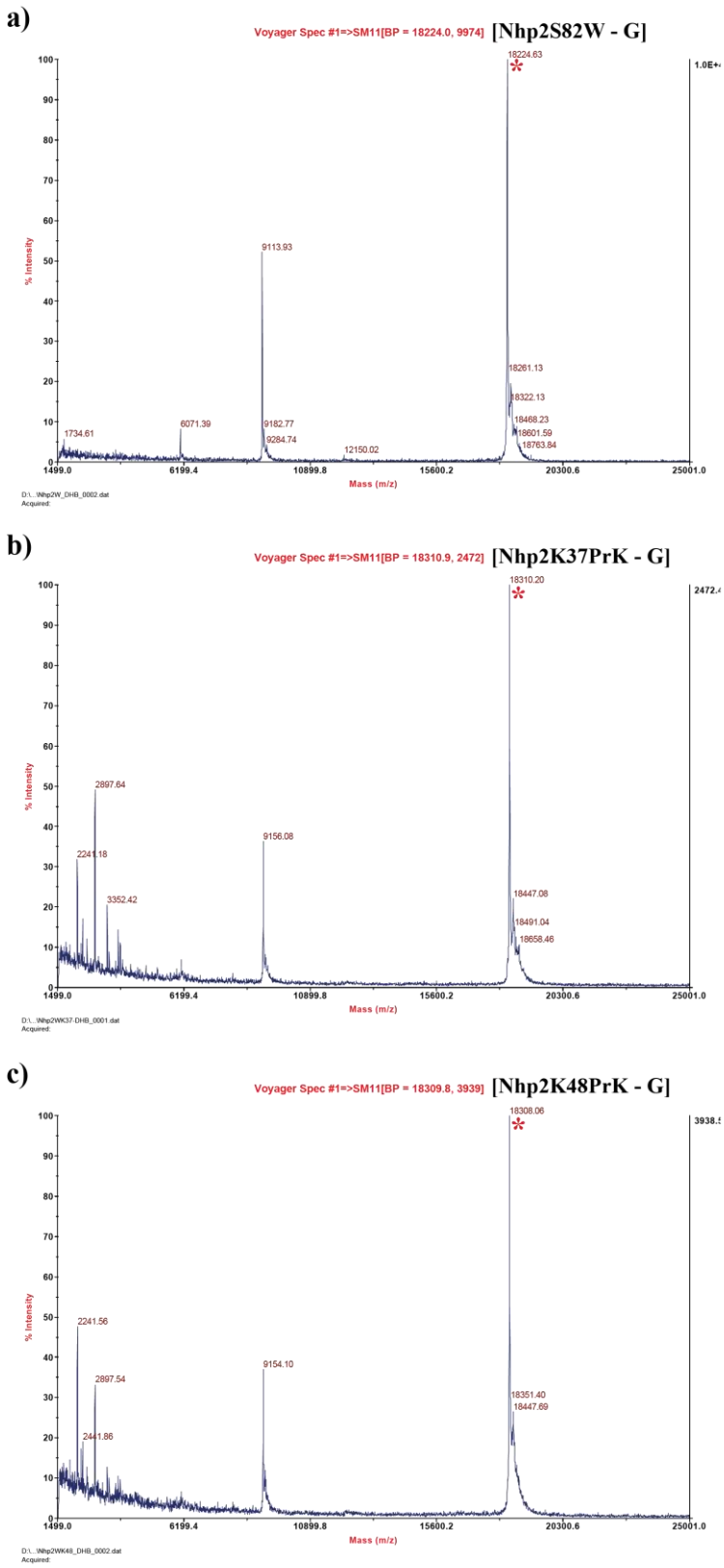


Figure 132: MALDI data for Nhp2 (a), Nhp2K37PrK (b) and Nhp2K48PrK (c).

4.10.5 Multiple turnover with fluorophore labeled Nhp2

Table 9: reaction kinetics raw data for labeled Nhp2

H5 W-Cy3 NCG

time [min]	assay 1
5	43.76
20	73.27
60	83.64
120	86.52

H5 W-Cy5 NCG

time [min]	assay 1
5	25.47
20	73.86
60	84.82
120	87.63

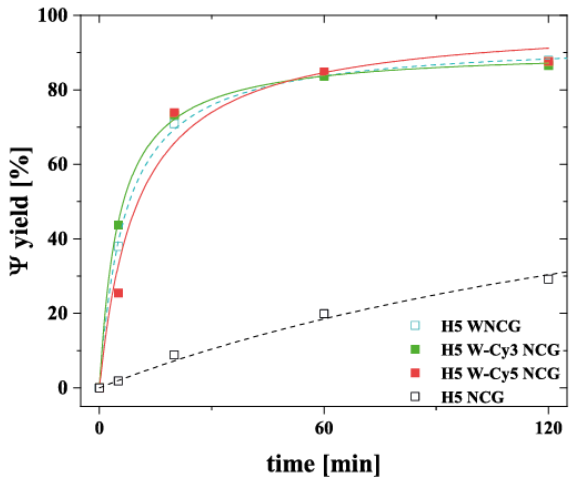


Figure 133: Time course of pseudouridylation reaction under multiple turnover conditions of H5 snoRNA with Nhp2K48Cy3 (W-Cy3) and NCG (green) and Nhp2K48Cy5 (W-Cy5) and NCG (red) in comparison to WNCG (cyan) and NCG (black).

4.10.6 Incorporation of PrK into Cbf5Δ L156

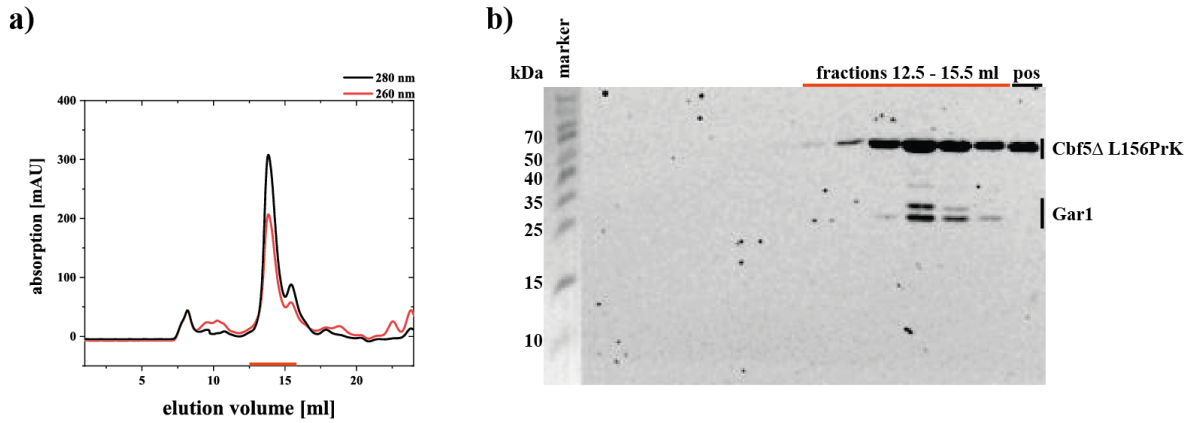


Figure 134: Size exclusion chromatography purification of Cbf5Δ L156PrK together with Nop10 and Gar1. a) SEC elution chromatogram using a Superdex 200 10/300 GL increase column. Fractions used for further analysis are marked with a red bar. b) Western blot analysis of fractions from SEC. Marker: Spectra Multicolor Broad Range Protein Ladder (Thermo Scientific). positive control: Cbf5Δ

4.10.7 Cbf5 L156Prk and Cbf5Δ L156PrK Click tests

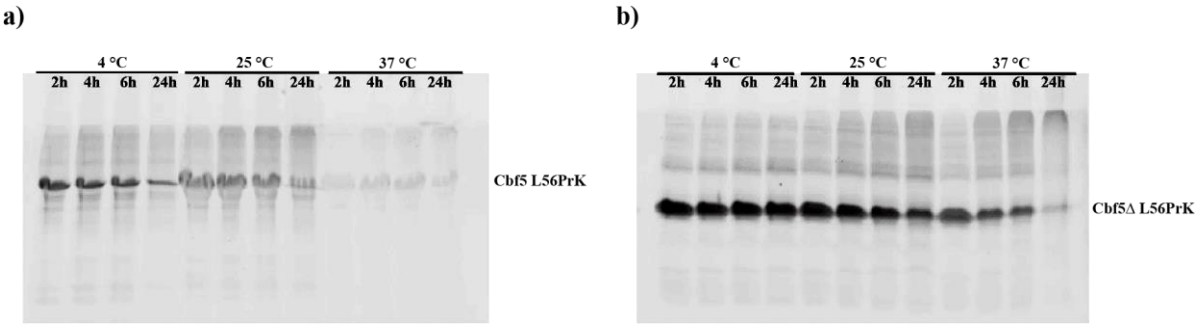


Figure 135: SDS PAGE fluorescence scan of test click reactions for Cbf5 L156PrK with sulfo-Cy3-azide (a) and Cbf5Δ L156PrK with sulfo-Cy3-azide.

4.10.8 Incorporation of PrK into Gar1 constructs

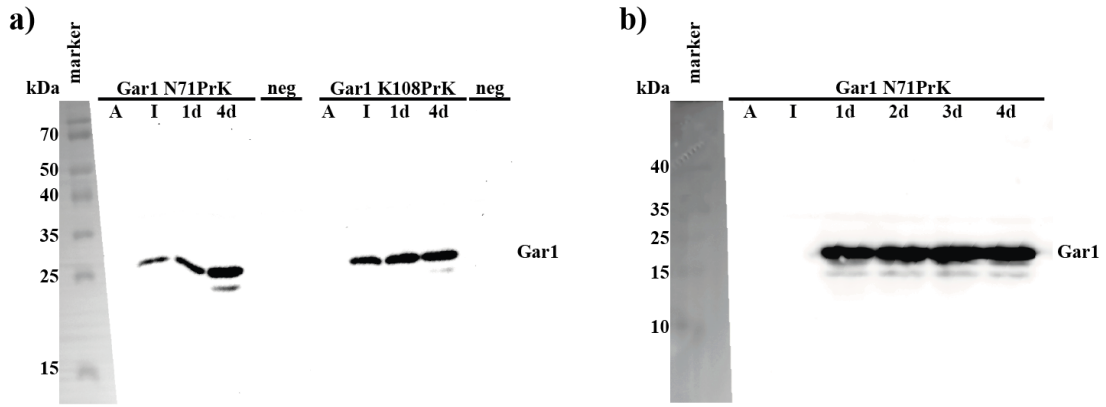


Figure 136: Anti-His Western blots of expression test of PrK incorporation into the Gar1 constructs N71 and K108. a) Incubation at 37 °C (a) and 20 °C (b) after arabinose induction. Marker: Spectra Multicolor Broad Range Protein Ladder (Thermo Scientific). negative control: no addition of PrK into expression medium.

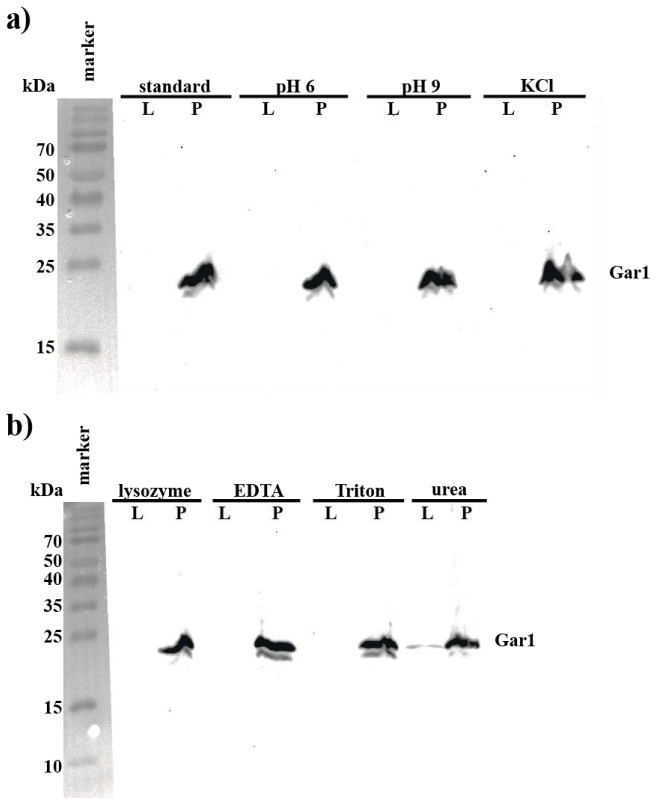


Figure 137: Lysis test of Gar1 N71PrK with different buffer conditions (a) and different additives (b). Gar1 N71PrK can only be identified in the cell debris pellet. Only the addition of urea leads to observable Gar1 in the lysate. P=pellet, L=lysate.

4.10.9 Fluorophore labeled RNA constructs

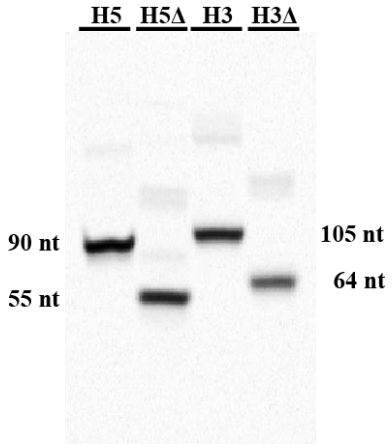


Figure 138: SDS denaturing 10% PAGE fluorescence scan of snap cooled, Cy5-labeled RNA constructs.

4.10.10 smFRET measurements with inverted labeling scheme

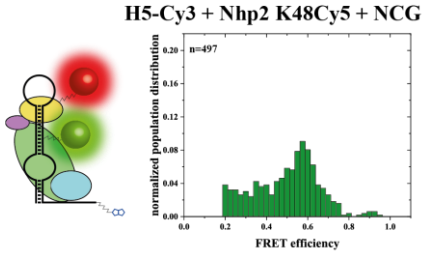


Figure 139: a) Histogram of fully reconstituted 5' hairpin complex H5-Cy3 + Nhp2 K48Cy5 + NCG. n represents the number of molecules observed for E-FRET ≥ 0.2. The labeling scheme for donor and acceptor molecules was inverted for this control experiment.

4.10.11 smFRET measurements without NCG and without Nop10

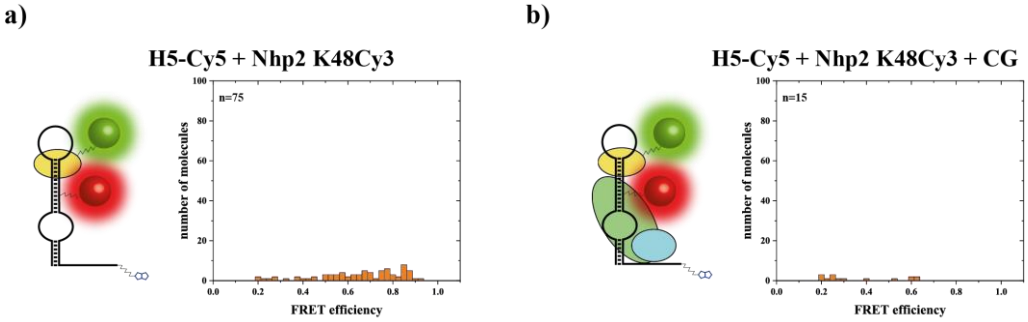


Figure 140: a) Histogram of 5' hairpin complex H5-Cy5 + Nhp2 K48Cy3 (omission of NCG). b) Histogram of 5' hairpin complex H5-Cy5 + Nhp2 K48Cy3 + CG (omission of Nop10). n represents the number of molecules observed for E-FRET ≥ 0.2. No populations can be observed for the two control experiments.

4.10.12 smFRET exemplary traces for H3Δ

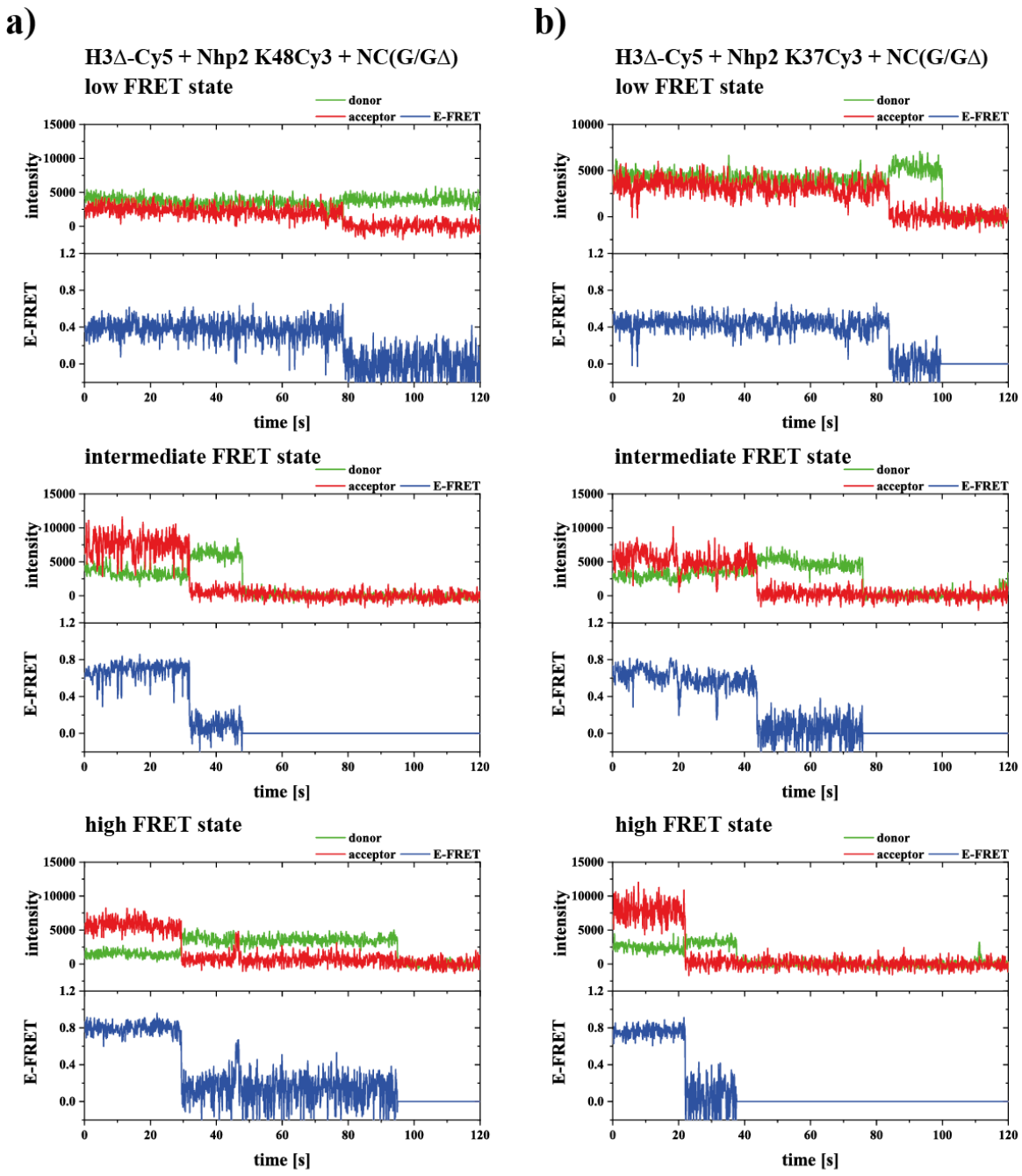


Figure 141: Exemplary time resolved FRET traces, representing donor and acceptor intensity as well as corresponding E-FRET. Traces for the three observed states of H3Δ-Cy5 with Nhp2 K48Cy3 (a) and Nhp2 K37Cy3 (b).

4.10.13 smFRET measurements in presence of substrate RNA

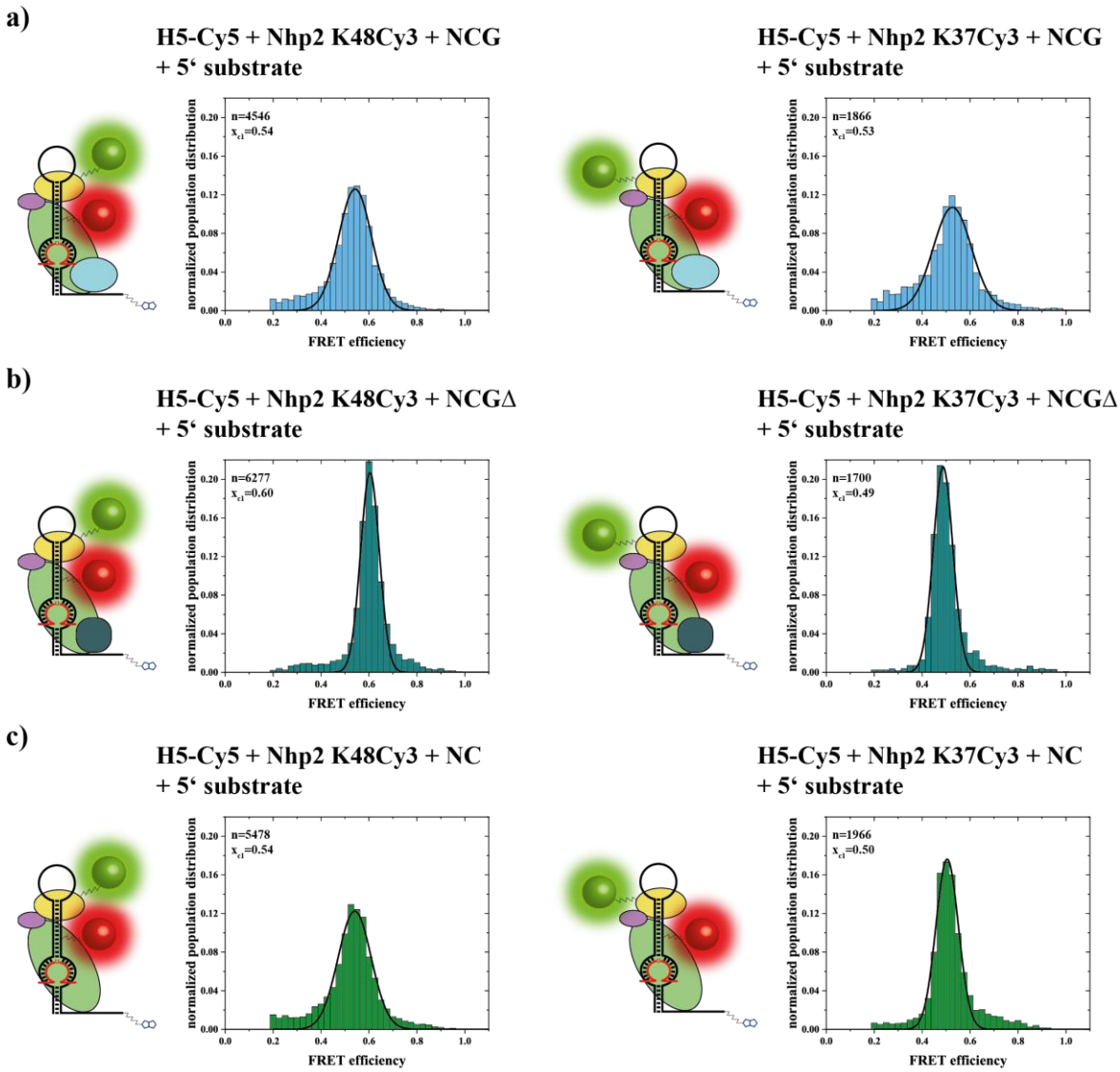


Figure 142: a) Histograms of fully reconstituted 5’ hairpin complexes H5-Cy5 + Nhp2 K48Cy3 + NCG (Ieft) and H5-Cy5 + Nhp2 K37Cy3 + NCG (right) in presence of 5’ substrate RNA. b) Histograms of GAR domain omission for the 3’ hairpin complexes H5-Cy5 + Nhp2 K48Cy3 + NCGΔ (Ieft) and H5-Cy5 + Nhp2 K37Cy3 + NCGΔ (right) in presence of 5’ substrate RNA. c) Histograms of Gar1 omission for the 3’ hairpin complexes H5-Cy5 + Nhp2 K48Cy3 + NC (Ieft) and H5-Cy5 + Nhp2 K37Cy3 + NC (right) in presence of 5’ substrate RNA. n represents the number of molecules observed for E-FRET ≥ 0.2. The states x_c including error were determined by analyzing 120 second movies. Gaussian fitting within the error of x_c was used to determine and plot the Gaussian distribution in the histogram.

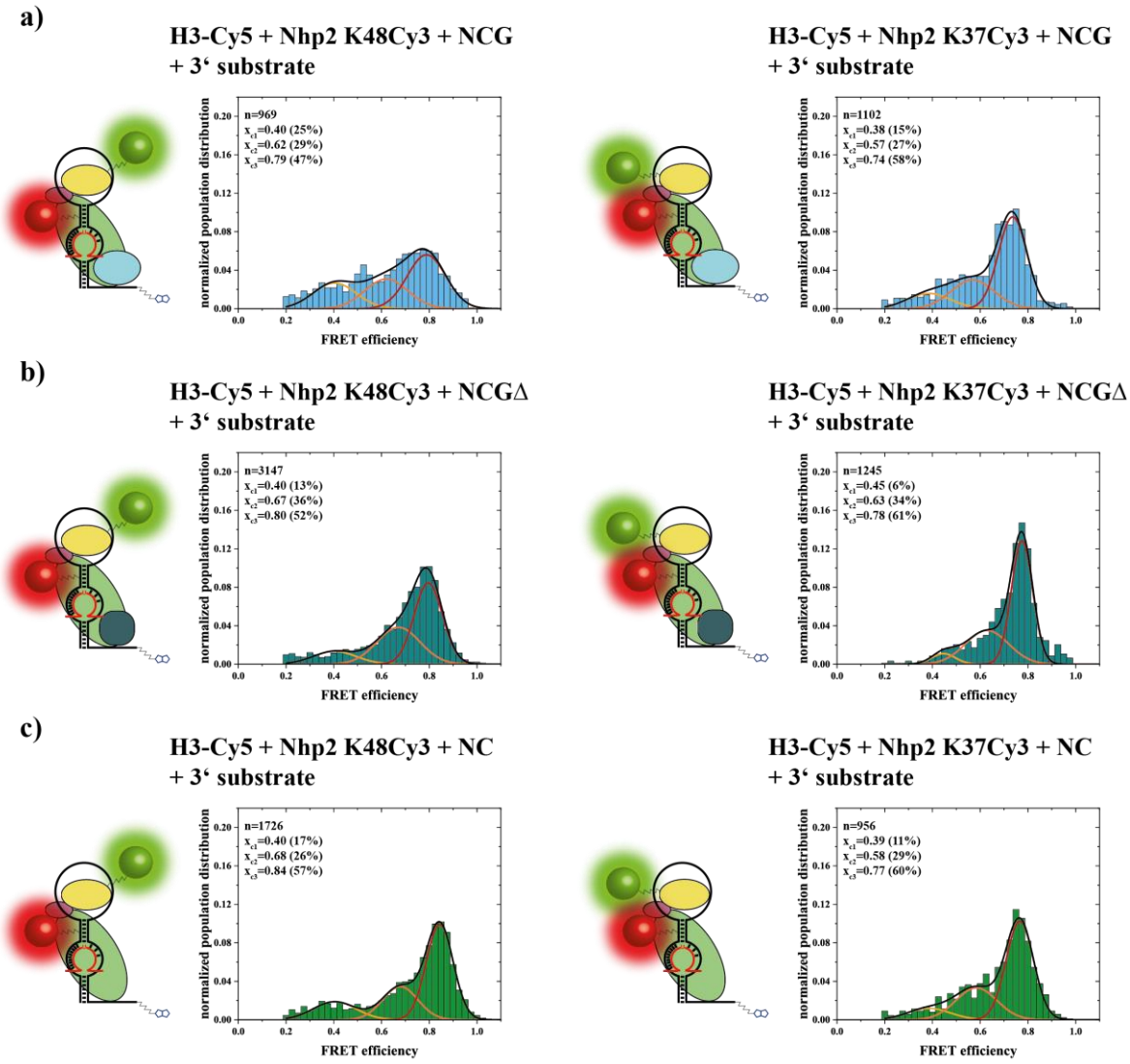


Figure 143: a) Histograms of fully reconstituted 3’ hairpin complexes H3-Cy5 + Nhp2 K48Cy3 + NCG (Ieft) and H3-Cy5 + Nhp2 K37Cy3 + NCG (right) in presence of 3’ substrate RNA. b) Histograms of GAR domain omission for the 3’ hairpin complexes H3-Cy5 + Nhp2 K48Cy3 + NCGΔ (Ieft) and H3-Cy5 + Nhp2 K37Cy3 + NCGΔ (right) in presence of 3’ substrate RNA. c) Histograms of Gar1 omission for the 3’ hairpin complexes H3-Cy5 + Nhp2 K48Cy3 + NC (Ieft) and H3-Cy5 + Nhp2 K37Cy3 + NC (right) in presence of 3’ substrate RNA. n represents the number of molecules observed for E-FRET ≥ 0.2. The three states x_{c1-3} including error were determined by analyzing 120 second movies (see Table 7) and Gaussian fitting within the error of x_c was used to determine and plot three Gaussian distributions in the histogram with a given peak width between 0.05 and 0.18 based on the single state constructs. Values in % are calculated by determining the area under the curve using the Gaussian distribution and represent the amount of molecules in one of the three states.

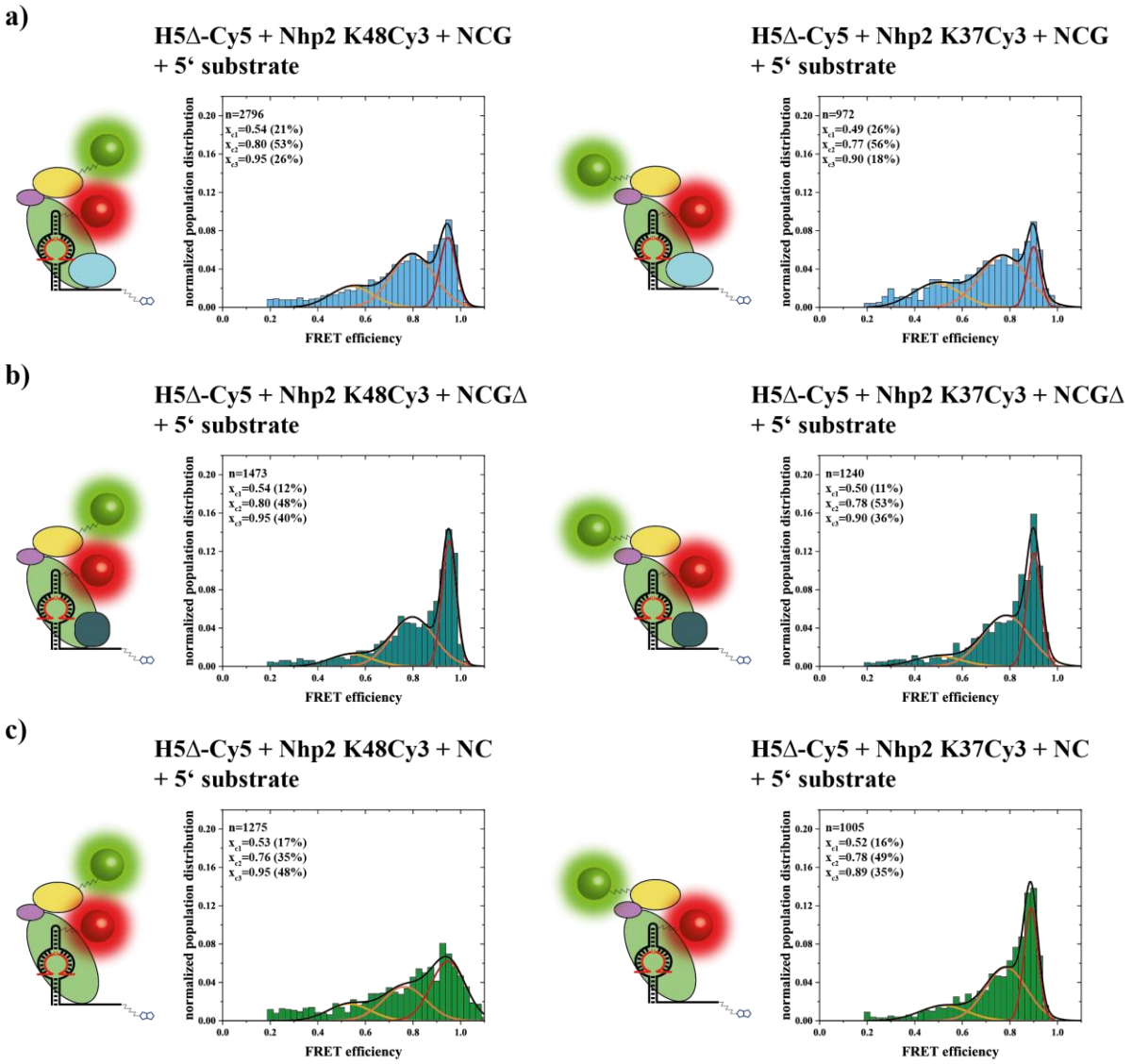


Figure 144: a) Histograms of fully reconstituted 5’ hairpin complexes H5Δ-Cy5 + Nhp2 K48Cy3 + NCG (Ieft) and H5Δ-Cy5 + Nhp2 K37Cy3 + NCG (right) in presence of 5’ substrate RNA. b) Histograms of GAR domain omission for the 3’ hairpin complexes H5Δ-Cy5 + Nhp2 K48Cy3 + NCGΔ (Ieft) and H5Δ-Cy5 + Nhp2 K37Cy3 + NCGΔ (right) in presence of 5’ substrate RNA. c) Histograms of Gar1 omission for the 3’ hairpin complexes H5Δ-Cy5 + Nhp2 K48Cy3 + NC (Ieft) and H5Δ-Cy5 + Nhp2 K37Cy3 + NC (right) in presence of 5’ substrate RNA. n represents the number of molecules observed for E-FRET ≥ 0.2. The three states x_{c1-3} including error were determined by analyzing 120 second movies (see Table 8) and Gaussian fitting within the error of x_c was used to determine and plot three Gaussian distributions in the histogram with a given peak width between 0.05 and 0.18 based on the single state constructs. Values in % are calculated by determining the area under the curve using the Gaussian distribution and represent the amount of molecules in one of the three states.

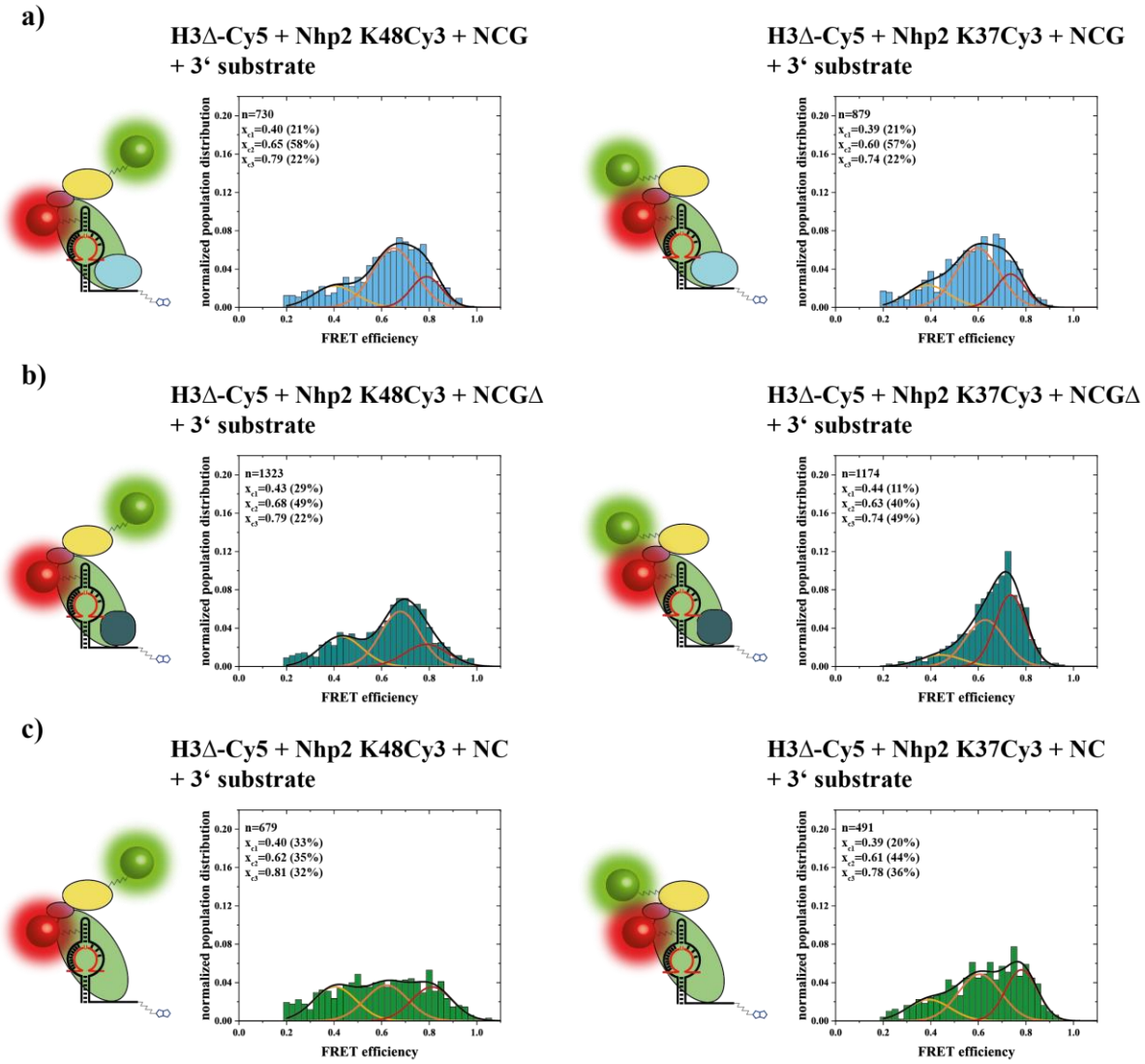


Figure 145: a) Histograms of fully reconstituted 3’ hairpin complexes H3Δ-Cy5 + Nhp2 K48Cy3 + NCG (Ieft) and H3Δ-Cy5 + Nhp2 K37Cy3 + NCG (right) in presence of 3’ substrate RNA. b) Histograms of GAR domain omission for the 3’ hairpin complexes H3Δ-Cy5 + Nhp2 K48Cy3 + NCGΔ (Ieft) and H3Δ-Cy5 + Nhp2 K37Cy3 + NCGΔ (right) in presence of 3’ substrate RNA. c) Histograms of Gar1 omission for the 3’ hairpin complexes H3Δ-Cy5 + Nhp2 K48Cy3 + NC (Ieft) and H3Δ-Cy5 + Nhp2 K37Cy3 + NC (right) in presence of 3’ substrate RNA. n represents the number of molecules observed for E-FRET ≥ 0.2. The three states x_{c1-3} including error were determined by analyzing 120 second movies (see Table 7) and Gaussian fitting within the error of x_c was used to determine and plot three Gaussian distributions in the histogram with a given peak width between 0.05 and 0.18 based on the single state constructs. Values in % are calculated by determining the area under the curve using the Gaussian distribution and represent the amount of molecules in one of the three states.

