# Structural dynamics of eukaryotic H/ACA RNPs from *Saccharomyces cerevisiae*

# &

# Structural dynamics of the Guanidine-II riboswitch from *Escherichia coli*

## Dissertation

zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Biochemie, Chemie und Pharmazie
der Johann Wolfgang Goethe–Universität
in Frankfurt am Main

von

## Christin Fuks

aus Aschaffenburg

Frankfurt 2023

(D 30)

Vom Fachbereich Biochemie, Chemie und Pharmazie der
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Clemens Glaubitz

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

Datum der Disputation:

# Table of Contents

# Chapter I

# General Introduction

# 1 General Introduction

## 1.1 Fluorescence

For understanding the principles of life, it is important to visualize molecules. Different techniques were developed for this purpose with some making use of fluorescence. A molecule that was not exited by light is found in the electronic ground state ($S_0$)[1]. With the absorption of energy in form of a photon it can enter an excited state (e. g. $S_1$ & $S_2$) (Figure 1). The probability of the absorption at a certain wavelength is defined by the extinction coefficient (ε). An outer electron is promoted into a different orbital. This process occurs within femtoseconds. The molecule can relax to the vibrational ground state of the exited states by transferring the energy to the solvent via vibrational relaxation. It is also possible that the molecule enters a lower exited state or the ground state without radiation in an internal conversion event. Both, vibrational relaxation and internal conversion happen in picoseconds. If the lowest vibrational state of the exited state is left to the ground state via the emission of a photon, this is called fluorescence. Since this process takes nanoseconds, thus is slower than the vibrational relaxation, the emitted photon has less energy and a higher wavelength due to the energy loss. This red-shift is also called Stokes shift. The molecule can also leave the excited singulet state in an intersystem crossing event to an excited triplet state ($T_1$). Here, the outer electron is not only boosted to a new orbital compared to the ground state, but also a forbidden spin reversal occurs. This can happen because the energy levels of the lowest $S_1$ state and $T_1$ overlap. The relaxation from $T_1$ to $S_0$ requires another spin reversal and is slow with microseconds or longer. This process is called phosphorescence. The ratio of the fluorescence pathway to all other non-radiative energy losses within the molecule defines the quantum yield of the fluorophore. Another way to get out of the exited state is by non-radiative transmission of the energy from one molecule to another molecule with a lower energetic $S_1$ state. This transfer is called Förster resonance energy transfer (FRET)[1]. The Jablonski diagram sums up the different relaxation pathways (Figure 1).

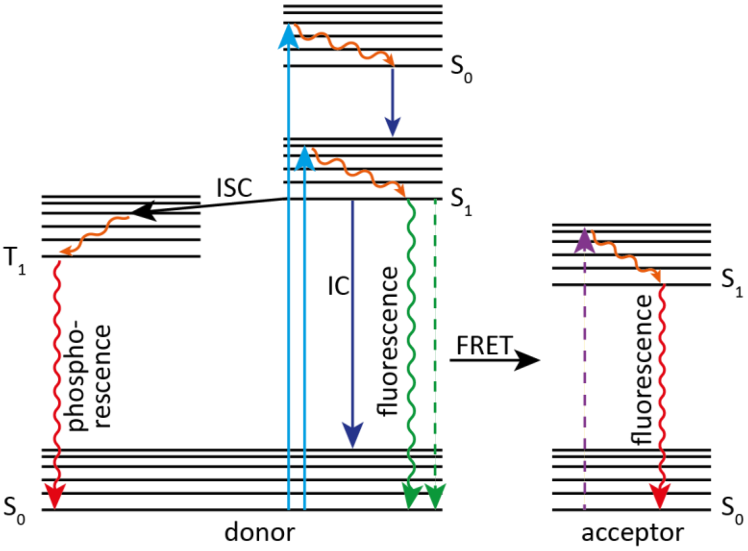


Figure 1: Jablonski diagram summing up different relaxation pathways from the exited states. IC: internal conversion; ISC: intersystem crossing.

## 1.2 Förster Resonance Energy Transfer (FRET)

FRET describes the non-radiative energy transfer from a donor fluorophore to an acceptor fluorophore[2]. This transfer is mediated via dipole-dipole interaction between a donor and an acceptor[3]. To be able to perform FRET, the two fluorophores must have a spectral overlap of the donor emission and acceptor absorption spectrum (Figure 2B). Then the dyes qualify as a FRET pair. The efficiency of this transfer (FRET efficiency; $E_{FRET}$) is defined as percentage of the fluorescence intensity of the acceptor molecule (equation 1). It is dependent on the 6th power of the inter-dye distance (R) between the fluorophores (Figure 2A). When the fluorophores have a large distance the FRET efficiency is low and when the fluorophores are close together, the transfer is more efficient. Equation 1 also contains the Förster radius ($R_0$) which is defined as the distance where the FRET efficiency is 0.5 [3]. $R_0$ is mostly defined by the choice of the FRET pair (equation 2). However, the orientation of the fluorophores and their respective dipolar moment also influences the transfer. For freely rotating molecules the orientation factor κ can be assumed with 2/3 which is not always the case in biological systems[4,5].

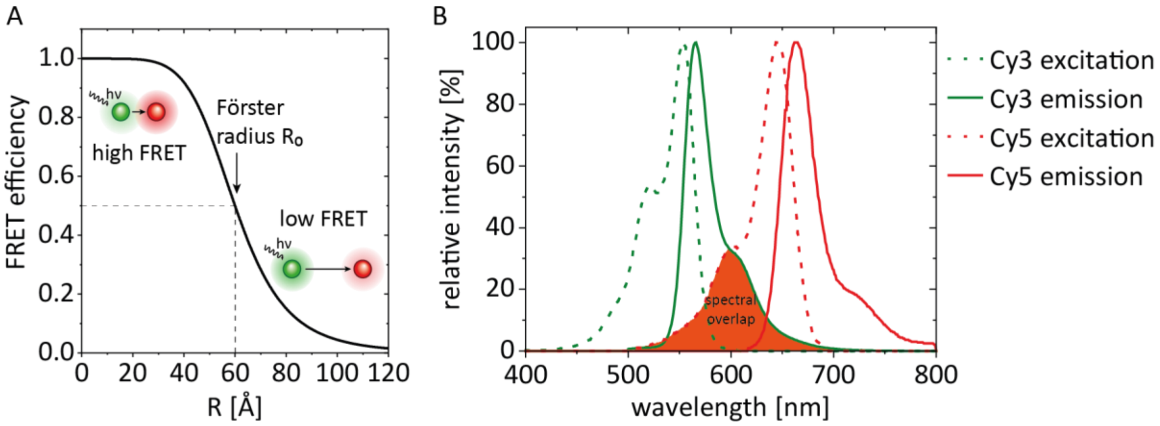


Figure 2: A) Distance dependence of the FRET efficiency. B) Excitation and emission spectra of the FRET pair Cy3 and Cy5. The spectral overlap between donor emission and acceptor excitation is shown in orange.

$$E_{FRET} = \frac{I_A}{I_A + I_D} = \frac{R_0^6}{R_0^6 + R^6} \tag{1}$$

$E_{FRET}$: FRET efficiency

$I_A$: acceptor fluorescence intensity

$I_D$: donor fluorescence intensity

$R_0$: Förster radius, distance with $E_{FRET} = 0.5$

R: inter-dye distance

$$R_0 = 0.211(\kappa^2 n^{-4} Q_D J(\lambda))^{\frac{1}{6}} \tag{2}$$

κ: orientation factor

n: refractive index of the optical medium

$Q_D$: donor quantum yield

J(λ): spectral overlap of donor emission and acceptor excitation spectrum

This relationship between distance and FRET efficiency makes FRET spectroscopy an important tool for the study of distances between 1 and 10 nm[6]. It can be used as a molecular ruler or for the observation of dynamic structures. FRET allows the time-resolved investigation of molecules from milliseconds to several minutes[3]. When working with FRET the choice of the fluorophores is crucial. The ideal fluorophores are bright with $Q > 0.1$ and $\varepsilon > 50{,}000\ M^{-1}cm^{-1}$ [3]. They should be small, photostable, water-soluble and show minimal aggregation effects. The fluorophores of the FRET pair should have a large spectral overlap of donor emission and acceptor excitation to allow a good transfer. Furthermore, it is advantageous when they have good detection efficiencies and similar quantum yields. Even though fluorescent proteins like variants of the green fluorescent protein (GFP) from *Aequorea victoria* are used for FRET measurements[7], their size could have adverse effects especially on smaller systems. For this reason small organic molecules are usually used for FRET spectroscopy[8]. A commonly used FRET pair consists of the cyanine dyes Cy3 ( Figure 3A) and Cy5 (Figure 3B)[3] which have a good spectral overlap (Figure 2B).

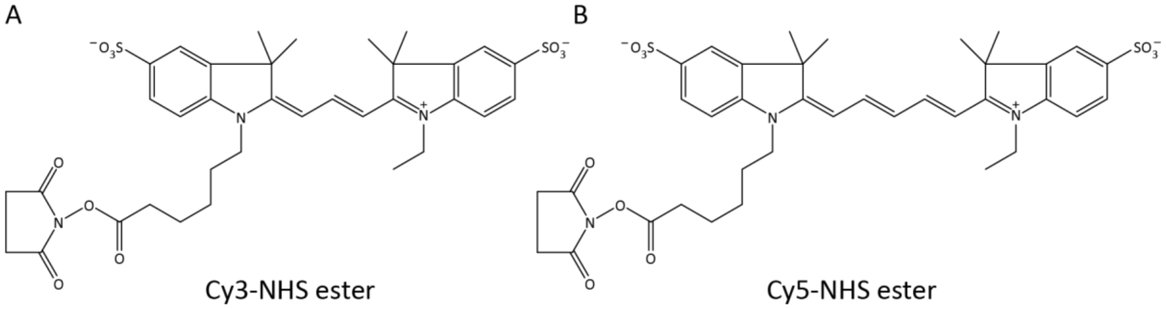


Figure 3: Chemical structure derivatives of A) Cy3 and B) Cy5 in the NHS ester reactive form.

## 1.3 Single-Molecule Techniques

FRET spectroscopy can be used in ensemble measurements which could give a hint on the distance of the dye. However, these bulk experiments have the disadvantage that smaller subpopulations cannot be resolved and donor only molecules e. g. due to acceptor photobleaching would also contribute to the FRET efficiency. These problems can be overcome by FRET analysis on a single-molecule level (smFRET). Here, the (time-resolved) donor and acceptor intensities and with that the FRET efficiency can be determined for individual molecules which allows for the investigation of short-lived states and smaller subpopulations[3]. For smFRET experiments a confocal fluorescence microscope can be used[9,10], but usually this is not the method of choice. Total internal reflection fluorescence (TIRF) microscopy is now the most common method used for smFRET studies[11]. TIRF makes use of the reflection of light at the phase boundary between two media with different optical densities. At a certain angle all light is reflected which is called total reflection. At total reflection an evanescent field can penetrate around 100 to 200 nm into the sample, illuminating only the molecules close to the microscope slide. In comparison to confocal microscopy, TIRF has a reduced background fluorescence, because the light beam cannot reach other planes[11]. Two different TRIF microscope types were developed, the objective- and the prism-type (Figure 4)[12].

In the objective-type TIRF microscope an oil objective with a high numerical aperture is used. The laser is located under the objective and passes it in a parallel beam. Total reflection then occurs at the glass-water interface. The molecules at the lower side of the microscope slide can be detected. The fluorescence light as well as the reflected excitation light passes again through the objective. Using a dichroic mirror, the excitation light can be separated from the fluorescence information[12].

In prism-type TIRF the laser beam comes from above and passes through a silica prism. The oil used between the prism and the objective carrier should have the same refractive index and quartz slides need to be used. Then total reflection happens at the quartz-water interface. In the case of prism-type microscope only the molecules at the upper side of the slide are measured. Fluorescence light passes through the objective[12]. Prism-type microscopy has the disadvantage that the prism needs to be assembled whenever changing the slide. However, since the excitation light is reflected and does not pass through the objective the background is reduced compared to objective-type setups.

Independent of whether an objective- or prism-type microscope is used, the detection pathway is similar. The scattered light from the excitation laser can be filtered out using a long-pass or band-pass filter. Afterwards, a split determines the image area on the camera. smFRET requires the simultaneous detection of donor and acceptor fluorescence. Using an arrangement of dichroic mirrors that reflect the donor but not the acceptor emission wavelength the channels can be separate to different areas of the camera. For smFRET usually electron multiplying charge-coupled device (EMCCD) cameras are used[11].

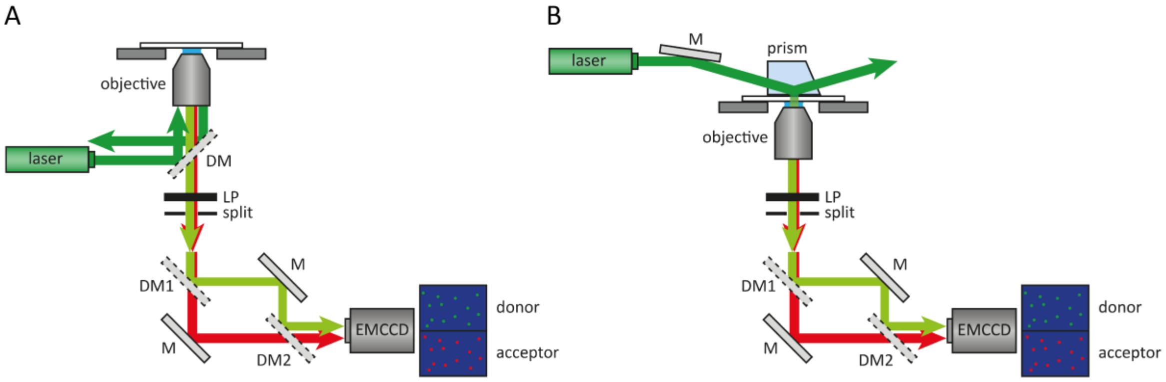


Figure 4: Setup of the A) objective-type and B) prism-type TIRF microscope. M: mirror; DM: dichroitic mirror; LP: long-pass filter.

Because in TIRF the evanescent field only penetrates about 100 to 200 nm of the sample, only molecules close to the surface can be measured. The molecules can be immobilized on the slide using affinity tags like the His-tag-Ni-NTA interaction or via streptavidin-biotin interaction[3]. The protein streptavidin from *Streptomyces avidinii* has one of the strongest binding affinity found in nature to biotin with a $K_D$ in the femtomolar range[13]. When investigating only nucleic acids without proteins the cleaned slides can be blocked using bovine serum albumin (BSA) with a fraction of biotinylated BSA. The tetrameric protein streptavidin can bind to the biotin moiety of the BSA-biotin and another biotin attached to the target molecule[3]. To avoid sticking of proteins to the surface of the slide, polyethylene glycol (PEG)-modified slides need to be used for experiments involving protein samples[14]. For this, surface-activated slides are aminosilanized. The aminosilane reacts with NHS-ester modified PEG with a small percentage of biotin-PEG-NHS ester. For single-molecule detection the molecules need to be spatially separated far enough, but at the same time a high number of molecules should be in the image area to reduce measurement time. Optimal molecule densities are around 0.1 - 0.2 molecules/µm$^2$ [3].

## 1.4 Oxygen Scavenging

Time-resolved smFRET experiments require the two fluorophores to be photostable over a long period. After several excitation cycles fluorophores can lose their fluorescence irreversibly due to photobleaching[15]. Photobleaching can be induced by radical oxygen species (ROS) in the measurement buffer[16]. For this reason, the sample needs to be protected from air and oxygen solved in the buffer needs to be removed (oxygen scavenging). One approach for the removal is the Gloxy system, a combination of the enzymes glucose oxidase and catalase (Figure 5A)[17,18]. Glucose oxidase can bind solvent oxygen by oxidation of glucose. The products of this reaction are glucuronic acid and hydrogen peroxide, which is a ROS itself. The peroxide is then inactivated using catalase. Since the product of the first reaction is an acid, the Gloxy system can lead to a drop in pH over time which might have adverse effects in long-time experiments. Another approach is the PCA/PCD system (Figure 5B). Here, protocatechuic acid (PCA) is oxidized by protocatechuate-3,4-dioxygenase (PCD) which binds the oxygen[19,20]. Fluorophore blinking can be avoided by quenching of the triplet state using the vitamin E analogue 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid (Trolox) (Figure 5C)[21].

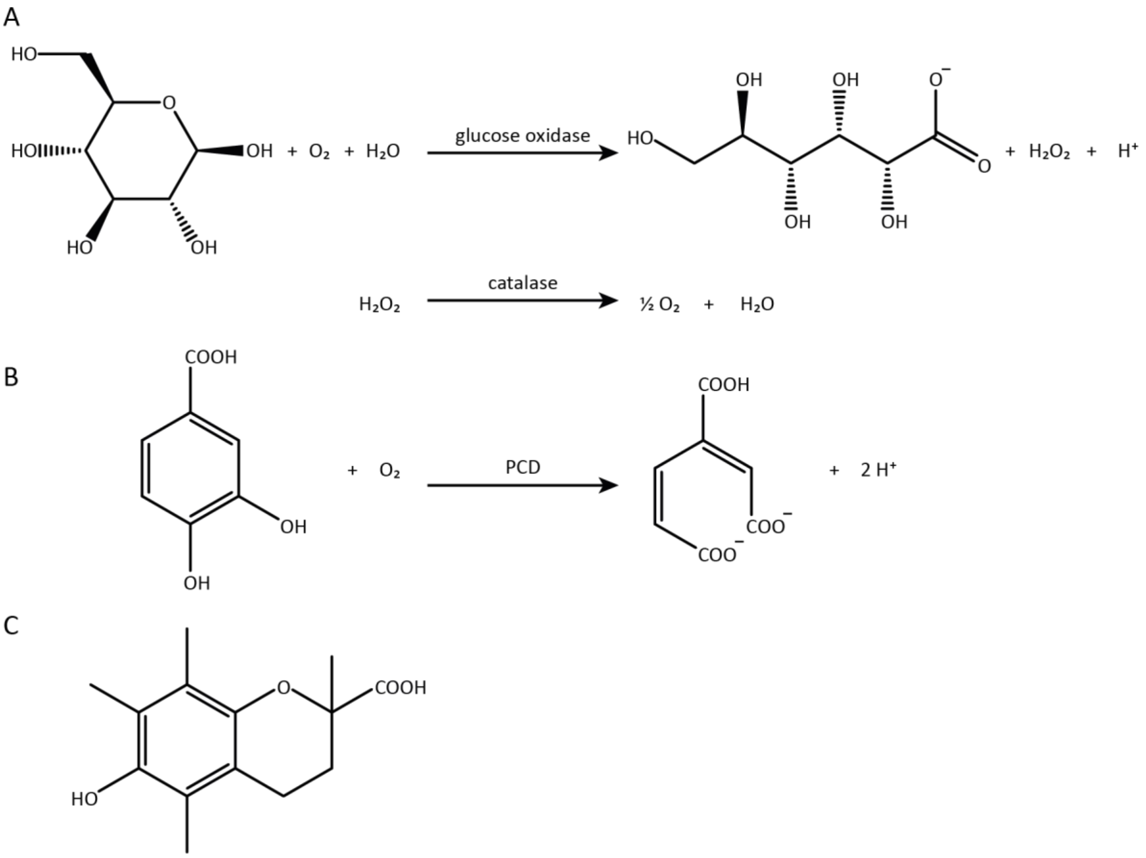


Figure 5: Oxygen scavenging scavenging systems. A) Enzymatic reaction of glucose oxidase oxidizing glucose to glucuronic acid followed by removal of hydrogen peroxide using catalase. B) Enzymatic reaction of protocatechuic acid (PCA) oxidation with protocatechuate-3,4-dioxygenase (PCD). C) Chemical structure of 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid (Trolox).

# Chapter II

# Eukaryotic H/ACA RNPs from *S. cerevisiae*

# 2 Introduction

## 2.1 RNA Modifications

While proteins consist of 20 different amino acid, RNA is built of four different nucleotides A, C, G and U. In 1950s a modified nucleotide, pseudouridine (Ψ), was discovered as the fifth nucleotide[22]. Today over 150 RNA modifications are known. Some modifications are exclusive to one or two kingdoms of life, while others are distributed in a wider range of organisms[23,24]. The research of RNA modification is also called epitranscriptomics. Modifications are introduced into the RNA posttranscriptionally to expand the variety of these building blocks. The types of modification range from addition of chemical groups to the base or the ribose (e. g. *N*6-methyladenosine, $m^6A$) to isomerisations (pseudouridine, Ψ) or oxygen substitutions in both base and phosphate backbone (e. g. 2-thiouridine, $s^2U$)[23] (Figure 6).

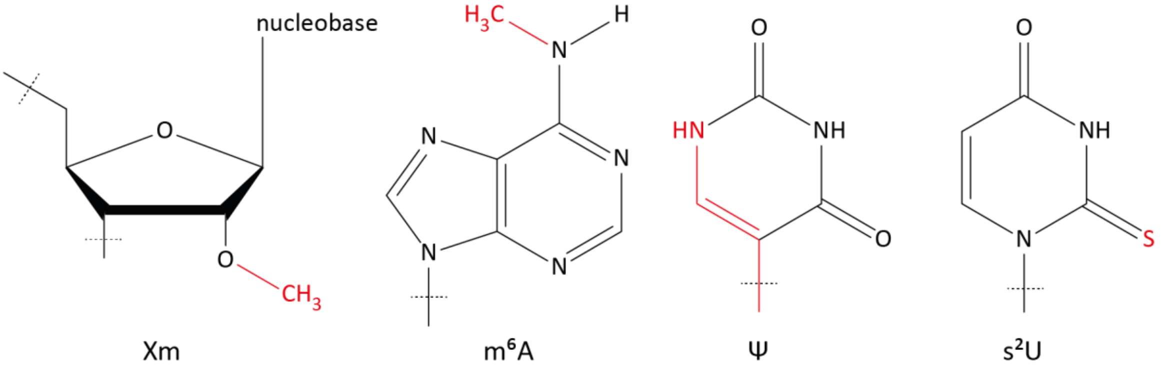


Figure 6: Examples of RNA modifications. Chemical structures of ribose methylation (Xm), N6-methyladenosine ($m^6A$), pseudouridine (Ψ) and 2-thiouridine ($s^2U$).

Different types of RNA can be modified including rRNA[25], tRNA[26], mRNA[27], snRNA[28], miRNA[29] and lncRNA[30]. Of those RNA types tRNA is the most extensively modified one. The level of modification depends on the organism and partially the growth phase. For instance eukaryotic tRNA has a modification degree of about 16.5%[31]. In general, these modifications are important for the overall stability[32,33], however for some a more specific function was discovered. Here the modification sites at position 34 and 37 are most important. The modification at position 34 in the anticodon loop is required for correct translation of wobble base pairs[34]. The nucleotide adjacent to the anticodon loop at position 37 is also necessary for codon recognition and prevents frameshifting[35,36]. The second most modified RNA class is rRNA with about 2% modifications[37]. The importance of the rRNA modifications is underlined by their accumulation in functional regions like the decoding site, the tRNA binding site, peptidyltranserfase center and the intersubunit interface[38,39]. While some individual modifications are essential for biogenesis and function, others only unfold severe effects in combination[40–42]. In mRNA $m^6A$ is the most prevalent modified nucleoside[43]. However, the biological relevance for many modifications remains unknown.

### 2.1.1 Pseudouridine (Ψ)

In 1951 a fifth nucleotide was found in addition to the known nucleotides A, C, G and U[44] which was later identified as 5-ribosyl uracil or pseudouridine (Ψ), the isomer of uridine[22] (Figure 7). Ψ is the most abundant modified nucleotide[23,24]. It is found in a variety of RNA species like rRNA[31], tRNA[31], snRNA[45], mRNA[46], telomerase RNA[47] and long non-coding RNA[46].

Ψ is introduced to the RNA in an isomerization reaction of uridine (Figure 7). The C-N-glycosidic bond between ribose and base is broken, followed by rotation of uracil by 180° and formation of a new C-C-glycosidic bond with C5 of uracil. This is either achieved by stand-alone pseudouridine synthases or via RNA guided pseudouridine synthases. The Watson-Crick interface is unaffected by the isomerization, hence A-Ψ basepairing remains possible. However, there is an additional hydrogen bond donor at N1. This can for instance increase the RNA's backbone rigidity compared to uridine by coordinating the phosphate through a water molecule[48]. Furthermore, the 3' endo preference of Ψ could lead to an increased base stacking[49].

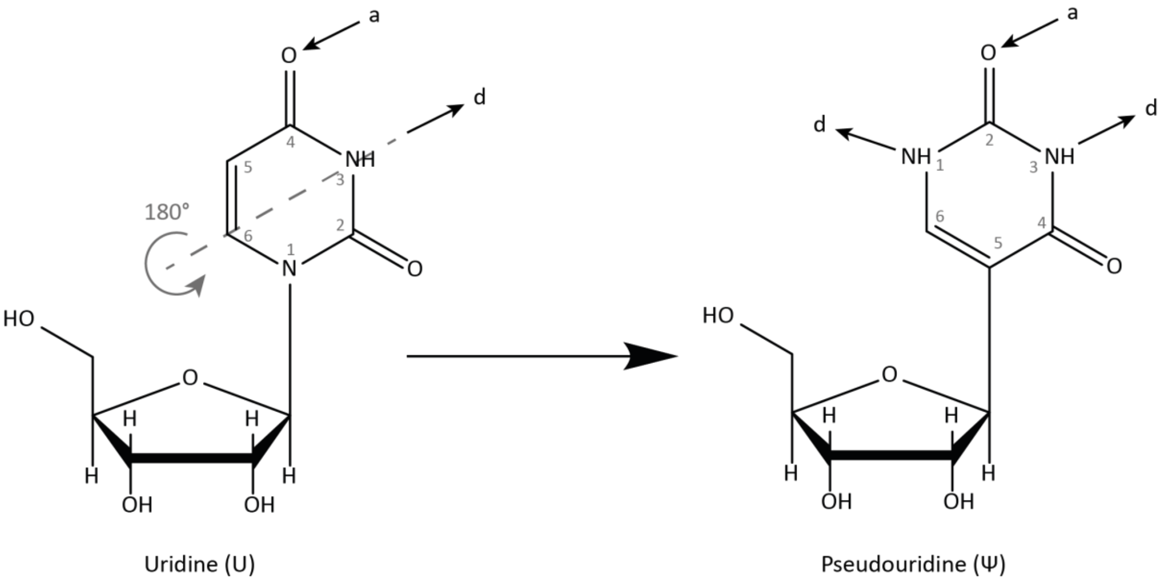


Figure 7: Isomerization of uridine to pseudouridine.

While some Ψ modifications affect RNA stability and severe effects only occur when several of these modifications are missing[40], other Ψ sites are more essential. For example, the conserved Ψ42 of the spliceosomal U2 snRNA is introduced by an H/ACA pseudouridylase using snR81 as a guide RNA. At this position the Ψ is important for Prp5 binding and the structure of the branch site recognition thus necessary for efficient spliceosome assembly[45]. A lack of this Ψ42 results in a reduction of the splicing efficiency. Apart from conserved Ψ sites, the modification can also be a mediator of a stress response. The same snoRNA that determines U42 isomerization in the U2 snRNA can also introduce Ψ93 during nutrient deprivation[50]. Ψ93 then alters the spliceosome dynamics leading to reduced splicing[51].

## 2.2 H/ACA RNPs

One way for the site-specific introduction of Ψ is the use of H/ACA ribonucleoproteins (RNPs). The target uridines are selected via the use of a guide RNA. These guide RNAs contain a helix-bulge-helix motif. In the bulge (also called pseudouridylation pocket) the substrate can bind via complementary base pairing (Figure 8). In the three-way junction between the distal helix and the Ψ-pocket one uridine remains unpaired and is pseudouridylated by the H/ACA RNP[52]. The hairpins are followed either by an H-box (ANANNA) sequence or an ACA-box[53]. In the nucleolus where rRNA is modified those guides are called small nucleolar RNAs (snoRNA) and for the pseudouridylation of snRNA in the Cajal bodies scaRNAs are used. scaRNAs have a CAB-box (UGAG) motif located in their loop which is necessary for localization into the Cajal bodies[54]. In addition to the RNA guide four more proteins are involved in the pseudouridylation reaction: the catalytic active centromere-binding factor 5 (Cbf5) which is also called dyskerin in humans or Nap57 in vertebrates, the nucleolar protein 10 (Nop10), the glycine and arginine-rich protein 1 (Gar1), and the nonhistone protein 2 (Nhp2; L7Ae in archaea). While the H/ACA RNPs are best known for their function in RNA pseudouridylation, additional functions were discovered. For instance, the U17/snR30 RNP are important for pre-rRNA processing and ribosome assembly[55–57]. Furthermore the TERT RNA contains an H/ACA motif and they work as part of the telomerase complex in telomere maintenance[58].

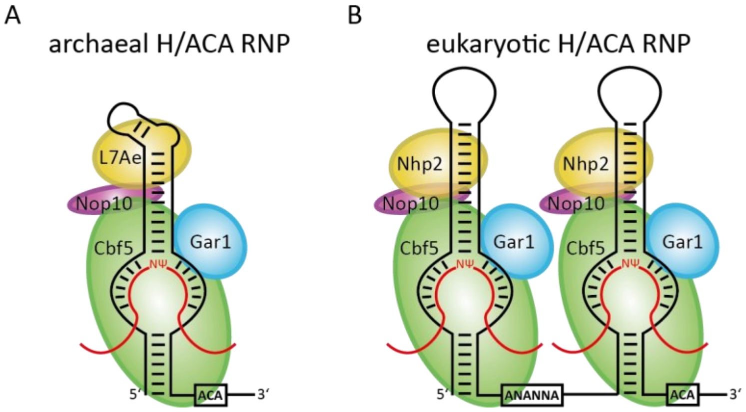


Figure 8: Schematic representation of H/ACA RNPs from A) archaea and B) eukaryotes. The guide RNAs are shown in black and the substrate RNA is shown in red.

### 2.2.1 Archaeal H/ACA RNPs

H/ACA RNPs are found in both, archaea and eukaryotes. Despite having a similar general structure and function there are several differences. Archaeal H/ACA complexes consist of one to three hairpin structures[59,60]. In or near the distal loop of the sRNA a kink-turn (K-turn) motif is found. Nhp2 is replaced by the structurally similar L7Ae which binds specifically to t K-turn motif[61] but does not interact with the other protein components in absence of RNA[62]. Archaeal Cbf5 can recognize and bind the ACA box specifically[62]. Direct binding of L7Ae and Cbf5 to the RNA then leads to an interaction of L7Ae with Nop10[63]. In this complex L7Ae is responsible for the correct placing of the target uridine in the catalytic site of Cbf5 site[64,65] which makes the protein essential for pseudouridylation in archaea[66]. Structural knowledge of H/ACA RNPs was gained through crystallization of fully assembled archaeal RNP[67].

### 2.2.2 Eukaryotic H/ACA RNPs

While in archaea one to three hairpins can form the guide RNA, in eukaryotic H/ACA RNPs two hairpins are conserved. Several reports showed that the bipartite architecture increases pseudouridylation in comparison to isolated hairpins[66,68,69]. The K-turn motif is not found in these guide RNA. The 5’ hairpin sequence is followed by the H-box and the 3’ hairpin by the ACA-box motif. Mutation of either the H-box or the ACA-box can reduce or completely inhibit substrate turnover even at the other hairpin[69,70]. The Ψ-sites are in almost all cases 14 nt upstream of those elements, which shows that target selection is not only dependent on the base pairing but also the correct distance for positioning of the substrate in the catalytic center of Cbf5[52]. The base-pairing interactions between the pseudouridylation pocket and the target RNA modulate the efficiency of the isomerization reaction but not the affinity with at least eight base pairs necessary for performing the pseudouridylation[71,72]. Eukaryotic H/ACA RNAs can bind to Cbf5, Gar1 and Nhp2 individually[73]. However, in contrast to L7Ae, Nhp2 can form a stable complex with Cbf5 and Nop10 (WNC complex) before interaction with the RNA[74]. Nhp2 alone binds with an affinity of 600 nM to the guide RNA[69]. Cbf5 can bind to RNA unspecifically. For the whole WNCG complex however binding affinities of 0.5 and 0.7 nM were reported for the snoRNAs snR34 and snR5, respectively, with half-lifes in the minute range[69].

### 2.2.3 Eukaryotic H/ACA proteins

In the H/ACA complex Cbf5 is the actual catalytic component[75]. It is a homologue of other pseudouridylases like the bacterial TruB[76]. The catalytic domain contains a conserved aspartate residue which is essential for pseudouridylation. Furthermore, Cbf5 has a thumb loop that coordinates substrate binding and release from the catalytic center. Additionally, Cbf5 contains a pseudouridine synthase and archaeosine transglycosylase (PUA) domain which is involved in ACA box binding[77] and positions the protein on the guide RNA that brings the catalytic domain to the 14 nt upstream target site[52]. In eukaryotes this protein also has N- and C-terminal extensions. Cbf5 alone is unstable in solution and needs other proteins, like the H/ACA accessory protein Nop10. While Cbf5-Nop10 interaction (NC complex) requires the Zinc binding domain in archaea, eukaryotic Nop10 lacks this domain[66].

Eukaryotic Nhp2 is structurally similar to L7Ae despite having different functions[78]. Nhp2 can associate with Cbf5 and Nop10 via Nop10 to form the WNC complex even in absence of RNA[74]. The exact role of Nhp2 in the complex is unknown. It is important for a fast and efficient catalysis[69] but in contrast to the archaeal counterpart L7Ae the activity is only reduced and not completely abolished[66].

The fourth protein in the complex is Gar1 and is involved in substrate release. Gar1 consists of a core domain that is also found in archaeal Gar1 but has an additional eukaryote-specific C-terminal extension (CTE). This extension is important for efficient pseudouridylation under multiple-turnover conditions[66]. Furthermore, the CTE increases the affinity of Gar1 to Cbf5 which allows the replacement of the assembly factor Naf1 that lacks this domain[66]. The core domain and CTE are flanked by domains with arginine-glycine-glycine repeats (RGG domains) in eukaryotes[79]. For yeast Gar1 there is one RGG domain on the N-terminal and two on the C-terminal side. RGG motifs are found in different proteins and affect the affinity and specificity of RNA binding[80]. In the case of Gar1 the presence of the RGG domains could either increase or decrease the pseudouridylation efficiency of the H/ACA RNP depending on the guide RNA and target[66,68] and alter the dynamics in smFRET measurements[68]. However, the Gar1 core domain is also capable of binding to the RNA in absence of the RGG domains[81]. The main interaction of Gar1 with the other complex components seems to be a protein-protein interaction with the thumb loop of Cbf5 instead of direct RNA binding.

### 2.2.4 Structure of eukaryotic H/ACA RNPs

Most structural knowledge of H/ACA complexes is based on archaeal constructs (Figure 9A)[67]. However, some information was gained on eukaryotic systems, as well. In 1998 a rough electron density map for the bipartite complex was reported[73]. In the electron microscopy data a structure with two lobes is visible. They most likely represent the two hairpins of the eukaryotic structure that are in an angled arrangement to each other. A more in-depth structure of the protein components delivered the crystallization of the NCG core complex[66] and the NMR structure of

Nhp2[78] (Figure 9B).

For the crystallization of the NCG complex only the core domains of Cbf5 and Gar1 were used without the eukaryotic specific NTE, CTE and RGG domains[66]. The overall arrangement is comparable to the studies on the archaeal complexes. Nop10 and Gar1 bind on different sites to Cbf5. Nop10 folds into a ribbon structure in complex with Cbf5. Gar1 forms a 6-stranded β-barrel. Binding to Cbf5 occurs in two different interfaces. One of them involves with the CTE and the eukaryote-specific α1 helix of Gar1 to the thumb loop of Cbf5 via hydrophobic contacts. Those contacts with the CTE are important for multiple-turnover activity and increases the affinity of the biding.

Koo *et al.* solved the solution NMR structure of the accessory protein Nhp2 with a deleted NTE[78]. The overall structure is similar to Nhp2 homologs like the archaeal L7Ae. Nhp2 has a secondary structure consisting of five α-helices and four β-strands in an alternating order. They fold into a three-dimensional α-β-α sandwich structure with the β-sheets in the middle and two or three α-helices on each site. The Nop10 binding interface is at the side formed by α2 and α3. Binding of Nop10 and snoRNA takes place according to similarities with L7Ae-Nop10 binding. The conserved proline 83 in α3 is prone to *cis*/*trans* isomerization with the *trans* conformation being preferred in solution. The S82W mutations shifts the equilibrium to the *cis* conformation which increases association with several snoRNAs as well as Cbf5 and might be the isomer present in the assembled H/ACA RNAP[78].

The only more detailed structural insights into an RNA bound bipartite complex are from cryo-electron microscopic structures of the human telomerase complex (Figure 9C)[82–84]. The human telomerase RNA (hTER) contains an H/ACA motif which binds the four H/ACA proteins dyskerin (the human Cbf5), Nop10, Gar1 and Nhp2. Transferring structural information from the telomerase structure to eukaryotic, bipartite H/ACA pseudouridylases should be taken with caution for two reasons: First, the loop of the 5' hairpin of the H/ACA domain in hTER is extended and interacts with other parts of the telomerase complex. Second, for modelling of the H/ACA proteins the archaeal structure was used, which lacks eukaryotic specific domains. In the human telomerase structure each hairpin is anchored by a dyskerin molecule binding to the lower stem and the H- or ACA box, respectively. For the 5' hairpin this is the only protein RNA interaction. The other components associate with dyskerin[82–84]. The two hairpins are in a bent conformation which is mediated by the dyskerin-dyskerin interaction at the base[82,83] and van-der-Waals stacking interactions of two nucleotides, one from the H-box and ACA box each[83]. In the 3' hairpin Nhp2 and Nop10 are involved in RNA binding in addition to dyskerin. Additionally the Cajal body localization factor Tcab1 is bound to the CAB box motif of the 3' hairpin[82,83]. The structure of the human telomerase shows the Nhp2 from the 5’ hairpin in close proximity to the Gar1 from the 3’ hairpin.

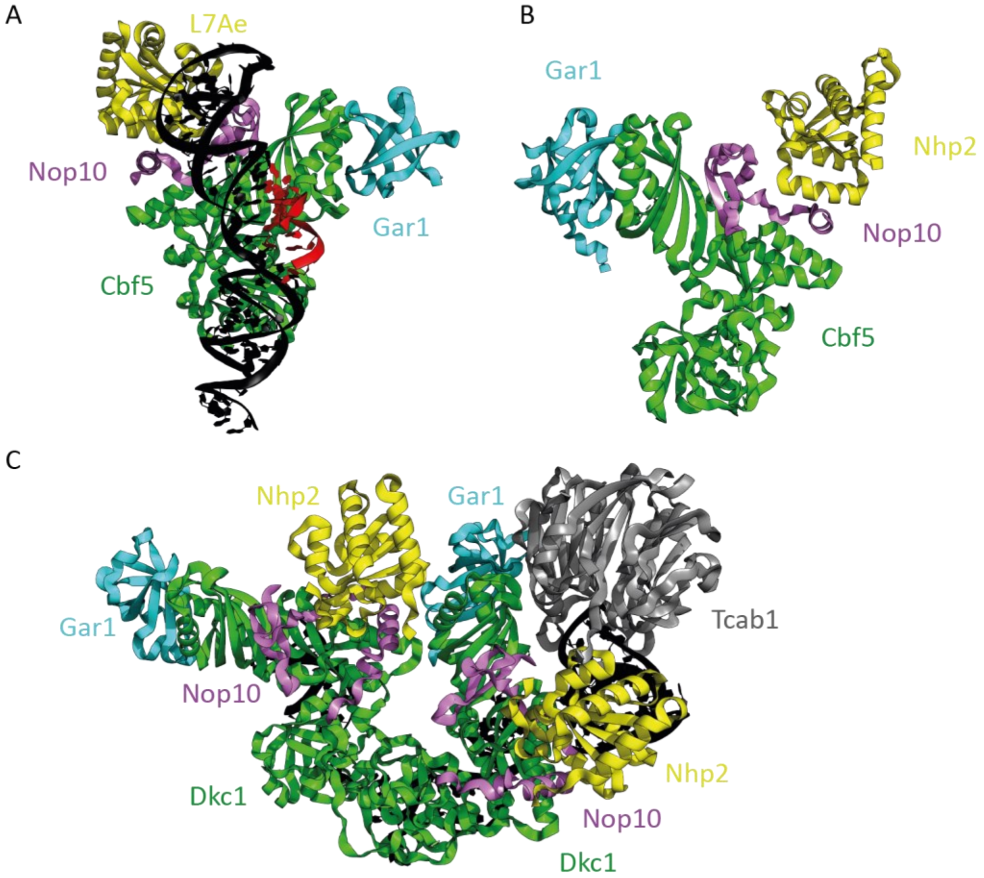


Figure 9: Structures of H/ACA complexes. A) Crystal structure of the archaeal complex (PDB: 3HAY)[67]. B) Crystal structure of core domains of the trimeric NCG complex (3U28)[66] and the solution NMR structure of Nhp2 (2LBX)[78]. Nhp2 was aligned according to the position in the archaeal complex. C) Cryo-EM structure of the H/ACA lobe of the human telomerase (7TRC)[84].

### 2.2.5 H/ACA assembly and catalysis

Recombinant archaeal and eukaryotic H/ACA ribonucleoproteins have been successfully reconstituted *in vitro* without any additional assembly factors[68]. *In vivo* however this is a multistep process involving several assembly factors. Free Cbf5 is unstable in solution. Shortly after translation it interacts with Shq1 which acts as a chaperone via the RNA binding surface of Cbf5[85]. The Cbf5-Shq1 complex is translocated to the nucleoplasm. Here, Naf1 that shares a similar fold with the Gar1 core domain binds to Cbf5 at the Gar1 binding site in the thumb loop[86]. Shq1 is released with the help of the R2TP complex that contains AAA+ ATPases[87]. Nhp2, Nop10 and Cbf5 can form a trimeric complex (WNC complex) prior to RNA binding. SnoRNAs are transcribed in intronic sequences[88], the assembly of the pre-snoRNPs occurs co-transcriptionally. To achieve this, Naf1 interacts with the C-termianl domain of the RNA polymerase II and recognizes the 3’ ends of the snoRNA sites[89,90]. The intronic snoRNAs are released through splicing and debranching reaction. The RNA-Cbf5-Nop10-Nhp2-Naf1 complex translocates to the Cajal body. Competitive replacement of Naf1 with Gar1 finishes the assembly of H/ACA RNP[86]. The eukaryotic-specific CTE of Gar1 is crucial for the replacement as it increases the affinity of Gar1 to Cbf5 compared to Naf1[66]. The mature snoRNA H/ACA RNP is translocated to the nucleolus where pseudouridylation of rRNA takes place.

Yang *et al.* proposed a thermodynamic and kinetic reaction pathway for substrate turnover[91]. In a first step the target RNA binds to the pseudouridylation pocket of the guide RNA via complementary base pairing. In eukaryotes binding affinities of 50-330 nM were determined[69]. After binding, the substrate RNA interacts with Cbf5, more specifically the thumb loop to place the target uridine in the reactive center where the conserved aspartate is located. The exact molecular mechanism of the actual isomerization of uridine into pseudouridine is not clear yet. Different mechanisms like a Michael-addition like reaction[92] or an acylal-[93] or glycal-[94,95] mechanism were proposed. According to the model of Yang *et al.* the structural difference of Ψ leads to the release of the thumb loop. In this case Gar1, which is bound to the thump loop, increases the product release 2-fold through destabilization of the substrate-bound complex[91].

In general, the target uridine is well defined through base-pairing of the surrounding sequences which leave the target uridine unpaired and through the 14 nt distance to the H/ACA box elements. However, it is possible that different, not optimal base-paired substrates can be pseudouridylated under stress conditions[50]. Furthermore, for some guides it is possible that the adjacent or the next but one uridine is pseudouridylated[96].

### 2.2.6 FRET studies on the H/ACA RNP

In previous smFRET studies both, archaeal[97] and eukaryotic[68] H/ACA RNPs were investigated. Schmidt *et al.* investigated the assembly, substrate binding and protein dynamics of the archaeal *Pyroccocus furiosus* H/ACA complex with different labeling and immobilization schemes[97]. FRET labels were placed at different sites of the pseudouridylation pocket or within the K-turn containing apical helix. smFRET data suggest, that L7Ae binding has no detectable effect on the RNA structure and that all four proteins are necessary for a conformational change and a quantitative opening of the pseudouridylation pocket. Introducing one fluorophore to the substrate RNA revealed that at least the LNC complex is necessary for substrate binding and addition of Gar1 lead to increased conformational dynamics as evidenced in a broader FRET peak. Immobilization of the complex through the RNA discovered and discriminating non-productive conformation when the target uridine is replaced with a mismatch. Furthermore, the smFRET study of the archaeal complex could show how Gar1 contributes to the mobility of the thump loop of Cbf5 when the FRET labels are placed in the thump loop and the substrate RNA.

In a different study the isolated hairpins of the eukaryotic yeast snR81 snoRNA were investigated[68]. Fluorophores were attached on different sites of the pseudouridylation pocket. For the H5 hairpin two discernable conformations were detected. The equilibrium between both states is dependent on the protein components. Addition of NCG could increase the fraction of molecules in a low FRET state. An interplay of all four proteins is necessary to bring most molecules into this open conformation. Labeling of Nhp2 resulted in one detectable binding mode to the H5 RNA. The H3 hairpin showed three peaks in the FRET data with double labeled RNA as well as in experiments with labeled Nhp2. A high FRET state is only populated when at least the NCG complex is present. Further addition of Nhp2 does not influence the equilibrium between the three states. The presence of the substrate RNA had no detectable influence in the smFRET experiments.

### 2.2.7 H/ACA in disease and therapy

In humans H/ACA RNPs are involved in ribosome biogenesis, pseudouridylation of other snRNAs, as well as telomere maintenance as part of the telomerase complex. It is not surprising that mutations and or malfunction of H/ACA proteins causes severe diseases including the X-linked bone marrow failure dyskeratosis congenita and several types of cancer[98–100]. Hence, investigation of H/ACA RNPs is crucial for understanding and possibly treating those diseases. Furthermore, artificial H/ACA RNPs might also be an interesting tool for the treatment of other diseases[101]. Nonsense mutations where a codon is mutated to a stop codon cause premature translation termination. All stop codons start with a uridine. When this uridine is replaced by Ψ this leads to a stop codon read through[102]. Instead of translation termination a serine, threonine, tyrosine or phenylalanine is inserted depending on the identity of the stop codon and translation progresses normally. Even though this could lead to one altered amino acid the function of the protein might be rescued. Making use of this feature Song *et al.* constructed an artificial guide RNA in a program named RESTART that can site-specifically pseudouridylate nonsense mutations and leads to partially full length proteins[103]. These artificial H/ACA might become a therapeutic tool for the treatment of nonsense mutation related diseases on a mRNA level[101].

# 3 Motivation and objective

Ψ is the most abundant natural RNA modification[23,104]. It is produced by the rotational isomerization of uridine. This process can be achieved by stand-alone pseudouridylases or in archaea and eukaryotes by RNA-guided H/ACA snoRNPs. H/ACA consist of a guide RNA and a set of four proteins: Cbf5 (dyskerin in human, Nap57 in vertebrates), Nop10, Gar1 and Nhp2 (L7Ae in archaea). The guide RNA consists of one to three hairpins in archaea and for eukaryotic snoRNAs two hairpins are conserved[59,60]. The hairpins have an internal pseudouridylation pocket where the substrate RNA can be bound via complementary base pairing[52]. Following the hairpins is either a H-box or ACA-box sequence motif[53]. Ψs are important for ribosome biogenesis and function, tRNA stability and codon-anticodon interaction, spliceosome function and are involved in mRNA stop codon read-through. With all these function it is not surprising that malfunctions of pseudouridylases can lead to severe and lethal diseases such as different types of cancer and the bone marrow failure syndrome dyskeratosis congenita[98–100]. It can also be a therapeutic tool for suppression of nonsense mutations[103]. For these reasons it is important to expand the knowledge in the field of H/ACA pseudouridylases.

In eukaryotic H/ACA RNPs a bipartite setup is conserved. Activity assays showed that presence of the second hairpin has positive effects on the pseudouridylation reaction[66,68,69]. However, most studies investigated only individual hairpins from archaeal or eukaryotic complexes. This work will expand the knowledge of the bipartite nature of the H/ACA complex. For this, several research questions will be addressed:

1) How does the second hairpin affect the conformation of the pseudouridylation pocket in the other hairpin?
2) How is the snoRNA arranged before protein assembly?
3) How could the three-dimensional architecture of the assembled complex look like?
4) What is the role of the individual proteins in the arrangement of the bipartite RNA?
5) How do the eukaryotic specific RGG domains of Gar1 influence the conformation of the complex?

To study the conformational dynamics of eukaryotic H/ACA RNPs smFRET is the weapon of choice. Using different labeling schemes can provide information of the three-dimensional arrangement. Furthermore, using smFRET the protein composition can be quickly changed to simulate complex assembly *in vitro*. As a model system for the study of the eukaryotic H/ACA RNP the snR81 guide RNA from *S. cerevisiae* was chosen. snR81 is involved in the site-specific pseudouridylation of the

U2 spliceosomal RNA at the conserved position Ψ42 and 25S ribosomal RNA at Ψ1051[50]. Furthermore, under nutrient deprivation position 93 in U2 snRNA can be pseudouridylated using the snR81 guide[50].

First of all, the H/ACA components needed to be prepared. On the protein side Nhp2 as well as the NC complex were already available from a previous project. The NCG complex was heterologously expressed in *E. coli* and purified. Since Gar1 is prone to degradation, an optimization of the synthesis was required. Furthermore, Gar1 was produced in six variants that had different combinations of the RGG domains removed. On the RNA side the RNA was prepared with and without fluorophore labeling. For the unlabeled RNA different shortenings of the apical region were synthesized. For a good spatial impression of the snoRNA arrangements, several FRET constructs were generated. Four different labeling sites were chosen in accordance with the fluorophore attachment sites used by Truck *et al*.[68]. From that six different labeling schemes were possible. All these combinations were produced. Two constructs were labeled within the H5 or H3 hairpin, respectively with the second but unlabeled hairpin present. With these constructs the influence of the other hairpin could be investigated when compared to individual hairpin FRET constructs. For the analysis of the three-dimensional conformation of the bipartite complex the fluorophores were placed on different hairpins. Using two constructs the conformational dynamics of the apical and basal parts of the hairpin towards each other can be followed. Two more labeled RNAs gave more distance vectors and insights into the structure.

After preparing all the components it needed to be shown that functional H/ACA complex can be reconstituted from heterologous expression. For that, analytical size-exclusion chromatography and $^{32}$P-activity assays were conducted. Then the conformational dynamics during assembly could be analyzed using smFRET. Here, the individual hairpins were compared with the bipartite RNAs to better understand the influence of the second hairpin. The conformational changes during assembly could be followed using the FRET RNAS with the different labeling schemes. First, without protein, then with a combination of proteins even though this combination might not occur *in vivo* and last the spatial arrangement of the fully assembled complex. With that, information could be gained on the RNA alone, the roles of individual proteins and protein combination as well as the three-dimensional conformations. The last set of experiments focused more on Gar1, specifically its RGG domains. Their role could be studied by replacing the full-length Gar1 with variants lacking different combinations of RGG domains. Overall, the smFRET approach in this work increased our understanding of the bipartite H/ACA pseudouridylase.

# 4 Results

## 4.1 Preparation of H/ACA components

### 4.1.1 Protein preparation

The H/ACA RNP consist of the four protein components Cbf5, Nop10, Gar1 and Nhp2. To study the assembly and structural dynamics of the H/ACA complex, the protein components needed to be expressed and purified. Cbf5, Nop10 and Gar1 were heterologous expressed as a complex in *E. coli* and co-purified. Previous studies had shown that Gar1 with an N-terminal His-tag is not stable and results in up to four bands on an SDS-PAGE[68]. To generate a more homogenous sample two different approaches were tested in this work. First, the N-terminal His-tag of Gar1 was removed (NCG no tag) and second, the His-tag was placed on the C-terminus (NCG Ctag).

As an example, the purification of NCG no tag is described here. Since Cbf5 binds RNA unspecifically[69], several purification steps were necessary to remove endogenous *E. coli* RNAs. The cells were disrupted by ultrasound in the presence of DNase I and RNase A. After centrifugation the supernatant was subjected to PEI precipitation to remove nucleic acids before loading on a Ni-loaded HisTrap. The chromatogram in Figure 10A shows how the cell lysate that was not binding was washed from the column resulting in a decrease of absorbance. The column was washed with buffer supplied with 1 M LiCl to further remove RNAs. After one column volume the absorbance increased to the detection limit which means that UV absorbing biomolecules were washed from the column by this buffer. After the LiCl washing step, the complex was eluted with increasing concentrations of imidazole (Figure 10B). Analysis of the fractions revealed that during the high salt LiCl wash step also some proteins including Cbf5 were washed away (Figure 10C). In the first 13 fractions a protein with the size of Cbf5 was detectable. A band possibly corresponding to Nop10 was found in fractions 1 – 7. In the collected fractions high impurities were found especially between 20 and 30 kDa. Gar1 would migrate in the gel in this range, thus it is not distinguishable from the impurities. Fractions 1 to 11 were pooled for further purification.

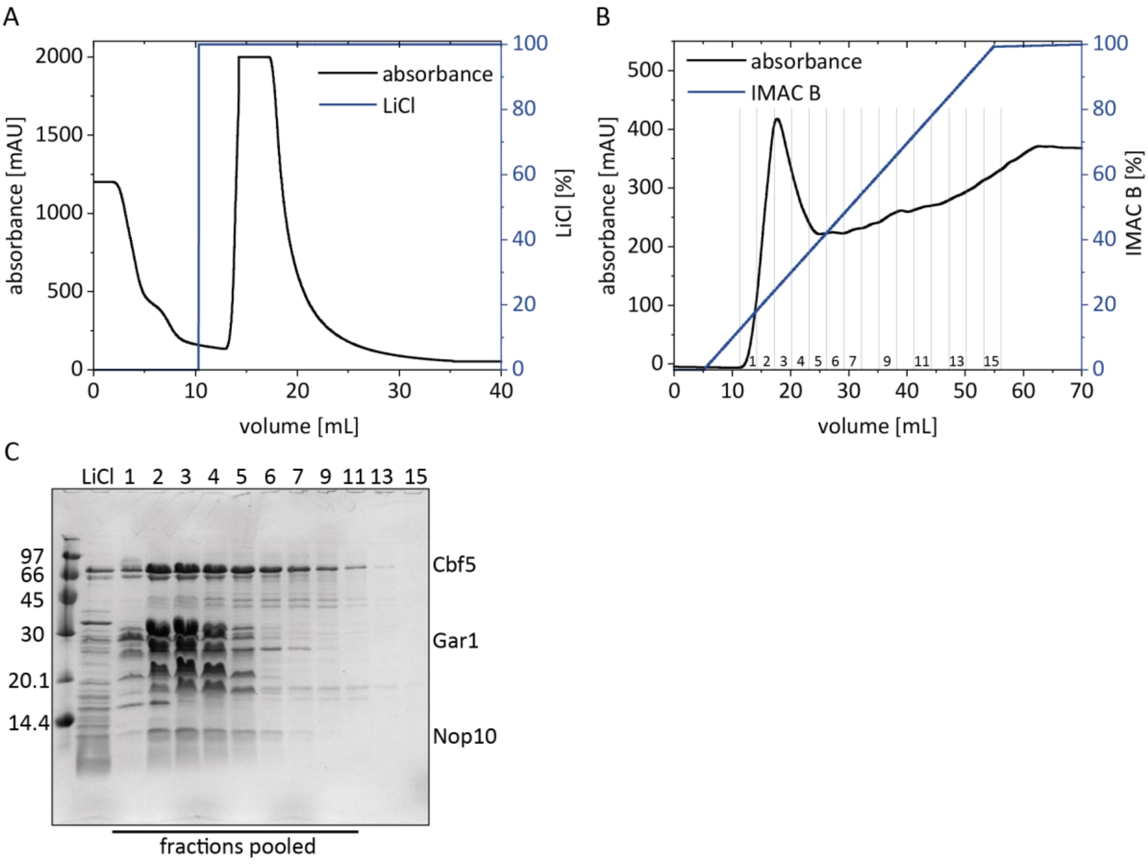


Figure 10: First His-Trap of the purification of NCG no tag. A) Chromatogram of the His-Trap during the column wash and LiCl wash. B) Chromatogram of the His-Trap during the elution. The collected fractions are indicated. C) SDS-PAGE analysis of the collected fractions. The gel was stained using Coomassie Blue.

For further removal of RNA, the proteins were incubated with RNase A overnight. To remove the RNase afterwards a second His-Trap was performed. The chromatogram of this purification step showed one peak (Figure 11A). The impurities around 20 and 30 kDa were still present in the gel (Figure 11A). The fractions containing Cbf5 (fractions 1-11) were pooled and concentrated for size exclusion chromatography.

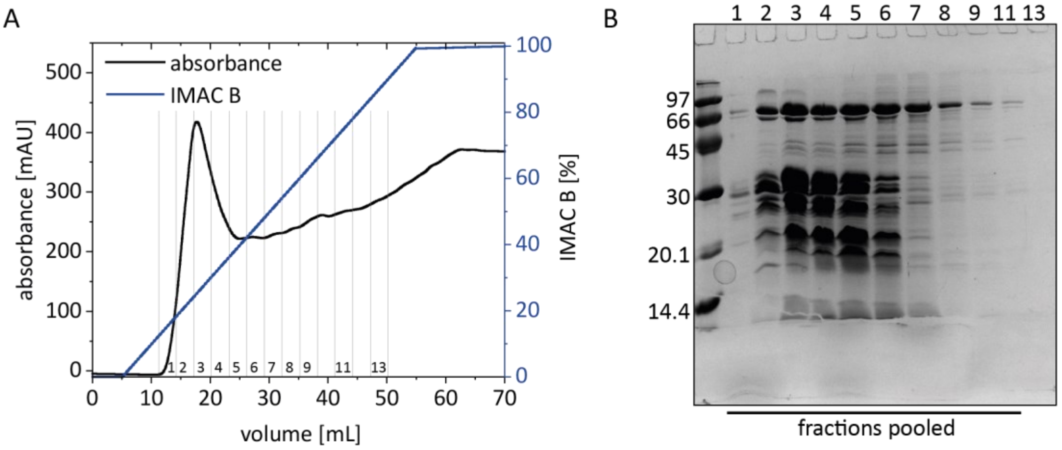


Figure 11: Second His-Trap of the purification of NCG no tag. A) Chromatogram of the His-Trap. The collected fractions are indicated. B) SDS-PAGE analysis of the collected fractions. The gel was stained using Coomassie Blue.

After the second His-Trap a size exclusion chromatography was performed using a Superdex 200 10/300 increase column. The corresponding chromatogram shows one peak around 8 mL of elution volume and a double peak around 14 and 16 mL (Figure 12A). The first peak with an elution volume of 8 mL is the void volume, where very large complexes and mostly aggregates can be found. In the gel several protein bands including one with the size of Cbf5 were detected (Figure 12B). In the fraction 11 to 13 Cbf5, Nop10 and several bands which might be degraded Gar1 were found. In later fractions the impurities eluted from the column. Fractions 12 and 13 were pooled and concentrated. In summary, the complex was successfully purified via IMAC and SEC through the His-tag of Cbf5. This means that both, Nop10 and Gar1 no tag, associate with sufficient strength with Cbf5 to survive the purification steps. However, the purified NCG no tag complex showed several bands corresponding to Gar1.

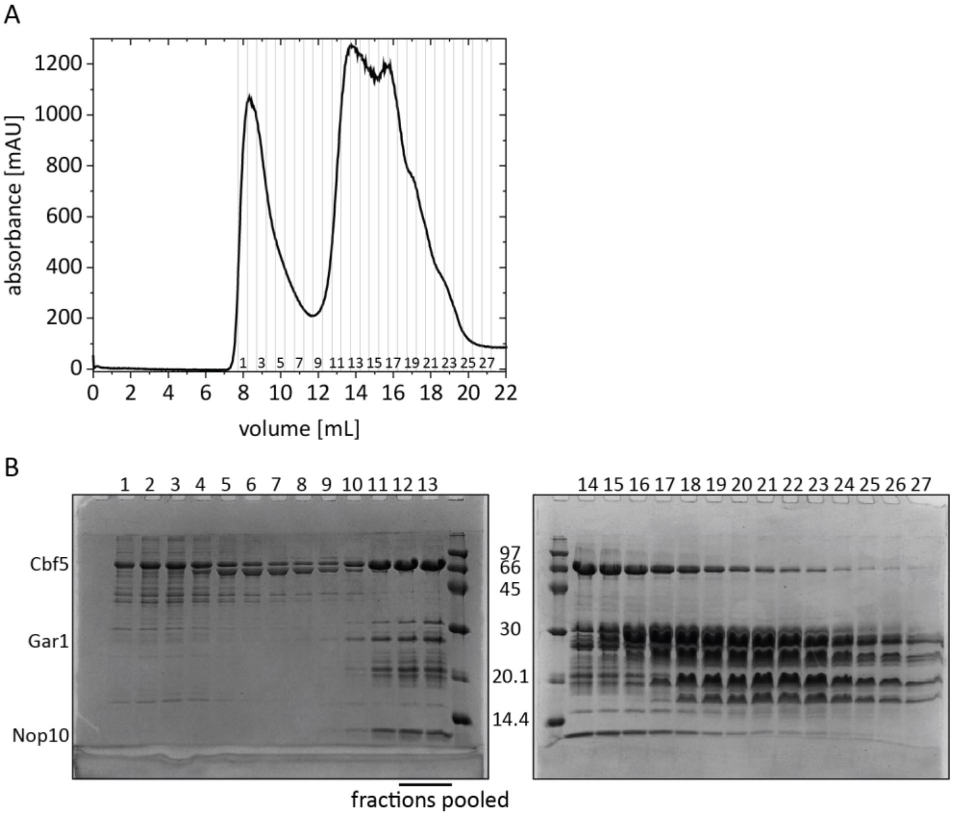


Figure 12: Size-exclusion chromatography of the purification of NCG no tag using a Superdex 200 10/300 increase column. A) Chromatogram of the SEC. The collected fractions are indicated. B) SDS-PAGE analysis of the collected fractions. The gels were stained using Coomassie Blue.

Since Gar1 no tag split in several bands, the His-tag was attached to the C-terminus of Gar1 in a second approach. The complex was purified through two tags, one on Cbf5 and one on Gar1 (NCG Ctag). Furthermore, the His-Trap purification steps were optimized to remove the impurities between 20 and 30 kDa already after the His-Traps (Appendix Figure 52 & Figure 53). At the end it was possible to purify the NCG Ctag complex without detectable impurities or degradation

(Appendix Figure 54). The expression of 6 L resulted in yields between 1 and 2 mg.

In addition to the NCG complex, several variants of Gar1 were synthesized. Gar1 has three RGG domains, one N-terminal of the core domain and two C-terminal. All truncation mutants of these RGG domains were produced. For this work the RGG domains are named 1 to 3. For example, Gar1 Δ3 has the first RGG domain, followed by the core domain and the second RGG domain, while the third RGG domain is missing. Gar1 Δ1,2,3 is only comprised of the core domain. If no further specification is made, the full-length construct was used. All Gar1 variants showed a high degree of precipitation during purification. However, it was possible to purify them in sufficient amounts for the smFRET measurements with only minute impurities (Figure 13, lanes 5-9). The proteins Nhp2 and the Cbf5-Nop10 complex were prepared by Dr. Sven Trucks. The full-length Gar1 protein was synthesized and purified by Nicole Mench. All other proteins and protein complexes were prepared in this work. Figure 13 shows all protein constructs used.

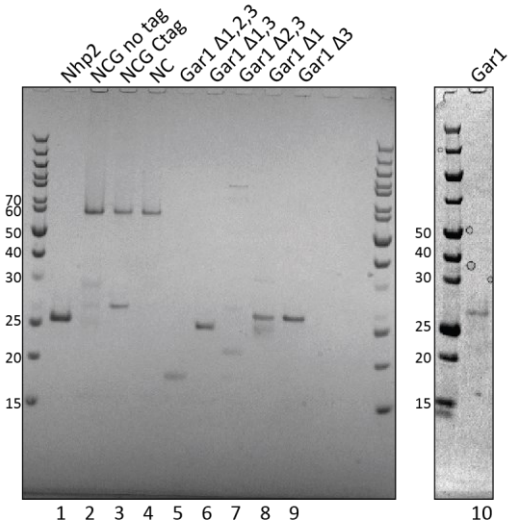


Figure 13: SDS-PAGE analysis of the H/ACA protein components used in this work. The gels were stained using Coomassie Blue.

### 4.1.2 Preparation of unlabeled RNA

For the reconstitution of the H/ACA RNPs the unlabeled yeast snoRNA snR81 needed to be prepared. The guide RNA was prepared as the full-length construct, as well as two constructs with deletions in both loop regions (Appendix Figure 55). In the first shortened construct the large distal loops of both hairpins were replaced by tetraloops (Δ1). In the second shortened constructs even more nucleotides were removed to correspond to the deletions made by Trucks *et al.* for the single hairpins[68] (Δ2). The plasmids coding for the two deletion constructs Δ1 and Δ2 were successfully generated via two cloning steps. To synthesize the RNA a polymerase chain reaction (PCR) was performed from the plasmid RNAs (Figure 14A). The final RNA constructs were transcribed using

T7 RNA polymerase and subsequently purified (Figure 14B). All RNAs showed a second minor band just underneath the main band on a denaturing gel.