4.10.14 Activity tests for Cbf5Δ L156Cy3

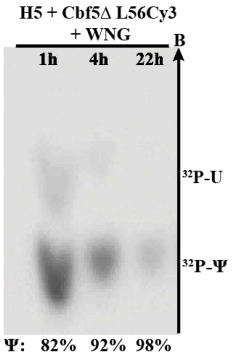


Figure 146: a) Thin layer chromatography of multiple turnover pseudouridylation activity tests performed with site specific ^{32}P-labeled 5' substrate RNA (standard multiple turnover conditions). Labeling of Cbf5Δ with Cy3 does not interfere with activity.

4.10.15 Preparation of Nhp2 K37NoK

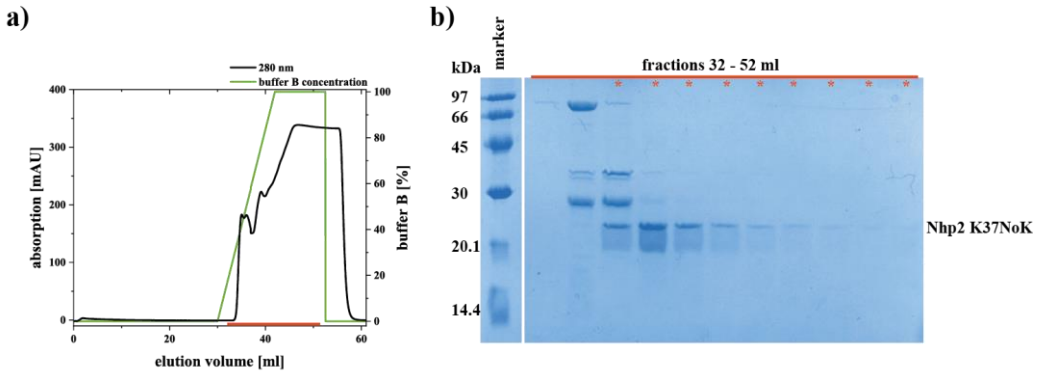


Figure 147: Affinity chromatography purification of Nhp2 K37NoK after RNase A treatment. a) Ni-NTA elution chromatogram using an imidazole gradient. Fractions used for analysis are marked with a red bar. b) SDS PAGE analysis of fractions from Ni-NTA. Fractions used for further purification are marked with a red asterisk. Marker: low molecular weight marker (GE Healthcare).

4.10.16 EPR measurements

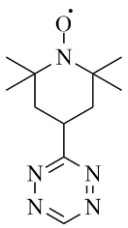


Figure 148: structural formula of tetrazine-TEMPO.

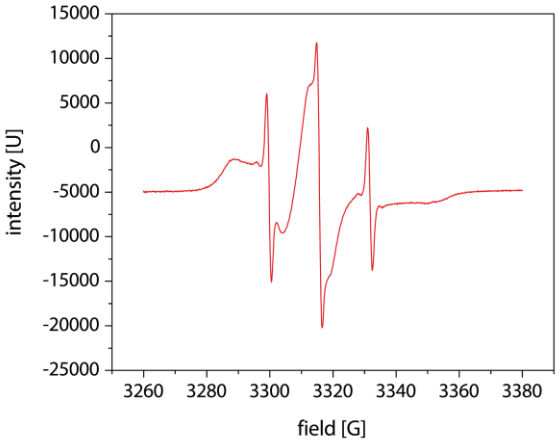


Figure 149: Spin counting of MTSSL labeled Nhp2 at native position C94.

4.10.17 Activity tests for MTSSL labeled constructs

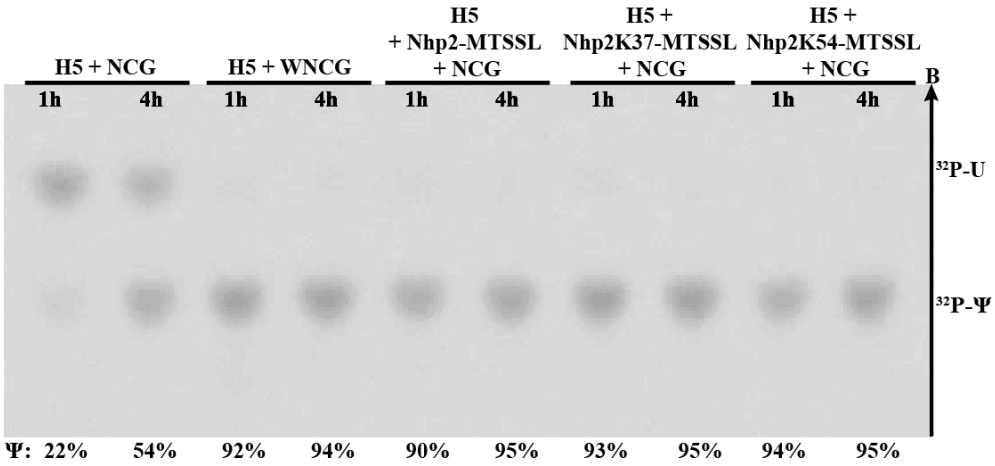


Figure 150: a) Thin layer chromatography of multiple turnover pseudouridylation activity tests performed with site specific ^{32}P-labeled 5' substrate RNA (standard multiple turnover conditions). Labeling of Nhp2 with MTSSL at different sites does not interfere with activity.

Chapter 5

Conclusion and Outlook

The main goal of this thesis was to gain new insights into eukaryotic H/ACA snoRNPs on a structural as well as dynamical level. For this purpose, the snR81 H/ACA snoRNP from *Saccharomyces cerevisiae* was established as a system of investigation. The snoRNA of this complex forms a bipartite double hairpin structure with the conserved H- and ACA-motifs and both hairpins contain a set of the four *S. cerevisiae* proteins Nhp2, Nop10, Cbf5 and Nop10.

Structurally, little is known about the bipartite complex architecture, and most structural knowledge stems from archaeal complexes, which mostly form a single hairpin structure[140]. Most recently, a cryo-EM structure of the H/ACA region of the human telomerase was presented, which however is also only part of a bigger structure[169]. Crystal structure data for eukaryotic complexes is only available for a Nop10-Cbf5Δ-Gar1Δ subcomplex[160], for Nhp2 only a standalone NMR structure exists[362]. Both of these structures suffer to some extent from the fact, that they were both modeled using the pyrococcal complex as a template.

In the first part of this thesis, the main goal was to prepare *in vitro* a functionally active snR81 snoRNP. The process of uridine to pseudoruridine formation should then be dissected with a set of catalysis assays, utilizing radioactively labeled substrate RNA to track the time resolved course of the reaction. In the second part of the thesis, different complex components should be labeled with fluorophore dyes to perform FRET spectroscopy on the complex on a single molecular level. Both methods should be used to gain new insights into the formation of an active complex, the requirements as well as the structural arrangements of different complex components and information about the dynamic of the complex during catalysis via dissection of its enzymatic kinetics in comparison to structural changes and omission of different complex parts.

In this chapter, the different accomplishments and findings of this thesis will be summarized and possible future aspects of the project will be outlined.

5.1 *In vitro* reconstitution of an active RNP and spectroscopic labeling

In this thesis, it was possible to establish a workflow for the efficient *in vitro* reconstitution of the eukaryotic snR81 H/ACA snoRNP from *Saccharomyces cerevisiae*. The different snoRNA construct could be prepared via transcription and subsequent purification and were obtained in high yield in a major conformation. For smFRET analysis, a part of the snoRNA was labeled with a fluorophore dye and the different snoRNAs were prepared via ligation[371]. Since in previous investigations of eukaryotic snoRNPs different behaviors of the two hairpins in terms of enzymatic catalysis were reported [159,160], standalone constructs of both hairpins were generated to individually dissect the different requirements of these hairpins during complex formation and catalysis.

The four proteins Nhp2, Nop10, Cbf5 and Gar1 could be recombinantly expressed from *E. coli* and purified in good quality and high yield. Endogenous *E. coli* RNA could be efficiently removed from the proteins prior to complex reconstitution with the snoRNA. While due to the

lack of different assembly factors a different reconstitution pathway was utilized than the complex normally undergoes *in vivo*, it was still possible to prepare a catalytically highly active H/ACA RNP, compared with other *in vivo*[96,356] and *in vitro* approaches[159,160]. The *in vitro* preparation was possible with the formation of a ternary Nop10-Cbf5-Gar1 complex, which then forms an active RNP with the snoRNA and Nhp2. The reconstituted complex could be used for catalytic assays and the information on complex preparation could be used for a fluorophore labeled complex for smFRET experiments.

Different non-canonical amino acids were chemically synthesized in good yield and amber codon suppression was utilized to readily attach reactive side chains at site-specific positions into the proteins. While the non-canonical amino acid incorporation was problematic for Gar1, the modified proteins Cbf5Δ and especially Nhp2 could be obtained in good yield and with different labeling positions. The modified proteins could be used to site-specifically attach fluorophore dyes as well as paramagnetic labels, and to reconstitute labeled RNP complexes applicable for spectroscopic measurements. The labeled complexes still showed enzymatic activity comparable to their wildtype counterparts. While for single molecule analysis only a small amount of labeled sample was needed, it was also possible to scale up the whole process to produce sufficient amounts of the complex to perform PELDOR measurements. The knowledge gained during the preparation and purification of the different complex components can be utilized for future complex or subcomplex analyses via NMR spectroscopy or cryo-EM. Especially the foundation for future cryo-EM experiments was laid with purification strategies of the H/ACA RNP after complex assembly, to separate unbound proteins from the reconstituted complex.

5.2 Different requirements and structural arrangements for 5’ and 3’ hairpins for an enzymatically active RNP

Eukaryotic H/ACA RNPs share the conserved feature of a bipartite complex structure, composed of a 5’ and 3’ hairpin, with the H- and ACA-motives in the hinge and tail regions. Both hairpins contain a pseudouridylation pocket that is responsible for the recruitment of a substrate RNA via Watson-Crick base-pairing, which results in the correct positioning of the target uridine for the enzymatic isomerization into pseudouridine. The main differences between the two hairpins of snR81 are: 1) The 5’ hairpins forms a long double stranded stem comprised of paired nucleotides and a small loop region, while the 3’ hairpin forms a very short stem and contains a large loop region with unpaired nucleotides. 2) The 5’ hairpin forms more Watson-Crick base-pairs with the substrate RNA than the 3’ hairpin, and in addition the base-pairing in the 5’ hairpin is more distributed between the two strands of the pseudouridylation pocket in comparison to the 3’ hairpin, where the binding is primarily contained to one strand.

In this thesis, both hairpins were analyzed both individually and in the full snR81 RNP complex for their catalytic activity. Interestingly, both hairpins showed different behavior in response to several “alterations” to the fully reconstituted RNP.

The 5’ hairpin in the full length snR81 construct (FL5) and the standalone hairpin (H5) showed similarly high enzymatic activity. This shows, that the 5’ hairpin is not reliant on the presence of the 3’ hairpin. The stable complex assembly of both constructs may be influenced by the relatively stable upper stem-loop structure, consisting of mainly double-paired nucleotides in a stem and a small unpaired loop region. Also, the removal of the upper stem-loop structure in the 5’ hairpin (H5Δ) did not have a big influence on complex activity.

The activity of the 3’ hairpin in the full length snR81 (FL3) was distinctively lower than the activity for the 5’ hairpin. Furthermore, the activity for the 3’ hairpin standalone construct (H3) got drastically reduced in absence of the 5’ hairpin. This indicates a dependency of the 3’ hairpin towards the 5’ hairpin. Possibly the 3’ hairpin requires the presence of the 5’ hairpin for correct folding or complex assembly. A major factor for this effect may be the unstructured loop region of the 3’ hairpin, that contains a large number of unpaired nucleotides and only a small double-stranded stem, resulting in a lower activity than the 5’ hairpin and a reliance on the 5’ hairpin for correct folding and stabilization. However, the big stem-loop structure of the 3’ hairpin is also strictly required for catalytic activity, as removal of the upper stem-loop structure of the 3’ hairpin (H3Δ) resulted in complete activity loss. This points to the conclusion that while the stem-loop structure of the 3’ hairpin slows down the adaption of the “correct” folding and RNP formation and thereby the activity, at the same time it appears to be mandatory for the correct folding and assembly with the H/ACA core proteins.

5.2.1 Influence of Gar1 and its eukaryotic specific GAR domains

The removal of the eukaryotic specific “GAR” domains of Gar1 had an activity reducing effect on both FL5 and H5. These glycine-arginine rich domains, that also exist in other proteins, are known to have RNA binding promoting capabilities[358,363–368]. In EMSA experiments, the influence of the GAR domains in promoting Gar1-snR81 interactions could be observed and it is possible, that in the full RNP the GAR domains can actually interact with the snoRNA (or the substrate RNA). While in archaeal H/ACA RNPs, Gar1 (which does not contain glycine-arginine rich domains) only interacts with Cbf5[140], in eukaryotic H/ACA RNPs eukaryotic specific domains of Gar1 may have an influence on the snoRNA and contribute to the correct folding or positioning of the snoRNA or substrate RNA, and may also have an effect on complex stability.

Interestingly, GAR domain omission actually resulted in a higher activity for the 3’ hairpin under multiple turnover conditions, while it decreased activity under single turnover conditions. This shows, that the GAR domains actually exhibits some kind of effect, which slows down substrate turnover on the 3’ hairpin. While the effect that the GAR domains have on the assembly, stability or substrate positioning is beneficial for multiple substrate turnover at the 5’ hairpin, it seems to have an inhibiting effect on the 3’ hairpin – possibly by interfering with the correct folding of the RNA and thereby the complex assembly or the substrate recruitment by unspecifically binding to the unstructured loop region of the 3’ hairpin. While overall, the conserved GAR domains of Gar1 may have an activity increasing

effect on the catalytic reaction, in some cases (like for the 3' hairpin of snR81) the GAR domains may have a negative effect on multiple substrate turnover that overshadows the positive effect.

Naf1, that is present in the place of Gar1 during the *in vivo* complex assembly route and gets replaced by Gar1 at the end, does not contain any glycine-arginine rich domains. This may point to the hypothesis, that the GAR domains are directly involved during the catalytic process, e.g. for either correct snoRNA or substrate RNA placement for the pseudouridine formation rather than playing a role in complex assembly.

5.2.2 Influence of Nhp2

Upon Nhp2 omission, the 5' hairpin showed reduced activity, but still pseudouridine formation was relatively high – even higher than for the 3' hairpin in the presence of Nhp2. Nhp2 seems to have some accelerating effect on the pseudouridylation, either by inducing a more stable complex or by actually supporting the correct substrate placement (like L7Ae[157]). However, it is not as strictly required as L7Ae and the complex can also function in multiple turnover reactions in its absence.

With fluorophore labeling and smFRET spectroscopy, a single binding state for Nhp2 could be identified for the 5' hairpin, showing a tight interaction of Nhp2 with its complex binding site – Nop10 and the upper-stem loop of the snoRNA. No such population could be observed in the absence of Nop10. For the shortened construct H5Δ, three different binding modes could be observed, showing that upon the disruption of its snoRNA binding site but in the presence of Nop10, Nhp2 is still able to bind but no longer specific for one state. Without the full binding site provided by both Nop10 and the snoRNA, Nhp2 seems to be able to bind to Nop10 in three different conformational arrangements. Since no transitions were observed between the different states, Nhp2 either gets locked in one of these three binding modes after complex assembly, or the transition between the states is extremely slow. The observation of three binding modes with no noticeable loss in activity points to the conclusion, that Nhp2s function is to stabilize the upper stem loop structure – a function that is no longer required after the removal of this structure. Since the omission of Nhp2 was not as drastic for H5Δ than it was for H5 and FL5, this actually supports this hypothesis. However, since the activity was still reduced for H5Δ without Nhp2, this protein might actually play additional roles in the complex: stabilization of the hairpin structure and a second function, which may be correct substrate placement, like it is performed by L7Ae.

The 3' hairpin on the other hand nearly lost its complete enzymatic activity under multiple turnover conditions in the absence of Nhp2. Only after 22 hours, a small amount of pseudouridine could be detected, and only for the full snR81 snoRNA, not for the standalone hairpin. This shows that Nhp2 seems to have either a different or an additional effect on the 3' hairpin than it has on the 5' hairpin, or the effect that it has is negligible for the 5' hairpin, but strictly required for the 3' hairpin. In that matter, the 3' hairpin shows a similar requirement for Nhp2 than archaeal H/ACA RNPs show for L7Ae. In several other eukaryotic RNPs

(snR5 and snR34), different hairpins also showed different requirements for Nhp2, however in no complex an effect as drastic as for the 3' hairpin of snR81 was observed[159,160]. If Nhp2 has a stabilizing effect on the hairpin structure, the different requirement may also have to do with the relatively large upper stem-loop structure – a structure element that is not present in snR5 and snR34.

Fluorophore labeled constructs and smFRET analysis showed that Nhp2 adopted three different binding modes for the full construct H3 as well as for shortened construct H3Δ. Since no transitions could be observed, similar to H5Δ it is not clear whether Nhp2 gets locked in one conformation or if very slow transitions between the states are possible. Since Nhp2 adopts three binding modes even for the full construct H3, Nhp2 seems to be impaired in its ability to interact with a stable binding site (in contrast to the single binding mode for the 5' hairpin). This may be heavily influenced by the previously described large loop region of unpaired nucleotides, that may be slow in adapting a conformation that provides the correct binding site and is unstructured, which also explains the generally lower activity for the 3' hairpin in comparison to the 5' hairpin. The strict requirement for Nhp2 at the 3' hairpin seems to show, that without Nhp2, the upper stem loop is not able to adopt the correct conformation inside of the fully reconstituted RNP. Upon binding of Nhp2 at the "correct" binding site, it may stabilize the upper stem-loop structure and thereby stabilize the whole complex, which is required for enzymatic activity. If the three different states represent different binding modes of Nhp2 to Nop10, then one of these states represents a binding to Nhp2 and additionally to the snoRNA.

Observations made for the GAR domain omission from Gar1 showed, that in the absence of the GAR domains, the high FRET state increased in population and at the same time, complex activity was increased. This supports the hypothesis that the low and intermediate FRET states represent Nhp2 that binds to Nop10 in a way that it cannot interact with the snoRNA, and the high FRET state represents Nhp2 that binds to Nop10 in way that it can bind towards and stabilize the snoRNA, forming a functional complex. Upon removal of the upper stem-loop structure, all three states were more evenly populated, showing that without the canonical snoRNA binding surface, none of the three states is majorly preferred. However, other than the 5' hairpin, where the removal of the upper stem-loop structure actually "relieved" Nhp2 from its function to stabilize the hairpin and hairpin removal did not lower the activity, for the 3' hairpin the removal of the upper-stem loop structure leads to a complete activity loss, assumingly due to misfolding of the complex.

Since no significant structural changes were observed in absence or presence of substrate RNA (which was added into the imaging buffer), the Nhp2 binding state populations seem independent from substrate RNA binding (note: with the described experiments, the actual binding of substrate RNA to the complex cannot be ascertained, however similar smFRET experiments with fluorophore labeled substrate RNA showed binding events of the substrate RNA to the RNP[371]) . If Nhp2 does play a role in substrate RNA placement, it seems to not have a strong effect on the conformation in which Nhp2 is bound to the complex. This could also show that Nhp2 does not have to undergo structural changes and may only play a

promoting or supporting role in correct substrate placement, while its major role lies in the stabilization of the complex discussed above.

5.3 Limitations of the single molecule approach on dynamical observations

While it was possible to observe different complex states during smFRET analysis, no transitions between different states could be observed in real time. It was possible to observe binding events of fluorophore labeled substrate RNA to the complex in a different work[371], but it was not possible to directly observe direct dynamic movements of neither Nhp2 nor the thumb loop domain of Cbf5. This can both be explained with two different limitations of the experimental approach of smFRET analysis. The complexes that are observed during the measurements are immobilized onto the glass surface via biotin-streptavidin interactions. This immobilization may simply prevent the complex from catalytic activity. Also, while the activity tests were performed for a complex concentration of 100 nM, the complex concentration for smFRET measurements was 100 - 1000-fold lower (100 pm - 1 nM). Maybe the low complex concentration in combination with the immobilization of the complexes only allows the observation of reconstituted, but inactive complexes. The second explanation lies in the time resolution of smFRET spectroscopy. A time resolution of 100 ms was applied and collected movies lasted from two seconds to two minutes. While Nhp2 may be locked in one conformation after complex reconstitution and the transition between Nhp2 states may happen on a much larger time-scale (if at all), the dynamic of the Cbf5 thumb loop domain on the other hand may happen on a sub-millisecond timescale and may simply be too fast for real time observation via smFRET spectroscopy.

5.4 Insights into the bipartite complex structure

Since neither a FRET signal between the two hairpins could be detected, nor a distance measurement by PELDOR could be performed between them, the complex binding sites of Nhp2 seem to be further apart than the amenable distances of both methods allow. While pointing towards some “angled” structure between both hairpins, exact conclusion will certainly be possible in the future. By adapting the workflow of preparation of complex components and performance of RNP assembly and purification, the work on dissecting the bipartite complex structure via cryo-EM has already begun.

Furthermore, one important conclusion can be drawn from the activity data of the full length snR81 in comparison to the two standalone hairpins. The presence/absence of either Nhp2 as well as the GAR domains of Gar1 showed the same influence on the hairpin under observation, regardless of the presence of the other hairpin. One can assume, that the observed inter-hairpin interactions on enzymatic activity likely are neither based on Nhp2/Nhp2 and Gar1/Gar1 interactions, nor on the interactions of the RNA of one hairpin with Nhp2 or Gar1 of the other hairpin.

For the future, it would also be interesting to perform smFRET measurements on a fully reconstituted snR81 H/ACA RNP – if not between the two hairpins, then at least on one

hairpin while the other hairpin is present and the RNP is fully assembled. With measurements similar to those performed in this thesis, it would be interesting to see the effects on Nhp2 binding and GAR domain influence in the presence of both hairpins. Since the 3' hairpin showed majorly decreased activity in absence of the 5' hairpin, and also the activity of the 5' hairpin was lower in absence of the 3' hairpin (however only by a smaller amount), there seems to be some interdependence of both hairpins, which could influence the complex assembly, the correct binding of proteins or the catalytic reaction itself.

In conclusion, during this thesis, it was possible to prepare the functionally active snR81 H/ACA snoRNP from *Saccharomyces cerevisiae* as well as substructures of this complex *in vitro*. With radioactively labeled substrate RNA it was possible to dissect the complex for its activity during uridine to pseudouridine formation and highlight the importance of different complex components during this enzymatic reaction. Furthermore, it was possible to label different complex parts with spectroscopic probes via the utilization of genetic code expansion and click chemistry. This allowed insights into the complex assembly on a single molecule level and allowed the study of structural and dynamical aspects of the complex and also inquire on the role of several eukaryotic specific complex components.

Summarizing, it was possible to gain new insights into the complex structure and the dynamical behavior of the still sparsely characterized eukaryotic H/ACA RNPs. Especially new insights into the rather ambiguous role that Nhp2 plays during complex assembly and catalysis as well as the importance of eukaryotic specific Gar1 domains were gained, including hairpin specific behavior on the bipartite complex structure.

Figure 151 shows a summary of structural as well as kinetic effects, that could be dissected from the two snR81 hairpins as well as the effects that different complex components seem to play during the complex assembly and the enzymatic reaction of the pseudouridine formation in the snR81 H/ACA snoRNP.

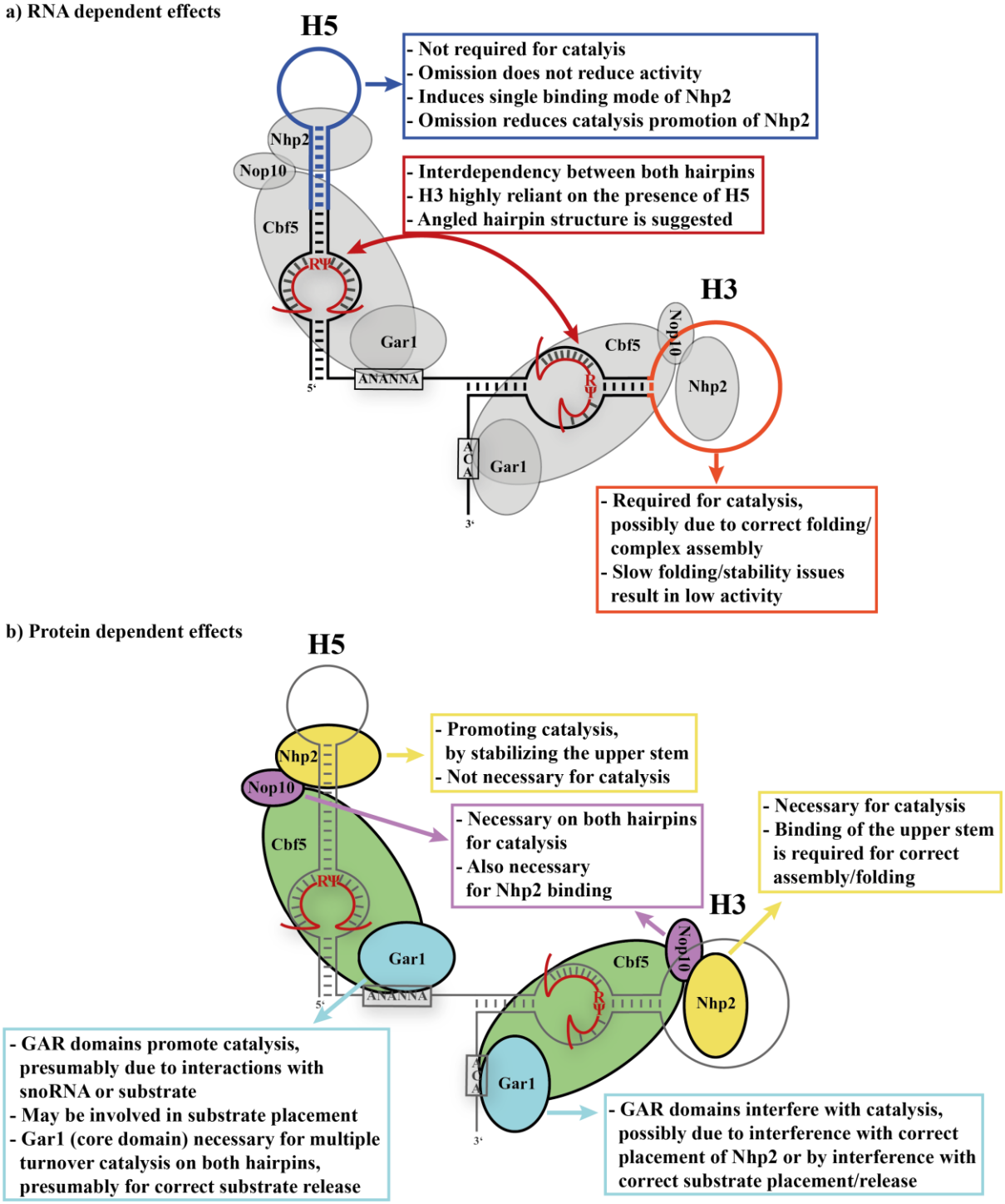


Figure 151: Summary of observed influences of different snR81 H/ACA snoRNP complex components on complex formation and catalytic activity. a) Observed snoRNA effects, including the role of the upper stem-loop structure. b) Observed protein effects of Nop10, Nhp2 and GAR domains of Gar1.

Chapter 6

Material and methods

6.1 Media and buffers

Chemicals used for media and buffers were ordered from Sigma Aldrich and Carl Roth.

6.1.1 Media for cell cultures

LB medium
Yeast extract (5 g/l)
Tryptone (10 g/l)
NaCl (5 g/l)

TB medium
Tryptone/Peptone (12 g/l)
Yeast extract (24 g/l)
Glycerol (4 ml/l)
K_2HPO_4 (72 mM)
KH_2PO_4 (72 mM)

LB-Agar
Yeast extract (5 g/l)
Tryptone (10 g/l)
NaCl (5 g/l)
Agar-Agar (15 g/l)

SOC outgrow medium
Peptone (2%)
Yeast extract (0.5%)
NaCl (10 mM)
KCl (2.5 mM)
$MgCl_2$ (10 mM)
$MgSO_4$ (10 mM)
Glucose (20 mM)

All media were autoclaved directly after preparation (121 °C, 15 min, 2 bar). Potassium salts were added after autoclaving, directly before the usage of the media. Agar plates were prepared by adding antibiotics to the LB-Agar medium at ~55 °C, adding the medium into petri dishes (~20 ml/dish) and letting it cool down. LB medium was purchased from Carl Roth. SOC medium was purchased from NEB.

6.1.2 Antibiotics used in media

Working concentration for antibiotics were 100 µg/ml for Ampicillin (Amp), 30 µg/ml for Kanamycin (Kan) and 34 µg/ml for Chloramphenicol (Cmp). Antibiotics were prepared as a 1000x stock solution in water (Amp and Kan) or ethanol (Cmp). Aliquots of antibiotic stocks were stored at -20 °C.

6.1.3 General buffers for gel electrophoresis

4 x Stacking gel buffer
Tris/HCl (1 M, pH 6.8)
SDS (0.4%)

4 x Resolving Gel buffer
Tris/HCl (1.5 M, pH 8.8)
SDS (0.4%)

SDS-PAGE buffer
Tris (25 mM, pH 8.3)
Glycine (250 mM)
SDS (0.1% (w/v)

TBE buffer	TAE buffer	TA buffer
Tris (90 mM, pH 8.3)	Tris (40 mM, pH 8.3)	Tris (40 mM, pH 8.3)
Boric acid (90 mM)	Acetic acid (20 mM)	Acetic acid (20 mM)
EDTA (2 mM)	EDTA (1 mM)	

2 x SDS loading dye
Stacking gel buffer (4x)
Glyerol (20% (v/v))
SDS (4% (w/v))
β-ME (1.4%)
bromophenol blue (0.05% (w/v))

Coomassie staining solution
Ethanol (10% (v/v))
AcOH (5% (v/v))
Coomassie Brillant Blue R250 (2.5‰ (w/v))
Coomassie Brillant Blue G250 (2.5‰ (w/v))

2 x FA buffer
TBE buffer (1x)
Formamide (90% (v/v))
Bromophenol blue (0.05% (w/v))
Xylene cyanol (0.05% (w/v))

PBS buffer
NaCl (137 mM)
KCl (2.7 mM)
Na_2HPO4 (100 mM)
KH_2PO4 (20 mM, pH 7.4)

Anode buffer
Tris/HCl (75 mM, pH 7.4)
Methanol (20% (v/v))

Kathode buffer
Tris/HCl (25 mM, pH 9.0)
ε-Aminocaproic acid (40 mM)
Methanol (20% (v/v))

6.1.4 Buffers for protein purification

IMAC A	IMAC B	SEC
NaCl (500 mM)	NaCl (500 mM)	NaCl (500 mM)
Tris/HCl (50 mM, pH 8.0)	Tris/HCl (50 mM, pH 8.0)	Tris/HCl (50 mM, pH 8.0)
Glycerol (10% (v/v))	Glycerol (10%(v/v))	Glycerol (10%(v/v))
β-ME (7 mM)	β-ME (7 mM)	
Imidazole (20mM)	Imidazole (500 mM)	

All buffers for protein purification were sterile filtered (0.2 µm).

6.1.5 Buffers for pseudouridylation activity assays

Ψ-buffer
NaCl (250 mM)
Tris/HCl (125 mM, pH 8.0)

P1 storage buffer	**P1 buffer for digestion**
NaCl (27.5 mM)	NH4OAc/AcOH (20 mM, pH 5.3)
NaOAc/AcOH (11 mM, pH 5.4)	ZnCl2 (500 mM)
ZnCl2 (2.75 mM)	
Glycerol (45%(v/v))	

TLC buffer A	**TLC buffer B**
2-propanol (70 % (v/v))	2-Methylpropanoic acid (70 % (v/v))
conc. HCl (15 % (v/v))	25 % ammonia solution (1.1 % (v/v))
H_2O (15 % (v/v))	H_2O (28.9 % (v/v))

All buffers were sterile filtered (0.2 µm). P1-endonuclease (lyophilized powder, ≥250 units, 1 mg, Sigma Aldrich) was added to the storage buffer (0.25 U/µl).

6.1.6 Buffers for smFRET spectoscropy

T50-buffer (50x)	**Gloxy (50x)**	**Imaging buffer**
NaCl (50 mM)	T50 (1x)	Gloxy (1x)
Tris/HCl (10 mM, pH 8.0)	Glucose oxidase (4 mg/ml)	D-Glucose (10% (w/v))
	Catalase (1 mg/ml)	Ψ-buffer (1x)
		Trolox (until saturation)

All buffers were sterile filtered (0.2 µm). Gloxy (50x) and Imaging buffer were prepared freshly before usage.

6.2 General biochemical methods

6.2.1 Gel electrophoresis methods

6.2.1.1 SDS PAGE for proteins

For SDS gels, a stacking gel solution (1x stacking gel buffer, 5% acrylamide/bis-acrylamide (29:1)) and a resolving gel solution (1x running gel buffer, 15% acrylamide/bis-acrylamide (29:1)) were prepared, polymerization was induced with TEMED (0.1% (v/v)) and APS (0.1% (w/v)) and gel solutions were cast between two glass plates to form a two layered gel (⅕ stacking gel, ⅘ resolving gel). Protein samples were mixed with 2x SDS loading dye (1:1) and heated at 95 °C for 10 min. PAGE was done with 1x SDS-PAGE running buffer at 80 V for 20 min and 200 V for 40 min in a Mini-PROTEAN Tetra Cell chamber (BioRad).

6.2.1.2 Denaturing TBE-PAGE for RNA

For TBE gels, a gel solution (1x TBE buffer, 12-20% acrylamide/bis-acrylamide (29:1), 7 M Urea) was prepared, polymerization was induced with TEMED (0.1% (v/v)) and APS (0.1% (w/v)) and the gel solution was cast between two glass plates. RNA samples were mixed with FA-buffer (1:1) and heated at 95 °C for 3 min. PAGE was done with 1x TBE-buffer at 220 V in a Multigel sytsem (Biometra) and stopped according to dye migration.

6.2.1.3 Native TBE-PAGE for RNA

For TBE gels, a gel solution (1x TBE buffer, 12-20% acrylamide/bis-acrylamide (29:1)) was prepared, polymerization was induced with TEMED (0.1% (v/v)) and APS (0.1% (w/v)) and gel solution was cast between two glass plates. RNA samples were mixed with glycerol (10% (v/v)) (1:1). PAGE was done with 1x TBE-buffer at 45 V and 4 °C in a Multigel system (Biometra) and stopped accordingly to dye migration.

6.2.1.4 EMSA

For EMSAs, a gel solution (1x TA buffer, 12-20% acrylamide/bis-acrylamide (29:1)) was prepared, polymerization was induced with TEMED (0.1% (w/v)) and APS (0.1% (v/v)) and gel solution was cast between two glass plates. RNP samples were prepared by mixing protein and RNA sampels in SEC buffer (1x) and incubated at 30 °C for 30 min. Afterwards, glycerol was added (17% (v/v) and PAGE was performed with 1x TA-buffer at 45 V and 4°C in a Multigel system (Biometra) and stopped according to dye migration.

6.2.1.5 TAE-agarose gel electrophoresis for DNA

For TAE gels, a gel solution (1x TAE buffer, 0.5-2% agarose) was prepared, heated up and cast onto an appropriate glass plate. RNA samples were mixed with 6x gel loading dye (1x, NEB). Agarose gel electrophoresis was done with 1x TAE-buffer at 120 V in a Mini-

Horizontal Unit Electrophoresis System (FisherBiotech) and stopped according to dye migration.

6.2.2 Visualization of biomolecules

6.2.2.1 Protein staining

SDS-PAGE gels were covered with Coomassie staining solution, microwave heated for 1 min and afterwards put on a shaking plate for 20 min. Afterwards, gels were covered with water and cellulose, microwave heated for 2 min and put on a shaking plate overnight for destaining of background. Gels were scanned with a Molecular Imager Gel Doc XR+ (BioRad).

6.2.2.2 DNA/RNA staining

RNA PA gels and DNA agarose gels were covered with GelRed solution (Biotium) for 15 min. GelRed fluorescence was detected via UV illumination with a Molecular Imager Gel Doc XR+ (BioRad).

6.2.2.3 Fluorescence labeled samples

Gels with fluorescence labeled proteins and RNA were scanned using a Typhoon 9400 (GE Healthcare). For visualization of Cy3 labeled samples a laser excitation of λ_{ex} = 532 nm with an emission filter of 580 nm was used. For visualization of Cy5 labeled samples a laser excitation of λ_{ex} = 633 nm with an emission filter of 670 nm was used.

6.2.2.4 Radioactive labeled samples

Gels and TLC plates with radioactive labeled RNA were visualized via autoradiography by storing the gel or the TLC plate in a storage phosphor screen (GE Healthcare) for several minutes up to several days, depending on the amount of radioactive sample used. The screen was scanned via autoradiography using a Typhoon 9400 (GE Healthcare) with a laser excitation of λ_{ex} = 633 nm and an emission filter of 390 nm.

6.2.2.5 Western Blot

For Western blotting of proteins, first SDS PAGE was performed (chapter 6.2.1.1). Afterwards, proteins were blotted onto an Immobilon-P PVDF Transfer Membrane (Merck Millipore) with a Fastblot B44 (Biometra) and Extra Thick Western Blotting Filter Paper (Thermo Scientific) with Anode/Kathode puffer. The membrane was put into blocking solution (5% powdered milk, 1x PBS buffer). Afterwards, the membrane was put into antibody 1 solution (6x-His Tag Monoclonal Antibody (HIS.H8) (c_{stock}=1 mg/ml, 1:3000 dilution, ThermoFisher Scientific), 5% powdered milk, 1x PBS buffer) and incubated overnight at 4 °C. Afterwards, the membrane was washed with washing solution (0.05% Tween, 1x PBS buffer) and put into antibody 2 solution (Peroxidase-conjugated AffiniPure

Goat Anti-Mouse IgG (H+L) (c_{stock}=0.8 mg/ml, 1:5000 dilution, dianova), 5% powdered milk, 1x PBS buffer) and incubated for 2 h at 4 °C. Membrane was washed again with washing solution and peroxide solution (1 ml) and luminol solution (1 ml) (SuperSignal West Pico PLUS Chemiluminescent Substrate, Thermo Scientific) were added. Oxidation induced chemiluminescence of luminol was detected with an Advanced Fluorescence Imager (Intas).

6.2.3 Quantification of biomolecules

6.2.3.1 Concentration determination via UV-vis spectroscopy

Protein concentration was determined with a Nanodrop one UV-vis spectrophotometer (Thermo Fisher Scientific) using the Lambert-Beer equation $A = \varepsilon \cdot l \cdot c$[370], with optical path length l (1 cm), absorption A at 280 nm (proteins) or 260 nm (DNA, RNA), molar attenuation coefficient ε according to Table 10 and concentration of the biomolecule c. Molar attenuation coefficient ε at 280 nm for proteins was calculated using ExPASy ProtParam tool[393]. For DNA and RNA, a standard coefficient was used (50 ng/µl · A for DNA, 40 ng/µl · A for RNA). For Gar1-constructs, a Bradford-Assay with ROTI Nanoquant (Carl Roth) was performed according to the manufacturer's protocol.

Table 10: molar extinction coefficient, molecular weight and number of amino acids for used protein constructs

protein	ε (M^{-1} cm^{-1})	Molecular weight (kDa)	Number of amino acids
Nhp2	4470	18.2	165
Nhp2W	9970	18.3	165
Nop10	4470	6.6	58
Cbf5	53860	55.5	489
Cbf5Δ	53860	44.9	400
Gar1(Ntag)	4470	23.6	225
Gar1Δ (Ntag)	4470	12.6	114
Gar1(Ctag)	4470	22.3	211
Gar1Δ (Ctag)	4470	11.2	100

6.2.3.2 Densitometric analysis

For densitometric analysis of RNA, DNA or protein bands on a gel, the ImageLab software (version 6.0, Bio-Rad) as well as ImageJ (version 1.50i) was utilized.

6.2.3.3 MALDI-TOF

MALDI-TOF mass spectrometry was performed by the mass spectrometry service unit of the Goethe university. Proteins were desalted prior to the measurement with C18 ZipTip pipette resins (Millipore) according to the manufacturer's protocol.

6.3 Protein based methods

6.3.1 Protein constructs

Table 11: Codon optimized amino acid sequences of used *S. cerevisiae* proteins.

Nhp2 (1-149)	MGKDNKEHKESKESKTVDNYEARMPAVLPFAKPLASKKLNKKVLKTVKK ASKAKNVKRGVKEVVKALRKGEKGLVVIAGDISPADVISHIPVLCEDHSVP YIFIPSKQDLGAAGATKRPTSVVFIVPGSNKKKDGKNKEEEYKESFNEVVKE VQAL
Nhp2W (1-149)	MGKDNKEHKESKESKTVDNYEARMPAVLPFAKPLASKKLNKKVLKTVKK ASKAKNVKRGVKEVVKALRKGEKGLVVIAGDIWPADVISHIPVLCEDHSV PYIFIPSKQDLGAAGATKRPTSVVFIVPGSNKKKDGKNKEEEYKESFNEVVK EVQAL
Nop10 (1-58)	MHLMYTLGPDGKRIYTLKKVTESGEITKSAHPARFSPDDKYSRQRVTLKK RFGLVPGQ
Cbf5 (1-483)	MAKEDFVIKPEAAGASTDTSEWPLLLKNFDKLLVRSGHYTPIPAGSSPLKR DLKSYISSGVINLDKPSNPSSHEVVAWIKRILRCEKTGHSGTLDPKVTGCLI VCIDRATRLVKSQQGAGKEYVCIVRLHDALKDEKDLGRSLENLTGALFQR PPLISAVKRQLRVRTIYESNLIEFDNKRNLGVFWASCEAGTYMRTLCVHLG MLLGVGGHMQELRRVRSGALSENDNMVTLHDVMDAQWVYDNTRDESY LRSIIQPLETLLVGYKRIVVKDSAVNAVCYGAKLMIPGLLRYEEGIELYDEI VLITTKGEAIAVAIAQMSTVDLASCDHGVVASVKRCIMERDLYPRRWGLG PVAQKKKQMKADGKLDKYGRVNENTPEQWKKEYVPLDNAEQSTSSSQE TKETEEEPKKAKEDSLIKEVETEKEEVKEDDSKKEKKEKKDKKEKKEKKE KKDKKEKKEKKEKKRKSEDGDSEEKKSKKSKK
Cbf5c (1-394)	MAKEDFVIKPEAAGASTDTSEWPLLLKNFDKLLVRSGHYTPIPAGSSPLKR DLKSYISSGVINLDKPSNPSSHEVVAWIKRILRCEKTGHSGTLDPKVTGCLI VCIDRATRLVKSQQGAGKEYVCIVRLHDALKDEKDLGRSLENLTGALFQR PPLISAVKRQLRVRTIYESNLIEFDNKRNLGVFWASCEAGTYMRTLCVHLG MLLGVGGHMQELRRVRSGALSENDNMVTLHDVMDAQWVYDNTRDESY LRSIIQPLETLLVGYKRIVVKDSAVNAVCYGAKLMIPGLLRYEEGIELYDEI VLITTKGEAIAVAIAQMSTVDLASCDHGVVASVKRCIMERDLYPRRWGLG PVAQKKKQMKADGKLDKYGRVNENTPEQWKKEYVPLDNAEQ
Gar1 (1-205)	MSFRGGNRGGRGGFRGGFRGGRTGSARSFQQGPPDTVLEMGAFLHPCEGD IVCRSINTKIPYFNAPIYLENKTQVGKVDEILGPLNEVFFTIKCGDGVQATSF KEGDKFYIAADKLLPIERFLPKPKVVGPPKPKNKKKRSGAPGGRGGASMG RGGSRGGFRGGRGGSSFRGGRGGSSFRGGSRGGSFRGGSRGGSRGGFRGG RR
Gar1c (32-124)	MGPPDTVLEMGAFLHPCEGDIVCRSINTKIPYFNAPIYLENKTQVGKVDEIL GPLNEVFFTIKCGDGVQATSFKEGDKFYIAADKLLPIERFLP

Nhp2 and Nhp2W were expressed from a pET-24b plasmid, containing a His_6-tag at the C-terminus. Cbf5 together with Nop10 and Cbf5Δ together with Nop10 were co-expressed from a pET-Duet plasmid, with Cbf5/Cbf5Δ containing a His_6-tag at the C-terminus. Gar1 and Gar1Δ were expressed from a pET-28b plasmid, containing a His_6-tag at the N-terminus. For generation of amber suppression mutants, Gar1 and Gar1Δ were expressed from a pET-24b plasmid, containing a His_6-tag at the C-terminus. For ternary Nop10-Cbf5-Gar1, constructs were co-expressed and co-purified.

6.3.2 Cryo stocks

For the generation of cryo stocks, 500 µl of an overnight culture were mixed with 50% glycerol (500 µl, 25% final conc.) and shock frozen in liquid nitrogen. Cryo stocks were stored at -80 °C.

6.3.3 Protein expression

6.3.3.1 Small scale expression tests

For validation of constructs, several test expressions were performed simultaneously. Test expressions were performed in a Greiner tube with an expression volume of 3 ml. Everything else was done accordingly to chapter 6.3.3.2 or 6.3.3.3. Samples of 100 µl were directly taken from the test expression cultures, cells harvested (17000 x g, 5 min, 4 °C) and mixed with 2x SDS loading dye (40 µl per OD_{600}). Test expressions were analyzed via SDS-PAGE and Western Blotting (chapter 6.2.2.5).

6.3.3.2 Expression of WT proteins

As overnight culture, 30 ml of LB medium in a 100 ml baffled Erlenmeyer flask with appropriate antibiotics (chapter 6.1.2) was inoculated with a single colony of *E. coli* BL21 (DE3) cells from an LB-agar plate or with a small amount of cells from a cryo stock. In a 5 l baffled Erlenmeyer flask, 1 l of TB medium was inoculated to an OD_{600} ~ 0.1 with overnight culture. Cells were incubated (130 rpm, 37 °C) to an OD_{600} ~ 0.6-0.8 and cooled on ice for ~ 15 min afterwards. Protein expression was induced by addition of 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and cells were incubated for ~ 20 h (130 rpm, 20 °C) and afterwards harvested (8000 x g, 30 min, 4 °C). If not purified immediately, pelleted cells were frozen in liquid nitrogen and stored at -80 °C.

6.3.3.3 Expression of proteins containing a non-canonical amino acid

For incorporation of non-canonical amino acids, the pEVOL plasmid (WT or AF mutant) was co-transformed together with the protein expressing plasmid into BL21 DE3 *E. coli* cells (chapter 6.4.1.2). As overnight culture, 30 ml of LB medium in a 100 ml baffled Erlenmeyer flask with appropriate antibiotics (chapter 6.1.2, including Cmp for the pEVOL plasmid) was inoculated with a single colony of *E. coli* BL21 (DE3) cells from an LB-agar plate or with a

small amount of cells from a cryo stock. In a 5 l baffled Erlenmeyer flask, 1 l of TB medium was inoculated to an OD_{600} ~ 0.1 with overnight culture. Cells were incubated (130 rpm, 37 °C) for 1 h, supplemented with an appropriate amount of non-canonical amino acid (1-5 mM PrK, 1 mM SCoK, 1 mM NoK) and incubated for an additional 1 h. Arabinose (0.2% (w/v) was added to induce the pEVOL plasmid and cells were incubated for an additional 5 h and cooled on ice for ~ 15 min afterwards. Protein expression was induced by addition of 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and cells were incubated for ~ 48 h (130 rpm, 20 °C) and afterwards harvested (8000 x g, 30 min, 4 °C). If not purified immediately, pelleted cells were frozen in liquid nitrogen and stored at -80 °C.

6.3.4 Protein purification

6.3.4.1 Standard protocol

Cell pellet from 1 l medium was resuspended in 30 ml IMAC A buffer, supplemented with one tablet EDTA-free protease inhibitor cocktail (Roche), 200 µg of RNase A (Invitrogen) and 20 U of Turbo DNase (Thermo Fisher Scientific). Lysis was performed via sonification with a Sonopuls ultrasonic homogenizer (duty cycle 6x, power 60%, 10 min, Bandelin), the lysate was separated from cell debris by ultracentrifugation (21000 x g, 1 h, 4 °C). Polyethyleniminine (PEI, 0.017% (w/v)) was added to the lysate and the lysate incubated at 4 °C for 15 min. Precipitated nucleic acids were removed via centrifugation (10000 x g, 20 min, 4 °C). PEI treatment was repeated once and the supernatant was filtered two times afterwards (0.45 µm and 0.2 µm) to remove any remaining particles. The supernatant was loaded onto a Ni^{2+}-NTA column (HisTrap HP, GE Healthcare, utilizing an Äkta start system), which was preequilibrated with IMAC A buffer. The loaded column was washed with IMAC A buffer (6 CV) and protein elution was performed with a linear gradient to IMAC B buffer (10 CV), finishing with pure IMAC B buffer (5 CV). Protein containing fractions (identified by absorption at 280 nm) were analyzed with SDS-PAGE (chapter 6.2.1.1), pooled and concentrated to a volume of 110 µl with a VivaSpin (2 or 20 (depending on initial volume), 10 or 50 kDa molecular weight cutoff (depending on protein construct), PES (Sartorius)). Afterwards, the protein was loaded onto a SEC column (Superdex 75 10/300 GL or Superdex 200 10/300 GL (depending on the protein construct), GE Healthcare, utilizing an Äkta purifier 900 system), which was preequilibrated with SEC buffer. Protein containing fractions (absorption at 280 nm) were analyzed with SDS-PAGE (chapter 6.2.1.1), pooled and concentrated to a volume of 100 µl.

6.3.4.2 Protocol for Nhp2 (WT and with incorporated non-canonical amino acid)

Since it was not possible to remove all RNA binding to Nhp2 with the standard protocol, additional measures had to be taken. The purification was done as described in chapter 6.3.4.1. However, after the first Ni^{2+}-NTA purification, protein containing fractions (absorption at 280 nm) were analyzed by SDS-PAGE (chapter 6.2.1.1), pooled and incubated with 400 µg RNase A (20 °C, 16 h) to remove copurified contaminating RNA. Afterwards,

the protein solution was diluted to an imidazole concentration of ~ 20 mM and RNase A was removed via a second Ni^{2+}-NTA column purification.

6.3.4.3 Protocol for Cbf5/Cbf5Δ with incorporated non-canonical amino acid together with Nop10 and Gar1

Expression of Cbf5/Cbf5Δ with an incorporated non-canonical amino acid (together with WT Nop10) was performed according to chapter 6.3.3.3. Expression of Gar1 (WT) was performed according to chapter 6.3.3.2. Cell Pellets of both expression cultures were mixed prior to lysis and purification of proteins was done according to standard protocol (chapter 6.3.4.1).

6.3.5 Click reactions for protein dye labeling

6.3.5.1 Cu^{+} catazalyzed alkyne-azide cycloaddition (CuAAC)

Copper catalyzed click reaction was utilized to label proteins with incorporated propargyl-lysine (PrK) with a fluorescence dye (sulfo-Cy3-azide or sulfo-Cy5-azide). Test reactions were usually done in a 5 µl reaction volume in SEC buffer (1x) with $CuSO_4$ (500 µM), THPTA (2.5 mM) (both were premixed prior to the reaction), PrK containing protein (50 µM), sulfo-Cy3/Cy5-azide (150 µM, Jena Bioscience), aminoguanidine (5 mM) and sodium ascorbate (5 mM). Test reactions were incubated at different temperatures and for different amounts of time under exclusion of light. Samples were mixed with 2x SDS loading dye (1:1) and analyzed via SDS-PAGE (chapter 6.2.1.1) and subsequently scanned (chapter 6.2.2.3).
Labeling reactions were usually done in a 110 µl reaction volume in SEC buffer (1x) with $CuSO_4$ (500 µM), THPTA (2.5 mM) (both were premixed prior to the reaction), PrK containing protein (100 µM), sulfo-Cy3/Cy5-azide (250 µM, Jena Bioscience), aminoguanidine (5 mM) and sodium ascorbate (5 mM). Samples were incubated for 3 h at 37 °C (Nhp2) or for 6 h at 4 °C (Cbf5). The reaction mixture was loaded onto a SEC column (Superdex 75 10/300 GL or Superdex 200 10/300 GL (depending on the protein construct), GE Healthcare, utilizing an Äkta purifier 900 system), which was preequilibrated with SEC buffer. Protein and dye containing fractions (identified by absorption at 280 nm and 550 nm) were analyzed with SDS-PAGE (chapter 6.2.1.1) and fluorescence detection (chapter 6.2.2.3), pooled and concentrated to a volume of 100 µl.

6.3.5.2 Strain promoted alkyne-azide cycloaddition (SPAAC)

A copper-free click reaction was utilized to label proteins with incorporated strained cyclooctyne-lysine (SCOK) with a fluorescence dye (sulfo-Cy3-azide or sulfo-Cy5-azide). Test reactions were usually done in a 5 µl reaction volume in SEC buffer (1x) with SCOK containing protein (50 µM) and sulfo-Cy3/Cy5-azide (150 µM, Jena Bioscience). Test reactions were incubated at different temperatures and for different amounts of time under exclusion of light. Samples were mixed with 2x SDS loading dye (1:1) and analyzed via SDS-PAGE (chapter 6.2.1.1) and subsequently scanned (chapter 6.2.2.3).

Labeling reactions were usually done in a 110 µl reaction volume in SEC buffer (1x) with SCOK containing protein (100 µM) and sulfo-Cy3/Cy5-azide (250 µM, Jena Bioscience). Samples were incubated for 24 h at 37 °C. The reaction mixture was loaded onto a SEC column (either Superdex 75 10/300 GL or Superdex 200 10/300 GL (depending on the protein construct), GE Healthcare, utilizing an Äkta purifier 900 system), which was preequilibrated with SEC buffer. Protein and dye containing fractions (absorption at 280 nm and 550 nm) were analyzed with SDS-PAGE (chapter 6.2.1.1) and fluorescence detection (chapter 6.2.2.3), pooled and concentrated to a volume of 100 µl.

6.3.6 Reactions for spin labeling

6.3.6.1 Strain promoted inverse electron Diels Alder cycloaddition (SPIEDAC)

Copper free click reaction was utilized to label proteins with incorporated norbornene-lysine (NoK) with tetrazine-TEMPO (see appendix, Figure 148). Labeling reactions were usually done in a 500 µl reaction volume in SEC buffer (1x) with NoK containing protein (40 µM) and spin label (400 µM). Samples were incubated for 1 h at 25 °C. The protein concentration was diluted to 5 µM (volume of 4 ml), the label concentration was brought to 200 µM and the reaction was continued for 1 h at 25 °C. Afterwards, samples were brought to a volume of 50 µl using a VivaSpin (500, 5 kDa molecular weight cutoff, PES (Sartorius)). The reaction mixture was purified with Protein Desalting Spin Columns (Thermo Scientific) to remove unincorporated spin label.

6.3.6.2 MTSSL labeling

Labeling reactions were usually done in a 4 ml reaction volume in SEC buffer (1x) with cysteine mutant of protein (5 µM) and MTSSL (200 µM). Samples were incubated for 2 h at 25 °C. Afterwards, samples were brought to a volume of 50 µl using a VivaSpin (500, 5 kDa molecular weight cutoff, PES (Sartorius)). The reaction mixture was purified with Protein Desalting Spin Columns (Thermo Scientific) to remove unincorporated MTSSL.

6.4 DNA based methods

6.4.1 Transformation

6.4.1.1 DH5α high efficiency *E. coli* cell line

5 µl of plasmid (10 - 50 ng, typically from a ligation reaction or a site directed mutagenesis) was mixed with 50 µl of cell line and stored for 30 min on ice. Heat shock (42 °C) was done for 30 sec and cell line was stored on ice afterwards for 5 min. SOC medium (300 µl) was added and cells were shaked (600 rpm) at 37 °C for 30 min. Cell line was spread on LB-Agar plates and cells were incubated at 37 °C over night or at room temperature over the weekend.

6.4.1.2 BL21 (DE3) *E. coli* cell line

1 µl of plasmid (typically 50 – 150 ng, purified and sequenced) was mixed with 50 µl of cell line and stored for 15 min on ice. Heat shock (42 °C) was done for 45 sec and cell line was stored on ice afterwards for 2 min. SOC medium (300 µl) was added and cells were shaked (600 rpm) at 37 °C for 30 min. Cell line was spread on LB-Agar plates and cells were incubated at 37 °C over night or at 22 °C over the weekend.

6.4.2 DNA isolation, preparation and sequencing

In a Greiner tube, 5 ml of LB medium with the appropriate antibiotic was inoculated with one colony of cells from a LB-Agar plate. Cells were shaked (160 rpm) at 37 °C over night. DNA preparation was done with QIAprep Spin Miniprep Kit (Qiagen) according to manufacturer's protocol. If a larger amount of sample was needed, a larger overnight culture in a baffled Erlenmeyer flask was prepared, and instead the QIAprep Spin Midiprep Kit (Qiagen) was used according to manufacturer's protocol. Sequencing of samples was done by Eurofins Genomics or SeqLab.

6.4.3 PCR based methods

6.4.3.1 Site directed mutagenesis

Site directed mutagenesis was done utilizing the QuikChange Site-directed Mutagenesis Kit (Agilent Genomics) or the Lightning QuikChange Site-directed Mutagenesis Kit (Agilent Genomics). Complementary primers were designed according to the manufacturer's protocol and ordered from Eurofins Genomics (Munich, Germany). Reactions were usually performed in a 20 µl reaction volume, with 10x buffer (1x), dNTP-Mix (0.2 mM), primers (0.3 µM each), plasmid template (10-100 ng) and Pfu polymerase (0.4 µl) (1 U). A peqSTAR PCR cycler (Peqlab) with the appropriate temperature program (Table 12) was used for the PCR. In case of the Lighting Kit, additionally Quik Lightning solution (0.6 µl) was include in the reaction mix.

Table 12: PCR conditions for site directed mutagenesis with QuikChange Site-directed Mutagenesis Kit (left) and Lightning QuikChange Site-directed Mutagenesis Kit (right)

Temperature	Time	No. of repeats
95 °C	30 s	0
95 °C	30 s	17
55 °C	1 min	
68 °C	7 min	
4 °C	end	0

Temperature	Time	No. of repeats
95 °C	2 min	0
95 °C	20 s	17
60 °C	10 s	
68 °C	5 min	
68 °C	5 min	0
4 °C	end	0

6.4.3.2 Amplification

For amplification of DNA, either as an insert for ligation or for generation of transcription template, PCR with homemade Phusion polymerase was utilized. Primers were designed accordingly, enclosing the sequence of interest. Reactions were usually performed in 100 µl reaction volume, with 5x HF buffer (1x), dNTP-Mix (0.2 mM), primers (3 µM each), plasmid template (10-100 ng) and Phusion polymerase (0.4 U/µl). A peqSTAR PCR cycler (Peqlab) with the appropriate temperature program (Table 13) was used for the PCR.

Table 13: PCR conditions for amplification

Temperature	Time	No. of repeats
98 °C	2 min	0
54 °C	15 s	34
68 °C	15-45 s	
98 °C	30 s	
4 °C	end	0

6.4.3.3 Amplification of whole plasmid

Whole plasmid PCR was used amplify the whole plasmid except a specific part of the plasmid, which was desired to be cut out. Primers were designed accordingly, enclosing the whole plasmid away from both sides of the region that was supposed to be cut out. Reactions were usually performed in 100 µl reaction volume, with 5x HF buffer (1x), dNTP-Mix (0.2 mM), primers (3 µM each), plasmid template (10-100 ng) and Phusion polymerase (0.4 U/µl). A peqSTAR PCR cycler (Peqlab) with the appropriate temperature program (Table 14) was used for the PCR.

Table 14: PCR conditions for whole plasmid amplification

Temperature	Time	No. of repeats
98 °C	30 s	0
98 °C	10 s	29
65 °C	20 s	
72 °C	7 min	
4 °C	end	0

6.4.3.4 Colony PCR

For colony PCR, first a ligation mix (chapter 6.4.7) was transformed into DH5α high efficiency *E. coli* cell line and cells were grown overnight (chapter 6.4.1.1). For primers, the T7 and T7-terminator primer were utilized. Reactions were usually performed in 20 µl reaction volume, with 5x HF buffer (1x), dNTP-Mix (0.2 mM), primers (3 µM each) and Phusion polymerase (0.4 U/µl). For DNA template, a single colony was picked from the LB-Agar plate and mixed with the reaction mixture prior to polymerase addition. A peqSTAR PCR cycler (Peqlab) with the appropriate temperature program (Table 15) was used for the PCR.

Table 15: PCR conditions for colony PCR

Temperature	Time	No. of repeats
95 °C	5 min	0
58 °C	30 s	29
70 °C	90 s	
95 °C	30 s	
70 °C	5 min	0
4 °C	end	0

6.4.4 DpnI digestion reaction

DpnI digestion reaction, if the removal of DNA template of bacterial origin after the PCR reaction was required, was performed by adding DpnI (20 U) to the reaction mixture and incubation at 37 °C for 1.5 h.

6.4.5 Restriction digestion reaction

Analytical restriction digestion reactions were usually performed in a 20 µl reaction volume in Cutsmart buffer (1x) with DNA (200 ng) and either one or two restriction enzymes (5 U per 200 ng of DNA, NEB). Restriction digestion reactions were incubated at 37 °C for 2 h. Afterwards, completion of digestion was tested with an agarose gel (chapter 6.2.1.5) and cleanup of reactions was done (chapter 6.4.6).

Preparative restriction digestion reactions were usually performed in a 50 - 200 µl reaction volume in Cutsmart buffer (1x) with DNA (2-20 µg) and two restriction enzymes (5-10 U per µg of DNA, NEB). For vector DNA, shrimp alkaline phosphatase (rSAP, 5-10 U per µg of DNA, NEB) for simultaneous dephosphorylation was added during restriction digestion reaction. Restriction digestion reactions were incubated at 37 °C for 2 h. Afterwards, completion of digestion was tested with an agarose gel (chapter 6.2.1.5) and cleanup of reactions was done (chapter 6.4.6).

6.4.6 DNA purification

If DNA purification was necessary for subsequent reactions, the PCR or digestion reaction mixture was purified with the QIAquick PCR purification kit (Quiagen) according to the manufacturer's protocol. DNA was eluted with H_20 (30 µl).

6.4.7 Ligation of insert into vector

For the ligation of an insert DNA into a vector DNA, first restriction digestion reaction had to be performed with the appropriate restriction enzymes (chapter 6.4.5). Ligation reactions were usually performed in a 40 µl reaction volume in DNA ligase buffer (1x) with vector DNA (186.9 ng, 2.9 nM), insert DNA (83.6 ng, 14.2 nM) and T4 DNA ligase (20 U/µl) at 25 °C for 45 min. Afterwards, transformation in DH5α high efficiency *E. coli* cells (chapter 6.4.1.1), colony PCR (chapter 6.4.3.4) and analysis via agarose gel electrophoresis (chapter 6.2.1.5) were performed.

6.4.8 Religation

For closing a shortened plasmid after whole plasmid PCR (chapter 6.4.3.3), plasmid was phosphorylated and religated. Ligation reactions were usually performed in a 50 µl reaction volume in DNA ligase buffer (1x) with plasmid DNA (~1 µg), T4 polynucleotide kinase (PNK, 10 U per µg of DNA, NEB) and T4 DNA ligase (2000 U per µg of DNA, NEB) at 25 °C for 2 h. Afterwards, transformation in DH5α high efficiency *E. coli* cells (chapter 6.4.1.1), colony PCR (chapter 6.4.3.4) and analysis via agarose gel electrophoresis (chapter 6.2.1.5) were performed.

6.5 RNA based methods

6.5.1 Standard *in vitro* transcription reaction

In vitro transcription reactions were usually performed in a 200 µl reaction volume in Tris-Buffer (1x) with NTPs (ATP, CTP, GTP, UTP, 5 mM of each), magnesium acetate (30 mM), Tris/HCl (100 mM), DTT (10 mM), spermidine (3 mM), DNA template (10-100 ng) and homemade T7 RNA polymerase (24 ng/µl). Samples were incubated at 37 °C for 3 h - 18 h. For overnight reactions, additionally RNasin (40 U, PROMEGA) was added to the reaction mixture. DNA was digested with Turbo DNase (0.02 U/µL, Thermo Fisher Scientific) at 37 °C for 1 h. Afterwards, pyrophosphate precipitation was dissolved with EDTA (40 mM final conc.) and enzymes were removed via phenol extraction (chapter 6.5.3). RNA sequences are summarized in Table 16.

Table 16: snoRNA constructs.

snR81 FL	GGGACUGCAAAAGAAGCGGCGAGGCAGCCCACAUCAAGUGGAACUAC ACAGACUUCCUUGUCGCGAUACUACGGUCCCAAGAGCAAUCCUAACA AGCAAUUACAUAUUCCCCCGCUGAACCUGUACAGUCCACGGAUGGUG CAGAAGUUAUAUGAUUUGGGGGAAGACGCUUUUUCACAUCUUCUUG CAUGAUAA
H5	GGGACUGCAAAAGAAGCGGCGAGGCAGCCCACAUCAAGUGGAACUAC ACAGACUUCCUUGUCGCGAUACUACGGUCCCAAGAGCAAUCCU
H5Δ	GGGACUGCAAAAGAAGCGGCUUGGGUCGCGAUACUACGGUCCCAAGA GCAAUCCU
H3	AACAAGCAAUUACAUAUUCCCCCGCUGAACCUGUACAGUCCACGGAU GGUGCAGAAGUUAUAUGAUUUGGGGGAAGACGCUUUUUCACAUCUU CUUGCAUGAUAA
H3Δ	AACAAGCAAUUACAUAUUCCCCCGCUUGGGGGAAGACGCUUUUUCAC AUCUUCUUGCAUGAUAA

6.5.2 GMP transcription reaction

GMP transcription was done according to the standard protocol (chapter 6.5.1), however NTP concentration was lowered to 3.75 mM of each NTP and additionally GMP (15 mM) was added to the reaction mixture.

6.5.3 Phenol extraction

Phenol extraction was performed to remove proteins from RNA solution. Aqua phenol (1 volume equivalent, Carl Roth) was added to RNA/protein containing sample and the mixture was vortexed and afterwards centrifuged for phase separation (2000g, 1 min). The aqueous and organic phases were separated and diethyl ether (3 volume equivalents, water saturated) was added to the aqueous phase. The mixture was again vortexed and centrifuged (2000g, 1 min) and the aqueous and organic phases were separated. The organic phase was discarded and the diethyl ether washing step was repeated once. Remaining organic solvent residues in

the aqueous phase were removed by vaporization at room temperature or with a Concentrator Plus Speedvac (setting 'V-Al', 5 min, 22 °C, Eppendorf).

6.5.4 Precipitation and purification of RNA

RNA from transcription reactions was precipitated with addition of NH_4OAc (0.5 M final conc.) and either ethanol (2.5 volume equivalents) or isopropanol (1 volume equivalent), subsequent storage at -20 °C for several hours or at -80 °C for 1 h, and centrifugation (17000 x g, 45 min, -4 °C). The supernatant was removed and excess solvent was removed with a Concentrator Plus Speedvac (setting 'V-Al', 5-10 min, 22 °C, Eppendorf) and pelleted RNA was resuspended in H_2O/formamide (1:1, v/v) and the product was separated from side products by denaturing urea-PAGE (chapter 6.2.1.2). The product was identified via UV shadowing, excised and eluted from the gel in NH_4OAc (0.5 M) by shaking (1000 rpm) at 22 °C overnight. RNA was again precipitated and excess organic solvent removed as described before. Afterwards, RNA was resuspended in H_2O.

6.5.5 Generation of radioactively labeled substrate RNA

6.5.5.1 In vitro transcription with [α-^{32}P] UTP

In vitro transcription to generate universally ^{32}P-labeled substrate RNAs was done as described in chapter 6.5.1. However, the UTP was substituted by [α-^{32}P] UTP. After the transcription reaction, the reaction mixture was purified with illustra MiroSpin G-25 columns (GE Healthcare) to remove unincorporated [α-^{32}P] UTP. Radioactive counts per minute were determinted using a Tri-Carb 2100TR Liquid Scintillation Analyzer (Packard) and incorporation rate of [α-^{32}P] UTP was determined by $\frac{cpm\,(before\ filtration)}{cpm\,(after\ filtration)}$. Concetration of the labeled RNA was determined by $c = \frac{incorporation\ rate \cdot concentration\ of\ ^{32}P-\alpha-\mathrm{UTP}}{numer\ of\ uridine\ nucleotides}$

6.5.5.2 Phosphorylation and splinted ligation with [γ-^{32}P] ATP

Splinted ligation was used to generate site specifically ^{32}P-labeled substrate RNAs. Oligonucleotides (Table 17) were purchased from Eurofins Genomics. Initial phosphorylation reactions were usually done in a 30 µl reaction volume in T4 ligase buffer (1x) with the oligonucleotide containing the target uridine at the 5' end (M172/M174, 166.7 µM), [γ-^{32}P] ATP (0.4 µM, Hartmann Analytic) and T4-Polynucleotidkinase (0.67 U/µl) at 37 °C for 30 min. Afterwards, non-radioactive ATP (10 mM) was added to a final volume of 40 µl and reaction mixture was incubated at 37 °C for 30 min. The reaction mixture was purified with illustra MiroSpin G-25 columns (GE Healthcare) to remove unincorporated [γ-^{32}P] ATP. Annealing was usually done in a 95 µl reaction volume in T4 ligase buffer (0.5x) with the now partially radioactive, fully phosphorylated 3' oligonucleotide (50 µM), the corresponding 5' oligonucleotide (M171/M173, 50 µM) and the corresponding DNA splint (M130/M132, 50 µM) at 75 °C for 2 min, with subsequent cooling to 22 °C and afterwards to 4 °C. For ligation, T4 ligase buffer was adjusted to 1x, T4-DNA-ligase (80 U/µl, NEB) was added and

reaction mixture was incubated at 16 °C overnight. Afterwards, Turbo DNase (0.11 U/µl, Invitrogen) was added and reaction mixture incubated at 37 °C for 1.5 h. Ligated products were separated from unreacted oligonucleotides with denaturing PAGE (chapter 6.2.1.2), identified by autoradiography (chapter 6.2.2.4), excised and eluted from the gel in NH_4OAc (0.5 M) by shaking (1000 rpm) at 22 °C overnight. RNA was precipitated by addition of ethanol (2.5 volume equivalents), storage at -20 °C for 2 h and centrifugation (17000 x g, 45 min, -4 °C). Afterwards, RNA was resuspended in H_2O.

Table 17: Oligonucleotides for ligation. Target uridine is marked in red.

5' substrate (5' end, M171)	5'-GGGAGUAGUAUC-3'
5' substrate (3' end, M172)	5'-**U**GUUCUUUUCAG-3'
3' substrate (5' end, M173)	5'-GGGAACUU-3'
3' substrate (3' end, M174)	5'-**U**AAAUAUGUAAGAA-3'

6.6 Pseudouridylation activity assays

Since RNPs were reconstituted with an excess of proteins, the limiting factor for enzymatic activity is the snoRNA concentration. Pseudouridylation activity assays were either performed as single turnover assays (snoRNA/substrate 2:1) or as multiple turnover assays (snoRNA/substrate 1:40). The ^{32}P-labeled substrate RNA was spiked with non-labeled substrate RNA due to the use of non-radioactive ATP for completion of the phosphorylation reaction after the initial phosphorylation with [γ-^{32}P] ATP. For multiple turnover reactions, additional non-labeled substrate RNA was used for spiking.

6.6.1 Single turnover assays

Pseudouridylation activity assays with radioactively labeled substrate RNA were done to test the enzymatic activity of different H/ACA RNP subcomplexes. For single turnover conditions, reconstitution of RNPs was usually performed in Ψ-buffer (1x) with snoRNA (1 µM), different proteins (2 µM for single hairpin constructs, 4 µM for double hairpin constructs) and ^{32}P-labeled substrate RNA (500 nM). After addition of substrate RNA, samples were immediately incubated at 30 °C. After 1-2 h, samples were mixed with water/aquaphenol (1:1) at 4 °C to stop the enzymatic reaction.

6.6.2 Multiple turnover assays

For multiple turnover conditions, reconstitution of RNPs was usually performed in Ψ-buffer (1x) with snoRNA (100 nM), different proteins (1 µM for single hairpin constructs, 2 µM for double hairpin constructs), ^{32}P-labeled substrate RNA (100 nM) and unlabeled substrate RNA (3.9 µM). After addition of substrate RNA, samples were immediately incubated at 30 °C. At different points of time, aliquots of 20 µl were taken from the samples and immediately mixed with water/aquaphenol (1:1) at 4 °C to stop the enzymatic reaction.

6.6.3 RNA digestion and thin layer chromatography

A phenol extraction was performed (chapter 6.5.3). RNA was precipitated with addition of NaOAc (0.3 M) and ethanol (2.5 volume equivalents), subsequent storage at -20 °C for several hours and centrifugation (17000 x g, 45 min, -4 °C). Afterwards, RNA was resuspended in P1 buffer and P1-endonuclease (0.0125 U/µl) was added. Samples were incubated at 55 °C for 3.5 h and afterwards at 70 °C for 30 min to fully digest all RNA. 2-3 µl of each sample was spotted on a thin layer chromatography cellulose plate (20 cm height for 1D TLC, 10x10 cm for 2D TLC, Merck) and thin layer chromatography was performed with TLC buffer A for 1D TLC or first with TLC buffer B and afterwards with TLC buffer A for 2D TLC. TLC plates were visualized by autoradiography chapter 6.2.2.4 and spots assigned to known R_f values[369]. Image analysis and integration of peaks was done with ImageJ, determination of pseudouridylation yield was done by comparison of intensity of U and Ψ spots.

6.7 Single molecule FRET spectroscopy

6.7.1 Preparation of cover slips and objective slides

For immobilization of biotinylated samples and subsequent smFRET experiments, PEG-passivated and biotin activated cover slips and objective slides to prepare a measurement chamber were prepared. Cover slips (24x60 mm, Carl Roth) were cleaned by being exposed to oxygen plasma for 15 min. Cover slip holding chambers were cleaned by MeOH and sonification for 5 min. Afterwards, silanization was performed in the holding chambers with 3-aminopropyltriethoxysilane (1 % (v/v)) and acetic acid (5% (v/v)) in methanol under sonication for 1 minute and incubation at room temperature for an additional 20 minutes. Cover slips were thoroughly washed with H_2O and dried in a nitrogen stream. Surface functionalization was performed with PEG/PEG-biotin (33 mM mPEG-succinimidyl valerate, MW = 5 kDa; 0.7 mM biotin-PEG-succinimidyl valerate, MW = 5 kDa; NANOCS) in sodium bicarbonate buffer (100 mM) at 25 °C overnight in a moisture chamber. Afterwards, cover slips were thoroughly washed with H_2O, dried in a nitrogen stream to remove excess PEG and stored under argon at -80 °C. Objective slides (26x76 mm, Carl Roth) were cleaned prior to smFRET measurement by being exposed to nitrogen plasma for 15 min and measurement channels were generated by combining cover slips and objective slides with double-sided sticky tape (Scotch).