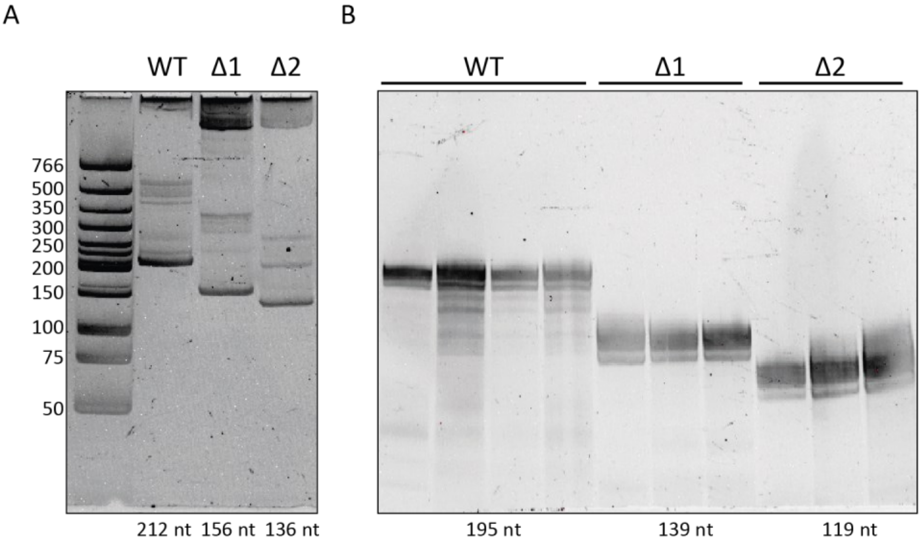


Figure 14: Preparation of unlabeled snR81 guide RNA and two deletion variants. A) Native gel analysis of the PCR reaction. B) Denaturing gel analysis of RNA after transcription and purification. The gels were stained with GelRed.

### 4.1.3 *In* vitro RNP reconstitution

After producing all H/ACA protein and RNA components it was tested whether they assemble into functional complexes *in vitro*. For this, the different snR81 constructs were incubated with NCGW at 30 °C for 1 h. The protein components were used in 4-fold excess to the RNA. This corresponds to double the amount of protein required for a stoichiometric assembly on the two hairpins. The reconstitution was analyzed using analytical SEC. The chromatogram at 280 nm showed two peaks for all RNA constructs used (Figure 15A). The retention volume of the first peak varied with 1.22, 1.29 and 1.33 mL for the full-length WT, Δ1 and Δ2 constructs, respectively. This is in agreement with the longest construct (WT) eluting first and the shortest (Δ2) last. The complexes eluting in the first peaks are larger than the protein or RNA only samples and have high absorbance at both 260 and 280 nm. This means, that some protein components have bound to the RNA. However, the SEC analysis cannot sufficiently differentiate between partially or fully reconstituted complexes. The second peak for all samples eluted at the same retention volume as the protein only sample. The shape of those second peaks with a shoulder or tailing had resemblance with the protein only sample, too. Furthermore, the second peak is not that pronounced in the chromatogram recorded at 260 nm (Figure 15B). For all these reasons it can be concluded that the second peak is mainly caused by unbound protein that was used in excess during reconstitution.

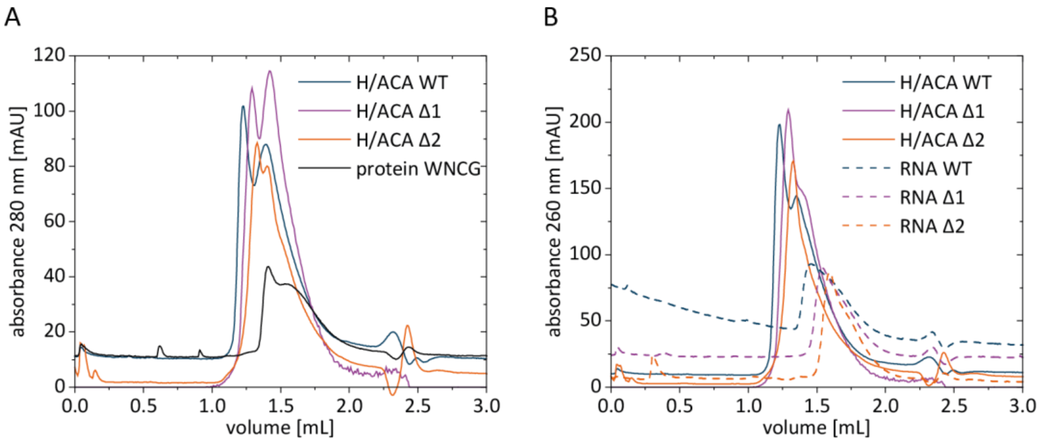


Figure 15: *in vitro* reconstitution of H/ACA complexes. The RNA to protein ratio was in all cases 1:4. Reconstitutions were analyzed using an analytical S200 3.2/300 SEC. The chromatograms detected at A) 280 nm and B) 260 nm are shown.

To assess whether the first peak in the SEC chromatograms is indeed the fully assembled and functional H/ACA RNP, the full-length RNA was reconstituted with an RNA concentration of 5 µM. During the following SEC fractions were collected (Figure 16A). The fractions were individually subjected to UV/Vis spectroscopy. From the absorbance at 260 nm and 280 nm the concentration was determined with an adjusted extinction coefficient which assumed an RNA to protein components ratio of 1:2 (Table 1). If the adjusted concentrations at 260 nm and 280 nm are the same (or the ratio is near 1) this is a good indicator of full assembly. This was the case for fraction 2, which was part of the first peak in the SEC chromatogram (Figure 16A). The 3rd fraction was collected at the end of the first peak which might have some overlap with the second (protein) peak. The concentration ratio of 0.95 reflects this (Table 1). However, it is still close enough to 1 to hypothesize that some molecules in this fraction are fully assembled. Late fractions contained an excess of protein in agreement with the first experiments. To test whether the possibly fully assembled RNP fractions are also functional, fractions 2 and 3 were subjected to $^{32}$P-activity assays under single-turnover conditions. The complexes of both fractions pseudouridylated the 5' substrate almost quantitatively (Figure 16B). The activity was slightly reduced for the 3' substrate with 82% and 77% for fraction 2 and 3, respectively.

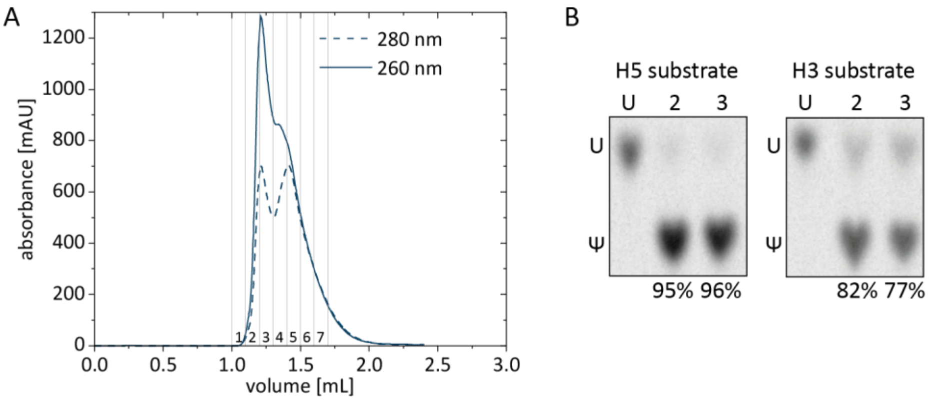


Figure 16: A) Preparative SEC of *in vitro* reconstitution of WT H/ACA complexes using a Superdex 200 3.2/300 increase column. The RNA to protein ratio was 1:4. B) $^{32}$P-radioactivty pseudouridylation activity assay. After the reaction the substrate RNAs were digested and analyzed using thin-layer chromatography. Radioactivity was detected using a storage phosphor screen and imaged on a Typhoon scanner.

Table 1: UV/Vis spectroscopy of the individual fractions of the preparative SEC. The concentrations were calculated according to an adjusted extinction coefficient assuming stoichiometric reconstitution.

| Fraction | Adjusted concentration 280 nm [µM] | Adjusted concentration 260 nm [µM] | Ratio c(260 nm)/ c(280 nm) |
|---|---|---|---|
| 1 | 0 | 0.007 | |
| 2 | 0.266 | 0.267 | 1.002 |
| 3 | 0.353 | 0.336 | 0.951 |
| 4 | 0.595 | 0.400 | 0.672 |
| 5 | 0.484 | 0.296 | 0.612 |
| 6 | 0.406 | 0.242 | 0.595 |
| 7 | 0.174 | 0.059 | 0.340 |

## 4.2 Preparation of FRET labeled snR81 constructs

The aim of this work is the study of the conformational dynamics of the eukaryotic bipartite H/ACA RNP using smFRET. FRET requires a FRET donor and acceptor placed within the complex in a site-specific fashion. To understand the influence of the second hairpin and the three-dimensional architecture, both fluorophores were placed in the snoRNA. As labeling sites the 5' end (G1, site 1), the upper part of the H5 hairpin (U61, site 2), the upper part of H3 (U111, site 3) and the lower part of H3 (U183, site 4) were chosen in accordance to the labeling sites of the isolated hairpins used by Trucks *et al.*[68]. For the maximum number of distance vectors with these labeling sites all combinations of labeling schemes were generated in addition to the isolated hairpin FRET RNAs that were already synthesized by Dr. Gerd Hanspach during his PhD thesis (Figure 17). In this work the constructs were named by their length H5 for the isolated 5' hairpin, H3 for the isolated 3' hairpin and FL for the full-length snR81 sequence. The labeling scheme is indicated in the index. For example, FL$_{1,3}$ contains the whole sequence the first and third labeling site fluorophore labeled.

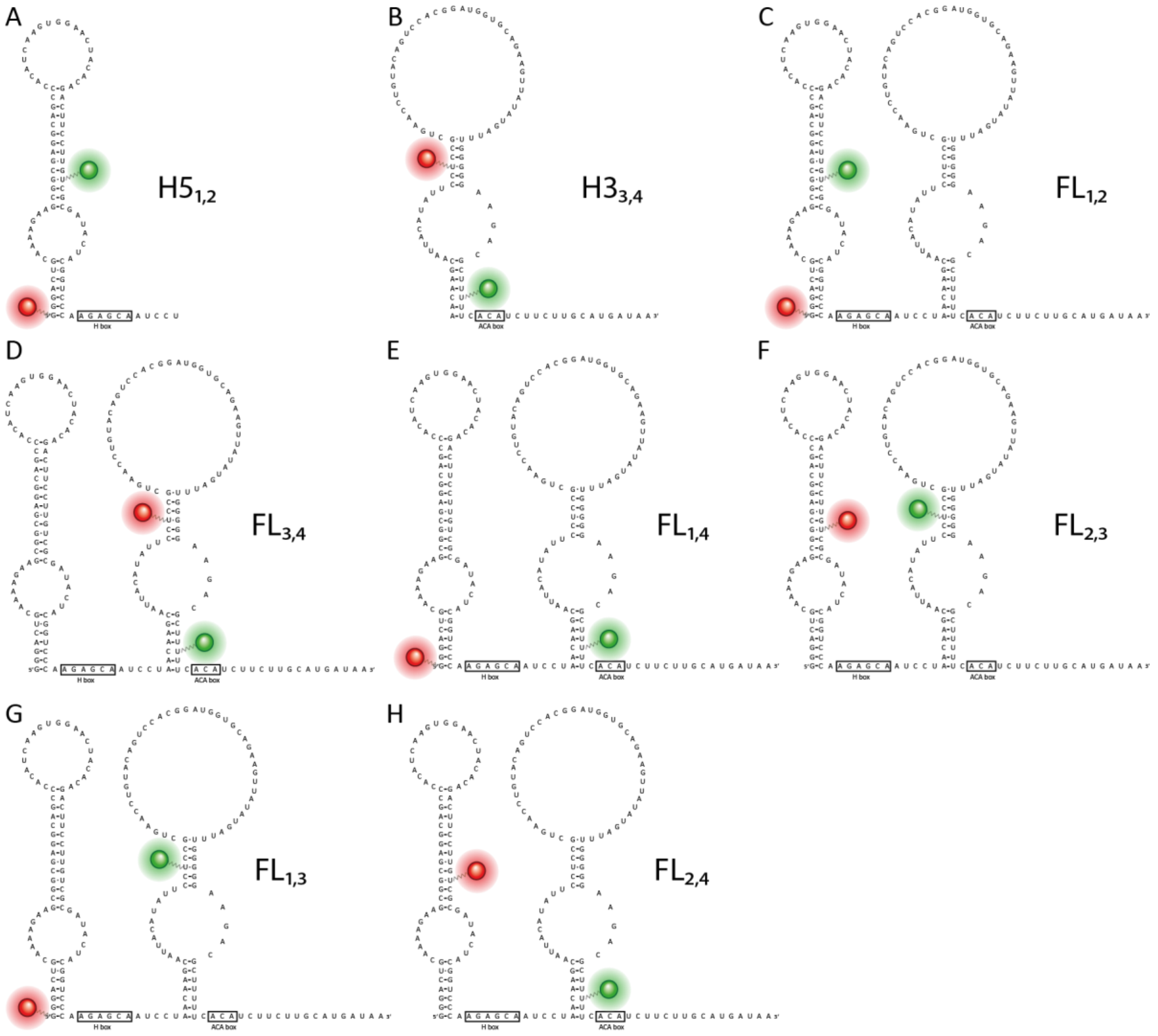


Figure 17: FRET constructs used in this work.

To synthesize these FRET constructs the sequence was split into 6 fragments. The RNA fragments were purchased from Dharmacon in the 2' ACE protected form. The fragments one (M164), two (M167), four (M439) and six (M411) had an amino modifier and the last fragment additionally a biotin modification. The fluorophores were attached to the individual fragments using NHS chemistry. One dye pack was used for three fragments with 30 nmol each. After deprotection the successfully labeled fragments were separated from the unlabeled RNA using RP-LC. Figure 18 shows the chromatograms of these purifications. For all fragments a peak with absorbance at 260 nm was detected around a retention volume of approximately 30 mL. Since there was no sign for absorbance at 550 nm or 650 nm these peaks most likely correspond to the unlabeled RNAs. At increased concentrations of acetonitrile a second peak was detected which had also absorbance at 550 nm or 650 nm for Cy3 or Cy5 labeled fragments, respectively. This second peak corresponds to RNA which is labeled with the fluorophore and was collected and analyzed. In the case of M441_Cy3 there was a third peak in between those two peaks (Figure 18C). The identity of this peak was not investigated further. Since this fragment should also contain the biotin modification, it is possible that some fraction of the delivered RNA lacked the biotin moiety and thus eluted earlier on the reverse-phase. From the chromatograms it was visible that in general the labeling with only one dye pack per three RNAs did not yield high labeling efficiencies as evidenced by a high unlabeled peak in comparison to the labeled peak at 260 nm. According to Figure 18 the worst labeling efficiency had the M439_Cy5 fragment (Figure 18E) and the best was achieved with M164_Cy5 (Figure 18F), which had the labeling site at the 5' end. However, the amount of labeled RNA was sufficient for further applications.

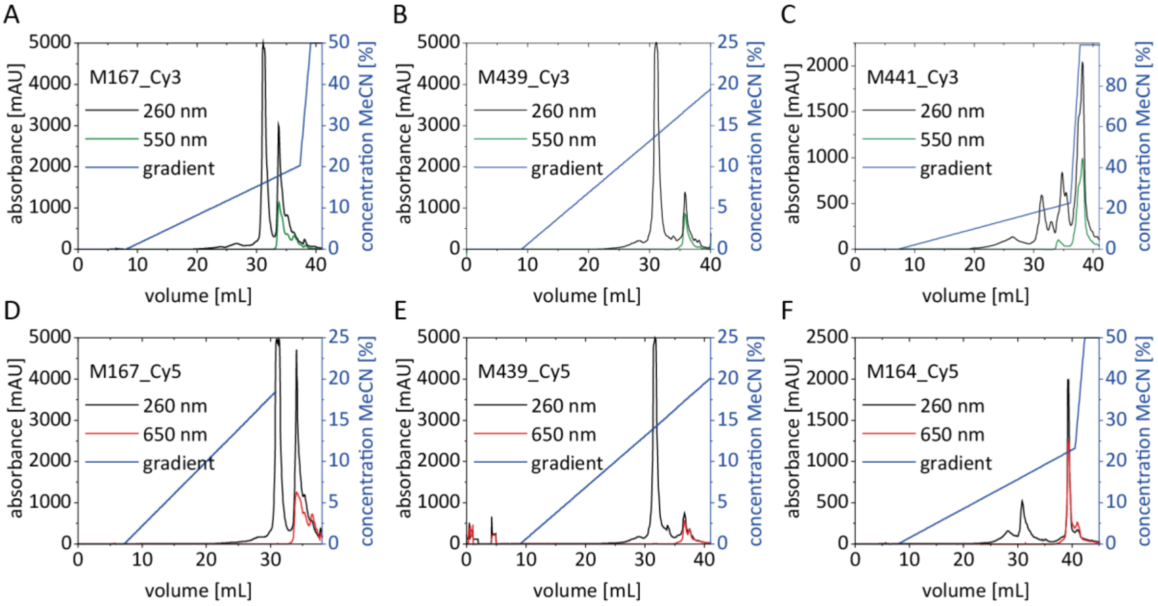


Figure 18: Chromatograms of the RP-LC purifications of the labeled fragments. The RNA was eluted with a gradient with increasing concentrations of acetonitrile. Absorbance at 260 nm and 550 nm or 650 nm was detected during the run.

To synthesize the six different full-length FRET constructs, different combinations of the labeled and unlabeled fragments were ligated in a splinted ligation. For this, two DNA splints without overlap were used. The first and second fragments could bind to the first splint, a part of the third fragment could anneal to each splint and the other three fragments anneals to the second splint. The reaction was catalyzed by T4 RNA ligase 2 followed by heat inactivation and DNase degradation of the DNA splint. Figure 19A shows the fluorescence scan of the analytical gel of these ligation reactions. After ligation (lanes 3 & 4 & 9-12) several bands were detectable in the Cy3 and Cy5 scans which is an indicator of incomplete ligation. Only the RNAs with all six fragments ligated can be used for the smFRET experiments. To prove that the highest bands on the gel indeed represent the completely ligated constructs, test ligations with different amounts of fragments were performed. The $FL_{3,4}$ construct, which has a Cy3 fluorophore at the 3' fragment, was chosen for the tests. The length of the highest band on the gel for each reaction was increased stepwise with every added fragment (Figure 19B). This proves that the highest band for the six-fragment reaction was indeed the full-length FRET snoRNA. Every reaction was referenced to this sample. This highest band of the reactions for the different FRET constructs was excised and the RNA was eluted from the gel. All six full-length constructs with Cy3 and Cy5 labeling were successfully synthesized and purified with high purity (Figure 19C). The yields for the splinted ligation and purification range from 0.27% to 1.29% (Table 2).

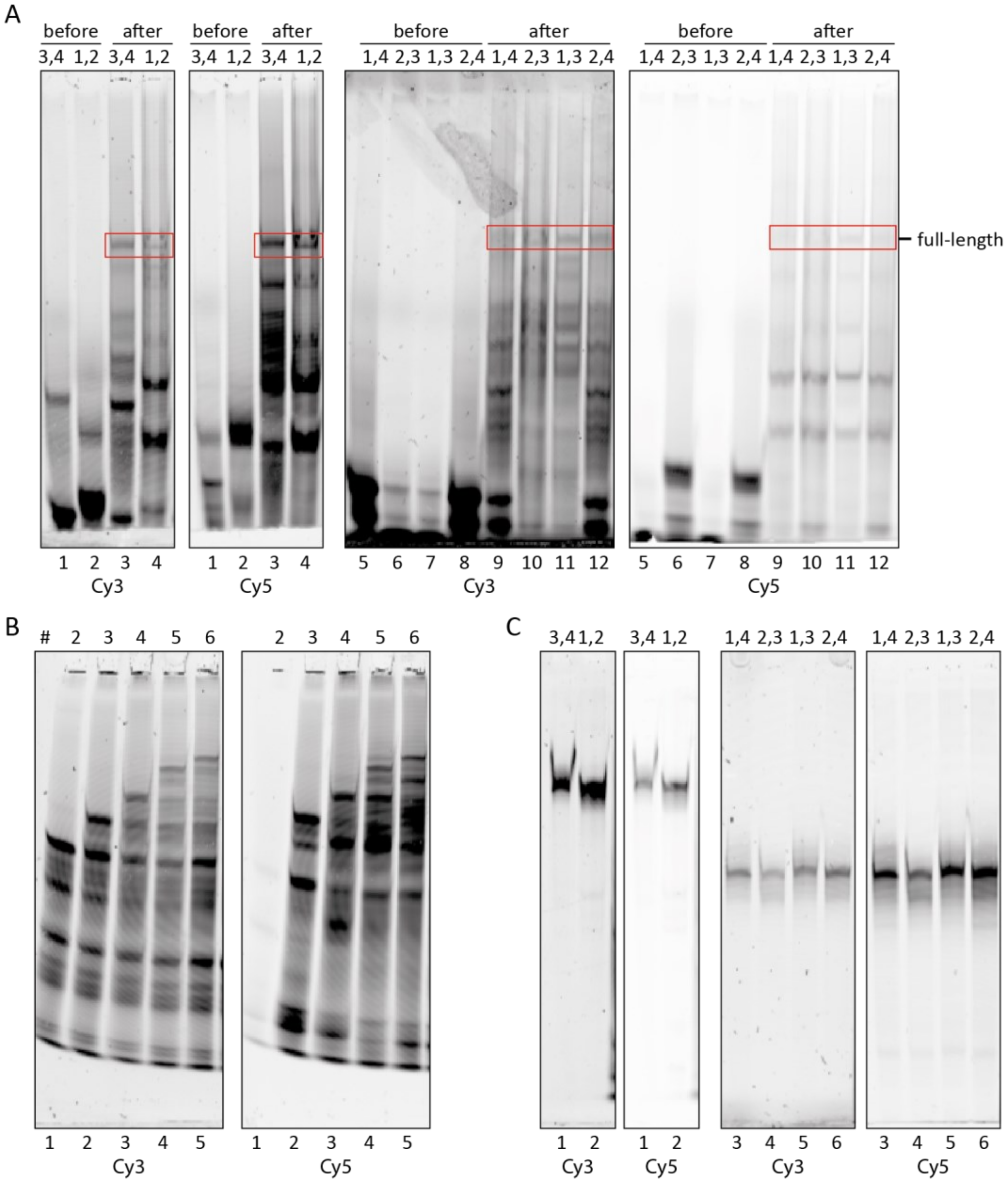


Figure 19: Denaturing PAGE analysis of the splinted ligation of the full-length snR81 FRET constructs. A) Large-scale splinted ligation for all full-length constructs. Samples were taken before addition of the T4 RNA ligase 2 and after ligase incubation and DNase treatment. The band sizes labeled as full-length were excised from the gel and eluted. B) Testligation of $FL_{3,4}$ using different numbers of fragments starting with the two 3' fragments to the complete 6 fragments necessary for the full-length construct. C) Analytical gel of the purified constructs. The labeling schemes are indicated above the gel lanes. All gels were scanned on a Typhoon scanner for Cy3 and Cy5 fluorescence.

Table 2: Yields of the splinted ligation and purification of the H/ACA FRET constructs. 500 pmol were used for each

reaction.

| Construct | Yield [pmol] | Yield [%] |
|---|---|---|
| $FL_{1,2}$ | 4.89 | 0.98 |
| $FL_{3,4}$ | 6.44 | 1.29 |
| $FL_{1,4}$ | 2.27 | 0.45 |
| $FL_{2,3}$ | 1.34 | 0.27 |
| $FL_{1,3}$ | 1.75 | 0.35 |
| $FL_{2,4}$ | 2.27 | 0.45 |

## 4.3 smFRET Results

The main aim of this project was the study of the conformational dynamics of yeast H/ACA RNP complexes using smFRET. After the preparation of both, the protein components, as well as the fluorophore labeled FRET constructs, the smFRET experiments were conducted. The work will focus on four main questions: 1) the structure of the RNA without proteins, 2) the influence of the second hairpin on the conformation of the other hairpin, 3) the three-dimensional architecture of the bipartite complex during assembly and full reconstitution and 4) the role of the Gar1 RGG domains. An overview of the histograms can be found in the Appendix Figure 56.

### 4.3.1 RNA only

As a first point in the assembly of the H/ACA RNP the snoRNA was analyzed in absence of the protein components. To understand the influence of the second hairpin smFRET measurements were conducted with the isolated hairpins used by Trucks *et al.*[68] ($H5_{1,2}$ and $H3_{3,4}$) and the bipartite counterparts with the same labeling sites and an unlabeled second hairpin ($FL_{1,2}$ and $FL_{3,4}$). In all those four constructs the fluorophores are attached at opposite sides of the pseudouridylation pocket. For the isolated hairpin $H5_{1,2}$ a single sharp peak was detected with a FRET efficiency of 0.59 (Figure 20A). Only a few molecules populated states with lower FRET values. The results of the H5 labeled RNA with the attached H3 hairpin ($FL_{1,2}$) were comparable to the isolated hairpin ($E_{FRET}$ = 0.57) (Figure 20B). In contrast the $H3_{3,4}$ construct did not show a homogenous behavior (Figure 20C). Here, two populations were detectable ($E_{FRET}$ = 0.37 & 0.57) with the high-FRET peak being higher. In the full-length RNA the histogram appeared broader, which is a sign for more dynamic or heterogeneous structures Figure 20 (D). Two peaks were fitted to the data. The high-

FRET peak had a similar efficiency ($E_{FRET}$ = 0.59) than the isolated H3 hairpin. The other peak with an efficiency of 0.29 suggests that the fluorophores are further apart in the bipartite RNA than in the low-FRET population of $H3_{3,4}$. The shape of the histogram allowed the possibility of additional populations.

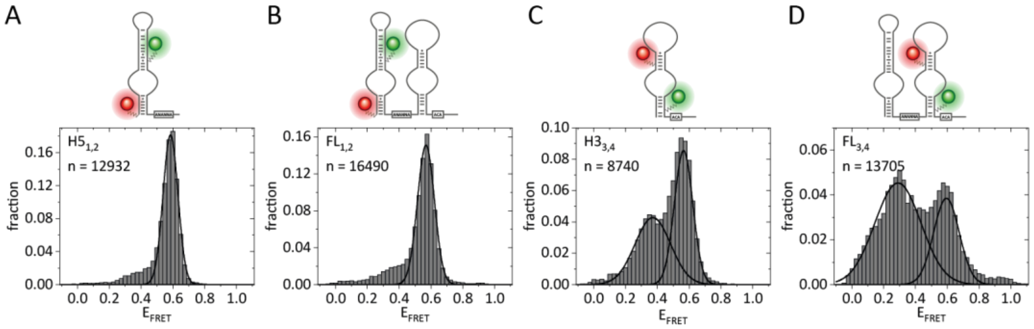


Figure 20: SmFRET analysis of snR81 RNAs labeled in the same hairpin in absence of proteins. A) The isolated hairpin $H5_{1,2}$, B) the full-length $FL_{1,2}$ construct labeled in H5, C) the isolated hairpin $H3_{3,4}$ and D) the full-length construct $FL_{3,4}$ labeled in H3. The Histograms were fitted with one or two Gaussian fits.

After observing differences within the H3 hairpin in presence of the other hairpin the fluorophores were placed on different hairpins. These labeling schemes could shed light on the three-dimensional arrangement of the bipartite structures. In the $FL_{1,4}$ RNA the fluorophores are both in the lower parts of the hairpins. In absence of proteins the RNA folds in different structures represented by three detectable peaks in the histogram (Figure 21A). The main peak had an efficiency of 0.67. The other two peaks were less populated with FRET efficiencies of 0.45 and 0.90. The majority of the molecules displayed FRET efficiencies above 0.5. This means that both fluorophores were rather closer together. The histogram also shows some molecules with FRET values around 0.1. They might either result from insufficient removal of the donor-only peak or could represent unfolded RNA, because in this case the fluorophores may be far from each other almost at opposite ends of the snoRNA.

Using the other two labeling sites in the distal helices of the hairpin structures the $FL_{2,3}$ RNA was analyzed. After removing of the donor-only peak a single low-FRET peak ($E_{FRET}$ = 0.18) was observable in the histogram (Figure 21B). The low FRET efficiency of the two fluorophores at the tops of each of the hairpins suggests a large distance to each other. This is in contrast to the schematical, parallel conformation of the two hairpins found in literature and in the pictograms in Figure 21B because the fluorophores would be in close proximity in this representation. In the histogram of the $FL_{1,3}$ construct two peaks were separable with FRET values of 0.26 and 0.43 (Figure 21C). In the $FL_{2,4}$ RNA three populations could be detected with the middle one being dominant (Figure 21D). FRET efficiencies of 0.1, 0.23 and 0.36 were found for this construct in absence of the protein components.

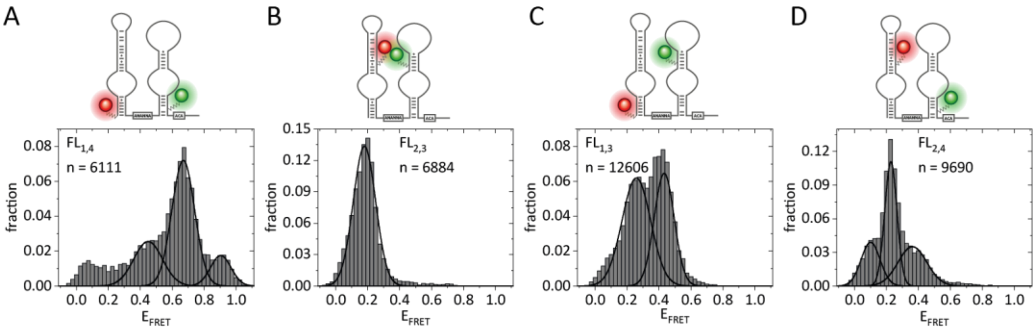


Figure 21: SmFRET analysis of snR81 RNAs labeled in opposite hairpins in absence of proteins. A) the full-length $FL_{1,4}$ construct labeled at the basal helices, B) the full-length $FL_{2,3}$ construct labeled at the apical helices, C) the full-length $FL_{1,3}$ labeled at the 5' end and in the upper part of the H3 hairpin and D) the full-length $FL_{2,4}$ construct labeled at the apical helix of H5 and the basal helix of the H3 hairpin. The histograms were fitted with one to three Gaussian fits.

### 4.3.2 Assembly of the H5 hairpin

The RNA alone is not capable to perform the pseudouridylation reaction. In eukaryotes at least the proteins Cbf5, Nop10 and Gar1 are necessary. The addition of Nhp2 further increases the activity of the complex. For this reason, in a second step the assembly of the RNP complex was investigated. The eight FRET constructs were tested with different combinations of the four protein components. Binding of the pseudouridylase Cbf5 and Nop10 to the isolated hairpin $H5_{1,2}$ resulted into a shift in FRET towards lower FRET efficiencies (Figure 22B). The most populated peak had a FRET value of 0.39. The tail of this peak suggested a second population with 0.49 in comparison to the 0.59 for the RNA alone. The tail vanished with the further addition of Gar1 or in the fully assembled complex. The peak around 0.39 was sharp and the only detectable peak at those two conditions (Figure 22C & D). The sharpness of the peak is a sign for a stable conformation without significant dynamics between the fluorophores.

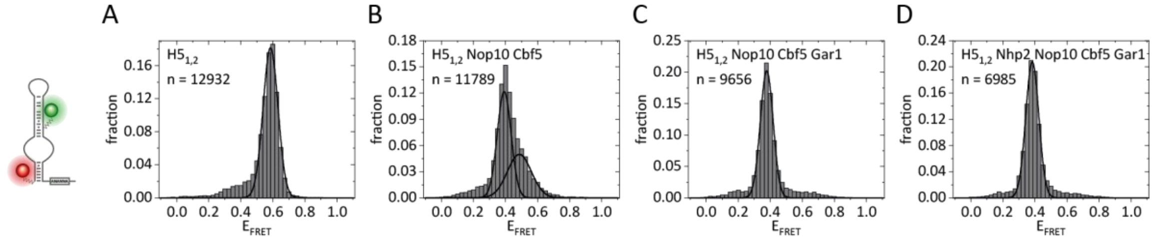


Figure 22: SmFRET analysis of the isolated hairpin $H5_{1,2}$ with different Cbf5 and Nop10 containing complexes. A) RNA only, B) Cbf5 and Nop10, C) Cbf5, Nop10 and Gar1 D) all H/ACA proteins. The histograms were fitted with one or two Gaussian fits.

With an attached second hairpin the construct labeled in the H5 hairpin, $FL_{1,2}$ showed a broad peak in the histogram in presence of Cbf5 and Nop10 with a peak center at 0.45 (Figure 23B). It is possible that there are two peaks similar to the measurement of $H5_{1,2}$ but with an increased fraction of the higher FRET peak (compare Figure 22B and Figure 23B). After binding of the NCG complex the equilibrium was shifted towards the peak at $E_{FRET} = 0.37$, but in contrast to the isolated hairpin this shift was not quantitative (Figure 23C). All four protein components were necessary for a homogenous peak at this FRET efficiency (Figure 23D). The presence of either Nhp2 alone or in combination with Gar1 did not lead to a detectable change in FRET for the $FL_{1,2}$ construct (Figure 24). In Summary, the equilibria of the partial assembled H/ACA complex might deviate for the isolated H5 and the bipartite RNA, but the FRET values for the RNA alone and the fully assembled RNP have a remarkable similarity for this labeling scheme.

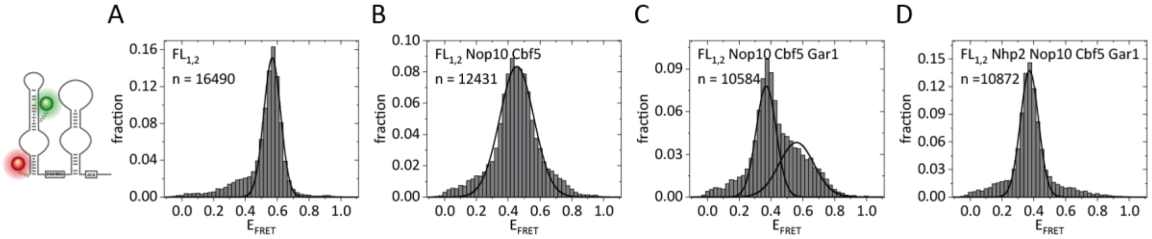


Figure 23: SmFRET analysis of the full-length snR81 RNA $FL_{1,2}$ labeled in the H5 hairpin with different Cbf5 and Nop10 containing complexes. A) RNA only, B) Cbf5 and Nop10, C) Cbf5, Nop10 and Gar1 D) all H/ACA proteins. The histograms were fitted with one or two Gaussian fits.

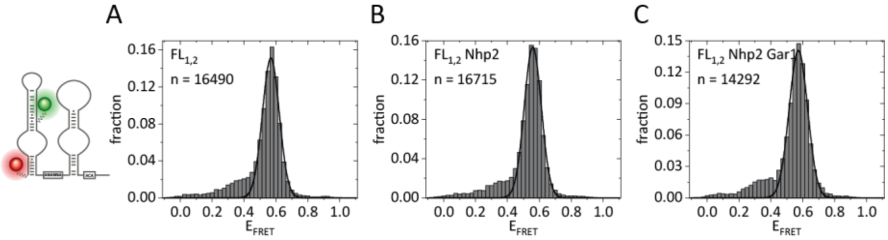


Figure 24 SmFRET analysis of the full-length snR81 RNA $FL_{1,2}$ labeled in the H5 hairpin A) without protein, B) with Nhp2 and C) with Nhp2 and Gar1. The histograms were fitted with a Gaussian fit.

### 4.3.3 Assembly of the H3 hairpin

The conformational changes within the H3 hairpin during H/ACA RNP assembly were also investigated using smFRET. Binding of Cbf5 and Nop10 reduced the fraction of the low FRET populations (Figure 25B). The highest number of molecules were found in the 0.675 bin. However, the peak showed an extensive amount of tailing which might coincide with the two FRET states in absence of proteins. The histogram of the measurements under NCG conditions looked similar except for a possible small shoulder towards higher FRET efficiencies (Figure 25C). In the fully assembled $H3_{3,4}$ construct the effects seen for the Gar1 addition further continued. The number of molecules in the low FRET area decreased, the peak with the highest point at a FRET value of 0.675 further manifests, and the shoulder around 0.8 became more pronounced (Figure 25D). However, the presence of all protein did not lead to a single sharp peak as evidenced for H5 (compare Figure 25D & Figure 23D).

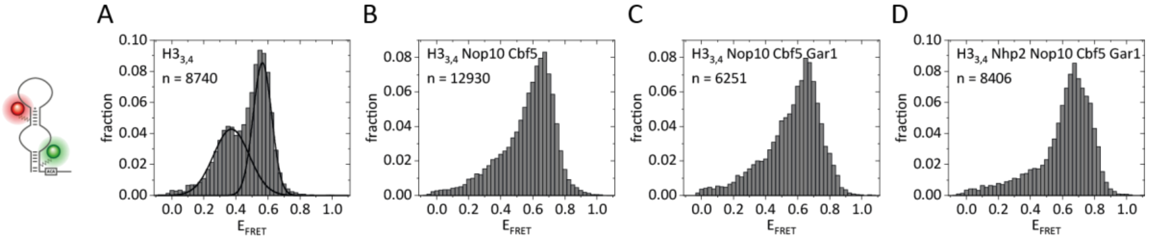


Figure 25: SmFRET analysis of the isolated hairpin $H3_{3,4}$ with different Cbf5 and Nop10 containing complexes. A) RNA only, B) Cbf5 and Nop10, C) Cbf5, Nop10 and Gar1 D) all H/ACA proteins.

The $FL_{3,4}$ construct has the same labeling scheme as $H3_{3,4}$ with the difference that the H5 hairpin is included in the sequence. Addition of Cbf5 and Nop10 to the measurement buffer also reduced the amount of low FRET populations (Figure 26B). The highest bins crowded around FRET values between 0.50 and 0.75, but also some molecules with higher FRET efficiencies were observed in the histogram. In presence of the NCG complex, the $FL_{3,4}$ adopted a variety of conformations for which it proved troublesome to identify distinct states (Figure 26C). With all protein components present the equilibrium shifted in favor of the higher FRET efficiencies (Figure 26D). However, the lower FRET state made up a small but relevant percentage of the molecules. Overall, the experiments with different protein combinations including Cbf5 showed that the bipartite architecture added dynamics and heterogeneity to the H3 hairpin (compare Figure 25 & Figure 26).

The effects of Nhp2 or Nhp2 with Gar1 were also analyzed for the $FL_{3,4}$ RNA. In both cases two peaks with peak centers around 0.3 and 0.57 could be discerned (Figure 27B & C), which were similar to those found in absence of proteins (Figure 27A). The peak with a higher FRET efficiency was higher than in the RNA only measurement shown in Figure 27A. Since the experiments were performed on different measurement days it is possible that the folding effects are not a result of the protein in the measurement buffer but might origin from other causes e. g. temperature.

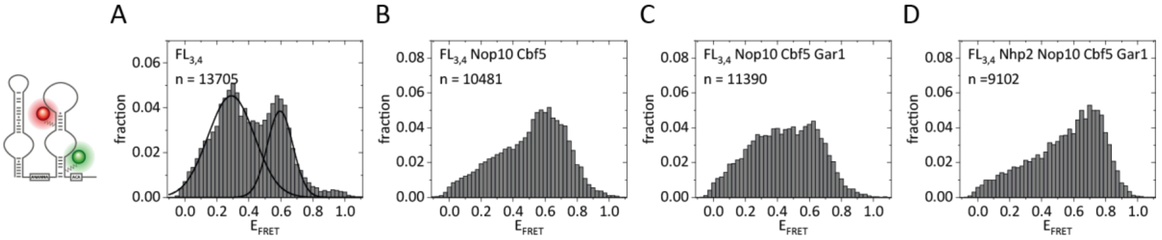


Figure 26: SmFRET analysis of the full-length snR81 RNA $FL_{3,4}$ labeled in the H3 hairpin with different Cbf5 and Nop10 containing complexes. A) RNA only, B) Cbf5 and Nop10, C) Cbf5, Nop10 and Gar1 D) all H/ACA proteins.

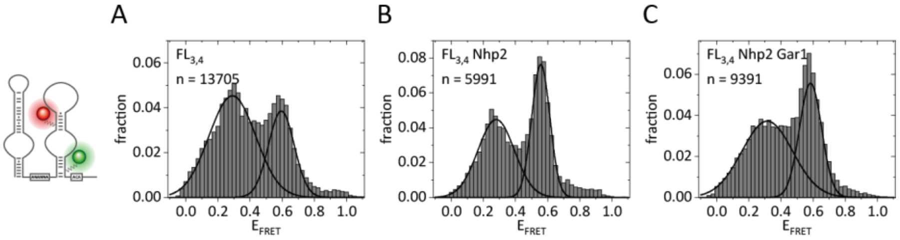


Figure 27: SmFRET analysis of the full-length snR81 RNA $FL_{3,4}$ labeled in the H3 hairpin A) without protein, B) with Nhp2 and C) with Nhp2 and Gar1. The histograms were fitted with one or two Gaussian fits.

### 4.3.4 Assembly of $FL_{1,4}$

Little is known about the three-dimensional architecture of the bipartite complex and especially the roles of the individual proteins. smFRET experiments with RNA constructs with labels at the different hairpins are important to obtain distance information for possibly different conformations. The first RNA under investigation labeled across the hairpins was the $FL_{1,4}$ construct. The RNA only samples showed three peaks with the middle one ($E_{FRET} = 0.67$) being dominant. The results in presence of Gar1 deviate between different measurements. At one measurement no clear difference compared to the protein-free sample was observable. The other two measurements show a peak broadening and a shift to higher FRET efficiencies as depicted in Figure 28B with $E_{FRET} = 0.76$. No structural changes to the RNA alone were detectable in presence of Nhp2 (Figure 28C). The combination of Nhp2 and Gar1 manifested the high FRET efficiency found similarly for Gar1 only ($E_{FRET} = 0.77$). Additionally, some molecules occupied intermediate FRET states in this condition.

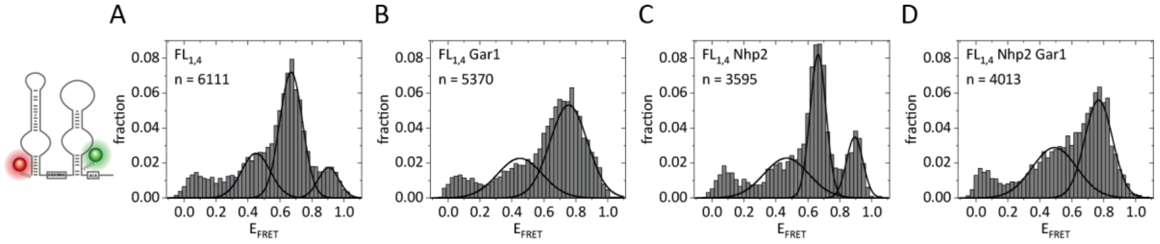


Figure 28: SmFRET analysis of the assembly of $FL_{1,4}$ with different Nhp2 and Gar1 combinations A) without protein, B) with Gar1, C) with Nhp2 and D) with Nhp2 and Gar1. The histograms were fitted with two or three Gaussian fits.

Binding of Cbf5 and Nop10 to the $FL_{1,4}$ RNA resulted in an overall decrease in FRET efficiencies (Figure 29B). The molecules occupied a broad range of states in the low and intermediate FRET range. This variety can be a sign for conformational dynamics. The small high FRET peak that was observed for the RNA only condition was completely absent with Cbf5 and Nop10. The most populated bin of $E_{FRET} = 0.40$ might represent a defined FRET state with this efficiency. Furthermore, there might be a FRET state around the FRET value of 0.6 hidden in the data. There was also a fraction of molecules found with FRET efficiencies under $E_{FRET} = 0.4$. Adding Gar1 to the

RNA-NC complex reduced the heterogeneity slightly. The state around $E_{FRET} = 0.6$ was not detectable anymore during this measurement (Figure 29C). Also, the percentage of low FRET molecules ($E_{FRET} < 0.2$) was reduced. Most molecules had conformations resulting in a broad peak centering at approximately 0.35. However, some dynamic properties remained with the NCG complex. In contrast, the measurement of the WNC complex with $FL_{1,4}$ produced sharper peaks (Figure 29D). A major low FRET peak and an intermediate FRET peak centered at $E_{FRET} = 0.28$ and 0.60, respectively. The shape of the histogram allowed the possibility of more hidden states in the histogram. In presence of all H/ACA proteins this RNA construct folded into conformations that resulted in a sharp peak with a FRET efficiency of 0.33 with some shoulders towards lower and intermediate FRET efficiencies (Figure 30B).

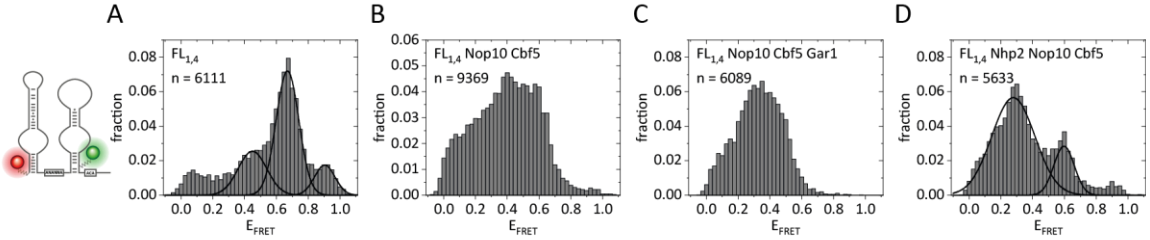


Figure 29: SmFRET analysis of the assembly of $FL_{1,4}$ with different Cbf5 and Nop10 containing complexes A) without protein, B) with Cbf5 and Nop10, C) with Cbf5, Nop10 and Gar1, and D) with Nhp2, Cbf5 and Nop10. When possible, the histograms were fitted with two or three Gaussian fits.

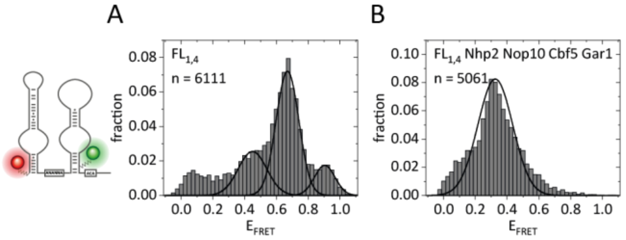


Figure 30: SmFRET analysis of the full assembly of $FL_{1,4}$ A) without and B) with all H/ACA protein components Nhp2, Cbf5, Nop10 and Gar1. The histograms were fitted with one or three Gaussian fits.

### 4.3.5 Assembly of $FL_{2,3}$

In the $FL_{2,3}$ RNA the upper helices of each hairpin are labeled. This allows for monitoring the relative positions of the hairpins towards each other. In absence of proteins the tops of the hairpins were far apart as evidenced by a low FRET efficiency (Figure 31A). In presence of either Gar1 or Nhp2 the same low FRET state was observed in the histogram analysis (Figure 31B & C). From this experiment alone it is not possible to assess whether either protein binds to the RNA. In one out of three replicate experimnets for the Gar1 condition the low FRET peak showed a small shoulder towards intermediate FRET efficiencies. In combination however, the two proteins caused a clear shift towards higher FRET efficiencies (Figure 31D). This means that the fluorophores come closer together which can be interpreted as a tethering together of the tops of the hairpin. The exact peak position, peak broadness and quantity varied between different measurements. However, an increase of the magnitude was only observed with a possible interplay between Nhp2 and Gar1.

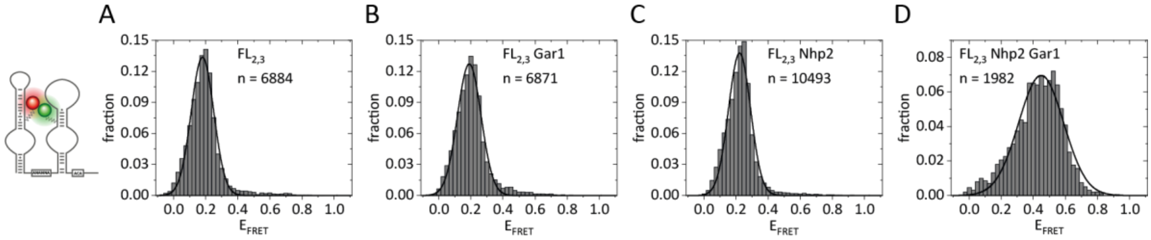


Figure 31: SmFRET analysis of the assembly of $FL_{2,3}$ with different Nhp2 and Gar1 combinations A) without protein, B) with Gar1, C) with Nhp2 and D) with Nhp2 and Gar1. The histograms were fitted with a Gaussian fit.

The experiments with Cbf5 and Nop10 lead to a slight increase in FRET efficiencies from 0.18 to 0.24 compared to the $FL_{2,3}$ RNA only sample (Figure 32B). The broadness of the peak is an indicator for dynamics of the complex. With further addition of Gar1 a smaller fraction at FRET efficiencies around 0.5 emerged (Figure 32C). In presence of the WNC complex a broad distribution of FRET values under 0.5 was detected in the histograms (Figure 32D). Figure 33B shows the FRET results of $FL_{2,3}$ in presence of all H/ACA proteins. Two peaks were identifiable. The low FRET peak had the same FRET efficiency as in the NC condition. The second, less populated peak centered at $E_{FRET} = 0.59$. The ratio between the peaks varied between different measurement days but the higher FRET state was always clearly identifiable as a peak, which was not the case in absence of Nhp2 (compare Figure 32C & Figure 33B).

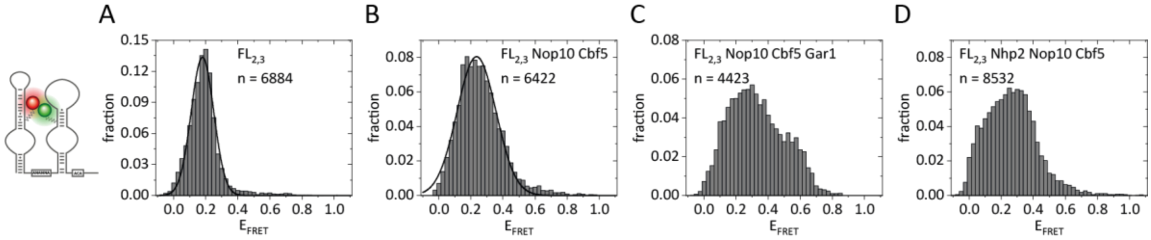


Figure 32: SmFRET analysis of the assembly of $FL_{2,3}$ with different Cbf5 and Nop10 containing complexes A) without protein, B) with Cbf5 and Nop10, C) with Cbf5, Nop10 and Gar1, and D) with Nhp2, Cbf5 and Nop10. When possible, the histograms were fitted with a Gaussian fit.

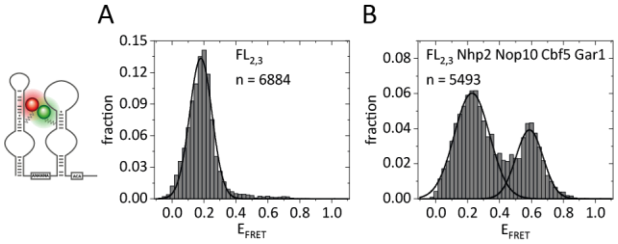


Figure 33: SmFRET analysis of the full assembly of $FL_{2,3}$ A) without and B) with all H/ACA protein components Nhp2, Cbf5, Nop10 and Gar1. The histograms were fitted with one or two Gaussian fits.

### 4.3.6 Assembly of $FL_{1,3}$

The $FL_{1,3}$ construct has the fluorophores attached in the lower helix of H5 and the upper helix of H3. The results for the RNA with Gar1 were not entirely reproducible on each measurement day. During some measurements no Gar1 dependent conformational changes were detectable in the histogram analysis, for others the FRET efficiencies increased to 0.66 as shown in Figure 34B. For Nhp2 alone the data also differed between measurements. In two out of three measurements no changes from the RNA only sample were detected (Figure 34C). In one measurement higher FRET efficiencies were found. Since no other tested construct showed clear signs of Nhp2 binding at the used concentration (1 µM) in absence of any other proteins the latter measurement might be an outlier. A combination of the proteins Nhp2 and Gar1 lead to a decreased distance between the flurophores at the labeling sites 1 and 3 as evidenced by an increase in FRET efficiencies (Figure 34D). The extent of the increase differed between the measurements.

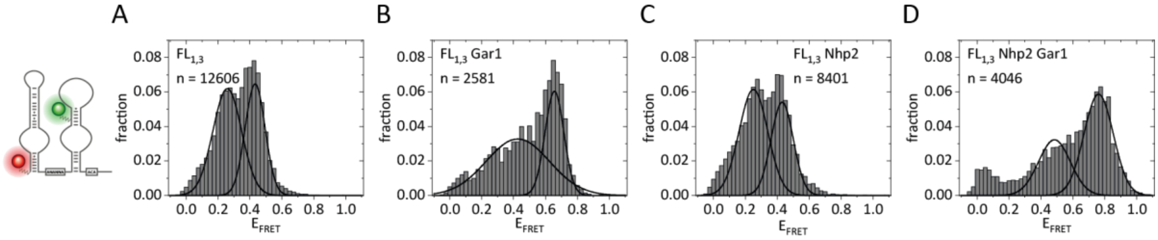


Figure 34: SmFRET analysis of the assembly of $FL_{1,3}$ with different Nhp2 and Gar1 combinations A) without protein, B) with Gar1, C) with Nhp2 and D) with Nhp2 and Gar1. The histograms were fitted with two Gaussian fits.

The $FL_{1,3}$ RNA showed two peaks in absence of proteins ($E_{FRET}$ = 0.26 & 0.43). The peak at higher FRET efficiency almost completely vanished after binding of Cbf5 and Nop10 (Figure 35B). Only a peak at $E_{FRET}$ = 0.20 remained which had a tail towards intermediate FRET efficiencies. With the NCG complex an additional peak centering around $E_{FRET}$ = 0.43 appeared in the histogram (Figure 35C). The measurement with the WNC complex resulted in a histogram similar to when only Cbf5 and Nop10 were present (Figure 35D). When all H/ACA components were in the sample buffer, a major peak with a FRET value of 0.15 and a minor peak around 0.5 manifested (Figure 36B).

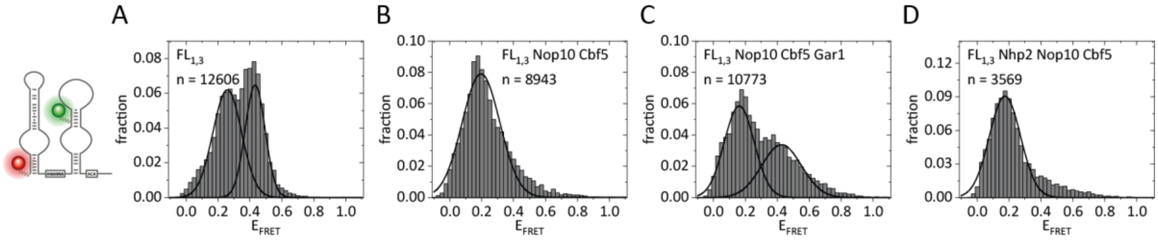


Figure 35: SmFRET analysis of the assembly of $FL_{1,3}$ with different Cbf5 and Nop10 containing complexes A) without protein, B) with Cbf5 and Nop10, C) with Cbf5, Nop10 and Gar1, and D) with Nhp2, Cbf5 and Nop10. When possible, the histograms were fitted with one or two Gaussian fits.

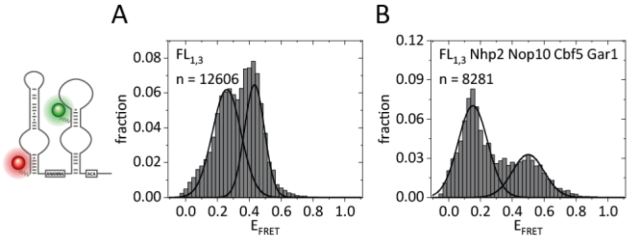


Figure 36: SmFRET analysis of the full assembly of $FL_{1,3}$ A) without and B) with all H/ACA protein components Nhp2, Cbf5, Nop10 and Gar1. The histograms were fitted with two Gaussian fits.

### 4.3.7 Assembly of $FL_{2,4}$

The last labeling scheme under investigation had one fluorophore in the apical helix of H5 and the other in the basal helix of H3. The $FL_{2,4}$ RNA folded into three discernable states in absence of proteins. A similar picture emerges with either the proteins Gar1 or Nhp2 present (Figure 37B & C). For Gar1 the two peaks flanking the major peak are higher than for the other conditions, which might not be representative for Gar1 at other measurements. Using both proteins, Nhp2 and Gar1, in the sample buffer a broad peak centering at $E_{FRET} = 0.4$ was observed in the histogram analysis (Figure 37D). This can be translated that the two fluorophores come closer together but there are also some conformational dynamics introduced into the RNA.

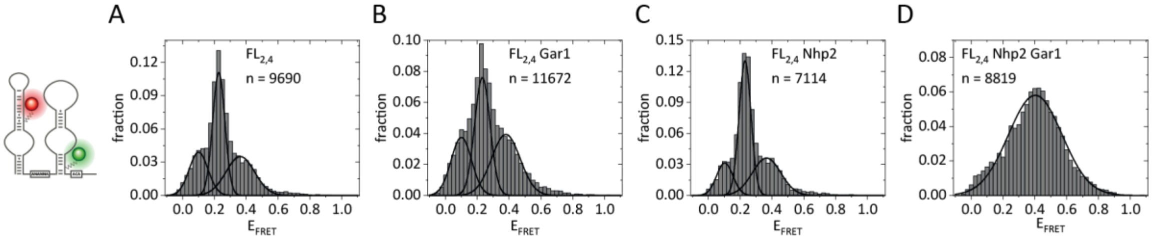


Figure 37: SmFRET analysis of the assembly of $FL_{2,4}$ with different Nhp2 and Gar1 combinations A) without protein, B) with Gar1, C) with Nhp2 and D) with Nhp2 and Gar1. The histograms were fitted with one or three Gaussian fits.

Fitting states to the $FL_{2,4}$ construct in presence of Cbf5 containing protein complexes proved troublesome. With Cbf5 and Nop10 in the buffer, the highest bin was around a FRET value of 0.2 (Figure 38B) similar to the RNA only measurement. The peak had a shoulder towards lower FRET. The ratio between this shoulder and the main peak was increased from the protein-free sample in favor of the lower FRET shoulder. Furthermore, $FL_{2,4}$ with Cbf5 and Nop10 had a long tail towards intermediate FRET efficiencies until approximately 0.7. With addition of Gar1 to the Cbf5 and Nop10 sample the ratios between the low FRET shoulder observed before and the intermediate tail changed (Figure 38C). The shoulder decreased in relation to the main peak and the tail increased. The $FL_{2,4}$ construct in presence of the WNC complex had the opposite effects. Here, the shoulder was more pronounced (Figure 38D). It is not clear whether the changed ratios between main peak, shoulder and tail are in fact reproducible effects of protein binding. In the complex with all protein components in the sample it seemed as if there were two main peaks in the histogram (Figure 39B). One of the peaks might still center around the FRET efficiency of 0.2. The highest populated bin however is in the second peak at the higher FRET efficiency of 0.325.

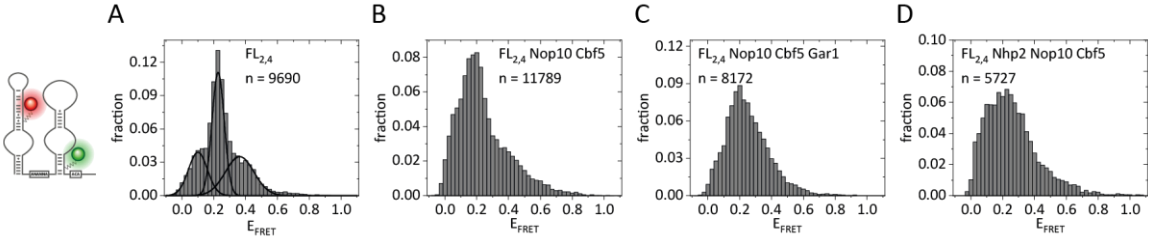


Figure 38: SmFRET analysis of the assembly of $FL_{2,4}$ with different Cbf5 and Nop10 containing complexes A) without protein, B) with Cbf5 and Nop10, C) with Cbf5, Nop10 and Gar1, and D) with Nhp2, Cbf5 and Nop10. When possible, the histograms were fitted with two or three Gaussian fits.

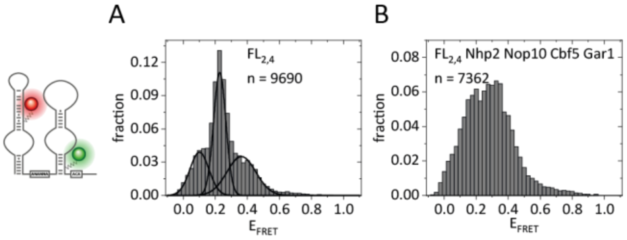


Figure 39: SmFRET analysis of the full assembly of $FL_{2,4}$ A) without and B) with all H/ACA protein components Nhp2, Cbf5, Nop10 and Gar1.

### 4.3.8 Influence of RGG domains

The previous experiments have shown that Gar1 and especially Gar1 in combination with Nhp2 shape the three-dimensional architecture of the bipartite snoRNA. It is not clear which properties of the two proteins are involved in these rearrangements. Eukaryotic Gar1 contains several RG or RGG domains. In S. *cerevisiae* the core domain of Gar1 is flanked by one RGG domain on the N-terminal side and two at the C-terminus. The following experiments aim to test the hypothesis that the RGG domains are involved in the inter-hairpin crosstalk. For this, Gar1 deletion mutants were expressed and purified which lack one (Gar1 Δ1 & Gar1 Δ3), two (Gar1 Δ1,3 & Gar1 Δ2,3) or all RGG domains (Gar1 Δ1,2,3).

The full-length Gar1 induced a peak broadening and shift to higher FRET efficiencies in the $FL_{1,4}$ RNA. In contrast the Gar1 core domain alone (Gar1 Δ1,2,3) did not lead to detectable changes to the RNA only sample (compare Figure 40A & E). From these data it is questionable whether Gar1 Δ1,2,3 is binding to the RNA at all. The histogram in Figure 40F shows that in a combination of Nhp2 and Gar1 Δ1,2,3 the minor peaks found for the RNA only sample disappeared, but a shift to higher FRET efficiencies like for full-length Gar1 did not occur. The FRET efficiencies found for $FL_{1,4}$ RNA only reduced in presence of Gar1 Δ1,2,3 and with all other H/ACA components (Figure

40G). However, the shape of the histogram resembles more the broad distribution caused by Cbf5 and Nop10 than the peak sharpening seen when all full-length components were present. For this reason the conformational changes might most likely be triggered rather by Cbf5 and Nop10 than by the Gar1 without RGG domains.

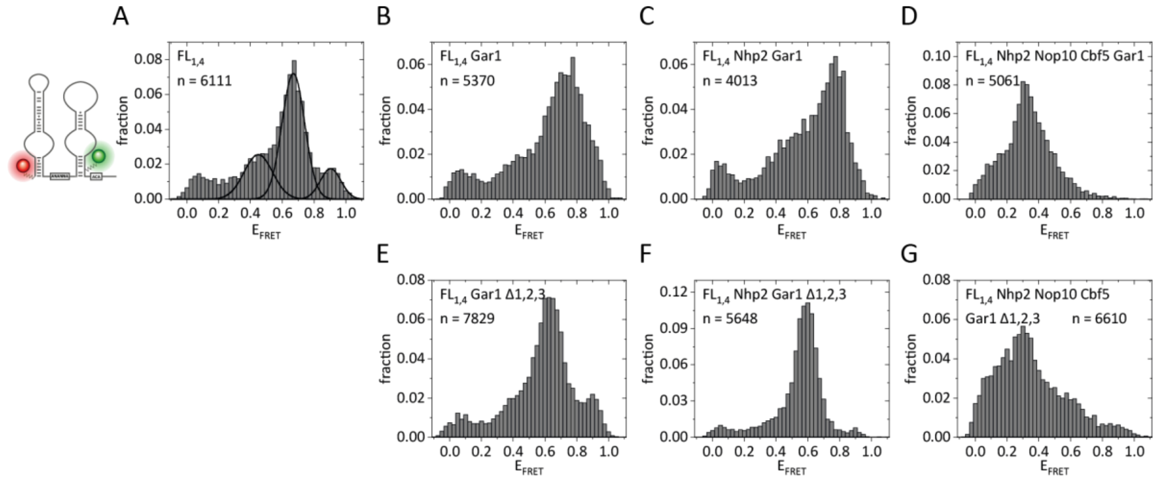


Figure 40: SmFRET analysis of the comparison of Gar1 and Gar1 Δ1,2,3 in the assembly of $FL_{1,4}$.