6.7.2 Preparation of fluorescent labeled RNP samples

Reconstitution of Cy3/Cy5 double labeled RNP complexes was usually done in a 4 µl reaction volume in Ψ buffer (1x).
For samples with fluorophore-label on snoRNA and Nhp2, fluorophore labeled (Cy5 or Cy3) biotinylated snoRNA (600 nM), fluorophore labeled (Cy3 or Cy5) Nhp2 (6 µM) and other proteins (Nop10, Cbf5, Gar1, Gar1Δ (depending on the constructs measured), 6 µM) were used.
For samples with fluorophore label on snoRNA and Cbf5/Cbf5Δ, fluorophore labeled (Cy5 or Cy3) biotinylated snoRNA (600 nM), fluorophore labeled (Cy3 or Cy5) Cbf5/Cbf5Δ (co-purified with Nop10 and Gar1, 6 µM) and Nhp2 (6 µM) were used.
For samples with fluorophore-label only on Nhp2, unlabeled biotinylated full length snoRNA (600 nM), Cy3-labeled Nhp2 (3 µM), Cy5-labeled Nhp2 (3 µM) and other proteins (Nop10, Cbf5, Gar1, 6 µM) were used.
Samples were incubated at room temperature for 5 min and placed on ice afterwards.

6.7.3 Immobilization of RNP sample on measurement channel

Prior to smFRET measurement, biotinylated RNP samples were immobilized on a passivated and biontinylated glass surface via biotin-streptadivin-biotin interactions. Each sample was diluted in Ψ buffer (1x) to a final snoRNA concentration of 1 nM – 100 pM (usually 250 pM). Measurement channels were prepared as described in chapter 6.7.1. For each sample, a channel was flushed consecutively with streptavidin (15 µL, 0.2 mg/ml in T50 buffer (1x)), Ψ

buffer (70 µL), diluted sample (5-20 µL) and freshly prepared Imaging buffer (40 µL). The amount of diluted sample was adjusted to yield a molecule density of ~200-300 molecules per field of view. For measurements with substrate, substrate RNA (100 nM) was added to the Imaging buffer.

6.7.4 smFRET microscopy setup

smFRET measurements were performed on a computer controlled, objective-type spinning-spot total internal reflection microscopy setup with a microscope (IXI71, Olympus), an electron multiplying charge coupled device (EMCCD) camera (iXon897, Andor Technology) a green (532 nm, 150 mW) and a red (633 nm, 140 mW) diode laser (Coherent, OBIS), an acousto-optical tunable filter (AOTFnC-400.650-TN, AA Opto-Electronic) a NA 1.4 100x oil immersion objective (UPLSAPO100XO, Olympus), and a dual view emission splitter (OTPOSPLIT II, Cairn Research).

6.7.5 smFRET data aquisition

smFRET measurements were performed with continuous green laser excitation at 22 °C. For data Acquisition, the software Andor SOLIS (version 4.21) was used. Data acquisition was performed with an integration time of 100 ms and an EM gain of 300. Data was saved as *.sif file-format. Before the measurement of the actual samples, the lasers were set to 10 mW for at least 1 h and the setup was adjusted to TIRF mode. For each day of measurement, a 20-frame calibration movie of a 1:250 dilution of TetraSpeck Microspheres (0.1 µM, Thermo Fisher Scientific) in T50 buffer (1x) was collected. Usually, for each sample 40 70-frame movies were collected, where the green laser was turned on at the start and the red laser was turned on after 25 frames (2.5s). Additionally, for each sample 2 1200-frame movies were collected to check for single-step photobleaching as well as possible FRET dynamics.

6.7.6 smFRET data analysis

Data that was collected in the *.sif file-format, which was converted into the 16-bit-integer *.dat file-format with the software Andor SOLIS (version 4.2), using the softwares batch conversion tool. For further processing of the data, the software IDL (version 6.2, Exelis) with scripts originally written by Hazen Babcock and Greg Bokinsky was utilized. First, the IDL-scripts were used to create a transformation map with the 20-frame calibration movie of TetraSpeck Microspheres for donor and acceptor channel alignment. Afterwards, with this transformation map, background corrected smFRET fluorescence intensities for donor and acceptor channel were generated for each movie by Gaussian point spread function fitting. The STD threshold was set to 3. This procedure generated a *.traces file from each *.dat file. For further processing of the data, the software MatLab (version R2019b, MathWorks) was utilized. For the generation of smFRET histograms out of all 70 frame *.traces files from a single sample, the intensity of donor (I_D) and acceptor (I_A) of the first 20 frames of each 70 frame movie was averaged and the FRET efficiency was determined as E-FRET = $\frac{I_A}{I_A+I_D}$ with a fretthreshold of 2000. E-FRET values of all molecules were binned into intervals of 0.025

and the center of the donor-only peak was adjusted to E-FRET = 0 using a correction factor. Afterwards, the I_D and I_A of the last 40 frames of each movie was averaged and subtracted from each other, and this value had to exceed a certain threshold for the molecule to be accepted for further analysis. This procedure was performed to only analyze molecules with an intact FRET pair and to exclude molecules carrying bleached dyes. For further processing of the data, the software OriginPro (version 2019, OriginLab) was utilized. Histograms were plotted from binned data. For fitting, the Gaussian function

$$y = y_0 + \frac{A}{w\sqrt{\frac{\pi}{2}}} e^{-2\frac{(x-x_c)^2}{w^2}}$$

with peak center x_c, peak width w and area under the curve A was utilized for E-FRET distribution of different states.

For the generation of smFRET traces, the 1200 frame *.traces files from a single sample was plotted as I_D, I_A and E-FRET against time (t) for each individual molecule. Data of each molecule that showed typical FRET behavior was manually exported as *.mat file and "stitched" together into a single *.dat file. For further processing of the data, the software HaMMy[381] (version 4.0) was utilized. Each *.dat file from a single sample measurement was used to determine the number of E-FRET states and the E-FRET value of each state utilizing Hidden Markov Modelling.

6.8 Synthesis of non-canonical amino acids

6.8.1 General synthesis methods

All reactions with substances volatile to moisture were carried out under inert conditions (argon or nitrogen atmosphere) and glass ware was heated under vacuum ($p = 10^{-2}$ mbar) to remove surface adsorbed water. Glass joints were sealed with silicon grease.
Organic solvents were ordered in absolute (THF, DMF, $CHCl_3$) or p.a. quality (THF, ethyl acetate, 1,4-dioxane, AcOH, DCM, MeOH, formic acid).
Chemicals for synthesis were ordered from Sigma Aldrich and Carl Roth.

6.8.2 Characterization and analysis of synthesized compounds

6.8.2.1 NMR spectroscopy

NMR spectra of synthesized compounds were recorded in deuterated solvents chloroform-*d* ($CDCl_3$) and DMSO-d_6, which were dried over molecular sieve (4 Å).
NMR spectroscopy was performed on Bruker NMR spectrometers.

- DPX-250 : δ(1H) = 250.1 MHz
- Avance-500 : δ(1H) = 500.2 MHz, δ(13C) = 125.8 MHz

Given data of δ is in ppm with positive values a low field shift. ^{1}H and ^{13}C spectra are calibrated to literature values of incompletely deuterated solvents.

- CDCl3 : δ (1H) = 7.26 ppm, δ(13C{1H}) = 77.16 ppm
- DMSO-*d*6 : δ (1H) = 2.50 ppm, δ(13C{1H}) = 39.52 ppm

For signal assignment, the following abbreviations are used:

s = singulett, d = dublett, dd = dublett from dubletts, t = triplett, m = multiplett

6.8.2.2 ESI

ESI mass spectrometry was performed by the mass spectrometry service unit of the Goethe university.

6.8.3 Synthesis of propargyl-lysine (PrK)

6.8.3.1 Synthesis of compound 6

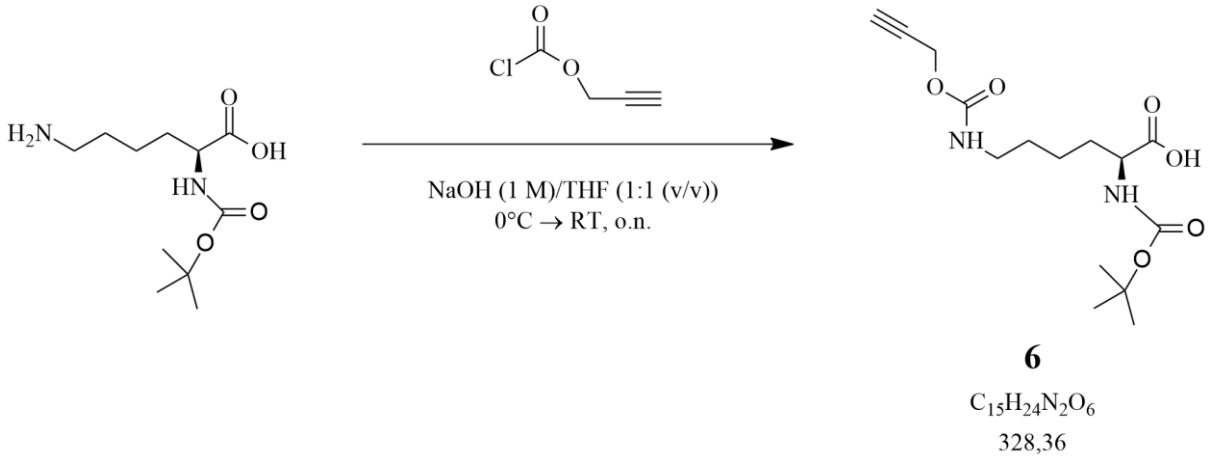
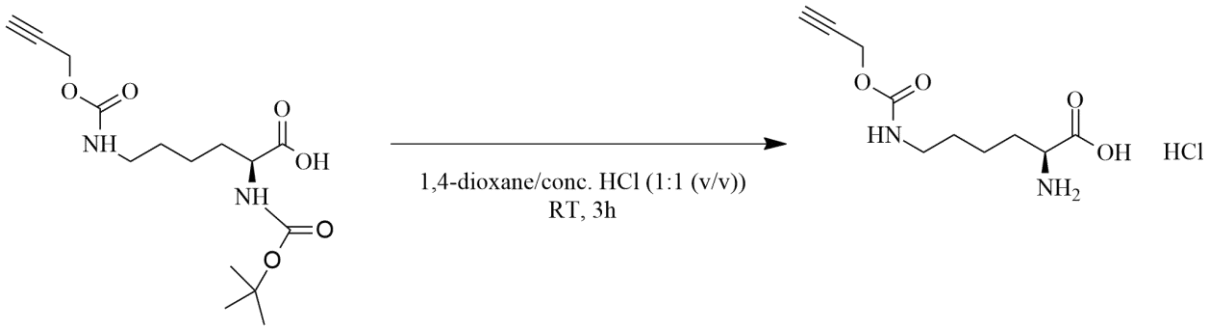


Figure 152: Reaction conditions for synthesis of compound 6.

Synthesis of compound 6 was performed according to literature[199].
Nα-(tert-Butoxycarbonyl)-L-lysine (20.0 g, 81.0 mmol, 1.20 eq) was dissolved in NaOH (1M)/THF (100 ml, 1:1 (v/v)). Reaction mixture was cooled on ice to 0 °C and propargyl chloride (6.60 ml, 8.02 g, 67.7 mmol, 1.00 eq) was added dropwise over the course of 40 min. Reaction mixture was allowed to warm up to RT and stirred at RT overnight. Afterwards, the reaction mixture was taken up in ethyl acetate (200 ml) and the aqueous phase was acidified to pH = 2 with conc. HCl. Phases were separated, the aqueous phase was washed with ethyl acetate (3 x 100 ml), organic phases were combined and dried with sodium sulfate. Organic solvents were removed with a rotary evaporator and product was dried under vacuum. Compound 6 was obtained as a colorless oil (23.3 g, 100%) and was directly used for the next reaction step.

6.8.3.2 Synthesis of compound 1



6 **1**

$C_{10}H_{17}ClN_2O_4$

264,71

Figure 153: Reaction conditions for synthesis of compound 1.

Synthesis of compound 1 was done according to literature[199].
Compound 6 (23.3 g, 71.0 mmol, 1.00 eq) was dissolved in 1,4-dioxane/conc. HCl (200 ml, 1:1 (v/v)). Reaction mixture was stirred at RT for 4 h. Organic solvents were removed with a rotary evaporator and product was dried under vacuum. Afterwards, the product was taken up in HCl (1 M, 100 ml), frozen in liquid nitrogen in lyophilized. HCl salt of compound 1 was obtained as a beige powder (18.2 g, 97%).

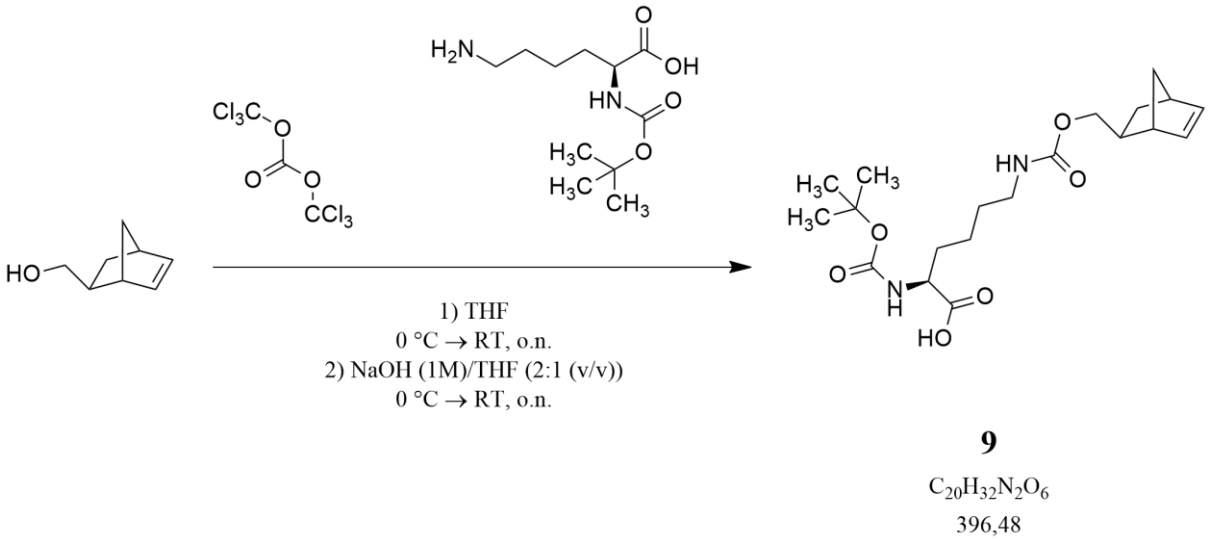
^{1}H-NMR (500.2 MHz, DMSO-d_6): δ = 8.29 (s, 2H, NH_2), 7.35 (t, 1H, $^3J_{HH}$ = 5.6 Hz, NH), 4.60 (d, 2H, $^4J_{HH}$ = 2.4 Hz, CH_2), 3.86 (m, 1H, α-CH^{lys}), 3.47 (t, 1H, $^4J_{HH}$ = 2.4 Hz, $CH^{propargyl}$), 2.98 (d, 2H, $^3J_{HH}$ = 6.0 Hz, ε-CH_2^{lys}), 1.82-1.71 (m, 2H, β-CH_2^{lys}), 1.44-1.36 (m, 3H, δ-CH_2^{lys}, γ-CH_2^{lys}), 1.32-1.22 (m, 1H, γ-CH_2^{lys}).

^{13}C{^{1}H}-NMR (125.78 MHz, DMSO-d_6): δ = 171.07 (C(O)ON), 155.27 (C(O)OC), 79.46 ($C^{propargyl}$), 77.08 ($CH^{propargyl}$), 51.84 (α-CH^{lys}), 51.40 (CH_2), 40.02 (ε-CH_2^{lys}), 29.62 (β-CH_2^{lys}), 28.77 (δ-CH_2^{lys}), 21.58 (γ-CH_2^{lys}).

R_f value: 0.60 (*n*-butanol/acetic acid/H_2O 5 : 2 : 3 (v/v))

6.8.4 Synthesis of norbornene-lysine (NoK)

6.8.4.1 Synthesis of compound 9



9

$C_{20}H_{32}N_2O_6$

396,48

Figure 154: Reaction conditions for synthesis of compound 9.

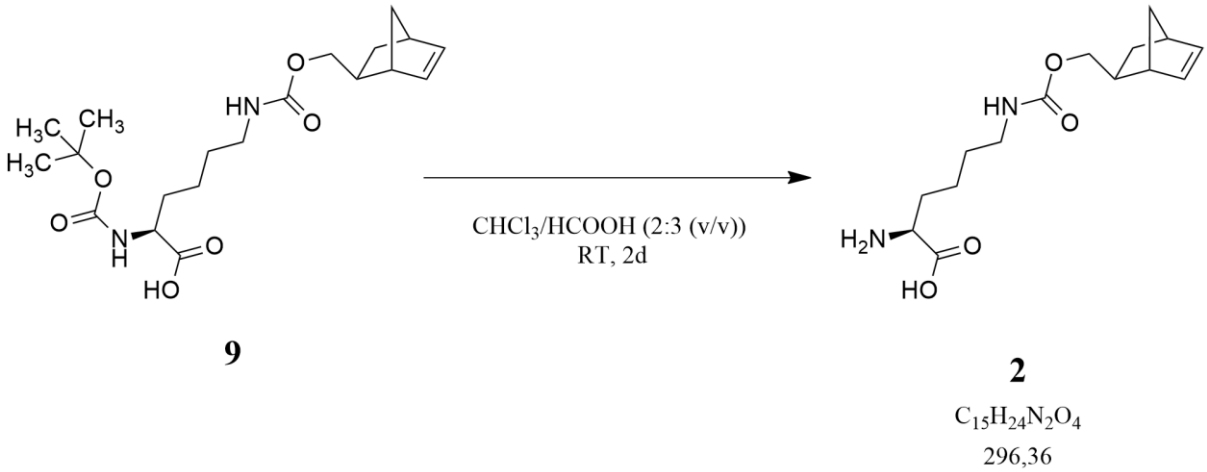
Synthesis of compound 9 was done according to literature[201].
Triphosgene (5.00 g, 16.8 mmol, 0.72 eq) was dissolved in dry THF (35 ml) and cooled on ice to 0 °C. 5-norbornene-2-methanol (2.90 g, 23.4 mmol, 1.00 eq) was dissolved in dry THF (10 ml) and added dropwise to the reaction mixture over the course of 3 h. During this time period, additional triphosgene (1.93 g, 6.50 mmol, 0.28 eq) was added to the reaction mixture. Reaction mixture was stirred at 0 °C for 2 h, allowed to warm up to RT and stirred at RT overnight. Organic solvents were removed with a rotary evaporator and product was dried under vacuum. The intermediate product was obtained as a colorless oil, which was dissolved in THF (10 ml). Nα-(tert-Butoxycarbonyl)-L-lysine (6.90 g, 28.0 mmol, 1.20 eq) was dissolved in NaOH (1 M)/THF (2:1 (v/v), 120 ml) and cooled on ice to 0 °C. The intermediate product was added dropwise to the reaction mixture over the course of 1 h. Reaction mixture was allowed to warm up to RT and stirred at RT overnight. Additional NaOH was added to the reaction mixture to maintain a basic milieu. Afterwards, reaction mixture was taken up in ethyl acetate (100 ml) and the aqueous phase was acidified to pH = 2 with conc. HCl. Phases were separated, the aqueous phase was washed with ethyl acetate (3 x 20 ml), organic phases were unified, washed with brine (saturated, 100 ml) and dried with sodium sulfate. Organic solvents were removed with a rotary evaporator and raw product was dried under vacuum. Afterwards, the raw product was purified by column chromatography (silica gel, DCM/MeOH/AcOH (96:3:1)) and compound 9 was obtained as a yellow oil (7.30 g, 79%) in an endo/exo ratio of 3:2.

R_f value: 0.64 (DCM/MeOH/AcOH 87:10:3 (v/v))

^{1}H-NMR (500.2 MHz, DMSO-d_6): δ = 7.11 (t, $^{3}J_{HH}$ = 5.0 Hz, 0.4 H, NH), 7.06 (t, $^{3}J_{HH}$ = 5.0 Hz, 0.6 H, NH*), 6.16 (dd, $^{3}J_{HH}$ = 5.0 Hz, 0.6 H, C=CH*), 6.11-6.08 (m, 0.8 H, HC=CH), 5.93 (dd, $^{3}J_{HH}$ = 5.0 Hz, 0.6 H, HC=C*), 4.02-3.98 (m, 0.4 H, CH_2), 3.86-3.79 (m, 1.4 H, CH_2, α-CH^{lys}), 3.68-3.65 (m, 0.6 H, CH_2*), 3.50-3.46 (m, 0.6 H, CH_2*), 2.94-2.92 (m, 2H, ε-CH_2^{lys}), 2.81-2.78 (m, 1.6 H, CH, CH*), 2.65 (m, 0.4 H, CH), 2.33-2.27 (m, 0.6H, CH*), 1.91 (s, 9H, CH_3), 1.80-1.75 (m, 0.6 H, CH_2*), 1.64-1.50 (m, 2.4 H, CH_2, β-CH_2^{lys}), 1.37-1.34 (m, 5H, CH_2 CH_2*, γ-CH_2^{lys}, δ-CH_2^{lys}), 1.22-1.24 (m, 1H, CH_2, CH_2*), 1.15-1.14 (m, 0.8H, CH_2), 0.49-0.45 (m, 0.6H, CH_2*). Endo signals are marked with *.

^{13}C{^{1}H}-NMR (125.78 MHz, DMSO-d_6): δ = 174.28 (COON), 172.04 (COO), 155.62 (COO), 137.26 (C=C*), 136.77 (C=C), 136.44 (C=C), 132.22 (C=C*), 67.54 (CH_2), 66.97 (CH_2*), 53.47 (α-CH^{lys}), 48.94 (CH_2), 44.62 (CH_2*), 43.42 (CH*), 43.18 (CH), 41.69 (CH), 41.07 (CH*), 40.11 (ε-CH_2^{lys}), 38.15 (CH), 37.88 (CH*), 30.39 (β-CH_2^{lys}), 29.17 (δ-CH_2^{lys}), 29.01 (CH_2), 28.65 (CH_2*), 22.92 (γ-CH_2^{lys}), 21.08 (CH_3). Endo signals are marked with *.

6.8.4.2 Synthesis of compound 2



9

2

$C_{15}H_{24}N_2O_4$

296,36

Figure 155: Reaction conditions for synthesis of compound 2.

Synthesis of compound 2 was done according to literature[201].
Compound 9 (7.30 g, 18.4 mmol) was dissolved in chloroform/formic acid (2:3 (v/v), 25 ml) and stirred for 48. DMF (25 ml) was added and organic solvents were removed with a rotary evaporator and product was dried under vacuum. Afterwards, the raw product was taken up in HCl (50 mM, 20 ml), frozen in liquid nitrogen in lyophilized. HCl salt of compound 2 was obtained as a beige powder (4.11 g, 75%) in an endo/exo ratio of 3:2.

R_f value: 0.02 (DCM/MeOH/AcOH 87:10:3 (v/v))

MS (ESI): m/z = 297.21 $[M+H]^+$

^{1}H-NMR (500.2 MHz, DMSO-d_6): δ = 7.12 (t, $^3J_{HH}$ = 5.0 Hz, 0.4 H, NH), 7.07 (t, $^3J_{HH}$ = 5.0 Hz, 0.6 H, NH*), 6.18 (dd, $^3J_{HH}$ = 5.0 Hz, 0.6 H, C=CH*), 6.12-6.08 (m, 0.8 H, HC=CH), 5.94 (dd, $^3J_{HH}$ = 5.0 Hz, 0.6 H, HC=C*), 4.03-3.99 (m, 0.4 H, CH_2), 3.87-3.83 (m, 0.4 H, CH_2), 3.69-3.66 (m, 0.6 H, CH_2*), 3.51-3.47 (m, 0.6 H, CH_2*), 3.41 (t, $^3J_{HH}$ = 5.0 Hz, 1 H, α-CH^{lys}), 2.96-2.94 (m, 2H, ε-CH_2^{lys}), 2.82-2.78 (m, 1.6 H, CH, CH*), 2.66 (m, 0.4 H, CH), 2.34-2.27 (m, 0.6H, CH*), 1.81-1.71 (m, 1.6 H, CH_2*, β-CH_2^{lys}), 1.67-1.58 (m, 1.4 H, CH_2, β-CH_2^{lys}), 1.38-1.32 (m, 5H, CH_2 CH_2*, γ-CH_2^{lys}, δ-CH_2^{lys}), 1.23-1.25 (m, 1H, CH_2, CH_2*), 1.16-1.15 (m, 0.8H, CH_2), 0.50-0.46 (m, 0.6H, CH_2*). Endo signals are marked with *.

$^{13}C\{^{1}H\}$-NMR (125.78 MHz, DMSO-d_6): δ = 170.53 (COO), 156.35 (COO), 137.29 (C=C*), 136.79 (C=C), 136.21 (C=C), 132.22 (C=C*), 67.57 (CH_2), 67.00 (CH_2*), 53.24 (α-CH^{lys}), 48.95 (CH_2), 44.62 (CH_2*), 43.43 (CH*), 43.18 (CH), 41.70 (CH), 41.07 (CH*), 40.11 (ε-CH_2^{lys}), 38.15 (CH), 37.88 (CH*), 30.34 (β-CH_2^{lys}), 29.10 (δ-CH_2^{lys}), 28.70 (CH_2), 28.23 (CH_2*), 22.10 (γ-CH_2^{lys}). Endo signals are marked with *.

6.9 NMR spectra of synthesized compounds

For the spectra of compounds 1 and 3 as well as the corresponding precursor compounds see appendix of my Master Thesis[359].

6.9.1 NMR spectra of compound 9

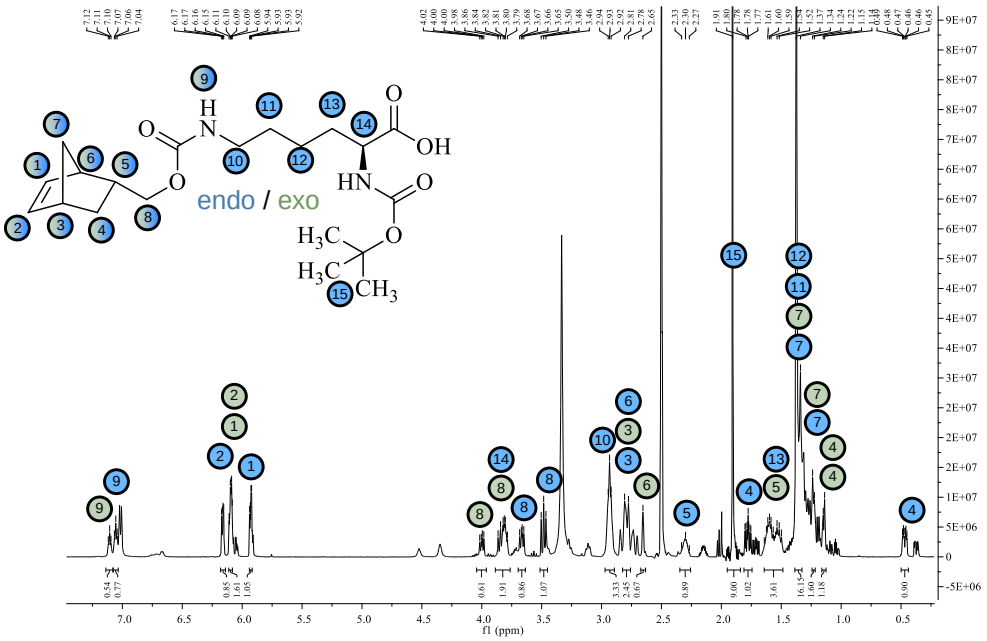


Figure 156: ^{1}H-1D-NMR spectrum of compound 9.

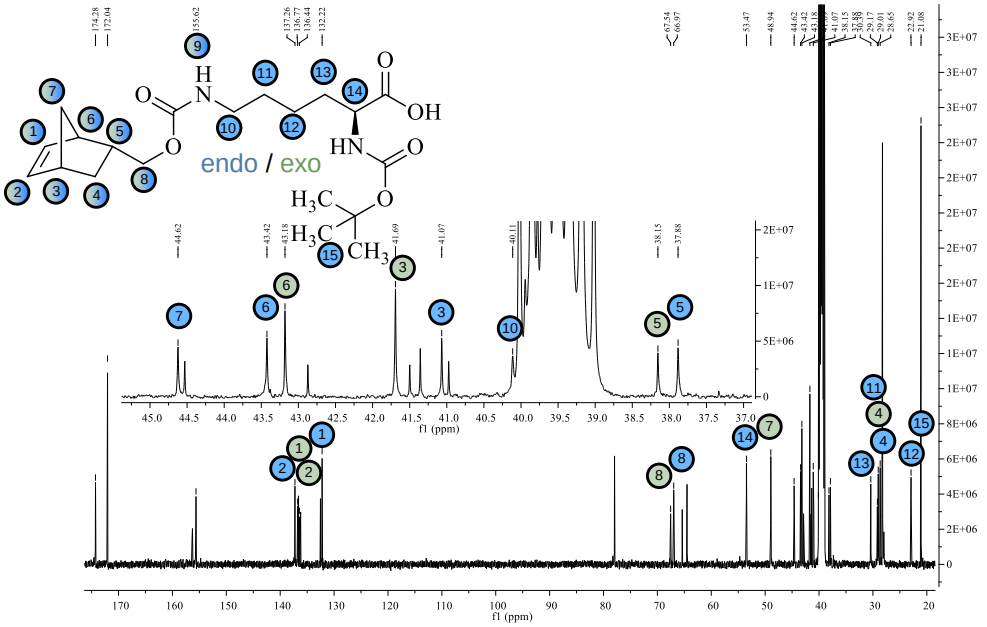


Figure 157: ^{13}C-1D-NMR spectrum of compound 9.

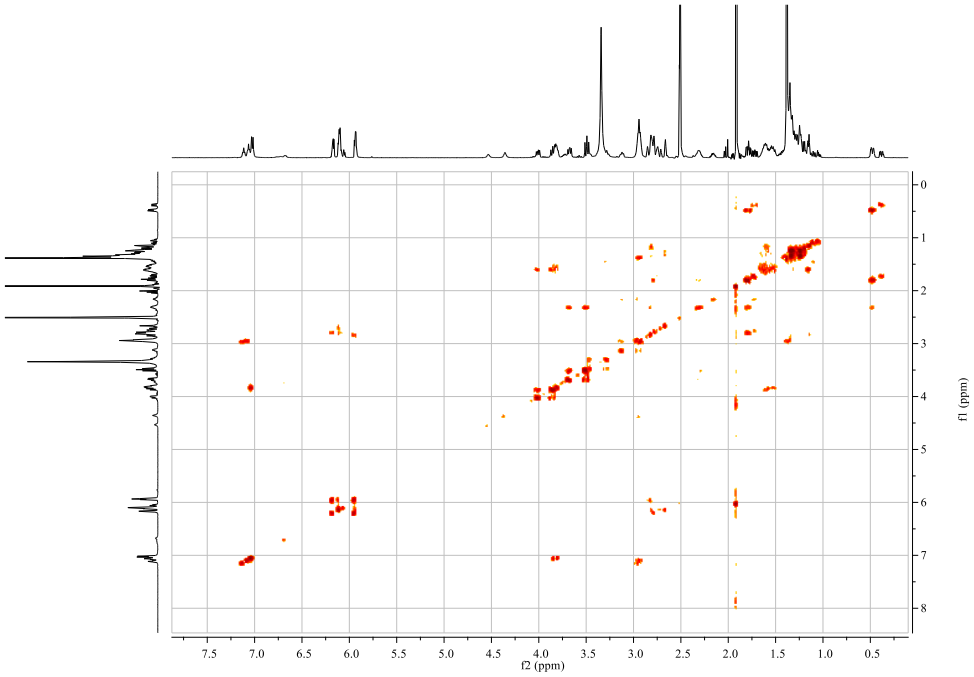


Figure 158: ^{1}H-^{1}H-2D-COSY-NMR spectrum of compound 9.

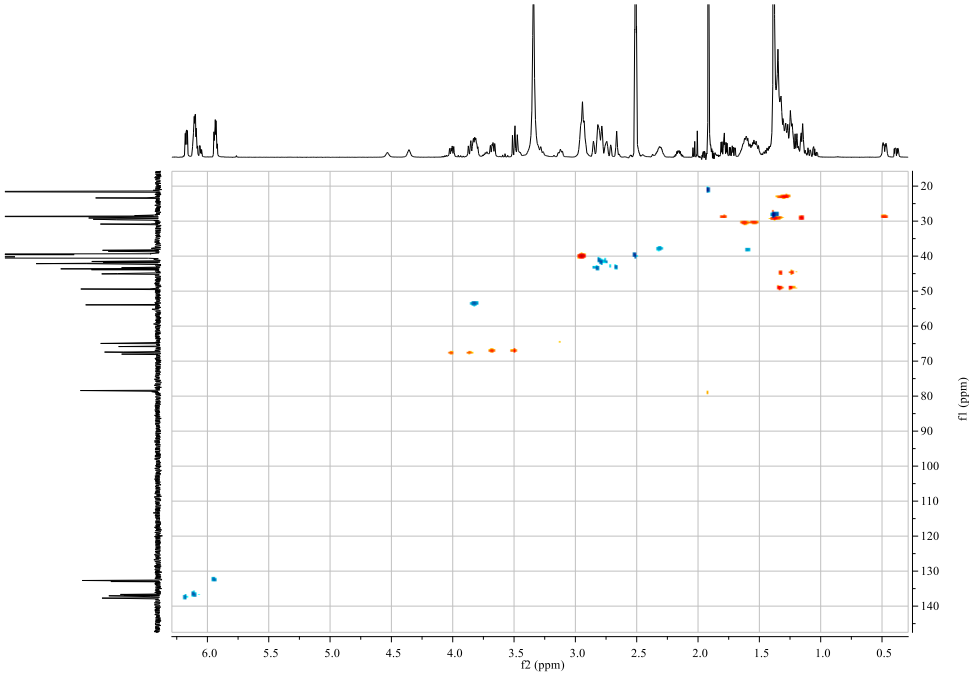


Figure 159: ^{1}H-^{13}C-2D-HSQC-NMR spectrum of compound 9.

6.9.2 NMR spectra of compound 2

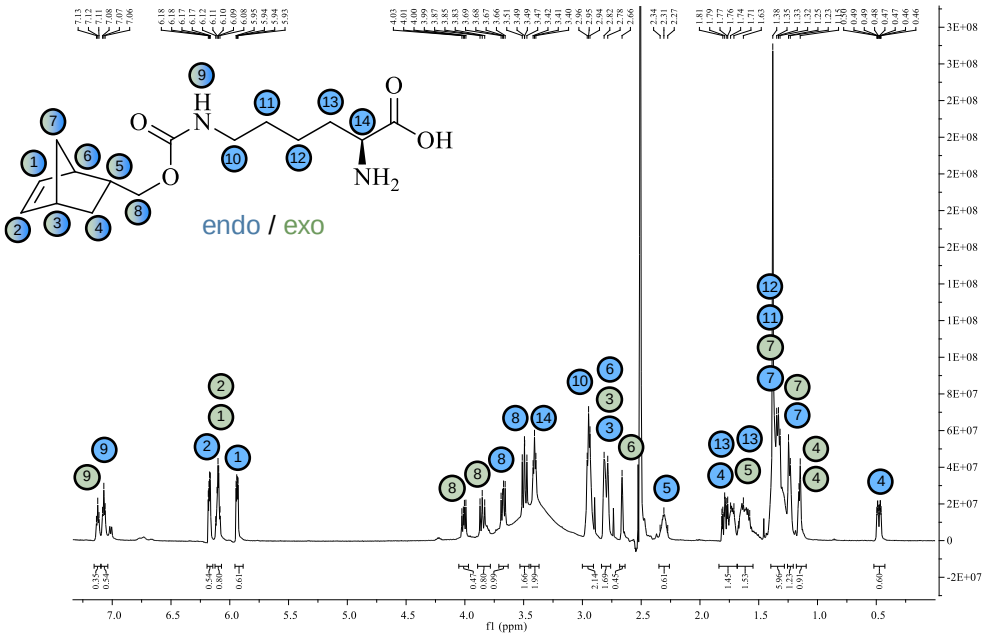


Figure 160: ^{1}H-1D-NMR spectrum of compound 2.

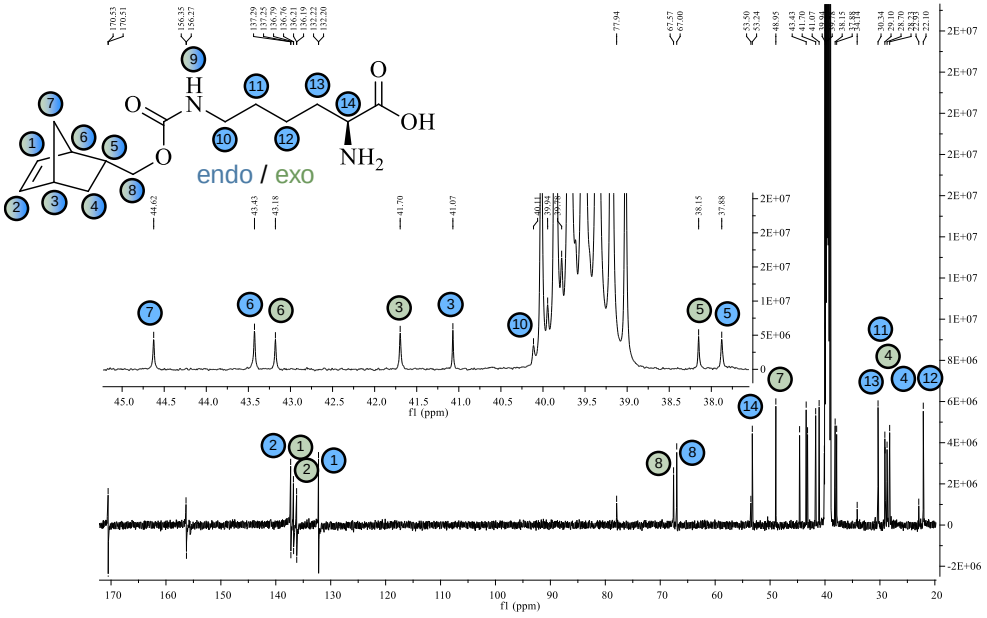


Figure 161: ^{13}C-1D-NMR spectrum of compound 2.

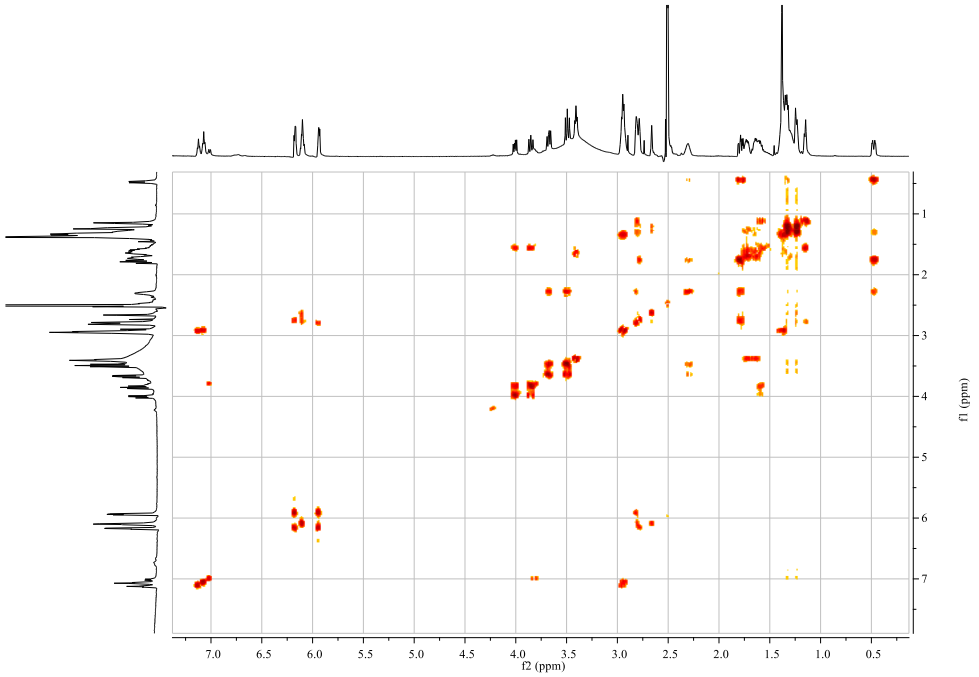


Figure 162: ^{1}H-^{1}H-2D-COSY-NMR spectrum of compound 2.

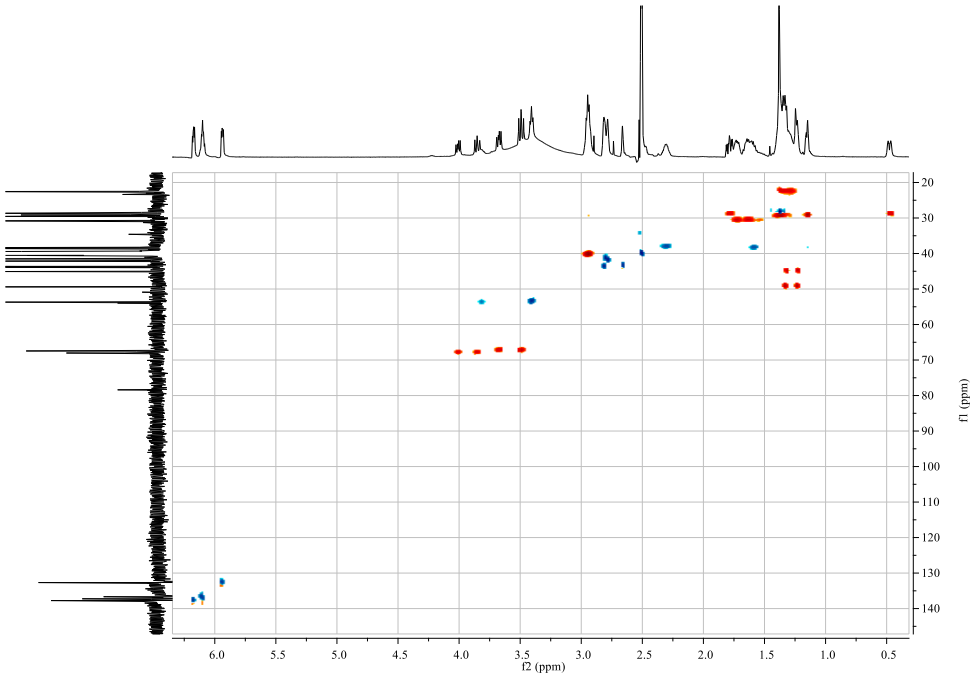


Figure 163: ^{1}H-^{13}C-2D-HSQC-NMR spectrum of compound 2.

Deutsche Zusammenfassung der Arbeit

Die post-transkriptionelle Modifikation der vier Grundbausteine der RNA – Adenosin, Cytidin, Guanosin und Uridin – konnte bereits in allen Domänen des Lebens (Bakterien, Eukaryoten, Archaeen) nachgewiesen werden, und über 150 verschiedene modifizierte Nukleoside wurden bereits sowohl in codierender als auch nicht-codierender RNA identifiziert[1,2]. Die unterschiedlichen Modifikationen werden dabei enzymatisch in den Nukleobasen-Teil oder den Ribose-Teil des Nukleosides eingebaut, wobei es auch zu mehreren Modifikationen an der gleichen Stelle kommen kann und modifizierte Nukleoside im Anschluss auch weiter modifiziert werden können[3,4]. Die Modifikationen haben verschiedene Effekte auf die RNA, wie Regulation der Funktion[30–32,52,53] und RNA-Reifung[51] oder induzieren strukturelle Änderungen, die zur Stabilität der RNA beitragen und die Resistenz gegenüber Degradation erhöhen[14,36–43]. Während einzelne RNA Modifikations-Stellen eher kleinere Einflüsse auf den Organismus haben, so ist die Gesamtheit aller Modifikationen doch überlebenswichtig[3,10–16]. Defekte in Enzymen, die für die Modifikationen verantwortlich sind konnten mit verschiedenen Krankheiten in Verbindung gebracht werden[17,18].

Die häufigste (und auch die erstmals beschriebene) post-transkriptionelle Modifikation ist dabei Pseudouridin „Ψ“, welches auch als „fünftes Nukleotid“ bezeichnet wird[63–65]. Pseudouridin stellt dabei ein Rotationsisomer von Uridin dar, welches durch einen Bindungsbruch zwischen dem Uracil und der Ribose, einer 180° Rotation der Uracil-Base um die N3-C6 Achse sowie einer C-C Bindungsknüpfung entsteht (Abbildung 164a). Pseudouridin erhält dadurch eine zusätzliche Amino-Gruppe, die im Vergleich zu Uridin eine zusätzliche Wasserstoffbrückenbindung eingehen kann[75,77,78], was zu lokalen Strukturänderungen und damit einhergehend zu erhöhter struktureller Stabilität führt[80,81]. Die Pseudouridylierung von Uridin konnte in verschiedenen RNA-Klassen nachgewiesen werden und induziert dabei verschiedene neue Eigenschaften. Während für tRNA verbesserte Interaktion mit der mRNA erreicht wird[79,88], so scheint Ψ sowohl regulatorische Eigenschaften in spliceosomaler snRNA zu haben[69] als auch die Translations-Effizienz in rRNA zu erhöhen[94]. In mRNA wird Pseudouridin zumal mit der Erweiterung des genetischen Codes in Verbindung gebracht[89,90].

Der exakte enzymatische Mechanismus der Pseudouridylierung ist noch nicht zweifelsfrei beschrieben worden[66]. Die Umwandlung von Uridin zu Pseudouridin kann durch zwei unterschiedliche Klassen von Enzymen erfolgen. Pseudouridylasen-Enzyme, die allesamt ein ähnliches Strukturmotiv aufweisen, können die enzymatische Reaktion ohne zusätzliche Nebenfaktoren durchführen und funktionieren dabei meist nur für eine bestimmte oder für mehrere strukturähnliche RNA Modifikationsstellen[70]. H/ACA Ribonucleoproteine (RNPs) hingegen bestehen aus mehreren Proteinen und unterscheiden sich durch ihre „guide“ RNA, die für die Rekrutierung der Substrat RNA verantwortlich ist (Abbildung 164b). Die guide RNA formt dabei eine Haarnadel-Struktur (Hairpin), die eine „Pseudouridylierungs-Tasche“ enthält[138–140]. In diesem Bereich wird das Substrat via Watson-Crick Basenpaarung gebunden und das Ziel-Uridin in Position gebracht. Während archäische H/ACA sRNAs zwischen einer und drei aufeinanderfolgende Hairpins ausbilden können, so bilden eukaryotische H/ACA snoRNAs immer eine zweigeteilte Komplexstruktur mit zwei

aufeinanderfolgenden Hairpins, welche jeweils eine Bindungsstelle für Substrate aufweisen[143]. Auf beide Hairpins folgen die konservierten H- und ACA-Motive, die zusammen mit dem RNA-Stamm eine Bindungsstelle für das Protein Cbf5 bereitstellen[101,149–151]. Cbf5 enthält das katalytische Zentrum für die Pseudouridylierung in Form eines konservierten Aspartates, welches während der enzymatischen Reaktion an das Uridin bindet sowie ein „Thumb Loop“ Strukturelement, das für Substrat-Rektrutierung und -Freisetzung verantwortlich ist[138,140]. Das Protein Gar1, welches an Cbf5 bindet, stellt dabei eine Bindungsstelle für diese Domäne bereit, und ist damit ebenfalls in den Umsatz des Substrates involviert[140,150,157]. Eukaryotisches Gar1 besitzt dabei im Vergleich zu archäischem Gar1 zusätzliche Glycin-Arginin-reiche „GAR“ Domänen, die mit der Verbesserung von RNA-Bindemotiven im Zusammenhang stehen[158,358,363–368]. Das Protein Nop10 bindet an Cbf5 und stellt zusammen mit der guide RNA eine Bindungsstelle für Nhp2 (in Eukaryoten) oder L7Ae (in Archaeen) bereit. In dieser Hinsicht unterscheiden sich archäische und eukaryotische Komplexe stark voneinander. L7Ae ist in die korrekte Orientierung des Substrates involviert und ist für die katalytische Reaktion zwingend notwendig[157]. Zusätzlich ist es auf ein K-turn-Motiv in der guide RNA angewiesen, um korrekt zu binden, ist dadurch aber auch in der Lage, in Abwesenheit der anderen Proteine an die guide RNA zu binden[126,143]. Nhp2 hingegen bindet unspezifisch an RNAs und erst durch die Formation eines trimeren Komplexes mit Nop10 und Cbf5 wird die Spezifität gegenüber einer guide RNA induziert[122,126,143,154–156]. Zusätzlich scheint die Pseudouridylierung auch in Abwesenheit von Nhp2 zu funktionieren[159,160].

Während archäische H/ACA RNPs strukturell bereits besser erforscht sind, so ist die Erkenntnislage auf dem Gebiet der eukaryotischen H/ACA RNPs noch vergleichsweise gering. Die genaue Struktur und Struktur-Funktionsbeziehungen eukaryotischer RNPs sind noch größtenteils unbekannt; während es für archäische RNPs komplette Kristallstrukturen mit sRNA, Proteinen und gebundenem Substrat gibt[140], so existieren für eukaryotische RNPs lediglich eine Teil-Kristallstruktur von Nop10-Cbf5-Gar1[160], eine NMR-Struktur von Nhp2[362] (Abbildung 164c) sowie seit kurzem eine Kryoelektronenmikroskopie-Struktur der nichtkanonischen H/ACA Domäne der Telomerase[70]. Ein besonderes Interesse in der Erforschung von eukaryotischen H/ACA RNPs liegt in deren Ähnlichkeit zu humanen H/ACA RNPs. Protein-Defekte in diesen Komplexen stehen mit einer Anfälligkeit für die Erbkrankheit Dyskeratosis congenita im Zusammenhang, die Knochenmarkversagen, verfrühte Alterung und erhöhtes Krebsrisiko zur Folge hat[147,161,170–174].

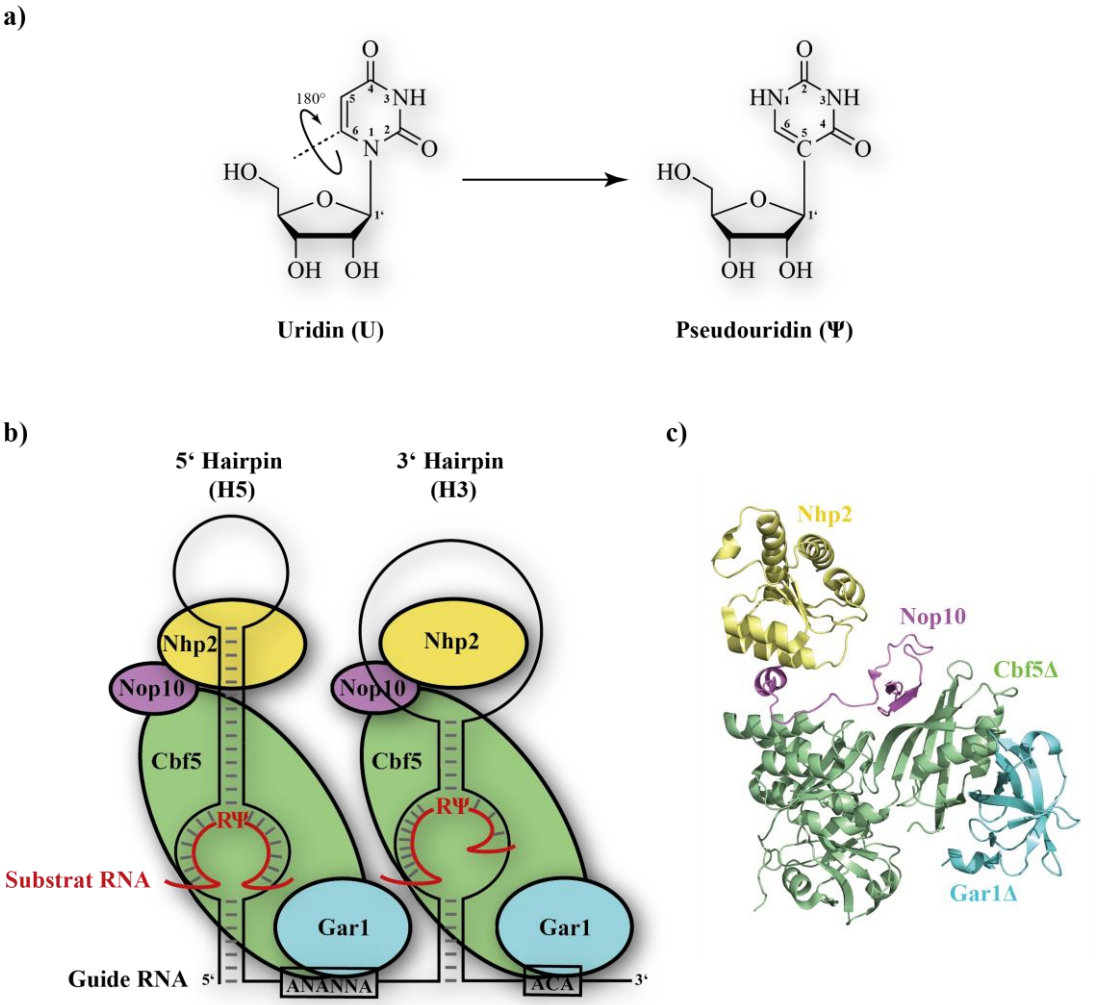


Abbildung 164: a) Schematische Darstellung der Umwandlung von Uridin zu Pseudouridin. b) Schematische Darstellung des snR81 H/ACA snoRNPs. c) Kristallstruktur von Nop10-Cbf5Δ-Gar1Δ (pdb: 3U28) und NMR-Struktur von Nhp2 (pdb: 2LBW).

Das Hauptziel dieser Arbeit lag in der Erforschung des snR81 H/ACA snoRNPs aus *Saccharomyces cerevisiae*, einem eukaryotischen snoRNP, der für die U zu Ψ Umwandlung von U42 in der U2 snRNA und U1051 der 25S rRNA zuständig ist[96,356]. Es sollten neue Erkenntnisse sowohl in Bezug auf die Struktur des Komplexes als auch auf die Dynamik während der enzymatischen Reaktionen gewonnen werden. Dabei bestand der erste Teil der Arbeit in einer *in vitro* Präparation des Komplexes sowie einer Untersuchung dessen katalytischer Aktivität. Spezieller Fokus sollte dabei den Proteinen Nhp2 und Gar1 gewidmet werden, die in eukaryotischen H/ACA RNPs eine Art „Sonderstellung“ im Vergleich zu archäischen H/ACA RNPs einnehmen: Nhp2 durch seine veränderten Eigenschaften im Vergleich zu L7Ae, obwohl beide Proteine eine starke strukturelle Ähnlichkeit aufweisen, und Gar1 durch seine zusätzlichen RNA-Bindedomänen. Im zweiten Teil der Arbeit sollten dann nicht-natürliche Aminosäuren (nnAs) synthetisiert und mittels Manipulation des genetischen Codes in verschiedene Proteine eingebaut werden, was das Anbringen von spektroskopischen Sonden wie Fluoreszenzfarbstoffen ermöglicht. Markierte Komplexe sollten dann zur

Untersuchung der Interaktion zwischen den Komplex-Partnern in Bezug auf Struktur und Dynamiken mittels Einzelmolekül FRET-Spektroskopie untersucht werden.

Für die *in vitro* Präparation des RNP Komplexes wurde die snoRNA via Transkription hergestellt. Zusätzlich wurden Einzel-Konstrukte beider Hairpins (H5 und H3) hergestellt, um den Einfluss des bipartiten Komplexes im Vergleich zu den einzelnen Hairpins zu untersuchen. Außerdem wurden von beiden Hairpins verkürzte Konstrukte hergestellt, bei denen die obere Stamm-Loop Struktur der RNA durch einen stabilen Tetraloop oberhalb der Pseudouridylierungs-Tasche ersetzt wurde, um den Einfluss dieses Strukturelements auf die Aktivität und auf die Nhp2-Bindeeigenschaften zu testen.
Die *Saccharomyces cerevisiae* Proteine Nhp2, Nop10, Cbf5 und Gar1 konnten in guten Ausbeuten und guter Reinheit mithilfe von rekombinanter (Co-)Expression aus *E. coli* hergestellt werden. Dabei war es auch möglich, endogene *E. coli* RNA, die vor allem an Nhp2 gebunden war, quantitativ zu entfernen. Zusätzlich wurde – anders als in der *in vivo* Komplex-Assemblierung – eine Co-Expression und Co-Aufreinigung des trimeren Nop10-Cbf5-Gar1 Komplexes durchgeführt. Durch eine Rekonstitution mit Nhp2 und snoRNA konnte dann ein enzymatisch aktiver H/ACA RNP Komplex *in vitro* gewonnen werden.
Für die Präparation der Fluorophor-markierten Proteine wurden mehrere nicht-natürliche Aminosäuren synthetisiert. Mithilfe von Amber Codon Suppression war es möglich, diese nnAs an verschiedenen Positionen in die Proteine Nhp2 und Cbf5 einzubauen. Diese modifizierten Proteine konnten dann verwendet werden, um mittels „Click“ Chemie sowohl Fluorophore als auch Nitroxidlabel an den Proteinen anzubringen und die vollständigen RNP Komplexe in aktiver Form mit den gelabelten Komponenten zu rekonstituieren.

Für die generierten RNP Konstrukte wurden sowohl single als auch multiple Turnover Reaktionen mit radioaktiv markiertem Substrat durchgeführt. Unter single Turnover Bedingungen konnte festgestellt werden, dass die 5‘ Hairpin (H5) des Komplexes höhere Aktivitäten erreichte als die 3‘ Hairpin (H3). Zusätzlich konnte gezeigt werden, dass auf ihre Hauptdomänen verkürzte Konstrukte von Gar1 und Cbf5 (Gar1Δ und Cbf5Δ) zu einer Aktivitätsverringerung unter single Turnover Bedingungen führten und auch die Abwesenheit von Nhp2 die Aktivität senkte, was in ähnlicher Weise für die snoRNPs mit den snoRNAs snR5 und snR34 beobachtet werden konnte[159,160]. In multiple Turnover Assays wurde der zeitliche Verlauf der enzymatischen Pseudouridylierung der verschiedenen „Sub-Komplexe“ von snR81 dann genauer untersucht. Während das Volllängen-Konstrukt von snR81 in Gegenwart von 5‘ Substrat (FL5) und H5 eine ähnlich hohe Aktivität aufzeigten, so war im Vergleich die Aktivität des Volllängen-Konstrukts in Gegenwart von 3‘ Substrat (FL3) verringert. Des Weiteren war die Aktivität von H3 im Gegenzug zu FL3 stark eingeschränkt. Dies zeigt, dass sowohl die 5‘ Hairpin einen stärkeren Substrat-Umsatz als die 3‘ Hairpin hat und dass zusätzlich die 3‘ Hairpin auf die Anwesenheit der 5‘ Hairpin angewiesen ist, während die 5‘ Hairpin auch ohne die 3‘ Hairpin starke enzymatische Aktivität aufweist. Dieses unterschiedlichen Verhalten der beiden Hairpins wird dabei vermutlich von ihren verschiedenen Strukturelementen beeinflusst: Während bei der 5‘ Hairpin die RNA-Struktur oberhalb der Pseudouridylierungs-Tasche aus einem größeren doppelsträngigen Stamm und einem kleinen Loop mit wenigen ungepaarten Nukleotiden besteht, so weist die 3‘ Hairpin

einen eher kleinen Doppelstrang und einen relativ großen Loop mit einer Vielzahl ungepaarter Nukleotide auf. Zusätzlich ist die Substrat-Bindung bei der 5‘ Hairpin relativ symmetrisch, während bei der 3‘ Hairpin die Substrat-Bindung auf eine Seite der Pseudouridylierungs-Tasche verlagert ist. Interessanterweise hat die Entfernung der oberen Stamm-Loop Struktur ebenfalls unterschiedliche Effekte bei der 5‘ Hairpin (H5Δ) als bei der 3‘ Hairpin (H3Δ). Während im Konstrukt H5Δ die Aktivität nur unmerklich beeinflusst wird, so wird H3Δ enzymatisch komplett inaktiv. Die obere Stamm-Loop Struktur verringert also die Aktivität von H3 im Vergleich zu H5, ist aber gleichzeitig in H3 auch zwangsläufig für Komplexaktivität nötig, da sie scheinbar für korrekte RNA-Faltung und/oder die Komplex-Assemblierung unabdingbar ist.

Für Gar1 konnte mittels EMSA-Bindungstest gezeigt werden, dass die Glycin-Arginin reichen GAR Domänen am C- und N-Terminus des Proteins die RNA-Bindeeigenschaften des Proteins beeinflussten. Dies zeigt, dass in eukaryotischen H/ACA RNPs (im Gegensatz zu archäischen RNPs) Gar1 unter Umständen nicht nur an Cbf5 bindet, sondern auch einen direkten Einfluss auf snoRNA oder Substrat-RNA nehmen kann. Während in single Turnover Aktivitätstests das Hauptdomänen-Konstrukt Gar1Δ, bei dem sämtliche GAR Domänen entfernt wurden, die enzymatische Aktivität für beide Hairpins verringerte, so hatte es in multiple Turnover Aktivitätstests eine aktivitätsverringernde Wirkung auf die 5‘ Hairpin, wirkte jedoch aktivitätssteigernd auf die 3‘ Hairpin. Dies verstärkt den Verdacht, dass die GAR Domänen durch RNA-bindende Eigenschaften einen Einfluss auf die Komplex Assemblierung oder die Substratplatzierung bzw. Produktfreisetzung haben, was zwar die Komplexaktivität erhöhen kann, unter bestimmten Umständen (wie im Falle der snR81 3‘ Hairpin) aber auch den multiplen Substratumsatz stark verlangsamen kann – möglicherweise ausgelöst durch unspezifische RNA-Bindung.

Die Rekonstitution der Konstrukte ohne Nhp2 hatte in multiple Turnover Reaktionen einen negativen Effekt auf die Aktivität der 5‘ Hairpin. Sowohl für FL5 als auch für H5 war vor allem die Startgeschwindigkeit der Pseudouridylierung stark reduziert, trotzdem erreichte die Ausbeute an Pseudouridin nach 22 Stunden einen relativ hohen Wert. Für FL3 und H3 hingegen wurde der Komplex unter multiple Turnover Bedingungen komplett inaktiv – lediglich für FL3 konnte nach 22 Stunden eine kleine Menge Ψ nachgewiesen werden. Dies zeigt, dass Nhp2 eine unterstützende Wirkung auf die Ψ-Generierung in der 5‘ Hairpin hat, möglicherweise durch Stabilisierung der Komplex-Struktur, und dass die Pseudouridylierung ohne Nhp2 verlangsamt abläuft. Für die 3‘ Hairpin ist Nhp2 jedoch zwingend notwendig für enzymatische Aktivität, was entweder darauf hindeutet, dass Nhp2 zusätzliche sekundäre Funktionen ausführt, oder dass die Stabilisierungs-Eigenschaften von Nhp2 dort für die obere Stamm-Loop Struktur zwingend notwendig sind.

Einzelmolekül FRET spektroskopischen Untersuchungen wurden mit Donor-Fluorophor gelabeltem Nhp2 und Akzeptor-Fluorophor gelabelter snoRNA vorgenommen. Nhp2 bindet dabei über eine Bindestelle mit Nop10 und der snoRNA an den Komplex. In Abwesenheit von Nop10 konnte keine Bindung von Nhp2 an den Komplex nachgewiesen werden. Für Hairpin H5 zeigte Nhp2 einen einzelnen Bindungsmodus, der eine spezifische Bindung in

Anwesenheit beider Komplex-Bindungsstellen nahelegt. Die Entfernung der GAR-Domänen an Gar1 führte zu schärferen FRET-States, was auf einen starreren Komplex schließen lässt, und die GAR-Domänen in H5 Dynamik zu induzieren scheinen. Mit der Verkürzung der oberen Stamm-Loop Struktur (H5Δ) hingegen konnte interessanterweise immer noch Nhp2-Bindung nachgewiesen werden, Nhp2 zeigte nun aber Bindung in drei verschiedenen Bindungsmodi. Die Anwesenheit der RNA-Bindestelle ist somit nicht notwendig für Nhp2-Bindung an den Komplex. Da kein Übergang zwischen den verschiedenen Bindungsmodi während zeitaufgelöster FRET Spektroskopie beobachtet werden konnte, ist davon auszugehen, dass Nhp2 nach Komplex-Assemblierung in einem Bindungsmodus festsitzt, oder dass der Übergang zwischen verschiedenen Bindungsmodi zumindest extrem langsam abläuft. Da Konstrukt H5Δ eine ähnliche Aktivität wie H5 und FL5 zeigte, und gleichzeitig die Entfernung von Nhp2 für dieses Konstrukt nicht so stark aktivitätsverringernd war wie für H5 und FL5, ist davon auszugehen, dass Nhp2 tatsächlich für die Stabilisierung der oberen RNA-Struktur zuständig ist – eine Funktion die in Abwesenheit dieses Strukturelementes nicht mehr benötigt wird, was zur unspezifischeren Bindung von Nhp2 führt. Gleichzeitig scheint Nhp2 aber auch noch zusätzliche Funktionen zu haben, da die Aktivität für H5Δ trotz Abwesenheit der oberen Strukturelemente immer noch reduziert war in Abwesenheit von Nhp2, und Nhp2 trotz der unterschiedlichen Bindungsmodi einen katalysierenden Effekt auf H5Δ hatte.

Für die 3‘ Hairpin zeigte die Einzelmolekül FRET Analyse, dass Nhp2 sowohl für Konstrukt H3 als auch für das verkürzte H3Δ Konstrukt in drei verschiedenen Bindungsmodi an den Komplex bindet, wobei sich die gemessenen FRET-Effizienzen der drei Populationen zwar von H5Δ unterscheiden, zwischen den beiden 3‘ Hairpin Konstrukten jedoch gleich sind. Die unspezifischere Bindung von Nhp2 im Vergleich zur 5‘ Hairpin liefert eine Erklärung für die reduzierte Aktivität der 3‘ Hairpin verglichen mit der 5‘ Hairpin. Während der „high FRET“ State einen „korrekten“ Bindungsmodus darzustellen scheint, so stellen die anderen beiden Populationen vermutlich einen katalytisch nicht aktiven Komplex dar. Durch die Entfernung der GAR Domänen wurde diese Vermutung verstärkt, da mit der damit verbundenen Aktivitätserhöhung auch die Population des high FRET States zunahm – Gar1 scheint also indirekt eine Auswirkung auf die korrekte Bindung von Nhp2 zu haben. Während für H3Δ zwar auch drei Bindungsmodi von Nhp2 nachgewiesen werden konnten, so war die Populationsverteilung zwischen den drei FRET States doch eher willkürlich, was zu der beobachteten Inaktivität dieses Konstruktes passt.

Da in Anwesenheit der Substrat RNA keine starken Veränderungen der FRET States oder der Populationsverteilung beobachtet werden konnte ist davon auszugehen, dass die Bindung von Nhp2 unabhängig vom Vorhandensein der Substrat RNA ist, und die katalysierende Wirkung von Nhp2 eher durch Komplex-Stabilisierung begründet ist und nicht durch den direkten Eingriff von Nhp2 in die Uridin zu Pseudouridin Umwandlung.

Durch die Donor-Fluoreszenzmarkierung der Thumb Loop Domäne von Cbf5 konnte ein einzelner FRET State in Verbindung mit Akzeptor-markierter snoRNA für beide Hairpins gezeigt werden. Es waren jedoch keine zusätzlichen Populationen oder Übergänge zwischen

verschiedenen States zu beobachten, wodurch ein Echtzeit-Betrachtung der Dynamik der Thumb Loop Domäne zwischen geöffnetem und geschlossenem Zustand nicht möglich war. Dies kann zwar auf eine Inaktivität des an die Glasoberfläche assemblierten Komplexes während der FRET Messung hindeuten, kann aber auch ein Indiz dafür sein, dass der geschwindigkeitsbestimmende Schritt bei der Pseudouridylierung die Substrat Bindung und Produkt Freisetzung ist, und die eigentliche katalytische Reaktion im sub-Millisekunden Bereich und damit außerhalb des Beobachtungsrahmens der Einzelmolekül FRET Spektroskopie liegt.

Sowohl FRET Messungen als auch PELDOR Abstandmessungen zwischen den beiden Komplex-Hairpins zeigten, dass der Abstand zwischen beiden Hairpins augenscheinlich mehr als 10 nm beträgt, und damit außerhalb des Beobachtungsrahmens beider Messmethoden liegt. Dies zeigt jedoch, dass auf Basis bekannter Kristallstrukturen und der Annahme, dass die Protein-Anordnung im Komplex zumindest ähnlich zu archäischen H/ACA Komplexen verläuft, die beiden Hairpins des Komplexes vermutlich eine gewinkelte Struktur zueinander annehmen.

Durch die Untersuchung des snR81 snoRNPs aus *Saccharomyces cerevisiae* war es in dieser Arbeit möglich, neue Einblicke in die Struktur und Dynamiken von eukaryotischen H/ACA snoRNPs zu gewinnen (Abbildung 165). Dabei wurde vor allem die relativ unerforschte Rolle des Proteins Nhp2 sowie die Funktion der eukaryotisch spezifischen Gar1 Domänen während der Komplex-Assemblierung und Katalyse in Bezug auf die bipartite Komplex-Struktur untersucht. Durch die Entwicklung eines Arbeitsablaufes zur rekombinanten Expression der verschiedenen Proteine und *in vitro* Komplex-Assemblierung sowie die Möglichkeit, unterschiedliche Komplex-Teile ortsspezifisch zu markieren, kann die Untersuchung des snR81 snoRNPs auf weitere spektroskopische Methoden ausgeweitet werden – oder auch die Proteine in Kombination mit einer anderen snoRNA untersucht werden. Vor allem eine Untersuchung des snR81 Volllängen-Komplexes sowohl mittels smFRET Spektroskopie als auch mittels Kryoelektronenmikroskopie stellen dabei die nächsten Herausforderungen zu einem besseren Verständnis von eukaryotischen H/ACA snoRNPs dar.

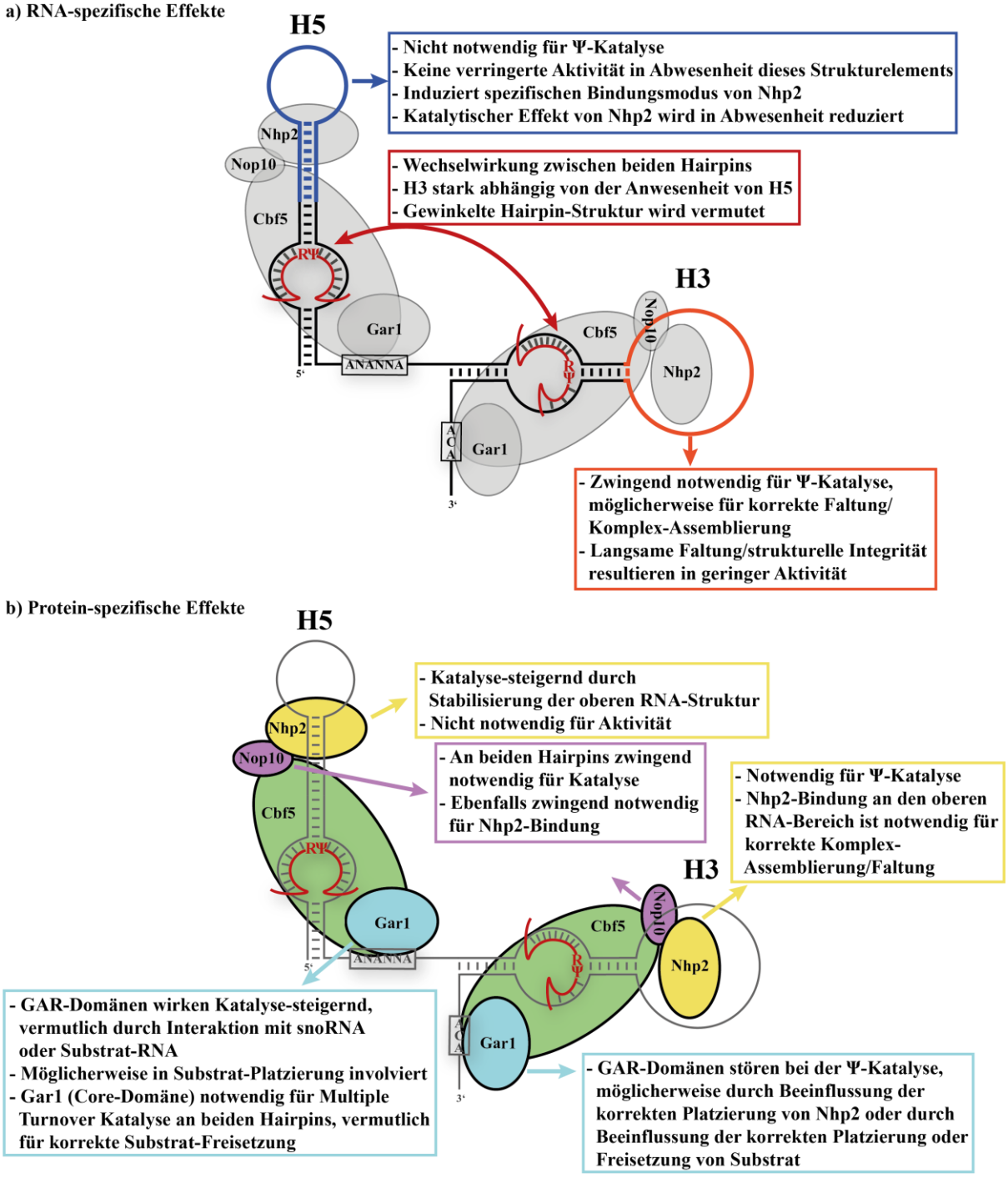


Abbildung 165: Zusammenfassung der beobachteten Effekte von snR81 H/ACA snoRNP Komplex-Komponenten auf die Komplex-Assemblierung und die katalytische Aktivität. a) snoRNA-spezifische Effekte inklusive der Rolle der oberen Stamm-Loop Struktur. b) Protein-spezifische Effekte von Nop10, Nhp2 und den GAR-Domänen von Gar1.

List of references

[1] Cantara, W. A.; Crain, P. F.; Rozenski, J.; McCloskey, J. A.; Harris, K. A.; Zhang, X.; Vendeix, F. A. P.; Fabris, D.; Agris, P. F. The RNA Modification Database, RNAMDB: 2011 update. *Nucleic acids research* **2011**, *39*, D195-201. DOI: 10.1093/nar/gkq1028.

[2] Boccaletto, P.; Machnicka, M. A.; Purta, E.; Piatkowski, P.; Baginski, B.; Wirecki, T. K.; Crécy-Lagard, V. de; Ross, R.; Limbach, P. A.; Kotter, A.*; et al.* MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic acids research* **2018**, *46*, D303-D307. DOI: 10.1093/nar/gkx1030.

[3] Helm, M.; Alfonzo, J. D. Posttranscriptional RNA Modifications: playing metabolic games in a cell's chemical Legoland. *Chemistry & biology* **2014**, *21*, 174–185. DOI: 10.1016/j.chembiol.2013.10.015.

[4] Motorin, Y.; Helm, M. RNA nucleotide methylation. *Wiley interdisciplinary reviews. RNA* **2011**, *2*, 611–631. DOI: 10.1002/wrna.79.

[5] Machnicka, M. A.; Milanowska, K.; Osman Oglou, O.; Purta, E.; Kurkowska, M.; Olchowik, A.; Januszewski, W.; Kalinowski, S.; Dunin-Horkawicz, S.; Rother, K. M.*; et al.* MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic acids research* **2013**, *41*, D262-7. DOI: 10.1093/nar/gks1007.

[6] Motorin, Y.; Helm, M. tRNA stabilization by modified nucleotides. *Biochemistry* **2010**, *49*, 4934–4944. DOI: 10.1021/bi100408z.