The interplay of Gar1 and Nhp2 was best visible with the $FL_{2,3}$ labeling scheme. This construct was also used to study the influence of the RGG domains. The RNA and the RNA with full-length Gar1 showed a single low FRET peak, which had a tail towards intermediated FRET efficiencies on some of the measurement days. The majority of molecules populated the same low FRET peak in presence of the Gar1 deletion construct (Figure 40). The measurements for Gar1 Δ3, Gar1 Δ2,3 and Gar1 Δ1,2,3 showed some small tail towards intermediate FRET (Figure 40B, D & F).

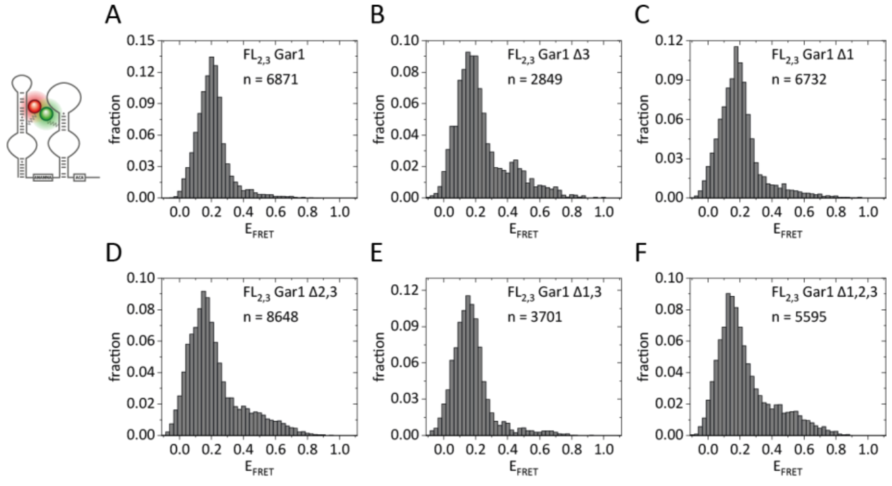


Figure 41: SmFRET analysis of the comparison of different Gar1 RGG deletion constructs on the $FL_{2,3}$ construct.

The combination of Nhp2 and full-length Gar1 lead to a tethering of the two snR81 hairpins detectable by a quantitative increase in FRET efficiency (Figure 42A). The extent of the increase and the peak broadness varied between measurements. Using the Gar1 variant lacking the third

RGG domain again a quantitative shift to a FRET value of 0.43 was detectable (Figure 42B). The peak was sharp, which can be interpreted as a rather static conformation. The Gar1 variants lacking either the N-terminal or both C-terminal RGG domains (Gar1 Δ1 & Gar1 Δ2,3) in combination with Nhp2 had the highest bin still at low FRET, but a large amount of molecules were also found in the intermediate FRET range (Figure 42C &D). For Gar1 Δ1,3 the histogram in Figure 42E showed a peak broadening but only negligible molecules in the intermediate FRET area. The experiment with Nhp2 and the Gar1 Δ1,2,3 construct also showed the main peak in the low FRET range (Figure 42F). This peak had a tail towards intermediate FRET, which was more pronounced in the sample without Nhp2 but not as elevated like the Gar1 Δ2,3 and Gar1Δ1 constructs.

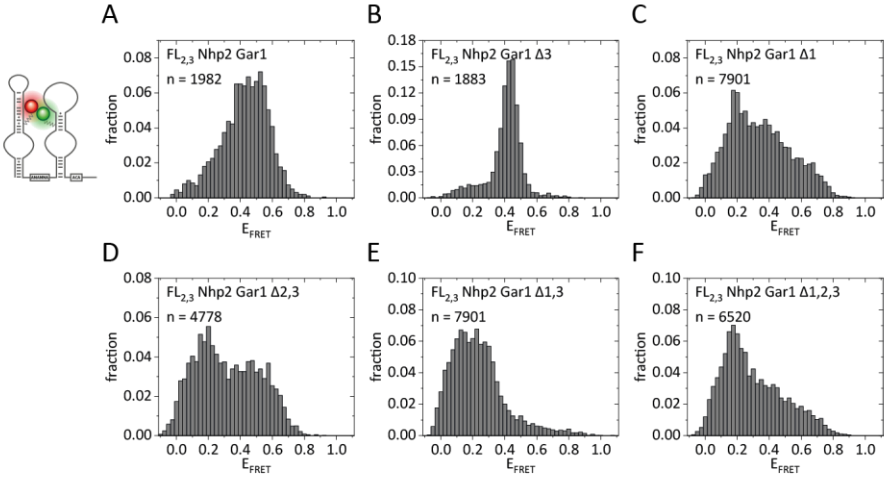


Figure 42: SmFRET analysis of the comparison of different Gar1 RGG deletion constructs in combination with Nhp2 on the $FL_{2,3}$ construct.

Next, the Gar1 variants were analyzed with the other three H/ACA proteins on the $FL_{2,3}$ construct. The full-length proteins had two peaks, one low FRET peak and one around $E_{FRET} = 0.59$. In the Gar1 Δ3 deletion construct with all proteins the same sharp FRET peak was found as for Gar1 Δ3 with only Nhp2 (Figure 43B). It is not clear whether this is a feature of this construct or if there were problems with Cbf5 and Nop10 during this particular measurement. For the other Gar1 deletions constructs in presence of all other proteins a broad low FRET peak was seen in the histogram analysis (Figure 43C-F).

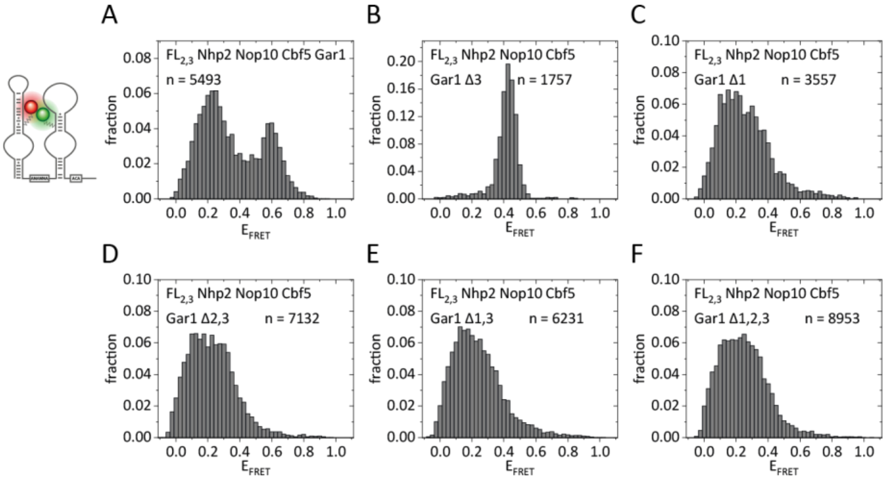


Figure 43: SmFRET analysis of the comparison of different Gar1 RGG deletion constructs on the fully assembled FL$_{2,3}$ construct.

FL$_{1,3}$ was used as another construct to understand the role of Gar1 RGG domains in the assembly of H/ACA RNPs. This FRET construct had two low FRET peaks in absence of proteins and some molecules shifted to FRET efficiencies around 0.66 when Gar1 was added. With Gar1 Δ3 in the measurement buffer most molecules divided between the two RNA only peaks (Figure 44B). A smaller portion was found in a tail towards intermediate FRET efficiencies. FL$_{1,3}$ with either Gar1 Δ1 or Gar1 Δ2,3 which both have either the full C- or N-terminal RGG domains present, showed a broad FRET distribution between around $E_{FRET} \approx 0.2 - 0.8$ (Figure 44C&D). However, it is possible that the RNA only peaks, as well as a peak around $E_{FRET} = 0.7$ are hidden in the histograms. For the proteins with only the second or no RGG domain (Gar1 Δ1,3 & Gar1 Δ1,2,3) no shift towards intermediate or higher FRET efficiencies was detectable (Figure 44E &F).

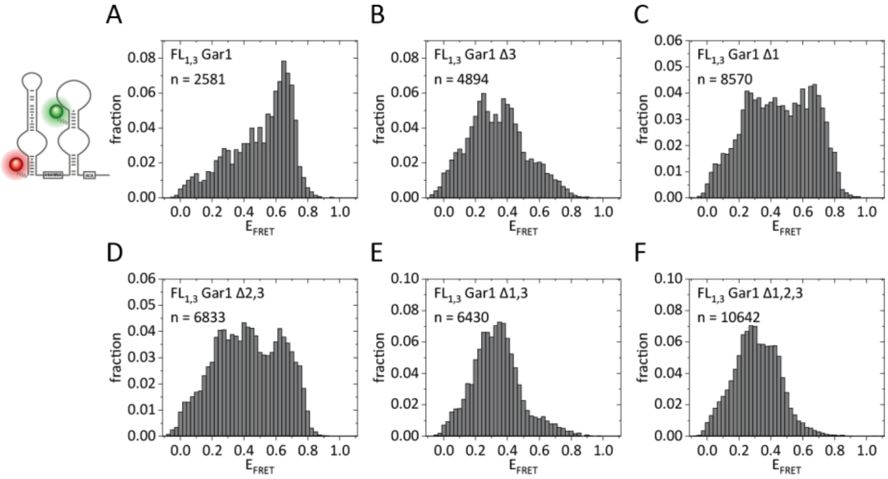


Figure 44: SmFRET analysis of the comparison of different Gar1 RGG deletion constructs on the FL$_{1,3}$ construct.

The results for FL$_{1,3}$ with different Gar1 variants in the NCG complex seemed relatively similar (Figure 45). Two low FRET peaks around FRET values of 0.2 and 0.4 with the 0.2 peak being more pronounced were detected. The histograms show different levels of tailing towards intermediate FRET efficiencies. Gar1 Δ2,3 and Gar1 Δ1,2,3 had the least tailing of all constructs (Figure 45D & F).

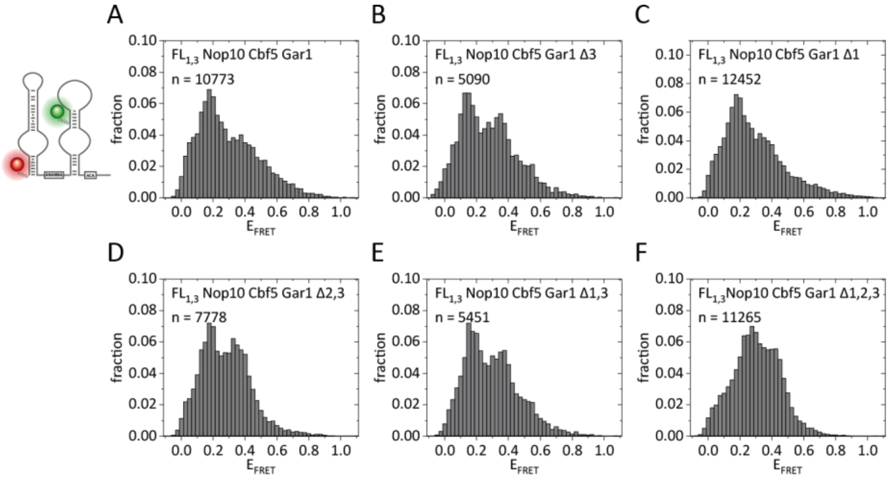


Figure 45: SmFRET analysis of the comparison of different Gar1 RGG deletion constructs in combination with Nop10 and Cbf5 on the $FL_{1,3}$ construct.

For the full assembled complex there were differences between the different Gar1 variants (Figure 46). All constructs had a rather sharp peak around $E_{FRET} = 0.15$. Only in the sample with the full-length Gar1 and Gar1 Δ3 additional minor peaks were found around FRET values of 0.5 and 0.4, respectively. In the construct with all proteins and Gar1 without any RGG domains the low FRET peak was broader than for the other constructs and had a tail towards intermediate FRET.

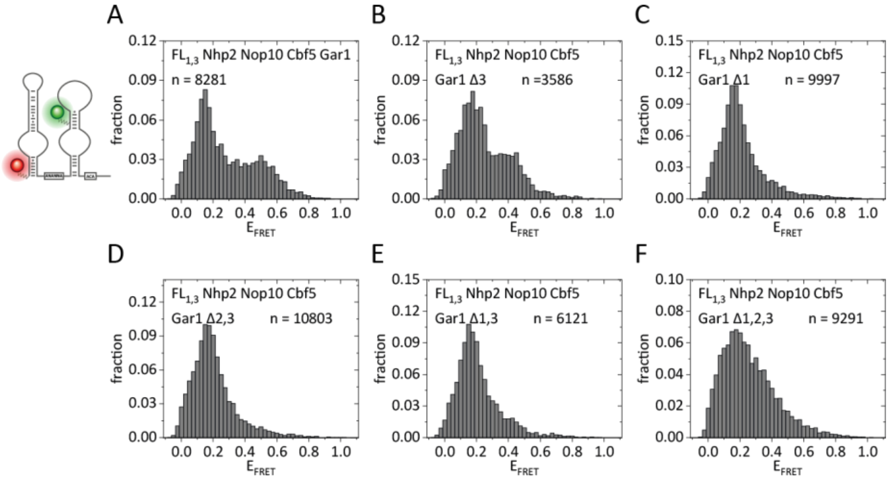


Figure 46: SmFRET analysis of the comparison of different Gar1 RGG deletion constructs on the fully assembled $FL_{2,3}$ construct.

In summary, the experiments with the Gar1 RGG deletion proteins have shown differences in the assembly depending on the presence of individual RGG domain. For some conditions, the results could not be fully reproduced in replicate measurements.

# 5 Discussion

The aim of this project was the study of the eukaryotic H/ACA RNP using smFRET. Several research questions were addressed including the role of the second hairpin, the role of individual protein components, the role of Gar1 RGG domains and the three-dimensional architecture of the bipartite complex. For this, all the individual components needed to be generated first.

## 5.1 Preparation of proteins

The H/ACA RNP consists of a snoRNA and the four proteins Nhp2, Cbf5, Nop10 and Gar1. For the investigation all components were needed. The protein components were synthesized in a heterologous expression in *E. coli* followed by purification. The ternary NCG complex was co-expressed. Since Cbf5 in particular can bind RNA unspecifically, several steps were necessary to remove *E. coli* RNAs. The RNA removal protocols were already optimized in the group and will not be discussed. Previously the complex was purified via a His-tag at the C-terminus of Cbf5 and an additional N-terminal His-tag for Gar1. This resulted in up to four bands corresponding to Gar1. Those bands could be either incomplete protein synthesis or post-translational degradation. In this work the purification for a more homogenous protein was optimized. First, the tag for Gar1 was completely removed. This attempt showed that Gar1 binds sufficiently strong to the other two proteins to be co-purified. However, this did not solve the problem with the bands for Gar1 in the SDS-PAGE. Using a C-terminal His-tag for Gar1 on the other hand resulted in a single band for Gar1 in the gel. The C-terminal tag might stabilize the protein or protect it from protease degradation. Another possibility is that with the C-terminal His-tag only those proteins were purified that did not have a premature translation abortion. In summary, the expression and purification of the NCG complex was optimized in this work and the proteins with this new tag proved to be active in pseudouridylation activity assays.

The purification of the different Gar1 variants alone proved troublesome. The main problem was aggregation of the proteins. Addition of the mild detergent DDM subjectively showed only minor improvements. To remove the aggregates during protein purification several centrifugation steps were necessary. Furthermore, Gar1 is able to induce liquid-liquid phase separation mediated by its RGG domains. Overall, all Gar1 variants were purified even though with major losses due to the aggregation.

## 5.2 Preparation of the FRET constructs

FRET measurements require a construct with a FRET donor and a FRET acceptor. To achieve different labeling schemes, the 195 nt snR81 snoRNA was divided into six fragments. The first, second, fourth and sixth fragments carried an amino modification that allowed fluorophore attachment. Here, Cy3 was used as a FRET donor and Cy5 served as the FRET acceptor. The fragments were labeled individually. RP-LC revealed that for most fragments the labeling efficiency was not good and the majority of the RNA did not carry a fluorophore label (Figure 18). The best results were achieved for 164_Cy5 that has the attachment site at the 5' end. A reason for the bad labeling could be that one dye pack was used for 90 nmol of RNA. Using more dye might improve the labeling efficiency. The main challenge in the preparation of the RNA was the splinted ligation. Because of the length of the RNA two splints were used. T4 RNA ligase 2 was used for the reaction because it proved successful for the ligation of the six fragments for the Gdn-II riboswitch. Performing test reactions adding one fragment at a time showed that all six fragments of the snoRNA could be ligated with the method as evidenced by a step in the denaturing gel with every new fragment (Figure 19B). However, the yields for the full-length H/ACA guide RNA were not as favorable as for the Gdn-II riboswitch. A secondary reason for this might be the twice as long sequence. Also, internal structures for the RNA fragments and the splint DNA might decrease the ligation efficiency. A main factor is probably the use of two DNA splints. Apart from the weak band corresponding to a successful ligation of all six fragments, depending on the labeling scheme other intense bands were visible in the denaturing gel (Figure 19A). This means that some ligation sites did ligate worse than others. Using combinatory it is possible to identify which combinations of fragments probably belong to the intense bands. All constructs with a Cy5 modification in H5 (all except $FL_{3,4}$) showed two intense bands in the scan. For the $FL_{1,2}$ construct that is additionally labeled with Cy3 in the 5' hairpin, bands with the same size were also visible in the Cy3 scan. This was not the case for constructs with a Cy3 at the fourth fragment ($FL_{1,3}$ & $FL_{2,3}$). From that it can be concluded that the two bands likely correspond to the first two or three 5' fragments with a length of 64 nt and 90 nt, respectively. All fragments with a label at the fourth attachment site in the 3' most fragment had a band in Cy3 migrating between the two bands mentioned above. This could be the 79mer consisting of the two 3' fragments. For $FL_{3,4}$ an intense band is detected in the Cy5 scan that migrates similarly to the band identified as the 90mer RNA. It is possible that this is the 93 nt long RNA consisting of M595, M439_Cy5 and 440. Interestingly, several of the incomplete ligations end either before or after the fragment M595. This fragment spans the cut between the two splints. This means first, that the least number of nucleotides stabilize the annealing to the splint, and second, that this fragment might be annealed to only one of the splints and not both at the same time. To improve the yields, it might help to perform an additional annealing step after certain time of ligation. However, all six full-length constructs could be synthesized in sufficient amounts and purity for single-molecule experiments.

## 5.3 Reconstitution of H/ACA RNPs *in vitro*

In the eukaryotic cells H/ACA assembly is a multistep process in different compartments involving several assembly factors like Shq1 and Naf1. For the smFRET studies in this work only the snoRNA and the four final protein components Nhp2, Cbf5, Nop10 and Gar1 were used. It is important that the proteins alone are capable to form the complete and active RNP *in vitro*. To show this, RNA and proteins were reconstituted with a 2-fold excess of protein and analyzed using analytical SEC (Figure 15). Two peaks were detected. The peak with the higher retention volume elutes at the same volume as the free proteins, also the 260/280 nm absorbance ratio suggests that this might be proteins. Changing the RNA protein ratio also changed the ratio between the peaks (data not shown). From that it could be concluded that the second peak represents the unbound proteins. The retention volume of the first peak varied depending on the size of RNA used in agreement with larger complexes eluting at an earlier time. This is an indicator that the first peak contained the reconstituted complexes. In the first fractions the absorbance ratio of 260/280 nm was as expected for the RNA with two sets of the four H/ACA proteins which is a first sign of the completely assembled RNP. Furthermore, activity assays revealed pseudouridylation activities under single turnover conditions of 95% for H5 and 85% for the reportedly less active 3' hairpin [68]. Lack of one of the H/ACA components would decrease or completely abolish the activity. Taken together with the SEC and absorbance results this is a strong sign that it is possible to reconstitute the full and especially active H/ACA complex *in vitro* from heterologous protein expression. This allows the further investigation of *in vitro* reconstituted proteins using smFRET.

## 5.4 Effects of the bipartite structure

In eukaryotes the H/ACA RNPs have a conserved bipartite structure. Activity data had shown that this dual structure increases the activity by a so far uncharacterized mechanism. What three-dimensional features are responsible for changed activity remains unclear. Previous work has analyzed the assembly of the snR81 isolated hairpins using smFRET[68]. For my experiments I used the same sequence and labeling sites, but additionally included constructs with both hairpins present.

### 5.4.1 H5 hairpin

Trucks *et al.* detected an open, low FRET and a closed, intermediate FRET state for the isolated hairpin $H5_{1,2}$ construct[68]. In their study the equilibrium shifted only slightly to the open conformation with the ternary NCG complex and had 85% of the molecules populating this conformation after full assembly. Repeating the experiments with their construct, my FRET efficiencies were in agreement with their two peaks (Figure 22). However, the ratios during assembly differed. A major shift towards the open conformation was already detected with the two proteins Nop10 and Cbf5. The 55 kDa protein Cbf5 binds to the H-box and places its catalytic center towards the target uridine in the pseudouridylation pocket. From my data it seems that Cbf5 and Nop10 are responsible for the reorientation and opening of the pocket. Since some molecules were still found in the intermediate FRET state, it is possible, that Cbf5 could allow both, the open and the closed conformation. It is more likely that those molecules did not have a bound Cbf5 thus remaining in the RNA only, intermediate FRET state. Using the NCG complex or all four H/ACA proteins, the shift to low FRET was quantitative. The quantitative shift of the equilibrium might be caused by the additional Cbf5 protein in the measurement buffer, rather than the effects of Gar1. It is surprising, that the literature and this work on the same construct generate different results, in particular for the NCG condition where the amount of low FRET is 28% to 100%. A reason for this might be the different sample preparation techniques used. Trucks *et al.* pre-annealed the RNA with the proteins, diluted the complex and immobilized it on the slide, before washing away unbound RNA and protein. This might lead to the problem of conformational discrimination during the immobilization on the slide that might overrepresent a certain conformation. Furthermore, there were no proteins in the buffer to rebind a hairpin that has lost the proteins. In this work on the other hand, the RNA was immobilized, and the proteins were used at the concentration of 1 µM in the measurement buffer. As all RNAs had the same starting condition during immobilization, selection processes may have been avoided. Additionally, the protein concentration during the measurement was always high which might have saturated the RNA better. Taken together, the only slight shifts of the equilibrium in literature might be caused by incomplete binding of Cbf5 to the RNA in the experiment. Having an excess of proteins in the measurement buffer seems to be the method of choice to increase the likelihood but does still not guarantee that the RNA has the respective protein bound.

The differences between the two techniques might give further evidence to the already proposed role of Nhp2 in anchoring the RNP. The H5 hairpin has a reduced activity without Nhp2. The low FRET state was proposed to represent the active conformation. With the dilution technique, Nhp2 was necessary for the shift of the equilibrium towards this conformation. Cbf5 and Nop10 alone were responsible for the shift in this work, which means that Nhp2 was less important in presence of an excess of Cbf5. It is possible that Nhp2 somehow prevents dissociation of Cbf5 and Nop10 and is important when reassociation is hindered e. g. by a lack of free Cbf5 in the buffer with the dilution technique.

For the bipartite $FL_{1,2}$ the peak centers in the histograms can be compared to the open and closed states in absence of the second hairpin (Figure 23). However, the equilibrium during assembly had changed. With Cbf5 and Nop10 a broad peak or two overlapping peaks were found. In the first case, the peak width would represent a dynamic structure, but with the peak shape it is more likely that under these conditions the two states were similar populated. In the bipartite complex the equilibrium improved towards the proposedly active, low FRET state with NCG but was not quantitatively shifted as in the individual hairpin. Nevertheless, in the fully assembled complex a homogenous peak was visible. The differences between the isolated hairpin and the bipartite RNA might originate from the second binding site at the opposite hairpin. Since the same protein concentrations were used for both experiments, it is possible that binding to the second hairpin removed free Cbf5 from the solution, which would make Nhp2 stabilized binding more important. It is also possible that the presence of the second hairpin blocks efficient Cbf5 binding e. g. by steric hindrance. Experiments using $FL_{1,2}$ in combination with Gar1 and Nhp2 (Figure 24) contributed to earlier conclusions that they cannot open the pseudouridylation pocket and that Cbf5 is required for that. In summary, the effects of the second hairpin for H5 are limited which is in agreement with the similar activity data for both RNA lengths[68].

### 5.4.2 H3 hairpin

As for the 5′ hairpin, the 3′ hairpin was also investigated with labeling sites across the pseudouridylation pocket. The conformations of the two H3 labeled constructs ($H3_{3,4}$ & $FL_{3,4}$) were generally less homogenous than the H5 hairpin even in absence of proteins. A reason for this might be the large apical loop of H3 which could potentially assume different structures or might even disrupt the helix-bulge-helix structure of H/ACA hairpins. $H3_{3,4}$ had two discernable conformations (Figure 25). In comparison to previous studies on this construct[68] more molecules populated the low FRET peak, but the intermediate peak was still the major peak. Also, the FRET efficiencies were shifted towards lower values in this measurement. The reason for this remains unclear. Binding of Cbf5 and Nop10 abolished the low FRET conformations and increased the overall FRET efficiencies. As witnessed for the H5 hairpin, in H3 this leads to rearrangements of the pseudouridylation pocket, possibly to allow substrate binding. Rearrangements and torsion effects probably lead to more proximity between the fluorophores. It is possible that Cbf5 also binds to conformations that are not active or aids the folding of the correct basepairing. In contrast to the published data, in my experiments only the combination of all four H/ACA proteins could bring a fraction of the molecules to a high FRET state. With just NCG only a shoulder for high FRET could be assumed. If the high FRET state was the active conformation this would explain why Nhp2 is essential for 3' substrate turnover. In the full-length snoRNA sequence the histograms were more heterogenous (Figure 26). Again, Cbf5 induced rearrangements decreased the low FRET state but also lead to a general peak broadening which is an indicator of dynamics. In the bipartite H/ACA containing telomerase structure the two Cbf5 molecules interact with each other. This might have an impact on the dynamics in the H3 pseudouridylation pocket but only minor effects on the H5 conformation. In the $FL_{3,4}$ construct all protein components including Nhp2 were necessary to shift the equilibrium in favor of higher FRET efficiencies. Previous work has shown that Nhp2 binds in three different conformations in the fully assembled isolated H3 hairpin[68]. It is possible that for the $FL_{3,4}$ construct also three conformations were hidden within the histogram. The effects of Nhp2 addition could be a result of binding to and rearrangement of the large H3 loop. Interestingly in the telomerase structure the Nhp2 from the H5 hairpin is in close proximity with the H3 protein components, and even interactions seem possible[82]. Maybe it is not only the H3 Nhp2 but also the H5 Nhp2 that triggers the conformational changes leading to the high FRET shift in the H3 pseudouridylation pocket. In contrast to the H5 hairpin constructs, there is an influence of the second hairpin in the assembled complex. According to literature, the full-length RNA should be more active[68]. However, from the FRET data it is not possible to conclude which of the many states is the catalytically active conformation. In summary, as expected the second hairpin does influence the conformational dynamics for H3, which might be the reason for the different activities of the two lengths.

## 5.5 snR81 conformation in absence of proteins

To better understand the three-dimensional structure of the bipartite complex which might help understand the changes in activity in presence of the second hairpin, the labeling sites were also placed on different hairpins. The aim was to analyze the conformational dynamics during assembly of the H/ACA RNP. The first point of the assembly to be investigated was the RNA alone without any protein factors. All eight FRET constructs were measured (Figure 20 & Figure 21). The first observation was the heterogeneity of the RNA conformations. The labeling sites located exclusively in H5 resulted in a narrow single peak. From that it could be concluded that the structural variability in the H5 region might be limited. For $FL_{2,3}$ also only a single peak low FRET was detected. In the low FRET regime, distance changes between the fluorophores result in only small changes in the FRET efficiency, compared to distances around the Förster radius, according to equation 1. For this reason, several high distance conformations might be described by the FRET peak. The other constructs all had more than one FRET peak. The labeling schemes comprising the third labeling site ($H3_{3,4}$, $FL_{3,4}$ & $FL_{1,3}$) had two peaks. Since the 1,2 labeling scheme showed a stable conformation it is likely that the local structural changes leading to the peak splitting into two peaks are the same for $FL_{1,3}$ and $FL_{3,4}$. For $FL_{1,3}$ the two FRET efficiencies are close together, which is a further indicator that $FL_{2,3}$ also might have two conformations which cannot be resolved in the low FRET range. The constructs with the fourth labeling site, $FL_{1,4}$ and $FL_{2,4}$ had three distinguishable populations. One main peak and two smaller peaks with higher and lower FRET efficiencies. For $FL_{3,4}$ the effects of the third and fourth labeling site might have come together. One of the peaks split by the third site structure deviations had two shoulders. It is possible that the ratios correspond to the three peaks originating from the fourth site. The different conformations in the region of the 3' hairpin could have different origins. H3 has a large apical loop that might disturb the helix-bulge-helix secondary structure expected for eukaryotic H/ACA hairpins or induce other basepairings. In absence of proteins that would organize the pseudouridylation pocket, it is possible that it can adopt different conformations. It is also possible that one or both of the helices in H3 do not form according to the proposed secondary structure. The different heights of the peaks might indicate different stabilities of the alternative structures.

In the construct $FL_{1,4}$ the basal helices of the hairpins were labeled. This construct showed a high FRET efficiency without proteins. If the secondary structure is folded as proposed, the fluorophores are separated mostly by the linker between the hairpins. The linker could be unstructured but might be extended due to electrostatic repulsion. The torsion angle between the two fluorophores might be favorable to generate the high FRET efficiency. The peak width does not suggest extensive dynamics of the linker. In literature the bipartite snoRNA structure is usually depicted in a parallel conformation of the two hairpins. In that case the label attachment sites for $FL_{2,3}$ would be in close proximity as illustrated in Figure 47A. A close proximity would result in a high FRET efficiency for the construct in that case. The opposite was the case in the smFRET measurements in absence of proteins. $FL_{2,3}$ had a low FRET peak with $E_{FRET} = 0.18$ under these conditions. This means that the apical parts of the hairpins are rather further apart. For this reason, I propose a conformation for the free snoRNA where the hairpins point in different directions like in the pictogram in Figure 47B. Then the fluorophores would be at greater distance and the conformation would fit better to the data. The sequence of snR81 with a shortened H3 loop and the secondary structure was subjected to the RNAcomposer online tool[105,106] to give an estimation about the three-dimensional structure. Here, the two hairpins point more or less in opposite directions (Figure 47). Using the FRET positioning and screening (FPS) software, for the $FL_{2,3}$ an average inter-dye distance of 109 Å was calculated with a FRET value estimation of 0.031. This might be far off the experimental FRET efficiency but it shows that the fluorophores can be far apart. With different rotation angles it would be realistic that the $E_{FRET} = 0.18$ for $FL_{2,3}$ represents a conformation with hairpins in opposite directions. The distances of $FL_{1,3}$ and $FL_{2,4}$ detected by FRET were in between the long distance of $FL_{2,3}$ and the short distance of $FL_{1,4}$ in agreement with the opposite conformation.

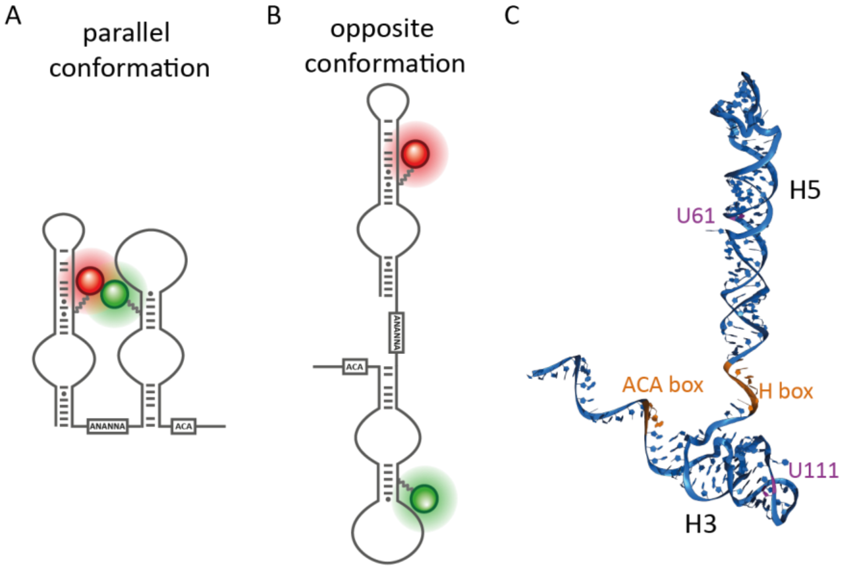


Figure 47: Model for snoRNA conformation in absence of proteins. A) Schematic representation of a parallel conformation as depicted often in literature. The fluorophores indicate the labeling sites 2 and 3 B) Schematic model of the hairpins pointing in opposite direction as suggested by the smFRET data. C) 3-dimenional structure calculation using snR81 without the H3 loop by RNAcomposer. The nucleotides labeled in $FL_{2,3}$ are shown in magenta, the H- and ACA box are shown in orange.

## 5.6 Roles of Cbf5

After proposing a conformation where the free RNA has its hairpins in opposite direction the effects of individual protein components on the three-dimensional orientations were investigated. Cbf5 is known as the catalytic subunit. It binds to the H- or the ACA-box and spans over the pseudouridylation pocket where it determines the unpaired target uridine at the apical part through distance dependence. The small protein Nop10 is important for Cbf5 stability and Nhp2 binding. It is very unlikely that Nop10 is the main factor for NC induced conformational changes. Therefore, the focus will be on Cbf5 with the ambiguity that Nop10 might also cause those differences. As discussed above, Cbf5 binding rearranges the pseudouridylation pocket in the constructs with the first two or last two labeling sites ($H5_{1,2}$, $FL_{1,2}$, $H3_{3,4}$ & $FL_{3,4}$). This made sense with the conservation of the distance between the H- and ACA- box and the target uridine and the role of Cbf5 as a molecular ruler. Additionally, the smFRET histograms of the constructs labeled at different hairpins change with NC in the buffer (Figure 48A), suggesting that Cbf5 might also play a role in the three-dimensional arrangement of the hairpins. In $FL_{1,4}$ the high FRET conformations were abolished in favor for a variety of low FRET states. This translates as a distance increase between the fluorophores and an introduction of dynamics. Also, the constructs $FL_{1,3}$ and $FL_{2,4}$ had a reduced FRET efficiency in presence of NC. Only the construct $FL_{2,3}$ where the apical parts of the hairpins are labeled, showed an increase in FRET efficiency and a decrease in distance. With the relative distance information of those four constructs, I propose a conformation where the two hairpins are bend in presence of Cbf5 (Figure 48B). Cbf5 binding in H5 to the H-box in the linker and the H5 stem might orientate the hairpins and distorts the linker. Such a reorientation might twist the fluorophores in $FL_{1,4}$ further away from each other leading to the lower FRET values. Comparing a completely extended conformation with the hairpins at opposite sites with a bent conformation, the distance between the apical parts of the hairpins would decrease. This is reflected well by the FRET data of $FL_{2,3}$ with an increase in FRET efficiency. In the structure of the human telomerase the sequences correlating to the H/ACA hairpins are also in a bend conformation with several overlapping electron densities that might represent protein-protein interactions. In the telomerase structure, the two molecules of the Cbf5 homolog dyskerin interact in the linker region. If this arrangement can be transferred the bipartite H/ACA complex of *S. cerevisiae,* this might explain how Cbf5 could lock the hairpins in a certain orientation towards each other. In general, the peak width increased with binding of Cbf5. This is found most pronounced in $FL_{1,4}$ but also in the other constructs. This means that the complex with Cbf5 is dynamic. While the protein-free RNA showed up to three different conformations it seemed as if Cbf5 binding does not discriminate between those conformations. This is best evidenced by the constructs $FL_{1,3}$ and $FL_{1,4}$ where the different former FRET states are only sparsely populated after Cbf5 binding. Either all those free RNA conformations are Cbf5 binding competent on their own, or Cbf5 works as a chaperone to refold the RNA. In summary, I propose that Cbf5 rearranges the pseudouridylation pocket, distorts the linker between the hairpins and bends the orientation of the hairpins towards each other.

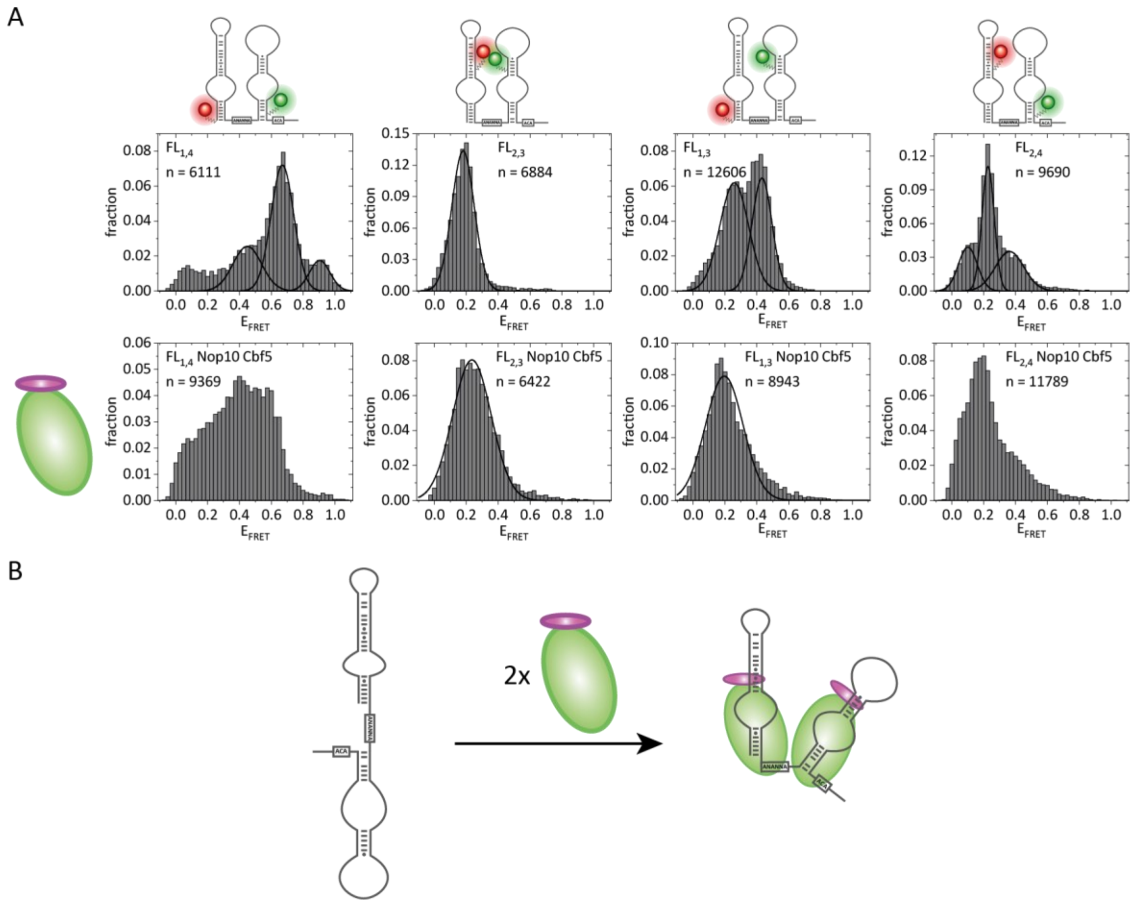


Figure 48: Effects of Cbf5 and Nop10 binding on the 3-dimensional structure of the bipartite complex. A) Histogram analysis with and without NC. B) Schematic model of the effects of NC binding. Green: Cbf5; magenta: Nop10.

## 5.7 Roles of Nhp2 and Gar1

In the eukaryotic cells, WNC bind to the snoRNA co-transcriptionally as a ternary complex. Gar1 binds at a later point in time in exchange for the assembly factor Naf1. However, to understand the involvement of the individual proteins, it might be interesting to analyze the assembly in a different order even though those complexes would probably not exist *in vivo*. In that regard several experiments with Gar1 were conducted, lacking some of the other protein components. Unfortunately, the reproducibility of the smFRET experiments involving Gar1 was challenging. At some measurement days adding Gar1 to the RNA or to an RNA-protein complex did result in detectable conformational changes while on other days the histograms remained as the previous condition or only partial changes were observed. This also includes peak width among the same conditions involving the protein Gar1. For that reason, the experiments with Gar1 should be considered carefully. Especially in cases where no change was detected this may not necessarily mean that Gar1 is not capable to induce a change at all. A reason for the inconsistencies could be that the used concentration of 1 µM was close to the $K_D$ under the respective conditions and small changes in concentration would have a large impact on the binding. Furthermore, Gar1 is also involved in liquid-liquid phase separation. It is possible that the stock was not sufficiently mixed and that either a Gar1 enriched or Gar1 reduced phase was pipetted resulting in different concentrations in the measurement buffer. Also, Gar1 aggregated during the purification. In the small volume of the aliquot used it would not be possible to see whether that was the case. This would also affect the actual concentration pipetted to the microscope slide. To still be able to gain insights from the experiments with Gar1, effects that were observed at least once will be considered for discussion. No conclusion will be made if Gar1 had no effect on the conformational dynamics because the effective Gar1 concentration might have been below the binding affinity.

For the constructs $FL_{1,4}$ and $FL_{1,3}$ conformational changes in the presence of Gar1 were detected by the smFRET analysis. In the case of $FL_{2,3}$ one measurements had a small shoulder. This shows that Gar1 is able to bind RNA and is bound to at least one of the hairpins independent of other protein factors. For all three mentioned constructs the distance between the fluorophores decreased. However, the binding site or the structural rearrangements induced by Gar1 could not be concluded from that. Gar1 is supposed to bind to the thump loop of Cbf5 in the fully assembled complex. From the FRET data it is unclear whether the direct Gar1-RNA interaction would persist in the full complex in addition to this Gar1-Cbf5 interaction.

For Nhp2 alone no reliable conformational changes were detected in the FRET histogram analysis. This means either that Nhp2 binding would not change the distances between the combinations of labeling sites used in this work, or that Nhp2 alone is not capable of binding in absence of other protein factors. For snR34 an affinity of Nhp2 alone was determined with 600 nM[69]. This value might be different for the snR81 sequenced used here. However, the affinity is in a similar working range with the working concentrations of 1 µM for the smFRET experiments. It is possible that the protein concentration of Nhp2 was not sufficient to bind a large fraction of the RNA.

Gar1 had a small impact on the three-dimensional structure of the snoRNA and Nhp2 had no influence. In combination the two proteins completely remodel the bipartite RNA according to the FRET experiments with the labeling schemes at the different hairpins (Figure 49). No shifts were observed for $FL_{1,2}$ and $FL_{3,4}$ that are labeled at the same hairpin. From that it can be concluded that the interplay of Nhp2 and Gar1 affects the orientation of the hairpins towards each other and might not have a big influence on the pseudouridylation pocket. In all constructs with a labeling scheme at different hairpins the FRET efficiency with WG increases compared to the free RNA. The two proteins bring the labels in proximity by somehow tethering the hairpins closer together.

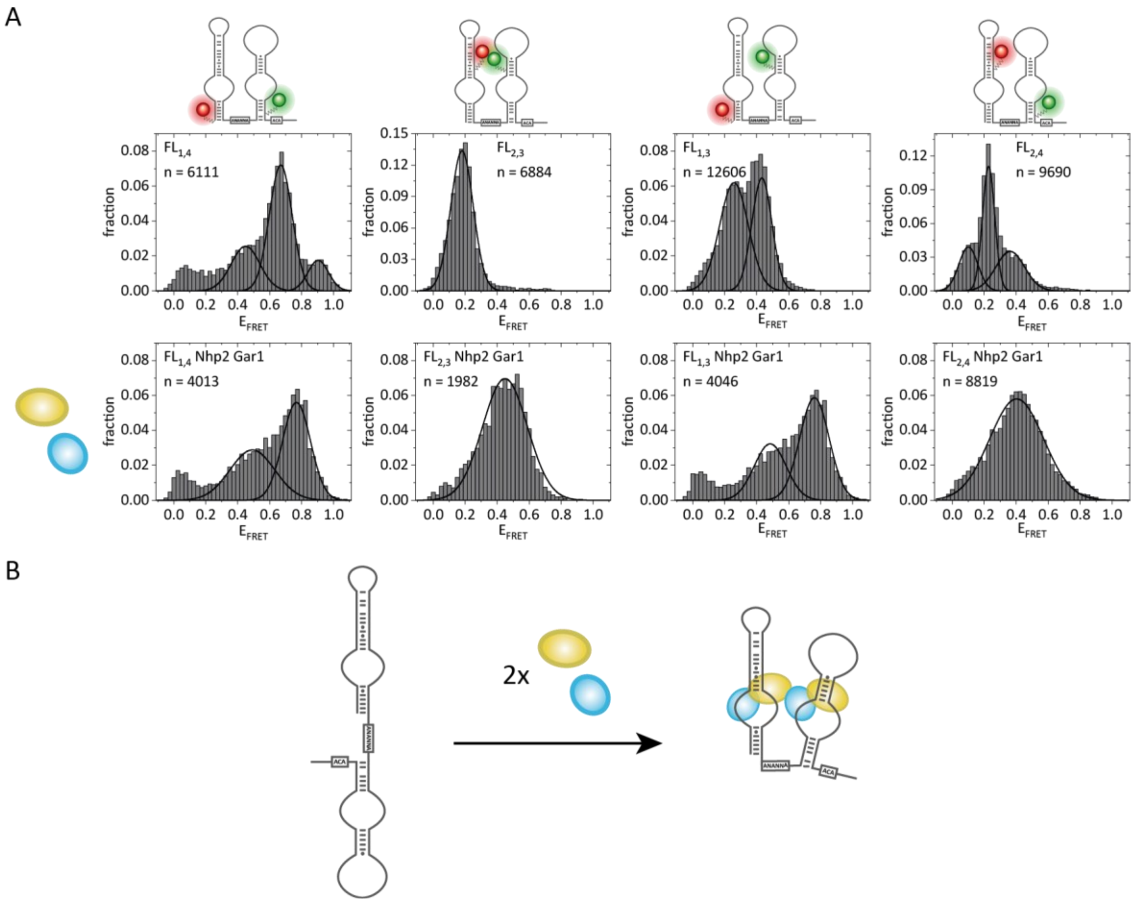


Figure 49: Effects of Nhp2 and Gar1 binding on the 3-dimensional structure of the bipartite complex. A) Histogram analysis with and without WG. B) Schematic model of the effects of WG binding. Blue: Gar1; yellow: Nhp2.

There are several possible explanations how the combination of Nhp2 and Gar1 could bring the hairpins in proximity. First, this might be an artifact of the above-mentioned problems when working with Gar1. In some measurements with some constructs the increase in FRET efficiencies was already partially observed with just Gar1. It is possible that the additional Gar1 is responsible for the effects instead of the combination. One way how the interplay between both proteins could occur is that one component aids the (correct) binding of the other and this second protein is responsible for the three-dimensional change in shape observed in the FRET histograms with WG in the buffer. Nhp2 is a bad binder in absence of other proteins alone as discussed above and evidenced by the low (600 nM) affinity for snR34. With all protein components present the $K_D$ dropped to the subnanomolar range[69]. If Gar1, which can bind to the RNA according to the FRET experiments, is sufficient to increase Nhp2 affinity, Nhp2 might tie the hairpins together. Nhp2 runs as a dimer on size exclusion chromatography[78]. Maybe this Nhp2-Nhp2 interaction remained when each monomer binds to one of the hairpins in the bipartite structure. In the human telomerase structure however, the two Nhp2 molecules bind far apart from each other[82]. Conversely, Nhp2 might help correct and efficient binding of Gar1. It was shown previously that

Nhp2 has three different binding modes for the H3 hairpin[68]. Nhp2 might help the folding of the upper stem, especially in H3 which might then select a conformation that can bind Gar1 better. Also, direct protein-protein interactions between Gar1 and Nhp2 are possible that might increase the affinity of Gar1. Some indicators that Nhp2 might amplify the effects that Gar1 alone is already to some extent capable of, are found in the smFRET data. Some FRET changes were visible with just Gar1 or the NCG complex even though just as a small shoulder in some cases and manifest to a peak or a quantitative shift under WG or WNCG conditions. If the conformation with the closer hairpins is the more active one, previous activity data might also be interpreted in the direction that Nhp2 aids efficient Gar1 binding. The H/ACA complex is capable of pseudouridylation in absence of Nhp2 but with an extremely low activity. Gar1 on the other hand is essential for activity. Possibly the activity under NCG conditions derives from the fraction of molecules where Gar1 is correctly bound and Nhp2 would add to the activity by inducing correct Gar1 binding to the other molecules. In this scenario Nhp2 would have the role of an anchor for the other protein components to the RNA. Another way of how the combination Nhp2 and Gar1 but not the individual proteins would tether the hairpins together is via direct protein-protein interactions between Gar1 of one hairpin and Nhp2 of the other. In the human telomerase structure Nhp2 of the 5’ structure and Gar1 of the 3’ hairpin are localized in proximity even though too far apart for protein-protein interactions. For the modelling of the telomerase the archaeal H/ACA hairpin was used. Archaeal H/ACA complexes contain the protein L7Ae instead of Nhp2 and Gar1 only consists of the core domain without the eukaryotic specific RGG domains that would increase the size of Gar1. Furthermore, the telomerase does not need the pseudouridylation function and in the telomerase RNA the region where the H5 loop would be, folds back into another lobe of the structure. With all these differences between the cryo-EM structure of the human telomerase and the H/ACA pseudouridylases it could be possible that the 5’ Nhp2 and the 3’ Gar1 are not only in proximity to each other but might actually directly interact resulting in a tethering of the two hairpins. If this is the case the Nhp2 from H5 might position the Gar1 in H3 correctly or might assist the Gar1 dependent opening of the Cbf5 thumb loop. This might explain why turnover of the 3’ substrate is more affected by the presence of the second hairpin. It cannot be excluded that two effects might work together in the Gar1-Nhp2 dependent interplay between the hairpins, such as the anchoring of Nhp2 of the same hairpin and the support of the Nhp2 from the other.

In summary, an interplay between Nhp2 and Gar1 tethers the two hairpins of the bipartite complex together. What kind of interactions induce the crosstalk remains unclear. Crosslinking experiments might give more insights into this question. The role of Nhp2 could be the correct folding of the upper stem of the hairpins as well as the anchoring of the other protein components.

## 5.8 Conformation of the fully assembled complex

The three-dimensional architecture of the fully assembled complex was also investigated using smFRET. For none of the constructs with a labeling scheme across the different hairpin a single peak without shoulders or additional peaks was found (Figure 50). This means that the RNP cannot be described by one distinct conformation. Interestingly, for the constructs $FL_{2,3}$, $FL_{1,3}$ and $FL_{2,4}$ under WNCG conditions the conformations might have similarities to the lower FRET state in presences of only NC and an additional increased FRET state like the increase of FRET efficiency to the RNA induced by WG even though the exact values did not match. The ratios between the two states for $FL_{2,3}$ and $FL_{1,3}$ were similar which might be an indication that lower FRET and the higher FRET peaks in each histogram could represent a similar conformation, respectively. For $FL_{1,4}$ that is labeled at the lower stems of the hairpin reproducible no additional increased FRET state was detected.

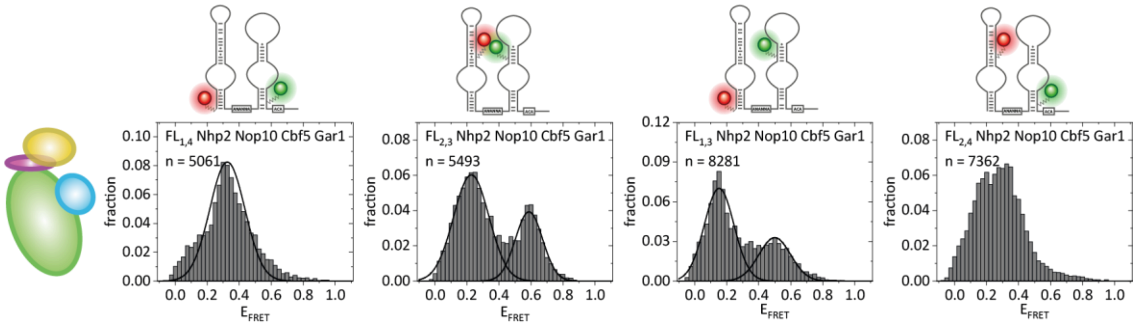


Figure 50: Histogram analysis of the full assembled H/ACA RNAP

From the experiments with $FL_{1,2}$ it can be concluded that Cbf5 binds quantitative at least to H5 under the experimental conditions. The experimental setup cannot differentiate between full and partially assembled complexes. The heterogeneity in the $FL_{3,4}$ RNA with all protein components hints that maybe some of the FRET states resulted from complexes lacking one or more components. However, even in the complete H/ACA complex different conformations might be adopted. The previously found three binding modes of Nhp2 might not all induce the same RNA conformations. Overall, the presence of all four proteins resulted in more defined and sharper peaks in the histogram. This means that the proteins stabilize different less dynamic structures.

Furthermore, it is unclear whether the individual FRET states represent active complexes. The higher of the two FRET peaks for $FL_{2,3}$ and $FL_{1,3}$ seemed to be stabilized by Nhp2. Assuming that this state represents a conformation where a Gar1 and Nhp2 dependent tethering of the hairpin occurs and knowing that the H/ACA RNP is more active with the second hairpin and with Nhp2 this might hint to an active state. However, it cannot be concluded that the other state would be less active or completely inactive in this case. It would be interesting to analyze whether the ratio between the different states changes in presence of a substrate RNA. Additionally, the substrate RNA could be immobilized on the slide and using unbiotinylated FRET constructs, the binding active FRET states could be selected.

With the information gained from the assembly experiments and the smFRET data for the full complex I would like to propose a model for the orientation of the hairpins. In this model Cbf5 adjusts the pseudouridylation pocket, distorts the linker and bends the orientation of the lower stems of the hairpins. When the linker conformation is locked by Cbf5, the effects of Gar1 and Nhp2 dependent tethering of the upper parts of the hairpins might not be transmitted to the lower parts anymore thus $FL_{1,4}$ has no WG-dependent high FRET state in the presence of all proteins. The second state under WNCG conditions was increased only when the third labeling site in the apical H3 stem was included in the labeling scheme. For $FL_{2,4}$ no increase was found and the H5 labeled $FL_{1,2}$ was not affected by Nhp2 and Gar1. It is possible that when Cbf5 dominantly determines the conformation of the lower parts, the upper part of H3 is bent and pulled towards H5. If only part of the hairpin is tethered it would also explain the different FRET values for the increased FRET states between the WG and WNCG condition with the $FL_{2,3}$ and $FL_{1,3}$ constructs. Further this would be in agreement with activity data suggesting that the second hairpin influences H3 more than H5. However, this is just one possibility of how the three-dimensional arrangement of the bipartite complex might be organized.

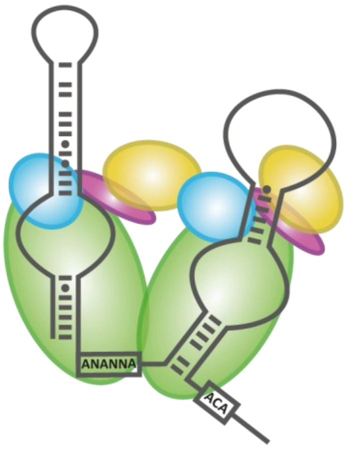


Figure 51: Schematic representation of the proposed model for the assembled bipartite H/ACA RNAP. Green: Cbf5; blue: Gar1; yellow: Nhp2; magenta: Nop10.

## 5.9 Role of RGG domains for the 3-dimensional structure

From the experiments with the full-length protein a model for tethering the top parts of the hairpins together was established that is dependent on the interplay between Nhp2 and Gar1.

How exactly those two proteins induce the conformational changes and which parts of the proteins are involved remains unclear. Gar1 itself is supposed to bind to the Cbf5 thumb loop to aid substrate release. Gar1 has three RGG domains that are known to bind RNA. Those RGG domains are a specific feature of eukaryotic Gar1 proteins and the bipartite structure is also conserved in eukaryotes, but not in archaea. One hypothesis is that the tethering of the two H/ACA hairpins is mediated through the RGG domains. This was investigated by smFRET using Gar1 deletion variants that had different combinations of the RGG domains removed.

In the experiments with full-length Gar1 the tethering of the two hairpins was best observed for the construct $FL_{2,3}$ that is labeled at the upper parts of the hairpin. A combination of Gar1 and Nhp2 lead to an increased FRET efficiency. A quantitative increase was also observed when just the third RGG domain was removed in Gar1 Δ3. This protein variant had one RGG domain on each side of the core domain. Those two protein constructs were the only ones that could induce a second peak for $FL_{1,3}$ with all other protein components present. Further removal of the first RGG domain (Gar1 Δ1,3) completely abolished the capability of tethering the hairpins together. When the N- or only the two C-terminal RGG domains were removed (Gar1 Δ1 or Gar1 Δ2,3) only a fraction of the molecules was able to fold into a conformation with the hairpins tethered. Interestingly those two proteins introduced a variety of FRET states and dynamics to the $FL_{1,3}$ construct in absence of other proteins. It is possible that the protein variants could bind to RNA through one of their domains, but this might be either at a wrong site, or counterbalance from the other RGG domain might be necessary. The dynamics with Gar1 Δ1 or Gar1 Δ2,3 were suppressed when Nhp2 was also present.

The smFRET experiments could suggest that for the efficient tethering effects at least one RGG domain has to be present at each side of the Gar1 core domain like in the full-length Gar1 and Gar1 Δ3. Those domains might act in a cooperative manner. The protein variants that only had either the complete N- or the C-terminal RGG domains (Gar1 Δ1 & Gar1 Δ2,3) were less but might introduce dynamics or alternative binding sites. For the other proteins variants, the least or no tethering effects were observed in the smFRET analysis.

In summary, the RGG domains of Gar1 impact the three-dimensional structure of the bipartite H/ACA RNA. Depending on the snoRNA the presence of the RGG domains could improve or reduce the pseudouridylation activity. This means that a reshaping of the conformation mediated by the Gar1 RGG domains might not be advantageous for all snoRNAs.

# 6 Summary and outlook

In this work, the H/ACA RNP form *S. cerevisiae* was investigated using smFRET. First, the protein components, and the labeled and unlabeled guide RNA were successfully produced. It could be shown that the RNP reconstitutes *in vitro* and is active. The influence of the second hairpin of the bipartite complex was analyzed. For the H5 hairpin no observable effects were detected, while the H3 hairpin seemed more dynamic in the conformations. Overall, Cbf5 seemed to be the main factor to align the pseudouridylation pocket. Furthermore, the spatial orientation of the hairpin towards each other was followed during assembly. Here, the snoRNA could be in an extended rather than in a parallel conformation with the hairpins pointing in different directions. Cbf5 might distort and arrange the linker between the individual hairpins and bend the two hairpins. Interestingly, a combination of Nhp2 and Gar1 seems to tether the apical parts of the hairpins together. If the lower parts are in a more fixed orientation because of Cbf5, this Nhp2-Gar1 tethering might mostly affect the H3 hairpin. The ability for tying the hairpins together could possibly be mediated through the RGG domains of Gar1. First results hint towards cooperative effects of the N- and C-terminal RGG domains and that lacking only one of them might pull the snoRNA in a different conformation. Overall, the FRET experiments with the all H/ACA RNP components showed that the complex adopts different conformations either due to incomplete reconstitution or structural heterogeneity.

The smFRET measurements provided helpful insights into the three-dimensional structure of the bipartite H/ACA RNP and the roles of the individual protein components during assembly. However, for a more sophisticated structure other techniques need to be applied. Due to the large size of the complex cryo-EM would be the most obvious choice. Additionally, it might be interesting how the substrate could change the conformations. This could be studied using this smFRET approach with the substrate RNAs in the measurement buffer. The ratio between the conformations in the fully assembled RNAP might change or the substrate could induce new conformations. Immobilization of the substrate RNA on the slide and an unbiotinylated FRET construct might give insights which of the conformations found in this work might be binding competent. Also, the roles of the RGG domains need to be studied more intensely. In this context, the assembly factor Naf1 could be analyzed. It has a similar fold to the Gar1 core domain but lacks the RGG domain extensions and is replaced by Gar1 in the Cajal bodies.

# 7 Appendix Chapter II

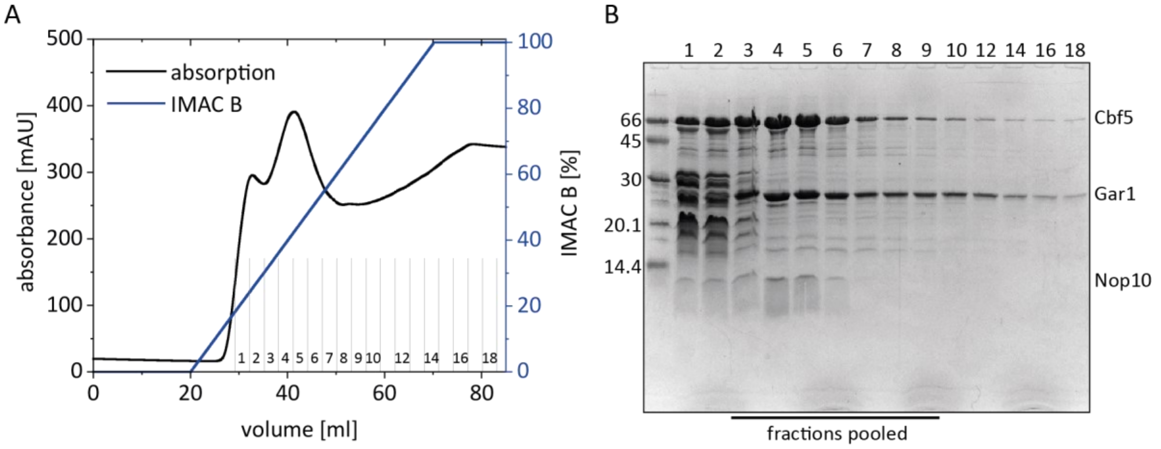


Figure 52: First His-Trap of the purification of NCG Ctag. A) Chromatogram of the His-Trap. The collected fractions are indicated. B) SDS-PAGE analysis of the collected fractions. The gel was stained using Coomassie Blue.

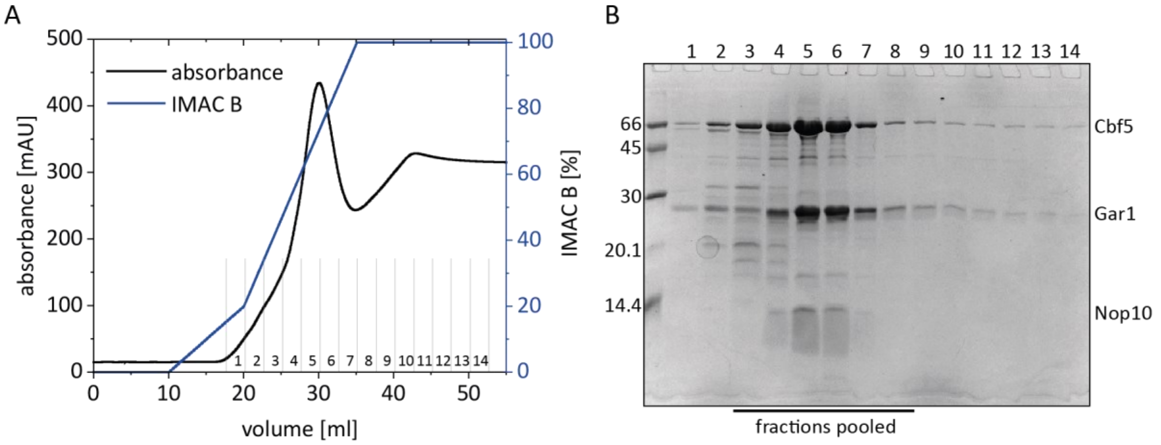


Figure 53: Second His-Trap of the purification of NCG Ctag. A) Chromatogram of the His-Trap. The collected fractions are indicated. B) SDS-PAGE analysis of the collected fractions. The gel was stained using Coomassie Blue.

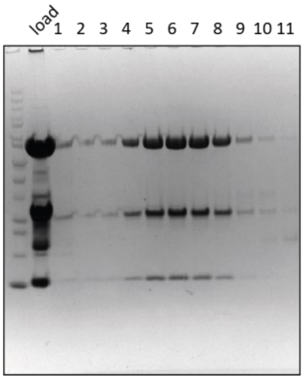


Figure 54: Analytical SDS-PAGE of the size-exclusion chromatography of the purification of NCG Ctag. The gel was stained using Coomassie Blue.