[7] Bujnicki, J. M.; Droogmans, L.; Grosjean, H.; Purushothaman, S. K.; Lapeyre, B. *Practical bioinformatics: Bioinformatics-guided identification and experimental characterization of novel RNA methyltransferases*; Springer: Berlin, Heidelberg, New York, Hong Kong, London, Milan, Paris, Tokyo, 2004.

[8] Czerwoniec, A.; Kasprzak, J. M.; Kaminska, K. H.; Rother, K.; Purta, E.; Bujnicki, J. M. *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution: Folds and Functions of Domains in RNA Modification Enzymes*; Landes Bioscience: Austin, 2009.

[9] Lorenz, C.; Lünse, C. E.; Mörl, M. tRNA Modifications: Impact on Structure and Thermal Adaptation. *Biomolecules* **2017**, *7*. DOI: 10.3390/biom7020035.

[10] Björk, G. R. Genetic dissection of synthesis and function of modified nucleosides in bacterial transfer RNA. *Progress in nucleic acid research and molecular biology* **1995**, *50*, 263–338. DOI: 10.1016/s0079-6603(08)60817-x.

[11] Alfonzo, J. D. Post-transcriptional RNA modification methods. *Methods (San Diego, Calif.)* **2016**, *107*, 1–2. DOI: 10.1016/j.ymeth.2016.08.007.

[12] Alexandrov, A.; Chernyakov, I.; Gu, W.; Hiley, S. L.; Hughes, T. R.; Grayhack, E. J.; Phizicky, E. M. Rapid tRNA decay can result from lack of nonessential modifications. *Molecular cell* **2006**, *21*, 87–96. DOI: 10.1016/j.molcel.2005.10.036.

[13] Chernyakov, I.; Whipple, J. M.; Kotelawala, L.; Grayhack, E. J.; Phizicky, E. M. Degradation of several hypomodified mature tRNA species in Saccharomyces cerevisiae is mediated by Met22 and the 5'-3' exonucleases Rat1 and Xrn1. *Genes & development* **2008**, *22*, 1369–1380. DOI: 10.1101/gad.1654308.

[14] Helm, M.; Brulé, H.; Degoul, F.; Cepanec, C.; Leroux, J. P.; Giegé, R.; Florentz, C. The presence of modified nucleotides is required for cloverleaf folding of a human

mitochondrial tRNA. *Nucleic acids research* **1998**, *26*, 1636–1643. DOI: 10.1093/nar/26.7.1636.

[15] Isel, C.; Lanchy, J. M.; Le Grice, S. F.; Ehresmann, C.; Ehresmann, B.; Marquet, R. Specific initiation and switch to elongation of human immunodeficiency virus type 1 reverse transcription require the post-transcriptional modifications of primer tRNA3Lys. *The EMBO journal* **1996**, *15*, 917–924.

[16] Muramatsu, T.; Nishikawa, K.; Nemoto, F.; Kuchino, Y.; Nishimura, S.; Miyazawa, T.; Yokoyama, S. Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature* **1988**, *336*, 179–181. DOI: 10.1038/336179a0.

[17] Corbett, A. H. Post-transcriptional regulation of gene expression and human disease. *Current opinion in cell biology* **2018**, *52*, 96–104. DOI: 10.1016/j.ceb.2018.02.011.

[18] Tong, J.; Flavell, R. A.; Li, H.-B. RNA m6A modification and its function in diseases. *Frontiers of medicine* **2018**, *12*, 481–489. DOI: 10.1007/s11684-018-0654-8.

[19] Ross, R.; Cao, X.; Yu, N.; Limbach, P. A. Sequence mapping of transfer RNA chemical modifications by liquid chromatography tandem mass spectrometry. *Methods (San Diego, Calif.)* **2016**, *107*, 73–78. DOI: 10.1016/j.ymeth.2016.03.016.

[20] Thüring, K.; Schmid, K.; Keller, P.; Helm, M. Analysis of RNA modifications by liquid chromatography-tandem mass spectrometry. *Methods (San Diego, Calif.)* **2016**, *107*, 48–56. DOI: 10.1016/j.ymeth.2016.03.019.

[21] Tserovski, L.; Marchand, V.; Hauenschild, R.; Blanloeil-Oillo, F.; Helm, M.; Motorin, Y. High-throughput sequencing for 1-methyladenosine (m(1)A) mapping in RNA. *Methods (San Diego, Calif.)* **2016**, *107*, 110–121. DOI: 10.1016/j.ymeth.2016.02.012.

[22] Motorin, Y.; Helm, M. Methods for RNA Modification Mapping Using Deep Sequencing: Established and New Emerging Technologies. *Genes* **2019**, *10*. DOI: 10.3390/genes10010035.

[23] Zaringhalam, M.; Papavasiliou, F. N. Pseudouridylation meets next-generation sequencing. *Methods (San Diego, Calif.)* **2016**, *107*, 63–72. DOI: 10.1016/j.ymeth.2016.03.001.

[24] Machnicka, M. A.; Dunin-Horkawicz, S.; Crécy-Lagard, V. de; Bujnicki, J. M. tRNAmodpred: A computational method for predicting posttranscriptional modifications in tRNAs. *Methods (San Diego, Calif.)* **2016**, *107*, 34–41. DOI: 10.1016/j.ymeth.2016.03.013.

[25] Nachtergaele, S.; He, C. The emerging biology of RNA post-transcriptional modifications. *RNA biology* **2017**, *14*, 156–163. DOI: 10.1080/15476286.2016.1267096.

[26] Phizicky, E. M.; Hopper, A. K. tRNA biology charges to the front. *Genes & development* **2010**, *24*, 1832–1860. DOI: 10.1101/gad.1956510.

[27] Weixlbaumer, A.; Murphy, F. V.; Dziergowska, A.; Malkiewicz, A.; Vendeix, F. A. P.; Agris, P. F.; Ramakrishnan, V. Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. *Nature structural & molecular biology* **2007**, *14*, 498–502. DOI: 10.1038/nsmb1242.

[28] El Yacoubi, B.; Bailly, M.; Crécy-Lagard, V. de. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annual review of genetics* **2012**, *46*, 69–95. DOI: 10.1146/annurev-genet-110711-155641.

[29] Gustilo, E. M.; Vendeix, F. A.; Agris, P. F. tRNA's modifications bring order to gene expression. *Current opinion in microbiology* **2008**, *11*, 134–140. DOI: 10.1016/j.mib.2008.02.003.

[30] Hori, H. Methylated nucleosides in tRNA and tRNA methyltransferases. *Frontiers in genetics* **2014**, *5*, 144. DOI: 10.3389/fgene.2014.00144.

[31] Rezgui, V. A. N.; Tyagi, K.; Ranjan, N.; Konevega, A. L.; Mittelstaet, J.; Rodnina, M. V.; Peter, M.; Pedrioli, P. G. A. tRNA tKUUU, tQUUG, and tEUUC wobble position modifications fine-tune protein translation by promoting ribosome A-site binding. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 12289–12294. DOI: 10.1073/pnas.1300781110.

[32] Tükenmez, H.; Xu, H.; Esberg, A.; Byström, A. S. The role of wobble uridine modifications in +1 translational frameshifting in eukaryotes. *Nucleic acids research* **2015**, *43*, 9489–9499. DOI: 10.1093/nar/gkv832.

[33] Dalluge, J. J.; Hashizume, T.; Sopchik, A. E.; McCloskey, J. A.; Davis, D. R. Conformational flexibility in RNA: the role of dihydrouridine. *Nucleic acids research* **1996**, *24*, 1073–1079. DOI: 10.1093/nar/24.6.1073.

[34] Oerum, S.; Dégut, C.; Barraud, P.; Tisné, C. m1A Post-Transcriptional Modification in tRNAs. *Biomolecules* **2017**, *7*. DOI: 10.3390/biom7010020.

[35] Anderson, J.; Phan, L.; Cuesta, R.; Carlson, B. A.; Pak, M.; Asano, K.; Björk, G. R.; Tamame, M.; Hinnebusch, A. G. The essential Gcd10p-Gcd14p nuclear complex is required for 1-methyladenosine modification and maturation of initiator methionyl-tRNA. *Genes & development* **1998**, *12*, 3650–3662. DOI: 10.1101/gad.12.23.3650.

[36] Kawai, G.; Yamamoto, Y.; Kamimura, T.; Masegi, T.; Sekine, M.; Hata, T.; Iimori, T.; Watanabe, T.; Miyazawa, T.; Yokoyama, S. Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. *Biochemistry* **1992**, *31*, 1040–1046. DOI: 10.1021/bi00119a012.

[37] Lee, C. H.; Tinoco, I. Studies of the conformation of modified dinucleoside phosphates containing 1,N6-ethenoadenosine and 2'-O-methylcytidine by 360-MHz 1H nuclear magnetic resonance spectroscopy. Investigation of the solution conformations of dinucleoside phosphates. *Biochemistry* **1977**, *16*, 5403–5414. DOI: 10.1021/bi00644a001.

[38] Watanabe, K.; Yokoyama, S.; Hansske, F.; Kasai, H.; Miyazawa, T. CD and NMR studies on the conformational thermostability of 2-thioribothymidine found in the T psi C loop of thermophile tRNA. *Biochemical and biophysical research communications* **1979**, *91*, 671–677. DOI: 10.1016/0006-291x(79)91574-2.

[39] Yokoyama, S.; Watanabe, K.; Miyazawa, T. Dynamic structures and functions of transfer ribonucleic acids from extreme thermophiles. *Advances in biophysics* **1987**, *23*, 115–147. DOI: 10.1016/0065-227x(87)90006-2.

[40] Voigts-Hoffmann, F.; Hengesbach, M.; Kobitski, A. Y.; van Aerschot, A.; Herdewijn, P.; Nienhaus, G. U.; Helm, M. A methyl group controls conformational equilibrium in human mitochondrial tRNA(Lys). *Journal of the American Chemical Society* **2007**, *129*, 13382–13383. DOI: 10.1021/ja075520.

[41] Helm, M.; Attardi, G. Nuclear control of cloverleaf structure of human mitochondrial tRNA(Lys). *Journal of molecular biology* **2004**, *337*, 545–560. DOI: 10.1016/j.jmb.2004.01.036.

[42] Helm, M.; Giegé, R.; Florentz, C. A Watson-Crick base-pair-disrupting methyl group (m1A9) is sufficient for cloverleaf folding of human mitochondrial tRNALys. *Biochemistry* **1999**, *38*, 13338–13346. DOI: 10.1021/bi991061g.

[43] Wakeman, C. A.; Winkler, W. C. Analysis of the RNA backbone: structural analysis of riboswitches by in-line probing and selective 2'-hydroxyl acylation and primer extension. *Methods in molecular biology (Clifton, N.J.)* **2009**, *540*, 173–191. DOI: 10.1007/978-1-59745-558-9_13.

[44] Jiang, H. Q.; Motorin, Y.; Jin, Y. X.; Grosjean, H. Pleiotropic effects of intron removal on base modification pattern of yeast tRNAPhe: an in vitro study. *Nucleic acids research* **1997**, *25*, 2694–2701. DOI: 10.1093/nar/25.14.2694.

[45] Pintard, L.; Lecointe, F.; Bujnicki, J. M.; Bonnerot, C.; Grosjean, H.; Lapeyre, B. Trm7p catalyses the formation of two 2'-O-methylriboses in yeast tRNA anticodon loop. *The EMBO journal* **2002**, *21*, 1811–1820. DOI: 10.1093/emboj/21.7.1811.

[46] Rider, L. W.; Ottosen, M. B.; Gattis, S. G.; Palfey, B. A. Mechanism of dihydrouridine synthase 2 from yeast and the importance of modifications for efficient tRNA reduction. *The Journal of biological chemistry* **2009**, *284*, 10324–10333. DOI: 10.1074/jbc.M806137200.

[47] Belin, S.; Beghin, A.; Solano-Gonzàlez, E.; Bezin, L.; Brunet-Manquat, S.; Textoris, J.; Prats, A.-C.; Mertani, H. C.; Dumontet, C.; Diaz, J.-J. Dysregulation of ribosome biogenesis and translational capacity is associated with tumor progression of human breast cancer cells. *PloS one* **2009**, *4*, e7147. DOI: 10.1371/journal.pone.0007147.

[48] Jack, K.; Bellodi, C.; Landry, D. M.; Niederer, R. O.; Meskauskas, A.; Musalgaonkar, S.; Kopmar, N.; Krasnykh, O.; Dean, A. M.; Thompson, S. R.*; et al.* rRNA pseudouridylation defects affect ribosomal ligand binding and translational fidelity from yeast to human cells. *Molecular cell* **2011**, *44*, 660–666. DOI: 10.1016/j.molcel.2011.09.017.

[49] Sharma, S.; Lafontaine, D. L. J. 'View From A Bridge': A New Perspective on Eukaryotic rRNA Base Modification. *Trends in biochemical sciences* **2015**, *40*, 560–575. DOI: 10.1016/j.tibs.2015.07.008.

[50] Yoon, A.; Peng, G.; Brandenburger, Y.; Brandenburg, Y.; Zollo, O.; Xu, W.; Rego, E.; Ruggero, D. Impaired control of IRES-mediated translation in X-linked dyskeratosis congenita. *Science (New York, N.Y.)* **2006**, *312*, 902–906. DOI: 10.1126/science.1123835.

[51] Zhao, B. S.; Roundtree, I. A.; He, C. Post-transcriptional gene regulation by mRNA modifications. *Nature reviews. Molecular cell biology* **2017**, *18*, 31–42. DOI: 10.1038/nrm.2016.132.

[52] Meyer, K. D.; Patil, D. P.; Zhou, J.; Zinoviev, A.; Skabkin, M. A.; Elemento, O.; Pestova, T. V.; Qian, S.-B.; Jaffrey, S. R. 5' UTR m(6)A Promotes Cap-Independent Translation. *Cell* **2015**, *163*, 999–1010. DOI: 10.1016/j.cell.2015.10.012.

[53] Zhou, J.; Wan, J.; Gao, X.; Zhang, X.; Jaffrey, S. R.; Qian, S.-B. Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature* **2015**, *526*, 591–594. DOI: 10.1038/nature15377.

[54] Wang, X.; Lu, Z.; Gomez, A.; Hon, G. C.; Yue, Y.; Han, D.; Fu, Y.; Parisien, M.; Dai, Q.; Jia, G.*; et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **2014**, *505*, 117–120. DOI: 10.1038/nature12730.

[55] Dominissini, D.; Moshitch-Moshkovitz, S.; Schwartz, S.; Salmon-Divon, M.; Ungar, L.; Osenberg, S.; Cesarkas, K.; Jacob-Hirsch, J.; Amariglio, N.; Kupiec, M.*; et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **2012**, *485*, 201–206. DOI: 10.1038/nature11112.

[56] Harvey, R.; Dezi, V.; Pizzinga, M.; Willis, A. E. Post-transcriptional control of gene expression following stress: the role of RNA-binding proteins. *Biochemical Society transactions* **2017**, *45*, 1007–1014. DOI: 10.1042/BST20160364.

[57] Carlile, T. M.; Rojas-Duran, M. F.; Zinshteyn, B.; Shin, H.; Bartoli, K. M.; Gilbert, W. V. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **2014**, *515*, 143–146. DOI: 10.1038/nature13802.

[58] Li, X.; Zhu, P.; Ma, S.; Song, J.; Bai, J.; Sun, F.; Yi, C. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nature chemical biology* **2015**, *11*, 592–597. DOI: 10.1038/nchembio.1836.

[59] Schwartz, S.; Bernstein, D. A.; Mumbach, M. R.; Jovanovic, M.; Herbst, R. H.; León-Ricardo, B. X.; Engreitz, J. M.; Guttman, M.; Satija, R.; Lander, E. S.*; et al.* Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **2014**, *159*, 148–162. DOI: 10.1016/j.cell.2014.08.028.

[60] Zheng, G.; Dahl, J. A.; Niu, Y.; Fedorcsak, P.; Huang, C.-M.; Li, C. J.; Vågbø, C. B.; Shi, Y.; Wang, W.-L.; Song, S.-H.*; et al.* ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Molecular cell* **2013**, *49*, 18–29. DOI: 10.1016/j.molcel.2012.10.015.

[61] Liu, J.; Yue, Y.; Han, D.; Wang, X.; Fu, Y.; Zhang, L.; Jia, G.; Yu, M.; Lu, Z.; Deng, X.*; et al.* A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nature chemical biology* **2014**, *10*, 93–95. DOI: 10.1038/nchembio.1432.

[62] Jia, G.; Fu, Y.; Zhao, X.; Dai, Q.; Zheng, G.; Yang, Y.; Yi, C.; Lindahl, T.; Pan, T.; Yang, Y.-G.*; et al.* N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature chemical biology* **2011**, *7*, 885–887. DOI: 10.1038/nchembio.687.

[63] Cohn, W. E.; Volklin, E. Nucleoside-5′-Phosphates from Ribonucleic Acid. *Nature* **1951**, *167*, 483–484. DOI: 10.1038/167483a0.

[64] Davis, F. F.; Allen, F. W. Ribonucleic acids from yeast which contain a fifth nucleotide. *The Journal of biological chemistry* **1957**, *227*, 907–915.

[65] Ge, J.; Yu, Y.-T. RNA pseudouridylation: new insights into an old modification. *Trends in biochemical sciences* **2013**, *38*, 210–218. DOI: 10.1016/j.tibs.2013.01.002.

[66] Spenkuch, F.; Motorin, Y.; Helm, M. Pseudouridine: still mysterious, but never a fake (uridine)! *RNA biology* **2014**, *11*, 1540–1554. DOI: 10.4161/15476286.2014.992278.

[67] Lovejoy, A. F.; Riordan, D. P.; Brown, P. O. Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. *PloS one* **2014**, *9*, e110799. DOI: 10.1371/journal.pone.0110799.

[68] Li, X.; Ma, S.; Yi, C. Pseudouridine: the fifth RNA nucleotide with renewed interests. *Current opinion in chemical biology* **2016**, *33*, 108–116. DOI: 10.1016/j.cbpa.2016.06.014.

[69] Karijolich, J.; Yi, C.; Yu, Y.-T. Transcriptome-wide dynamics of RNA pseudouridylation. *Nature reviews. Molecular cell biology* **2015**, *16*, 581–585. DOI: 10.1038/nrm4040.

[70] Hamma, T.; Ferré-D'Amaré, A. R. Pseudouridine synthases. *Chemistry & biology* **2006**, *13*, 1125–1135. DOI: 10.1016/j.chembiol.2006.09.009.

[71] Charette, M.; Gray, M. W. Pseudouridine in RNA: what, where, how, and why. *IUBMB life* **2000**, *49*, 341–351. DOI: 10.1080/152165400410182.

[72] Yu, C. T.; Allen, F. W. Studies on an isomer of uridine isolated from ribonucleic acids. *Biochimica et biophysica acta* **1959**, *32*, 393–406. DOI: 10.1016/0006-3002(59)90612-2.

[73] Pomerantz, S. C.; McCloskey, J. A. Detection of the common RNA nucleoside pseudouridine in mixtures of oligonucleotides by mass spectrometry. *Analytical chemistry* **2005**, *77*, 4687–4697. DOI: 10.1021/ac058023p.

[74] Durairaj, A.; Limbach, P. A. Mass spectrometry of the fifth nucleoside: a review of the identification of pseudouridine in nucleic acids. *Analytica chimica acta* **2008**, *623*, 117–125. DOI: 10.1016/j.aca.2008.06.027.

[75] Arnez, J. G.; Steitz, T. A. Crystal structure of unmodified tRNA(Gln) complexed with glutaminyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. *Biochemistry* **1994**, *33*, 7560–7567. DOI: 10.1021/bi00190a008.

[76] Auffinger, P.; Westhof, E. *Modification and Editing of RNA: Effects of pseudouridylation on tRNA hydration and dynamics: A theoretical approach*; ASM Press: Washington DC, 1998.

[77] Davis, D. R.; Poulter, C. D. 1H-15N NMR studies of Escherichia coli tRNA(Phe) from hisT mutants: a structural role for pseudouridine. *Biochemistry* **1991**, *30*, 4223–4231. DOI: 10.1021/bi00231a017.

[78] Kierzek, E.; Malgowska, M.; Lisowiec, J.; Turner, D. H.; Gdaniec, Z.; Kierzek, R. The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic acids research* **2014**, *42*, 3492–3501. DOI: 10.1093/nar/gkt1330.

[79] Davis, D. R.; Veltri, C. A.; Nielsen, L. An RNA model system for investigation of pseudouridine stabilization of the codon-anticodon interaction in tRNALys, tRNAHis and tRNATyr. *Journal of biomolecular structure & dynamics* **1998**, *15*, 1121–1132. DOI: 10.1080/07391102.1998.10509006.

[80] Davis, D. R. Stabilization of RNA stacking by pseudouridine. *Nucleic acids research* **1995**, *23*, 5020–5026. DOI: 10.1093/nar/23.24.5020.

[81] Durant, P. C.; Davis, D. R. Stabilization of the anticodon stem-loop of tRNALys,3 by an A+-C base-pair and by pseudouridine. *Journal of molecular biology* **1999**, *285*, 115–131. DOI: 10.1006/jmbi.1998.2297.

[82] Newby, M. I.; Greenbaum, N. L. Investigation of Overhauser effects between pseudouridine and water protons in RNA helices. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 12697–12702. DOI: 10.1073/pnas.202477199.

[83] Hudson, G. A.; Bloomingdale, R. J.; Znosko, B. M. Thermodynamic contribution and nearest-neighbor parameters of pseudouridine-adenosine base pairs in oligoribonucleotides. *RNA (New York, N.Y.)* **2013**, *19*, 1474–1482. DOI: 10.1261/rna.039610.113.

[84] Lane, B. G.; Ofengand, J.; Gray, M. W. 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. *Biochimie* **1995**, *77*, 7–15. DOI: 10.1016/0300-9084(96)88098-9.

[85] Saponara, A. G.; Enger, M. D. The isolation from ribonucleic acid of substituted uridines containing alpha-aminobutyrate moieties derived from methionine. *Biochimica et biophysica acta* **1974**, *349*, 61–77. DOI: 10.1016/0005-2787(74)90009-4.

[86] Brand, R. C.; Klootwijk, J.; Planta, R. J.; Maden, B. E. Biosynthesis of a hypermodified nucleotide in Saccharomyces carlsbergensis 17S and HeLa-cell 18S ribosomal ribonucleic acid. *The Biochemical journal* **1978**, *169*, 71–77. DOI: 10.1042/bj1690071.

[87] Grosjean, H.; Sprinzl, M.; Steinberg, S. Posttranscriptionally modified nucleosides in transfer RNA: their locations and frequencies. *Biochimie* **1995**, *77*, 139–141. DOI: 10.1016/0300-9084(96)88117-x.

[88] Dao, V.; Guenther, R.; Malkiewicz, A.; Nawrot, B.; Sochacka, E.; Kraszewski, A.; Jankowska, J.; Everett, K.; Agris, P. F. Ribosome binding of DNA analogs of tRNA requires base modifications and supports the "extended anticodon". *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 2125–2129. DOI: 10.1073/pnas.91.6.2125.

[89] Fernández, I. S.; Ng, C. L.; Kelley, A. C.; Wu, G.; Yu, Y.-T.; Ramakrishnan, V. Unusual base pairing during the decoding of a stop codon by the ribosome. *Nature* **2013**, *500*, 107–110. DOI: 10.1038/nature12302.

[90] Tomita, K.; Ueda, T.; Watanabe, K. The presence of pseudouridine in the anticodon alters the genetic code: a possible mechanism for assignment of the AAA lysine codon as asparagine in echinoderm mitochondria. *Nucleic acids research* **1999**, *27*, 1683–1689. DOI: 10.1093/nar/27.7.1683.

[91] Bakin, A.; Lane, B. G.; Ofengand, J. Clustering of pseudouridine residues around the peptidyltransferase center of yeast cytoplasmic and mitochondrial ribosomes. *Biochemistry* **1994**, *33*, 13475–13483. DOI: 10.1021/bi00249a036.

[92] Bakin, A.; Ofengand, J. Four newly located pseudouridylate residues in Escherichia coli 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry* **1993**, *32*, 9754–9762. DOI: 10.1021/bi00088a030.

[93] Brimacombe, R.; Mitchell, P.; Osswald, M.; Stade, K.; Bochkariov, D. Clustering of modified nucleotides at the functional center of bacterial ribosomal RNA. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **1993**, *7*, 161–167. DOI: 10.1096/fasebj.7.1.8422963.

[94] Cunningham, P. R.; Richard, R. B.; Weitzmann, C. J.; Nurse, K.; Ofengand, J. The absence of modified nucleotides affects both in vitro assembly and in vitro function of the 30S ribosomal subunit of Escherichia coli. *Biochimie* **1991**, *73*, 789–796. DOI: 10.1016/0300-9084(91)90058-9.

[95] Massenet, S.; Mougin, A.; Branlant, C. *Modification and Editing of RNA: Posttranscriptional modifications in the U small nuclear RNAs*; ASM Press: Washington DC, 1998.

[96] Wu, G.; Xiao, M.; Yang, C.; Yu, Y.-T. U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. *The EMBO journal* **2011**, *30*, 79–89. DOI: 10.1038/emboj.2010.316.

[97] Watanabe, Y.; Gray, M. W. Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in Euglena gracilis, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaebacteria. *Nucleic acids research* **2000**, *28*, 2342–2352. DOI: 10.1093/nar/28.12.2342.

[98] Sivaraman, J.; Sauvé, V.; Larocque, R.; Stura, E. A.; Schrag, J. D.; Cygler, M.; Matte, A. Structure of the 16S rRNA pseudouridine synthase RsuA bound to uracil and UMP. *Nature structural biology* **2002**, *9*, 353–358. DOI: 10.1038/nsb788.

[99] Hoang, C.; Ferre-D'Amare, A. R. Crystal structure of the highly divergent pseudouridine synthase TruD reveals a circular permutation of a conserved fold. *RNA (New York, N.Y.)* **2004**, *10*, 1026–1033. DOI: 10.1261/rna.7240504.

[100] Hoang, C.; Chen, J.; Vizthum, C. A.; Kandel, J. M.; Hamilton, C. S.; Mueller, E. G.; Ferré-D'Amaré, A. R. Crystal structure of pseudouridine synthase RluA: indirect sequence readout through protein-induced RNA structure. *Molecular cell* **2006**, *24*, 535–545. DOI: 10.1016/j.molcel.2006.09.017.

[101] Hoang, C.; Ferré-D'Amaré, A. R. Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell* **2001**, *107*, 929–939. DOI: 10.1016/s0092-8674(01)00618-3.

[102] Foster, P. G.; Huang, L.; Santi, D. V.; Stroud, R. M. The structural basis for tRNA recognition and pseudouridine formation by pseudouridine synthase I. *Nature structural biology* **2000**, *7*, 23–27. DOI: 10.1038/71219.

[103] Kamalampeta, R.; Keffer-Wilkes, L. C.; Kothe, U. tRNA binding, positioning, and modification by the pseudouridine synthase Pus10. *Journal of molecular biology* **2013**, *425*, 3863–3874. DOI: 10.1016/j.jmb.2013.05.022.

[104] Kaya, Y.; Del Campo, M.; Ofengand, J.; Malhotra, A. Crystal structure of TruD, a novel pseudouridine synthase with a new protein fold. *The Journal of biological chemistry* **2004**, *279*, 18107–18110. DOI: 10.1074/jbc.C400072200.

[105] Koonin, E. V. Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. *Nucleic acids research* **1996**, *24*, 2411–2415. DOI: 10.1093/nar/24.12.2411.

[106] Czudnochowski, N.; Ashley, G. W.; Santi, D. V.; Alian, A.; Finer-Moore, J.; Stroud, R. M. The mechanism of pseudouridine synthases from a covalent complex with RNA, and alternate specificity for U2605 versus U2604 between close homologs. *Nucleic acids research* **2014**, *42*, 2037–2048. DOI: 10.1093/nar/gkt1050.

[107] Alian, A.; DeGiovanni, A.; Griner, S. L.; Finer-Moore, J. S.; Stroud, R. M. Crystal structure of an RluF-RNA complex: a base-pair rearrangement is the key to selectivity of RluF for U2604 of the ribosome. *Journal of molecular biology* **2009**, *388*, 785–800. DOI: 10.1016/j.jmb.2009.03.029.

[108] Friedt, J.; Leavens, F. M. V.; Mercier, E.; Wieden, H.-J.; Kothe, U. An arginine-aspartate network in the active site of bacterial TruB is critical for catalyzing pseudouridine formation. *Nucleic acids research* **2014**, *42*, 3857–3870. DOI: 10.1093/nar/gkt1331.

[109] Wright, J. R.; Keffer-Wilkes, L. C.; Dobing, S. R.; Kothe, U. Pre-steady-state kinetic analysis of the three Escherichia coli pseudouridine synthases TruB, TruA, and RluA reveals uniformly slow catalysis. *RNA (New York, N.Y.)* **2011**, *17*, 2074–2084. DOI: 10.1261/rna.2905811.

[110] Arluison, V.; Buckle, M.; Grosjean, H. Pseudouridine synthetase Pus1 of Saccharomyces cerevisiae: kinetic characterisation, tRNA structural requirement and real-time analysis of its complex with tRNA. *Journal of molecular biology* **1999**, *289*, 491–502. DOI: 10.1006/jmbi.1999.2789.

[111] Veerareddygari, G. R.; Singh, S. K.; Mueller, E. G. The Pseudouridine Synthases Proceed through a Glycal Intermediate. *J. Am. Chem. Soc.* **2016**, *138*, 7852–7855. DOI: 10.1021/jacs.6b04491.

[112] Gu, X.; Liu, Y.; Santi, D. V. The mechanism of pseudouridine synthase I as deduced from its interaction with 5-fluorouracil-tRNA. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 14270–14275. DOI: 10.1073/pnas.96.25.14270.

[113] McDonald, M. K.; Miracco, E. J.; Chen, J.; Xie, Y.; Mueller, E. G. The handling of the mechanistic probe 5-fluorouridine by the pseudouridine synthase TruA and its consistency with the handling of the same probe by the pseudouridine synthases TruB and RluA. *Biochemistry* **2011**, *50*, 426–436. DOI: 10.1021/bi101737z.

[114] Huang, S.; Mahanta, N.; Begley, T. P.; Ealick, S. E. Pseudouridine monophosphate glycosidase: a new glycosidase mechanism. *Biochemistry* **2012**, *51*, 9245–9255. DOI: 10.1021/bi3006829.

[115] Silva, R. G.; Schramm, V. L. Uridine phosphorylase from Trypanosoma cruzi: kinetic and chemical mechanisms. *Biochemistry* **2011**, *50*, 9158–9166. DOI: 10.1021/bi2013382.

[116] Wu, J. C.; Santi, D. V. Kinetic and catalytic mechanism of HhaI methyltransferase. *The Journal of biological chemistry* **1987**, *262*, 4778–4786.

[117] Santi, D. V.; McHenry, C. S.; Sommer, H. Mechanism of interaction of thymidylate synthetase with 5-fluorodeoxyuridylate. *Biochemistry* **1974**, *13*, 471–481. DOI: 10.1021/bi00700a012.

[118] Kunitani, M. G.; Santi, D. V. On the mechanism of 2'-deoxyuridylate hydroxymethylase. *Biochemistry* **1980**, *19*, 1271–1275. DOI: 10.1021/bi00548a001.

[119] Phannachet, K.; Elias, Y.; Huang, R. H. Dissecting the roles of a strictly conserved tyrosine in substrate recognition and catalysis by pseudouridine 55 synthase. *Biochemistry* **2005**, *44*, 15488–15494. DOI: 10.1021/bi050961w.

[120] Phannachet, K.; Huang, R. H. Conformational change of pseudouridine 55 synthase upon its association with RNA substrate. *Nucleic acids research* **2004**, *32*, 1422–1429. DOI: 10.1093/nar/gkh287.

[121] Miracco, E. J.; Mueller, E. G. The products of 5-fluorouridine by the action of the pseudouridine synthase TruB disfavor one mechanism and suggest another. *Journal of the American Chemical Society* **2011**, *133*, 11826–11829. DOI: 10.1021/ja201179f.

[122] Kiss, T.; Fayet-Lebaron, E.; Jády, B. E. Box H/ACA small ribonucleoproteins. *Molecular cell* **2010**, *37*, 597–606. DOI: 10.1016/j.molcel.2010.01.032.

[123] Watkins, N. J.; Bohnsack, M. T. The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley interdisciplinary reviews. RNA* **2012**, *3*, 397–414. DOI: 10.1002/wrna.117.

[124] Narayanan, A.; Lukowiak, A.; Jády, B. E.; Dragon, F.; Kiss, T.; Terns, R. M.; Terns, M. P. Nucleolar localization signals of box H/ACA small nucleolar RNAs. *The EMBO journal* **1999**, *18*, 5120–5130. DOI: 10.1093/emboj/18.18.5120.

[125] Darzacq, X.; Jády, B. E.; Verheggen, C.; Kiss, A. M.; Bertrand, E.; Kiss, T. Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *The EMBO journal* **2002**, *21*, 2746–2756. DOI: 10.1093/emboj/21.11.2746.

[126] Wang, C.; Meier, U. T. Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *The EMBO journal* **2004**, *23*, 1857–1867. DOI: 10.1038/sj.emboj.7600181.

[127] Wang, C.; Query, C. C.; Meier, U. T. Immunopurified small nucleolar ribonucleoprotein particles pseudouridylate rRNA independently of their association with phosphorylated Nopp140. *Molecular and cellular biology* **2002**, *22*, 8457–8466. DOI: 10.1128/mcb.22.24.8457-8466.2002.

[128] Watkins, N. J.; Ségault, V.; Charpentier, B.; Nottrott, S.; Fabrizio, P.; Bachi, A.; Wilm, M.; Rosbash, M.; Branlant, C.; Lührmann, R. A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* **2000**, *103*, 457–466. DOI: 10.1016/s0092-8674(00)00137-9.

[129] Reichow, S. L.; Hamma, T.; Ferré-D'Amaré, A. R.; Varani, G. The structure and function of small nucleolar ribonucleoproteins. *Nucleic acids research* **2007**, *35*, 1452–1464. DOI: 10.1093/nar/gkl1172.

[130] Kuhn, J. F.; Tran, E. J.; Maxwell, E. S. Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic acids research* **2002**, *30*, 931–941. DOI: 10.1093/nar/30.4.931.

[131] Lin, J.; Lai, S.; Jia, R.; Xu, A.; Zhang, L.; Lu, J.; Ye, K. Structural basis for site-specific ribose methylation by box C/D RNA protein complexes. *Nature* **2011**, *469*, 559–563. DOI: 10.1038/nature09688.

[132] Xue, S.; Wang, R.; Yang, F.; Terns, R. M.; Terns, M. P.; Zhang, X.; Maxwell, E. S.; Li, H. Structural basis for substrate placement by an archaeal box C/D ribonucleoprotein particle. *Molecular cell* **2010**, *39*, 939–949. DOI: 10.1016/j.molcel.2010.08.022.

[133] Omer, A. D.; Ziesche, S.; Ebhardt, H.; Dennis, P. P. In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proceedings of the National*

Academy of Sciences of the United States of America **2002**, *99*, 5289–5294. DOI: 10.1073/pnas.082101999.

[134] Aittaleb, M.; Rashid, R.; Chen, Q.; Palmer, J. R.; Daniels, C. J.; Li, H. Structure and function of archaeal box C/D sRNP core proteins. *Nature structural biology* **2003**, *10*, 256–263. DOI: 10.1038/nsb905.

[135] Bleichert, F.; Gagnon, K. T.; Brown, B. A.; Maxwell, E. S.; Leschziner, A. E.; Unger, V. M.; Baserga, S. J. A dimeric structure for archaeal box C/D small ribonucleoproteins. *Science (New York, N.Y.)* **2009**, *325*, 1384–1387. DOI: 10.1126/science.1176099.

[136] Balakin, A. G.; Smith, L.; Fournier, M. J. The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell* **1996**, *86*, 823–834. DOI: 10.1016/s0092-8674(00)80156-7.

[137] Zoysa, M. D. de; Yu, Y.-T. Posttranscriptional RNA Pseudouridylation. *The Enzymes* **2017**, *41*, 151–167. DOI: 10.1016/bs.enz.2017.02.001.

[138] Liang, B.; Zhou, J.; Kahen, E.; Terns, R. M.; Terns, M. P.; Li, H. Structure of a functional ribonucleoprotein pseudouridine synthase bound to a substrate RNA. *Nature structural & molecular biology* **2009**, *16*, 740–746. DOI: 10.1038/nsmb.1624.

[139] Ganot, P.; Bortolin, M. L.; Kiss, T. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **1997**, *89*, 799–809. DOI: 10.1016/s0092-8674(00)80263-9.

[140] 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*, 427–439. DOI: 10.1016/j.molcel.2009.05.005.

[141] 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. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 6655–6660. DOI: 10.1073/pnas.0701534104.

[142] Kiss, A. M.; Jády, B. E.; Darzacq, X.; Verheggen, C.; Bertrand, E.; Kiss, T. A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. *Nucleic acids research* **2002**, *30*, 4643–4649. DOI: 10.1093/nar/gkf592.

[143] Rozhdestvensky, T. S.; Tang, T. H.; Tchirkova, I. V.; Brosius, J.; Bachellerie, J.-P.; Hüttenhofer, A. Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic acids research* **2003**, *31*, 869–877. DOI: 10.1093/nar/gkg175.

[144] Watkins, N. J.; 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 (New York, N.Y.)* **1998**, *4*, 1549–1568. DOI: 10.1017/s1355838298980761.

[145] Terns, M.; Terns, R. Noncoding RNAs of the H/ACA family. *Cold Spring Harbor symposia on quantitative biology* **2006**, *71*, 395–405. DOI: 10.1101/sqb.2006.71.034.

[146] Meier, U. T.; Blobel, G. NAP57, a mammalian nucleolar protein with a putative homolog in yeast and bacteria. *The Journal of cell biology* **1994**, *127*, 1505–1514. DOI: 10.1083/jcb.127.6.1505.

[147] Heiss, N. S.; Knight, S. W.; Vulliamy, T. J.; Klauck, S. M.; Wiemann, S.; Mason, P. J.; 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*, 32–38. DOI: 10.1038/ng0598-32.

[148] Henras, A.; Henry, Y.; Bousquet-Antonelli, C.; Noaillac-Depeyre, J.; Gélugne, J. P.; Caizergues-Ferrer, M. Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *The EMBO journal* **1998**, *17*, 7078–7090. DOI: 10.1093/emboj/17.23.7078.

[149] Lafontaine, D. L.; Bousquet-Antonelli, C.; Henry, Y.; Caizergues-Ferrer, M.; Tollervey, D. The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes & development* **1998**, *12*, 527–537. DOI: 10.1101/gad.12.4.527.

[150] Li, L.; Ye, K. Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* **2006**, *443*, 302–307. DOI: 10.1038/nature05151.

[151] Zebarjadian, Y.; King, T.; Fournier, M. J.; Clarke, L.; Carbon, J. Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. *Molecular and cellular biology* **1999**, *19*, 7461–7472. DOI: 10.1128/mcb.19.11.7461.

[152] Khanna, M.; Wu, H.; Johansson, C.; Caizergues-Ferrer, M.; Feigon, J. Structural study of the H/ACA snoRNP components Nop10p and the 3' hairpin of U65 snoRNA. *RNA (New York, N.Y.)* **2006**, *12*, 40–52. DOI: 10.1261/rna.2221606.

[153] Hamma, T.; Reichow, S. L.; Varani, G.; Ferré-D'Amaré, A. R. The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nature structural & molecular biology* **2005**, *12*, 1101–1107. DOI: 10.1038/nsmb1036.

[154] Nottrott, S.; Hartmuth, K.; Fabrizio, P.; Urlaub, H.; Vidovic, I.; Ficner, R.; Lührmann, R. Functional interaction of a novel 15.5kD U4/U6.U5 tri-snRNP protein with the 5' stem-loop of U4 snRNA. *The EMBO journal* **1999**, *18*, 6119–6133. DOI: 10.1093/emboj/18.21.6119.

[155] 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*, 2733–2746. DOI: 10.1093/nar/29.13.2733.

[156] Charpentier, B.; Muller, S.; Branlant, C. Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic acids research* **2005**, *33*, 3133–3144. DOI: 10.1093/nar/gki630.

[157] Liang, B.; Xue, S.; Terns, R. M.; Terns, M. P.; Li, H. Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex. *Nature structural & molecular biology* **2007**, *14*, 1189–1195. DOI: 10.1038/nsmb1336.

[158] Girard, J. P.; Lehtonen, H.; Caizergues-Ferrer, M.; Amalric, F.; Tollervey, D.; Lapeyre, B. GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast. *The EMBO journal* **1992**, *11*, 673–682.

[159] Caton, E. A.; Kelly, E. K.; Kamalampeta, R.; Kothe, U. Efficient RNA pseudouridylation by eukaryotic H/ACA ribonucleoproteins requires high affinity binding and correct positioning of guide RNA. *Nucleic acids research* **2018**, *46*, 905–916. DOI: 10.1093/nar/gkx1167.

[160] Li, S.; Duan, J.; Li, D.; Yang, B.; Dong, M.; Ye, K. Reconstitution and structural analysis of the yeast box H/ACA RNA-guided pseudouridine synthase. *Genes & development* **2011**, *25*, 2409–2421. DOI: 10.1101/gad.175299.111.

[161] Rashid, R.; Liang, B.; Baker, D. L.; Youssef, O. A.; He, Y.; Phipps, K.; Terns, R. M.; Terns, M. P.; Li, H. Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Molecular cell* **2006**, *21*, 249–260. DOI: 10.1016/j.molcel.2005.11.017.

[162] Yang, X.; Duan, J.; Li, S.; Wang, P.; Ma, S.; Ye, K.; Zhao, X. S. Kinetic and thermodynamic characterization of the reaction pathway of box H/ACA RNA-guided pseudouridine formation. *Nucleic acids research* **2012**, *40*, 10925–10936. DOI: 10.1093/nar/gks882.

[163] Darzacq, X.; Kittur, N.; Roy, S.; Shav-Tal, Y.; Singer, R. H.; Meier, U. T. Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. *The Journal of cell biology* **2006**, *173*, 207–218. DOI: 10.1083/jcb.200601105.

[164] Fatica, A.; Dlakić, M.; Tollervey, D. Naf1 p is a box H/ACA snoRNP assembly factor. *RNA (New York, N.Y.)* **2002**, *8*, 1502–1514.

[165] Godin, K. S.; 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*, 231–244. DOI: 10.1016/j.jmb.2009.04.076.

[166] 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*, 1769–1778. DOI: 10.1128/mcb.24.4.1769-1778.2004.

[167] Morrissey, J. P.; Tollervey, D. Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis. *Molecular and cellular biology* **1993**, *13*, 2469–2477. DOI: 10.1128/mcb.13.4.2469.

[168] Mitchell, J. R.; 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*, 567–576. DOI: 10.1128/mcb.19.1.567.

[169] Nguyen, T. H. D.; Tam, J.; Wu, R. A.; Greber, B. J.; Toso, D.; Nogales, E.; Collins, K. Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature* **2018**, *557*, 190–195. DOI: 10.1038/s41586-018-0062-x.

[170] Ashbridge, B.; Orte, A.; Yeoman, J. A.; Kirwan, M.; Vulliamy, T.; Dokal, I.; Klenerman, D.; Balasubramanian, S. Single-molecule analysis of the human telomerase RNA.dyskerin interaction and the effect of dyskeratosis congenita mutations. *Biochemistry* **2009**, *48*, 10858–10865. DOI: 10.1021/bi901373e.

[171] Kirwan, M.; Dokal, I. Dyskeratosis congenita, stem cells and telomeres. *Biochimica et biophysica acta* **2009**, *1792*, 371–379. DOI: 10.1016/j.bbadis.2009.01.010.

[172] 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*, 825–836. DOI: 10.1093/hmg/ddp551.

[173] Walne, A. J.; Dokal, I. Advances in the understanding of dyskeratosis congenita. *British journal of haematology* **2009**, *145*, 164–172. DOI: 10.1111/j.1365-2141.2009.07598.x.

[174] Grozdanov, P. N.; Fernandez-Fuentes, N.; Fiser, A.; Meier, U. T. Pathogenic NAP57 mutations decrease ribonucleoprotein assembly in dyskeratosis congenita. *Human molecular genetics* **2009**, *18*, 4546–4551. DOI: 10.1093/hmg/ddp416.

[175] Müller, D.; Trucks, S.; Schwalbe, H.; Hengesbach, M. Genetic Code Expansion Facilitates Position-Selective Modification of Nucleic Acids and Proteins. *ChemPlusChem* **2020**, *85*, 1233–1243. DOI: 10.1002/cplu.202000150.

[176] Stryer, L. *Biochemistry*, 3. ed., internat. student ed.; Freeman: New York, NY, 1988.

[177] Hughes, A. B. *Origins and synthesis of amino acids*; s.l., 2010.

[178] Renner, C.; Alefelder, S.; Bae, J. H.; Budisa, N.; Huber, R.; Moroder, L. Fluoroprolines as Tools for Protein Design and Engineering. *Angewandte Chemie (International ed. in English)* **2001**, *40*, 923–925. DOI: 10.1002/1521-3773(20010302)40:5<923:AID-ANIE923>3.0.CO;2-#.

[179] Larregola, M.; Moore, S.; Budisa, N. Congeneric bio-adhesive mussel foot proteins designed by modified prolines revealed a chiral bias in unnatural translation. *Biochemical and biophysical research communications* **2012**, *421*, 646–650. DOI: 10.1016/j.bbrc.2012.04.031.

[180] Doerfel, L. K.; Wohlgemuth, I.; Kubyshkin, V.; Starosta, A. L.; Wilson, D. N.; Budisa, N.; Rodnina, M. V. Entropic Contribution of Elongation Factor P to Proline Positioning at the Catalytic Center of the Ribosome. *Journal of the American Chemical Society* **2015**, *137*, 12997–13006. DOI: 10.1021/jacs.5b07427.

[181] Dedkova, L. M.; Fahmi, N. E.; Paul, R.; del Rosario, M.; Zhang, L.; Chen, S.; Feder, G.; Hecht, S. M. β-Puromycin selection of modified ribosomes for in vitro incorporation of β-amino acids. *Biochemistry* **2012**, *51*, 401–415. DOI: 10.1021/bi2016124.

[182] Maini, R.; Chowdhury, S. R.; Dedkova, L. M.; Roy, B.; Daskalova, S. M.; Paul, R.; Chen, S.; Hecht, S. M. Protein Synthesis with Ribosomes Selected for the Incorporation of β-Amino Acids. *Biochemistry* **2015**, *54*, 3694–3706. DOI: 10.1021/acs.biochem.5b00389.

[183] Katoh, T.; Suga, H. Ribosomal Incorporation of Consecutive β-Amino Acids. *Journal of the American Chemical Society* **2018**, *140*, 12159–12167. DOI: 10.1021/jacs.8b07247.

[184] Porter, E. A.; Weisblum, B.; Gellman, S. H. Mimicry of host-defense peptides by unnatural oligomers: antimicrobial beta-peptides. *Journal of the American Chemical Society* **2002**, *124*, 7324–7330. DOI: 10.1021/ja0260871.

[185] Bain, J. D.; Wacker, D. A.; Kuo, E. E.; Chamberlin, A.R. Site-specific incorporation of non-natural residues into peptides: Effect of residue structure on suppression and translation efficiencies. *Tetrahedron* **1991**, *47*, 2389–2400. DOI: 10.1016/S0040-4020(01)81776-2.

[186] Heckler, T. G.; Roesser, J. R.; Xu, C.; Chang, P. I.; Hecht, S. M. Ribosomal binding and dipeptide formation by misacylated tRNA(Phe),S. *Biochemistry* **1988**, *27*, 7254–7262. DOI: 10.1021/bi00419a012.

[187] Tan, Z.; Forster, A. C.; Blacklow, S. C.; Cornish, V. W. Amino acid backbone specificity of the Escherichia coli translation machinery. *Journal of the American Chemical Society* **2004**, *126*, 12752–12753. DOI: 10.1021/ja0472174.

[188] Maini, R.; Nguyen, D. T.; Chen, S.; Dedkova, L. M.; Chowdhury, S. R.; Alcala-Torano, R.; Hecht, S. M. Incorporation of β-amino acids into dihydrofolate reductase by ribosomes having modifications in the peptidyltransferase center. *Bioorganic & medicinal chemistry* **2013**, *21*, 1088–1096. DOI: 10.1016/j.bmc.2013.01.002.

[189] Gademann, K.; Hintermann, T.; Schreiber, J. V. Beta-peptides: twisting and turning. *Current medicinal chemistry* **1999**, *6*, 905–925.

[190] Seebach, D.; Overhand, M.; Kühnle, F. N. M.; Martinoni, B.; Oberer, L.; Hommel, U.; Widmer, H. β-Peptides: Synthesis by Arndt-Eistert homologation with concomitant peptide coupling. Structure determination by NMR and CD spectroscopy and by X-ray crystallography. Helical secondary structure of a β-hexapeptide in solution and its stability towards pe. *HCA* **1996**, *79*, 913–941. DOI: 10.1002/hlca.19960790402.

[191] Kubyshkin, V.; Budisa, N. Anticipating alien cells with alternative genetic codes: away from the alanine world! *Current opinion in biotechnology* **2019**, *60*, 242–249. DOI: 10.1016/j.copbio.2019.05.006.

[192] Ambrogelly, A.; Palioura, S.; Söll, D. Natural expansion of the genetic code. *Nature chemical biology* **2007**, *3*, 29–35. DOI: 10.1038/nchembio847.

[193] Böck, A.; Forchhammer, K.; Heider, J.; Leinfelder, W.; Sawers, G.; Veprek, B.; Zinoni, F. Selenocysteine: the 21st amino acid. *Molecular microbiology* **1991**, *5*, 515–520. DOI: 10.1111/j.1365-2958.1991.tb00722.x.

[194] Polycarpo, C.; Ambrogelly, A.; Bérubé, A.; Winbush, S. M.; McCloskey, J. A.; Crain, P. F.; Wood, J. L.; Söll, D. An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 12450–12454. DOI: 10.1073/pnas.0405362101.

[195] Srinivasan, G.; James, C. M.; Krzycki, J. A. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science (New York, N.Y.)* **2002**, *296*, 1459–1462. DOI: 10.1126/science.1069588.

[196] Uy, R.; Wold, F. Posttranslational covalent modification of proteins. *Science (New York, N.Y.)* **1977**, *198*, 890–896. DOI: 10.1126/science.337487.

[197] Fowden, L. Plant amino acid research in retrospect: from Chinball to Singh. *Amino acids* **2001**, *20*, 217–224. DOI: 10.1007/s007260170040.

[198] Lu, Y.; Freeland, S. On the evolution of the standard amino-acid alphabet. *Genome biology* **2006**, *7*, 102. DOI: 10.1186/gb-2006-7-1-102;

[199] Milles, S.; Tyagi, S.; Banterle, N.; Koehler, C.; VanDelinder, V.; Plass, T.; Neal, A. P.; Lemke, E. A. Click strategies for single-molecule protein fluorescence. *Journal of the American Chemical Society* **2012**, *134*, 5187–5195. DOI: 10.1021/ja210587q.

[200] Plass, T.; Milles, S.; Koehler, C.; Schultz, C.; Lemke, E. A. Genetically encoded copper-free click chemistry. *Angewandte Chemie (International ed. in English)* **2011**, *50*, 3878–3881. DOI: 10.1002/anie.201008178.

[201] Plass, T.; Milles, S.; Koehler, C.; Szymański, J.; Mueller, R.; Wiessler, M.; Schultz, C.; Lemke, E. A. Amino acids for Diels-Alder reactions in living cells. *Angewandte Chemie (International ed. in English)* **2012**, *51*, 4166–4170. DOI: 10.1002/anie.201108231.

[202] Schneider, S.; Gattner, M. J.; Vrabel, M.; Flügel, V.; López-Carrillo, V.; Prill, S.; Carell, T. Structural insights into incorporation of norbornene amino acids for click

modification of proteins. *Chembiochem : a European journal of chemical biology* **2013**, *14*, 2114–2118. DOI: 10.1002/cbic.201300435.

[203] Inada, H.; Furukawa, K.; Shibuya, M.; Yamamoto, Y. One-pot, two-step synthesis of unnatural α-amino acids involving the exhaustive aerobic oxidation of 1,2-diols. *Chemical communications (Cambridge, England)* **2019**, *55*, 15105–15108. DOI: 10.1039/C9CC07889D.

[204] Furukawa, K.; Inada, H.; Shibuya, M.; Yamamoto, Y. Chemoselective Conversion from α-Hydroxy Acids to α-Keto Acids Enabled by Nitroxyl-Radical-Catalyzed Aerobic Oxidation. *Organic letters* **2016**, *18*, 4230–4233. DOI: 10.1021/acs.orglett.6b01964.

[205] Cellitti, S. E.; Jones, D. H.; Lagpacan, L.; Hao, X.; Zhang, Q.; Hu, H.; Brittain, S. M.; Brinker, A.; Caldwell, J.; Bursulaya, B.*; et al.* In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *Journal of the American Chemical Society* **2008**, *130*, 9268–9281. DOI: 10.1021/ja801602q;

[206] Deiters, A.; Geierstanger, B. H.; Schultz, P. G. Site-specific in vivo labeling of proteins for NMR studies. *Chembiochem : a European journal of chemical biology* **2005**, *6*, 55–58. DOI: 10.1002/cbic.200400319.

[207] Braun, T.; Drescher, M.; Summerer, D. Expanding the Genetic Code for Site-Directed Spin-Labeling. *International journal of molecular sciences* **2019**, *20*. DOI: 10.3390/ijms20020373;

[208] Kugele, A.; Braun, T. S.; Widder, P.; Williams, L.; Schmidt, M. J.; Summerer, D.; Drescher, M. Site-directed spin labelling of proteins by Suzuki-Miyaura coupling via a genetically encoded aryliodide amino acid. *Chemical communications (Cambridge, England)* **2019**, *55*, 1923–1926. DOI: 10.1039/c8cc09325c;

[209] Kugele, A.; Silkenath, B.; Langer, J.; Wittmann, V.; Drescher, M. Protein Spin Labeling with a Photocaged Nitroxide Using Diels-Alder Chemistry. *Chembiochem : a European journal of chemical biology* **2019**, *20*, 2479–2484. DOI: 10.1002/cbic.201900318;

[210] Schmidt, M. J.; Fedoseev, A.; Bücker, D.; Borbas, J.; Peter, C.; Drescher, M.; Summerer, D. EPR Distance Measurements in Native Proteins with Genetically Encoded Spin Labels. *ACS chemical biology* **2015**, *10*, 2764–2771. DOI: 10.1021/acschembio.5b00512.

[211] Widder, P.; Berner, F.; Summerer, D.; Drescher, M. Double Nitroxide Labeling by Copper-Catalyzed Azide-Alkyne Cycloadditions with Noncanonical Amino Acids for Electron Paramagnetic Resonance Spectroscopy. *ACS chemical biology* **2019**, *14*, 839–844. DOI: 10.1021/acschembio.8b01111;

[212] Lim, B. J.; Ackermann, B. E.; Debelouchina, G. T. Targetable Tetrazine-Based Dynamic Nuclear Polarization Agents for Biological Systems. *Chembiochem : a European journal of chemical biology* **2020**, *21*, 1315–1319. DOI: 10.1002/cbic.201900609.

[213] Pham, N. D.; Parker, R. B.; Kohler, J. J. Photocrosslinking approaches to interactome mapping. *Current opinion in chemical biology* **2013**, *17*, 90–101. DOI: 10.1016/j.cbpa.2012.10.034;

[214] Kanamori, T.; Nishikawa, S.; Shin, I.; Schultz, P. G.; Endo, T. Probing the environment along the protein import pathways in yeast mitochondria by site-specific photocrosslinking. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 485–490. DOI: 10.1073/pnas.94.2.485;

[215] Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichiacoli. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 11020–11024. DOI: 10.1073/pnas.172226299.

[216] Dziuba, D.; Hoffmann, J.-E.; Hentze, M. W.; Schultz, C. A Genetically Encoded Diazirine Analogue for RNA-Protein Photo-crosslinking. *Chembiochem : a European journal of chemical biology* **2020**, *21*, 88–93. DOI: 10.1002/cbic.201900559.

[217] Reille-Seroussi, M.; Mayer, S. V.; Dörner, W.; Lang, K.; Mootz, H. D. Expanding the genetic code with a lysine derivative bearing an enzymatically removable phenylacetyl group. *Chemical communications (Cambridge, England)* **2019**, *55*, 4793–4796. DOI: 10.1039/c9cc00475k.

[218] Summerer, D.; Chen, S.; Wu, N.; Deiters, A.; Chin, J. W.; Schultz, P. G. A genetically encoded fluorescent amino acid. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 9785–9789. DOI: 10.1073/pnas.0603965103.

[219] Wang, J.; Xie, J.; Schultz, P. G. A genetically encoded fluorescent amino acid. *Journal of the American Chemical Society* **2006**, *128*, 8738–8739. DOI: 10.1021/ja062666k;

[220] Cowie, D. B.; Cohen, G. N. Biosynthesis by Escherichia coli of active altered proteins containing selenium instead of sulfur. *Biochimica et biophysica acta* **1957**, *26*, 252–261. DOI: 10.1016/0006-3002(57)90003-3.

[221] Yang, W.; Hendrickson, W. A.; Kalman, E. T.; Crouch, R. J. Expression, purification, and crystallization of natural and selenomethionyl recombinant ribonuclease H from Escherichia coli. *The Journal of biological chemistry* **1990**, *265*, 13553–13559.

[222] Yang, W.; Hendrickson, W. A.; Crouch, R. J.; Satow, Y. Structure of ribonuclease H phased at 2 A resolution by MAD analysis of the selenomethionyl protein. *Science (New York, N.Y.)* **1990**, *249*, 1398–1405. DOI: 10.1126/science.2169648;

[223] van Hest, J. C.M.; Tirrell, D. A. Efficient introduction of alkene functionality into proteins in vivo. *FEBS Letters* **1998**, *428*, 68–70. DOI: 10.1016/s0014-5793(98)00489-x.

[224] Lehner, F.; Kudlinzki, D.; Richter, C.; Müller-Werkmeister, H. M.; Eberl, K. B.; Bredenbeck, J.; Schwalbe, H.; Silvers, R. Impact of Azidohomoalanine Incorporation on Protein Structure and Ligand Binding. *Chembiochem : a European journal of chemical biology* **2017**, *18*, 2340–2350. DOI: 10.1002/cbic.201700437;

[225] Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 19–24. DOI: 10.1073/pnas.012583299;

[226] Budisa, N.; Steipe, B.; Demange, P.; Eckerskorn, C.; Kellermann, J.; Huber, R. High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic

acid, selenomethionine, telluromethionine and ethionine in Escherichia coli. *European journal of biochemistry* **1995**, *230*, 788–796. DOI: 10.1111/j.1432-1033.1995.tb20622.x.

[227] Link, A. J.; Mock, M. L.; Tirrell, D. A. Non-canonical amino acids in protein engineering. *Current opinion in biotechnology* **2003**, *14*, 603–609. DOI: 10.1016/j.copbio.2003.10.011.

[228] Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A. Biosynthesis of Proteins Incorporating a Versatile Set of Phenylalanine Analogues. *Chembiochem : a European journal of chemical biology* **2002**, *3*, 235–237. DOI: 10.1002/1439-7633(20020301)3:2/3<235:AID-CBIC235>3.0.CO;2-7.

[229] Pratt, E. A.; Ho, C. Incorporation of fluorotryptophans into proteins of escherichia coli. *Biochemistry* **1975**, *14*, 3035–3040. DOI: 10.1021/bi00684a037.

[230] Budisa, N.; Rubini, M.; Bae, J. H.; Weyher, E.; Wenger, W.; Golbik, R.; Huber, R.; Moroder, L. Global Replacement of Tryptophan with Aminotryptophans Generates Non-Invasive Protein-Based Optical pH Sensors. *Angew. Chem. Int. Ed.* **2002**, *41*, 4066–4069. DOI: 10.1002/1521-3773(20021104)41:21<4066:AID-ANIE4066>3.0.CO;2-6.

[231] Muir, T. W. Semisynthesis of proteins by expressed protein ligation. *Annual review of biochemistry* **2003**, *72*, 249–289. DOI: 10.1146/annurev.biochem.72.121801.161900;

[232] Muir, T. W.; Sondhi, D.; Cole, P. A. Expressed protein ligation: a general method for protein engineering. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 6705–6710. DOI: 10.1073/pnas.95.12.6705;

[233] Hackenberger, C. P. R.; Schwarzer, D. Chemoselective ligation and modification strategies for peptides and proteins. *Angew. Chem. Int. Ed.* **2008**, *47*, 10030–10074. DOI: 10.1002/anie.200801313.

[234] Flavell, R. R.; Muir, T. W. Expressed protein ligation (EPL) in the study of signal transduction, ion conduction, and chromatin biology. *Accounts of chemical research* **2009**, *42*, 107–116. DOI: 10.1021/ar800129c.

[235] Liu, C. C.; Schultz, P. G. Adding new chemistries to the genetic code. *Annual review of biochemistry* **2010**, *79*, 413–444. DOI: 10.1146/annurev.biochem.052308.105824.

[236] Leinfelder, W.; Zehelein, E.; Mandrand-Berthelot, M. A.; Böck, A. Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. *Nature* **1988**, *331*, 723–725. DOI: 10.1038/331723a0.

[237] Robertson, S. A.; Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. The use of 5'-phospho-2 deoxyribocytidylylriboadenosine as a facile route to chemical aminoacylation of tRNA. *Nucleic acids research* **1989**, *17*, 9649–9660. DOI: 10.1093/nar/17.23.9649.

[238] Wan, W.; Huang, Y.; Wang, Z.; Russell, W. K.; Pai, P.-J.; Russell, D. H.; Liu, W. R. A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli. *Angew. Chem. Int. Ed.* **2010**, *49*, 3211–3214. DOI: 10.1002/anie.201000465.

[239] Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science (New York, N.Y.)* **1989**, *244*, 182–188. DOI: 10.1126/science.2649980;

[240] Bain, J. D.; Diala, E. S.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R. Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide. *J. Am. Chem. Soc.* **1989**, *111*, 8013–8014. DOI: 10.1021/ja00202a052.

[241] Wang, L.; Xie, J.; Schultz, P. G. Expanding the genetic code. *Annual review of biophysics and biomolecular structure* **2006**, *35*, 225–249. DOI: 10.1146/annurev.biophys.35.101105.121507.

[242] Bossi, L.; Roth, J. R. Four-base codons ACCA, ACCU and ACCC are recognized by frameshift suppressor sufJ. *Cell* **1981**, *25*, 489–496. DOI: 10.1016/0092-8674(81)90067-2.

[243] Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **2010**, *464*, 441–444. DOI: 10.1038/nature08817.

[244] Hohsaka, T.; Ashizuka, Y.; Taira, H.; Murakami, H.; Sisido, M. Incorporation of nonnatural amino acids into proteins by using various four-base codons in an Escherichia coli in vitro translation system. *Biochemistry* **2001**, *40*, 11060–11064. DOI: 10.1021/bi0108204.

[245] Hohsaka, T.; Sisido, M. Incorporation of non-natural amino acids into proteins. *Current opinion in chemical biology* **2002**, *6*, 809–815. DOI: 10.1016/S1367-5931(02)00376-9.

[246] Rodriguez, E. A.; Lester, H. A.; Dougherty, D. A. In vivo incorporation of multiple unnatural amino acids through nonsense and frameshift suppression. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 8650–8655. DOI: 10.1073/pnas.0510817103.

[247] Anderson, J. C.; Wu, N.; Santoro, S. W.; Lakshman, V.; King, D. S.; Schultz, P. G. An expanded genetic code with a functional quadruplet codon. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 7566–7571. DOI: 10.1073/pnas.0401517101.

[248] CRICK, F. H. On protein synthesis. *Symposia of the Society for Experimental Biology* **1958**, *12*, 138–163.

[249] Chapeville, F.; Lipmann, F.; Ehrenstein, G. von; Weisblum, B.; Ray, W. J.; Benzer, S. On the role of soluble ribonucleic acid in coding for amino acids. *Proceedings of the National Academy of Sciences of the United States of America* **1962**, *48*, 1086–1092. DOI: 10.1073/pnas.48.6.1086;

[250] Heckler, T. G.; Chang, L. H.; Zama, Y.; Naka, T.; Chorghade, M. S.; Hecht, S. M. T4 RNA ligase mediated preparation of novel "chemically misacylated" tRNAPheS. *Biochemistry* **1984**, *23*, 1468–1473. DOI: 10.1021/bi00302a020.

[251] Hendrickson, T. L.; Crécy-Lagard, V. de; Schimmel, P. Incorporation of nonnatural amino acids into proteins. *Annual review of biochemistry* **2004**, *73*, 147–176. DOI: 10.1146/annurev.biochem.73.012803.092429.

[252] Mendel, D.; Cornish, V. W.; Schultz, P. G. Site-directed mutagenesis with an expanded genetic code. *Annual review of biophysics and biomolecular structure* **1995**, *24*, 435–462. DOI: 10.1146/annurev.bb.24.060195.002251.

[253] Normanly, J.; Kleina, L. G.; Masson, J.-M.; Abelson, J.; Miller, J. H. Construction of Escherichia coli amber suppressor tRNA genes. *Journal of molecular biology* **1990**, *213*, 719–726. DOI: 10.1016/S0022-2836(05)80258-X.

[254] Cload, S. T.; Liu, D. R.; Froland, W. A.; Schultz, P. G. Development of improved tRNAs for in vitro biosynthesis of proteins containing unnatural amino acids. *Chemistry & biology* **1996**, *3*, 1033–1038. DOI: 10.1016/s1074-5521(96)90169-6.

[255] Saks, M. E.; Sampson, J. R.; Nowak, M. W.; Kearney, P. C.; Du, F.; Abelson, J. N.; Lester, H. A.; Dougherty, D. A. An engineered Tetrahymena tRNAGln for in vivo incorporation of unnatural amino acids into proteins by nonsense suppression. *The Journal of biological chemistry* **1996**, *271*, 23169–23175. DOI: 10.1074/jbc.271.38.23169.

[256] Nowak, M. W.; Kearney, P. C.; Sampson, J. R.; Saks, M. E.; Labarca, C. G.; Silverman, S. K.; Zhong, W.; Thorson, J.; Abelson, J. N.; Davidson, N. Nicotinic receptor binding site probed with unnatural amino acid incorporation in intact cells. *Science (New York, N.Y.)* **1995**, *268*, 439–442. DOI: 10.1126/science.7716551.

[257] Beene, D. L.; Dougherty, D. A.; Lester, H. A. Unnatural amino acid mutagenesis in mapping ion channel function. *Current Opinion in Neurobiology* **2003**, *13*, 264–270. DOI: 10.1016/S0959-4388(03)00068-0.

[258] Liu, D. R.; Magliery, T. J.; Schultz, P. G. Characterization of an 'orthogonal' suppressor tRNA derived from E. coli tRNA2Gln. *Chemistry & biology* **1997**, *4*, 685–691. DOI: 10.1016/s1074-5521(97)90224-6.

[259] Liu, D. R.; Magliery, T. J.; Pastrnak, M.; Schultz, P. G. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 10092–10097. DOI: 10.1073/pnas.94.19.10092.

[260] Liu, D. R.; Schultz, P. G. Progress toward the evolution of an organism with an expanded genetic code. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 4780–4785. DOI: 10.1073/pnas.96.9.4780.

[261] Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *J. Am. Chem. Soc.* **2002**, *124*, 9026–9027. DOI: 10.1021/ja027007w.

[262] Chen, P. R.; Groff, D.; Guo, J.; Ou, W.; Cellitti, S.; Geierstanger, B. H.; Schultz, P. G. A facile system for encoding unnatural amino acids in mammalian cells. *Angew. Chem. Int. Ed.* **2009**, *48*, 4052–4055. DOI: 10.1002/anie.200900683.

[263] Nguyen, D. P.; Lusic, H.; Neumann, H.; Kapadnis, P. B.; Deiters, A.; Chin, J. W. Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA Synthetase/tRNA(CUA) pair and click chemistry. *J. Am. Chem. Soc.* **2009**, *131*, 8720–8721. DOI: 10.1021/ja900553w.

[264] Tharp, J. M.; Ad, O.; Amikura, K.; Ward, F. R.; Garcia, E. M.; Cate, J. H. D.; Schepartz, A.; Söll, D. Initiation of Protein Synthesis with Non-Canonical Amino Acids In Vivo. *Angew. Chem. Int. Ed.* **2020**, *59*, 3122–3126. DOI: 10.1002/anie.201914671.

[265] Ryu, Y.; Schultz, P. G. Efficient incorporation of unnatural amino acids into proteins in Escherichia coli. *Nature methods* **2006**, *3*, 263–265. DOI: 10.1038/nmeth864.

[266] Chatterjee, A.; Sun, S. B.; Furman, J. L.; Xiao, H.; Schultz, P. G. A versatile platform for single- and multiple-unnatural amino acid mutagenesis in Escherichia coli. *Biochemistry* **2013**, *52*, 1828–1837. DOI: 10.1021/bi4000244.

[267] Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. An enhanced system for unnatural amino acid mutagenesis in E. coli. *Journal of molecular biology* **2010**, *395*, 361–374. DOI: 10.1016/j.jmb.2009.10.030.

[268] Guo, J.; Melançon, C. E.; Lee, H. S.; Groff, D.; Schultz, P. G. Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids. *Angew. Chem. Int. Ed.* **2009**, *48*, 9148–9151. DOI: 10.1002/anie.200904035.

[269] Ohtsuki, T.; Yamamoto, H.; Doi, Y.; Sisido, M. Use of EF-Tu mutants for determining and improving aminoacylation efficiency and for purifying aminoacyl tRNAs with non-natural amino acids. *Journal of biochemistry* **2010**, *148*, 239–246. DOI: 10.1093/jb/mvq053.

[270] Gan, R.; Perez, J. G.; Carlson, E. D.; Ntai, I.; Isaacs, F. J.; Kelleher, N. L.; Jewett, M. C. Translation system engineering in Escherichia coli enhances non-canonical amino acid incorporation into proteins. *Biotechnology and bioengineering* **2017**, *114*, 1074–1086. DOI: 10.1002/bit.26239.

[271] Smolskaya, S.; Andreev, Y. A. Site-Specific Incorporation of Unnatural Amino Acids into Escherichia coli Recombinant Protein: Methodology Development and Recent Achievement. *Biomolecules* **2019**, *9*. DOI: 10.3390/biom9070255.

[272] Wang, K.; Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nature biotechnology* **2007**, *25*, 770–777. DOI: 10.1038/nbt1314.

[273] Lajoie, M. J.; Rovner, A. J.; Goodman, D. B.; Aerni, H.-R.; Haimovich, A. D.; Kuznetsov, G.; Mercer, J. A.; Wang, H. H.; Carr, P. A.; Mosberg, J. A.*; et al.* Genomically recoded organisms expand biological functions. *Science (New York, N.Y.)* **2013**, *342*, 357–360. DOI: 10.1126/science.1241459.

[274] Lesser, C. F.; Guthrie, C. Mutational analysis of pre-mRNA splicing in Saccharomyces cerevisiae using a sensitive new reporter gene, CUP1. *Genetics* **1993**, *133*, 851–863.

[275] Korostelev, A.; Asahara, H.; Lancaster, L.; Laurberg, M.; Hirschi, A.; Zhu, J.; Trakhanov, S.; Scott, W. G.; Noller, H. F. Crystal structure of a translation termination complex formed with release factor RF2. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 19684–19689. DOI: 10.1073/pnas.0810953105.

[276] Karijolich, J.; Yu, Y.-T. Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature* **2011**, *474*, 395–398. DOI: 10.1038/nature10165.

[277] Huisgen, R. *1,3-dipolar cycloaddition chemistry*; Wiley: New York, NY, 1984.

[278] Huisgen, R. Kinetics and reaction mechnanisms: selected examples from the experience of forty years. *Pure & Appl. Chem.* **1989**, *61*, 613–628.

[279] Huisgen, R. 1,3-Dipolare Cycloadditionen Rückschau und Ausblick. *Angewandte Chemie* **1963**, *75*, 604.

[280] Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021. DOI: 10.1002/1521-3773(20010601)40:11<2004:aid-anie2004>3.3.co;2-x.

[281] Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Ed.* **2002**, *41*, 2596–2599. DOI: 10.1002/1521-3773(20020715)41:14<2596:AID-ANIE2596>3.0.CO;2-4.

[282] Tornøe, C. W.; Christensen, C.; Meldal, M. Peptidotriazoles on solid phase: 1,2,3-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *The Journal of organic chemistry* **2002**, *67*, 3057–3064. DOI: 10.1021/jo011148j.

[283] Presolski, S. I.; Hong, V. P.; Finn, M. G. Copper-Catalyzed Azide-Alkyne Click Chemistry for Bioconjugation. *Current protocols in chemical biology* **2011**, *3*, 153–162. DOI: 10.1002/9780470559277.ch110148.

[284] Zhang, W. H.; Otting, G.; Jackson, C. J. Protein engineering with unnatural amino acids. *Current opinion in structural biology* **2013**, *23*, 581–587. DOI: 10.1016/j.sbi.2013.06.009.

[285] Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. A strain-promoted 3 + 2 azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **2004**, *126*, 15046–15047. DOI: 10.1021/ja044996f.

[286] Evans, H. L.; Slade, R. L.; Carroll, L.; Smith, G.; Nguyen, Q.-D.; Iddon, L.; Kamaly, N.; Stöckmann, H.; Leeper, F. J.; Aboagye, E. O.*; et al.* Copper-free click--a promising tool for pre-targeted PET imaging. *Chemical communications (Cambridge, England)* **2012**, *48*, 991–993. DOI: 10.1039/c1cc16220a.

[287] Laughlin, S. T.; Baskin, J. M.; Amacher, S. L.; Bertozzi, C. R. In vivo imaging of membrane-associated glycans in developing zebrafish. *Science (New York, N.Y.)* **2008**, *320*, 664–667. DOI: 10.1126/science.1155106;

[288] Neef, A. B.; Schultz, C. Selective fluorescence labeling of lipids in living cells. *Angew. Chem. Int. Ed.* **2009**, *48*, 1498–1500. DOI: 10.1002/anie.200805507.

[289] Nikić, I.; Kang, J. H.; Girona, G. E.; Aramburu, I. V.; Lemke, E. A. Labeling proteins on live mammalian cells using click chemistry. *Nature protocols* **2015**, *10*, 780–791. DOI: 10.1038/nprot.2015.045.

[290] Chang, P. V.; Prescher, J. A.; Sletten, E. M.; Baskin, J. M.; Miller, I. A.; Agard, N. J.; Lo, A.; Bertozzi, C. R. Copper-free click chemistry in living animals. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 1821–1826. DOI: 10.1073/pnas.0911116107.

[291] Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V.; Miller, I. A.; Lo, A.; Codelli, J. A.; Bertozzi, C. R. Copper-free click chemistry for dynamic in vivo imaging. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 16793–16797. DOI: 10.1073/pnas.0707090104.

[292] Atkins, P. W.; Paula, J. de. *Physikalische Chemie*, 4., vollst. überarb. Aufl.; Wiley-VCH: Weinheim, 2006.

[293] Calvert, J. G.; Pitts, J. N. *Photochemistry*; Wiley: New York, 1966.

[294] Eisenberg, D.; Crothers, D. *Physical chemistry with applications to the life sciences*; Benjamin/Cummings Publ: Menlo Park, Calif., 1979.

[295] Naik, A. K.; Hanay, M. S.; Hiebert, W. K.; Feng, X. L.; Roukes, M. L. Towards single-molecule nanomechanical mass spectrometry. *Nature nanotechnology* **2009**, *4*, 445–450. DOI: 10.1038/nnano.2009.152.

[296] Robertson, J. W. F.; Rodrigues, C. G.; Stanford, V. M.; Rubinson, K. A.; Krasilnikov, O. V.; Kasianowicz, J. J. Single-molecule mass spectrometry in solution using a solitary nanopore. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 8207–8211. DOI: 10.1073/pnas.0611085104.

[297] Francis, L. W.; Lewis, P. D.; Wright, C. J.; Conlan, R. S. Atomic force microscopy comes of age. *Biology of the cell* **2009**, *102*, 133–143. DOI: 10.1042/BC20090127.

[298] Neuman, K. C.; Nagy, A. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nature methods* **2008**, *5*, 491–505. DOI: 10.1038/nmeth.1218.