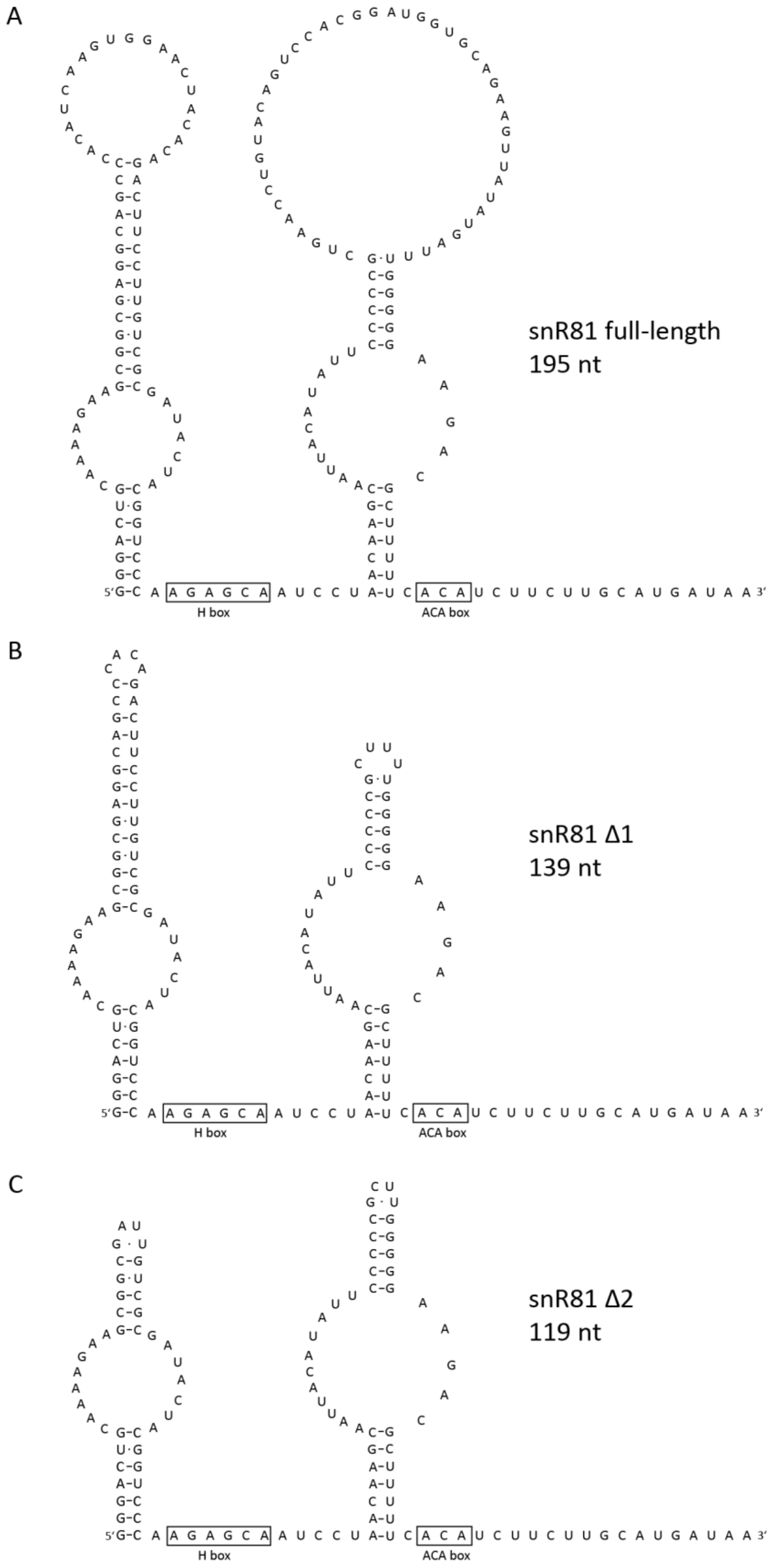
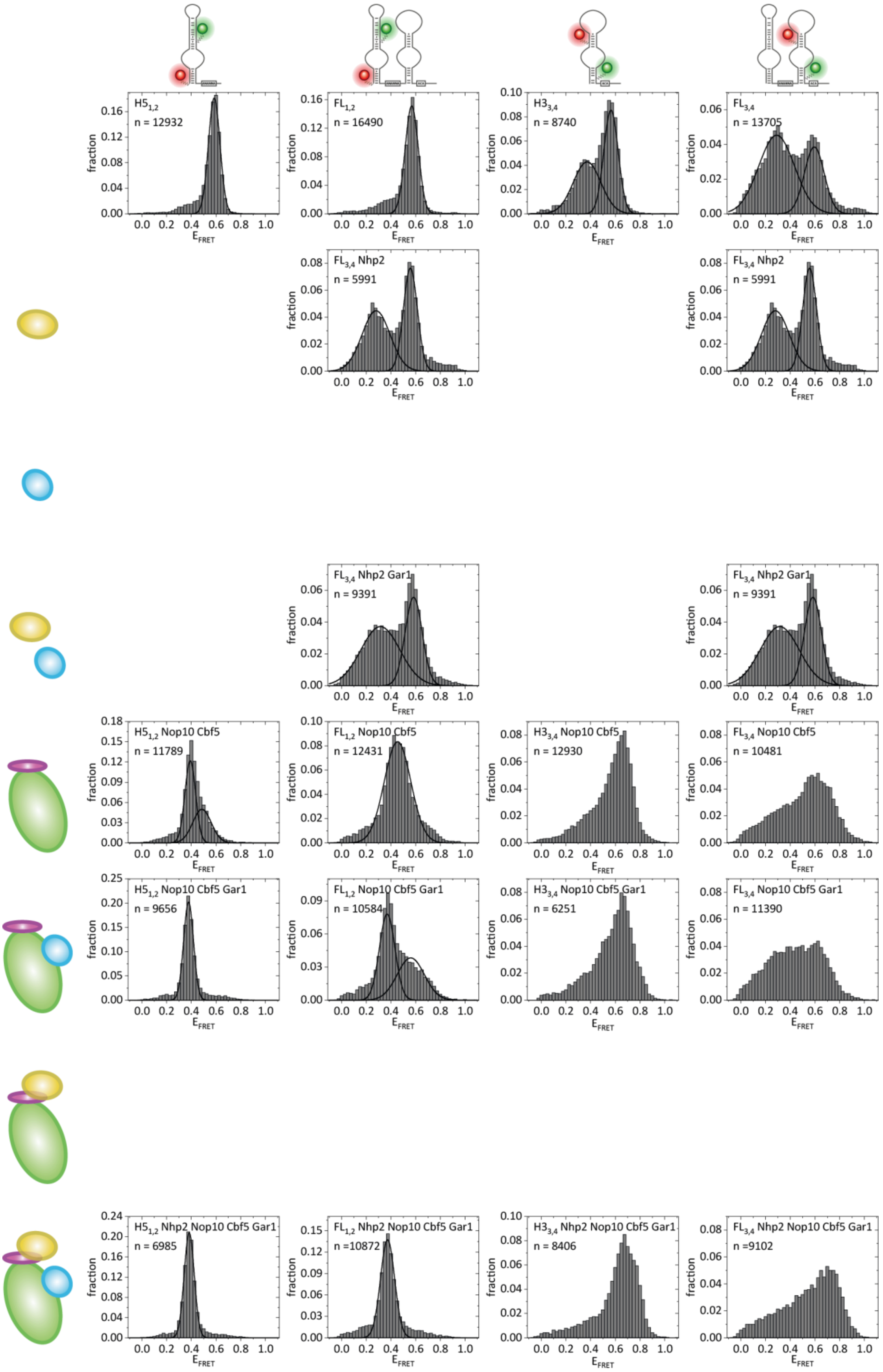


Figure 55: Sequences of the guide RNAs used for reconstitution.


$H5_{1,2}$
n = 12932
$FL_{1,2}$
n = 16490
$H3_{3,4}$
n = 8740
$FL_{3,4}$
n = 13705
$FL_{3,4}$ Nhp2
n = 5991
$FL_{3,4}$ Nhp2
n = 5991
$FL_{3,4}$ Nhp2 Gar1
n = 9391
$FL_{3,4}$ Nhp2 Gar1
n = 9391
$H5_{1,2}$ Nop10 Cbf5
n = 11789
$FL_{1,2}$ Nop10 Cbf5
n = 12431
$H3_{3,4}$ Nop10 Cbf5
n = 12930
$FL_{3,4}$ Nop10 Cbf5
n = 10481
$H5_{1,2}$ Nop10 Cbf5 Gar1
n = 9656
$FL_{1,2}$ Nop10 Cbf5 Gar1
n = 10584
$H3_{3,4}$ Nop10 Cbf5 Gar1
n = 6251
$FL_{3,4}$ Nop10 Cbf5 Gar1
n = 11390
$H5_{1,2}$ Nhp2 Nop10 Cbf5 Gar1
n = 6985
$FL_{1,2}$ Nhp2 Nop10 Cbf5 Gar1
n =10872
$H3_{3,4}$ Nhp2 Nop10 Cbf5 Gar1
n = 8406
$FL_{3,4}$ Nhp2 Nop10 Cbf5 Gar1
n =9102
fraction
$E_{FRET}$
0.0 0.2 0.4 0.6 0.8 1.0

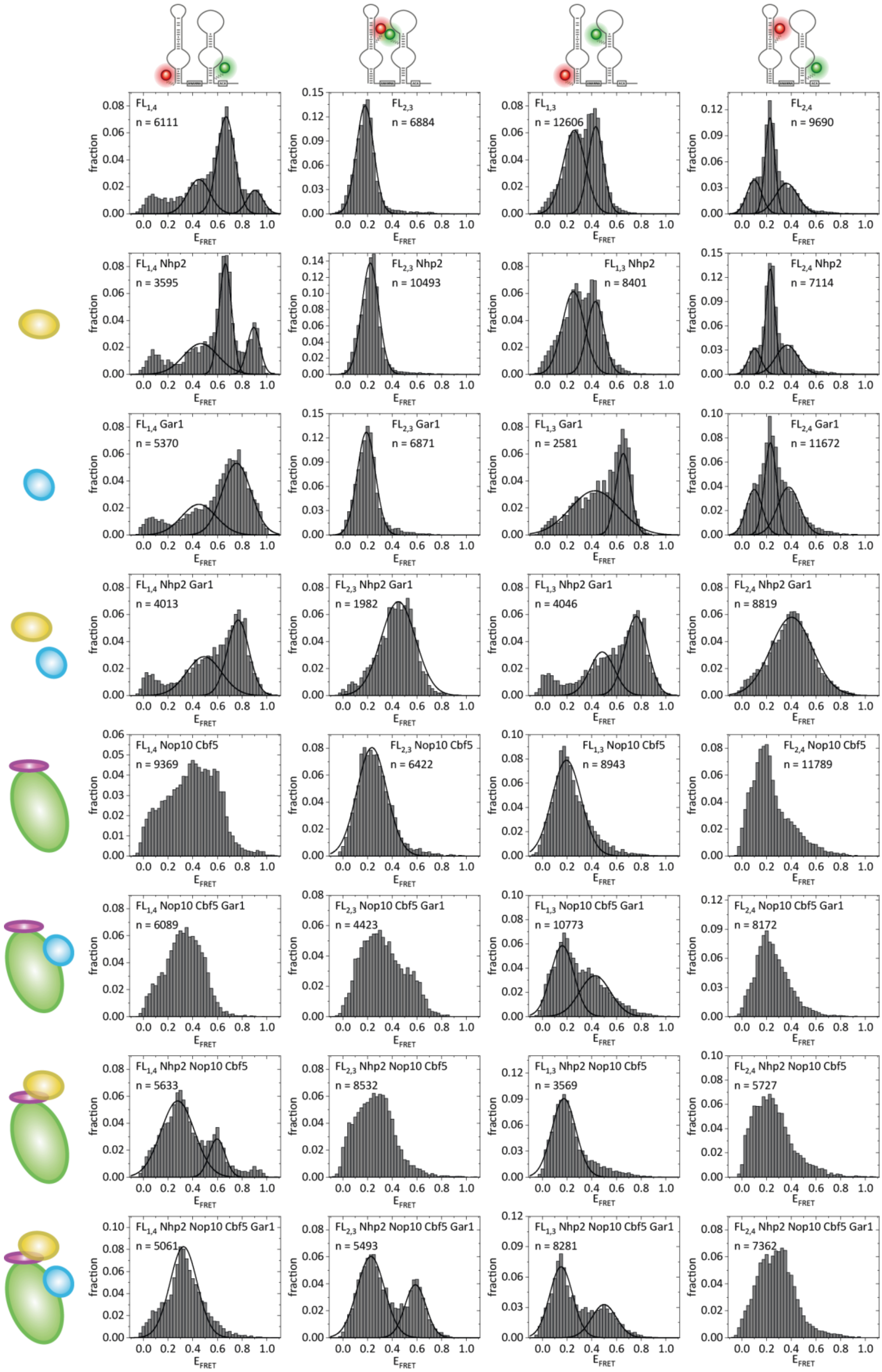


Figure 56: Overview of H/ACA FRET data with different RNA constructs and different combinations of proteins. Yellow: Nhp2, blue: Gar1, green: Cbf5, purple: Nop10.

# Chapter III

# Guanidine-II riboswitch from *E. coli*

# 8 Introduction

## 8.1 Riboswitches

Protein synthesis is an energy and resource intensive process. Different strategies have evolved to adapt protein production to the cellular needs. In bacteria gene expression can be controlled by riboswitches. Riboswitches are *cis*-regulatory RNA elements in the 5' untranslated region (5' UTR) of bacterial mRNA. They can recognize their ligands specifically and control the expression of the downstream genes. Examples of riboswitches range from ions[107,108], cofactors[109], amino acids[110], nucleotides[111] and tRNA[112]. Today, riboswitch candidates are usually discovered using bioinformatics approaches[113]. In general, they are composed of an aptamer domain that is overlapping with an expression platform. In the aptamer region ligand binding takes places and leads to a conformational change in the RNA. This structural rearrangement affects the expression platform in a way that either allows (ON-switch) or prohibits (OFF-switch) gene expression. Riboswitches can act on transcriptional (Figure 57A), translation (Figure 57B), RNA decay levels (Figure 57C) - or in the case of eukaryotic riboswitches - via alternative splicing. While most riboswitches are found in bacteria, TPP binding riboswitches are also found in eukaryotes like fungi[114], plants[115] and algae[116]. However, since most riboswitch classes are exclusive to bacteria riboswitches might be a promising target for the development of novel antibiotic therapies[117].

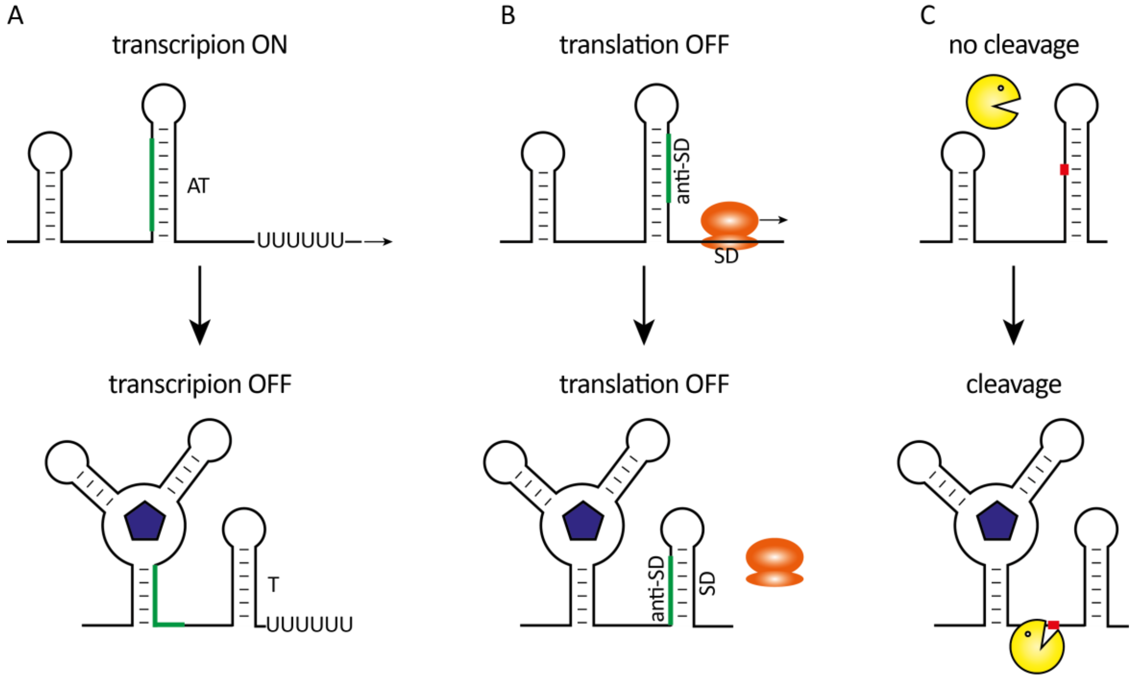


Figure 57: Schematic represantation of the different riboswitch types. A) transcriptional OFF switch. AT: antiterminator; T: terminator. B) translational OFF riboswitch. Orange: ribosome. C) RNA degradation. Yellow: RNase.

### 8.1.1 Transcription termination

Transcriptional riboswitches can cause a ligand-dependent premature transcription termination. In addition to the aptamer domain, transcriptional riboswitches form either an anti-terminator hairpin or a terminator hairpin. These two hairpin structures compete with each[118,119]. After leaving the RNA polymerase exit tunnel, the mRNA starts to fold co-transcriptionally. The sequence necessary for antiterminator formation is synthesized first. The switching has to occur faster than the elongation speed of the RNA polymerase. The RNA has to fold, bind the ligand and either stabilize the antiterminator (ON-switch)[119] or destabilize it (OFF switch)[118] before the full terminator sequence is transcribed. Pause sites within the riboswitch slow down the polymerase and increase the time for the structural arrangements[120,121]. The terminator hairpin is followed by a U-rich sequence which leads to Rho-independent dissociation of the RNA polymerase and transcription termination, when the terminator conformation is formed.

### 8.1.2 Translation initiation

Most bacterial mRNAs have a purine rich sequence element of approximately six nucleotides upstream of the start codon AUG. This ribosomal binding site (RBS) is also called Shine-Dalgarno (SD) sequence and can form Watson-Crick base pairing with the 3' end of the 16S ribosomal RNA[122]. Ribosome binding to the mRNA can also be assisted by the ribosomal protein S1[123]. Translation initiation requires ribosome binding to a free SD sequence that is not involved in protein binding or base pairing. Translational riboswitches use this feature for their regulation. Some riboswitches have the SD sequence located within the aptamer. For example, in the AdoCbl riboswitch, the SD sequence is involved in kissing loop formation upon ligand binding and therefore is inaccessible for the ribosome[124]. In many translational riboswitches the SD sequence is located downstream of the aptamer. In these cases the conformational change within the aptamer is transferred to the expression platform through the release or sequestering of a sequence complementary to the SD sequence, the anti-SD sequence[125]. There are also riboswitches with both, a transcriptional and a translational regulation. Here, the SD sequence is part of the terminator hairpin and serves as a second checkpoint to switch off gene expression[126]. In contrast to transcriptional riboswitches, translational switches are not that dependent on the kinetic folding competing with the transcription rate. They are usually under thermodynamic control and can adapt to environmental changes during the mRNA lifetime[127,128]. An exception is the *thiM* riboswitch from *E. coli* where the switching is determined during transcription[129].

### 8.1.3 RNA degradation

Even though most riboswitches act on either transcriptional or translational level there are examples where the riboswitches control gene expression via RNA stability. In at least three riboswitch containing mRNAs half-life is altered by exposing or sequestering RNase cleavage sites upon ligand-induced conformational change as a second regulation mechanism, additional to the transcriptional or translational control. In the SAM-I (S-Box) *yitJ* riboswitch from *B. subtilis*, which is a transcriptional OFF-switch, ligand binding leaves an RNase Y cleavage site located in the antiterminator hairpin unpaired, resulting in subsequent degradation[130]. In the translational-OFF switch *lysC* from *E. coli,* lysine binding reveals cleavage sites for RNase E[131]. In contrast, the translational ON-switch Gdn-III *sugE* riboswitch from *L. pneumophilia* protects the RNA by blocking RNase E scanning when an H-type pseudoknot is formed in presences of guanidine[132]. The *glmS* riboswitch works on a completely different mechanism. It is a self-cleaving ribozyme where the ligand glucosamine-6-phosphate (GlcN6p) acts as a cofactor[133]. After self-cleavage the RNA is further degraded by 3' exonucleases. The last example, the Vc2 riboswitch of the *tfoY* gene works as an OFF-switch[134]. Binding of cyclic di-GMP increases the stability of a non-coding sRNA which can then downregulate the gene expression.

### 8.1.4 smFRET analysis of riboswitches

Riboswitches have to adopt dynamically to environmental changes with a conformational change. Furthermore, they a rather small and no additional protein factors are required. This makes them a target for smFRET investigation. A spectrum of both translational[129,135,136] and transcriptional[137–141] riboswitches have been employed to smFRET studies. Strategic labeling is important. Fluorophore attachment sites are chosen where a distance change is expected upon ligand binding. In most smFRET studies the aptamer region was labeled. For example attaching the fluorophore to two loop regions results in an increased FRET efficiency when a kissing loop conformation is formed upon ligand binding[138]. In another example two stems of a 4-way junction were labeled that experience a distance change during ligand-dependent reorientation[141]. For translational riboswitches, smFRET was used to understand how the expression platform is affected by ligand binding. For this, labeling sites were selected in the aptamer and near the SD-sequence[136]. In the case of transcriptional riboswitches only FRET-constructs without the full terminator hairpin sequence are used. While riboswitches are classically described as a two state system with an ON- and OFF state (with one of them usually the ligand bound state), smFRET experiments have in one case uncovered up to 5 discernable states that evade ensemble techniques[142]. Using a combination of different labeling schemes for the same riboswitch allows detailed insights into the three-dimensional architecture[135,143] and can discriminate between states that would overlap due to similar inter-dye distances[142]. It is also possible to calculate apparent affinities based on FRET data[135]. With smFRET experiments not only the effects of the ligand to the folding have been analyzed but also the dependence of other ions such as $Mg^{2+}$ [138,143], temperature[141] and pressure[140]. Information gained from smFRET give new insights into the folding mechanism of riboswitches[143], such as (short lived) folding intermediates[138] or ion induced prefolding states[141]. Analysis of smFRET data can also dissect whether the riboswitch acts via induced fit or conformational selection[139]. Furthermore, cotranscriptional folding pathways can be investigated[129]. Apart from distance and mechanistical information kinetic data can also be extracted from long-time smFRET measurements[135,137,138]. In some cases ligand binding changes the folding dynamics between different conformations and intermediates. This is especially important for transcriptional riboswitches that are under kinetic control as folding time competes with transcription rate. Another fluorescence microscopy base method to study riboswitches is Single Molecule Analysis of RNA Transient Structure (SiM-KARTS)[144]. In translational riboswitches the accessibility of the SD sequence is observed with a dye-labeled probe competing with the anti-SD sequence. In summary, smFRET is an excellent tool for the study of riboswitches and provides a variety of information.

## 8.2 Guanidinium Riboswitches

### 8.2.1 Guandinium

Guandine (Figure 58A) is a small molecule that was first synthesized in a Strecker synthesis in 1861[145]. With a $pK_a$ of 13.6 it is present as the cationic guanidinium ($Gdm^+$) (Figure 58B) under physiological conditions[146]. The guanidine structure element is part of several biomolecules like the amino acid arginine and the nucleobase guanine. In the RNA world hypothesis it is believed to be part of a prebiotic pathway for purine synthesis[147]. But it can also be produced in living organisms. Bacteria can produce $Gdm^+$ endogenously when grown in minimal medium[108]. In a proposed guanidine cycle[148] that recycles nitrogen in reactions similar to the urea cycle, $Gdm^+$ is produced in the enzymatic reaction of canavanine to homoserine[149]. $Gdm^+$ is also the byproduct of the ethylene-forming enzyme (EFE) in bacteria that catalyzes the δ-hydroxylation of arginine[150].

Independent of whether $Gdm^+$ is from endogenous production or taken up from the environment it is toxic for the cell. It is best known for its function as protein denaturant in elevated levels above 1 M[151]. However, depending on the organism $Gdm^+$ acts as a more specific protein inhibitor at lower concentrations. In bacteria, the arginase enzymes can be inhibited by as little as 10 mM

$Gdm^+$ [152]. Additionally the chaperone Hsp100 is inhibited also in the low mM range[153]. For the eukaryotic chaperone Hsp104 from *S. cerevisiae* this already starts at concentrations as low as 30 µM[154]. In higher eukaryotes 3 mM $Gdm^+$ sufficiently inhibits voltage-dependent potassium channels in the neuromuscular junction[155].

These negative effect of $Gdm^+$ show the need for detoxification strategies for the cells. Bacteria have developed different mechanisms that might aim at $Gdm^+$ detoxification by direct export or through chemical modification. To relate and fine-tune the energy-consuming production of detoxifying proteins to the requirements determined by intracellular levels, four structural and mechanistically different riboswitch classes have evolved that specifically recognize $Gdm^+$: guanidine-I – IV[108,156–159].

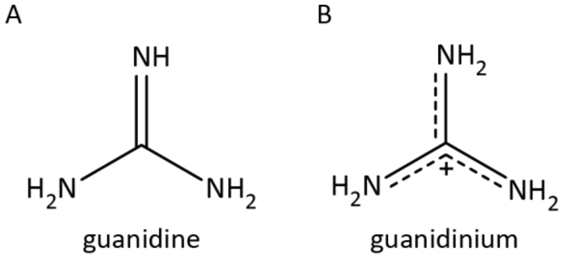


Figure 58: Chemical structure of A) guanidine and B) guanidinium.

### 8.2.2 Guanidine-I

The first class of guanidine riboswitches was identified as a potential riboswitch candidate in 2004 and was named *ykkC* element[160] before identification of $Gdm^+$ as their ligand in 2017[108]. The approximately 2,000 members are found across Actinobacteria, Firmicutes, Proteobacteria and Cyanobacteria[108]. The guanidine-I riboswitch class can be further divided into two subtypes. Subtype 1 controls the expression of a variety of proteins: proteins associated with nitrogen metabolism, such as urea carboxylases, urea carboxylase-associated protein family, allophanate hydrolases and proteins of the arginase superfamily. Genes for nitrate, sulfate and bicarbonate transporters are also found downstream of guanidine-I riboswitches. Additional regulatory targets are multidrug resistance transporters like the small multidrug resistance (SMR) transporter family including SugE and EmrE and the YkkC/YkkD heterodimeric transporter which resulted in the original name of this riboswitch class. The guanidine-I riboswitch is built by four coaxial helices P1, P1a, P1b and P2 (Figure 59)[161,162]. In the ligand-bound crystal structure a fifth helix (P3) interacts with the P1a helix through tertiary interactions. Ligand binding takes places in a small pocket in this interface. A metal ion blocks the funnel from solvent. Ligand binding is achieved with all six hydrogens of $Gdm^+$ being involved in hydrogen bonding. Two hydrogen bonds each are formed with two guanine bases in the P1a and P3 helices and two further with oxygens of the phosphate backbone. Binding of the cationic ligand is further stabilized as it is sandwiched between two guanines via cation-π interactions and through electrostatic interaction with the RNA backbone. All these interactions result an affinity between 39 and 60 µM[108,162] and in a high specificity for Gdm$^+$. Derivates are not able to bind to the riboswitch[108]. The conserved P3 hairpin region is overlapping with the terminator hairpin[160]. Ligand-dependent stabilization of the P3 hairpin through tertiary interaction with P1a might compete with terminator formation. Thus, the guanidine-I riboswitch is a transcriptional ON-switch.

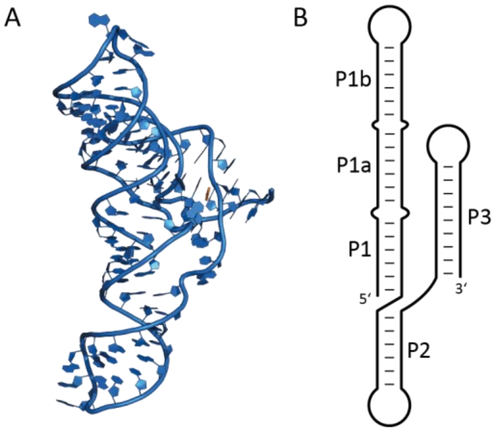


Figure 59: A) Crystal structure of the guanidine-I riboswitch from *Dickeya dadantii* (PDB: 5U3G)[162]. B) Schematic represantation of the guanidine-I secondary structure.

In addition to subtype 1 there is a second subtype which comprises approximately a quarter of the guanidine-I riboswitch class[108]. The subtypes differ in 3 otherwise conserved nucleotide regions and by the genes they control. All these nucleotides are either directly involved in ligand binding or within one nucleotide according to the crystal structure of subtype 1[161]. Furthermore, guanidine is unable to trigger a structural response in subtype 2 as revealed by in-line probing[108]. The subtype 2 is further divided into 4 subclasses. Subclass 2a senses guanosine tetraphosphate (ppGpp)[163] which is a signal molecule in bacterial starvation pathways. The genes regulated by this riboswitch are involved in branched amino acid synthesis, glutamate synthases and ABC transporters. Phosoribosyl pyrophosphate (PRPP) was identified as the ligand for subclass 2b[164]. PRPP is an intermediate of purine biosynthesis. The PRPP riboswitch controls genes that are either involved in purine biosynthesis or inhibited by purines. Sometimes this ON-riboswitch is in tandem with an upstream guanine OFF-switch[164]. (d)ADP and (d)CDP are sensed by the subclass 2c which mostly regulates expression of nucleoside diphosphate linked to X (NUDIX) hydrolases which can cleave the phosphoanhydride linkage of NDPs[165]. The ligand for the 2d subclass remains unknown.

### 8.2.3 Guanidine-II

The second guanidine-binding riboswitch class is the shortest of the four RNA elements. It was found in a bioinformatics study in 2007 and was named mini-*ykkC* based on the gene similarities with the guanidine-I riboswitch class[166]. The approximately 800 representatives are mainly found in Proteobacteria with examples also in Actionbacteria, Chrysiogenetes, Cyanobacteria, Planctomycetes and Verrucomicrobia[156]. Genes under control of the guanidine-II riboswitch are mostly SMR transporters, urea carboxylases, urea carboxylase associated proteins 1 and 2 as well as the $Ca^{2+}/H^+$ antiporter ChaA[156]. Sometimes it can be found in the same operon that is under control of guanidine-I[156]. The aptamer domain of the guanidine-II riboswitch is composed of two covariant GC-rich hairpins P1 and P2 which are each capped by a ACGR sequenced tetraloop where R can be either adenosine or guanosine (Figure 60)[156,166]. Both hairpins are connected with a variable linker with a length between 7 and 40 nucleotides[166]. The ribosomal binding site is usually around 33 nucleotides downstream of P2[166].

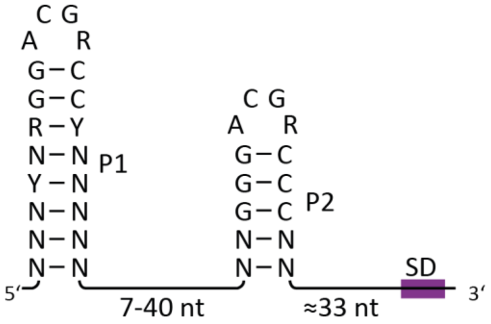


Figure 60: Consensus sequence of the Gdn-II riboswitch[166]. N: variable nucleotide, Y: C or U, R: A or G.

The guanidine-II aptamer domain can bind two molecules of $Gdm^+$, one at each of the loops[167]. Binding takes place with apparent $K_D$s around 33-300 µM[156,167,168] in a cooperative manner with a Hill coefficient of 1.4[156]. In contrast to guanidine-I, small derivates of $Gdm^+$ like methylguanidine and aminoguanidine bind with only slightly lower affinity, and agmatine and ethylguanidine are also binding competent[156]. Urea however is not recognized by guanidine-II. Ligand binding leads to formation of a kissing loop interaction. In-line probing experiments showed that there are also structural changes within the linker[156]. The loop-loop interaction can occur as homodimers[169,170] or heterodimers[171] of isolated hairpins or presumably in the full-length aptamer[167].

Crystal structures of the isolated hairpin reveal that both hairpins are coaxially aligned and rotated by 180° to form the interaction (Figure 61A)[170]. A $Mg^{2+}$ binding site is found in the penultimate base pair before the loop[168,169]. The hairpins interact in a kissing loop interaction with Watson-Crick base pairing between the loop CGs from the other stem-loop (Figure 61B). The first adenine nucleobase of the ACGR motif is stacked onto those base paired nucleotides and forms hydrogen bonds with the sugar backbone of the opposite hairpin[169,170]. The fourth loop nucleotide, which is an adenosine in all solved crystal structures so far, is flipped out of the loop and stacks onto its equivalent base from the other hairpin (Figure 61C). Both loops form a binding cavity for the Gdm$^+$ ligand with the solvent accessible openings facing each other and thus enable steric discrimination against bulky ligands. The ligand is bound with five out of six possible hydrogen bonds (Figure 61D). Two H-bonds are formed with the Hoogsteen edge of the guanosine nucleobase and three more with the phosphate backbone of the first three loop nucleotides[169,170]. In contrast to the first class of Gdm$^+$-binding riboswitches the sixth hydrogen is solvent accessible, which might explain the binding capabilities of close Gdm$^+$ derivatives. In some crystal structures this hydrogen is bridged through two water molecules with its partner from the other strand[170]. Binding of the cationic ligand is further stabilized with electrostatic interactions with the negatively charged phosphate backbone and with a cation-π interaction with the last guanine base before the loop (Figure 61B)[170]. All these interactions explain the binding specificity for Gdm$^+$ and how the RNA can discriminate against urea.

The hairpin is able to form the kissing loop interaction in absence of ligand e. g. in high ammonium buffer[169]. Molecular dynamic simulations revealed that the riboswitch switches between a bound-like state and an unbound state where the guanine nucleobase is collapsed into the binding pocket and therefore no longer is capable of forming Watson-Crick base pairing[172].

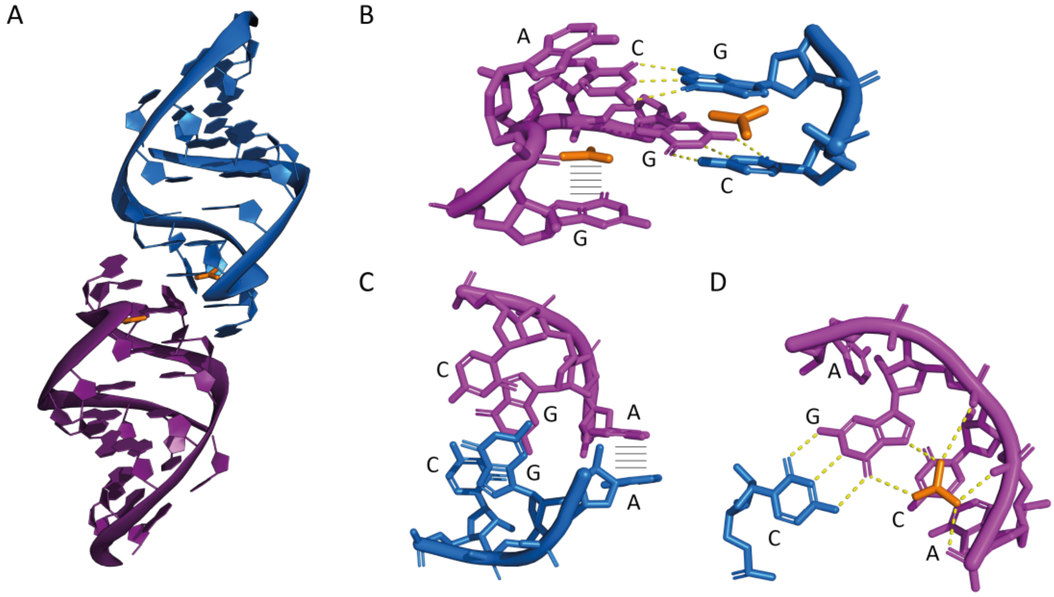


Figure 61: Crystal structure of *E. coli* Gdn-II P1 homodimer (PDB: 5NDI)[169]. The Gdm$^+$ ligand is shown in orange. A) Whole crystal structure. B) Kissing loop basepairs and cation-π interaction of the ligand and the last stem G. C) Stacking of the fourth loop nucleotides. C) Hydrogen bonding of Gdm$^+$.

Reiss and Strobel proposed a regulation mechanism for the guanidine-II riboswitch as a translational ON-switch (Figure 62)[170]. According to the model, in absence of ligand the two hairpins are in an open conformation without interaction. The anti-SD sequence which is located in the linker between the hairpins or also partially before the P1 hairpin is basepaired with the SD sequence downstream of the aptamer, thus preventing ribosome binding and translation. According to the model, upon binding of two $Gdm^+$ molecules to the hairpins a kissing loop interaction is formed. In this conformation the linker including the anti-SD sequence is sequestered. The unpaired SD sequence then allows ribosome binding and translation of the adjacent protein[170].

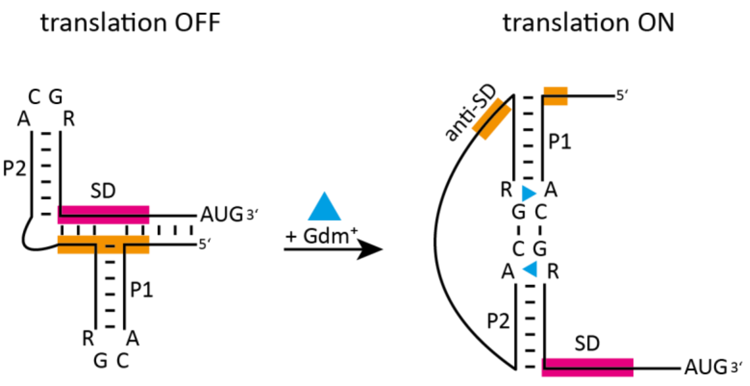


Figure 62: Proposed regulation model of the Gdn-II riboswitch by Reiss and Strobel[170]. In absence of $Gdm^+$ ligand (blue) the SD sequence (magenta) pairs with the anti-SD sequence (orange) in the linker region and nucleotides before P1. After $Gdm^+$ binding to both loops, the hairpins form a kissing loop interaction through their CG base pairs. The anti-SD:SD interaction is disrupted in this conformation and the free SD sequence allows translation initiation.

The loop motif for both hairpins is ACGR. For the P1 hairpin the last purine is in almost all cases an adenine[166]. In the P2 position it is equally distributed between adenine and guanine. The variant with adenines in both hairpin (A/A) was the most affine with a drop of 2 to 3-times in affinity with every loop carrying a guanine[168]. In-line probing experiments revealed that mutation of this position with an uracil base decreases affinity for $Gdm^+$ [156]. Another study reported that mutations of the fourth loop nucleotide of P2 to either cytosine or uridine showed a phenotype corresponding to a consitutively ON state[173]. In *E. coli* the guanidine-II riboswitch has a A in P1 and G in P2 (A/G) and controls the expression of a gene annotated as *sugE*[166] which is a SMR transporter specifically exporting $Gdm^+$ out of the bacterial cell[174].

### 8.2.4 Guanidine-III

The third class of guanidine riboswitches was first found in a bioinformatics study in 2010 and was named *ykkC*-III because of the similarity of genes found downstream of this element compared to the other two riboswitch classes[175]. Later it was renamed after its ligand into guanidine-III. Most members of this class are from Actinobacteria with a few represenatives also in the phylla

Proteobacteria, Deinococcus-Thermus, Planctomycetes and Firmicutes. It mostly controls the gene expression of the SMR transporters EmrE and SugE but also of arginase superfamily proteins and *N*-acetyltransferases[157]. In contrast to the other classes, guanidine/urea- carboxylases cannot be found under the regulation of the guanidine-III riboswitch. Most members of guanidine-III have a P0 hairpin with variable length and sequence which is not necessary for structural changes triggered by ligand binding[157]. The riboswitch folds into an H-type pseudoknot with helices P1 and P2[157]. Directly adjacent to P1 is a conserved ACGA motif comparable to the Gdm$^+$-binding ACGR motif of guanidine-II. Crystal structures revealed that the region between P1 and P2 stems folds into a helical ramp with involvement of Gdm$^+$ and inserts into the major groove of P2[176]. The ligand is bound via five out of six possible hydrogen bonds with the CG of the ACGA motif and a guanine of P2. Further stabilization is provided through cation-π interactions between the cationic ligand and the cytosine of the same motif. The guanidine-III aptamers show apparent $K_D$s between 20 and 200 µM in range with the other guanidine riboswitch classes[157]. Since one hydrogen of the ligand is not involved in specific interactions, close derivates like methylguanidine and aminoguanidine bind only 3-fold poorer than the native ligand. Ethylguanidine and agmatine can also bind but urea is still excluded form recognition[157]. The aptamer is within 1-33 nucleotides upstream of the SD-sequence[175] therefore the guanidine-III riboswitch class was proposed as translational ON-switches. Nevertheless, Richards and Belasco showed a second level of regulation through mRNA stability[132]. Scanning of RNaseE from the 5' monophosphate end is blocked by the coaxial basepairing of the H-type pseudoknot in the ligand bound conformation. The stability of this structure thus protects the mRNA from RNaseE dependent degradation.

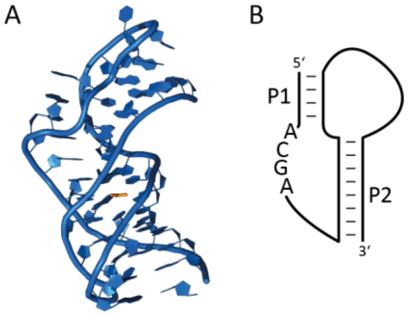


Figure 63: A) Crystal structure of the guanidine-III riboswitch from *Thermobifida fusca* (PDB: 5NWQ)[176]. B) Schematic represenation of the guanidine-III riboswitch.

### 8.2.5 Guanidine-IV

The fourth class of guanidine riboswitches was found most recently. Two separate bioinformatic studies discovered this structural motive by searching for novel riboswitch candidates in the bacterium *Clostridium botulinum* Strain A Hall[158] and specifically within the 5' UTR of *sugE* genes[159].

Most representatives are found in the phylum Firmicutes[159] but are also present in Actinobacteria, Fusobacteria, Spirochaetes, Synergistetes[158,159], Proteobacteria[159] and Tenericutes[158]. The guanidine-IV riboswitches control the expression of SugE family of multidrug efflux pumps, as well as transporter from the PnuC-like transporter family and Multidrug And Toxic compound Extrusion (MATE) transporter family which have not been reportedly involved in guanidine metabolism yet[158,159]. Furthermore, protein expression of the GCN5-related *N*-acetyltransferase family (GNAT) and the Phenylalanin- and lysidine-tRNA synthetase domain superfamily (B3/B4) is controlled by this riboswitch with the latter also being under the regulation of guanidine-I riboswitches[158,159]. *In vitro* transcription assays and *in vivo* fluorescence experiments have shown that this riboswitch class is a transcriptionally controlled ON-switch with guanidinium as a ligand[159]. For the OFF state, both studies proposed a covariant hairpin P1 with a conserved RCCG tetraloop and a terminator hairpin connected by a linker and followed by several Us as a polymerase pause site (Figure 64)[158,159]. The terminator loop contains a conserved GGY sequence. Lenkeit *et al.* suggest that ligand binding might take place in this region[159]. In a proposed mechanism, ligand binding during transcription leads to basepairing between the RCC of the P1 loop and the GGY in the terminator[158,159]. A minimum length of the linker between P1 and terminator is crucial for this pseudoknot formation[158]. The rigidity of the guanidine-bound kissing hairpin conformation prevents closing of the lower parts of the terminator hairpin sterically, thus allowing transcription elongation through the pause site[158,159]. In-line probing revealed apparent $K_D$s for guanidinium between 64 µM[158] and 210 µM[159] which is in the same range as the other guanidine riboswitch classes. Like guanidine-II and –III, derivates like aminoguanidine, methylguanidine and ethylguanidine can also trigger conformational changes to the riboswitch RNA but with affinities decreased to the mM range, while urea and other derivates are not recognized[158,159].

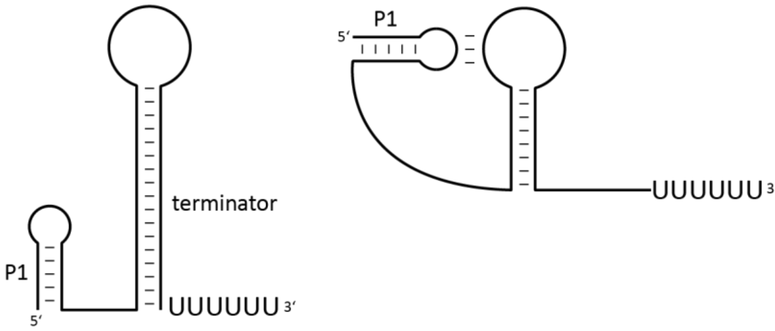


Figure 64: Schematic representation of the guanidine-IV riboswitch.

### 8.2.6 Genes under guanidine riboswitch control

$Gdm^+$ is a considered to be toxic for the cells because of its denaturing effects[151]. With the discovery of $Gdm^+$ as a ligand for guanidine riboswitches the genes under riboswitch control got more into focus. It was believed that proteins are involved in detoxification either by direct export out of the cell as evidenced by the SMR transporters or via chemical modifications[108]. It is also possible that the latter is more a $Gdm^+$ utilization than a detoxification pathway. $Gdm^+$ might serve as a nitrogen source for bacteria found in nitrogen-scarce environments like fresh water, marine or soil as the utilization pathway is more prevalent here[177]. In bacteria preferring nitrogen-rich environments the $Gdm^+$ export pathway is more commonly regulated by guanidine riboswitches.

For the utilization pathway genes for proteins like ABC transporters are found in the same operon as urea carboxylases, the urea carboxylase-associated protein family and allophanate hydrolases[177]. The function of the ABC transporter could be import of bicarbonate[160] which is necessary for carboxylation or could import $Gdm^+$ as a nitrogen source from the environment[177]. The ATP-dependent urea carboxylase can also use $Gdm^+$ as a substrate with a 40-times higher affinity most likely due to a conserved aspartate in the active center preferring a positively charged molecule[108,178] which resulted in the renaming to guanidine carboxylase. The urea carboxylase-associate protein carboxyguanidine deiminase (CgdAB) might then transform carboxyguanidine to allophanate (*N*-carboxyurea)[178]. Finally the riboswitch controlled allophanate hydrolases converts allophante into $CO_2$ and ammonium which serves as a building block for the bacteria. In some cases the $Gdm^+$ utilization pathway is controlled by two different guanidine riboswitches. Here, the transcriptional ON-switch guanidine-I is located at the beginning of the operon with a guanidine-II riboswitch as a second checkpoint before translation of the 135 kDa urea carboxylase[156]. Guanidine riboswitches classes I and III control expression of annotated agmatinases or enzymes of the arginase superfamily. These agmatinases catalyze the reaction from agmatine to putrescine and ammonium but prefer hydrolysis of $Gdm^+$ into urea and ammonium which can then be used as nitrogen source[179]. In comparison to the carboxylation pathway this reaction is slower but ATP independent.

Bacteria in nitrogen-rich environment are more likely to employ the guanidine detoxification pathway by direct export of the toxic molecule out of the bacteria cell. This is achieved using exporter from the small multidrug resistance family which are under control of different guanidine riboswitch classes. The SMR family is divided into the small multidrug pumps (SMP) like EmrE that is capable to export a broad spectrum of positively charged molecules, the paired small multidrug resistance (PSMR) proteins including YkkC and YkkD, and the suppressor of *groEL* mutation (SUG) proteins with SugE[180]. Members of the SMR protein family are characterized by their small size of 100 – 140 amino acids which fold into 4 transmembrane helices and their coupled $H^+$ antiport. Representatives that are not under riboswitch control like the well studied EmrE can export a wide range of quaternary ammonium compounds and cationic dyes[180]. Most SMR transporters are regulated by one of the guanidine riboswitch classes. Those proteins do not show this broad substrate spectrum and have $Gdm^+$ as their main target which resulted in the renaming into guanidine exporter (Gdx)[174]. $Gdm^+$ export is coupled to $H^+$ import in a 1:2 stoichiometry[174]. SugE was originally found and named after suppression of effects of mutations in the chaperonin GroEL[181]. Suppression might be achieved by exporting the protein denaturant $Gdm^+$ and thereby reducing chaperone necessity. The SugE protein from *E. coli* is regulated by a guanidine-II riboswitch[156]. With less affinity it is also able to transport methylguanidine and aminoguanidine which can be recognized by the guanidine-II riboswitch, too[174]. Furthermore, SugE can increase tolerance to disinfectants when bacteria are grown in biofilm via transport of the antiseptic detergents cetyltrimethylammonium (CTAB) and cetylpyridinium (CTP)[182].

In summary, the abovementioned roles of $Gdm^+$ in bacteria between nitrogen source and toxic properties via protein denaturation and specific inhibition show the necessity of fine-tuning cellular $Gdm^+$ concentrations. To achieve this, four structurally and mechanistically different riboswitch classes have evolved. It is important to obtain more and deeper insights into guanidine riboswitches to understand how bacteria maintain $Gdm^+$ homeostasis.

# 9 Motivation and objective

Bacteria can adapt their protein expression dependent on the requirements from the environment. One strategy for this is the use of riboswitches that can bind a ligand specifically, perform a conformational change and switch gene expression on or off. With their ability to respond to the binding of small molecules they might be an interesting target for future drug development. The Gdn-II riboswitch is one of four $Gdm^+$-binding riboswitch classes[156]. Gdn-II class riboswithces regulate the expression of small multidrug resistance efflux pumps, urea carboxylases, urea carboxylase associated proteins 1 and 2 as well as the $Ca^{2+}/H^+$ antiporter ChaA[156]. Especially in the context of multidrug resistance, manipulation of this riboswitch might be interesting. The Gdn-II riboswitch consists of two hairpins, both with an ACGR loop motif[166]. The fourth loop position is usually an A in the P1 hairpin and approximately equally distributed between A and G in P2. However, the identity of this nucleotide seems to influence the ligand binding affinities[168]. The two hairpins are connected by a linker that contains the anti-SD sequence. The SD sequence is located downstream of P2. Both loops bind $Gdm^+$ specifically and lead to the formation of a kissing loop interaction[169,170]. In-line probing experiments[156] as well as crystallography approaches[169,170] with the isolated hairpins support the kissing loop structure. This conformation is supposed to sequester the anti-SD sequence in the linker, thus freeing the SD sequence and allowing translation of the downstream gene[170]. Despite this proposed mechanism the Gdn-II riboswitch was never investigated with a full-length construct, and experimental evidence concerning the regulation mechanism is scarce. Published experimental data is limited to the loop regions[172], the isolated hairpins[169,170], or the aptamer domain[156,168].

This work aims to gain more insights into the regulatory mechanism. For this, smFRET experiments and *in vivo* experiments will be conducted focusing on the following research questions:

1) How are the conformational dynamics of the aptamer domain affected by $Mg^{2+}$ and the $Gdm^+$ ligand?
2) How do the conformations of Gdn-II change in full-length constructs containing the SD sequence?
3) How is the concentration dependency of this riboswitch?
4) Which influence has the identity of the fourth loop nucleotide?
5) How does the riboswitch behave in an *in vivo* context?

For the smFRET measurements different FRET constructs needed to be synthesized. Four different RNA length were chosen for the analysis. A short aptamer domain (47mer), a longer aptamer domain with the transcription start site (TSS) (59mer), a construct with addition of the SD sequence (74mer) and a construct containing SD sequence and also the start codon (93mer). Three labeling sites were chosen: the first in the P1 helix, the second in the linker and the third between P2 and the SD sequence. The three labeling sites were used for three different labeling schemes. Different RNA fragments were purchased and labeled via NHS-chemistry. The final RNA constructs were generated using splinted ligation. This required optimization to ligate the total of six fragments in case of the 93mer.

After the synthesis of the FRET constructs, the smFRET experiments could be conducted. In cooperation with Sebastian Falkner from the Schwierz-Neumann group the smFRET experiments were combined with coarse-grained simulations to characterize the conformations of the 47mer aptamer region in regard to $Mg^{2+}$ response. The 59mer aptamer domain was investigated with all three labeling schemes towards its conformational changes induced by $Mg^{2+}$ and $Gdm^{+}$. To investigate a possible anti-SD:SD interaction the longer constructs were investigated using smFRET in absence and presence of $Mg^{2+}$ and ligand. Mutations were introduced to the anti-SD and SD sequence aiming at disrupting a possible interaction. Furthermore, the role of the fourth loop nucleotides could be analyzed by mutating them in the 59mer and 93mer RNAs. Additionally, the concentration dependence of the conformational changes was addressed in rough ligand titrations using smFRET.

To study the riboswitch in its native *in vivo* environment, a riboswitch construct under control of a synthetic *E. coli* promoter was cloned and transformed into the bacteria. Bacteria were grown in absence and presence of $Gdm^{+}$ and the extracted RNA content was analyzed using denaturing PAGE. With this approach changes in RNA levels could compared between the conditions, different time points and with Gdn-II riboswitch constructs containing mutations at the fourth loop position.

# 10 Results

## 10.1 Preparation of smFRET constructs

Aim of this project was to study the structural dynamics of the Gdn-II riboswitch from *E. coli* to gain greater insights into its regulation mechanism. More specifically, the conformational changes of the aptamer domain were to be investigated and also how the expression platform including the SD sequence affects the riboswitch. Furthermore, more light will be shed on the role of the loop nucleotides at position four. To analyze environment dependent structural changes of the riboswitch RNA, smFRET was used. FRET requires the site-specific modification of the RNA with two different fluorophores that act as FRET donor and acceptor. Here, the cyanine dyes Cy3 and Cy5 were selected. Additionally, for the experimental setup the RNAs have to carry a biotin moiety for immobilization.

In this work different Gdn-II FRET constructs were used which differ in length, labeling sites and mutations (Figure 65). A 47mer, 59mer, 74mer and 93mer were used. The 47mer starts directly at the start of the P1 hairpin and ends with the last nucleotide of the P2 hairpin. All other constructs start at the transcription start site of the *sugE* mRNA. The 59mer constructs are used to study the aptamer alone and cover the RNA until one nucleotide after P2. The 74mer constructs ends 4 nucleotides after the proposed SD sequence and can give insights into the influence of the SD:anti-SD interaction. RNAseq data have shown that there is a drop in cellular RNA levels for this mRNA after 93 nucleotides in *E. coli* [183]. Therefore a 93mer construct was also chosen which contains the start codon as well as the first few codons of SugE.

3 labeling sites were selected to gain a better three-dimensional understanding of the riboswitch folding. Labeling sites are in the P1 stem (site 1), the linker (site 2) and between P2 and the SD sequence (site 3). Mutations were introduced in the fourth loop nucleotides of P1 and P2 hairpins or at the proposed SD and anti-SD sequence. The different constructs are named after their length, labeling site positions and mutation. For example $59_{1,2}$(GA) is a 59mer with labeling sites 1 and 2 with a G in P1 and an A in P2 and $59_{1,2}$(aSD) is mutated at the anti-SD sequence. If no mutation is defined in brackets the wildtype was used.

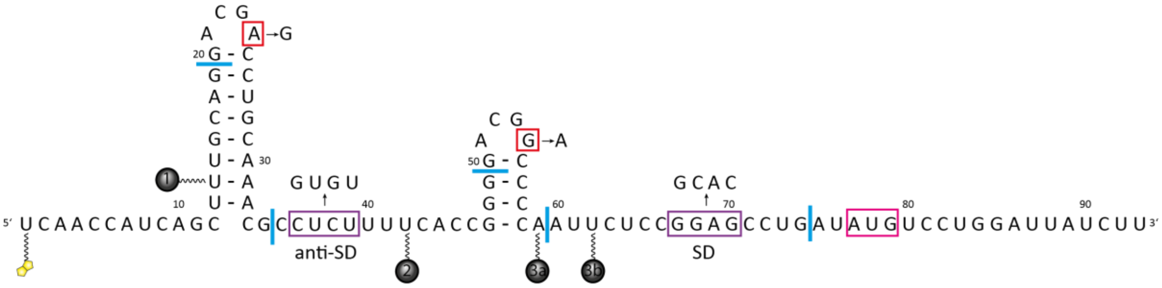


Figure 65: Sequences and ligation sites (blue) of the Gdn-II riboswitch constructs used in this work. The 59mers, 74mers and 93mers consist of the first four, five or six fragments, respectively. The fluorophore attachment sites are indicated as grey balls. For 59mer constructs the labeling site 3a and for longer constructs 3b was used. For mutations the RNA was mutated at the fourth loop position (red) or the anti-SD or SD sequence (purple). The biotin modifier is shown in yellow.

The RNA was divided into 6 fragments and purchased from Dharmacon in the 2'-ACE protected form. The first fragment had a biotin modifier and the nucleotides at the labeling sites an amino modification attached to the C5 of a uridine or the 3' phospate. The RNA was fluorophore labeled using NHS chemistry and subsequently deprotected. In some cases an additional RP-LC purification step was performed (Figure 66). During this chromatography step the fragments M412_Cy5, M415_Cy5 and M417_Cy5 showed a single peak with absorbance at both, 260 nm and 650 nm corresponding to the labeled RNA. M418_Cy5 showed an additional peak with only absorbance at 260 nm around 19% MeCN (Figure 66D). This is most likely unlabeled RNA. However, the major peak around 41% MeCN also could be detected at 650 nm. Due to the 650 nm absorbance and since the fluorophore added hydrophobicity to the RNA, resulting in a later elution, this peak might be the labeled RNA. Because the labeling occurred in most cases quantitatively with no unlabeled RNA detectable in analytical RP-LC, the additional preparative purification step was omitted for other labeled fragments.

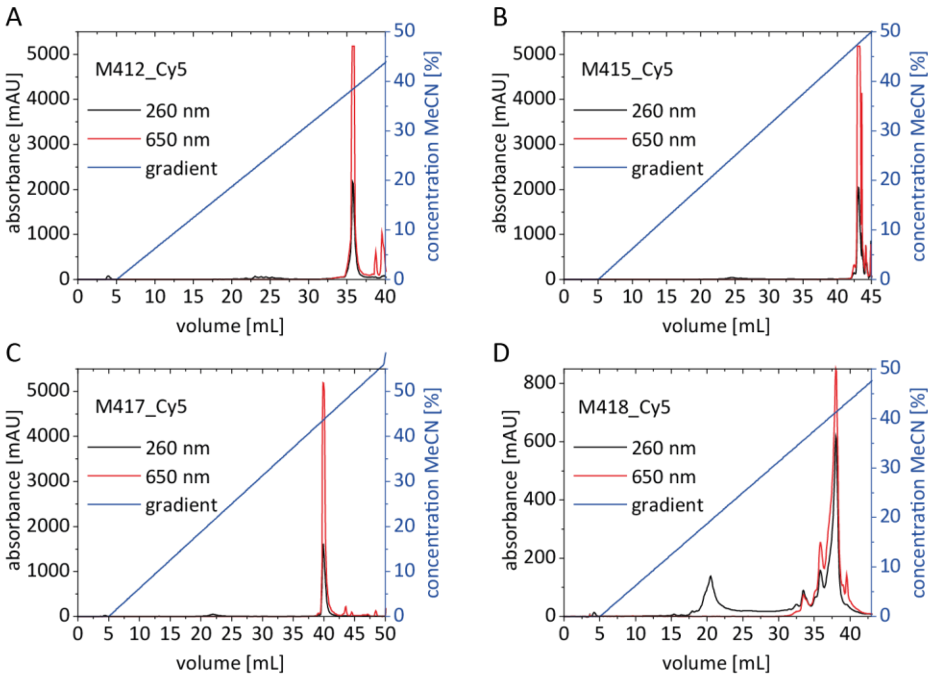


Figure 66: Reverse phase liquid chromatography of Cy5 labeled RNA fragments. A gradient from TEAA buffer to acetonitril (MeCN) was applied. A) M412_Cy5 B) M415_Cy5 C) M417_C5 D) M418_Cy5

After all labeled fragments were synthesized, the fragments were connected in a splinted ligation to yield the final FRET constructs. Either 4, 5 or 6 fragments were used to generate the 59mers, 74mers and 93mers, respectively. First, the splinted ligation was optimized for the 6 fragment ligation of the different $93_{1,3}$ constructs. In a first attempt the fragments were ligated with T4 DNA ligase at room temperature. The gel shows that in comparison to the negative control (Figure 67, lanes 6-9), both the Cy3 and Cy5 channels have two additional bands after T4 DNA ligase incubation (lanes 10-13). The more intense band in the Cy5 channel is most likely an RNA where the two fragments M418_Cy5 and M420 were ligated as it migrates similar to the reference with only these two fragments (lane 5). The weaker band might additionally have the previous fragment ligated in accordance with lane 4. The more intense band in the Cy3 channel is most likely the M408_Cy3 with the M410/411 RNAs as they run as the reference of the same nucleotide length. The other band could be the first 3 fragments from the 5' end. However, no other bands were visible that would represent a fully ligated 93mer with all 6 fragments present. Thus, after incubation with T4 DNA ligase a second step with T4 RNA ligase 2 was performed (lanes 14-17), followed by heat inactivation and digestion of the DNA splint with DNase (lanes 18-21). Now, several larger RNAs were detectable on the denaturing gel and the amount of unligated fluorophore labeled RNA had decreased visibly. The highest bands for each construct are fluorescent in both channels. This means that at least the first five 5' fragments or all 6 fragments were ligated. To confirm that these bands represent the full 93mer RNAs, several ligation reactions were performed as references where only a part of the fragments were used, starting at the 3’ end (lanes 1-5). Adding of more fragments resulted in a stepwise increase in RNA length of the longest ligation product. This analysis confirmed that indeed the ligation reaction was successful with all 6 fragments ligated. The target bands were excised and extracted from the gel and the purified FRET constructs were received. For further ligations only the T4 RNA ligase 2 was used, as it has proven to be sufficient to ligate 6 fragments.

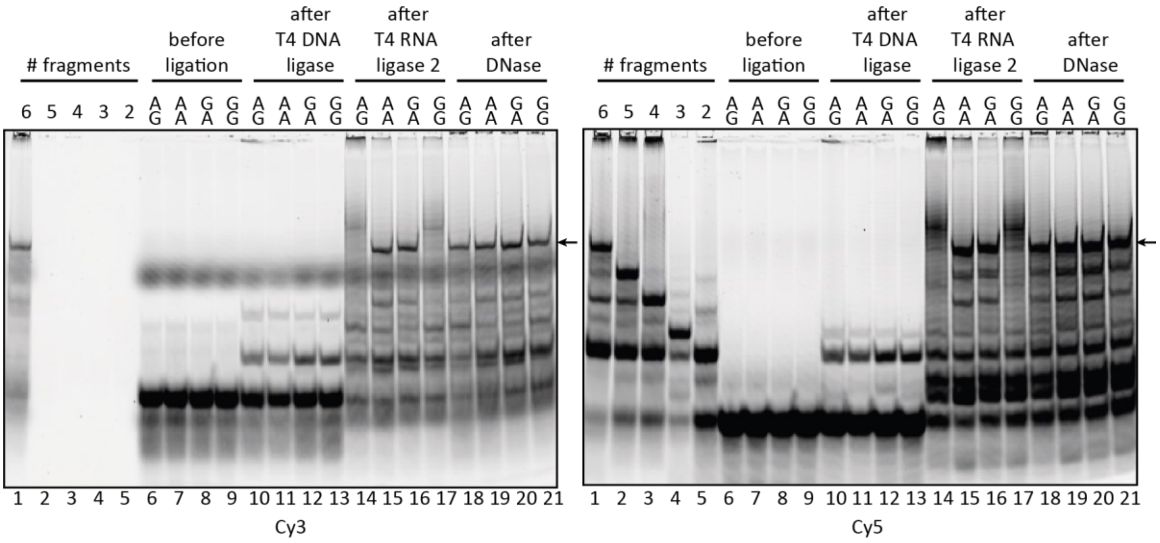


Figure 67: Optimization of the splinted ligation of the $93_{1,3}$ loop mutants analyzed on a denaturing gel. Lanes 1-5: ligation reaction with different numbers of fragments starting at the 3’ end. Lanes 6-9: Ligation reaction before addition of ligase. Lanes 10-13: Ligation reaction after incubation with T4 DNA ligase. Lanes 14-17: Ligation reaction after incubation with T4 DNA ligase and T4 RNA ligase 2. Lanes 18-21: Ligation reaction after digestion of the DNA splint with DNase. The fluorescent gel was imaged using a Typhoon scanner. The full-length 93mer is indicated by an arrow. Left: Cy3 scan, right: Cy5 scan.

The constructs $93_{1,3}$, $93_{1,3}$(AA), $93_{1,3}$(GA) and $93_{1,3}$(GG) where synthesized by Alena Sens during her supervised internship, $59_{1,2}$(AA), $59_{1,2}$(GA), $59_{1,2}$(GG) and $59_{1,2}$(AU) by Stefan Wildt during his supervised internship and $59_{1,2}$, $74_{1,2}$ and $93_{1,2}$ by Alexander Kinner during his supervised bachelor thesis. The construct $47_{1,3}$ was already synthesized during my master thesis. Figure 68 and Supplemantary Figure 91 show the analytical gels of all purified FRET constructs generated during this work. Some show impurities which could be neglected, because they are only a small percentage or do not carry both fluorophores. To summarize, all FRET constructs were successfully generated in sufficient amounts for single-molecule experiments using splinted ligation.

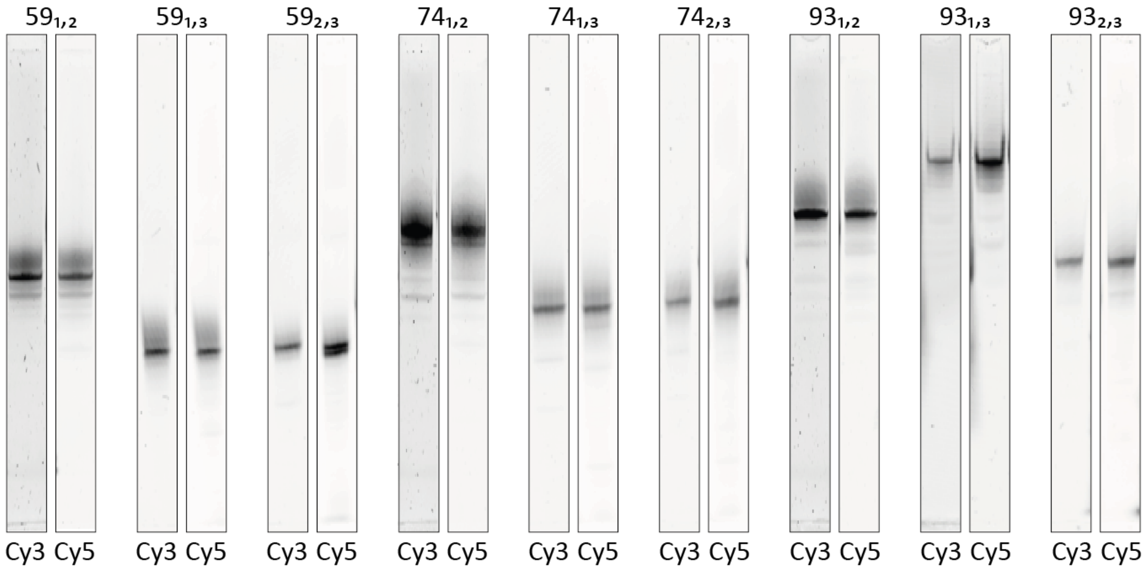


Figure 68: Analysis of unmutated purified smFRET constructs. The RNA was analyzed using denaturing gel. Different percentages of acrylamid/bisacrylamid solution were used and the gels were performed for different times. The fluorescent gels were imaged using a Typhoon scanner.

## 10.2 Analysis of aptamer folding

To reach a deeper understanding of the regulatory mechanism of the Gdn-II riboswitch the RNA was subdivided into the aptamer domain and the expression platform. In a first step the behavior of the aptamer alone was analyzed at varying concentration of $Mg^{2+}$ ions and the $Gdm^{+}$ ligand. The first construct under investigation was the $47_{1,3}$ RNA which consists only of the two hairpins P1 and P2 and the connecting linker. Using smFRET the RNA was tested under varying $Mg^{2+}$ concentrations (Figure 69A). In the absence of ions the RNA folded into a low FRET conformation of $E_{FRET} = 0.4$. This means that both FRET dyes and therefore the two hairpins have a large distance to each other. A possible structure might be the partially unfolded RNA where the hairpins are in an open conformation facing different directions. The FRET efficiencies were increased towards an intermediate FRET state in presence of near physiological $Mg^{2+}$ concentrations ($E_{FRET} = 0.57$). The fluorophores were in closer proximity. An intermediate FRET efficiency was already found in earlier studies of the same construct in the presence of $Gdm^{+}$ during my master thesis. In presence of the ligand a kissing loop formation would be expected. At concentrations of 5 mM and above in addition to the intermediate FRET state a second high FRET conformation ($E_{FRET} = 0.80\text{-}0.83$) emerged. The results from these experiments suggest a three-state model for the $47_{1,3}$ riboswitch. smFRET histograms can only shed light on inter-dye distances, and three-dimensional structures can be proposed using these information. Sebastian Falkner applied coarse-grained simulations to the Gdn-II riboswitch aptamer during his master thesis in the group of Nadine Schwierz-Neumann. From his structures he calculated the distances for the FRET labeling sites and the FRET efficiencies expected form those distances (Figure 69B). In absence of $Mg^{2+}$ he observed a low FRET state ($E_{FRET}$ = 0.40) which is in very good agreement to the experimental determined FRET efficiency of 0.41 at this condition. A representative structure shows that both hairpins were formed but they were in an extended conformation with no interaction between them (Figure 69C & D). At intermediate $Mg^{2+}$ concentrations Sebastian Falkner also found an intermediate FRET efficiency. With a Gaussian fit center of 0.65 this was slightly higher than the 0.57 found in smFRET measurements. However, based on the broad peak width with both techniques this still may well represent the same structure. According to the simulation at 1 mM $Mg^{2+}$ the Gdn-II aptamer forms a kissing loop interaction with Watson-Crick base pairing between the loop CGs. In contrast to the smFRET experiments where the high FRET state is already populated at 5 mM, it took 10 mM $Mg^{2+}$ to form this state in the simulations. Nevertheless, with FRET values of 0.80-0.83 in the smFRET measurements and 0.81 in the simulation it is likely that the data from both methods display the same RNA structure. The coarse-grained simulations revealed a secondary structure which was not expected from the kissing loop model in Figure 62 (Figure 69D). While the P1 hairpin was completely folded, the P2 hairpin folded back to the linker forming 8 base pairs (including 3 G-U base pairs) with a one nucleotide bulge. The anti-SD sequence in the linker is now base paired and not accessible to ribosome binding. In this structure derived from the coarse-grained simulations, the Cy5 labeled 3’ end is in close proximity to the Cy3 in P1 which nicely explains the high FRET efficiency.

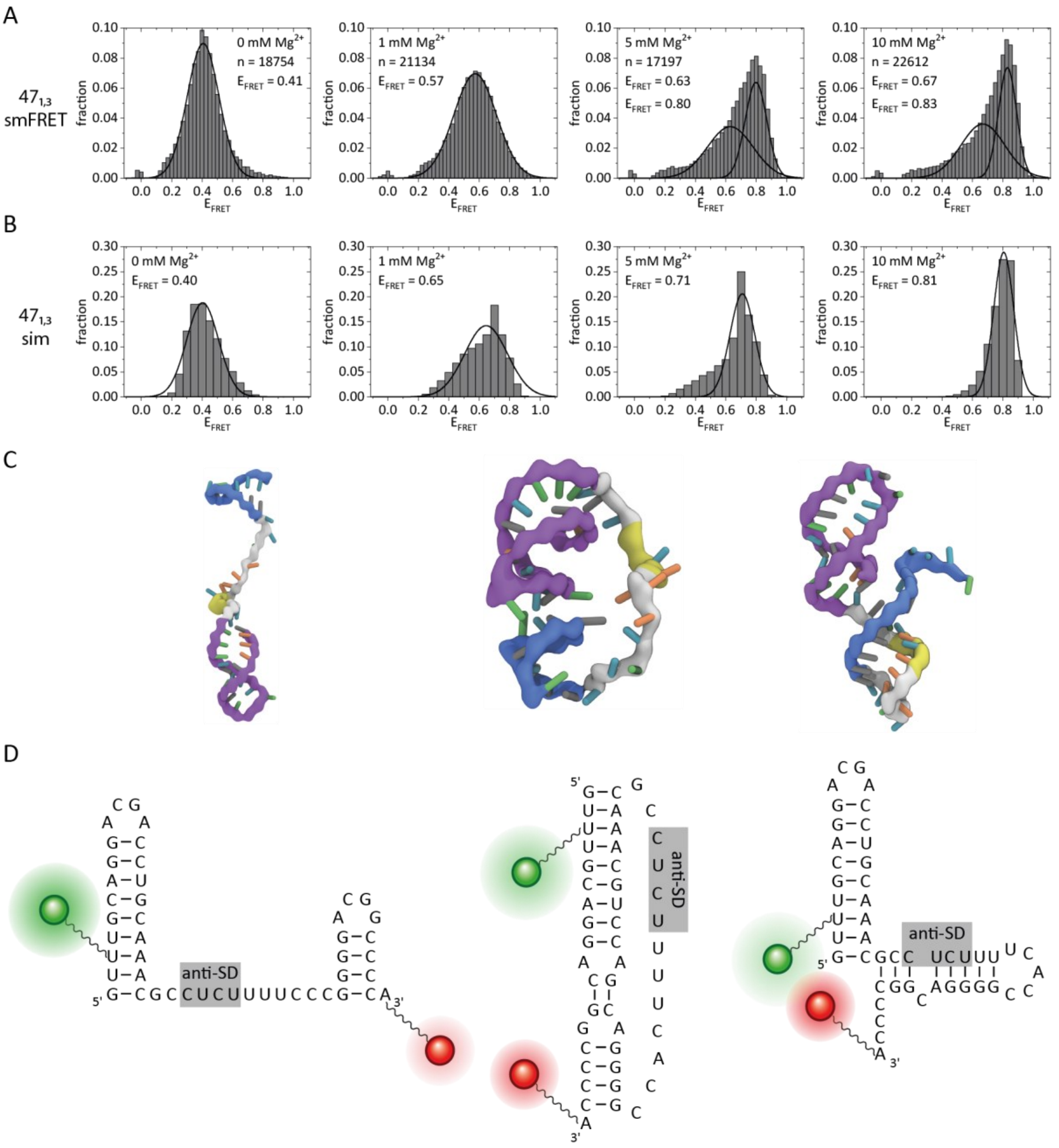


Figure 69: A) smFRET studies of the $47_{1,3}$ aptamer construct at varying $Mg^{2+}$ concentrations B) histogram analysis of calculated FRET efficiencies from coarse-grained simulations at different $Mg^{2+}$ concentrations performed by Sebastian Falkner (AK Schwierz-Neumann)[184] C) representative structures from Sebastian Falkners coarse-grained simulations[184] D) Secondary structures based on the simulation data in B. The location of the fluorophores is indicated by green (Cy3) and red (Cy5) balls.

For the following experiments the 47mer aptamer was extended to a 59mer sequence. This new aptamer construct starts at the transcription start site and also includes an adenosine at the 3' end after the P2 hairpin that does not base pair. The third labeling site for the 59mer constructs was chosen at this 3' nucleotide. The $59_{1,3}$ construct was tested for structural changes with $Mg^{2+}$and the $Gdm^+$ ligand (Figure 70). As observed for the construct with only 47 nucleotides in absence of both ions, a low FRET population was observed ($E_{FRET} = 0.36$). With 10 mM $Mg^{2+}$ only an intermediate FRET state of 0.70 was found. This is in contrast to the 47mer measurements at

10 mM $Mg^{2+}$ which showed an additional high FRET population. The measurements in presence of ligand appeared to be similar with a FRET value of 0.69. Based on the mechanistic model[170], crystal structures[169] and coarse-grained simulation (Figure 69B & C) this might be the kissing loop conformation.

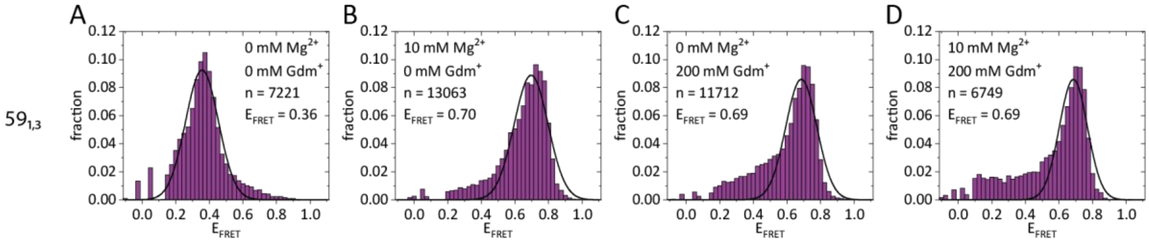


Figure 70: smFRET experiments of the $59_{1,3}$ FRET construct at varying $Mg^{2+}$ and $Gdm^+$ concentrations. The histograms were fitted with a Gaussian fit.

For a better three-dimensional characterization of the Gdn-II riboswitch the RNA was labeled at the linker and the 3’ end ($59_{2,3}$). In absence of $Mg^{2+}$ and $Gdm^+$ a sharp FRET peak at 0.81 was detected in the histograms (Figure 71A). This high FRET state with a small inter dye distance would be expected if the linker is in an extended conformation. Addition of $Mg^{2+}$ lead to a shift in conformations (Figure 71B). Two different states were observed with FRET efficiencies of 0.57 and 0.70, respectively. It is likely that one of these states corresponds to the kissing loop conformation. Without additional experimental data it is however not possible to assign the FRET states to a distinct RNA fold. Independent of the presence of $Mg^{2+}$, at 200 mM ligand a single sharp peak ($E_{FRET}$ = 0.70-0.71) appeared in the FRET histograms. The small width of this peak (w = 0.10) is a sign that the structure is apparently quite static at least in the fluorophore region. The $Gdm^+$ state represents the same inter-dye distance and might belong to the same RNA fold as the peak in the $Mg^{2+}$ measurement with the same FRET efficiency.

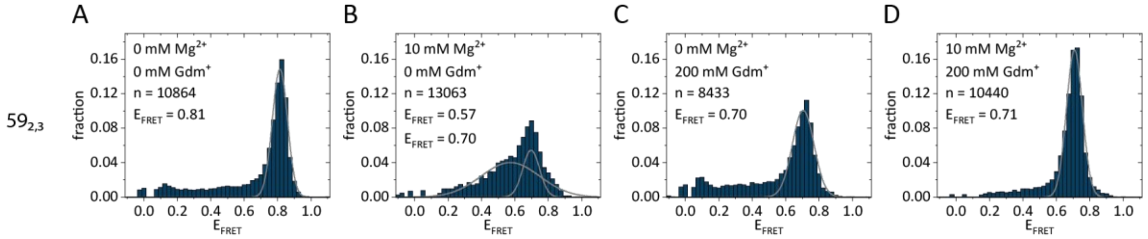


Figure 71: smFRET experiments of the $59_{2,3}$ FRET construct at varying $Mg^{2+}$ and $Gdm^+$ concentrations. The histograms were fitted with one or two Gaussian fits.

Next, the $59_{1,2}$ FRET construct was used with the same conditions as the other labeling sites. In the absence of both $Mg^{2+}$ and $Gdm^+$, a single sharp peak with a FRET efficiency of 0.60 was detectable (Figure 72A). This FRET value of 0.60 is higher than for the $59_{1,3}$ construct with 0.36. This was expected since the distance between labeling sites 1 and 2 is shorter than between sites 1 and 3.

Labeling site 2 in the linker is closer to the P2 hairpin than to the P1 hairpin which explains that the FRET efficiency is lower in the $59_{1,2}$ construct than the 0.81 for the $59_{2,3}$ RNA. Taken all 3 constructs into account the conformation in absence of $Mg^{2+}$ ions or ligand fits well into the model of an open conformation. With addition of $Mg^{2+}$ the FRET efficiency shifts only slightly from 0.60 to 0.66 (Figure 72B). This means that the fluorophores have a similar distance in both conformation. However, the data from the other two 59mer constructs show a stronger shift so it can be assumed that the free state and the $Mg^{2+}$-state represent two different structures. Surprisingly, the FRET efficiency of the $59_{1,2}$ construct increased to 0.89-0.90 when $Gdm^+$ was present (Figure 72C&D). This increase in FRET is due to a closer proximity of the two fluorophores in the P1 stem and the linker when the ligand is present.

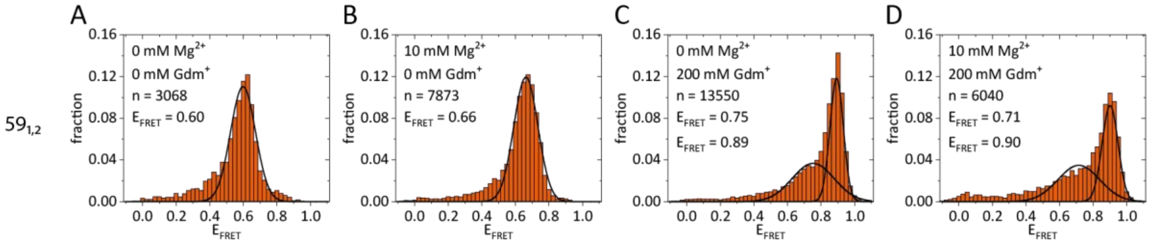


Figure 72: smFRET experiments of the $59_{1,2}$ FRET construct at varying $Mg^{2+}$ and $Gdm^+$ concentrations. The histograms were fitted with one or two Gaussian fits.

## 10.3 Influence of SD sequence

After establishing the three labeling schemes with the aptamer sequence and understanding the conformational changes triggered by environmental changes, the transfer to the expression platform was investigated. In translational riboswitches like the Gdn-II riboswitch the regulation occurs via accessibility of the SD sequence. To analyze a possible anti-SD:SD interaction, a 74mer sequence which contains the SD sequence and a 93mer with the SD sequence and the start-codon were analyzed. Without ions the $74_{1,3}$ and the $93_{1,3}$ both showed a low FRET peaks with an efficiency of 0.29 and 0.14, respectively (Figure 73A & E). Both are lower than the value of 0.36 observed for the $59_{1,3}$ RNA (Figure 70A). Addition of $Mg^{2+}$ shifted the equilibrium to different higher FRET efficiencies (0.46 and 0.25) (Figure 73B & F). However, a high FRET peak evident in the 59mer at these conditions was not detectable with any of the longer constructs. With high ligand concentrations but no $Mg^{2+}$, two different states were visible in the histograms (Figure 73C & G). The first showed FRET efficiencies comparable to the respective $Mg^{2+}$ states for that length. The second peaks were found at FRET efficiencies of 0.62 and 0.66 for the $74_{1,3}$ and the $93_{1,3}$, respectively. These second peaks had the same FRET efficiencies as the peaks when 10 mM $Mg^{2+}$ and 200 mM $Gdm^+$ were in the solution (Figure 73D & H). In this case there was no additional $Mg^{2+}$

conformation.

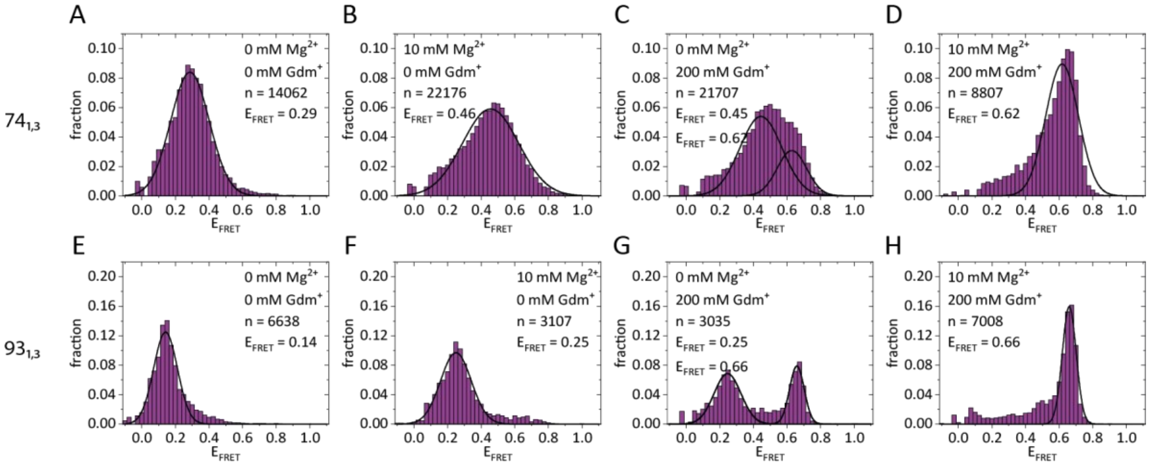


Figure 73: smFRET experiments of the $74_{1,3}$ FRET construct (A-D) and the $93_{1,3}$ FRET construct (E-H) at varying $Mg^{2+}$ and $Gdm^{+}$ concentrations. The histograms were fitted with one or two Gaussian fits.

The next labeling scheme used for analyzing the effects of the SD sequence had the attachment sites in the linker and near the SD sequence. The $59_{2,3}$ had shown a high FRET population if neither $Mg^{2+}$ or ligand were present which was shifted to lower FRET efficiencies with ions (Figure 71). In the $74_{2,3}$ RNA the high FRET state was only slightly populated (Figure 74A) and completely absent for the $93_{2,3}$ construct (Figure 74E). Instead a low FRET state of 0.40 ($74_{2,3}$) or 0.34 ($93_{2,3}$) were the major peaks under this condition. The large difference in FRET values depending on the presence of the SD sequence suggests that a significantly different structure was formed, affecting the region between the two fluorophore attachment sites. The FRET values remained similar when $Mg^{2+}$ was added (Figure 74B & F). The small fraction of high FRET remained for $74_{2,3}$ (Figure 74B). For this labeling scheme again two discernible populations were observed under only high $Gdm^{+}$ concentrations (Figure 74C & G). The lower of the two states ($E_{FRET} = 0.44/0.41$) was for both construct lengths' slightly increased from the $Mg^{2+}$ state. The other state FRET efficiencies of 0.70 and 0.71 had resemblance with the ligand bound state found in the aptamer only construct (Figure 71D). Again, this was the only populated state when $Mg^{2+}$ was present in addition to $Gdm^{+}$ (Figure 74D & H).

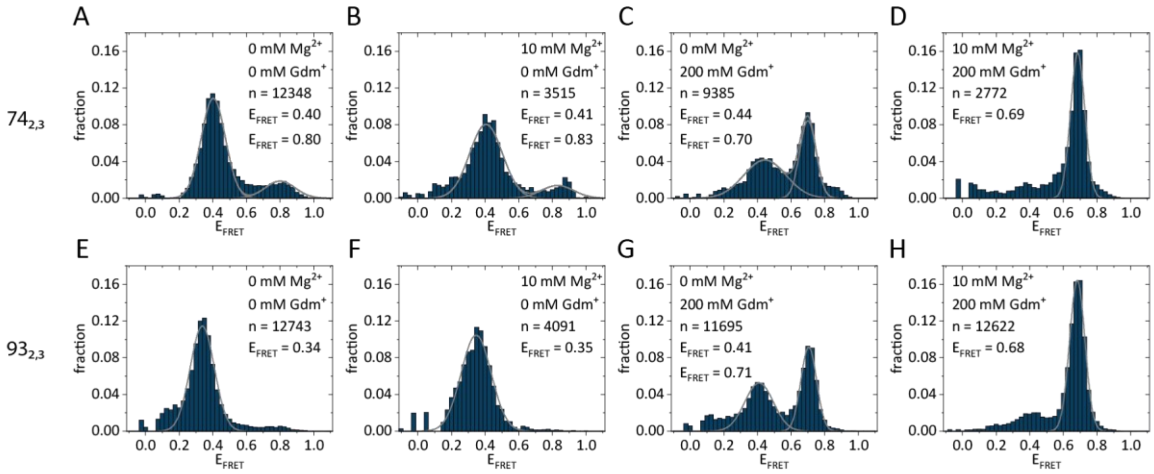


Figure 74: smFRET experiments of the $74_{2,3}$ FRET construct (A-D) and the $93_{2,3}$ FRET construct (E-H) at varying $Mg^{2+}$ and $Gdm^{+}$ concentrations. The histograms were fitted with one or two Gaussian fits.

The results for the previous two labeling schemes showed differences for the two tested lengths in absence of the ligand (Figure 73 & Figure 74). In contrast, for the $74_{1,2}$ and $93_{1,2}$ the centers of the FRET fit were comparable. Without $Mg^{2+}$ or $Gdm^{+}$ ions a FRET peak of 0.63 was visible (Figure 75A & E). This peak shifted slightly to 0.68-0.70 with $Mg^{2+}$ (Figure 75B & F). Figure 75F shows that the measurement of $93_{1,2}$ had an increased background, especially in the low FRET region. It is not possible to state whether another population was hidden underneath. With only $Gdm^{+}$ present, this labeling sites also had two different peaks ($E_{FRET}$ = 0.73-0.74 and $E_{FRET}$ = 0.90-0.91) (Figure 75C & G). The high FRET peak became the predominant peak when $Mg^{2+}$ and ligand were used in the FRET experiment (Figure 75D & H). Overall, the FRET efficiencies between the three tested lengths with this labeling scheme did not deviate much. This could be explained that the distances in the 5’ region were coincidentally similar in the different folds, or that the major structural changes induced by the SD sequence affected mostly the region after the second labeling site located in the linker which would then not be reported as a change in distance between the FRET pair.

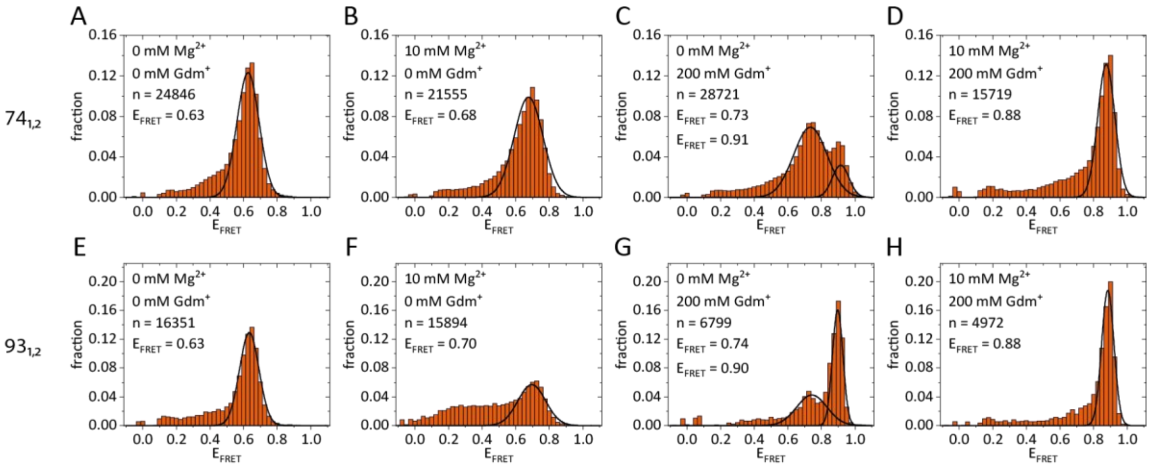


Figure 75: smFRET experiments of the $74_{1,2}$ FRET construct (A-D) and the $93_{1,2}$ FRET construct (E-H) at varying $Mg^{2+}$ and $Gdm^{+}$ concentrations. The histograms were fitted with one or two Gaussian fits.

In general, the peak width for the $Mg^{2+}$ conformation (Figure 73B & F, Figure 74B & F & Figure 75B & F) was wider than the state in absence of $Mg^{2+}$ (Figure 73A & E, Figure 74A & E & Figure 75A & E) for all tested constructs. This is a sign for a more dynamic structure. The ligand-bound conformation was always present in a narrow peak in comparison which means that this might be a more static structure.

## 10.4 Analysis of anti-SD:SD interaction

RNAs that contained the SD sequence had a different fold than the aptamer only constructs, as evidenced by different FRET efficiencies between the labeling schemes 1 and 3 as well as between 2 and 3 in absence of ligand. In the context of a translational ON-switch the SD sequence would be inaccessible for ribosome binding without ligand. One possibility for a sequestered SD sequence is an interaction with a complementary anti-SD sequence. To test whether such an interaction is the cause for the different conformations, either the SD sequence or the anti-SD sequence were mutated with the intent to disrupt this interaction. The presumed anti-SD sequence starting at nucleotide 36 (CUCU) was mutated to GUGU and the SD sequence from nucleotide 67 from GGAG to GCAC. Mutation of both sites simultaneously was then supposed to restore the interaction.

First, the $59_{1,3}$(aSD) was analyzed whether mutation of the anti-SD in the linker between the two hairpins would affect the conformations identified before. The measurement in absence of $Mg^{2+}$ and $Gdm^{+}$ revealed a FRET efficiency of 0.43 (Figure 76A) which is a small increase from the unmutated $59_{1,3}$ ($E_{FRET}$ = 0.36, Figure 70A). Addition of $Mg^{2+}$ lead to a peak center at $E_{FRET}$ = 0.57 (Figure 76B). This FRET value is lower than the high FRET state of 0.70 of $59_{1,3}$ (Figure 70B). In Figure 76A&B some molecules showed lower FRET efficiencies. It is not clear whether this is solely background or a hidden FRET state. Supplying the measurement buffer with the ligand lead to a high FRET efficiency independent of the presence of 10 mM $Mg^{2+}$ (Figure 76C & D). This is in agreement with the ligand-dependent high FRET states observed before (Figure 70D, Figure 73D & H).

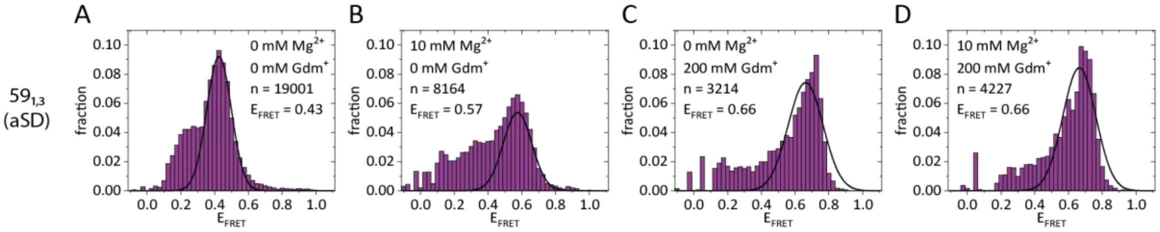


Figure 76: smFRET experiments of the $59_{1,3}$(aSD) FRET construct which is mutated at the anti-SD sequence at varying $Mg^{2+}$ and $Gdm^+$ concentrations. The histograms were fitted with a Gaussian fits.

The $74_{1,3}$(aSD) FRET RNA showed a high level of acceptor bleaching during the measurement and consequently a high donor only peak which proved troublesome to remove during analysis. For this reason the results from this construct should be considered with care. All three mutated $74_{1,3}$ constructs populated a low FRET state (0.20-0.24) at 0 mM $Mg^{2+}$ and 0 mM $Gdm^+$ (Figure 77A, E & I). However, the peaks differed in their peaks width. Mutation of the SD sequence alone resulted in a peak width of 0.16. A mutated anti-SD sequence lead to a peak width of 0.27 in the FRET histogram. Introducing two mutation sites into the RNA, it became more dynamic evidenced by a peak width of 0.37. $Mg^{2+}$ induced an increase in FRET efficiencies to 0.32 and 0.53 for $74_{1,3}$(SD) and $74_{1,3}$(aSD/SD), respectively (Figure 77F & J). At high $Gdm^+$ concentrations without $Mg^{2+}$ again two populations were visible in the FRET histograms (Figure 77G & K). The lower one in each case had a similar FRET efficiency to the measurement with $Mg^{2+}$. The high FRET state with peak centers at 0.69 ($74_{1,3}$(SD)) and 0.63 ($74_{1,3}$(aSD/SD)) was consistent with other measurements in presence of ligand for this labeling scheme. In presence of $Mg^{2+}$ and $Gdm^+$ the high FRET peak between 0.63 and 0.66 became dominant for all mutated $74_{1,3}$ constructs (Figure 77D, H & L) in agreement with all other RNAs with this labeling scheme.

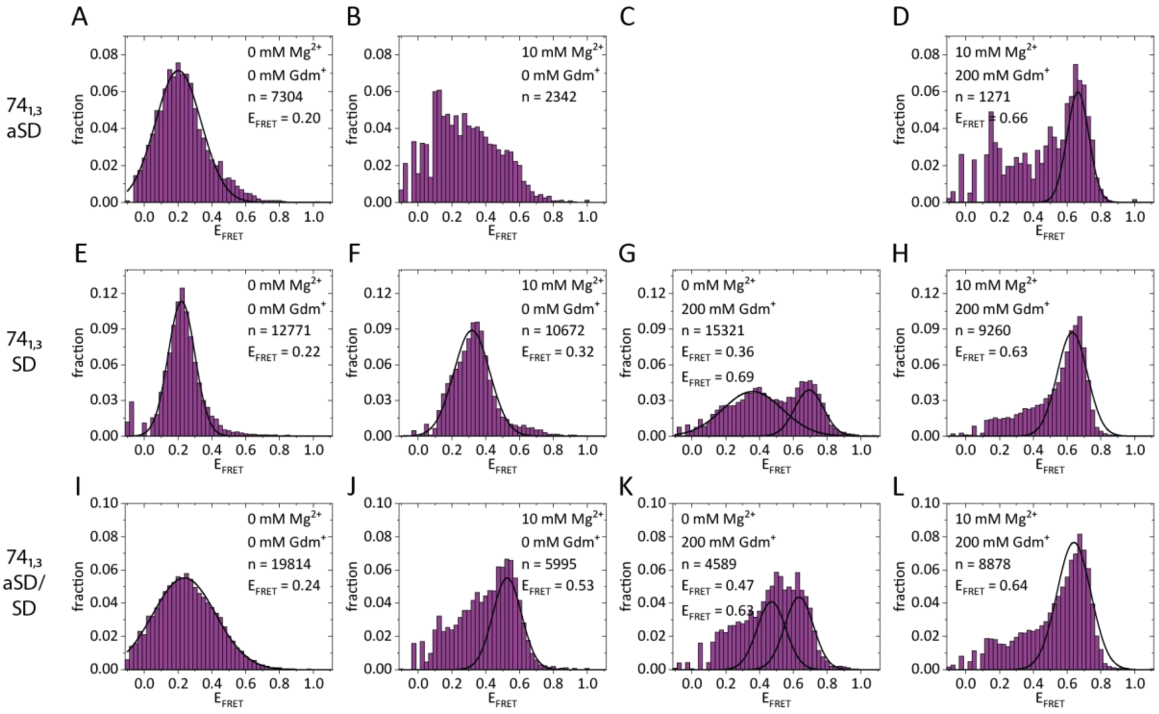


Figure 77: smFRET experiments of $74_{1,3}$ FRET constructs with mutations either at the anti-SD sequence(A-D), the SD sequence (E-H) or at both sites (I-L) at varying $Mg^{2+}$ and $Gdm^{+}$ concentrations. The histograms were fitted with one or two Gaussian fits.

Extending the $59_{2,3}$ construct to the $74_{2,3}$ sequence resulted in a different behavior in the FRET experiment (Figure 71 & Figure 74). For this reason, the $59_{2,3}$ and $74_{2,3}$ constructs were also used for anti-SD/SD mutation studies. The histograms of the $59_{2,3}$(aSD) RNA all showed high background levels (Figure 78). However under each condition a peak was identifiable. Without ligand and $Mg^{2+}$ a high FRET state ($E_{FRET} = 0.79$) was populated (Figure 78A). The same peak center was found when $Mg^{2+}$ was present (Figure 78B). With ligand present with or without $Mg^{2+}$ FRET efficiencies of 0.69-0.71 were found (Figure 78C & D).

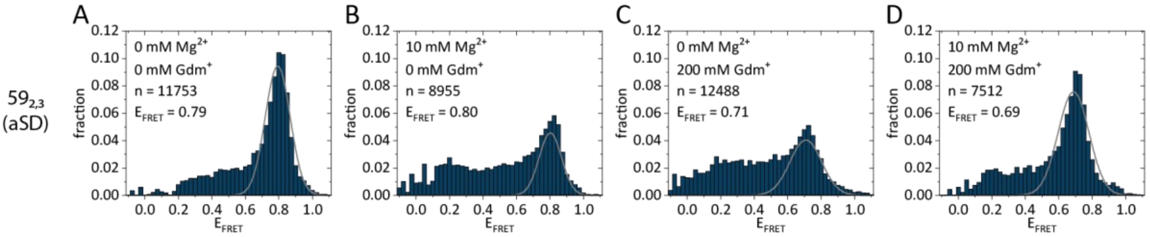


Figure 78: smFRET experiments of the $59_{2,3}$(aSD) FRET construct which is mutated at the anti-SD sequence at varying $Mg^{2+}$ and $Gdm^{+}$ concentrations. The histograms were fitted with a Gaussian fits.

For the $74_{2,3}$(aSD) RNA two peaks were observed in the FRET analysis under $Mg^{2+}$- and ligand-free as well as at $Mg^{2+}$ conditions (Figure 79A & B). The peak centers of 0.39-0.40 and 0.80-0.86 of the two peaks were comparable with the results for the $74_{2,3}$ RNA (Figure 74A & B). However, equilibrium between both peaks shifted towards the high FRET population when the anti-SD sequence is present: from 18% to 28% in absence of $Mg^{2+}$ (Figure 74A & Figure 79A) and from 15% to 27% in the presence of $Mg^{2+}$ (Figure 74B & Figure 79B). At high ligand concentrations with or without $Mg^{2+}$ the FRET analysis of $74_{2,3}$(aSD) revealed a peak at $E_{FRET}$ = 0.66-0.67 (Figure 79C & D). The $74_{2,3}$(SD) construct was mutated in the SD sequence. In absence of $Mg^{2+}$ and $Gdm^{+}$ ions, the molecules populated a FRET state of 0.56 (Figure 79E). This FRET efficiency did not match with any of the two states found in the unmutated construct ($E_{FRET}$ = 0.40 & 0.80, Figure 74A). The peak center for $74_{2,3}$(SD) with 10 mM $Mg^{2+}$ remained similar to the ion-free condition (Figure 79F). With 10 mM $Mg^{2+}$ and 200 mM $Gdm^{+}$ a FRET peak of 0.69 emerged (Figure 79H). In a high ligand, no $Mg^{2+}$ environment a mixture of the ligand-dependent and ligand-free populations were found (Figure 79G). The FRET results for the double-mutated $74_{2,3}$(aSD/SD) had generally slightly lower FRET efficiencies but were overall comparable to the $74_{2,3}$(SD) construct (Compare Figure 79I-L to Figure 79E-H).

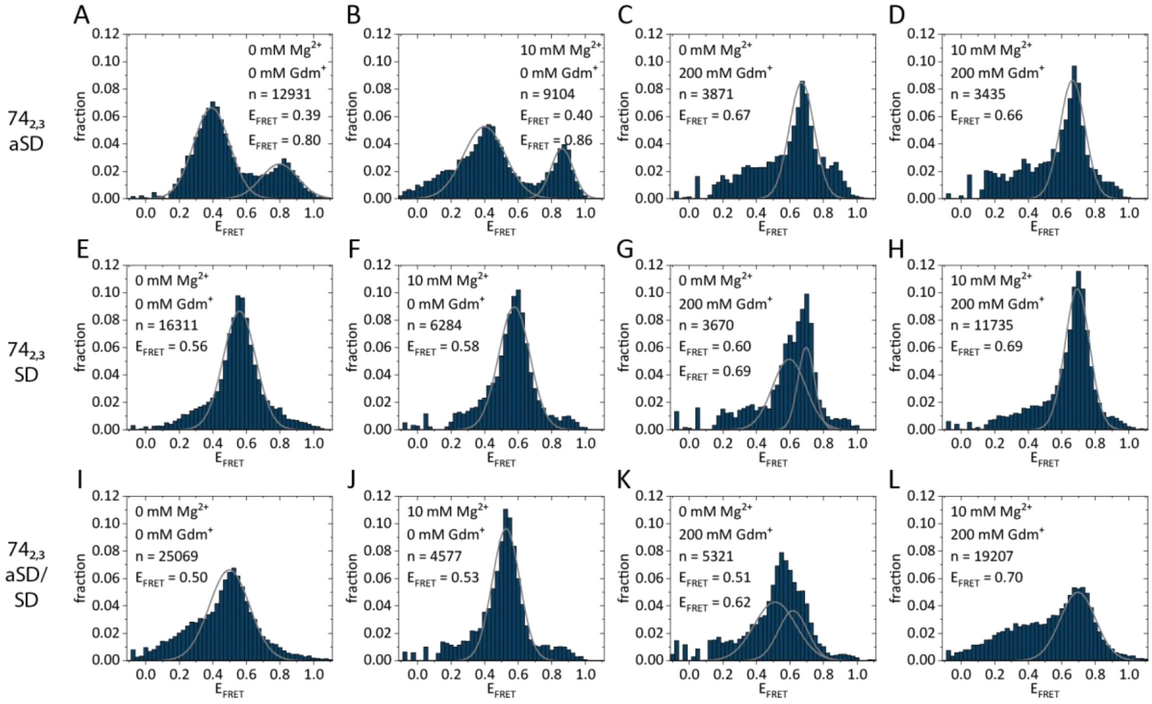


Figure 79: smFRET experiments of $74_{2,3}$ FRET constructs with mutations either at the anti-SD sequence(A-D), the SD sequence (E-H) or at both sites (I-L) at varying $Mg^{2+}$ and $Gdm^{+}$ concentrations. The histograms were fitted with one or two Gaussian fits.

## 10.5 Influence of loop nucleotides

The Gdn-II riboswitch motif contains two hairpins with an ACGR tetraloop each. The identity of the fourth loop nucleotide depends on the organism. In most cases this is an A in the P1 and approximately equally distributed between A and G in P2[166]. In the *sugE* riboswitch from *E. coli* this is an A in P1 and a G in P2. Crystal structures were solved using homodimers with an ACGA loop[169,170]. The fourth loop nucleotides were found to be twisted out of the binding pocket and forming a stacking interaction with its counterparts from the opposite hairpin. Higgins *et al.* observed a phenotypical ON-state in *E. coli* when the nucleotide is mutated to either pyrimidine even in absence of ligand[173]. The following experiments aim to understand the role of the conserved purines in the loop.

To analyze this, the $93_{1,3}$ construct was chosen. FRET RNA with all combinations of purines in the two hairpins were synthesized as described above, and used in smFRET experiments. Analysis of FRET histograms showed that two constructs each exhibited similar tendencies: RNAs with an A in the P2 hairpin ($93_{1,3}$(AA) & $93_{1,3}$(GA)) and with a G in the P2 hairpin ($93_{1,3}$(AG) & $93_{1,3}$(GG)) (Figure 80). As described for the $93_{1,3}$(AG) construct above (Figure 73E-H), the $93_{1,3}$(GG) had a low FRET peak ($E_{FRET}$ = 0.16) (Figure 80E) which shifted to a higher efficiency (0.27) in presence of 10 mM $Mg^{2+}$ (Figure 80F). At 200 mM $Gdm^+$ two populations were present (Figure 80G): The one at lower FRET had a similar efficiency as detected in presence of $Mg^{2+}$ (0.24) and the higher FRET state was comparable to the population found under high $Mg^{2+}$ and ligand concentrations ($E_{FRET}$ = 0.68-0.70) (Figure 80H). The two constructs with an A in P2 ($93_{1,3}$(AA) & $93_{1,3}$(GA)) populated two different FRET states in absence of either $Mg^{2+}$ or $Gdm^+$ (Figure 80I & M). One peak had a FRET efficiency of 0.15 similar to the other two variants and the second novel peak centered at an efficiency of 0.30. Also with $Mg^{2+}$ the riboswitch split into two detectable FRET states with the lower one at 0.28-0.32 comparable to the other RNAs and an additional peak between 0.56 and 0.59 (Figure 80J & N). However, while a G in P2 had two populations with 200 mM $Gdm^+$ (Figure 80C & G), theses high amounts of ligand were sufficient to almost quantitatively switch to the ligand-dependent high FRET state in case of an A in P2 (Figure 80K & O).

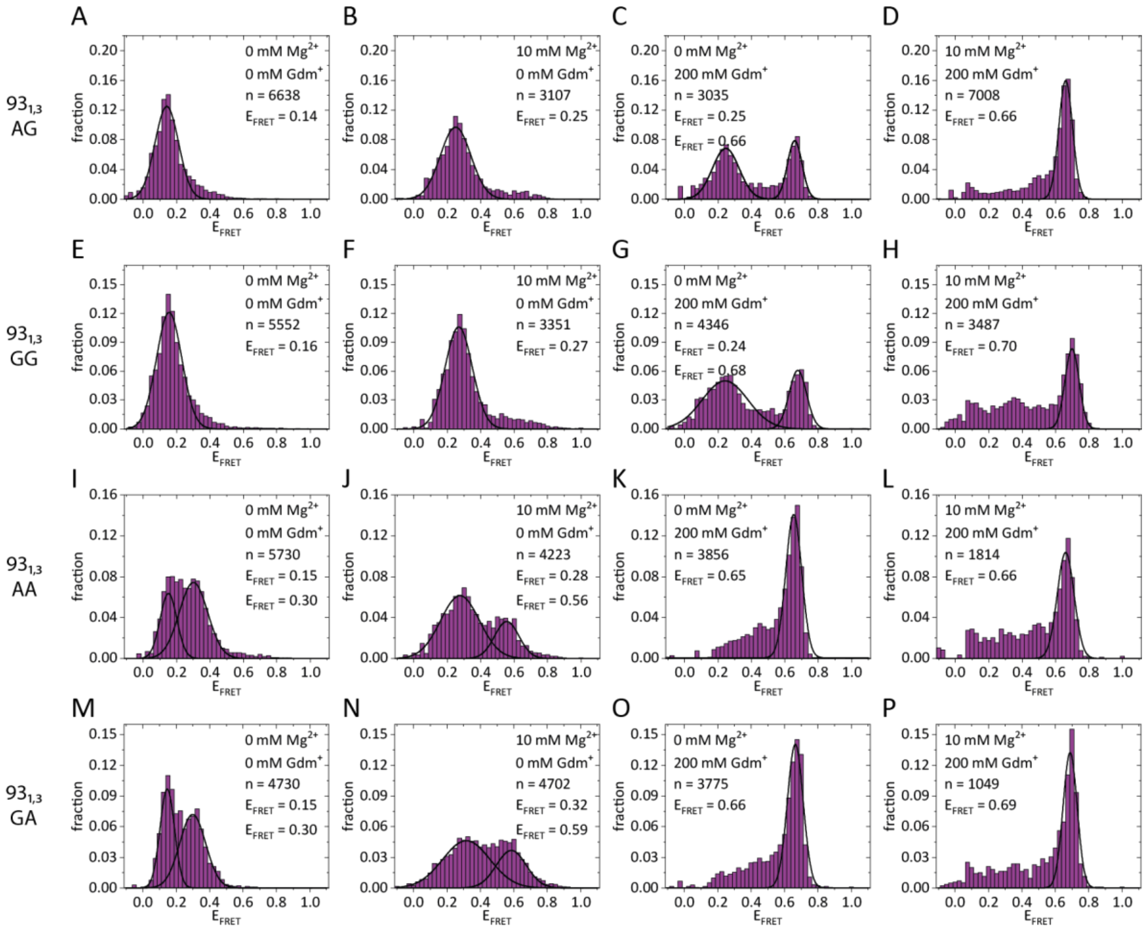


Figure 80: smFRET experiments of $93_{1,3}$ FRET constructs with mutations at the fourth loop position in P1 and P2 at varying $Mg^{2+}$ and $Gdm^{+}$ concentrations. The histograms were fitted with one or two Gaussian fits.

In a next step the influence of the fourth loop nucleotide was investigated with the $59_{1,2}$ labeling. For this, the four purine variants and an additional $59_{1,2}$(AU) were prepared. FRET experiments showed that no new FRET states emerged from the mutations (Figure 81). With a few exceptions, the riboswitch variants had a similar response to the buffer conditions used in the measurements. While all other mutants displayed a single peak with an efficiency around 0.65-0.68 at 10 mM $Mg^{2+}$, the shape of $59_{1,2}$(AU) suggests a second conformation (Figure 81R). It is possible that in the AU variant the $Mg^{2+}$-free and $Mg^{2+}$-dependent states were overlapping. Another difference between the mutants was observed at high $Gdm^{+}$ concentration. In case of an A at the fourth loop position ($59_{1,2}$(AA) & $59_{1,2}$(GA)) nearly all molecules were present in the ligand-dependent peak (Figure 81K & O). The other tested variants showed a shoulder or tailing of that peak. This was most evident with the $59_{1,2}$(GG) construct. This could be attributed to background, or be a sign for not fully ligand-saturating conditions. In summary, all variations of the fourth loop nucleotide under investigation were able to switch into the ligand-dependent FRET state.

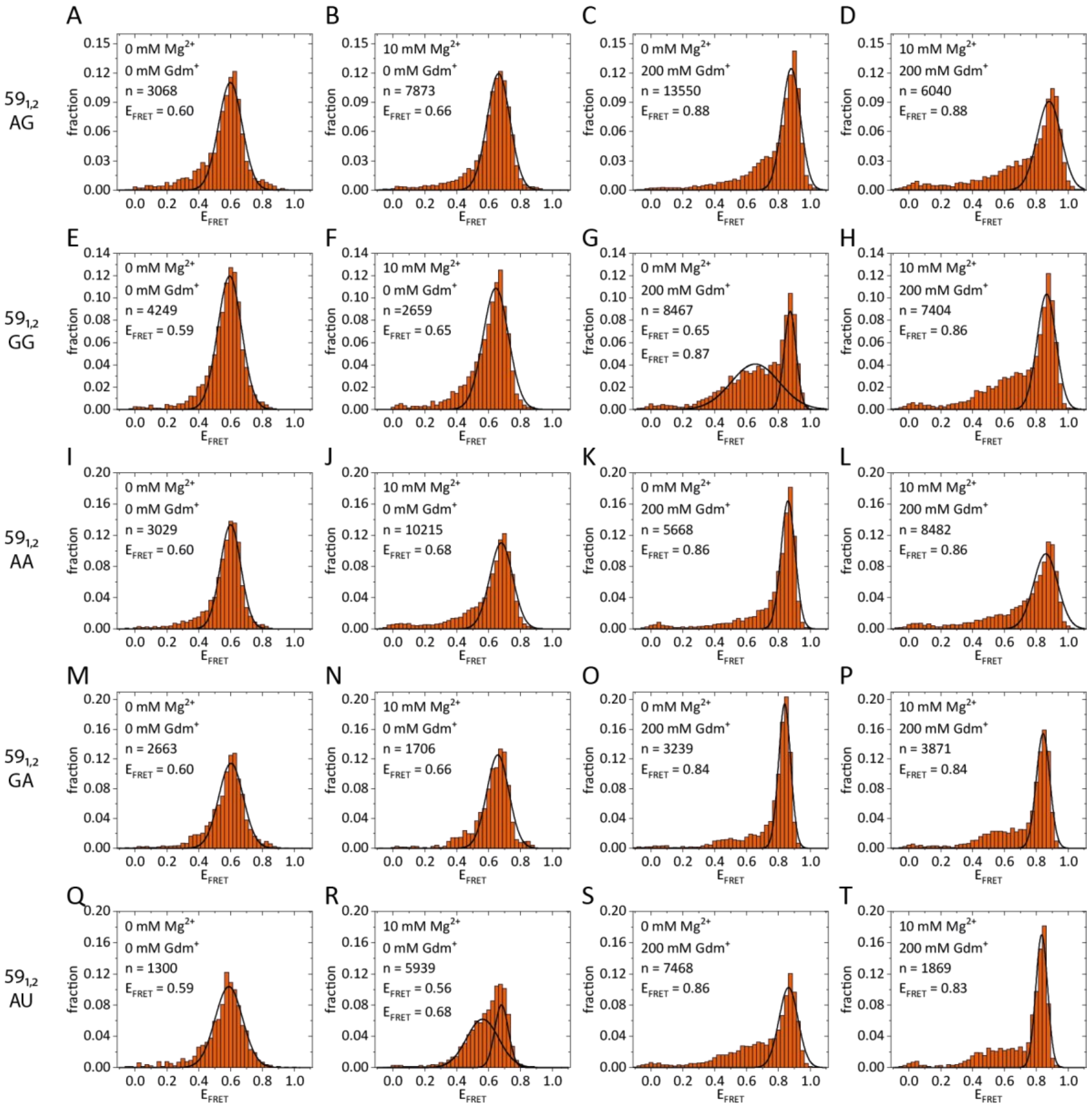


Figure 81: smFRET experiments of $59_{1,2}$ FRET constructs with mutations at the fourth loop position in P1 and P2 at varying $Mg^{2+}$ and $Gdm^+$ concentrations. The histograms were fitted with one or two Gaussian fits.

## 10.6 Titration experiments

The previous experiments revealed that mutations of the fourth loop nucleotide did not significantly influence the switch to the ligand-dependent conformation at high $Mg^{2+}$ and $Gdm^+$ concentrations. It is possible that the sequence conservation at these positions can fine-tune the amount of the toxic $Gdm^+$ tolerable before a regulation towards detoxification is necessary for each organism. To test this hypothesis coarse ligand titration experiments were performed at constant, near physiological amounts of 1 mM $Mg^{2+}$ for the $59_{1,2}$(AG) WT and the $59_{1,2}$(GG) RNA.

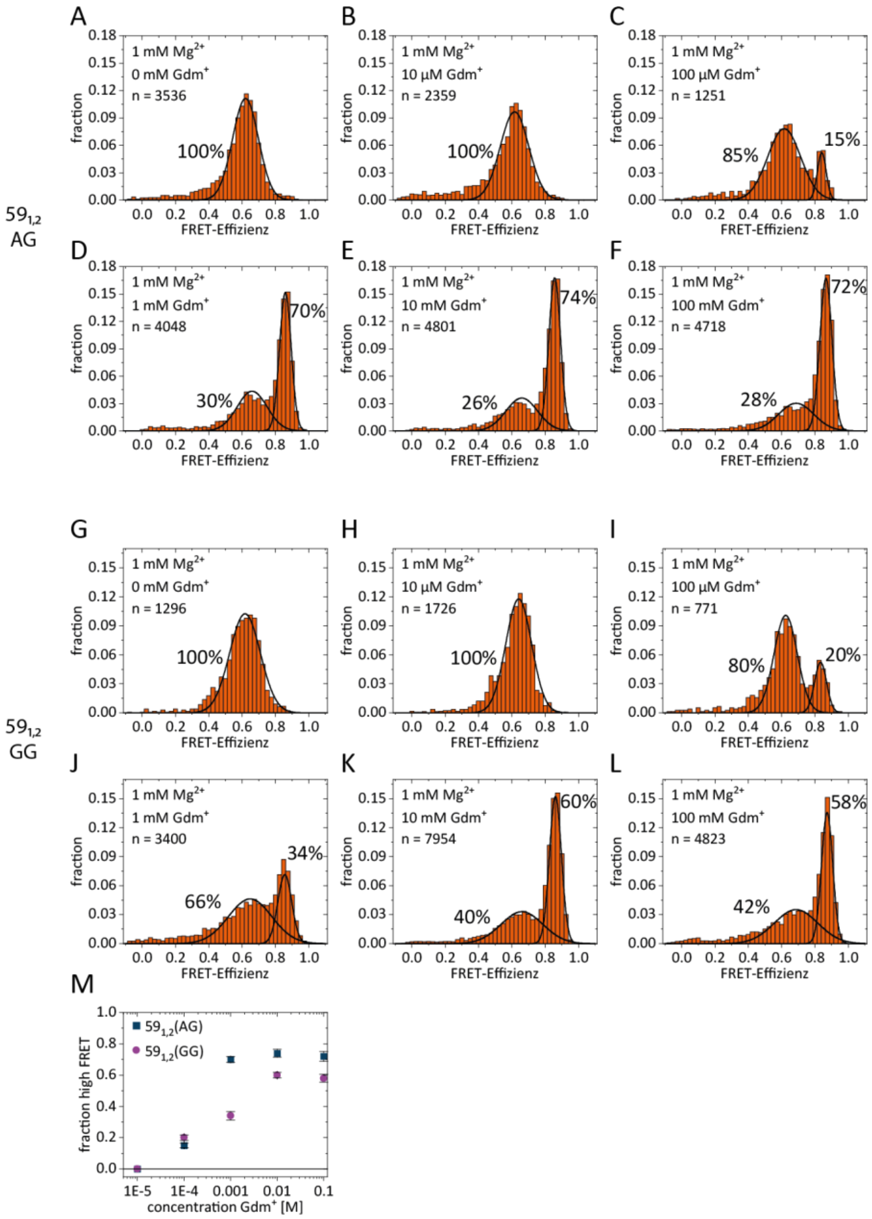


Figure 82: smFRET histograms of the Gdm+ titration of the $59_{1,2}$(AG) construct (A-F) and the $59_{1,2}$(GG) construct (G-L) at 1 mM $Mg^{2+}$. The peaks were fitted with one or two Gaussian fits. M) Fraction of molecules populating the high FRET state for the two constructs. Percentages were calculated using the area under the fits.

The fraction of the ligand-dependent high FRET was used as a read-out (Figure 82M). Both loop variants did not show a high FRET peak at 10 µM or without $Gdm^+$ (Figure 82A, B, G & H). Starting at 100 µM the high FRET state appeared in addition to the $Mg^{2+}$ peak (Figure 82C & I). Fitting of the two peaks revealed that 15% of the molecules populated the ligand-state in case of $59_{1,2}$(AG) and 20% in case of $59_{1,2}$(GG). For $59_{1,2}$ the amount of high FRET increased to 70% at 1 mM $Gdm^+$ and plateaued above 70% for the following titration steps (Figure 82D-F). In comparison, the fraction of $59_{1,2}$(GG) molecules in the ligand-dependent state was only 34% at 1 mM ligand (Figure 82J). A plateau was reached at approximately 60% with 10 mM (Figure 82K-L). Plotting the fractions in dependence of the $Gdm^+$ concentration for each construct apparent affinities could be estimated (Figure 82M). The AG variant showed an apparent $K_D$ in the range between 100 µM and 1 mM. Between 1 mM and 10 mM Gdm were necessary to raise the amount of ligand-bound molecules above the 50% mark for the $59_{1,2}$(GG) construct. The apparent $K_D$ of $59_{1,2}$(GG) is higher than for the WT ($59_{1,2}$(AG)). More titration steps are needed for an exact $K_D$ determination.

## 10.7 *In vivo* experiments

*In vitro* smFRET experiments revealed structural changes within the riboswitch RNA that can suggest a regulation model on a molecular level. However, the RNA might behave differently in an *in vivo* context. Since the Gdn-II riboswitch under investigation is from *E. coli*, the growth behavior of this bacterium was analyzed in regard to the toxic molecule $Gdm^+$. LB medium with and without addition of 50 mM $Gdm^+$ were inoculated with an overnight culture of *E. coli* DH5α cells with a kanamycin resistence plasmid and the cells were grown at 37 °C. Figure 83 shows the growth of these bacteria measured through the optical density at 600 nm. Bacteria grown in standard LB medium exhibited an exponential growth for about 5 h before entering the saturation phase. In comparison, with $Gdm^+$ cell division seemed to be impaired for this time frame. Afterwards bacteria in $Gdm^+$ supplemented LB medium also revealed an exponential growth. The OD reached after 24 h was comparable within error to the other sample. This leads to the conclusion that $Gdm^+$ can extend the lag phase but does not impair growth after equilibration.

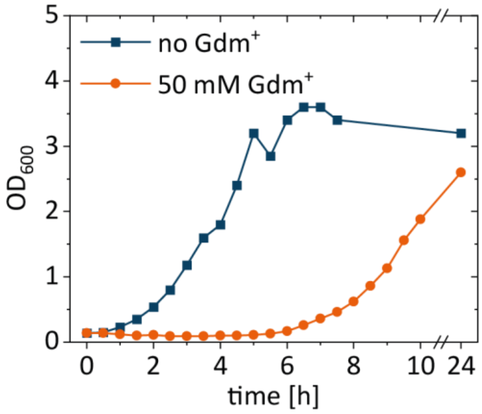


Figure 83: *E. coli* DH5α growth curve with (orange) and without (blue) $Gdm^+$. LB medium was inoculated from an overnight culture to an OD of 0.1 and grown at 37 °C and 180 rpm shaking. Samples for OD measurement were taken every 30 min and after 24 h.

After observing that $Gdm^+$ influenced the overall cell growth of the bacteria, this leads to the question how the Gdn-II riboswitch is affected by its ligand *in vivo*. A riboswitch construct was designed for the analysis (Table 10). At the 5’ end it contained the 93mer riboswitch sequence used for the smFRET experiments. The sequence is followed by ten more nucleotides from the WT *sugE* sequence. In total the first 9 amino acids of SugE were coded by the *in vivo* construct. The codons of a $His_6$ tag, followed by a stop codon were connected to the sequence. The 3’ end of the construct was a 24-2min aptamer. The total construct length was 204 nucleotides. The whole construct was ordered with the native *E. coli sugE* promoter in a pEX plasmid. Two more constructs were successfully cloned with the promoter exchanged to a T7A1 promoter[185] or a synthetic J23119+UP promoter[186]. In an additional control construct the 49 nt long aptamer region was deleted.

To investigate the effects of $Gdm^+$ on the riboswitch, cells were transfected either with the riboswitch plasmid, the control plasmid or used untransfected. LB medium with and without $Gdm^+$ was inoculated and cells were grown at 37 °C. After 1, 4 and 24 h the OD was measured, and samples were taken corresponding to the same number of bacteria. Subsequently the total RNA was extracted using phenol/ether extraction, and the transcriptiome was analyzed via denaturing PAGE (Figure 84). Without an additional plasmid it is apparent that the RNA content differs between the tested time points and dependend on whether $Gdm^+$ was preent in the medium. The latter could be either a direct effect of the toxic molecule, or a result of the different growth phases since $Gdm^+$ delayed initial growth. The RNA profile of bacteria with the control plasmid without the $Gdm^+$ binding aptamer had no detectable differences to the samples without plasmid. This was not the case for the Gdn-II riboswitch construct. With $Gdm^+$ supplied to the medium additional bands appeared on the gel (Figure 84). At four RNA lengths these differences were most pronounced, and those sites are marked with an arrow in Figure 84. The first site is a long RNA migrating between the SsrRNA (184 nt)[104] and the 208mer reference. The RNA at 2 appeared between the intense 5S RNA (120 nt)[187] and the tRNA bands (74 – 95 nt) and had a similar migration behavior as the 96mer on the gel. The RNA in bands 3 and 4 were smaller than the tRNAs. Sites 2 and 3 had the most intense bands of the novel $Gdm^+$-dependent bands. To a lesser extent they were also visible in the medium without additional $Gdm^+$. All $Gdm^+$-dependent RNA bands had reduced intensities after 24 h. There is a possibility for all constructs that changes in RNA pattern existed but cannot be detected either because of the small amount of RNA or because of overlap with endogenous RNA. With this experiment it was shown that addition of Gdm+ in presence of the riboswitch RNA influenced the transcriptome.

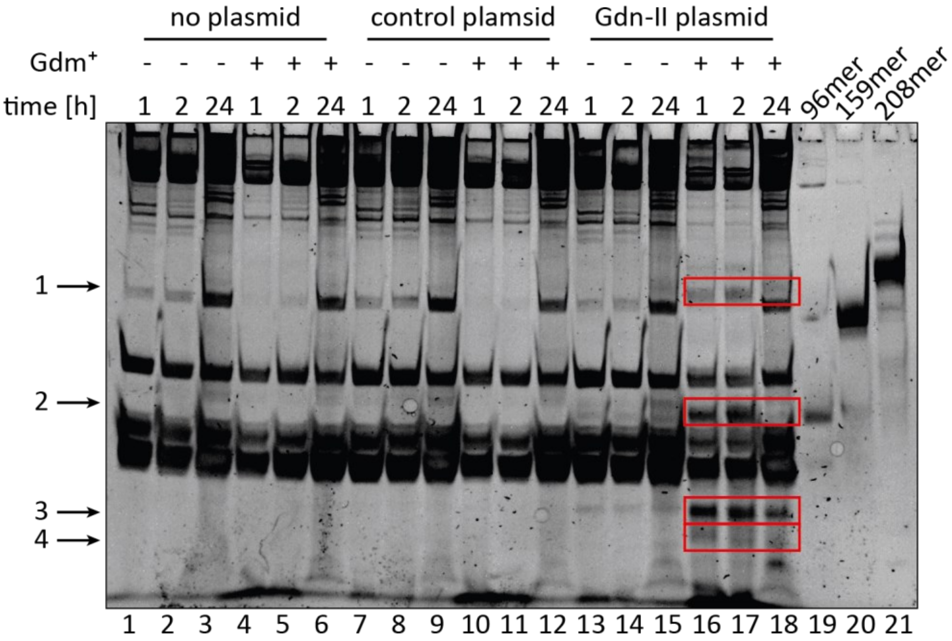


Figure 84: RNA band pattern analysis of *in vivo* expressed Gdn-II riboswitch construct under T7A1 promoter control. *E. coli* were grown in LB medium with or without 50 mM Gdm⁺. The cells carried either no plasmid, a control construct with a deleted aptamer or the Gdn-II riboswitch construct. Samples were taken at different time points and RNA amounts corresponding to the same ODs of bacterial culture were loaded onto a denaturing gel. RNA was visualized using GelRed.

After establishing the *in vivo* assay, different promoters were tested. The strong A1 promoter from the bacteriophage T7 (T7A1) and the strong synthetic promoter J23119 with an additional UP element showed similar results with respect to the band intensities at the four sites described above (Appendix Figure 92). The native *sugE* promoter from *E. coli* lead to reduced amounts of the investigated RNA. Only sites 2 and 3 were above the detection limit. For this reason, the J23119+UP was used for the following experiments.

Furthermore, the Gdm⁺ concentration necessary to trigger these RNA changes was investigated. Addition of 1 mM Gdm⁺ to the LB medium already reduced bacteria density after 4 h (Appendix Figure 93B). The RNA 1 was clearly visible at 10 mM or above (Appendix Figure 93A). It is possible that a reduced level was also present at 1 mM Gdm⁺. There was a shadow at a length of RNA 2 present in all gel lanes. RNA levels increased visibly at 10 mM and above. Small amounts for RNA 3 were also detectable in all samples. 50 mM of ligand were necessary for an intense band. RNA 4 could only be observed when 50 mM Gdm⁺ were supplied to the medium. Taken together the results from the Gdm⁺ concentration row, 50 mM delivered the most obvious changes and was thus used for further experiments.