[299] Li, P. T. X.; Vieregg, J.; Tinoco, I. How RNA unfolds and refolds. *Annual review of biochemistry* **2008**, *77*, 77–100. DOI: 10.1146/annurev.biochem.77.061206.174353.

[300] Tinoco, I.; Gonzalez, R. L. Biological mechanisms, one molecule at a time. *Genes & development* **2011**, *25*, 1205–1231. DOI: 10.1101/gad.2050011.

[301] Tinoco, I.; Li, P. T. X.; Bustamante, C. Determination of thermodynamics and kinetics of RNA reactions by force. *Quarterly reviews of biophysics* **2006**, *39*, 325–360. DOI: 10.1017/S0033583506004446.

[302] Shank, E. A.; Cecconi, C.; Dill, J. W.; Marqusee, S.; Bustamante, C. The folding cooperativity of a protein is controlled by its chain topology. *Nature* **2010**, *465*, 637–640. DOI: 10.1038/nature09021.

[303] Galburt, E. A.; Grill, S. W.; Wiedmann, A.; Lubkowska, L.; Choy, J.; Nogales, E.; Kashlev, M.; Bustamante, C. Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* **2007**, *446*, 820–823. DOI: 10.1038/nature05701.

[304] Wen, J.-D.; Lancaster, L.; Hodges, C.; Zeri, A.-C.; Yoshimura, S. H.; Noller, H. F.; Bustamante, C.; Tinoco, I. Following translation by single ribosomes one codon at a time. *Nature* **2008**, *452*, 598–603. DOI: 10.1038/nature06716.

[305] Ha, T.; Enderle, T.; Ogletree, D. F.; Chemla, D. S.; Selvin, P. R.; Weiss, S. Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proceedings of the National Academy of Sciences of the United States of America* **1996**, *93*, 6264–6268. DOI: 10.1073/pnas.93.13.6264.

[306] Moerner, W. E.; Fromm, D. P. Methods of single-molecule fluorescence spectroscopy and microscopy. *Review of Scientific Instruments* **2003**, *74*, 3597–3619. DOI: 10.1063/1.1589587.

[307] Roy, R.; Hohng, S.; Ha, T. A practical guide to single-molecule FRET. *Nature methods* **2008**, *5*, 507–516. DOI: 10.1038/nmeth.1208.

[308] Buschmann, V.; Weston, K. D.; Sauer, M. Spectroscopic study and evaluation of red-absorbing fluorescent dyes. *Bioconjugate chemistry* **2003**, *14*, 195–204. DOI: 10.1021/bc025600x.

[309] Widengren, J.; Schwille, P. Characterization of Photoinduced Isomerization and Back-Isomerization of the Cyanine Dye Cy5 by Fluorescence Correlation Spectroscopy. *J. Phys. Chem. A* **2000**, *104*, 6416–6428. DOI: 10.1021/jp000059s.

[310] Hanspach, G.; Trucks, S.; Hengesbach, M. Strategic labelling approaches for RNA single-molecule spectroscopy. *RNA biology* **2019**, *16*, 1119–1132. DOI: 10.1080/15476286.2019.1593093.

[311] Hanne, J.; Zila, V.; Heilemann, M.; Müller, B.; Kräusslich, H.-G. Super-resolved insights into human immunodeficiency virus biology. *FEBS Letters* **2016**, *590*, 1858–1876. DOI: 10.1002/1873-3468.12186.

[312] Huang, B.; Bates, M.; Zhuang, X. Super-resolution fluorescence microscopy. *Annual review of biochemistry* **2009**, *78*, 993–1016. DOI: 10.1146/annurev.biochem.77.061906.092014.

[313] Dyba, M.; Hell, S. W. Focal spots of size lambda/23 open up far-field fluorescence microscopy at 33 nm axial resolution. *Physical review letters* **2002**, *88*, 163901. DOI: 10.1103/PhysRevLett.88.163901.

[314] Hell, S. W.; Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics letters* **1994**, *19*, 780–782. DOI: 10.1364/ol.19.000780.

[315] Klar, T. A.; Hell, S. W. Subdiffraction resolution in far-field fluorescence microscopy. *Optics letters* **1999**, *24*, 954–956. DOI: 10.1364/ol.24.000954.

[316] Sydor, A. M.; Czymmek, K. J.; Puchner, E. M.; Mennella, V. Super-Resolution Microscopy: From Single Molecules to Supramolecular Assemblies. *Trends in cell biology* **2015**, *25*, 730–748. DOI: 10.1016/j.tcb.2015.10.004.

[317] Donnert, G.; Keller, J.; Medda, R.; Andrei, M. A.; Rizzoli, S. O.; Lührmann, R.; Jahn, R.; Eggeling, C.; Hell, S. W. Macromolecular-scale resolution in biological fluorescence microscopy. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 11440–11445. DOI: 10.1073/pnas.0604965103.

[318] Willig, K. I.; Kellner, R. R.; Medda, R.; Hein, B.; Jakobs, S.; Hell, S. W. Nanoscale resolution in GFP-based microscopy. *Nature methods* **2006**, *3*, 721–723. DOI: 10.1038/nmeth922.

[319] Hess, S. T.; Girirajan, T. P. K.; Mason, M. D. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical journal* **2006**, *91*, 4258–4272. DOI: 10.1529/biophysj.106.091116.

[320] Betzig, E.; Patterson, G. H.; Sougrat, R.; Lindwasser, O. W.; Olenych, S.; Bonifacino, J. S.; Davidson, M. W.; Lippincott-Schwartz, J.; Hess, H. F. Imaging intracellular fluorescent proteins at nanometer resolution. *Science (New York, N.Y.)* **2006**, *313*, 1642–1645. DOI: 10.1126/science.1127344.

[321] Rust, M. J.; Bates, M.; Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods* **2006**, *3*, 793–795. DOI: 10.1038/nmeth929.

[322] Kim, J.-Y.; Kim, C.; Lee, N. K. Real-time submillisecond single-molecule FRET dynamics of freely diffusing molecules with liposome tethering. *Nature communications* **2015**, *6*, 6992. DOI: 10.1038/ncomms7992.

[323] Axelrod, D. Total internal reflection fluorescence microscopy in cell biology. *Methods in enzymology* **2003**, *361*, 1–33. DOI: 10.1016/s0076-6879(03)61003-7.

[324] Ha, T.; Rasnik, I.; Cheng, W.; Babcock, H. P.; Gauss, G. H.; Lohman, T. M.; Chu, S. Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase. *Nature* **2002**, *419*, 638–641. DOI: 10.1038/nature01083.

[325] Sofia; Premnath; Merrill. Poly(ethylene oxide) Grafted to Silicon Surfaces: Grafting Density and Protein Adsorption. *Macromolecules* **1998**, *31*, 5059–5070. DOI: 10.1021/ma971016l.

[326] Andreou, A. Z.; Harms, U.; Klostermeier, D. Single-stranded regions modulate conformational dynamics and ATPase activity of eIF4A to optimize 5'-UTR unwinding. *Nucleic acids research* **2019**, *47*, 5260–5275. DOI: 10.1093/nar/gkz254.

[327] Hengesbach, M.; Kim, N.-K.; Feigon, J.; Stone, M. D. Single-molecule FRET reveals the folding dynamics of the human telomerase RNA pseudoknot domain. *Angew. Chem. Int. Ed.* **2012**, *51*, 5876–5879. DOI: 10.1002/anie.201200526.

[328] Kramm, K.; Endesfelder, U.; Grohmann, D. A Single-Molecule View of Archaeal Transcription. *Journal of molecular biology* **2019**, *431*, 4116–4131. DOI: 10.1016/j.jmb.2019.06.009.

[329] Schuler, B.; Soranno, A.; Hofmann, H.; Nettels, D. Single-Molecule FRET Spectroscopy and the Polymer Physics of Unfolded and Intrinsically Disordered Proteins. *Annual review of biophysics* **2016**, *45*, 207–231. DOI: 10.1146/annurev-biophys-062215-010915.

[330] Meiser, N.; Fuks, C.; Hengesbach, M. Cooperative Analysis of Structural Dynamics in RNA-Protein Complexes by Single-Molecule Förster Resonance Energy Transfer Spectroscopy. *Molecules (Basel, Switzerland)* **2020**, *25*. DOI: 10.3390/molecules25092057.

[331] Blanco, M. R.; Martin, J. S.; Kahlscheuer, M. L.; Krishnan, R.; Abelson, J.; Laederach, A.; Walter, N. G. Single Molecule Cluster Analysis dissects splicing pathway conformational dynamics. *Nature methods* **2015**, *12*, 1077–1084. DOI: 10.1038/nmeth.3602.

[332] Beier, D. H.; Carrocci, T. J.; van der Feltz, C.; Tretbar, U. S.; Paulson, J. C.; Grabowski, N.; Hoskins, A. A. Dynamics of the DEAD-box ATPase Prp5 RecA-like domains provide a conformational switch during spliceosome assembly. *Nucleic acids research* **2019**, *47*, 10842–10851. DOI: 10.1093/nar/gkz765.

[333] Tippana, R.; Hwang, H.; Opresko, P. L.; Bohr, V. A.; Myong, S. Single-molecule imaging reveals a common mechanism shared by G-quadruplex-resolving helicases. *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113*, 8448–8453. DOI: 10.1073/pnas.1603724113.

[334] Chen, W.-F.; Rety, S.; Guo, H.-L.; Dai, Y.-X.; Wu, W.-Q.; Liu, N.-N.; Auguin, D.; Liu, Q.-W.; Hou, X.-M.; Dou, S.-X.*; et al.* Molecular Mechanistic Insights into Drosophila DHX36-Mediated G-Quadruplex Unfolding: A Structure-Based Model. *Structure (London, England : 1993)* **2018**, *26*, 403-415.e4. DOI: 10.1016/j.str.2018.01.008.

[335] Parks, J. W.; Kappel, K.; Das, R.; Stone, M. D. Single-molecule FRET-Rosetta reveals RNA structural rearrangements during human telomerase catalysis. *RNA (New York, N.Y.)* **2017**, *23*, 175–188. DOI: 10.1261/rna.058743.116.

[336] Bhattacharjee, A.; Wang, Y.; Diao, J.; Price, C. M. Dynamic DNA binding, junction recognition and G4 melting activity underlie the telomeric and genome-wide roles of human CST. *Nucleic acids research* **2017**, *45*, 12311–12324. DOI: 10.1093/nar/gkx878.

[337] Lai, W.-J. C.; Ermolenko, D. N. Ensemble and single-molecule FRET studies of protein synthesis. *Methods (San Diego, Calif.)* **2018**, *137*, 37–48. DOI: 10.1016/j.ymeth.2017.12.007.

[338] Munro, J. B.; Altman, R. B.; O'Connor, N.; Blanchard, S. C. Identification of two distinct hybrid state intermediates on the ribosome. *Molecular cell* **2007**, *25*, 505–517. DOI: 10.1016/j.molcel.2007.01.022.

[339] Prokhorova, I.; Altman, R. B.; Djumagulov, M.; Shrestha, J. P.; Urzhumtsev, A.; Ferguson, A.; Chang, C.-W. T.; Yusupov, M.; Blanchard, S. C.; Yusupova, G. Aminoglycoside interactions and impacts on the eukaryotic ribosome. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114*, E10899-E10908. DOI: 10.1073/pnas.1715501114.

[340] Wang, L.; Wasserman, M. R.; Feldman, M. B.; Altman, R. B.; Blanchard, S. C. Mechanistic insights into antibiotic action on the ribosome through single-molecule fluorescence imaging. *Annals of the New York Academy of Sciences* **2011**, *1241*, E1-16. DOI: 10.1111/j.1749-6632.2012.06839.x.

[341] Cui, T. J.; Klein, M.; Hegge, J. W.; Chandradoss, S. D.; van der Oost, J.; Depken, M.; Joo, C. Argonaute bypasses cellular obstacles without hindrance during target search. *Nature communications* **2019**, *10*, 4390. DOI: 10.1038/s41467-019-12415-y.

[342] Warnasooriya, C.; Ling, C.; Belashov, I. A.; Salim, M.; Wedekind, J. E.; Ermolenko, D. N. Observation of preQ1-II riboswitch dynamics using single-molecule FRET. *RNA biology* **2019**, *16*, 1086–1092. DOI: 10.1080/15476286.2018.1536591.

[343] Bains, J. K.; Blechar, J.; Jesus, V. de; Meiser, N.; Zetzsche, H.; Fürtig, B.; Schwalbe, H.; Hengesbach, M. Combined smFRET and NMR analysis of riboswitch structural dynamics. *Methods (San Diego, Calif.)* **2019**, *153*, 22–34. DOI: 10.1016/j.ymeth.2018.10.004.

[344] Li, Z.; Li, W.; Lu, M.; Bess, J.; Chao, C. W.; Gorman, J.; Terry, D. S.; Zhang, B.; Zhou, T.; Blanchard, S. C.*; et al.* Subnanometer structures of HIV-1 envelope trimers on aldrithiol-2-inactivated virus particles. *Nature structural & molecular biology* **2020**, *27*, 726–734. DOI: 10.1038/s41594-020-0452-2.

[345] Lu, M.; Ma, X.; Castillo-Menendez, L. R.; Gorman, J.; Alsahafi, N.; Ermel, U.; Terry, D. S.; Chambers, M.; Peng, D.; Zhang, B.*; et al.* Associating HIV-1 envelope glycoprotein structures with states on the virus observed by smFRET. *Nature* **2019**, *568*, 415–419. DOI: 10.1038/s41586-019-1101-y.

[346] Fessl, T.; Adamec, F.; Polívka, T.; Foldynová-Trantírková, S.; Vácha, F.; Trantírek, L. Towards characterization of DNA structure under physiological conditions in vivo at the single-molecule level using single-pair FRET. *Nucleic acids research* **2012**, *40*, e121. DOI: 10.1093/nar/gks333.

[347] Sustarsic, M.; Kapanidis, A. N. Taking the ruler to the jungle: single-molecule FRET for understanding biomolecular structure and dynamics in live cells. *Current opinion in structural biology* **2015**, *34*, 52–59. DOI: 10.1016/j.sbi.2015.07.001.

[348] Clamme, J.-P.; Deniz, A. A. Three-color single-molecule fluorescence resonance energy transfer. *Chemphyschem : a European journal of chemical physics and physical chemistry* **2005**, *6*, 74–77. DOI: 10.1002/cphc.200400261.

[349] Heinze, K. G.; Jahnz, M.; Schwille, P. Triple-color coincidence analysis: one step further in following higher order molecular complex formation. *Biophysical journal* **2004**, *86*, 506–516. DOI: 10.1016/S0006-3495(04)74129-6.

[350] Hohng, S.; Joo, C.; Ha, T. Single-molecule three-color FRET. *Biophysical journal* **2004**, *87*, 1328–1337. DOI: 10.1529/biophysj.104.043935.

[351] Benesch, R. E.; Benesch, R. Enzymatic removal of oxygen for polarography and related methods. *Science (New York, N.Y.)* **1953**, *118*, 447–448. DOI: 10.1126/science.118.3068.447.

[352] Rasnik, I.; McKinney, S. A.; Ha, T. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nature methods* **2006**, *3*, 891–893. DOI: 10.1038/nmeth934.

[353] Orte, A.; Clarke, R.; Balasubramanian, S.; Klenerman, D. Determination of the fraction and stoichiometry of femtomolar levels of biomolecular complexes in an excess of monomer using single-molecule, two-color coincidence detection. *Analytical chemistry* **2006**, *78*, 7707–7715. DOI: 10.1021/ac061122y.

[354] Ren, X.; Li, H.; Clarke, R. W.; Alves, D. A.; Ying, L.; Klenerman, D.; Balasubramanian, S. Analysis of human telomerase activity and function by two color single molecule coincidence fluorescence spectroscopy. *J. Am. Chem. Soc.* **2006**, *128*, 4992–5000. DOI: 10.1021/ja056613z.

[355] Li, H.; Ying, L.; Green, J. J.; Balasubramanian, S.; Klenerman, D. Ultrasensitive coincidence fluorescence detection of single DNA molecules. *Analytical chemistry* **2003**, *75*, 1664–1670. DOI: 10.1021/ac026367z.

[356] Ma, X.; Yang, C.; Alexandrov, A.; Grayhack, E. J.; Behm-Ansmant, I.; Yu, Y.-T. Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism. *The EMBO journal* **2005**, *24*, 2403–2413. DOI: 10.1038/sj.emboj.7600718.

[357] Schattner, P.; Decatur, W. A.; Davis, C. A.; Ares, M.; Fournier, M. J.; Lowe, T. M. Genome-wide searching for pseudouridylation guide snoRNAs: analysis of the Saccharomyces cerevisiae genome. *Nucleic acids research* **2004**, *32*, 4281–4296. DOI: 10.1093/nar/gkh768.

[358] Bagni, C.; Lapeyre, B. Gar1p binds to the small nucleolar RNAs snR10 and snR30 in vitro through a nontypical RNA binding element. *The Journal of biological chemistry* **1998**, *273*, 10868–10873. DOI: 10.1074/jbc.273.18.10868.

[359] Trucks, S. Einbau von nicht-natürlichen Aminosäuren in Proteine als spektroskopische Sonden. Master Thesis, Goethe-Universität, Frankfurt am Main, 2015.

[360] Schmidt, A. Investigation of structural dynamics of H/ACA RNPs by smFRET spectroscopy & Design and applications of versatile small mole-cule dependent RNA of interest release platform. Dissertation, Goethe-Universität, Frankfurt am Main, 2019.

[361] Schmidt, A.; Altincekic, N.; Gustmann, H.; Wachtveitl, J.; Hengesbach, M. The Protein Microenvironment Governs the Suitability of Labeling Sites for Single-Molecule Spectroscopy of RNP Complexes. *ACS chemical biology* **2018**, *13*, 2472–2483. DOI: 10.1021/acschembio.8b00348.

[362] Koo, B.-K.; Park, C.-J.; Fernandez, C. F.; Chim, N.; Ding, Y.; Chanfreau, G.; Feigon, J. Structure of H/ACA RNP protein Nhp2p reveals cis/trans isomerization of a conserved proline at the RNA and Nop10 binding interface. *Journal of molecular biology* **2011**, *411*, 927–942. DOI: 10.1016/j.jmb.2011.06.022.

[363] Kiledjian, M.; Dreyfuss, G. Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *The EMBO journal* **1992**, *11*, 2655–2664.

[364] Ghisolfi, L.; Joseph, G.; Amalric, F.; Erard, M. The glycine-rich domain of nucleolin has an unusual supersecondary structure responsible for its RNA-helix-destabilizing properties. *The Journal of biological chemistry* **1992**, *267*, 2955–2959.

[365] Ghisolfi, L.; Kharrat, A.; Joseph, G.; Amalric, F.; Erard, M. Concerted activities of the RNA recognition and the glycine-rich C-terminal domains of nucleolin are required for efficient complex formation with pre-ribosomal RNA. *European journal of biochemistry* **1992**, *209*, 541–548. DOI: 10.1111/j.1432-1033.1992.tb17318.x.

[366] Lee, W. C.; Xue, Z. X.; Mélèse, T. The NSR1 gene encodes a protein that specifically binds nuclear localization sequences and has two RNA recognition motifs. *The Journal of cell biology* **1991**, *113*, 1–12. DOI: 10.1083/jcb.113.1.1.

[367] Jong, A. Y.; Clark, M. W.; Gilbert, M.; Oehm, A.; Campbell, J. L. Saccharomyces cerevisiae SSB1 protein and its relationship to nucleolar RNA-binding proteins. *Molecular and cellular biology* **1987**, *7*, 2947–2955. DOI: 10.1128/mcb.7.8.2947.

[368] Russell, I. D.; Tollervey, D. NOP3 is an essential yeast protein which is required for pre-rRNA processing. *The Journal of cell biology* **1992**, *119*, 737–747. DOI: 10.1083/jcb.119.4.737.

[369] Keith, G. Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. *Biochimie* **1995**, *77*, 142–144. DOI: 10.1016/0300-9084(96)88118-1.

[370] Lottspeich, F.; Engels, J. W., Eds. *Bioanalytik*, 2. Aufl.; Spektrum Akad. Verl.: München, 2006.

[371] Hanspach, G. Structure, Assembly and Dynamics of H/ACA-Ribonucleoproteins. Dissertation, Goethe-Universität, Frankfurt am Main, 2020.

[372] Henras, A. K.; 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 (New York, N.Y.)* **2004**, *10*, 1704–1712. DOI: 10.1261/rna.7770604.

[373] Jiang, W.; Middleton, K.; Yoon, H. J.; Fouquet, C.; Carbon, J. An essential yeast protein, CBF5p, binds in vitro to centromeres and microtubules. *Mol. Cell. Biol.* **1993**, *13*, 4884–4893. DOI: 10.1128/mcb.13.8.4884.

[374] Hamma, T.; Ferré-D'Amaré, A. R. The box H/ACA ribonucleoprotein complex: interplay of RNA and protein structures in post-transcriptional RNA modification. *The Journal of biological chemistry* **2010**, *285*, 805–809. DOI: 10.1074/jbc.R109.076893.

[375] Ballarino, M.; Morlando, M.; Pagano, F.; Fatica, A.; Bozzoni, I. The cotranscriptional assembly of snoRNPs controls the biosynthesis of H/ACA snoRNAs in Saccharomyces cerevisiae. *Molecular and cellular biology* **2005**, *25*, 5396–5403. DOI: 10.1128/MCB.25.13.5396-5403.2005.

[376] *OligoAnalyzer*; Integrated DNA Technologies (https://eu.idtdna.com/calc/analyzer, last visited 21.09.2020).

[377] Kelly, E. K.; Czekay, D. P.; Kothe, U. Base-pairing interactions between substrate RNA and H/ACA guide RNA modulate the kinetics of pseudouridylation, but not the affinity of substrate binding by H/ACA small nucleolar ribonucleoproteins. *RNA (New York, N.Y.)* **2019**, *25*, 1393–1404. DOI: 10.1261/rna.071043.119.

[378] Zoysa, M. D. de; Wu, G.; Katz, R.; Yu, Y.-T. Guide-substrate base-pairing requirement for box H/ACA RNA-guided RNA pseudouridylation. *RNA (New York, N.Y.)* **2018**, *24*, 1106–1117. DOI: 10.1261/rna.066837.118.

[379] Majumder, M.; Bosmeny, M. S.; Gupta, R. Structure-function relationships of archaeal Cbf5 during in vivo RNA-guided pseudouridylation. *RNA (New York, N.Y.)* **2016**, *22*, 1604–1619. DOI: 10.1261/rna.057547.116.

[380] Kalinin, S.; Peulen, T.; Sindbert, S.; Rothwell, P. J.; Berger, S.; Restle, T.; Goody, R. S.; Gohlke, H.; Seidel, C. A. M. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nature methods* **2012**, *9*, 1218–1225. DOI: 10.1038/NMETH.2222.

[381] McKinney, S. A.; Joo, C.; Ha, T. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophysical journal* **2006**, *91*, 1941–1951. DOI: 10.1529/biophysj.106.082487.

[382] Samuelsson, T. Interactions of transfer RNA pseudouridine synthases with RNAs substituted with fluorouracil. *Nucleic acids research* **1991**, *19*, 6139–6144. DOI: 10.1093/nar/19.22.6139.

[383] Zhao, X.; Yu, Y.-T. Incorporation of 5-fluorouracil into U2 snRNA blocks pseudouridylation and pre-mRNA splicing in vivo. *Nucleic acids research* **2007**, *35*, 550–558. DOI: 10.1093/nar/gkl1084.

[384] Polyhach, Y.; Bordignon, E.; Jeschke, G. Rotamer libraries of spin labelled cysteines for protein studies. *Physical chemistry chemical physics : PCCP* **2011**, *13*, 2356–2366. DOI: 10.1039/c0cp01865a.

[385] Nikić, I.; Plass, T.; Schraidt, O.; Szymański, J.; Briggs, J. A. G.; Schultz, C.; Lemke, E. A. Minimal tags for rapid dual-color live-cell labeling and super-resolution microscopy. *Angew. Chem. Int. Ed.* **2014**, *53*, 2245–2249. DOI: 10.1002/anie.201309847.

[386] Corey, E. J.; Weinshenker, N. M.; Schaaf, T. K.; Huber, W. Stereo-controlled synthesis of prostaglandins F-2a and E-2 (dl). *J. Am. Chem. Soc.* **1969**, *91*, 5675–5677. DOI: 10.1021/ja01048a062.

[387] Hagendorn, T.; Bräse, S. A Route to Cyclooct-2-ynol and Its Functionalization by Mitsunobu Chemistry. *Eur. J. Org. Chem.* **2014**, *2014*, 1280–1286. DOI: 10.1002/ejoc.201301375.

[388] Tharp, J. M.; Liu, W. R. Using Amber and Ochre Nonsense Codons to Code Two Different Noncanonical Amino Acids in One Protein Gene. *Methods in molecular biology (Clifton, N.J.)* **2018**, *1728*, 147–154. DOI: 10.1007/978-1-4939-7574-7_9.

[389] Murphy, M. C.; Rasnik, I.; Cheng, W.; Lohman, T. M.; Ha, T. Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophysical journal* **2004**, *86*, 2530–2537. DOI: 10.1016/S0006-3495(04)74308-8.

[390] Wolber, P. K.; Hudson, B. S. An analytic solution to the Förster energy transfer problem in two dimensions. *Biophysical journal* **1979**, *28*, 197–210. DOI: 10.1016/S0006-3495(79)85171-1.

[391] Wu, P.; Brand, L. Resonance energy transfer: methods and applications. *Analytical biochemistry* **1994**, *218*, 1–13. DOI: 10.1006/abio.1994.1134.

[392] Schiemann, O.; Prisner, T. F. Long-range distance determinations in biomacromolecules by EPR spectroscopy. *Quarterly reviews of biophysics* **2007**, *40*, 1–53. DOI: 10.1017/S003358350700460X.

[393] Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research* **2003**, *31*, 3784–3788. DOI: 10.1093/nar/gkg563.

List of abbreviations

2D TLC	two-dimensional thin layer chromatography
5FU	5-fluoro-uridine
A	adenosine
aaRS	aminoacyl tRNA synthetase
AcOH	acetic acid
AFM	atomic force microscopy
Amp	ampicillin
ATP	adenosine triphosphate
bp	base-pairs
C	Cbf5 (Centomere Binding Factor 5)
C	cytidine
c	concentration
CG	Cbf5-Gar1
CMP	chloramphenicol
CTP	cytidine triphosphate
CuAAC	copper(I)-catalyzed azide-alkyne cycloaddition
CV	Column Volumes
DC	Dykeratosis congenita
DCM	dichloromethane
DMF	dimethylformamide
DNA	deoxyribonucleic acid
DNP	dynamic nuclear polarization
DMSO	dimethylsulfoxide
E. coli	*Escherichia coli*
E-FRET	FRET efficiency
e.g.	exempli gratia (“for example”)
EDTA	ethylenediaminetetraacetic acid
EF-Tu	elongation factor Tu
EM	electron microscopy
EM-CCD	electron-multiplying charged couple device
EMSA	electrophoretic mobility shift assay
EPL	Expressed Protein Ligation
EPR	electron paramagnetic resonance
eq.	equivalent
ESI	Electrospray ionization
et al.	et alia (“and others”)
etc.	et cetera (“and so forth”)
FPALM	Fluorescence Photoactivation Localization Microscopy
FRET	Förster resonance energy transfer
FWHM	full width at half maximum
G	Gar1 (Glycine Arginine Rich 1)
G	guanosine

GTP	guanosine triphosphate
HPLC	high performance liquid chromatography
I	intensity
i.a.	inter alia (“among other things”)
i.e.	id est (“that is”)
IPTG	Isopropyl β-D-1-thiogalactopyranoside
λ	wavelength (nm)
K	lysine
K-	kink-
Kan	Kanamycin
K_d	dissociation constant
MALDI	Matrix-assisted laser desorption/ionization
MeOH	methanol
mRNA	messenger RNA
MTSSL	(1-Oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate
m/z	mass-to-charge ratio
N	Nop10 (Nucleolar Protein 10)
N	any nucleotide
NC	Nop10-Cbf5
NCG	Nop10-Cbf5-Gar1
NEB	New England Biolabs
NHS	*N*-Hydroxysuccinimide
Ni-NTA	nickel(II) nitrilotriacetic acid
NMR	nuclear magnetic resonance
NoK	norbornene-lysine
nt	nucleotides
NTP	nucleotide triphosphate
OD_{600}	optical density measured at λ = 600 nm
OTIR	objective type total internal reflection
Ψ	pseudouridine
PAA	polyacrylamide
PAGE	polyacrylamide gel electrophoresis
PALM	Phoactivated Localization Microscopy
PCR	polymerase chain reaction
PDB	protein database
PELDOR	Pulsed Electron Electron Double Resonance
P. furiosus / *P.f.*	*Pyrococcus furiosus*
ppm	parts per million
PrK	propargyl-lysine
PTIR	prism type total internal reflection
PUA	pseudouridine synthase archaeosine transglycolase
PUS	pseudouridine synthase
R	dye distance

R	adenosine or guanosine
RESOLFT	Reversible Saturable Optically Linear Fluorescence Transition
RNA	ribonucleic acid
RNase	ribonuclease
RNP	ribonucleoprotein
rRNA	ribosomal RNA
S. cerevisiae / *S.c.*	*Saccharomyces cerevisiae*
SCOK	strained cyclooctyne-lysine
SDS	sodium dodecyl sulfate
SEC	size exclusion chromatography
sRNA	small RNA
sRNP	small ribonucleoprotein
snRNA	small nuclear RNA
snoRNA	small nucleolar RNA
snoRNP	small nucleolar ribonucleoprotein
smFRET	single molecule FRET
SPAAC	strain-promoted azide-alkyne cycloaddition
SPIEDAC	srain-promoted inverse-electron-demand Diels-Alder cycloaddition
SRM	super resolution microscopy
STED	Stimulated Emission Depletion
STORM	Stochastic Optical Reconstruction microscopy
sub	substrate
THF	tetrahydrofuran
TIR	total internal reflection
TIRF	total internal reflection fluorescence
TLC	thin layer chromatography
tRNA	transfer RNA
U	uridine
UTP	uridine triphosphate
UV	ultraviolet
v/v	volume per volume
vol. eq.	volume equivalent
v_{start}	initial velocity
vis	visible
W	Nhp2 (Non-Histone Protein 2)
WNCG	Nhp2-Nop10-Cbf5-Gar1
wt	wildtype
w/v	weight per volume
x_c	peak center

Danksagung

Hiermit möchte ich mich ganz besonders bei **Dr. Martin Hengesbach** bedanken für die Möglichkeit, meine Doktorarbeit in seiner Nachwuchsgruppe anfertigen zu können. Die stetige Unterstützung, die gewährten Freiheiten und die immer offene Tür wusste ich sehr zu schätzen. Das interessante Forschungsthema und die damit verbundene Erlernung von vielen neuen Techniken im Bereich Chemie und Biochemie haben mich in meiner wissenschaftlichen Ausbildung extrem vorangebracht. Vielen, vielen Dank Martin!

Äußerster Dank gilt auch **Prof. Dr. Harald Schwalbe** für die großzügige Unterstützung während meiner Dissertation, die Übernahme des Zweitgutachtens und natürlich auch für die interessanten Vorlesungen während meines Studiums.

Natürlich danke ich auch **Prof. Dr. Mike Heilemann** für die Bereitstellung des TIRF-Mikroskops, an dem ich meine smFRET-Experimente durchführen konnte, sowie dem gesamten **Arbeitskreis Heilemann** für die Hilfe, falls es mal technische Schwierigkeiten gab. Weiterhin danke ich auch **Dr. Burkhard Endeward** für die Durchführung der EPR Messungen und **Dr. Thomas Halbritter** für die Bereitstellung des Nitroxid-Spinlabels.

Ganz besonderen Dank möchte ich **Dr. Andreas Schmidt**, **Gerd Hanspach** und **Christin Fuks** aussprechen für die tolle Zusammenarbeit an den verschiedenen Projekten, die interessanten und konstruktiven wissenschaftlichen Diskussionen und die gegenseitige Unterstützung im Labor.

Ein großes Dankeschön geht auch an die Mitglieder des Horizon Rooms **Nathalie Meiser**, **Linda Schulte**, **György Pintér**, **Vijay Nimbarte**, **Tatjana Schamber** und **Branco Gilbert** für die gute Atmosphäre im Büro, die lustigen aber auch konstruktiven Unterhaltungen und eine insgesamt sehr schöne Zeit während der Dissertation. Auch danke an die Mitglieder vom **AK Smash**, die ab und zu den Horizon Room unsicher machten und natürlich an **Mopso**, der ebenfalls für tatkräftige Unterstützung im Laufe der Jahre gesorgt hat.

Bei den gesamten **Arbeitskreisen Hengesbach**, **Fürtig und Schwalbe** bedanke ich mich für das tolle Gruppenklima, die stets positive Atmosphäre im und um das Labor sowie die spaßigen Grouptrips. In diesem Zusammenhang auch ein super Dankeschön an **Kerstin Dathe** für die Hilfe bei allen organisatorischen Dingen und die Rügen, falls ich mal wieder zu viele Kaffeetassen gebunkert hatte.

Außerdem möchte ich mich auch bei meinen Bachelorstudenten **Nicole Mench** und **Christoph Nocker** sowie meinen Masterpraktikanten **Aysun Özdemir**, **Daniel Hart**, **Daniel Ries**, **Jonathan Zöller** und **Mirko Brodsky** für die tatkräftige Unterstützung an den Projekten und die gute Zusammenarbeit bedanken.

Unbedingt möchte ich mich auch bei meinen Freunden bedanken, die mich durch mein Studium begleitet haben, unter anderem **Dr. Katja Barth**, **Dr. Sven Kirschner** und **Michael Gärtner**, sowie **Dr. Nai Wei Liu** für den ein oder anderen erfolgreichen Taschenmonster Raid. Einen ganz besonderen Dank möchte ich an **Alexander Gessing** und **Thorben Bürgel** richten, deren Freundschaft seit der Schulzeit Bestand hat und die mir immer mit Rat und Tat zur Seite standen.

Chemie und Kochen gehört zusammen, das hat die BARNS-Gruppe gezeigt: Danke and **Betül Ceylan**, **Albrecht Völklein**, **Robin Krishnatas** und **Nadide Altınçekiç** für die vielen leckeren Gerichte, die wir zusammen gekocht haben, für die Cocktailabende und die gemeinsamen Kaffeepausen.

Hier möchte ich auch noch einmal einen ganz besonderen Dank an **meine Familie** und vor allem an meine **Mutter** und meinen **Vater** aussprechen, die mich bisher bei allem, was ich getan habe, unterstützt haben und die immer für mich da waren. Dafür danke ich euch von ganzem Herzen.

Publications

Peer-reviewed and published:

Hanspach, G.*; Trucks, S.*; Hengesbach, M. Strategic labelling approaches for RNA single-molecule spectroscopy. *RNA biology* **2019**, *16*, 1119–1132.

Müller, D.*; Trucks, S.*; Schwalbe, H.; Hengesbach, M. Genetic Code Expansion Facilitates Position-Selective Modification of Nucleic Acids and Proteins. *ChemPlusChem* **2020**, *85*, 1233–1243.

(*****shared co-first authorship)

Submitted:

Trucks, S.*; Hanspach, G.*; Hengesbach, M. Eukaryote specific RNA and protein features facilitate assembly and catalysis of H/ACA snoRNPs. (*submitted to Nucleic Acids Research*)

Kubatova, N.*; Qureshi, N. S.*; Altincekic1, N.; Bains, J. K.; Ceylan, B.; Fuks, C.; Hargittay, B.; Hutchison, M. T.; Jesus, V. d.; Kutz, F.; Martin, M. W.; Meiser, N.; Linhardt, V.; Pyper, D. J.; Trucks, S.; Fürtig, B.; Hengesbach, M.; Löhr, F.; Richter, C.; Saxena, K.; Schlundt, A.; Schwalbe, H.; Sreeramulu, S.; Wacker, A.; Weigand, J.; Wöhnert, J. ^{1}H, ^{13}C, and ^{15}N backbone chemical shift assignments of coronavirus-2 non-structural protein Nsp10 (*submitted to Biomolecular NMR Assignments*)

(*****shared co-first authorship)

Scientific Activities

10/2018	10th RNA Biochemistry Meeting, Bonn, Germany (**Poster Presentation**)
09/2017	VIII. Nucleinsäurechemietreffen der DNG, Mainz, Germany (**Oral and Poster Presentation**)
07/2017	Integrated Research Training Group Summer School 2017 of CRC 902 "Molecular Principles of RNA-based Regulation", Wernigerode, Germany (**Oral Presentation**)
10/2016	9th RNA Biochemistry Meeting, Bonn, Germany (**Poster Presentation**)
09/2016	Integrated Research Training Group Summer School 2018 of CRC 902 "Molecular Principles of RNA-based Regulation", Hirschegg, Austria (**Oral Presentation**)

Students mentored

2019 Christoph Nocker, Bachelor Thesis
2018 Jonathan Zöller, M. Sc. Internship
2017 Nicole Mench, Bachelor Thesis
2017 Daniel Ries, M. Sc. Internship
2017 Aysun Özdemir, M. Sc. Internship
2016 Mirko Brodsky, M. Sc. Internship
2015 Daniel Hart, M. Sc. Internship

List of Academic Teachers

Dr. Markus Braun, Dr. Georg Bruls, Prof. Dr. Irene Burghardt, Dr. Stephanie Cronje, Prof. Dr. Ernst Egert, Prof. Dr. Clemens Glaubitz, Prof. Dr. Michael Göbel, Prof. Dr. Martin Grininger, Prof. Dr. Alexander Heckel, Dr. Rainer Hegger, Dr. Martin Hengesbach, Prof. Dr. Max Holthausen, Dr. Eltahmash Israr, Dr. Martin Kind, Dr. H.-Wolfram Lerner, Prof Dr. Arnim Lühken, Prof. Dr. Georg Manolikakes, Dr. Steffen Metz, Prof. Dr. Thomas Prisner, Prof. Dr. Matthias Rehahn, Prof. Dr. René Reifarth, Prof. Dr. Martin Schmidt, Prof. Dr. Harald Schwalbe, Prof. Dr. Andreas Terfort, Prof. Dr. Josef Wachtveitl, Prof. Dr. Matthias Wagner