The previous experiments showed that the lag phase of the bacterial growth was extended when Gdm⁺ was supplied directly at the inoculation. The resulting question was how Gdm⁺ influences bacteria growth and the RNA pattern after adding Gdm⁺ in the mid-exponential phase. For this, bacteria were grown for 3.5 h to the mid-exponential phase (OD 1.4) in absence of external Gdm$^+$ and were then split and diluted with the same volume of fresh LB medium. After dilution one sample had 50 mM Gdm$^+$ and the control had no additional Gdm$^+$. The cell growth in the exponential phase was similar in both samples for the 3 h following dilution (Figure 85D). Gdm$^+$ had no observable impact. Afterwards both cell cultures started transitioning into the stationary phase. Here, slightly higher ODs were measured in the sample without external Gdm$^+$. The time-resolved RNA levels are shown in Figure 85A&B for the medium with and without Gdm$^+$, respectively. In the sample taken 30 min after addition of Gdm$^+$ a band showed up on the gel with a higher mobility than the tRNA. Quantification of this band over the whole measurement period revealed that the RNA amount peaked during the first hour after Gdm$^+$ addition (Figure 85C). Afterwards the levels dropped about half and remained constant within error. They increased again after the stationary growth phase was reached. In the sample taken 24 h after inoculation the intensity of this band was reduced again.

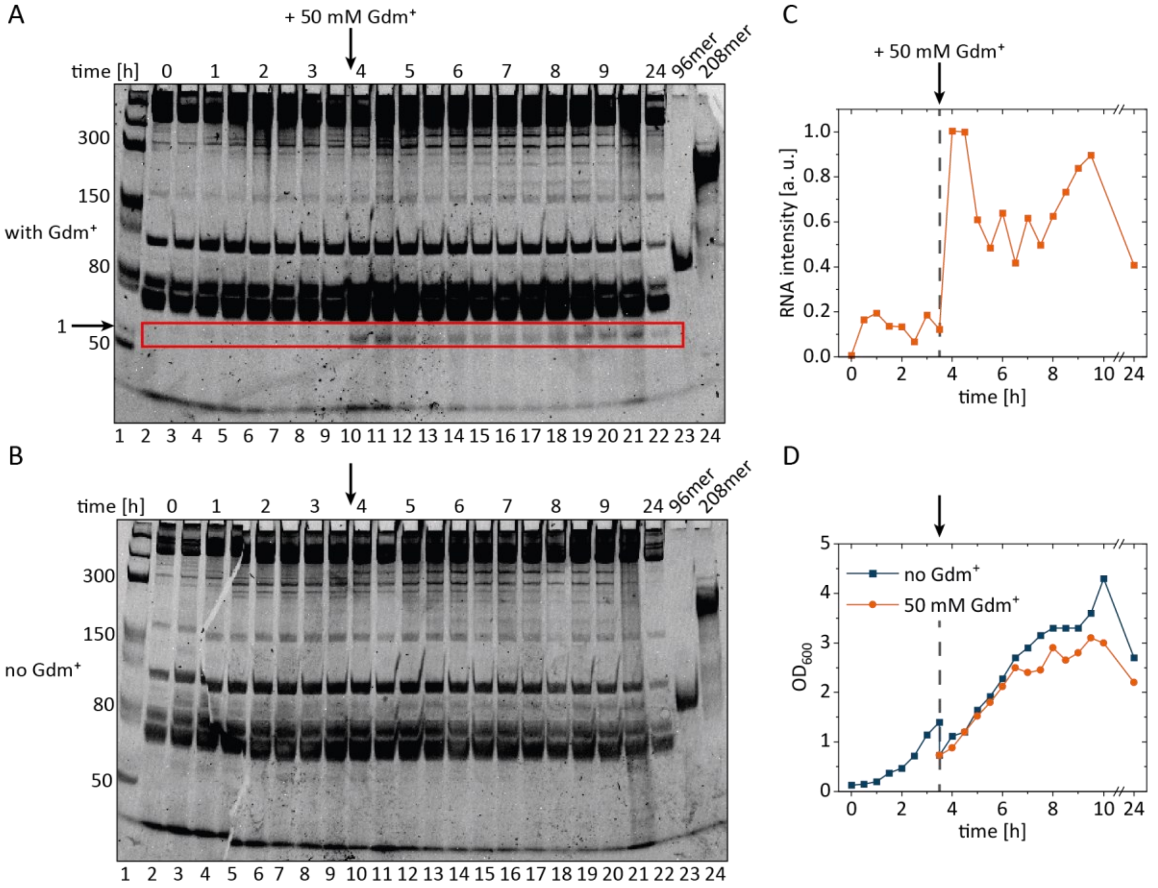


Figure 85: Time-resolved RNA pattern analysis of the *in vivo* expressed Gdn-II riboswitch construct under J23119+UP promoter control. The cells were grown in LB medium for 3.5 h and were then split and diluted into fresh LB medium with and without an end concentration of 50 mM Gdm$^+$ (indicated by an arrow). RNA from the same ODs of bacterial culture was extracted and loaded a denaturing gel and visualized using GelRed. A) PAGE analysis of bacteria diluted into Gdm$^+$ containing medium. B) PAGE analysis of bacteria diluted into LB medium without external Gdm$^+$. C) Time-dependent intensity of the RNA band marked in A with a red box. D) Bacterial growth curves.The previous

experiments were conducted in LB medium. Nelson *et al.* found that Gdm$^+$ can be produced endogenously in minimal medium[108]. To analyze the Gdn-II riboswitch construct in minimal medium the *in vivo* assay was transferred to M9 medium. As expected even without Gdm$^+$ the *E. coli* grew slower than in LB (Figure 86D). In presence of the toxic molecule, the culture did not reach the exponential phase during the experiment (26h). However, both samples had grown overnight. In both gels two of the bands seen in the previous experiments were detected that changed their intensity over time (Figure 86A-C). In contrast to all experiments with LB medium, in absence of external Gdm$^+$ the amount of both bands increased over time but dropped down to the starting point overnight. In presence of 50 mM Gdm$^+$ the RNA levels at both sites first raised, and peaked at 2 and 2.5 h, respectively. Afterwards the RNA levels decreased again. In the sample taken after 26 h the two RNA bands were clearly visible (Figure 86A, lane 22). When comparing the band intensities of the other RNAs in this sample, it cannot be excluded that more RNA was loaded onto this lane.

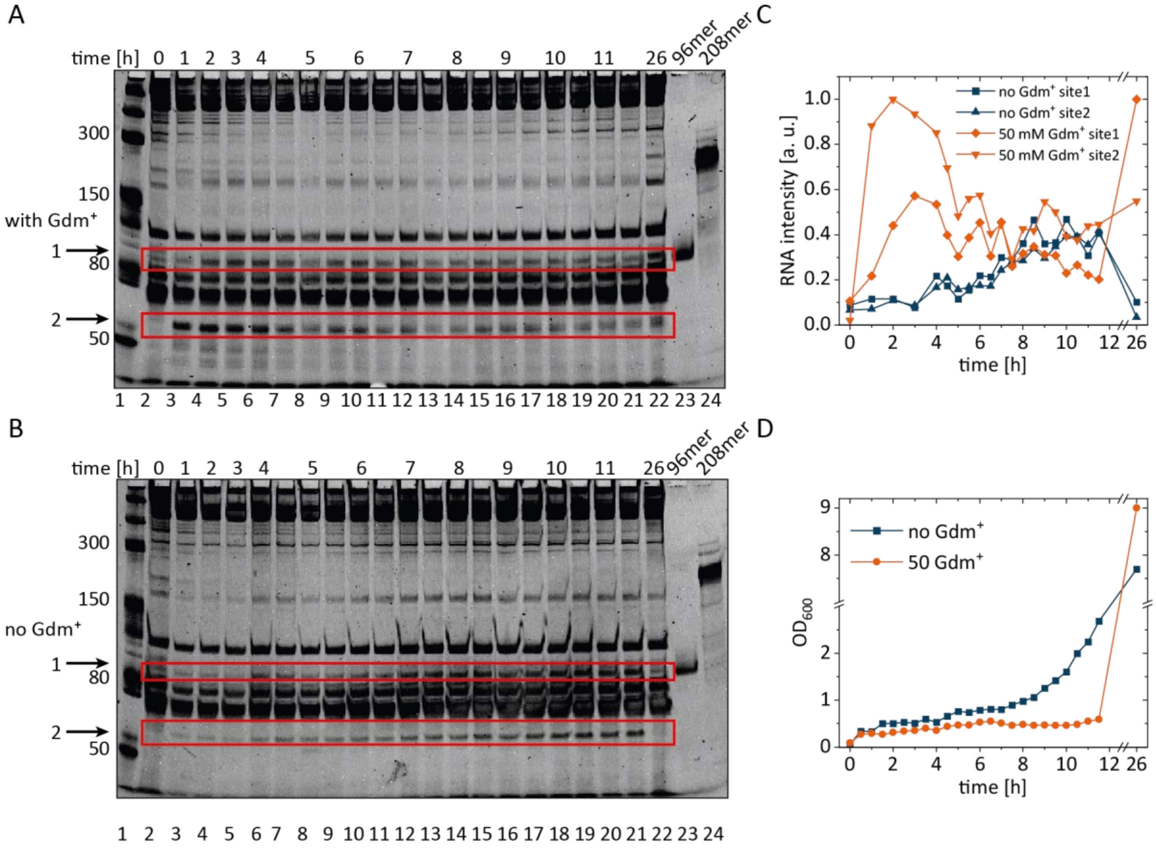


Figure 86: Time-resolved RNA pattern analysis of the *in vivo* expressed Gdn-II riboswitch construct under J23119+UP promoter control in M9 minimal medium. RNA from the same ODs of bacterial culture was extracted and loaded on a denaturing gel and visualized using GelRed. A) PAGE analysis of bacteria grown in Gdm$^+$ containing M9 medium. Sites with varying RNA levels were marked with an arrow and a red box. B) PAGE analysis of bacteria grown in Gdm$^+$ free M9 medium. C) Time-dependent intensity of the RNA bands marked in A and B. D) Bacterial growth curves.

The Gdn-II riboswitch has an ACGR motif in both loops. smFRET experiments suggested that the identity of the fourth loop nucleotide has effects on the structure as well as the affinity of the riboswitch. To test whether the loop nucleotides also influence the *in vivo* assay, mutations were introduced to the *in vivo* construct via site directed mutagenesis. All six variants were grown with and without Gdm$^+$ and whole RNA was extracted. Some differences in the RNA pattern were observed between the mutants (Figure 87). For instance, the two constructs with a G in P1 (GA and GG) had elevated RNA levels at site 4 in absence of Gdm$^+$ while in presence of Gdm$^+$ this band was as faint as in the other variants. In the AU version the Gdm$^+$ dependent RNAs were generally more intense than in the other mutants. The smallest response to Gdm$^+$ in site 3 and 4 was found for the AC construct. The two constructs with a pyrimidine moiety in P2 (AU and AC) had an additional band when grown in Gdm$^+$ supplemented medium which migrated between RNA 3 and the *E. coli* tRNA (lanes 10 & 12, RNA 5). The experiment showed that the loop nucleotides also affect the behavior of the riboswitch RNA *in vivo*.

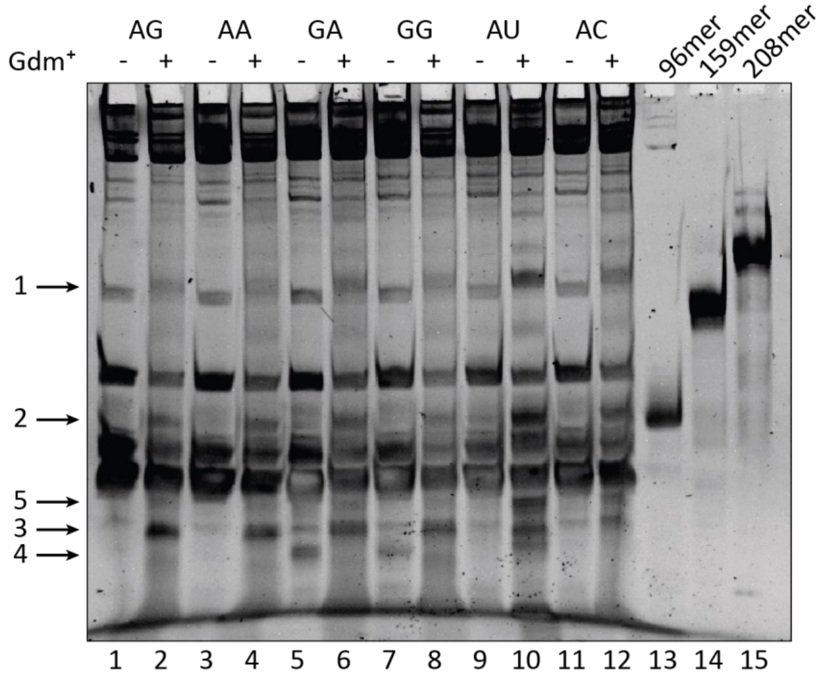


Figure 87: RNA pattern analysis of *in vivo* expressed Gdn-II riboswitch constructs with mutations in the fourth loop nucleotids under J23119+UP promoter control. *E. coli* were grown in LB medium with or without 50 mM Gdm$^+$. Samples were taken after 1 h and RNA amounts corresponding to the same ODs of bacterial culture were loaded onto a denaturing gel. RNA was visualized using GelRed. RNA bands that showed differences were indicated with an arrow.

The analysis via denaturing gel can only detect differences in RNA pattern, and only if the RNAs under investigation do not overlap with other endogenous RNAs. Experiments with a control plasmid or without a plasmid strongly suggest that the RNA fragments seen on the gel belonged to the riboswitch construct under constitutive promoter control (Figure 84). The RNAs with different length could be either premature transcription termination or specific degradation products. The accumulation of these RNAs in presence of $Gdm^+$ could be caused by an increased production under these conditions or an increase in RNA stability. To test whether the RNA fragments indeed are part of the riboswitch construct three antisense probes were designed which target a sequence involving either the TSS, the SD sequence or the 24-2min aptamer (Appendix Figure 94). Those probes were annealed to the whole RNA extract and analyzed via EMSA. In the sample from cells grown in presence of $Gdm^+$ five bands with varying intensities were detectable using the TSS probe in the EMSA (Figure 88, lane 9). Since all of the monomeric probe was bound it is possible that other bands were not detected due to a lack of probe. Detecting five bands in the EMSA does not necessarily mean that five different RNA length were present. Different folds of the same RNA could increase the number of bands. Most of these bands were also found in the sample without external $Gdm^+$ (lane 8) but were neglectably weak. The use of the probe targeting the sequence around the SD sequence visualized only a weak band (lanes 11 & 12). Since the same probe bound to two different reference RNAs (lane 5) in the EMSA, the weak bands are likely not due to a dysfunctional antisense probe. This means that either the fold in the *in vivo* extract impaired binding, or - more likely - that the whole target sequence was not present at elevated levels. The probe designed to detect the 24-2min aptamer stained two bands on the EMSA (lanes 14 & 15). The smaller RNA had an increased intensity when the cells were grown in LB with $Gdm^+$. These data showed that the transcription of the *in vivo* Gdn-II riboswitch construct had at least partially occurred to an RNA length that was capable of binding the 24-2min probe. It is possible that the RNA was then either processed or degraded to the RNA fragments seen in the gels. The fact that defined RNA bands could be detected renders the possibility that degradation occurred unspecifically highly unlikely. To asses whether these processes require additional factors or the ionic environment found in bacteria, the *in vivo* construct was T7 transcribed, and the resulting RNA was incubated at 37 °C with and without $Mg^{2+}$ and $Gdm^+$ as well as combinations thereof. In all cases degradation occurred as evidenced by a series of smaller RNA bands (Appendix Figure 95). The strongest degradation was observed for the sample in presence of both $Mg^{2+}$ and $Gdm^+$, and the least amount of degradation was found when both ions were absent. Nevertheless, no specific RNA band accumulated in either combination that would represent the product of a site-specific degradation reaction *in vitro*.

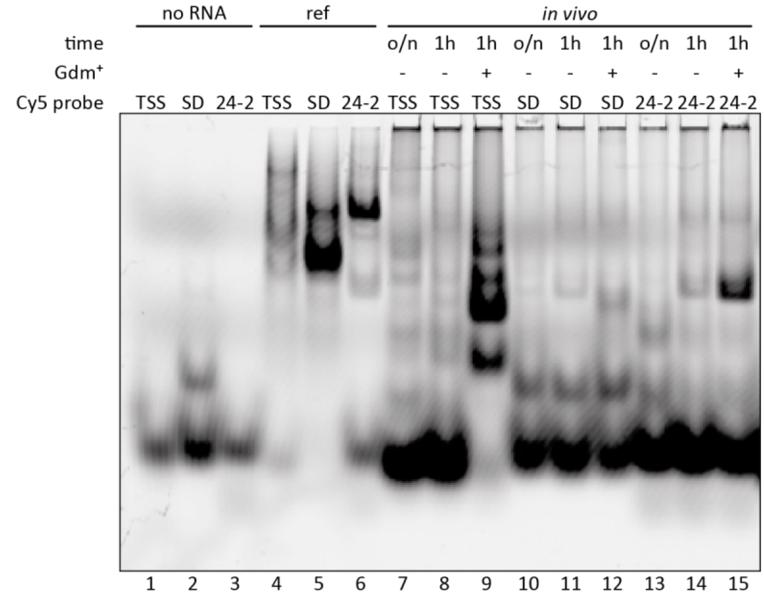


Figure 88: EMSA analysis with Cy5-labeled probes targeting either a region at the TSS, the SD sequence or in the 24-2min aptamer. *E. coli* were grown in LB medium with (+) or without (-) 50 mM $Gdm^+$. Samples were taken before incubation (o/n) and after 1 h and RNA amounts corresponding to the same ODs were used for the EMSA analysis.

# 11 Discussion

The Gdn-II riboswitch is an important regulator of the detoxification of Gdm$^+$. So far, only parts of the riboswitch were investigated, such as the aptamer domain, isolated hairpins or even just the loop region, but the expression platform was never the target of analyis. The only structural data were derived from the isolated hairpins that are not sufficient to explain the structure of switching competent constructs. Reiss and Strobel proposed a regulatory mechanism as a translational riboswitch [161] but did not deliver evidence supporting their model. In this work, using smFRET as well as *in vivo* assays, the conformational dynamics of the aptamer domain and the full-length riboswitch were investigated and more insights into the possible regulation were gained.

## 11.1 Preparation of labeled RNA

The study of conformational dynamics using smFRET requires a construct with a FRET donor and a FRET acceptor. To achieve this, the RNA sequence was split into fragments which could be labeled individually and set together in a building block principle in a splinted ligation. This allows the possibility to easily introduce mutations or different labeling sites by switching of one fragment. Splinted ligation is a standard procedure to generate site-specifically labeled FRET constructs[135,188]. The fluorophore labeling of the individual fragments was almost quantitative as evidenced by the lack of a 260 nm RNA only peak during analytical RP-LC. The challenge of this riboswitch was to ligate a total of six fragments for the full-length 93mer construct at five ligation sites. Two different ligases were tested. First the T4 DNA ligase and second the T4 RNA ligase 2. The T4 DNA ligase proved successful to ligate the $47_{1,3}$ construct from two fragments during my Master's thesis. However, for the splinted ligation of the 6-fragments 93mers it was insufficient. Figure 67 shows that after T4 DNA ligase incubation (lanes 10-13) three bands were found in the Cy3 and the Cy5 scan each. In both scans the lowest band is the monomeric, and unligated labeled fragments 408_Cy3 and 418_Cy5, respectively. As expected, they migrate on the same height as before the ligation. The second strongest band in the Cy5 scan had the same migration behavior as the 34 nt ligation with the 2 3' fragments (lane 5). This band is most likely exactly this ligation product of 418_Cy5 and 420. Other combinations of fragments with an included Cy5 would be either too small (414/416 & 418_Cy5, 25 nt) or too big (412, 414/416 & 418_Cy5, 40 nt). The other, weaker band of the Cy5 scan was migrating comparable to the 3 fragment 414/416, 418_Cy5 and 420 ligation (lane 4) of 44 nt. Since the difference of 4 nt might not be resolved in the gel, it is also possible, but less likely that the band originates from the 40 nt long 412, 414/416 and 418_Cy5 reaction. In the Cy3 scan one band also migrated similarly to the 34 nt reference. Since this is found in the Cy3 channel and the construct had the Cy3 labeling site in the 5' fragment, this band has to originate from the 408_Cy3 and 410/411 ligation. The other band in Cy3 was found between the 44 nt (lane 4) and 59 nt (lane 3) references. This is in agreement with the first three 5' fragments with a length of 49 nt. From the experiment it seemed that ligation with T4 DNA ligase could ligate the first three or last three fragments. At the same time, T4 DNA ligase was unable to ligate the nick between the fragments 412 and 414/416. This could have had structural reasons on both, the splint DNA or RNA site, a bad annealing of the 414/416 fragments which are the shortest with 10 nt or problems with one of the fragments such as partial degradation. The next approach was the use of T4 RNA ligase 2. This enzyme had the disadvantages that it might have some nuclease activity for unpaired RNA and could also ligate fragments with a free 5' phosphate that are not annealed to the splint DNA. The first problem was avoided by using a DNA splint that was longer than the 93mer sequence, thus leaving no parts of the RNA unpaired. The latter issue was observed when the DNA splint was degraded which resulted in several higher bands (data not shown). In this case the fragments could be in a random order and cannot be used for studies of the riboswitch. This problem was overcome by including a heat inactivation step for the ligase before DNase treatment. To prove that T4 RNA ligase 2 was able to ligate all six fragments a ligation ladder was created with only part of the fragments. Adding of fragments resulted in stepwise increase in RNA length (Figure 67, lanes 1-5). It can be concluded that the highest bands for the six fragment ligation were indeed the full-length 93mer RNAs. Despite the presence of several bands corresponding to only partially ligations, it was not necessary to optimize the preparation of the FRET labeled RNA any further as smFRET only needs small amounts of RNA. In summary, 24 different Gdn-II riboswitch FRET constructs were successfully prepared with this method.

## 11.2 Combining smFRET and coarse-grained simulation for aptamer analysis

Using smFRET, the conformational dynamics of the aptamer, the full-length riboswitch including the expression platform and the influence of the loop nucleotides could be analyzed. For the analysis of the aptamer conformations two different construct length were chosen. The first, $47_{1,3}$, includes only the two hairpins P1 and P2 and the linker between the hairpins. The other construct is a 59mer construct starting at the TSS and ending one nucleotide after P2. The labeling site 1 is the same in both constructs whereas the third labeling site differs by one nucleotide.

The shorter construct was analyzed by coarse-grained simulation by Sebastian Falkner from the Schwierz-Neumann group and in my work using smFRET experiments on its $Mg^{2+}$ dependence in absence of ligand. The coarse-grained simulation revealed three different structures dependent on the $Mg^{2+}$ concentration (Figure 69B). Calculating the theoretical FRET efficiencies from the distance information three populations were found. A low FRET state in absence of $Mg^{2+}$, an intermediate FRET state between 0.5 and 5 mM $Mg^{2+}$ and a high FRET state at 10 mM. smFRET experiments also suggested a three-state model (Figure 69A). The low FRET state was nearly identical between simulation and experiment with $E_{FRET}$ = 0.40 and 0.41, respectively. The intermediate FRET varied between FRET values of 0.57 and 0.67 for the smFRET or 0.65 and 0.71 for the simulation. Since the uncertainty in FRET is that high even with the same method it is possible to assume that the found FRET efficiencies are similar within error. Also, the high FRET efficiencies found at elevated $Mg^{2+}$ (0.81 and 0.83) are in good agreement, with the exception that the high FRET state was already populated at 5 mM $Mg^{2+}$ in the experiment. Another difference was that the shift to the high FRET state was quantitative for the simulations while the smFRET experiments suggested an equilibrium between intermediate and high FRET conformations. This likely is an artifact of the force field used in the simulations where a single state might get overrepresented. Altogether the simulation-based FRET efficiencies were comparable to the experimental data, both in the peak center and in the tendencies to populate certain states in dependence of increasing $Mg^{2+}$ concentration.

Crystallography approaches were able to obtain kissing loop hairpins at high salt concentration [169]. Furthermore, MD simulations showed that the loop region can switch between an unbound and a bound-like structure in absence of $Gdm^{+}$ [172]. Sebastian Falkners simulations at intermediate $Mg^{2+}$ levels and in absence of ligand found kissing loop structures which produced the calculated intermediate FRET efficiencies (Figure 69B-D). This confirms that the ligand is not a necessary discriminator for kissing loop formation. According to the switching model by Reiss and Strobel [170], the ligand triggers the folding of the riboswitch into a kissing loop conformation. Previous data from my Master thesis revealed that the ligand also stabilized an intermediate FRET state ($E_{FRET}$ = 0.69). With the knowledge that the kissing loop structure is accessible in absence of ligand, the good agreement of coarse-grained and smFRET data in absence of ligand and the FRET efficiency with ligand, it could be concluded that the intermediate FRET state found for $47_{1,3}$ might also represent the kissing loop conformation. This shows how combining coarse-grained simulation and smFRET experiments could correlate FRET efficiencies with structural data and was further used to characterize the other FRET states.

In absence of ions smFRET experiments revealed a FRET state at 0.41 (Figure 69A). A low FRET state means that the two fluorophores have a large distance. The Reiss and Strobel model proposes an open conformation in absence of ligand with the two hairpins folded, but no interaction between them. Since there are no ions to shield the negatively charged phosphate backbone the linker would be extended due to electrostatic repulsion. Such a structure was found in the coarse-grained simulation (Figure 69C). The hairpins point in different directions, which equals a large distance between the FRET labeling sites. These structures have the same calculated FRET efficiency as the experimental data with $E_{FRET} = 0.4$. Thus, the low FRET state could be an open conformation as suggested by the model.

The model only included a two state system, but both, smFRET and coarse-grained simulation, detected a third, high FRET, state (Figure 69A & B). According to the simulations this high FRET state derived from structures that had a folded P1 hairpin but the sequence of P2 folded back to the linker with eight base pairs and a single nucleotide bulge (Figure 69C & D). The 3’ end where Cy5 was attached would be close to the P1 helix carrying the Cy3 modification. The high FRET efficiency measured is a representation of the proximity. Such a high $Mg^{2+}$, high FRET conformation was not proposed previously but can be found using online secondary structure prediction tools[189,190] (Appendix Figure 96). Previous in-line probing experiments of the *E. coli* Gdn-II riboswitch conducted at 20 mM $Mg^{2+}$ detected only minor cleavage of the P2 loop even in absence of ligand in comparison to representatives from other species. That is not compativle with an open conformation under these conditions. If the loop nucleotides were unpaired, they would likely show higher cleavage. In the high FRET structure the loop nucleotides are paired with nucleotides from the linker. Furthermore, the linker region experienced more cleavage in in-line probing with increasing ligand and increased kissing loop formation. This would fit to an opening of base pairing during the transition from the high FRET structure to a kissing loop conformation. In total, the in-line probing experiments could be explained by an equilibrium between a low FRET, open conformation and a high FRET conformation where P2 binds back to the linker. Increasing $Gdm^+$ would shift the equilibrium towards the kissing loop conformation. From the data it is not clear whether this high FRET structure is found *in vivo* or has any physiological relevance in the regulation mechanism of Gdn-II. Additionally, this might be a specific feature of the *E. coli* sequence and might not be transferred to other species.

In summary the experiments with $47_{1,3}$ revealed three states which could be characterized by coarse-grained simulations including a novel conformation.

## 11.3 59mer aptamer conformations

The results of the short 47mer aptamer domain were then transferred to the extended 59mer aptamer domain. Furthermore, different labeling schemes were supposed to create better three-dimensional insights. The 59mer is extend by 11 nt on the 5’ side and by one nt 3’ of P2. Surprisingly, the $59_{1,3}$ did not behave like the $47_{1,3}$ RNA with the same labeling scheme in the smFRET experiments. The $47_{1,3}$ high FRET state was not populated by $59_{1,3}$ even at 10 mM $Mg^{2+}$. Instead, the intermediate FRET efficiency of 0.7 was detected under these conditions, similar to the FRET efficiency in presence of $Gdm^+$ or the intermediate $Mg^{2+}$ state of $47_{1,3}$. In some way the extension prevented the high FRET conformation at the tested $Mg^{2+}$ concentration. Since P1 was always folded in all coarse-grained simulation structures it is very likely that the 3′ extension by one nt is the discriminator between the high FRET and the intermediate FRET conformation. It is possible that the additional nt stabilized the P2 stem that only consists of four base pairs. A stabilization of P2 would prevent an unfolding which is required to form the new base pairs of the $47_{1,3}$ high FRET state. The stabilization in the 59mer either might be sufficient to completely prohibit the population of the high FRET conformation. It is also possible that even higher levels than 10 mM $Mg^{2+}$ would be necessary to overcome the stabilization and would result in the formation of the high FRET structure. The sequence used for in-line probing had a 3′ extension[156]. The results at 20 mM $Mg^{2+}$ might be explained by a fraction of the molecules adopting the high FRET conformation in absence of ligand as discussed above. The findings with the 59mer make it even less likely that the high FRET structure is of physiological relevance *in vivo*.

While the $59_{1,3}$ experiments could not confirm the high FRET conformation, the data were in agreement with a low FRET state in absence of either $Mg^{2+}$ or $Gdm^+$ and an intermediate FRET state in presence of $Mg^{2+}$ with $E_{FRET} = 0.70$ that might represent the kissing loop conformation and is also found with the ligand. Changing the labeling scheme to one fluorophore in the linker and one at the 3′ end ($59_{2,3}$) also suggested a two state system with a high FRET peak ($E_{FRET} = 0.81$) in absence of ions and a FRET peak around 0.7 in presence of either $Mg^{2+}$ or with $Gdm^+$. For this construct the fluorophore attachment sites were in close proximity when P2 is formed in an open conformation. This is well represented by the high FRET peak in the experimental smFRET data. Kissing loop formation would bend the linker. It is possible that the bending and torsion would separate the fluorophores and lead to the drop in FRET efficiency measured with $Mg^{2+}$and $Gdm^+$. The tail in the histogram for the 10 mM $Mg^{2+}$ might indicate some dynamics.

The two state system cannot be confirmed using the data from the third labeling scheme ($59_{1,2}$). In this case there were differences between the $Mg^{2+}$ and the ligand condition. In absence of both analyzed ions the FRET efficiency was 0.60. $Mg^{2+}$ addition reproducibly increased the efficiency by only 0.06. Based on this construct alone this small increase might be within error, but the other constructs had more drastic FRET changes. From this it can be concluded that the two FRET efficiencies were a result of two different conformations that only coincidentally had a similar distance between the fluorophores with the $59_{1,2}$ labeling scheme resulting in the similar FRET efficiency. In contrast to the other labeling schemes $Gdm^+$ triggered a FRET change in comparison to the $Mg^{2+}$ condition. Here, high FRET efficiencies around 0.9 were found. This means that the fluorophores are very close together. In a more structural context assuming a folded P1, the linker might somehow be pulled towards the P1 helix axis in presence of $Gdm^+$.

The FRET technique describes a distance vector between site-specific introduced fluorophores. It is possible that fluorophores coincidentally have the same distance despite originating from different structures which would result in the same FRET efficiencies. This was the case twice for the Gdn-II aptamer FRET constructs. First, the $Mg^{2+}$ and $Gdm^+$ conditions were the same for the $59_{1,3}$ and $59_{2,3}$ and then the ion-free an $Mg^{2+}$ were similar in $59_{1,2}$. Only the use of another labeling schemes was able to tell the conformations apart. This showcases the importance of using multiple labeling schemes. A similar effect was observed by Manz *et al.* for the SAM-I riboswitch which allowed the establishing of a 5-state regulation model [142].

The proposed two-state regulation model by Reiss and Strobel with an open and a kissing loop conformation was unable to explain the three-state FRET system for the aptamer domain found in this work or the three-states detected in NMR studies [168]. For this reason I would like to expand and propose a new three-state model for the aptamer only Gdn-II riboswitch (Figure 89). Subsequently, I will also characterize the proposed conformations in more detail.

With the additional nucleotide 3’ of P2 the hairpin is stabilized. The P1 hairpin is always folded in all conformations of this model. In absence of the $Gdm^+$ ligand and with a reduced ionic strength the RNA is in an extended, open conformation. Addition of ions such as $Mg^{2+}$ leads to the folding of a loose kissing loop interaction. The loop CG nucleotides form Watson-Crick base pairs with the counterparts from the other hairpin. In the loose kissing loop conformation, the linker in which the anti-SD sequence is located and which is between 7 – 40 nt long, is free. Binding of $Gdm^+$ to the loose kissing loop structure stabilizes the kissing loop (“stabilized kissing loop interaction”). The stabilization could be some kind of interaction between the linker with the helices of the hairpin such as an insertion into the major grove as described in the Gdn-III structure [176]. It is more likely that nucleotides from the linker interact with the fourth loop nucleotides which are twisted outwards in the crystal structures. In either way the linker would be tethered towards the helix axis resulting in a sequestering of the anti-SD sequence and consequently would allow translation initiation.

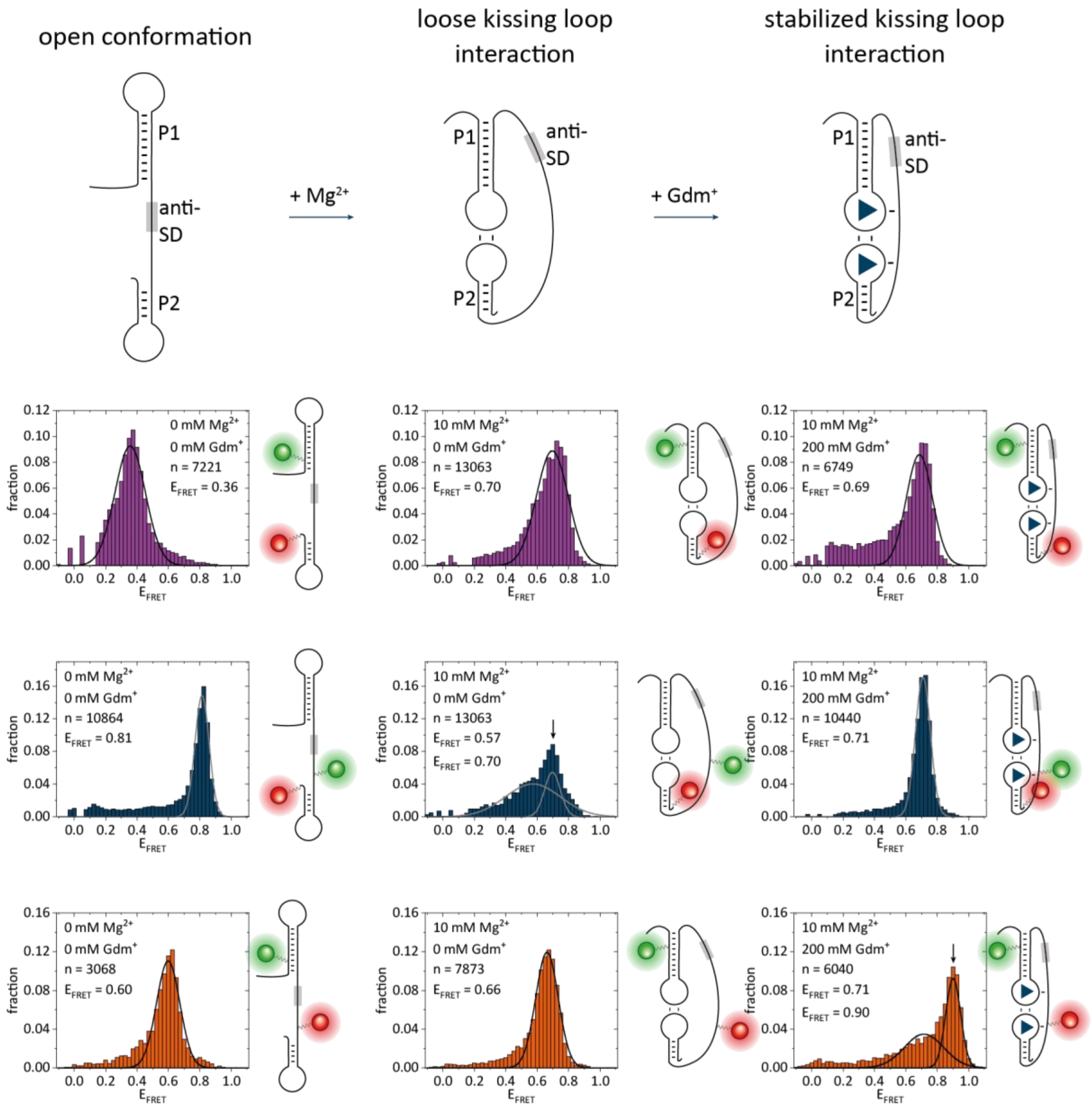


Figure 89: 3-state switching model for the Gdn-II aptamer domain.

The first conformation of the model is the open conformation which occurs in absence of $Mg^{2+}$ or ligand. The two hairpins are formed but do not interact with each other. With the lack of positively charged counter-ions to shield the phosphate backbone, the linker would be extended due to electrostatic repulsion. Such a structure was found in the coarse-grained simulations under these conditions (Figure 69C & D). In-line probing data of the Gdn-II RNA from *G. violaceus* showed that in absence of ligand the two hairpins are folded and only the loop nucleotides and the linker undergo cleavage[156]. The smFRET data of the aptamer support an open structure (Figure 89). Labeling at the different hairpins ($59_{1,3}$) revealed a low FRET efficiency translating into a large distance between the fluorophores. The large distance fits to an extended linker. Placing the second label in the linker and closer to P1 ($59_{1,2}$) increased the FRET efficiency, in agreement with the model of an extended structure. The highest FRET efficiency was detected using the $59_{2,3}$

construct. The second attachment site is in the linker but closer to P2 than to P1. The high FRET efficiency suggests that P2 is folded, thus bringing the two fluorophores in close proximity.

The second conformation of the model is the loose kissing loop conformation. I propose a structure that has intramolecular Watson-Crick base pairing with the CG nucleotides of each loop. The linker would not be part of the stabilization and might be flexible depending on its length. While the kissing loop conformation would not be accessible in the Reiss and Strobel model in absence of ligand, there are several indications including the smFRET data in this work that the aptamer domain can pre-form such a structure at high salt conditions. The FRET results of $59_{1,3}$ show a similar FRET efficiency in presence of $Mg^{2+}$ as found for the condition with $Gdm^+$ where a kissing loop should be present (Figure 89). This means that the fluorophore distances are similar as well. This labeling scheme has the fluorophores in or close to the hairpins which is an indicator that the hairpins might be in the right orientation for the kissing loop. For one of the peaks of $59_{2,3}$ with $Mg^{2+}$ the FRET efficiency is also similar as for the ligand condition. The FRET efficiency is lower than in the open conformation. This could be rationalized by bending of the linker and torsion effects which could increase the inter-dye distance. In the histogram of $59_{2,3}$ at 10 mM $Mg^{2+}$ a second peak or tail towards intermediate FRET efficiencies is visible. This could represent some dynamics that were not observed for the other construct. One of the labeling sites is located in the linker even though it is close to P2. These findings fit to the loose kissing loop conformation with flexibility of the linker but rather static conformations between the hairpins. At 10 mM $Mg^{2+}$, the $59_{1,2}$ construct did not show the same high FRET peak as in presence of $Gdm^+$. Thus the structures must be different between both conditions, for example represented by a loose and a stabilized kissing loop conformation. With $Mg^{2+}$ the FRET efficiency is only increased by 0.06 compared to the open conformation. This could be rationalized by a rotation of P1, which is labeled in the lower part of the helix towards the linker as expected for the loose kissing loop conformation. Whether the riboswitch is competent to form the loose kissing loop conformation is also dependent on the linker sequence, as mutations like in $59_{1,3}$ (aSD) and $59_{2,3}$ (aSD) show. Overall, the smFRET data of the three constructs would fit nicely to the proposed loose kissing loop structure. Additional methodological approaches also reason for the existence of a kissing loop conformation despite the absence of ligand. Using coarse-grained simulations with different $Mg^{2+}$ concentrations ranging from 0.5 to 5 mM show kissing loop structures (Figure 69C). Furthermore, crystallography of the isolated hairpins showed that kissing loop structured crystals grew also in high salt buffer without $Gdm^+$ [169]. More evidence that the kissing loop can form in absence of ligand derived from MD simulations of the loop region. Here, the simulated molecules were able to switch between a collapsed/unbound to a bound like structure[172].

So far, the mechanistic order of ligand binding and kissing loop formation is unclear. In theory, two different ways for ligand binding to biomolecules are discriminated: the conformational selection, where a ligand-bound-like conformation is pre-folded and ligand binding stabilizes this conformation or the induced fit, where binding triggers the folding to the ligand-bound state. The conformational selection requires the accessibility of the bound-like state in absence of ligand. The data of the loose kissing loop conformation suggest that this prerequisite is given for Gdn-II. However, the results were not able to forecast that the loose kissing loop conformation is binding competent at all or whether binding might also happen at the stage of the open conformation.

The last conformation of my proposed model is the stabilized kissing loop. In this conformation the two kissing loop CG base pairs are formed as for the loose kissing loop. The difference is the linker is tethered to the helix axis formed by P1 and P2 and stabilizes the whole structure. One possibility is the formation of basepairing between the two fourth loop nucleotides and the linker. I will focus on evidence for this possibility despite other structures like an insertion into the major major grove like for the Gdn-III riboswitch cannot be excluded. The smFRET data showed that $59_{1,2}$ folds into a high FRET state in presence of $Gdm^+$ which is not the case for the loose kissing loop conformation in $Mg^{2+}$ buffer (Figure 89). The fluorophores come closer together which means the linker comes into closer proximity to P1. Also, for all 59mer FRET constructs, but especially for $59_{2,3}$, the peak width decreases with added ligand which argues for a more stable conformation than the loose kissing loop. In coarse-grained simulations in absence of ligand some simulations showed short-lived interactions between the fourth loop nucleotides and two uridines in the linker, close to P2 (Nadine Schwierz-Neumann, personal communication). It is possible that the presence of the ligand increases the longevity of these interactions and thus would stabilize the stabilized kissing loop interaction. Not all, but the majority of representatives of the Gdn-II class have a UU, UC or UUC motif in the linker[166] which would be capable of Watson-Crick basepairing with the fourth loop purines. The smFRET results with $59_{1,2}$ (A/U) which also had the high FRET state in presence of $Gdm^+$ suggest that also non-canonical interactions with the linker would be possible. In in-line probing experiments with Gdn-II from *G. violaceus* a reduced cleavage was detected for two uridines in the linker close to P2 with increasing $Gdm^+$ concentrations[156]. Decreased cleavage efficiency for in-line probing is a sign of basepairing of those nucleotides. This is in agreement with the proposed structure of the stabilized kissing loop. The loop nucleotides-linker interactions might also be the reason for the conservation of purines at the fourth loop position and might be a factor for changed affinities in the loop mutants detected by NMR[168].

In summary, several sets of data from this work and the literature argue for the three-state model with an open conformation, a loose kissing loop and a stabilized kissing loop for the aptamer domain. In a regulatory context for the translational riboswitch a sequestering of the anti-SD sequence would switch ON translation. It is very likely that in the stabilized kissing loop the linker is tightly bound and the anti-SD sequence would be inaccessible for the small subunit of the ribosome. In the loose kissing loop conformation the linker would be less tightly bound. It is not clear if the flexibility would suffice for ribosome binding *in vivo* or if that structure needs to be opened before binding.

## 11.4 Expression platform

For the function of a riboswitch it is not only necessary to recognize the ligand in the aptamer followed by a conformational change. For the regulation it is crucial that this information is passed to the expression platform. In case of translational riboswitches, including proposedly the Gdn-II riboswitch, this happens through accessibility of the SD sequence. So far, the Gdn-II riboswitch was not investigated apart from the aptamer region. To get better insights into the regulation mechanism two construct lengths were designed that include the SD sequence. It would be expected that in absence of the ligand the SD sequence would bind back to the anti-SD sequence in the linker (Figure 90). The P2 hairpin could either be folded as depicted by Reiss and Strobel or opens during the anti-SD:SD interaction. Ligand binding would then disrupt the anti-SD:SD interaction and the stabilized kissing loop would form. Under conditions with 10 mM $Mg^{2+}$ and 200 mM $Gdm^{+}$ all construct lengths, all labeling schemes and all mutations showed FRET efficiencies that were comparable with the results of the 59mer with the respective labeling scheme. Based on that it can be concluded that the ligand successfully switches the riboswitch and the stabilized kissing loop conformation might always be the most stable conformation under these conditions. This was expected of a functioning translational riboswitch and will not be discussed further.

On the other side, the FRET efficiencies in absence of ligand were not that comparable between the 59mer and the longer constructs. Without $Gdm^{+}$ and $Mg^{2+}$, the FRET efficiencies of the constructs including the third labeling site were decreased. There were also different peaks detected between the 74mer and the 93mers which means that the additional nucleotides including the start codon also had an effect on the structure of the riboswitch. The distances for these constructs were greater for the 93mers than the 74mers with the third labeling site. Interestingly, a small percentage of $74_{2,3}$ molecules populated the high FRET state found for the 59mer (Figure 74A). For the $74_{1,3}$ construct where the differences in FRET efficiencies between $59_{1,3}$ and $74_{1,3}$ are not so large, this small fraction would not be detectable due to peak overlap. This fraction of high FRET state for $74_{2,3}$ could either indicate that those molecules adopt the open conformation or only in these structures the P2 is folded but the rest of the sequence remains in the 74mer conformation.

The constructs with the first and second labeling sites displayed a FRET efficiency similar to the 59mer aptamer construct (compare Figure 72A & Figure 75A & E). This might be a coincidentally similar distance in the two conformations, or might be a sign that the differences between the aptamer and the longer constructs are mostly located in the region 3’ of the second labeling site.

For the $74_{1,3}$ and $93_{1,3}$ constructs addition of $Mg^{2+}$ resulted in an increase in FRET efficiency but the differences between the two construct lengths remained (Figure 73B & F). This was not the case for the constructs labeled at the sites 2 and 3 which had FRET efficiencies similar to those in absence of $Mg^{2+}$ (Figure 74). The $Mg^{2+}$ induced refolding could either lead to a similar distance between the fluorophores or the structural rearrangements might be located closer to the second labeling site. In total, none of the smFRET data for the SD sequence including constructs showed any signs of the $Mg^{2+}$ induced loose kissing loop conformation witnessed for the aptamer only RNA.

From these data it is not possible to assess whether any of these constructs show an anti-SD:SD paired conformation with or without $Mg^{2+}$. For a better estimation of this question the SD sequence and the assumed anti-SD sequence in the linker were mutated at two nucleotides. Ideally it would be expected that mutation of one site would result in the open conformation or loose kissing loop conformations of the 59mer and double mutations would restore the 74mer conformation. This was however not the case. Mutation of the anti-SD sequence destabilized the new 74mer conformation as evidenced by an increase fraction of the $59_{2,3}$ specific high FRET population in $74_{2,3}$ (Figure 79A). However, the mutation was not able to completely disrupt the 74mer conformation. There are three possible reasons for this: First, the assumed anti-SD sequence was wrong and another sequence basepairs with the SD sequence. Another possibility is that the sequence was right but the two point mutations were not sufficient to destabilize the anti-SD:SD interaction. The last possibility is that in the chosen constructs under *in vitro* conditions the RNA folds into structures that do not involve the anti-SD sequence. The fact that the third labeling site is located 5’ of the SD sequence and the 74mer and 93mer RNAS behaved differently in the smFRET experiments speaks for more complex structures.

On the contrary, mutating the SD sequence, either alone or in combination with the anti-SD sequence, had a huge impact on the structures. It did not lead to the 59mer conformation, but for the $74_{2,3}$ (SD) and $74_{2,3}$ (aSD/SD) a new FRET state around $E_{FRET} \approx 0.5$ emerged, which was not accessible before (Figure 79E & I). This new structure after mutation could either mean that the anti-SD:SD interaction gets destabilized allowing the folding of a less stable structure. In that case the anti-SD sequence would have been misassigned because double mutations did not restore the starting structure. It is also possible that the new FRET state originates from a stabilization of this alternative conformation by the specific mutations. Most likely both effects were present, a destabilization of one and a stabilization of the other structure. As for the unmutated versions, the smFRET data do not allow a prognosis on the 74mer structure. However, because of the changes when the SD sequence is mutated it can be concluded that either in WT 74mer or in the 74mer (SD) or in both the SD sequence is involved in basepairing and therefore not accessible for ribosome binding. More experiments need to be performed to estimate the structures of the longer constructs. This could either be smFRET experiments with different labeling schemes or more mutations as well as other structural biological techniques.

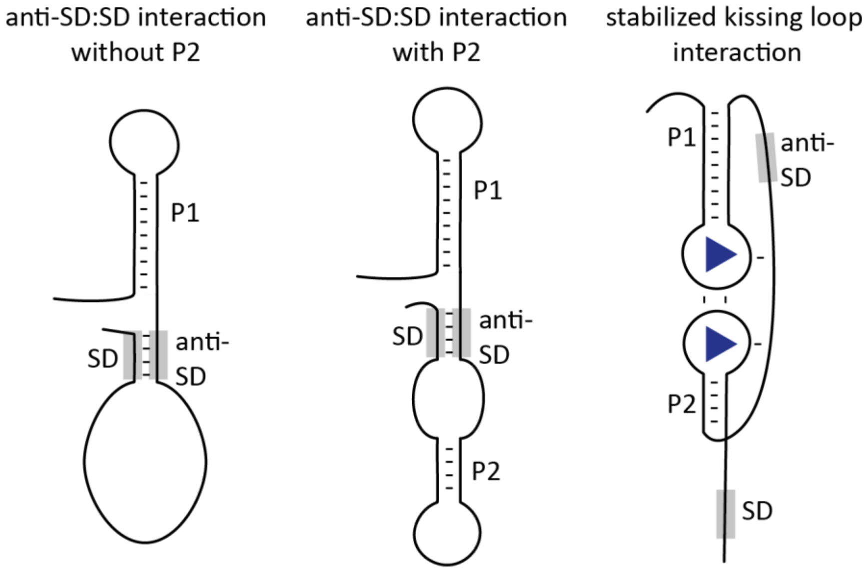


Figure 90: Potential conformations of the full-length Gdn-II riboswitch.

## 11.5 Loop nucleotides

In this work I proposed a stabilized kissing loop conformation. In this conformation the hairpins interact potentially through the fourth loop nucleotides with the linker. In the literature different groups have investigated the effects of the fourth loop nucleotides. Mutation of one of these nucleotides to a U reduced but did not abolish switching in in-line probing experiments[156]. Gdn-II riboswitches with a pyrimidine in that position in P2 resulted in an constitutively ON phenotype[173]. Furthermore, apparent affinities differed dependent on the purine used[168]. Using smFRET the effects were further investigated.

First of all, none of these mutation could disrupt the FRET states corresponding to the stabilized kissing loop in any of the tested constructs in a high $Mg^{2+}$ and $Gdm^{+}$ buffer (Figure 80D, H, L, P).

For the $93_{1,3}$ variants an A in P2 stabilized the overall structures (Figure 80). The FRET efficiency that was only accessible with $Mg^{2+}$ for $93_{1,3}$(AG) and $93_{1,3}$(GG) (Figure 80A & E) was partially occupied already in the absence of ions for $93_{1,3}$(AA) and $93_{1,3}$(GA) (Figure 80I & M). These two constructs had a second peak at intermediate FRET under $Mg^{2+}$ conditions while the other did not. Another indicator for the overall stability of the constructs with an A in P2 was observed when 200 mM $Gdm^{+}$ but no $Mg^{2+}$ was used. The ligand concentration was sufficient to almost quantitatively shift the equilibrium of these 93 nt long RNAs to a kissing loop conformation (Figure 80K & O). In case of a G in P2, only part of the molecules were shifted to the kissing loop FRET state (Figure 80C & G). This speaks for a higher affinity of the riboswitch variants with an A. Additionally the constructs behaved differently during the splinted ligation reaction (Figure 67). The AA and GA constructs did not anneal to the splint during the denaturing gel before DNase treatment. Since the same splint was used for all constructs and the splint was designed to complement a G in P2 it could simply be affinity reasons. This is not very likely as it was only one or two out of 93 base pairs and should not have such tremendous effects on the affinity and the GG version also had a mismatch, but remained annealed to the splint. It is more likely that the A in P2 stabilizes secondary structures that compete with the annealing to the splint. All these indicators argue towards stabilizing effects of an A in P2 for the 93mer RNA. As discussed above, it is not clear how the structures of the 93mer look like and whether P2 is formed as a hairpin in any of those structures. It is also possible that the A allowed different base pairings resulting in structures that do not include an ACGA tetraloop in P2. The new FRET states at 10 mM $Mg^{2+}$ are a sign for this.

Other smFRET loop mutation experiments were conducted with the $59_{1,2}$ construct (Figure 81). Here, the differences were less pronounced as for the $93_{1,3}$ RNA. The aptamer alone did not fold detectably into different conformations at the concentrations tested. Only two differences were detected. First, $59_{1,2}$(AU) had two populations at 10 mM $Mg^{2+}$ while the other only had one. The additional conformation had a similar FRET efficiency to that in absence of ions. This could either be a completely different conformation or more likely a mix of the open conformation and the loose kissing loop structure found in the other constructs in absence and presence of $Mg^{2+}$, respectively. In the latter case this means that more ionic strength is needed to stabilize the AU variant. Unless the second FRET state found for $59_{1,2}$(AU) was a structure that sequestered the anti-SD sequence, the smFRET data were not able to explain the always ON phenotype found for pyrimidine mutants *in vivo* [173].

The other difference of the loop mutant 59mers was in presence of high ligand concentration but absence of $Mg^{2+}$. As for the 93mers, an A in P2 stabilized the stabilized kissing loop conformation and shifted the equilibrium quantitatively to that FRET state. For $59_{1,2}$ (GG) the equilibrium was the most on the loose kissing loop side of all constructs which indicated a higher apparent $K_D$. This affinity tendency was confirmed by titration experiments (see below).

Differences between the RNAs with different loop nucleotides were also observed in *in vivo* experiments (Figure 87). Constructs with a G in P1 had additional bands in absence of $Gdm^+$ which was not the case for the other variants. The nature of these bands is not clear. There might be a possible connection to why the fourth loop nucleotides in P1 is usually conserved as an A. For the two constructs with a pyrimidine at that position in P2 (AU & AC) a new $Gdm^+$ dependent band was found in *in vivo* experiments. This new band length might be a hint to the always ON phenotype for these two mutants[173].

## 11.6 Concentration dependence

Most biomolecular binding events are concentration dependent. This is also the case for the Gdn-II riboswitch from *E. coli*. The smFRET data revealed some tendencies of ligand and ion dependencies. First of all, the condition where 200 mM $Gdm^+$ and no $Mg^{2+}$ was used, showed effects dependent on the RNA length and loop nucleotides. For the 59mer constructs the molecules almost exclusively populated the same stabilized kissing loop state as when $Mg^{2+}$and the ligand were both present. For the longer constructs there was an equilibrium between the FRET state found in the $Mg^{2+}$ condition and the stabilized kissing loop state. $Gdm^+$ is a cationic ligand at a physiological pH. It seemed like $Gdm^+$ at 200 mM could compensate the effects of the missing divalent ion. A role of cations in RNA folding is often to work as a counterion and shield the negatively charged backbone thus aid more compact folds that might otherwise be hindered by electrostatic repulsion. In the case of the Gdn-II riboswitch in presence of only $Gdm^+$, the ion might work as both, the cationic counterion and the ligand. If $Gdm^+$ is bound to the backbone it is not accessible to bind to the loop as a ligand anymore. In this way the effectively free $Gdm^+$ concentration would be reduced. The smFRET data for the longer constructs in presence of 200 mM $Gdm^+$ suggest that the free $Gdm^+$ concentration is closer to the $K_D$ than for the 10 mM $Mg^{2+}$ and 200 mM $Gdm^+$ condition where $Mg^{2+}$ could serve the role of the counterion instead or in addition to $Gdm^+$. Two reasons or a combination of both could explain why the equilibrium at 200 mM $Gdm^+$ is on the ligand bound side for the 59mer but not for the longer constructs. First, the longer constructs might have specific $Mg^{2+}$ binding sites that cannot be compensated well by the monovalent cation $Gdm^+$. Second, the ligand affinities could be different for the three construct lengths. The fact that the starting states in absence of ligand are different between the RNA lengths is a strong clue that this might be the case for Gdn-II. Schamber *et al.* claimed that

$Mg^{2+}$ binding is essential for kissing loop formation. In the NMR study higher RNA and lower $Gdm^+$ concentrations were used. The statement might be true for their working concentrations, but the smFRET data in presence of only $Gdm^+$ argues against this beeing a mechanistic feature. The less favorable $Gdm^+$ to RNA ratio in NMR might not be sufficient to saturate the phosphate backbone in contrast to the single-molecule experiments.

As discussed above, Gdn-II variants with an A in the P2 loop might have smaller $K_D$s. NMR experiments have resolved apparent $K_D$s for the aptamer domain with AA<AG<GA<GG. The aptamer variants $59_{1,2}$(AG) and $59_{1,2}$(GG) were subjected to a rough titration using smFRET. The results suggest apparent $K_D$s between 100 µM and 1 mM for the AG variant and between 1 mM and 10 mM for the GG variant. The first one is in the range of NMR and in-line probing experiments with 133 µM [168] and 300 µM [156], respectively. The rough apparent $K_D$ determined in this work for $59_{1,2}$(GG) is higher than the 264 µM found by NMR [168]. However, the mutants showed the same tendencies for the loop nucleotide dependence. In summary, the loop nucleotides influence the affinity of the Gdn-II riboswitch to its ligand possibly through interactions with the linker. The fourth loop nucleotide in P1 is mostly conserved as an A while for P2 both purines are approximately equally distributed [166]. Since the affinities vary, it is possible that the nucleotide identity has evolved to finetune the requirements for the particular riboswitches in the particular organisms. It should perform the switch to the ON state before toxic effects appear. In *E. coli* cell growth was impaired starting at 1 mM $Gdm^+$ (Appendix Figure 93). The WT Gdn-II riboswitch controls the expression of the detoxifying efflux pump SugE and acts below that concentration, thus fulfilling the requirement.

For future experiments it would be necessary to do more titration steps for a more precise $K_D$ estimation. Also the other loop nucleotide combinations should be tested, especially the AU variant, since there is no $K_D$ found in the literature. It would be very interesting to perform titration experiments in the future, comparing the aptamer constructs to the SD sequence containing RNAs.

## 11.7 *In vivo* effects of $Gdm^+$ and the Gdn-II riboswitch

$Gdm^+$ is best known for its denaturing properties above 1 M, but can also inhibit proteins in a more specific fashion. In this work the effects on bacterial cell growth were investigated. The growth was already impaired at $Gdm^+$ concentrations of 1 mM or above (Appendix Figure 93B). This was evidenced by an extended lag phase after nutrient depletion in overnight cultures (Figure 83 & Figure 86D). After exiting the lag phase, the exponential growth was normal and similar maximum

ODs were detected as without Gdm$^+$. There are two possible mechanisms how the bacteria might escape the Gdm$^+$ impaired lag phase. They could adapt the detoxification mechanisms such as the Gdn-II riboswitch regulated expression of the efflux pump SugE. They could also remove Gdm$^+$ below toxic levels from the medium by uptake and metabolizing to less or non-toxic molecules. When Gdm$^+$ is supplied in the exponential growth phase, cell growth is not impaired (Figure 85D). In the exponential growth phase, proteins, such as SugE, could be produced fast. Metabolizing of all Gdm$^+$ in the medium might take longer and might affect cell growth in the meantime. This argues more to a quick adaptation via efflux pumps. A better way to address the fate of Gdm$^+$ in the medium in future experiments could be metabolomic NMR experiments using isotope labeled Gdm$^+$. In case of export out of the cells, the levels in the medium should remain constant with low Gdm$^+$ levels in cell lysate. If the toxic molecule is metabolized, the metabolites could be tracked, and the medium concentrations would decrease.

To test how the Gdn-II riboswitch RNA behaves in an *in vivo* environment, a riboswitch containing construct was introduced to the *E. coli* cells. After incubation with and without Gdm$^+$ the total RNA was analyzed via denaturing PAGE. Several bands were visible on the gel even in absence of the riboswitch construct which belonged to endogenous *E. coli* RNAs (Figure 84). The tRNAs, which have a length between 74 and 95 n,t were visualized in two bands. Another intense band was the 5S ribosomal RNA which should have 120 nt[187]. Another 184 nt long RNA accumulates over time. This could be the non-coding 6S SsrS RNA[104]. Other very abundant RNAs in *E. coli* are the 16S and 23S rRNAs seen in the upper part of the gel. Due to their large size, they are not relevant for the further analysis. When the cells with the reporter construct are grown in presence of Gdm$^+$, several bands accumulated which were not or only weakly visible in absence of Gdm$^+$. The longest band migrated between the 164 nt 6S RNA and the 208 nt long reference. The second Gdm$^+$ dependent RNA under investigation ran just above the 96mer mark. This band had a similar intensity as a band that corresponds to an RNA which is smaller than the tRNAs. A fourth smaller RNA could be detected in a weak band when Gdm$^+$ was present. It cannot be excluded that other RNAs originating from the reporter plasmid were present in the RNA extract but were below detection limit or hidden by endogenous RNAs of the same length. The different bands might indicate different degradation or processing intermediates of the same RNA species. The distribution of which RNA lengths were detectable also depended on which medium was used (Figure 86) and at which growth phase Gdm$^+$ was added (Figure 85). This might be dependent on which processing factors or RNases were present at such growth phase. In initial *in vivo* experiments the amounts of the longer Gdm$^+$ dependent RNA constructs decreased after 24 h in favor of smaller RNA fragments (Figure 84). Effects of Gdm$^+$ supplementation were already visible in the whole RNA extract analysis after the first measurement time point of 30 min (Figure 85). After a while the amount of $Gdm^+$ specific RNAs decreased both in LB and in M9 medium (Figure 85 & Figure 86). As discussed above the intracellular $Gdm^+$ content probably decreases after a time due to different detoxification mechanisms. This would be reflected by the decrease RNA levels. This effect could point towards a delayed biosensor for the $Gdm^+$ concentrations in the cells. It is not clear how fast the RNA concentrations drop after removal of $Gdm^+$. Interestingly, two of the $Gdm^+$ dependent RNA bands were detectable in M9 minimal medium even if no external $Gdm^+$ was added. In contrast to the other sample, in absence of external $Gdm^+$ the bands increased over time. Considering the role of the experiment as a biosensor for $Gdm^+$, this would reflect increased $Gdm^+$ concentrations in the cells. This is in agreement with findings by Nelson *et al.* that $Gdm^+$ can be endogenously produced by bacteria in minimal medium[108].

The accumulation of RNAs can occur via different mechanisms. It might originate by an increased RNA production with the given conditions. Since the control plasmid with the same promoter but without the riboswitch did not change the RNA pattern this might rather be a secondary effect in this case. What is possible is the specific transcription termination or processing that would lead to the specific RNAs detected in the gel. Another way to accumulate RNA is via protection of RNA degradation. Different riboswitches use an RNA stability mechanism for regulation of gene expression. One example is another class of $Gdm^+$ binding riboswitches, the Gdn-III riboswitch, that protects the mRNA from RNase E scanning by forming a pseudoknot[132]. In a stabilize kissing loop conformation like characterized for the Gdn-II riboswitch with smFRET, scanning by RNase E or other RNases might also be blocked. Or the stable, basepaired RNA structure might evade degradation in other ways.

To get a first hint about the identity of the $Gdm^+$ dependent RNA bands, EMSAs were conducted using probes for the TSS, the SD sequence region and part of the 24-2 min aptamer. Several bands were targeted by the TSS probe for RNA extracts from cells grown with $Gdm^+$. It is very likely that some of the bands were the same as seen in the denaturing gels. However, the amount of probe was not sufficient in the experiment and some bands might have evaded the analysis. Furthermore, the EMSA uses native conditions and different bands might correspond to the same RNA length with different folds. For the SD region probe not such big differences were observed between the presence and absence of $Gdm^+$. Either a complementary sequence was not found in amounts above the detection limit, or the 24-2 min probe is somehow hindered to bind to the RNAs in the whole RNA cell extract. The 24-2 min probe also detected a band which was stronger under $Gdm^+$ conditions. This proves that $Gdm^+$ dependent RNAs at least partially originate from the *in vivo* assay plasmid as the 24-2 min sequence is synthetic. The 24-2 min probe should anneal far away from the $Gdm^+$ stabilized aptamer region and represents a part of the sequence where the expression regulation target would locate. This can be translated by a $Gdm^+$ dependent increase in mRNA coding the target protein. Since increased mRNA levels in general lead to more protein production, it seems that the ligand might somehow act as an RNA stabilizing factor in addition to the translational regulation mechanism.

For a better identification of the $Gdm^+$ dependent RNAs other methods than EMSAs such as Northern blotting need to be applied. Furthermore, it would be interesting for future experiments whether the altered RNA stability for Gdn-II involves the RNase E nuclease as for Gdn-III.

# 12 Summary and outlook

In summary, the smFRET and *in vivo* experiments performed in this work gave more details into the structure and regulation of the Gdn-II riboswitch from *E. coli*. A new three-state switching model for the aptamer domain was developed with an open, a loose kissing loop and a stabilized kissing loop conformation. Furthermore, for the first time the effects of the expression platform were analyzed. Possibly a sequestering of the SD sequence could be observed in absence of ligand, but also the longer constructs switched the riboswitch to a stabilized kissing loop conformation which would represent the ON-state. Additionally to the proposed translational mechanism, indications for a possible ligand-dependent RNA stability mechanism were introduced. This research also analyzed the effects of mutations at the fourth loop position.

In future experiments the structure of the Gdn-II riboswitch with its expression platform should be investigated with different structural biological approaches. Using different labeling schemes and mutations in smFRET experiments might also give more insights into the structure of this conformation. Further work needs to be done to determine and compare the $K_D$s of the full-length riboswitch with the aptamer only domain. This can be done for example by FRET titrations with more titration points. Also, the affinities of the loop mutants could be determined in future FRET experiments. Another focus to understand the full regulation mechanism of the Gdn-II riboswitch should be the investigation of a possible RNA stability mechanism. In that regard, Northern blotting experiments need to be conducted and a potential role of RNase E need to be determined.

# 13 Appendix Chapter III

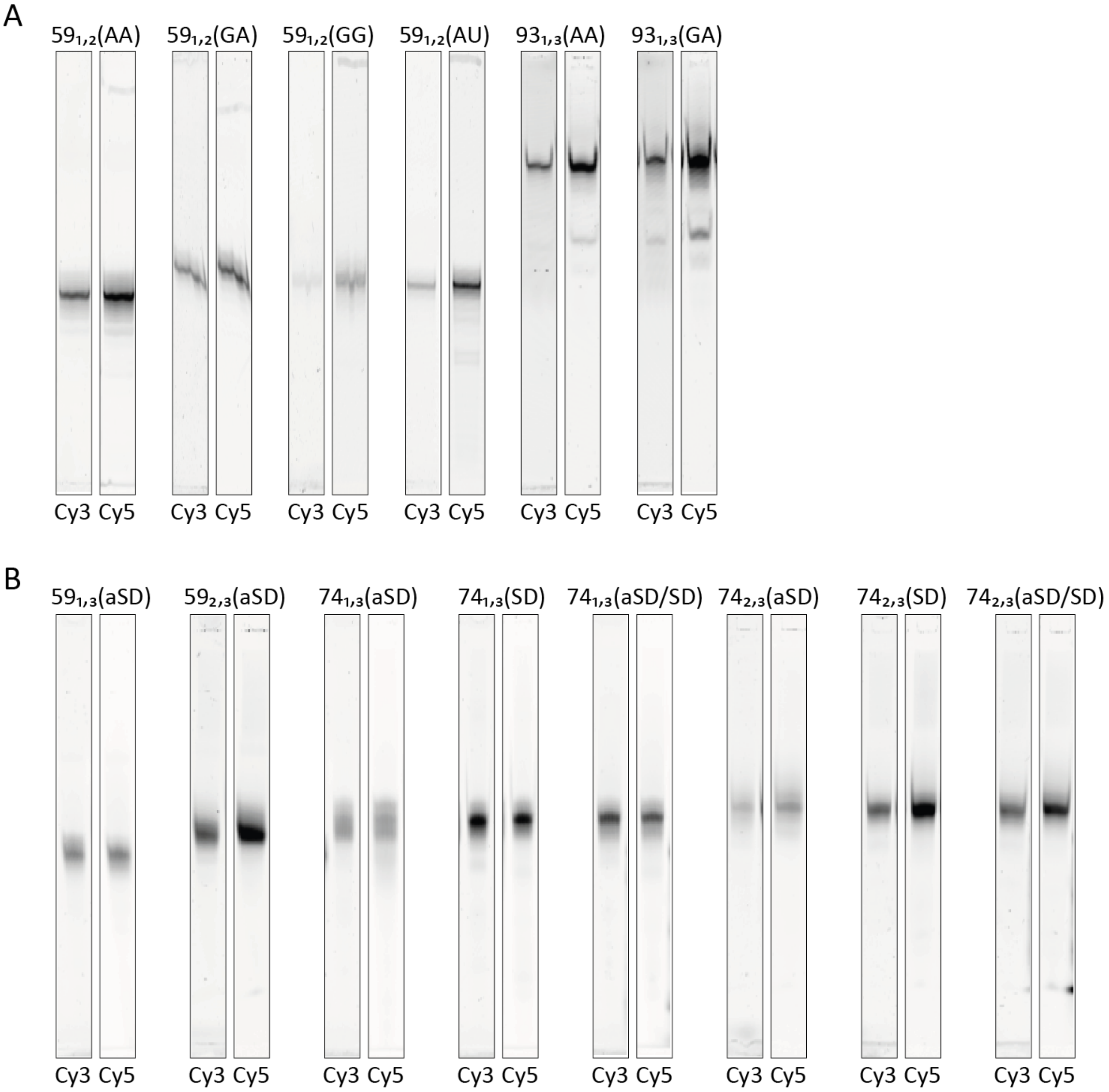


Figure 91: Analysis of unmutated purified smFRET constructs. A) Loop mutations B) SD mutations. The RNA was analyzed using denaturing gel. Different percentages of acrylamid/bisacrylamid solution were used and the gels were performed for different times. The fluorescent gels were imaged using a Typhoon scanner.

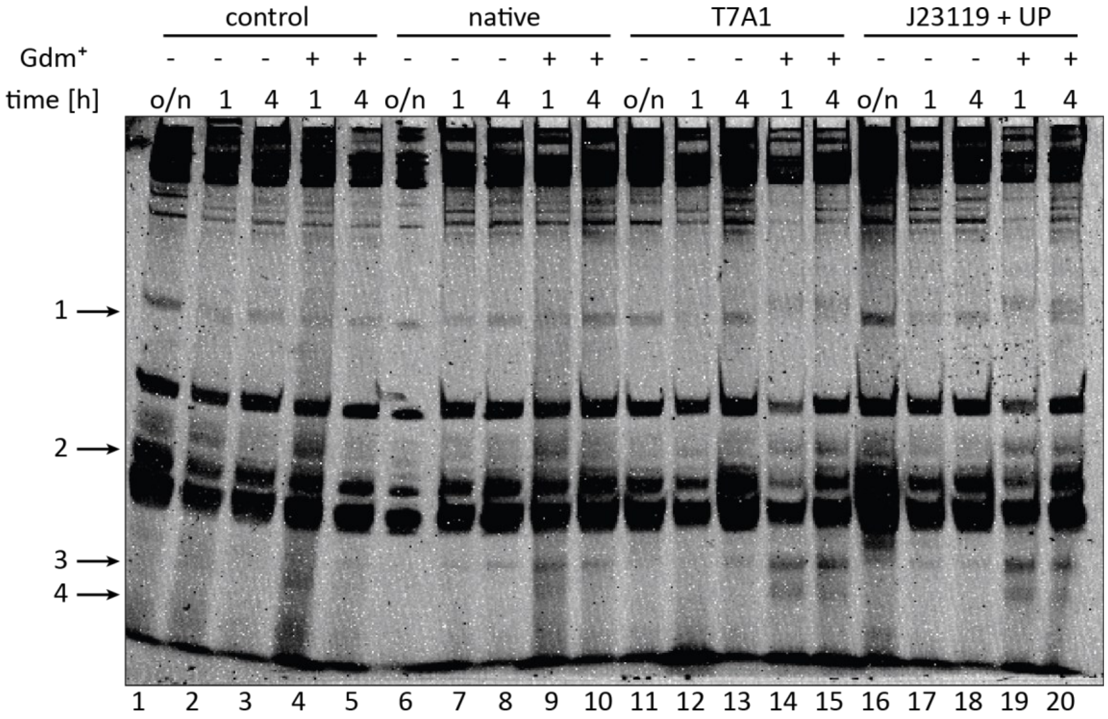


Figure 92: RNA pattern analysis of *in vivo* expressed Gdn-II riboswitch construct under different promoter control and a control plasmid without the riboswitch construct. *E. coli* were grown in LB medium with or without 50 mM Gdm⁺. Samples were taken at different time points and RNA amounts corresponding to the same ODs of bacterial culture were loaded onto a denaturing gel. RNA was visualized using GelRed.

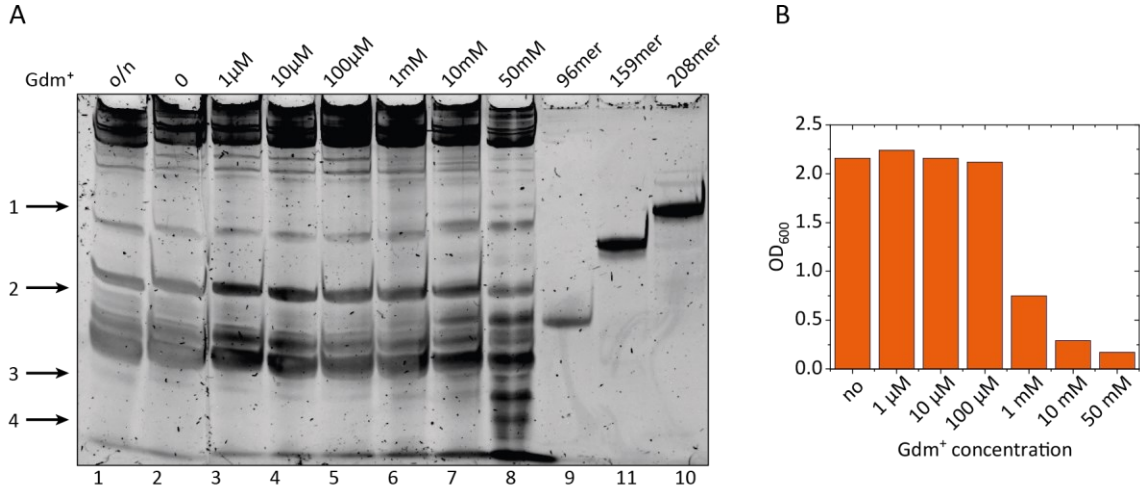


Figure 93: A) RNA pattern analysis of *in vivo* expressed Gdn-II riboswitch construct under J23119+UP promoter control. *E. coli* were grown in LB medium with varying amounts of Gdm⁺. Samples were taken after 4 h and RNA amounts corresponding to the same ODs of bacterial culture were loaded onto a denaturing gel. RNA was visualized using GelRed. B) ODs of bacterial culture after 4 h.

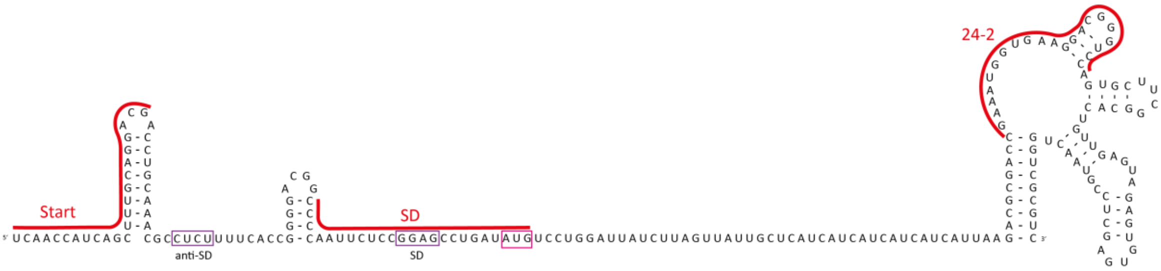


Figure 94: Sequence and EMSA annealing sites of the Gdn-II *in vivo* construct.

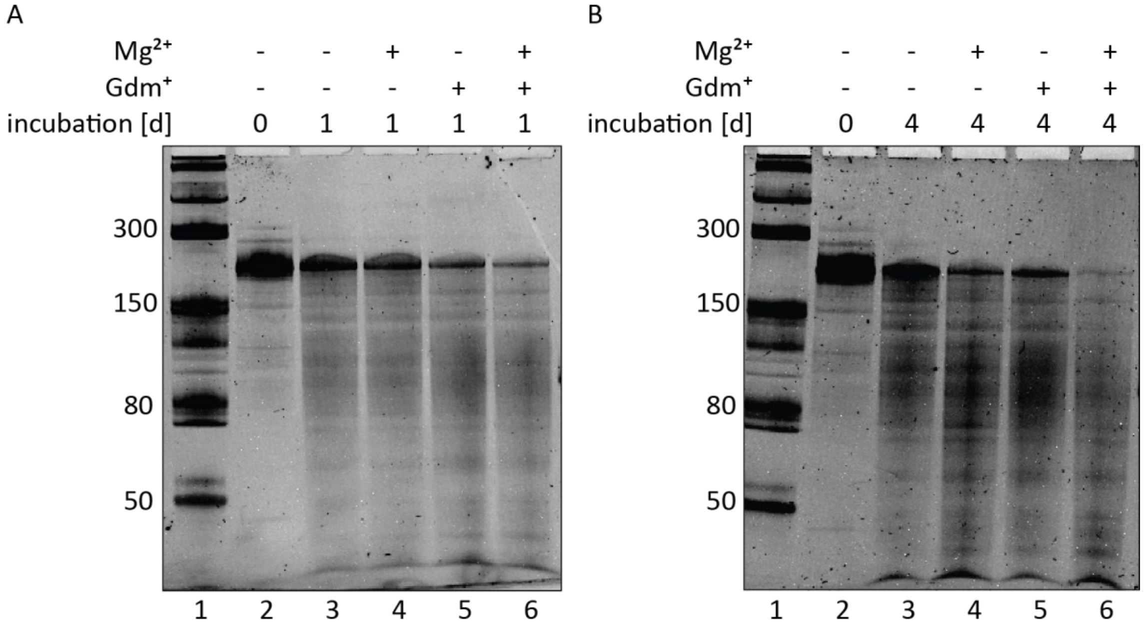


Figure 95: RNA stability assay. *In vitro* transcribed 208mer RNA with the *in vivo* construct sequence was incubated at 37 °C for A) 1 or B) 4 days. Samples were prepared in 1x transcription buffer with or without 5 mM $Mg^{2+}$ and 50 mM $Gdm^{+}$. RNA degradation was analyzed using denaturing PAGE. RNA was visualized using GelRed.

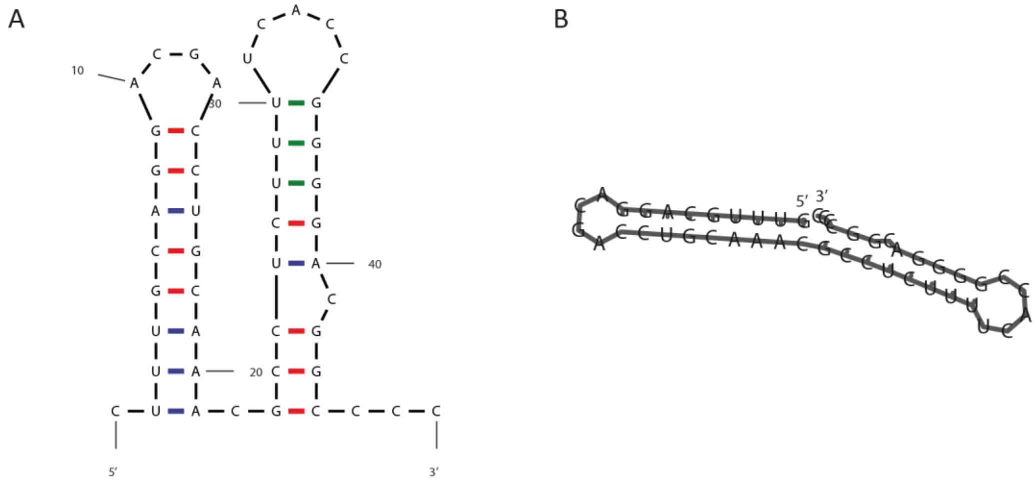


Figure 96: Secondary structure prediction for Gdn-II showing structures similar to the high FRET structures found by coarse-grained simulations using B) Mfold RNA folding Form Version 2.3[190] or C) Vienna RNA fold[189].

# Chapter IV

# Material and Methods

# 14 Material and Methods

## 14.1 Media

Table 3: Composition of media used in this work.

| Medium name | Components |
|---|---|
| LB medium | 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, pH 7.0 |
| LB agar | 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, 15g/L agar-agar, pH 7.0 |
| SB medium | 32 g/L tryptone, 20 g/L yeast extract, 5 g/L NaCl, pH 7.0 |
| TB medium | 20 g/L tryptone, 24 g/L yeast extract, 4 mL/L glyderol, 17 mM $KH_2PO_4$, 72 mM $K_2HPO_4$ |
| M9 medium | 6 g/L $Na_2HPO_4$, 3 g/L $KH_2PO_2$, 0.5 g/L NaCl, 1 g/L $NH_4Cl$, 2 g/L D-glucose, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 50 µM niacin, 50 µM thiamin, 50 µM biotin, 5 µM viatimin B12, 10 µM $FeCl_3$, 2 µM $MnCl_2$, 2 µM $ZnSO_4$, 0.4 µM $CoCl_2$, 0.4 µM $Na_2MoO_4$, 0.4 µM $H_2BO_3$, 0.4 µM $NiSO_4$, 0.4 µM $CuCl_2$, 0.4 µM $NaSeO_3$ |

## 14.2 Buffer

Table 4: Composition of buffers used in this work

| Buffer name | Components |
|---|---|
| Ψ buffer | 125 mM Tris-Cl, 250 mM NaCl, pH 8.0 |
| Coomassie staining solution | 10% (v/v) ethanol, 5% (v/v) HOAc, 2.5% (w/v) Coomassie Brillant Blue R250, 2.5% (w/v) Coomassie Brillant Blue G250, |
| Deprotection buffer | 100 mM HOAc, pH 3.4 – 3.8 (TEMED adjusted) |
| Destaining solution | 10% (v/v) ethanol, 5% (v/v) HOAc |
| FA buffer | 1x TBE buffer, 90% (v/v) formamide, 0.1% (w/v) bromphenol blue, 0.1% (w/v) xylene cyanol |
| IMAC A | 50 mM Tris-Cl, 500 mM NaCl, 10% (v/v) glycerol, 20 mM imidazol, 7 mM β-mercapto ethanol, pH 8.0 |

| **Buffer name** | **Components** |
|---|---|
| IMAC B | 50 mM Tris-Cl, 500 mM NaCl, 10% (v/v) glycerol, 500 mM imidazol, 7 mM β-mercapto ethanol, pH 8.0 |
| Imaging buffer (Gdn-II) | 50 mM Tris-Cl, 2.5 mM PCA, 50 nM PCD, Trolox (Sigma Aldrich) saturated, (+ Gdm-HCl, + $MgCl_2$) |
| Imaging buffer (H/ACA) | 1x Ψ buffer, 2.5 mM PCA, 50 nM PCD, Trolox (Sigma Aldrich) saturated, (+ proteins) |
| MES buffer | 50 mM 2-(N-morpholino)ethanesulfonic acid, 50 mM Tris, 0.1 % SDS, 0.1 mM EDTA, pH 7.3 |
| P1 buffer | 20 mM $NH_4Oac$, 0.5 mM $ZnCl_2$, pH 5.3 |
| Purple loading dye, no SDS | 2.5% (w/v) Ficoll-400, 10 mM EDTA, 3.3 mM Tris-Cl, 0.02% dye 1 (pink/red), 0.001% dye 2 (blue), pH 8.0 |
| SEC buffer | 50 mM Tris-Cl, 500 mM NaCl, 10% (v/v) glycerol, pH 8.0 |
| 2x SDS loading dye | 4x SDS stacking gel buffer, 20% (v/v) glycerol, 4% (w/v) SDS, 1.4% (v/v) β-ME, 0.05% (w/v) bromophenol blue |
| SDS stacking gel buffer | 250 mM Tris-Cl, 0.4% (w/v) SDS, pH 6.8 |
| SDS resolving gel buffer | 375 mM Tris-Cl, 0.4% (w/v) SDS, pH 8.8 |
| SDS running buffer | 25 mM Tris-Cl, 250 mM glycine, 0.1% (w/v) SDS |
| SSC buffer | 150 mM NaCl, 15 mM sodium citrate, pH 7.0 |
| STE buffer | 10 mM Tris-Cl, 100 mM NaCl, 1 mM EDTA, 0.5% (w/v) SDS, PH 8.0 |
| T4 PNK buffer | 70 mM Tris-Cl, 10 mM $MgCl_2$, 5 mM DTT, pH 7.6 |
| T4 DNA ligase buffer | 50 mM Tris-Cl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, pH 7.5 |
| T50 buffer | 10 mM Tris-Cl, 50 mM NaCl, pH 8.0 |
| T7 Script | 250 mM Tris/glutamate, pH 8.3 |
| TA buffer | 50 mM Tris-acetate, 100 mM NaOAc, pH 8.3 |
| TBE buffer | 90 mM Tris, 90 mM boric acid, 2 mM EDTA, pH 8.0 |
| TEAA buffer | 100 mM triethylammonium acetate, pH 7.0 |
| TLC buffer | 2-propanol/HCl(37%)/$H_2O$ (70:15:15 [v:v:v]) |

## 14.3 H/ACA sequences

### 14.3.1 Protein sequences

Table 5: H/ACA protein Sequences

| Name | Sequence | MW |
|---|---|---|
| Cbf5 | MAKEDFVIKP EAAGASTDTS EWPLLLKNFD KLLVRSGHYT PIPAGSSPLK RDLKSYISSG VINLDKPSNP SSHEVVAWIK RILRCEKTGH SGTLDPKVTG CLIVCIDRAT RLVKSQQGAG KEYVCIVRLH DALKDEKDLG RSLENLTGAL FQRPPLISAV KRQLRVRTIY ESNLIEFDNK RNLGVFWASC EAGTYMRTLC VHLGMLLGVG GHMQELRRVR SGALSENDNM VTLHDVMDAQ WVYDNTRDES YLRSIIQPLE TLLVGYKRIV VKDSAVNAVC YGAKLMIPGL LRYEEGIELY DEIVLITTKG EAIAVAIAQM STVDLASCDH GVVASVKRCI MERDLYPRRW GLGPVAQKKK QMKADGKLDK YGRVNENTPE QWKKEYVPLD NAEQSTSSSQ ETKETEEEPK KAKEDSLIKE VETEKEEVKE DDSKKEKKEK KDKKEKKEKK EKKDKKEKKE KKEKKRKSED GDSEEKKSKK SKKHHHHHH | 55.5 kDa |
| Nop10 | MHLMYTLGPD GKRIYTLKKV TESGEITKSA HPARFSPDDK YSRQRVTLKK RFGLVPGQ | 6.6 kDa |
| Nhp2 | MGKDNKEHKE SKESKTVDNY EARMPAVLPF AKPLASKKLN KKVLKTVKKA SKAKNVKRGV KEVVKALRKG EKGLVVIAGD IWPADVISHI PVLCEDHSVP YIFIPSKQDL GAAGATKRPT SVVFIVPGSN KKKDGKNKEE EYKESFNEVV KEVQALGLEH HHHHH | 18.3 kDa |
| Gar1 Ctag | MSFRGGNRGG RGGFRGGFRG GRTGSARSFQ QGPPDTVLEM GAFLHPCEGD IVCRSINTKI PYFNAPIYLE NKTQVGKVDE ILGPLNEVFF TIKCGDGVQA TSFKEGDKFY IAADKLLPIE RFLPKPKVVG PPKPKNKKKR SGAPGGRGGA SMGRGGSRGG FRGGRGGSSF RGGRGGSSFR GGSRGGSFRG GSRGGSRGGF RGGRRHHHHH H | 22.3 kDa |
| Gar1 no tag | MSFRGGNRGG RGGFRGGFRG GRTGSARSFQ QGPPDTVLEM GAFLHPCEGD IVCRSINTKI PYFNAPIYLE NKTQVGKVDE ILGPLNEVFF TIKCGDGVQA TSFKEGDKFY IAADKLLPIE RFLPKPKVVG PPKPKNKKKR SGAPGGRGGA SMGRGGSRGG FRGGRGGSSF RGGRGGSSFR GGSRGGSFRG GSRGGSRGGF RGGRR | 21.5 kDa |

| Name | Sequence | MW |
|---|---|---|
| Gar1<br>Δ3 | MGSSHHHHHH SSGLVPRGSH MSFRGGNRGG RGGFRGGFRG GRTGSARSFQ QGPPDTVLEM GAFLHPCEGD IVCRSINTKI PYFNAPIYLE NKTQVGKVDE ILGPLNEVFF TIKCGDGVQA TSFKEGDKFY IAADKLLPIE RFLPKPKVVG PPKPKNKKKR SGAPGGRGGA SMGRGGSRGG FRGGRGGSSF RGGRGGSSF | 21.1 kDa |
| Gar1<br>Δ1 | MGSSHHHHHH SSGLVPRGSH MGPPDTVLEM GAFLHPCEGD IVCRSINTKI PYFNAPIYLE NKTQVGKVDE ILGPLNEVFF TIKCGDGVQA TSFKEGDKFY IAADKLLPIE RFLPKPKVVG PPKPKNKKKR SGAPGGRGGA SMGRGGSRGG FRGGRGGSSF RGGRGGSSFR GGSRGGSFRG GSRGGSRGGF RGGRR | 20.5 kDa |
| Gar1<br>Δ2,3 | MGSSHHHHHH SSGLVPRGSH MSFRGGNRGG RGGFRGGFRG GRTGSARSFQ QGPPDTVLEM GAFLHPCEGD IVCRSINTKI PYFNAPIYLE NKTQVGKVDE ILGPLNEVFF TIKCGDGVQA TSFKEGDKFY IAADKLLPIE RFLP | 15.7 kDa |
| Gar1<br>Δ1,3 | MGSSHHHHHH SSGLVPRGSH MGPPDTVLEM GAFLHPCEGD IVCRSINTKI PYFNAPIYLE NKTQVGKVDE ILGPLNEVFF TIKCGDGVQA TSFKEGDKFY IAADKLLPIE RFLPKPKVVG PPKPKNKKKR SGAPGGRGGA SMGRGGSRGG FRGGRGGSSF RGGRGGSSF | 18.0 kDa |
| Gar1<br>Δ1,2,3 | MGSSHHHHHH SSGLVPRGSH MGPPDTVLEM GAFLHPCEGD IVCRSINTKI PYFNAPIYLE NKTQVGKVDE ILGPLNEVFF TIKCGDGVQA TSFKEGDKFY IAADKLLPIE RFLP | 12.6 kDa |

### 14.3.2 RNA sequences

Table 6: H/ACA RNA sequences

| Name | Sequence | Use |
|---|---|---|
| snR81 | GGG ACU GCA AAA GAA GCG GCG AGG CAG CCC ACA UCA AGU GGA ACU ACA CAG ACU UCC UUG UCG CGA UAC UAC GGU CCC AAG AGC AAU CCU AAC AAG CAA UUA CAU AUU CCC CCG CUG AAC CUG UAC AGU CCA CGG AUG GUG CAG AAG UUA UAU GAU UUG GGG GAA GAC GCU UUU UCA CAU CUU CUU GCA UGA UAA | snoRNA |
| M164 | (N6G)GG ACU GCA AAA GAA GCG G | FRET RNA fragment 1 |
| M167 | P-CGA GGC AGC CCA CAU CAA GUG GAA CUA CAC AGA CUU CCU UG(5N-U) CGC | FRET RNA fragment 2 |
| M595 | P-GAU ACU ACG GUC CCA AGA GCA AUC CU | FRET RNA fragment 3 |
| M439 | P-AAC AAG CAA UUA CAU AUU CC(5-LC-N-U) CCG CU | FRET RNA fragment 4 |
| M440 | P-GAA CCU GUA CAG UCC ACG GAU GGU GCA GAA GUU AUA UGA UU | FRET RNA fragment 5 |
| M441 | P-UGG GGG AAG ACG CUU (5-LC-N-U)UU CAC AUC UUC UUG CAU GAU AA-biotin | FRET RNA fragment 6 |
| M171 | GGG AGU AGU AUC | 5’ substrate fragment 1 |
| M172 | UGU UCU UUU CAG | 5’ substrate fragment 2 |
| M173 | GGG AAC UU | 3’ substrate fragment 1 |
| M174 | UAA AUA UGU AAG AA | 3’ substrate fragment 2 |

### 14.3.3 DNA sequences

Table 7: H/ACA DNA splints for splinted ligation

| Name | Sequence |
|---|---|
| M599 | GGA CCG TAG TAT CGC GAC AAG GAA GTC TGT GTA GTT CCA CTT GAT GTG GGC TGC CTC GCC GCT TCT TTT GCA GTC CC |
| M600 | AAA AAG CGT CTT CCC CCA AAT CAT ATA ACT TCT GCA CCA TCC GTG GAC TGT ACA GGT TCA GCG GAG GAA TAT GTA ATT GCT TGT TAG GAT TGC TCT TG |

## 14.4 Gdn-II sequences

### 14.4.1 RNA sequences

Table 8: Gdn-II FRET RNA fragments

| Name | Sequence | Use |
|---|---|---|
| M408 | UCA ACC AUC AGC U(5NU)U GCAG | fragment 1 |
| M410 | P-GAC GAC CUG CAA ACG | fragment 2 |
| M411 | P-GAC GGC CUG CAA ACG | fragment 2 (P1G) |
| M412 | P-CCU CUU U(5NU)C ACC GGG | fragment 3 |
| M413 | P-CCU CUU UUC ACC GGG | fragment 3 |
| M616 | P-CGU GUU U(5-LC-N-U)C ACC GGG | fragment 3 (aSD) |
| M617 | P-CGU GUU UUC ACC GGG | fragment 3 (aSD) |
| M415 | P-GAC GGC CCC A N6 | fragment 4 |
| M416 | P-GAC GAC CCC A | fragment 4 (P2A) |
| M417 | P-GAC GAC CCC A N6 | fragment 4 (P2A) |
| M630 | P-GAC GUC CCC A | fragment 4 (P2U) |
| M418 | P-AU(5NU) CUC CGG AGC CUG | fragment 5 |
| M419 | P-AUU CUC CGG AGC CUG | fragment 5 |
| M618 | P-AU(5-LC-N-U) CUC CGC ACC CUG | fragment 5 (SD) |
| M619 | P-AUU CUC CGC ACC CUG | fragment 5 (SD) |
| M420 | P-GAC GAC CUG CAA ACG | fragment 6 |

### 14.4.2 DNA sequences

Table 9: Gdn-II DNA splints for splinted ligation

| Name | Sequence |
|---|---|
| M519 | TAA TAC GAC TCA CTA TAG GTC AAC CAT CAG CTT TCA GGA CGA CCT GCA AAC GCC TCT TTT CAC CGG GGA CGG CCC CAA TTC TCC GGA GCC TGA TAT GTC CTG GAT TAT CTT |
| M613 | AAG ATA ATC CAG GAC ATA TCA GGG TGC GGA GAA TTG GGG CCG TCC CCG GTG AAA AGA GGC GTT TGC AGG TCG TCC TGC AAA GCT GAT GGT TGA |
| M614 | AAG ATA ATC CAG GAC ATA TCA GGC TCC GGA GAA TTG GGG CCG TCC CCG GTG AAA ACA CGC GTT TGC AGG TCG TCC TGC AAA GCT GAT GGT TGA |
| M615 | AAG ATA ATC CAG GAC ATA TCA GGG TGC GGA GAA TTG GGG CCG TCC CCG GTG AAA ACA CGC GTT TGC AGG TCG TCC TGC AAA GCT GAT GGT TGA |

Table 10: DNA sequences of *in vivo* constructs. The full constructs contain one promoter and one construct. The nucleoties for loop mutations are highlighted.

| Name | Sequence |
|---|---|
| native promoter | GTA ATA GTT GAA ATT CCC CTG CCA CCT GGC AAA ATA TCC GT |
| T7A1 promoter | TTA TCA AAA AGA GTA TTG ACT TAA AGT CTA ACC TAT AGG ATA CTT ACA GCC |
| J2319+UP promoter | GGA AAA TTT TTT TAA AAA AAA AAC TTG ACA GCT AGC TCA GTC CTA GGT ATA ATG CTA GC |
| Gdn-II *in vivo* construct | TCA ACC ATC AGC TTT GCA GGA CG**A** CCT GCA AAC GCC TCT TTT CAC CGG GGA CG**G** CCC CAA TTC TCC GGA GCC TGA TAT GTC CTG GAT TAT CTT AGT TAT TGC TCA TCA TCA TCA TCA TCA TTA AGA CGC GAC CGA AAT GGT GAA GAA CGG GTC CAG TGC TTC GGC ACT GTT GAG TAG AGT GTG AGC TCC GTA ACT GGT CGC GTC |
| Δaptamer control construct | TCA ACC ATC AAT TCT CCG GAG CCT GAT<br>ATG TCC TGG ATT ATC TTA GTT ATT GCT CAT CAT CAT CAT CAT CAT TAA GAC GCG ACC GAA ATG GTG AAG GAC GGG TCC AGT GCT TCG GCA CTG TTG AGT AGA GTG TGA GCT CCG TAA CTG GTC GCG TC |

Table 11: DNA probes for EMSA

| Name | Sequence | Probe |
|---|---|---|

| | | |
|---|---|---|
| M596 | CGT CCT GCA AAG CTG ATG GTT GA | TSS |
| M597 | CAT ATC AGG CTC CGG AGA ATT GGG G | SD |
| M598 | TGG ACC CGT CCT TCA CCA TTT CG | 24-2 min |

## 14.5 General biochemical methods

### 14.5.1 Native polyacrylamide gel electrophoresis (PAGE)

DNA e. g. from PCR was analyzed using native polyacrylamide gel electrophoresis (PAGE). For the gel preparation 10% rotiphoreses acrylamide/bisacrylamid 29:1 (Carl Roth) in 1x TBE buffer was polymerized with 0.1% (w/v) APS and 0.1% (v/v) TEMED between two glass plates. DNA samples were mixed with 6x purple loading dye, no SDS (NEB). Gel electrophoresis was performed with TBE buffer and a voltage of 200 V.

### 14.5.2 Denaturing PAGE

Denaturing polyacrylamide gel electrophoresis (PAGE) was used for analysis of RNA and for preparative RNA purification. Gels were polymerized with varying concentrations of rotiphorese acrylamide/ bisacrylamide 29:1 in 1x TBE with 0.1% (w/v) APS and 0.1% (v/v) TEMED between two glass plates. Samples were mixed with FA buffer and loaded onto the gel. Gel electrophoresis was performed in 1x TBE buffer at 220 V. The progression of the PAGE was followed based on bromphenol blue and xylene cyanol bands.

### 14.5.3 Analysis of PAGE

After electrophoresis gel bands were visualized with different methods depending on the labeling of the RNA and the analytical or preparative character of the gels. Analytical gels with fluorophores were scanned with a Typhoon 9400 scanner (GE Healthcare). Analytical gels of nucleic acids without fluorophores were transferred to a GelRed solution (biotium) and incubated for 15 min and DNA and RNA bands were imaged using a GelDoc XR+ Imager.

Preparative Gels with unlabeled RNA were analyzed with UV shadowing. For this, the gel was transferred to a fluorescent thin layer chromatography plate and illuminated with UV light. RNA containing bands remained dark. The target bands were excised from the gel. For preparative gels with fluorophore labeled RNAs an analytical lane was scanned with a Typhoon 9400 scanner. The images were printed out without rescaling and used as a template for excision on the preparative lanes. For preparative gels with radioactive labeled RNA the gel was covered with plastic wrap and exposed to a storage phosphor screen (GE Healthcare) for 30 s. The screen was scanned using a Typhoon 9400 scanner (GE Healthcare). The resulting image was printed out and used as a template for excision of the target RNA bands with small radioactive markers as positioning aids.

### 14.5.1 Electrophoretic mobility shift assay (EMSA)

For detection of the Gdn-II *in vivo* construct from the extracted RNA, an electrophoretic mobility shift assay (EMSA) was used. The RNA was mixed with 150 nM of the Cy5-labeled probe targeting either the TSS, SD sequence region or the 24-2 min aptamer (Table 11 & Figure 94) in 1x SSC buffer and incubated at 65 °C for 5 min. Glycerol was added to the samples to a final concentration of 10% (v/v). A native PAGE in 1x TA buffer was performed at 4 °C. The gel was run with <1 W. The Cy5 fluorescence was imaged using a Typhoon 9400 imager (GE healthcare).

### 14.5.2 SDS PAGE

Protein samples were analyzed using SDS-polyacrylamide gel electrophoresis (SDS-PAGE). Either self-made gels or commercial available gels were used. Self-made gels were divided into a stacking and running gel. The running gel was prepared by polymerizing 15% rotiphorese acrylamide/bisacrylamid 29:1 with 0.1% (w/v) APS and 0.1% (v/v) TEMED in running gel buffer between two glas plates. The solution was coated with isopropanol. After polymerization the isopropanol was removed and a layer of stacking gel with 5% polyacrylamide in stacking gel buffer was applied.

Protein samples were mixed with 2x SDS loading buffer and loaded onto the gel. Gel electrophoresis with self-made gels was performed in SDS buffer for 20 min at 80 V afterwards for 40 min at 200 V. Gel electrophoresis with purchased gels was performed in MES buffer for 30 min at 200 V.

For staining the gels were transferred into Coomassie staining solution and heated in a microwave. After 20 min the staining solution was replaced with destaining solution and heated again. After destaining for >16 h gels were imaged using an Image Doc XR+ (BioRad).

### 14.5.1 UV/Vis spectroscopy

The concentration of biomolecules except for Gar1 was determined using UV/Vis spectroscopy. Absorption was measured using a NanoDrop One (ThermoFisher). DNA and RNA was measured at 260 nm, proteins at 280 nm, Cy3 at 550 nm, and Cy5 at 650 nm. The concentrations were determined according to Lambert-Beer's Law (equation 3).

$$A = \varepsilon_{\lambda} \cdot c \cdot d \tag{3}$$

A = absorption at wavelength λ

$\varepsilon_{\lambda}$ = extinction coefficient at wavelength λ

c = concentration

d = path length of the light beam

For the components for H/ACA reconstitution experiments additionally the 260/280 nm ratio was determined. An adjusted $\varepsilon_{260}$ and $\varepsilon_{280}$ was calculated with equations 4 and 5 assuming a protein:RNA molar ratio of 2 for full reconstitution.

$$\varepsilon_{260} = \varepsilon_{260}(RNA) + 2 \cdot \varepsilon_{280}(protein) \cdot \frac{260}{280}(protein) \tag{4}$$

$$\varepsilon_{280} = 2 \cdot \varepsilon_{280}(protein) + \varepsilon_{260}(RNA) \cdot \frac{260}{280}(RNA)^{-1} \tag{5}$$

### 14.5.2 Bradford assay

For protein concentration determination of Gar1 and Gar1 mutants a Bradford assay was used. To produce Bradford reagent 50 mg Coomassie Brilliant Blue G250 were dissolved in 25 mL 95% EtOH. 50 mL of 85% phosphoric acid were added and the solution was diluted with water to a final volume of 500 mL. Bradford reagent was stored at 4 °C until further use. For the assay 250 µL of freshly filtered Bradford reagent were mixed with 5 µL of protein sample in a clear bottom 96-well plate. After incubation for 15 min at room temperature absorption was measured at 595 nm in a Tecan Spark plate reader. Absorption was compared to a BSA protein standard (Sigma Aldrich).

## 14.6 DNA based methods

### 14.6.1 PCR

dsDNA templates for T7 transcription were synthesized via polymerase chain reaction (PCR). PCR was also used for blunt end cloning. 0.5 µM of forward and reverse primer (Eurofins gemomics) were mixed with 10 ng of template DNA (ssDNA, dsDNA, or plasmid DNA), 200 µM dNTP mix and 1% (v/v) of home-made Phusion polymerase in 1x HF buffer. In a thermocycler the reaction mixture was heated to 98 °C for 30 s. In 30 cycles the mixture was first melted at 98 °C for 10 s, then DNA was annealed at 40-65 °C (depending on $T_m$ of the primers) for 15 s followed by elongation at 72 °C for 30 – 120 s (depending on the length of the PCR product). In a last step the reaction was incubated at 72 °C for 2 min. PCR products were analyzed using native PAGE.

### 14.6.2 Cloning

To remove the loops of the H/ACA guide RNA and the exchange of promoters for Gdn-II *in vivo* and *in vitro* transcriptions, blunt end cloning of full-plasmid PCR products. First primers were phosphorylated. 10 µM of each primer was incubated with 666.7 U/mL T4 polynucleotide kinase (NEB) in T4 DNA ligase buffer at 37 °C for 30 min. Primers were used in a standard phusion PCR reaction in 50 µL scale with an elongation time of 2 min. To digest the DNA template 30 U of DpnI (NEB) were added and incubated at 37 °C for 1 h. The DNA was purified with a PCR purification kit (Macherey Nagel) according to manufacturer’s protocol and the resulting DNA concentration was determined using UV/Vis spectroscopy. Approximately 500 ng of DNA were ligated overnight with 26,667 U/mL T4 DNA ligase in T4 DNA ligase buffer at 16 °C. The ligase was heat inactivated at 60 °C for 10 min and the ligation product was transformed into DH5α high efficiency cells (NEB). Plasmids with the target sequence were selected through sequencing.

Single nucleotide mutations were introduced using the Lightning QuikChange Site-directed Mutagenesis Kit (Agilent Genomics). Overlapping primers carrying the target mutation were designed and purchased from Eurofins Genomics. The reaction mixture contained 1x reaction buffer, 10 ng of dsDNA template, 0.3 µM of each primer, 0.6 µL of Quik Solution reagent and 0.4 µL of Pfu polymerase in a total volume of 20 µL. PCR was performed in a thermocycler. Initial melting took place at 95 °C. In 18 cycles the DNA was melted for 20 s at 95 °C, primers were annealed at 60 °C for 10 min and elongation took place at 68 °C for 3 min. PCR was finished after a final elongation step of 5 min. 1 µL of DpnI was added and incubated for 1 h at 37 °C. The DNA was transformed into DH5α high efficiency cells (NEB). Plasmids with the target mutation were selected through sequencing.

### 14.6.3 DNA isolation and sequencing

Plasmid DNA was amplified in DH5α cells. SB medium with antibiotic selection was inoculated with the transformed bacteria. The cells were grown overnight at 37 °C with vigorous shaking. The DNA from 4 mL of overnight culture was extracted using a QIAprep Spin Miniprep kit (Quiagen) according to manufacturer’s protocol. Concentration of the plasmid DNA was determined using UV/Vis spectroscopy. 20 µL (40-200 ng/µL) were sent to Eurofins Genomics for sequencing.

## 14.7 RNA based methods

### 14.7.1 T7 Transcription

Unlabeled RNA was prepared via T7 transcription. 1x T7 Script buffer, 5 mM of each NTP, 30 mM of MgOAc, 2 mM spermidine, 10 mM of DTT, 10% (v/v) DNA from the PCR, and 0.04 mg/mL home-made T7 polymerase were incubated for 2-5 h at 37 °C. 20 U/mL of Turbo DNase were added followed by further incubation for 30 min. Precipitated magnesium pyrophosphate was resolubilized with sufficient amounts of EDTA.

### 14.7.2 Ethanol precipitation

For precipitation RNA was mixed with ammonium acetate ($NH_4OAc$) to a final concentration of 0.5 M, and 2.5 Vol of cold, absolute ethanol (EtOH) were added. The mixture was incubated for <15 min at -80 °C or overnight at -20 °C before centrifugation at 17,000 x g for >30 min. The supernatant was carefully removed and the pellet was dried using a SpeedVac vacuum concentrator.

### 14.7.3 RNA labeling and deprotection

RNA fragments were purchased from Dharmacon in a 2’-ACE protected form. Fragments that were to be fluorophore labeled contained an amino modification either at their 5’ or 3’ end or an amino modified uridine (5-Aminohexylacrylamino-uridine) within their sequence. For fluorophore labelling 30 nmol of aminomodified RNA were precipitated using 0.3 M NaOAc instead of $NH_4OAc$. The pellet was resuspended in 20 µL fresh 0.1 M $NaHCO_3$ buffer. 0.5-1 Cy3 or Cy5 maleimide mon-reactive dye packs (Cytiva) were dissolved in 20 µL DMSO and mixed with the RNA. The NHS-ester reaction took place for 90 min at room temperature in the dark. Labeled RNA was recovered using EtOH precipitation

2’-ACE protected RNA was deprotected by incubation with deprotection buffer (Dharmacon) at 60 °C for 30 min or 2 h for biotinylated oligonucleotides. Deprotection was followed by EtOH precipitation. Pellets were resuspended in $H_2O$ and concentrations were determined using UV/Vis spectroscopy.

### 14.7.4 Reverse Phase Chromatography

To separate labeled RNA fragments from unlabeled RNA a reverse phase liquid chromatography (RP-LC) was performed. RNA fragments were diluted 1:1 with TEAA buffer and injected onto a C8 column (250 x 4.6 mm, Kromasil) mounted to an Äkta Basic$^{TM}$ system. Separation was achieved using a gradient of TEAA buffer to acetonitrile in 80 mL with a flow rate of 1 mL/min. Absorption was detected at 260 nm, 550 nm and 650 nm for RNA, Cy3 and Cy5, respectively. The eluate was collected in 300 µL fractions and analyzed via denaturing PAGE. Fractions containing the labelled RNA fragments were pooled and EtOH precipitated. Pellets were resuspended in $H_2O$ and concentrations were determined using UV/Vis spectroscopy.

### 14.7.5 Splinted ligation

Deprotected RNA fragments were ligated to obtain the smFRET constructs in a splinted ligation. RNA oligonucleotides and the single-stranded DNA splints were mixed in equimolar ratios to a concentration of 2-5 µM in 0.5 x T4 DNA Ligase buffer. The mixture was incubated for 4 min at 85 °C and subsequently slowly cooled to 37 °C over 15 min followed by incubation at 37 °C for 10 min. Another 0.5 x T4 DNA Ligase buffer and 10% (v/v) of home-made T4 RNA ligase 2 were added. The ligation reaction was carried out for 2-3 h at 37 °C, 300 rpm in the dark. Subsequently, the ligase was heat inactivated for 15 min at 65 °C. The splint DNA was removed by adding Turbo DNase (ThermoFisher) to a final concentration of 0.2 U/µL and incubation at 37 °C for 45 min.

### 14.7.6 Phenol/ether extraction

After ligation, enzymes were removed in a phenol/ether extraction. 1 Vol Eq of Aqua-Phenol (Carl Roth) was added to the reaction mixture, vortexed and phases were separated via centrifugation (3,000 rpm, 2 min). The upper aqueous phase was transferred to a new tube and phenol extraction was repeated. Subsequently, the aqueous phase was extracted three times with 3 Vol Eq of water saturated diethyl ether. Residual ether was removed via incubation at 45 °C for 20 min with an open lid.

### 14.7.7 RNA gel excision

Gel pieces that contain the target RNA length in the preparative lane were cut out of the gel and incubated in a 0.5 M $NH_4OAc$ or 0.3 M NaOAc solution overnight at 16 °C. The supernatant was filtered using centrifugal filter units and subsequently precipitated. RNA was resuspended in $H_2O$ and concentrations were determined using UV/Vis spectroscopy.

## 14.1 Cell based methods

### 14.1.1 Transformation

Genetic material was introduced into *E. coli* bacteria via transformation. Autoclaved LB agar was liquefied in a microwave oven. 15 mL of lukewarm LB agar was supplied with the respective antibiotic and transferred to a petri dish. The LB agar plate was cooled and hardened. For protein expression the BL21(DE3) strain for all other applications DH5α cells were used in their competent form purchased from NEB. Cells were thawed on ice for 5 min. 1-5 µL (1 pg – 100 ng) of plasmid DNA were added to the bacteria and incubated for 30 min on ice. A heat shock at 42 °C was performed for 10 s (BL21(DE3)) or 30 s (DH5α) in a water bath. Bacteria were immediately incubated on ice for 5 min. 300 µL of SOC medium were added and the cells were incubated for 1 h at 37 °C with vigorous shaking. 150 µL of transformed bacteria were spread in a dilution smear on a LB agar plate. Agar plates were incubated at 37 °C overnight or at room temperature for up to 3 days. Agar plates were sealed with parafilm and stored at 4 °C for up to 2 weeks.

### 14.1.2 Cryo Stocks

2 mL of LB medium with the respective antibiotic were inoculated with transformed *E. coli* from an agar plate and bacteria were grown at 37 °C overnight. 500 µL of dense bacteria culture were thoroughly mixed with 500 µL of 50% glycerol (v/v) and frozen in liquid nitrogen. Glycerol stocks were stored at -80 °C.

### 14.1.3 *In vivo* expression experiments

For expression of Gdn-II riboswitch RNA *in vivo* pEX Vectors with different constructs were transformed into DH5α *E. coli* bacteria. The constructs consist of one of the three promoters (native, T7A1 or J23119+UP) followed by either the *in vivo* riboswitch construct with or without loop mutations or the Δaptamer control construct (Table 10). Cells were grown overnight in LB medium with the respective antibiotic. Depending on the experiment, pH 7.0 adjusted LB medium or M9 medium with or without $Gdm^+$ was inoculated from the overnight culture to a final $OD_{600}$ of 0.1. Cells were grown at 37 °C and samples were taken at different time points. For dilution experiments *E. coli* were grown to the mid-logarithmic $OD_{600}$ of 1.4 and were diluted with the same volume of fresh, pre-warmed LB medium with or without 100 mM $Gdm^+$ (final concentration 50 mM) before further incubation.

### 14.1.4 Bacterial RNA extraction

To extract total bacterial RNA from the *in vivo* expression, samples, with an amount of bacteria corresponding to 1 mL $OD_{600}$ of 1 were taken. They were pelleted by centrifugation at 13,000 rpm for 5 min. The supernatant was discarded and the bacteria pellet was stored at -80 °C. For further extraction, the pellet was resuspended in 50 µL STE buffer. 50 µL of Aqua-phenol were added and mixed by vortexing. After centrifugation at 13,000 rpm for 5 min the supernatant was further used for denaturing PAGE and EMSAs.

## 14.2 Protein based methods

### 14.2.1 Protein expression

LB medium (30 mL for 1 L main culture) with antibiotic(s) (100 µg/mL ampicillin and or 30 µg/mL kanamycin) was inoculated with BL21(DE3) cells carrying the desired plasmids either from agar plate or from cryo stock. Bacteria were grown overnight to the stationary phase at 37 °C and 160 rpm shaking. $OD_{600}$ was determined. 1 L of TB medium was prepared in 5 L shaking flasks. Antibiotics and 2 drops of Antifoam Y-30 emulsion (Sigma Aldrich) were added. The TB medium was inoculated with the overnight culture to a final $OD_{600}$ of 0.1 – 0.2. Bacteria were incubated at 37 °C and at 120 rpm shaking until an $OD_{600}$ of 0.6 – 2 was reached. Then a cold shock was performed by incubation of the shaking flasks in ice cold water for 15 min. 100 µM of Isopropyl-β-D-thiogalactopyranosid (IPTG) were added. The cells were subsequently incubated at 18 °C and 120 rpm shaking overnight. Bacteria were harvested by centrifugation at 4,000 x g for 20 min. The supernatant was discarded and the cell pellets were frozen in liquid nitrogen and stored at -80 °C until purification.

### 14.2.2 Protein purification

**Cell lysis**

Frozen cell pellets were resuspended in 30 mL IMAC A buffer supplied with 10 µL Turbo DNase and 20 µL RNase A. Cell suspension was sonicated with a Sonoplus ultrasonic homogenizer for >7 min in 30 s intervals with >1 min breaks on ice. Disrupted cells were centrifuged at 15,000 rpm at 4 °C for 1 h with a Beckman SW 32 Ti rotor in a Beckman Colter Optima L-90K ultracentrifuge.

**PEI precipitation**

To protein complexes containing Cbf5 0.05% of polyethylenimine was added to the supernatant and the mixture was incubated for 10 min at 4 °C before centrifugation at 8,000 x g for 10 min. This step was repeated once. For all proteins the supernatant was filtered with 0.45 µm and 0.2 µm syringe filters.

**Primary IMAC**

The cell lysate was loaded onto a 5 mL Ni-loaded HisTrap pre-equilibrated with IMAC A. Afterwards the column was washed with IMAC A buffer until UV absorption was constant. For constructs containing Cbf5 the column was washed with IMAC A supplemented with 1 M LiCl. Proteins were eluted from the column with a gradient from IMAC A to IMAC B in 10 CV. Fractions were collected and analyzed using SDS PAGE. Fractions containing the target proteins were pooled and incubated overnight with 30 µL RNase A at 4 °C.

**Secondary IMAC**

Protein aggregates were removed by centrifugation (8,000 x g, 10 min, 4 °C). Supernatant was diluted with SEC buffer to an imidazole concentration of 20 mM and loaded onto a HisTrap. After washing with IMAC A the protein was eluted and fractionated with a gradient from 0 – 20% IMAC B in 2 CV and then to 100% IMAC B in 3 CV. Fractions were analyzed using SDS PAGE.

**Size Exclusion Chromatography (SEC)**

Fractions containing the target protein were pooled and concentrated to a volume <200 µL with a VivaSpin centrifugal concentrator (30 kDa Molecular Weight Cutoff for Cbf5 containing complexes, 5 kDa for Gar1 and Gar1 derivates). Size exclusion chromatography was performed using an Äkta purifier system. For Cbf5 complexes a Superdex 200 increase 10/300 (flow rate 0.5 mL/min), for Gar1 Δ1,2,3 a Superdex 75 HiLoad 26/600 (flow rate 2.5 mL/min) and for all other Gar1 variants a Superdex 75 10/300 column (flow rate 0.4 mL/min) were used. Fractions were analyzed using SDS PAGE. Pure protein containing fractions were pooled, concentrated, aliquoted, frozen in liquid nitrogen and stored at -80 °C until further use.

## 14.3 H/ACA reconstitution and activity

### 14.3.1 RNP reconstitution and analytical SEC

H/ACA complexes with different guide RNAs were reconstituted *in vitro* and analyzed with analytical SEC. Reconstitution was achieved after incubation of 1-10 µM of RNA with varying protein ratios in SEC buffer for 30 min at 30 °C. 50 µL were subjected to an analytical Superdex 200 Increase 3.2/300 on an Äkta Purifier system with a flow rate of 0.075 mL/min. As a buffer SEC buffer was used. During the run absorbance at 260 nm and 280 nm was measured. The eluate was fractioned in 100 µL fractions.

### 14.3.2 Pseudouridylation activity assays

**$^{32}$P RNA preparation**

For pseudouridylation activity assays the H/ACA substrate RNAs were generated with a $^{32}$P radioactive labeled phosphate 5’ of the pseudouridylation site. Substrate RNAs were purchased in 2 fragments with the split before the target uridine. 5 nmol of 3’ fragments were phosphorylated in a 30 µL reaction with 0.4 U/µL T4 polynucleotide kinase (NEB) in 1x T4 PNK buffer with $\gamma$-$^{32}$P ATP (2 µM, 2.2 MBq/µL) for 1 h at 37°C. For final phosphorylation, non-radioactive ATP was added to a final concentration of 10 mM and the mixture was incubated for 1 h. Unincorporated ATP was removed with an Illustra Micro-Spin G-25 column (GE Healthcare). Phosphorylated RNA fragments were ligated to the 5’ fragments in a splinted ligation. RNA and DNA components were mixed with 0.5x T4 DNA ligase buffer in a final concentration of 50 µM. After heating to 85 °C for 5 min they were cooled down to room temperature. Another 0.5x T4 DNA ligase buffer was added and 80 U/µL T4 DNA ligase. The reaction was carried out at 16°C for 6 h or overnight followed by EtOH precipitation. Ligated RNA was purified with preparative denaturing PAGE.

**Pseudouridylation activity assay**

Pseudouridylation activity was tested for selected fractions after analytical SEC of reconstituted H/ACA complex. Fractions of reconstituted complex were used in a final concentration of 0.1 µM and incubated with 0.1 µM of radioactive substrate in 1x SEC-buffer at 30 °C for 1 h. The protein components were removed via phenol-ether extraction. The RNA was EtOH precipitated in presence of additional unlabeled, unrelated RNA to aid precipitation. Afterwards the RNA was resuspended in P1 buffer and digested with P1 ribonuclease (0.0125 U/µL) for 2 h at 55 °C. P1 was inactivated at 70 °C for 30 min. 2x 1.5 µL were transferred to a 20 cm cellulose thin layer chromatography plate. Thin layer chromatography was performed with TLC buffer for 20-24 h. Subsequently the plate was air dried and exposed to a storage phosphor screen (GE Healthcare) for 1 d. The phosphor screen was analyzed with a Typhoon 9400 imager (GE Healthcare). Pseudouridylation efficiency was calculated by comparing intensities of Ψ and U spots using the Fiji ImageJ software package[191].

## 14.4 smFRET spectroscopy methods

### 14.4.1 Slide Preparation

For smFRET measurements micro-Tec quartz micorosope slides (76.2 x 25.4 x 1 mm) were used. Slides were cleaned by first boiling them in water and removing residual coverslips and parafilm. Then the slides were scrubbed with Alconox detergens (Sigma Aldrich). Slides were subsequently sonicated for 10 min in a glass slide holder, first with Alconox solution, then with acetone, then with 1 M KOH solution and finally with water. In between they were rinsed with water. After sonication in water, slides were dried on a bunsen burner. Cleaned slides can be used for RNA only FRET measurements.

For FRET samples with protein the quartz slides have to be PEGgylated. An empty glass slide holder with methanol was sonicated for 5 min. 100 mL of methanol were mixed with 5 mL of acetic acid and 1 mL (3-aminopropyl)-triethoxy silane (Sigma Aldrich) and replaced the methanol in the glass slide holder. The slide holder was filled with clean quartz slides and sonicated for 1 min and incubated for 20 min. Afterwards slides were rinsed with water and dried with nitrogen. For five slides 200 mg of PEG (mPEG-succinimidyl valerate, MW 5,000) and 2 mg of PEG-Biotin (Biotin-PEG-succinimidyl valerate, MW 5,000) were resuspended by vortexing in 800 µL and 400 µL fresh, sterile 0.1 M $NaHCO_3$ buffer, respectively. Solutions were mixed and centrifuged at top speed. 150 µL of the mixture were transferred to each of the microscope slides and covered with a coverslip (22 x 50 mm). Slides were incubated overnight in a moisture chamber in the dark. At the next day coverslips were removed and slides rinsed with water and dried with nitrogen. PEGgylated slides were stored in an argon atmosphere at -80 °C.

To build the sample chambers stripes of parafilm were arranged on the cleaned or PEGglyated slides to generate 3-7 channels. The slides prepared with parafilm were covered with a coverslip (22 x 40 mm). Overhanging Parafilm was removed. The slides were put coverslip side down on a pre-heated aluminium block at 85 °C for 30-60s. Afterwards, the coverslip was carefully pressed to the slide.

### 14.4.2 FRET sample preparation

The Gdn-II riboswitch FRET constructs were measured on BSA slides and the H/ACA RNP experiments were conducted using PEG-slides.

For Gdn-II RNA folding, 20 µL of 50-500 pM riboswitch FRET construct in 50 mM TRIS-Cl pH 7.4 with the respective MgCl and Gdn-HCl concentrations were incubated at 95 °C for 5 min and immediately placed on ice for >5 min. The pre-prepared channels on the cleaned slides were filled with 15 µL of 1 mg/mL BSA biotin (Sigma Aldrich) in T50 buffer. After 2 min, the channel was washed with 50 µL T50 buffer. The buffer was exchanged with 30 µL of 0.2 mg/mL streptavidin (Invitrogen) and followed by incubation for 2 min. The channel was equilibrated with 50 mM TRIS-Cl pH 7.4 buffer with the respective MgCl and Gdn-HCl concentration. The folded RNA was added to the channel and incubated for 2 min. 50 µL of Imaging buffer were prepared by adding PCA (final concentration 2.5 mM) and PCD (final concentration 50 nM) to the pre-imaging buffer. The imaging buffer was applied to the channels and incubated for at least 5 min before measurement.

The H/ACA samples were prepared by incubating 0.1 – 1 nM of the FRET constructs at 30 °C for 5 min before placing the RNA on ice. The pre-prepared channels on the PEGgylated slides were filled with 15 µL of 0.2 mg/mL streptavidin (Invitrogen) followed by incubation for 2 min. The channel was equilibrated with 50 µL 1x Ψ-buffer. The folded RNA was added to the channel and incubated for 2 min. 50 µL of imaging buffer were prepared by adding PCA (final concentration

2.5 mM), PCD (final concentration 50 nM) and 1 µM of the respective proteins to the pre-imaging buffer. The imaging buffer was applied to the channels and incubated for at least 10 min before measurement.

### 14.4.3 FRET data aquisition

smFRET measurements were conducted on a prism-type TIRF microscope (Olympus). A green (532 nm) laser (gem 532, 500 mW, Laser Quantum) and a red (637 nm) laser (OBIS LX, 140 mW, Coherent) were used with laser beam intensities of 5 – 15 mW. For imaging an EMCCD camera was used (Evolve 512 Delta with CCD97 sensor, Photometrics).

To be able to align the green and red channel during the analysis, for every measurement day one slide with TetraSpeck microspheres (0.1 µm, fluorescent blue/green/orange/dark red, Invitrogen) was measured (20 frames, 100 ms exposure time, EM gain 10). The samples were measured with a molecule density of 100 – 500 molecules per image. For histogram data without donor only reduction a Tif stack of 20 frames with green laser excitation was acquired (100 ms exposure time, EM gain 200). For histogram data with donor only reduction the molecules were illuminated with the green laser for the first 20 frames and then it was switched to the red laser for the last 20 frames (100 ms exposure time, EM gain 200). The scripts to switch the lasers using the LabView software were provided by Nathalie Meiser.

### 14.4.4 FRET analysis

The Tif Stacks were transformed to 16-int raw files using the Andor Solis (v4.21) software. Image processing was continued with the IDL (student edition v6.2, Exelis) software. To align the green and red channel the measurement of the TetraSpecks was analyzed with the scripts 'beads_co2' and 'beads_map' to create a transformation map. To obtain donor and acceptor intensities for the individual sample molecules the scripts 'dat_all2', 'ffp_dat2', 'ap_dat2' and 'ave_tr_dat' were used. The generated .trace files were further analyzed using MatLab (R2021a, MathWorks).

For analyis the scripts readhisttraceandor.m or for direct donor only removal readhisttraceandor_acceptorfilter.m were used. In both cases the acceptor intensitiy was corrected due to donor leakage with $I_{A,cor} = I_A - x \cdot I_D$. x was chosen to set the donor only peak to $E_{FRET} = 0$. Molecules with a dyesum below 2,000 were removed from analysis. In case of donor only removal also all molecules with an acceptor intensity below a red_threshold (500 - 4,000) in the last 20 frames were removed from analysis. For all remaining molecules the intensity of the first 20 frames was averaged and binned with a bin size of 0.025. The binned data was exported to

OriginPro 2021b (OriginLab).

If the donor only peak was not removed in MatLab, the peak was fitted with one or two Gaussian (equation 6) fits and substracted from the data. The binned data was normalized. If possible the histogram was fitted using one to three Gaussian fits.

$$y = \frac{A}{w\sqrt{\frac{\pi}{2}}} e^{-2\frac{(x-x_c)^2}{w^2}} \tag{6}$$

A = amplitude
w = peak width
$x_c$ = center of maximum

The scripts for IDL and MatLab were provided by Dr. Martin Hengesbach. The MatLab script readhisttraceandor_acceptorfilter.m was edited for removal of donor only peaks based on readhisttraceandor.m

# 15 Appendix of Chapter IV

**Readhisttraceandor.m**

```
%Program to read binary traces file output by Hazens IDL program written
by
%Michael Stone 01/17/2004
clear all;
total_frets = [];
total_dyesum = [];
n_molecules = [];


s = pwd;
files = dir(s);

for i = 3:length(files)

fileName = files(i).name;
s = strcat(s,'\');
fullFileName = strcat(s,fileName);
FID = fopen(fullFileName);

% data = importdata(fullFileName);
accepted_traces = []; % an array to hold good molecule data for future
analysis
fretthreshold = 2000; % this is a threshold value to prevent crazy fret
traces


%find number of frames and peaks from traces header file
[n_fr,z1] = fread(FID,1,'int32');    %z1 is just the number of indicated
data types read (should be 1)
[n_traces,z2] = fread(FID,1,'int16');

rate_fr = 10;%input('At what frame rate was the data collected?');
n_peaks = n_traces/2;
n_molecules = [n_molecules n_peaks];
exp_length =  n_fr/rate_fr; %(seconds)
time = (1/rate_fr:(1/rate_fr):exp_length);


%% Now read in the rest of the data from the traces file
[total_data,z3] = fread(FID,[n_traces+1,n_fr],'int32');
fclose=(FID);
total_data = total_data(2:n_traces,1:n_fr);   %% this gets rid of the
frame number col, which is the first colum in a traces file

%RemoveBleachedMolecules


total_donor = [];
total_acceptor = [];
total_acceptor_corrected = [];
```

```
total_donor(1:n_peaks-1,1:n_fr) = total_data(2*(0:n_peaks-2)+1,1:n_fr);
%%remember matlab starts arrays at 1
total_acceptor(1:n_peaks-1,1:n_fr) = total_data(2*(0:n_peaks-
2)+2,1:n_fr);


%correct the acceptor intensities for leakage from the donor channels
%using 12% for the 630 nm dichroics

total_acceptor_corrected(1:n_peaks-1,1:n_fr)= total_acceptor(1:n_peaks-
1,1:n_fr)-0.142*total_donor(1:n_peaks-1,1:n_fr);
%total_acceptor_corrected = total_acceptor;

for i = 1:n_peaks-1
    i;
    dyesum = total_donor(i,:)+total_acceptor_corrected(i,:);
    FRET = total_acceptor_corrected(i,:)./dyesum;
    [rows] = find(dyesum <fretthreshold);
    FRET(rows) = [];
    total_frets = [total_frets mean(FRET)];
    %remove FRET below 0.1 and above 1.1:
    [rowsoutsidelow] = find(total_frets <-0.1);
    total_frets(rowsoutsidelow) = [];
    [rowsoutsidehigh] = find(total_frets >1.1);
    total_frets(rowsoutsidehigh) = [];
    total_dyesum = [total_dyesum mean(dyesum)];
    total_dyesum (rowsoutsidelow) = [];
    total_dyesum (rowsoutsidehigh) = [];

end

end

bins = [-0.1:0.025:1.1];
bins = bins';
% and this makes the vector for exporting the histogram figures (to
Origin)
binned_data = hist(total_frets,bins);
binned_data = binned_data';
% finally, this gives the raw data in a vector for export
export_frets = total_frets';

    figure
    hist(total_frets,bins)
    % plot A
    title(fileName)
```

### Readhisttraceandor_acceptorfilter.m

```
clear all;
total_frets = [];
n_molecules = [];

n_fr_analysis = 20 ;%input ('How many frames should be analyzed?');
red_on = 30 ; %input ('Which frame was red laser turned on?')
red_threshold = input ('What is your acceptor threshold?')

s = pwd;
files = dir(s);

for i = 3:length(files)

fileName = files(i).name;
s = strcat(s,'\');
fullFileName = strcat(s,fileName);
FID = fopen(fullFileName);

% data = importdata(fullFileName);
accepted_traces = []; % an array to hold good molecule data for future
analysis
fretthreshold = 2000; % this is a threshold value to prevent crazy fret
traces

%find number of frames and peaks from traces header file
[n_fr,z1] = fread(FID,1,'int32');    %z1 is just the number of indicated
data types read (should be 1)
[n_traces,z2] = fread(FID,1,'int16');

rate_fr = 10;%input('At what frame rate was the data collected?');
n_peaks = n_traces/2;
n_molecules = [n_molecules n_peaks];
exp_length =  n_fr/rate_fr; %(seconds)
time = (1/rate_fr:(1/rate_fr):exp_length);


%% Now read in the rest of the data from the traces file
[total_data,z3] = fread(FID,[n_traces+1,n_fr],'int32');
fclose=(FID);
total_data = total_data(2:n_traces+1,1:n_fr);   %% this gets rid of the
frame number col, which is the first colum in a traces file


%RemoveBleachedMolecules


total_donor = [];
total_acceptor = [];
total_acceptor_corrected = [];

total_donor(1:n_peaks-1,1:n_fr) = total_data(2*(0:n_peaks-2)+1,1:n_fr);
%%remember matlab starts arrays at 1
```

```
total_acceptor(1:n_peaks-1,1:n_fr) = total_data(2*(0:n_peaks-
2)+2,1:n_fr);


%correct the acceptor intensities for leakage from the donor channels
%using 12% for the 630 nm dichroics

total_acceptor_corrected(1:n_peaks-1,1:n_fr)= total_acceptor(1:n_peaks-
1,1:n_fr)-0.148*total_donor(1:n_peaks-1,1:n_fr);
%total_acceptor_corrected = total_acceptor;




for i = 1:n_peaks-1
    i;
    acceptor_intensity = mean (total_acceptor_corrected
(i,red_on:n_fr)); % intensity after switching on the red laser
    if acceptor_intensity > red_threshold; % only molecules that have
acceptor signal
        dyesum =
total_donor(i,1:n_fr_analysis)+total_acceptor_corrected(i,1:n_fr_analysi
s);
        FRET = total_acceptor_corrected(i,1:n_fr_analysis)./dyesum;
        [rows] = find(dyesum <fretthreshold);
        FRET(rows) = [];
        total_frets = [total_frets mean(FRET)];
    %remove FRET below 0.1 and above 1.1:
        [rowsoutsidelow] = find(total_frets <-0.1);
        total_frets(rowsoutsidelow) = [];
        [rowsoutsidehigh] = find(total_frets >1.1);
        total_frets(rowsoutsidehigh) = [];
    end
end

end

%this creates the vector for binning into 0.025 FRET steps
bins = [-0.1:0.025:1.1];
bins = bins';
% and this makes the vector for exporting the histogram figures (to
Origin)
binned_data = hist(total_frets,bins);
binned_data = binned_data';
% finally, this gives the raw data in a vector for export
export_frets = total_frets';

    figure
    hist(total_frets,bins)
    % plot A
    title(fileName)
```

# 16 List of References

1. Lichtman, J. W. & Conchello, J.-A. Fluorescence microscopy. *Nature methods* **2,** 910–919; 10.1038/nmeth817 (2005).
2. Forster, T. Energiewanderung und Fluoreszenz. *Naturwissenschaften* **33,** 166–175; 10.1007/BF00585226 (1946).
3. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nature methods* **5,** 507–516; 10.1038/nmeth.1208 (2008).
4. 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* **13,** 2472–2483; 10.1021/acschembio.8b00348 (2018).
5. Hengesbach, M. *et al.* RNA intramolecular dynamics by single-molecule FRET. *Current protocols in nucleic acid chemistry* **Chapter 11,** Unit 11.12; 10.1002/0471142700.nc1112s34 (2008).
6. Shrestha, D., Jenei, A., Nagy, P., Vereb, G. & Szöllősi, J. Understanding FRET as a research tool for cellular studies. *International journal of molecular sciences* **16,** 6718–6756; 10.3390/ijms16046718 (2015).
7. Brasselet, S., Peterman, E. J. G., Miyawaki, A. & Moerner, W. E. Single-Molecule Fluorescence Resonant Energy Transfer in Calcium Concentration Dependent Cameleon. *The Journal of Physical Chemistry B* **104,** 3676–3682; 10.1021/jp993954o (2000).
8. Kapanidis, A. N. & Weiss, S. Fluorescent probes and bioconjugation chemistries for single-molecule fluorescence analysis of biomolecules. *The Journal of Chemical Physics* **117,** 10953–10964; 10.1063/1.1521158 (2002).
9. Schuler, B. & Eaton, W. A. Protein folding studied by single-molecule FRET. *Current opinion in structural biology* **18,** 16–26; 10.1016/j.sbi.2007.12.003 (2008).
10. Hillger, F., Nettels, D., Dorsch, S. & Schuler, B. Detection and analysis of protein aggregation with confocal single molecule fluorescence spectroscopy. *Journal of fluorescence* **17,** 759–765; 10.1007/s10895-007-0187-z (2007).
11. Michalet, X. *et al.* Detectors for single-molecule fluorescence imaging and spectroscopy. *Journal of modern optics* **54,** 239; 10.1080/09500340600769067 (2007).
12. Axelrod, D. Total internal reflection fluorescence microscopy in cell biology. *Methods in enzymology* **361,** 1–33; 10.1016/s0076-6879(03)61003-7 (2003).
13. Green, N. M. Avidin and streptavidin. *Methods in enzymology* **184,** 51–67; 10.1016/0076-6879(90)84259-J (1990).

14. Sofia, S. J., Premnath, V. & Merrill, E. W. Poly(ethylene oxide) Grafted to Silicon Surfaces: Grafting Density and Protein Adsorption. *Macromolecules* **31,** 5059–5070; 10.1021/ma971016l (1998).

15. Eggeling, C., Widengren, J., Rigler, R. & Seidel, C. A. Photobleaching of Fluorescent Dyes under Conditions Used for Single-Molecule Detection: Evidence of Two-Step Photolysis. *Analytical chemistry* **70,** 2651–2659; 10.1021/ac980027p (1998).

16. Hübner, C. G., Renn, A., Renge, I. & Wild, U. P. Direct observation of the triplet lifetime quenching of single dye molecules by molecular oxygen. *The Journal of Chemical Physics* **115,** 9619–9622; 10.1063/1.1421382 (2001).

17. Benesch, R. E. & Benesch, R. Enzymatic removal of oxygen for polarography and related methods. *Science (New York, N.Y.)* **118,** 447–448; 10.1126/science.118.3068.447 (1953).

18. Aitken, C. E., Marshall, R. A. & Puglisi, J. D. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophysical journal* **94,** 1826–1835; 10.1529/biophysj.107.117689 (2008).

19. Swoboda, M. *et al.* Enzymatic oxygen scavenging for photostability without pH drop in single-molecule experiments. *ACS nano* **6,** 6364–6369; 10.1021/nn301895c (2012).

20. Patil, P. V. & Ballou, D. P. The use of protocatechuate dioxygenase for maintaining anaerobic conditions in biochemical experiments. *Analytical biochemistry* **286,** 187–192; 10.1006/abio.2000.4802 (2000).

21. Rasnik, I., McKinney, S. A. & Ha, T. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nature methods* **3,** 891–893; 10.1038/nmeth934 (2006).

22. Cohn, W. E. 5-Ribosyl uracil, a carbon-carbon ribofuranosyl nucleoside in ribonucleic acids. *Biochimica et biophysica acta* **32,** 569–571; 10.1016/0006-3002(59)90644-4 (1959).

23. Boccaletto, P. *et al.* MODOMICS: a database of RNA modification pathways. 2021 update. *Nucleic Acids Res* **50,** D231-D235; 10.1093/nar/gkab1083 (2022).

24. Cantara, W. A. *et al.* The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res* **39,** D195-201; 10.1093/nar/gkq1028 (2011).

25. Begik, O. *et al.* Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nature biotechnology* **39,** 1278–1291; 10.1038/s41587-021-00915-6 (2021).

26. Behrens, A., Rodschinka, G. & Nedialkova, D. D. High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. *Molecular Cell* **81,** 1802-1815.e7; 10.1016/j.molcel.2021.01.028 (2021).

27. Meyer, K. D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149,** 1635–1646; 10.1016/j.cell.2012.05.003 (2012).

28. Pendleton, K. E. *et al.* The U6 snRNA m6A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. *Cell* **169,** 824-835.e14; 10.1016/j.cell.2017.05.003 (2017).

29. Kawahara, Y. *et al.* Frequency and fate of microRNA editing in human brain. *Nucleic Acids Res* **36,** 5270–5280; 10.1093/nar/gkn479 (2008).

30. Jin, D. *et al.* m6A mRNA methylation initiated by METTL3 directly promotes YAP translation and increases YAP activity by regulating the MALAT1-miR-1914-3p-YAP axis to induce NSCLC drug resistance and metastasis. *J Hematol Oncol* **12,** 135; 10.1186/s13045-019-0830-6 (2019).

31. Machnicka, M. A., Olchowik, A., Grosjean, H. & Bujnicki, J. M. Distribution and frequencies of post-transcriptional modifications in tRNAs. *RNA biology* **11,** 1619–1629; 10.4161/15476286.2014.992273 (2014).

32. Vermeulen, A., McCallum, S. A. & Pardi, A. Comparison of the global structure and dynamics of native and unmodified tRNAval. *Biochemistry* **44,** 6024–6033; 10.1021/bi0473399 (2005).

33. Serebrov, V., Vassilenko, K., Kholod, N., Gross, H. J. & Kisselev, L. Mg2+ binding and structural stability of mature and in vitro synthesized unmodified Escherichia coli tRNAPhe. *Nucleic Acids Res* **26,** 2723–2728; 10.1093/nar/26.11.2723 (1998).

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

35. Hoffer, E. D. *et al.* Structural insights into mRNA reading frame regulation by tRNA modification and slippery codon–anticodon pairing. *eLife Sciences Publications, Ltd* (2020).

36. Stuart, J. W., Koshlap, K. M., Guenther, R. & Agris, P. F. Naturally-occurring modification restricts the anticodon domain conformational space of tRNA(Phe). *Journal of molecular biology* **334,** 901–918; 10.1016/j.jmb.2003.09.058 (2003).

37. Sloan, K. E. *et al.* Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA biology* **14,** 1138–1152; 10.1080/15476286.2016.1259781 (2017).

38. Ben-Shem, A. *et al.* The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science (New York, N.Y.)* **334,** 1524–1529; 10.1126/science.1212642 (2011).

39. Decatur, W. A. & Fournier, M. J. rRNA modifications and ribosome function. *Trends in biochemical sciences* **27,** 344–351; 10.1016/s0968-0004(02)02109-6 (2002).

40. King, T. H., Liu, B., McCully, R. R. & Fournier, M. J. Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Molecular Cell* **11,** 425–435; 10.1016/S1097-2765(03)00040-6 (2003).

41. Liang, X., Liu, Q. & Fournier, M. J. rRNA modifications in an intersubunit bridge of the ribosome strongly affect both ribosome biogenesis and activity. *Molecular Cell* **28,** 965–977; 10.1016/j.molcel.2007.10.012 (2007).

42. Baudin-Baillieu, A. *et al.* Nucleotide modifications in three functionally important regions of the Saccharomyces cerevisiae ribosome affect translation accuracy. *Nucleic Acids Res* **37,** 7665–7677; 10.1093/nar/gkp816 (2009).

43. Zaccara, S., Ries, R. J. & Jaffrey, S. R. Reading, writing and erasing mRNA methylation. *Nat Rev Mol Cell Biol* **20,** 608–624; 10.1038/s41580-019-0168-5 (2019).

44. Cohn, W. E. Some results of the applications of ion-exchange chromatography to nucleic acid chemistry. *Journal of cellular physiology. Supplement* **38,** 21–40; 10.1002/jcp.1030380405 (1951).

45. Wu, G. *et al.* Pseudouridines in U2 snRNA stimulate the ATPase activity of Prp5 during spliceosome assembly. *The EMBO journal* **35,** 654–667; 10.15252/embj.201593113 (2016).

46. Schwartz, S. *et al.* Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159,** 148–162; 10.1016/j.cell.2014.08.028 (2014).

47. Kim, N.-K., Theimer, C. A., Mitchell, J. R., Collins, K. & Feigon, J. Effect of pseudouridylation on the structure and activity of the catalytically essential P6.1 hairpin in human telomerase RNA. *Nucleic Acids Res* **38,** 6746–6756; 10.1093/nar/gkq525 (2010).

48. Charette, M. & Gray, M. W. Pseudouridine in RNA: what, where, how, and why. *IUBMB life* **49,** 341–351; 10.1080/152165400410182 (2000).

49. Davis, D. R. Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res* **23,** 5020–5026; 10.1093/nar/23.24.5020 (1995).

50. 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* **30,** 79–89; 10.1038/emboj.2010.316 (2011).

51. van der Feltz, C., DeHaven, A. C. & Hoskins, A. A. Stress-induced Pseudouridylation Alters the Structural Equilibrium of Yeast U2 snRNA Stem II. *Journal of molecular biology* **430,** 524–536; 10.1016/j.jmb.2017.10.021 (2018).

52. Ganot, P., Bortolin, M. L. & Kiss, T. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **89,** 799–809; 10.1016/S0092-8674(00)80263-9 (1997).

53. Ganot, P., Caizergues-Ferrer, M. & Kiss, T. The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes & development* **11,** 941–956; 10.1101/gad.11.7.941 (1997).

54. Richard, P. *et al.* A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *The EMBO journal* **22,** 4283–4293; 10.1093/emboj/cdg394 (2003).

55. Morrissey, J. P. & Tollervey, D. Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis. *Molecular and cellular biology* **13,** 2469–2477; 10.1128/mcb.13.4.2469-2477.1993 (1993).

56. 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* **24,** 1769–1778; 10.1128/MCB.24.4.1769-1778.2004 (2004).

57. Fayet-Lebaron, E., Atzorn, V., Henry, Y. & Kiss, T. 18S rRNA processing requires base pairings of snR30 H/ACA snoRNA to eukaryote-specific 18S sequences. *The EMBO journal* **28,** 1260–1270; 10.1038/emboj.2009.79 (2009).

58. Mitchell, J. R., Wood, E. & Collins, K. A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* **402,** 551–555; 10.1038/990141 (1999).

59. Tang, T.-H. *et al.* Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 7536–7541; 10.1073/pnas.112047299 (2002).

60. Muller, S. *et al.* Combined in silico and experimental identification of the Pyrococcus abyssi H/ACA sRNAs and their target sites in ribosomal RNAs. *Nucleic Acids Res* **36,** 2459–2475; 10.1093/nar/gkn077 (2008).

61. Rozhdestvensky, T. S. *et al.* Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res* **31,** 869–877; 10.1093/nar/gkg175 (2003).

62. Baker, D. L. *et al.* RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes & development* **19,** 1238–1248; 10.1101/gad.1309605 (2005).

63. Li, L. & Ye, K. Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* **443,** 302–307; 10.1038/nature05151 (2006).

64. Liang, B., Xue, S., Terns, R. M., Terns, M. P. & Li, H. Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex. *Nat Struct Mol Biol* **14,** 1189–1195; 10.1038/nsmb1336 (2007).

65. Liang, B. *et al.* Long-distance placement of substrate RNA by H/ACA proteins. *RNA* **14,** 2086–2094; 10.1261/rna.1109808 (2008).

66. Li, S. *et al.* Reconstitution and structural analysis of the yeast box H/ACA RNA-guided pseudouridine synthase. *Genes & development* **25,** 2409–2421; 10.1101/gad.175299.111 (2011).

67. 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* **34,** 427–439; 10.1016/j.molcel.2009.05.005 (2009).

68. Trucks, S., Hanspach, G. & Hengesbach, M. Eukaryote specific RNA and protein features facilitate assembly and catalysis of H/ACA snoRNPs. *Nucleic Acids Res* **49,** 4629–4642; 10.1093/nar/gkab177 (2021).

69. 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 Res* **46,** 905–916; 10.1093/nar/gkx1167 (2018).

70. Bortolin, M. L., Ganot, P. & Kiss, T. Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. *The EMBO journal* **18,** 457–469; 10.1093/emboj/18.2.457 (1999).

71. 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* **24,** 1106–1117; 10.1261/rna.066837.118 (2018).

72. 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* **25,** 1393–1404; 10.1261/rna.071043.119 (2019).

73. Watkins, N. J. *et al.* 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.)* **4,** 1549–1568; 10.1017/S1355838298980761 (1998).

74. Wang, C. & Meier, U. T. Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *The EMBO journal* **23,** 1857–1867; 10.1038/sj.emboj.7600181 (2004).

75. 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* **12,** 527–537; 10.1101/gad.12.4.527 (1998).

76. Ramamurthy, V., Swann, S. L., Paulson, J. L., Spedaliere, C. J. & Mueller, E. G. Critical aspartic acid residues in pseudouridine synthases. *Journal of Biological Chemistry* **274,** 22225–22230; 10.1074/jbc.274.32.22225 (1999).

77. Normand, C. *et al.* Analysis of the binding of the N-terminal conserved domain of yeast Cbf5p to a box H/ACA snoRNA. *RNA (New York, N.Y.)* **12,** 1868–1882; 10.1261/rna.141206 (2006).

78. Koo, B.-K. *et al.* Structure of H/ACA RNP protein Nhp2p reveals cis/trans isomerization of a conserved proline at the RNA and Nop10 binding interface. *Journal of molecular biology* **411,** 927–942; 10.1016/j.jmb.2011.06.022 (2011).

79. Girard, J. P. *et al.* GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast. *The EMBO journal* **11,** 673–682; 10.1002/j.1460-2075.1992.tb05099.x (1992).

80. Ozdilek, B. A. *et al.* Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. *Nucleic Acids Res* **45,** 7984–7996; 10.1093/nar/gkx460 (2017).

81. Bagni, C. & Lapeyre, B. Gar1p binds to the small nucleolar RNAs snR10 and snR30 in vitro through a nontypical RNA binding element. *Journal of Biological Chemistry* **273,** 10868–10873; 10.1074/jbc.273.18.10868 (1998).

82. Nguyen, T. H. D. *et al.* Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature* **557,** 190–195; 10.1038/s41586-018-0062-x (2018).

83. Ghanim, G. E. *et al.* Structure of human telomerase holoenzyme with bound telomeric DNA. *Nature* **593,** 449–453; 10.1038/s41586-021-03415-4 (2021).

84. Liu, B. *et al.* Structure of active human telomerase with telomere shelterin protein TPP1. *Nature* **604,** 578–583; 10.1038/s41586-022-04582-8 (2022).

85. 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* **390,** 231–244; 10.1016/j.jmb.2009.04.076 (2009).

86. Leulliot, N. *et al.* The box H/ACA RNP assembly factor Naf1p contains a domain homologous to Gar1p mediating its interaction with Cbf5p. *Journal of molecular biology* **371,** 1338–1353; 10.1016/j.jmb.2007.06.031 (2007).

87. Machado-Pinilla, R., Liger, D., Leulliot, N. & Meier, U. T. Mechanism of the AAA+ ATPases pontin and reptin in the biogenesis of H/ACA RNPs. *RNA* **18,** 1833–1845; 10.1261/rna.034942.112 (2012).

88. Schattner, P., Barberan-Soler, S. & Lowe, T. M. A computational screen for mammalian pseudouridylation guide H/ACA RNAs. *RNA (New York, N.Y.)* **12,** 15–25; 10.1261/rna.2210406 (2006).

89. 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* **25,** 5396–5403; 10.1128/MCB.25.13.5396-5403.2005 (2005).

90. Yang, P. K. *et al.* Cotranscriptional recruitment of the pseudouridylsynthetase Cbf5p and of the RNA binding protein Naf1p during H/ACA snoRNP assembly. *Molecular and cellular biology* **25,** 3295–3304; 10.1128/MCB.25.8.3295-3304.2005 (2005).

91. Yang, X. *et al.* Kinetic and thermodynamic characterization of the reaction pathway of box H/ACA RNA-guided pseudouridine formation. *Nucleic Acids Res* **40,** 10925–10936; 10.1093/nar/gks882 (2012).

92. Czudnochowski, N. *et al.* The mechanism of pseudouridine synthases from a covalent complex with RNA, and alternate specificity for U2605 versus U2604 between close homologs. *Nucleic Acids Res* **42,** 2037–2048; 10.1093/nar/gkt1050 (2014).

93. 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* **96,** 14270–14275; 10.1073/pnas.96.25.14270 (1999).

94. Veerareddygari, G. R., Singh, S. K. & Mueller, E. G. The Pseudouridine Synthases Proceed through a Glycal Intermediate. *Journal of the American Chemical Society* **138,** 7852–7855; 10.1021/jacs.6b04491 (2016).

95. 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* **50,** 426–436; 10.1021/bi101737z (2011).

96. Jády, B. E., Ketele, A., Moulis, D. & Kiss, T. Guide RNA acrobatics: positioning consecutive uridines for pseudouridylation by H/ACA pseudouridylation loops with dual guide capacity. *Genes & development* **36,** 70–83; 10.1101/gad.349072.121 (2022).

97. Schmidt, A., Hanspach, G. & Hengesbach, M. Structural dynamics govern substrate recruitment and catalytic turnover in H/ACA RNP pseudouridylation. *RNA biology* **18,** 1300–1309; 10.1080/15476286.2020.1842984 (2021).

98. McMahon, M., Contreras, A. & Ruggero, D. Small RNAs with big implications: new insights into H/ACA snoRNA function and their role in human disease. *Wiley Interdisciplinary Reviews: RNA* **6,** 173–189; 10.1002/wrna.1266 (2015).

99. Trahan, C., Martel, C. & Dragon, F. Effects of dyskeratosis congenita mutations in dyskerin, NHP2 and NOP10 on assembly of H/ACA pre-RNPs. *Hum Mol Genet* **19,** 825–836; 10.1093/hmg/ddp551 (2010).

100. Liang, J. *et al.* Small Nucleolar RNAs: Insight Into Their Function in Cancer. *Front. Oncol.* **9,** 587; 10.3389/fonc.2019.00587 (2019).

101. Morais, P., Adachi, H. & Yu, Y.-T. Suppression of Nonsense Mutations by New Emerging Technologies. *International journal of molecular sciences* **21**; 10.3390/ijms21124394 (2020).

102. Karijolich, J. & Yu, Y.-T. Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature* **474,** 395–398; 10.1038/nature10165 (2011).

103. Song, J. *et al.* CRISPR-free, programmable RNA pseudouridylation to suppress premature termination codons. *Molecular Cell* **83,** 139-155.e9; 10.1016/j.molcel.2022.11.011 (2023).

104. Chen, J. *et al.* 6S RNA Mimics B-Form DNA to Regulate Escherichia coli RNA Polymerase. *Molecular Cell* **68,** 388-397.e6; 10.1016/j.molcel.2017.09.006 (2017).

105. Popenda, M. *et al.* Automated 3D structure composition for large RNAs. *Nucleic Acids Res* **40,** e112; 10.1093/nar/gks339 (2012).

106. Antczak, M. *et al.* New functionality of RNAComposer: an application to shape the axis of miR160 precursor structure. *Acta biochimica Polonica* **63,** 737–744; 10.18388/abp.2016_1329 (2016).

107. Baker, J. L. *et al.* Widespread genetic switches and toxicity resistance proteins for fluoride. *Science (New York, N.Y.)* **335,** 233–235; 10.1126/science.1215063. (2012).

108. Nelson, J. W., Atilho, R. M., Sherlock, M. E., Stockbridge, R. B. & Breaker, R. R. Metabolism of Free Guanidine in Bacteria Is Regulated by a Widespread Riboswitch Class. *Molecular Cell* **65,** 220–230; 10.1016/j.molcel.2016.11.019 (2017).

109. Panchapakesan, S. S. S., Corey, L., Malkowski, S. N., Higgs, G. & Breaker, R. R. A second riboswitch class for the enzyme cofactor NAD. *RNA* **27,** 99–105; 10.1261/rna.077891.120 (2021).

110. Sudarsan, N., Wickiser, J. K., Nakamura, S., Ebert, M. S. & Breaker, R. R. An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes & development* **17,** 2688–2697; 10.1101/gad.1140003. (2003).

111. Mandal, M. & Breaker, R. R. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature structural & molecular biology* **11,** 29–35; 10.1038/nsmb710 (2004).

112. Green, N. J., Grundy, F. J. & Henkin, T. M. The T box mechanism: tRNA as a regulatory molecule. *FEBS letters* **584,** 318–324; 10.1016/j.febslet.2009.11.056. (2010).

113. Weinberg, Z. *et al.* Detection of 224 candidate structured RNAs by comparative analysis of specific subsets of intergenic regions. *Nucleic Acids Res* **45,** 10811–10823; 10.1093/nar/gkx699 (2017).

114. Kubodera, T. *et al.* Thiamine-regulated gene expression of Aspergillus oryzae thiA requires splicing of the intron containing a riboswitch-like domain in the 5'-UTR. *FEBS letters* **555,** 516–520; 10.1016/s0014-5793(03)01335-8 (2003).

115. Sudarsan, N., Barrick, J. E. & Breaker, R. R. Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA (New York, N.Y.)* **9,** 644–647; 10.1261/rna.5090103 (2003).

116. Croft, M. T., Moulin, M., Webb, M. E. & Smith, A. G. Thiamine biosynthesis in algae is regulated by riboswitches. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 20770–20775; 10.1073/pnas.0705786105 (2007).

117. Panchal, V. & Brenk, R. Riboswitches as Drug Targets for Antibiotics. *Antibiotics* **10,** 45; 10.3390/antibiotics10010045 (2021).

118. Helmling, C. *et al.* NMR Structural Profiling of Transcriptional Intermediates Reveals Riboswitch Regulation by Metastable RNA Conformations. *Journal of the American Chemical Society* **139,** 2647–2656; 10.1021/jacs.6b10429 (2017).

119. Watters, K. E., Strobel, E. J., Yu, A. M., Lis, J. T. & Lucks, J. B. Cotranscriptional folding of a riboswitch at nucleotide resolution. *Nat Struct Mol Biol* **23,** 1124–1131; 10.1038/nsmb.3316 (2016).

120. Steinert, H. *et al.* Pausing guides RNA folding to populate transiently stable RNA structures for riboswitch-based transcription regulation. *eLife* **6**; 10.7554/eLife.21297 (2017).

121. Wickiser, J. K., Winkler, W. C., Breaker, R. R. & Crothers, D. M. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Molecular Cell* **18,** 49–60; 10.1016/j.molcel.2005.02.032 (2005).

122. Shine, J. & Dalgarno, L. The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proceedings of the National Academy of Sciences of the United States of America* **71,** 1342–1346; 10.1073/pnas.71.4.1342 (1974).

123. Subramanian, A. R. Structure and functions of ribosomal protein S1. *Progress in nucleic acid research and molecular biology* **28,** 101–142; 10.1016/S0079-6603(08)60085-9 (1983).

124. Johnson, J. E., Reyes, F. E., Polaski, J. T. & Batey, R. T. B12 cofactors directly stabilize an mRNA regulatory switch. *Nature* **492,** 133–137; 10.1038/nature11607 (2012).

125. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419,** 952–956; 10.1038/nature01145 (2002).

126. Rodionov, D. A., Vitreschak, A. G., Mironov, A. A. & Gelfand, M. S. Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. *Journal of Biological Chemistry* **277,** 48949–48959; 10.1074/jbc.M208965200 (2002).

127. Rieder, R., Lang, K., Graber, D. & Micura, R. Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control. *Chembiochem : a European journal of chemical biology* **8,** 896–902; 10.1002/cbic.200700057 (2007).

128. Marton Menendez, A. & Nesbitt, D. J. Lysine-Dependent Entropy Effects in the B. subtilis Lysine Riboswitch: Insights from Single-Molecule Thermodynamic Studies. *The journal of physical chemistry. B* **126,** 69–79; 10.1021/acs.jpcb.1c07833 (2022).

129. Uhm, H., Kang, W., Ha, K. S., Kang, C. & Hohng, S. Single-molecule FRET studies on the cotranscriptional folding of a thiamine pyrophosphate riboswitch. *Proceedings of the National Academy of Sciences of the United States of America* **115,** 331–336; 10.1073/pnas.1712983115 (2018).

130. Shahbabian, K., Jamalli, A., Zig, L. & Putzer, H. RNase Y, a novel endoribonuclease, initiates riboswitch turnover in Bacillus subtilis. *The EMBO journal* **28,** 3523–3533; 10.1038/emboj.2009.283 (2009).

131. Caron, M.-P. *et al.* Dual-acting riboswitch control of translation initiation and mRNA decay. *Proceedings of the National Academy of Sciences of the United States of America* **109,** E3444-53; 10.1073/pnas.1214024109 (2012).

132. Richards, J. & Belasco, J. G. Widespread Protection of RNA Cleavage Sites by a Riboswitch Aptamer that Folds as a Compact Obstacle to Scanning by RNase E. *Molecular Cell* **81,** 127-138.e4; 10.1016/j.molcel.2020.10.025 (2021).

133. Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. & Breaker, R. R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428,** 281–286; 10.1038/nature02362 (2004).

134. Pursley, B. R., Fernandez, N. L., Severin, G. B. & Waters, C. M. The Vc2 Cyclic di-GMP-Dependent Riboswitch of Vibrio cholerae Regulates Expression of an Upstream Putative Small RNA by Controlling RNA Stability. *Journal of bacteriology* **201**; 10.1128/JB.00293-19 (2019).

135. Warhaut, S. *et al.* Ligand-modulated folding of the full-length adenine riboswitch probed by NMR and single-molecule FRET spectroscopy. *Nucleic Acids Res* **45,** 5512–5522; 10.1093/nar/gkx110 (2017).

136. Warnasooriya, C. *et al.* Observation of preQ1-II riboswitch dynamics using single-molecule FRET. *RNA biology* **16,** 1086–1092; 10.1080/15476286.2018.1536591 (2019).

137. Manz, C. *et al.* Single-molecule FRET reveals the energy landscape of the full-length SAM-I riboswitch. *Nat Chem Biol* **13,** 1172–1178; 10.1038/nchembio.2476 (2017).

138. Lemay, J.-F., Penedo, J. C., Tremblay, R., Lilley, D. M. J. & Lafontaine, D. A. Folding of the adenine riboswitch. *Chemistry & Biology* **13,** 857–868; 10.1016/j.chembiol.2006.06.010 (2006).

139. Sung, H.-L. & Nesbitt, D. J. Single-Molecule FRET Kinetics of the Mn2+ Riboswitch: Evidence for Allosteric Mg2+ Control of "Induced-Fit" vs "Conformational Selection" Folding Pathways. *The journal of physical chemistry. B* **123,** 2005–2015; 10.1021/acs.jpcb.8b11841 (2019).

140. Sung, H.-L. & Nesbitt, D. J. High pressure single-molecule FRET studies of the lysine riboswitch: cationic and osmolytic effects on pressure induced denaturation. *Phys. Chem. Chem. Phys.* **22,** 15853–15866; 10.1039/D0CP01921F (2020).

141. Sung, H.-L. & Nesbitt, D. J. Sequential Folding of the Nickel/Cobalt Riboswitch Is Facilitated by a Conformational Intermediate: Insights from Single-Molecule Kinetics and Thermodynamics. *The journal of physical chemistry. B* **124,** 7348–7360; 10.1021/acs.jpcb.0c05625 (2020).

142. Manz, C., Kobitski, A. Y., Samanta, A., Jäschke, A. & Nienhaus, G. U. The multi-state energy landscape of the SAM-I riboswitch: A single-molecule Förster resonance energy transfer spectroscopy study. *The Journal of Chemical Physics* **148,** 123324; 10.1063/1.5003783 (2018).

143. St-Pierre, P. *et al.* A structural intermediate pre-organizes the add adenine riboswitch for ligand recognition. *Nucleic Acids Res* **49,** 5891–5904; 10.1093/nar/gkab307 (2021).

144. Rinaldi, A. J., Lund, P. E., Blanco, M. R. & Walter, N. G. The Shine-Dalgarno sequence of riboswitch-regulated single mRNAs shows ligand-dependent accessibility bursts. *Nat Commun* **7,** 8976; 10.1038/ncomms9976 (2016).

145. Strecker, A. Untersuchungen über die chemischen Beziehungen zwischen Guanin, Xanthin, Theobromin, Caffeïn und Kreatinin. *Ann. Chem. Pharm.* **118,** 151–177; 10.1002/jlac.18611180203 (1861).

146. Selig, P. *Guanidines as reagents and catalysts II* (Springer, Cham, Switzerland, 2017).

147. Becker, S. *et al.* A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science (New York, N.Y.)* **352,** 833–836; 10.1126/science.aad2808 (2016).

148. Natelson, S. & Sherwin, J. E. Proposed mechanism for urea nitrogen re-utilization: relationship between urea and proposed guanidine cycles. *Clinical Chemistry* **25,** 1343–1344; 10.1093/clinchem/25.7.1343 (1979).

149. Kihara, H., Prescott, J. M. & Snell, E. E. THE BACTERIAL CLEAVAGE OF CANAVANINE TO HOMOSERINE AND GUANIDINE. *Journal of Biological Chemistry* **217,** 497–503; 10.1016/S0021-9258(19)57198-8 (1955).

150. Eckert, C. *et al.* Ethylene-forming enzyme and bioethylene production. *Biotechnol Biofuels* **7,** 33; 10.1186/1754-6834-7-33 (2014).

151. Greene, R. F. & Pace, C. N. Urea and Guanidine Hydrochloride Denaturation of Ribonuclease, Lysozyme, α-Chymotrypsin, and b-Lactoglobulin. *Journal of Biological Chemistry* **249,** 5388–5393; 10.1016/S0021-9258(20)79739-5 (1974).

152. Bewley, M. C., Jeffrey, P. D., Patchett, M. L., Kanyo, Z. F. & Baker, E. N. Crystal structures of Bacillus caldovelox arginase in complex with substrate and inhibitors reveal new insights into activation, inhibition and catalysis in the arginase superfamily. *Structure* **7,** 435–448; 10.1016/S0969-2126(99)80056-2 (1999).

153. Zeymer, C., Werbeck, N. D., Schlichting, I. & Reinstein, J. The molecular mechanism of Hsp100 chaperone inhibition by the prion curing agent guanidinium chloride. *The Journal of biological chemistry* **288,** 7065–7076; 10.1074/jbc.M112.432583 (2013).

154. Grimminger, V., Richter, K., Imhof, A., Buchner, J. & Walter, S. The prion curing agent guanidinium chloride specifically inhibits ATP hydrolysis by Hsp104. *Journal of Biological Chemistry* **279,** 7378–7383; 10.1074/jbc.M312403200 (2004).

155. Lundh, H. & Thesleff, S. The mode of action of 4-aminopyridine and guanidine on transmitter release from motor nerve terminals. *European Journal of Pharmacology* **42,** 411–412; 10.1016/0014-2999(77)90176-5 (1977).

156. Sherlock, M. E., Malkowski, S. N. & Breaker, R. R. Biochemical Validation of a Second Guanidine Riboswitch Class in Bacteria. *Biochemistry* **56,** 352–358; 10.1021/acs.biochem.6b01270 (2017).

157. Sherlock, M. E. & Breaker, R. R. Biochemical Validation of a Third Guanidine Riboswitch Class in Bacteria. *Biochemistry* **56,** 359–363; 10.1021/acs.biochem.6b01271 (2017).

158. Salvail, H., Balaji, A., Yu, D., Roth, A. & Breaker, R. R. Biochemical Validation of a Fourth Guanidine Riboswitch Class in Bacteria. *Biochemistry* **59,** 4654–4662; 10.1021/acs.biochem.0c00793 (2020).

159. Lenkeit, F., Eckert, I., Hartig, J. S. & Weinberg, Z. Discovery and characterization of a fourth class of guanidine riboswitches. *Nucleic acids research* **48,** 12889–12899; 10.1093/nar/gkaa1102 (2020).

160. Barrick, J. E. *et al.* New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 6421–6426; 10.1073/pnas.0308014101 (2004).

161. Reiss, C. W., Xiong, Y. & Strobel, S. A. Structural Basis for Ligand Binding to the Guanidine-I Riboswitch. *Structure (London, England : 1993)* **25,** 195–202; 10.1016/j.str.2016.11.020 (2017).

162. Battaglia, R. A., Price, I. R. & Ke, A. Structural basis for guanidine sensing by the ykkC family of riboswitches. *RNA* **23,** 578–585; 10.1261/rna.060186.116 (2017).

163. Sherlock, M. E., Sudarsan, N. & Breaker, R. R. Riboswitches for the alarmone ppGpp expand the collection of RNA-based signaling systems. *Proceedings of the National Academy of Sciences of the United States of America* **115,** 6052–6057; 10.1073/pnas.1720406115 (2018).

164. Sherlock, M. E., Sudarsan, N., Stav, S. & Breaker, R. R. Tandem riboswitches form a natural Boolean logic gate to control purine metabolism in bacteria. *eLife Sciences Publications, Ltd* (2018).

165. Sherlock, M. E., Sadeeshkumar, H. & Breaker, R. R. Variant Bacterial Riboswitches Associated with Nucleotide Hydrolase Genes Sense Nucleoside Diphosphates. *Biochemistry* **58,** 401–410; 10.1021/acs.biochem.8b00617 (2019).

166. Weinberg, Z. *et al.* Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res* **35,** 4809–4819; 10.1093/nar/gkm487 (2007).

167. Huang, L., Wang, J., Wilson, T. J. & Lilley, D. M. J. Structure-guided design of a high-affinity ligand for a riboswitch. *RNA* **25,** 423–430; 10.1261/rna.069567.118 (2019).

168. Schamber, T., Binas, O., Schlundt, A., Wacker, A. & Schwalbe, H. Characterization of Structure and Dynamics of the Guanidine-II Riboswitch from Escherichia coli by NMR Spectroscopy and Small-Angle X-ray Scattering (SAXS). *ChemBioChem* **23,** e202100564; 10.1002/cbic.202100564 (2022).

169. Huang, L., Wang, J. & Lilley, D. M. J. The Structure of the Guanidine-II Riboswitch. *Cell Chemical Biology* **24,** 695-702.e2; 10.1016/j.chembiol.2017.05.014 (2017).

170. Reiss, C. W. & Strobel, S. A. Structural basis for ligand binding to the guanidine-II riboswitch. *RNA* **23,** 1338–1343; 10.1261/rna.061804.117 (2017).

171. Wuebben, C., Vicino, M. F., Mueller, M. & Schiemann, O. Do the P1 and P2 hairpins of the Guanidine-II riboswitch interact? *Nucleic Acids Res* **48,** 10518–10526; 10.1093/nar/gkaa703 (2020).

172. Steuer, J., Kukharenko, O., Riedmiller, K., Hartig, J. S. & Peter, C. Guanidine-II aptamer conformations and ligand binding modes through the lens of molecular simulation. *Nucleic Acids Res* **49,** 7954–7965; 10.1093/nar/gkab592 (2021).

173. Higgins, D. A. *et al.* Guanidine Riboswitch-Regulated Efflux Transporters Protect Bacteria against Ionic Liquid Toxicity. *Journal of bacteriology* **201**; 10.1128/JB.00069-19 (2019).

174. Kermani, A. A., Macdonald, C. B., Gundepudi, R. & Stockbridge, R. B. Guanidinium export is the primal function of SMR family transporters. *Proceedings of the National Academy of Sciences of the United States of America* **115,** 3060–3065; 10.1073/pnas.1719187115 (2018).

175. Weinberg, Z. *et al.* Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome biology* **11,** R31; 10.1186/gb-2010-11-3-r31 (2010).

176. Huang, L., Wang, J., Wilson, T. J. & Lilley, D. M. Structure of the Guanidine III Riboswitch. *Cell Chemical Biology* **24,** 1407-1415.e2; 10.1016/j.chembiol.2017.08.021 (2017).

177. Sinn, M., Hauth, F., Lenkeit, F., Weinberg, Z. & Hartig, J. S. Widespread bacterial utilization of guanidine as nitrogen source. *Molecular microbiology* **116,** 200–210; 10.1111/mmi.14702 (2021).

178. Schneider, N. O. *et al.* Solving the Conundrum: Widespread Proteins Annotated for Urea Metabolism in Bacteria Are Carboxyguanidine Deiminases Mediating Nitrogen Assimilation from Guanidine. *Biochemistry* **59,** 3258–3270; 10.1021/acs.biochem.0c00537 (2020).

179. Wang, B. *et al.* A guanidine-degrading enzyme controls genomic stability of ethylene-producing cyanobacteria. *Nat Commun* **12,** 5150; 10.1038/s41467-021-25369-x (2021).

180. Bay, D. C., Rommens, K. L. & Turner, R. J. Small multidrug resistance proteins: a multidrug transporter family that continues to grow. *Biochimica et biophysica acta* **1778,** 1814–1838; 10.1016/j.bbamem.2007.08.015 (2008).

181. Greener, T., Govezensky, D. & Zamir, A. A novel multicopy suppressor of a groEL mutation includes two nested open reading frames transcribed from different promoters. *The EMBO journal* **12,** 889–896; 10.1002/j.1460-2075.1993.tb05729.x (1993).

182. Slipski, C. J., Jamieson, T. R., Zhanel, G. G. & Bay, D. C. Riboswitch-Associated Guanidinium-Selective Efflux Pumps Frequently Transmitted on Proteobacterial Plasmids Increase Escherichia coli Biofilm Tolerance to Disinfectants. *Journal of bacteriology* **202**; 10.1128/JB.00104-20 (2020).

183. Thomason, M. K. *et al.* Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in Escherichia coli. *Journal of bacteriology* **197,** 18–28; 10.1128/JB.02096-14 (2015).

184. Fuks, C., Falkner, S., Schwierz, N. & Hengesbach, M. Combining Coarse-Grained Simulations and Single Molecule Analysis Reveals a Three-State Folding Model of the Guanidine-II Riboswitch. *Frontiers in molecular biosciences* **9,** 826505; 10.3389/fmolb.2022.826505 (2022).

185. Kadesch, T. R., Rosenberg, S. & Chamberlin, M. J. Binding of Escherichia coli RNA polymerase holoenzyme to bacteriophage T7 DNA. Measurements of binding at bacteriophage T7 promoter A1 using a template competition assay. *Journal of molecular biology* **155,** 1–29; 10.1016/0022-2836(82)90489-2 (1982).

186. Yan, Q. & Fong, S. S. Study of in vitro transcriptional binding effects and noise using constitutive promoters combined with UP element sequences in Escherichia coli. *Journal of biological engineering* **11,** 33; 10.1186/s13036-017-0075-2 (2017).

187. Ammons, D., Rampersad, J. & Fox, G. E. 5S rRNA gene deletions cause an unexpectedly high fitness loss in Escherichia coli. *Nucleic Acids Res* **27,** 637–642; 10.1093/nar/27.2.637 (1999).

188. Hanspach, G., Trucks, S. & Hengesbach, M. Strategic labelling approaches for RNA single-molecule spectroscopy. *RNA biology* **16,** 1119–1132; 10.1080/15476286.2019.1593093 (2019).

189. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms for molecular biology : AMB* **6,** 26; 10.1186/1748-7188-6-26 (2011).

190. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31,** 3406–3415; 10.1093/nar/gkg595 (2003).

191. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature methods* **9,** 676–682; 10.1038/nmeth.2019 (2012).

# 17 List of Abbreviations

| | |
|---|---|
| Ψ | Pseudouridine |
| APS | Ammonium persulfate |
| BSA | Bovine serum albumine |
| CTE | C-terminal extension |
| CV | Column volume |
| DMSO | Dimethyl sulfoxide |
| DNA | Desoxyribonucleic acid |
| *E.coli* | *Escherichia coli* |
| EDTA | Ethylenediaminetetraacetic acid |
| $E_{FRET}$ | FRET efficiency |
| EMSA | Electromobility shift assay |
| EtOH | Ethanol |
| FRET | Förster resonance energy transfer |
| $Gdm^+$ | Guanidinium |
| Gdn | Guanidine |
| IMAC | Immobilized metal affinity chromatography |
| IPTG | Isopropyl β-D-1-thiogalactopyranoside |
| lncRNA | long-noncoding RNA |
| miRNA | Micro RNA |
| mRNA | Messenger RNA |
| NHS | N-Hydrosysuccinimide |
| nt | Nucleotide |
| NTA | Nitrilotiracetic acid |
| NTE | N-terminal extension |
| Oac | Acetate |
| PAGE | Polyacrylamide gel electrophoresis |
| PCA | Protocatechuic acid |
| PCD | protocatechuate-3,4-dioxygenase |
| PCR | Polymerase chain reaction |
| PEG | Polyethylene glycol |
| RBS | Ribosomal binding site |
| RNA | Ribonucleic acid |
| RNP | Ribonucleoprotein |
| ROS | Reactive oxygen species |

| | |
|---|---|
| RP-LC | Reverse phase liquid chromatography |
| rpm | Rounds per minute |
| rRNA | Ribosomal RNA |
| *S. cerevisiae* | *Saccharomyces cerevisiae* |
| SD | Shine-Dalgarno |
| SDS | Sodium dodecyl sulfate |
| SEC | Size exclusion chromatography |
| smFRET | Single-molecule FRET |
| snoRNA | Small nucleolar RNA |
| snRNA | Small nuclear RNA |
| TEMED | Tetramethylethylenediamine |
| TIRF | Total internal reflection fluorescence |
| TRIS | Tris(hydroxymethyl)aminomethane |
| tRNA | Transfer RNA |
| Trolox | 6-Hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid |
| TSS | Transcription start site |
| UTR | Untranslated region |
| V | Volume |
| v/v | Volume per volume |
| w/v | Weight per volume |
| WNCG | Nhp2, Nop10, Cbf5, Gar1 |
| WT | Wildtype |

# 18 Deutschsprachige Zusammenfassung

Während dieser Doktorarbeit wurden zwei verschiedene Projekte bearbeitet. Zum einen die H/ACA Ribonukleoproteine aus Hefe und zum anderen der Guanidin-II Riboschalter aus dem gramnegativen Bakterium *E. coli*. Bei beiden Projekten lag der Fokus auf den Strukturdynamiken, die mittels Einzelmolekül Förster Energie Resonanz Transfer (smFRET) untersucht wurden. Mittels smFRET lassen sich (relative) Abstände und deren zeitliche Veränderungen zwischen zwei Fluorophoren bestimmen. Diese Einzelmolekülmethode ermöglicht es auch, Konformationsinformationen über kleinere Teilpopulationen zu gewinnen.

Während RNA lange Zeit nur als Blaupause für die Proteinproduktion angesehen wurde, weiß man heute, dass sie auch viele katalytische und regulatorische Aufgaben in der Zelle übernimmt. Die RNA besteht hauptsächlich aus den vier Bestandteilen Adenosin, Guanosin, Cytidin und Uridin, doch auch verschiedene Modifikationen sind für die Funktionen der Zelle wichtig. Die häufigste RNA Modifikation ist Pseudouridin (Ψ), das Rotationsisomer von Uridin[23,24]. Die Umwandlung kann entweder durch selbstständige Pseudouridylasen oder über RNA-geleitete H/ACA Pseudouridylasen erfolgen. Die H/ACA Ribonukleoproteine kommen sowohl in Archaeen als auch in Eukaryoten vor. Sie können sequenzspezifische Ψ in die RNA einführen, indem sie das Substrat über komplementäre Basenpaarung binden[52]. Diese Paarung erfolgt in einer internen Ausbuchtung in einer der Haarnadelstrukturen der Leit-RNA. Bei Eukaryoten sind genau zwei Haarnadelstrukturen pro Leit-RNA konserviert auf die entweder ein H-Box oder ein ACA-Box Sequenzmotiv folgt[53]. Bei Archaeen ist es meist eine Haarnadelstruktur, es können aber auch bis zu drei sein[53,60]. Jede Haarnadelstruktur bindet einen Satz aus vier Proteinen. Die eigentlich katalytisch aktive Pseudouridylase Cbf5 (Dyskerin im Menschen), Nop10, Gar1 und Nhp2 (L7Ae in Archaeen). Fehlfunktionen von H/ACA Pseudouridylasen können beim Menschen zu letalen Krankheitsbildern wie verschiedenen Krebsarten oder dem Knochenmarksversagen Dyskeratosis congenita führen[98–100]. Aber sie bieten auch die Möglichkeit der Therapie durch das Einführen von gezielten Ψ bei Patienten mit bestimmten Mutationen, die zum Abbruch der Proteinsynthese führen[103]. Aus diesen Gründen ist es wichtig das Wissen über die Struktur und Funktionsweise von H/ACA Ribonukleoproteinen zu erweitern.

Frühere Studien konnten zeigen, dass der Eukaryoten-spezifische zweiteilige Aufbau der H/ACA Ribonukleoproteine die Aktivität der Isomerisation steigert im Vergleich zu einzelnen Haarnadelstrukturen[66,68,69]. Strukturinformationen für den zweiteiligen Aufbau sind eher spärlich. Es gibt Kristallstrukturen des vollständigen Archaeen Komplex[63,67], sowie Kristall-[66] oder NMR-Strukturen[78] der Kerndomänen der einzelnen Proteinkomponenten. Außerdem enthält die menschliche Telomerase eine H/ACA Domäne, wovon es mehrere Kryo-EM Strukturen gibt[82–84]. Diese Arbeit beschäftigt sich hauptsächlich mit dem zweiteiligen Aufbau in Eukaryoten und der daraus resultierende drei-dimensionale Architektur, sowie die Rolle der einzelnen Proteinkomponenten hierfür. Als Modell wurde die snR81 Leit-RNA aus der Hefe verwendet, die die U2 snRNA des Splicosoms (Ψ42 & Ψ93), sowie die 25S rRNA (Ψ1051) pseudouridylieren kann.

Um sich die Strukturdynamiken des H/ACA Komplexes anschauen zu können mussten zunächst alle Komponenten hergestellt und auf ihre Funktionalität überprüft werden. Das Protein Nhp2, der Cbf5-Nop10 Komplex, sowie das Volllängen-Gar1 wurden von Dr. Sven Trucks beziehungsweise Nicole Mench bereitgestellt. Der ternäre Komplex aus Nop10, Cbf5 und Gar1 (NCG) wurde durch heterologe Expression in *E. coli* hergestellt und über Nickel-Affinitätschromatographie sowie Größenausschlusschromatographie gereinigt. Hierbei waren mehrere Schritte zur Entfernung der bakteriellen RNAs von Nöten, da vor allem Cbf5 auch unspezifisch RNAs binden kann. Die Synthese von NCG musste optimiert werden, da Gar1 zu Degradation neigt. Durch die Änderung des His-Affinitätstags zum C-Terminus des Proteins konnte dies verhindert werden. Außerdem wurden verschiedene verkürzte Gar1 Proteine hergestellt, denen verschiedene Eukaryoten-spezifische RGG-Domänen fehlten. Unmarkierte snR81 RNA konnte mittels Polymerase-Kettenreaktion und T7 Transkription hergestellt werden. Für die FRET Messungen wurde eine RNA mit sequenzspezifisch eingebauten Cy3 und Cy5 RNA benötigt. Hierfür wurden gekaufte RNA-Fragmente über NHS-Chemie mit den Farbstoffen markiert und gereinigt. Im Anschluss wurden die Fragmente über DNA-geschiente Ligation miteinander zur vollständigen RNA verknüpft. Die Verwendung von zwei separaten DNA Schienen führte zu geringen Ausbeuten. Letztendlich konnten jedoch sechs FRET Konstrukte mit verschiedenen Markierungsschemas erhalten werden. Mittels analytischer Größenausschlusschromatographie, UV/Vis Spektroskopie und $^{32}$P-Radioaktivitätsassays ließ sich nachweisen, dass die H/ACA Ribonukleoproteine *in vitro* rekonstituiert werden könne und an beiden Haarnadelstrukturen Aktivität aufweisen.

Mittels smFRET Experimenten wurde zunächst der Einfluss der zweiten Haarnadelstruktur auf die Konformationen während der Assemblierung analysiert. Dabei wurde die isolierten Haarnadelstrukturen mit ihren Volllängen Konstrukten mit dem gleichen Markierungsschema verglichen. Hier konnte gezeigt werden, dass Cbf5 die hauptsächliche Komponente für Distanzänderungen in der Pseudouridylierungstasche war. Die 5‘ Haarnadelstruktur (H5) zeigte keine messbaren Veränderungen durch die Anwesenheit von H3, wie annährend identische FRET Effizienzen und ähnliche Proteinabhängigkeiten zeigten. Im Gegensatz dazu scheint H3 durch die zweite Haarnadelstruktur dynamischer und heterogener zu werden. Dies passt zu Aktivitätsdaten für die gewählte Leit-RNA snR81, die die Effekte auf H3 als bedeutender beschreiben. Als nächstes wurde sich dem Zusammenbau des H/ACA Komplexes gewidmet. Im ersten Schritt, der proteinfreien RNA, konnten die smFRET Ergebnisse mit Konstrukten, die an verschiedenen Haarnadelstrukturen markiert waren, zeigen, dass diese wahrscheinlich nicht in der oft in der Literatur dargestellten parallelen Konformation vorliegen, sondern eher in verschiedene Richtungen zeigen. Neben der bereits beschriebenen Rolle von Cbf5 im Ausrichten der Pseudouridylierungstasche, konnte mit den Konstrukten auch eine Cbf5-vermittelte Neuorientierung der Haarnadelstrukturen festgestellt werden. Diese strukturellen Veränderungen betreffen die unteren Bereiche der RNA am meisten. Cbf5 könnte an die H-Box im Linker binden, diesen dadurch ausstrecken und Haarnadelstrukturen in eine angewinkelte Konformation bringen. Dabei könnte dies sowohl durch die Art der Bindung auf der RNA, als auch über Protein-Protein Interaktion bewerkstelligt werden. Nhp2 alleine scheint nach den smFRET Daten keinen eindeutigen Einfluss auf die Struktur zu haben, wobei fraglich ist, ob Nhp2 bei den gegebenen Messbedingungen überhaupt an die RNA bindet. Bei Gar1 alleine gab es hingegen Hinweise auf Veränderungen, die allerdings schwer zu reproduzieren waren. Interessanterweise haben die beiden Proteine, Nhp2 und Gar1, zusammen größere Effekte auf die RNA, obwohl diese Kombination in der Zelle nicht vorkommen sollte. Sie scheinen die beiden Haarnadelstrukturen, insbesondere die oberen Teile, näher zusammenzubringen, was durch eine Erhöhung der FRET Effizienz deutlich wurde. Ob dabei direkte Interaktionen der beiden Partner involviert sind, oder ein Protein die Bindung des anderen verändert bleibt für zukünftige Studien offen. Es wurde getestet, ob die RGG Domänen von Gar1, die in Eukaryoten, aber nicht in Archaeen vorhanden sind, das Zusammenziehen der Haarnadelstrukturen beeinflussen. Die FRET Ergebnisse zeigten Hinweise für eine Hierarchie der RGG Domänen. Die N- und C-Terminalen Seiten könnten kooperativ agieren. Ist nur eine der beiden Seiten vollständig sind die Effekte abgeschwächt und die Komplexe dynamischer. Ist keine der Seiten vollständig vorhanden oder keine der RGG Domänen scheinen die Nhp2 und Gar1 vermittelten Veränderungen nicht aufzutreten. Letztendlich wurden auch die vollständig rekonstituierten Komplexe mittels smFRET untersucht. Die Ergebnisse legten nahe, dass keine einheitliche Konformation vorlag. Dies kann entweder auf die Abwesenheit bestimmter Komponenten oder einer allgemeinen strukturellen Heterogenität zurückzuführen sein. Anhand der Daten aus den Assemblierungsexperimenten und den vollständigen Komplexen konnte eine Orientierung der Haarnadelstrukturen des zweiteiligen Ribonukleoproteins vorgeschlagen werden. In diesem Modell winkelt Cbf5 die unteren Bereiche der Haarnadelstrukturen an und Nhp2 und Gar1 zusammen bringen die oberen Teile näher. Vermutlich wirkt sich das Zusammenziehen der oberen Bereiche besonders auf H3 aus, der auch stärker von der zweiten Haarnadelstruktur beeinflusst wird.

Das zweite Projekt dieser Arbeit beschäftigte sich mit dem Guanidinium ($Gdm^+$) bindenden Riboschalter aus *E. coli* (Gdn-II). Riboschalter sind regulatorische RNA Elemente in der 5ʹ untranslatierten Region der mRNA. Sie binden ihren Liganden spezifisch in der Aptamer Domäne und können durch eine Konformationsänderung das Signal an die Expressionsplattform weiterleiten, wodurch die Genexpression an- oder abgeschaltet wird. Es gibt vier bekannte Klassen and Riboschaltern, die das toxische Molekül $Gdm^+$ binden können[108,156–159]. Sie regulieren zum Großteil die Expression von Proteinen die zur Detoxifizierung oder zum Metabolismus von $Gdm^+$ gehören wie beispielsweise Mehrmedikamenten-Effluxpumpen oder Ureacarboxylasen.

Hier wurde die zweite Klasse untersucht. Gdn-II besteht aus zwei Haarnadelstrukturen mit GC-reichen Stämmen und einem ACGR Motiv in der Schleife, wobei R eine Purinbase ist[166]. Die beiden Strukturen sind mit einem Linker verbunden, der die Anti-Shine-Dalgarno Sequenz (anti-SD Sequenz), eine Sequenz komplementär zur Ribosomenbindungsstelle, enthält. *In-line probing* Experimente konnten zeigen, dass der Ligand $Gdm^+$ dazu führt, dass die beiden Schleifen der Haarnadelstrukturen jeweils mit ihren CG Basen, Watson-Crick Basenpaarungen mit der anderen Schleife eingehen[156]. Auch durch kristallographische Methoden konnte für isolierte Haarnadelstrukturen eine solche Schleifen-Schleifen Interaktion strukturell gezeigt werden[169,170]. Der Ligand wird hierbei spezifisch in der Schleifenregion gebunden. Die Identität des vierten Purins in den beiden Schleifen scheint einen Einfluss auf die Bindungsaffinität des Liganden zu haben, wie NMR Studien zeigen konnten[168]. Bisher wurde lediglich die $Gdm^+$-bindende Aptamerdomäne des Gdn-II Riboschalters untersucht. Dennoch wurde ein Regulationsmodell entsprechend eines translationalen Riboschalters aufgestellt[170]. Entsprechend diesen Modells führt die Bindung von $Gdm^+$ and die Schleifen beider vorgeformter Haarnadelstrukturen zur Stabilisierung einer Schleifen-Schleifen Interaktion. Dadurch würde die anti-SD Sequenz unzugänglich für die SD Sequenz werden. Die freie SD Sequenz erlaubt dann die Bindung des Ribosoms und damit die Translation.

Dieses Projekt teilt sich in zwei methodische Teilprojekte auf. Zum einen wurde mit smFRET die Strukturdynamiken von der Aptamerdomäne und des Volllängen Riboschalters analysiert. Zusätzlich wurde das Verhalten eines Gdn-II Konstruktes *in vivo* beobachtet. Für die smFRET Studien musste Fluorophor-markierte RNA Konstrukte hergestellt werden. Hierfür wurden RNA Fragmente an drei möglichen Positionen markiert. Außerdem wurden verschiedene Längen an RNA synthetisiert. Ein 59mer soll die Untersuchungen der Aptamerdomäne ermöglichen. Ein 74mer und ein 93mer enthielten zusätzlich noch die SD Sequenz. Außerdem wurden verschiedene

Mutationen eingeführt, einmal an der vierten Position der Schleifen und einmal in der anti-SD oder SD Sequenz. Die einzelnen Fragmente wurden mittels DNA-geschienter Ligation zu den entsprechenden Längen zusammengefügt. Die Ligation von den bis zu sechs Fragmenten im Falle des 93mers bedurfte Optimierung. Letztendlich konnte die vollständig Ligation mit einem Wechsel von der T4 DNA Ligase zur T4 RNA Ligase 2 erreicht werden. Insgesamt wurden 24 FRET Konstrukte generiert und zusammen mit einem Konstrukt aus meiner Masterarbeit alle mittels FRET analysiert.

In einer Kooperation mit Sebastian Falkner aus der Gruppe von Nadine Schwierz-Neumann, der *Coarse-Grained* Simulationen durchgeführt hat, konnten drei $Mg^{2+}$-abhängige Konformationen für das 47mer Aptamerkonstrukt festgestellt und identifiziert werden. Dabei handelt es sich wahrscheinlich um eine offene Konformation mit vorgeformten Haarnadelstrukturen in Abwesenheit von $Mg^{2+}$, einer Schleifen-Schleifen Interaktion bei mittleren $Mg^{2+}$ Konzentrationen und einer Struktur, bei der die zweite Haarnadel nicht vorhanden war und stattdessen mit dem Linker basenpaart, wenn viel $Mg^{2+}$ in der Lösung vorhanden war. Dieser letzte Zustand trat bei längeren Konstrukten wie dem 59mer nicht mehr auf. Durch drei verschiedene Markierungsschematas konnte ein Modell für die Aptamerdomäne in Abhängigkeit von $Mg^{2+}$ und dem Liganden $Gdm^+$ aufgestellt werden. In Abwesenheit von beiden Ionen liegt demnach eine offene Konformation vor, bei dem die Haarnadeln geformt sind, allerdings räumlich getrennt sind und nicht miteinander interagieren. Durch $Mg^{2+}$ wird dem aufgestellten Modell zufolge eine lose Schleifen-Schleifeninteraktion ausgelöst. Der Linker mit der Anti-SD Sequenz ist hierbei flexibel. Der Ligand hingegen könnte zu einer stabilisierten Schleifen-Schleifen Konformation führen, bei dem der Linker an die Helixachse der Haarnadelstrukturen gezogen wird, beispielsweise durch direkte Basenpaarungen mit den vierten Schhleifennukleotiden.

Bei den 74mer und 93mer RNAs ist die SD Sequenz vorhanden. Während die FRET Effizienzen im ligandgebundenem Zustand mit dem Aptamer vergleichbar sind, scheinen die längeren Konstrukte eine andere Konformation in Abwesenheit von $Gdm^+$ einzunehmen. Mutationen in der SD Sequenz geben einen Hinweis darauf, dass diese Sequenz in Basenpaarung involviert sein könnte, da sich neue FRET Effizienzen ergeben, dies kann allerdings anhand der FRET Daten nicht eindeutig gesagt werden.

Außerdem wurde der Einfluss der Schleifennukleotide an Position vier analysiert. Alle getesteten Mutationen waren fähig FRET Effizienzen zu produzieren, die einer stabilisierten Schleifen-Schleifen Interaktion entsprechen, wenn Ligand vorhanden war. Nichtsdestotrotz, konnten Unterschiede ermittelt werden bezüglich der Konzentrationen an Ionen, die für die Einnahme einer gewissen Konformation nötig waren. Mittels grober Affinitätsbestimmungen wurden $K_D$s im oberen mikromolaren bis zum millimolaren Bereich für das Aptamer gefunden.

In einem anderen Teilprojekt der Untersuchung des Gdn-II Riboschalters wurde sich mit dem Verhalten *in vivo* beschäftigt. *E. coli* wurden mit einem Plasmid transformiert, dass ein Volllängen Riboschalter Konstrukt enthält und unter Kontrolle eines artifiziellen *E. coli* Promotors steht. Die Zellen wurden in Medium mit oder ohne externes $Gdm^+$ wachsen gelassen und nach gewissen Zeiten wurde der RNA Inhalt per denaturierender Gelelektrophorese analysiert. Dabei zeigte sich, dass mindestens vier RNAs, die von dem *in vivo* Riboschalter Konstrukt stammen, sich akkumulieren, wenn $Gdm^+$ hinzugefügt wurde. Die Level dieser RNAs veränderten sich mit der Zeit. In Minimalmedium nahmen ihre Konzentrationen in Abwesenheit von $Gdm^+$ über die Messung zu, was im Einklang mit der Literatur steht, dass unter diesen Bedingungen die Bakterien $Gdm^+$ selbst produzieren. Der Riboschalter könnte somit als eine Art Biosensor für $Gdm^+$ fungieren. Schlussendlich konnten auch in den *in vivo* Experimente leichte Unterschiede festgestellt werden, je nachdem welches Nukleotid an der vierten Stelle in der Schleife vorlag.